Dissertation

# Search for lepton-flavour violating decays of the Higgs boson with the ATLAS detector

Katharina Schleicher


Fakultät für Mathematik und Physik
Albert-Ludwigs-Universität Freiburg

# Search for lepton-flavour violating decays of the Higgs boson with the ATLAS detector

Dissertation

zur Erlangung des Doktorgrades der

Fakultät für Mathematik und Physik der

Albert-Ludwigs-Universität

Freiburg im Breisgau

vorgelegt von

Katharina Schleicher

Februar 2023

| | |
|---|---|
| Dekan: | Prof. Dr. Michael Thoss |
| Betreuer der Arbeit: | Prof. Dr. Markus Schumacher |
| Referent: | Prof. Dr. Markus Schumacher |
| Koreferent: | apl. Prof. Dr. Ulrich Landgraf |
| Prüfer: | Prof. Dr. Gregor Herten |
| | Prof. Dr. Stefan Dittmaier |
| | Prof. Dr. Markus Schumacher |

Datum der mündlichen Prüfung: 25.04.2023

# Contents

# Introduction

The Standard Model (SM) of particle physics is a relativistic quantum field theory forming the foundation of modern particle physics. It describes the entirety of all currently known elementary particles and their interactions apart from gravity and was developed from an interplay of theoretical and experimental efforts over several decades and was finalised in the 1970's. In the year 2012 it was crowned by the discovery of the last particle which was predicted within the SM but not yet confirmed experimentally, the Higgs boson.

The SM describes the existence and interaction of so-called *fermions* with spin $1/2$ of which a subset forms the basis for the matter surrounding us and of so-called *gauge bosons* with spin 1 which are the mediators of the fundamental interactions (electromagnetic, weak and strong). Members of both classes of particles are experimentally known to be massive. However, local gauge invariance, which is a fundamental principle in the formulation of the SM, does not allow for the introduction of mass terms and hence requires the particles to be massless. To solve this contradiction, a mechanism deploying the principle of spontaneous symmetry breaking was suggested in 1964 independently by Robert Brout and François Englert [1], by Peter Higgs [2] and by Gerald Guralnik, Carl Richard Hagen and Tom Kibble [3] and is today mainly referred to as the *Higgs mechanism*. It implies the existence of a neutral scalar field, the Higgs field. Massive bosons and charged fermions acquire mass through interactions with this field. The Higgs boson manifests itself as excitations of this field. Thus, its long awaited discovery was an important milestone to confirm the validity of the introduction of the Higgs mechanism into the SM and of the SM itself. The ATLAS and CMS experiments, located at the Large Hadron Collider (LHC) at CERN[1], announced the observation of a neutral boson with a mass of approximately 125 GeV in the search for the Higgs boson on July 4th 2012 [4, 5]. Subsequent measurements confirmed that the newly found boson is compatible with the Higgs boson predicted by the SM. So far many predictions of the SM could be verified experimentally and the SM has proven very successful.

However, the SM predicts neutrinos to be massless which was falsified by the observation of neutrino oscillations [6–8] which require neutrinos to have differing masses. In addition, neutrino oscillations imply that lepton flavour is not conserved in nature in contrast to the prediction by the SM. This raises the question if lepton-flavour violation is also realised in the charged lepton sector. One approach to address this question is searching for lepton-flavour violating (LFV) decays of the Higgs boson. The small decay width of the Higgs boson allows for decays beyond the SM to have branching ratios sufficiently sized for detection and LFV decays of the Higgs boson emerge naturally in different models beyond the SM [9–16]. Models extending the SM aim for providing solutions to deficiencies of the SM which, for example, are the inability to explain the abundance of *Dark Matter*, the size of the *matter-antimatter asymmetry* in the universe and the non-vanishing masses of neutrinos. Possible LFV decays of the Higgs boson are $H \to e\mu$, $H \to e\tau$ and $H \to \mu\tau$. Indirect limits constrain the branching ratio of $H \to e\mu$ to be smaller than $\mathcal{O}(10^{-8})$ [17, 18] which is currently out of reach for the experiments at the LHC while the branching

[1] Conseil Européen pour la Recherche Nucléaire

ratios of $H \to e\tau$ or $H \to \mu\tau$ are allowed to be of the order of 10 % [17, 18]. Direct limits on the latter two decays by the ATLAS and CMS collaborations [19, 20] outperform the indirect limits. The most stringent limits on the branching ratios $\mathcal{B}$, neglecting the findings of this thesis, are by the CMS collaboration using the dataset of Run 2 of the LHC: $\mathcal{B}(H \to e\tau) < 0.22\,\%\,(0.16\,\%)$ and $\mathcal{B}(H \to \mu\tau) < 0.15\,\%\,(0.15\,\%)$ [20], where the values in brackets are the expected limits.

The analysis presented in this thesis aims at either finding evidence for one of the two LFV decays $H \to e\tau$ and $H \to \mu\tau$ or at further constraining their branching ratios by analysing the full Run 2 dataset of proton-proton collisions delivered by the LHC from 2015 to 2018 at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. This dataset was recorded by the ATLAS detector and corresponds to an integrated luminosity of $138.4\,\text{fb}^{-1}$. Only leptonic decays of the $\tau$-lepton leading to different-flavour final states are considered. The requirement of different-flavour final states allows for the suppression of background from Drell-Yan production. In order to further gain sensitivity in the search for a process with an expectedly small cross section, two signal regions are defined to exploit different event topologies and background compositions. Furthermore, neural networks are employed to obtain a probability for each event indicating its similarity to a signal event. The resulting probability distributions are utilised as final discriminants in the statistical analysis. The majority of the background contributions is estimated using the *Symmetry Method* [21] which exploits the fact that, at the energies prevalent at the LHC, SM processes are symmetric with respect to an exchange of electrons with muons and vice versa[2]. The two LFV decays $H \to e\tau_\mu$ and $H \to \mu\tau_e$, in contrast, break this symmetry as long as their branching ratios differ. This implies that the *Symmetry Method* is only sensitive to the difference of the branching ratios of the two decays. It is, however, expected that these LFV decays do not simultaneously show sizeable branching ratios when neglecting interferences of other new LFV sources in $\mu \to e\gamma$ [18]. The notable advantage of the *Symmetry Method* lies in the fact that simulations of collision events only play a minor role in this background estimation method since it is primarily data-driven.

This thesis is structured as follows. An introduction to the theoretical background relevant for this thesis is given in Chapter 1. The Standard Model is briefly discussed, followed by experimental results on the measurement of the Higgs-boson properties and by the topic of lepton-flavour violation with a focus on charged lepton-flavour violation within the Higgs sector. Chapter 2 provides a description of the LHC and the ATLAS detector and gives an overview of the different data taking periods. The strategy of the analysis including a discussion of the basic principles of the *Symmetry Method* is illustrated in Chapter 3. Chapter 4 describes the generation of simulated events and discusses the various signal and background processes and their characteristics. The reconstruction and identification of physics objects is discussed in Chapter 5. Chapter 6 gives a general overview of neural networks in the context of signal and background separation. The event selection as well as the estimation of the different background contributions is presented in Chapter 7. The separation of these background contributions from the signal by means of the definition of two signal regions and the application of neural networks is discussed in Chapter 8. Chapter 9 discusses the statistical as well as the systematic uncertainties relevant to this analysis while Chapter 10 introduces the statistical concepts required to derive the final results. The results obtained by the analysis presented in this thesis are discussed and contextualised in Chapter 11. In addition, the results obtained from combinations with two partner analyses [22] are presented. Finally, Chapter 12 recaps the analysis and its results and gives an outlook into the future of searches for lepton-flavour violating decays of the Higgs boson with the *Symmetry Method*.

[2]The differing coupling strength of the Yukawa-couplings for electrons and muons can be neglected as the expected rates are low and same-flavour final states are produced.

## Personal Contributions

The work presented in this thesis was performed within the ATLAS collaboration and relies on the preceding work of many other people as well as on joint efforts during the development and finalisation of the analysis.

The author of this thesis contributed small studies to the analysis of LFV decays of the Higgs boson using only a subset of the Run 2 dataset corresponding to $36.1\,\text{fb}^{-1}$. The analysis does not use the *Symmetry Method* for background estimation. It is published in Ref. [19].

The results of the analysis performed on the full Run 2 dataset are published in Ref. [22] and will soon be available as a paper in the Journal of High Energy Physics (JHEP). This publication includes results of the analysis presented in this thesis using the *Symmetry Method*, results derived by the partner analyses mentioned above as well as a combination of all three analyses. Contributions to this analysis and publication are listed in the following:

- Production of ntuples used to determine the fake factors for the the data-driven fake estimate with the xTauFramework
- Implementation, development and validation of the full analysis chain, starting at ntuple level, in the Common Analysis Framework (CAF) used to prepare inputs for the statistical analysis; apart from the calculation of the electron offline efficiencies and of the calculation of the fake faktors used in the data-driven fake estimate
- Development and validation of the event selection
- Validation and partial development of the background estimation methods: data-driven fake estimate, fake estimate based on simulated events and efficiency corrections which are part of the *Symmetry Method*
- Implementation, development and validation of the neural network training and the application of the networks
- Implementation, development and validation of the statistical analysis in TRExFitter; in direct interplay with the development of the preceding steps
- Derivation of the final results
- Studies on the interpretation of the results
- Performing and validating the combinations of the results of this analysis (using the *Symmetry Method* for background estimation) with the results of the partner analyses

The author provided the inputs for the extrapolation of the analysis discussed in the thesis to the High-Luminosity LHC (HL-LHC) scenario as well as technical support and discussions. The results of this extrapolation are published in Ref. [23].

The author also contributed to the measurement of the efficiencies of the hadronic $\tau$-lepton trigger by analysing $t\bar{t}$-events and performing a combination with the measurements performed with $Z \to \tau\tau$-events. This work was performed within the context of a *qualification task* in order to obtain ATLAS authorship. It is based on data recorded in the year 2016. The results are published in Ref. [24]. The combination of the results of the $t\bar{t}$- and $Z \to \tau\tau$-analyses was pursued for the results of the full Run 2 dataset. The work on the $\tau$-trigger efficiency measurements is not documented in this thesis.

# Chapter 1

# Theoretical Background

This chapter introduces the theoretical background relevant for this thesis. First, an overview of the Standard Model of particle physics is given in Section 1.1 including the mechanism of electroweak symmetry breaking which entails the Higgs boson. The production processes of the Higgs boson in proton-proton collisions and its decay modes as well as the latest experimental results are discussed in Section 1.2, followed by a brief description of the $\tau$-lepton in Section 1.3. Finally, the topic of lepton-flavour violation with a focus on charged lepton-flavour violation in the Higgs sector is reviewed in Section 1.4.

## 1.1 The Standard Model of Particle Physics

The Standard Model (SM) of particle physics describes the entirety of all currently known elementary particles and their interactions (excluding gravitation). It was developed over several decades and finalised in the 1970's. An overview over the particle content of the SM and its fundamental forces and interactions is given in Section 1.1.1.

The SM is based on relativistic *Quantum Field Theory* (QFT) while exploiting the principle of local gauge (phase) invariance. Particles are described by fields, and excitations of these fields are identified as the "physical" particles. Similar to classical mechanics, a Lagrangian density is constructed serving as basis to derive equations of motions by applying the Euler-Lagrange equations. The Lagrangian density $\mathscr{L}$ is typically abbreviated as "Lagrangian". The Lagrangian allows for deriving the corresponding Feynman rules [25] and hence the interactions of the particles and their kinematic properties. The transition probability of an initial state to a certain final state can be calculated by making use of perturbation theory. The process can be visualised in the form of so-called Feynman diagrams [25]. The different components of the Lagrangian of the SM are discussed in Sections 1.1.2–1.1.4, based on [26–29].

So far many aspects of the SM were confirmed experimentally and no significant deviations from the SM predictions were observed at colliders. The observation of neutrino oscillations, however, is for example not in agreement with the SM. This aspect is further discussed in Section 1.4.

### 1.1.1 The Particle Content of the SM

The elementary particles can be divided into two categories depending on their spin. All particles with half-integer spin are called fermions and obey the Fermi-Dirac-statistics. In the SM all fermions have spin 1/2 and are further sub-divided into leptons and quarks. The former participate only in electroweak, i.e. electromagnetic and weak, interactions while the latter additionally carry colour-charge and hence also participate in the strong interaction. There are six leptons and six quarks which can be grouped in three generations.

The generations differ only in the masses of their members and in their flavour. The masses increase with the generations and each of the twelve fermions has an associated flavour. For leptons, however, both members of one generation have the same lepton flavour quantum number. The first generation of leptons consists of the electron ($e$) and the electron-neutrino ($\nu_e$) with electric elementary charges of $q = -1$ and $q = 0$, respectively. The electron lepton flavour number is 1 while the muon and tau lepton flavour numbers are 0. The second generation is built by the muon ($\mu, q = -1$) and the muon-neutrino ($\nu_\mu, q = 0$) where the muon lepton flavour number is 1 while the other two are 0. The third generation is built by the tau ($\tau, q = -1$) and the tau-neutrino ($\nu_\tau, q = 0$) where the tau lepton flavour number is 1 while the other two are 0. Each generation of quarks consists of an *up*-type and a *down*-type quark which have electric elementary charges of $+2/3$ and $-1/3$, respectively. The first generation is built by the up ($u$) and the down ($d$) quark, the second by the charm ($c$) and the strange ($s$) quark and the third by the top ($t$) and the bottom ($b$) quark. Each quark carries colour charge which can either be red ($r$), green ($g$) or blue ($b$). In addition, there exists an antiparticle of opposite quantum numbers, e.g. electric charge and lepton flavour number , for each particle.

All particles with an integer spin are called bosons and obey the Bose-Einstein-statistics. The bosons with spin 1, also called gauge bosons, are the mediators of the interactions. The massless photon ($\gamma$) mediates the electromagnetic interactions by coupling to the electric charge $Q$. It does not carry an electric charge and hence does not interact with itself. In addition, it is also neutral with respect to the weak and strong interactions. Weak interactions are mediated by three massive gauge bosons, $W^\pm$ and $Z$, which couple to the weak isospin $\boldsymbol{I}$ and to the weak isospin and hypercharge $Y$, respectively. They have masses of 80.4 GeV and 91.2 GeV [30], respectively, and interact with themselves. The strong interaction is mediated by eight massless gluons $g$ which also carry colour charge and hence interact with themselves. The gluons are neutral with respect to electroweak interactions. An additional boson exists, the Higgs boson, with spin 0. It has an electric and colour charge of 0 and couples to all massive elementary particles thereby providing them with mass.

The particles and a selection of their properties are summarised in Tables 1.1 and 1.2. All matter surrounding us is made of electrons, protons and neutrons. Protons and neutrons are composite particles made of three valence quarks of up- and down-flavour. Hence, only the first generation of fermions contributes to our everyday matter. The members of the higher generations, created in high-energetic events or collider experiments, decay quickly into particles of the first generation.

### 1.1.2 Quantum Electrodynamics

The concept of requiring local gauge invariance in order to derive the components of the SM Lagrangian describing the dynamics of the fermions and gauge bosons is explained in detail in this section on the basis of the Lagrangian of quantum electrodynamics (QED). This concept is similarly applied in Sections 1.1.3 and 1.1.4 to obtain the Lagrangians of quantum chromodynamics and of the electroweak unification, respectively.

Quantum electrodynamics describe the propagation and interaction of electrically charged particles mediated by the photon. The Lagrangian of a free spin-1/2 particle (fermion) of mass $m$ is

$$\mathscr{L} = i\overline{\psi}\gamma^\mu\partial_\mu\psi - m\overline{\psi}\psi\,, \tag{1.1}$$

with $\psi$ being a Dirac spinor describing the field of the particle and $\overline{\psi}$ the adjoint spinor. Both depend on $x = x^\mu$ which are the space-time coordinates with $\mu \in \{0, 1, 2, 3\}$. The matrices $\gamma^\mu$ are the Dirac matrices and $\partial_\mu$ is the short form of the partial derivative $\partial/\partial_{x^\mu}$. The Einstein summation notation applies, which is the implicit summation over indices

| Generation | Flavour | | $L_e$ | $L_\mu$ | $L_\tau$ | Electric Charge | Mass/GeV |
|---|---|---|---|---|---|---|---|
| | | | **Leptons**, Spin 1/2 | | | | |
| I | $e$ | (electron) | 1 | 0 | 0 | -1 | $0.5110 \times 10^{-3}$ |
| | $\nu_e$ | (electron-neutrino) | 1 | 0 | 0 | 0 | 0 |
| II | $\mu$ | (muon) | 0 | 1 | 0 | -1 | 0.1057 |
| | $\nu_\mu$ | (muon-neutrino) | 0 | 1 | 0 | 0 | 0 |
| III | $\tau$ | (tau) | 0 | 0 | 1 | -1 | 1.777 |
| | $\nu_\tau$ | (tau-neutrino) | 0 | 0 | 1 | 0 | 0 |
| | | | **Quarks**, Spin 1/2 | | | | |
| I | $u$ | (up) | 0 | 0 | 0 | +2/3 | $2.16 \times 10^{-3}$ |
| | $d$ | (down) | 0 | 0 | 0 | -1/3 | $4.67 \times 10^{-3}$ |
| II | $c$ | (charm) | 0 | 0 | 0 | +2/3 | 1.27 |
| | $s$ | (strange) | 0 | 0 | 0 | -1/3 | $93.4 \times 10^{-3}$ |
| III | $t$ | (top) | 0 | 0 | 0 | +2/3 | 173 |
| | $b$ | (bottom) | 0 | 0 | 0 | -1/3 | 4.18 |

**Table 1.1:** Overview of the fermions of the SM with a selection of their properties. $L_e$, $L_\mu$ and $L_\tau$ denote the electron, muon and tau lepton flavour quantum numbers. The electric charge $q$ is given in units of the elementary charge $e$ and the values for the masses $m$ are taken from [30]. The quark masses have sizeable uncertainties and are hence to be understood as estimates. For the neutrino masses the theoretical values in the SM of 0 GeV are listed.

| | | | **Gauge Bosons**, Spin 1 | | |
|---|---|---|---|---|---|
| Interaction | Mediators | | Couples to | Electric Charge | Mass/GeV |
| weak | $W^\pm$ | ($W$-bosons) | weak isospin | $\pm1$ | 80.38 |
| | $Z$ | ($Z$-boson) | weak isospin and hypercharge | 0 | 91.19 |
| electromagnetic | $\gamma$ | (photon) | electric charge | 0 | 0 |
| strong | $g$ | (8 gluons) | colour charge | 0 | 0 |

**Table 1.2:** Overview of the gauge bosons of the SM with a selection of their properties. The electric charge $q$ is given in units of the elementary charge $e$ and the values for the masses $m$ are from [30]. For the photon and gluon masses the theoretical values of 0 GeV are listed.

appearing twice.

The Lagrangian is invariant under a global phase or gauge transformation $\psi \to e^{i\alpha}\psi$ with $\alpha$ being a real constant which implies $\overline{\psi} \to e^{-i\alpha}\overline{\psi}$ and $\partial_\mu\psi \to e^{i\alpha}\partial_\mu\psi$. This means the Lagrangian is not altered by this transformation and hence also the equations of motion stay the same. When requiring a more general, local gauge invariance, i.e. $\psi \to e^{i\alpha(x)}\psi$ with $\alpha$ depending on the space-time coordinates, the Lagrangian as given in Eq. 1.1 ceases to be invariant. In QED this transformation is modified to be $\psi \to e^{i\alpha(x)Q}\psi = e^{i\alpha(x)q}\psi$ where $Q$ is the electric charge operator with eigenvalue $q$ representing the electric charge of the fermion. When applying this transformation, an additional term from the derivative $\partial_\mu\psi$ remains in the Lagrangian:

$$\mathscr{L} \to \mathscr{L} - q\,(\partial_\mu\alpha(x))\,\overline{\psi}\gamma^\mu\psi\,. \tag{1.2}$$

In order to obtain a modified Lagrangian that is invariant under the local gauge transfor-

mation, the derivative $\partial_\mu$ is replaced by the covariant derivative $D_\mu$, which itself transforms as $D_\mu \to e^{i\alpha(x)q} D_\mu$,

$$D_\mu = \partial_\mu + iqA_\mu \,. \tag{1.3}$$

The second term of $D_\mu$ compensates the additional term in Eq. 1.2 if $A_\mu$ transforms as:

$$A_\mu \to A_\mu - \partial_\mu \alpha(x) \,, \tag{1.4}$$

with $A_\mu$ being a vector field, also called gauge field. It can be identified as the photon field and a term representing the kinetic energy of the photon must be added to the Lagrangian in addition: $-\frac{1}{4} F_{\mu\nu} F^{\mu\nu}$ with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ being the gauge invariant field strength tensor. Then, the full modified Lagrangian becomes

$$\mathscr{L}_{\text{QED}} = i\overline{\psi}\gamma^\mu D_\mu \psi - m\overline{\psi}\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{1.5}$$

$$= i\overline{\psi}\gamma^\mu \partial_\mu \psi - m\overline{\psi}\psi - q\overline{\psi}\gamma^\mu A_\mu \psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \,, \tag{1.6}$$

which is the Lagrangian of QED. The first term in Eq. 1.6 corresponds to the kinetic term of the fermion field, the second term is its mass term. The third term represents the interaction between the fermion and the photon field, proportional to $q$, and the fourth term is the kinetic term of the photon. When applying the Euler-Lagrange equations to the first two terms, the Dirac equation is obtained; applying them to the latter two terms, the Maxwell equations are obtained. Demanding invariance under the local gauge transformations prohibits adding a mass term of the form $m^2 A_\mu A^\mu$ for the photon. Hence, the photon is massless which is in accordance with experimental observations and the electromagnetic force has infinite range. The coupling strength of QED can be reparametrised as $\alpha_{\text{QED}} = e^2/(4\pi)$. It depends on the energy of the underlying process and has at least a value of $\sim 1/137$ [30]. The higher the energy or the lower the distance between the interacting particles, the larger the value of $\alpha_{\text{QED}}$.

The phase transformations $e^{i\alpha Q}$ form the unitary Abelian group $U(1)_Q$. Following Noether's theorem [31], the $U(1)_Q$ invariance implies the existence of a conserved current: $\partial_\mu j^\mu = 0$. Here, this current is the electromagnetic charge current density $j^\mu = q\overline{\psi}\gamma^\mu\psi$ and hence also the electric charge $q = \int d^3x \; j^0$ is conserved. This allows for replacing the third term in Eq. 1.6 with $-j^\mu A^\mu$.

### 1.1.3 Quantum Chromodynamics

Quantum chromodynamics (QCD) is a non-Abelian gauge theory with symmetry group $SU(3)_C$ describing the strong interactions between particles carrying colour charge $C$ (quarks and gluons), mediated by gluons. By imposing the same principle of requiring local gauge invariance of the Lagrangian as in QED, the Lagrangian of QCD is derived starting from the Lagrangian of a free fermion (Eq. 1.1). However, due to the three colour states of a quark, the Dirac spinor $\psi$ in Eq. 1.1 representing the fermion field is replaced by a column vector of three Dirac spinors, $\boldsymbol{\psi} = (\psi_r, \psi_g, \psi_b)$, with $r, g, b$ representing the three different colour states red, green and blue. This vector transforms under local $SU(3)$ transformations as follows:

$$\boldsymbol{\psi} \to e^{i\frac{g_s}{2}\alpha_a(x)\lambda_a} \boldsymbol{\psi} \,, \tag{1.7}$$

where summation over repeated indices is implied. The matrices $\lambda_a$ with $a = 1, ..., 8$ are the linearly independent, traceless $3 \times 3$ Gell-Mann matrices [32] serving as the generators of the $SU(3)_C$ group and $\alpha_a$ are the gauge phases. The $SU(3)$ group is non-Abelian as the commutator of any two matrices $\lambda$ does not vanish, instead it is a linear combination of all $\lambda$: $[\lambda_a, \lambda_b] = 2i f_{abc} \lambda_c$ with $f_{abc}$ being the structure constants.

In analogy to QED, the derivative $\partial_\mu$ in the free Lagrangian is replaced by the covariant derivative

$$D_\mu = \partial_\mu + i\frac{g_s}{2}\lambda_a G^a_\mu \,, \tag{1.8}$$

where the eight gauge fields $G^a_\mu$ transform as

$$G^a_\mu \to G^a_\mu - \partial_\mu \alpha_a - g_s f_{abc}\alpha_b G^c_\mu \,. \tag{1.9}$$

In addition, a kinetic term for each of the eight gauge fields of the form

$$-\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a \quad \text{with} \quad G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f_{abc} G^b_\mu G^c_\nu \tag{1.10}$$

is added to the Lagrangian. Here, $G^a_{\mu\nu}$ is the field strength tensor which has an additional term compared to the field strength tensor $F_{\mu\nu}$ of QED due to the non-Abelian nature of $SU(3)_C$. Finally, the QCD Lagrangian which is gauge invariant under local colour phase transformations ($SU(3)_C$) is:

$$\mathscr{L}_{\text{QCD}} = i\overline{\psi}\gamma^\mu\partial_\mu\psi - m\overline{\psi}\psi - \frac{g_s}{2}\left(\overline{\psi}\gamma^\mu\lambda_a\psi\right)G^a_\mu - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a \,. \tag{1.11}$$

The gauge fields $G^a_\mu$ correspond to eight massless gluons which couple to the quark fields with the coupling strength $g_s$. The gluons are required to be massless due to the potential mass terms not being invariant under the local gauge transformations.

The last term in Eq. 1.11 is not a purely kinetic term for the gluons but also adds self-interaction terms for the gluons due to the last term in Eq. 1.10 arising from the non-Abelian nature of $SU(3)_C$. The expansion of the kinetic term shows that there are three- and four-gluon vertices implying that the gluons themselves carry colour charge.

Similar to QED, the coupling strength of QCD can be reparametrised as $\alpha_{\text{QCD}} = g_s^2/(4\pi)$ and also depends on the energy of the underlying process. In QCD, however, the coupling strength decreases for increasing energies or decreasing distances resulting in so-called *asymptotic freedom* [33, 34] where the quarks and gluons are essentially free. Whereas at increasing distances of the particles, the coupling strength increases, resulting in increased potential energy. Once the potential energy e.g. between two quarks is sufficiently high, an additional quark-antiquark pair is created. This proceeds until the kinetic energy does not suffice anymore to take the particles further apart. Colourless clusters of the created quarks and gluons, called hadrons, are formed. This is referred to as *confinement* [27].

### 1.1.4 Electroweak Unification

The weak interactions, known for example from the $\beta$-decay, are formulated in combination with the electromagnetic interactions within the electroweak theory also known as the Glashow-Salam-Weinberg model [35–37]. The underlying symmetry group is $SU(2)_{I,L} \otimes U(1)_Y$ with the weak isospin operator $\boldsymbol{I}$ and the weak hypercharge operator $Y$ serving as generators of the groups $SU(2)$ and $U(1)$, respectively. The fermions are organised in multiplets of the weak isospin. Left-chiral fermions, indicated by the subscript $L$, are grouped in doublets with weak isospin $I = 1/2$. They participate in charged weak interactions via the exchange of a $W^+$- or $W^-$-boson within each doublet as well as in neutral weak and electromagnetic interactions via the exchange of a $Z$-boson and a photon. Neutrinos only interact weakly as they have no electric charge. Right-chiral fermions, indicated by the subscript $R$, have a weak isospin of $I = 0$ and are hence organised in singlets. It is experimentally known that neutrinos only exist left-chiral (and anti-neutrinos only right-chiral); therefore no neutrino singlets are added. Right-chiral fermions (and left-chiral anti-fermions) participate only in neutral electroweak interactions through the

exchange of a $Z$-boson or a photon. An overview of the different doublets and singlets and their quantum numbers is given in Table 1.3.

| Multiplets | | | $Q$ | $I_3$ | $Y$ |
|---|---|---|---|---|---|
| Leptons | | | | | |
| $\begin{pmatrix}\nu_e\\ e\end{pmatrix}_L$ | $\begin{pmatrix}\nu_\mu\\ \mu\end{pmatrix}_L$ | $\begin{pmatrix}\nu_\tau\\ \tau\end{pmatrix}_L$ | $0$ | $+1/2$ | $-1$ |
| | | | $-1$ | $-1/2$ | $-1$ |
| $e_R$ | $\mu_R$ | $\tau_R$ | $-1$ | $0$ | $-2$ |
| Quarks | | | | | |
| $\begin{pmatrix}u\\ d'\end{pmatrix}_L$ | $\begin{pmatrix}c\\ s'\end{pmatrix}_L$ | $\begin{pmatrix}t\\ b'\end{pmatrix}_L$ | $+2/3$ | $+1/2$ | $+1/3$ |
| | | | $-1/3$ | $-1/2$ | $+1/3$ |
| $u_R$ | $c_R$ | $t_R$ | $+2/3$ | $0$ | $+4/3$ |
| $d_R$ | $s_R$ | $b_R$ | $-1/3$ | $0$ | $-2/3$ |

**Table 1.3:** Overview of the fermion doublets and singlets of the SM with their values of the electric charge $Q$, the third component of the weak isospin $I_3$ and the hypercharge $Y$. All fermions listed are the weak eigenstates.

The weak hypercharge $Y$ is connected to the third component of the weak isospin $I_3$ and to the electric charge $Q$ via the Gell-Mann-Nishijima relation [38, 39]:

$$Y = 2(Q - I_3)\,. \tag{1.12}$$

The multiplets of the anti-particles have opposite quantum numbers, i.e. opposite chirality, $Q$, $Y$ and $I_3$. The components of each doublet are basically described as two states of the same particle where the third component of the weak isospin, $I_3$, is $+1/2$ for the first components of the doublets and $-1/2$ for the second components. This implies that the masses in this description are the same (vanishing). The down-type quarks $d'$, $s'$ and $b'$ are the weak eigenstates and are related to the mass eigenstates $d$, $s$ and $b$ via the Cabibbo–Kobayashi–Maskawa (CKM) matrix [40, 41]:

$$\begin{pmatrix}d'\\ s'\\ b'\end{pmatrix} = V_{\mathrm{CKM}} \begin{pmatrix}d\\ s\\ b\end{pmatrix}\,. \tag{1.13}$$

The left- and right-chiral fermion fields, $\chi_L$ and $\psi_R$, transform under the $SU(2)_{I,L} \otimes U(1)_Y$ symmetry as follows:

$$\chi_L \to e^{ig\boldsymbol{\alpha}(x)\cdot\boldsymbol{I} + i\frac{g'}{2}\beta(x)Y}\chi_L\,, \tag{1.14}$$

$$\psi_R \to e^{i\frac{g'}{2}\beta(x)Y}\psi_R \tag{1.15}$$

with the coupling strengths $g$ and $g'$ and the local phases $\boldsymbol{\alpha}(x)$ and $\beta(x)$. Usually, the three Pauli matrices $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ are used for the weak isospin operator: $\boldsymbol{I} = \boldsymbol{\sigma}/2$. In analogy to the procedure for the QED and QCD Lagrangians, invariance under these local gauge transformations is required for the Lagrangian of a free field, leading to couplings with the gauge fields. The Lagrangian can be written as the sum of the terms for the freely propagating left- and right-chiral fields:

$$\mathscr{L} = i\overline{\chi}_L\gamma^\mu\partial_\mu\chi_L + i\overline{\psi}_R\gamma^\mu\partial_\mu\psi_R\,, \tag{1.16}$$

where no mass-term for the fermionic fields are considered. In order to obtain an invariant

Lagrangian, the derivative $\partial_\mu$ is replaced by the covariant derivative $D_\mu$

$$D_\mu = \partial_\mu + ig\boldsymbol{I}\boldsymbol{W}_\mu + i\frac{g'}{2}YB_\mu \tag{1.17}$$

by introducing three gauge fields $\boldsymbol{W}_\mu = (W^1_\mu, W^2_\mu, W^3_\mu)$ for the group $SU(2)_{I,L}$, one gauge field $B_\mu$ for the group $U(1)_Y$ and two couplings $g$ and $g'$. The replacement of $\partial_\mu$ by $D_\mu$ in Eq. 1.16 and the addition of the kinetic terms for the gauge fields yield the locally invariant electroweak Lagrangian for one generation of fermions:

$$\begin{aligned}\mathscr{L}_{\mathrm{EW}} = {} & \overline{\chi}_L\gamma^\mu\left(i\partial_\mu - g\boldsymbol{I}\boldsymbol{W}_\mu + \frac{g'}{2}YB_\mu\right)\chi_L + \overline{\psi}_R\gamma^\mu\left(i\partial_\mu - \frac{g'}{2}YB_\mu\right)\psi_R \\ & - \frac{1}{4}\boldsymbol{W}_{\mu\nu}\boldsymbol{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}\,,\end{aligned} \tag{1.18}$$

where

$$\begin{aligned} W^i_{\mu\nu} &= \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g\epsilon^{ijk}W^j_\mu W^k_\nu \quad \text{with} \quad i,j,k = 1,2,3\,, \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \end{aligned} \tag{1.19}$$

are the field-strength tensors of the gauge fields with $\epsilon^{ijk}$ being the Levi-Civita tensor. Owing to the non-Abelian structure of $SU(2)_{I,L}$, Eq. 1.19 has an additional term leading to self-interactions of the $W^i_\mu$ fields. The physical neutral fields, the photon field $A_\mu$ and the $Z$-boson field $Z_\mu$, are obtained from a mixture of the gauge fields $W^3_\mu$ and $B_\mu$:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_w & -\sin\theta_w \\ \sin\theta_w & \cos\theta_w \end{pmatrix}\begin{pmatrix} W^3_\mu \\ B_\mu \end{pmatrix} \tag{1.20}$$

with the weak mixing angle $\theta_w$ defined by

$$\cos\theta_w = \frac{g}{\sqrt{g^2 + (g')^2}} \quad \text{and} \quad e = g\sin\theta_w = g'\cos\theta_w \tag{1.21}$$

being the relation between the electric unit charge $e$, the weak mixing angle $\theta_w$ and the couplings $g$ and $g'$. This mixture allows the photon field as well as the $Z$-boson field to interact with left- and right-chiral particles. The coupling strength to the photon field is the same for left- and right-chiral particles while the coupling strength to the $Z$-boson field differs for left- and right-chiral particles. However, it is ensured that only the $Z$-boson field interacts with left-(right-)chiral (anti-)neutrinos while the coupling between photon and neutrino fields vanishes. The fields of the known charged $W$-bosons which are eigenstates of the charge operator are obtained by combinations of the two gauge fields $W^1_\mu$ and $W^2_\mu$:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}\left(W^1_\mu \mp iW^2_\mu\right)\,. \tag{1.22}$$

The $W^1_\mu$- and $W^2_\mu$-bosons and hence also the $W^\pm$-bosons couple only to left-chiral fermions. The weak coupling strength $\alpha_w$ is defined as $\alpha_w = g^2/(4\pi)$.

### Electroweak Symmetry Breaking

However, up to now no mass terms for the gauge bosons $W^\pm$ and $Z$, which are experimentally known to be heavy [42–44], were introduced and the fermions are still assumed to be massless as well. Adding naive mass terms of the form $M^2W^i_\mu W^{i,\mu}$ for the gauge bosons or $m(\overline{\psi}_R\chi_L + \overline{\chi}_L\psi_R)$ for the fermions would violate the gauge invariance, lead to a non-renormalisable theory and violate the unitarity bound in scattering processes at high energies. In order to introduce mass terms for the gauge bosons and fermions while preserving gauge

invariance, the $SU(2)_{I,L} \otimes U(1)_Y$ symmetry is spontaneously broken to a $U(1)_Q$ symmetry by introducing a weak isospin doublet $\Phi$ with weak isospin 1/2 and weak hypercharge 1. This mechanism was proposed in 1964 independently by Robert Brout and François Englert [1], by Peter Higgs [2] and by Gerald Guralnik, Carl Richard Hagen and Tom Kibble [3] and is today mainly referred to as *Higgs mechanism*.

The weak isospin doublet consists of complex scalar fields and can be expressed as

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \Phi_1 + i\Phi_2 \\ \Phi_3 + i\Phi_4 \end{pmatrix} . \tag{1.23}$$

The corresponding Lagrangian is

$$\mathscr{L}_H = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi) , \tag{1.24}$$

where the potential $V(\Phi)$ is given by

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda \left(\Phi^\dagger \Phi\right)^2 \tag{1.25}$$

and $D_\mu$ is defined as given in Eq. 1.17. The potential consists of a mass term and a self interaction term for $\Phi$ and is the most general form that fulfils gauge invariance and renormalisability. The fields $\Phi^+$ and $\Phi^0$ have electric charges of +1 and 0, respectively. The free parameter $\lambda$ is required to be larger than 0 due to vacuum stability. For $\mu^2 > 0$, the potential would have a purely parabolic shape with its minimum at $\Phi_{\min} = 0$. For $\mu^2 < 0$, however, the potential has the shape shown in Figure 1.1. In this case, the potential has

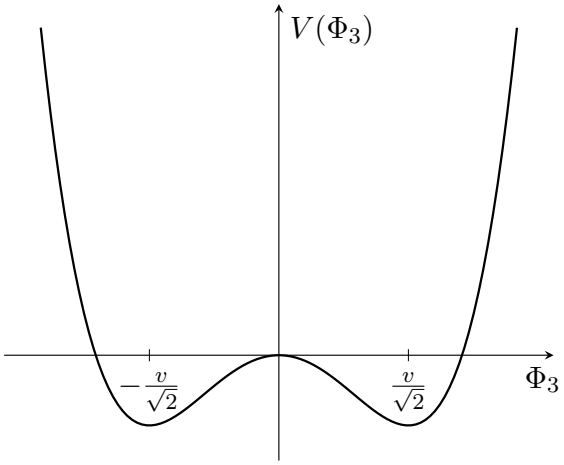



**Figure 1.1:** The Higgs potential $V(\Phi)$ for one degree of freedom ($\Phi_3$).

an infinite number of degenerate groundstates $\Phi_{\min}$, differing from 0. For presentation purposes, the potential in Figure 1.1 is shown only for a single degree of freedom and hence possesses only two groundstates. The number of groundstates becomes infinite for two or more degrees of freedom. All combinations of the fields $\Phi_1, \Phi_2, \Phi_3, \Phi_4$ fulfilling

$$\Phi^\dagger_{\min} \Phi_{\min} = \frac{-\mu^2}{2\lambda} = \frac{v^2}{2} \quad \text{with} \quad v^2 \equiv \frac{-\mu^2}{\lambda} \tag{1.26}$$

minimise $V(\Phi)$. Without loss of generality, one specific ground state, also called vacuum expectation value (vev) can be chosen. Considering that the first component $\Phi^+$ of $\Phi$ must

vanish since the vacuum is neutral, one possible choice for the ground state $\Phi_{\text{min}}$ is:

$$\Phi_{\text{min}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} . \tag{1.27}$$

By the choice of one specific ground state, the $SU(2)_{I,L} \otimes U(1)_Y$ symmetry is broken. However, $\Phi_{\text{min}}$ remains invariant under $U(1)_Q$ transformations with $Q = I_3 + Y/2$ (Eq. 1.12) due to being electrically neutral which results in the preservation of the masslessness of the photon. The expansion of $\Phi$ around the ground state $\Phi_{\text{min}}$ results in:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \Phi^+ \\ v + H + i\eta \end{pmatrix} \quad \xrightarrow{\text{unitary gauge}} \quad \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix} , \tag{1.28}$$

with the real fields $H$ and $\eta$ and the complex field $\Phi^+$. The former is the Higgs field, the latter two correspond to Goldstone fields with three massless Goldstone bosons [45, 46]. By choosing an appropriate gauge, the *unitary gauge* [47, 48], the fields $\Phi^+$ and $\eta$ vanish and the released degrees of freedom are taken up by the longitudinal polarisations of the three massive gauge bosons $W^i_\mu$ and only the Higgs field remains. Hence, the Higgs field describes excitations above the vacuum which result in a scalar boson called *Higgs boson*. By inserting the righ term of Eq. 1.28 for $\Phi$ into Eq. 1.24, making use of Eqs. 1.20–1.22 and remembering that the weak isospin $I$ and the weak hypercharge $Y$ for $\Phi$ are $1/2$ and 1, respectively, the Lagrangian becomes:

$$\begin{aligned} \mathscr{L}_H = \frac{1}{2} \left(\partial_\mu H\right) \left(\partial^\mu H\right) + \frac{g^2}{8 \cos^2 \theta_w} \left(v + H\right)^2 Z_\mu Z^\mu + \frac{g^2}{4} \left(v + H\right)^2 W^+_\mu W^{-,\mu} \\ - \frac{\mu^2}{2} \left(v + H\right)^2 - \frac{\lambda}{4} \left(v + H\right)^4 . \end{aligned} \tag{1.29}$$

It includes terms bilinear in the fields $Z$, $W^\pm$ and $H$ while the photon field $A$ is absent. The factors in front of the bilinear (mass) terms allow for identifying terms for the masses of the bosons:

$$M_{W^\pm} = \frac{gv}{2}, \quad M_Z = \frac{gv}{2 \cos \theta_w} = \frac{M_{W^\pm}}{\cos \theta_w}, \quad m_H = \sqrt{2\lambda v^2} = \sqrt{-2\mu^2} . \tag{1.30}$$

The ratio of the $W$- and $Z$-boson masses is determined by the weak mixing angle. The vacuum expectation value $v$ is obtained from Fermi's constant $G_F$ via the relation $v^2 = (\sqrt{2} G_F)^{-1}$ to be approximately 246 GeV[1] with $G_F$ being determined experimentally. Further terms in Eq. 1.29 are interaction terms of the gauge fields $W^\pm$ or $Z$ and the Higgs field $H$ of the form $HVV$ and $H^2VV$ where $V$ stands for either of the three gauge bosons. The strength of these interactions is proportional to $M^2_V$. The last two terms in Eq. 1.29 corresponding to the potential, provide triple and quartic self-interaction terms of the Higgs field proportional to $m^2_H$ as well as the mass term and contributions to the cosmological constant.

[1] using $G_F = 1.1663787 \times 10^{-5}\,\text{GeV}^{-2}$ [30]

### Yukawa Couplings

The description of fermion masses is incorporated into the theory by introducing an additional, gauge invariant term to the Lagrangian which features the same Higgs doublet:

$$\mathscr{L}_{\text{Yukawa}} = -\sum_{i,j=1}^{3} \overline{\chi}_{L,L,i} G_{\ell,ij} \psi_{R,\ell,j} \Phi + \overline{\chi}_{L,Q,i} G_{u,ij} \psi_{R,u,j} \Phi_c + \overline{\chi}_{L,Q,i} G_{d,ij} \psi_{R,d,j} \Phi + \text{h.c.}\,, \tag{1.31}$$

where $\chi_{L,L,i}$ denotes the left-chiral $SU(2)$-doublet of lepton generation $i$, $\psi_{R,\ell,j}$ the right-chiral singlet of charged lepton generation $j$ and $\chi_{L,Q,i}$, $\psi_{R,u,j}$ and $\psi_{R,d,j}$ equivalently are the quark doublet and the up- and down-type quark singlets for quark generation $i$ or $j$, respectively. The indices $i, j$ run over the three fermion generations. The matrices $G_f$, $f = \ell, u, d$, are arbitrary, complex $3 \times 3$ matrices with the matrix elements $G_{f,ij}$. The charge-conjugate $\Phi_c = i\sigma_2 \Phi^* = ((\phi^0)^*, -\Phi^-)^T$ Higgs doublet is necessary to generate masses for the upper components of the quark doublets and has opposite quantum numbers compared to $\Phi$. The abbreviation "h.c." stands for hermitian conjugate and indicates that the hermitian conjugate version of all previous terms is added in addition.

Using the term for $\Phi$ after symmetry breaking and in the unitary gauge as in Eq. 1.28, the Lagrangian of Eq. 1.31 becomes:

$$\mathscr{L}_{\text{Yukawa}} = -\frac{1}{\sqrt{2}} \sum_{i,j=1}^{3} G_{\ell,ij} \overline{\psi}_{L,\ell,i} \psi_{R,\ell,j} (v + H) + G_{u,ij} \overline{\psi}_{L,u,i} \psi_{R,u,j} (v + H) + G_{d,ij} \overline{\psi}_{L,d,i} \psi_{R,d,j} (v + H) + \text{h.c.}\,, \tag{1.32}$$

where $\psi_{L,f,i}$ denotes the respective component of the left-chiral $SU(2)$-doublet of fermion generation $i$. Expanding the term of the Lagrangian further, mass terms for the charged fermions proportional to $\overline{\psi}_{L,f,i} \psi_{R,f,j}$ and interaction terms with the Higgs field proportional to $\overline{\psi}_{L,f,i} \psi_{R,f,j} H$ become apparent. By reading off the proportionality factors, three mass matrices $M_f$ with elements $M_{f,ij} = v/\sqrt{2} G_{f,ij}$ and three coupling matrices $Y_f$, called Yukawa matrix, with elements $Y_{f,ij} = 1/\sqrt{2} G_{f,ij}$ are obtained.

The off-diagonal elements of the matrices $G_f$ allow for a mixture of the left- and right-chiral components of different fermions' generations. However, the matrices $G_f$ can be diagonalised by a transformation with the unitary matrices $U^{f_H}$ with $H = L, R$ such that the mass matrix as well as the Yukawa matrix become diagonal:

$$M_{\text{f,diag}} = \frac{v}{\sqrt{2}} U^{f_L} G_f (U^{f_R})^\dagger \quad \text{and} \quad Y_{\text{f,diag}} = \frac{1}{\sqrt{2}} U^{f_L} G_f (U^{f_R})^\dagger\,, \tag{1.33}$$

where a separate matrix for the charged leptons, up-type and down-type quarks as well as for right- and left-chiral fields is needed. Consequently, the fermion fields are transformed from the weak into the mass basis by:

$$\hat{\psi}_{H,f} = U^{f_H} \psi_{H,f}\,, \tag{1.34}$$

By applying these transformations and using $M_{\text{diag}} = \text{diag}\{m_{f,1}, m_{f,2}, m_{f,3}\}$ where the elements are the masses $m_{f,i}$ of the fermions of generation $i$, the Lagrangian of Eq. 1.31 becomes [29]:

$$\mathscr{L}_{\text{Yukawa}} = -\sum_f m_f \left(\overline{\psi}_{L,f} \psi_{R,f} + \overline{\psi}_{R,f} \psi_{L,f}\right) \left(1 + \frac{H}{v}\right)\,, \tag{1.35}$$

where the sum is over all fermion flavours of all generations. The hats, indicating the mass basis, are dropped. By reordering the two terms in Eq. 1.33, it becomes apparent that the coupling strength of the fermions to the Higgs boson is:

$$Y_f = \frac{m_f}{v}\,. \tag{1.36}$$

This means the coupling of the Higgs boson to fermions is proportional to the mass of the fermion which can be investigated experimentally. Furthermore, it is flavour conserving in the SM.

The same change of basis (Eq. 1.34) must coherently be performed in all components of the SM Lagrangian. When doing so, the only additional change is the appearance of the CKM-matrix, $V_{\mathrm{CKM}} = U^{u_L}(U^{d_L})^\dagger$ introduced in Eq. 1.13, which performs the transformation of the lower components of the left-chiral quark doublets in Eq. 1.18 from the mass basis into the weak basis. No equivalent matrix for the left-chiral leptons appears as the charged leptons and the neutrinos can be transformed with the same unitary matrix $U^{\ell_L}$ which results in $U^{\ell_L}(U^{\ell_L})^\dagger = 1$. The matrices also vanish for the right-chiral fields. The transformation of charged leptons and neutrinos with the same matrix is possible due to neutrinos being massless in the SM and hence are diagonal in any basis. Once non-vanishing masses for the neutrinos are assumed, the transformation matrices for the charged leptons and the neutrinos differ, as it is the case for the up- and down-type quarks. This leads to the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix [49] for leptons equivalent to the CKM-matrix for quarks. More details are discussed in Section 1.4.

### 1.1.5 The Complete Standard Model Lagrangian

The complete Lagrangian of the Standard Model is obtained as the sum of the Lagrangians of QCD, the electroweak interaction, the Higgs-mechanism and the Yukawa interactions. The single parts need small modifications in terms of notation in order to obtain one coherent term for the SM Lagrangian and the fermion mass term included by hand in $\mathscr{L}_{\mathrm{QCD}}$ must be dropped.

## 1.2 The Higgs Boson

A neutral scalar boson, the Higgs boson, was predicted as a consequence of the electroweak symmetry breaking discussed in Section 1.1.4. It was the last particle of the SM to be observed which succeeded in 2012. Two experiments at the LHC, ATLAS and CMS, announced the observation of a particle in the searches for the Higgs boson of the SM. Thenceforth, both experiments continued to investigate the properties of the newly found particle and up to date all results are in agreement with the SM predictions. The mass of the Higgs boson is a free parameter in the SM but once it is measured, the production cross sections and the branching ratios of the Higgs boson can be calculated. Hence, the following discussion of the production processes of the Higgs boson at the LHC and its decay modes assumes a mass of $m_H = 125\,\mathrm{GeV}$. Afterwards, an overview of the latest experimental results is given.

### 1.2.1 Production Processes at the LHC

The total production cross section of the Higgs boson in proton-proton ($pp$) collisions at the LHC is several magnitudes below the cross sections of other processes of the SM as can be seen in Figure 1.2 which results in an a-priori low signal-over-background ratio and thus requires exceptional particle detectors and sophisticated analysis techniques to be enhanced. Figure 1.2 also shows that to date, all of the measured SM cross sections are in agreement

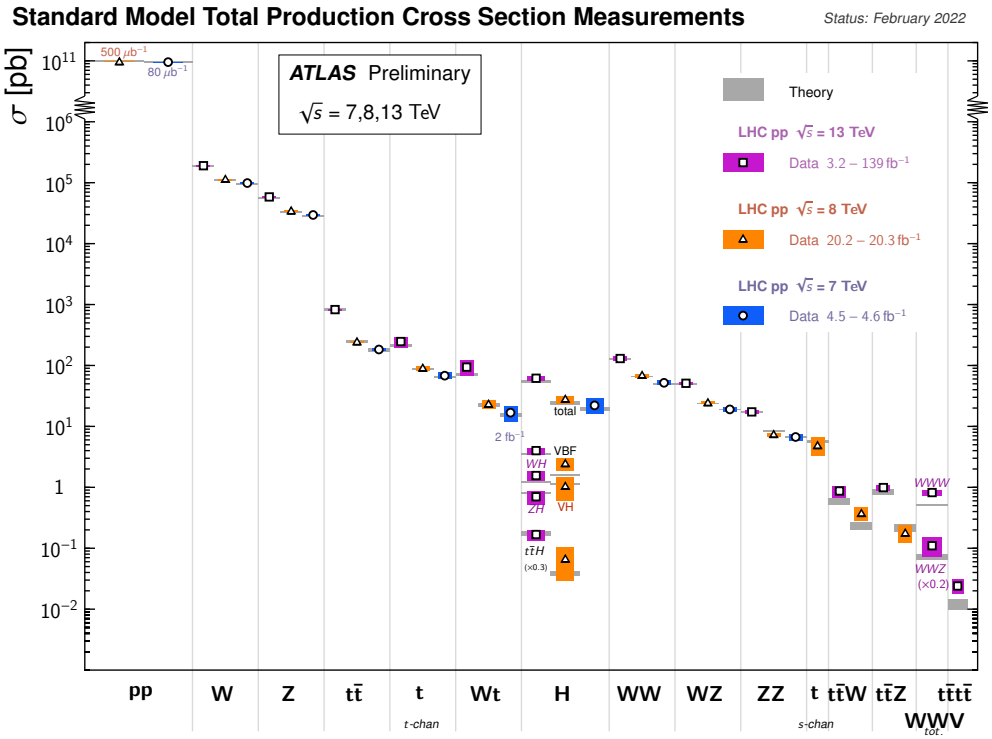



**Figure 1.2:** Measurements of total cross sections of several Standard odel processes by the ATLAS collaboration for different center-of-mass energies in proton-proton collisions [50]. For comparison, the theoretical predictions are indicated in gray.

with the prediction. The Higgs boson is produced in *pp*-collisions via several processes which differ in their cross sections and phenomenology. Partons, gluons and quarks, which are the constituents of protons serve as the actual collision partners in *pp*-collisions. Further details on the calculation of these cross sections is given in Section 4.1. The four main production processes are *gluon-gluon fusion* (ggF), *vector-boson fusion* (VBF), *Higgs strahlung* (VH) and the production in association with a top-quark pair (ttH). The corresponding Feynman diagrams at leading order and the cross sections as a function of the center-of-mass energy are shown in Figures 1.3(a)–1.3(d) and Figure 1.3(e), respectively. The values of the cross sections and their uncertainties quoted in the following are for a center-of-mass energy of 13 TeV and are taken from [51].

The most dominant production process is ggF which accounts for 87 % of all produced Higgs bosons. The coupling of the massless gluons to the Higgs boson is mediated by a heavy-quark loop which is dominated by the top quark due to the proportionality of the Yukawa couplings to the mass of the fermions. The cross-section is 48.58 pb, calculated at next-to-next-to-next-to-leading order ($N^3LO$) in the strong coupling strength $\alpha_s$ [52]. Electroweak (EW) corrections calculated at next-to-leading order (NLO) in the electromagnetic coupling strength $\alpha_{QED}$ are included as well [53–56]. The theoretical uncertainties are $+4.56\,\%/-6.72\,\%$, including QCD scale uncertainties, and $\pm 3.2\,\%$ from the parton density functions and $\alpha_s$. In contrast to the other production processes, no additional particles appear in the final state at leading order. Owing to momentum conservation, the Higgs boson has zero transverse momentum in these cases. Additional partons can arise in the final state, however, by radiation of gluons off the incoming gluons or the quarks in the loop which results in a transverse momentum of the Higgs boson different from zero.

The second largest production process is VBF with a cross section of 3.78 pb which is more than one order of magnitude below the ggF cross section. Thus, the VBF production

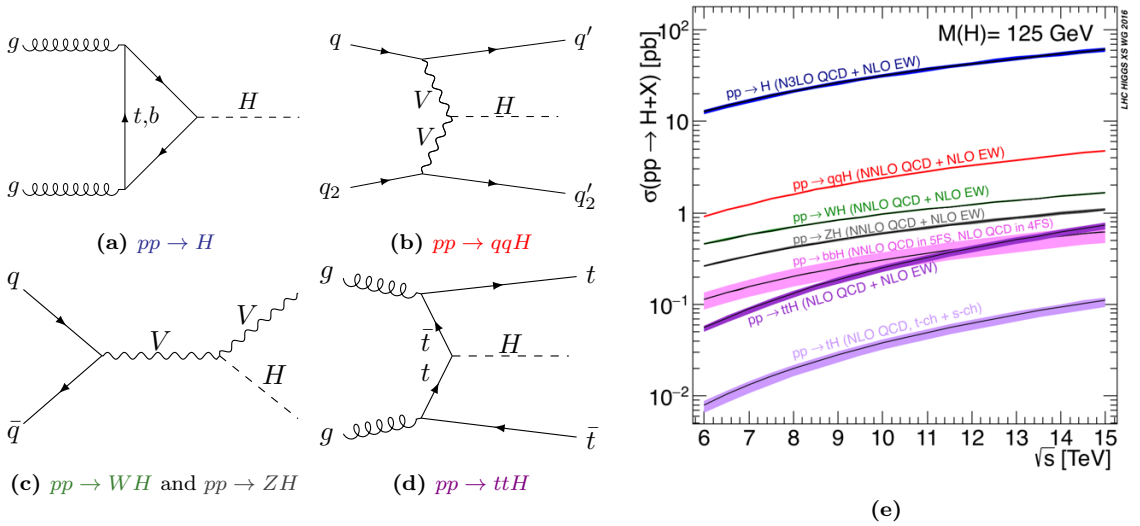



**Figure 1.3:** Feynman diagrams at leading order of the four main production processes of the Higgs boson in $pp$-collisions (a-d) with $V = W, Z$ and the cross sections of all production processes as a function of the center-of-mass energy (e) for $m_H = 125\,\text{GeV}$ [51].

accounts for 7 % of all produced Higgs bosons. The calculation of the cross section includes NNLO corrections in $\alpha_s$ (QCD corrections) and NLO corrections in $\alpha_{\text{QED}}$ (EW corrections) [57–59]. The QCD scale uncertainties are $+0.43\,\%/-0.33\,\%$ and uncertainties from the parton density functions and $\alpha_s$ are $\pm 2.1\,\%$. In the VBF process, two quarks originating from each of the protons radiate off a $W$- or $Z$-boson which fuse to produce a Higgs boson. Hence, the VBF production is characterised by two additional high-energetic quarks in the final state which manifest themselves as jets, one in the forward and the other one in the backward region of the detector. This distinct pattern can be exploited to suppress background processes. Furthermore, VBF production allows for investigating the coupling of the Higgs boson to vector bosons.

The VH process accounts for 4 % of all produced Higgs bosons. A quark and an anti-quark annihilate to produce a $W$- or $Z$-boson which radiates off a Higgs boson. The signature with the additional vector boson in the final state can be utilised to suppress background processes, e.g. when requiring leptonic decays of the vector boson since charged light leptons can be measured precisely in the detector. The cross section of the Higgs boson produced in association with a $W$-boson is 1.37 pb with $+0.5\,\%/-0.7\,\%$ QCD scale uncertainties and $\pm 1.7\,\%$ pdf and $\alpha_s$ uncertainties. The production in association with a $Z$-boson has a cross section of 0.88 pb and uncertainties of $+3.8\,\%/-3.1\,\%$ and $\pm 1.6\,\%$, respectively. The calculation of the cross section includes NNLO corrections in $\alpha_s$ (QCD corrections) and NLO corrections in $\alpha_{\text{QED}}$ (EW corrections) [60–62].

The cross section of the ttH production process is two orders of magnitudes smaller than the one of ggF. It is 0.51 pb with uncertainties of $+5.8\,\%/-9.2\,\%$ from the QCD scales and of $\pm 3.6\,\%$ from the pdfs and $\alpha_s$. The cross section is calculated at NLO accuracy in $\alpha_s$ and $\alpha_{\text{QED}}$ [63–69]. The ttH production accounts for approximately 1 % of all produced Higgs bosons. It allows for investigating the coupling of the Higgs boson to the top quark which is not possible in decays of the Higgs boson as it is lighter than the top quark.

Other production mechanisms are bbH where the Higgs boson is produced in associaten with a bottom-quark pair in analogy to the ttH production and tH where the Higgs boson is produced in association with a single top quark. The former is of the size of ttH production but plagued with considerable background contributions. The cross section of the latter is almost another order of magnitude smaller than ttH.

### 1.2.2 Decay

The prediction in the SM for the total decay width of the Higgs boson with a mass of $m_H = 125\,\text{GeV}$ is $\Gamma_H = 4.07\,\text{MeV}$ [30] with a relative uncertainty of $+4.0\,\%/-3.9\,\%$. This is equivalent to a lifetime of $1.6 \times 10^{-22}\,\text{s}$, meaning that the Higgs boson rapidly decays after its production. The total decay width is the sum of the partial widths corresponding to the different decay channels of the Higgs boson. The so-called *branching ratios* ($\mathcal{B}$) are defined as the ratio of the partial widths over the total width. The branching ratios as a function of the Higgs boson mass are shown in Figure 1.4. The coupling strength of the

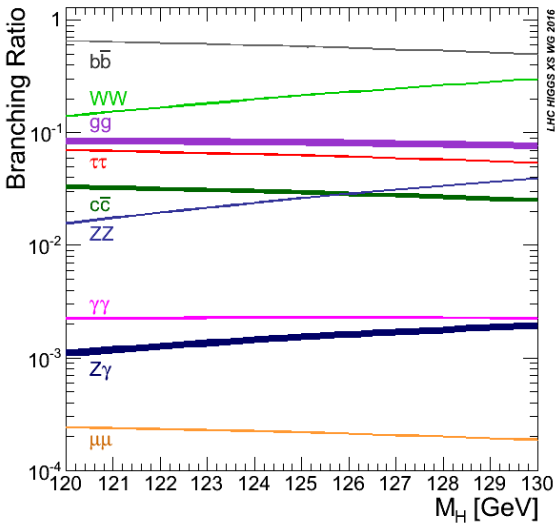



**Figure 1.4:** Branching ratios of Higgs boson decay channels as a function of the Higgs boson mass. The width of the curves indicates the uncertainties [51].

Higgs boson to fermions is proportional to their mass while the coupling strength to gauge bosons is proportional to the square of their mass. This means the decay probability into third generation fermions is considerably larger than into first generation fermions. The available phase space plays another important role in the size of the branching ratios.

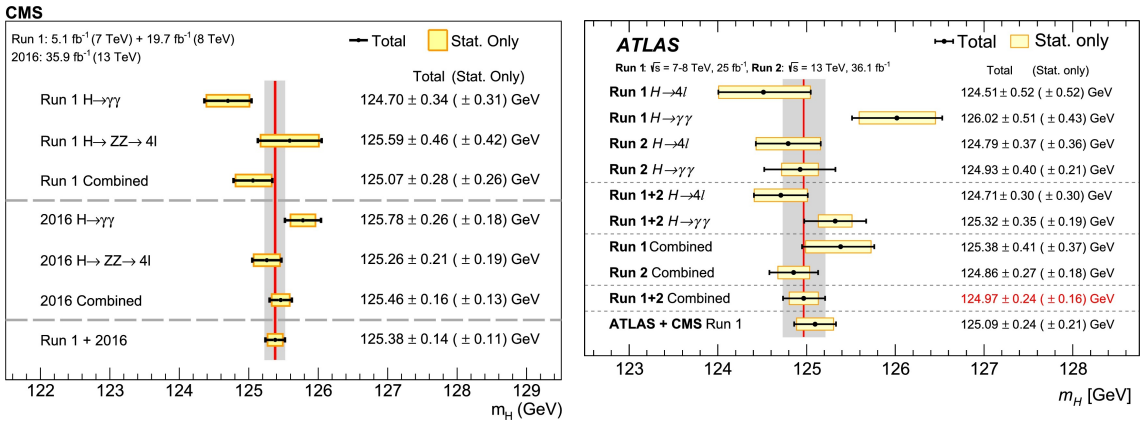
At $m_H = 125\,\text{GeV}$ the decay mode with the largest branching ratio of 58 % is $H \to b\bar{b}$. The decay into two $W$-bosons has the second largest branching ratio of 21 %. This is only allowed as the $W$-boson is short-lived and hence can be produced virtually. The decay into two gluons has a branching ratio of 8.2 % and is mediated by a heavy-quark loop, dominated by the top quark, analogously to the ggF production process. The decay into $\tau$-leptons with a branching ratio of 6.3 % is the most probable decay into leptons. It is followed by the decay into a pair of charm quarks ($\mathcal{B} = 2.9\,\%$) and a pair of $Z$-bosons ($\mathcal{B} = 2.6\,\%$). In the latter, one of the $Z$-bosons must be produced virtually. Although the decay into two $Z$-bosons has a low branching ratio, it is experimentally well accessible when both bosons decay leptonically such that four light charged leptons are in the final state. Another decay with a clean signature in the detector but an even lower branching ratio of 0.23 % is the decay into two photons. Finally, the decay into a $Z$-boson and a photon has a branching ratio of 0.15 % and the decay into two muons of 0.02 %. The branching ratios of the decays into first generation fermions are consequently even lower.

### 1.2.3 Experimental Results

The observation of a neutral boson with a mass of approximately 125 GeV in the search for the Higgs boson was announced on July 4th 2012 by the ATLAS and CMS collaborations [4, 5]. The observation was based on data taken at center-of-mass energies of $\sqrt{s} = 7\,\text{TeV}$ and

$\sqrt{s} = 8\,\text{TeV}$ and the background-only hypothesis could be rejected with $5.9\,\sigma$ (ATLAS) and $5.0\,\sigma$ (CMS), driven by the two search channels with the best mass resolution: $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$. These two decay modes indicate that the new particle is a neutral boson. And, according to the Landau-Yang theorem [70, 71], no massive spin-1 particle can decay into two photons. Hence, the excess found in the $H \to \gamma\gamma$ decay channel disfavours the observed new boson to have a spin of 1, in accordance with the SM prediction. In addition, the SM predicts that the Higgs boson is even under transformations of charge and parity (CP). Measurements in diboson decays of the Higgs boson excluded several alternative spin and CP hypotheses in favour of the SM prediction at more than 99.9 % confidence level [72].

The currently most precise measurement of the Higgs-boson mass is performed by the CMS collaboration by combining the results of Run 1 of the LHC with results obtained from data taken in 2016 [73]. The mass of the Higgs boson is obtained by reconstructing the invariant mass of its decay products. Hence, the measurement is solely based on the two decay channels $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$ where no neutrinos, which escape the detector undetected, and only objects which have a clear signature in the detector are in the final state. The measured value of the Higgs-boson mass is $125.38 \pm 0.14\,\text{GeV}$. The values of the single measurements contributing to the combined measurement are listed in Figure 1.5(a). Figure 1.5(b) summarises the measurements performed by the ATLAS




**(a)** **(b)**

**Figure 1.5:** Measured Higgs-boson mass values of the two collaborations, CMS (a) [73] and ATLAS (b) [74], with different datasets and in two decay channels and various combinations thereof.

collaboration and also quotes the value obtained from a combination of the Run 1 results of both collaborations.

The total decay width of the Higgs boson is not accessible via direct measurements as the value predicted by the SM ($\Gamma_H = 4.07\,\text{MeV}$ [30]) is three orders of magnitude below the experimental resolution of the width. However, if a deviation of the decay width from the SM prediction were found, this would be a sign for new physics beyond the SM (BSM). It was suggested [75–77] that the ratio of the cross section of the off-shell production of the Higgs boson to the on-shell production is proportional to the decay width. This was exploited by the ATLAS and CMS collaborations in the decay channels $H \to ZZ^* \to 4\ell$ and $H \to ZZ^* \to 2\ell 2\nu$ which resulted in the following measured values for the width of the Higgs boson: $\Gamma_H = 4.6^{+2.6}_{-2.5}\,\text{MeV}$ [78] (ATLAS) and $\Gamma_H = 3.2^{+2.4}_{-1.7}\,\text{MeV}$ [79] (CMS) which is in agreement with the SM prediction. Simultaneously, evidence for the off-shell production of the Higgs boson was given.

Depending on the total width, the branching ratios of the different (known) decay modes may vary. Hence, a combined measurement of the different decay modes is pursued. Although the production and decay of the Higgs boson are independent processes, they

cannot be investigated fully disentangled. Therefore, a signal strength modifier is defined: $\mu_{if} = (\sigma_i/\sigma_i^{\mathrm{SM}}) \cdot (\mathcal{B}_f/\mathcal{B}_f^{\mathrm{SM}})$ [80] for a production process $i$ and a decay mode $f$. Perfect agreement with the SM prediction is obtained for $\mu_{if} = 1$. The presentation of the following results focuses on measurements performed by the ATLAS collaboration with the full Run 2 dataset [80]. The results of similar measurements performed by the CMS collaboration can be found in [81].

The values of $\mu_{if}$ are measured for different combinations of production processes and decay channels under varying assumptions. Assuming that all production processes and decay modes scale with the same signal strength modifier $\mu$, both collaborations perform a fit to the data of all production modes and decay channels. The resulting values of $\mu$ are both in agreement with the SM: $\mu = 1.05 \pm 0.06$ (ATLAS) and $\mu = 1.002 \pm 0.057$ (CMS).

Individual $\mu_{if}$ for different combinations of production processes and decay modes are utilised to obtain the values shown in Figure 1.6. All of the measurements are in

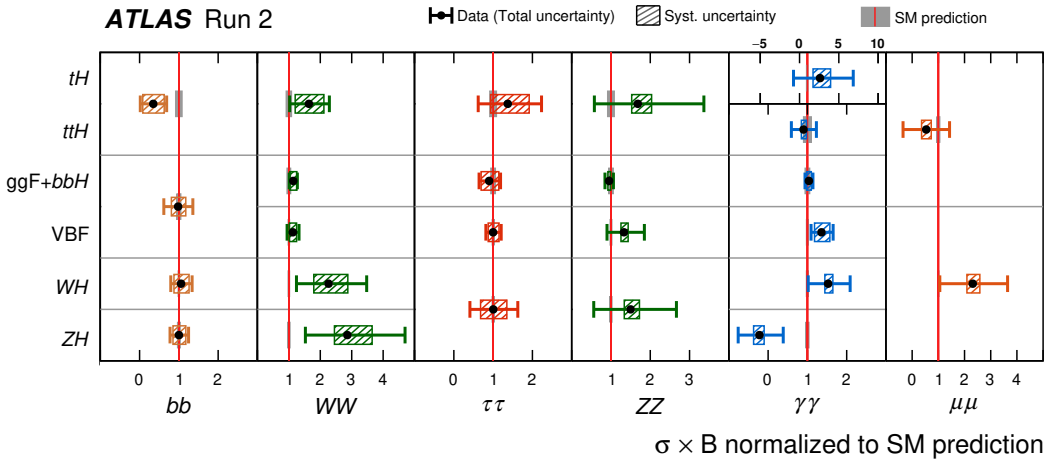



**Figure 1.6:** Values of the signal strength modifiers $\mu_{if}$ for different combinations of production processes and decay modes, obtained with the full Run 2 dataset by the ATLAS collaboration [80]. The gray bands indicate the SM prediction, the horizontal bars the 68 % confidence intervals.

agreement with the SM predictions. In addition, values for the different $\mu_i$ ($\mu_f$) were obtained when fixing the branching ratios (production cross sections) to their SM prediction. The production processes WH, ZH and ttH and tH combined are now also observed with significances of at least $5\,\sigma$ after ggF and VBF were already observed in Run 1. In addition to the decays into $\gamma\gamma$, $ZZ$, $WW$ and $\tau\tau$ which were already observed in Run 1, the decay into a pair of bottom ($b$) quarks is now also established. The rare decays into a pair of second-generation muons is measured with a significance of $2.0\,\sigma$ and the decay into $Z\gamma$ with $2.3\,\sigma$. The CMS collaboration found evidence for the decay of the Higgs boson into a pair of muons with a significance of $3.0\,\sigma$ [82].

In order to extract the values of particular Higgs-boson coupling strengths, a simultaneous fit within the $\kappa$-framework [83] is performed. The coupling strength modifier $\kappa_p$ for the coupling to a certain particle $p$ is defined as $\kappa_p^2 = \sigma_p/\sigma_p^{\mathrm{SM}}$ for a production process and as $\kappa_p^2 = \Gamma_p/\Gamma_p^{\mathrm{SM}}$ for a decay mode [80]. It is taken into account that the total decay width of the Higgs boson on one hand depends on the decay modes that are considered in the measurements to data but on the other hand also on yet undetected or invisible decays and decays not predicted by the SM [80]. The latter contribution is divided into decays to invisible particles ($\mathcal{B}_{\mathrm{inv.}}$) and decays which are undetected due to large backgrounds ($\mathcal{B}_{\mathrm{u}}$).

Depending on the assumptions, different models can be constructed and the values of the corresponding coupling strength modifiers extracted. In order to obtain the values for the modifiers shown in Figure 1.7, separate modifiers for each of the $W$, $Z$, $t$, $b$, $c$, $\tau$ and $\mu$

particles are assumed while the modifiers of the first-generation fermions are set to their SM-values and the second-generation modifiers are set to the ones of the third generation. Furthermore, effective modifiers for the loop-induced processes are used: $\kappa_g$, $\kappa_\gamma$ and $\kappa_{Z\gamma}$ for

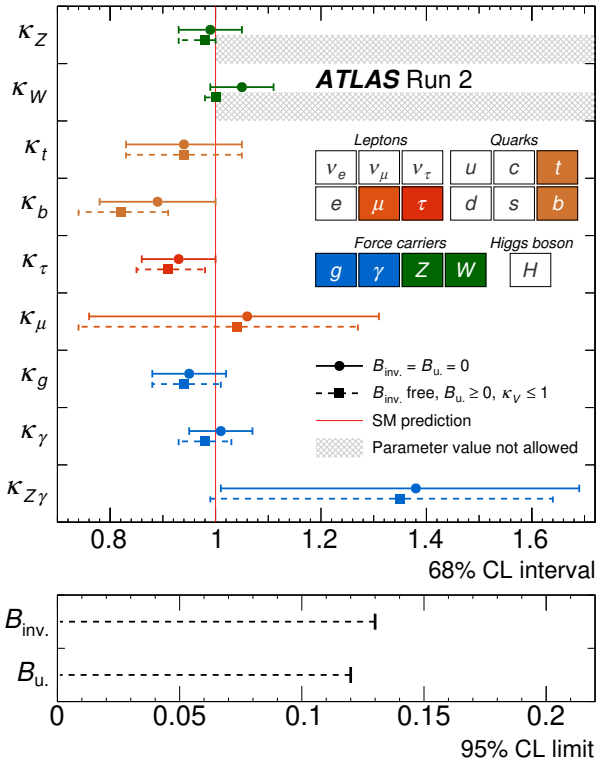



**Figure 1.7:** The measured values of the coupling strength modifiers when excluding non-SM decays from the decay width (circles and solid lines) and when allowing them (squares and dashed lines) [80]. The horizontal bars indicate the 68 % confidence intervals.

the couplings to two gluons, two photons and a $Z$-boson and a photon, respectively. This allows for non-SM particles in the loops. In addition, the $\kappa$-values are shown for two cases: excluding non-SM decays of the Higgs boson ($\mathcal{B}_{\text{inv.}} = 0$ and $\mathcal{B}_{\text{u}} = 0$) and allowing them to contribute. For the latter $\mathcal{B}_{\text{u}} \geq 0$ and $\kappa_V \leq 1$ with $V = W, Z$ is assumed. The values of all measured $\kappa$-parameters are in agreement with the SM prediction. No significant change in the values is observed whether invisible and undetected non-SM decays are allowed or not. When allowing these decays to contribute to the total width, upper limits on their branching ratios at 95 % confidence level are set: $\mathcal{B}_{\text{u}} < 0.12$ and $\mathcal{B}_{\text{inv.}} < 0.13$.

## 1.3 The Charged Leptons

In addition to the Higgs boson the charged leptons are of particular interest in this thesis. All values quoted in the course of this section are taken from [30]. The $\tau$-lepton is the heaviest of all leptons with a mass of $m_\tau = 1776.86 \pm 0.12\,\text{MeV}$. It has a short mean lifetime of $\tau_\tau = (290.3 \pm 0.5) \times 10^{-15}\,\text{s}$ corresponding to a proper decay length of $c\tau = 87.03\,\mu\text{m}$. Hence, $\tau$-leptons decay within the beam pipe of the LHC and must be reconstructed through their decay products. The decay is driven solely by the weak force meaning that the $\tau$-lepton always decays into a $\tau$-neutrino and a virtual $W$-boson. The $W$-boson subsequently decays either hadronically into quarks or leptonically into an electron or muon and the respective anti-neutrino. If it decays hadronically, the quarks hadronise and form colour-neutral mesons such that at least one charged meson is in the final state and optionally also neutral

ones. These mesons are predominantly pions. In 49.04 % of all $\tau$-lepton decays, there is one charged meson in the final state, in 15.2 % there are three charged mesons in the final state while in 17.82 % and 17.39 % of the decays an electron or a muon are in the final state. The branching ratios of the decays with more than three charged mesons are very small.

The muon has a mass of $m_\mu = 105.66\,\text{MeV}$ and a mean lifetime of $\tau_\mu = 2.20 \times 10^{-6}\,\text{s}$ corresponding to a proper decay length of $c\tau = 658.64\,\text{m}$ meaning that it traverses the full depth of the ATLAS detector and can be detected directly. Muons decay almost exclusively into an electron, an electron anti-neutrino and a muon neutrino.

The electron has a mass of $m_e = 0.511\,\text{MeV}$ and a mean lifetime of at least $6.6 \times 10^{28}$ years and hence is stable and therefore can also be detected directly.

## 1.4 Lepton-Flavour Violation

Lepton-flavour is an accidental symmetry and hence conserved in the SM. The Yukawa-interaction matrix can be diagonalised in the mass basis of the leptons as was shown in Section 1.1.4. However, oscillations of neutrinos were observed by the Super-Kamiokande experiment [7] in 1998 and the Sudbury Neutrino Observatory [8] in 2001 after first evidence was found by the Homestake experiment [6] in 1968 which was rewarded with the 2015 and 2002 Nobel Prize of Physics. The discovery that neutrinos oscillate, meaning that they can transmute from one flavour state into another, implies that lepton flavour is not conserved in nature and hence not an exact symmetry. Therefore, the questions arises whether lepton-flavour violation is also realised in the charged lepton sector. The oscillations imply that neutrinos have non-vanishing masses in order to allow for different mixtures of mass eigenstates at different points in time (in fact it is required that they have differing masses). This necessitates some form of extension of the SM. Once non-vanishing neutrino masses are incorporated, the mass matrices of the neutrinos and the charged leptons cannot be diagonalised with the same matrix $U^{\ell_L}$ anymore which results in the appearance of the PMNS-matrix and hence in mixing within the lepton sector, as discussed in Section 1.1.4. The PMNS-matrix can also give rise to charged lepton-flavour violation (CLFV) which can, however, in this context only occur in loop diagrams [84]. An example of such a CLFV process is shown in Figure 1.8. The corresponding branching ratio of $\mu \to e\gamma$ in a minimally

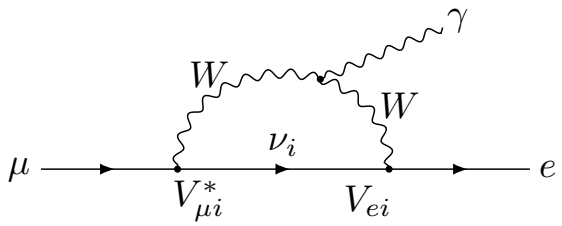



**Figure 1.8:** Diagram of $\mu \to e\gamma$ mediated by neutrinos $\nu_i$ with non-vanishing masses in the SM. The $\nu_i$ with $i = 1, 2, 3$ are mass eigenstates, ordered by mass.

extended SM incorporating neutrino masses can be calculated as follows [85, 86]:

$$\mathcal{B}(\mu \to e\gamma) = \frac{3\alpha}{32\pi} \left| \sum_{i=2,3} V^*_{\mu i} V_{ei} \frac{\Delta m^2_{i1}}{M^2_W} \right|^2 < 10^{-54}\,, \tag{1.37}$$

where $\alpha$ is the fine structure constant, $V_{\beta i}$ are the elements of the PMNS-matrix, $\Delta m^2_{ij}$ are the differences of the squared neutrino masses and $M_W$ is the mass of the $W$-boson. The branching ratio is highly suppressed by the GIM-mechanism [87] and hence the observation of such a decay would clearly be an indication of new physics beyond the SM (BSM). The

values obtained for the branching ratios of similar processes, e.g. $\tau \to e\gamma$ and $\tau \to \mu\gamma$, are of similar order of magnitude.

However, numerous models extending the SM were and are developed in order to address the mentioned and other shortcomings of the SM, e.g. the lack of explanation for the evidence of dark matter and the baryon asymmetry in the universe. In many of them (C)LFV processes occur naturally, including LFV decays of the Higgs boson which are the topic of this thesis. Owing to the small decay width of the Higgs boson, decays of the Higgs boson corresponding to new BSM physics could have a sufficiently sized branching ratio in order to be detected. In addition, the current upper limits on the branching ratios of non-SM decays, obtained in combined fits of the coupling strength modifiers $\kappa$ discussed in Section 1.2.3, do not yet exclude such decays.

The following Section 1.4.1 discusses the appearance of CLFV in an effective field theory ansatz of the SM (SMEFT) and in type-III two-Higgs-doublet models. Section 1.4.2 gives an overview of the current indirect and direct limits on LFV decays of the Higgs.

To date, no hints or evidence for CLFV were found although many searches in numerous channels were and are extensively performed.

### 1.4.1 LFV in Charged Lepton Sector in Extensions of the SM

Charged lepton-flavour violation arises naturally in many models extending the SM, e.g. two-Higgs-doublet models (2HDM) [9–11], composite Higgs-models [12, 13], minimal supersymmetric standard model (MSSM) [14], Ranadall-Sundrum models [15] and the Froggatt-Nielsen mechanism [16]. However, it can be useful to investigate beyond-SM phenomena within effective theories in order to set model-independent limits on parameters of the effective theory within a so-called bottom-up approach. Hence, CLFV is discussed first in the framework of the SM effective field theory in the next section. Afterwards it is reviewed in the context of one specific model, the type-III 2HDM.

#### LFV Interactions of the Higgs Boson in SMEFT

Effective field theories (EFT) are used to describe effects of new physics in a model-independent way at the "low-energy" scale accessible by an experiment. In this context new physics is characterised by new heavy particles at some high-energy scale $\Lambda$. In the framework of the Standard Model effective field theory (SMEFT), it is assumed that the SM is valid up to the energy scale $\Lambda$ above which the effective theory (the SM) must be replaced by the full (unknown) fundamental theory. The energy scale $\Lambda$ is required to be above the electroweak scale $v = 246\,\text{GeV}$ by several orders of magnitude. At the electroweak scale only the SM particles are relevant and for $\Lambda \to \infty$ the full theory should reduce to the SM. The SMEFT aims to provide a systematic parametrisation of all potentially possible new physics residing at scale $\Lambda$ at the electroweak scale $v$. The Lagrangian describing the SMEFT can be written as

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \frac{1}{\Lambda}\sum_k C_k^{(5)} Q_k^{(5)} + \frac{1}{\Lambda^2}\sum_k C_k^{(6)} Q_k^{(6)} + \mathcal{O}\left(\frac{1}{\Lambda^3}\right), \tag{1.38}$$

where $\mathscr{L}_{\text{SM}}$ denotes the usual SM Lagrangian of mass-dimension 4, the sum runs over the number of higher-dimensional operators $Q_k$ which are built from the usual SM fields and are gauge-invariant under $SU(3) \otimes SU(2)_{I,L} \otimes U(1)_Y$. The numbers in brackets denote the dimension of the respective operators. The summands are suppressed by inverse powers of the new physics scale. The new physics at high energies is absorbed into the dimensionless Wilson coefficients $C_k$ while the higher-dimensional operators represent the effect at the low-energy scale. This separation of scales is possible due to the decoupling theorem [88]. The Wilson coefficients are unknown but could in principle be calculated from the full

theory or any other BSM model affecting the respective coefficient. On the other hand experimental data can be used to set bounds on the coefficients and thus on corresponding models (bottom-up approach).

There are many operators per dimension, a complete set of independent operators of dimension 5 and 6 is listed in [89]. After requiring SM gauge invariance, only one operator of dimension 5 remains. It is the only operator allowing for massive neutrinos and their mixing but does not introduce LFV interactions of the Higgs boson [10, 89]. Hence this operator is not considered in the following. Among the 59 independent dimension 6 operators[2] for one fermion generation only one gives direct rise to modifications in the Yukawa-couplings and the fermion masses [18] and hence to tree-level flavour-changing neutral currents mediated by the Higgs-boson [90]. Following the notation in [89] this operator is called $Q_{e\Phi}$. It is of the form

$$Q_{e\Phi} = (\Phi^\dagger \Phi)(\bar{\chi}_{L,L,i}\psi_{R,\ell,j}\Phi)\,. \tag{1.39}$$

The notation of the fermion fields is as introduced in Eq. 1.31: $\chi_{L,L,i}$ denotes the left-chiral $SU(2)$-doublet of lepton generation $i$ and $\psi_{R,\ell,j}$ the right-chiral singlet of charged lepton generation $j$. The SM Higgs doublet is denoted by $\Phi$. Adding the Lagrangian corresponding to this operator to the lepton-related part of the Yukawa-Lagrangian of the SM (Eq. 1.32) and denoting the corresponding Wilson coefficients by $C_{ij}$, the resulting Lagrangian after electroweak symmetry breaking is:

$$\mathscr{L}^{\text{modif.}}_{\text{Yuk.},\ell} = -\frac{1}{\sqrt{2}}\sum_{i,j=1}^{3} G_{\ell,ij}\overline{\psi}_{L,\ell,i}\psi_{R,\ell,j}(v+H) + \frac{C_{ij}}{2\Lambda^2}(v+H)^2\overline{\psi}_{L,\ell,i}\psi_{R,\ell,j}(v+H) + \text{h.c.} \tag{1.40}$$

$$\begin{aligned} &= -\frac{1}{\sqrt{2}}\sum_{i,j=1}^{3}\left(G_{\ell,ij}v + \frac{C_{ij}}{2\Lambda^2}v^3\right)\overline{\psi}_{L,\ell,i}\psi_{R,\ell,j} + \left(G_{\ell,ij} + \frac{C_{ij}}{2\Lambda^2}3v^2\right)\overline{\psi}_{L,\ell,i}\psi_{R,\ell,j}H \\ &\qquad + \frac{C_{ij}}{2\Lambda^2}3v\overline{\psi}_{L,\ell,i}\psi_{R,\ell,j}H^2 + \frac{C_{ij}}{2\Lambda^2}\overline{\psi}_{L,\ell,i}\psi_{R,\ell,j}H^3 + \text{h.c.}\,, \end{aligned} \tag{1.41}$$

where the first term in Eq. 1.41 corresponds to the mass terms of the leptons and the second term to the Yukwawa interactions. The last terms represent interaction terms of two leptons with either two or three Higgs bosons. The mass matrix $M$ as well as the Yukawa matrix $Y$, with elements $M_{ij}$ and $Y_{ij}$, gain additional terms compared to the SM only case discussed in Section 1.1.4 and become

$$M = \frac{1}{\sqrt{2}}\left(G_\ell v + \frac{C}{2\Lambda^2}v^3\right)\,, \tag{1.42}$$

and

$$Y = \frac{1}{\sqrt{2}}\left(G_\ell + \frac{C}{2\Lambda^2}3v^2\right)\,. \tag{1.43}$$

Again, as discussed in Section 1.1.4 the mass matrix can be diagonalised by a transformation with new unitary matrices $U'_{\ell,H}$ with $H = L, R$ [18]:

$$M_{\text{diag}} = \frac{v}{\sqrt{2}}U'_{\ell,L}\left(G_\ell + \frac{v^2}{2\Lambda^2}C\right)(U'_{\ell,R})^\dagger\,, \tag{1.44}$$

[2] 64 when allowing for violation of baryon number conservation

and since $Y = M/v + v^2/(\sqrt{2}\Lambda^2)C$, it cannot be diagonalised simultaneously. Instead, $Y$ expressed in the mass basis becomes:

$$\begin{aligned}\hat{Y} = U'_{\ell,L} Y (U'_{\ell,R})^\dagger = U'_{\ell,L}\left(\frac{M}{v} + \frac{v^2}{\sqrt{2}\Lambda^2}C\right)(U'_{\ell,R})^\dagger &= \frac{M_{\text{diag}}}{v} + \frac{v^2}{\sqrt{2}\Lambda^2}U'_{\ell,L}C(U'_{\ell,R})^\dagger \\ &= \frac{M_{\text{diag}}}{v} + \frac{v^2}{\sqrt{2}\Lambda^2}C'\,,\end{aligned} \tag{1.45}$$

which corresponds to the SM coupling plus an additional, not necessarily diagonal term that is suppressed by the squared scale of new physics. Thus, the coupling strength of charged leptons $i$ and $j$ to the Higgs boson is

$$\hat{Y}_{ij} = \frac{m_i}{v}\delta_{ij} + \frac{v^2}{\sqrt{2}\Lambda^2}C'_{ij}\,, \tag{1.46}$$

and hence introduces lepton-flavour violating interactions of the Higgs boson. Measuring or constraining the values of the elements of the matrix $C'$ is the aim of experimental searches for LFV decays of the Higgs boson.

### LFV Interactions of the Higgs Boson in 2HDM

Two Higgs-doublet models (2HDM) extend the SM by adding an additional Higgs doublet with hypercharge $Y = 1$ and can, for example, generate a sufficient baryon asymmetry. Naturally, 2HDMs come with sizeable contributions of tree-level flavour-changing neutral currents (FCNCs) and thus also lepton-flavour violation [11]. These can be suppressed by the requirement of discrete symmetries which is applied in type-I and type-II 2HDMs. In the following, however, a 2HDM allowing FCNCs is used which is typically referred to as type-III or general 2HDM. The subsequent description is based on [91] with additional background information from [11]. Both Higgs doublets, $\Phi_1$ and $\Phi_2$, can couple to all fermion types in the type-III 2HDM. Furthermore, the *Higgs basis* in which only one Higgs doublet acquires a vacuum expectation value $v$ can be used. Then, the scalar potential of the model reads

$$\begin{aligned}V = {} & M_{11}^2\Phi_1^\dagger\Phi_1 + M_{22}^2\Phi_2^\dagger\Phi_2 - M_{12}^2\left(\Phi_1^\dagger\Phi_2 + \Phi_2^\dagger\Phi_1\right) + \frac{\lambda_1}{2}\left(\Phi_1^\dagger\Phi_1\right)^2 + \frac{\lambda_2}{2}\left(\Phi_2^\dagger\Phi_2\right)^2 \\ & + \lambda_3\left(\Phi_1^\dagger\Phi_1\right)\left(\Phi_2^\dagger\Phi_2\right) + \lambda_4\left(\Phi_1^\dagger\Phi_2\right)\left(\Phi_2^\dagger\Phi_1\right) + \frac{\lambda_5}{2}\left[\left(\Phi_1^\dagger\Phi_2\right)^2 + \left(\Phi_2^\dagger\Phi_1\right)^2\right] \\ & + \lambda_6\left[\left(\Phi_1^\dagger\Phi_1\right)\left(\Phi_1^\dagger\Phi_2\right) + \left(\Phi_1^\dagger\Phi_1\right)\left(\Phi_2^\dagger\Phi_1\right)\right] \\ & + \lambda_7\left[\left(\Phi_2^\dagger\Phi_2\right)\left(\Phi_1^\dagger\Phi_2\right) + \left(\Phi_2^\dagger\Phi_2\right)\left(\Phi_2^\dagger\Phi_1\right)\right]\end{aligned} \tag{1.47}$$

with all parameters being real (assuming CP-conservation[3]) and the $\lambda_i$ being the Higgs self-coupling constants. The Higgs doublets can be expressed as

$$\Phi_1 = \begin{pmatrix} G^+ \\ \frac{1}{\sqrt{2}}\left(v + \Phi_1^0 + iG^0\right)\end{pmatrix} \quad \text{and} \quad \Phi_2 = \begin{pmatrix} H^+ \\ \frac{1}{\sqrt{2}}\left(\Phi_2^0 + iA\right)\end{pmatrix}, \tag{1.48}$$

where $G^+$ and $G^0$ are Goldstone bosons, $\Phi_1^0$, $\Phi_2^0$ and $A$ are neutral scalars and $H^+$ is a charged scalar. The scalar $A$ is a mass eigenstate and odd under the combined charge and parity transformation (CP). Assuming CP conservation, it does not mix with the CP-even

[3] and hence no additional baryon asymmetry

states $\Phi_1^0$, $\Phi_2^0$. The latter two are mixed to obtain the mass eigenstates $h$ and $H$:

$$\begin{pmatrix} h \\ H \end{pmatrix} = \begin{pmatrix} \sin(\beta-\alpha) & \cos(\beta-\alpha) \\ \cos(\beta-\alpha) & -\sin(\beta-\alpha) \end{pmatrix} \begin{pmatrix} \Phi_1^0 \\ \Phi_2^0 \end{pmatrix} . \tag{1.49}$$

Here, $\beta-\alpha$ is the mixing angle and $m_h < m_H$ with $h$ being identified with the Higgs boson discovered in 2012. The corresponding Yukawa Lagrangian for the leptons can be written in terms of the Higgs basis for the Higgs doublets and the mass basis for the leptons:

$$\mathscr{L}_{\text{Yuk.},\ell}^{\text{2HDM}} = -\frac{\sqrt{2}}{v} \sum_{i,j=1}^{3} \overline{\chi}_{L,L,i} M_{\text{diag},ij} \psi_{R,\ell,j} \Phi_1 + \overline{\chi}_{L,L,i} \rho_{ij} \psi_{R,\ell,j} \Phi_2 + \text{h.c.} , \tag{1.50}$$

where the sum runs over the lepton flavours, $M_{\text{diag}}$ is the diagonal mass matrix and $\rho$ is a general complex hermitian $3\times3$ matrix in flavour space with elements $\rho_{ij}$. After electroweak symmetry breaking, using Eqs. 1.48 and 1.49, the Lagrangian becomes:

$$\begin{aligned} \mathscr{L}_{\text{Yuk.},\ell}^{\text{2HDM}} = &- \sum_{i,j=1}^{3} \overline{\psi}_{L,\ell,i} \left( \frac{1}{v} M_{\text{diag},ij} \sin(\beta-\alpha) + \frac{1}{\sqrt{2}} \rho_{ij} \cos(\beta-\alpha) \right) \psi_{R,\ell,j} h \\ &+ \overline{\psi}_{L,\ell,i} \left( \frac{1}{v} M_{\text{diag},ij} \cos(\beta-\alpha) - \frac{1}{\sqrt{2}} \rho_{ij} \sin(\beta-\alpha) \right) \psi_{R,\ell,j} H \\ &+ \frac{i}{\sqrt{2}} \overline{\psi}_{L,\ell,i} \rho_{ij} \psi_{R,\ell,j} A + \overline{\psi}_{L,\nu,i} \left( V_{ik}^{\dagger} \rho_{kj} \right) \psi_{R,\ell,j} H^{+} + \text{h.c.} , \end{aligned} \tag{1.51}$$

with $V$ being the PMNS matrix due to assuming the neutrinos in the mass basis. Hence, the coupling $Y$ matrix with elements $Y_{ij}$ of the light scalar boson $h$ to the leptons is:

$$Y = \frac{M_{\text{diag}}}{v} \sin(\beta-\alpha) + \frac{1}{\sqrt{2}} \rho \cos(\beta-\alpha) , \tag{1.52}$$

which is not necessarily diagonal. The second term, introducing off-diagonal elements, becomes zero for $\beta-\alpha = \pi/2$ for which the first, solely diagonal, term is maximal. In this scenario, the SM case (Eq. 1.36) is reproduced and the alignment limit for the 2HDM is reached [92]. For small deviations of $\sin(\beta-\alpha)$ from 1, however, lepton-flavour violating interactions of the SM-like Higgs boson are introduced. But the mixing of the two scalar bosons $h$ and $H$, steered by $\beta-\alpha$, must still be in agreement with the measurements of the couplings of the Higgs boson and further constraints, e.g. from low-energy measurements discussed in Section 1.4.2. A random scan of the parameters $m_H, m_A, m_{H^\pm}$ and $\sin(\beta-\alpha)$ is performed in [91] with a focus on $h \to \tau\mu$ while fixing $m_h = 125.1\,\text{GeV}$, thereby respecting the aforementioned experimental constraints. It was found "that the type-III 2HDM can induce $h \to \tau\mu$ rates close to the current bound while being in agreement with all experimental constraints" [91], where the "current bound" refers to the latest limits by direct searches for $h \to \mu\tau$ by the ATLAS and CMS collaborations discussed in Section 1.4.2. In these cases, $\sin(\beta-\alpha)$ is close to 1 (but obviously not exactly 1; $\sim 0.9$ in [91]).

The decoupling limit [93] offers one possibility to realise the alignment limit. However, not the exact alignment ($\sin(\beta-\alpha)=1$) is desired but an approximate alignment ($\sin(\beta-\alpha) \lesssim 1$) where LFV decays of the Higgs boson are allowed but bounded. In the decoupling limit, the light scalar $h$ corresponds to the SM Higgs-boson while the remaining non-SM-like scalars, $H, A, H^\pm$, are very heavy which results in $\cos(\beta-\alpha)$ being close to 0 [93] and consequently in $\sin(\beta-\alpha)$ being close to 1. In this case the existence of heavy scalars only become noticeable at the electroweak scale through small deviations of the couplings of the light scalar $h$ from SM predictions. This clearly allows interpreting this 2HDM setup within SMEFT. Comparing the 2HDM $h\ell_i\ell_j$-coupling (Eq. 1.52) to the one obtained in the

previous section from the SMEFT ansatz (Eq. 1.45) shows a very similar but not identical structure. In principle, the Wilson coefficient matrix $C'$ could be expressed in terms of the 2HDM parameters, $\Lambda$ and $M_{\text{diag}}$.

### 1.4.2 Current Limits on LFV Decays of the Higgs Boson

This section gives an overview of indirect as well as direct limits on the branching ratios of LFV decays of the Higgs boson, i.e. on $\mathcal{B}(H \to e\mu)$, $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$.

#### Indirect Limits

Indirect limits can for example be derived from low-energy measurements searching for the decays $\mu \to e\gamma$, $\tau \to e\gamma$ and $\tau \to \mu\gamma$ (radiative LFV decays). There are 1-loop and 2-loop contributions of the Higgs-boson with CLFV couplings to these decays. The corresponding dominant diagrams of the 1-loop contributions are shown in Figure 1.9 for $\tau \to \mu\gamma$, but are equivalent for the other two decays. The 2-loop diagrams can be found in [18]. By

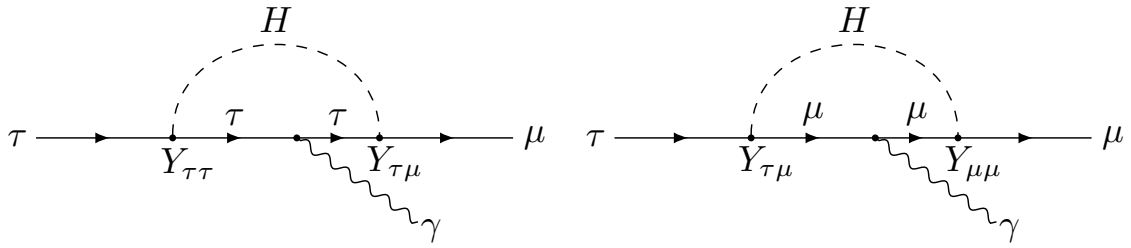



**Figure 1.9:** Example diagrams of the 1-loop contributions to the radiative LFV decay $\tau \to \mu\gamma$ that are mediated by a Higgs boson and include the LFV off-diagonal Yukawa couplings $Y_{\tau\mu}$ and $Y_{\mu\tau}$.

evaluating these diagrams and using the latest upper limits on the branching ratios of $\mu \to e\gamma$, $\tau \to e\gamma$ and $\tau \to \mu\gamma$, upper bounds $B$ on the off-diagonal Yukawa couplings $Y_{ij}$ with $i \neq j$ and $i, j = e, \mu, \tau$ of the form

$$\sqrt{|Y_{ij}|^2 + |Y_{ji}|^2} < B \tag{1.53}$$

can be derived [17, 18]. The strongest upper limits on these radiative LFV decays are obtained by the MEG experiment and by the BaBar collaboration and are summarised in Table 1.4. The MEG experiment is specialised in the search for $\mu \to e\gamma$-decays. It uses

| Process | Limit | Reference | Experiment |
|---|---|---|---|
| $\mu \to e\gamma$ | $4.2 \times 10^{-13}$ | [94] | MEG |
| $\tau \to e\gamma$ | $3.3 \times 10^{-8}$ | [95] | BaBar |
| $\tau \to \mu\gamma$ | $4.4 \times 10^{-8}$ | [95] | BaBar |

**Table 1.4:** Upper limits on the branching ratios of the respective processes at 90 % CL.

a continuous beam of positively charged muons which are stopped by a thin target such that the muons decay at rest. In contrast to the MEG experiment, the Babar experiment investigates many different processes originating from the collisions of electrons and positrons of different energy.

The upper bounds $B$ on $\sqrt{|Y_{ij}|^2 + |Y_{ji}|^2}$ from [18], obtained using the values in Table 1.4, are summarised in Table 1.5. It is assumed that "the Higgs is the only additional degree of freedom with mass $\mathcal{O}(100\,\text{GeV})$ and that the $Y_{ij}$'s are the only source of flavor violation" [18],

| Coupling | Bound | $H$ LFV Decay | Bound on $\mathcal{B}$ | From Process |
|---|---|---|---|---|
| $\sqrt{\lvert Y_{e\mu}\rvert^2 + \lvert Y_{\mu e}\rvert^2}$ | $< 3.6 \times 10^{-6*}$ | $H \to e\mu$ | $1.6 \times 10^{-8}$ | $\mu \to e\gamma$ |
| $\sqrt{\lvert Y_{e\tau}\rvert^2 + \lvert Y_{\tau e}\rvert^2}$ | $< 0.014$ | $H \to e\tau$ | 0.19 | $\tau \to e\gamma$ |
| $\sqrt{\lvert Y_{\mu\tau}\rvert^2 + \lvert Y_{\tau\mu}\rvert^2}$ | $< 0.016$ | $H \to \mu\tau$ | 0.24 | $\tau \to \mu\gamma$ |
| $\sqrt[4]{\lvert Y_{e\tau}Y_{\tau\mu}\rvert^2 + \lvert Y_{\tau e}Y_{\mu\tau}\rvert^2}$ | $< 2.2 \times 10^{-4}$ | - | - | $\mu \to e\gamma$ |

**Table 1.5:** Upper bounds on the off-diagonal Yukawa couplings from [18], obtained from the limits on the branching ratios of the radiative LFV processes, and the corresponding upper bounds on the branching ratios $\mathcal{B}$ of the LFV decays of the Higgs boson. The * indicates that this bound was obtained with a weaker bound on $\mu \to e\gamma$ of $2.4 \times 10^{-12}$ [96]. Thus, the current bound on $\sqrt{\lvert Y_{e\mu}\rvert^2 + \lvert Y_{\mu e}\rvert^2}$ is stronger.

with the consequence that only the 1-loop and 2-loop diagrams mediated by the Higgs boson contribute to the radiative LFV decays. Further, it is assumed that the flavour diagonal Yukawa couplings $Y_{ii}$ equal their SM prediction, that the second 1-loop diagram of Figure 1.9 can be neglected (for $Y_{\mu\mu} \ll Y_{\tau\tau}$) and it is used that $m_\mu \ll m_\tau \ll m_H$. In order to obtain upper bounds on the branching ratios of the Higgs-boson LFV decays from the bounds on the couplings, the following relations are utilised:

$$\Gamma(H \to ij) = \frac{m_H}{8\pi}\left(|Y_{ij}|^2 + |Y_{ji}|^2\right) \quad \text{and} \quad \mathcal{B}(H \to i\tau) = \frac{\Gamma(H \to ij)}{\Gamma_{\text{SM}}(H) + \Gamma(H \to ij)}\,. \quad (1.54)$$

The resultings bounds on the branching ratios $\mathcal{B}$ of the Higgs boson LFV decays are listed in Table 1.5. The bound on $\mathcal{B}(H \to e\mu)$ is very strong and below the sensitivity that can be reached by the experiments at the LHC. Whereas the bounds on the other two Higgs-boson LFV decays are rather weak allowing for branching ratios around 20 %.

In addition, a bound on the product of the couplings $Y_{\mu\tau}$ and $Y_{\tau e}$ can be derived from $\mu \to e\gamma$, which is also listed in Table 1.5. The corresponding 1-loop diagram is shown in Figure 1.10. The 2-loop contribution which is proportional to $Y_{\mu e}$ and $Y_{e\mu}$ is neglected. The

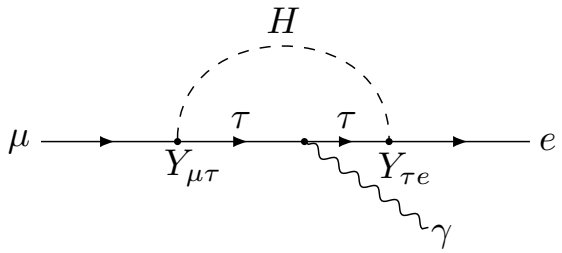



**Figure 1.10:** Diagram of the 1-loop contribution to the radiative LFV decay $\mu \to e\gamma$ that is mediated by a Higgs boson and features two LFV off-diagonal Yukawa couplings: $Y_{\mu\tau}$ and $Y_{\tau e}$.

consequence of this bound is that either $H \to e\tau$ or $H \to \mu\tau$ are allowed to be "large" but no sizeable contribution from both at once is expected. The largest value the two branching ratios, $H \to e\tau$ and $H \to \mu\tau$, can take on simultaneously, is $8.28 \times 10^{-3}$ %. The constraint from this bound on the branching ratios is visualised in Figure 11.8 in the discussion of the results.

Indirect upper limits on Higgs boson LFV decays can also be obtained from measurements of e.g. $\mu/\tau \to 3e$, $\tau \to \mu\gamma$, $\mu \to e$ conversions in nuclei and from the electron and muon electric and magnetic dipole moments [17, 18]. However, the limits obtained from the radiative LFV decays are the most stringent amongst these limits.

### Direct Limits

Direct limits on the branching ratios of the LFV decays of the Higgs boson were derived in dedicated searches by the ATLAS and CMS collaborations. Both collaborations placed limits at 95 % confidence level (CL) on all three $H$ LFV decays. The latest limits by each collaboration are summarised in Table 1.6. The strongest direct limit on $\mathcal{B}(H \to e\mu)$ is

| | ATLAS | | | CMS | | |
|---|---|---|---|---|---|---|
| Process | Limit on $\mathcal{B}$ | | Dataset | Limit on $\mathcal{B}$ | | Dataset |
| $H \to e\mu$ | $6.1 \times 10^{-5}$ | [97] | $139\,\text{fb}^{-1}$ | $3.5 \times 10^{-4}$ | [98] | $19.7\,\text{fb}^{-1*}$ |
| $H \to e\tau$ | $4.7 \times 10^{-3}$ | [19] | $36.1\,\text{fb}^{-1}$ | $2.2 \times 10^{-3}$ | [20] | $137\,\text{fb}^{-1}$ |
| $H \to \mu\tau$ | $2.8 \times 10^{-3}$ | [19] | $36.1\,\text{fb}^{-1}$ | $1.5 \times 10^{-3}$ | [20] | $137\,\text{fb}^{-1}$ |

**Table 1.6:** Most stringent direct upper limits on the branching ratios of $H$ LFV decays at 95 % CL by the ATLAS and CMS collaborations. All limits are obtained from data recorded at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$, except the one indicated by $*$ which used $\sqrt{s} = 8\,\text{TeV}$.

achieved by the ATLAS collaboration but is still almost three orders of magnitudes weaker than the indirect limit obtained from $\mu \to e\gamma$. The strongest direct limits on $H \to e\tau$ and $H \to \mu\tau$ are from the CMS collaboration, using the full Run 2 dataset. These limits outperform the indirect limits obtained by the low-energy measurements.

# Chapter 2

# The LHC and the ATLAS experiment

This chapter introduces the experimental setup. Data from proton-proton collisions created by the *Large Hadron Collider* (LHC) described in Section 2.1 and recorded by the ATLAS detector introduced in Section 2.2 are the basis for this thesis. Section 2.3 gives a brief overview of the luminosity measurement at the ATLAS experiment. An overview over the different data taking periods and the dataset used in the analysis presented in this thesis is given in Section 2.4.

## 2.1 The LHC

The LHC [99, 100] is the world's largest and most powerful circular particle accelerator to date. With a circumference of 26.7 km it is located underground at the France-Swiss border between Lake Geneva and the Jura Mountains in a depth of 45 to 175 m. It was built between 1998 and 2008 in the tunnel of the former *Large Electron-Positron Collider* by the *European Organization for Nuclear Research* (CERN)[1].

Two particle beams circulate in opposite directions in two beam pipes with an ultrahigh vacuum. The particles accelerated by the LHC are mainly protons ($p$) and, for a short run time period, also lead ions (Pb) and other heavy ions. In this thesis, however, only data from collisions of protons is analysed. At four points of the LHC, these two beams are brought to collision. One experiment is housed at each of the four points: ALICE (A Large Ion Collider Experiment) [101], ATLAS (A Toroidal LHC ApparatuS) [102] which is described in more detail in the next section, CMS (Compact Muon Solenoid) [103] and LHCb (Large Hadron Collider beauty) [104]. ATLAS and CMS are both multipurpose detectors while ALICE is sepcialized in the investigation of quark-gluon-plasma states originating from Pb-Pb collisions and LHCb is set up to study decays of hadrons that include $b$- or $c$-quarks (bottom or beauty and charm quarks).

The LHC is not a perfect circle but consists of alternating eight straight segments and eight arcs. 1323 superconducting dipole electromagnets bend the particles along the arcs and reach field strengths up to 8.3 T. The particle beams are focused by 392 quadrupole magnets and multiple additional magnets are used to influence and correct the particles' trajectories. The hadrons are accelerated and then kept at their energy by eight superconducting radiofrequency (RF) cavities per beam. The particles are, however, not solely accelerated by the LHC but travel through several pre-accelerators. The whole acceleration system is depicted in Figure 2.1.

The protons are obtained by ionising hydrogen atoms from a simple gas cylinder and are linearly accelerated to an energy of 50 MeV by LINAC2 (Linear Accelerator 2) which injects them into the PS Booster (Proton Synchroton Booster) where they reach 1.4 GeV.

[1] In French: Conseil européen pour la recherche nucléaire

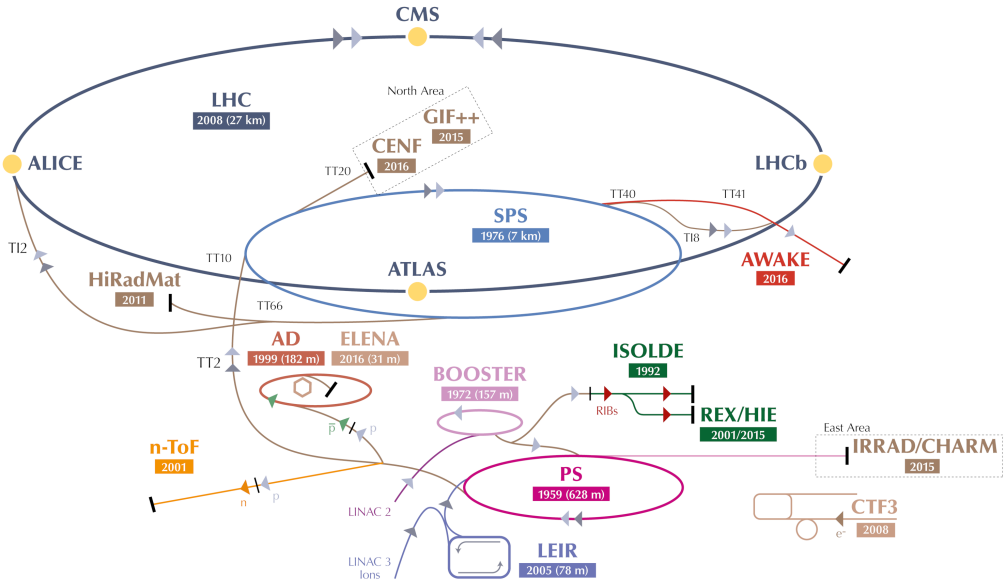



**Figure 2.1:** Schematic of the accelerator system [105].

They are further accelerated to 25 GeV by the PS (Proton Synchroton) and then to 450 GeV by the SPS (Super Proton Synchroton) from which the protons are finally transferred to the LHC. It takes the LHC twenty minutes until the protons reach their final energy of 6.5 TeV which corresponds to a center-of-mass energy of 13 TeV at collision.

In the case of $pp$-collisions each beam in the LHC during Run 2 consisted of up to 2556 bunches formed by up to $1.25 \times 10^{11}$ protons with a distances of 25 ns [106]. This bunch structure is already prepared by the pre-accelerators.

Besides the center-of-mass energy, the luminosity is an important measure of the performance of an accelerator. It has the dimension of events per time and area and is given by the LHC's beam parameters [99]:

$$\mathcal{L} = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \varepsilon_n \beta^*} F \,. \tag{2.1}$$

With $N_b$ being the number of particles per bunch, $n_b$ the number of bunches per beam and $f_{\text{rev}}$ the revolution frequency of the particles. Further, $\gamma_r$ is the relativistic gamma factor, $\varepsilon_n$ the normalised transverse beam emittance, $\beta^*$ the beta function, describing the transverse spread of the beams, at the collision point and $F$ a geometric reduction factor due to the crossing angle and other effects at the interaction point. Furthermore, the luminosity integrated over time $\mathcal{L}_{\text{int}} = \int \mathcal{L}\, dt$ is proportional to the number of produced events $N$ for a specific process which are given by:

$$N = \sigma \cdot \mathcal{L}_{\text{int}} \,. \tag{2.2}$$

Here, $\sigma$ is the cross section of the respective process.

Therefore, the aim is to maximise $\mathcal{L}_{\text{int}}$. Luminosities of up to $2.1 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$ were reached during Run 2, yielding an integrated luminosity of $138.4\,\text{fb}^{-1}$ from 2015-2018 used for analyses. The measurement of the integrated luminosity performed by ATLAS is described in Section 2.3.

## 2.2 The ATLAS Detector

The ATLAS detector is a multipurpose detector designed not only to precisely investigate the standard model of particle physics but also to search for new phenomena. It is built cylindrically around the beam axis with multiple layers of different sub-detectors going from the inside to the outside, see Figure 2.2. It weighs 7 000 tons, has a length of 44 m and a

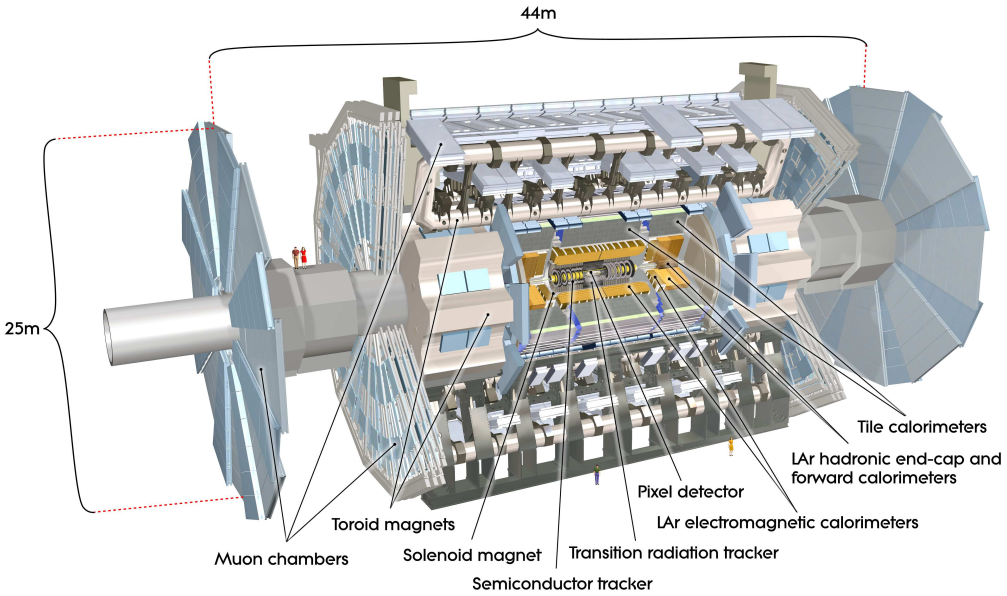



**Figure 2.2:** Schematic of the ATLAS experiment [102].

height of 25 m. Each sub-detector is optimised for detecting and measuring different particles or properties of particles such that when combining the information of all sub-detectors, different particles can be distinguished and finally, the events can be reconstructed. This is visualised in Figure 2.3.

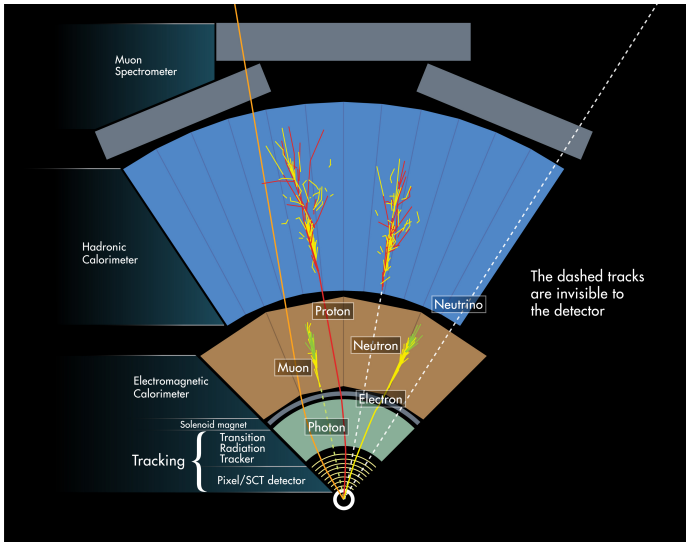



**Figure 2.3:** Diagram showing possible particle paths in the ATLAS detector, visualising how they are detected [107].

The *Inner Detector*, further described in Section 2.2.2, is used to determine decay vertices, the sign of charge, the momentum and the trajectory of charged particles by detecting their tracks. The calorimeters, surrounding the *Inner Detector*, measure the

energy of all particles which do not only interact via the weak interaction and their direction of travel. The *Electromagnetic Calorimeter* is primarily used for the detection of electrons and photons, the *Hadronic Calorimter* for the detection of hadrons. Both are described in more detail in Section 2.2.3. Muons pass these sub-detectors almost unhampered due to their large mass compared to electrons and therefore little interactions with the detector materials and since they do not interaction strongly. They are detected in the *Muon System*, described in Section 2.2.4, which also measures their momentum and charge.

Each sub-detector consists of three regions, the barrel region in the central part of the detector and the two endcaps in the forward and backward regions.

The description of the ATLAS detector follows [102] if not indicated otherwise.

### 2.2.1 Coordinate System

A right-handed Cartesian coordinate system with its origin at the point of collisions is used to describe the trajectories of the objects (particles, jets, missing transverse energy). The $x$-axis points from the origin to the center of the LHC ring, the $y$-axis from the origin to the earth surface and the $z$-axis along the beam line.

The azimuth angle $\phi$ is defined in the transverse plane (the $x$-$y$-plane) and gives the angle around the beam axis. The polar angle $\theta$ indicates the angle between the trajectory of the particle and the beam axis ($z$-axis). The rapidity $y$ is given by

$$y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right) , \tag{2.3}$$

where $E$ is the energy and $p_z$ the $z$-component of the momentum of the particle. An advantage of this observable is the Lorentz invariance of differences in rapidity under a boost in $z$-direction. In experimental particle physics, the rapidity is typically replaced by pseudorapidity $\eta$ which solely depends on the polar angle $\theta$:

$$\eta = - \ln \left( \tan \left( \theta / 2 \right) \right) . \tag{2.4}$$

For high energetic particles for which $m \ll p$ holds with $p$ being the momentum, the pseudorapidity converges to the rapidity. The values of $|\eta|$ range from 0 (perpendicular to the beam axis) to $\infty$ (parallel to the beam axis).

The angular separation of two objects $\Delta R$ is also used frequently and defined as follows:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} . \tag{2.5}$$

When the partons of the protons are colliding, the total momentum in $z$-direction is not known since the partons carry only parts of the proton's momentum. However, the momenta of the partons in the transverse plane are zero before they collide and therefore the total transverse momentum of all final state particles after collision must be zero as well. This means that the projection of the particle's momentum onto the transverse plane is of particular importance. It is denoted by $p_T$ and its absolute value is calculated from the $x$- and $y$-components as follows:

$$p_T = \sqrt{p_x^2 + p_y^2} . \tag{2.6}$$

Another important argument is its invariance under boosts in $z$-direction.

### 2.2.2 Inner Detector

The *Inner Detector* (ID) consists of three independent but complementary sub-detectors. The *Pixel* detector which is closest to the beam line is surrounded by the *Semiconductor*

*Tracker* (SCT). Both are silicon based and must withstand a high-radiation environment. The *Transition Radiation Tracker* (TRT) makes for the outer layers of the inner detector. It consists of gas-filled straw tubes interlaced with transition radiation material. The whole inner detector is contained within a cylindrical envelope of approximately seven metres length and 2.3 metres diameter and permeated by a 2 T magnetic field which is produced by the central super-conducting solenoid and is parallel to the beam axis. A sketch of a quarter-section of the ID showing all three sub-detectors is given in Figure 2.4.

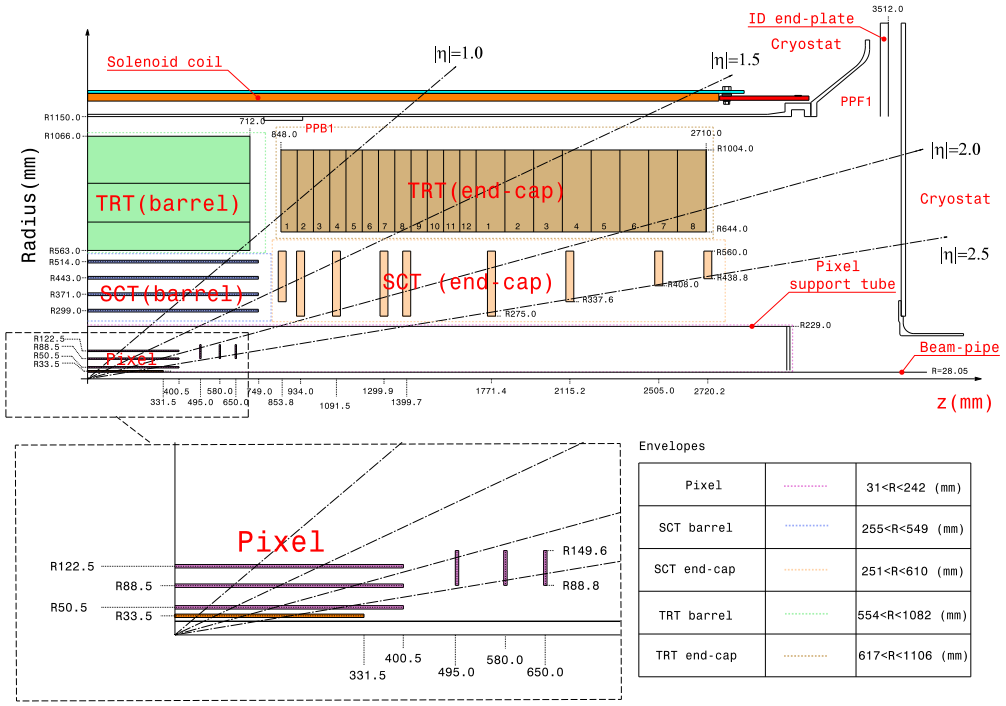



**Figure 2.4:** Sketch of a quarter-section of the ATLAS Inner Detector [108].

With this setup the ID is capable of providing reliable pattern recognition and tracking of charged particles, measurement of primary and secondary vertices and precise determination of momentum and charge. The latter are derived from the radius and direction of curvature of the particle's trajectory induced by the magnetic field. The target resolution of the ID for the transverse momentum measurement is $\sigma_{p_T}/p_T = 0.05\,\% p_T \oplus 1\,\%$.

The *Pixel* and SCT sensors are constructed on concentric cylinders around the beam axis in the barrel region and on disks perpendicular to the beam axis in the endcap regions. They cover the region $|\eta| < 2.5$.

The *Pixel* detector has three cylindrical layers and two times three disks at the endcaps. The layers are segmented in $R$-$\phi$ and $z$ with a pixel size of 50 µm × 400 µm. In total, there are 1 744 pixel sensors, each with 47 232 pixels which results in approximately 80 million readout channels. The intrinsic hit accuracy is 10 µm in $R$-$\phi$ and 115 µm in $z$ for the barrel and in $R$ for the endcaps. During *Long Shutdown 1*, between Run 1 and Run 2 of the LHC, a forth innermost layer was added to the *Pixel* detector in the barrel region: the *Insertable B-Layer* (IBL) [109]. The additional layer, its even finer granularity and the smaller distance to the interactions resulted in improved tracking and vertex reconstruction. The pixels have a size of 50 µm × 250 µm in $R$-$\phi \times z$.

The SCT consists of four silicon microstrip layers in the barrel and two times nine disk layers in the endcaps. Each of the layers is composed by a pair of strips with one strip being aligned to the beam line in the barrel and perpendicular to it in the endcaps. The other

strip is placed at a stereo angle of 40 mrad to enable the measurement of the $z$-coordinate in the barrel and of the $R$-coordinate in the endcaps. In total, there are 768 active strips with a length of 12 cm per sensor and a strip pitch of 80 µm. The intrinsic accuracy is 17 µm in $R$-$\phi$ and 580 µm in $z$ for the barrel and in $R$ for the endcaps. There are approximately 6.3 million readout channels. The SCT exhibits a coarser granularity compared to the *Pixel* detector and therefore a worse resolution but covers a larger sensitive area.

The TRT is the outermost and therefore also largest sub-detector of the ID. It is built by approximately 300 000 gaseous straw tubes with a diameter of 4 mm interlaced with stabilising carbon fiber structures. Straw tube elements are interleaved with transition radiation material consisting of polypropylene fibres. In the barrel, these straws are installed parallel to the beam line and have a length of 144 cm, in the endcaps they are 37 cm long and are arranged radially. Approximately 350 000 readout channels are utilised. The gas mixture used in the tubes is Xenon-based, although in some tubes it is replaced by a cheaper Argon-based mixture due to ongoing leaks [110]. A sensitive region $|\eta| < 2.0$ is covered and only $R$-$\phi$ information with an intrinsic accuracy of 130 µm is provided. However, a large number of hits per track, typically 36, is achieved which results in continuous tracking information. Furthermore, the TRT contributes significantly to the identification of mainly low-$p_T$ electrons by the detection of X-ray photons originating from the transition radiation.

### 2.2.3 Calorimeters

Outside the ID and the subsequent solenoid, the *Electromagnetic Calorimeter* (ECAL) is installed followed by the *Hadronic Calorimeter* (HCAL). They are utilised to detect and measure the energy deposit by electrons and photons (ECAL) and by hadrons (ECAL and HCAL). The calorimeters cover the region $|\eta| < 4.9$ while they extend symmetrically over the full $\phi$-range and radially up to 4.25 m. It is crucial that the particles (electrons, photons, hadrons) deposit all their energy in the calorimeters to prevent punch-through into the muon system to not confound the measurements there. All ATLAS calorimeters are sampling calorimeters, i.e. they are built by alternating absorbing and active detector material.

The ECAL uses liquid argon (LAr) as active material and lead as absorbing material. It is divided into a barrel part, covering $|\eta| < 1.475$ and two endcaps which cover $1.375 < |\eta| < 3.2$. A visualisation is shown in Figure 2.5. In order to get a good energy resolution, the thickness of the lead parts was optimised as a function of $\eta$. The final depth of the ECAL in the barrel is $> 22\,X_0$ (radiation lengths) and in the endcaps $> 24\,X_0$. The ECAL exhibits a fine granularity between 0.003 and 0.075 in $\Delta\eta$ and between 0.025 and 0.1 in $\Delta\phi$ in the pseudorapidity range matched to the ID ($|\eta| < 2.5$) such that the ID is ideally complemented for precision measurements of electrons and photons. The coarser granularity of $0.1 \times 0.1$ ($\Delta\eta \times \Delta\phi$) in the remaining $|\eta|$-range of the ECAL and also of the other calorimeters is sufficient for the requirements on the jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ (missing transverse energy, see Section 5.6 for further details) reconstruction. In addition, a *Presampler* is installed between the solenoid and the ECAL in the region $|\eta| < 1.8$ in order to measure and account for the energy of the particles lost up to this point. It consists of an active LAr layer of thickness 1.1 cm in the barrel and 0.5 cm in the endcaps. The total number of readout channels of the ECAL and the presampler is 170 000, the design energy resolution for electrons and photons is $\sigma_E/E = 10\,\%/\sqrt{E} \oplus 0.7\,\%$.

The HCAL, which is placed around the ECAL, is built by a *Scintillator-Tile Calorimeter* in the region $|\eta| < 1.7$. Steel is used as absorber and scintillating tiles as active material. There is a larger central barrel covering up to $|\eta| < 1.0$ and a smaller extended barrel on each side that ranges from 0.8 to 1.7 in $|\eta|$. The total thickness at $\eta = 0$ is $9.7\,\lambda$ (hadronic interaction lengths). The granularity of $0.1 \times 0.1$ ($\Delta\eta \times \Delta\phi$) in the first two layers and of $0.2 \times 0.1$ in the last layer, is coarser than in the main part of the ECAL. The $|\eta|$-range

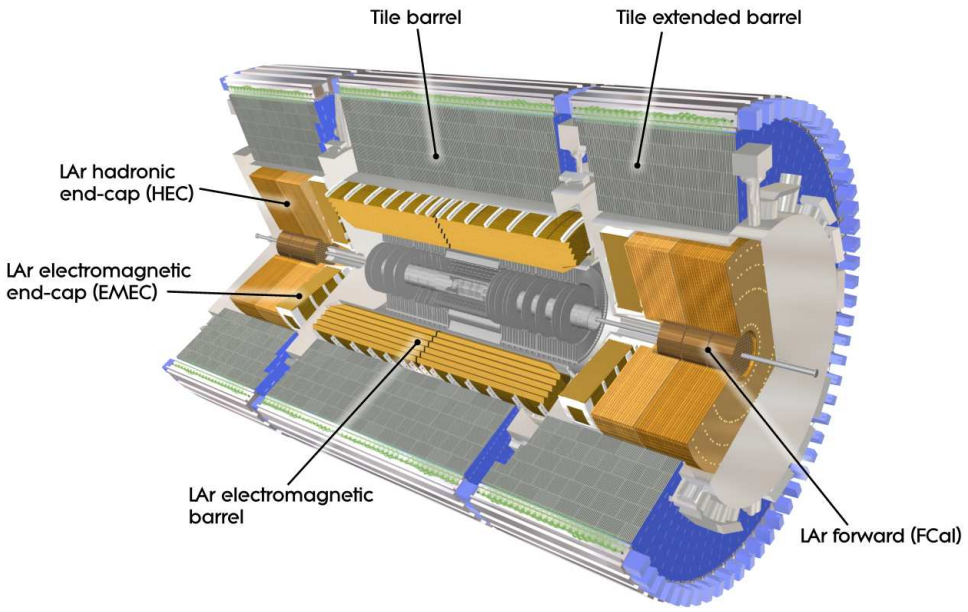



**Figure 2.5:** Cut-away view of the ATLAS calorimeters [102].

between 1.5 and 3.2 is covered by the *Hadronic Endcap Calorimeter* (HEC). It overlaps with the *Tile Calorimter* to reduce the drop in material density. Two independent wheels per endcap are located directly behind the ECAL endcaps and share the same LAr cryostats since they also utilise LAr as active material. Copper is used as absorber. It has thinner layers closer to the interaction point and thicker ones further away. The copper layers are interleaved with 8.5 mm thin LAr gaps. Below $|\eta| = 2.5$, the granularity in $\Delta\eta \times \Delta\phi$ is $0.1 \times 0.1$, above it is $0.2 \times 0.2$. The HCAL has in total 15 500 readout channels and a design energy resolution of $\sigma_E/E = 50\,\%/\sqrt{E} \oplus 3\,\%$ for the measurement of jets.

The *LAr Forward Calorimter* (FCAL) is integrated into the endcap cryostats as well and is placed within the HEC, but with a distance of 1.2 m to the ECAL front face. This reduces the neutron albedo from the ID cavity but also limits the depth of the calorimeter. Therefore, a high density design is needed which achieves a total depth of $10\,\lambda$. The FCAL consists of three modules in each endcap: the first uses copper as absorbing material, optimised for measuring electromagnetic showers. The remaining two use tungsten, predominantly deployed for hadronic showers. Approximately 3 500 readout channels are used and the design energy resolution for jets is $\sigma_E/E = 100\,\%/\sqrt{E} \oplus 10\,\%$.

### 2.2.4 Muon System

The outermost part of the ATLAS detector is built by the *Muon System* (MS). Its purpose is the detection of charged particles which traversed all previous detector layers, i.e. mainly muons, and the measurement of their transverse momenta within $|\eta| < 2.7$. This measurement utilises the magnetic deflection of muon trajectories due to large superconducting air-core toroid magnets. In the region $|\eta| < 1.4$ the magnetic field is created by the large barrel toroid consisting of eight coils arranged radially and symmetrically around the beam line. A bending power, i.e. field integral, of 1.5 to 5.5 Tm is achieved. Two smaller endcap toroids, inserted into both ends of the barrel toroid and built up of eight coils likewise, produce the magnetic field with a bending power of 1 to 7.5 Tm in the endcap regions $1.6 < |\eta| < 2.7$. The transition region in between, $1.4 < |\eta| < 1.6$, is covered by a superposition of the two fields. The resulting field is mainly orthogonal to the muon tracks.

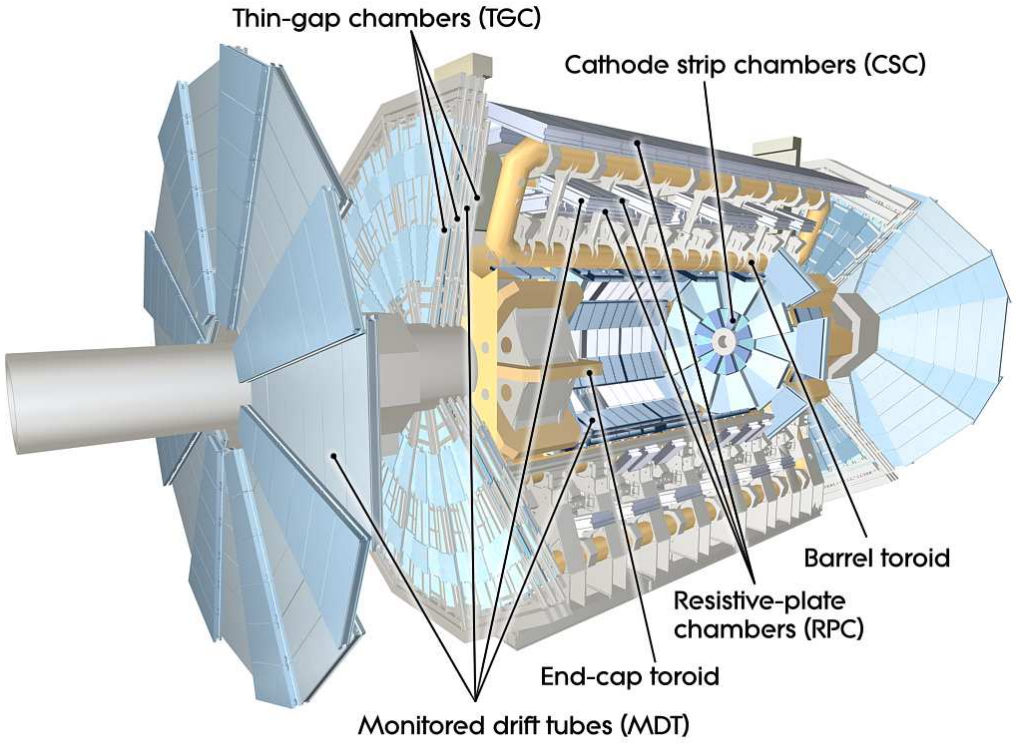



**Figure 2.6:** Cut-away view of the ATLAS muon system [102].

The MS is assembled by separate fast trigger and high-precision tracking chambers. Each of them uses two different types of chambers: *Monitored Drift Tube*s (MDTs) and *Cathode Strip Chamber*s (CSCs) are optimised for high-precision tracking while *Resistive Plate Chamber*s (RPCs) and *Thin Gap Chamber*s (TGCs) are well suited for a fast response required by the trigger within $|\eta| < 2.4$ and have a high rate capability. A cut-away view of the MS is given in Figure 2.6. In addition, the trigger chambers provide the measurement of the second coordinate ($\phi$) of the muon tracks, orthogonal to that ($\eta$) measured by the tracking chambers and also identify the bunch-crossings. In the barrel, chambers are installed in three cylindrical layers around the beamline. In the endcaps and the transition region, they are placed in three layers in large wheels perpendicular to the beamline. There is a gap in the chamber coverage at $\eta \approx 0$ to facilitate services for the inner detector elements. It prevents the measurement of straight tracks within $\pm 4.8°$, seen from the interaction point. The detector support structures (feet) create two additional gaps in the coverage at $\phi = 240°$ and $\phi = 300°$.

The MDTs consist of three to eight layers of drift tubes with a diameter of almost 30 mm. They are operated with 93 % Argon and 7 % $CO_2$ at an absolute pressure of 3 bar. The track coordinate in the principal bending direction of the magnetic field is measured with an average resolution of 35 µm per chamber. The CSCs, deployed at large pseudorapidities, are multiwire proportional chambers with cathodes segmented ino strips. They have a higher granularity compared to the MDTs as well as better rate capabilities. Hence, they are used in the innermost plane where a higher flux of particles is expected. The gas mixture exploited for the CSCs consists of 80 % Argon and 20 % $CO_2$. The wires are oriented in radial direction. One of the cathodes is oriented parallel to the wires, providing the transverse $\phi$-coordinate, the other perpendicular to the wires, providing the precision $\eta$-coordinate. Thus, both cathodes are segmented. The achieved resolution per chamber is 40 µm in the bending plane ($R$-$z$-plane) and ∼5 mm in the transverse plane.

The whole MS is equipped with approximately 1 million readout channels. The design

resolution of the transverse momentum for an MS standalone measurement is $\sigma_{p_T}/p_T = 10\,\%$ at $p_T = 1\,\text{TeV}$. In order to achieve these precision measurements, an excellent alignment of the muon chambers as well as precise knowledge of the magnetic field strengths is mandatory. Therefore, approximately 12 000 optical alignment sensors monitor the internal deformations and relative positions of the MDT chambers and ∼1 800 Hall sensors continuously gauge the magnetic field.

### 2.2.5 Trigger and Data Acquisition

During Run 2 an instantaneous peak luminosity of almost $21 \times 10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}$ was reached. The bunch-crossing rate was 40 MHz with an average number of interactions per bunch crossing of approximately 35. It is impossible to readout and save such a huge stream of data. Therefore, the *Trigger and Data Acquisition* (TDAQ) system of ATLAS [111, 112] identifies interesting characteristics in real time and stores only these for offline analysis. A sketch of this system is shown in Figure 2.7. It consists of two stages: the hardware-based

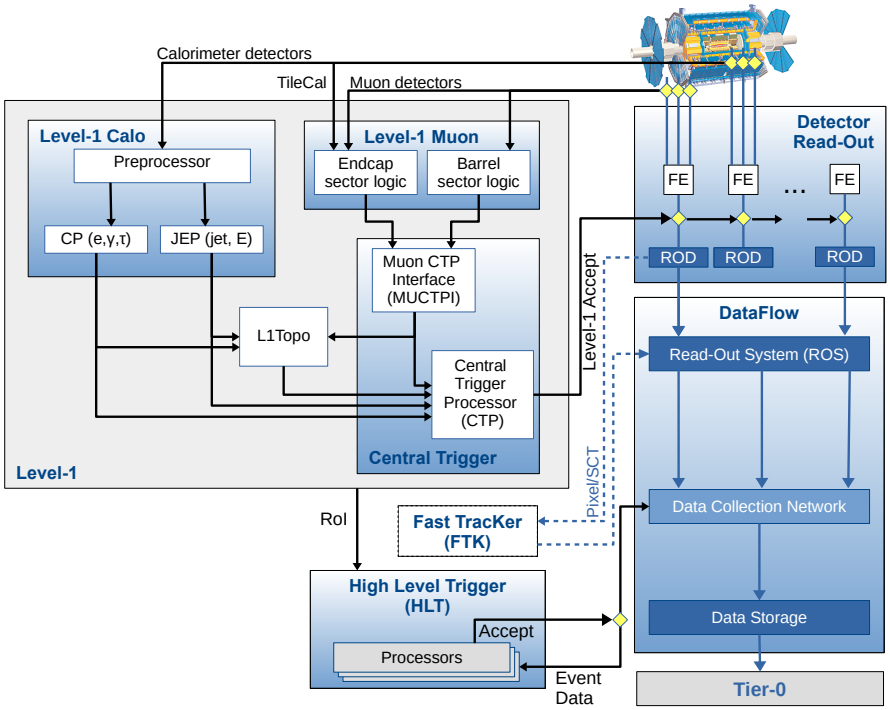



**Figure 2.7:** The Run 2 TDAQ system of ATLAS with relevant components for triggering, detector read-out and data flow [112].

*Level-1 Trigger* (L1 Trigger) and the software-based *High-Level Trigger* (HLT). The L1 Trigger utilises custom made electronics located on the detector components. For its decision whether to keep an event or not, it takes into account only a subset of information from the calorimeters and the muon system such that this decision can be made within 2.5 µs. In case of an interesting signature the full event information is read out and passed on to the HLT which can accept events at a rate of 100 kHz. In addition, the L1 Trigger selects regions of interest (RoIs) in $\eta$ and $\phi$ which are forwarded to the HLT as additional information. The HLT performs two steps: first, a fast trigger algorithm based on the RoIs from L1 Trigger is deployed for an efficient further rejection before the second, more CPU-intense step. The second step is more precise and already close to the offline reconstruction. A dedicated computing farm with ∼40 000 cores is available to fulfil this task. Finally, an output rate of 1.2 kHz is achieved which corresponds to 1.2 GB/s that are saved to permanent storage.

## 2.3 Luminosity Measurement

Precise knowledge of the integrated luminosity recorded by the ATLAS detector is crucial for performing high precision cross section measurements. The measurement of the luminosity is described in [113, 114]. In the following a brief summary of the measurement is provided. The main luminosity detector used in Run 2 was LUCID-2 [115]. It is a Cherenkov detector consisting of 16 *Photo Multiplier Tube*s (PMTs) in each forward arm of the ATLAS detector at a distance of $\pm 17\,\text{m}$ from the interaction point. Cherenkov light from energetic particles is induced by the quartz windows of the PMTs and measured by the PMTs. The raw hit counts are converted into a visible number of interactions per bunch crossing $\mu_{\text{vis}}$ which is proportional to the instantaneous luminosity. The instantaneous luminosity per bunch crossing $\mathcal{L}_b$ is given by

$$\mathcal{L}_b = \frac{\mu_{\text{vis}} f_{\text{rev}}}{\sigma_{\text{vis}}}\,, \tag{2.7}$$

with $f_{\text{rev}}$ being the revolution frequency (11 246 Hz for protons) and $\sigma_{\text{vis}} = \varepsilon\sigma_{\text{inel}}$ the visible cross section. Here, $\varepsilon$ is the efficiency of the respective luminosity detector and $\sigma_{\text{inel}}$ the inelastic cross-section of $pp$ collisions. Therefore, also $\mu_{\text{vis}} = \varepsilon\mu_{\text{inel}}$ holds.

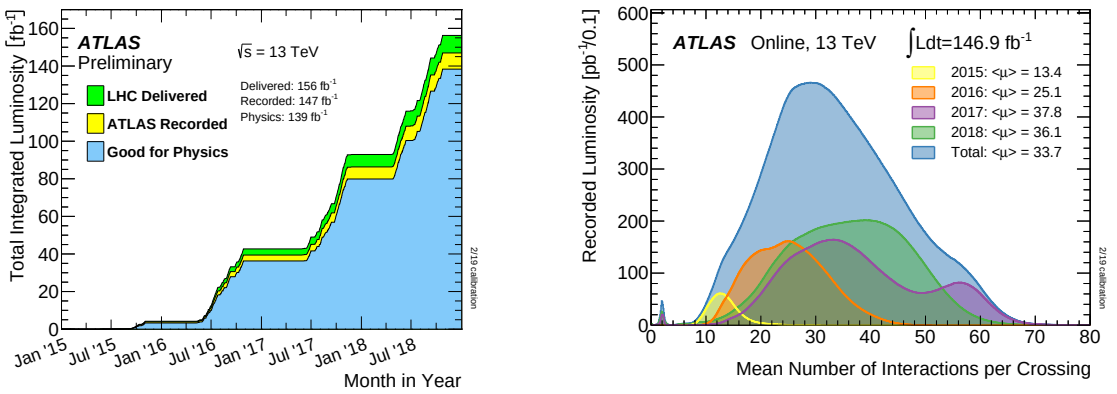
The total instanteaneous luminosity is obtained by summing $\mathcal{L}_b$ over all colliding bunch pairs. The revolution frequency $f_{\text{rev}}$ is given by the LHC, $\mu_{\text{vis}}$ can be directly measured during data taking while $\sigma_{\text{vis}}$ is obtained from dedicated runs using the *van der Meer* [116] (vdM) method. During these runs, so-called vdM scans are performed in a low-luminosity environment to reduce uncertainties. The absolute calibration of the LUCID-2 luminometer at the conditions of the vdM runs needs to be transferred to physics conditions at high luminosity. LUCID-2 shows significant non-linearities in the extrapolation from low to high $\mu$ which are corrected using the so-called *track-counting* luminosity measurement. The number of reconstructed tracks in the *Inner Detector* is assumed to be proportional to the luminosity over the full luminosity range given a dedicated track selection. The uncertainty in this so-called *calibration transfer* procedure is the largest single uncertainty component for the final integrated luminosity measurement.

## 2.4 Data Taking

The LHC started delivering collisions for data taking for the first time in the year 2010 at a center-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$. Until the end of 2011, $5.08\,\text{fb}^{-1}$ of $pp$-collision data were recorded by the ATLAS detector. In 2012, the center-of-mass energy was increased to $\sqrt{s} = 8\,\text{TeV}$ and $22.8\,\text{fb}^{-1}$ of data were provided by the LHC. This period is referred to as Run 1 and was followed by the *Long Shutdown 1* which lasted until spring 2015 and was used to perform maintenance and upgrade work of the LHC and of the detectors.

Data taking resumed in spring 2015 at $\sqrt{s} = 13\,\text{TeV}$ and lasted until the end of 2018. The period from 2015 to 2018 is referred to as Run 2 and is the basis for the analysis presented here. $156\,\text{fb}^{-1}$ of data from $pp$-collisions were delivered by the LHC. Figure 2.8(a) shows the cumulative integrated luminosity versus time during Run 2 data taking. Because of ramp-up phases of the sub-detectors or data acquisition issues, the actual recorded amount of data is slightly reduced: $147\,\text{fb}^{-1}$ were recorded by the ATLAS detector. The data that is declared *good for physics* and is utilised in this analysis is further reduced to $138.4\,\text{fb}^{-1}$ to ensure sufficient data quality. This includes proper trigger operation, fully functional sub-detectors and no ongoing luminosity or other calibration activities.

The data were recorded with 2556 bunches of protons in each beam in the LHC during Run 2, where bunches consisted of up to $1.25 \times 10^{11}$ protons and were spaced with distances of 25 ns. Instantaneous luminosities of up to $2.1 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$ were reached (see Section 2.1). High luminosities are crucial to produce a large amount of data, i.e.




**(a)** Cumulative integrated luminosity versus time during Run 2 delivered by the LHC (green), recorded by the ATLAS detector (yellow) and *good for physics*, i.e. with sufficient data quality(blue) [117].

**(b)** The luminosity-weighted distribution of the mean number of interactions per crossing for the different years of Run 2 and combined. All data recorded by ATLAS detector during stable beams is included [117].

**Figure 2.8:** Cumulative integrated luminosity versus time during Run 2 (a) and the luminosity-weighted distribution of the mean number of interactions per crossing (b).

a high integrated luminosity. However, high luminosities come with a large number of simultaneous interactions per bunch crossing and therefore with an increased activity in the detector per event. The integrated-luminosity weighted distribution of the mean number of interactions per bunch crossing or Run 2 is shown in Figure 2.8(b). The increase from 2015 to 2017/18 is clearly visible.

According to schedule, the LHC started Run 3 on July 5, 2022 with stable beam collisions at a center-of-mass energy of $\sqrt{s} = 13.6\,\text{TeV}$, after the *Long Shutdown 2* which was used again for maintenance and upgrade work. Run 3 is expected to last until 2025 and to provide $\sim 250\,\text{fb}^{-1}$ of data to the ATLAS and CMS experiments. Another 3-year long shutdown is planned from 2026 onwards to prepare for the *High Luminosity LHC* (HL-LHC) starting in 2029. An integrated luminosity of $\sim 3\,000\,\text{fb}^{-1}$ 12 years after the start is targeted at $\sqrt{s} = 14\,\text{TeV}$.

# Chapter 3

# Analysis Strategy

The analysis presented in this thesis searches for *lepton-flavour violating* decays of the Higgs boson using the *Symmetry Method* [21] for background estimation. The analysis comprises the searches for the decays $H \to e\tau$ and $H \to \mu\tau$ and aims at either observing evidence for these decays or to improve the direct limits on their branching ratios. The decay $H \to e\mu$ is not considered due to a stringent indirect limit on its branching ratio. An overview of the latest indirect and direct upper limits on the respective branching ratios is given in Section 1.4.2.

This analysis focuses on the leptonic decays of the $\tau$-lepton. The statistical combination with an analysis targeting the hadronic decays of the $\tau$-lepton is performed and presented at the end of this thesis. Only different flavour final states are considered due to the overwhelmingly large Drell-Yan background for same-flavour final states. This means for $H \to e\tau$ only $\tau$-leptons decaying into a muon and two neutrinos and for $H \to \mu\tau$ only $\tau$-leptons decaying into an electron and two neutrinos are taken into account. Hence, the final state at leading order in QCD consists of an electron, a muon and two neutrinos plus additional jets in the case of production via VBF for both signals. More details about the signal processes are given in Section 4.2.1. A sketch of the decay is given in Figure 3.1. In

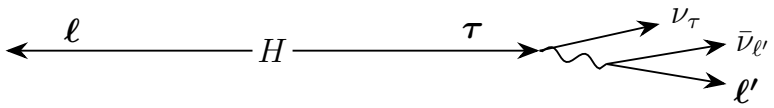



**Figure 3.1:** Sketch of a Higgs boson LFV decay in the Higgs boson rest frame.

this sketch and throughout the thesis, $\ell$ refers to the light leptons, electron or muon, unless stated otherwise. Irreducible background contributions, i.e. processes with the same final state objects, arise from $Z \to \tau\tau$ decays, diboson production (mainly $WW$), $H \to \tau\tau$ and $H \to WW$ decays. Additional contributions are from $t\bar{t}$-production and from other objects mis-identified as electrons or muons (*fakes*). These background processes are discussed in more detail in Section 4.2.2.

The *Basic Event Selection*, detailed in Section 7.1.2, requires exactly one electron and one muon in the final state while a veto is applied on hadronically decaying $\tau$-leptons and on jets originating from $b$-flavour hadrons to suppress contributions from $t\bar{t}$-production. Kinematic requirements, for example on the transverse momentum, $p_{\mathrm{T}}$, and the pseudorapidity, $\eta$, are defined symmetrically for electrons and muons. Different requirements for electrons and muons would be possible from a reconstruction point of view and would lead to a higher acceptance of events but violates assumptions made within the *Symmetry Method* illustrated below. In order to search for the two LFV signals, $H \to e\tau_\mu$ and $H \to \mu\tau_e$, suitable and mutually exclusive sub-datasets for each of the searches are defined which aim

at containing the majority of the respective signal events. The full dataset is split into two datasets, the $e\tau$- and the $\mu\tau$-dataset, depending on the transverse momenta of the two leptons in an approximate Higgs-boson rest frame:

**$e\tau$-dataset:** $p_T^e > p_T^\mu$ in an approximate Higgs-boson rest frame,

**$\mu\tau$-dataset:** $p_T^\mu > p_T^e$ in an approximate Higgs-boson rest frame.

For signal events it is assumed that the light lepton originating from the decay of the $\tau$-lepton has a lower transverse momentum in the approximate Higgs-boson rest frame compared to the light lepton originating directly from the Higgs boson (due to the additional neutrinos in the $\tau$-lepton decay). Therefore, the $H \to e\tau_\mu$-events are expected to be in the $e\tau$-dataset while the $H \to \mu\tau_e$-events are expected to be in the $\mu\tau$-dataset. The definition of the approximate Higgs-boson rest frame as well as further information on the procedure can be found in Section 7.1.3.

After the *Basic Selection*, the events are split into two signal regions to be able to exploit the distinct topology of the VBF production of the Higgs boson which exhibits two additional jets in the final state. These jets typically show a large separation in pseudorapidity and a large invariant mass (Section 8.1). This is done simultaneously for both the $e\tau$- and the $\mu\tau$-dataset. In order to reduce the statistical uncertainties, the selection of events by cuts on single variables is limited to the *Basic Selection* and the split into the two signal regions. Further separation of signal and background is achieved with the help of a multivariate analysis technique. *Neural Network*s (NNs) are optimised, trained and applied in each of the signal regions independently (Section 8.2). The NN output distributions in each of the two regions serve as final discriminants in the statistical analysis (Chapter 10).

All background contributions apart from *fakes* are estimated with the data-driven *Symmetry Method*, which relies on two assumptions:

1. Standard model processes are symmetric with respect to an exchange of electrons with muons and vice versa, resulting in approximately the same kinematic distributions for both lepton flavours.

2. The two LFV Higgs-boson decays $H \to e\tau$ and $H \to \mu\tau$, on the contrary, break this symmetry as long as their branching ratios differ.

Owing to the high energies present at the LHC, the different masses of electrons and muons can be neglected in the first assumption. The different Yukawa-couplings of electrons and muons can be neglected since different flavour final states are required in the analysis and both branching ratios are expected to be small: $2.176 \times 10^{-2}\,\%$ [83] for $H \to \mu\mu$ and $5.1 \times 10^{-7}\,\%$ [118] for $H \to ee$. The second assumption is only valid as long as $H \to e\tau$ and $H \to \mu\tau$ do not have the same branching ratio which is well motivated as detailed in Section 1.4.2 when neglecting interferences of other new LFV sources in $\mu \to e\gamma$. As a consequence of the first assumption, both sub-datasets introduced above are expected to have the same size and exhibit the same kinematic distributions for SM processes (background). This means that e.g. the $\mu\tau$-dataset can serve as an estimate of the $e\tau$-dataset. When in addition assuming a $H \to e\tau_\mu$ signal with a branching ratio of 10 % and a non-existent $H \to \mu\tau_e$ signal, the $e\tau$-dataset is enriched with signal while the $\mu\tau$-dataset only consists of background. This is depicted in Figure 3.2 which shows the reconstructed Higgs-boson mass in the collinear mass approximation (Eq. 8.3 in Section 8.2.3). The peak at around 90 GeV on the left corresponds to the $Z$-boson mass. Both datasets are in agreement in the sideband regions, i.e. outside the Higgs-boson mass range, while the bump around the Higgs-boson mass of 125 GeV is only present in the $e\tau$-dataset. Hence, the $\mu\tau$-dataset can serve as background estimate of the $e\tau$-dataset and vice versa for the presence of a $H \to \mu\tau$

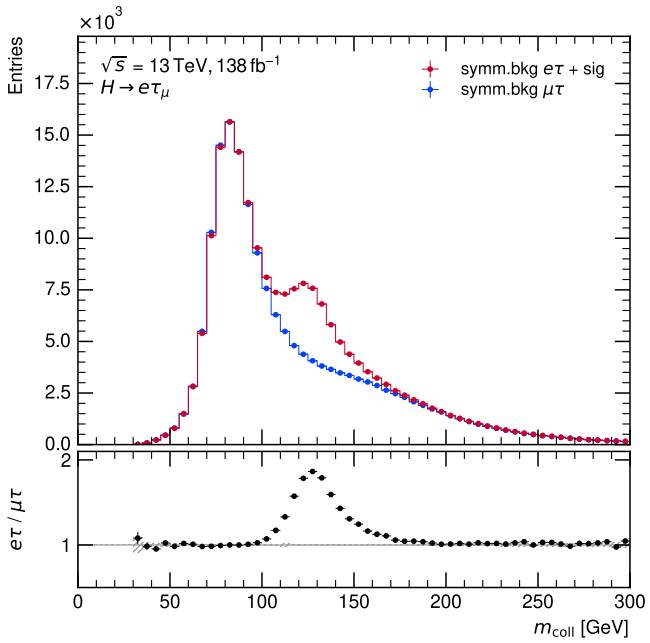



**Figure 3.2:** Reconstructed Higgs-boson mass. Red indicates the $e\tau$-dataset enriched with $H \to e\tau_\mu$ signal (branching ratio 10 %), blue indicates the background-only $\mu\tau$-dataset. The distributions are obtained from simulated events after the *Basic Selection* detailed in Section 7.1.2. "Symm.bkg" in the legend means "$e/\mu$-symmetric background". The lower panel shows the ratio $e\tau/\mu\tau$. Only simulated events are utilised here.

signal and the absence of a $H \to e\tau$ signal. The background contribution estimated in this way will be denoted as *$e/\mu$-symmetric background*. It comprises all background processes apart from *fakes*. Evidently, the *Symmetry Method* is only sensitive to a difference in branching ratios of the two LFV Higgs-boson decays $H \to e\tau$ and $H \to \mu\tau$. Each of the sub-datasets can be used as $e/\mu$-symmetric background estimate or as data. Hence, the difference in branching ratios can be determined twice (with opposite signs) by two correlated measurements; once in the search for $H \to e\tau$ using the $e\tau$-dataset as data and the $\mu\tau$-dataset as $e/\mu$-symmetric background estimate and once in the search for $H \to \mu\tau$ using the $\mu\tau$-dataset as data and the $e\tau$-dataset as $e/\mu$-symmetric background estimate. Under the assumption that the branching ratio of $H \to e\tau$ is zero, the measurement in the search for $H \to \mu\tau$ can be interpreted as the measurement of the branching ratio of $H \to \mu\tau$. On the contrary, assuming that the branching ratio of $H \to \mu\tau$ is zero allows for interpreting the measurement in the search for $H \to e\tau$ as the measurement of the branching ratio of $H \to e\tau$. The notable advantage of this method is its reduced dependence on Monte-Carlo simulations and the inevitable uncertainties associated with them. The leading uncertainties in the *Symmetry Method* stem from the available amount of data statistics and therefore the sensitivity of the method improves with an increasing amount of collected data.

However, electrons and muons manifest in different ways in the ATLAS detector (Chapter 5) leading to differing detection efficiencies and energy resolutions which violates the first assumption of the *Symmetry Method*. The effect of the energy resolution is negligible in the distribution of the reconstructed Higgs-boson mass. Thus, this effect is neglected. The detection efficiencies of electrons and muons are measured and correction factors are applied to the $e/\mu$-symmetric background contribution (Section 7.2.1). In addition, the probability of other objects being mis-identified as electrons or muons (the *fake rate*) differs between the two lepton flavours which again violates the symmetry assumption. Hence, this

background contribution must be estimated separately. The estimation is also primarily based on a data-driven method (Section 7.2.2).

In the statistical analysis where the final results are derived, the background estimation is performed as detailed above. During the development and optimisation of the analysis also simulated events are used to estimate the $e/\mu$-symmetric background contributions for cross checking. Furthermore, simulated events are used partially in the estimation of the *fakes*, to determine the detection efficiencies and for the prediction of the signal processes.

While developing the analysis, sensitive bins in the distributions of observables are *blinded* such that optimisation processes of the analysis are not biased. In this context *blinding* means that the data value in a bin is not shown if the signal-over-background ratio exceeds 5 % when assuming a branching ratio of 1 % for signal.

There are three publications regarding the searches for the LFV decays $H \to e\tau / H \to \mu\tau$ by the ATLAS collaboration prior to this thesis: two Run 1 and an early Run 2 (2015 + 2016 data) publication. The Run 1 publications [119, 120] estimate all background processes apart from fakes based on templates from Monte-Carlo simulations in the hadronic decay channel of the $\tau$-lepton while the *Symmetry Method* is used for the leptonic decay channel. The statistical model used for the analysis based on the *Symmetry Method* is less sophisticated compared to the one used in this thesis. The early Run 2 publication [19], on the other hand, is solely based on templates from Monte-Carlo simulations for background estimation (except estimation of fakes) for both the hadronic and the leptonic decay channels of the $\tau$-lepton.

The *Symmetry Method* can also be utilised in the hadronic decay channel of the $\tau$-lepton, but was not optimised and ready for the full Run 2 publication of which the analysis presented here is part of. However, a proof of concept is published in a master's thesis [121].

# Chapter 4

# Physics Processes and their Simulation

The relevant signal and background processes in the search for $H \to e\tau$ and $H \to \mu\tau$ were briefly described in Chapter 3. This chapter discusses them in more detail and gives an overview of their simulation. The simulations comprise the theoretical prediction of the cross sections for the processes and of their kinematic distributions described by the four-vectors of the final state objects in each event. Section 4.1 introduces the methods to simulate the processes while Section 4.2 discusses the signatures and characteristics of the signal and background processes and provides further information on their simulation.

## 4.1 Generation of Simulated Events

Chapter 3 explicated the primarily data-driven ansatz to estimate the majority of the contributing background processes. Nevertheless, simulated events representing the individual background processes are an important ingredient to accomplish multiple steps in the analysis: on one hand they are used to validate the method in various occasions. On the other hand they serve as inputs to the neural network training which targets the separation of signal and background and as inputs to the derivation of corrections applied to the data-driven methods. In any case the estimation of the signal processes can only be obtained by simulation.

The different components needed for the simulations are discussed in the following, inspired by [122–129].

**Parton Distribution Functions and the Hard Scattering**
The number of expected events of a certain process, e.g. $pp \to X$, for a given integrated luminosity $\mathcal{L}_{\text{int}}$ is obtained by multiplying $\mathcal{L}_{\text{int}}$ with the respective cross section $\sigma$ as given in Eq. 2.2. Thanks to the *factorisation theorems* [130], the cross section $\sigma_{p_1p_2\to X}$ of protons $p_1$ and $p_2$ can be split into two parts: on one hand there is the partonic cross section $\sigma_{ij\to X}$ of the hard-scattering process $ij \to X$ with partons $i$ and $j$ originating from $p_1$ and $p_2$, respectively, which takes place at high energies (at a short time scale/distance). And on the other hand there are the lower-energetic dynamics within the proton described by the *parton distribution functions* (pdf) $f_i, f_j$ which express the probability to find parton type $i, j$ with a proton momentum fraction $x_1, x_2$ inside the proton probed at the energy scale $\mu_F$. The factorisation scale $\mu_F$ separates both regimes, i.e. the long and short distance physics and is typically set to the energy $Q$ of the hard-scattering process. Then, the cross

section $\sigma_{p_1p_2\to X}$ can be written as:

$$\sigma_{p_1p_2\to X} = \sum_{i,j} \int_0^1 \int_0^1 dx_1 dx_2 f_i\left(x_1, \mu_F\right) f_j\left(x_2, \mu_F\right) \sigma_{ij\to X}\left(x_1, x_2, \mu_R, \mu_F, \alpha_s(\mu_R)\right) . \quad (4.1)$$

The sum goes over all possible parton flavours. The momentum fractions carried by the partons inside a proton must add up to 1 to fulfil momentum conservation.

The dependence of $f_i$ on $x$ is based on non-perturbative QCD dynamics because the strong coupling strength $\alpha_s$ gets too large for low energies which does not allow for an expansion in $\alpha_s$. Thus, the dependence of $f_i$ on $x$ must be extracted from data. Typically, data from deep-inelastic scattering collected e.g. at the HERA collider but also data from the Tevatron and now the LHC [131], is utilised. The $f_i$ are parametrised for example as $f_i(x, Q_0) \propto A_i \cdot x^{-\alpha_i} \cdot (1-x)^{\beta_i}$ and the parameters are extracted from a global fit to data at different ranges of $x$ and different reference scales $Q_0$. Then, the $f_i$ can be evolved to any value of $Q$ by means of the DGLAP[1] evolution equations [132–134]. They describe the dependency of the pdfs on $Q^2$ and depend on the splitting functions that describe (at leading order) the probability for a quark to split in a quark and a gluon or for a gluon to split in two quarks or two gluons. The splitting functions are calculated perturbatively and are computed up to $\mathcal{O}(\alpha_s^4)$.

The distributions of the different $f_i$ multiplied by $x$ obtained by the NNPDF Collaboration is shown in Figure 4.1 for two different values of $Q^2$, $10\,\mathrm{GeV}^2$ and $10^4\,\mathrm{GeV}^2$. The two valence quark types ($u$, $d$) are peaking at larger values of $x$ while the sea partons increase with smaller values of $x$ with gluons having the highest probability for small $x$. At $Q^2 = 10^4\,\mathrm{GeV}^2$ (close to the Higgs-boson mass) compared to $Q^2 = 10\,\mathrm{GeV}^2$, the shape of the valence quarks stays approximatey the same while the peak is decreased and the sea partons are substantially increased for low $x$ where the gluons are even more dominating.

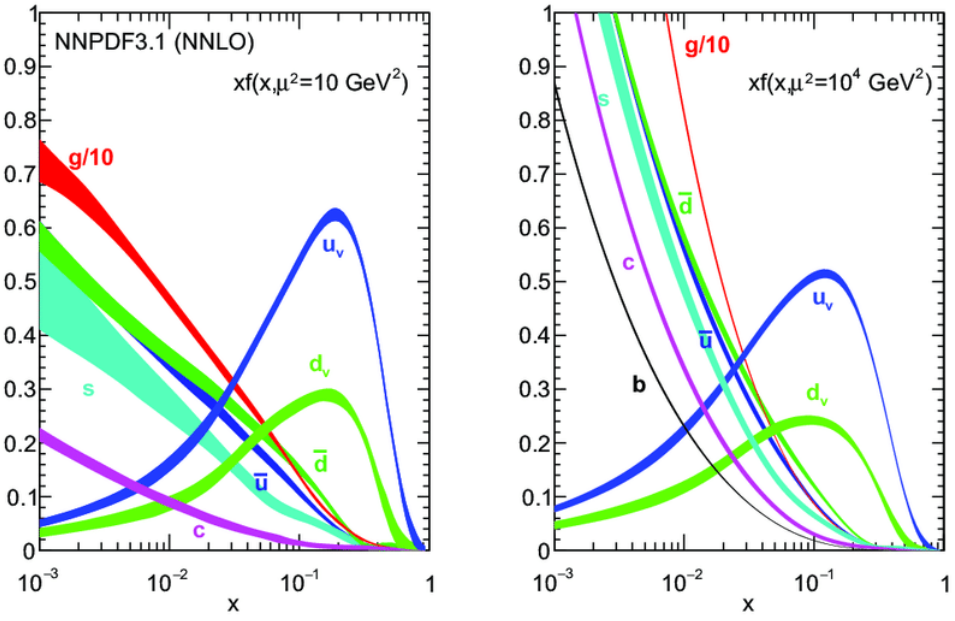



**Figure 4.1:** The pdfs (multiplied by $x$) from a global fit by the NNPDF Collaboration for different factorisation scales of $10\,\mathrm{GeV}^2$ (left) and $10^4\,\mathrm{GeV}^2$ (right). They are shown as a function of the momentum fraction $x$. The width of the line indicates the corresponding uncertainties [135].

[1] the DGLAP equations were formulated independently by Yuri **D**okshitzer, by Wladimir Naumowitsch **G**ribow and Lew Nikolajewitsch **L**ipatow and by Guido **A**ltarelli und Giorgio **P**arisi

The partonic cross section $\sigma_{ij\to X}$ of the hard-scattering process is obtained from integrating the squared matrix element over the respective phase space $d\Omega$. The energies involved are large, i.e. the time scale is short, and hence, due to *asymptotic freedom* [136, 137], the partons can be treated as free particles and the cross section can be calculated with perturbation theory by expansion in orders of ${\alpha_s}^2$. The matrix element, which represents the transition amplitude from the initial state $ij$ to the final state $X$, is calculated up to some order $n$ in $\alpha_s$ following the Feynman rules [25], i.e. taking into account all Feynman diagrams contributing to this order. Higher orders in $\alpha_s$ involve either additional particles in the final state (real emissions) or virtual loop contributions. Loop corrections in higher-order Feynman diagrams can lead to divergences when integrating over the infinite loop momenta. These divergences can be regularised which necessitates the introduction of an (unphysical) *renormalisation scale* $\mu_R$. Absorbing the renormalisation scale into the coupling strength $\alpha_s$ results in $\alpha_s$ becoming dependent on $\mu_R$: $\alpha_s(\mu_R)$. Typically, $\mu_R$ is chosen to be equal to the energy scale $Q$ of the hard process. The value of the cross section obtained by these calculations evidently depends on the choice of $\mu_R$ (and $\mu_F$). However, if all orders were included in the perturbative calculations, the cross section would not depend on their choice. Hence, variations of $\mu_R$ and $\mu_F$ can be utilised to estimate uncertainties due to missing higher orders.

In order to obtain simulated events, four-vectors of the final state particles are generated according to the differential cross section $d\sigma_{p_1p_2\to X}/d\Omega$ using Monte-Carlo (MC) techniques. The phase-space information of the event is contained in $d\Omega$. So-called parton-level event generators are utilised to fulfil this task. Thereby, they must correctly sum over the different colour states, helicities and sub-processes. Depending on the MC technique, so-called *event weights* can be assigned to the individual events such that they correctly emulate the distribution of the differential cross section.

However, there are more components to consider in order to obtain a complete picture of a *pp*-collision event as visualised in Figure 4.2. So far, the part illustrated in blue, corresponding to the initial state partons and hence to the description of the parton distribution functions, was discussed as well as the part of the red part which indicates the hard scattering.

**Parton Shower and Hadronisation**

The matrix element of the hard-scattering process is typically only evaluated up to a finite number of orders in $\alpha_s$. In order to include the effects of all higher orders at leading log accuracy, *parton shower* (PS) algorithms are deployed. The partons, initial state as well as final state partons, radiate gluons which subsequently leads to further gluon radiation and parton multiplication [125] and thus a shower evolves. Typically, the radiation happens at small angles leading to a narrow shower in the direction of the original parton. The energy of the partons reduces from splitting to splitting with the probablity of the $1 \to 2$ splitting being described by the splitting functions already mentioned above. Starting from the scale of the hard interaction, the evolution proceeds down to some low-energy scale of around 1 GeV at which the processes of non-perturbative confinement occur. *Hadronisation models* are used to describe these processes of confinement where the partons form colour-neutral hadrons. These non-perturbative models must be tuned to external data. There are two main classes of hadronisation models: the class of string models where the Lund model [139, 140] is common and the class of cluster models [141, 142].

**Higher-Order Corrections and Matching and Merging**

The parton shower represents approximate higher-order corrections due to parton emissions. If the partonic cross section is calculated and simulated at higher orders, it must be

[2] in addition, expansion in $\alpha_w$ is possible to obtain electroweak corrections

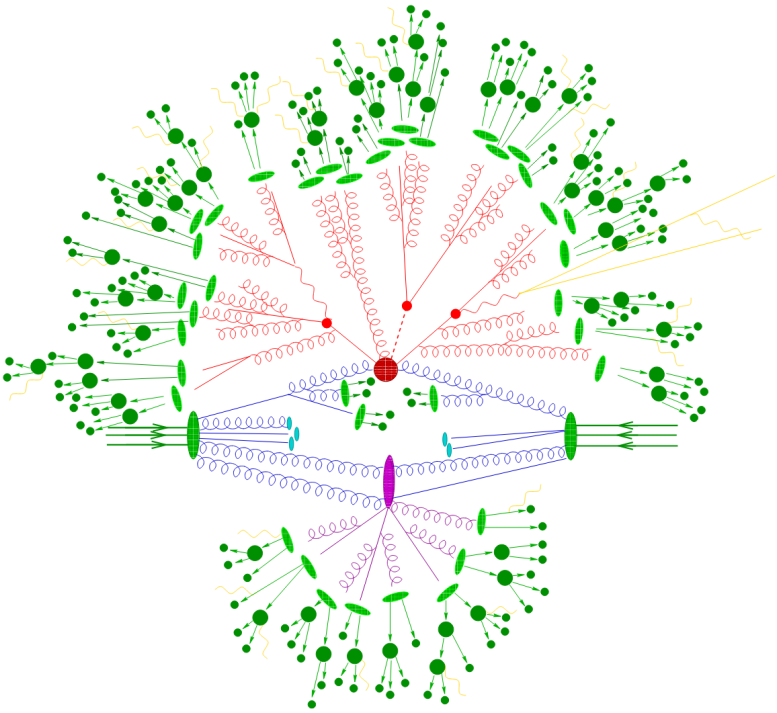


**Figure 4.2:** Sketch of all components, apart from pile-up, relevant for simulating a *pp*-collision event [138]. The initial state partons (constituents of the colliding protons) are depicted in blue. Some of them contribute to the hard interaction (big dark red blob). Subsequent decays (small red blobs), outgoing partons and additional hard QCD radiation is illustrated in red as well. The hadronisation of the final state partons is indicated by the light green ellipses, the subsequent decay of the hadrons is indicated by the dark green blobs and arrows. The initial state partons not contributing to the hard scattering form the underlying event (purple). Yellow illustrates photon radiation from electromagnetic charged objects which can occur at any moment.

ensured that no double-counting between the hard-scatter representation and the PS occurs. Furthermore, the former is better suited for the description of emissions at large angles and high energies while the latter describes soft and collinear emissions. Common methods to address this issue are based on *matching* or *merging*. The CKKW [143] and the MLM [144] method are examples for the merging approach. Here, a threshold scale is introduced below which the jets are described by the parton shower and above which they are described by the matrix element calculation. As a result of the matching and merging procedures some events can obtain negative event weights such that they are subtracted when illustrated in a histogram. NLO simulations of the hard process can also yield negative event weights in certain phase space regions due to destructive interference.

### Decay

The decay of hadrons produced in the hadronisation process (depicted in dark green in Figure 4.2) or of other unstable particles, e.g. $\tau$-leptons, into particles stable within the detector reach is simulated either by dedicated tools or by the same generator that is also used for the parton shower and hadronisation. The decay products of the unstable hadrons originating from one parton develop into a so-called jet which can be reconstructed from its signatures in the detector as explained in Section 5.4.

### Underlying Event

Furthermore, the behaviour of the remnants of the colliding protons, the *underlying event* (UE) (shown in purple in Figure 4.2), needs to be considered. These partons undergo soft

but also semi-hard interactions and remain linked to the hard scatter partons due to their colour charge. Hence, they can be distributed within the active detector regions and do not necessarily disappear undetected in the beam line. There are different models to describe the UE, of which models based on multiple parton interactions (MPI) are widely used. The simulation includes perturbatively calculated cross sections augmented by procedures similar to the parton shower and the modelling of non-perturbative effects. Although the partons in the UE are usually not reconstructed as separate jets in the detector, they can contribute to jets from the hard interaction and bias their measurement.

### Pile-Up

Another important component in the simulation of $pp$-collisions that is not depicted in Figure 4.2 is *pile-up* (PU) [145]. It comprises the interactions of other protons apart from the hard scatter event in the same proton bunch crossing (*in-time pile-up*) and in adjacent bunch crossings (*out-of-time pile-up*) leading to several interaction vertices along the beam line. The average number of interactions per bunch crossing $\langle\mu\rangle$, a measure for the size of pile-up, ranged from 13.4 during data taking in 2015 to 37.8 in 2017 as shown in Figure 4.3. Reconstructed jets stemming from pile-up can be suppressed by vertex reconstruction but nevertheless make the event reconstruction more complicated and enhance the uncertainties. Pile-up events are simulated by single soft QCD interactions that are overlaid on simulated events for the hard process of interest. The number of overlaid pile-up vertices are chosen according to the pile-up profile measured in data and shown in Figure 4.3 separate for the different data taking years of Run 2.

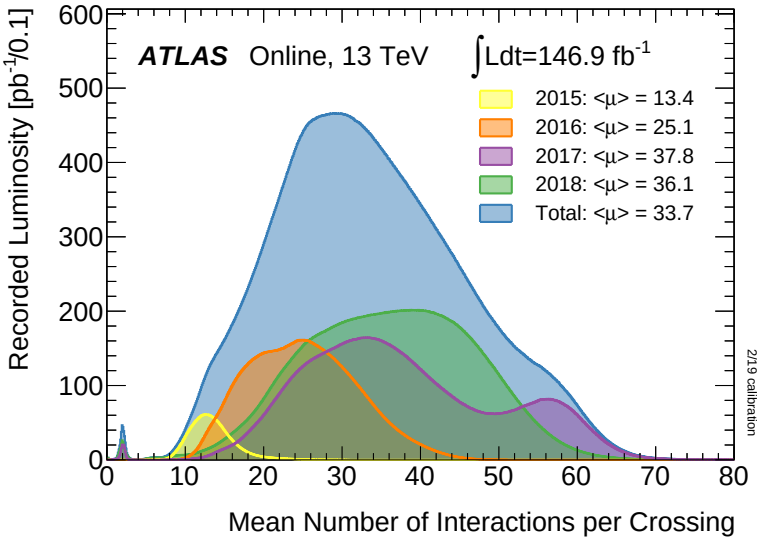



**Figure 4.3:** The luminosity-weighted distribution of the mean number of interactions per bunch crossing for the different years of Run 2 and combined. All data recorded by the ATLAS detector during stable beams is included [117].

### Detector Simulation

After the full event generation, each event is passed on to a sophisticated simulation of the detector response to the finale-state particles based on GEANT4 [146] which includes a full description of the ATLAS detector. Thus, all subsequent steps such as the reconstruction of the particle objects can be performed in the same way for data and simulation from here on since both are stored in the same data format.

## 4.2 Signal and Background Processes

The following sections discuss the physics processes relevant to the search for LFV Higgs-boson decays and their simulation. The signature and properties of the signal processes are given in Section 4.2.1 while the background contributions are presented separately for each considered process in Section 4.2.2.

### 4.2.1 Signal Processes

The two signal processes of interest are the *lepton-flavour violating* decays $H \to e\tau$ and $H \to \mu\tau$. The main processes to produce a Higgs boson at the LHC were introduced in Section 1.2.1. However, only the three dominant ones, *gluon-gluon fusion* (ggF), *vector-boson fusion* (VBF) and *Higgs strahlung* (VH), are considered in this analysis. The production in association with two top quarks is neglected due to its low cross section and its suppression as a result of the applied event selection (veto on $b$-tagged jets), detailed in Section 7.1. The Feynman diagram at leading order for the most abundant production process ggF together with the Higgs-boson LFV decay is given in Figure 4.4. In this analysis, only

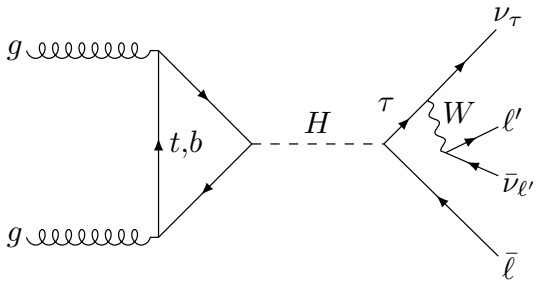



**Figure 4.4:** The production of a Higgs boson via *gluon-gluon fusion* together with a subsequent *lepton-flavour violating* decay into a different flavour light-lepton final state.

leptonic decays of the $\tau$-lepton are considered[3]. In addition, as described in Chapter 3, only different-flavour final states are considered to suppress background contributions from Drell-Yan production. Hence, there is always an electron, a muon and two neutrinos in the final state; independent of which of the two LFV decays is investigated. Owing to conservation of electric charge and the Higgs boson being neutral, the two leptons in the final state have opposite-sign electric charges. The neutrinos are weakly interacting particles and escape the detector without leaving a trace which results in the fact that they cannot be reconstructed independently from each other. The transverse component of the system of the two neutrinos is approximated by the measurement of the *missing transverse energy* $E_{\mathrm{T}}^{\mathrm{miss}}$. More details on $E_{\mathrm{T}}^{\mathrm{miss}}$ can be found in Section 5.6.

To be able to distinguish whether an event originates from a $H \to e\tau$ or from a $H \to \mu\tau$ decay, a lepton assignment procedure is applied where the leptons are labelled $\ell_H$ (originating directly from the Higgs boson) and $\ell_\tau$ (originating from the $\tau$-lepton) accordingly. The procedure is mainly based on the fact that the lepton $\ell_\tau$ from the $\tau$-decay is expected to have a lower momentum in the Higgs boson rest-frame compared to lepton $\ell_H$ from the $H$-decay due to the additional neutrinos in the $\tau$-decay. Further details are given in Section 7.1.3.

The decay products of the Higgs boson are boosted as a result of their mass difference to the Higgs boson, leading to likewise strongly boosted decay products of the $\tau$-lepton. This implies that the angular difference $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ of $\ell_\tau$ and $E_{\mathrm{T}}^{\mathrm{miss}}$ is small. Furthermore, in the case of ggF production of the Higgs boson without any additional jets in the final

[3]The hadronic decays are considered in a separate analysis (not performed by me) since both decay modes come with their own challenges and hence each requires customised analysis techniques.

state as depicted in Figure 4.4, the two leptons are expected to be almost back-to-back and hence have a large $\Delta\phi(\ell_\tau, \ell_H)$ separation (and therefore also a large $\Delta\phi(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$) since the Higgs boson is produced almost at rest in the transverse plane. However, if there are additional jets in the final state corresponding to higher-order diagrams, they recoil against the Higgs boson thereby providing it with a boost in the transverse plane.

In the production via VBF there are two jets already at leading order that show a large separation in pseudorapidity $\eta$ and a large invariant mass $m_{jj}$. This distinct signature allows for an effective separation from background processes. A diagram of the VBF production at leading order is shown in Figure 1.3(b).

The ggF production of the Higgs boson is simulated using `Powheg-Box v2` [147–150] at next-to-next-to-leading order (NNLO) accuracy in $\alpha_s$ which is obtained with the `NNLOPS` [151] approach based on `HJ-MINLO` [152] by reweighting the events in the rapidity of the Higgs boson. It includes matching to the parton shower. The cross section is calculated with an N$^3$LO accuracy in $\alpha_s$ and includes NLO electroweak (EW) corrections [153, 154].

The VBF and VH production processes are simulated at NLO accuracy in $\alpha_s$ using `Powheg-Box v2` [147–149, 155] and the `MiNLO` [156, 157] approach. The cross sections are calculated at NNLO accuracy in $\alpha_s$ and with NLO EW corrections [57, 59–62] included.

For all production processes the `PDF4LHC15` [158] parametrisation for the parton distribution functions (pdf) is used, at NNLO accuracy for ggF and at NLO accuracy for VBF and VH. The parton shower, the underlying event and the hadronisation are simulated with `Pythia 8.2` [159]. Non-pterurbative effects are modelled using the `AZNLO` tune [160] with the `CTEQ6L1` [161] pdf set. The decay of the $\tau$-leptons is simulated with `Pythia`. A Higgs-boson mass of 125.09 GeV is used and a branching ratio of $H \to e\tau$ and $H \to \mu\tau$ of 1% is assumed if not given otherwise. Table 4.1 summarises the generators and pdf sets used for all processes and also includes their cross sections.

### 4.2.2 Background Processes

This section gives an overview over the different processes that show a similar signature as the signal processes and therefore contribute background events in this analysis. Commonly, one distinguishes *irreducible* and *reducible* background contributions. The former consist of processes creating the same final state objects as the signal, i.e. at least one electron and one muon of opposite-sign charges and potentially $E_{\mathrm{T}}^{\mathrm{miss}}$. The latter are also called *fake* backgrounds, indicating that at least one of the two leptons is "faked", i.e. another object is mis-identified as a lepton (electron or muon). Most processes can contribute to both categories due to detector deficiencies; however, the majority of their events contributes either to one or the other category. Typically, the cross sections of the background processes are substantially larger than that of signal. Hence, they must be suppressed by several orders of magnitude by a dedicated event selection (Section 7.1) and in the case of this analysis separated from the signal with the help of a multivariate technique (Section 8.2) in order to achieve an adequate sensitivity.

At the end of each process description a short summary of the event generators used to simulate the respective process is given. Table 4.1 summarises this information for all processes and also includes their cross sections.

#### $Z/\gamma^* \to \tau\tau/\ell\ell$ + Jets Production

The main irreducible background in this analysis stems from $Z/\gamma^* \to \tau\tau$. If both $\tau$-leptons decay leptonically they can reproduce the same final state, one electron, one muon and $E_{\mathrm{T}}^{\mathrm{miss}}$, as the signal process. Example Feynman diagrams of $Z/\gamma^*$-production at the LHC are shown in Figure 4.5. The first diagram, Figure 4.5(a), given at leading order also includes the subsequent decay into two $\tau$-leptons or two light leptons ($\ell$). The second diagram, Figure 4.5(b), shows a higher-order contribution with an additional gluon in

the final state while the third one in Figure 4.5(c) shows the electro-weak production of $Z/\gamma^* + 2\,\text{jets}$ which is similar to the VBF-production of the Higgs boson. Quark and gluons

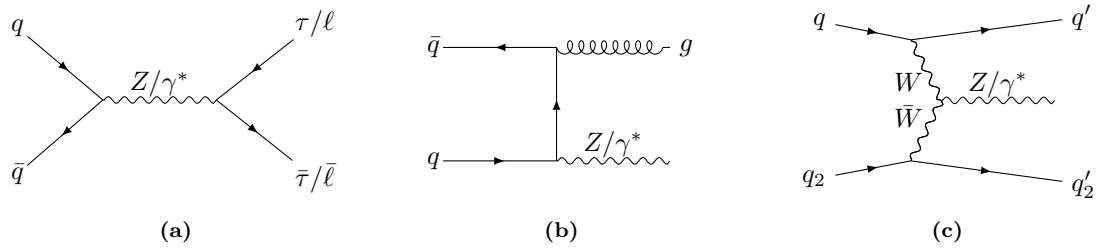



**Figure 4.5:** Example diagrams of $Z/\gamma^*$-production at the LHC; leading order (a) and higher order (b) as well as an electro-weak production diagram (c) are shown. In (a) the subsequent decay of $Z/\gamma^*$ into a pair of $\tau$-leptons or light leptons is shown in addition.

originating from higher-order contributions as well as the electro-weak production are reconstructed as jets and hence contribute additional objects to the final state. However, this is equivalent to the signal process as well as many other processes and therefore does not help in suppressing the contribution from $Z/\gamma^* \to \tau\tau$. There are two properties of the process that can be utilised to distinguish it from the signal processes: on one hand there are more neutrinos in the final state than in the signal process (four instead of two) which, in addition, originate from two particles instead of just one. On the other hand the invariant mass of the di-$\tau$-system peaks at the $Z$-boson mass.

The two processes $Z \to ee$ and $Z \to \mu\mu$, together referred to as $Z \to \ell\ell$, are reducible background contributions since both leptons in the final state have the same flavour. However, it can occur that one muon of a $Z \to \mu\mu$-event is mis-identified as an electron and hence the event can pass the different-flavour requirement of the event selection, see Section 7.2.2. The contribution of $Z \to ee$ is very small because the probability that an electron is mis-identified as a muon is almost zero since it usually does not reach the muon spectrometer of the ATLAS detector.

Both, the $Z$-boson as well as the photon $\gamma$, contribute indistinguishably to the discussed processes. However, in the following they will be abbreviated by $Z \to \tau\tau/\ell\ell$.

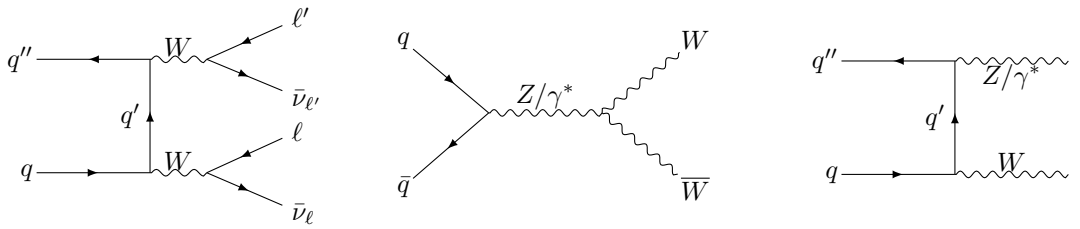
The simulated events of $Z \to \tau\tau$ and $Z \to ee$ are generated with `Sherpa 2.2.1` [162] using `Amegic` [163], `Comix` [164] and `OpenLoops` [165] for matrix element calculation and the `NNPDF3.0 NNLO` [166] pdf set is used. The accuracy of the matrix element is NLO in QCD for up to 2 partons and LO in QCD for up to 4 partons. The matching to the parton shower is performed using the `MC@NLO` and the `MEPS@NLO` [167] prescription. The parton shower, underlying event and hadronisation are also simulated with `Sherpa 2.2.1`. The `NNPDF3.0 NNLO` pdf set using a `Sherpa` tune is used to model non-perturbative effects. The electroweak production is simulated with up to 4 partons at LO. The decays of the $\tau$-leptons are simulated with `Sherpa`. The cross sections are calculated at NNLO accuracy in QCD using `FEWZ` [168].

These samples are sliced in $\max(H_T, p_T(V))$ and in $b$- or $c$-flavour of the jets which means that separate samples for different ranges in $\max(H_T, p_T(V))$ and the flavours are produced. $H_T$ represents the scalar sum of the transverse momenta of all parton-level jets with $p_T > 20\,\text{GeV}$ while $p_T(V)$ is the transverse momentum of the vector boson. The lowest slice of $\max(H_T, p_T(V))$ of the $Z \to \mu\mu$-sample has very low statistics in the *Basic Event Selection* (Section 7.1.2) used in this analysis, which yields large statistical uncertainties. Therefore, it was decided to use $Z \to \mu\mu$-samples that were generated using `Powheg-Box` [147–149, 169] at NLO accuracy in QCD and the `CT10 NLO` pdf set. The parton shower, underlying event and hadronisation is simulated with `Pythia 8.186`. The `CTEQ6L1` pdf set with the `AZNLO` tune is used for the simulation of non-perturbative effects. The decays of $b$- and $c$-flavour hadrons are simulated with `EvtGen 1.2.0` [170], of $\tau$-leptons

with Pythia and bremsstrahlung corrections with PHOTOS [171, 172]. The cross sections are calculated at NNLO accuracy in QCD using FEWZ [168]. These samples consist of only two slices in the invariant mass $m_{\ell\ell}$ of the two leptons and provide better statistics for the selection used in this analysis.

**Diboson Production**

The second largest background contribution stems from diboson production of which the creation of two $W$-bosons contributes the most. Both $W$-bosons can decay leptonically and hence produce a signal-like different-flavour final state with opposite-sign charges. Exemplary leading order Feynman diagrams are given in Figure 4.6. Figure 4.6(a) includes the subsequent leptonic decays of both $W$-bosons which are omitted in the other diagrams. In addition, small contributions are from $WZ$- and $ZZ$-production. The former can only




**(a)** **(b)** **(c)**

**Figure 4.6:** Example leading order diagrams of diboson production at the LHC. Diagrams (a) and (b) show possible production modes of two $W$-bosons, diagram (c) the production of $Z/\gamma^*$ and $W$. In diagram (a) the subsequent leptonic decays of both $W$-bosons which are omitted in the other diagrams is depicted.

pass the event selection if e.g. one of the leptons originating from the $Z$-boson is not identified or if the $Z$-boson decays into $\tau$-leptons and one of them decays hadronically (but is not identified as such) and the other one leptonically. Similar reasons allow an even smaller amount of the latter to pass the selection. Also, $V\gamma$ where $V$ stands for $W$ or $Z$, contributes a smaller amount, mainly through mis-identification of the photon as an electron or of a hadronically decaying $\tau$-lepton (if $V$ decays into one) as an electron.

The distribution of the reconstructed mass of the two leptons plus $E_{\mathrm{T}}^{\mathrm{miss}}$ behaves rather flat which provides at least some separation power. The momenta of the two leptons are more similar to each other compared to signal for which one shows a clearly harder momentum than the other. And also the angular correlations between the final state objects differ.

The simulation of the diboson events uses the same generators as the simulation of the $Z \to \tau\tau$ and $Z \to ee$ events. The matrix element is, however, calculated at NLO accuracy in QCD only for up to one parton and at LO accuracy only for up to three partons. The cross sections are calculated at NLO accuracy in QCD.

**Top-quark Production**

The production of a top anti-top pair ($t\bar{t}$-production) contributes a larger background in the *basic event selection* (Section 7.1) although it exhibits additional objects in the final state. The top quark almost exclusively decays into a $b$-quark and a $W$-boson. In the case of $t\bar{t}$-production this can lead to a final state consisting of two different-flavour leptons of opposite-signed electric charges if both $W$-bosons decay leptonically, neutrinos and two jets originating from the two $b$-quarks. This is depicted in Figure 4.7(a) which also shows part of the subsequent decay of the top quarks.

In the case of single-top production, only the one in association with a $W$-boson contributes as irreducible background while the other production modes show at most one lepton in the final state and hence only contribute as a reducible background component.

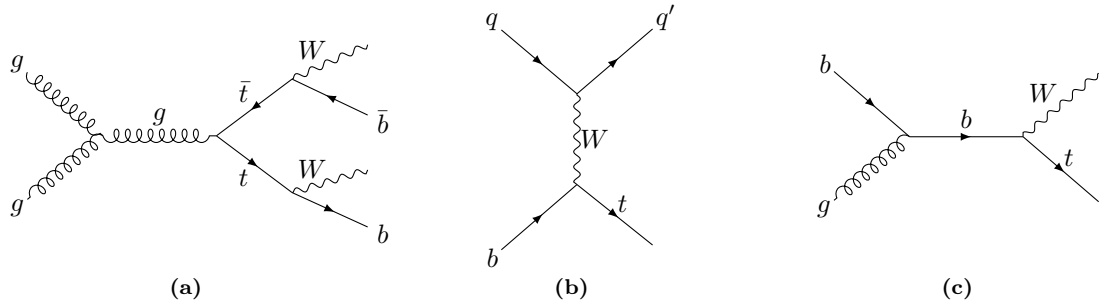



**Figure 4.7:** Example leading order diagrams of $t\bar{t}$-production (a), single-top production via t-channel (b) and in association with a $W$-boson (c). The first diagram (a) also depicts the decay of the top quarks into a $W$-boson and a $b$-quark.

The top-quark contributions[4] are suppressed by requiring a veto on $b$-jets[5] but due to its large cross section and the imperfect $b$-tagging there is nevertheless a considerable amount of events that passes the *basic event selection*. A further separation from the signal can also be achieved due to differences in kinematics and due to a rather flat distribution of the reconstructed mass of the two leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$.

The generation of $t\bar{t}$-events is performed with `Powheg-Box v2` [147–149] using the `NNPDF3.0 NNLO` [166] pdf set. They are interfaced to `Pythia 8.230` [159] for the simulation of the parton shower, the underlying event and the hadronisation. The `NNPDF2.3 LO` [166] pdf set with the `A14` [173] tune is used to model non-perturbative effects. `EvtGen 1.6.0` [170] is used to simulate the decays of $b$- and $c$-flavour hadrons.

The simulation of a single top-quark in association with a $W$-boson is performed using `Powheg-Box` [147–149] with the `NNPDF3.0 NLO` [166] pdf set. `Pythia 8` [159] is used to simulate the parton shower, the underlying event and the hadronisation. Non-perturbative effects are modelled by the `NNPDF2.3 LO` [166] pdf set with the `A14` [173] tune. `EvtGen 1.6.0` [170] is used to simulate the decays of $b$- and $c$-flavour hadrons.

The decays of $\tau$-leptons are simulated with `Pythia`. The cross section of $t\bar{t}$ is calculated at NNLO+NNLL (next-to-next-to leading order + next-to-next-to leading log) accuracy in QCD, the cross section of $Wt$ at NLO in QCD.

### $W \to \tau\nu/\ell\nu$ + Jets Production

The production of a $W$-boson and its subsequent decay into a $\tau$ or light lepton and a neutrino contributes to the reducible background component since the second lepton is missing. However, jets originating from additional quarks or gluons in the final state can be mis-identified as a lepton and non-prompt leptons from semi-leptonic decays of heavy-flavour quarks can pass the lepton identification criteria. Figure 4.8 shows a leading order diagram together with the leptonic decay of the $W$-boson as well as diagrams contributing to higher orders in QCD. Events from $W$ + jets are mainly suppressed by the appropriate choice of lepton identification and isolation requirements described in Chapter 5. Furthermore, separation power is obtained from the distribution of the reconstructed mass of the two leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$ and from the transverse mass of $\ell_\tau$ and $E_{\mathrm{T}}^{\mathrm{miss}}$.

The simulation of the $W$ + jets-events uses the same generators as the simulation of the $Z \to \tau\tau$ and $Z \to ee$ events, and the same accuracy in the calculation of the cross section.

### QCD Multijet Production

QCD multijet production does not create any prompt leptons but nevertheless contributes to the event yields in the signal region. It contributes solely through mis-identification

[4] The production of both, $t\bar{t}$ and single top, will be referred to together as *top-quark production*.
[5] More details on tagging jets as $b$-jets are given in Section 5.4.3.

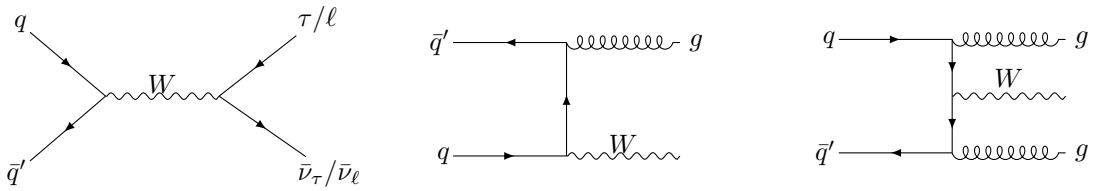



**Figure 4.8:** Example Feynman diagrams for the production of a $W$-boson at leading order (a), with 1 additional gluon (b) and two additional gluons (c) in the final state (contributing to higher orders in QCD). Diagram (a) also depicts the leptonic decay of the $W$-boson while this is omitted in the other diagrams.

of jets as leptons and through selection of non-prompt leptons from semi-leptonic decays of heavy-flavour quarks, thereby needing to fulfil the different flavour and opposite-sign charge requirements of the leptons. The probability for this to happen is rather low; however, the cross section of QCD multijet production at the LHC is many orders of magnitude larger than for Higgs-boson production such that a considerable amount remains. This background contribution is mainly suppressed by the appropriate choice of lepton identification and isolation requirements described in Chapter 5. In addition, the distribution of the reconstructed mass of the two leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$ is flatter compared to signal and shifted to lower masses. Exemplary Feynman diagrams representing multijet production are given in Figure 4.9. There are many more possible diagrams including ones with more partons in the final state, e.g. through subsequent radiation.

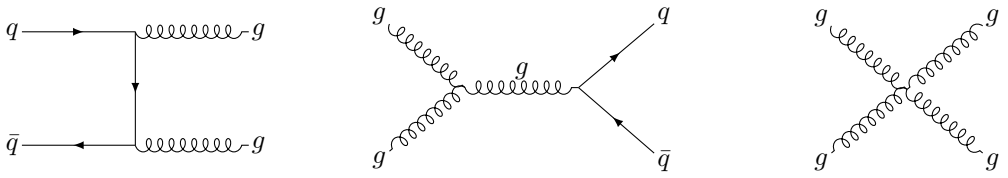



**Figure 4.9:** Example Feynman diagrams for QCD multijet production with two partons in the final state. Subsequent radiation can create more final state partons.

There are no simulated samples representing QCD multijet production used in this analysis since a huge amount of simulated events would be needed due to their large cross section but small selection probability. Hence, this contribution is solely estimated from data (Section 7.2.2).

### $H \to \tau\tau$ and $H \to WW$ Production

The Higgs-boson production with subsequent decay into two $\tau$-leptons or two $W$-bosons evidently also shows a similar signature as the signal processes if the $\tau$-leptons or $W$-bosons decay into an electron and a muon. Differences compared to the signal, such as the relation of the resulting $E_{\mathrm{T}}^{\mathrm{miss}}$ with the leptons, the $p_{\mathrm{T}}$-spectra of the leptons as well as differences in the distribution of the reconstructed mass of the two leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$, are utilised to separate the SM Higgs-boson production from the signal.

The simulation of the $H \to \tau\tau$- and $H \to WW$-events of the three main production channels uses the same generators as the simulation of the signal events. The $t\bar{t}H$-production uses `Powheg-Box v2` [147–149] and the `NNPDF3.0 NNLO` pdf set. This is interfaced to `Pythia 8.230` [159] for the simulation of the parton shower, the underlying event and the hadronisation. The `NNPDF2.3 LO` [166] pdf set with the `A14` tune is used for the modelling of non-perturbative effects. `EvtGen 1.6.0` [170] is used for the simulation of the decays of

$b$- and $c$-flavour hadrons. The cross section is calculated at NLO accuracy in QCD with NLO EW corrections included.

| Process | Generator | | PDF set | | Tune | Order of Cross-Section $\sigma$ | $\sigma$/pb |
|---|---|---|---|---|---|---|---|
| | ME | PS+UE | ME | PS+UE | | | |
| (sig + H bkg) ggF | Powheg-Box v2 NNLOPS | Pythia 8.212 | PDF4LHC15 NNLOv | CTEQ6L1 | AZNLO | $N^3LO$ QCD + NLO EW | 48.58 |
| (sig + H bkg) VBF | Powheg-Box v2 MiNLO | Pythia 8.230 | PDF4LHC15 NLO | CTEQ6L1 | AZNLO | NNLO QCD + NLO EW | 3.78 |
| (sig + H bkg) $WH$ | Powheg-Box v2 MiNLO | Pythia 8.212 | PDF4LHC15 NLO | CTEQ6L1 | AZNLO | NNLO QCD + NLO EW | 1.37 |
| (sig + H bkg) $ZH$ | Powheg-Box v2 MiNLO | Pythia 8.212 | PDF4LHC15 NLO | CTEQ6L1 | AZNLO | NNLO QCD + NLO EW | 0.88 |
| (H bkg) $t\bar{t}H$ | Powheg-Box v2 | Pythia 8.230 | NNPDF3.0 NNLO | NNPDF2.3 LO | A14 | NLO QCD + NLO EW | 0.51 |
| $Z \to \mu\mu$+jets | Powheg-Box | Pythia 8.186 | CT10 NLO | CTEQ6L1 | AZNLO | NNLO QCD | 3 762.91* |
| other $Z$+jets | Sherpa 2.2.1 | Sherpa 2.2.1 | NNPDF3.0 NNLO | NNPDF3.0 NNLO | Sherpa | NNLO QCD | 7 247.42* |
| $W$+jets | Sherpa 2.2.1 | Sherpa 2.2.1 | NNPDF3.0 NNLO | NNPDF3.0 NNLO | Sherpa | NNLO QCD | 59 828.15* |
| diboson | Sherpa 2.2.1 | Sherpa 2.2.1 | NNPDF3.0 NNLO | NNPDF3.0 NNLO | Sherpa | NLO QCD | 1 527.80* |
| $t\bar{t}$ | Powheg-Box v2 | Pythia 8.230 | NNPDF3.0 NLO | NNPDF2.3 LO | A14 | NNLO+NNLL QCD | 831.76 |
| single $t$ | Powheg-Box | Pythia 8 | NNPDF3.0 NLO | NNPDF2.3 LO | A14 | NLO QCD | 71.72 |

**Table 4.1:** Overview over the different MC event generators used for the simulation of the signal and background processes as well as over the different parametrisations used for the parton distribution functions (pdf). The abbreviations "sig" and "bkg" denote "signal" and "background", respectively. More details are given in the text at the end of the paragraph of each process. The cross sections in the last column represent the production cross sections. However, the values marked with a * are the product of production cross section, the branching ratios of the indicated decay or requiring at least one lepton in the final state and the efficiencies of the filters applied during simulation. All cross sections are given for $\sqrt{s} = 13\,\text{TeV}$.

# Chapter 5

# Reconstruction and Identification of Physics Objects

This chapter summarises the procedures of reconstructing, calibrating and identifying physics objects, such as muons, from detector signals. In the search for lepton-flavour violating Higgs-boson decays presented in this thesis, electrons and muons as well as missing transverse energy representing neutrinos, are of particular importance. Whereas hadronically decaying $\tau$-leptons and jets originating from $b$-quarks are vetoed in order to suppress background processes. Jets originating from quarks or gluons are also crucial, on one hand for a proper reconstruction of missing transverse energy, on the other hand for the definition of the two signal regions, *nonVBF* and *VBF* (Section 8.1).

The full processing chain is split into reconstruction with calibration, identification and also isolation from adjacent activity in the detector in the case of electrons and muons. Several *working point*s (WPs) which differ in signal efficiency and background rejection capabilities are provided for the latter two and for the tagging of jets originating from $b$-flavour hadrons. A higher signal efficiency inevitably comes at the expense of a lower background rejection, hence each analysis must choose its own optimal choice of WPs.

In addition, two types of reconstructed light leptons, *baseline* and *signal*, are defined. *Baseline* leptons are used in the overlap removal procedure, to select exactly two light leptons and to calculate the missing transverse energy. *Signal* leptons must fulfil stricter quality criteria and are utilised in all subsequent steps of the event reconstruction and selection. The overlap removal aims to solve possible ambiguities between identified objects as they are not mutually exclusive.

The requirements imposed on light leptons partially deviate from the standard ATLAS recommendations to ensure the best possible symmetric treatment of electrons and muons which is required in this analysis. In addition, it must be permitted that the light leptons can originate from $\tau$-lepton decays and hence do not stem for the primary vertex.

Simulated events and data are treated equally during the processing chain. However, small differences due to imperfect simulations can occur, leading to slight deviations in the efficiencies. Therefore, *scale factor*s ($SF$s) are derived as the ratio of efficiencies measured in data and in simulated events. These are subsequently applied to the simulated events in order to correct them to agree with data.

Section 5.1 starts with a description of track finding and vertex reconstruction which is a key component for the reconstruction of all physics objects. It is followed by a discussion of the reconstruction and identification of electrons and muons in Sections 5.2 and 5.3, respectively. Jets are discussed in Section 5.4 followed by a brief overview of the applied algorithms to tag jets originating from $b$-quarks and of hadronically decaying $\tau$-leptons in Section 5.5. Finally, the reconstruction of the missing transverse energy and the overlap removal procedure are introduced in Sections 5.6 and 5.7, respectively.

## 5.1 Track and Vertex Reconstruction

The reconstruction of tracks (i.e. the trajectories of charged particles) and vertices in the *Inner Detector* (ID) lays the foundation for most of all subsequent object reconstruction. First, *clusters* are built from raw measurements by grouping pixels (from the Pixel detector) or strips (from SCT) in a given sensor if the deposited energy exceeds a certain limit [174, 175]. Sets of three of these clusters in different layers of the ID form a track seed which has to fulfil additional basic quality criteria. Track candidates are built by successively incorporating additional clusters from remaining layers in the direction of the track seed. Depending on the evaluation of this additional cluster based on a combinatorial Kalman filter [176], it is either added to the track candidate or rejected. If there are several clusters on the same layer compatible with the track seed, multiple track candidates are formed. This leads to a collection of track candidates that comprises fake tracks and tracks overlapping with each other. Therefore, an ambiguity solver is deployed. This tool scores the tracks based on simple measures of the track quality where higher scores correspond to better quality and handles them in descending order. Several additional criteria are imposed on the candidates. Amongst others [175], a minimum transverse momentum of 500 MeV is required [177]. The tracks found in the silicon detectors are then extended outwards into the TRT if compatible hits are found. This procedure is called *inside-out* and relies on finding track seeds in the silicon detectors. However, tracks from prolonged decays of particles with a longer lifetime or from photon conversions may not leave enough hits within the silicon detectors. Therefore, the *outside-in* sequence is deployed on all TRT hits that remain after the *inside-out* algorithm. Pattern recognition is used to find track segments in the TRT which are propagated into the silicon detectors utilising residual clusters [174]. The fraction of tracks that fail to be reconstructed in the core of jets is measured to be $0.061 \pm 0.006\,(\text{stat.}) \pm 0.014\,(\text{syst.})$ and $0.093 \pm 0.017\,(\text{stat.}) \pm 0.021\,(\text{syst.})$ for transverse momenta of the jets of 200 to 400 GeV and 1 400 to 1 600 GeV, respectively [175].

Primary vertices are reconstructed from the track collection in order to find the spatial coordinates of the hard-scattering interaction [178]. The position of the vertex seed defines the starting point of the primary vertex finding procedure. The transverse position of the seed is taken from the beam spot position and its $z$-coordinate is obtained from the mode of the tracks' $z$-coordinates (at the point where they are closest to the beam spot). After finding the optimal vertex position through an iterative $\chi^2$-minimisation, tracks less compatible with the position are down-weighted and the vertex coordinates are recalculated. This is repeated until certain criteria [175] are met. Tracks incompatible with the final vertex position are removed and are utilised in sub-sequent vertex findings. The vertices must have at least two associated tracks. The primary vertex with the largest sum of squared transverse momenta of associated tracks, $\sum p_T^2$, is chosen as the hard-scatter primary vertex. The remaining ones are categorised as pile-up vertices. The efficiency of vertex reconstruction in low-$\mu$ (low average number of interactions per bunch crossing) data is shown in Figure 5.1. It is obtained from data taken in one run in the year 2015.

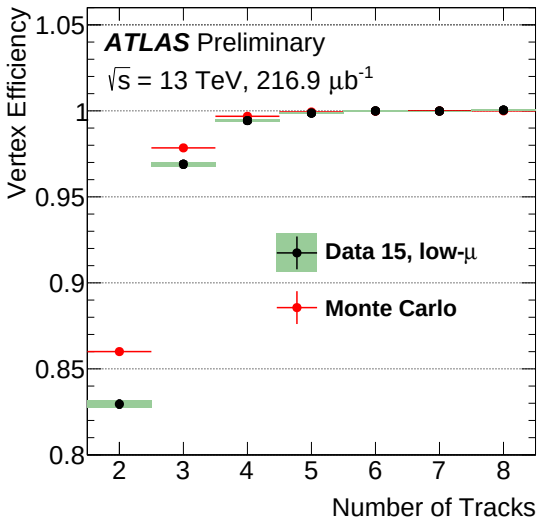



**Figure 5.1:** Vertex reconstruction efficiency in low-$\mu$ data (black) as a function of the number of tracks associated with this vertex. The efficiency obtained from Monte Carlo simulations is given in red [179].

## 5.2 Electrons

Electrons induce a signal in the ID and the ECAL. Through interactions with the detector material, they deposit a significant amount of their energy due to bremsstrahlung. This triggers a subsequent cascade of photons and electron-positron pairs. A relatively narrow shower in the calorimeter is expected. Photons behave in a similar way. But an unconverted photon does not show any tracks in the ID while a converted one comes with a conversion vertex. However, since only electrons are of direct interest to this analysis, the following description is limited to those.

### 5.2.1 Reconstruction

The electron reconstruction combines clusters of energy deposition in the ECAL originating from particle showers initiated by electrons with tracks in the ID [180]. First, so-called *topo-clusters* that are clusters spreading over topologically connected ECAL and HCAL cells are formed [180–182]. Cells with a large signal to noise ratio, i.e. $|\zeta_{\text{cell}}| \geq 4$, serve as seeds for the formation of *proto-clusters*. The cell signal significance $\zeta_{\text{cell}}$ is defined as

$$\zeta_{\text{cell}} = \frac{E_{\text{cell}}}{\sigma_{\text{noise,cell}}}\,, \tag{5.1}$$

with $E_{\text{cell}}$ being the energy deposit and $\sigma_{\text{noise,cell}}$ the expected noise in this cell. Neighbouring cells with $|\zeta_{\text{cell}}| \geq 2$ are added until they fail the requirement. If two proto-clusters share a cell, these two are merged. Finally, all nearest-neighbour cells of a proto-cluster are added, independent of their $\zeta_{\text{cell}}$-value. Proto-clusters are split if they contain more than one local maximum, i.e. the cell energy $E_{\text{cell}}$ must be larger than 500 MeV, it must have at least four neighbours and none of the neighbours must have a larger energy. When the formation of proto-clusters is finalised, all proto-clusters are topo-clusters. However, for the subsequent steps of the electron reconstruction, only the energy of cells in the ECAL ($E_{\text{EM}}$) is used and only topo-clusters with $E_{\text{EM}} > 400$ MeV are considered. In addition, the fraction of $E_{\text{EM}}$ on the total energy of the topo-cluster is required to be larger than 0.5. This rejects 60 % of pile-up clusters without reducing the efficiency to select true electron topo-clusters [180]. Tracks are reconstructed in the ID following the procedure described in Section 5.1. If there are no resulting tracks for a silicon track seed in regions-of-interest

(ROIs) created from fixed-sized clusters in the calorimeter that exhibit a typical EM shower profile, a modified pattern-recognition algorithm is deployed. It accounts for enhanced energy loss from interactions with the detector material due to bremsstrahlung. Tracks loosely matched to these fixed-sized clusters are refitted to obtain more accurate track parameters using a Gaussian sum filter (GSF) algorithm [183], also taking into account the enhanced energy loss due to bremsstrahlung. Finally, the refitted tracks are matched to the EM topo-clusters. If there is more than one track assigned to a cluster, the tracks are ranked by several quality and matching criteria and the highest ranked one is chosen.

In order to create *superclusters*, all topo-clusters are sorted in descending order by their transverse EM energy $E_T$. Each of them is checked whether it fulfils the requirements to be a seed candidate. It must be matched to a track with at least four hits in the silicon detectors and its $E_T$ must be larger than 1 GeV. Next, so-called *satellite clusters* within a window of $\Delta\eta \times \Delta\phi = (3 \times 5) \cdot 0.025$ around the seed cluster barycentre are identified. If the best-matched track of the satellite candidate is identical with the best-matched track of the seed, the window size is allowed to be $(5 \times 12) \cdot 0.025$. All topo-clusters found to be satellite clusters are removed from the topo-cluster list such that they cannot serve as seed clusters themselves. The track information is included in this procedure to suppress electrons from conversions, pile-up noise and unrelated clusters. The satellite clusters are results of secondary EM showers from the same initial electron. The combination of the seed and its satellite clusters form the supercluster.

Then, a matching of tracks and superclusters as well as an initial energy calibration are performed. However, the same seed cluster can lead to both, an electron and a photon candidate. If it is a cluster without a good associated track (following the criteria described in [180]), it is reconstructed as a photon. If it is a cluster with a good associated track, it is reconstructed as an electron. Otherwise an electron and a photon object, which are marked as ambiguous, are created for analysis. Then, the energies are recalibrated while considering the matched track in order to represent the original electron energy (via the procedure described in [184]). Finally, the $p_{\mathrm{T}}$ of the electron is obtained by the energy measurement in the calorimeter while $\eta$ and $\phi$ are taken from the properties of the associated track.

The efficiency of the electron reconstruction as a function of the true $E_T$ determined with simulated events is given in Figure 5.2 (purple triangles). The other curves show the efficiencies after the single steps in the reconstruction chain.

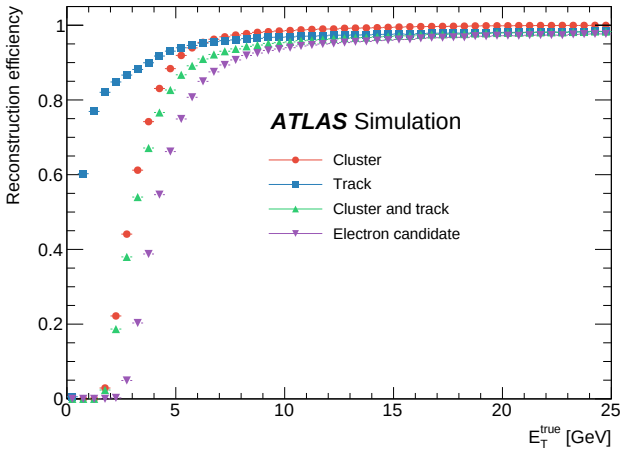



**Figure 5.2:** The efficiencies of the different electron reconstruction steps and their combinations as a function of the true transverse energy determined in simulated events [180].

### 5.2.2 Identification

The identification algorithm aims at an improved purity of real electrons and therefore a reduction of classification of other objects as electrons.

The identification of prompt leptons is based on a likelihood discriminant $d_L$ [180, 185]. It is defined as the natural logarithm of the ratio of the signal and the background likelihood, $L_S$ and $L_B$. The signal likelihood describes the probability for a reconstructed electron to be prompt while the background likelihood describes its probability to originate from hadronic jets, converted photons, heavy-flavour decays such as $b$- and $c$-flavour hadrons or hadronic $\tau$-lepton decays. The likelihoods are constructed from a set of discriminating quantities which are properties of the primary electron track, lateral and longitudinal development of the EM shower in the calorimeter and the spatial compatibility of the electron track and its reconstructed cluster. For each variable, probability density functions are derived from smoothed histograms, separately for signal and background and for 9 bins in $|\eta|$ and 7 bins in $E_T$. The signal histograms are filled with $Z \to ee$-events[1] recorded in the years 2015 and 2016 while the background histograms are primarily obtained from dijet events. In order to select either signal- or background-like data, a tag-and-probe method is used. The single pdfs are combined into the likelihoods $L_S$ and $L_B$. Finally, a cut on the discriminant $d_L$ can be imposed. Depending on the cut value and additional criteria, three different *working point*s (WPs) for the electron identification are defined. The working points, *loose*, *medium* and *tight*, differ in their efficiency to identify prompt electrons: *loose* has the highest efficiency at the price of the lowest background rejection while it is the opposite for *tight*. The improvement in background rejection is approximately a factor 2.0 for *medium* and 3.5 for *tight* compared to *loose*. The efficiencies as a function of $E_T$ and $\eta$ are shown in Figure 5.3. The efficiencies increase with increasing $E_T$ from $E_T = 15\,\text{GeV}$

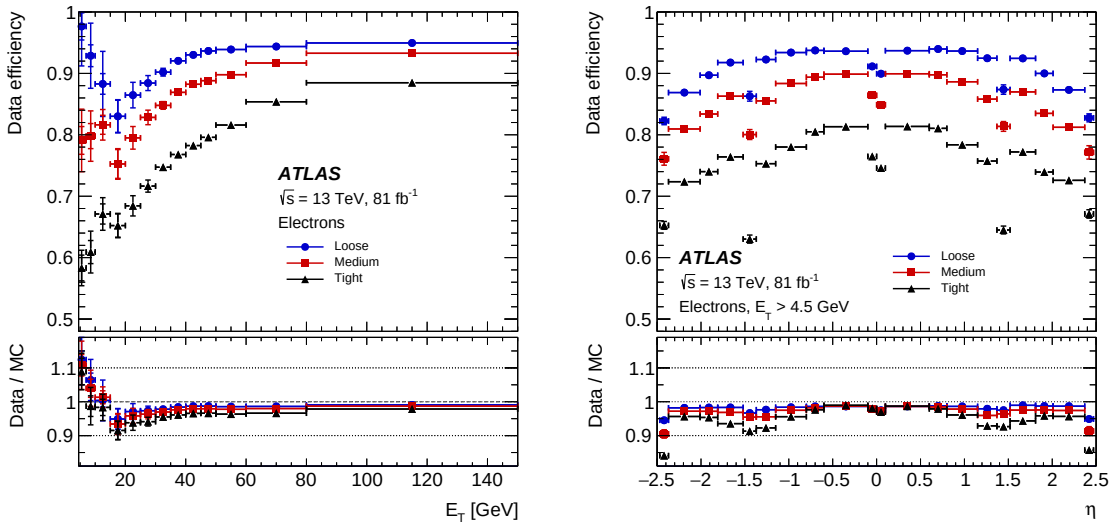



**Figure 5.3:** Electron identification efficiencies in data as a function of $E_T$ (a) and $\eta$ (b). The three different colors represent the three different working points. The outer error bars indicate the total uncertainties. The bottom panels show the efficiency ratio of data and simulation (scale factors) [180].

onwards. A plateau is reached at around 50 GeV, a bit later for the *tight* working point. At $E_T = 100\,\text{GeV}$, efficiencies of 95 % (*loose*), 93 % (*medium*) and 88 % (*tight*) are reached. The increase in efficiency below $E_T = 15\,\text{GeV}$ is due to a known mismodelling of the variables entering the likelihoods at this energy scale. The electron identification with

[1] $Z \to ee$-events are used for $E_T > 15\,\text{GeV}$ while $J/\psi \to ee$-events are used for $E_T < 15\,\text{GeV}$. However, in this analysis only electrons with $p_T > 15\,\text{GeV}$ are considered.

respect to $\eta$ is most efficient in the central region, i.e. for lower $|\eta|$-values. Low efficiencies are experienced in the transition regions of the calorimeter between barrel and endcaps, $1.37 < |\eta| < 1.52$ and perpendicular to the beam axis ($\eta \approx 0$). In addition, scale factors that are simulation-to-data corrections, are derived in bins of $p_{\mathrm{T}}$ and $\eta$ and shown in the lower panels of Figure 5.3. They are all below 1 (for $p_{\mathrm{T}} > 15\,\mathrm{GeV}$) and above 0.9 except for the highest bins in $|\eta|$ where the scale factors for the *tight* working point are below 0.9.

In this analysis, the *loose* identification WP is chosen for *baseline* electrons and the *medium* identification WP for *signal* electrons. In addition, $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| < 2.47$ are required with a veto of electrons inside the calorimeter transition region, $1.37 < |\eta| < 1.52$. The absolute value of the longitudinal impact parameter of the electron track, calculated with respect to the primary vertex, and multiplied by $\sin\theta$ of the track is required to be $|z_0 \sin\theta| < 0.5\,\mathrm{mm}$ with $\theta$ being the polar angle. The significance of the transverse impact parameter $d_0$ which corresponds to its value divided by its resolution is required to be $|d_0|/\sigma_{d_0} < 10$. This requirement is less stringent than the usual recommendation for electrons but allows for the origin from a leptonically decaying $\tau$-lepton due to the additional decay length. The same cut value is chosen, independently whether the electron is assumed to originate directly from the $H$-boson or from the $\tau$-lepton, in order to keep a symmetric selection for all electrons which is required for the usage of the *Symmetry Method* (Chapter 3).

### 5.2.3 Isolation

The isolation of an electron from other objects can help to further classify it as prompt or not. If it is non-prompt or mis-identified it is typically accompanied by many more particles and the activity around it is enhanced. In general, the isolation is measured by the amount of additional activity in a cone around the particle of interest, here an electron. Therefore, two types of isolation variables are defined, on the one hand describing the additional activity in the ECAL on the other hand in the ID [180, 185].

In the calorimeter, a variable called $E_T^{\mathrm{cone20}}$ is utilised. A cone of size $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.2$ around the electron cluster barycentre is defined. The transverse energies of all topological clusters whose barycentres are within this cone are summed. In order to only get a measure of the activity around the electron, its energy in a rectangular of $\Delta\eta \times \Delta\phi = (5 \times 7) \cdot 0.025$ around its barycentre is subtracted from the summed energies. However, not all of the energy of the electron is removed this way and hence a leakage correction depending on $E_T$ and $|\eta|$ must be considered as well as a correction for contributions from pile-up and underlying event [186].

In the ID, a variable called $p_T^{\mathrm{varcone20}}$ is utilised. Again, a cone around the electron is defined. The cone is centred around the electron track direction and has a variable cone size depending on the transverse momentum of the electron:

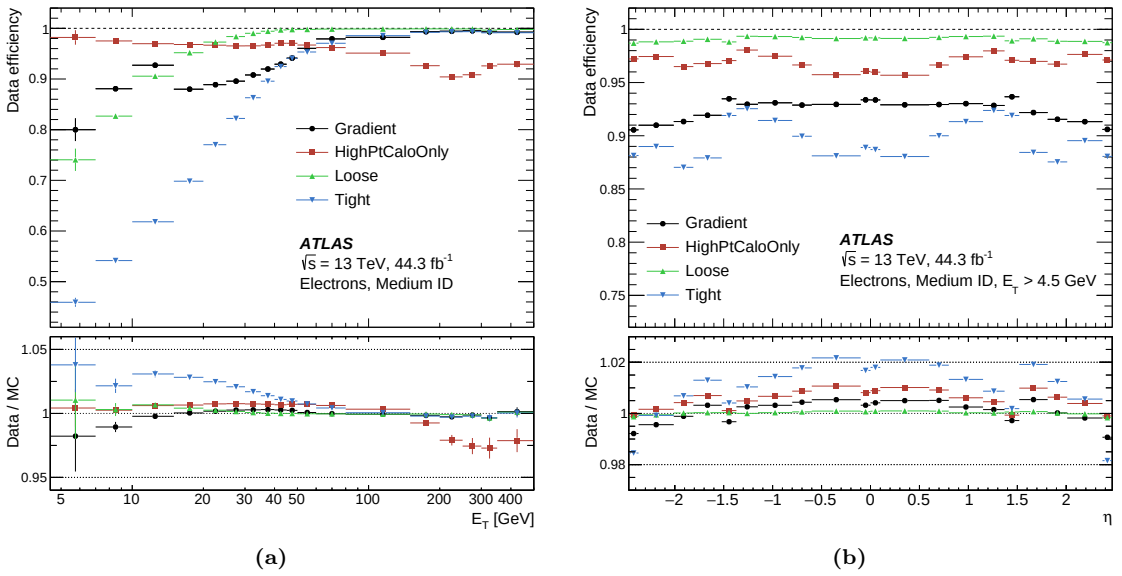
$$\Delta R = \min\left(\frac{10}{p_T[\mathrm{GeV}]}, \Delta R_{\mathrm{max}}\right), \tag{5.2}$$

with $\Delta R_{\mathrm{max}}$ being the maximum cone size of 0.2. The cone size is reduced with increasing $p_T$ since the tracks of other decay products from the same boosted primary particle can be very close. The transverse momenta of all tracks fulfilling certain criteria within this cone but the one matched to the electron are summed. These criteria comprise a minimum $p_T$ of 1 GeV and $|\eta| < 2.5$, several requirements on the hits and a loose vertex association.

Isolation *working point*s can either be obtained by imposing fixed cuts on $E_T^{\mathrm{cone20}}$ and/or $p_T^{\mathrm{varcone20}}$ or by aiming at a fixed value of efficiency and hence having varying cut values. The latter method is used for the *gradient working point* which is chosen in this analysis. It is designed to reach an efficiency of 90 % at $p_T = 25\,\mathrm{GeV}$ and of 99 % at $p_T = 60\,\mathrm{GeV}$, i.e.

it is parametrised as a function of $p_T$ but aims for uniformity in $\eta$. The respective varying cut values on the two isolation variables are obtained from simulated $Z \to ee$-events[1] that pass the *tight* identification requirements.

The resulting efficiencies measured in data recorded in the year 2017 are shown in Figure 5.4 for the *gradient* isolation WP and other WPs defined by fixed cuts on the isolation variables. A data sample enriched with $Z \to ee$ events and electrons passing the




**(a)** **(b)**

**Figure 5.4:** Electron isolation efficiencies in data as a function of $E_T$ (a) and $\eta$ (b) of electrons passing the *medium* identification requirement. The different colours represent different working points. The total uncertainties are indicated by the vertical bars. The bottom panels show the efficiency ratio of data and simulation (scale factors) [180].

*medium* identification requirements were used. In general, the efficiencies vary for different processes. This can be observed at $E_T = 15\,\text{GeV}$ in Figure 5.4(a) for the *gradient* isolation WP. Below $E_T = 15\,\text{GeV}$, $J/\psi \to ee$ simulated events were used to define the variable cuts on $E_T^{\text{cone20}}$ and $p_T^{\text{varcone20}}$ while $Z \to ee$ events are used to measure the efficiencies displayed in the figure. In the range from $E_T = 15\,\text{GeV}$ to $E_T = 40\,\text{GeV}$ it is clearly visible that the *gradient* WP provides a good compromise between the *Loose* (and *HighPtCaloOnly*) and *Tight* isolation WPs, keeping higher event statistics but still benefiting from a reasonable background suppression. The efficiency as a function of $\eta$ is close to flat with a slight decrease in the more forward directions. The scale factors which are the simulation-to-data corrections, deviate by less than 1 % from 1.0 above $E_T = 15\,\text{GeV}$.

## 5.3 Muons

Muons are minimal-ionising particles that manifest themselves in the detector as tracks in the *Inner Detector* (ID), small energy deposits in the calorimeters and tracks in the *Muon System* (MS) while traversing the whole detector.

### 5.3.1 Reconstruction

The reconstruction of tracks in the ID is done as described in Section 5.1. In the MS (Section 2.2.4) local straight-line track segments are formed within each chamber from hits gleaned by a tailored pattern recognition algorithm [187, 188]. Preliminary track candidates are built by combining the segments from the different layers. They must roughly point into the direction of the *interaction point* (IP) and be in agreement with the expected

bending in the $R$-$z$-plane for muons due to the magnetic field. The precision measurement of $\eta$ in the bending plane is combined with the $\phi$-measurement of the trigger chambers to construct a three-dimensional picture. Then, a global $\chi^2$-fit on the track candidate hits is performed. Afterwards, additional hits that lie on the resulting trajectory but were not taken into account before are added while hits deviating from the trajectory are removed and the fit is repeated. Possible ambiguities are resolved depending on the quality of the tracks and the remaining tracks are re-fitted including a loose IP constraint. Eventually, the tracks are extraploated to the beam line.

For the global muon reconstruction, measurements from the ID, the MS and also the calorimeters are combined and five different reconstruction strategies are pursued which are detailed in the following.

**Combined Muons (CB)** A combined fit of hits corresponding to matched tracks in the MS and the ID is performed while also taking into account the energy loss in the calorimeters. The MS tracks associated with the resulting trajectory are again updated and the fit is repeated.

**Inside-out Combined Muons (IO)** Again, a combined fit of hits in the ID and the MS is performed. But to locate the relevant hits, the track in the ID is extrapolated into the MS and at least three loosely-aligned MS hits must be found. That means no independently reconstructed track in the MS is necessary.

**MS Extrapolated Muons (ME)** An MS track is extrapolated to the beamline if no ID track can be matched. This is mainly beneficial in the $|\eta|$-region outside the ID acceptance: $2.5 < |\eta| < 2.7$.

**Segment-tagged Muons (ST)** If an ID track can be extrapolated to the MS by fulfilling tight angular matching requirements with at least one reconstructed MS segment, the ID track is utilised as muon candidate and provides its properties.

**Calorimeter-tagged Muons (CT)** If an ID track can be extrapolated to the calorimeters by matching energy deposits satisfying requirements of minimal-ionising particles, the ID track is utilised as muon candidate and provides its properties. Due to an overwhelming background contamination at low transverse momenta, a threshold of $p_T = 5\,\text{GeV}$ is required.

In addition, a calibration of the momentum scale and resolution is performed and corrections are applied to simulated events to compensate for an imperfect description of the detector response due to different areas of detector technology, inaccuracies in the description of the magnetic field integral and misalignment of the detector components [188]. The correction factors are binned in $p_\text{T}$, $\eta$ and $\phi$, reflecting different detector regions and are derived from comparing the invariant mass distributions of $Z \to \mu\mu$ and $J/\Psi \to \mu\mu$ events in simulation and data. Only CB-muons are considered for this procedure.

### 5.3.2 Identification

In order to improve the purity of the muon candidates and to enhance their quality, additional criteria are imposed on them. These include the number of hits in the ID and MS, properties of the track fit and compatibility of the measurements in the sub-detectors [187, 188]. Depending on the strictness of these criteria, different *working point*s (WPs) are defined: *loose*, *medium*, *tight*, *low-$p_T$* and *high-$p_T$*. As for the electrons, the *loose* WP has the highest efficiency and the lowest background rejection while the *tight* WP has the lowest efficiency but the best background rejection. The *loose* WP is designed for a specific analysis targeting Higgs-boson decays in the four-muon final state and is used in

this analysis only for *baseline* muons. The *high-$p_T$* WP is intended for $W'$ and $Z'$ searches while the *low-$p_T$* WP targets muons with very low transverse momenta not relevant for this analysis.

Within the $\eta$-range relevant for this analysis ($|\eta| < 2.47$), only CB and IO muons are used for the *medium* WP which is chosen in this analysis for *signal* muons. The muon candidates are required to have at least three hits in at least two layers of the precision chambers while one layer is sufficient for $|\eta| < 0.1$. Furthermore, a loose compatibility between the ID and MS measurements is ensured by requiring the $q/p$-compatibility to be smaller than seven. The latter is defined as the absolute value of the difference of $q/p$ measured independently in the ID and the MS divided by the square root of the corresponding uncertainties added in quadrature.

The *loose* WP, in addition, also accepts CT and ST muons for $|\eta| < 0.1$ and has some looser requirements for muon candidates with $p_T < 7\,\text{GeV}$ which are not considered within this analysis.

The efficiencies of the three WPs *loose*, *medium* and *tight* as a function of $p_T$ and $\eta$ are shown in Figure 5.5. All ID tracks that are associated with a reconstructed muon are in the denominator of the efficiency calculation while in the numerator they have to pass the WP requirements in addition. The efficiencies are derived from simulated $t\bar{t}$-events. For

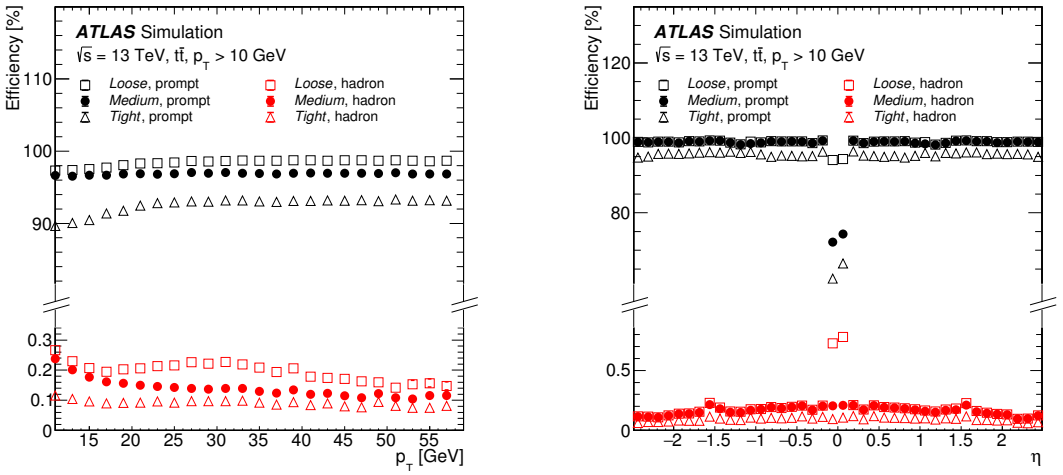
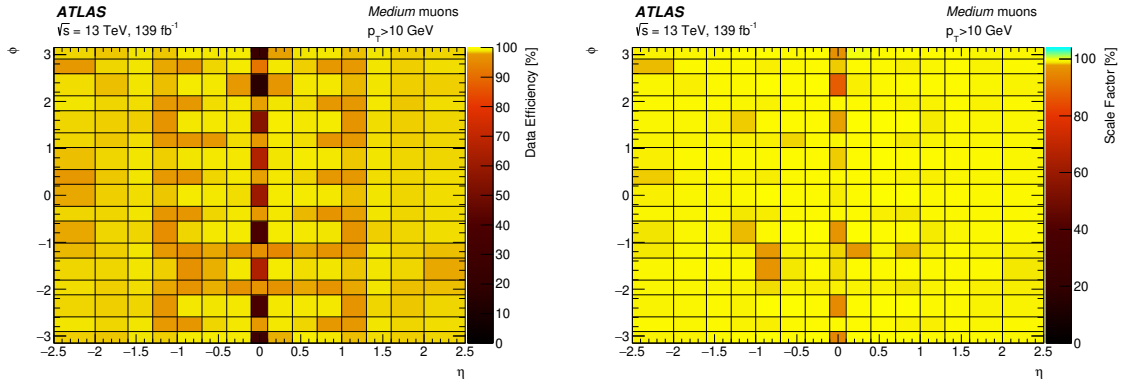



**Figure 5.5:** Muon reconstruction and identification efficiencies in simulated $t\bar{t}$-events as a function of $p_T$ (a) and $\eta$ (b) for the different working points. Black shows the efficiencies for prompt muons, red for muons from light hadron decays [187].

prompt muons passing the *medium* WP, $\eta$-inclusive efficiencies of up to 97 % are reached. Whereas the maximum efficiency for muons from light hadron decays is 0.24 %.

The scale factors, i.e. the simulation to data correction factors are parametrised as a function of $\eta$ and $\phi$. Figure 5.6 shows the data efficiencies as a function of these two variables and the resulting scale factors. Here, the efficiencies are obtained from $Z \to \mu\mu$-events. Overall, data and simulation are in good agreement and the scale factors are all close to 1 (100 %). The largest deviations from 1 are for $|\eta| < 0.1$ and at around $\phi = -1.2$ and $\phi = -2.0$, corresponding to the support structure of the ATLAS detector.

In this analysis, *baseline* muons must pass the *loose* identification WP and fulfil $p_\text{T} > 10\,\text{GeV}$ and $|\eta| < 2.47$. *Signal* muons must pass the *medium* identification WP and the cut on the transverse momentum is tightened to $p_\text{T} > 15\,\text{GeV}$ to be symmetric with the cut for electrons. Also due to symmetry reasons, muons in the transition region of the calorimeters, $1.37 < |\eta| < 1.52$, are vetoed and the same impact parameter cuts as for




**(a)** **(b)**

**Figure 5.6:** Muon efficiencies measured in data enriched with $Z \to \mu\mu$-events (a) and the corresponding scale factors (b) as a function of $\eta$ and $\phi$ for the *medium* working point [187].

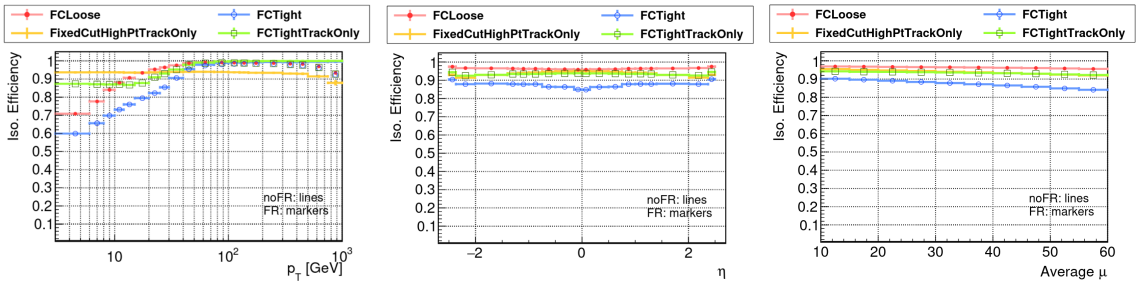
electrons are applied: $|z_0 \sin\theta| < 0.5\,\text{mm}$ and $|d_0|/\sigma_{d_0} < 10$. As for electrons, the latter is less stringent than generally recommended for muons to allow them to originate from leptonically decaying $\tau$-leptons due to their non-negligible decay length.

### 5.3.3 Isolation

The activity in the vicinity of non-prompt muons from heavy-flavour hadron decays is enhanced compared to prompt muons. This difference in isolation is utilised to further suppress the heavy-flavour background. The muon isolation is defined by the amount of energy or momentum reconstructed in a cone of $\Delta R$ around the muon, divided by the transverse momentum of the muon. This can be calculated by either using the tracks in the ID (transverse momentum), the energy deposit in the calorimeters (transverse energy) or by both. For the track-based isolation, the isolation cone can either have a fixed size with $\Delta R = 0.2$ leading to the observable $p_T^{\text{cone20}}$ or by a varying cone size $\Delta R = \min(10\,\text{GeV}/p_T^{\mu}, 0.3)$ leading to the observable $p_T^{\text{varcone30}}$. The transverse momenta of all ID tracks (but not the muon) within this cone that are associated with the primary vertex are considered. For the calorimeter-based isolation, a fixed cone size of $\Delta R = 0.2$ is chosen leading to the observable $E_T^{\text{topocone20}}$. All energy-deposits, disregarding the ones associated to the muon, are considered and a correction for pile-up effects is performed. The advantages of the track-based isolation are the lower pile-up dependence and a better resolution whereas the calorimeter-based isolation takes into account neutral particles and particles with transverse momenta below the threshold required in the ID (0.5 GeV or 1 GeV depending on the working point).

Different isolation working points are defined, either making use of only the track-based isolation or of both. In this analysis, the *FCTightTrackOnly_FixedRad* working point is applied. It is defined only by the track-based isolation with the fixed and the variable cone size observables: $p_T^{\text{varcone30}}/p_{\text{T}}^{\mu} < 0.06$ for $p_{\text{T}}^{\mu} \leq 50\,\text{GeV}$ and $p_T^{\text{cone20}}/p_{\text{T}}^{\mu} < 0.06$ for $p_{\text{T}}^{\mu} > 50\,\text{GeV}$, requiring a tight association of the tracks to the vertex and the transverse momenta of the tracks larger than 1 GeV. The resulting efficiencies obtained from simulated diboson events are visualised in green in Figure 5.7.

It can be seen that the *FCTightTrackOnly_FixedRad* WP provides a trade-off between the *FCLoose_FixedRad* and the *FCTight_FixedRad* WPs. The efficiencies for *FCTightTrackOnly_FixedRad* are almost flat in $\eta$ and have a reduced dependency on the average interactions per bunch-crossing compared to *FCTight_FixedRad* which, in addition, also makes use of the calorimeter-based isolation. Scale factors for the isolation efficiencies, binned in $p_{\text{T}}$, are provided as well and applied to simulated events in the analysis.




**(a)** **(b)** **(c)**

**Figure 5.7:** Muon isolation efficiencies obtained from simulated diboson events as a function of $p_{\mathrm{T}}$ (a), $\eta$ (b) and the average number of bunch-crossings per event (c) for muons passing the *medium* identification [189]. (b) and (c) include all muons with $p_{\mathrm{T}} \in [12, 50]\,\mathrm{GeV}$. The efficiencies of eight different working points are shown, the markers indicate the working points with _FixedRad.

## 5.4 Jets

Partons from the hard-scatter interaction hadronise and are detected as jets which are the output of a *jet algorithm*. They manifest in the detector by tracks in the ID and large energy deposits in the calorimeters. Due to the high production rate of jets, their efficient and precise reconstruction is crucial. They are an important ingredient to the calculation of missing transverse energy and in this analysis their multiplicity is utilised to define different event categories.

### 5.4.1 Reconstruction

Jets are reconstructed from energy deposits in the calorimeters, taking into account information from tracks in the ID by utilising the particle flow algorithm[190]. Tracks are reconstructed as described in Section 5.1 and are required to be within $|\eta| < 2.5$ and have $p_T > 500\,\mathrm{MeV}$, topo-clusters are formed from the energy deposits in the calorimeters in the same way as for electrons, discussed in Section 5.2.1. The former are rejected if they are matched to a candidate electron or muon. The latter are initially calibrated to the electromagnetic scale (as for electrons and photons) when entering the particle flow algorithm. The algorithm then matches tracks to topo-clusters and subtracts the energy corresponding to the matched tracks cell by cell thereby preventing that energy measured in the different sub-detectors is double-counted. The reason for combining information from the ID and the calorimeters and not solely taking the energy measurement of the latter is the superior momentum and $\eta$-$\phi$-resolution in the ID, the vertexing-capability and the possibility to take low-energetic charged particles into account which otherwise are below the noise threshold of the calorimeters. Therefore, only tracks with a transverse momentum below 40 GeV are considered. Higher-energetic particles are often less isolated, leading to an inaccurate energy-subtraction. In addition, the quality of the energy measurement in the calorimeters is comparable to the one in the ID at these energies. The pairs of tracks and (modified) topo-clusters, or only the latter if no tracks are assigned, are forwarded to the anti-$k_T$-algorithm [191] performing the jet-finding step. Before, all tracks failing $|z_0 \sin\theta| < 2\,\mathrm{mm}$ are rejected together with their associated calorimeter energy to suppress pile-up activity. And the $\eta$ and $\phi$ components of the topo-clusters are re-computed with respect to the hard-scatter primary vertex instead of the detector origin.

The anti-$k_T$-algorithm [191] is a jet clustering algorithm that is infrared and collinear safe and therefore the jet building is insensitive to splitting of partons. First, the distances $d_{ij}$ of all input 'particles' (either track and modified topo-cluster pairs or topo-clusters only)

to each other as well as their distances $d_{iB}$ to the beam line are calculated as

$$\begin{aligned} d_{ij} &= \min\left(p_{T,i}^{-2}, p_{T,j}^{-2}\right) \frac{\Delta R_{ij}^2}{R^2}\,, \\ d_{iB} &= p_{T,i}^{-2}\,, \end{aligned} \tag{5.3}$$

with $R$ being the cone radius (in $\eta$-$\phi$-space) of the resulting jet. A value of $R = 0.4$ is used in this analysis. The smallest distance measure of all $d_{ij}$ and $d_{iB}$ is identified and if it is a $d_{ij}$ the two corresponding particles are merged by adding their four-vectors and the distances are recalculated. If it is a $d_{iB}$, particle $i$ is defined to be a jet and removed from the pool of particles. This procedure is repeated until no particles are left. For energetic particles without another energetic particle within a distance of $2R$, a perfectly conical jet is built while the shape can differ from a perfect cone otherwise.

A calibration of the jet energy is performed as described in the following Section 5.4.2. Baseline jets are required to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 4.5$ and pass the *LooseBad* cleaning working point. All jets that additionally pass the overlap removal, detailed in Section 5.7, have to fulfil the requirements of the *tight* WPs of the *jet* and *forward jet vertex tagger* (JVT and fJVT) which were developed to further suppress pile-up jets. The JVT-quantity [192] exploits information of tracks matched to the jets to assign likelihoods whether they originate from the hard scatter vertex or not (pile-up). It is applied to jets with transverse momenta between 20 GeV and 60 GeV with $|\eta| < 2.4$. Forward jets with $2.5 < |\eta| < 4.5$ are judged by the fJVT-quantity [193] which cannot directly make use of track-based information since the ID only extends up to $|\eta| = 2.5$. Instead, it exploits momentum conservation at each reconstructed vertex. The *tight* JVT WP, corresponding to JVT $> 0.5$, has an average efficiency of 96 % [194] with an approximate pile-up fake rate of 3 %. The *tight* fJVT WP, corresponding to fJVT $< 0.4$, has an average efficiency below 76 % for hard-scatter jets and below 49 % for pile-up jets. In addition, JVT and fJVT scale factors are applied to simulated events. However, the fJVT calibration was not yet available for particle-flow jets when processing the events used in this analysis. Therefore, the fJVT decision and scale factors are taken from matched EM topo jets instead, which correspond to the former algorithm to obtain the jet constituents.

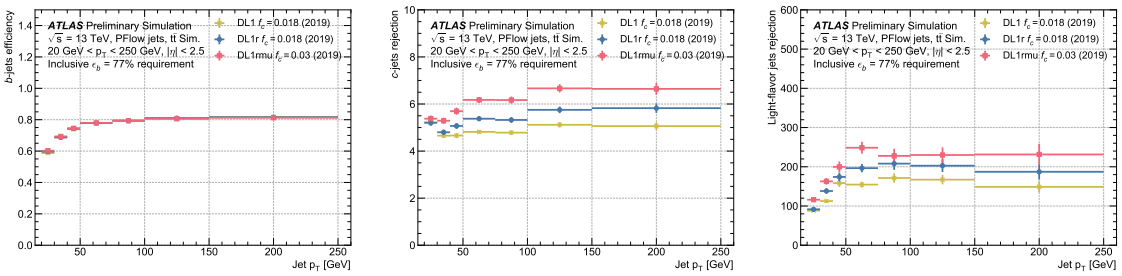
### 5.4.2 Jet Energy Calibration

Different energy scale corrections [195] are applied in the course of the jet reconstruction. First, the energy of the particle flow objects (tracks and topo-clusters) which are at the scale of the tracks is corrected such that it conforms with the electromagnetic scale of the cluster-only objects. Then, the jet energy scale (JES) of the final jet objects is restored in several steps by comparing the energy of the reconstructed jet to the one of a matched jet reconstructed at particle level (*truth* jet). The first step corrects for any remaining pile-up contamination. Then, an absolute calibration is performed, derived from simulated dijet-events. It corrects for effects due to different calorimeter technologies and varying granularity, for energy loss in passive material and for out-of-cone effects. Thereby, the jet four-momentum is aligned to the one of the respective *truth* jet by applying the inverse jet energy response as a function of the energy. Due to a remaining difference between the reconstructed and true pseudorapidity after this correction, a second correction, directly using the observed difference, is applied as a function of energy and pseudorapidity. This only affects $p_T$ and $\eta$ of the jet. However, the energy response also depends on the flvaour of the jet and therefore on its shower shape. The largest difference is observed between quark- and gluon-initiated jets. The *global sequential calibration* aims at reducing the dependence of the jet energy resolution on the flavour without changing the average jet energy response. It therefore applies six independent corrections, obtained from simulated events, to the

four-momenta of the jets. Each is derived from one of six observables describing the shower shape and has a different impact on the jet response. Finally, after the reconstructed jets are calibrated to particle level, an *in situ* calibration is applied, accounting for differences in the jet response between data and simulation. These differences originate from inaccurate descriptions of the detector material and response and of the physics processes in the simulations. Hence, the jet response in simulation and data is derived separately and the ratio is applied as a correction to data.

### 5.4.3 *b*-Jet Identification

An efficient algorithm to identify jets initiated by *b*-flavour quarks (*b*-tagging) is crucial in this analysis to suppress events from top-quark production. The *DL1r* [196, 197] *b*-tagging algorithm utilised in this analysis is based on a deep feed-forward neural network. The input variables exploited by the network are on one hand $p_T$ and $\eta$ of the respective jet and on the other hand the outputs of several low-level *b*- and *c*-flavour jet tagging algorithms exploiting the special signatures of *b*- and *c*-flavour jets. *b*-flavour hadrons have an approximate lifetime of 1.5 ps [198] leading to a decay length of a few millimeters assuming a momentum of several tenth of GeV. That means larger impact parameters as well as a secondary vertex (and even a tertiary vertex if the *b*-hadron decays into a *c*-hadron) can be observed. This information combined with track and energy qantities and the decay topologies enter the different low-level taggers [196, 197]. The final high-level *DL1r* tagger, which combines the results of the low-level taggers, provides a three-dimensional output. Each output component corresponds to the probabilities for a jet to be a *b*-flavour jet, a *c*-flavour jet or a light-flavour jet, respectively. Hence, this algorithm can in principle be used to identify either of the jet flavours. However, it is used here for *b*-tagging and therefore the final discriminant is defined as $D = \ln(p_b/(f_c \cdot p_c + (1 - f_c) \cdot p_{\text{light}}))$ with $p$ being the different probabilities and $f_c = 0.018$ representing the *c*-flavour jet fraction used to tune whether a better *c*-flavour jet or a better light-flavour jet suppression is the goal. Depending on the cut value on $D$, four different working points with different *b*-tagging efficiencies are defined: 60 %, 70 %, 77 % and 85 %. The *b*-flavour jet efficiency and the *c*-flavour jet and light-flavour jet rejection for the 77 % working point are shown in Figure 5.8. In this analysis the 85 % WP is chosen to veto *b*-jets. Further, simulation-to-data correction factors (scale factors) are applied to all simulated events.




**(a)** **(b)** **(c)**

**Figure 5.8:** *b*-jet tagging efficiency (a), *c*-jet rejection (b) and light-flavour jet rejection (c) measured in simulated $t\bar{t}$ events as a function of $p_{\text{T}}$ for the 77 % working point. The rejections are defined as the inverse efficiency. The blue markers indicate the tagger used in this analysis [199].

## 5.5 Hadronically Decaying $\tau$-Leptons

$\tau$-leptons can either decay hadronically or leptonically as described in Section 1.3. The decay happens before reaching active detector material and hence $\tau$-leptons must be reconstructed through their decay products. They manifest as an electron or muon with additional $E_{\mathrm{T}}^{\mathrm{miss}}$ if the $\tau$-lepton decays leptonically, and as one or three tracks in the ID together with energy depositions in the calorimeters and $E_{\mathrm{T}}^{\mathrm{miss}}$ depending on whether one or three charged hadrons are among the decay products if the $\tau$-lepton decays hadronically.

In the following, the reconstruction and identification of the visible decay products $\tau_{\text{had-vis}}$ of hadronically decaying $\tau$-leptons is discussed [200]. Jets obtained from the anti-$k_T$-algorithm, described in Section 5.4.1, are the basis for the reconstruction of $\tau$-lepton candidates. They are calibrated using a local hadronic calibration [201] and are required to have $p_T > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$, reflecting the coverage of the ID, while candidates in the transition region, $1.37 < |\eta| < 1.52$, are rejected. The corresponding vertex is chosen as the vertex candidate with the largest momentum fraction from all tracks matched to the jet within $\Delta R < 0.2$. The tracks are required to have $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ and to pass several quality criteria. Further requirements are placed on the transverse and longitudinal impact parameters. The $\tau$-lepton candidate is defined to be massless, hence its momentum and energy are identical. A dedicated tau energy calibration is performed such that it agrees with the one at generator level (from simulations). The $\eta$- and $\phi$-coordinates are determined from the vectorial sum of all topo-clusters within $\Delta R = 0.2$.

In order to separate hadronic $\tau$-leptons from quark or gluon initiated jets, a dedicated algorithm based on a *recurrent neural network* (RNN) is deployed [202]; one for tau lepton decays with one track and one for decays with three tracks. It uses a mixture of low and high level input variables that describe track and cluster properties and was trained on simulated $\gamma^* \to \tau\tau$-events versus dijet events. Jets originating from $\tau_{\text{had-vis}}$ are typically more narrow compared to quark- or gluon-initiated jets due to boosted decay products and due to the lower number of hadrons and have either one or three tracks associated. The output of the RNN is a value between zero and one that is tuned to be flat in the transverse momentum of true hadronic $\tau$-leptons as well as in the average number of interactions per bunch crossing. The resulting rejection power for jets misidentified as hadronic $\tau$-leptons against the efficiency to select true hadronic taus is shown in Figure 5.9. In addition, four

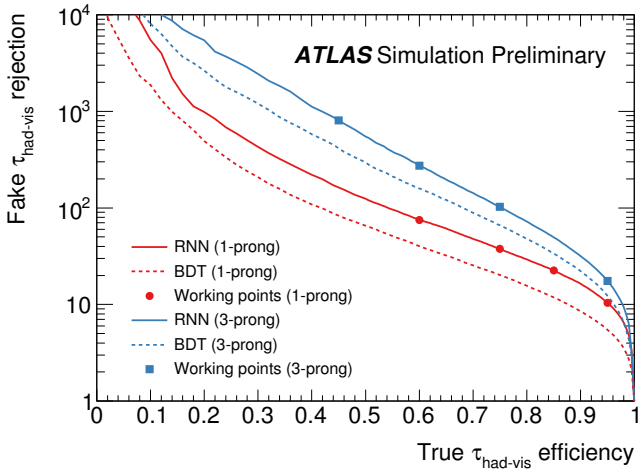



**Figure 5.9:** Rejection power for jets misidentified as hadronically decaying $\tau$-leptons against the efficiency to select true hadronic taus resulting from the RNN algorithm. In red, the curve for hadronic taus with one track are shown, in blue for hadronic taus with three tracks. The dashed lines indicate the curves of the previously used boosted decision tree algorithm [202].

different working points, *very loose*, *loose*, *medium* and *tight*, are marked, given in the order

of decreasing true $\tau_{\text{had-vis}}$ efficiency and increasing background rejection. In this analysis, the *medium* working point is required, corresponding to a signal efficiency of 75 %(60 %) and a background rejection of 35 (240) for hadronic taus with one (three) tracks.

Furthermore, *Boosted Decision Trees* are used to reject electrons that are mis-identified as hadronic $\tau$-lepton decays with one track. Three working points, *loose*, *medium* and *tight* are provided, listed in decreasing signal efficiency. The *loose* working point is utilised in this analysis, corresponding to a signal efficiency of 95 %. In addition, a muon veto criterion is applied. This selection defines the *baseline* hadronic tau leptons that enter the overlap removal procedure and that are used afterwards to veto events with hadronically decaying $\tau$-leptons.

## 5.6 Missing Transverse Energy

Weakly interacting particles like neutrinos travel through the detector and escape it without leaving a trace, and as such cannot be detected directly. However, the total momentum in the transverse plane is known to sum up to zero due to momentum conservation. The same does not hold for the longitudinal direction where the initial total momentum is unknown. Therefore, when assuming perfect resolution, the momentum of the system of the invisible particles in the transverse plane can be determined indirectly as *missing transverse momentum* ($\boldsymbol{p}_T^{\text{miss}}$) [203]:

$$\sum \boldsymbol{p}_T = \sum_{i=1}^{N_{\text{hard objects}}} \boldsymbol{p}_{T,i} + \sum_{j=1}^{N_{\text{soft term objects}}} \boldsymbol{p}_{T,j} + \boldsymbol{p}_T^{\text{miss}} = 0 \tag{5.4}$$

$$\Longrightarrow \quad \boldsymbol{p}_T^{\text{miss}} = -\sum_{i=1}^{N_{\text{hard objects}}} \boldsymbol{p}_{T,i} - \sum_{j=1}^{N_{\text{soft term objects}}} \boldsymbol{p}_{T,j}\,. \tag{5.5}$$

Here, the hard objects include all *baseline* muons, electrons, photons, hadronically decaying $\tau$-leptons and jets although in this analysis hadronically decaying $\tau$-leptons are rejected. An overlap removal as described in Section 5.7 between the objects is performed. The soft term comprises all charged signals from the primary vertex that are not associated with reconstructed objects. Its correct measurement is challenging since it is easily contaminated by signal from pile-up and underlying event. Here, a soft term calculation based on tracks only is utilised which is found to perform best (although contributions from neutral particles are neglected) since signal from out-of-time pile-up[2] can be suppressed. The tracks that are included in the soft term have to fulfil several track quality as well as track-to-vertex-association criteria. Only tracks matched to particle-flow objects that are not part of the hard objects are considered.

Usually, the absolute value of $\boldsymbol{p}_T^{\text{miss}}$ is used which is denoted with $E_{\text{T}}^{\text{miss}}$, the *missing transverse energy*. It presumes that the masses of the invisible particles can be neglected. In practice, $E_{\text{T}}^{\text{miss}}$ is also often used for the vector size $\boldsymbol{p}_T^{\text{miss}}$.

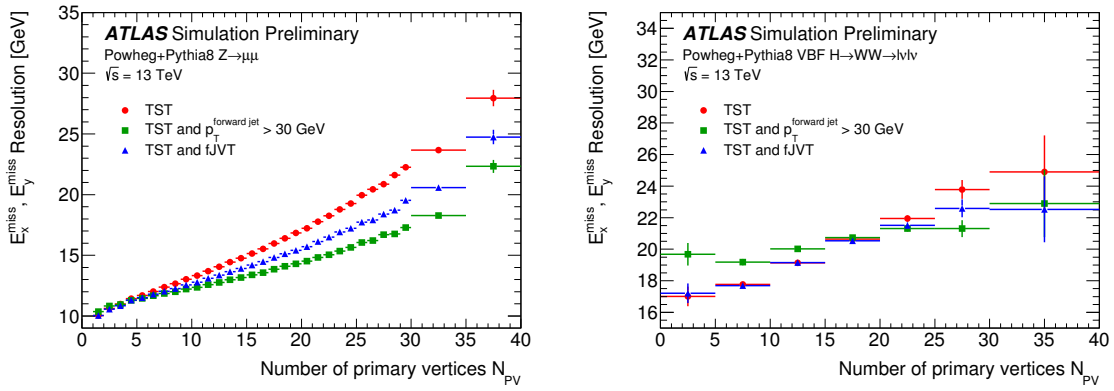
Two working points for the $E_{\text{T}}^{\text{miss}}$ reconstruction are defined based on the jet selection due to its large impact on the $E_{\text{T}}^{\text{miss}}$ performance. The *loose* working point includes all jets with $p_T > 20\,\text{GeV}$ in the hard objects term which in addition pass the JVT criteria when they have $|\eta| < 2.4$ and $p_T < 60\,\text{GeV}$. The *tight* working point excludes forward jets with $|\eta| > 2.4$ and $20\,\text{GeV} < p_T < 30\,\text{GeV}$ which results in an improved $E_{\text{T}}^{\text{miss}}$ resolution at high pile-up and in general a reduced pile-up dependency. The drawback is that also hard-scatter jets may be removed. In this analysis, the *tight* working point is chosen.

A non-zero $E_{\text{T}}^{\text{miss}}$ in events where no $E_{\text{T}}^{\text{miss}}$ is expected can arise from imprecisely

[2] signal from collisions in previous bunch crossings

measured momenta of the contributing objects, from miscalibration and from detector areas that are not or insufficiently instrumented as well as from pile-up.

Figure 5.10 shows the resolution of $E_{\mathrm{T}}^{\mathrm{miss}}$[3] with respect to the number of primary vertices for different working points and different processes. The resolution of both processes clearly shows the dependency on the number of primary vertices and also its reduction for the *tight* compared to the *loose* working point. Especially for the VBF $H \to WW$-events that have more hard scatter forward jets, it is visible that the resolution for the *tight* working point is worse for a low number of primary vertices. No true $E_{\mathrm{T}}^{\mathrm{miss}}$ is expected for $Z \to \mu\mu$-events. The resolution ranges from 10 GeV for a low number of primary vertices to 25 GeV.




**(a)** **(b)**

**Figure 5.10:** The resolution of $E_{\mathrm{T}}^{\mathrm{miss}}$ with respect to the number of primary vertices, determined with simulated $Z \to \mu\mu$- (a) and VBF $HWW$-events (b). Red corresponds to the *loose*, green to the *tight* working point [203].

## 5.7 Overlap Removal

Reconstructed and identified objects are not necessarily exclusive but can have shared hits or tracks in the ID or shared energy clusters in the calorimeters. Typically, if one object overlaps another within a certain $\Delta R$-cone one of them is removed following a pre-defined scheme which considers the probablities of each object type to be mis-identified as another. Hereby, the *standard* working point of the *AssociationUtils* package [204] is exerted. The details are summarised in Table 5.1. The removal is performed in the order given in the Table. For example, after the electron-electron decision based on the size of their transverse momenta, a hadronic $\tau$-lepton is rejected in favour of an electron if the $\tau$-lepton is within a cone of $\Delta R = 2$ around the electron. The overlap removal is performed between all *baseline* objects, before the missing transverse energy is calculated, the $b$-tagging is performed and hadronic $\tau$-leptons are vetoed.

[3] the previous EM topo (electromagnetic topological) jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ definitions were used here instead of the newer particle-flow definitions

| Reject | Against | Criteria |
|---|---|---|
| electron | electron | shared track, $\mathrm{p}_T^1 < \mathrm{p}_T^2$ |
| $\tau_{\mathrm{had}}$ | electron | $\Delta R < 0.2$ |
| $\tau_{\mathrm{had}}$ | muon | $\Delta R < 0.2$ |
| muon | electron | is CT muon and shared track |
| electron | muon | shared ID track |
| jet | electron | $\Delta R < 0.2$ |
| electron | jet | $\Delta R < 0.4$ |
| jet | muon | Number of tracks $< 3$ and ghost-associated or $\Delta R < 0.2$ |
| muon | jet | $\Delta R < 0.4$ |
| jet | $\tau_{\mathrm{had}}$ | $\Delta R < 0.2$ |

**Table 5.1:** The overlap removal criteria applied to all *baseline* objects in the order given in the table. The criteria correspond to the *standard* working point of the *AssociationUtils* package [204]. The criterium "shared track" removes electrons that share an ID track with muons, but vetoes overlapping CT (calorimeter-tagged) muons. The term "ghost-associated" means that the muon ID track is associated to the jet.

# Chapter 6

# Introduction to Neural Networks for Signal and Background Separation

One crucial component in the search for small[1] signals is the ability to obtain the best possible separation of signal from the expected background contributions. There are various tools available to tackle this challenge. For this analysis, it was decided to use *Fully Connected Feed Forward Neural Network*s (NNs). Several reasons led to this decision:

- in general, NNs can approximate any continuous function up to any desired accuracy and hence, in principle, also the *Bayes Optimal Classifier* [205],
- the information of many input features can be mapped onto a lower-dimensional observable,
- easy construction of multiclass classifiers by design,
- possibility to exploit unknown underlying correlations between the input features,
- in connection with the previous point, the usage of mainly low-level input features is promising.

Put simply, neural networks exhibit more potential for experimentation compared to e.g. *Boosted Decision Tree*s (BDTs). The basics of NNs are explained in Section 6.1, followed by an introduction of $k$-fold cross validation to enhance training statistics in Section 6.2. Section 6.3 describes the preprocessing and selection of input features as well as a method to obtain an importance ranking of the input features. The chapter ends with a discussion of the optimisation of the *hyperparameters* which define the architecture and the training process of an NN in Section 6.4.

This chapter gives a general overview of NNs for classification, i.e. in the context of signal and background separation. The details of the implementation specific for the analysis presented in this thesis as well as different studies that were performed are given in Section 8.2.

## 6.1 Feedforward Neural Networks

*Artificial Neural Network*s (ANNs), also called *Neural Network*s (NNs), are inspired by our understanding of the structure of the human or animal brain: artificial neurons, called *nodes*, are connected via artificial synapsis. The aim of an NN is to map the information of many input features $x_i$ (describing a single element or event of the dataset) onto a

[1] with respect to the expected background contributions

lower-dimensional observable which allows for obtaining simplified, reduced and condensed information hidden in the complexity of the input features. In NNs the nodes are typically arranged in layers, where each layer consists of a certain number of nodes. The nodes of successive layers are connected to each other. A basic example of such an NN where the information is only passed forward (*Feedforward Neural Network*) is depicted in Figure 6.1. The example network has three layers of nodes of which the first layer, shown in green, has

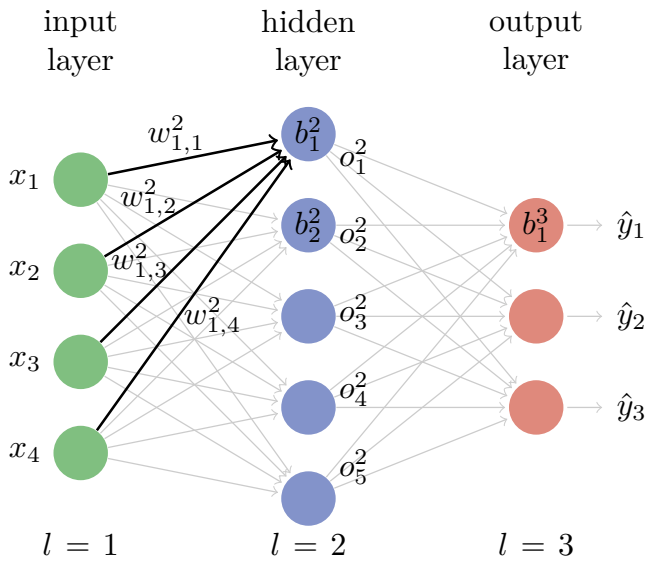



**Figure 6.1:** Sketch of a neural network architecture with an input layer (green), one hidden layer (blue) and an output layer (red). Four input features (four nodes in input layer) of each event are used to classify it into one of three classes (three nodes in output layer).

four nodes corresponding to four input features, one hidden layer (blue) with five nodes and an output layer (red) with three nodes. In the analysis presented in this thesis, NNs are used for classification, i.e. the categorisation of events into different classes. Therefore, all further description focuses on classification tasks and partially follows the presentation in [206]. The three output nodes in the example correspond to three individual classes, e.g. signal, background 1 and background 2. The number of hidden layers and the number of nodes per layer have impact on the categorisation performance of the NN and as such are subject to manual or automatic tuning (Section 6.4). In contrast to that, the number of nodes in the input and output layers are given by the task at hand. Details on how to choose appropriate input features for the input layer are given in Section 6.3.

### 6.1.1 Notation

The output values $\hat{y}_j$, corresponding to the different classes $j$, are a function of the input features $x_i$ and the parameters $w^l_{j,i}$ and $b^l_j$, called weights and biases where $l$ denotes the corresponding layer (not to be interpreted as an exponent). For example, the output value $o^2_1$ of node 1 in layer 2 (see Figure 6.1) is given by

$$o^2_1 = \sigma\left(\sum_i \left(w^2_{1,i} x_i\right) + b^2_1\right) , \tag{6.1}$$

with $\sigma$ being the so-called *activation function*. The output $o^2_1$ subsequently becomes the input to all nodes in the next layer. Therefore, the output $\hat{y}_1 = o^3_1$ of output node 1 can be

obtained from

$$\hat{y}_1 = o_1^3 = \sigma\left(\sum_i \left(w_{1,i}^3 o_i^2\right) + b_1^3\right) , \tag{6.2}$$

which depends on the previous $o_1^2$ and therefore on the $x_i$. More generally, the output of node $j$ in layer $l$ can be written as

$$o_j^l = \sigma\left(\sum_i \left(w_{j,i}^l o_i^{l-1}\right) + b_j^l\right) , \tag{6.3}$$

where $i$ denotes node $i$ in layer $l-1$. A more compact notation can be achieved by absorbing the indices $i$ and $j$ into a vector notation. The vector of all outputs $\boldsymbol{o}^l$ in layer $l$ can be written as

$$\boldsymbol{o}^l = \sigma\left(W^l \boldsymbol{o}^{l-1} + \boldsymbol{b}^l\right) = \sigma\left(\boldsymbol{z}^l\right) , \tag{6.4}$$

where $w_{j,i}^l$ is the $j,i$-th element of the weight matrix $W^l$ in layer $l$ and $\boldsymbol{o}^{2-1} = \boldsymbol{x}$ are the input features. Additionally, the argument of the activation function $\sigma$, also called *weighted input*, is typically defined as $z_j^l$ for node $j$ in layer $l$ as given in the second part of Eq. 6.4.

Such an NN can, in general, approximate any continuous function up to any desired accuracy. This is qualitatively visualised in chapter 4 of [206].

### 6.1.2 Activation Functions

Various functions qualify for use as an activation function. For classification, a common choice for the activation function $\sigma\left(\boldsymbol{z}^L\right)$ of the last layer $L$ (the output layer) is the *softmax function*. For output node $j$ which is equivalent to class $j$ it is defined by

$$\hat{y}_j = \sigma_j\left(\boldsymbol{z}^L\right) = \frac{e^{z_j^L}}{\sum_{k=1}^{Cl} e^{z_k^L}} , \tag{6.5}$$

which depends on the $z_k^L$ of all classes with $Cl$ denoting the number of classes. The sum over all classes adds up to 1,

$$\sum_{j=1}^{Cl} \sigma_j(\boldsymbol{z}^L) = \sum_{j=1}^{Cl} \hat{y}_j = 1 , \tag{6.6}$$

and all $\sigma_j$ are between 0 and 1 which allows for interpreting the individual $\sigma_j$ as probabilities. That means $\hat{y}_j$ gives the probability for an event with input feature vector $\boldsymbol{x}$ to belong to class $j$. An exemplary visualisation of the *softmax activation function* for three output classes is shown in Figure 6.2. It illustrates that the sum of the three curves adds up to 1 for each value of $z_3$. Each curve represents the output probability of one of the three classes as a function of $z_3$ with fixed $z_1$ and $z_2$. While $\sigma_3$ increases with increasing $z_3$ the other two decrease, consequently.

The activation function associated with the hidden layers is subject to the optimisation procedure described in Section 6.4.1. Possible functions among many more are the *sigmoid function* and the *(Leaky) ReLU function* (Rectified Linear Unit) [207, 208]. The sigmoid function is defined by

$$\sigma_j\left(z_j^l\right) = \frac{1}{1 + e^{z_j^l}} , \tag{6.7}$$

and is visualised in Figure 6.3(a) together with its gradient. The shape of the sigmoid

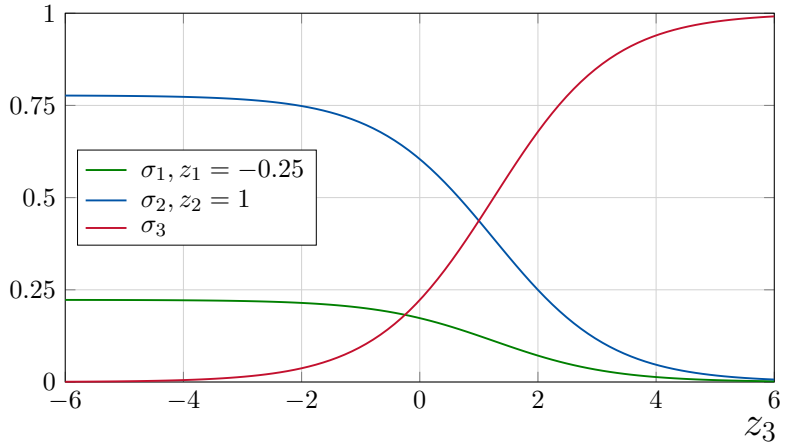



**Figure 6.2:** The activations $\sigma_1, \sigma_2$ and $\sigma_3$ of three output classes are plotted against the free variable $z_3$ which is the weighted input of the third class. The weighted inputs of the other two classes $z_1$ and $z_2$ are fixed to the arbitrary values of $z_1 = -0.025$ and $z_2 = 1$. Adapted from [206].

function is comparable to the softmax activation function. It maps values in the range $(-\infty, \infty)$ onto $(0, 1)$. Its derivative, given by $\sigma(z)(1-\sigma(z))$, is largest for $z = 0$ and vanishes for increasing $|z|$. The (leaky) ReLU function is defined by

$$\sigma_j\left(z_j^l\right) = \begin{cases} z_j^l & \text{if } z_j^l > 0 \\ \alpha \cdot z_j^l & \text{otherwise} \end{cases} \tag{6.8}$$

and its distribution and gradient are shown in Figure 6.3(b). For $\alpha = 0$, it is denoted by *ReLU function* only and maps values between $(-\infty, \infty)$ onto $[0, \infty)$ and has a slope of 1 for $z_j^l$ larger than 0 and a value and slope of 0 for $z_j^l \leq 0$. For $\alpha > 0$, the function is called *leaky ReLU function*. It has a small slope $\alpha$ below $z_j^l = 0$ which allows for a small positive gradient in this range. In this case the values are mapped onto a range of $(-\infty, \infty)$. The function value of the (leaky) ReLU function as well as its derivative are easy and fast to compute. All discussed activation functions are not linear, as otherwise the NN could only model linear functions. The discussion of the importance of the gradients of the activation functions as well as of the advantages and disadvantages of both activation functions is part of the following Section 6.1.3.

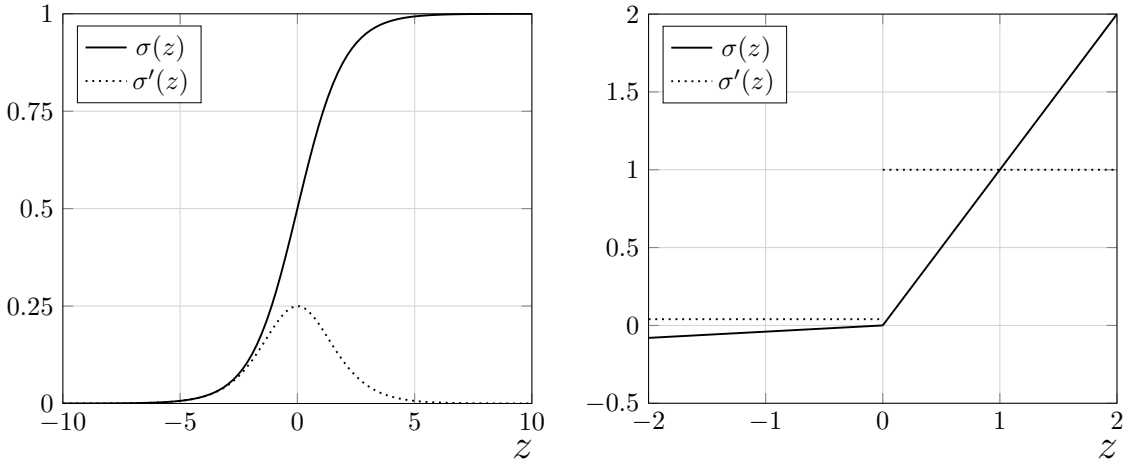



**Figure 6.3:** Two possible activation functions (solid lines) together with their derivatives (dotted lines): sigmoid activation function (a) and leaky ReLU activation function (b).

### 6.1.3 Training of a Neural Network

In order to apply an NN to classify events, appropriate values for the weights $w^l_{j,i}$ and biases $b^l_j$ must be found first. The process of finding the weights and biases is called *training*. To perform the training, the network is fed with the values of the input features $\boldsymbol{x}$ of each *training event* for which also the true class it belongs to is known (supervised learning). For example, $\boldsymbol{y} = (0, 0, 1)$ indicates the event belongs to class 3. The cost function used for this optimisation problem is introduced in the next section.

#### Cost Function

The optimal values of the weights and biases are found by minimising the cost (or loss) function $C$ with respect to all $w^l_{j,i}$ and $b^l_j$, hereby averaging over all $N$ training events:

$$C = \frac{1}{N} \sum_{n=1}^{N} C_n \,, \tag{6.9}$$

with $C_n$ being the cost of event $n$. Formulating the cost function as an average is crucial for the *Backpropagation Algorithm* which is discussed later. A typical choice for the cost in classification problems is the *categorical cross-entropy*. Its definition is given as

$$C = \frac{1}{N} \sum_{n=1}^{N} \sum_{c=1}^{Cl} -y_{n,c} \cdot \log \hat{y}_{n,c} \,, \tag{6.10}$$

with $N$ being the number of training events, $Cl$ the number of classes, $y_{n,c}$ being 0 or 1 depending on whether the event $n$ belongs to class $c$ (1) or not (0) and $\hat{y}_{n,c}$ being the predicted probability for event $n$ to be in class $c$. This means that if $y_{n,c}$ is 0, the corresponding $n, c$-term in $C$ is 0 as well and therefore does not contribute. Hence, only the $\hat{y}_n$ values that correspond to the class $c_{\text{true}}$ to which the event is known to belong to ($y_{n,c_{\text{true}}} = 1$) contribute to the cost function. This reduces the cost function to

$$C = \frac{1}{N} \sum_{n=1}^{N} -\log \hat{y}_{n,c_{\text{true}}} \,, \tag{6.11}$$

with the subscript $c_{\text{true}}$ indicating that only the $\hat{y}_n$ of the corresponding true class remain. Figure 6.4 shows the term $-\log \hat{y}$ plotted against $\hat{y}$. Due to the softmax activation function in the last layer, the range of $\hat{y}$ is $(0, 1)$. It shows that if the predicted $\hat{y}_{n,c_{\text{true}}}$ of an event $n$ is equal to the true value $y_{n,c_{\text{true}}}$ ($y_{n,c_{\text{true}}} = 1$), the corresponding summand in the cost function yields $0 = -\log(1)$ and thus minimises the cost function. The further the predicted probability $\hat{y}_{n,c_{\text{true}}}$ moves away from the ideal value 1, the larger the cost gets; the penalty increases exponentially. Thus, the categorical cross-entropy is a suitable cost function for classification.

#### Minimisation of the Cost Function

The cost function is a function of many parameters ($w^l_{j,i}$ and $b^l_j$)[2] and hence it is typically not possible to minimise the cost function analytically. Therefore, an iterative optimisation algorithm called *Gradient Descent* [209, 210] is used to find a local minimum. The gradient of the cost function with respect to all $w^l_{j,i}$ and $b^l_j$ at the current parameter space point is

[2]The actual number of parameters is $\sum_{l=1}^{L-1} \mathcal{N}^l \cdot \mathcal{N}^{l+1} + \mathcal{N}^{l+1}$ with $L$ being the number of layers and $\mathcal{N}^l$ the number of nodes in layer $l$.

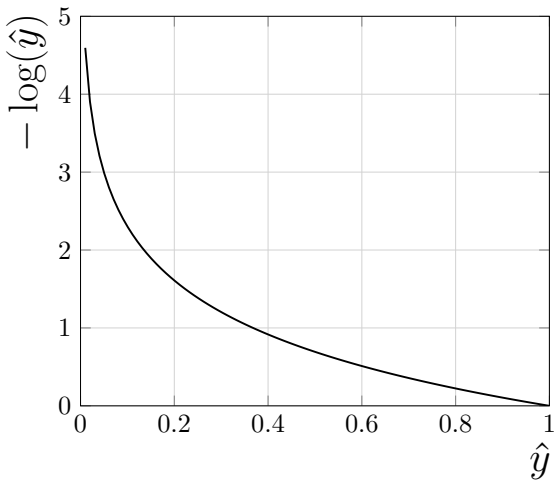



**Figure 6.4:** The distribution of the relevant part of the categorical cross-entropy cost function with respect to the predicted probability $\hat{y}$ of the corresponding true class.

calculated[3] and the parameter values are updated in the direction of the negative gradient which indicates the steepest descent. This can be expressed with the following update rules for the parameters, i.e. for the weights and biases:

$$w^l_{j,i} \longrightarrow w^{l\,\prime}_{j,i} = w^l_{j,i} - \eta \frac{\partial C}{\partial w^l_{j,i}} = w^l_{j,i} - \frac{\eta}{N} \sum_{n=1}^{N} \frac{\partial C_n}{\partial w^l_{j,i}} \,, \tag{6.12}$$

$$b^l_j \longrightarrow b^{l\,\prime}_j = b^l_j - \eta \frac{\partial C}{\partial b^l_j} \qquad = b^l_j - \frac{\eta}{N} \sum_{n=1}^{N} \frac{\partial C_n}{\partial b^l_j} \,. \tag{6.13}$$

The updated values of the parameters are indicated by $w^{l\,\prime}_{j,i}$ and $b^{l\,\prime}_j$ while $\frac{\partial C}{\partial p} (p = w^l_{j,i} \text{ or } b^l_j)$ denotes the gradient of the cost function with respect to the respective parameter $p$. When replacing $C$ by its average over all training events, as defined in Eq. 6.9, the expressions on the right hand side are obatined. The length of the step that is taken for each parameter in the direction of the negative gradient is represented by $\eta$, also called *learning rate*. The choice of the value of $\eta$ is very decisive for the training success of the network: if it is chosen too small the network only learns very slowly, if it is chosen too large it may lead to an increase of the cost function value instead of a decrease. The learning rate is one of the hyperparameters and as such part of the hyperparameter optimisation procedure described in Section 6.4.

Evidently, many gradients (for each weight and bias and for each training event) need to be computed, which is time-consuming. Therefore, in order to decrease training time, *Stochastic Gradient Descent* (SGD) is often used instead. The gradient is calculated only based on a random subset of size $M$ (also called *batch*) of all $N$ training events, i.e. $N$ is replaced by $M$ in Eqs. 6.12 and 6.13. This speeds up the calculation of the gradients for each iteration and more updates of the weights and biases can be achieved in the same time. If $M$ is large enough, the gradient obtained from the subset is assumed to be a good approximation of the true gradient. Small fluctuations in the update history are acceptable since it is more important to move in the approximately correct direction within reasonable time to decrease $C$ than to know the exact gradient in each iteration. After the iteration based on the first subset, another iteration based on the next (stochastically independent) subset of same size is performed until each training event was used once, which corresponds

[3]This implies that the cost function must be differentiable with respect to the weights and biases which is true for the categorical cross entropy.

to one *epoch*. The training is finished for a pre-defined number of epochs. If the usual gradient descent instead of SGD were used ($M = N$), one update per epoch would be performed. Otherwise, $N/M$ iterations per epoch are performed[4]. There are various factors that impact the appropriate choice of $M$: as mentioned before it should be large enough such that the fluctuations remain within reasonable limits, but it should also be small enough to gain from reduced computation times. Importantly, events of all classes should be represented in each batch and the sum of weights in the batch for each class should be positive (see Section 6.3.1 for details). Because the batch size $M$ determines how often the weights and biases are updated per epoch, the choice of $M$ directly impacts the choice of the number of epochs if the total number of iterations should be kept constant.

Although SGD provides a substantial improvement in training performance, the computation of the gradients remains costly and hence, a fast algorithm to obtain the gradients is crucial to perform the training of complex neural networks in a reasonable amount of time. The *Backpropagation Algorithm* is used to perform this task and it is introduced in the following paragraph.

### Backpropagation Algorithm

The backpropagation algorithm [211, 212] consists of four equations which allow the calculation of the gradients with respect to the weights and biases in all layers and nodes. The four equations are given in Eqs. 6.14–6.17 [206] and are motivated afterwards followed by an explanation of how this is applied to train an NN.

$$\frac{\partial C_n}{\partial b_j^l} = \delta_j^l \tag{6.14}$$

$$\frac{\partial C_n}{\partial w_{j,i}^l} = o_i^{l-1}\delta_j^l \tag{6.15}$$

$$\delta_j^l = \sum_i w_{i,j}^{l+1}\delta_i^{l+1}\sigma'\left(z_j^l\right) \rightarrow \boldsymbol{\delta}^l = \left(\left(W^{l+1}\right)^T \boldsymbol{\delta}^{l+1}\right) \odot \boldsymbol{\sigma}'\left(\boldsymbol{z}^l\right) \tag{6.16}$$

$$\delta_j^L = \frac{\partial C_n}{\partial o_j^L}\sigma'\left(z_j^L\right) \qquad \rightarrow \boldsymbol{\delta}^L = \boldsymbol{\nabla}_o C_n \odot \boldsymbol{\sigma}'\left(\boldsymbol{z}^L\right) \tag{6.17}$$

The quantity $\delta_j^l$ denotes the *error* of node $j$ in layer $l$, i.e. it quantifies by how much the cost changes if the weighted input $z_j^l$ to neuron $j$ in layer $l$ is changed by a small amount. Therefore, $\delta_j^l$ is defined as [206]:

$$\delta_j^l := \frac{\partial C_n}{\partial z_j^l}\,, \tag{6.18}$$

and Eqs. 6.14 and 6.15 can be obtained by combining the definition of $z_j^l$ given in Eq. 6.4, the chain rule and the definition of $\delta_j^l$. Eq. 6.16 is an instruction on how to backpropagate the error through the network from layer $l+1$ to layer $l$. First, the transposed weight matrix is applied to move the error backwards to obtain its impact on the output of the previous layer $l$. Then, the derivative of the activation function is applied to move it further backwards through the activation function of layer $l$ to obtain the error of the weighted input $z^l$ to layer $l$ [206]. Finally, Eq. 6.17 defines the error in the last layer $L$ which is needed as input to the previous Eq. 6.16. Both Eqs. 6.16 and 6.17 can, again, be proofen by applying the chain rule. The formulas in Eqs. 6.16 and 6.17 are also given in vector

[4] If $N$ is not divisible by $M$, the number of iterations is $\lceil N/M \rceil$ with the last batch being smaller than the previous ones.

notation for better readability where the symbol $\odot$ indicates the element-wise product and $\boldsymbol{\nabla}_o$ is a vector where each component is the partial derivative with respect to $o_j^L$.

The steps needed to apply backpropagation in the course of gradient descent during the training of a neural network are detailed below. Steps one to four are performed for each training event in the current batch. Their results are combined within step five in order to obtain the updated values for the weights and biases from this batch.

1. The output $\boldsymbol{o}^l = \sigma\left(\boldsymbol{z}^l\right) = \sigma\left(W^l \boldsymbol{o}^{l-1} + \boldsymbol{b}^l\right)$ is calculated for each layer $l \in [2, 3, ..., L]$ in a feedforward manner, starting with $\boldsymbol{o}^{2-1} = \boldsymbol{x}$ being the input features.

2. The vector of output errors $\boldsymbol{\delta}^L$ of the last layer is calculated using Eq. 6.17.

3. The error is backpropagated into the previous layers by evaluating Eq. 6.16 for each $l \in [L-1, L-2, ..., 2]$ and hence all $\boldsymbol{\delta}^l$ are obtained.

4. The derivatives of the cost function with respect to all weights and biases ($\partial C_n / \partial w_{j,i}^l$ and $\partial C_n / \partial b_j^l$) are calculated using Eqs. 6.14 and 6.15, respectively.

5. Finally, the results of all training events from the previous step are combined according to Eqs. 6.12 and 6.13 to obtain the updates for the weights and biases from the current batch.

In conclusion, the back propagation algorithm allows updating the weights and biases without explicitly calculating the derivatives of the cost function with respect to them. Only the derivatives of the cost function with respect to the final outputs $o_j^L = \hat{y}_j$ and the derivatives $\sigma'(z_j^l)$ of the activation function with respect to the $z_j^l$ are needed, which both are easy to compute. Therefore, the properties of the derivative of the activation function $\sigma$ are of high importance with respect to the performance of the backpropagation algorithm. For instance, the derivative of the sigmoid activation function in Figure 6.3(a) shows that the overall small value of its derivative, which is at most 0.25 and decreases further due to continuing multiplication across the layers (see Eq. 6.16) can be a disadvantage. Furthermore, the derivative almost vanishes for values of $|z|$ larger than $\sim 5$, for which it is 0.007, which leads to a slow learning progress. The ReLU activation does not show these problems, as the gradient is always either 1 or 0. One disadvantage of ReLU can be the unlimited value of its output. Additionally, the gradient of 0 for $z < 0$ can lead to *dying nodes* (i.e. if $z < 0$ for all inputs, the gradient is always zero and hence no learning occurs) which can be undesired. In order to overcome this problem, leaky ReLU was introduced with a small positive slope $\alpha$ below 0 which leads to a small but non-zero gradient.

### Alternative Optimiser

Other optimisers instead of *(Steepest) Gradient Descent* can be chosen for minimising the cost function. The *Adam* (Adaptive moment estimation) optimiser [213] is tested as an alternative in this work. It is an extension of *Gradient Descent*, incorporating first- and second-order moments of the gradients of the loss function with respect to the parameters $w_{j,i}^l$ and $b_j^l$. The update rules for the weights and biases (Eqs. 6.12 and 6.13) are adapted to the following terms, where $w_{j,i}^l$ and $b_j^l$ are replaced by a common $p$:

$$p_{t-1} \longrightarrow p_t = p_{t-1} - \eta \frac{\hat{m}_{p,t}}{\sqrt{\hat{v}_{p,t}} + \epsilon} \tag{6.19}$$

with

$$\hat{m}_{p,t} = \frac{m_{p,t}}{1 - (\beta_1)^t} \quad \text{and} \quad m_{p,t} = \beta_1 \cdot m_{p,t-1} + (1 - \beta_1) \cdot \frac{\partial C}{\partial p} \tag{6.20}$$

and

$$\hat{v}_{p,t} = \frac{v_{p,t}}{1-(\beta_2)^t} \quad \text{and} \quad v_t = \beta_2 \cdot v_{p,t-1} + (1-\beta_2) \cdot \left(\frac{\partial C}{\partial p}\right)^2 \tag{6.21}$$

with $t$ indicating the $t$-th iteration or time step. A new *batch* for evaluation is used at each step $t$. $m_{p,t}$ and $v_{p,t}$ are called running averages of the first- and second-order moments of the gradient, respectively. At time step $t = 0$, the $m_{p,0}$ and $v_{p,0}$ are initialised with 0 which leads to a bias of the running averages towards zero. This is circumvented by a correction that results in the terms $\hat{m}_{p,t}$ and $\hat{v}_{p,t}$. The parameters $\beta_1$ and $\beta_2$ steer the averaging time of the two moments, respectively and are $\in [0, 1)$. The small parameter $\epsilon$ prevents division by zero and is suggested in reference [213] to be set to $10^{-8}$. In reference [213] it is further recommended that $\beta_1$ and $\beta_2$ are set to 0.9 and 0.999, respectively.

*Adam* does not just take a step into the direction of steepest descent but rather takes the momentum of past iterations into account by averaging over the previous gradients. This can lead to a smoother path in the parameter space of all $\boldsymbol{p}$. By replacing $\hat{v}_{p,t}$ in Eq. 6.19 with the help of the variance $\sigma^2_{p,t} = \hat{v}_{p,t} - \hat{m}^2_{p,t}$ [214], one obtains useful insights into the *Adam* update rule. Eq. 6.19 becomes:

$$p_{t-1} \longrightarrow p_t = p_{t-1} - \eta \frac{\hat{m}_{p,t}}{\sqrt{\sigma^2_{p,t} + \hat{m}^2_{p,t}} + \epsilon}\,. \tag{6.22}$$

Assuming the variance of the gradient over the last steps is small, i.e. $\sigma^2_{p,t} \ll \hat{m}^2_{p,t}$, the update rule becomes $p_{p,t-1} \longrightarrow p_t = p_{t-1} - \eta \cdot \mathrm{sgn}(\hat{m}_{p,t})$. When, however, the variance is large, i.e. $\sigma^2_{p,t} \gg \hat{m}^2_{p,t}$, the update rule becomes $p_{t-1} \longrightarrow p_t = p_{t-1} - \eta \frac{\hat{m}_{p,t}}{|\sigma_{p,t}|}$ with $\left|\frac{\hat{m}_{p,t}}{\sigma_{p,t}}\right|$ being small. Hence, the learning rate is effectively adapted individually for each parameter $p$ and is bound by the size of $\eta$.

### 6.1.4 Regularisation of Neural Networks

A common concern in NN training is *overtraining* which describes the state where the network is not able to generalise to unseen data but instead learns specific features exclusive to the training dataset. This is especially an issue in networks with a large number of free parameters (weights and biases) combined with limited training statistics. One option to prevent overtraining is *L2 weight regularisation*, which is used in this analysis. The key idea is to add an additional term which is proportional to the sum of all weights to the loss function [206] which penalises large values of the weights $w^l_{j,i}$ in the network:

$$C = C_0 + \frac{\lambda}{2N} \sum_{i,j,l} \left(w^l_{j,i}\right)^2. \tag{6.23}$$

The parameter $\lambda$, called *regularisation strength*, steers the impact of the additional term with respect to the original loss function $C_0$ given in Eq. 6.10. A larger value $\lambda$ corresponds to a stronger regularisation, meaning that the training will be more focused on keeping the weights small compared to reducing $C_0$. With the regularisation term included, the update rule for the weights (Eq. 6.24) becomes:

$$w^l_{j,i} \longrightarrow w^l_{j,i} - \frac{\eta}{N} \sum_{n=1}^{N} \frac{\partial C_{0,n}}{\partial w^l_{j,i}} - \eta \frac{\partial \left(\frac{\lambda}{2N} \sum_w w^2\right)}{\partial w^l_{j,i}} = w^l_{j,i} - \frac{\eta\lambda}{N} w^l_{j,i} - \frac{\eta}{N} \sum_{n=1}^{N} \frac{\partial C_{0,n}}{\partial w^l_{j,i}}. \tag{6.24}$$

From the adapted update rule it is apparent that the weights are reduced by a factor $(1 - \eta\lambda/N)$, although the last term can increase the weight's value if this is beneficial for

the reduction of the overall cost $C$. A qualitative explanation of why smaller weights are advantageous for better generalisation is the following: smaller weights also cause a smaller change in $z_j = w_{j,i}x_i + b_j$ (and hence a smaller change in the outputs) for a small change in an input feature $x_i$ due to e.g. noise. If, in contrast, $w_{j,i}$ is large, $z_j$ also becomes larger even if the change in $x_i$ is very small. Therefore, the network is prone to pick up insignificant specific features of the training dataset. The parameter $\lambda$ is an additional *hyperparameter* and must be tuned for each classification task.

### 6.1.5 Metrics for the Performance of Neural Networks

Various metrics are available for judging the performance of neural networks. The ones made use of in this analysis are introduced in the following.

#### Loss

The value of the previously introduced loss function (Eq. 6.10) is an obvious metric, where smaller values indicate better performance. The evolution of the loss over the epochs of an NN-training is called *learning curve* and its shape gives insight into the learning progress.

#### Accuracy

In classification, an additional metric that can be monitored during the training process is the *accuracy* $A$. It is defined as:

$$A = \frac{1}{N}\sum_{n=1}^{N} \delta_{\arg\max_c\{y_{n,c}\},\arg\max_c\{\hat{y}_{n,c}\}}\,, \tag{6.25}$$

with $\delta$ being the Kronecker delta. The sum goes over all events, $c$ indicates the respective class while $y$ denotes the true value whether the event belongs to class $c$ ($y_{n,c} = 1$) or not ($y_{n,c} = 0$). The predicted value $\hat{y}$ can be any value between 0 and 1 representing the probability for event $n$ to belong to class $c$. The classes with the highest values for $y_n$ and $\hat{y}_n$ are obtained via the arg max operation. If the true class and the predicted class for an event coincide, the value of the summand of event $n$ is 1, otherwise it is 0. Thus, the accuracy is the ratio of events which are correctly classified over all events.

#### Area-Under-Curve

Another metric, typically only evaluated after training, is the *area under the ROC-curve* (AUC), where ROC stands for *receiver operating characteristic* [215]. It is derived from the individual output node probability distributions, i.e. for each class separately and is a measure for the discrimination power of the probability distributions. The curve is obtained by scanning the probability distribution $\hat{y}_{\text{signal}}$ in small steps (at different cut values) from right to left (from large to small values), each time calculating the *true-positive rate* (TPR) and the *false-positive rate* (FPR). The TPR is plotted versus the FPR, thus the ROC-curve evolves from left to right. This is visualised in Figure 6.5. The TPR (signal acceptance) is the ratio of the signal that is classified as signal over the total signal while the FPR (1 - background rejection) is the background that is classified as signal over the total background. In this context, signal corresponds to the class of the respective output node and background corresponds to the sum of all other classes.

A good separation of signal and background and hence a good performance of the network is achieved when the TPR is large while keeping the FPR small. Ideally, the ROC-curve would extend up into the top left corner which is equivalent to an AUC of 1.

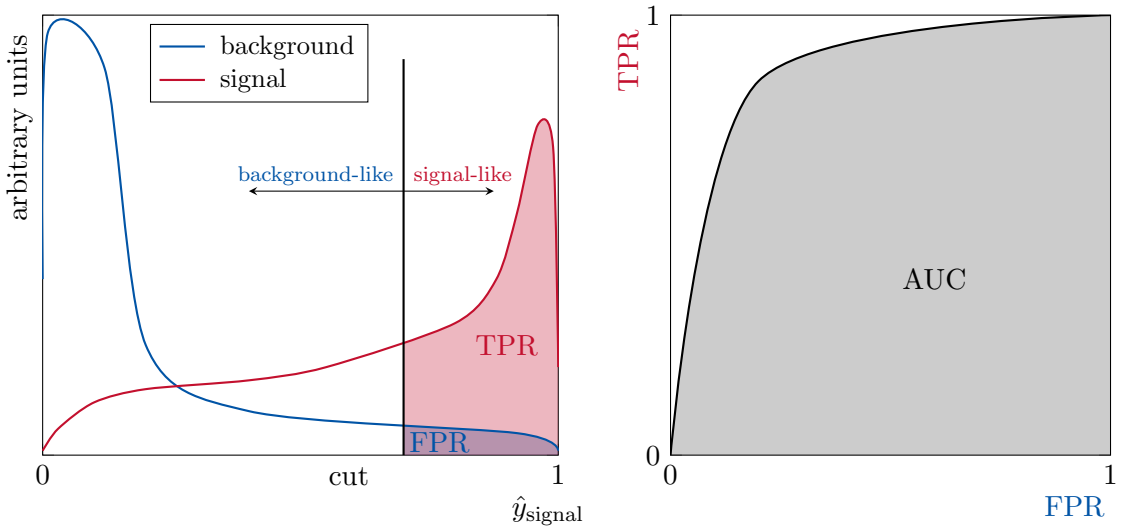



**Figure 6.5:** Visualisation of the *true-positive rate* (TPR) (red shaded area) and the *false-positive rate* (FPR) (blue shaded area) on the basis of the signal and background distributions of the signal output node probability $\hat{y}_{\text{signal}}$ (a) and the resulting ROC-curve (b). Note that the ROC-curve is an example and not related to the curves in (a). The *area-under-curve* (AUC) is shaded in gray.

### Full Statistical Analysis

The most accurate metric for comparing the performances of different networks in the context of the $H$-LFV search is the full statistical analysis described in Chapter 10. However, this is computationally expensive and furthermore, it is evaluated when estimating the background processes with the *Symmetry Method* and not when estimating the background processes from simulated events (as it is done for the training of the NNs as described in Section 8.2). Hence, this metric is only utilised for networks that were pre-selected with the help of other metrics.

### Binned Asimov Significance

The *Binned Asimov Significance* [216] is a fast method to calculate a metric that is closer to the final *Statistical Analysis* than the accuracy or the AUC. It yields the same result as the *Statistical Analysis* performed without considering any uncertainties on the estimation of the different processes. The *Binned Asimov Significance* $Z$ is defined in terms of *signal* ($s$) and *background* ($b$) (defined as described for the Area-Under-Curve) for a binned distribution of $\hat{y}_{\text{signal}}$:

$$Z_i = \sqrt{2 \cdot ((s_i + b_i) \cdot \ln(1 + s_i/b_i) - s_i)}, \tag{6.26}$$

$$Z = \sqrt{\sum_{i=1}^{N_{\text{bins}}} Z_i^2}, \tag{6.27}$$

with $N_{\text{bins}}$ being the number of bins and $s_i$ and $b_i$ denoting the number of signal and background events in bin $i$, respectively. The uncertainty on $Z_i$ based on the uncertainty of signal, $s_{s,i}$, and background, $s_{b,i}$, is given by:

$$s_{Z,i} = \frac{1}{Z_i} \cdot \sqrt{(\ln(1 + s_i/b_i) \cdot s_{s,i})^2 + ((\ln(1 + s_i/b_i) - s_i/b_i) \cdot s_{b,i})^2}. \tag{6.28}$$

## 6.2 $k$-fold Cross Validation

It is good practice to validate the performance of a neural network on statistically independent data (the so-called *validation set*) and to use another independent dataset (the so-called *test set*) to perform the final statistical analysis in order to avoid any bias. But this implies that all available data must be split, leading to reduced statistics for each step. $k$-fold cross validation can partially circumvent this. All available data is randomly split into $k$ equally sized slices. In this work $k = 10$ is used. Then, $k = 10$ individual neural networks (with the same hyperparameters) are trained, each using $k - 2 = 8$ slices, corresponding to 80 % of the data. The remaining 2 slices (10 % of the dataset, respectively) of each neural network are used for validation and testing. The usage of each slice is interchanged for the different neural networks such that each slice is used eight times for training, once for validation and once for testing. When adding up the distributions of all $k = 10$ neural networks, in total all available data is used for both, validation and testing. The concept of the cross validation is visualised in Figure 6.6. To apply the neural networks to recorded

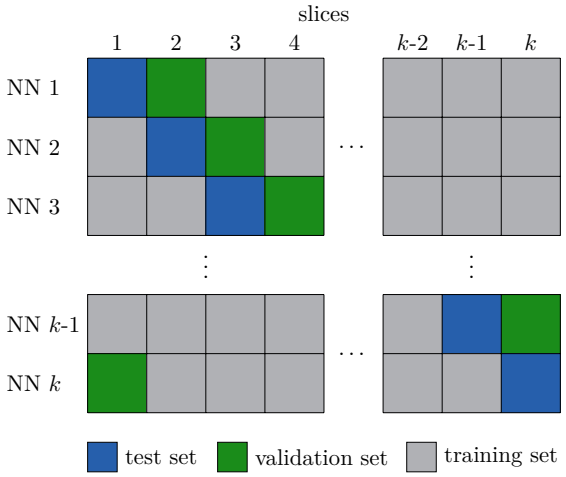



**Figure 6.6:** Visualisation of the concept of $k$-fold cross validation.

ATLAS data, the data is also split into 10 equally sized slices. To ensure that the data slices are always the same during repeated processing of the data, the splitting is done via 'event number modulo 10'.

## 6.3 Input Features

The preprocessing and selection of the input features that are utilised for the neural networks are of high importance for the performance of the trained NN.

### 6.3.1 Preprocessing

The input features or input variables utilised to classify the event can differ substantially in their magnitude, range and spread of values. This would inevitably require different orders of magnitude of the individual weights and biases such that the weighted inputs $z_j^l$ have comparable orders of magnitude when entering the activation function. However, the weights and biases are typically initialised with small random values and hence the training can be unstable or slow. Therefore, to stabilise the training process, input features are *standardised* such that they are in comparable ranges. To do so, the input features of the training events are shifted by their mean and divided by their standard deviation, such that they are centered around 0 and have a standard deviation of 1. The mean and standard

deviation used for scaling must also be applied to the events which are later classified using the network.

In particle physics, the events in the training dataset usually have associated event weights, e.g. MC weights coming from the simulation process (Section 4.1), from scale factors determined to correct for differences observed in simulation and data (Chapter 5), and from corrections applied in the course of the background estimation techniques (Section 7.2). These weights alter the shape (and also normalisation) of the distribution of the observables (i.e. input features) differently for the different classes. Hence, it is important to not neglect these weights in the training process. This is achieved by weighting each term $C_n$ corresponding to event $n$ in the loss function (Eq. 6.9) with the respective event weight which modifies the importance of the single $C_n$ terms. To some degree this is also possible if an event weight is negative given there are enough events with a positive weight in the same phase space corner such that their sum remains positive. This can also be steered by the choice of the batch size. But if the number of events with a negative event weight in one class is much larger than the number of events with a positive event weight (even if the sum of weights is positive), the training is at risk of becoming unstable. In such cases it is advisable to combine two event classes into a single one (e.g. if both classes comprise background processes). Negative event weights can either arise from negative MC weights due to higher-order corrections or if a whole process needs to be subtracted in the course of the background estimation.

There is not necessarily one class per physics process in the neural network training. Because on one hand there is no clear definition to which detail the physics processes should be split. And on the other hand the available statistics per class should not be too small. Hence, the classes defined for the neural network training can each comprise several physics processes. The number and the choice of these classes is also subject to an optimisation procedure. The processes that are combined into one class are weighted according to their respective cross sections. Furthermore, it should be ensured that all classes are weighted equally for a stable training process. These class weights are obtained by comparing the sum of weights of each class: the class with the largest sum of weights $N^{\text{largest}}_{\text{weighted}}$ gets a class weight of 1, the class weights of the other classes $i$ are calculated as $N^{\text{largest}}_{\text{weighted}}/N^{i}_{\text{weighted}}$. This ensures equal importance of each class in the training. A higher class weight can intentionally be assigned to some classes in order to bestow more importance on them in the training. However, equal weighting of the classes is set as default.

### 6.3.2 Selection

The selection of the input features is decisive for the performance of the network. There are various options for compiling a set of input features: either a set of input features mainly based on high-level features such as the collinear mass (reconstructed Higgs-boson mass, detailed in Section 8.2.3) or ratios of the momenta of final state particles could be chosen. Alternatively, a set consisting of only low-level features such as the four-vectors of the final state particles, or a combination of low-level input features and a small number of well selected high-level features could be chosen. The advantage of a set consisting only of low-level input features is that the network can identify high-level patterns by itself, which are (yet) unknown and which it possibly would not be able to learn when other pre-calculated high-level information is available. The downside of this elegant approach is that in order to find features and patterns in the training dataset, the statistics must be large enough to provide sufficient separation power between the classes. Hence, in this work, different sets of input features are evaluated. The set of selected input features should be large enough to provide sufficient separation power but should at the same time be as small as possible to prevent overtraining. Furthermore, only variables that show a good modelling

of the data by the simulated samples and by the full background estimation should be included such that a good modelling of the resulting NN output node distribution(s) is ensured. The *Permutation Feature Importance* described in the following Section 6.3.3 can help to reduce the number of input features for sets consisting mainly of high-level features by ordering them by their importance in a specific network. However, this method does not properly take correlations between features into account, which can influence the importance ranking. In addition, the correlations between input features and the differences in the correlations between the classes can be utilised to decide which input variables could be removed. If two variables show a high correlation that is similar for the different classes, one of the variables could be removed. But if the correlation differs between the classes this may provide separation power and both variables should be kept.

### 6.3.3 Ranking

Knowledge about the importance of the individual input features on the performance of the trained NN is useful for two reasons. On one hand it aids in improving the understanding of the underlying model. On the other hand a ranking of the input features can be used as a basis for removing unimportant features which do not contribute to the classification power of the NN. There are several different methods to obtain a ranking of the input features. One option is the successive removal of one input feature, followed by re-training the NN and evaluating the changes in performance, e.g. via an appropriate metric such as the value of the loss/cost function. However, this method is computionally expensive since the NN must be re-trained for each removed feature. In addition, for a fair comparison, hyperparameters may have to be re-tuned for each training. An alternative option is the *Permutation Feature Importance* which is introduced below. It does not require re-training the NN but instead provides an importance metric based on the change in loss or another performance metric when the events' values of an input feature are permuted and hence the relation between the input feature and the true class is broken.

#### Permutation Feature Importance

The *Permutation Feature Importance* is used to obtain an approximate ranking of the importance of the input features for a specific model, in this work a trained NN. The implementation and explanation detailed below follows the description given in [217].

A metric $M_{\text{original}}$ (e.g. the loss or another performance metric) of the trained NN is calculated with the validation data set (more details on the validation dataset are given in Section 6.2). Then, the values of one input feature $x_i$ are shuffled between all events such that the connection of the information of this input feature with the true classes of the events is destroyed. The metric is recalculated as $M^{x_i}_{\text{shuffled}}$. In principle, $M^{x_i}_{\text{shuffled}}$ is expected to indicate a worse performance than $M_{\text{original}}$ since information is removed by the shuffling. The importance $\mathcal{I}^{x_i}$ of one input feature $x_i$ is defined as $\mathcal{I}^{x_i} = M^{x_i}_{\text{shuffled}}/M_{\text{original}}$. When a smaller value of the metric indicates a better performance, $\mathcal{I}^{x_i}$ is expected to be larger than 1, and the larger $\mathcal{I}^{x_i}$ is, the more important the corresponding input feature is. For a more accurate result, the shuffling is repeated five times per input feature and the mean of the resulting importance is quoted in combination with its standard deviation as uncertainty. Finally, a ranking of the input features is obtained by sorting them in descending order following their assigned importances $\mathcal{I}^{x_i}$.

In case of 10-fold cross validation (detailed in Section 6.2) the *Feature Importance* is calculated as described above for each of the 10 NNs separately. Comparing the individual rankings of the 10 NNs gives insights into the stability and similarity of the single networks and of the feature importance ranking itself. A combined ranking over all 10 NNs is obtained by computing the mean of $\mathcal{I}^{x_i}$ of all 10 NNs. The uncertainty is given by the standard

deviation of the $\mathcal{I}^{x_i}$ of all 10 NNs.

However, there are also some concerns regarding *Permutation Feature Importance*: if the input variables are (highly) correlated, the importance can be biased such that both variables seem less important due to their shared importance. Furthermore, the shuffling can introduce unphysical combinations of input features.

## 6.4 Hyperparameter Selection and Optimisation

There are various parameters of the neural network that either describe its architecture or steer the training process. These so-called *hyperparameters* can be chosen freely (and are not determined in the training process). However, they are, in addition to the choice and preprocessing of the input features, pivotal for the network's performance. The hyperparameters relevant in this thesis are:

- number of output classes $Cl$
- one NN with $Cl$ output classes vs $Cl-1$ NNs with two output classes each
- number of hidden layers $L_h$
- number of nodes per hidden layer $\mathcal{N}^{l_h}$
- activation function in hidden layers and corresponding parameters
- optimiser used in training
- learning rate $\eta$
- regularisation of the network and corresponding parameters
- batch size $s_b$
- number of epochs $N_{ep}$
- initialisation of the weights and biases

Identifying the optimal choice or value for the single hyperparameters is a challenging and costly task. Some hyperparameters are chosen manually, for others systematic optimisation using the hyperparameter optimisation framework *Optuna* [218] is performed. A short description of *Optuna* is given in the following Section 6.4.1 while the details of the hyperparameter selection and optimisation specifically for the HLFV analysis are given in Section 8.2.4.

### 6.4.1 Optuna

The introduction to the *Optuna* framework in this section follows the description in [218, 219]. The selection of the optimal hyperparameter values is again a minimisation process. Hereby, an *objective function* $f(\boldsymbol{h})$ is minimised with $\boldsymbol{h} \in \mathcal{H}$ representing a set of hyperparameter values of the hyperparameter configuration space $\mathcal{H}$ such that $\boldsymbol{h}^* = \arg\min_{\boldsymbol{h}\in\mathcal{H}} f(\boldsymbol{h})$ represents the best performing set. The *objective function* is defined by the user of the framework and takes the hyperparameter values as inputs and returns the *objective value* which is used to judge the performance of the respective set of hyperparameters. Usually, the *objective function* is of unknown structure and the single hyperparameters can be of different types: categorical, continuous, discrete. Each evaluation of the *objective function*, i.e. calculation of the *objective value* for one set of hyperparameter values, is called *trial* within *Optuna*. Typical choices for the *objective value* can be the loss, the accuracy, the AUC or the binned significance (for details see Section 6.1.5). The *objective function* used in this analysis is introduced in Section 8.2.4.

The evaluation of $f(\boldsymbol{h})$ is typically very costly since a full NN training must be performed. Hence, an efficient method for finding $\boldsymbol{h}^*$ is desired. Methods such as a grid scan or the more sophisticated random search [220] do not use the outcomes of previous trials to steer the search in the hyperparameter space and thus may evaluate the *objective function* $f(\boldsymbol{h})$

in non-promising regions of the hyperparameter space. Furthermore, badly performing trials are not stopped prematurely (so-called *pruning*). Therefore, *Optuna* allows the user to make use of *Bayesian Optimisation* [221], which approximates $f(\boldsymbol{h})$ with a probabilistic surrogate model $S(\mathcal{T})$ (prior) which is based an all previous *trials* $\mathcal{T}$. The surrogate model is faster to compute than $f(\boldsymbol{h})$. It is the basis for the acquisition function $a(\boldsymbol{h}|\mathcal{T})$ which is used to obtain the next hyperparameter set $\boldsymbol{h}_{\text{new}}$ to be tested. The acquisition function $a(\boldsymbol{h}|\mathcal{T})$ (often the *Expected Improvement* [221, 222] is used) performs a tradeoff between exploitation, i.e. selecting a set $\boldsymbol{h}_{\text{new}}$ from a region in $\mathcal{H}$ where the surrogate predicts good performance, and exploration, i.e. selecting a set $\boldsymbol{h}_{\text{new}}$ from a region in $\mathcal{H}$ where the uncertainty on the surrogate is still large. The acquisition function $a(\boldsymbol{h}|\mathcal{T})$ and the model $S(\mathcal{T})$ are chosen such that their evaluation time is short. The new hyperparameter set $\boldsymbol{h}_{\text{new}}$ that maximises[5] $a(\boldsymbol{h}|\mathcal{T})$ is chosen and $f(\boldsymbol{h}_{\text{new}})$ is evaluated. The outcome is utilised to update $S(\mathcal{T})$ (posterior) and the steps are repeated. Following this procedure, different promising sets of hyperparameters $\boldsymbol{h}$ are evaluated and the best performing one is returned. The implementation using *Tree-structured Parzen Estimator*s (TPEs) [223–225] for *Bayesian Optimisation* is used in this analysis.

In addition, *Optuna* provides the possibility to *prune* non-promising trials. Different pruning algorithms are provided. For this analysis, the *Median Pruner* is used, which prunes "if the trial's best intermediate result is worse than median of intermediate results of previous trials at the same step" [226]. Here, the loss of the validation dataset represents the intermediate result and the training epochs are the steps.

There are also other frameworks that provide tools based on *Bayesian Optimisation* to optimise the hyperparameter values. But *Optuna* has the advantage that it lets the user define the hyperparameter search space in a *define-by-run* [218] style. That means that the value space of one hyperparameter can depend on the value of another hyperparameter. This allows for a flexible creation of the hyperparameter search space.

The usage of *Optuna* in the selection process of the hyperparameter values in this thesis is detailed in Section 8.2.4.

[5]Higher values of the acquisition function $a(\boldsymbol{h}|\mathcal{T})$ indicate more promising hyperparameter sets.

# Chapter 7

# Event Selection and Background Estimation

This chapter gives an overview of the event selection and the methods to estimate the individual background contributions. The event selection applies basic selection criteria to enhance the expected signal contribution compared to background predictions and is discussed in Section 7.1. The estimation of background contributions uses data-driven methods for the largest background contributions and simulated events for smaller contributions and is described in Section 7.2.

## 7.1 Event Selection

The primary selection of events is performed by the ATLAS trigger system described in Section 2.2.5. An overview of the triggers utilised for the $H$-LFV analysis is given in Section 7.1.1. The event selection aims at efficiently selecting the signal events while reducing the contributions from background processes. It does so by applying simple requirements on the objects' properties in the event which are motivated by the signal signature. This is described in Section 7.1.2. Finally, the splitting of the events into two datasets, each containing the majority of one of the two signal processes $H \to e\tau$ or $H \to \mu\tau$, is discussed in Section 7.1.3.

### 7.1.1 Trigger Selection

Events used in this analysis are triggered based on the presence either of an electron or a muon in the event with so-called single lepton triggers or of both with so-called dilepton triggers. These triggers apply requirements on the transverse momenta of the particles as well as identification and isolation criteria to select events relevant for analysis.

The trigger items used in this analysis are listed in Table 7.1, where triggers in 2016-2018 used tighter criteria compared to 2015 due to increased instantaneous luminosity and therefore increased pile-up (Figure 2.8(b)). The listed triggers are combined with a logical `OR`. Details on the ATLAS trigger system are given in Section 2.2.5. The naming convention of the triggers is as follows. The term `HLT` denotes the *High-Level Trigger*, `e` or `mu` indicates whether this trigger branch is an electron or muon trigger, the number afterwards gives the $p_T$-threshold required for the trigger object at HLT-level, followed by the quality criteria at HLT-level. The quality criteria comprise identification criteria (`lhmedium` or `lhloose`) including the information whether the requirement on $d_0$ (transverse distance between primary vertex and track) is dropped (`nod0`) as well as isolation criteria (`ivarloose` and `ivarmedium`). The quality criteria are defined similar to the offline[1] criteria discussed in

[1] Offline refers to the processing of the data after the events were triggered and stored while online refers

| Trigger Menu | Data Period | Trigger Chain Name | recommended $p_T$-thresholds | used $p_T$-thresholds |
|---|---|---|---|---|
| Single Electron | 2015 | `HLT_e24_lhmedium_L1EM20VH`<br>`HLT_e60_lhmedium`<br>`HLT_e120_lhloose` | 25 GeV | 25 GeV |
| | 2016-18 | `HLT_e26_lhtight_nod0_ivarloose`<br>`HLT_e60_lhmedium_nod0`<br>`HLT_e140_lhloose_nod0` | 27 GeV | 27 GeV |
| Single Muon | 2015 | `HLT_mu20_iloose_L1MU15`<br>`HLT_mu50` | 21 GeV | 25 GeV |
| | 2016-18 | `HLT_mu26_ivarmedium`<br>`HLT_mu50` | 27 GeV | 27 GeV |
| Dilepton | 2015 | `HLT_e17_lhloose_mu14` | 18 GeV, 15 GeV | 18 GeV, 18 GeV |
| | 2016-18 | `HLT_e17_lhloose_nod0_mu14` | 18 GeV, 15 GeV | 18 GeV, 18 GeV |

**Table 7.1:** Triggers used in this analysis, combined via a logical `OR`. Symmetric offline $p_T$-thresholds are used instead of the *recommended* ones to avoid introducing an artificial asymmetry between electrons and muons. For data taking during 2015 the *used $p_T$-thresholds* do not represent the complete picture; consult the text for more information.

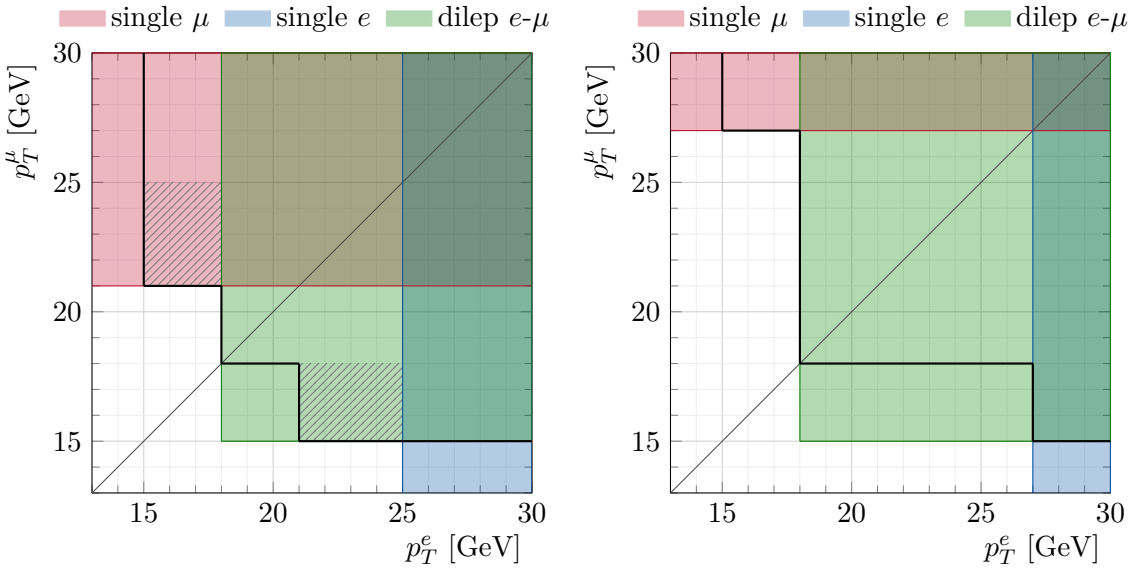



**Figure 7.1:** Visualisation of the trigger offline $p_T$-thresholds for data-taking during 2015 (a) and during 2016-2018 (b). The coloured areas indicate the phase space spanned by the *recommended* $p_T$-thresholds of the individual triggers: single-muon trigger (red), single-electron trigger (blue) and dilepton electron-muon trigger (green). Black solid lines indicate the symmetric thresholds used in the analysis. The offline object selection requires the leptons to have $p_T^\ell > 15\,\text{GeV}$, independent of trigger thresholds. This is incorporated in the diagrams as well. The shaded areas in (a) are not covered by the symmetric $p_T$-thresholds listed in Table 7.1 but are recovered by additional requirements; consult the text for more information.

Chapter 5 but are typically less costly in order to allow for sufficiently fast processing. The addition of an `L1*` requirement indicates that non-default requirements for the *Level-1*

to the processing during datataking.

*Trigger* were used. The term `L1EM20VH` indicates that instead a $p_{\mathrm{T}}$-threshold of 20 GeV measured by the electromagnetic calorimeter (`EM`) which varies with $\eta$ to account for energy loss (`V`) is required. In addition, hadronic core isolation (`H`) is required. The term `L1MU15` indicates a $p_T$-threshold of 15 GeV for the muon candidate for the *Level-1 Trigger*.

Dedicated groups within the ATLAS collaboration study the efficiencies of the triggers and recommend $p_{\mathrm{T}}$-thresholds to be used in analyses above which the triggers are expected to be fully efficient. Below the recommended $p_{\mathrm{T}}$-thresholds the agreement of data and simulated events is often insufficient. These recommended $p_{\mathrm{T}}$-thresholds are summarised in Table 7.1 together with the thresholds that are actually used in the $H$-LFV analysis which are at least as large as the recommended ones. The $p_{\mathrm{T}}$-threshold values chosen in the analysis differ from the recommended ones to ensure a symmetric selection of electrons and muons. A symmetric selection is crucial for the *Symmetry Method* as discussed in Chapter 3. The overall trigger strategy is visualised in Figure 7.1 for 2015 and the other datataking years separately. Both, the standard recommended $p_{\mathrm{T}}$-requirements and the adjusted symmetric ones used in this analysis are indicated. Their values depend on the trigger which fired (Table 7.1) and are indicated by the coloured lines. For 2016-2018 data taking the identification of the symmetric thresholds is straight-forward (Table 7.1 last column and Figure 7.1(b)). The green area of the dilepton trigger for $p_T^\mu < 18\,\mathrm{GeV}$ and $p_T^e < 27\,\mathrm{GeV}$ is discarded as a result of the symmetry-considerations essential for the analysis method. For 2015 data taking the choice of the thresholds is more subtle. The shaded areas in Figure 7.1(a) are not included by fully symmetric requirements of $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ for the single lepton triggers and $p_{\mathrm{T}} > 18\,\mathrm{GeV}$ for the dilepton trigger, as listed in Table 7.1. They are recovered with the following additional criteria:

- if single-$\mu$ trigger fired: $p_T^\mu > 21\,\mathrm{GeV}$
- if dilep trigger fired: $p_T^e > 21\,\mathrm{GeV}$ and $p_T^\mu > 15\,\mathrm{GeV}$

In this way, only the area of the dilepton trigger with $p_T^\mu < 18\,\mathrm{GeV}$ and $p_T^e < 21\,\mathrm{GeV}$ needs to be discarded as a result of symmetric thresholds.

The different efficiencies of the single triggers introduce asymmetries between reconstructed electrons and muons; this is, however, taken into account as described below in Section 7.2.4.

The `TrigGlobalEfficiencyCorrectionTool` [227] provides the correctly calculated efficiencies and *scale factor*s (simulation-to-data correction factors) for the `OR`-combination of the triggers.

### 7.1.2 Basic Event Selection

The events taken into account in this analysis must be part of a so-called *Good Runs List* (GRL). The list is designed to contain only data where the ATLAS detector was fully operational and the recorded data is complete and not corrupted. The *bad batman* event cleaning is applied to data recorded in the years 2015 and 2016 to remove problematic events based on the number of high $p_T$ clusters of poor quality in the electromagnetic end-cap calorimeter (EMEC) [228]. This is recommended for analyses susceptible to mis-identified $E_{\mathrm{T}}^{\mathrm{miss}}$ and selecting jets with $p_T < 50\,\mathrm{GeV}$. This reduces the number of events in data from 2016 by 1.66 % and those from 2015 by a negligible amount. The number of primary vertices must be at least one in order to suppress random, non-collision events such as cosmic rays.

The expected final state objects are required or non-expected ones are vetoed and basic kinematic cuts are applied. The object definitions themselves are discussed in Chapter 5. Two *baseline* light leptons[2] that pass the *overlap removal* (Section 5.7) are required while

[2]See Sections 5.2 and 5.3 for the definition of *baseline* and *signal* for electrons and muons.

hadronically decaying $\tau$-leptons fulfilling the requirements defined in Section 5.5 are vetoed. The two light leptons subsequently need to fulfil the *signal* lepton[2] requirements, need to have different flavour (i.e. one electron and one muon are required) and opposite-sign charges (OC). The transverse momentum $p_{\mathrm{T}}$ of the lepton with the larger transverse momentum, called leading lepton $\ell_0$, must be larger than 35 GeV while the transverse momentum of the other lepton, the sub-leading lepton $\ell_1$, must be larger than 15 GeV. The invariant mass $m_{\ell\ell}$ of the two leptons must be between 30 GeV and 150 GeV to suppress background contributions, primarily from top-quark production and diboson production, outside this mass window populated by signal. A veto on $b$-jets (see Section 5.4.3) is applied in order to suppress background contributions from top-quark production. The event selection is summarised in Table 7.2. Distributions of the kinematic observables used to define the

| Cut | Selection Criteria |
|---|---|
| 0 | $\in$ GRL + cleaning (data only) |
| 1 | $n_{\mathrm{pvx}} > 0$ |
| 2 | $n_\mu = 1,\ n_e = 1$ |
| 3 | $n_{\tau_{\mathrm{had}}} = 0$ |
| 4 | $q_\mu \cdot q_e = -1$ |
| 5 | $p_T^{\ell 0} > 35\,\mathrm{GeV}$ |
| 6 | $p_T^{\ell 1} > 15\,\mathrm{GeV}$ |
| 7 | $30\,\mathrm{GeV} < m_{\ell\ell} < 150\,\mathrm{GeV}$ |
| 8 | $n_{b\text{-jets}} = 0$ |

**Table 7.2:** The *Basic Event Selection*. $n$ stands for *number*, $q$ for *charge* and "pvx" for *primary vertex*.

afore listed requirements, also called *cuts*, are shown right before these cuts are applied in Figure 7.2[3]. Requirements on the transverse momenta of the two leptons were, however, already applied at an earlier stage of the sample processing and hence bias the displayed distributions: $p_T^\mu > 10\,\mathrm{GeV}$ and $p_T^e > 15\,\mathrm{GeV}$. Additionally, the $p_T$-thresholds required for being near the trigger-efficiency plateau (Section 7.1.1) are applied beforehand. The background contributions in the figures are estimated here based on simulated events and split into individual processes that were introduced in Section 4.2.2 such that their individual distributions can be seen. The reduction of contributions from top-quark production by the applied $b$-veto is apparent when comparing Figure 7.2(c) and Figure 7.2(d). The cuts at $p_T^{\ell 0}$ and on $m_{\ell\ell}$ are kept loose in order to maintain larger statistics for the subsequent training of *Neural Network*s discussed in Section 8.2.

The *Basic Event Selection* is abbreviated by *Basic Selection* in the following and the region defined by this selection is called *Basic Selection Signal Region* or in short *Basic SR*.

### 7.1.3 Lepton Assignment and Dataset Classification

The signal processes in this analysis are $H \to e\tau_\mu$ and $H \to \mu\tau_e$. Hence, one of the final state light leptons originates directly from the Higgs boson while the other originates from the $\tau$-lepton. The lepton assignment procedure assigns each of the two light leptons to the Higgs boson or to the $\tau$-lepton, leading to the labels $\ell_H$ and $\ell_\tau$, respectively. This is visualised in Figure 7.3, showing a sketch of a LFV decay of the Higgs boson in its rest frame. The assignment is necessary in order to:

[3]Most of the kinematic distributions shown in this thesis were created with the `uhepp` python package: `https://uhepp.org/`.

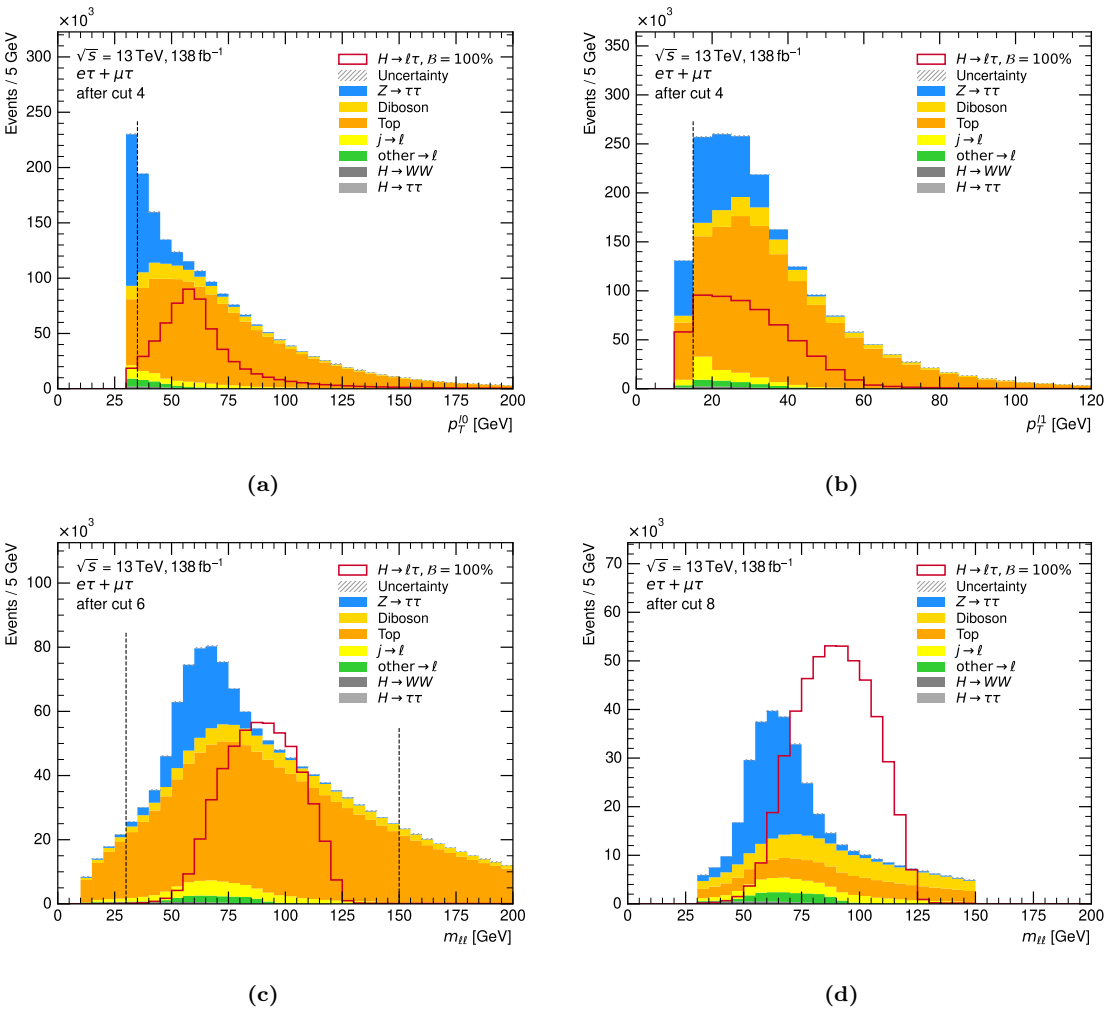



**Figure 7.2:** Distributions of the observables to which selection requirements are applied in the context of the *Basic Event Selection*. $p_{\mathrm{T}}^{\ell 0}$ (a) and $p_{\mathrm{T}}^{\ell 1}$ (b) are shown before cut 5 and 6, respectively (Table 7.2) while the invariant mass of the two leptons, $m_{\ell\ell}$, is shown before cut 7 (c). The latter is again shown after cut 8 (d) to visualise the effect from applying the $b$-veto. The cuts that are applied to the observables are indicated by black dashed lines. The coloured and filled histograms correspond to the background processes that were introduced in Section 4.2.2. The *irreducible* contributions are estimated directly from simulated events (Section 4.1) and are scaled according to their cross section and to the respective luminosity. The estimation of the *reducible fake*-backgrounds $j \to \ell$ and other $\to \ell$ is described in Section 7.2.2. The $H \to \ell\tau$-signal is the sum of the $H \to e\tau$ and the $H \to \mu\tau$ signals, for better visibility it is displayed with a branching ratio of 100 %.

- classify the event either as $H \to e\tau_\mu$ or as $H \to \mu\tau_e$ depending on whether the electron or the muon is labelled as $\ell_H$ or $\ell_\tau$
- improve the mass reconstruction due to the correct assignment of the neutrinos to the light lepton from the $\tau$-lepton
- exploit the different kinematics of $\ell_H$ and $\ell_\tau$, e.g. in order to separate the $H \to \ell\tau$ signals from $H \to \tau\tau$

As briefly discussed in Chapter 3 and Section 4.2.1, the lepton with the larger transverse momentum ($p_{\mathrm{T}}$) is assumed to originate from the Higgs boson while the one with the lower $p_{\mathrm{T}}$ is expected to originate from the $\tau$-lepton where the additional neutrinos in the $\tau$-lepton decay carry away a fraction of the $\tau$-lepton momentum, leaving less for the light decay lepton (Figure 7.3). Previously,[4] the transverse momenta were compared in the laboratory

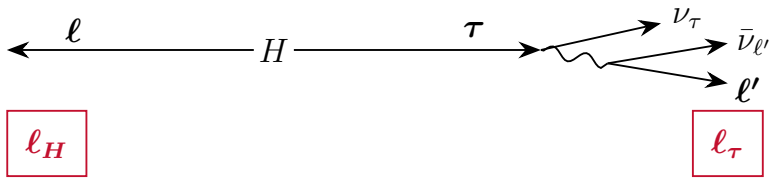



**Figure 7.3:** Sketch of a LFV decay of the Higgs boson.

frame[5] which results in $\ell_H = \ell_0$ and $\ell_\tau = \ell_1$. However, the assumption is only valid in the Higgs boson's rest frame which differs from the laboratory frame when the Higgs boson is boosted. Therefore, the aim is to boost the system back into the Higgs boson's rest frame in order to make the assumption valid again and hence enhance the number of events for which the assignment is performed correctly. Owing to the measurement of only the transverse component of the combined neutrino system as $E_{\mathrm{T}}^{\mathrm{miss}}$, the four-vector of the Higgs boson $p_H^\mu$ is not entirely known; its correct longitudinal or $\eta$ component is missing. In order to approximate this component, an estimation of the missing $\eta$-component of $E_{\mathrm{T}}^{\mathrm{miss}}$ is needed. Ideally, $\eta(\ell_\tau)$ would be used for $\eta(E_{\mathrm{T}}^{\mathrm{miss}})$ due to the approximate collinearity of $\ell_\tau$ and the two neutrinos. However, $\ell_\tau$ is still unknown (since the aim of the procedure is to determine it), therefore the missing $\eta$-component of $E_{\mathrm{T}}^{\mathrm{miss}}$ is approximated by $\eta$ of the electron-muon system:

$$\eta\left(p^\mu_{E_{\mathrm{T}}^{\mathrm{miss}}}\right) = \eta\left(p_e^\mu + p_\mu^\mu\right) , \tag{7.1}$$

where $p^\mu_{E_{\mathrm{T}}^{\mathrm{miss}}}$, $p_e^\mu$ and $p_\mu^\mu$ are the four-vectors of $E_{\mathrm{T}}^{\mathrm{miss}}$, the electron and the muon, respectively. This approximation defines the full momentum-vector $\boldsymbol{p}_H$ of the Higgs boson. In addition the energy of the Higgs boson is adapted by constraining its mass to 125 GeV:

$$125^2\,\mathrm{GeV}^2 = m_H^2 = \left(p_H^\mu\right)^2 \quad \Longrightarrow \quad E_H^2 = \boldsymbol{p}_H^2 + 125^2\,\mathrm{GeV}^2 . \tag{7.2}$$

With the help of the approximated four-vector of the Higgs boson, the system is boosted back into the approximate rest frame of the Higgs boson. The lepton with the larger transverse momentum in this rest frame is denoted as $\ell_H$ while the other is denoted as $\ell_\tau$.

A comparison of the assignment accuracies after the *Basic Selection* using different options is summarised in Table 7.3. All options listed in the table are based on $p_{\mathrm{T}}$-ordering. But different assumptions on the frames are made in which the transverse momenta are compared:

1. laboratory frame: the laboratory frame is used
2. approx. rest frame w/o $\eta$-, $m_H$-constr.: the approximate Higgs-boson rest-frame is used without the assumptions in Eqs. 7.1 and 7.2, i.e. the four-vector of the Higgs boson used to boost back into the rest frame is the sum of the four-vectors of electron, muon and $E_{\mathrm{T}}^{\mathrm{miss}}$ as they were measured (no $z$- or $\eta$-component for $E_{\mathrm{T}}^{\mathrm{miss}}$)
3. approx. rest frame w/o $\eta$-constr.: the Higgs-boson four-vector uses the constraint of Eq. 7.2 but not the assumption of Eq. 7.1, i.e. no $z$- or $\eta$-component for $E_{\mathrm{T}}^{\mathrm{miss}}$ is assumed
4. approx. rest frame: the assumptions of both Eqs. 7.1 and 7.2 are used; this is the option which is used in the analysis

[4] in the Run 1 [120] and in the early Run 2 [19] publications

[5] which is the reference frame usually used for all observables and in which also all observables in the scope of this thesis are measured in if not stated otherwise

| | $H \to e\tau_\mu$ | | | $H \to \mu\tau_e$ | | |
|---|---|---|---|---|---|---|
| option | ggF | VBF | VH | ggF | VBF | VH |
| 1. laboratory frame | 88.7 | 78.0 | 77.1 | 85.1 | 71.1 | 71.1 |
| 2. approx. rest frame w/o $\eta$-, $m_H$-constr. | 92.4 | 88.4 | 82.3 | 92.0 | 88.4 | 80.6 |
| 3. approx. rest frame w/o $\eta$-constr. | 92.9 | 89.4 | 82.6 | 92.3 | 89.1 | 80.8 |
| 4. approx. rest frame | 93.6 | 91.3 | 84.6 | 92.8 | 90.5 | 82.3 |

**Table 7.3:** Accuracy of the lepton assignment in % for the different $H$-boson production modes obtained by $p_{\mathrm{T}}$-ordering in the laboratory and in the approximate $H$-boson rest frame. The accuracies when using the approximate $H$-boson rest frame without the $\eta$-constraint on $E_{\mathrm{T}}^{\mathrm{miss}}$ and without the constraint on the $H$-boson mass are listed as well. The accuracy is defined as the percentage of signal events for which the lepton assignment is performed correctly such that the event is classified as the correct signal.

The gain by option 2 compared to option 1 is 4.2 % and 8.1 % for the ggF production process for the signals $H \to e\tau_\mu$ and $H \to \mu\tau_e$, respectively. The VBF production process gains 13 % and 24 %, respectively, while the improvemt for the VH production process is 6.7 % and 13 %. The improvement due to the additional assumptions (options 3 and 4) is smaller, up to 3 % in addition, but still desirable. Overall, when comparing option 4 to option 1, the VBF production process exhibits the largest gains, followed by the VH production, which is the same pattern as observed for option 2 versus option 1. The comparison of the accuracies of the different production modes within option 4 shows that the ggF production process achieves the highest accuracy, followed by VBF and VH. The reason why the VH production process has the lowest accuracy originates from the additional $V = Z, W$-boson. If it decays leptonically, there can be up to two additional leptons in the final sate which can be accepted in the *Basic Selection* instead of the leptons originating from the Higgs boson. This can deteriorate the accuracy. The assignment performs better for the $H \to e\tau_\mu$ than for the $H \to \mu\tau_e$signal.

Using the above assignment method, each event can be classified into one of the two datasets:

**$e\tau$-dataset:** $\ell_H$ is the electron, $\ell_\tau$ is the muon,

**$\mu\tau$-dataset:** $\ell_H$ is the muon, $\ell_\tau$ is the electron,

where the majority of the $H \to e\tau_\mu$-signal events is part of the $e\tau$-dataset and the majority of the $H \to \mu\tau_e$-signal events is part of the $\mu\tau$-dataset according to Table 7.3.

The expected yields of the different processes and the observed data after the *Basic Selection* for the $e\tau$- and $\mu\tau$-datasets are listed in Table 7.4. Apart from one *fakes* component, the *other* $\to \ell$ contribution which is discussed in Section 7.2.2, there are more events in the $e\tau$-dataset than in the $\mu\tau$-dataset. In the $e\tau$-dataset, the muon has the smaller transverse momentum in the laboratory frame for the majority of the events (if $\ell_\tau = \ell_1$). Compared to electrons, muons have a steeper and earlier turn-on curve for the offline efficiencies with respect to the transverse momentum (Chapter 5 and Section 7.2.3). Therefore, more events can be reconstructed for which the muon has a small transverse momentum than for which the electron has a small transverse momentum. Whereas the trigger efficiencies for electrons are higher than for muons (Section 7.2.3). Typically, the leading lepton in the event fires the trigger which leads to more $e\tau$- than $\mu\tau$-events as well.

| | | Yields | |
|---|---|---|---|
| | **Process** | $e\tau$ | $\mu\tau$ |
| Signal ($\mathcal{B} = 1\,\%$) | $H \to e\tau_\mu$ | 2 457 ± 4 | 182.70 ± 1.10 |
| | $H \to \mu\tau_e$ | 183.30 ± 1.20 | 2 155 ± 4 |
| Fake Contributions | $j \to \ell$ | 18 300 ± 160 | 17 060 ± 120 |
| | other $\to \ell$ | 5 870 ± 70 | 11 150 ± 90 |
| Symmetric Background | $Z \to \tau\tau$ | 81 340 ± 140 | 70 600 ± 130 |
| | Diboson | 43 840 ± 80 | 39 470 ± 70 |
| | Top | 41 400 ± 50 | 37 360 ± 50 |
| | $H \to WW$ | 1 775 ± 4 | 1 488 ± 4 |
| | $H \to \tau\tau$ | 810.60 ± 1.60 | 722.00 ± 1.50 |
| Total Background | | 193 340 ± 240 | 177 850 ± 210 |
| $s/\sqrt{s+b}$ | | 5.55 | 5.08 |
| | Data | 189 117 | 175 642 |

**Table 7.4:** Yields after the *Basic Selection* with statistical uncertainties, for the $e\tau$- and $\mu\tau$-datasets. All values except for the $j \to \ell$-fakes (Section 7.2.2) and data are estimated from simulated events. The signal yields are given for a branching ratio of 1 %.

## 7.2 Background Estimation

The estimation of the background contributions is based on the *Symmetry Method* introduced in Chapter 3 which implicitly requires a separate estimation of the *fakes* since they are one of the sources for the violation of the $e/\mu$-symmetry assumption.

In the following, a more detailed introduction to the *Symmetry Method* and its implementation is given in Section 7.2.1. Section 7.2.2 describes the estimation of the *fakes*, which are events where other objects are mis-identified as prompt leptons. The determination of the detection efficiencies utilised in the *Symmetry Method* is introduced in Section 7.2.3. Finally, Section 7.2.4 concludes by combining the individual steps to obtain the full background estimation and the modelling of the data by the background contributions is investigated.

### 7.2.1 Implementation of the Symmetry Method

The *Symmetry Method* is based on two assumptions, as described in Chapter 3:

1. Standard model processes are symmetric with respect to an exchange of electrons with muons and vice versa.

2. The two LFV Higgs-boson decays $H \to e\tau$ and $H \to \mu\tau$, on the contrary, break this symmetry as long as their branching ratios differ.

In the following, the focus is on the first assumption, neglecting any possible LFV decays. A consequence of this assumption is that the number of produced events $n_{\text{prod}}$ from SM processes in the $e\tau$-dataset is approximately the same as the number of produced events in the $\mu\tau$-dataset:

$$n^{e\tau}_{\text{prod}} = n^{\mu\tau}_{\text{prod}} = n\,. \tag{7.3}$$

This symmetry also holds for individual bins of kinematic distributions for both datasets, resulting in the agreement of differential distributions of the kinematic observables for both datasets.

After the events are produced, they need to be detected by the detector to be available for analysis. However, the efficiency $\varepsilon$ of the ATLAS detector to detect e.g. electrons and muons is not 100 %, and the reduced efficiency needs to be taken into account. On the other hand, other objects can mimic the signatures of electrons or muons and hence be incorrectly detected as these. This leads to additional events with one detected electron and one detected muon. These additional events are called *fakes* (as introduced in Chapter 3) and are abbreviated in the following equations by $f$. With these two considerations, the number of detected events $n_{\text{det}}$ in both of the datasets can be written as:

$$n_{\text{det}}^{e\tau} = \varepsilon^{e\tau} \cdot n + f^{e\tau}\,, \tag{7.4}$$

$$n_{\text{det}}^{\mu\tau} = \varepsilon^{\mu\tau} \cdot n + f^{\mu\tau}\,. \tag{7.5}$$

Here, $\varepsilon^{e\tau}$ and $\varepsilon^{\mu\tau}$ denote the detection efficiency of the $e\tau$- and $\mu\tau$-events, respectively. Although the events of both datasets contain one electron and one muon each, the detection efficiencies differ ($\varepsilon^{e\tau} \neq \varepsilon^{\mu\tau}$) as they depend on the kinematics of the electron and the muon in each event and those differ. In one dataset the electron has the lower transverse momentum in the majority of the events and in the other the muon.

The expected yields of the *fakes* in the $e\tau$- and the $\mu\tau$-datasets differ as well since the probability that other objects mimic electrons or muons also depends on the kinematics of these objects (and also whether an electron or a muon is mimicked).

Solving now for $n$ and expressing the number of detected events by those of the other dataset gives:

$$n_{\text{det}}^{e\tau} = \frac{\varepsilon^{e\tau}}{\varepsilon^{\mu\tau}} \cdot \left(n_{\text{det}}^{\mu\tau} - f^{\mu\tau}\right) + f^{e\tau}\,, \tag{7.6}$$

$$n_{\text{det}}^{\mu\tau} = \frac{\varepsilon^{\mu\tau}}{\varepsilon^{e\tau}} \cdot \left(n_{\text{det}}^{e\tau} - f^{e\tau}\right) + f^{\mu\tau}\,. \tag{7.7}$$

This allows one dataset to serve as independent estimate of the other dataset. For example, the $e\tau$-dataset[6] can be estimated by the difference of the $\mu\tau$-dataset and the asymmetric $\mu\tau$-*fake* contribution, corrected by the ratio of detection efficiencies to mimic the symmetric part of the $e\tau$-dataset. And the asymmetric $e\tau$-*fake* contribution is added.

Technically, the detection efficiencies are considered for each event individually due to their kinematic dependency. Hence, after expansion, Eqs. 7.6 and 7.7 become:

$$n_{\text{det}}^{e\tau} = \sum_{i=1}^{n_{\text{det}}^{\mu\tau}} \frac{\varepsilon_i^{e\tau}}{\varepsilon_i^{\mu\tau}} - \sum_{i=1}^{f^{\mu\tau}} \frac{\varepsilon_i^{e\tau}}{\varepsilon_i^{\mu\tau}} + f^{e\tau} = \sum_{i=1}^{n_{\text{det}}^{\mu\tau}} \mathcal{R}_i^{\mu\tau} - \sum_{i=1}^{f^{\mu\tau}} \mathcal{R}_i^{\mu\tau} + f^{e\tau}\,, \tag{7.8}$$

$$n_{\text{det}}^{\mu\tau} = \sum_{i=1}^{n_{\text{det}}^{e\tau}} \frac{\varepsilon_i^{\mu\tau}}{\varepsilon_i^{e\tau}} - \sum_{i=1}^{f^{e\tau}} \frac{\varepsilon_i^{\mu\tau}}{\varepsilon_i^{e\tau}} + f^{\mu\tau} = \sum_{i=1}^{n_{\text{det}}^{e\tau}} \mathcal{R}_i^{e\tau} - \sum_{i=1}^{f^{e\tau}} \mathcal{R}_i^{e\tau} + f^{\mu\tau}\,. \tag{7.9}$$

While the efficiency in the denominator corresponds to the efficiency of the respective event $i$ in the sum, the meaning of the efficiency in the numerator is not obvious: for example, in the sum over the events of the $\mu\tau$-dataset, there is no $e\tau$-event and hence also the kinematic properties needed to obtain the value of $\varepsilon_i^{e\tau}$ are not available. Therefore, the so-called *switch* event is introduced which is a virtual event made up by the switched properties of the electron and the muon of the real event $i$. The ratio of the efficiencies of the *switch* and of the real event is abbreviated by $\mathcal{R}$ as indicated in Eqs. 7.8 and 7.9. The superscript of $\mathcal{R}$ denotes the type ($e\tau$ or $\mu\tau$) of the real event in the ratio.

The relations in Eqs. 7.8 and 7.9 are utilised to estimate the background contributions in the $e\tau$- and $\mu\tau$-dataset, respectively.

[6] If not noted otherwise, *dataset* refers to the dataset of detected events.

### 7.2.2 Estimation of the *Fake*-Lepton Contribution

In addition to the $e/\mu$-symmetric, irreducible background contributions, events where either one or both leptons are mimicked of *faked* by another object contribute to the backgrounds in this analysis. This contribution is called *fakes* and is not symmetric for electrons and muons. The background contribution from *fakes* after the *Basic Selection* accounts for approximately 13 % of all background events in the $e\tau$-dataset, and 16 % in the $\mu\tau$-dataset. When using the *Symmetry Method*, the *fake*-estimate of both datasets contributes to the total background estimate of each dataset (Eqs. 7.6 and 7.7). This makes a precise estimation of this contribution very important.

The objects falsely passing the electron selection criteria are jets, converted photons, hadronically decaying $\tau$-leptons and muons. The objects falsely passing the muon selection criteria are mainly non-prompt muons, i.e. muons that do not originate from the primary vertex but from decays within jets or hadronically decaying $\tau$-leptons. The largest contribution to both, electrons and muons, are jets mimicking light leptons that mainly originate from $W + \text{jets} \to \ell\nu/\tau\nu + \text{jets}$. A smaller contribution where both leptons are imitated by jets originates from QCD multijet production. More details on both processes can be found in Section 4.2.2.

Jets, including non-prompt leptons within jets, faking light leptons are denoted as $j \to \ell$-fakes and are estimated in a data-driven way, other sources of *fakes* are denoted as *other* $\to \ell$-fakes and are determined from simulations.

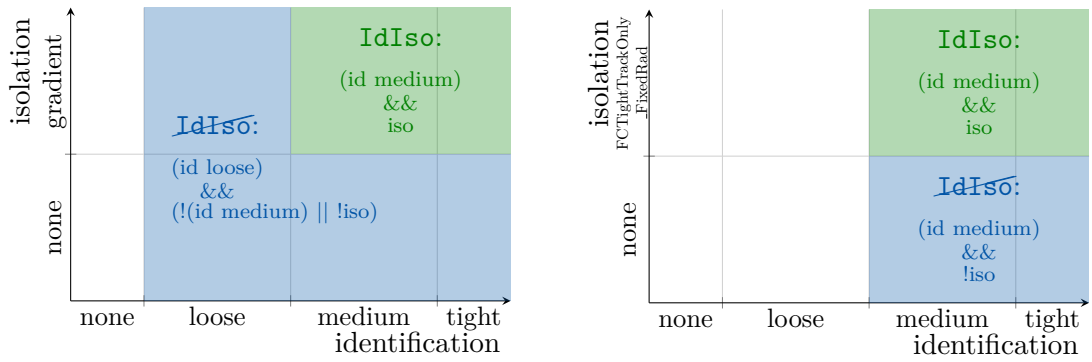
#### Estimation of Jets Faking Light Leptons

A data-driven method, the so-called *Fake Factor Method*, is utilised to estimate the background contribution from jets faking light leptons, abbreviated by $j \to \ell$-fakes. It is based on the assumption that the fake probability is a property of the fake lepton depending on its (kinematic) quantities but independent of the underlying process it stems from. The fake probablity is larger for low-energetic fake leptons and depends on the detector region it traverses. Hence, the probability is typically determined as a function of the transverse momentum $p_{\mathrm{T}}$ of the fake lepton, its pseudorapidity $\eta$ and separate for fake electrons and muons.

The *Fake Factor Method* does not directly determine and apply the actual fake probability but introduces the *Fake Factor* (FF). It is a measure of the ratio between the number of events with fake leptons fulfilling the respective lepton identification and isolation criteria (`IdIso`) and the number of events with fake leptons partially failing them, i.e. fulfilling the ~~`IdIso`~~ criteria. The definition of these criteria is shown in Figure 7.4. The `IdIso` criteria are the identification and isolation criteria defined in Chapter 5. Electrons must fail either the identification or the isolation criteria or both but still fulfil a minimum identification criterium (`id loose`) to fulfil the ~~`IdIso`~~ criteria. Muons must fail the isolation criterium to fulfil the ~~`IdIso`~~ criteria. The regions corresponding to either of the two criteria are orthogonal and hence statistically independent. The objects fulfilling the ~~`IdIso`~~ criteria are of low lepton quality and as such are likely to be *fakes* instead of real leptons. In order to determine the FFs, an *Extraction Region* (ER) enriched in events with one fake lepton is defined which is statistically independent of the *Signal Region* (SR) defined by the *Basic Selection*. In this ER region, the FF is extracted as follows:

$$\mathrm{FF} = \frac{n_{\mathrm{ER,IdIso}}}{n_{\mathrm{ER,\cancel{IdIso}}}}\,, \tag{7.10}$$

where $n_{\mathrm{ER,IdIso}}$ denotes the events where the fake lepton passes the lepton identification and isolation criteria, and $n_{\mathrm{ER,\cancel{IdIso}}}$ those events where the fake lepton fails these criteria.




**(a)** Electrons **(b)** Muons

**Figure 7.4:** Visualisation of the lepton identification and isolation criteria and their inversion for electrons (a) and muons (b). The green area corresponds to the criteria applied for *signal* leptons (Chapter 5) which is the default in the analysis, further denoted by `IdIso` (numerator of the FF) while the blue area corresponds to the ~~`IdIso`~~ criteria (denominator of the FF).

**Definition of the *Extraction Region*** The *Extraction Region* is constructed by requiring three leptons in the events, two of which must be compatible with the decay of a $Z$-boson, such that the third lepton is likely a fake lepton. Thus, the ER targets the selection of $Z$ + jets-events where the third lepton originates from a jet being mis-identified as lepton. The selection applied to obtain data enriched with $Z$ + jets-events is detailed in Table 7.5. Only requirements (denoted as *cuts*) differing from the *Basic Selection* are listed. The requirements on the leptons are the same as for the *Basic Selection* and are detailed in the respective sections of Chapter 5.

| **Extraction Region Selection** |
|---|
| *Baseline Cuts:* |
| $N_\mu + N_e = 3$<br>$p_T^{\ell i} > 15\,\text{GeV}$<br>$E_\text{T}^\text{miss} < 60\,\text{GeV}$ |
| *Z-tagging Requirements:* |
| $q\left(\ell_{Z,0}\right) \cdot q\left(\ell_{Z,1}\right) = -1$<br>$\text{flavour}\left(\ell_{Z,0}\right) = \text{flavour}\left(\ell_{Z,1}\right)$<br>$80\,\text{GeV} < m_{\ell_{Z,0}\ell_{Z,1}} < 100\,\text{GeV}$ |
| *Requirements on the probe lepton:* |
| $m_T\left(\ell_\text{p}, E_\text{T}^\text{miss}\right) < 40\,\text{GeV}$ |

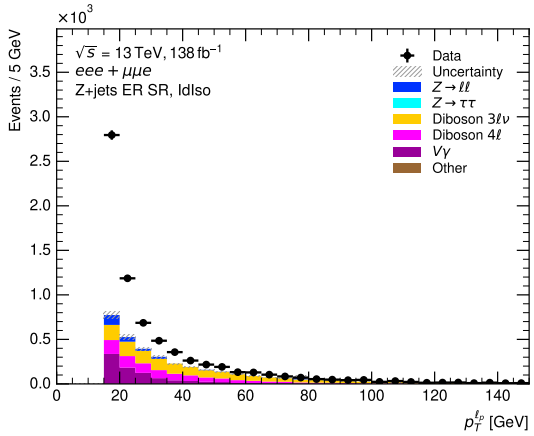
**Table 7.5:** The *Extraction Region Selection*. Only cuts differing from the *Basic Selection* (Table 7.2) are listed. $N$ stands for *number*, $q$ for *charge*, the two leptons fulfilling the $Z$-tagging requirements are denoted by $\ell_{Z,0}$ and $\ell_{Z,1}$ while $\ell_p$ denotes the third, probe lepton. The definition of $m_T$ is given in Eq. 8.1.

Two of the three leptons, $\ell_{Z,0}$ and $\ell_{Z,1}$, are assigned to originate from the $Z$-boson by requiring opposite charge, same flavour and an invariant mass between 80 GeV and 100 GeV. If more than one lepton pair fulfils these requirements, the two leptons with their invariant mass, $m_{\ell\ell}$, closest to the $Z$-boson mass are $Z$-tagged. The third lepton is the *probe* lepton, $\ell_p$, which is likely to be faked by a jet and that is utilised to determine the FF-values. The cuts on $E_\text{T}^\text{miss}$ and $m_T\left(\ell_\text{p}, E_\text{T}^\text{miss}\right)$ (defined in Eq. 8.1) are applied to reduce contributions from $WZ$-events. The single electron and single muon triggers listed in Table 7.1 are used to select the events of interest. For the trigger-matched lepton the

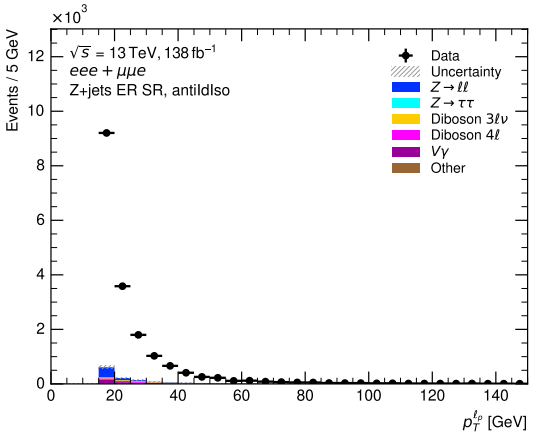
recommended $p_{\mathrm{T}}$-thresholds are applied. One of the $Z$-tagged leptons is required to have triggered the event such that the *probe* lepton is unbiased by possible tighter trigger object criteria which might be in conflict with the ~~IdIso~~-requirements.

The events passing the ER selection criteria are split into four datasets depending on the number of electrons and muons in the event. The naming of the datasets is of the form "$\ell_{Z,0}\ell_{Z,1}\ell_p$". Hence, the $eee$- and $\mu\mu e$-datasets are used to derive the electron FFs, while the $\mu\mu\mu$- and $ee\mu$-datasets are used to derive the muon FFs.

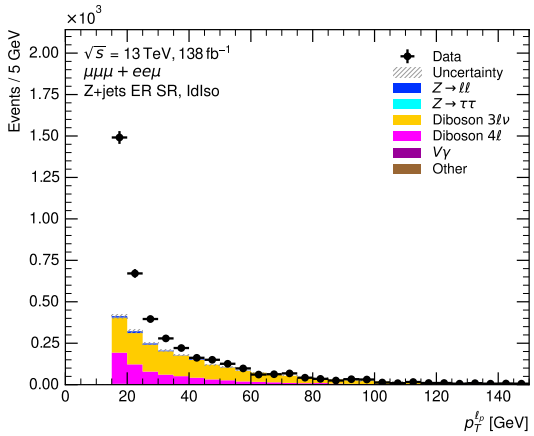
The distributions of the transverse momentum of the *probe* lepton in the $(eee + \mu\mu e)$- and in the $(\mu\mu\mu + ee\mu)$-datasets in the ER are shown in Figure 7.5. The contribution



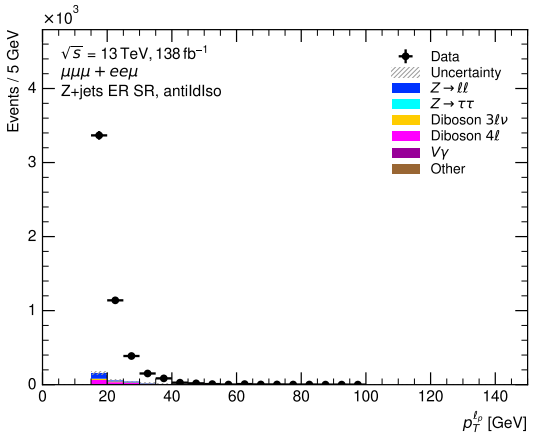

**(a)** $(eee+\mu\mu e)$-datasets with $\ell_p$ passing the `IdIso` criteria.




**(b)** $(eee+\mu\mu e)$-datasets with $\ell_p$ passing the ~~IdIso~~ criteria.




**(c)** $(\mu\mu\mu + ee\mu)$-datasets with $\ell_p$ passing the `IdIso` criteria.




**(d)** $(\mu\mu\mu + ee\mu)$-datasets with $\ell_p$ passing the ~~IdIso~~ criteria.

**Figure 7.5:** The transverse momentum of the *probe* lepton, $p_{\mathrm{T}}^{\ell_p}$, in the *Extraction Region* when $\ell_p = e$ (a), (b), when $\ell_p = \mu$ (c), (d) and whether it is passing the `IdIso` criteria (a), (c) or the ~~IdIso~~ (antiIdIso) criteria (b), (d). The coloured histograms represent the contributing background processes, i.e. processes where none of the leptons is mimicked by a jet, which are estimated from simulations. The "Other" category comprises very small contributions from $t\bar{t}W, t\bar{t}Z, WH$ and Triboson production. Credits to Valerie Lang for the inputs [229].

from processes where none of the three leptons is mimicked by a jet (coloured histograms) is – as expected – larger when the `IdIso` criteria are applied for the *probe* lepton than when the ~~IdIso~~ criteria are applied. The white area between the coloured histograms and the data corresponds to events where the third lepton is likely to be faked by a jet. Hence, the FFs are the ratios of these white areas in the `IdIso`- and the ~~IdIso~~-selection (Eq. 7.10). Furthermore, due to the smaller ~~IdIso~~ region for muons compared to electrons

(Figure 7.4), the yields for the $(\mu\mu\mu + ee\mu)$-datasets in the ER region are lower compared to the $(eee + \mu\mu e)$-datasets which increases the statistical uncertainties on the FF for muons. The main (real lepton) background processes represented by the coloured histograms are: $WZ$, $ZZ$. The first contributes three real leptons, the second contributes four real leptons although one can be undetected. Additional backgrounds are: $V\gamma$, $Z \to \ell\ell$ and $Z \to \tau\tau$, where the third lepton must be faked, though not originating from a jet faking a lepton. This is ensured through a truth matching that is applied with the help of the *IFFTruthClassifier*-tool[7] [230]. In particular, for electron fakes, the $V\gamma$-process where the third lepton originates from a photon faking an electron is an important process to consider properly. The estimation of the $\gamma \to e$-fakes from simulated events was found to provide a better description of data in fake-dominated regions than an estimation in combination with the data-driven $j \to \ell$-fakes. Hence, the modelling of this process from simulated events was chosen.

**Extraction and Application of the *Fake Factor*s** The values of the FFs are extracted for electrons and muons individually and in bins of $p_{\mathrm{T}}^{\ell_p}$ and $\Delta\phi(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}})$ for electrons, and in bins of $p_{\mathrm{T}}^{\ell_p}$ only for muons. The dependency of the FF on $\Delta\phi(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}})$ is flat for muons, and hence no binning in $\Delta\phi(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}})$ is needed for them. A binning of the FFs for electrons in $p_{\mathrm{T}}^{\ell_p}$ and $|\eta^{\ell_p}|$ was also investigated. More details can be found in [231]. But the estimation of the $\gamma \to e$-fakes from simulated events, subtracting its contribution from the $j \to \ell$-fakes in the ER, led to the disappearance of the dependence of the electron-FF on $|\eta^{\ell_p}|$ and instead a dependence on $\Delta\phi(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}})$.

Subtracting the contributions from other backgrounds except $j \to \ell$-fakes and considering the binning in $p_{\mathrm{T}}^{\ell_p}$ and $\Delta\phi(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}})$, the definition of the FFs becomes:

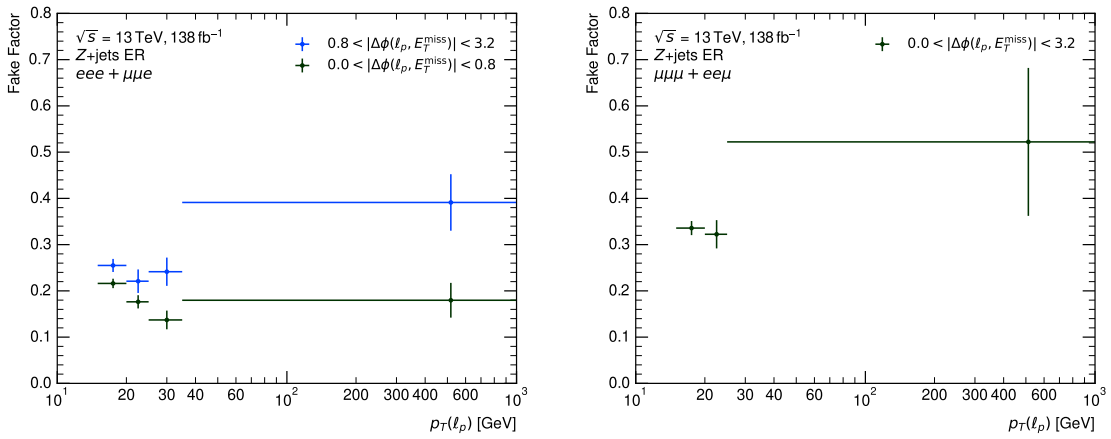
$$\mathrm{FF}\left(p_{\mathrm{T}}^{\ell_p}, \Delta\phi\left(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}}\right)\right) = \frac{n_{\mathrm{ER},\texttt{IdIso}}^{\mathrm{data}}\left(p_{\mathrm{T}}^{\ell_p}, \Delta\phi\left(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}}\right)\right) - n_{\mathrm{ER},\texttt{IdIso}}^{\mathrm{MC}}\left(p_{\mathrm{T}}^{\ell_p}, \Delta\phi\left(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}}\right)\right)}{n_{\mathrm{ER},\cancel{\texttt{IdIso}}}^{\mathrm{data}}\left(p_{\mathrm{T}}^{\ell_p}, \Delta\phi\left(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}}\right)\right) - n_{\mathrm{ER},\cancel{\texttt{IdIso}}}^{\mathrm{MC}}\left(p_{\mathrm{T}}^{\ell_p}, \Delta\phi\left(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}}\right)\right)}. \tag{7.11}$$

The bin edges used for $p_{\mathrm{T}}^{\ell_p}$ in GeV are $[15, 20, 25, 35, \infty]$ for electrons and $[15, 20, 25, \infty]$ for muons. The bin edges used for $\Delta\phi\left(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}}\right)$ for electrons are $[0.0, 0.8, 3.2]$. The resulting values of the FFs for electrons and muons are shown in Figure 7.6. Overall, the FF-values for electrons are smaller than for muons due to the larger yield in the ER-selection with the $\cancel{\texttt{IdIso}}$-requirements for electrons. The origin of the uncertainties is discussed further below.

In order to obtain the *fake*-estimate in the *Signal Region*, the FFs are applied to events passing the *Basic Selection* where at least one of the two leptons in the final state must fulfil the $\cancel{\texttt{IdIso}}$- instead of the usual `IdIso`-requirements. Depending on whether the $\cancel{\texttt{IdIso}}$-lepton is identified as electron or muon, the respective FF-value is applied, thereby also considering its $p_{\mathrm{T}}$- and $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$-values. This approach implies that events where both, electron and muon, result from a jet faking the lepton, are counted twice. Thus, the contribution of these so-called *double-fakes* needs to be subtracted once.

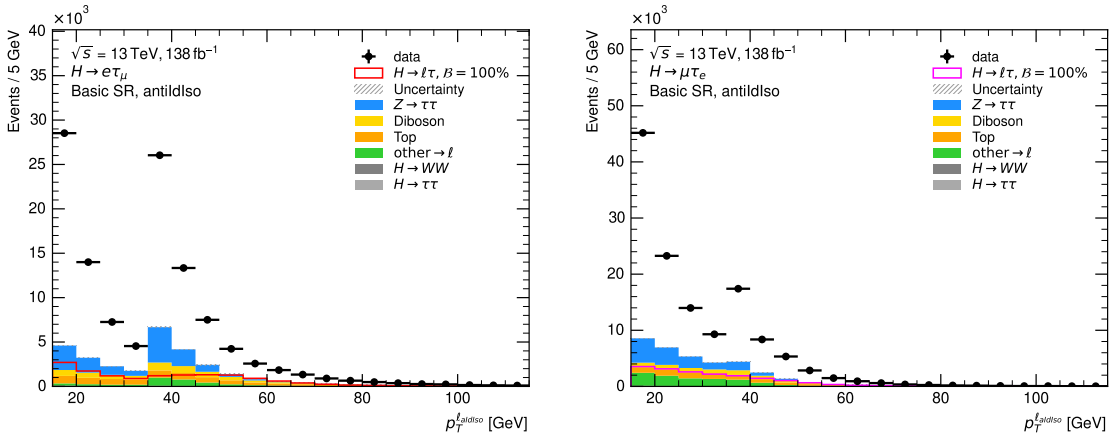
Contributions from events where none of the leptons originate from a jet mis-identified as lepton must be subtracted in the $\cancel{\texttt{IdIso}}$-SR equivalently to the subtraction performed in the ER. To do so, the same truth matching requirements as in the ER are imposed. The respective distributions of the $\cancel{\texttt{IdIso}}$-lepton of the $e\tau$- and $\mu\tau$-datasets in the SR are shown in Figure 7.7. In the $e\tau$-dataset, the majority of the $\cancel{\texttt{IdIso}}$-leptons are muons while in the

[7] Each lepton must belong to either of the following *IFFTruthClassifier*-classes: 2 (prompt electrons), 3 (charge-flip electrons), 4 (prompt muons), 5 (prompt photon-conversion), 6 (electrons from muons), 7 (electrons and muons from hadronic tau-decays).




**(a)** **(b)**

**Figure 7.6:** The extracted values for the FFs for electrons (a) and muons (b) with the statistical and systematic uncertainties added in quadrature. Credits to Valerie Lang for the inputs [229].




**(a)** **(b)**

**Figure 7.7:** The distributions of the ~~IdIso~~-lepton of the $e\tau$-dataset (a) and the $\mu\tau$-dataset (b) in the *Basic Selection* SR where at least one of the two leptons in the final state must fulfil the ~~IdIso~~-criteria. If both leptons in the event fulfil these criteria, the $p_{\mathrm{T}}$ of the lepton with the smaller $p_{\mathrm{T}}$ is chosen to be shown in the distributions. Contributions from events where none of the leptons originates from a jet mis-identified as lepton (estimated from simulations) are displayed as coloured histograms. The uncertainty band solely consists of statistical uncertainties on the prediction.

$\mu\tau$-dataset they are electrons. That is because the probability for fake-leptons is larger with smaller $p_{\mathrm{T}}$ and in the majority of events in the $e\tau(\mu\tau)$-dataset the muon (electron) has the smaller $p_{\mathrm{T}}$. The double-peak structure arises from the two different $p_{\mathrm{T}}$-thresholds for the leading ($p_{\mathrm{T}} > 35\,\mathrm{GeV}$) and sub-leading ($p_{\mathrm{T}} > 15\,\mathrm{GeV}$) lepton since both can be the ~~IdIso~~-lepton. The events corresponding to the white areas between the background processes and the data serve as templates to which the FFs are applied. The *fake*-estimate

in the SR where both leptons fulfil the IdIso-requirements is obtained as follows:

$$
\begin{aligned}
f_{\mathrm{SR,IdIso}} = & \underbrace{\left( \sum_{i=1}^{n^{\mathrm{data}}_{\mathrm{SR,\cancel{IdIso}\text{-}e}}} \mathrm{FF}_i^e - \sum_{i=1}^{n^{\mathrm{MC}}_{\mathrm{SR,\cancel{IdIso}\text{-}e}}} \mathrm{FF}_i^e \right)}_{\text{electron fakes}} \\
& + \underbrace{\left( \sum_{i=1}^{n^{\mathrm{data}}_{\mathrm{SR,\cancel{IdIso}\text{-}\mu}}} \mathrm{FF}_i^\mu - \sum_{i=1}^{n^{\mathrm{MC}}_{\mathrm{SR,\cancel{IdIso}\text{-}\mu}}} \mathrm{FF}_i^\mu \right)}_{\text{muon fakes}} \\
& - \underbrace{\left( \sum_{i=1}^{n^{\mathrm{data}}_{\mathrm{SR,\cancel{IdIso}\text{-}e\mu}}} \mathrm{FF}_i^e \cdot \mathrm{FF}_i^\mu - \sum_{i=1}^{n^{\mathrm{MC}}_{\mathrm{SR,\cancel{IdIso}\text{-}e\mu}}} \mathrm{FF}_i^e \cdot \mathrm{FF}_i^\mu \right)}_{\text{double fakes}}
\end{aligned}
\tag{7.12}
$$

with $\mathrm{FF}_i^\ell$ being the short form of $\mathrm{FF}^\ell\left(p_{T,i}^\ell, \Delta\phi_i\left(\ell, E_{\mathrm{T}}^{\mathrm{miss}}\right)\right)$ and $\cancel{\texttt{IdIso}}$-$\{e, \mu, e\mu\}$ indicating which of the leptons in the $\cancel{\texttt{IdIso}}$-SR fulfils the $\cancel{\texttt{IdIso}}$ criteria. The number of data events in the $\cancel{\texttt{IdIso}}$-SR are denoted by $n^{\mathrm{data}}_{\mathrm{SR},\cancel{\texttt{IdIso}}}$, the number of events where none of the leptons originate from a jet mis-identified as lepton (estimated from simulated events) is denoted by $n^{\mathrm{MC}}_{\mathrm{SR},\cancel{\texttt{IdIso}}}$. The last line takes care of removing the *double-fake* contribution once.

**Uncertainties on the *Fake Factor*s** Different sources of uncertainties contribute to the overall uncertainty on the FFs displayed in Figure 7.6:

- statistical uncertainties on $n^{\mathrm{data}}_{\mathrm{ER,IdIso}}$ and on $n^{\mathrm{data}}_{\mathrm{ER},\cancel{\texttt{IdIso}}}$
- statistical uncertainties on $n^{\mathrm{MC}}_{\mathrm{ER,IdIso}}$ and on $n^{\mathrm{MC}}_{\mathrm{ER},\cancel{\texttt{IdIso}}}$
- systematic uncertainties on $n^{\mathrm{MC}}_{\mathrm{ER,IdIso}}$ and on $n^{\mathrm{MC}}_{\mathrm{ER},\cancel{\texttt{IdIso}}}$ [232]:
  - variation of the cross-section of $WZ$ process by $\pm 7.1\,\%$
  - variation of the cross-section of $ZZ$ process by $\pm 6.0\,\%$

The statistical uncertainties on the single components of the ratio (Eq. 7.11) are propagated to the final FF-values. For each systematic uncertainty source, the variation is applied simultaneously in the numerator and denominator of Eq. 7.11 and the FFs are re-calculated. The difference between this varied FF-value and the nominal value is taken as the uncertainty. The statistical uncertainties on the electron-FFs range from 5 % in the lowest $p_{\mathrm{T}}$-bins to 15 % in the highest $p_{\mathrm{T}}$-bins while they range from 4 % in the lowest $p_{\mathrm{T}}$-bin to 15 % in the highest $p_{\mathrm{T}}$-bin for the muon-FFs. The systematic uncertainties due to the variation of the $WZ$ cross-section range from 0.4 to 14 % for the electron-FFs and from 1.3 % to 26 % in the last $p_{\mathrm{T}}$-bin for the muon-FFs. The systematic uncertainties due to the variation of the $ZZ$ cross-section range from 0.3 to 5 % for the electron-FFs and from 0.9 % to 7 % for the muon-FFs. In Figure 7.6, the statistical and systematic components are added in quadrature and the square-root is taken to obtain the overall uncertainty.

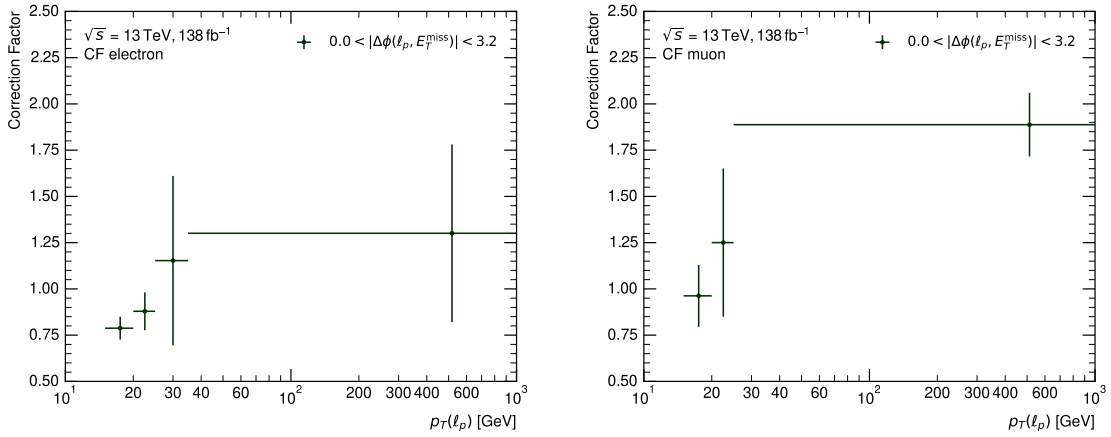
**Correction Factors** One additional effect that needs to be considered is the possibility of a different flavour composition of the jets faking leptons in the $Z$ + jets-ER and the SR. This was studied and more details can be found in Ref. [233].

The differences in the flavour composition can lead to differing values of the FFs in the $Z$ + jets-ER and the SR since it is expected that the values of the FFs vary for each flavour.

Hence, so-called *Correction Factor*s (CFs) are derived from simulated events where the origin of different fake flavours can be traced. These are calculated as the ratio of the FF in the SR and in the ER where the FF-values are determined from MC-simulated samples:

$$\mathrm{CF} = \frac{\mathrm{FF}_{\mathrm{SR}}^{\mathrm{MC}}}{\mathrm{FF}_{\mathrm{ER}}^{\mathrm{MC}}} \,. \tag{7.13}$$

Although FFs derived from simulations are of limited trustworthiness, it is assumed that the flavour composition and hence the ratio of the FFs is modelled sufficiently well. In addition, a systematic uncertainty on the CF-values is determined by deriving the same CFs with samples that were simulated with another MC-generator. The difference to the nominal CFs is used as uncertainty. The statistical uncertainties on the samples used to calculate the nominal CFs are propagated to the CFs. The resulting CFs together with the systematic and statistical uncertainties added in quadrature are visualised in Figure 7.8. The CFs are derived independently for electron- and muon-fakes due to differing flavour




**(a)** **(b)**

**Figure 7.8:** The extracted values for the CFs for electrons (a) and muons (b) with the statistical and systematic uncertainties added in quadrature. Credits to Mattias Birman for the inputs [234].

compositions and are only binned in $p_{\mathrm{T}}$ using the same binning as for the FFs determined in data in the $Z$ + jets-ER. They are applied in addition to the FFs in Eq. 7.12 to obtain the *fake*-estimate in the SR. Their values vary around 1 for electrons, although for small $p_{\mathrm{T}}$ (and hence for the majority of the fake electrons) they are below 1. For muons, the CF-values in the first two bins are compatible with 1 while it is almost 2 for $p_{\mathrm{T}} > 25\,\mathrm{GeV}$. In addition, large uncertainties of up to 40 % for electrons and up to 30 % for muons are visible. Apart from the last bin of the muon CFs, the systematic component is always larger than the statistical component.

**Validation of the jet $\rightarrow \ell$-fake estimate** The quality of the jet $\rightarrow \ell$-fake estimate is investigated by checking how well the data in different regions is modelled by the prediction.

One straight-forward check is the so-called *closure test* in the ER. The FFs obtained from the ER are applied to the events in the ER that are in the denominator of the ratio used to obtain the FFs (Eq. 7.11). The sum of the resulting contribution (the *fake*-estimate) and of all processes with three leptons of which none originates from a jet (estimated from MC-simulations) in the `IdIso`-ER are compared to data. By construction, the modelling should be perfect in the distribution of $p_{\mathrm{T}}^{\ell_p}$ when the same binning as for the FFs is chosen. Other distributions should demonstrate decent agreement if the chosen binning of the

FFs capture the kinematic properties of the fake background. Several observables are investigated and displayed in Ref. [233] and good agreement is found.

After the discussion of the estimation of the *other* $\rightarrow \ell$-fakes in the following paragraph, the *fake*-estimate is validated in a control region that is orthogonal to the *Basic Selection* SR and in the SR itself.

Overall, good performance of the $j \rightarrow \ell$-fake background is seen in the different validation regions and the SR, indicating a valid fake background estimate.

#### Estimation of Other Objects Faking Light Leptons

The background from hadronically decaying $\tau$-leptons that are mis-identified as leptons ($\tau_{\text{had}} \rightarrow \ell$), from muons that are mis-identified as electrons ($\mu \rightarrow e$) and from photons that are mis-identified as electrons ($\gamma \rightarrow e$) is estimated from MC-simulations and is denoted as *other* $\rightarrow \ell$-fakes. It corresponds to the remaining part of the fake backgrounds which is not covered by the data-driven *Fake Factor Method* which only describes $j \rightarrow \ell$-fakes. The $\tau_{\text{had}} \rightarrow \ell$- and $\mu \rightarrow e$-fakes are absent in the *Extraction Region* used for the *Fake Factor Method*. For $\gamma \rightarrow e$-fakes, the modelling of the data by the prediction in control regions is improved when not including the them in the *Fake Factor Method* but instead estimating them separately from MC-simulations – also see the discussion above.

Events for the *other* $\rightarrow \ell$-fake background estimate are selected according to the *Basic Selection*, based on the simulation of the following processes: $Z \rightarrow \tau\tau$, $t\bar{t}$, single-top, diboson (including $V\gamma$), $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$. Contributions from $H \rightarrow \tau\tau$ and $H \rightarrow WW$ are neglected as their contribution is small. A truth-matching[8] is applied to the simulated events to ensure that only the three respective classes of *fakes* ($\tau_{\text{had}} \rightarrow \ell$, $\mu \rightarrow e$, $\gamma \rightarrow e$) are selected. The same truth-matching is applied as part of the *Fake Factor Method* to ensure that these contributions are subtracted as background such that they do not enter the $j \rightarrow \ell$-fake estimate to avoid double-counting. At least one lepton has to fulfil the truth-matching requirements in order for the event to be classified as *other* $\rightarrow \ell$-fake.

The corresponding yields for each class of *other* $\rightarrow \ell$-fakes after the *Basic Selection*-SR are summarised in Table 7.6. The largest contribution to the *other* $\rightarrow \ell$-fakes comes from $Z \rightarrow \mu\mu$, where one muon (typically the one with the lower transverse momentum) is mis-identified as an electron. Thus, the $Z \rightarrow \mu\mu$-process contributes more events to the $\mu\tau$-dataset than to the $e\tau$-dataset. Other processes only contribute little to the $\mu \rightarrow e$-fakes. The $\tau_{\text{had}} \rightarrow \ell$-fakes contribute more to the $e\tau$- than to the $\mu\tau$-dataset. Their largest contribution is from $Z \rightarrow \tau\tau$ and $V\gamma$. The $\gamma \rightarrow e$-fakes are dominated by $V\gamma$ and contribute more to the $\mu\tau$-dataset where the electron is sub-leading in most of the events.

**$Z \rightarrow \mu\mu$ Scaling Uncertainty** The largest contribution to the *other* $\rightarrow \ell$-fakes is from $Z \rightarrow \mu\mu$-events as discussed above. Hence, the modelling of its normalisation is studied in more detail.

The process mainly contributes to the $\mu\tau$-dataset where the electron is sub-leading in the majority of the events (if $\ell_1 = \ell_\tau$, see Section 7.1.3). Therefore, only the $\mu\tau$-sample is utilised to study the $Z \rightarrow \mu\mu$-background.

In order to check whether the $\mu \rightarrow e$-fake rate is well modelled by the prediction of the MC-simulations, a dedicated *Validation Region* (VR) is defined. A high purity of $Z \rightarrow \mu\mu$-events is desired such that the expected yield can be compared to the corresponding data yield. In addition to the *Basic Selection* (Table 7.2), the following criteria are required for the $Z \rightarrow \mu\mu$-VR:

- $35\,\text{GeV} < p_{\text{T}}^{\ell_0} < 45\,\text{GeV}$

[8] each lepton must belong to either of the following *IFFTruthClassifier*-classes: 5 (prompt photon-conversion), 6 (electrons from muons), 7 (electrons and muons from hadronic tau-decays).

| | $\mu \to e$-fakes | | $\tau_{\text{had}} \to \ell$-fakes | | $\gamma \to e$-fakes | |
|---|---|---|---|---|---|---|
| Process | $e\tau$ | $\mu\tau$ | $e\tau$ | $\mu\tau$ | $e\tau$ | $\mu\tau$ |
| Top | 2.10 ± 0.35 | 5.70 ± 0.60 | 5.40 ± 0.70 | 2.60 ± 0.40 | 5.10 ± 0.60 | 8.50 ± 0.80 |
| Diboson | 4.10 ± 0.50 | 8.30 ± 0.80 | 40.40 ± 1.30 | 36.10 ± 1.30 | 0.02 ± 0.02 | 0.00 ± 0.00 |
| $V\gamma$ | 31 ± 6 | 65 ± 12 | 1 580 ± 50 | 1 290 ± 40 | 786 ± 33 | 2 140 ± 60 |
| $Z \to \tau\tau$ | 2.20 ± 0.60 | 2.20 ± 0.70 | 2 021 ± 25 | 726 ± 11 | 0.00 ± 0.00 | 0.00 ± 0.00 |
| $Z \to \mu\mu$ | 1 363 ± 24 | 6 770 ± 40 | 0.60 ± 0.40 | 2.00 ± 0.90 | 20.40 ± 2.30 | 91 ± 12 |
| $Z \to ee$ | 0.00 ± 0.00 | 0.00 ± 0.00 | 9 ± 7 | −0.10 ± 0.50 | 0.00 ± 0.00 | 0.00 ± 0.00 |
| $H \to \tau\tau$ | 0.01 ± 0.01 | 0.03 ± 0.01 | 0.02 ± 0.01 | 0.01 ± 0.00 | 0.07 ± 0.02 | 0.05 ± 0.01 |
| $H \to WW$ | 0.04 ± 0.01 | 0.03 ± 0.01 | 0.00 ± 0.00 | 0.00 ± 0.00 | 0.01 ± 0.01 | 0.03 ± 0.01 |

**Table 7.6:** Contributions of the individual processes to the *other* $\to \ell$-fake estimate after the *Basic Selection*, separate for $\tau_{\text{had}} \to \ell$-, $\mu \to e$- and $\gamma \to e$-fakes and for the $e\tau$- and $\mu\tau$-datasets. $V\gamma$ is listed separately from the other Diboson processes ($WW, WZ, ZZ$). The contributions from $H \to \tau\tau$ and $H \to WW$ are neglected in the analysis.

- $75\,\text{GeV} < m_{\ell\ell} < 100\,\text{GeV}$ with $m_{\ell\ell}$ being the invariant mass of the two leptons and the selected window covering the expected $Z$-peak
- $1.25 <$ trk-$p_\text{T}$/cluster-$p_\text{T}$ $< 3$ for electrons, with trk-$p_\text{T}$ being the $p_\text{T}$ measured in the ID and cluster-$p_\text{T}$ being the transverse energy measured in the ECAL; the ratio of these variables shows excellent separation power between $Z \to \mu\mu$ and the remaining processes. A muon mis-identified as electron deposits enough of its energy in the ECAL such that it is identified as an electron. But its trk-$p_\text{T}$ is larger than its cluster-$p_\text{T}$ as it does not deposit all of its energy in the ECAL and hence trk-$p_\text{T}$/cluster-$p_\text{T}$ $> 1$, see Figure 7.9.
- $m_\text{coll} < 115\,\text{GeV}$ with $m_\text{coll} = \sqrt{2p_\text{T}^{\ell_H}\left(p_\text{T}^{\ell_\tau} + E_\text{T}^\text{miss}\right)(\cosh\Delta\eta - \cos\Delta\phi)}$ aiming to reconstruct the Higgs-boson mass of the LFV decays; more details are given in Section 8.2.3

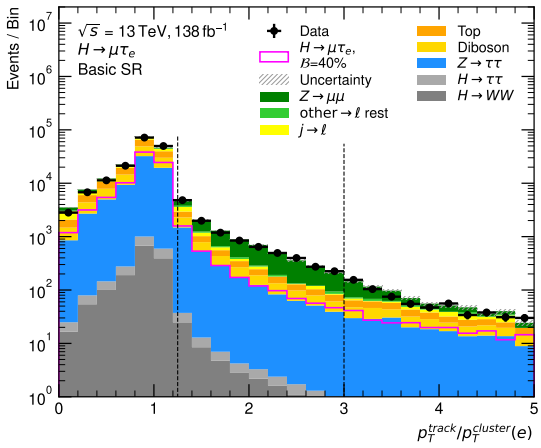



**Figure 7.9:** Distribution of the trk-$p_T$/cluster-$p_T$-variable in the $\mu\tau$-dataset, passing the *Basic Selection*. "*other* $\to \ell$ rest" in the legend is *other* $\to \ell$-fakes without $Z \to \mu\mu$ which is shown separately. Only statistical uncertainties are displayed.

The purity of $Z \to \mu\mu$-events in the VR is calculated as (data−(bkg w/o $Z \to \mu\mu$))/data where bkg stands for background. A value of $(68 \pm 5)\,\%$ is reached with the above selection. The yields of the single processes are summarised in Table 7.7.

Different kinematic distributions in the $Z \to \mu\mu$-VR are shown in Figure 7.10. Overall, an overshoot of the prediction compared to data is observed. The required normalisation correction, defined as (data−(bkg w/o $Z \to \mu\mu$))/$Z \to \mu\mu$, corresponds to $0.84 \pm 0.06$. The difference to unity (16 %) is applied as systematic uncertainty on the $Z \to \mu\mu$-normalisation. To be conservative, this uncertainty is applied to the total *other* $\to \ell$-fake contribution and not only to its $Z \to \mu\mu$-part.

### Validation of the Estimation of the *Fake*-Lepton Contribution

The *fake*-estimate is validated in a control region that is orthogonal to the *Basic Selection* SR by requiring the two leptons to have the same electric charge (SC) instead of the opposite charge (OC). All other requirements are the same as for the *Basic Selection*. The jet $\to \ell$-fake estimate in this SC-SR is obtained equivalently to the default OC-SR by utilising Eq. 7.12. The CFs, which are applied as well, are specifically derived for the SC-SR, again in the same way as for the OC-SR. The ratio of jet $\to \ell$- (*other* $\to \ell$)-fakes with respect to the total background is 53 % (12 %) in the $e\tau$-dataset and 51 % (25 %) in the $\mu\tau$-dataset. The modelling is validated by comparing the sum of the jet $\to \ell$-fake estimate,

| Process | $Z \to \mu\mu$-VR, $\mu\tau$ Yield |
|---|---|
| $H \to e\tau_\mu$ | $0.17 \pm 0.04$ |
| $H \to \mu\tau_e$ | $1.22 \pm 0.10$ |
| Top | $6.72 \pm 0.67$ |
| Diboson | $36.57 \pm 2.07$ |
| $Z \to \tau\tau$ | $127.79 \pm 5.01$ |
| $j \to \ell$ | $11.89 \pm 7.69$ |
| other $\to \ell$ w/o $Z \to \mu\mu$ | $24.51 \pm 6.18$ |
| $H \to \tau\tau$ | $1.57 \pm 0.07$ |
| $H \to WW$ | $0.24 \pm 0.05$ |
| $Z \to \mu\mu$ | $532.99 \pm 11.61$ |
| Data | 658 |

**Table 7.7:** Contributions of the individual processes in the $Z \to \mu\mu$-VR.

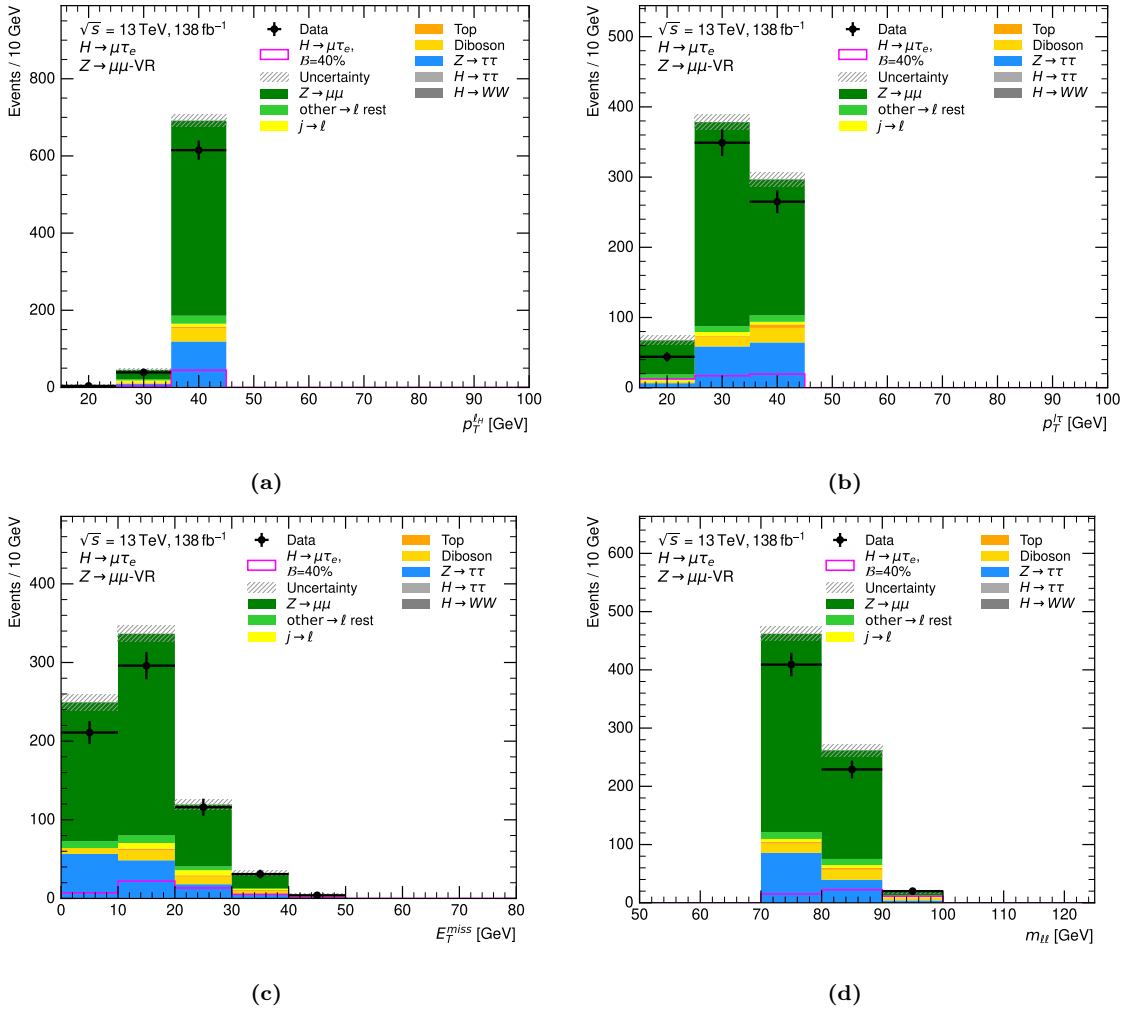



**Figure 7.10:** Kinematic distributions in the $Z \to \mu\mu$-VR in the $\mu\tau$-dataset: $p_{\mathrm{T}}^{\ell_H}$ (a), $p_{\mathrm{T}}^{\ell_\tau}$ (b), $E_{\mathrm{T}}^{\mathrm{miss}}$ (c) and $m_{\ell\ell}$ (d). The "*other* $\to \ell$ rest" label in the legend refers to the *other* $\to \ell$-fakes contribution without the $Z \to \mu\mu$-contribution which is shown separately. Only statistical uncertainties are displayed.

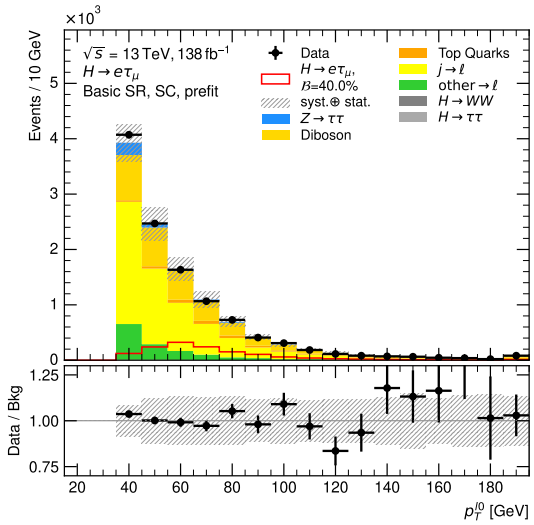
the *other* $\rightarrow \ell$-fake estimate and the contribution of events with two real leptons to data. The events with two real leptons are estimated from MC-simulations[9] and not yet with the *Symmetry Method* to disentangle the validation of the *fake*-estimate from the *Symmetry Method*. The normalisation of the predictions from MC-simulations is only based on the theoretical cross-sections for the individual processes, though. No specific normalisation factors are derived in dedicated control regions.

In addition, the same validation of modelling is conducted in the OC-SR (the default signal region). This region counts as validation region nevertheless since in the final statistical analysis the contributions from events with two real leptons are estimated with the *Symmetry Method* rather than from simulations.

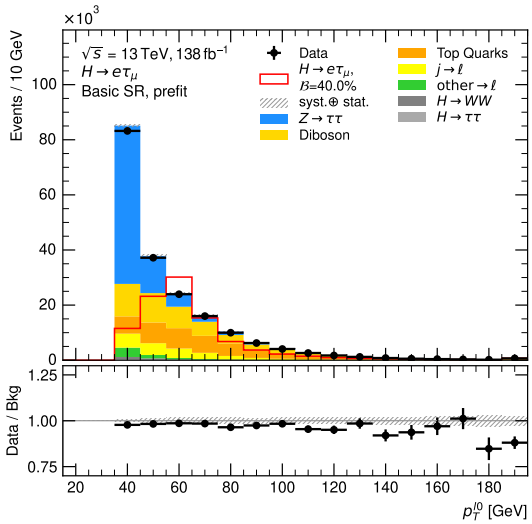
Kinematic distributions in both regions are shown in Figures 7.11 and 7.12 for the $e\tau$-dataset (where the majority of the fake leptons are muons) and in Figures 7.13 and 7.14 for the $\mu\tau$-dataset (where the majority of the fake leptons are electrons). More distributions are shown in Appendix A. Blinding of bins is performed as described in Chapter 3.

The overall modelling is convincing. Small trends in the ratio of data over background prediction shown in the lower panels of the figures are visible for some distributions in the SC-SR. For example the distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$ (Figures 7.11(e) and 7.13(e)), $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$ (Figures 7.12(c) and 7.14(c)) and $m_{\ell\ell}$ (Figures 7.12(a) and 7.14(a)) show slopes in the range of the lower values of the observables but are mostly within uncertainties. For some of the distributions similar slopes are visible in the OC-SR, e.g. in $m_{\ell\ell}$ (Figures 7.14(b) and 7.14(b)) but in general the agreement between prediction and data in the OC-SR is good within uncertainties, especially when considering that no systematic uncertainties on the background processes estimated from simulations are considered.

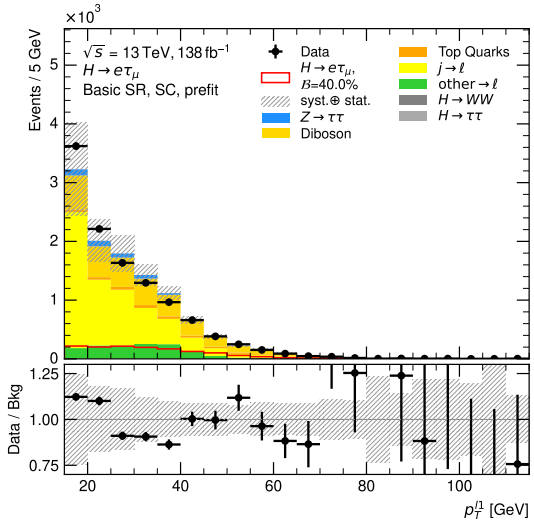
[9] each lepton must belong to either of the following *IFFTruthClassifier*-classes [230]: 2 (prompt electrons), 3 (charge-flip electrons), 4 (prompt muons) to ensure the selection of two real leptons



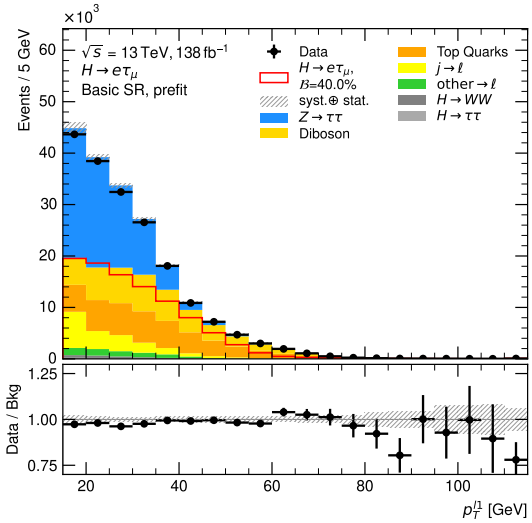

**(a)** $p_{\mathrm{T}}(\ell_0)$, Basic SR, SC requirement



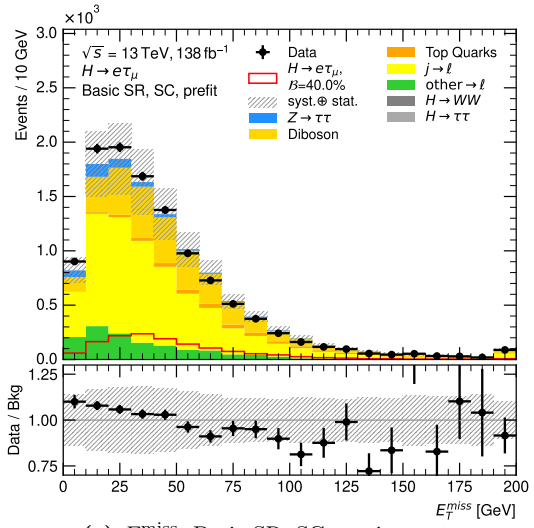

**(b)** $p_{\mathrm{T}}(\ell_0)$, Basic SR, OC requirement



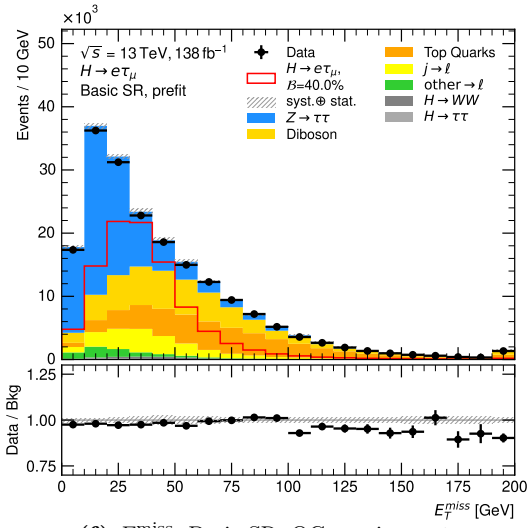

**(c)** $p_{\mathrm{T}}(\ell_1)$, Basic SR, SC requirement




**(d)** $p_{\mathrm{T}}(\ell_1)$, Basic SR, OC requirement

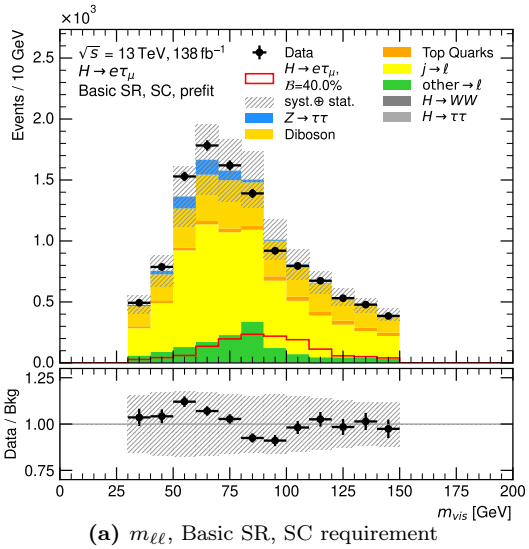



**(e)** $E_{\mathrm{T}}^{\mathrm{miss}}$, Basic SR, SC requirement

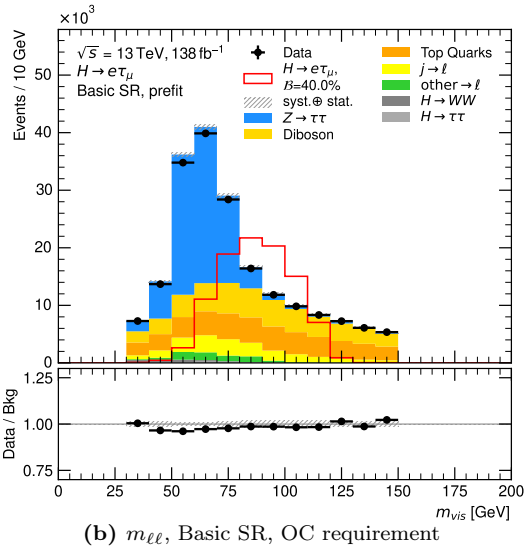



**(f)** $E_{\mathrm{T}}^{\mathrm{miss}}$, Basic SR, OC requirement

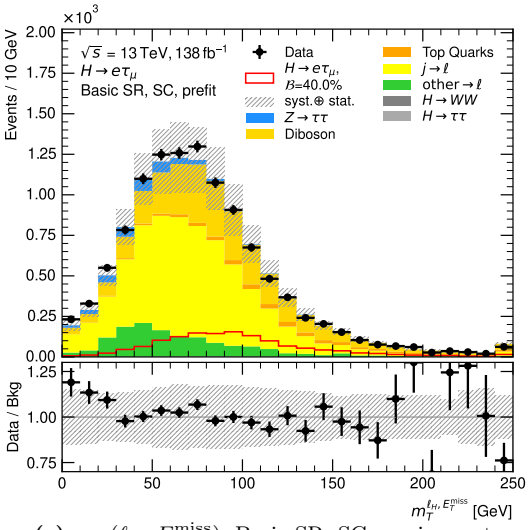
**Figure 7.11:** Distributions of $p_{\mathrm{T}}^{\ell_0}$ (a,b), $p_{\mathrm{T}}^{\ell_1}$ (c,d), $E_{\mathrm{T}}^{\mathrm{miss}}$ (e,f) in the $e\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining *other* $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate as discussed in the text.



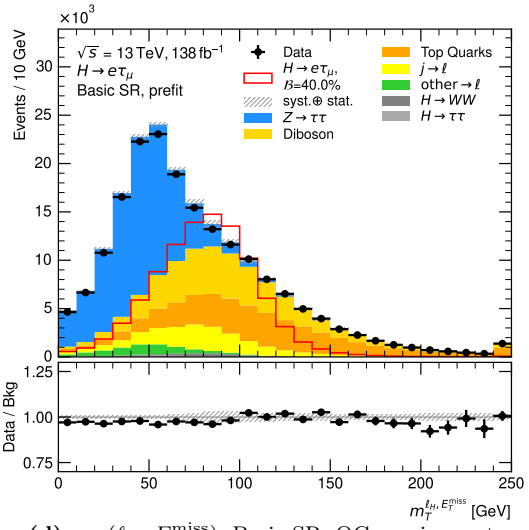

**(a)** $m_{\ell\ell}$, Basic SR, SC requirement



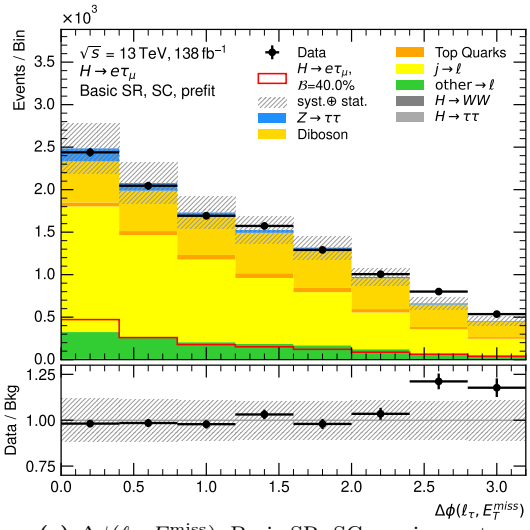

**(b)** $m_{\ell\ell}$, Basic SR, OC requirement



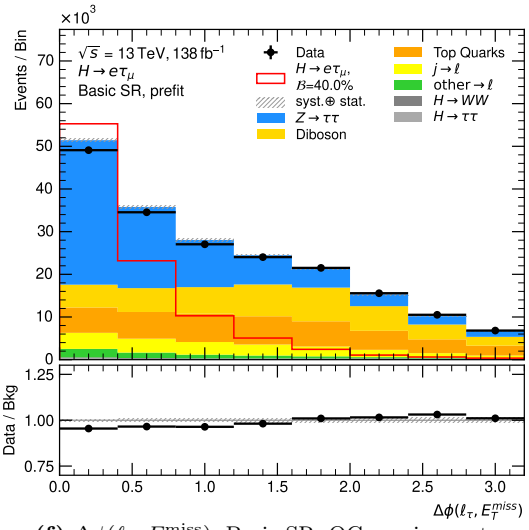

**(c)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, SC requirement




**(d)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, OC requirement

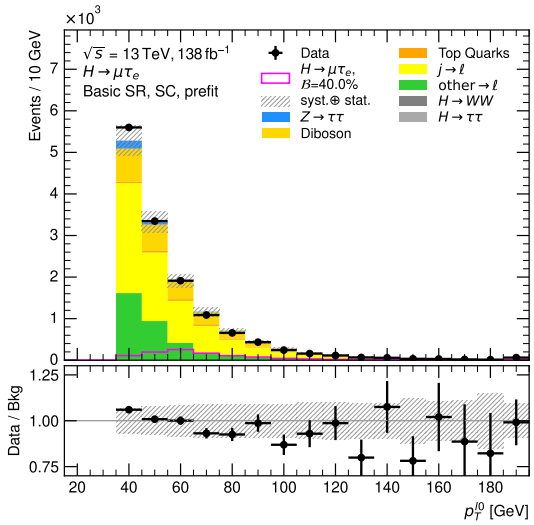



**(e)** $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, SC requirement

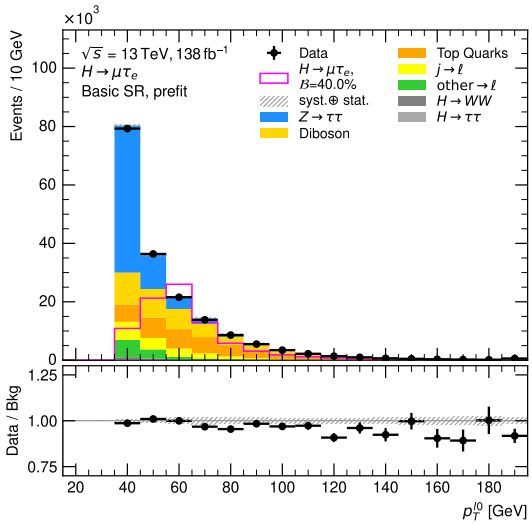



**(f)** $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, OC requirement

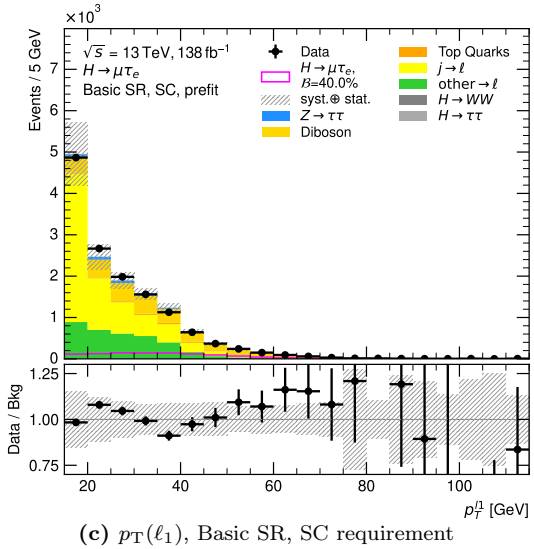
**Figure 7.12:** Distributions of $m_{\ell\ell}$ (a,b), $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$ (c,d), $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ (e,f) in the $e\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining *other* $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate as discussed in the text.



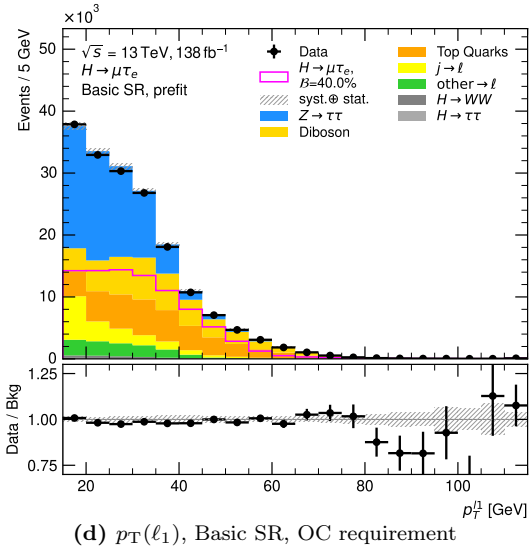

**(a)** $p_{\mathrm{T}}(\ell_0)$, Basic SR, SC requirement



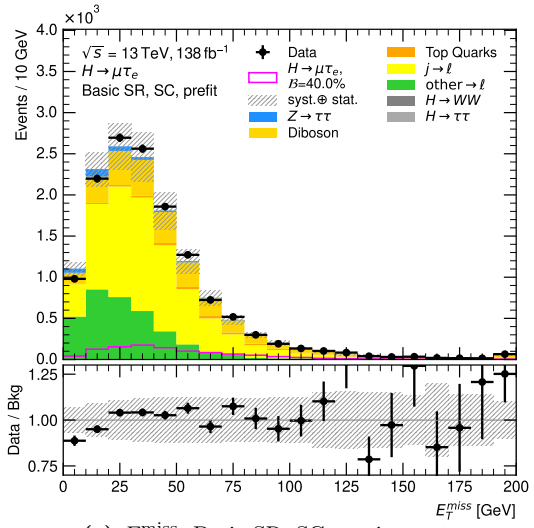

**(b)** $p_{\mathrm{T}}(\ell_0)$, Basic SR, OC requirement



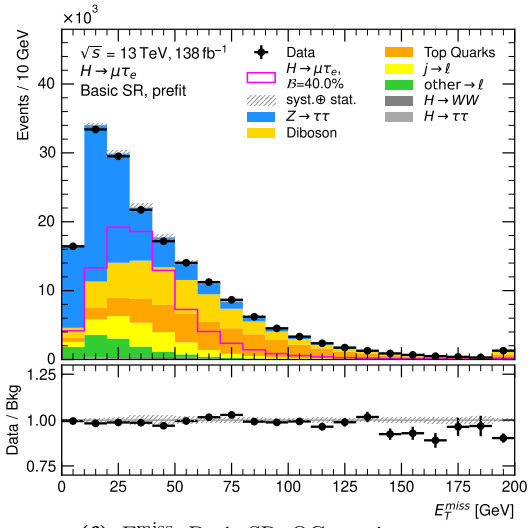

**(c)** $p_{\mathrm{T}}(\ell_1)$, Basic SR, SC requirement




**(d)** $p_{\mathrm{T}}(\ell_1)$, Basic SR, OC requirement

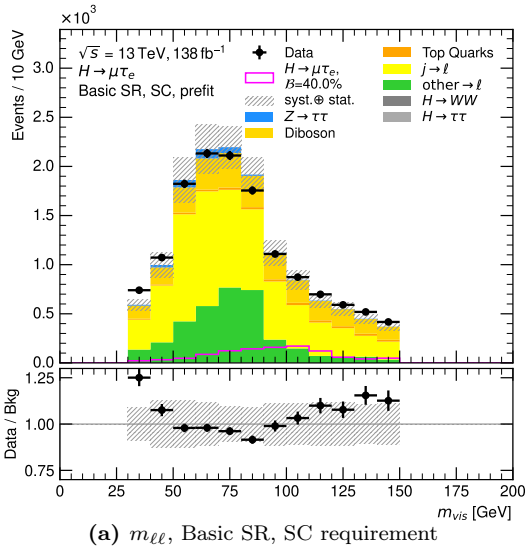



**(e)** $E_{\mathrm{T}}^{\mathrm{miss}}$, Basic SR, SC requirement

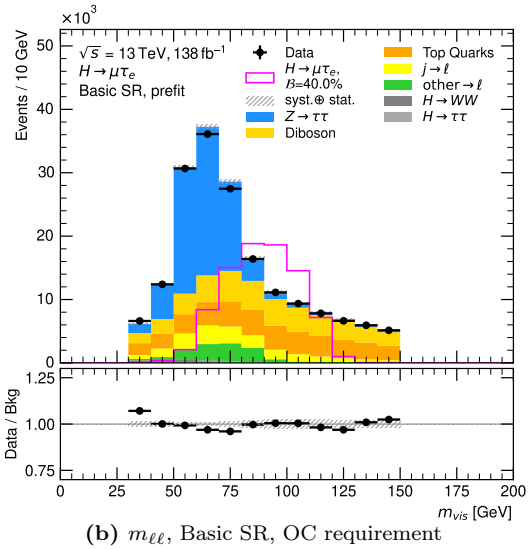



**(f)** $E_{\mathrm{T}}^{\mathrm{miss}}$, Basic SR, OC requirement

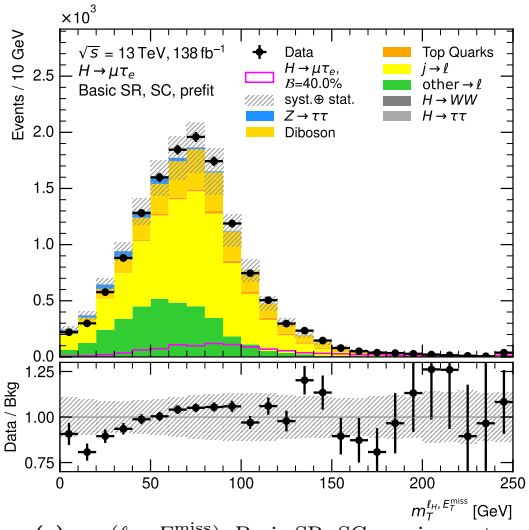
**Figure 7.13:** Distributions of $p_{\mathrm{T}}^{\ell_0}$ (a,b), $p_{\mathrm{T}}^{\ell_1}$ (c,d), $E_{\mathrm{T}}^{\mathrm{miss}}$ (e,f) in the $\mu\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining *other* $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate as discussed in the text.



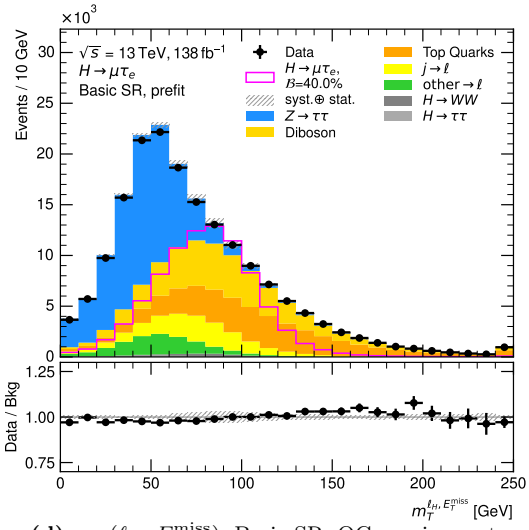

**(a)** $m_{\ell\ell}$, Basic SR, SC requirement



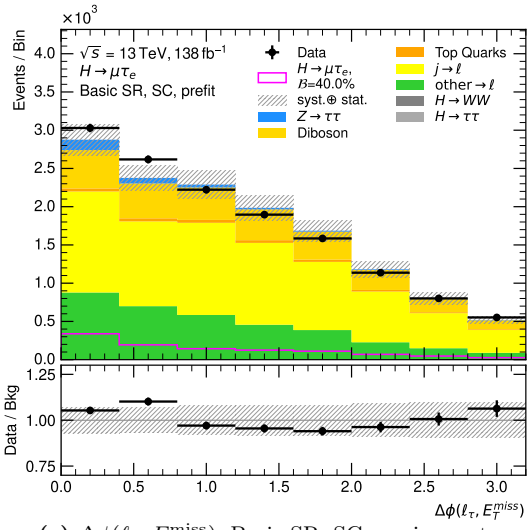

**(b)** $m_{\ell\ell}$, Basic SR, OC requirement



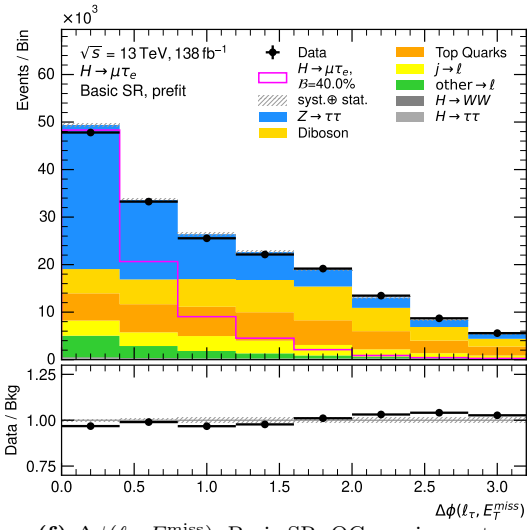

**(c)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, SC requirement




**(d)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, OC requirement




**(e)** $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, SC requirement




**(f)** $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, OC requirement

**Figure 7.14:** Distributions of $m_{\ell\ell}$ (a,b), $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$ (c,d), $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ (e,f) in the $\mu\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining *other* $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate as discussed in the text.

### 7.2.3 Light Lepton Detection Efficiencies

The detection efficiencies introduced in Section 7.2.1 are the product of the trigger efficiency of the event and the offline[10] efficiencies of the leptons in the event (one electron and one muon). For each lepton, the offline efficiency is a product of the reconstruction (reco), identification (id) and isolation (iso) efficiencies. In total, the detection ($d$) efficiency of an event can be written as:

$$\varepsilon_d = \varepsilon_{\text{trigger}} \cdot \varepsilon^e_{\text{reco}} \cdot \varepsilon^e_{\text{id}} \cdot \varepsilon^e_{\text{iso}} \cdot \varepsilon^\mu_{\text{reco}} \cdot \varepsilon^\mu_{\text{id}} \cdot \varepsilon^\mu_{\text{iso}} \,. \tag{7.14}$$

The determination of the individual efficiencies is based on a so-called *tag-and-probe* (T&P) method. In this method, the events of a well known process, here often $Z$-boson production with decay to either electrons or muons, are selected. One object in the event serves as *tag*-object which is required to have fired the trigger and fulfil strict selection criteria. Another object in the event, a lepton in this case, is utilised as *probe*-object that is unbiased with respect to the quantity for which the efficiency should be measured, while the kinematic selection of the event and the *tag*-object ensure a high purity of true leptons for the *probe*-lepton.

All efficiencies except the electron reconstruction, identification and isolation efficiencies are determined and provided centrally by the ATLAS collaboration and are expected to be valid in the *Basic Selection* and any derived selection untilised for the $H$-LFV analysis. The electron efficiencies need to be derive in the specific selection of interest (here the *Basic Selection*) since centrally determined ones are not valid for all selections. Usually, the analyses only make use of the *scale factors* which are the ratio of the efficiencies determined in data and in simulation and which serve as simulation-to-data correction factors. These are expected to be valid for all selections, also in the case of electrons.

The following sections give a short overview of the determination of the individual efficiencies and of the respective values.

#### Muon Trigger Efficiencies

The muon trigger efficiencies are measured with respect to muons that were reconstructed offline making use of the T&P-method. The efficiency is the ratio of number of *probe*-leptons that are matched to a trigger object within a cone $\Delta R = 0.01$ and of the number of all probe leptons.

$Z \to \mu\mu$-events are used to determine the efficiencies of the muon triggers used in this analysis (see Table 7.1) for $p^\mu_{\text{T}} < 100\,\text{GeV}$ while $t\bar{t}$- and $W$ + jets-events are used for $p^\mu_{\text{T}} \gtrsim 100\,\text{GeV}$ [235].

The $Z \to \mu\mu$-events are selected [235] by requiring two muons compatible with originating from a $Z$-boson which results in a pure sample of $Z \to \mu\mu$-events with a negligible background contribution. The *tag*-muon must have $p_{\text{T}} > 28\,\text{GeV}$, fulfil the *Loose* isolation [188] requirement, and must be matched to a trigger object within a cone $\Delta R = 0.01$.

$t\bar{t}$- and $W$ + jets-events are selected [235] by a trigger requiring large $E^{\text{miss}}_{\text{T}}$ and hence must fulfil $E^{\text{miss}}_{\text{T}} > 200\,\text{GeV}$ and must have exactly one muon with $p_{\text{T}} > 27\,\text{GeV}$ without isolation requirements. Different numbers of jets are required for $t\bar{t}$- and $W$ + jets-events keeping the selections orthogonal by requiring at least one $b$-jet or vetoing $b$-jets, respectively.

Either selection, the *probe*-lepton must fulfil the offline identification criteria of interest, e.g. *medium* identification. The resulting efficiencies of the `HLT_mu14` and the `HLT_mu26_ivarmedium OR HLT_mu50` triggers with respect to the offline $p_{\text{T}}$ of the *probe*-muon fulfilling the *medium* identification WP are shown in Figure 7.15. The dependence of the efficiencies on the muon-$p_{\text{T}}$ is small after the turn-on region and hence the efficiencies

[10] Offline refers to the processing of the data after the events were triggered and stored.

are not provided as a function of $p_{\mathrm{T}}$ but instead of $\eta$ and $\phi$ to account for the geometrical structure of the *Muon System*, if a muon-$p_{\mathrm{T}}$ above the turn-on in $p_{\mathrm{T}}$ (within the plateau region) is ensured. The lower efficiencies in the turn-on region occur because of the missing isolation requirement on the offline muon while there is an isolation requirement for the `HLT_mu26_ivarmedium` trigger. The reduced efficiency in the barrel region compared to the endcaps originates from a lower L1 Trigger efficiency. Overall, efficiencies of up to $\sim 70\,\%\,(87\,\%)$ in the barrel (endcaps) are reached for the `HLT_mu26_ivarmedium OR HLT_mu50` triggers and $\sim 77\,\%\,(88\,\%)$ in the barrel (endcaps) for the `HLT_mu14` trigger. The efficiency of the latter is higher due to less stringent L1 Trigger requirements.

Only statistical uncertainties are considered for efficiencies measured in data which are mostly below $1\,\%$ for the events selected in the $H$-LFV analysis.

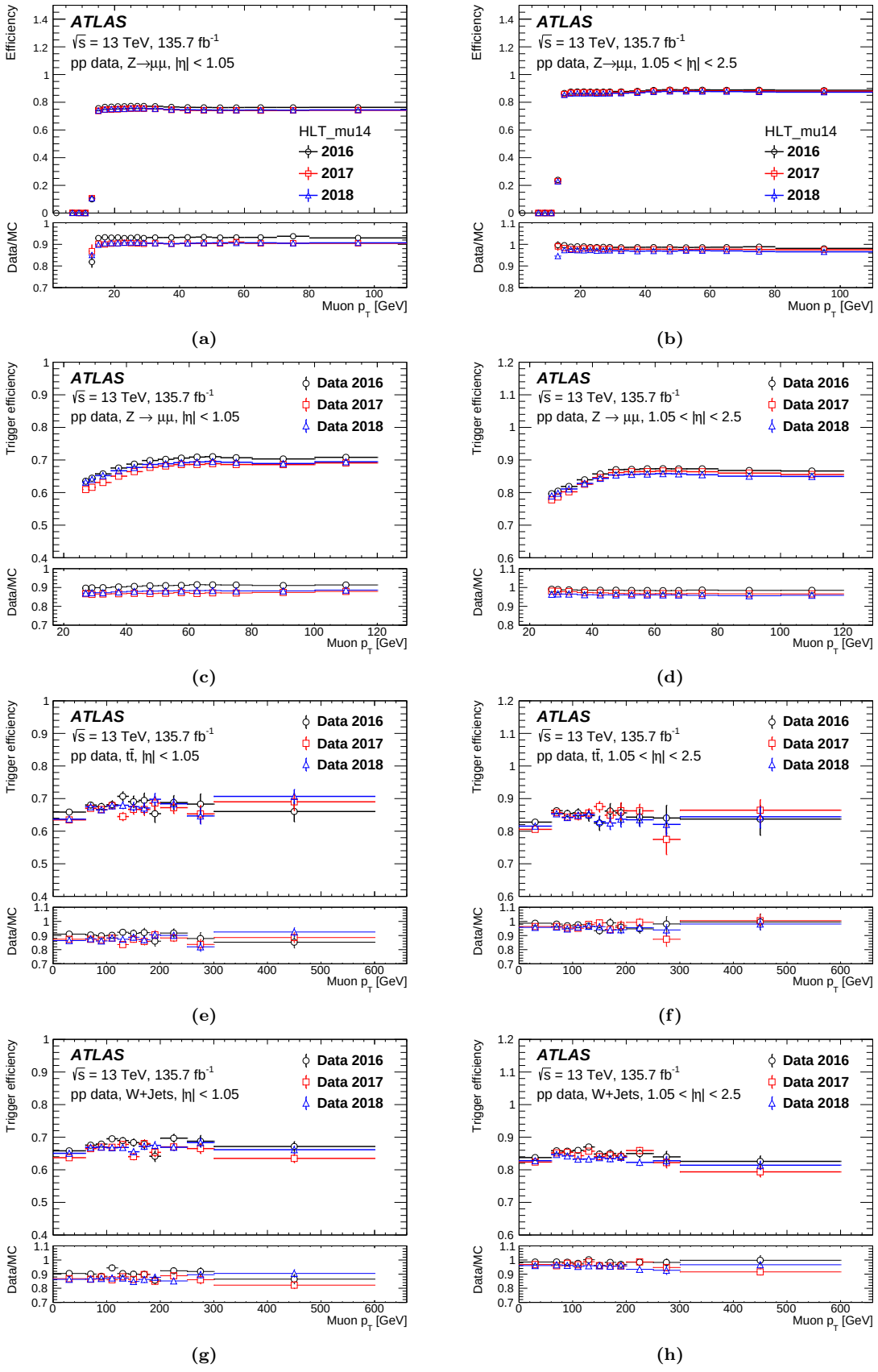



**Figure 7.15:** The efficiencies of the `HLT_mu14` (a),(b) and of the `HLT_mu26_ivarmedium OR HLT_mu50` (c)-(h) triggers with respect to the offline $p_{\mathrm{T}}$ of the *probe*-muon fulfilling the *medium* identification WP, measured in 2016–2018 data with $Z \to \mu\mu$-events (a)-(d), with $t\bar{t}$-events (e),(f) and with $W$+jets-events (g),(h) in the barrel (left) and the endcaps (right) [235]. Only statistical uncertainties are shown. The lower panels show the ratio of data and simulation.

### Muon Offline Efficiencies

In order to derive the muon reconstruction, identification and isolation efficiencies, pure samples of $Z \to \mu\mu$-decays are selected and an analysis using the T&P-method is performed, similar to the measurement of muon trigger efficiencies.

Again, the events must be compatible with originating from $Z \to \mu\mu$-decays. The *tag*-muon must have fired a single muon trigger, have $p_{\mathrm{T}} > 27\,\mathrm{GeV}$, $|\eta| < 2.5$, fulfil the *Tight* isolation requirement [187] and vertex association criteria. For the measurement of the reconstruction and identification efficiencies (which are measured together), the *probe*-muon must have $p_{\mathrm{T}} > 10\,\mathrm{GeV}$, $|\eta| < 2.5$ and fulfil certain isolation criteria [187]. A $\Delta R$-matching between the *probe*-muon and a muon candidate which is reconstructed and identified with the WP of interest is required ($\Delta R < 0.05$). The range of the invariant mass of the two muons is considered from 61 to 121 GeV. A purity of $Z \to \mu\mu$-events of 99.9 % in the probe sample is reached. For the measurement of the isolation efficiency, the *probe*-muon must have $p_{\mathrm{T}} > 3\,\mathrm{GeV}$, $|\eta| < 2.5$ and fulfil the *loose* identification criteria (Section 5.3.2) and vertex association requirements [187]. The separation of the two muons must be larger than $\Delta R = 0.3$ to ensure that the *tag*-muon is not within the isolation cone of the *probe*-muon. The considered range of the invariant mass of the two muons is from 81 to 101 GeV.

As mentioned above, a combined measurement of the reconstruction and identification efficiencies is performed, calculated by the product of independent efficiencies of several terms [187]. The resulting efficiencies were already shown in Figure 5.6(a) of Section 5.3 describing the methods to reconstruct and identify muons. The efficiencies are derived as a function of the muon's $\eta$- and $\phi$-values. The dependence on $p_{\mathrm{T}}$ is found to be flat above the turn-on curve as shown in Figure 5.5(a). Overall, the efficiency for reconstruction and identification with the *medium* WP is $\sim 98\,\%$ for $|\eta| > 0.1$. For $|\eta| < 0.1$ the efficiency is low due to a service access installed in this area. Furthermore, reduced efficiencies can be observed at the ATLAS detector's support structure positions at $\phi = -1.2$ and $\phi = -2.0$ and for $1.0 < |\eta| < 1.3$ due to imperfectly aligned MDT chambers. In addition to statistical uncertainties, several systematic uncertainties are considered: the parameters of the fit to extract the signal and background contributions are varied by their respective fit uncertainties. The fitted non-prompt muon background is varied by its fit uncertainty and the $\Delta R$-based matching of the *probe*-muon is varied. The luminosity uncertainty on background contributions estimated from MC simulations is considered as well as their cross section uncertainties. The single sources are assumed to be independent and added in quadrature. The statistical as well as the total systematic uncertainties are below 0.5 % for most of the events, overall they are at the level of ∼0.1 to 0.2 %.

The definition of the muon isolation was also already discussed in Section 5.3 and its resulting efficiencies are shown in Figure 5.7. For the isolation WP used in this analysis, *FCTightTrackOnly_FixedRad*, only visualisations of the efficiencies measured from simulated diboson events are available. Since they are almost flat in $\eta$, the efficiencies are only provided as a function of $p_{\mathrm{T}}$, and later also as a function of $\Delta R$ of the muon and the closest jet, but not yet when the processing of data for the analysis presented here was finalised. In addition to statistical uncertainties, several systematic uncertainties are considered which are partially the same as for the reconstruction and identification measurement. The statistical uncertainties are below 0.1 % for the majority of the events while the total systematic uncertainties are below 0.5 % and below 0.3 % for most of the events selected in the $H$-LFV analysis.

### Electron Trigger Efficiencies

The electron trigger efficiencies are measured with respect to specific offline identification and isolation WPs. A T&P-analysis is performed with $Z \to ee$-events [236] which are

selected with the requirement that the electron-pair is compatible with originating from a $Z$-boson. Events in an invariant mass window of $\pm 15\,\text{GeV}$ around the $Z$-boson mass are considered. The *tag*-electron is required to fulfil the criteria of the *tight* identification WP (Section 5.2.2), must be matched to an object that fired an unprescaled single-electron trigger, must have $E_T > 27\,\text{GeV}$ and $|\eta| < 2.47$ excluding the calorimeter transition region ($1.37 < |\eta| < 1.52$). The *probe*-electrons must fulfil the same identification and isolation WP criteria which are used in the analysis utilising the efficiencies. If both electrons in the event fulfil the *tag*-requirements, both are used as *probe*-electrons successively. The efficiency is calculated as the ratio of *probe*-electrons which are matched to a trigger object that fired the respective electron trigger over all *probe*-electrons, after subtracting background contributions.

Statistical as well as systematic uncertainties are considered. The systematic uncertainties include variations of the background subtraction method, the invariant mass window, the offline requirements on the *tag*-electron and the regions outside the mass window that are used in the background estimation.

The efficiencies are determined as functions $p_\text{T}$ and $\eta$ and separate for the individual data-taking periods. The resulting efficiencies for the combination of the single-electron triggers used in this analysis (Table 7.1) are exemplarily shown for data taken during 2018 in Figure 7.16. The displayed efficiencies are calculated with respect to offline electrons

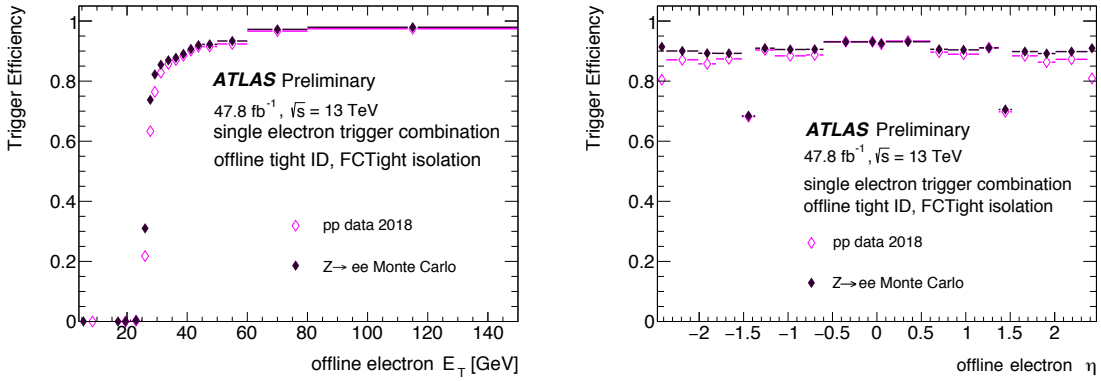
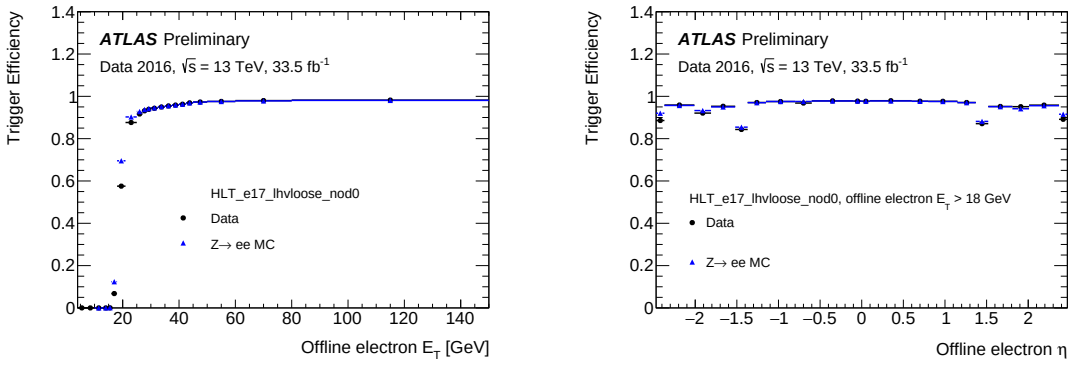



**Figure 7.16:** The efficiencies of the `HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e140_lhloose_nod0` trigger combination with respect to $E_T$ (a) and $\eta$ (b) for data recorded in 2018 [237]. The efficiencies are calculated with respect to offline electrons fulfilling the *tight* identification and the *FCTight* isolation criteria. The pink open markers indicate the efficiencies measured in data, the black closed markers the ones measured with simulated $Z \to ee$-events. The vertical error bars represent statistical and systematic uncertainties added in quadrature.

fulfilling the *tight* identification and the *FCTight* [180] isolation criteria and not with respect to the *medium* identification and the *gradient* isolation criteria used in this analysis (Sections 5.2.2 and 5.2.3). Above $p_\text{T} = 60\,\text{GeV}$ (where the identification WP of the trigger is loosened to *medium* and the isolation requirement is removed), the efficiency is larger than 95 %. An efficiency larger than 90 % is reached from $p_\text{T} = 40\,\text{GeV}$ onwards while it is around 65 % just above the threshold of the lowest trigger. The efficiency with respect to $\eta$ is almost flat but drops in the transition region of the calorimeter ($1.37 < |\eta| < 1.52$) and for largest $|\eta|$. The *FCTight* isolation WP is tighter than the isolation criteria used for the `HLT_e26_lhtight_nod0_ivarloose` trigger. Hence, for the *medium* identification and the *gradient* isolation criteria used in this analysis (which are both looser than *tight* identification and *FCTight* isolation, respectively) lower efficiencies are expected overall.

Figure 7.17 shows the efficiencies of the `HLT_e17_lhloose` trigger measured with data taken during 2016. This trigger is part of the dilepton trigger used in this analysis

(Table 7.1). The efficiencies are determined with respect to electrons fulfilling the *loose*




**(a)** **(b)**

**Figure 7.17:** The efficiencies of the `HLT_e17_lhloose` trigger which is part of the dilepton trigger used in this analysis with respect to $E_T$ (a) and $\eta$ (b) for data taken in 2016 [237]. The efficiencies are calculated with respect to offline electrons fulfilling the *loose* identification criteria. Black dots indicate the efficiencies measured in data, blue triangles the ones measured with simulated $Z \to ee$-events. The vertical error bars represent statistical uncertainties.

identification (Section 5.2.2) as the trigger itself only has *loose* identification requirements and are overall larger compared to Figure 7.16. For the requirements used in this analysis, *medium* identification and *gradient* isolation, the electron trigger efficiencies are larger than the ones displayed, given that the applied offline WPs are stricter.

The combined statistical and systematic uncertainties of the data efficiencies are below 0.3 % and about 0.1 % for most of the events in the $H$-LFV analysis.

### Electron Offline Efficiencies

The electron offline (reconstruction, identification and isolation) efficiencies are determined within the analysis team of the $H$-LFV analysis as it cannot be guaranteed that the centrally provided ones are valid in the selection of this analysis. The strategy foresees to determine the combined efficiencies of reconstruction, identification and isolation from simulated events in the *Basic Selection* and multiply them with the centrally provided *scale factor*s ($SF$s)[11], which are assumed to be valid for all selections, in order to obtain the efficiencies in data:

$$\varepsilon^{\text{data}}_{\text{reco,id,iso}} = \varepsilon^{\text{simulation}}_{\text{reco,id,iso}} \cdot SF_{\text{reco}} \cdot SF_{\text{id}} \cdot SF_{\text{iso}} \,. \tag{7.15}$$

A description of these $SF$s was given in Section 5.2.

The determination of $\varepsilon^{\text{simulation}}_{\text{reco,id,iso}}$ is also based on the T&P-method. A similar selection as the *Basic Selection* (Table 7.2) is applied. Differences are that the $p_{\text{T}}^{\ell_0}$-requirement is lowered to 15 GeV[12], no trigger matching is performed and the *tag*-leptons are only required to be *baseline* leptons (Chapter 5). The $p_{\text{T}}$ of the *tag*-lepton must be at least 20 GeV.

In order to enhance statistics, not only events with an electron and a muon in the final state are considered but also events with two electrons. In case of the latter, both electrons can serve as *tag*- and as *probe*-electron successively while in the former case the muon always serves as *tag*- and the electron as *probe*-lepton.

Simulated events of the major processes contributing two real leptons in the *Basic Selection* are taken into account, weighted according to their respective cross sections (and luminosity). These processes are $Z \to \tau\tau$, Diboson and top-quark production. Although

[11] $SF = \varepsilon^{\text{data}}/\varepsilon^{\text{simulation}}$

[12] After checking that the efficiencies with and without this cut are compatible.

considering events with two electrons, $Z \to ee$-events are not included as they do not contribute two real leptons in the *Basic Selection*.

The leptons are required to be truth-matched with the help of the *IFFTruthClassifier* [230]: the muons must be *prompt muons* (class 4) while the electrons must be *prompt electrons* (classes 2, 3).

The efficiencies are obtained as ratio of the *probe*-electrons that fulfil the reconstruction, identification and isolation requirements of the $H$-LFV analysis (Section 5.2) in the numerator and of all *probe*-electrons in the denominator. The *probe*-electrons in the denominator consist of all clusters that fulfil the truth-matching and have $p_{\mathrm{T}} > 15\,\mathrm{GeV}, |\eta| < 2.47$ while the calorimeter transition region ($1.37 < |\eta| < 1.52$) is vetoed. In general, only electrons leaving a cluster in the ECAL are considered which is a minor bias as the cluster efficiency is very high for the energies relevant in this analysis (see Figure 5.2). The parametrisation of the efficiencies is consistent with the one used for the provided $SF$s. Efficiencies determined for $-\eta$ and $+\eta$ are consistent and hence are merged, such that they only depend on $|\eta|$. To enhance statistics further, the different data-taking years are combined, given that no significant differences in the efficiencies between the datataking years were observed. The efficiencies, however, depend on whether the *probe*-electron is leading or sub-leading in $p_{\mathrm{T}}$. The combined electron reconstruction, identification and isolation efficiencies are therefore binned in $p_{\mathrm{T}}$, $|\eta|$ and leading/sub-leading electron.

In order to derive systematic uncertainties on the offline electron efficiencies, two variations are performed and compared to the nominal case:

- whether the *tag*-lepton is an electron or a muon
- re-introducing $p_{\mathrm{T}}^{\ell_0} > 35\,\mathrm{GeV}$.

For both variations, the absolute differences to the nominal case are utilised as up- and as down-uncertainties. When varying the flavour of the *tag*-lepton, the variation (electron- or muon-*tag*) yielding the larger difference to the nominal case, which uses both, electron- and muon-*tag*, is utilised. Both contributions are added in quadrature and the square-root is taken. The latter variation only impacts the efficiencies for $p_{\mathrm{T}} < 35\,\mathrm{GeV}$. The resulting efficiencies with the statistical and systematic uncertainties added in quadrature are displayed in Figure 7.18. The statistical uncertainties over the data events selected by the *Basic Selection* are on average 1 % and vary from 0.1 to 5 % for the single $p_{\mathrm{T}}$-$|\eta|$-bins in which they are measured. The systematic uncertainties are on average 1.4 % and vary from 0.1 to 6 % for the single $p_{\mathrm{T}}$-$|\eta|$-bins.

### Validation of the Restoration of the $e/\mu$-Symmetry by Applying the Detection Efficiencies

The application of the light lepton detection efficiencies in order to restore the symmetry assumption between electrons and muons is validated by applying the efficiencies to simulated events. The efficiency-ratio $\mathcal{R}^{e\tau}$, defined in Eq. 7.8, is applied to the $e/\mu$-symmetric background processes of the $e\tau$-dataset, estimated from simulations, and the sum of these processes is compared to the sum of the same processes of the $\mu\tau$-dataset where the efficiency-ratio is not applied. The considered processes are $Z \to \tau\tau$, top-quark and diboson production, $H \to \tau\tau$ and $H \to WW$. The efficiency-ratio is calculated using the trigger and offline efficiencies of electrons and muons discussed above and applied in an event-by-event basis. The detailed event-by-event formula is given in the next section in Eq. 7.17. The distributions of several observables are shown in Figures 7.19 and 7.20 where in addition to the $\mu\tau$- and the efficiency-corrected $e\tau$-dataset also the $e\tau$-dataset without $\mathcal{R}^{e\tau}$ applied is displayed to visualise the impact of the application of the efficiency-ratio $\mathcal{R}^{e\tau}$. Figures 7.19 and 7.20 show that the description of the $\mu\tau$-dataset by the efficiency-corrected $e\tau$-dataset is

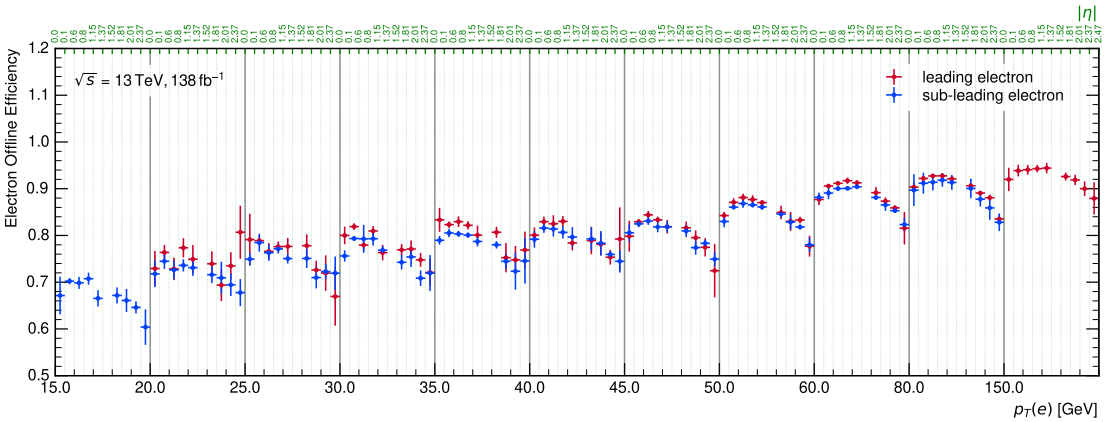
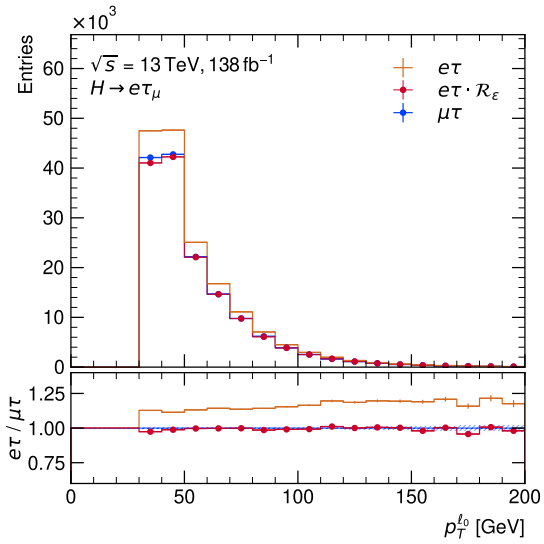


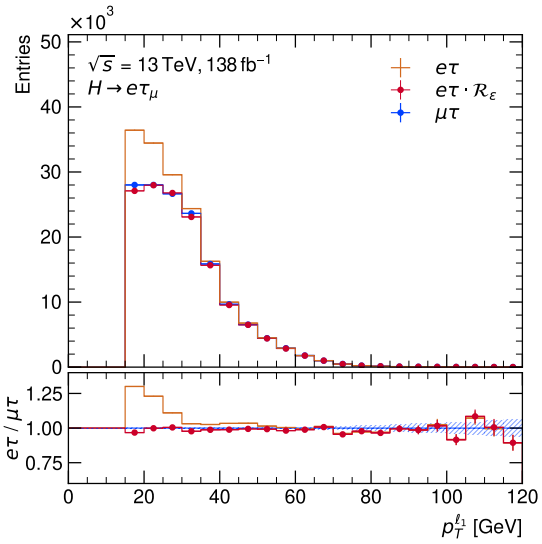

**Figure 7.18:** The electron offline efficiencies (reconstruction, identification, isolation) as determined for this analysis with respect to $p_{\mathrm{T}}$ and $|\eta|$ of the electron, separate for when the electron is sub-leading (blue) or leading (red). For each $p_{\mathrm{T}}$-bin on the lower $x$-axis, the efficiency values are displayed for each $|\eta|$-bin. The bin edges in $|\eta|$ are: [0.0, 0.1, 0.6, 0.8, 1.15, 1.37, 1.52, 1.81, 2.01, 2.37, 2.47], also indicated on the upper $x$-axis. The efficiencies for the leading electron are only relevant from $p_{\mathrm{T}} > 35\,\mathrm{GeV}$. The last bin in $p_{\mathrm{T}}$ is open towards larger values, i.e. $p_{\mathrm{T}} > 150\,\mathrm{GeV}$ when the electron is leading and $p_{\mathrm{T}} > 80\,\mathrm{GeV}$ when it is sub-leading. The vertical error bars represent the statistical and systematic uncertainties added in quadrature. Credits to Mattias Birman [234] for the inputs to this figure and the way of presentation.

convincing. In the regions of the distributions with sufficient statistics the deviation between the $\mu\tau$- and the efficiency-corrected $e\tau$-dataset is at most 4 %. Systematic uncertainties of the simulations and of the efficiency-corrections are not displayed. The restoration of the symmetry assumption works as intended, which indicates validity of the method.

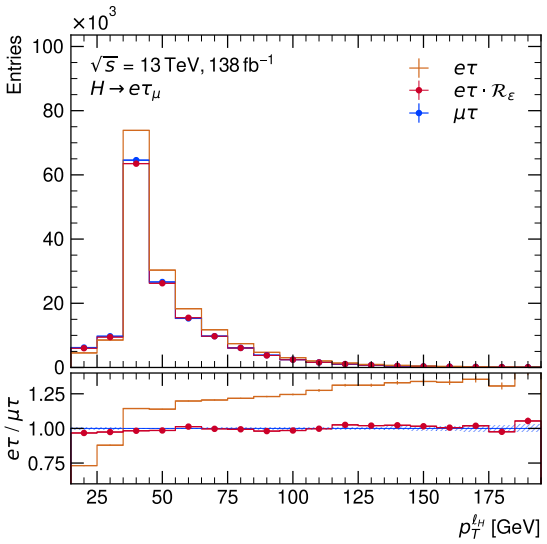
The same comparison of datasets can be performed the other way around, where the inverse of $\mathcal{R}^{e\tau}$, which is $\mathcal{R}^{\mu\tau}$, is applied to the the $\mu\tau$-dataset and compared to $e\tau$-dataset without $\mathcal{R}^{e\tau}$ applied. Since this does not provide additional insides, it is not shown here.



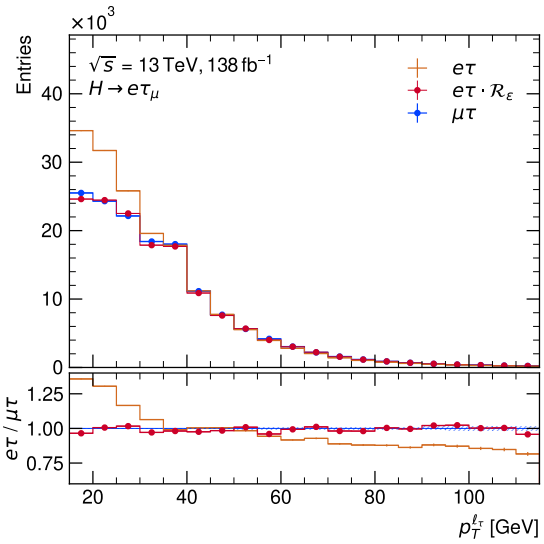

**(a)** $p_{\mathrm{T}}(\ell_0)$, Basic SR



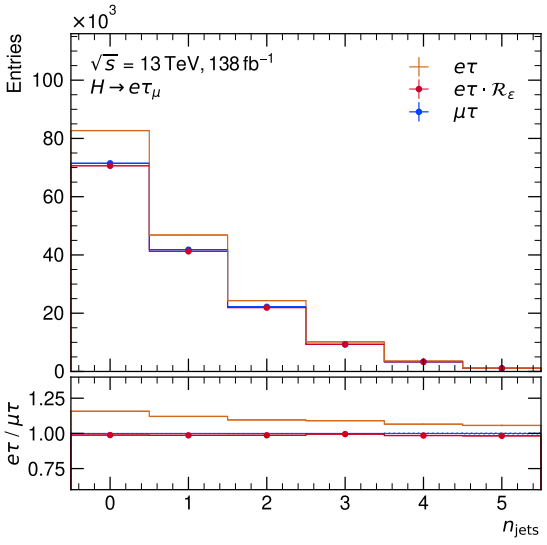

**(b)** $p_{\mathrm{T}}(\ell_1)$, Basic SR



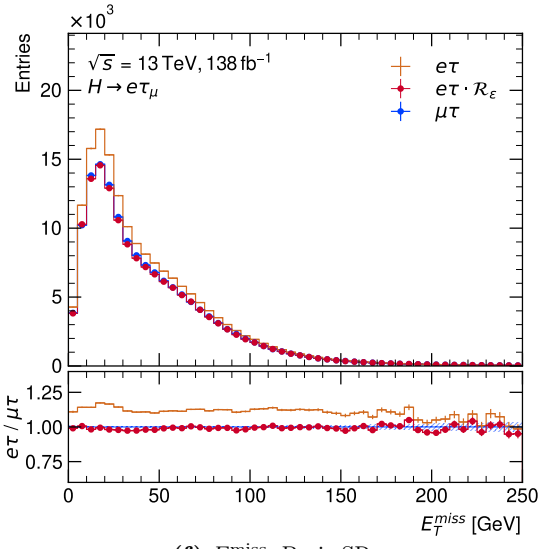

**(c)** $p_{\mathrm{T}}(\ell_H)$, Basic SR




**(d)** $p_{\mathrm{T}}(\ell_\tau)$, Basic SR

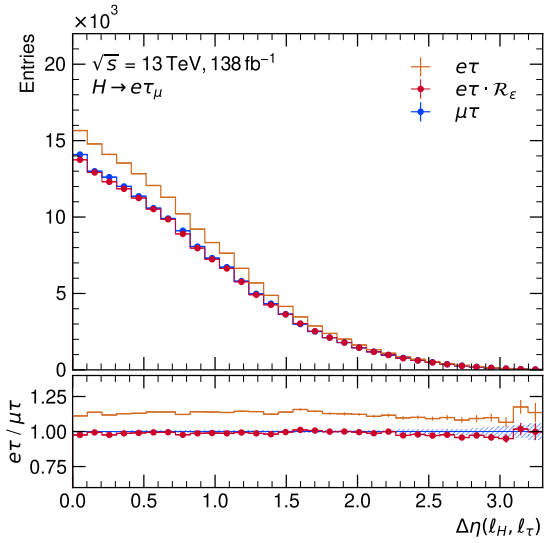



**(e)** $n_{\mathrm{jets}}$ with $p_{\mathrm{T}}(\mathrm{jet}) \geq 30\,\mathrm{GeV}$, Basic SR

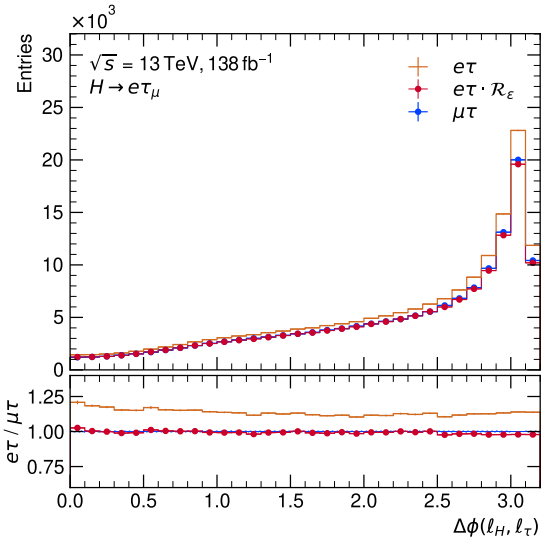



**(f)** $E_{\mathrm{T}}^{\mathrm{miss}}$, Basic SR

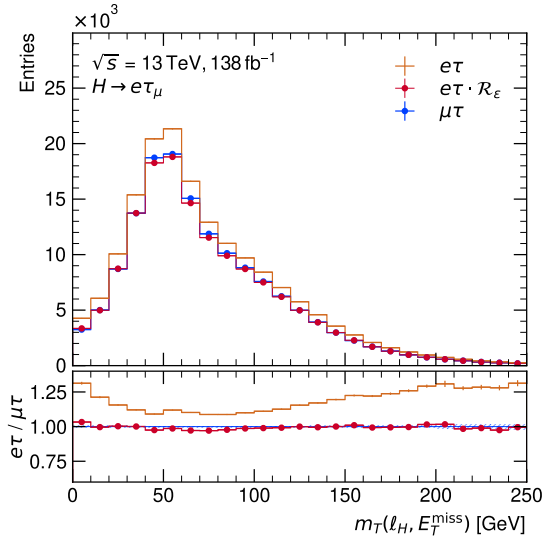
**Figure 7.19:** Distributions comparing the $e\tau$- and the efficiency-corrected ($\cdot\mathcal{R}_\varepsilon$) $e\tau$-dataset to the $\mu\tau$-dataset using all $e/\mu$-symmetric background processes estimated from simulations: $Z \to \tau\tau$, top-quark and diboson production, $H \to \tau\tau$, $H \to WW$. The efficiency-ratio $\mathcal{R}_\varepsilon$ is denoted as $\mathcal{R}^{e\tau}$ in the text. The indicated uncertainties only represent statistical uncertainties on the simulated events.



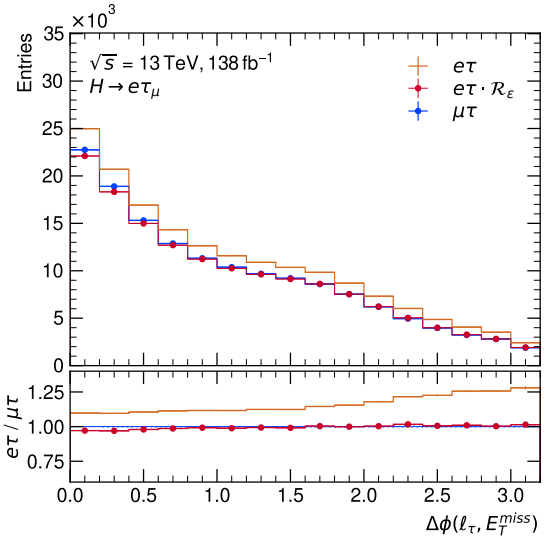

**(a)** $\Delta\eta(\ell_H, \ell_\tau)$, Basic SR



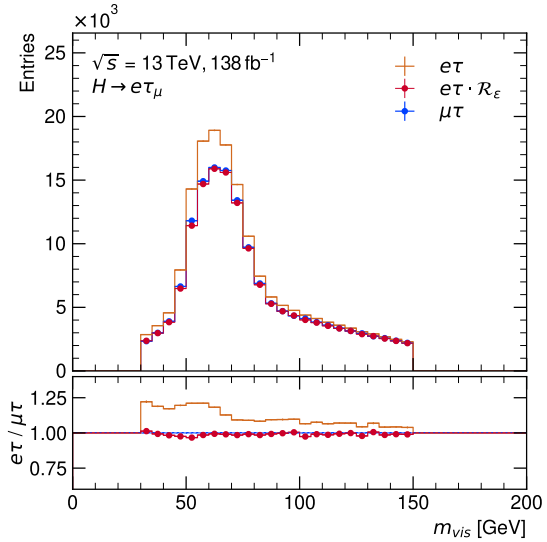

**(b)** $\Delta\phi(\ell_H, \ell_\tau)$, Basic SR



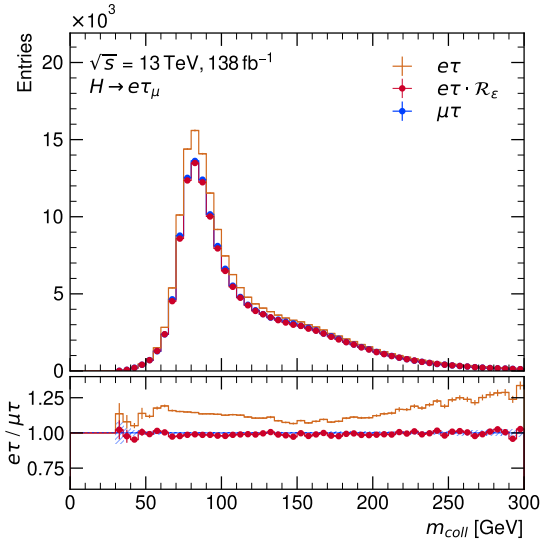

**(c)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR




**(d)** $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR




**(e)** $m_{\ell\ell}$, Basic SR




**(f)** $m_{\mathrm{coll}}$, Basic SR

**Figure 7.20:** Distributions comparing the $e\tau$- and the efficiency-corrected ($\cdot\mathcal{R}_\varepsilon$) $e\tau$-dataset to the $\mu\tau$-dataset using all $e/\mu$-symmetric background processes estimated from simulations: $Z \to \tau\tau$, top-quark and diboson production, $H \to \tau\tau$, $H \to WW$. The efficiency-ratio $\mathcal{R}_\varepsilon$ is denoted as $\mathcal{R}^{e\tau}$ in the text. The indicated uncertainties only represent statistical uncertainties on the simulated events.

### 7.2.4 Combining the Ingredients for the Background Estimation with the *Symmetry Method*

Combining the individual ingredients for the background estimation with the *Symmetry Method*, Eqs. 7.8 and 7.9 need to be expanded in an event-by-event notation. The number of detected events of the $e\tau$-dataset can be expressed as:

$$
\begin{aligned}
n^{e\tau}_{\mathrm{SR,det}} = \underbrace{\sum_{i=1}^{n^{\mu\tau}_{\mathrm{SR,det}}} \mathcal{R}^{\mu\tau}_i - \sum_{i=1}^{n^{\mu\tau}_{\mathrm{SR,\cancel{IdIso}}}} \mathcal{R}^{\mu\tau}_i \cdot w_i^{\mathrm{FF,CF}} - \sum_{i=1}^{n^{\mu\tau}_{\mathrm{SR},other\to\ell}} \mathcal{R}^{\mu\tau}_i \cdot w_i^{\mathrm{MC}}}_{e/\mu\text{-symmetric background estimate } n^{\sim\mu\tau}} \\
+ \underbrace{\sum_{i=1}^{n^{e\tau}_{\mathrm{SR,\cancel{IdIso}}}} w_i^{\mathrm{FF,CF}}}_{j\to\ell\text{-fakes } f_{j\to\ell}} + \underbrace{\sum_{i=1}^{n^{e\tau}_{\mathrm{SR},other\to\ell}} w_i^{\mathrm{MC}}}_{other\to\ell\text{-fakes } f_{other\to\ell}}
\end{aligned}
\tag{7.16}
$$

The same expression holds for the number of detected events of the $\mu\tau$-dataset when interchanging $e\tau$ and $\mu\tau$. The number of events $n_{\mathrm{SR,\cancel{IdIso}}}$ comprise the events in the $\cancel{\texttt{IdIso}}$-SR where either the electron, the muon or both fulfil the $\cancel{\texttt{IdIso}}$-criteria. The event weight $w_i^{\mathrm{FF,CF}}$ is the product of the respective *Fake Factor* (FF)(s) and the *Correction Factor* (CF). In principle, the $w_i^{\mathrm{FF,CF}}$ would need to be expanded into the different single- and double-fake terms discussed in Eq. 7.12; for simplicity and readability this is omitted in Eq. 7.16. For the *other* $\to \ell$-fakes, $w_i^{\mathrm{MC}}$ contains the standard weights related to MC-simulations and $n_{\mathrm{SR},other\to\ell}$ is the number of events of the contributing processes. The efficiency-ratio $\mathcal{R}^{\mu\tau}$ is obtained from Eqs. 7.8 and 7.14 and corresponds to:

$$
\begin{aligned}
\mathcal{R}^{\mu\tau}_i\left(k^\mu_i, k^e_i\right) &= \frac{\varepsilon^{e\tau}_i\left(k^\mu_i, k^e_i\right)}{\varepsilon^{\mu\tau}_i\left(k^\mu_i, k^e_i\right)} \\
&= \frac{\varepsilon^{e(k^\mu_i),\mu(k^e_i)}_{\mathrm{trigger}} \cdot \varepsilon^{e(k^\mu_i)}_{\mathrm{reco,id,iso}} \cdot SF^{e(k^\mu_i)}_{\mathrm{reco}} \cdot SF^{e(k^\mu_i)}_{\mathrm{id}} \cdot SF^{e(k^\mu_i)}_{\mathrm{iso}} \cdot \varepsilon^{\mu(k^e_i)}_{\mathrm{reco,id}} \cdot \varepsilon^{\mu(k^e_i)}_{\mathrm{iso}}}{\varepsilon^{e(k^e_i),\mu(k^\mu_i)}_{\mathrm{trigger}} \cdot \varepsilon^{e(k^e_i)}_{\mathrm{reco,id,iso}} \cdot SF^{e(k^e_i)}_{\mathrm{reco}} \cdot SF^{e(k^e_i)}_{\mathrm{id}} \cdot SF^{e(k^e_i)}_{\mathrm{iso}} \cdot \varepsilon^{\mu(k^\mu_i)}_{\mathrm{reco,id}} \cdot \varepsilon^{\mu(k^\mu_i)}_{\mathrm{iso}}}.
\end{aligned}
\tag{7.17}
$$

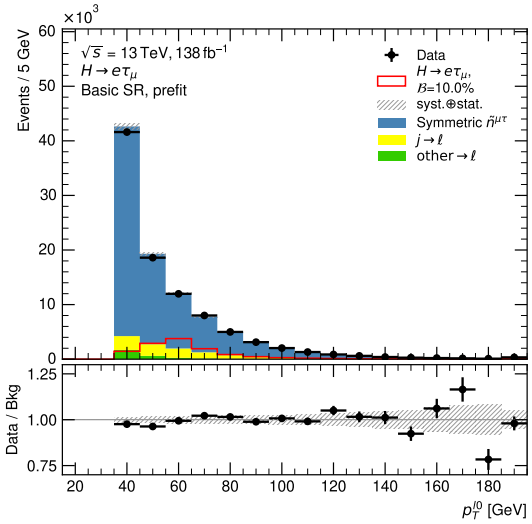
The variables $k^e_i$ and $k^\mu_i$ represent the kinematic properties of the electron and muon in the $i$-th $\mu\tau$-event. The kinematic properties considered here are the variables in which the efficiencies are parametrised in. The efficiencies in the denominator correspond to the real event while the ones in the numerator correspond to the virtual *switch* event (compare Section 7.2.1). The term for the efficiency-ratio $\mathcal{R}^{e\tau}$ is obtained equivalently by interchanging $e\tau$ and $\mu\tau$.

Distributions of the same kinematic observables as in Figures 7.11–7.14 but with the backgrounds estimated according to Eq. 7.16 are shown in Figures 7.21 and 7.22. The colours of the single contributions in the distributions correspond to the coloured boxes in Eq. 7.16. The displayed uncertainty band includes statistical uncertainties and full systematic uncertainties on the background predictions added in quadrature. The systematic uncertainties are summarised in Chapter 9.

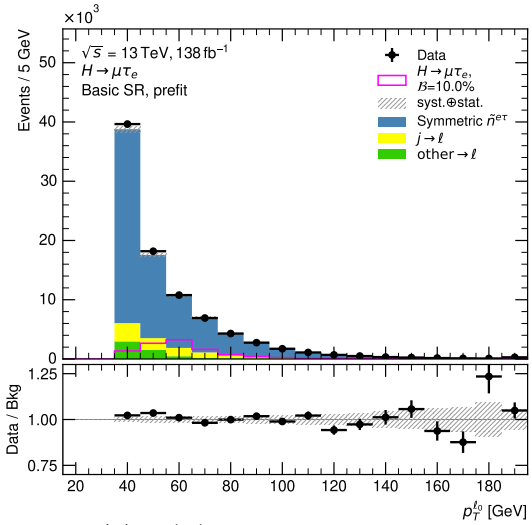
Small deviations in the ratio of the data over the background prediction from unity are visible for example in the $m_{\ell\ell}$-distribution (Figures 7.22(a) and 7.22(b)), similar to Figures 7.12(b) and 7.14(b), and in the distribution of $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ (Figures 7.22(e) and 7.22(f)), however, they do not exceed 8 %. Overall, the modelling is reasonable within uncertainties. Figure 7.23 shows the comparison of the modelling with the *Symmetry Method*, (a) and (b), compared to a direct estimate from simulated events, (c) and (d), in the distribution of the number of jets in the event, $n_{\mathrm{jets}}$. With simulated events, the

$n_{\text{jets}}$-distribution is typically not well modelled, while the modelling is very good with the *Symmetry Method*. This demonstrates the advantage of the background estimation mainly based on data.

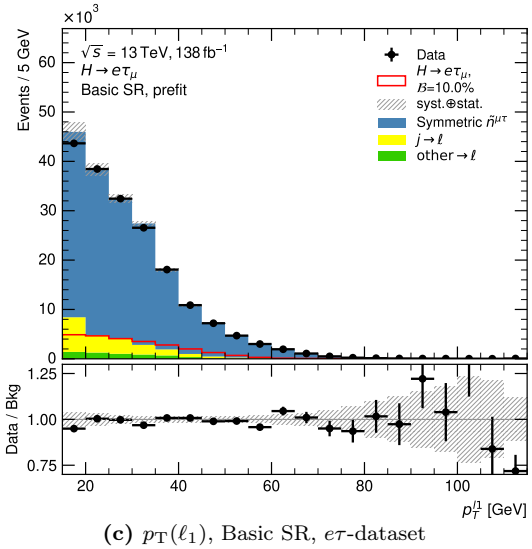
The correlation of the $e/\mu$-symmetric background $\widetilde{n}$ in one dataset with the data of the other dataset can be observed in all distributions. Each fluctuation in one dataset is visible in the other dataset with the opposite sign. This is particularly visible in the first three bins of the distributions of the invariant mass of the two leptons $m_{\ell\ell}$ in Figures 7.22(a) and 7.22(b): When the data of the $e\tau$-dataset shows a down-fluctuation with respect to the background prediction (a down-fluctuation in the data-over-background ratio), this results in an under-prediction of the data in the $\mu\tau$-dataset and hence an up-fluctuation in the data-over-background ratio in the $\mu\tau$-dataset (conmpare Eq. 7.16). It is, however, not apparent whether the $e\tau$- or the $\mu\tau$-dataset or both are responsible for the fluctuation.



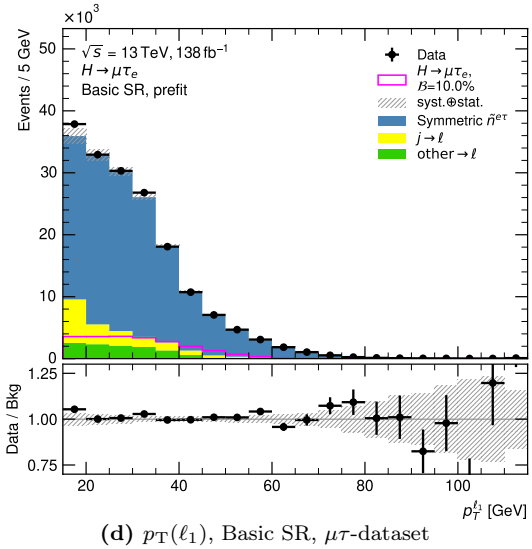

**(a)** $p_{\mathrm{T}}(\ell_0)$, Basic SR, $e\tau$-dataset



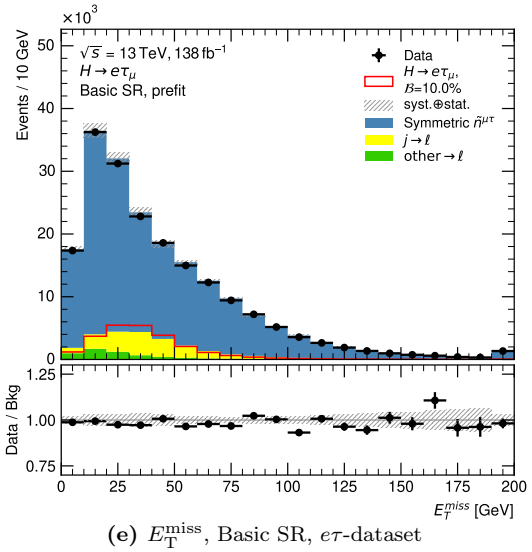

**(b)** $p_{\mathrm{T}}(\ell_0)$, Basic SR, $\mu\tau$-dataset



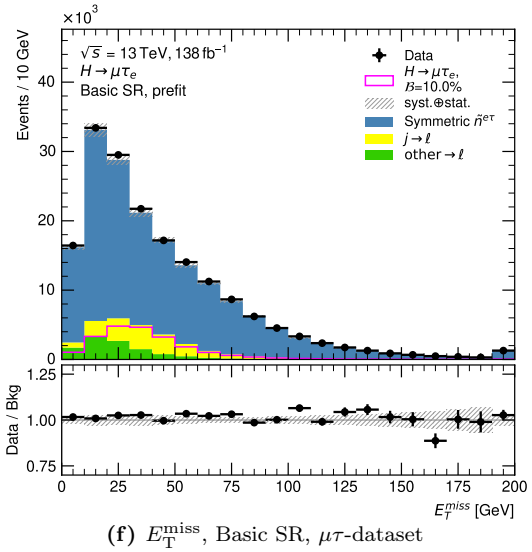

**(c)** $p_{\mathrm{T}}(\ell_1)$, Basic SR, $e\tau$-dataset




**(d)** $p_{\mathrm{T}}(\ell_1)$, Basic SR, $\mu\tau$-dataset

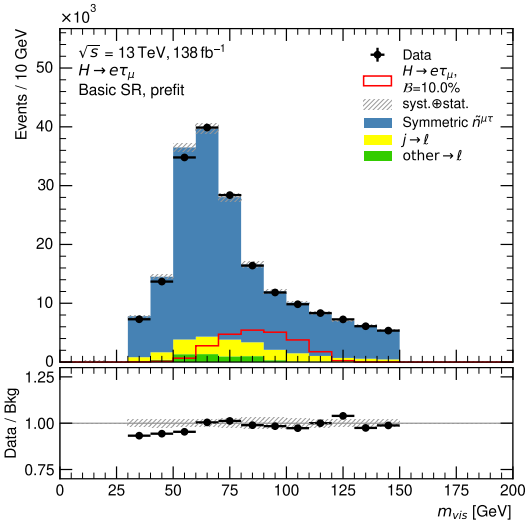



**(e)** $E_{\mathrm{T}}^{\mathrm{miss}}$, Basic SR, $e\tau$-dataset

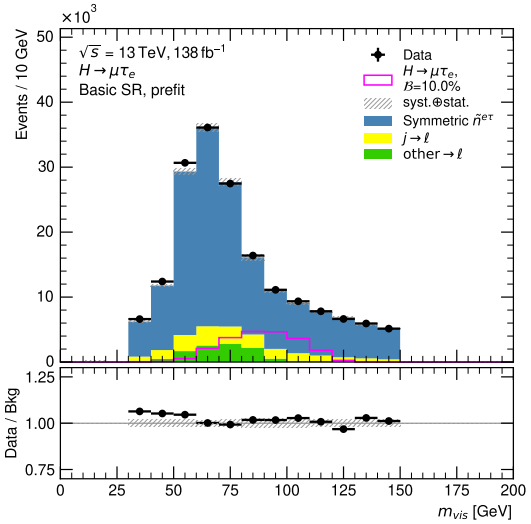



**(f)** $E_{\mathrm{T}}^{\mathrm{miss}}$, Basic SR, $\mu\tau$-dataset

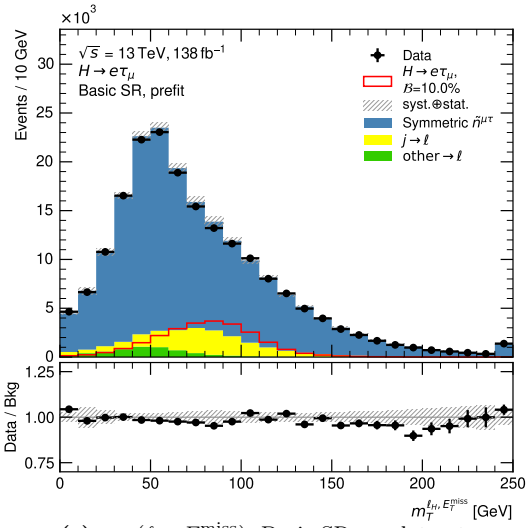
**Figure 7.21:** Kinematic distributions with the $e/\mu$-symmetric background contribution estimated with the *Symmetry Method* for the $e\tau$-dataset (left) and the $\mu\tau$-dataset (right) after the *Basic Selection* SR. The $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining *other* $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical and full systematic uncertainties on the background predictions added in quadrature. The signal prediction assuming a branching ratio of 10 % is overlaid.



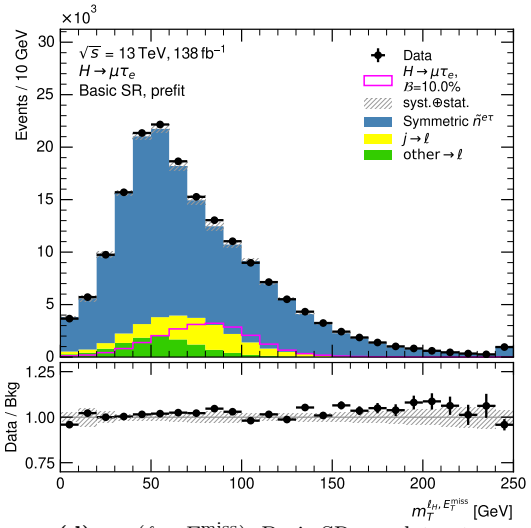

**(a)** $m_{\ell\ell}$, Basic SR, $e\tau$-dataset



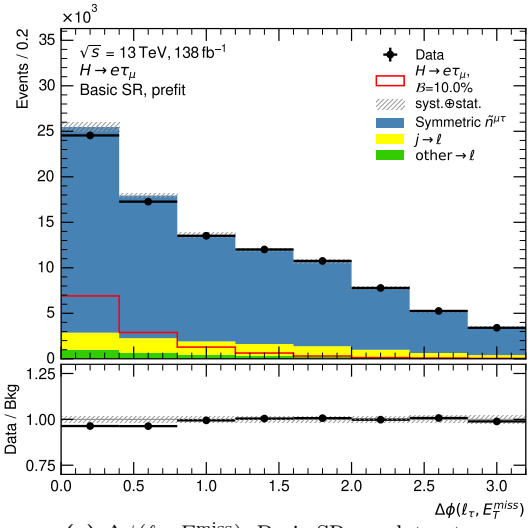

**(b)** $m_{\ell\ell}$, Basic SR, $\mu\tau$-dataset



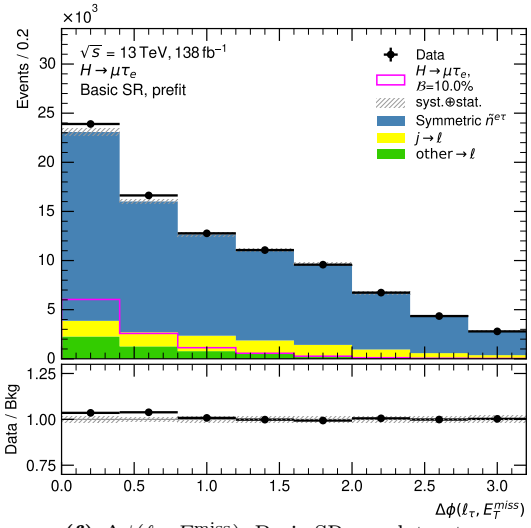

**(c)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, $e\tau$-dataset




**(d)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, $\mu\tau$-dataset

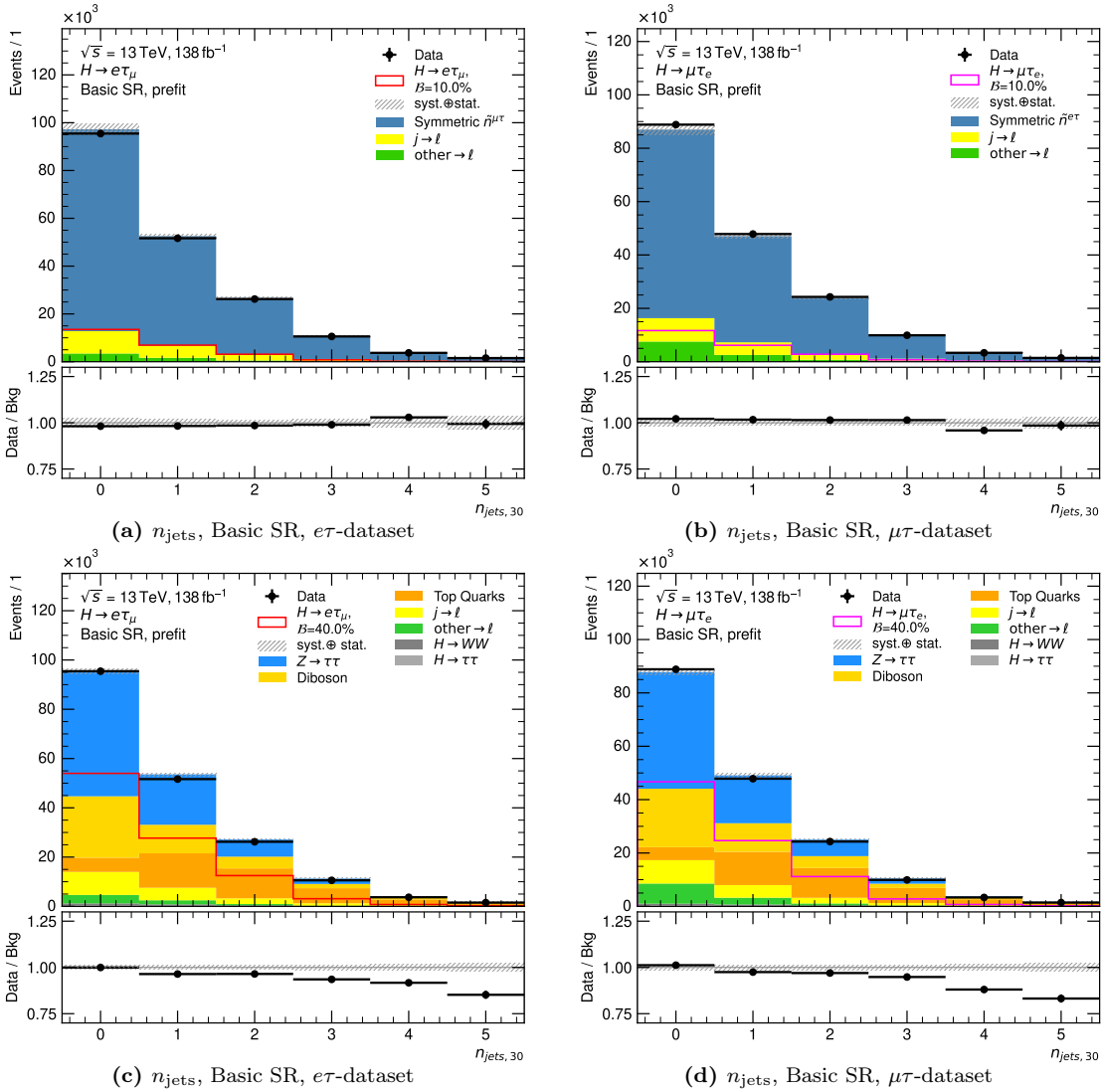



**(e)** $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, $e\tau$-dataset




**(f)** $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, $\mu\tau$-dataset

**Figure 7.22:** Kinematic distributions with the $e/\mu$-symmetric background contribution estimated with the *Symmetry Method* for the $e\tau$-dataset (left) and the $\mu\tau$-dataset (right) after the *Basic Selection* SR. The $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining *other* $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical and full systematic uncertainties on the background predictions added in quadrature. The signal prediction assuming a branching ratio of 10 % is overlaid.




**(a)** $n_{\text{jets}}$, Basic SR, $e\tau$-dataset

**(b)** $n_{\text{jets}}$, Basic SR, $\mu\tau$-dataset

**(c)** $n_{\text{jets}}$, Basic SR, $e\tau$-dataset

**(d)** $n_{\text{jets}}$, Basic SR, $\mu\tau$-dataset

**Figure 7.23:** Distributions of the number of jets for events with $p_{\text{T}} > 30\,\text{GeV}$ in the $e\tau$- (left) and the $\mu\tau$-datasets and with the symmetric background contributions estimated with the *Symmetry Method* (top) and from simulations (bottom). The $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining *other* $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background predictions and systematic uncertainties added in quadrature. The systematic uncertainties in the (a) and (b) comprise all systematic uncertainties while in (c) and (d) they only consist of uncertainties on the $j \to \ell$-fakes. The signal prediction assuming a branching ratio of 10 % is overlaid.

# Chapter 8

# Separation of Signal and Background

This chapter describes the methods to further separate the signal from the background contributions. First, the region defined by the *Basic Selection* is split into two *Signal Region*s (SRs) which are introduced in Section 8.1. Afterwards, *Neural Network*s (NNs) are trained for each region such that the resulting output node distributions can act as final discriminants in the statistical analysis. The selected input variables, the optimisation of the hyperaparameters and the results of the NN-training are discussed in Section 8.2.

## 8.1 Different Signal Regions

The *Basic Selection Signal Region* is split into two *Signal Region*s: the *Vector-Boson-Fusion Signal Region* (VBF SR) and the *Non-Vector-Boson-Fusion Signal Region* (nonVBF SR) in order to exploit the special topology of the VBF-signal. The VBF-signal is characterised by two additional jets in the final state which exhibit a large separation in the pseudorapidity ($\Delta\eta_{jj}$) and a large invariant mass ($m_{jj}$). Hence, the VBF SR is defined by requiring at least two jets in the final state which are further required to satisfy the following criteria regarding their $p_{\mathrm{T}}$, $\Delta\eta_{jj}$ and $m_{jj}$:

- $n_{\mathrm{jets}} \geq 2$
- $p_{\mathrm{T}}^{j_0} > 40\,\mathrm{GeV}$
- $p_{\mathrm{T}}^{j_1} > 30\,\mathrm{GeV}$
- $\Delta\eta_{jj} > 3$
- $m_{jj} > 400\,\mathrm{GeV}$

Normalised distributions of the latter four observables after requiring at least two jets (with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$) in the final state are shown in Figure 8.1. The above requirements, indicated as dashed vertical lines, clearly select the VBF production signal compared to the ggF production signal or the background predictions.

The above mentioned requirements split the events into the VBF SR (requirements fulfilled) and the nonVBF SR (requirements not fulfilled). The yields of the individual processes in the respective regions are listed in Tables 8.1 and 8.2. In the VBF SR, the yield of the VBF signal is the largest of all Higgs-boson production processes. The yield of the ggF signal is 55 % of the VBF signal yield and still contributes to the sensitivity in this region. In the nonVBF SR, the ggF signal is the prevalent signal contribution while the VBF and VH signals both contribute less than 6 % of the ggF signal. The contribution from top-quarks constitutes the largest background contribution in the VBF SR while $Z \to \tau\tau$ is

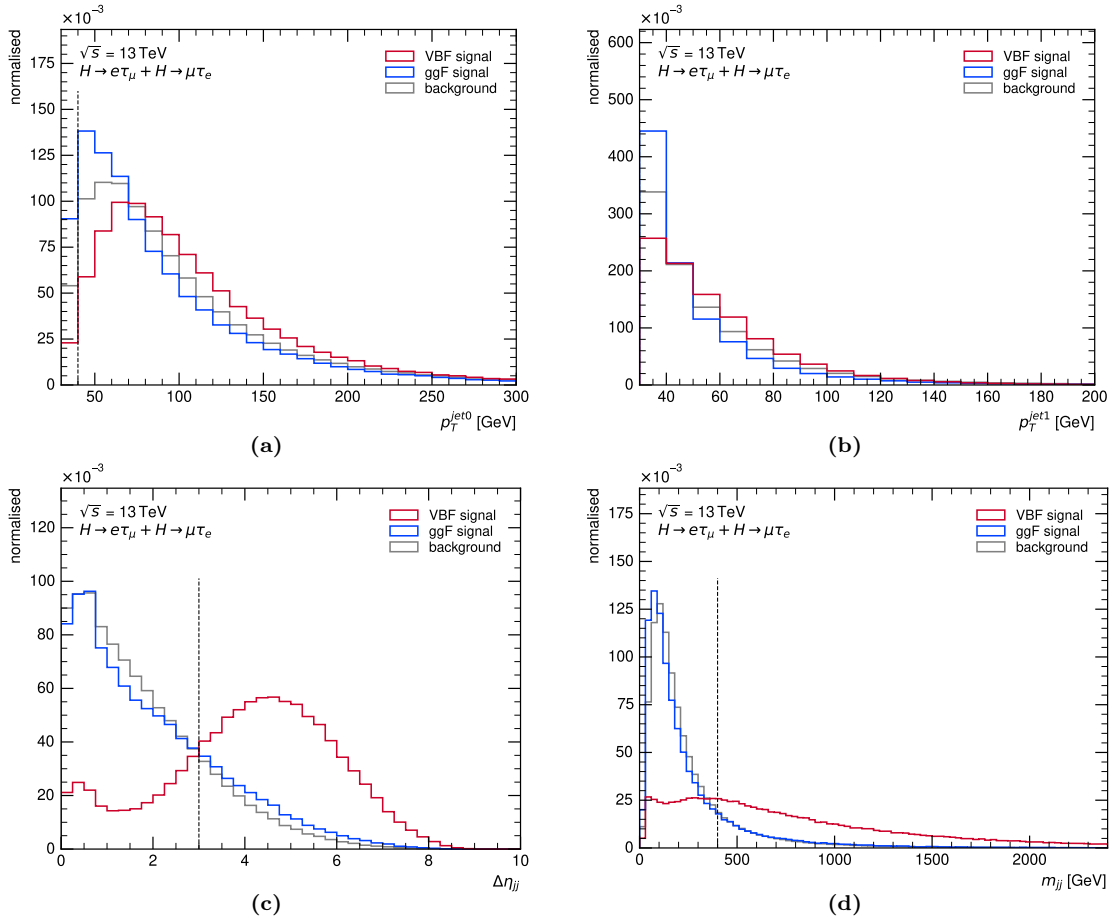



**Figure 8.1:** Normalised distributions of the observables ($p_{\mathrm{T}}^{j_0}$ (a), $p_{\mathrm{T}}^{j_1}$ (b), $\Delta\eta_{jj}$ (c), $m_{jj}$ (d)) defining the VBF SR after at least two jets (with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$) are required. The $e\tau$- and $\mu\tau$-datasets are combined. For signal, only the correctly classified signal in each dataset is considered. The VBF signal of interest and the *gluon-gluon fusion* (ggF) signal are shown individually. The background processes comprise all symmetric background processes estimated from MC simulations and both *fake*-components, $j \to \ell$- and *other* $\to \ell$-fakes. The dashed black lines indicate the value at which a cut is placed on the observables in order to define the VBF SR.

dominant in the nonVBF SR. The $s/\sqrt{s+b}$-significance in the VBF SR is smaller compared to the nonVBF SR while the $s/b$-ratio is higher.

The tables also list the number of signal events which are wrongly classified into the $e\tau$- or $\mu\tau$-dataset, i.e. for which the lepton assignment method described in Section 7.1.3 does not yield the correct assignment of the light leptons to the Higgs boson and the $\tau$-lepton. The prediction for the $e/\mu$-symmetric background is listed once estimated from MC-simulations and once estimated with the *Symmetry Method*. The events of the contributions estimated with the *Symmetry Method* are from the respective other dataset and are corrected with the efficiency-ratio $\mathcal{R}_\varepsilon$ (compare Eq. 7.16). The $\mathcal{R}_\varepsilon \cdot j \to \ell$- and $\mathcal{R}_\varepsilon \cdot \mathit{other} \to \ell$-contributions need to be subtracted and are hence listed with a negative sign. The total background prediction with the *Symmetry Method* is close to data, the deviation in the VBF SR is around 1 % and 1.7 % in the nonVBF SR. The deviations of the background prediction with simulated events is 10 % in the VBF SR and 2.1 % (1.1 %) for the $e\tau$-($\mu\tau$-)dataset in the nonVBF SR. The MC-smulated processes are, however, only normalised to their theoretical cross-section predictions and no dedicated normalisation factors are derived.

In order to find phase-space regions with an increased significance, *Neural Networks* are trained as discussed in the following.

| | | Yields | |
|---|---|---|---|
| | **Process** | $e\tau$ | $\mu\tau$ |
| Signal ($\mathcal{B} = 1\,\%$) | ggF$H \to e\tau_\mu$ | 34.70 ± 0.50 | 4.04 ± 0.17 |
| | VBF$H \to e\tau_\mu$ | 62.83 ± 0.27 | 5.31 ± 0.09 |
| | VH$H \to e\tau_\mu$ | 1.66 ± 0.07 | 0.41 ± 0.04 |
| | ggF$H \to \mu\tau_e$ | 4.24 ± 0.17 | 31.80 ± 0.50 |
| | VBF$H \to \mu\tau_e$ | 5.10 ± 0.08 | 57.44 ± 0.26 |
| | VH$H \to \mu\tau_e$ | 0.40 ± 0.04 | 1.33 ± 0.07 |
| Fake Contributions | $j \to \ell$ | 334 ± 22 | 314 ± 17 |
| | *other* $\to \ell$ | 80 ± 7 | 104 ± 10 |
| Symmetric Background from MC simulations | $Z \to \tau\tau$ | 1 098 ± 11 | 982 ± 10 |
| | Diboson | 757 ± 7 | 727 ± 6 |
| | Top | 2 226 ± 11 | 2 110 ± 10 |
| | $H \to WW$ | 103.20 ± 0.60 | 91.00 ± 0.60 |
| | $H \to \tau\tau$ | 43.69 ± 0.24 | 40.64 ± 0.24 |
| Symmetric Background from *Symmetry Method* | $R_\varepsilon \cdot$ Data | 4 330 ± 70 | 3 860 ± 60 |
| | $R_\varepsilon \cdot j \to \ell$ | −353 ± 19 | −290 ± 21 |
| | $R_\varepsilon \cdot$ *other* $\to \ell$ | −117 ± 12 | −71 ± 6 |
| Total Bkg. with Sym.Bkg. from MC simulations | | 4 643 ± 29 | 4 367 ± 25 |
| Total Bkg. with Sym.Bkg. from *Symmetry Method* | | 4 280 ± 80 | 3 920 ± 70 |
| $s/b$ | | 0.02 | 0.02 |
| $s/\sqrt{s+b}$ | | 1.50 | 1.43 |
| | Data | 4 230 | 3 967 |

**Table 8.1:** Yields after the VBF SR selection with statistical uncertainties, for the $e\tau$- and $\mu\tau$-datasets. The expected yields of the symmetric background contributions are given once estimated from MC simulations and once estimated with the *Symmetry Method*. For the latter, all three components (Data, $j \to \ell$- and *other* $\to \ell$-fakes) are of the respective other dataset, corrected with the efficiency ratio $R_\varepsilon$. The signal yields are given for a branching ratio of 1 %. The $s$ in the second to last line stands for the sum of the respective correctly classified signal events while the $b$ represents the total background prediction with the symmetric background (Sym.Bkg.) estimated with the *Symmetry Method*.

| | | Yields | |
|---|---|---|---|
| | **Process** | $e\tau$ | $\mu\tau$ |
| Signal ($\mathcal{B} = 1\,\%$) | ggF$H \to e\tau_\mu$ | 2 153 ± 4 | 144.90 ± 1.10 |
| | VBF$H \to e\tau_\mu$ | 112.50 ± 0.40 | 11.33 ± 0.12 |
| | VH$H \to e\tau_\mu$ | 92.10 ± 0.60 | 16.70 ± 0.24 |
| | ggF$H \to \mu\tau_e$ | 144.40 ± 1.10 | 1 882 ± 4 |
| | VBF$H \to \mu\tau_e$ | 11.48 ± 0.12 | 99.91 ± 0.34 |
| | VH$H \to \mu\tau_e$ | 17.71 ± 0.25 | 82.90 ± 0.50 |
| Fake Contributions | $j \to \ell$ | 17 970 ± 160 | 16 750 ± 120 |
| | $other \to \ell$ | 5 790 ± 70 | 11 050 ± 80 |
| Symmetric Background from MC simulations | $Z \to \tau\tau$ | 80 250 ± 140 | 69 620 ± 130 |
| | Diboson | 43 080 ± 80 | 38 740 ± 70 |
| | Top | 39 170 ± 50 | 35 250 ± 50 |
| | $H \to WW$ | 1 672 ± 4 | 1 398 ± 4 |
| | $H \to \tau\tau$ | 766.90 ± 1.60 | 681.30 ± 1.40 |
| Symmetric Background from *Symmetry Method* | $R_\varepsilon \cdot$ Data | 197 000 ± 500 | 160 900 ± 400 |
| | $R_\varepsilon \cdot j \to \ell$ | −19 900 ± 140 | −14 800 ± 140 |
| | $R_\varepsilon \cdot other \to \ell$ | −12 920 ± 110 | −5 150 ± 60 |
| Total Bkg. with Sym.Bkg. from MC simulations | | 188 690 ± 240 | 173 480 ± 210 |
| Total Bkg. with Sym.Bkg. from *Symmetry Method* | | 188 000 ± 600 | 168 700 ± 400 |
| $s/b$ | | 0.01 | 0.01 |
| $s/\sqrt{s+b}$ | | 5.40 | 5.00 |
| | Data | 184 887 | 171 675 |

**Table 8.2:** Yields after the nonVBF SR selection with statistical uncertainties, for the $e\tau$- and $\mu\tau$-datasets. The expected yields of the symmetric background contributions are given once estimated from MC simulations and once estimated with the *Symmetry Method*. For the latter, all three components (Data, $j \to \ell$- and $other \to \ell$-fakes) are of the respective other dataset, corrected with the efficiency ratio $R_\varepsilon$. The signal yields are given for a branching ratio of 1 %. The $s$ in the second to last line stands for the sum of the respective correctly classified signal events while the $b$ represents the total background prediction with the symmetric background (Sym.Bkg.) estimated with the *Symmetry Method*.

## 8.2 Application of Neural Networks

Classification *Neural Network*s (NNs) are utilised in this analysis to further separate the signal from the background processes. An introduction to NNs and the concepts relevant to the following discussions is given in Chapter 6. This section focuses on the implementation of these ideas within the scope of the $H$-LFV analysis. First, the strategy is summarised in Section 8.2.1, followed by the description of the training data in Section 8.2.2 including the choice of the respective classes. The definition and selection of the input variables is discussed in Section 8.2.3. The choice and optimisation of the *hyperparameters* is discussed in Section 8.2.4 and the results of the NN-trainings are presented in Section 8.2.5. Section 8.2.6 describes the binning strategy of the NN output node probability distributions for the final statistical analysis. The chapter concludes in Section 8.2.7 with a selection of studies performed leading to the final NN structure and setup.

### 8.2.1 Strategy

The NNs are trained after the loose *Basic Selection* in order to keep the number of events large for training. Good sensitivity in the statistical analysis (Chapter 10) is achieved by the separation power of the NN output node probability distributions which serve as final discriminants. Individual NNs are trained for the VBF and nonVBF SRs to be able to include the additional information from the two jets in the final state of the VBF signal topology. Two different approaches were tested for both regions. On one hand one NN with three output nodes corresponding to one signal and two background classes was trained and the signal node probability distribution was used in the statistical analysis. On the other hand three binary NNs, i.e. with one signal and one background class, were trained. For each of these NNs signal is trained versus another background class and the signal node distributions of each of these NNs are linearly combined into one distribution. Studies showed that the multi-class approach was superior for the NN in the nonVBF SR while the combination of binary NNs performed better in the VBF SR.

### 8.2.2 Training Data

As described in Section 6.1.3, supervised learning requires labelled data. That means for each event in the training dataset the class it belongs to must be known. The three natural classes in the $H$-LFV analysis are the LFV signal, the $e/\mu$-symmetric background contribution and the contribution from *fakes*. Due to their separate estimation, the labels for all events of these classes are known.

The signal events are estimated from MC simulations. For the $e/\mu$-symmetric background contribution events from simulated data are used for the training of the NNs as well instead of events from the estimation with the *Symmetry Method*. This is due to several reasons:

- The estimation with the *Symmetry Method* is based on the subtraction of the respective *fake*-contributions which leads to a considerable amount of events with negative event weights. This can be problematic as described in Section 6.3.1.
- The number of unweighted events of the estimation with simulated events is larger compared to the estimation with the *Symmetry Method*.
- The events can be split into more distinct classes when estimated from simulations (e.g. $Z \to \tau\tau$, top-quark production, diboson production) compared to the estimation with the *Symmetry Method* where only one class, the total $e/\mu$-symmetric contribution, is available. A finer splitting of the classes can yield better separation power also for signal versus total background.

Estimating the $e/\mu$-symmetric background contribution in the training from simulations, however, requires reasonable modelling of the data in the distributions of the input variables used in the NN training (in addition to the modelling with the *Symmetry Method*). The normalisation of the prediction compared to data is not relevant but the shapes must be reproduced correctly. When considering the sum of two processes, e.g. of top-quark and diboson production, the shape of the sum can be altered if their normalisation with respect to each other is inaccurate. This can introduce a small bias which implies that the trained NNs may perform not as well for the background estimation with the *Symmetry Method* as for the MC-simulations that they were trained on. Although the $e/\mu$-symmetric background estimate used for the NN training (MC-simulations) and the one used in the statistical analysis to derive the final results (*Symmetry Method*) are statistically independent, the usage of 10-fold cross validation (Section 6.2) is still a necessity due to the fact that the estimates of the signal and the *fakes* in the training and the statistical analysis are the same.

The estimate of the *fake* contributions for the training is the same as described in Section 7.2.2. The $other \rightarrow \ell$-fakes are estimated from simulations.

#### The Choice of the Classes

The three classes of the multi-class NN in the nonVBF SR – the *nonVBF NN*– correspond to the three natural classes of this analysis: LFV signal, $e/\mu$-symmetric background and *fakes*. The $e/\mu$-symmetric background consists of the simulated events of $Z \rightarrow \tau\tau$, top-quark, diboson, $H \rightarrow WW$ and $H \rightarrow \tau\tau$ production. In addition, the $other \rightarrow \ell$-fakes are added to this class rather than to the *fakes* class, given that the majority of the $other \rightarrow \ell$-fakes originates from $Z$-boson decays and hence their distributions are more similar to the ones of the $e/\mu$-symmetric background than of the $j \rightarrow \ell$-fakes. The *fakes* class in the NN training only consists of the $j \rightarrow \ell$-fakes.

In the VBF SR, each of the three binary NNs discriminates the LFV signal against one background class. The names of the three NNs and the respective background classes are:

- $VBF_{Z\rightarrow\tau\tau}$ *NN*: $Z \rightarrow \tau\tau + other \rightarrow \ell$-fakes $+ H \rightarrow \tau\tau$
- $VBF_{top}$ *NN*: top-quark production + diboson production $+ H \rightarrow WW$
- $VBF_{fakes}$ *NN*: $j \rightarrow \ell$-fakes.

Compared to the *nonVBF NN*, the $e/\mu$-symmetric background is split into two classes while the remaining classes are the same. Both setups, binary and multi-class NN, were tested in both SRs, and different sets (and numbers) of classes were investigated in either SR. The setups described above showed the best performance in either region. One caveat is that the NN hyperparameters were not retuned for every variation which can in principle impact the performance. The effect is, however, expected to be sufficiently small such that the computational overhead for the optimisation of the hyperparameters was considered too large, and the comparisons were made using the same hyperparameters in most cases.

In all cases, the signal class consists of the LFV signal produced via the three production processes that are considered in this analysis (ggF, VBF and VH).

#### Training Statistics

In order to enhance the training statistics, the $e\tau$- and $\mu\tau$-datasets are trained together. One caveat is that in particular the *fakes* are expected to have different shapes in the $e\tau$- and $\mu\tau$-datasets and could benefit from a training performed separately in both datasets. The shapes of the input variable distributions for the two datasets of the $j \rightarrow \ell$-fakes are

compared in Figures 8.2 and 8.3 for the nonVBF SR and in Figures 8.4–8.6 for the VBF SR. The definitions of the single input variables are given in Section 8.2.3. Small shape differences are visible in the ratio of the $j \to \ell$-fakes of the two datasets for a few variables. There are larger differences in the distribution of the transverse mass of the lepton from the Higgs boson and $E_{\mathrm{T}}^{\mathrm{miss}}$, $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, in Figure 8.2(j) for larger values of $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$. In this mass range, however, the contributions of the $j \to \ell$-fakes is small. Overall, the differences are considered small enough to justify a combined training.

The available training statistics – number of unweighted events – for each of the SRs are listed in Table 8.3. In order to enhance the statistics for the training of the *VBF NN*s, the requirement on the invariant mass of the two jets was loosened to $m_{jj} > 300\,\mathrm{GeV}$ instead of $m_{jj} > 400\,\mathrm{GeV}$ in the definition of the VBF SR (compare Section 8.1).

| **Process** | **nonVBF SR** | **VBF SR** |
|---|---|---|
| LFV signal | 856 722 | 148 990 |
| $Z \to \tau\tau$ | 3 639 955 | 143 062 |
| $other \to \ell$-fakes | 157 082 | 3 670 |
| $H \to \tau\tau + H \to WW$ | 1 281 650 | 292 887 |
| Top-quarks | 1 648 630 | 141 006 |
| Diboson | 4 103 124 | 204 069 |
| $j \to \ell$-fakes | 2 096 805 | 130 460 |
| $j \to \ell$-fakes data | 238 844 | 6 240 |
| $j \to \ell$-fakes MC | 1 857 961 | 124 220 |

**Table 8.3:** The available statistics – number of unweighted events – for training, passing the nonVBF SR selection (left) or passing the VBF SR selection (right). The events of the $e\tau$- and $\mu\tau$-datasets are combined. "Signal" comprises all the considered production modes. Only the events of both signals, $H \to e\tau$ and $H \to \mu\tau$, which are correctly classified to each of the datasets are considered. The individual components of the $j \to \ell$-fakes, the data part and the subtracted prompt lepton contribution estimated from MC simulations, are listed individually as well as combined.

The LFV signal is the sum of the $H \to e\tau$- and $H \to \mu\tau$-signals. But only the signal events that are correctly classified following the lepton assignment in Section 7.1.3 are considered.

For the $j \to \ell$-fakes two contributions are listed separately in Table 8.3: the data and the subtracted prompt lepton contribution estimated from MC simulations. The subtraction is realised by negative event weights and the total sum-of-weights of the data and the simulated events is positive. The number of events of the subtracted contribution with negative event weights is, however, substantially larger than the number of data events. This can lead to an unstable NN training as described in Section 6.3.1. In order to facilitate a stable training process, only the data part of the $j \to \ell$-fakes is utilised in the training when the $j \to \ell$-fakes form their own class. This is the case for the *nonVBF NN* as well as for the $VBF_{fakes}$ *NN*. In order to see the impact of the neglected prompt lepton contribution, the shapes of the input variable distributions are compared for the full $j \to \ell$-fake estimate and its data part only. Figures 8.7 and 8.8 display the normalised input variable distributions for the nonVBF SR and Figures 8.9–8.11 for the VBF SR. The definitions of the single input variables are given in Section 8.2.3. For most distributions, the agreement in the shapes is very good such that the data-only component can be seen as a decent approximation of the shape of the full $j \to \ell$-fake background estimate. In a few distributions, larger impacts of the subtracted MC contributions on the shape are observed, e.g. in the distribution of the transverse mass of $\ell_H$ and $E_{\mathrm{T}}^{\mathrm{miss}}$, $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, (Figure 8.7(j)), in the distribution of the ratio of both leptons, $p_{\mathrm{T}}^{\ell_H}/p_{\mathrm{T}}^{\ell_\tau}$, (Figure 8.7(d)) and in the distribution of the reconstructed

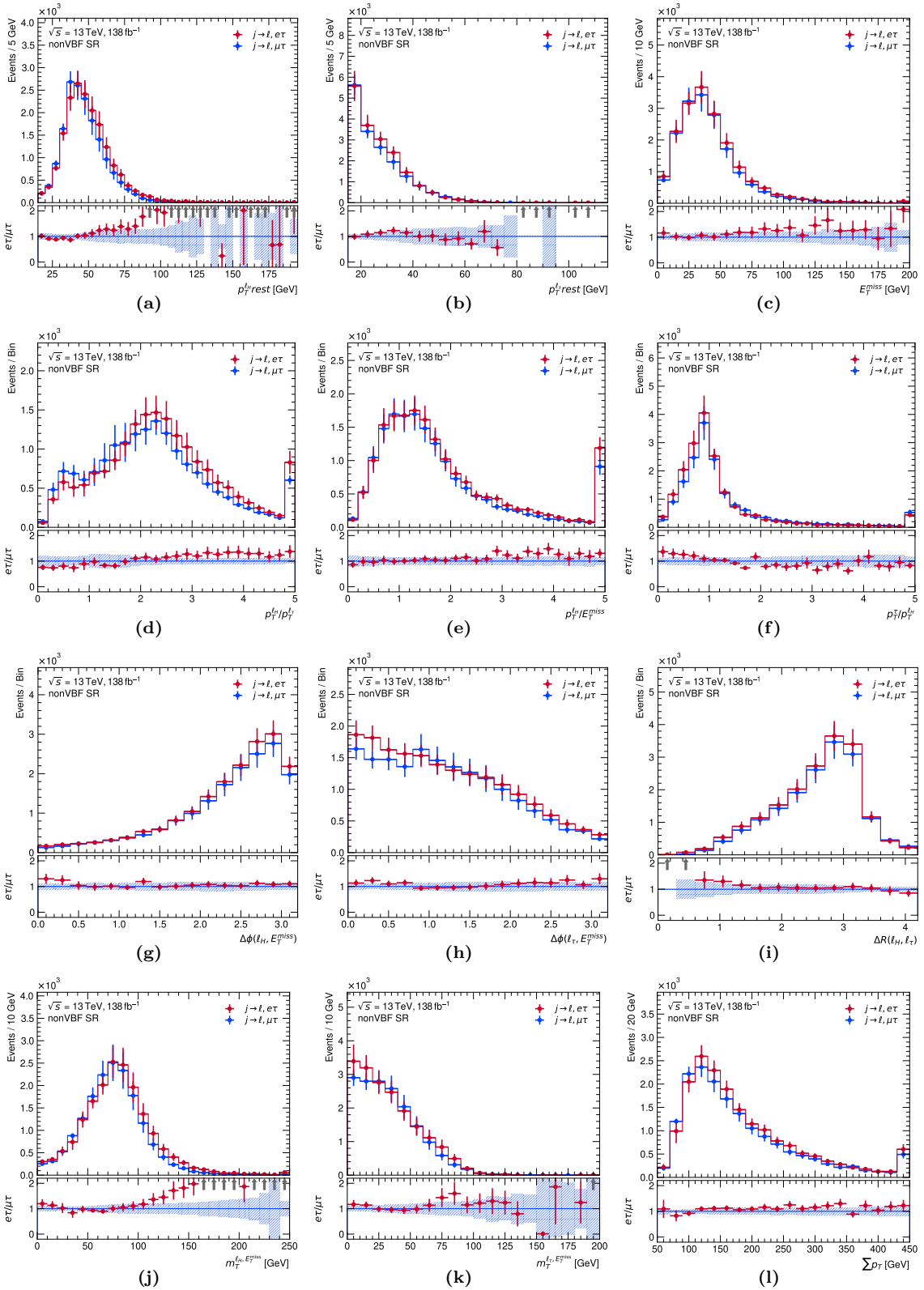



**Figure 8.2:** Distributions of input variables to the *nonVBF NN* separately for the $j \to \ell$-fakes in the $e\tau$- and $\mu\tau$-dataset. The lower panels show the ratio of the $j \to \ell$-fakes of both datasets. The uncertainties (bars in the upper panels, bands in the lower panels) comprise the statistical uncertainties as well as the systematic uncertainties of the *fake* estimate.

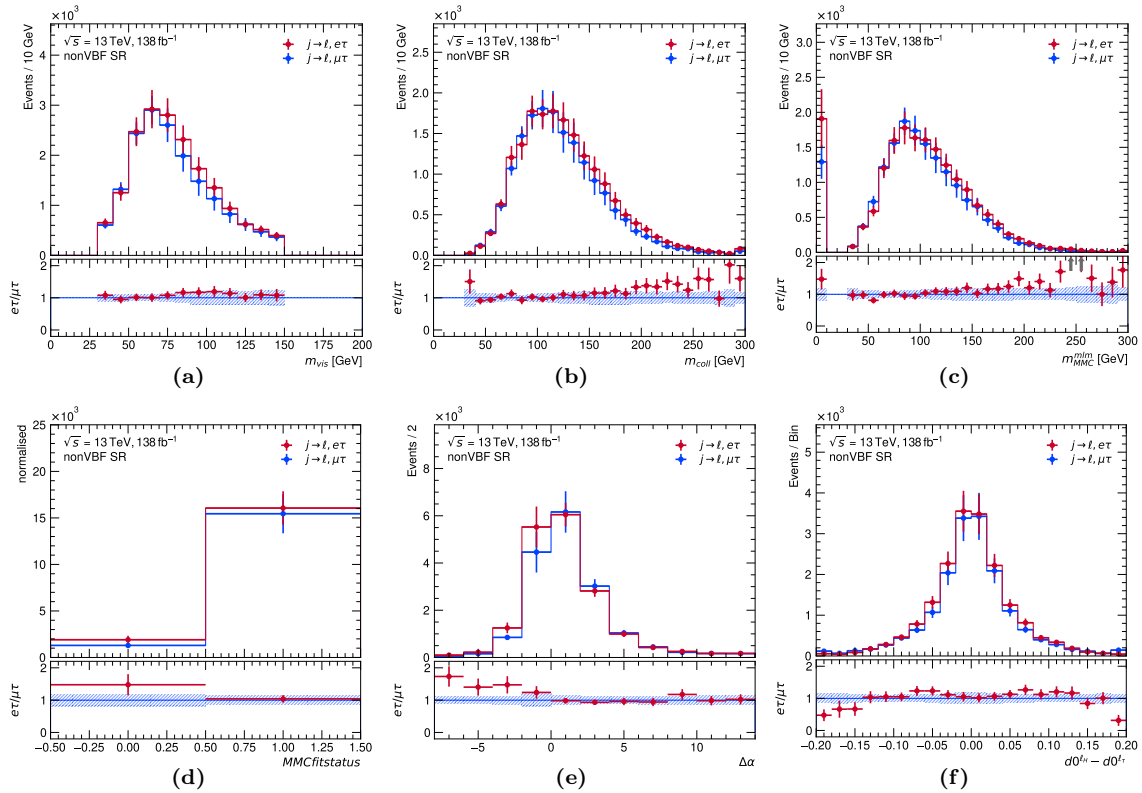



**Figure 8.3:** Distributions of input variables to the *nonVBF NN* separately for the $j \to \ell$-fakes in the $e\tau$- and $\mu\tau$-dataset. The lower panels show the ratio of the $j \to \ell$-fakes of both datasets. The uncertainties (bars in the upper panels, bands in the lower panels) comprise the statistical uncertainties as well as the systematic uncertainties of the *fake* estimate.

Higgs-boson mass in the collinear approximation, $m_{\text{coll}}$, (Figure 8.8(b)) in the nonVBF SR. The uncertainties in the full $j \to \ell$-fake estimate include those of the data-only part, thus the uncertainties of both distributions need to be considered as correlated which is not reflected in the figures. This lets the distributions appear slightly more in agreement than they actually are. The main consequence of neglecting the subtracted MC contribution in the NN training is a potentially reduced separation power when the NN is applied to the full $j \to \ell$-fake estimate. A combination of the $j \to \ell$-fakes with other processes was tested to see whether this would allow for keeping the subtracted MC part, but having a separate class with only the data part of the $j \to \ell$-fakes showed a better performance and was thus chosen.

### Preprocessing

The preprocessing of the training data ensures that the physics processes which are combined into one class are weighted by their respective cross section. Furthermore, it ensures that the classes are all weighted equally, i.e. that the sum-of-weights of each class is the same. A detailed description of the weighting can be found in Section 6.3.1. One study with larger weights for one class is described in Section 8.2.7, but otherwise equal weighting across output classes is used in the following.

As detailed in Section 6.2, 10-fold cross validation is used to split the available data into a *training*, *validation* and *test set*. As a consequence, 10 single NNs with identical *hyperparameters* are trained for each setup. A reproducible random number decides for each event which of the NNs is used. In the following, one NN refers to one setup, i.e. one set of *hyperparameters* and input features, which consists of these 10 NNs. Although

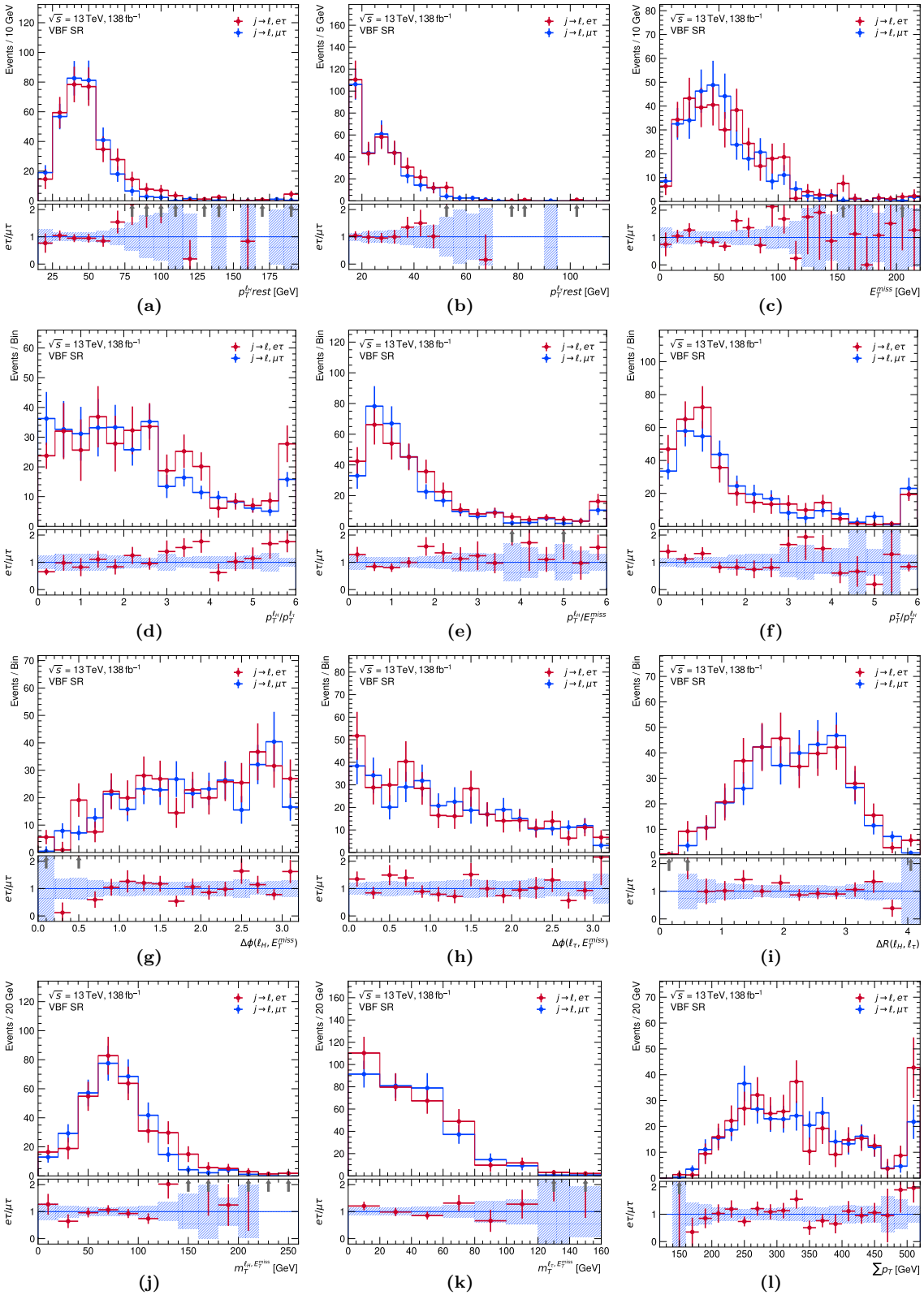



**Figure 8.4:** Distributions of input variables to the *VBF NN*s separately for the $j \to \ell$-fakes in the $e\tau$- and $\mu\tau$-dataset. The lower panels show the ratio of the $j \to \ell$-fakes of both datasets. The uncertainties (bars in the upper panels, bands in the lower panels) comprise the statistical uncertainties as well as the systematic uncertainties of the *fake* estimate.

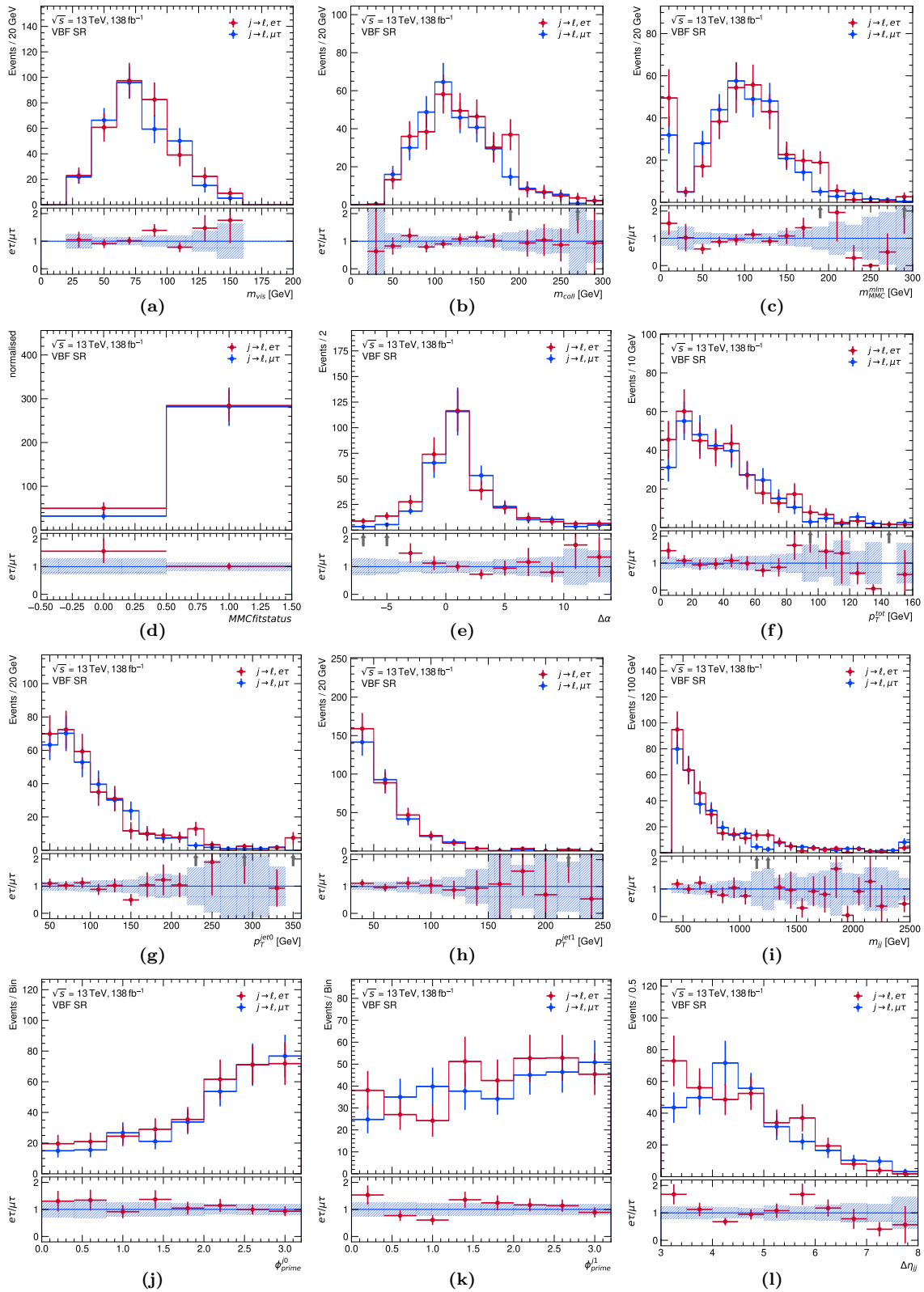
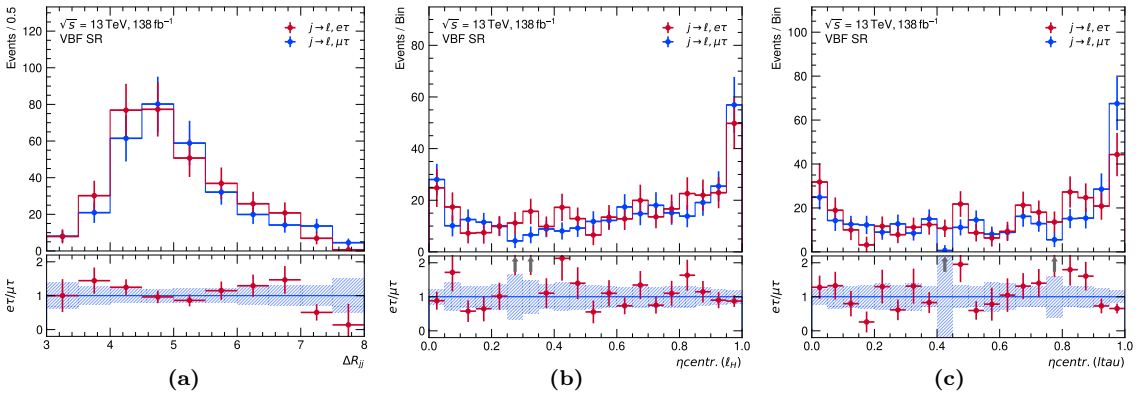



**Figure 8.5:** Distributions of input variables to the *VBF NN*s separately for the $j \to \ell$-fakes in the $e\tau$- and $\mu\tau$-dataset. The lower panels show the ratio of the $j \to \ell$-fakes of both datasets. The uncertainties (bars in the upper panels, bands in the lower panels) comprise the statistical uncertainties as well as the systematic uncertainties of the *fake* estimate.




**(a)** **(b)** **(c)**

**Figure 8.6:** Distributions of input variables to the *VBF NN*s separately for the $j \to \ell$-fakes in the $e\tau$- and $\mu\tau$-dataset. The lower panels show the ratio of the $j \to \ell$-fakes of both datasets. The uncertainties (bars in the upper panels, bands in the lower panels) comprise the statistical uncertainties as well as the systematic uncertainties of the *fake* estimate.

the events used for the $e/\mu$-symmetric background in the training are independent of the events used in the final statistical analysis (estimated with the *Symmetry Method*), cross validation is needed since the estimates of the *fakes* and the LFV signal are the same in the training and the statistical analysis. In addition, cross validation allows for defining not only a *test set* but also a *validation set*.

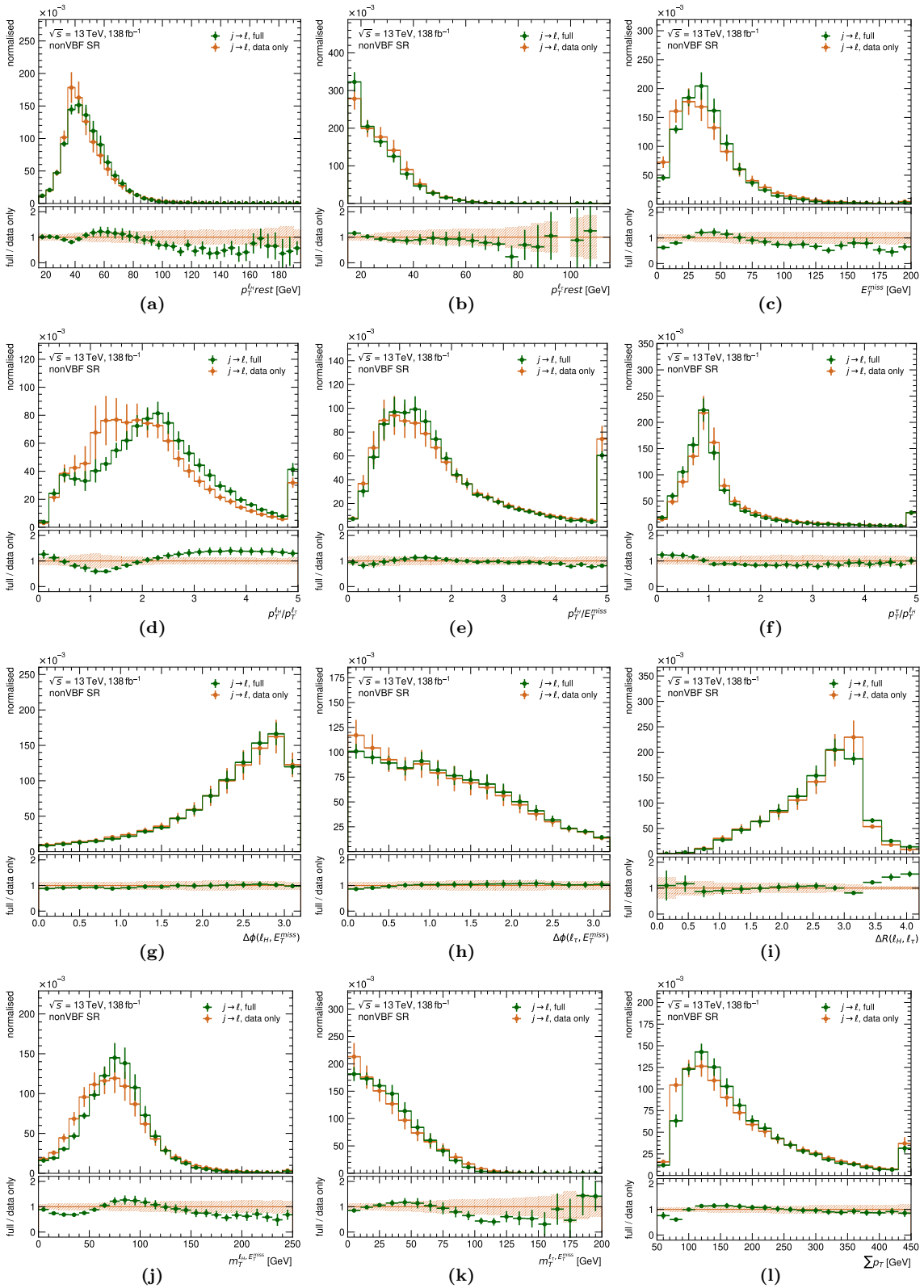



**Figure 8.7:** Normalised distributions of the input variables to the *nonVBF NN* shown separately for the full estimate of the $j \to \ell$-fakes and its data-only part. The events of the $e\tau$- and $\mu\tau$-dataset are summed up. The lower panels show the ratio of the full estimate and the data part only. The uncertainties (bars in the upper panels, bands in the lower panels) comprise the statistical uncertainties as well as the systematic uncertainties on the *fake* estimate. A correlation of the systematic uncertainties between both distributions is not considered.

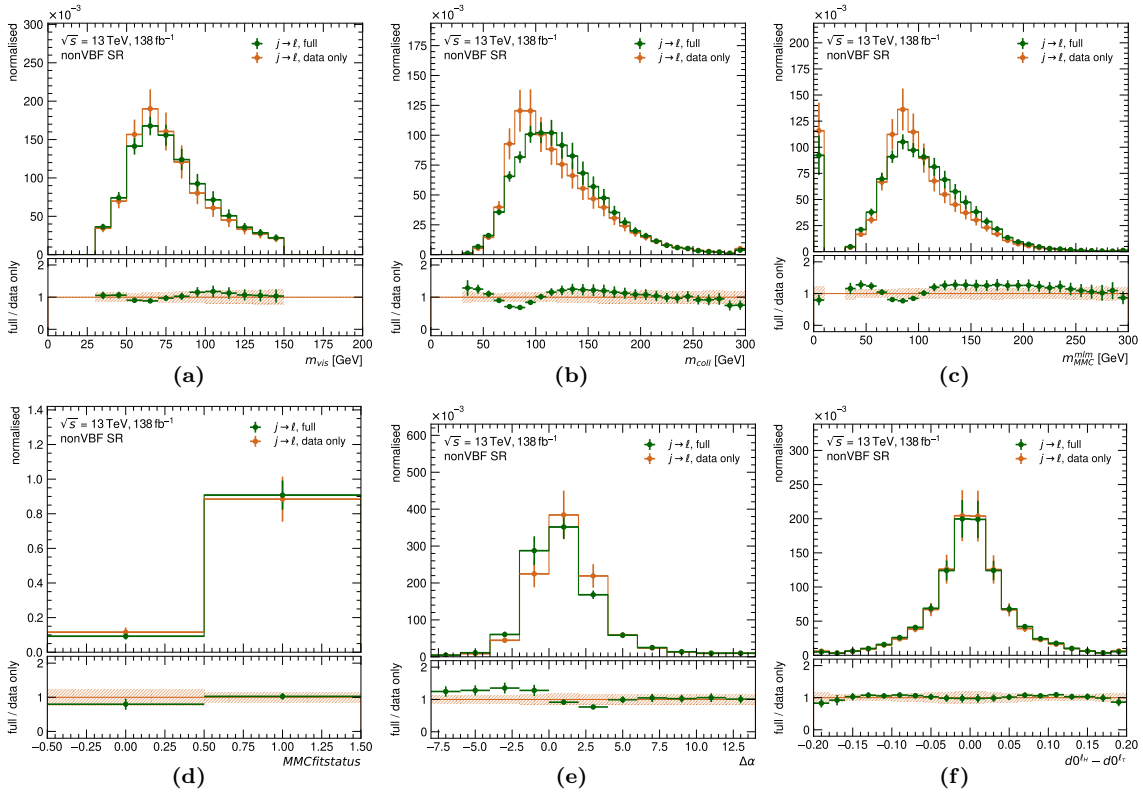



**Figure 8.8:** Normalised distributions of the input variables to the *nonVBF NN* shown separately for the full estimate of the $j \to \ell$-fakes and its data-only part. The events of the $e\tau$- and $\mu\tau$-dataset are summed up. The lower panels show the ratio of the full estimate and the data part only. The uncertainties (bars in the upper panels, bands in the lower panels) comprise the statistical uncertainties as well as the systematic uncertainties on the *fake* estimate. A correlation of the systematic uncertainties between both distributions is not considered.

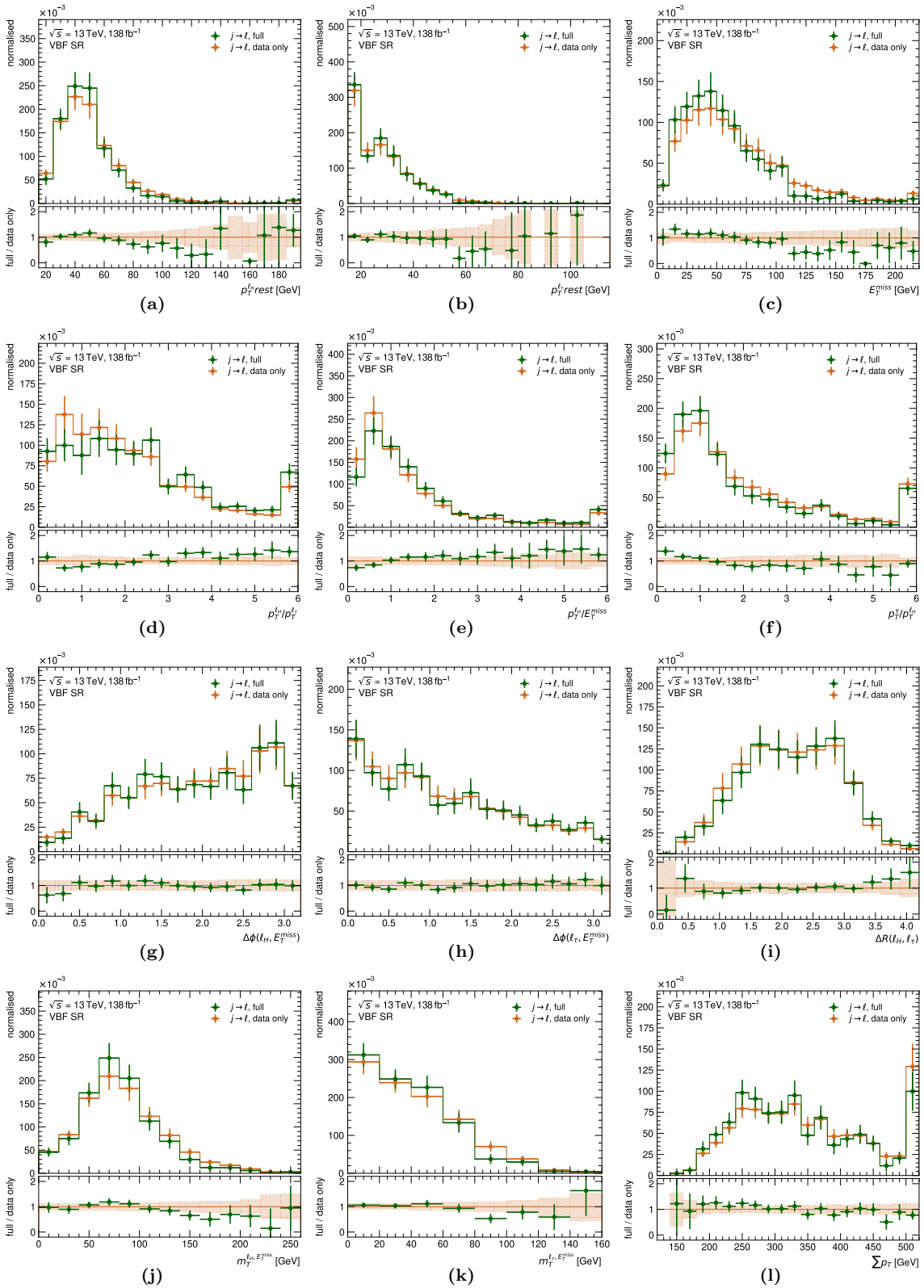



**Figure 8.9:** Normalised distributions of the input variables to the *VBF NN*s shown separately for the full estimate of the $j \to \ell$-fakes and its data-only part. The events of the $e\tau$- and $\mu\tau$-dataset are summed up. The lower panels show the ratio of the full estimate and the data part only. The uncertainties (bars in the upper panels, bands in the lower panels) comprise the statistical uncertainties as well as the systematic uncertainties on the *fake* estimate. A correlation of the systematic uncertainties between both distributions is not considered.

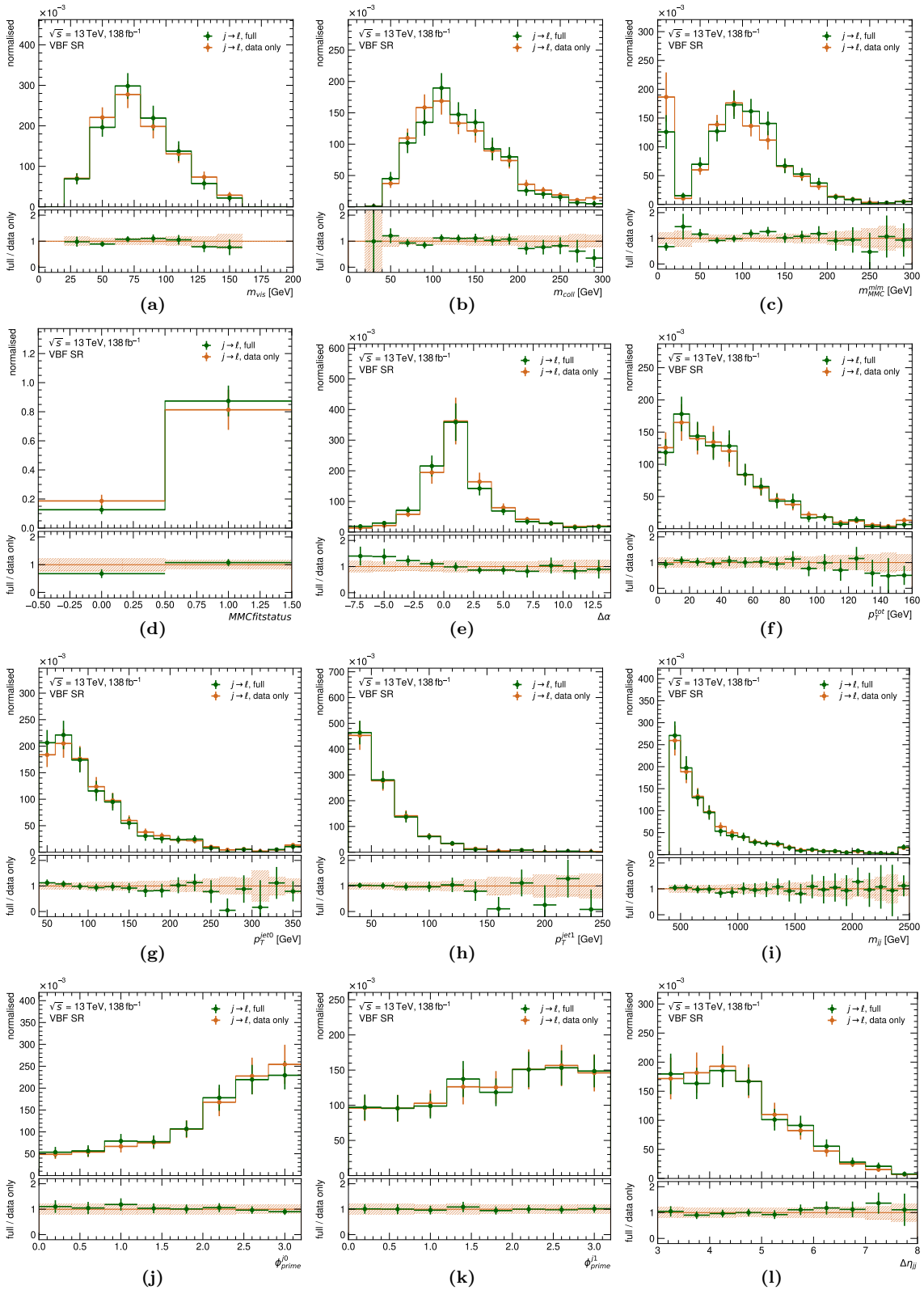
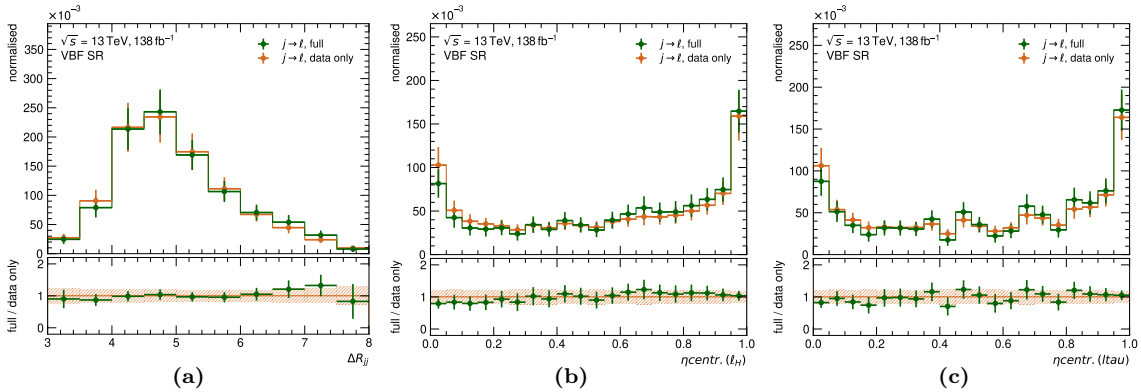



**Figure 8.10:** Normalised distributions of the input variables to the *VBF NN*s shown separately for the full estimate of the $j \to \ell$-fakes and its data-only part. The events of the $e\tau$- and $\mu\tau$-dataset are summed up. The lower panels show the ratio of the full estimate and the data part only. The uncertainties (bars in the upper panels, bands in the lower panels) comprise the statistical uncertainties as well as the systematic uncertainties on the *fake* estimate. A correlation of the systematic uncertainties between both distributions is not considered.




(a) (b) (c)

**Figure 8.11:** Normalised distributions of the input variables to the *VBF NN*s shown separately for the full estimate of the $j \to \ell$-fakes and its data-only part. The events of the $e\tau$- and $\mu\tau$-dataset are summed up. The lower panels show the ratio of the full estimate and the data part only. The uncertainties (bars in the upper panels, bands in the lower panels) comprise the statistical uncertainties as well as the systematic uncertainties on the *fake* estimate. A correlation of the systematic uncertainties between both distributions is not considered.

### 8.2.3 Input Variables

Sets of input features, also called input variables, can include low-level or high-level variables only, or a mixture, as discussed in Section 6.3. Results based on low-level input features only are summarised in Section 8.2.7. Given the limited training statistics, a set of low- and high-level input features with rather shallow NNs (Section 8.2.4) performed best and was chosen for the analysis. This is discussed in the following.

The *nonVBF NN* uses 18 input variables, the *VBF NN*s use 17 of these 18 variables plus 10 additional, VBF-specific ones. All three *VBF NN*s use the same variables. The variables and a short description are listed in Table 8.4. More detailed explanations, the definitions and motivation for why they were chosen is given below.

Distributions of the input variables in different visual representations are shown in Figures 8.12, 8.14, 8.16, 8.18 and 8.20 in the nonVBF SR and in Figures 8.13, 8.15, 8.17, 8.19, 8.21, 8.22 and 8.23 in the VBF SR. The left and middle columns show the comparison of the prediction to data and the ratio of data over background prediction in the lower panels of the figures. In the left column, the $e/\mu$-symmetric background is estimated from MC simulations with each process normalised to its predicted cross section. In the middle column, the $e/\mu$-symmetric background is estimated with the *Symmetry Method* which is used in the statistical analysis to derive the final results. Hence, good modelling of the data by the prediction with the *Symmetry Method* is required. The right column shows the normalised distributions for each of the individual classes used in the NN-training to visualise the potential separation power of the individual input variables. However, separation power of one variable can also arise only in conjunction with other variables which is not perceptible from its distribution. The left and middle columns are shown for the $e\tau$-dataset while the right column considers the sum of both datasets. The same set of figures for the $\mu\tau$-dataset is shown in Appendix C where the figures in the third column are identical to the ones shown here. The data-over-background ratio in the middle columns for the $\mu\tau$-dataset often shows the opposite behaviour to the one observed in the $e\tau$-dataset, thus showing only one here is sufficient. This "mirror"-feature between the $e\tau$- and $\mu\tau$-datasets is intrinsic to the *Symmetry Method*. If the data of the $\mu\tau$-dataset shows an upwards fluctuation, this propagates to the $e/\mu$-symmetric background estimate $\tilde{n}$ for the $e\tau$-dataset and leads to an overprediction of the data and hence a downward fluctuation in the data over background ratio in the $e\tau$-dataset, and vice versa.

The prediction based on simulated events is known to overestimate the data in the VBF SR. The partner analyses of this analysis, based on MC-templates for background estimation [22], have derived normalisation factors for the $Z \to \tau\tau$ and top-quark production processes which have been determined to be well below 1. The normalisation factors are not applied in the figures here. The data-over-background ratios in the figures of the VBF SR, with the $e/\mu$-symmetric background estimated from simulated events, are therefore expected to be below 1. In addition, neglecting the normalisation factors also impacts the shape of the total background prediction and hence the shape of the data-over-background ratios. Thus, only limited attention should be given to the modelling of these distributions.

The signal-enhanced bins of the distributions are blinded whenever the expected signal-over-background ratio ($s/b$) exceeds 5 % (assuming a branching ratio of 1 % for signal) such that the judgement is not biased by information from the signal-sensitive bins.

**$p_{\mathrm{T}}^{\ell_H}$ rest, $p_{\mathrm{T}}^{\ell_\tau}$ rest:** The transverse momenta $p_{\mathrm{T}}$ of the two leptons ($\ell_H$ and $\ell_\tau$) in the approximate Higgs-boson rest frame (see Section 7.1.3) show a better separation power than the ones in the laboratory frame (Figures 8.12(a)–8.12(f) for the *nonVBF NN* and Figures 8.13(a)–8.13(f) for the *VBF NN*s). In the case of signal, $p_{\mathrm{T}}^{\ell_H}$ rest is expected to be considerably larger than $p_{\mathrm{T}}^{\ell_\tau}$ rest, which can be exploited by the NNs. This $p_{\mathrm{T}}$-imbalance is larger in the approximate Higgs-boson rest-frame than in the laboratory frame, since in the

| | |
|---|---|
| ***nonVBF NN* and *VBF NN*s:** | |
| $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_H),\ p_{\mathrm{T}}^{\mathrm{rest}}(\ell_\tau)$ | transverse momenta of the two leptons in the approximate Higgs-boson rest frame |
| $E_{\mathrm{T}}^{\mathrm{miss}}$ | missing transverse energy (Section 5.6) |
| $p_{\mathrm{T}}(\ell_H)/p_{\mathrm{T}}(\ell_\tau)$ | ratio of the transverse momenta of the two leptons |
| $p_{\mathrm{T}}(\ell_H)/E_{\mathrm{T}}^{\mathrm{miss}}$ | ratio of the transverse momenta of the lepton from the $H$-boson and the missing transverse energy |
| $p_{\mathrm{T}}(\tau)/p_{\mathrm{T}}(\ell_H)$ | ratio of the transverse momentum of the $\tau$-lepton and the lepton from the $H$-boson |
| $\Delta\phi(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}}),\ \Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ | absolute values of the differences of the azimuthal angles of the respective lepton and the missing transverse energy direction |
| $\Delta R(\ell_H, \ell_\tau)$ | angular separation of the two leptons |
| $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}}),\ m_T(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ | transverse masses of the respective lepton and the missing transverse energy |
| $\sum p_{\mathrm{T}}$ | scalar sum of transverse momenta of all high-$p_{\mathrm{T}}$ objects in the event |
| $m_{\ell\ell}$ | invariant mass of the two leptons |
| $m_{\mathrm{coll}}$ | collinear mass; aims to reconstruct the Higgs-boson mass |
| $m_{\mathrm{MMC}}^{\mathrm{mlm}}$, MMC fit status | MMC mass and fit status; aims to reconstruct the Higgs-boson mass |
| $\Delta\alpha$ | measure of the collinearity of the $\tau$-lepton decay products |
| $d0(\ell_H) - d0(\ell_\tau)$ (*nonVBF* only) | difference of the transverse impact parameters of the two leptons |
| ***VBF NN*s only:** | |
| $p_{\mathrm{T}}(j_0),\ p_{\mathrm{T}}(j_1)$ | transverse momenta of the two leading jets |
| $\Delta\phi(j_0, E_{\mathrm{T}}^{\mathrm{miss}}),\ \Delta\phi(j_1, E_{\mathrm{T}}^{\mathrm{miss}})$ | absolute values of the differences of the azimuthal angles of the respective jet and the missing transverse energy energy |
| $m_{jj}$ | invariant mass of the two leading jets |
| $\Delta\eta_{jj},\ \Delta R_{jj}$ | absolute value of the difference in $\eta$ and angular separation of the two leading jets |
| $p_T^{\mathrm{tot}}$ | absolute value of the vectorial sum of the transverse momenta of the two leptons, $E_{\mathrm{T}}^{\mathrm{miss}}$ and the two leading jets |
| $\eta$-centr.$(\ell_H),\ \eta$-centr.$(\ell_\tau)$ | $\eta$-centrality of the respective lepton relative to the $\eta$-values of the two leading jets |

**Table 8.4:** Input variables used for the *Neural Network*s with a brief description. A detailed description and the definitions of the variables is given in the text.

latter the effect is smeared out. Both variables show a good modelling of the data by the background prediction based on the *Symmetry Method*.

The $p_{\mathrm{T}}^{\ell_H}$ rest- and $p_{\mathrm{T}}^{\ell_\tau}$ rest-variables are the only two variables measured in the approximate Higgs-boson rest-frame. All other variables used are measured in the laboratory frame.

**$p_{\mathrm{T}}(\ell_H)/p_{\mathrm{T}}(\ell_\tau)$, $p_{\mathrm{T}}(\ell_H)/E_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}(\tau)/p_{\mathrm{T}}(\ell_H)$:** The transverse momenta of the leptons in the laboratory frame enter the training via several ratios: $p_{\mathrm{T}}(\ell_H)/p_{\mathrm{T}}(\ell_\tau)$ (Figures 8.12(j)–8.12(l), Figures 8.13(j)–8.13(l)) is expected to be larger for the LFV signal than for e.g. $Z \to \tau\tau$ where both leptons have neutrinos associated from the $\tau$-lepton decay. The same argument holds for the ratio $p_{\mathrm{T}}(\ell_H)/E_{\mathrm{T}}^{\mathrm{miss}}$ (Figures 8.14(a)–8.14(c), Figures 8.15(a)–8.15(c)). The transverse momentum of the $\tau$-lepton in the ratio $p_{\mathrm{T}}(\tau)/p_{\mathrm{T}}(\ell_H)$ (Figures 8.14(d)–8.14(f), Figures 8.15(d)–8.15(f)) is obtained as the transverse momentum of the $\ell_\tau$+$E_{\mathrm{T}}^{\mathrm{miss}}$-system. And hence this should be approximately 1 for the LFV signal, which is the case in the nonVBF SR while it is shifted to slightly smaller values in the VBF SR. In addition, the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ itself serves as input variable (Figures 8.12(g)–8.12(i), Figures 8.13(g)–8.13(i)). The ratios as well as $E_{\mathrm{T}}^{\mathrm{miss}}$ show a convincing modelling by the background predictions. A small slope is visible in the data over background ratio of $p_{\mathrm{T}}(\ell_H)/E_{\mathrm{T}}^{\mathrm{miss}}$ when the $e/\mu$-symmetric background is estimated from simulated events. In this region – at low values of $p_{\mathrm{T}}(\ell_H)/E_{\mathrm{T}}^{\mathrm{miss}}$ – the predictions are dominated by diboson and top-quark production, which are estimated by data via the *Symmetry Method* in the statistical analysis for which good modelling is observed. The small slope therefore has little effect.

**$\Delta\phi(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$:** The angular separation of $\ell_H$ from $E_{\mathrm{T}}^{\mathrm{miss}}$ in the transverse plane, $\Delta\phi(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$ (Figures 8.14(g)–8.14(i), Figures 8.15(g)–8.15(i)), is expected to be large for signal and equal to $\pi$ if the Higgs-boson is at rest with respect to the transverse direction. Consequently, $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ (Figures 8.14(j)–8.14(l), Figures 8.15(j)–8.15(l)) is expected to be small for signal. Both is clearly visible in the respective figures. In the VBF SR, however, $\Delta\phi(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$ also prefers larger values for signal but does not peak at the highest value and instead shows a plateau from 1.5 onwards. Since $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ peaks at small values for signal in the VBF SR, this hints that the leptons themselves are closer together in $\phi$ in this SR. While $Z \to \tau\tau$-events in the nonVBF SR also prefer larger values it is the opposite in the VBF SR. The modelling of both distributions is very good, a small increasing slope in the data over background ratio in the nonVBF SR can be observed when the $e/\mu$-symmetric background is estimated from simulated events. This is also the case for $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ when the $e/\mu$-symmetric background is estimated with the *Symmetry Method* but it is acceptable within uncertainties.

**$\Delta R(\ell_H, \ell_\tau)$:** The angular separation $\Delta R(\ell_H, \ell_\tau) = \sqrt{(\Delta\eta(\ell_H, \ell_\tau))^2 + (\Delta\phi(\ell_H, \ell_\tau))^2}$ between the two leptons (Figures 8.16(a)–8.16(c), Figures 8.17(a)–8.17(c)) is peaking at higher values, around $\pi$ for all processes in the nonVBF SR. In the VBF SR it is condensed in a hill-shape around 2.0 with $Z \to \tau\tau$-events preferring slightly smaller values. In each region, the behaviour observed for the $\Delta\phi$-distributions is reflected. The modelling is very good in both regions.

**$m_T\left(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}}\right)$, $m_T\left(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}}\right)$:** The transverse mass of a lepton $\ell$ and of the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined as:

$$m_T\left(\ell, E_{\mathrm{T}}^{\mathrm{miss}}\right) = \sqrt{2p_{\mathrm{T}}^{\ell} E_{\mathrm{T}}^{\mathrm{miss}} \left(1 - \cos\left\{\Delta\phi\left(\ell, E_{\mathrm{T}}^{\mathrm{miss}}\right)\right\}\right)}\,. \tag{8.1}$$

The transverse mass of $\ell_H$ (Figures 8.16(d)–8.16(f), Figures 8.17(d)–8.17(f)) is expected to be at larger values for signal since $\Delta\phi\left(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}}\right)$ as well as $p_{\mathrm{T}}^{\ell_H}$ are at larger values as

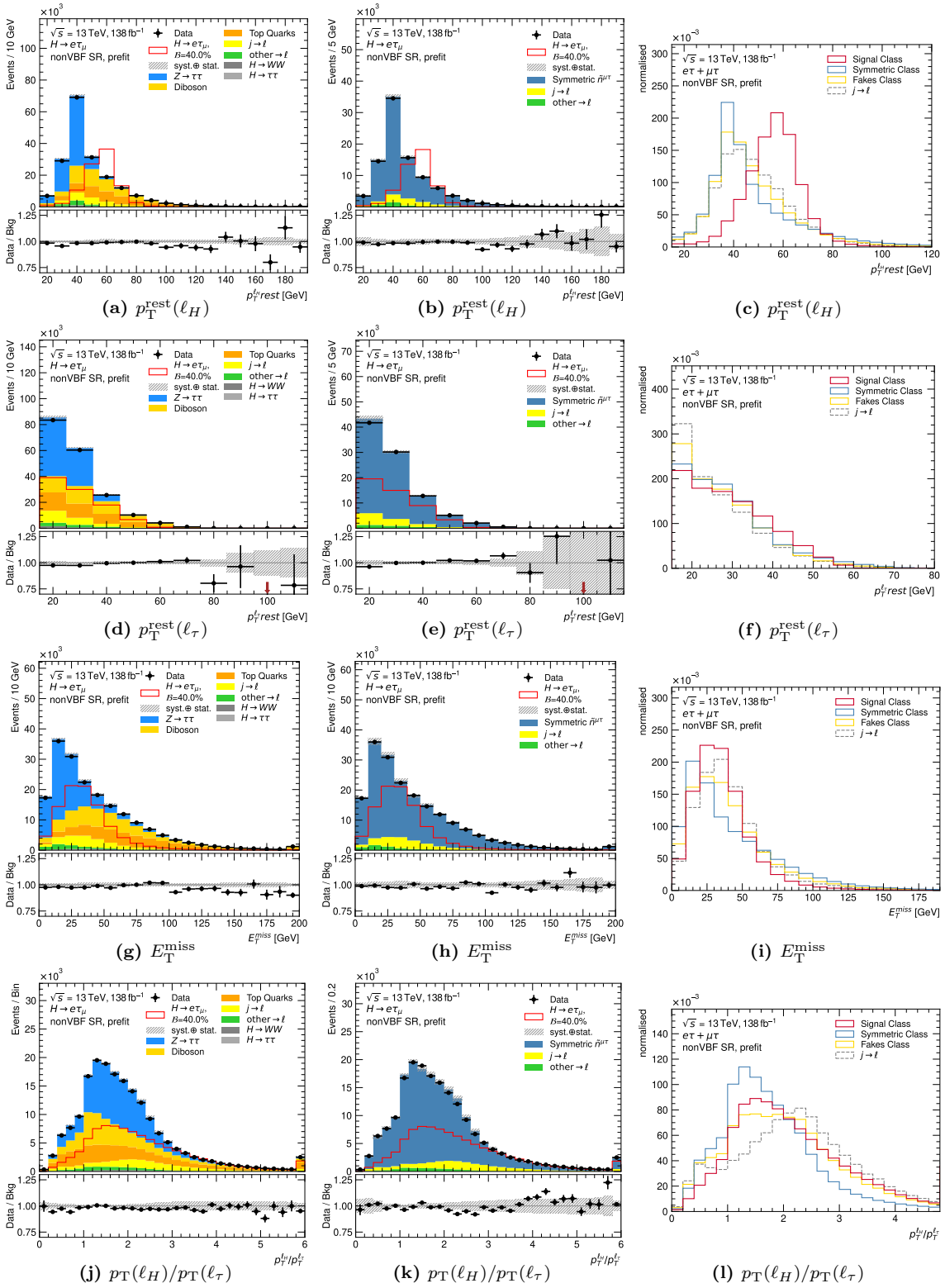
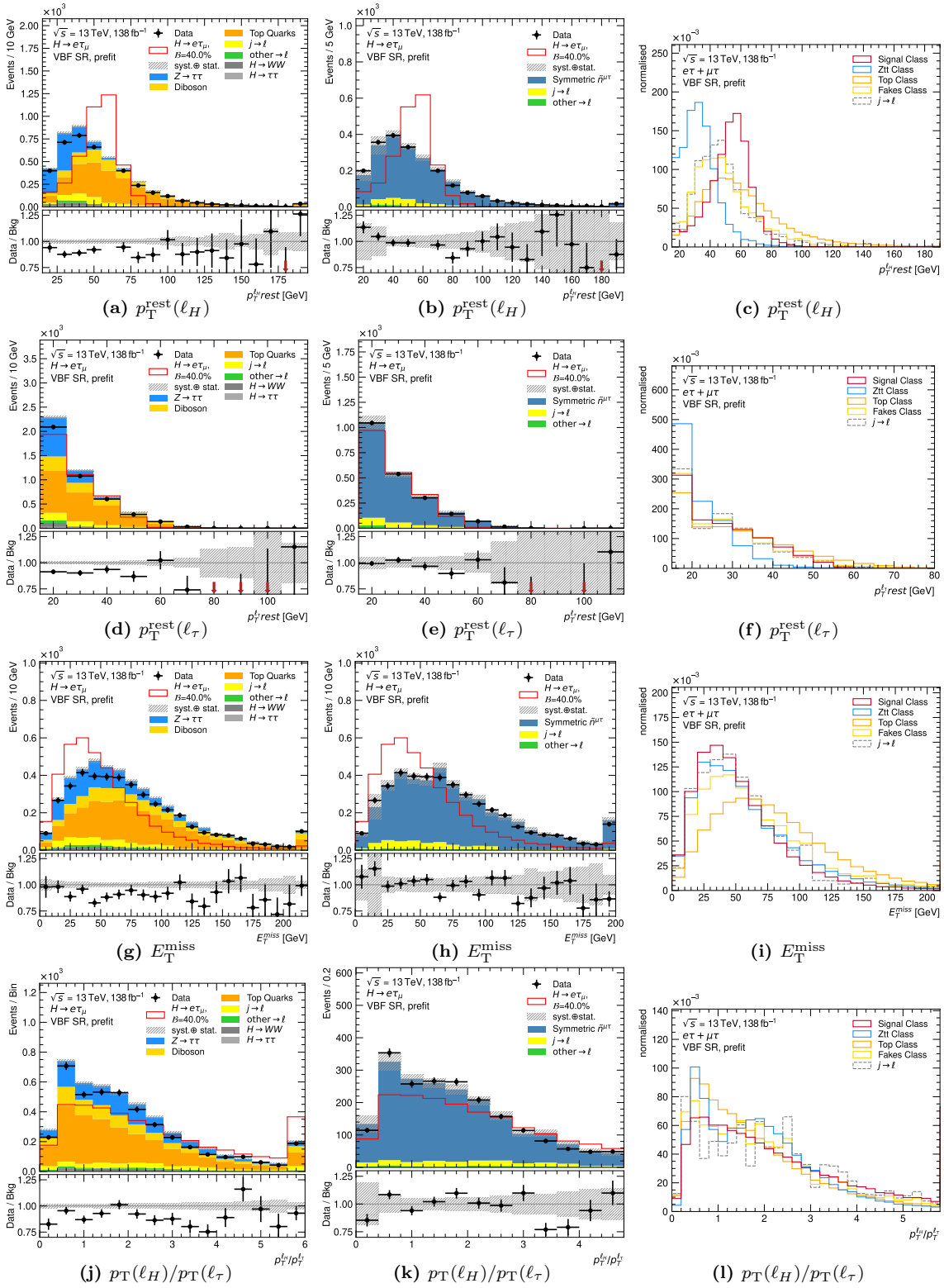



**Figure 8.12:** Distributions of input variables to the *nonVBF NN*. The left and middle columns compare the prediction to the data in the $e\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datasets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.




**(a)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_H)$ **(b)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_H)$ **(c)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_H)$

**(d)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_\tau)$ **(e)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_\tau)$ **(f)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_\tau)$

**(g)** $E_{\mathrm{T}}^{\mathrm{miss}}$ **(h)** $E_{\mathrm{T}}^{\mathrm{miss}}$ **(i)** $E_{\mathrm{T}}^{\mathrm{miss}}$

**(j)** $p_{\mathrm{T}}(\ell_H)/p_{\mathrm{T}}(\ell_\tau)$ **(k)** $p_{\mathrm{T}}(\ell_H)/p_{\mathrm{T}}(\ell_\tau)$ **(l)** $p_{\mathrm{T}}(\ell_H)/p_{\mathrm{T}}(\ell_\tau)$

**Figure 8.13:** Distributions of input variables to the *VBF NN*s. The left and middle columns compare the prediction to the data in the $e\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datasets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

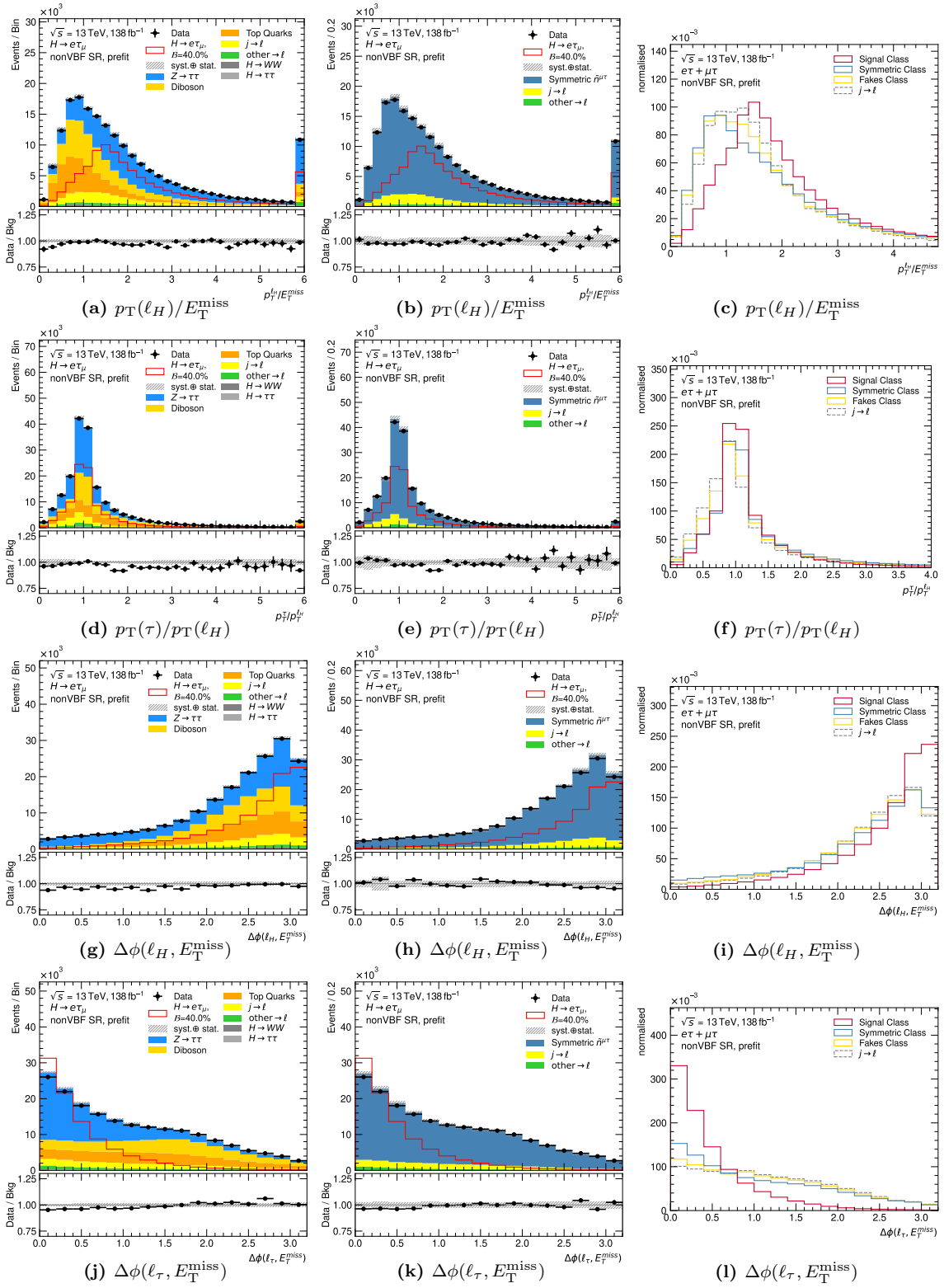



**Figure 8.14:** Distributions of input variables to the *nonVBF NN*. The left and middle columns compare the prediction to the data in the $e\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datasets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

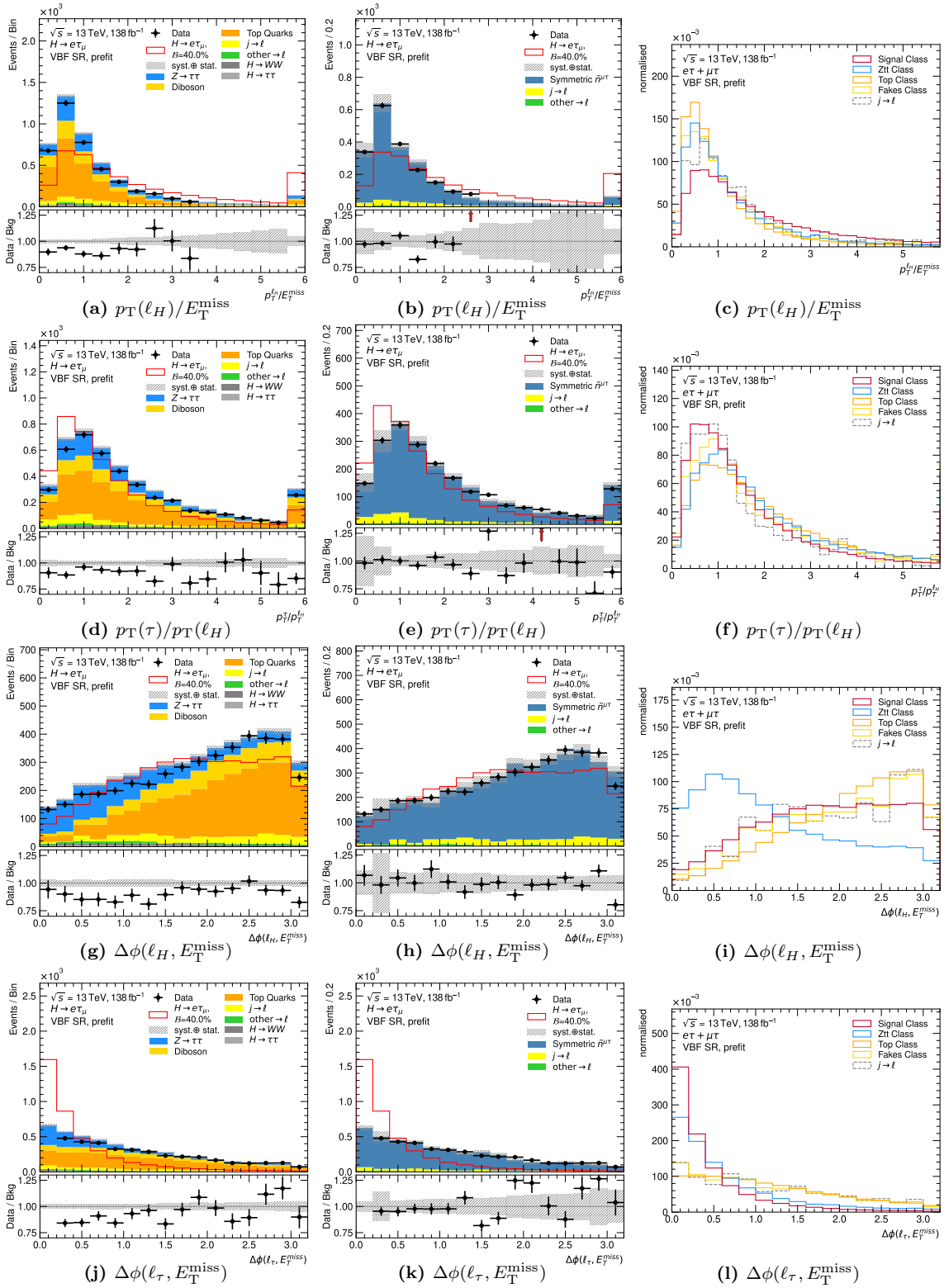
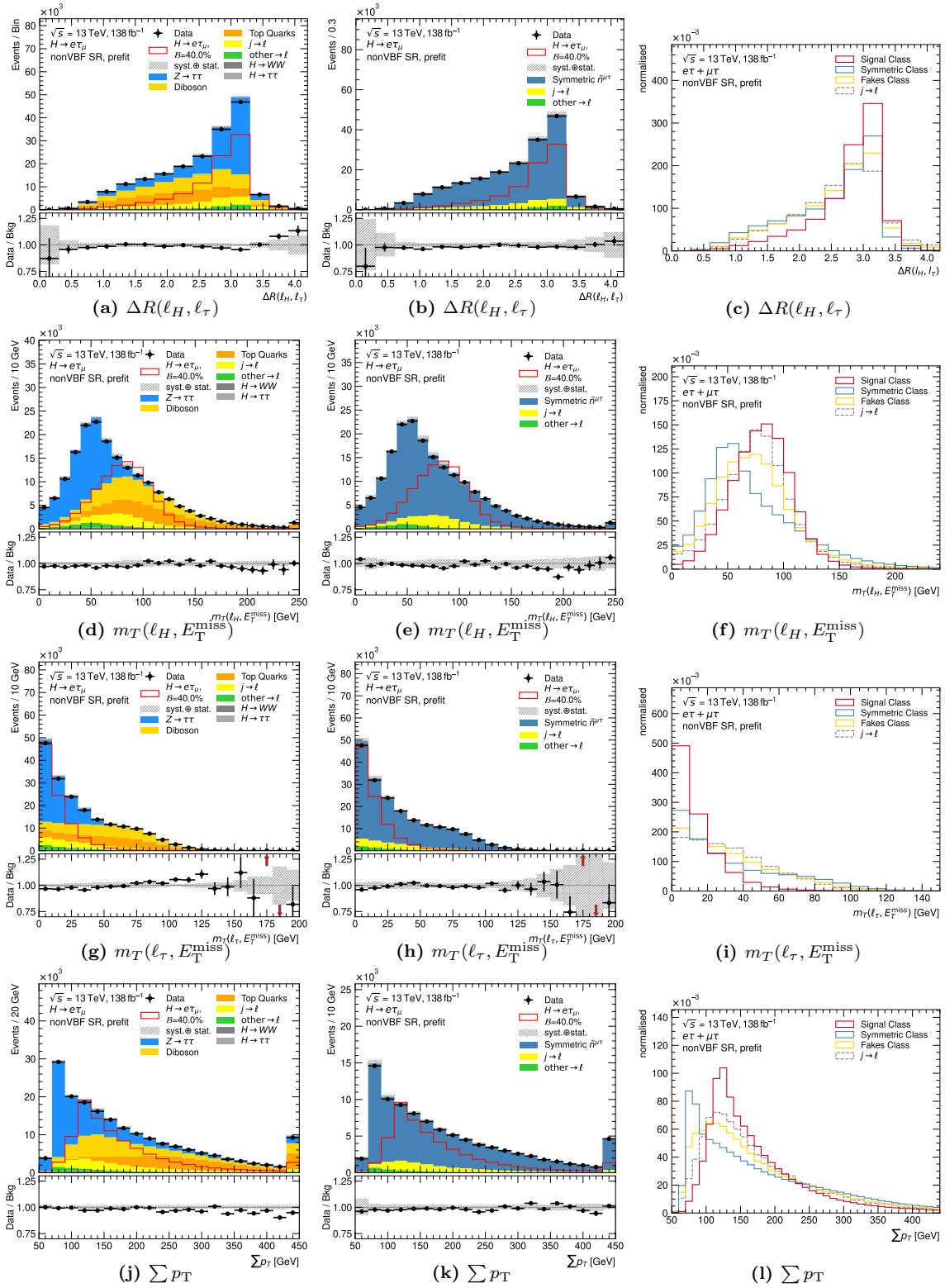



**Figure 8.15:** Distributions of input variables to the *VBF NN*s. The left and middle columns compare the prediction to the data in the $e\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datasets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.



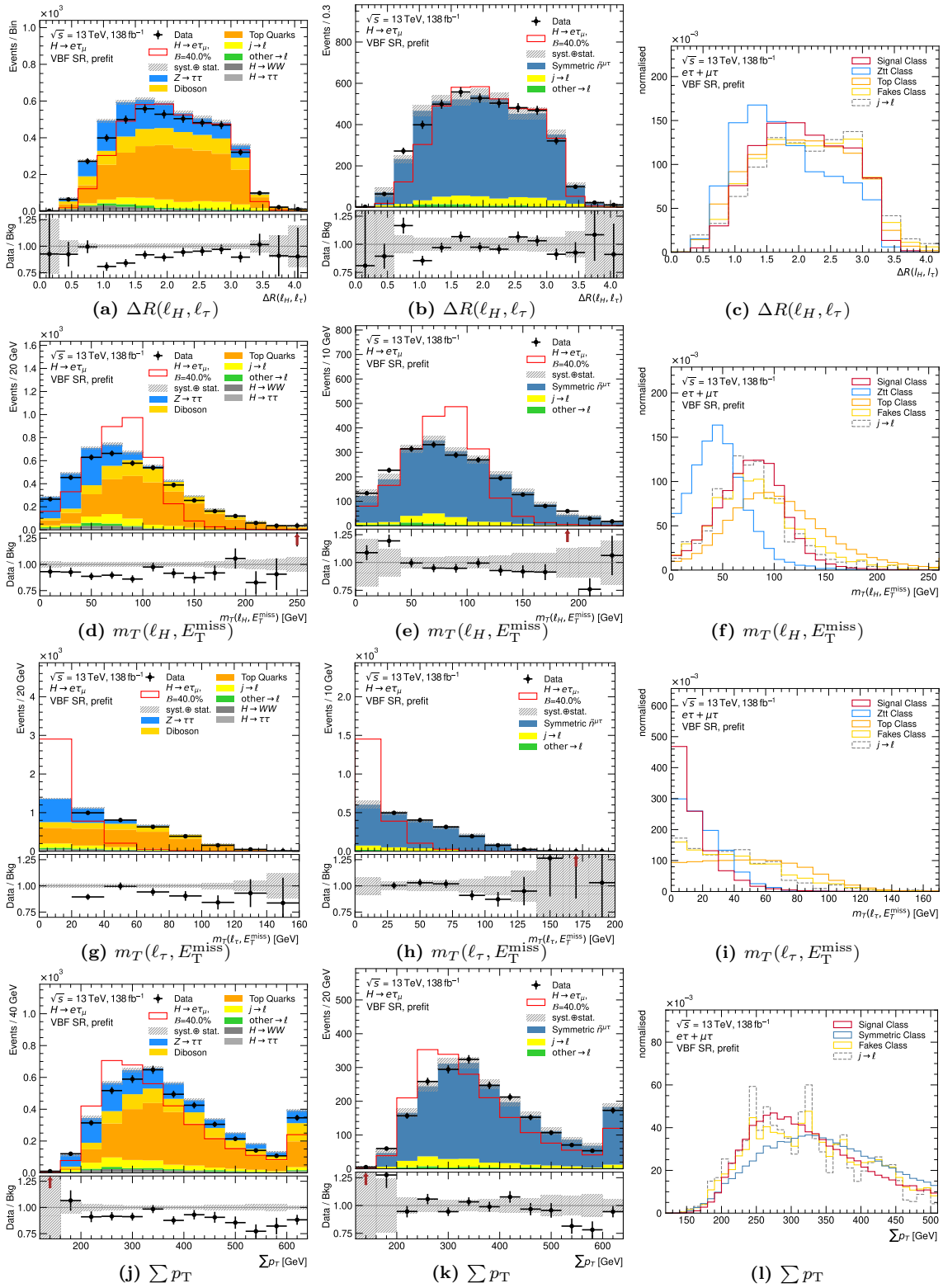

**(a)** $\Delta R(\ell_H, \ell_\tau)$ **(b)** $\Delta R(\ell_H, \ell_\tau)$ **(c)** $\Delta R(\ell_H, \ell_\tau)$

**(d)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$ **(e)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$ **(f)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$

**(g)** $m_T(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ **(h)** $m_T(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ **(i)** $m_T(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$

**(j)** $\sum p_{\mathrm{T}}$ **(k)** $\sum p_{\mathrm{T}}$ **(l)** $\sum p_{\mathrm{T}}$

**Figure 8.16:** Distributions of input variables to the *nonVBF NN*. The left and middle columns compare the prediction to the data in the $e\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datasets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.




(a) $\Delta R(\ell_H, \ell_\tau)$ (b) $\Delta R(\ell_H, \ell_\tau)$ (c) $\Delta R(\ell_H, \ell_\tau)$

(d) $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$ (e) $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$ (f) $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$

(g) $m_T(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ (h) $m_T(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ (i) $m_T(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$

(j) $\sum p_{\mathrm{T}}$ (k) $\sum p_{\mathrm{T}}$ (l) $\sum p_{\mathrm{T}}$

**Figure 8.17:** Distributions of input variables to the *VBF NN*s. The left and middle columns compare the prediction to the data in the $e\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datasets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

discussed above. Consequently, $m_T\left(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}}\right)$ (Figures 8.16(g)–8.16(i), Figures 8.17(g)–8.17(i)) peaks at small values for signal. Both SRs show a good modelling of the data within uncertainties. A small increasing slope is visible in the data over background ratio of $m_T\left(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}}\right)$ up to 50 GeV in the nonVBF SR.

**$\sum p_{\mathrm{T}}$:** The scalar sum $\sum p_{\mathrm{T}}$ (Figures 8.16(j)–8.16(l), Figures 8.17(j)–8.17(l)) of the transverse momenta of all high-$p_{\mathrm{T}}$ objects in the event is defined by:

$$\sum p_{\mathrm{T}} = p_{\mathrm{T}}^{\ell_H} + p_{\mathrm{T}}^{\ell_\tau} + E_{\mathrm{T}}^{\mathrm{miss}} + \sum_{i=0}^{n_{\mathrm{jets}}-1} p_{\mathrm{T}}^{j_i}. \tag{8.2}$$

All jets in the event with $p_{\mathrm{T}} > 20$ GeV are considered, although the information is only available for up to three jets (ordered by $p_{\mathrm{T}}$). Given the $p_{\mathrm{T}}$-requirements on the objects, the minimum value of $\sum p_{\mathrm{T}}$ is 50 GeV in the nonVBF SR and 120 GeV in the VBF SR. In the nonVBF SR the signal peaks at larger values than the overall background (at around 125 GeV) and in the VBF SR at lower values (at around 270 GeV). The prediction models the data well within its uncertainties in both SRs.

The mass of the Higgs boson cannot be directly reconstructed from its decay products as the longitudinal component of the neutrinos as well as their individual contributions are not known.

**$m_{\ell\ell}$:** The simplest mass observable is the invariant mass of the two leptons, $m_{\ell\ell}$ (Figures 8.18(a)–8.18(c), Figures 8.19(a)–8.19(c)), also denoted as $m_{\mathrm{vis}}$. It does not aim at reconstructing the full Higgs-boson mass as it neglects the neutrinos, but still provides a good discriminator to background processes. As expected, the signal-distribution as well as the $Z \to \tau\tau$-distribution peak at lower values compared to their respective resonance masses and show broadened distributions. Except for the range between 30 GeV and 65 GeV, where the prediction overestimates the data, the modelling of the data is good in both SRs.

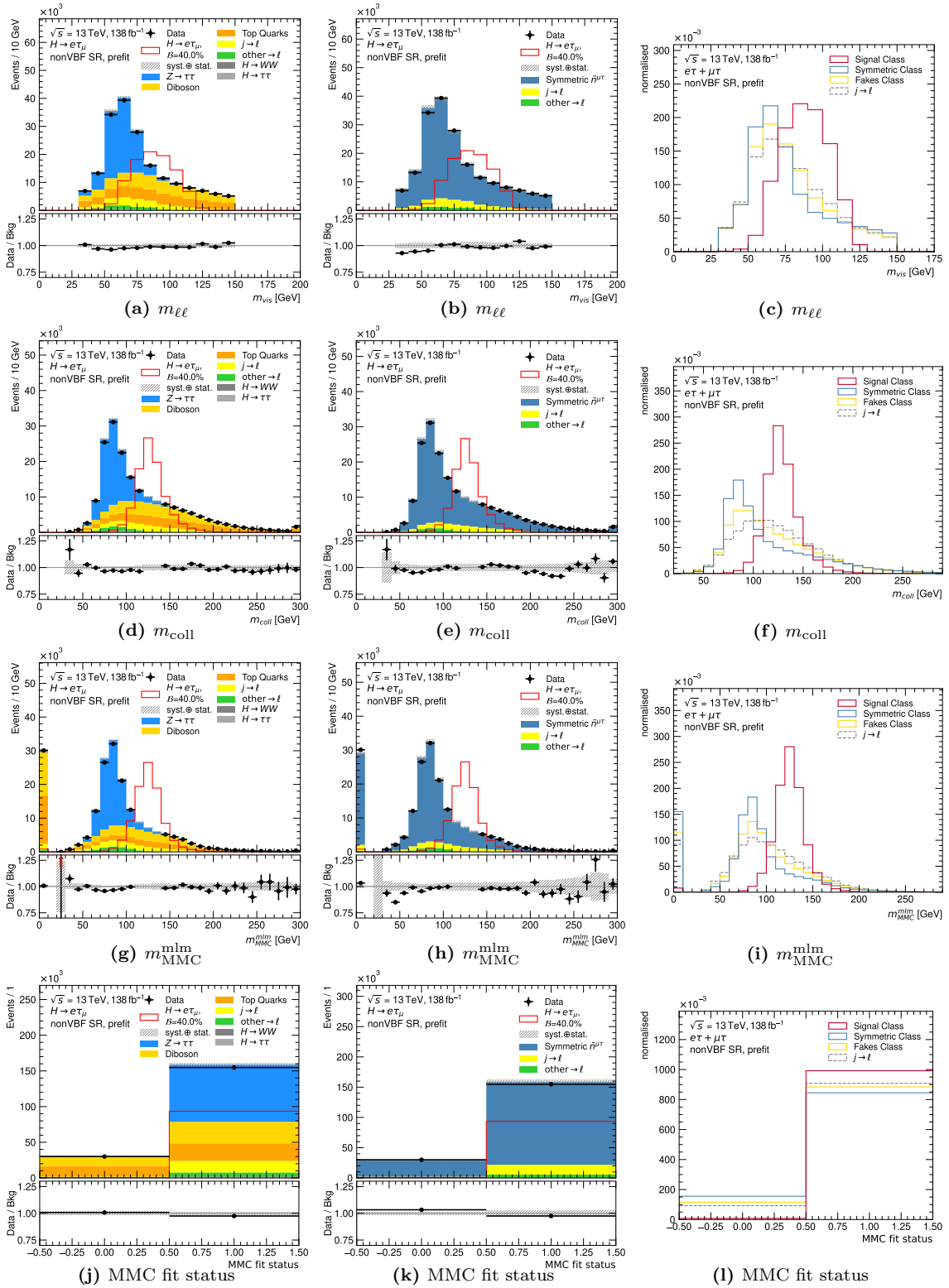
There are different methods that aim at providing an approximate estimate of the Higgs-boson mass taking into account the neutrinos in the final state. In the analysis here, the *collinear mass* $m_{\mathrm{coll}}$ [238] as well as the *missing mass calculator* (MMC) [239] which is the basis for $m_{\mathrm{MMC}}^{\mathrm{mlm}}$ are utilised. Both were developed for the reconstruction of the Higgs-boson mass in $H \to \tau\tau$-decays where the neutrinos originate from both subsequent $\tau$-lepton decays. In the $H$-LFV analysis, the calculation is adjusted for the presence of only one $\tau$-lepton decay. The modified versions are introduced in the following. Both methods, *collinear mass* and MMC, depend on the resolution of $E_{\mathrm{T}}^{\mathrm{miss}}$.

**$m_{\mathrm{coll}}$:** The *collinear mass* (Figures 8.18(d)–8.18(f), Figures 8.19(d)–8.19(f)) is based on the assumption that the decay products of the $\tau$-lepton travel in the same direction as the $\tau$-lepton (collinear) and that all missing transverse energy in the event originates from the neutrinos in the $\tau$-lepton decay. By applying these assumptions, the following equation is obtained:

$$m_{\mathrm{coll}} = \sqrt{2p_{\mathrm{T}}^{\ell_H}\left(p_{\mathrm{T}}^{\ell_\tau} + E_{\mathrm{T}}^{\mathrm{miss}}\right)\left(\cosh\left\{\Delta\eta\left(\ell_H, \ell_\tau\right)\right\} - \cos\left\{\Delta\phi\left(\ell_H, \ell_\tau\right)\right\}\right)}. \tag{8.3}$$

The modelling of data by the background prediction is reasonable within uncertainties, although the prediction slightly overestimates data in the ranges 50 to 90 GeV and 190 to 240 GeV in the nonVBF SR.

**$m_{\mathrm{MMC}}^{\mathrm{mlm}}$, MMC fit status:** The MMC performs a scan over the unknown elements of the di-neutrino four-momentum and the Higgs-boson mass is calculated at each scan point. Assuming that the the measured $E_{\mathrm{T}}^{\mathrm{miss}}$ in the events solely originates from the two neutrinos



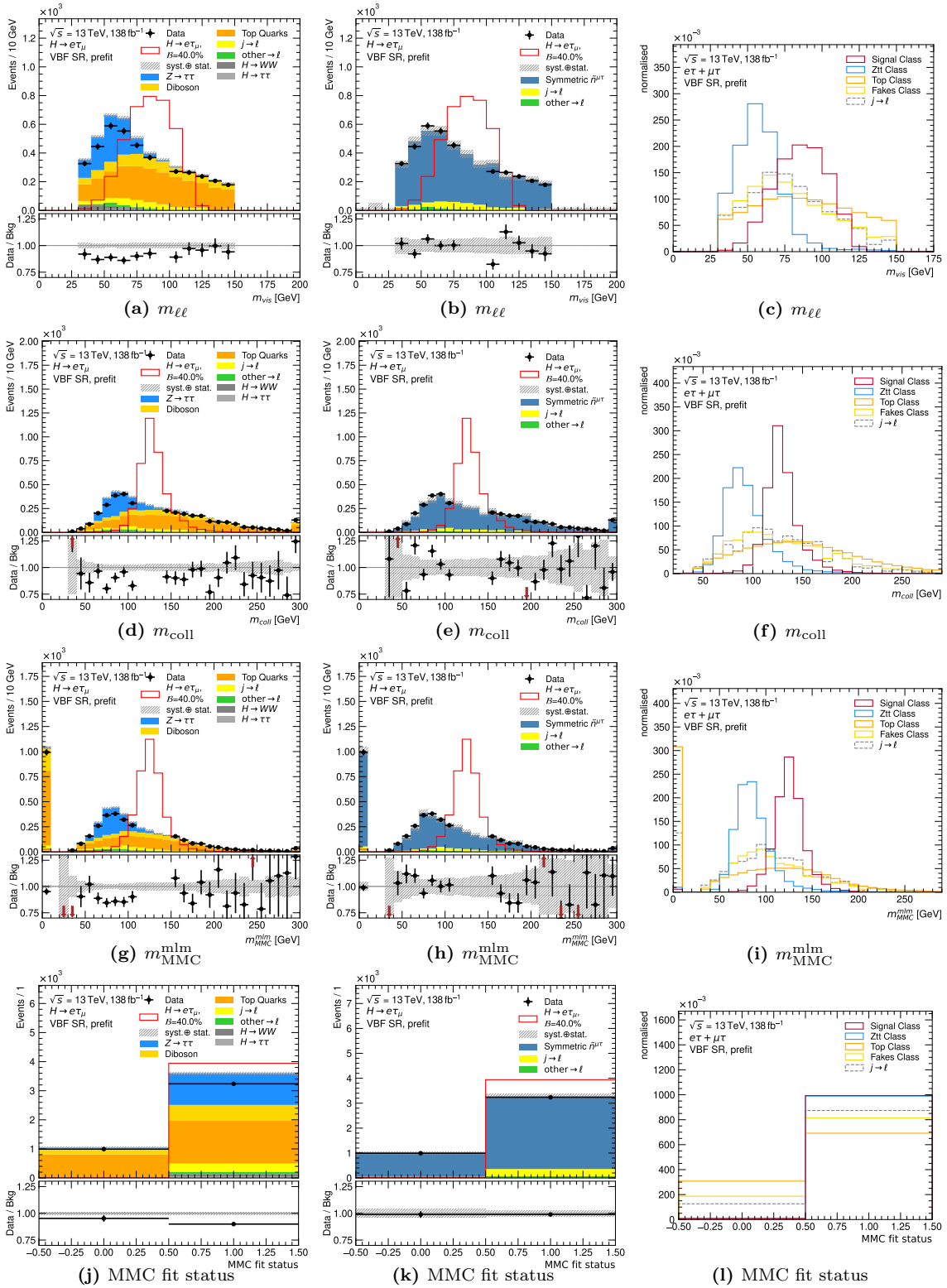

(a) $m_{\ell\ell}$ (b) $m_{\ell\ell}$ (c) $m_{\ell\ell}$

(d) $m_{\text{coll}}$ (e) $m_{\text{coll}}$ (f) $m_{\text{coll}}$

(g) $m_{\text{MMC}}^{\text{mlm}}$ (h) $m_{\text{MMC}}^{\text{mlm}}$ (i) $m_{\text{MMC}}^{\text{mlm}}$

(j) MMC fit status (k) MMC fit status (l) MMC fit status

**Figure 8.18:** Distributions of input variables to the *nonVBF NN*. The left and middle columns compare the prediction to the data in the $e\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datasets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.




**(a)** $m_{\ell\ell}$ **(b)** $m_{\ell\ell}$ **(c)** $m_{\ell\ell}$

**(d)** $m_{\text{coll}}$ **(e)** $m_{\text{coll}}$ **(f)** $m_{\text{coll}}$

**(g)** $m^{\text{mlm}}_{\text{MMC}}$ **(h)** $m^{\text{mlm}}_{\text{MMC}}$ **(i)** $m^{\text{mlm}}_{\text{MMC}}$

**(j)** MMC fit status **(k)** MMC fit status **(l)** MMC fit status

**Figure 8.19:** Distributions of input variables to the *VBF NN*s. The left and middle columns compare the prediction to the data in the $e\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datasets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

of the $\tau$-lepton decay results in two known elements of the di-neutrino four-momentum. The squared four-momentum of the system of the two neutrinos and the light lepton from the $\tau$-decay equals the invariant mass squared of the $\tau$-lepton (on-shell condition). This reduces the number of unknown elements of the di-neutrino four-momentum to only one. Therefore, only one variable needs to be considered in the scan; the invariant mass of the di-neutrino system is used. However, as the resolution of $E_{\mathrm{T}}^{\mathrm{miss}}$ impacts the reconstructed Higgs-boson mass, also a scan over its two components is performed, constrained by their resolution. Preassigned probability distributions of the scanned variable depending on the resolution of $E_{\mathrm{T}}^{\mathrm{miss}}$ and on the $\tau$-decay topology are used in order to weight the resulting mass of each scan point accordingly when filling it into a histogram. The histogram of the mass for the different scan points is obtained for each event and the value of the mass at its maximum is used as a proxy for the Higgs-boson mass, denoted as $m_{\mathrm{MMC}}^{\mathrm{mlm}}$ (Figures 8.18(g)–8.18(i), Figures 8.19(g)–8.19(i)). However, the MMC does not find a solution for every event. If no solution is found, a retry is performed where the other light lepton is assigned to the $\tau$-lepton. This improves the success rate, but there are still events where the MMC fails to find a solution. The outcome is encoded in the "MMC fit status" (Figures 8.18(j)–8.18(l), Figures 8.19(j)–8.19(l)). It is 1 if a solution is found and 0 if not and hence is a binary input variable. When no solution is found, $m_{\mathrm{MMC}}^{\mathrm{mlm}}$ is set to 0. For the LFV-signal and $Z \to \tau\tau$-background, the ratio of events for which no solution is found is below 1 % in the *Basic Selection Signal Region*. For top-quark production this ratio is 35.5 %, 28.3 % for diboson production and 7.4 % for $j \to \ell$-fakes. Hence, the "MMC fit status" itself provides separation power which, however, is also encoded in $m_{\mathrm{MMC}}^{\mathrm{mlm}}$ by setting it to 0 if no solution is found.

The modelling of data by the background prediction in the distribution of both observables is good within uncertainties, although the prediction slightly overestimates data in the range 40 to 50 GeV in the nonVBF SR.

**$\Delta\alpha$:** Another observable that targets the collinearity of the $\tau$-lepton's decay products is $\Delta\alpha$ (Figures 8.20(a)–8.20(c), Figures 8.21(a)–8.21(c)). It is originally used in the search for *lepton-flavour violating* decays of the $Z$-boson [240]. For $H$-LFV decays it is defined as:

$$\Delta\alpha = \frac{m_H^2 - m_\tau^2}{2p^{\ell_H} p^{\ell_\tau}} - \frac{p_{\mathrm{T}}^{\ell_H}}{p_{\mathrm{T}}^{\ell_\tau}} \,. \tag{8.4}$$

If the decay products of the $\tau$-lepton are collinear and if the transverse momentum of the Higgs boson can be neglected, it is expected to be 0 in the case of signal.

Following the first assumption, the four-vector $p^\mu$ of the $\tau$-lepton can be written as

$$p_\tau^\mu = p_{\ell_\tau}^\mu + p_\nu^\mu + p_{\bar{\nu}}^\mu \equiv \alpha_1 \; p_{\ell_\tau}^\mu \,. \tag{8.5}$$

In combination with

$$\begin{aligned} m_H^2 = \left(p_H^\mu\right)^2 = \left(p_\tau^\mu + p_{\ell_H}^\mu\right)^2 = (p_\tau^\mu)^2 + \left(p_{\ell_H}^\mu\right)^2 + 2 \; p_\tau^\mu \; p_{\ell_H,\mu} \approx m_\tau^2 + 0 + 2 \; p_\tau^\mu \; p_{\ell_H,\mu} \\ \implies m_H^2 - m_\tau^2 = 2 \; p_\tau^\mu \; p_{\ell_H}^\mu \end{aligned} \tag{8.6}$$

one obtains:

$$m_H^2 - m_\tau^2 = 2 \; \alpha_1 \; p_{\ell_\tau}^\mu \; p_{\ell_H}^\mu \implies \alpha_1 = \frac{m_H^2 - m_\tau^2}{2 \; p_{\ell_\tau}^\mu \; p_{\ell_H}^\mu} \,. \tag{8.7}$$

Following the second assumption ($p_{\mathrm{T}}^H = 0$) and taking into account Eq. 8.5 leads to:

$$\alpha_2 \; p_{\mathrm{T}}^{\ell_\tau} = p_{\mathrm{T}}^{\ell_H} \implies \alpha_2 = \frac{p_{\mathrm{T}}^{\ell_H}}{p_{\mathrm{T}}^{\ell_\tau}} \,. \tag{8.8}$$

And hence, $\Delta\alpha$ can be derived from Eqs. 8.7 and 8.8:

$$\Delta\alpha = \alpha_1 - \alpha_2 \,. \tag{8.9}$$

The distribution of $\Delta\alpha$ indeed peaks around 0 for signal. The bin from 0 to 2 is populated more than the bin from $-2$ to 0 but overall the distribution of the signal is negatively skewed which is the opposite for the symmetric background class (*nonVBF NN*) and the $Z \to \tau\tau$ class (*VBF NN*s). The modelling of the observable is good in the VBF SR and acceptable for the nonVBF SR where a decreasing slope in the data over background ratio is visible.

**$d0(\ell_H) - d0(\ell_\tau)$:** The impact parameters of the light lepton tracks with respect to the primary vertex are expected to be larger when the light lepton originates from a $\tau$-lepton decay due to the additional decay length of the $\tau$-lepton. Hence, the difference of the transverse impact parameters of $\ell_H$ and $\ell_\tau$ (with respect to the beam line), $d0(\ell_H) - d0(\ell_\tau)$, is expected to be slightly wider for signal events compared to background events. Also other impact parameter observables show a good separation power but typically suffer from a mis-modelling of data by simulated events or show an asymmetric behaviour between the $e\tau$- and $\mu\tau$-datasets. The modelling by simulated events and the *Symmetry Method* are sufficient in the case of $d0(\ell_H) - d0(\ell_\tau)$ in the nonVBF SR (Figures 8.20(d)–8.20(f)) but not in the VBF SR. For this reason why the variable is used in the *nonVBF NN* but omitted in the *VBF NN*s.

The remaining variables are only used for the training of the *VBF NN*s.

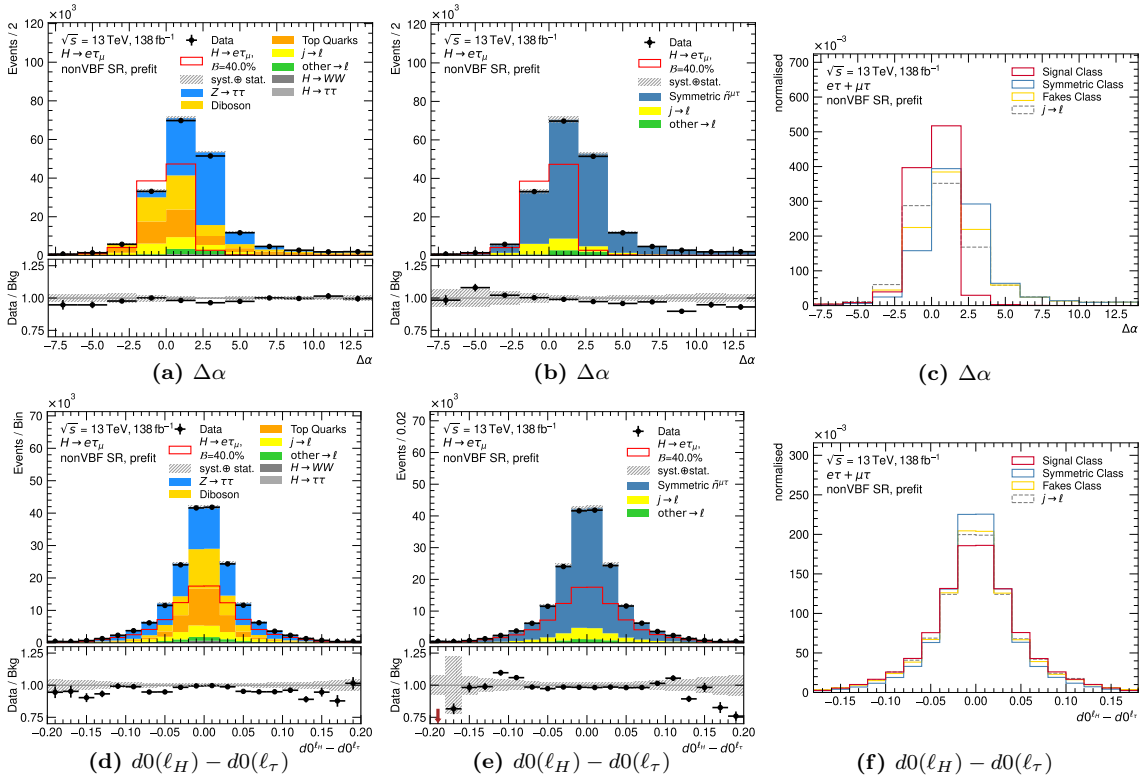
**$p_{\mathrm{T}}^{\mathrm{tot}}$:** The transverse momentum of the system of the two leptons, $E_{\mathrm{T}}^{\mathrm{miss}}$ and the two leading jets, denoted by $p_{\mathrm{T}}^{\mathrm{tot}}$ (Figures 8.21(d)–8.21(f)), is expected to be close to zero for (leading order) VBF signal events assuming the Higgs-boson and the two jets balance each other in the transverse plane. Hence, this observable can effectively act as a veto on a third jet. It shows a good modelling of the data within the prediction uncertainties.

**$p_{\mathrm{T}}^{j_0}$, $p_{\mathrm{T}}^{j_1}$:** The transverse momenta of the two leading jets, $p_{\mathrm{T}}^{j_0}$ (Figures 8.21(g)–8.21(i)) and $p_{\mathrm{T}}^{j_1}$ (Figures 8.21(j)–8.21(l)), do not show a distinct separation power by themselves but can still contribute through correlations. Apart from few fluctuations, the prediction of the background contributions describes data well in both distributions.

**$\Delta\phi(j_0, E_{\mathrm{T}}^{\mathrm{miss}})$, $\Delta\phi(j_1, E_{\mathrm{T}}^{\mathrm{miss}})$:** Similar observations concerning the separation power as for $p_{\mathrm{T}}^{j_0}$ and $p_{\mathrm{T}}^{j_1}$ apply to the differences in $\phi$ between $E_{\mathrm{T}}^{\mathrm{miss}}$ and the two leading jets, respectively (Figures 8.22(a)–8.22(c) and Figures 8.22(d)–8.22(f)). A difference in the shape of the distributions is only observed for $\Delta\phi(j_0, E_{\mathrm{T}}^{\mathrm{miss}})$ and the $Z \to \tau\tau$-background. Both distributions are modelled well.

**$m_{jj}$:** In the distribution of the invariant mass of the two leading jets, $m_{jj}$ (Figures 8.22(g)–8.22(i)), all processes decrease from left to right. However, the signal is flatter compared to the background processes and therefore is enhanced for higher values of $m_{jj}$. This is also highlighted by the fact that the bins at higher $m_{jj}$-values are blinded. The modelling in the non-blinded bins is very good.

**$\Delta\eta_{jj}$:** The separation in pseudorapidity of the two jets, denoted by $\Delta\eta_{jj}$ (Figures 8.22(j)–8.22(l)), has a clear difference in shape between signal and the background processes. It is shifted to larger values for the signal processes and is even further shifted for the VBF-signal only as was shown in Figure 8.1(c). This is again indicated by the fact that the bins at higher values are blinded. The modelling of the data by the background prediction in the remaining bins is very good.




(a) $\Delta\alpha$ (b) $\Delta\alpha$ (c) $\Delta\alpha$

(d) $d0(\ell_H) - d0(\ell_\tau)$ (e) $d0(\ell_H) - d0(\ell_\tau)$ (f) $d0(\ell_H) - d0(\ell_\tau)$

**Figure 8.20:** Distributions of input variables to the *nonVBF NN*. The left and middle columns compare the prediction to the data in the $e\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datasets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

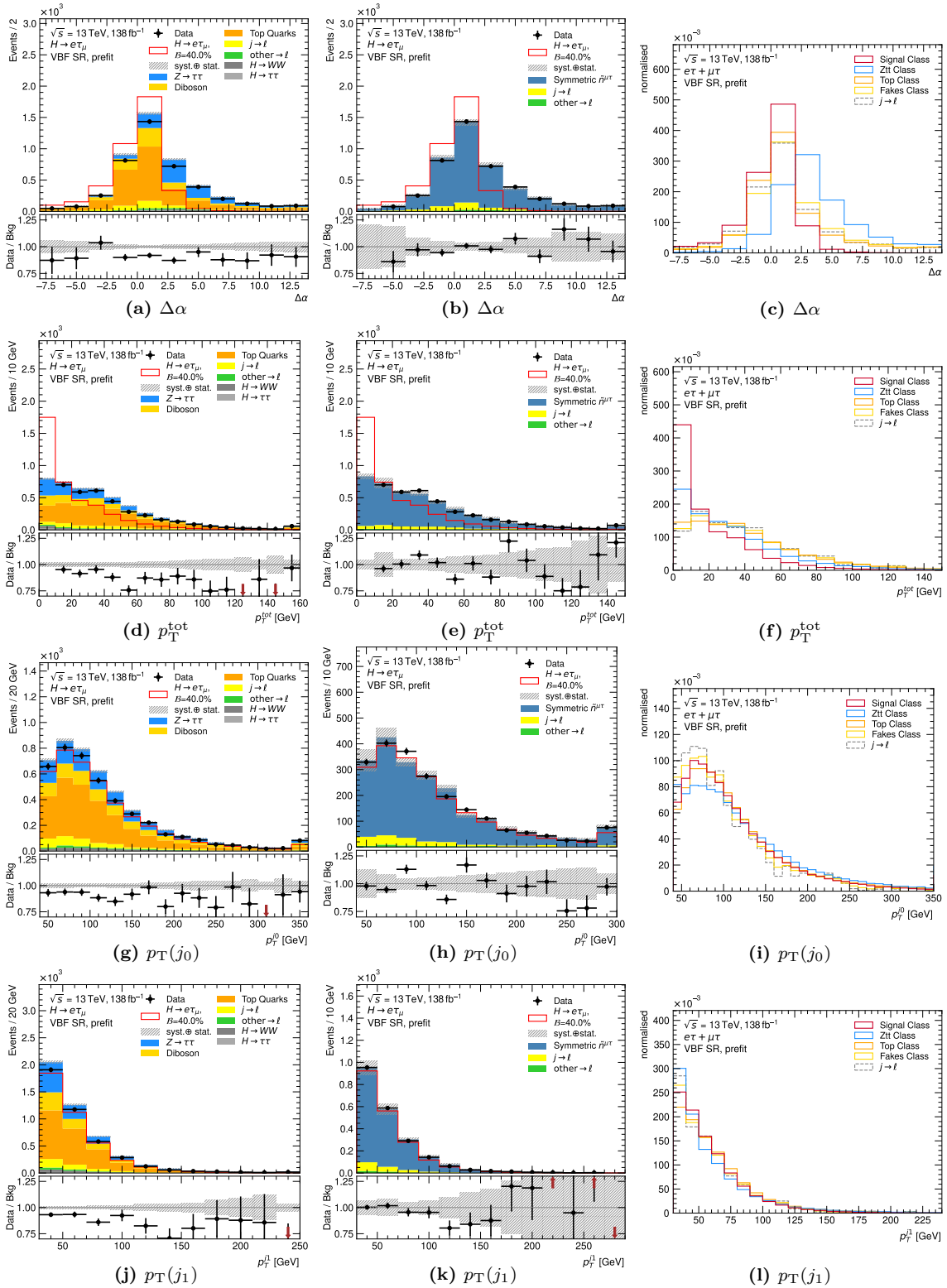
**$\Delta R_{jj}$:** Similar behaviour as for $\Delta\eta_{jj}$ can be observed for the angular separation of the two jets, denoted by $\Delta R_{jj}$ (Figures 8.23(a)–8.23(c)). The modelling of the data by the background prediction is very good.

**$\eta$-centr.($\ell_H$), $\eta$-centr.($\ell_\tau$):** The $\eta$-centrality of each of the leptons with respect to the $\eta$-values of the two leading jets is defined as:

$$\eta\text{-centr.}(\ell) = \exp\left(\frac{-4}{(\eta_{j_0} - \eta_{j_1})^2}\left(\eta_\ell - \frac{\eta_{j_0} + \eta_{j_1}}{2}\right)^2\right). \tag{8.10}$$

It is equal to 1 if the lepton lies exactly between the two jets, $1/e$ if the lepton is aligned with one of the jets and $< 1/e$ if it is outside the two jets. The VBF-signal is expected to accumulate close to 1 which is reflected in the distributions of $\eta$-centr.($\ell_H$) (Figures 8.23(d)–8.23(f)) and $\eta$-centr.($\ell_\tau$) (Figures 8.23(g)–8.23(i)), respectively. The background processes are shifted more to lower values of the $\eta$-centrality, especially $Z \to \tau\tau$. The data is described well by the background prediction in both distributions.

**Correlations of the Input Variables** The Pearson correlation coefficients in percent for all input variables, separate for the nonVBF SR and the VBF SR, are shown in Figures 8.24 and 8.25, respectively. They are calculated and shown separately for the signal and the



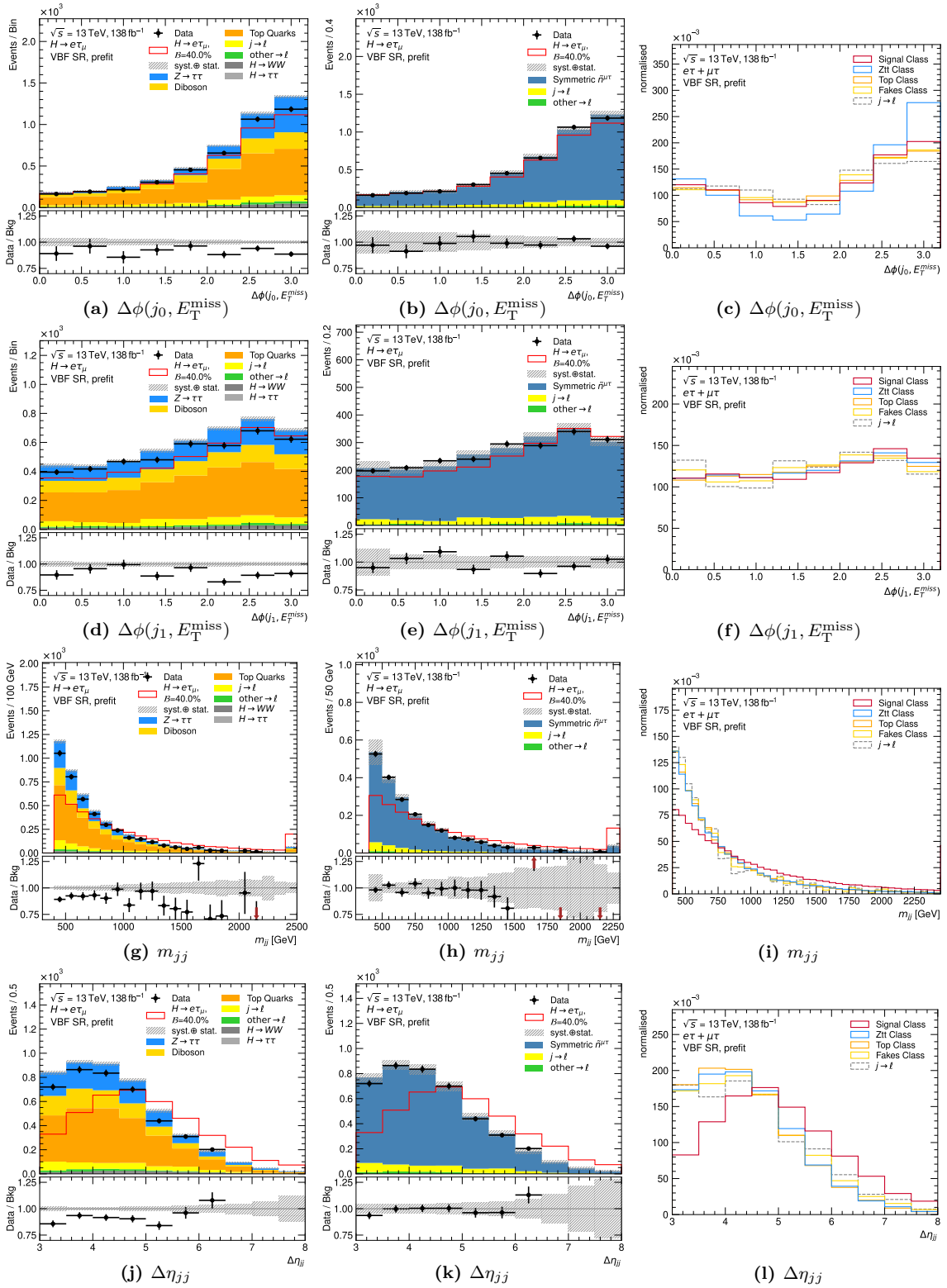

(a) $\Delta\alpha$ (b) $\Delta\alpha$ (c) $\Delta\alpha$

(d) $p_{\mathrm{T}}^{\mathrm{tot}}$ (e) $p_{\mathrm{T}}^{\mathrm{tot}}$ (f) $p_{\mathrm{T}}^{\mathrm{tot}}$

(g) $p_{\mathrm{T}}(j_0)$ (h) $p_{\mathrm{T}}(j_0)$ (i) $p_{\mathrm{T}}(j_0)$

(j) $p_{\mathrm{T}}(j_1)$ (k) $p_{\mathrm{T}}(j_1)$ (l) $p_{\mathrm{T}}(j_1)$

**Figure 8.21:** Distributions of input variables to the *VBF NN*s. The left and middle columns compare the prediction to the data in the $e\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datasets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.



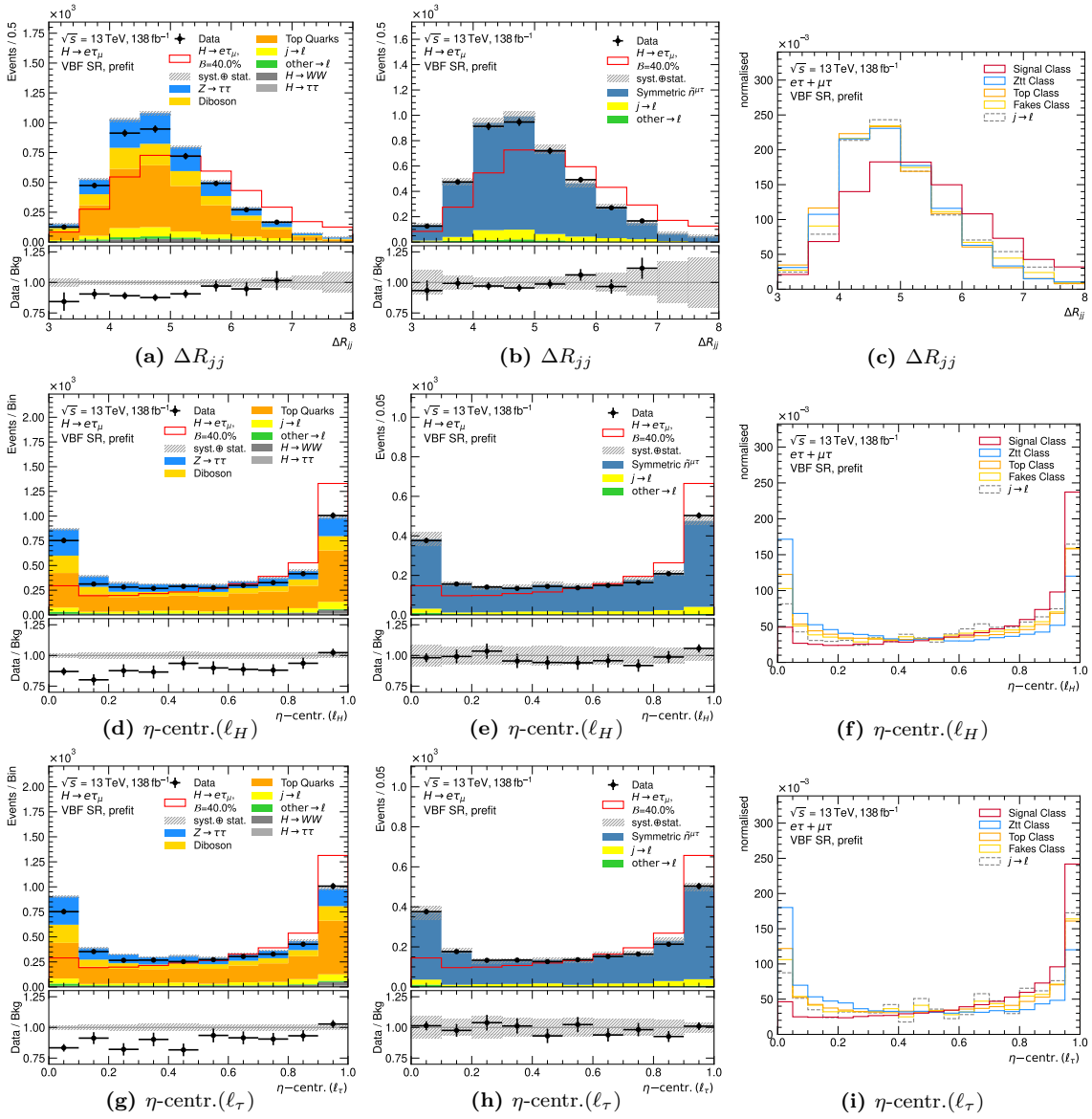

(a) $\Delta\phi(j_0, E_{\mathrm{T}}^{\mathrm{miss}})$ (b) $\Delta\phi(j_0, E_{\mathrm{T}}^{\mathrm{miss}})$ (c) $\Delta\phi(j_0, E_{\mathrm{T}}^{\mathrm{miss}})$

(d) $\Delta\phi(j_1, E_{\mathrm{T}}^{\mathrm{miss}})$ (e) $\Delta\phi(j_1, E_{\mathrm{T}}^{\mathrm{miss}})$ (f) $\Delta\phi(j_1, E_{\mathrm{T}}^{\mathrm{miss}})$

(g) $m_{jj}$ (h) $m_{jj}$ (i) $m_{jj}$

(j) $\Delta\eta_{jj}$ (k) $\Delta\eta_{jj}$ (l) $\Delta\eta_{jj}$

**Figure 8.22:** Distributions of input variables to the *VBF NN*s. The left and middle columns compare the prediction to the data in the $e\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datasets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.




**(a)** $\Delta R_{jj}$ **(b)** $\Delta R_{jj}$ **(c)** $\Delta R_{jj}$

**(d)** $\eta$-centr.$(\ell_H)$ **(e)** $\eta$-centr.$(\ell_H)$ **(f)** $\eta$-centr.$(\ell_H)$

**(g)** $\eta$-centr.$(\ell_\tau)$ **(h)** $\eta$-centr.$(\ell_\tau)$ **(i)** $\eta$-centr.$(\ell_\tau)$

**Figure 8.23:** Distributions of input variables to the *VBF NN*s. The left and middle columns compare the prediction to the data in the $e\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datasets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

background processes. Additionally, the differences of the correlation coefficients of signal and background are given which can be exploited by the NNs.

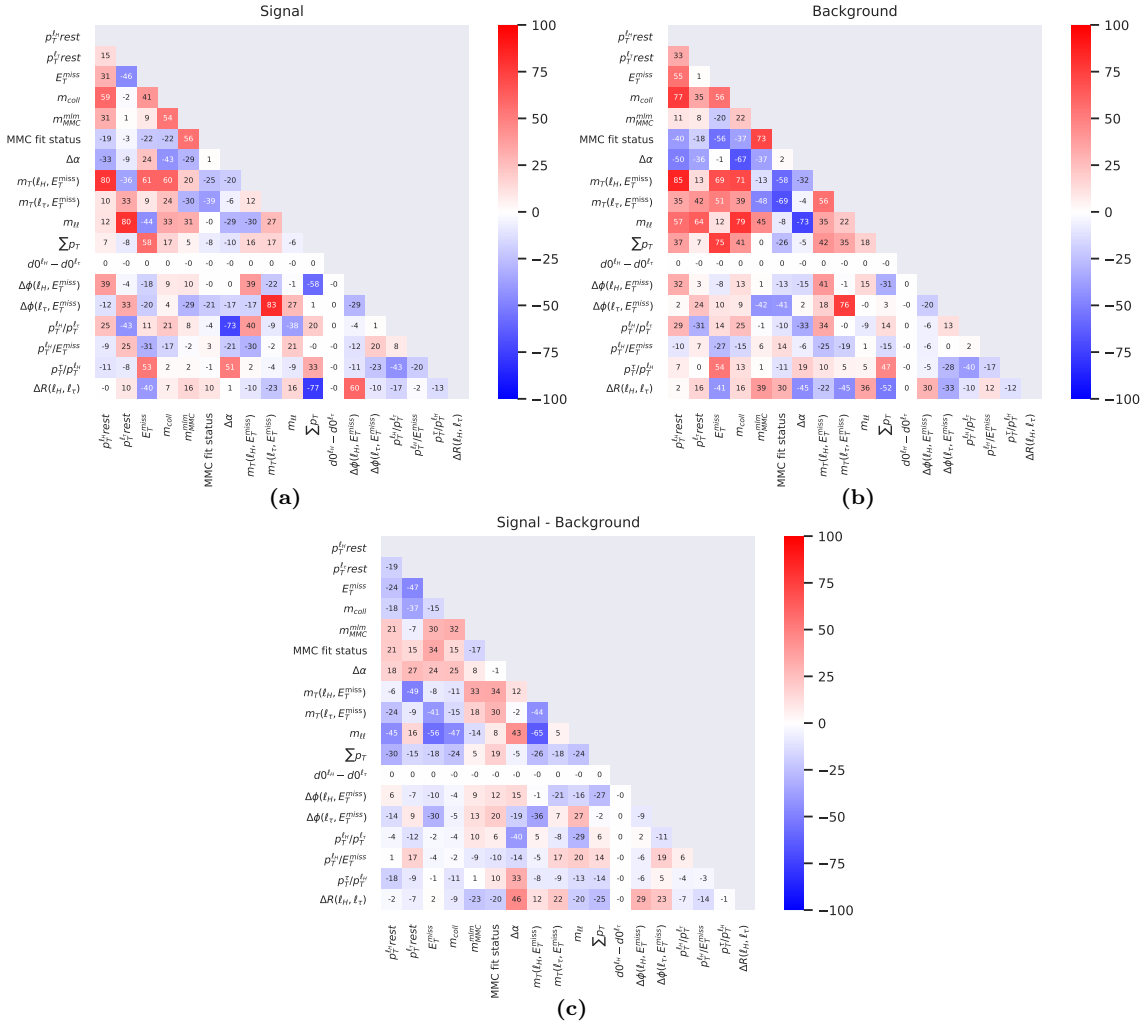



**Figure 8.24:** Matrices of Pearson correlation coefficients in percent of all input variables of the *nonVBF NN* for signal (a) and the sum of all backgrounds (b). The difference of the signal and background correlations is given in (c).

The strongest correlation in the nonVBF SR for signal is with 83 % (76 % for background) between $\Delta\phi(\ell_\tau, E_\mathrm{T}^\mathrm{miss})$ and $m_T(\ell_\tau, E_\mathrm{T}^\mathrm{miss})$. For background, the largest correlation is found with 85 % (80 % for signal) between $m_T(\ell_H, E_\mathrm{T}^\mathrm{miss})$ and $p_\mathrm{T}^{\ell_H}$ rest. These high correlations are reasonable when recalling the definition of $m_T(\ell, E_\mathrm{T}^\mathrm{miss})$ (Eq. 8.1). Other strong (anti-)correlations for signal are observed between $m_{\ell\ell}$ and $p_\mathrm{T}^{\ell_\tau}$ rest (80 %), between $\Delta R(\ell_H, \ell_\tau)$ and $\sum p_\mathrm{T}$ (−77 %) and between $p_\mathrm{T}^{\ell_H}/p_\mathrm{T}^{\ell_\tau}$ and $\Delta\alpha$ (51 %). The latter correlation is attributed to Eq. 8.8 which contains the ratio of the transverse momenta of the two leptons as part of the definition of $\Delta\alpha$. For background, further strong (anti-)correlations are obtained between $m_{\ell\ell}$ and $m_\mathrm{coll}$(79 %), between $m_\mathrm{coll}$ and $p_\mathrm{T}^{\ell_H}$ rest (77 %), between $\sum p_\mathrm{T}$ and $E_\mathrm{T}^\mathrm{miss}$ (75 %) and between $m_{\ell\ell}$ and $\Delta\alpha$ (−73 %). The input variable $d0^{\ell_H} - d0^{\ell_\tau}$ shows almost no correlations with respect to any other input variable. This variable is nearly symmetric around zero in contrast to any other variable. If its absolute value was used, non-zero correlations would appear. The largest difference in correlations of signal and background are observed between $m_{\ell\ell}$ and $m_T(\ell_H, E_\mathrm{T}^\mathrm{miss})$ (−65 %), between $m_{\ell\ell}$ and $E_\mathrm{T}^\mathrm{miss}$(−56 %), between $m_T(\ell_H, E_\mathrm{T}^\mathrm{miss})$ and $p_\mathrm{T}^{\ell_\tau}$ rest (−49 %) and between $\Delta R(\ell_H, \ell_\tau)$ and $\Delta\alpha$ (46 %).

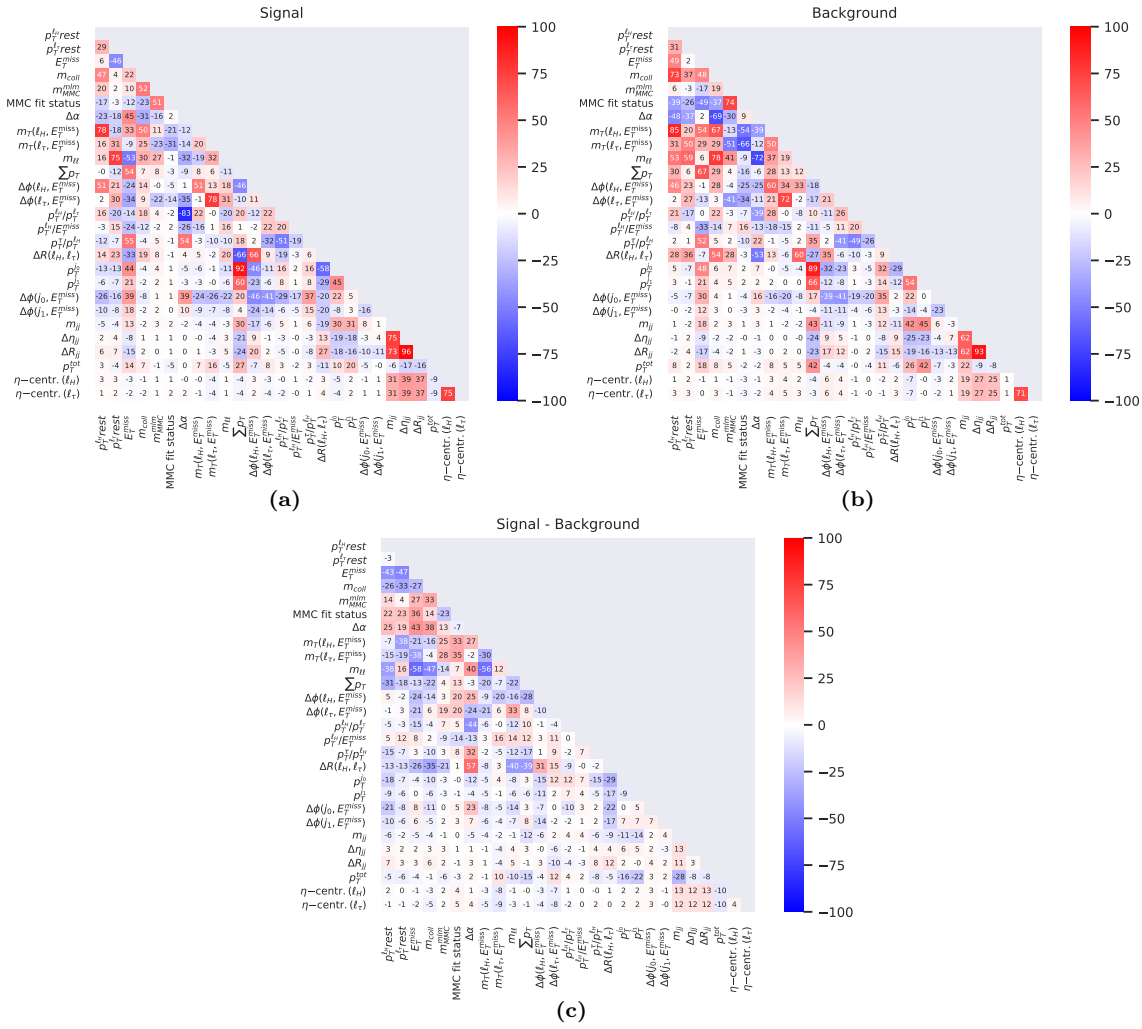



**Figure 8.25:** Matrices of Pearson correlation coefficients in percent of all input variables of the *VBF NN* for signal (a) and all backgrounds (b). The difference of the signal and background correlations is given in (c).

In the VBF SR, the strongest correlation for signal and background is observed with 96 % and 93 % between $\Delta R_{jj}$ and $\Delta\eta_{jj}$ as expected. Other strong (anti-)correlations that are also observed for the nonVBF SR are between $m_T(\ell_H, E_T^{\text{miss}})$ and $p_T^{\ell_H}$rest (78 %/85 %), between $m_{\ell\ell}$ and $p_T^{\ell_\tau}$rest (75 %/59 %) and between $p_T^{\ell_H}/p_T^{\ell_\tau}$ and $\Delta\alpha$ (54 %/−39 %) for signal/background. Strong correlations including input variables that are specific for the VBF SR are between $p_T^{j_0}$ and $\sum p_T$ (92 %/89 %), between the $\eta$-centralities of the two leptons (75 %/71 %), respectively, between $\Delta\eta_{jj}$ and $m_{jj}$ (75 %/62 %) and consequently between $\Delta R_{jj}$ and $m_{jj}$ (96 %/93 %). The largest difference in correlations of signal and background are between $m_{\ell\ell}$ and $E_T^{\text{miss}}$(−58 %), between $\Delta R(\ell_H, \ell_\tau)$ and $\Delta\alpha$ (57 %) and between $m_{\ell\ell}$ and $m_T(\ell_H, E_T^{\text{miss}})$ (−56 %). All three input variable pairs also show larger differences in correlations in the nonVBF SR.

### Preprocessing of the Input Variables

The distributions of the input variables are *standardised* as motivated in Section 6.3.1. That means that the input variable values of the single events are shifted by the respective distribution's mean and divided by its standard deviation such that each input variable distribution is centered around 0 with a standard deviation of 1.

### 8.2.4 Hyperparameters

The choice of the values for the *hyperparameters* of the NN-architecture and of the training process is crucial for obtaining a well-performing NN. Some of these hyperparameters are set to fixed values a priori as a result of experience gained from exploratory studies. They are summarised in Table 8.5. The remaining *hyperparameters* listed in Table 8.6 are subject to the optimisation procedure discussed in the following.

| Hyperparameter | Choice |
|---|---|
| architecture | fully connected NN |
| $Cl$ (# output nodes/classes) | 3 for *nonVBF NN*, 2 for *VBF NN*s |
| arrangement of nodes | # nodes in each layer is half of # of nodes in previous layer (triangle shape): $\mathcal{N}^{l_h+1} = 1/2\mathcal{N}^{l_h}$ |
| weight and bias initialisation | Glorot uniform for the weights and zeros for the biases |
| loss function | categorical cross entropy |
| activation function | softmax for last layer, leaky ReLU for inner layers |
| $N_{\text{ep}}$ (# epochs) | 100 |

**Table 8.5:** Hyperparameters fixed a priori for the networks used in this analysis.

Fully connected NNs with three output nodes in the case of the *nonVBF NN* and with two output nodes in the cases of the *VBF NN*s are trained. The nodes of the inner layers are arranged in a triangle shape [241]. The number of nodes in each hidden layer is half of the number of nodes in the previous hidden layer. In this way only the number of nodes of the first hidden layer needs to be optimised and the number of nodes in the others follows automatically. Another option, for example, would have been to use the same number of nodes in each hidden layer. But by choosing the triangle shape over this rectangular shape, the number of free parameters (weights and biases) can be reduced. And it is assumed that a reduced degree of freedom in later layers is sufficient and no information is lost; and if so, this is compensated in the optimisation process either by more layers or by more nodes overall.

The weights of each layer are initialised following the Glorot uniform prescription [242]. A random number from a uniform distribution in the range $[-a_l, a_l]$ is drawn for each weight. The parameter $a_l$ is defined for each layer $l$ as $a_l = \sqrt{6/(\mathcal{N}^{l-1} + \mathcal{N}^l)}$ with $\mathcal{N}^l$ and $\mathcal{N}^{l-1}$ being the number of nodes in layer $l$ and the previous layer $l-1$, respectively. The biases are initialised with zeros.

The *categorical cross entropy* loss or cost function as well as the chosen activation functions were discussed in detail in Sections 6.1.3 and 6.1.2, respectively.

The number of training epochs listed is fixed to 100 for the optimisation of the free *hyperparameters* but can be adapted for the final re-training.

#### Optimisation

The optimisation of the remaining *hyperparameters* is performed with *Optuna* [218] which was introduced in Section 6.4.1. For the optimisation, sets of allowed values for each *hyperparameter* to be optimised are defined and passed to the framework. A list of these *hyperparameters* and their respective set of values are given in Table 8.6. All 7 *hyperparameters* which are optimised were introduced in Chapter 6.

The number of hidden layers must be between 2 and 6. The number of nodes of the first hidden layer can be 64, 128, 256 or 512. The allowed values for the $\lambda$-parameter steering the strength of the *L2 weight regularisation* is continuous between $10^{-6}$ and $10^{-3}$. The values are drawn from a log-uniform distribution. This ensures that the sampling probability is

| Hyperparameter | Possible Choices |
|---|---|
| $L_h$ (# hidden layers) | $\in \{2, 3, 4, 5, 6\}$ |
| $\mathcal{N}^{0_h}$ (# nodes first hidden layer) | $\in \{64, 128, 256, 512\}$ |
| $\lambda$ (of *L2 weight regul.*) | log-uniform $\in (10^{-6}, 10^{-3})$ |
| $\alpha$ (leaky ReLU slope below 0) | uniform $\in (0.01, 0.1)$ |
| optimiser | Adam or SGD |
| $\eta$ (learning rate) | log-uniform $\in (10^{-5}, 10^{-1})$ |
| $s_b$ (batch size) | $\in \{128, 512, 1024, 2048, 4096, 8192\}$ |

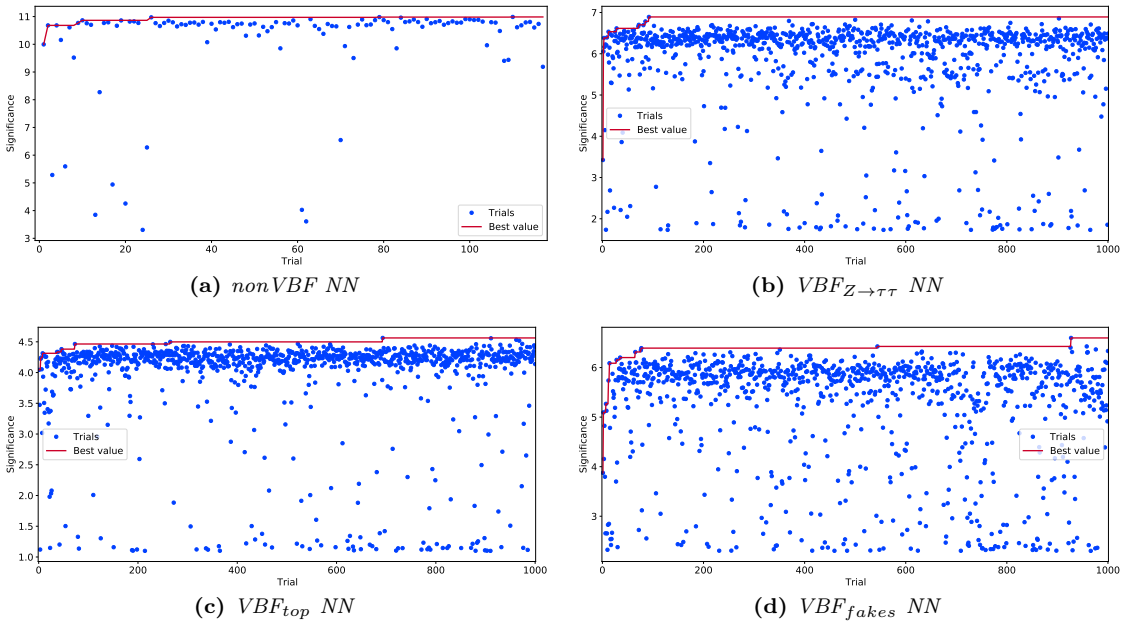
**Table 8.6:** Hyperparameters to be optimised in the *Optuna* optimisation.

the same for each order of magnitude. The slope parameter $\alpha$ of the leaky ReLU activation function is allowed to have values between 0.01 and 0.1 following a uniform distribution. The optimiser can either be Adam or SGD with learning rates $\eta$ between $10^{-5}$ and $10^{-1}$ following a log-uniform distribution. An individual learning rate parameter is implemented for each of the two optimisers. This ensures that the best performing outcome of one does not impact the choice for the other, as Adam typically needs smaller learning rates than SGD. Finally, the batch size is allowed to adopt either of the values in 128, 512, 1024, 2048, 4096 and 8192.

If the optimisation process showed that the most favoured values of one of the *hyperparameters* were at the edge of the allowed range, the optimisation procedure was performed again with a larger range for the respective *hyperparameter*. The individual ranges were chosen experimentally to be exhaustive enough to ensure that the search space includes a solution with acceptable performance. At the same time the ranges were kept as narrow as possible to avoid an overly expensive *hyperparameter* optimisation.

In order to judge the performance of the different trials (sets of hyperparameters), an objective function that returns an objective value is defined. Different metrics for the performance of NNs were introduced in Section 6.1.5. The *Binned Asimov Significance* $Z$ (Eq. 6.27) is used as objective value for the hyperparameter optimisation as it is fast to calculate and closest to the full statistical analysis. The *Binned Asimov Significance* is calculated for each trial using 10 equidistant bins of the signal node probability distribution. A branching ratio of 1 % is assumed for the signal. The objective function only considers the first two NNs of the 10-fold cross validation discussed in Section 6.2 in order to limit the required resources and time. That means the events of two *validation folds* (corresponding to the two NNs) are considered in the evaluation of $Z$. Each of the *validation folds* corresponds to 10 % of the available dataset, hence the objective value ($Z$) for each trial is evaluated based on 20 % of the full dataset. This is expected to provide a sufficiently precise estimate of the true objective value in order to compare the different trials. The value of $Z$ also depends on the number of events considered and increases when evaluated with the full set of validation folds, which is done for the final NNs after the optimisation is finished.

The histories of the four optimisations with *Optuna* are shown in Figure 8.26 for the *nonVBF NN* and the three *VBF NN*s. It is apparent that many more trials for the *VBF NN*s are performed than for the *nonVBF NN*. The reason is that each of the four optimisations was run for the same amount of time (2 days) and the *VBF NN*s are much faster to train as the number of processed events is lower. Hence, the number of parameters (weights and biases) to be trained is typically also lower for well performing trials compared to the *nonVBF NN*, which further reduces the average training time per trial. Overall, the majority of the trials exhibits a good performance. This is a result of the exploitation versus exploration tradeoff performed by the acquisition function in *Bayesian Optimisation* of which *Optuna* makes use of (Section 6.4.1).




**(a)** *nonVBF NN* **(b)** $VBF_{Z\to\tau\tau}$ *NN*

**(c)** $VBF_{top}$ *NN* **(d)** $VBF_{fakes}$ *NN*

**Figure 8.26:** History of the *Optuna* optimisations for the nonVBF SR (a) and for the VBF SR (b-d). The binned Asimov significance $Z$ (objective value) is shown as a function of the number of trials. The red line indicates the current best objective value.

An earlier version of the analysis was used to perform the optimisation for the *nonVBF NN*. The differences to the final version are as follows: the $j \to \ell$-fake estimate included the contribution of electron-fakes from photon conversion and used a binning of the electron *Fake Factor*s in $p_{\mathrm{T}}$ and $|\eta|$ instead of the final binning in $p_{\mathrm{T}}$ and $\Delta\phi(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}})$. Furthermore, the set of input features was slightly different at that time. A new optimisation was performed with the final $j \to \ell$-fake estimate and choice of input variables; but the best performing trial of the new optimisation obtained only a lower sensitivity in the final statistical analysis and hence it was decided to use the trial of the previous optimisation whose history is shown in Figure 8.26(a).

The objective values of the best trials (ranked by the objective value) of each optimisation are close together; thus, a more precise comparison of the two best performing trials is performed. Both trials are checked to exhibit reasonable loss curves, i.e. the loss evaluated on the *validation set* is meant to decrease with an increasing number of epochs to indicate that no *overtraining* is found. If overtraining is visible, the next best performing trial with a reasonable loss curve is chosen. The 10 networks of the 10-fold cross validation are trained for the two best performing trials with the respective hyperparameter sets and are evaluated on the full *validation set*. In addition, the full statistical analysis using the *pre-fit Asimov* dataset (see Section 10.3) is performed and the final trial is chosen based on the outcome. The latter is discussed in detail in Chapter 10. For the *VBF NN*s this procedure is done separately for each of the three NNs ($VBF_{Z\to\tau\tau}$ *NN*, $VBF_{top}$ *NN* and $VBF_{fakes}$ *NN*).

The finally chosen hyperparameters are listed in Table 8.7 for the four networks (*nonVBF NN*, $VBF_{Z\to\tau\tau}$ *NN*, $VBF_{top}$ *NN* and $VBF_{fakes}$ *NN*). The *nonVBF NN* is rather shallow with only two inner layers but has more nodes per layer compared to the *VBF NN*s. Overall, it has 141 827 parameters (weights and biases) to be trained compared to only 14 482, 13 986 and 14 482 for the three *VBF NN*s, respectively, which is in accordance with the higher available statistics of the training dataset for the *nonVBF NN* (see Table 8.3). The $\lambda$-parameter that steers the strength of the *L2 weight regularisation* is larger for the *VBF NN*s, the slope $\alpha$ of the leaky ReLU activation function is comparable for all networks,

| Hyperparameter | Choice | | | |
|---|---|---|---|---|
| | nonVBF NN | VBF$_{Z\to\tau\tau}$ NN | VBF$_{top}$ NN | VBF$_{fakes}$ NN |
| $L_h$ (# hidden layers) | 2 | 4 | 3 | 4 |
| $\mathcal{N}^{0_h}$ (# nodes first hidden layer) | 512 | 128 | 128 | 128 |
| $\lambda$ (of L2 weight regul.) | $4.8 \times 10^{-5}$ | $2.92 \times 10^{-4}$ | $9.4 \times 10^{-5}$ | $3.56 \times 10^{-4}$ |
| $\alpha$ (leaky ReLU slope below 0) | $8.0537 \times 10^{-2}$ | $1.9614 \times 10^{-2}$ | $6.2515 \times 10^{-2}$ | $8.4219 \times 10^{-2}$ |
| optimiser | SGD | Adam | Adam | Adam |
| $\eta$ (learning rate) | $2.5810 \times 10^{-2}$ | $1.42 \times 10^{-4}$ | $2.15 \times 10^{-4}$ | $3.507 \times 10^{-3}$ |
| $s_b$ (batch size) | 128 | 128 | 512 | 1024 |

**Table 8.7:** Chosen hyperparameters for the four networks.

only a bit smaller for the *VBF$_{Z\to\tau\tau}$ NN*. The *nonVBF NN* favours the SGD optimiser and hence also a larger learning rate $\eta$ while it is smaller for the *VBF NN*s which favour the *Adam* optimiser. The batch size $s_b$ is largest for the *VBF$_{fakes}$ NN* which is justified by the large difference in available training events between the two classes, signal and $j \to \ell$-fakes data (compare to Table 8.3). This larger difference requires a larger batch size to ensure that a sufficient amount of events of both classes are included per batch. There are approximately 24 times more signal events than $j \to \ell$-fake data events available, which means that there are approximately 983 events of the former and 41 events of the latter in one batch of size 1024.

### 8.2.5 Results of the NN Training

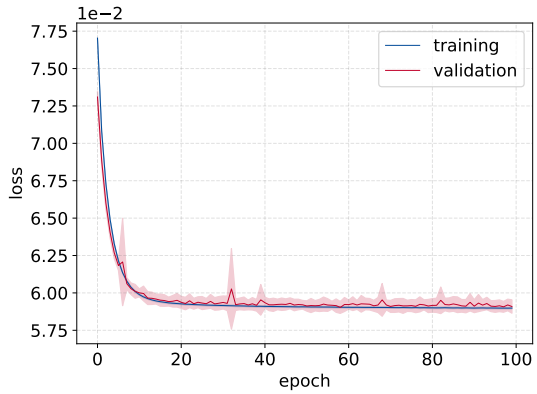
This section starts with a discussion of the performance of the *Neural Network*s trained with the input features and hyperparameters as described above and inspects the ranking of the input features. Afterwards the compatibility of the $e\tau$- and the efficiency-corrected $\mu\tau$-dataset in the distributions of the signal node probabilities, similar to Figures 7.19 and 7.20, is validated with simulated events. The section concludes with the investigation of the modelling of the data by the predictions in these distributions.

#### Performance of the *Neural Network*s

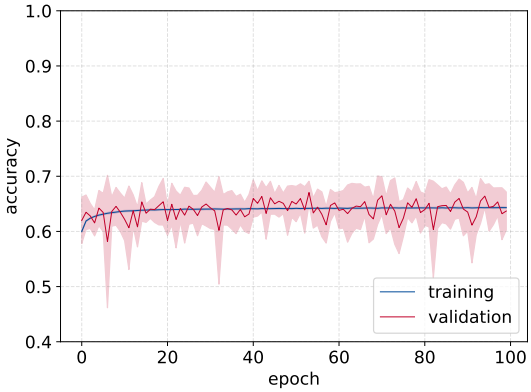
All results shown in this section are based on evaluating each of the 10 *Neural Network*s of the 10-fold cross validation with the events of the corresponding *validation fold*. Hence, 100 % of the dataset is used to obtain the results.

The loss and accuracy curves are given in Figure 8.27. They display the evolution of the loss and accuracy (see Section 6.1.5) and hence the performance of the networks as the training progresses. In addition to the curves evaluated on the validation set, the curves evaluated on the training sets are shown. If the validation curves diverge from the training curves, it indicates that the NN does not generalise well on independent data and as such is a sign for overtraining. The loss curves of all four networks however do not show any signs of overtraining. The same applies to the accuracy curves, although they show a higher variability. It is apparent that the statistics of the validation set (100 %) is considerably smaller than for the total training set (800 %).

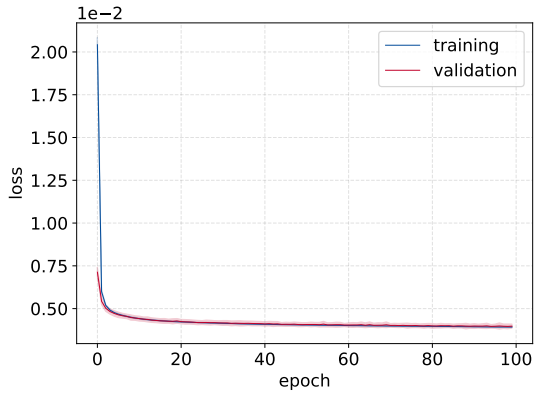
*Confusion matrices* are a measure for the classification power of a *Neural Network*. Each event in the *validation set* is classified to belong to the class for which it has the highest predicted probability. The confusion matrix gives an overview over which ratio of events of each true class are predicted correctly and which ratios are predicted to belong to wrong classes. The sum over the ratios over all predicted classes sums up to 1 for each true class. The confusion matrices of all four NNs are shown in Figure 8.28. A ratio of 87 % of all signal events in the nonVBF SR are predicted to belong to the signal class



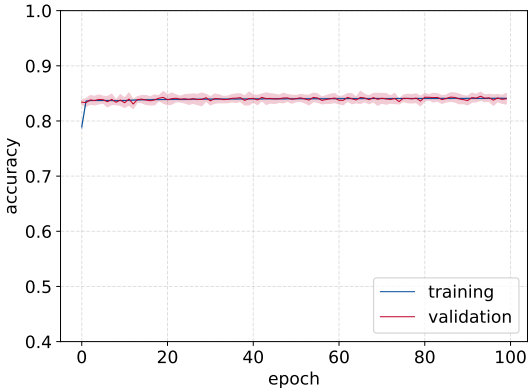

**(a)** *nonVBF NN* Loss



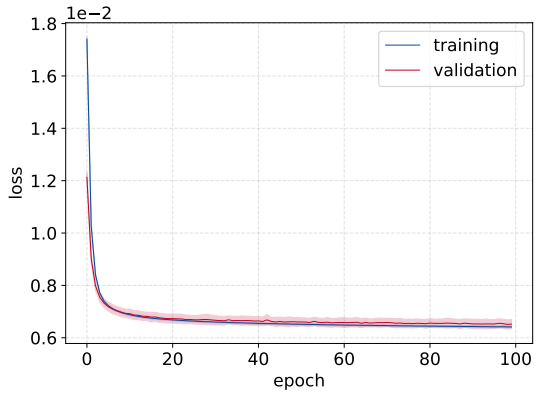

**(b)** *nonVBF NN* Accuracy



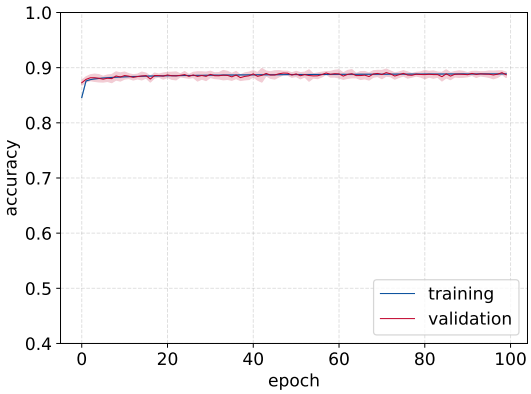

**(c)** $VBF_{Z\to\tau\tau}$ *NN* Loss



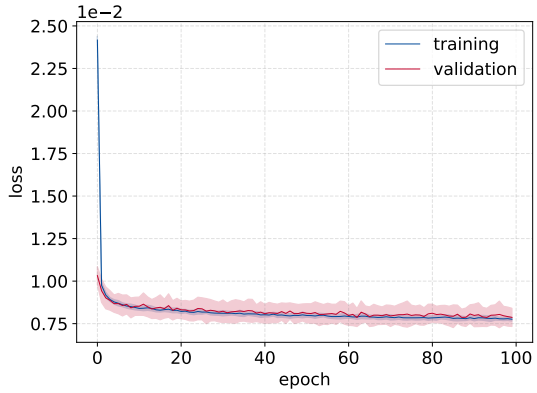

**(d)** $VBF_{Z\to\tau\tau}$ *NN* Accuracy



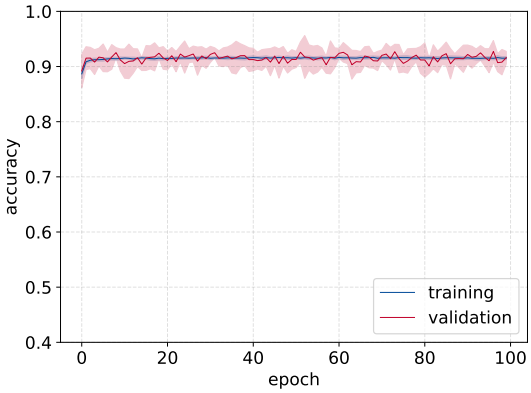

**(e)** $VBF_{top}$ *NN* Loss




**(f)** $VBF_{Z\to\tau\tau}$ *NN* Accuracy

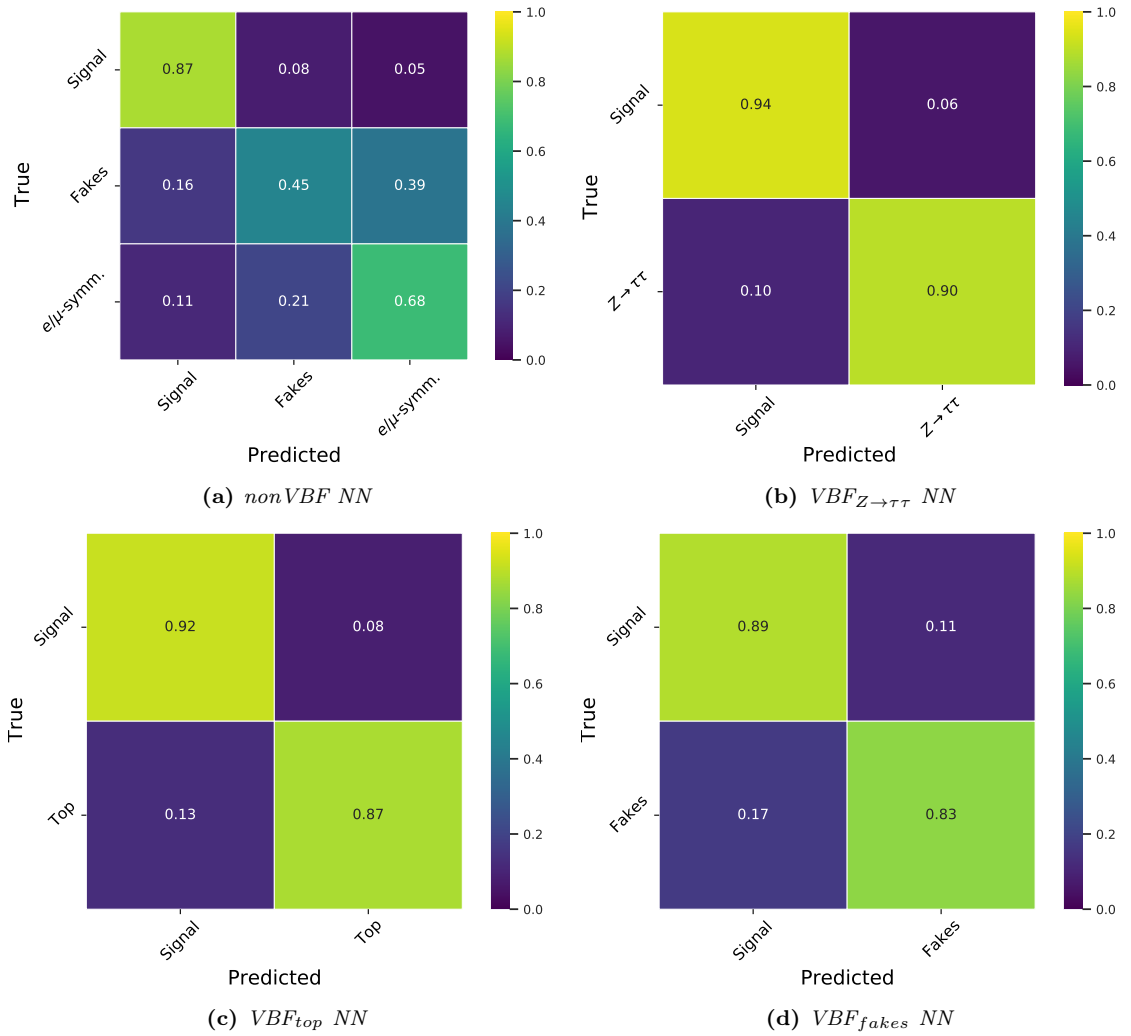



**(g)** $VBF_{fakes}$ *NN* Loss




**(h)** $VBF_{fakes}$ *NN* Accuracy

**Figure 8.27:** Loss and accuracy curves with respect to the number of training epochs for the nonVBF SR (a,b) and for the VBF SR (b-h). Blue represents the loss/accuracy evaluated with the training set and red with the validation set. The solid line indicate the mean over all 10 networks of the 10-fold cross validation, the shaded area represents the standard deviation.




**(a)** *nonVBF NN* **(b)** $VBF_{Z\to\tau\tau}$ *NN*

**(c)** $VBF_{top}$ *NN* **(d)** $VBF_{fakes}$ *NN*

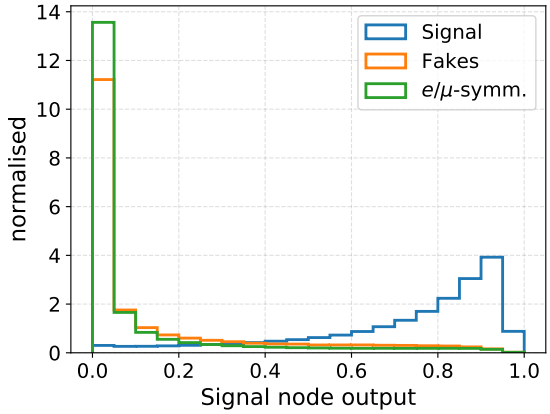
**Figure 8.28:** Confusion matrices of the four NNs normalised for each true class, obtained with the *validation set*.

while 8 and 5 % are predicted to belong to the *fakes* and to the $e/\mu$-symmetric background class, respectively. Only 45 % of the *fakes* are correctly predicted while it is 68 % for the $e/\mu$-symmetric background class. For the *VBF NN*s, 94, 92 and 89 % of the signal events in the VBF SR are correctly classified.

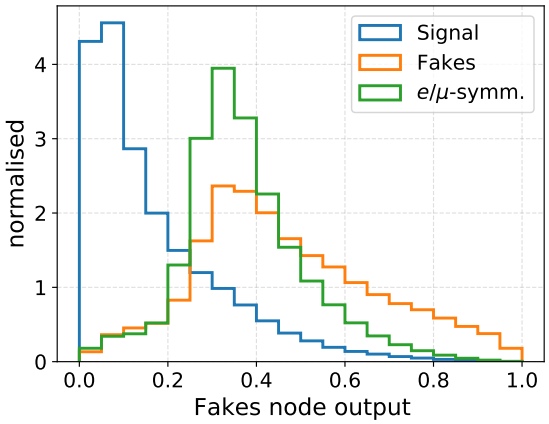
The normalised distributions of the output node probabilities for the individual true classes are shown in Figure 8.29 for the *nonVBF NN* and in Figure 8.30 for the three *VBF NN*s.

The distributions are shown for 20 equidistant bins and were obtained with the *validation set*. For the *VBF NN*s the distributions of the two nodes are the same (just mirrored) as only two classes or output nodes are used and their probabilities add up to 1. The signal node distributions of all four networks show a good separation between signal and background. The separation of the *fakes* and the $e/\mu$-symmetric background seems to be challenging for the *nonVBF NN*. However, this does not pose a problem as only the signal node distribution is utilised in the final statistical analysis.

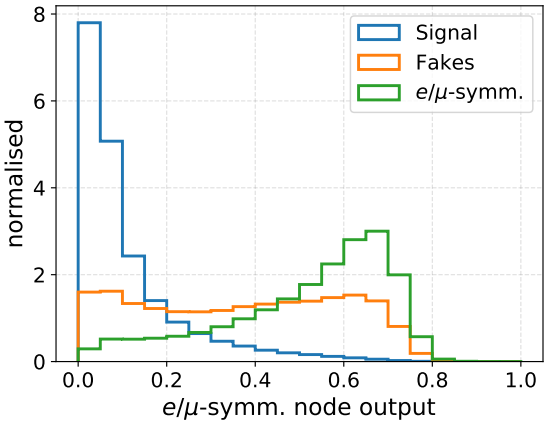
The ROC-curves and the areas under these curves (AUC) are obtained for each output node and for each *Neural Network* as explained in Section 6.1.5. The resulting curves and their AUC-values are shown in Figure 8.31 for all four networks. The *true-positive rate* (TPR) is displayed on the $y$-axis, the *false-positive rate* (FPR) on the $x$-axis. The former corresponds to the "signal" acceptance for a given cut value on the respective output node




**(a)** signal node distribution




**(b)** *fakes* node distribution




**(c)** $e/\mu$-symmetric background node distribution

**Figure 8.29:** Normalised distributions of the output node probabilities of the *nonVBF NN* with 20 equidistant bins, obtained with the *validation set*.

probablity distribution while the latter corresponds to 1-"background" rejection. Here, "signal" means the class of the corresponding output node and "background" all other processes. Ideally the AUC-values are close to 1. Hence, the AUC-values for the signal nodes of all four networks indicate good performance while the values for the *fakes* and $e/\mu$-symmetric background of the *nonVBF NN* are smaller. The values of the background nodes of the *VBF NN*s are the same as for the signal node since their distributions are the same as for the signal nodes but mirrored.

The *Binned Asimov Significance Z* (Section 6.1.5), assuming a branching ratio of 1 % for the signal, is calculated for the signal node distributions with 10 equidistant bins for all four networks. The numbers are listed in Table 8.8. However, in this table $Z$ is only

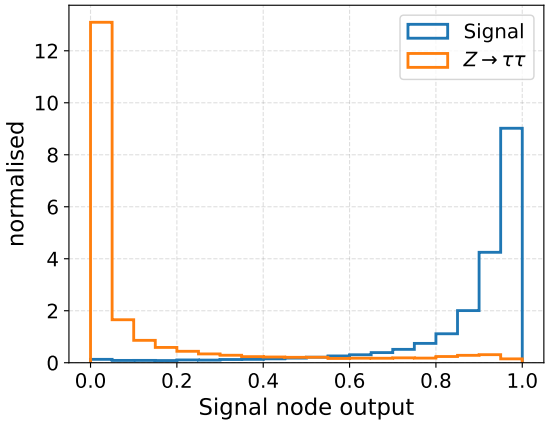
| *nonVBF NN* | $VBF_{Z\to\tau\tau}$ *NN* | $VBF_{top}$ *NN* | $VBF_{fakes}$ *NN* |
|---|---|---|---|
| $24.3 \pm 1.3$ | $14.6 \pm 5.0$ | $9.7 \pm 0.8$ | $15.0 \pm 7.9$ |

**Table 8.8:** The *Binned Asimov Significance Z* for each of the signal node probability distributions with 10 equidistant bins.

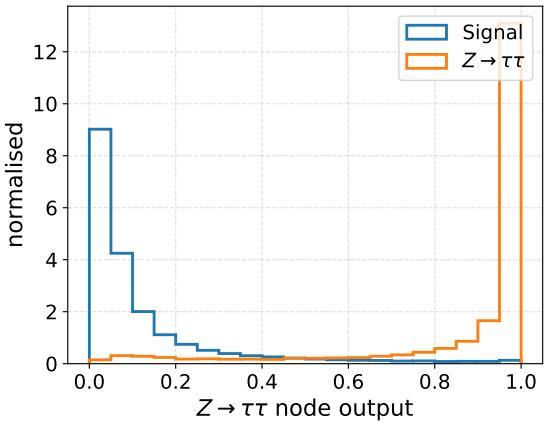
calculated with respect to the background processes that were also used in the training which particularly impacts $Z$ of the *VBF NN*s. Besides that, $Z$ is calculated for the combined $e\tau$- and $\mu\tau$-datasets in accordance with the training process.

Finally, the ranking of the input variables based on the *Permutation Feature Importance* described in Section 6.3.3 is shown in Figure 8.32 for all four networks. The metric used to

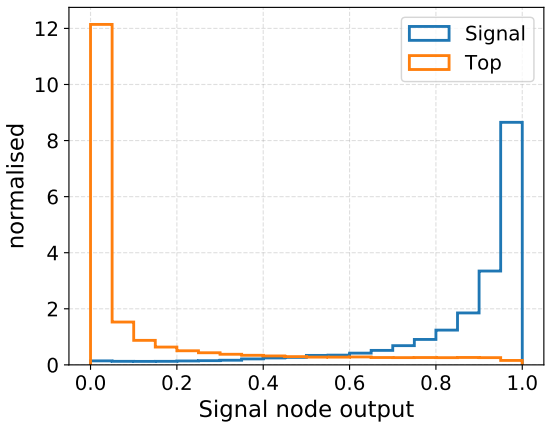
obtain this ranking is, again, the *Binned Asimov Significance* $Z$ with 10 equidistant bins. In general, the importance is calculated as the ratio of the metric obtained after the data is shuffeled ($Z_{\text{shuff}}$) over the original metric ($Z_{\text{orig}}$). As the original metric is expected to be larger than the shuffeled metric when using $Z$ as metric, both metric values (original and shuffeled) are transformed before calculating their ratio according to: $Z' = 1 - Z/20$. This assumes that $Z$ is smaller than 20 which is the case for all four NNs when evaluated for each validation fold of the 10 networks (in the course of 10-fold cross validation) individually. Now, the larger the ratio (i.e. the importance) of one input variable is, the higher it is ranked. For a few very low-ranked variables the importance is slightly smaller than 1, i.e. the corresponding network performs better when the value of the respective variable is assigned randomly compared to when its original value is used. Within the uncertainty given for each importance these are, however, in agreement with 1. Overall, the mass observables ($m_{\ell\ell}$, $m_{\text{coll}}$, $m_T(\ell_\tau, E_{\text{T}}^{\text{miss}})$) are of high importance in all four networks. The mass obtained from the *missing mass calculator* ($m_{\text{MMC}}^{\text{mlm}}$) is less important. Another higher ranked variable in all networks is the angular separation of both leptons ($\Delta R(\ell_H, \ell_\tau)$). After the first three to five observables the importance drops. Otherwise, the ranking varies between the networks which is reasonable as the considered processes vary. The *Permutation Feature Importance* ranking gives an indication of the importance of the variables, but can be biased by correlations between the input variables. The correlations can result in lower importance values of the respective variables as they basically share the importances. In addition, the shuffeling of the values of the input variables over the different events can lead to unphysical combinations per event.



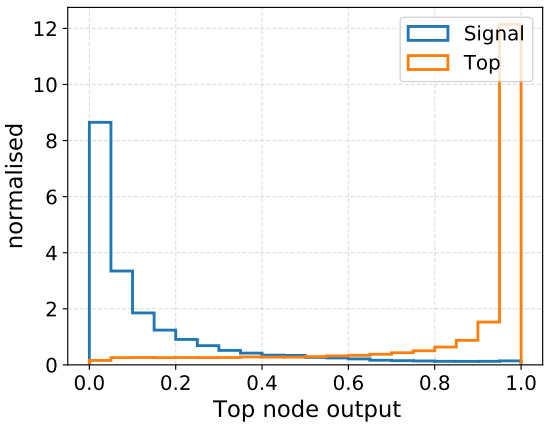

**(a)** $VBF_{Z\to\tau\tau}$ *NN* signal node distribution



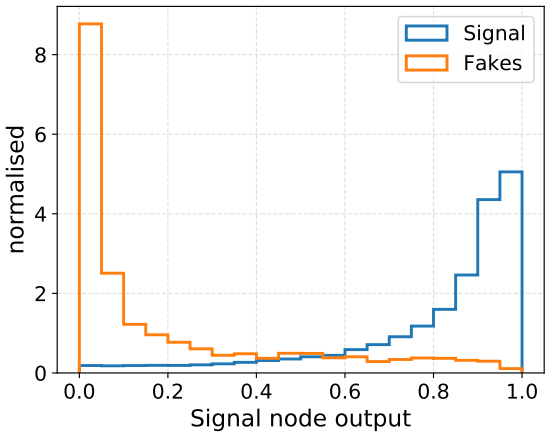

**(b)** $VBF_{Z\to\tau\tau}$ *NN*background node distribution



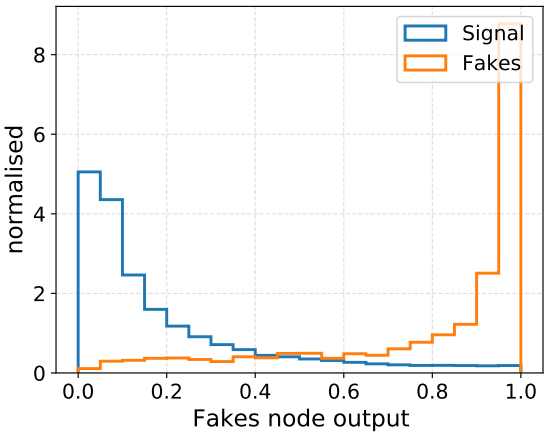

**(c)** $VBF_{top}$ *NN* signal node distribution




**(d)** $VBF_{top}$ *NN* background node distribution

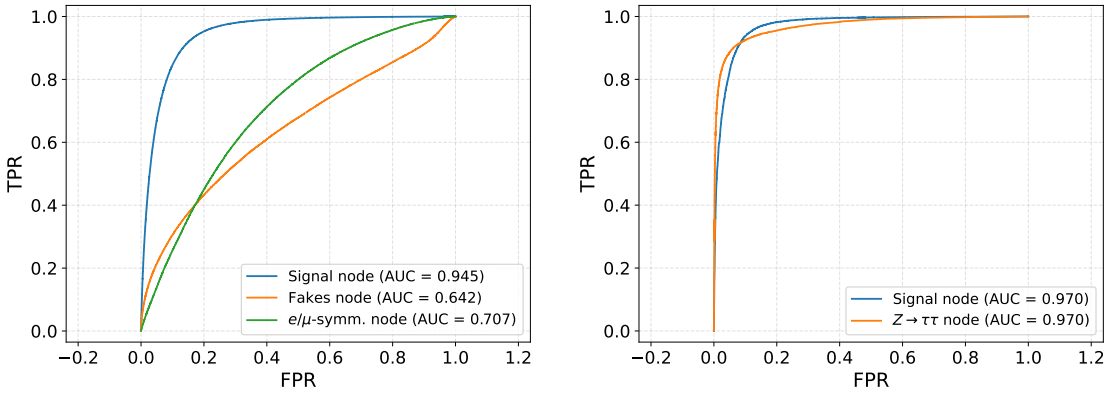



**(e)** $VBF_{fakes}$ *NN* signal node distribution




**(f)** $VBF_{fakes}$ *NN* background node distribution

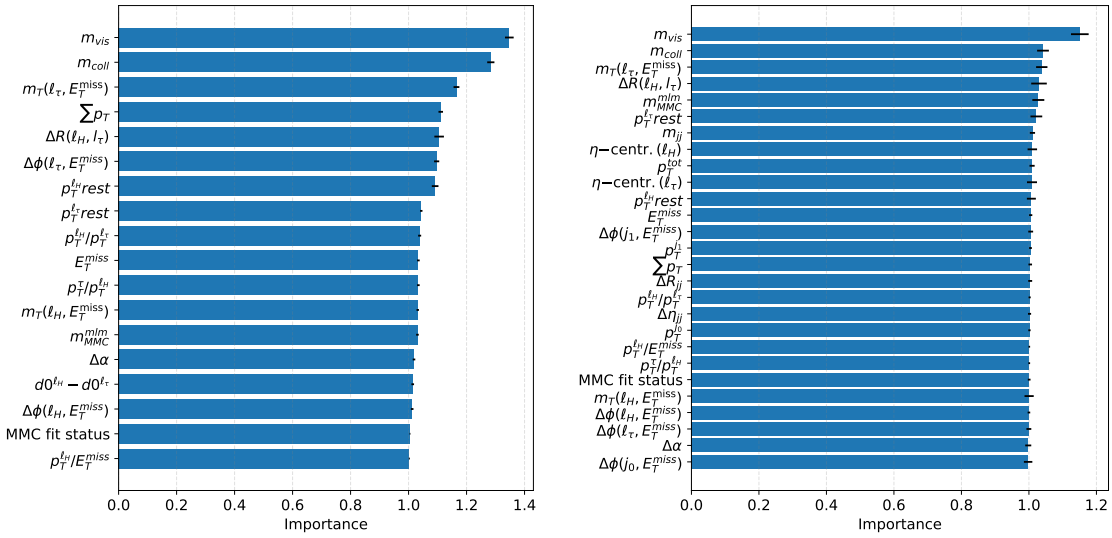
**Figure 8.30:** Normalised distributions of the output node probabilities of the *VBF NN*s with 20 equidistant bins, obtained with the *validation set*.




**(a)** *nonVBF NN* **(b)** $VBF_{Z\to\tau\tau}$ *NN*

**(c)** $VBF_{top}$ *NN* **(d)** $VBF_{fakes}$ *NN*

**Figure 8.31:** ROC-curves and the corresponding AUC-values of all output nodes of the four NNs, obtained with the *validation set*.




**(a)** *nonVBF NN* **(b)** $VBF_{Z\to\tau\tau}$ *NN*

**(c)** $VBF_{top}$ *NN* **(d)** $VBF_{fakes}$ *NN*

**Figure 8.32:** Ranking of the input features of the four networks, obtained with the *validation set*. It is based on the *Permutation Feature Importance* with the *Binned Asimov Significance Z* with 10 equidistant bins used as metric. The displayed importance values are obtained from averaging the importance values over all 10 networks of the 10-fold cross validation. The uncertainties on the mean values are indicated by the black bars.

**Combination of the three *VBF NN* signal node probabilities into one observable**

In order to obtain one discriminating observable for the VBF SR, the probabilities $p$ of the signal nodes of the three *VBF NN*s are linearly combined:

$$p_{\text{VBFNN}} = \frac{\sum_i c_i \cdot p_{\text{VBFNN}_i}}{\sum_i c_i} \quad \text{with } i \in \{Z \to \tau\tau,\ \text{Top},\ fakes\}\,. \tag{8.11}$$

The final observable $p_{\text{VBFNN}}$ is calculated per event with $p_{\text{VBFNN}_i}$ being the probability of the event to belong to the signal class of the respective NN and $c_i$ being the corresponding coefficient which is between 0 and 1. The values of the individual $c_i$ are optimised by performing a scan over all possible coefficient combinations with values between 0.0 and 1.0 in steps of 0.1 per coefficient $c_i$. The *Binned Asimov Significance Z* evaluated using 20 equidistant bins is used as metric. This optimisation is performed separately for the $e\tau$- and $\mu\tau$-dataset. The resulting values of the coefficients are summarised in Table 8.9. The values

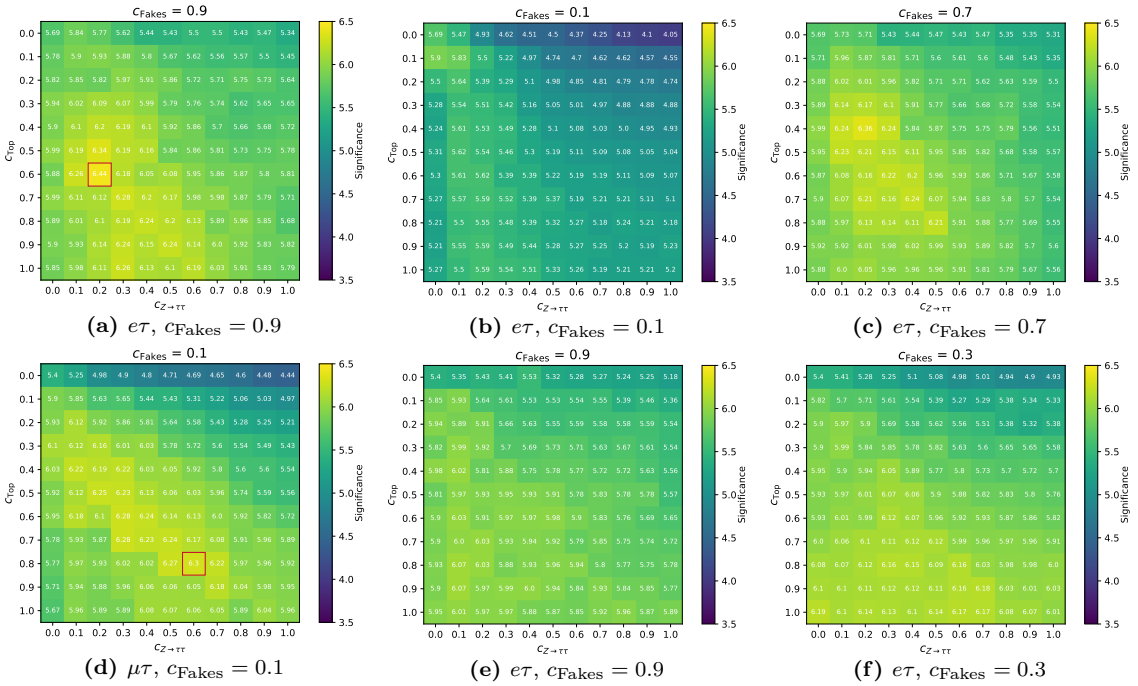
| | $VBF_{Z\to\tau\tau}$ *NN* | $VBF_{top}$ *NN* | $VBF_{fakes}$ *NN* |
|---|---|---|---|
| $e\tau$ | 0.2 | 0.6 | 0.9 |
| $\mu\tau$ | 0.6 | 0.8 | 0.1 |

**Table 8.9:** Coefficients of the single *VBF NN*s for their linear combination in order to obtain one final discriminant, separate for the $e\tau$- and $\mu\tau$-datasets.

are different for the $e\tau$- and $\mu\tau$-dataset even though a roughly symmetric behaviour could be expected. However, the $j \to \ell$-fake contribution behaves differently for both datasets which can justify the different importance given to the $VBF_{fakes}$ *NN* for the respective datasets. Furthermore, many different sets of coefficient values lead to similar values of the *Binned Asimov Significance Z* in the optimisation procedure. The values of $Z$ with respect to the coefficient values of the $VBF_{top}$ *NN* on the vertical and the $VBF_{Z\to\tau\tau}$ *NN* on the horizontal axis are visualised in Figure 8.33. The figures are shown for the best value and for two other exemplary values of the coefficient corresponding to the $VBF_{fakes}$ *NN*. Values of $Z$ close to the best value can be obtained with other sets of coefficients.

**Validation of the restoration of the *Symmetry*-assumption in the NN distributions**

The restoration of the *Symmetry*-assumption (Chapter 3) in the distributions of the *nonVBF NN* and the combined *VBF NN* are validated in the same way as done previously for several kinematic observables in the *Basic* SR (end of Section 7.2.3). The efficiency-ratio $\mathcal{R}^{\mu\tau}$ (Eqs. 7.8 and 7.17) is applied to the $e/\mu$-symmetric background processes estimated from simulation in the $e\tau$-dataset and compared to the sum of the same processes in the $\mu\tau$-dataset. The corresponding figures are given in Figures 8.34 and 8.35 for the *nonVBF NN* and the combined *VBF NN* output distributions, respectively. The agreement of the efficiency-corrected ($\cdot\mathcal{R}_\varepsilon$) $e\tau$-dataset and the $\mu\tau$-dataset is good in all of the distributions. Only the bins at higher values of the *nonVBF NN fakes* node distribution show a discrepancy of up to $\sim 13\,\%$. On one hand there are only a few events in these bins (140 and 27 for $\mu\tau$ in the last two bins), on the other hand this could hint at issues of the applicability of the efficiencies for these very $j \to \ell$-fake like events, meaning that the efficiencies are not valid in this extreme but small phase space. However, the *fakes* node is not utilised in subsequent steps of the analysis and to derive the final results and hence the observed discrepancy is potentially interesting but not important to the analysis. Events which are assigned to belong to the $j \to \ell$-fakes class with such a high probability consequently have a low signal probability value and hence are not in the sensitive region.




**(a)** $e\tau$, $c_{\text{Fakes}} = 0.9$ **(b)** $e\tau$, $c_{\text{Fakes}} = 0.1$ **(c)** $e\tau$, $c_{\text{Fakes}} = 0.7$

**(d)** $\mu\tau$, $c_{\text{Fakes}} = 0.1$ **(e)** $e\tau$, $c_{\text{Fakes}} = 0.9$ **(f)** $e\tau$, $c_{\text{Fakes}} = 0.3$

**Figure 8.33:** Values of the *Binned Asimov Significance* with respect to the coefficient values of the $VBF_{top}$ *NN* on the vertical and the $VBF_{Z\to\tau\tau}$ *NN* on the horizontal axis for the $e\tau$- (top) and $\mu\tau$-dataset (bottom). (a) and (d) show the values when the coefficient corresponding to the $VBF_{fakes}$ *NN* ($c_{\text{Fakes}}$) is fixed to its optimal value. The best values of $c_{Z\to\tau\tau}$ and $c_{\text{Top}}$ are indicated by red boxes. (b) and (e) show the values when $c_{\text{Fakes}}$ is fixed to the optimal value of the respective other dataset. (c) and (f) show the values when $c_{\text{Fakes}}$ is fixed to an (arbitrary) value close to its optimal value.

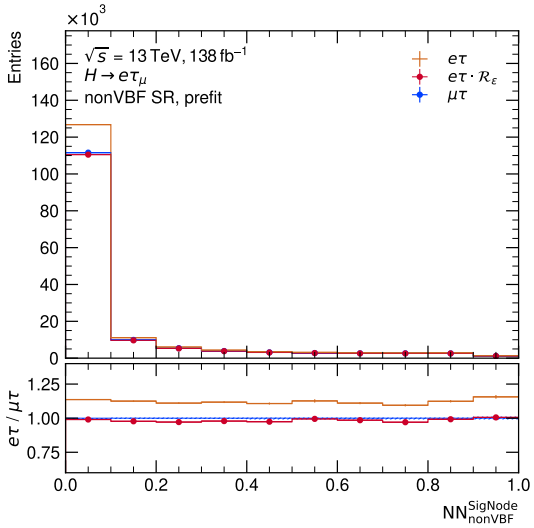
### Validation of the modelling of the data by the predictions in the NN distributions

A good modelling of the data by the prediction of the processes in the probability output node distributions of the NNs is crucial in order to be able to trust the final results derived from these distributions. Hence, the modelling of the signal node distribution of the *nonVBF NN* and the modelling of the combined distribution of the three *VBF NN* signal node distributions are most important as these distributions are used in the final statistical analysis (Chapter 10). The modelling is less important in the case when the $e/\mu$-symmetric background is estimated from MC simulations as this is not used in the statistical analysis and no correcting normfactors are derived (as mentioned above) which makes it a less precise estimate. It is mainly utilised to see the distributions of the single contributing processes. Consequently, a good modelling in the case when the $e/\mu$-symmetric background is estimated with the *Symmetry Method* is necessary.

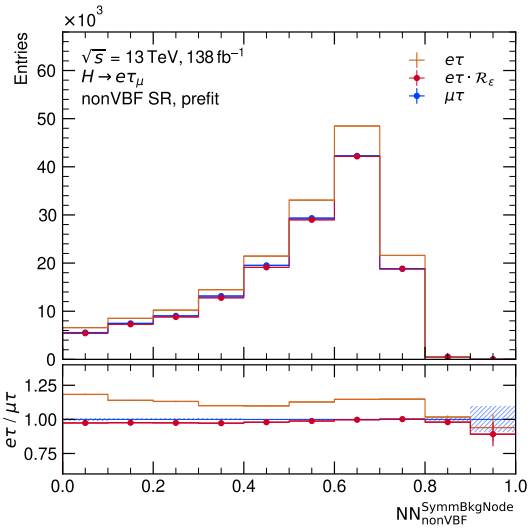
The distributions are obtained by using the *test set*. The signal-enhanced bins are blinded whenever the signal-over-background ratio ($s/b$) exceeds 5 % when assuming a branching ratio of 1 % for signal. The blinded versions of the distributions are shown here such that the judgement is not biased by information from the signal-sensitive bins.

Figures 8.36–8.38 show the distributions of the *nonVBF NN*, of the single *VBF NN*s and of the combined *VBF NN*, respectively, with the $e/\mu$-symmetric background estimated from simulations. Figures 8.39–8.41 show the same distributions but with the $e/\mu$-symmetric background estimated with the *Symmetry Method*.

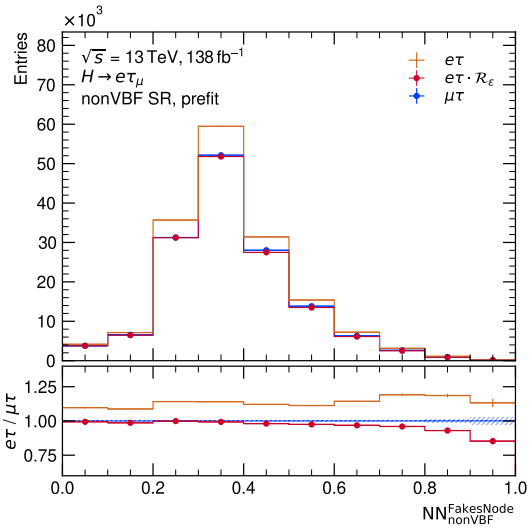
**Modelling using simulations for the $e/\mu$-symmetric background** The overall modelling of the NN distributions with the $e/\mu$-symmetric background estimated from simulations in the nonVBF SR is reasonable, in particular when considering that not all systematic uncertainties are included in these figures. However, for high values of the




**(a)** signal node distribution




**(b)** $e/\mu$-symmetric background node distribution



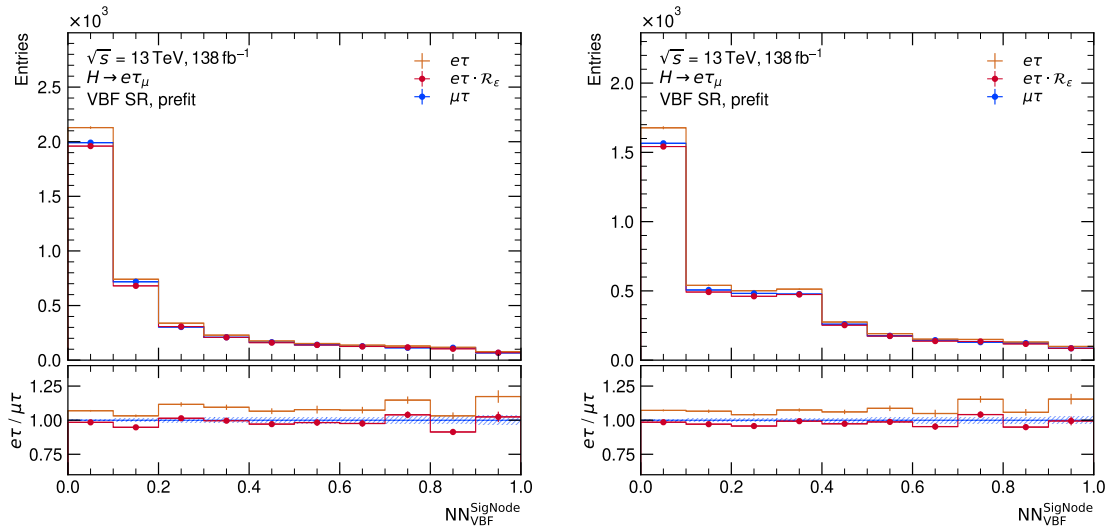

**(c)** *fakes* node distribution

**Figure 8.34:** Distributions of the output node probabilities of the *nonVBF NN*, comparing the $e\tau$- and the efficiency-corrected ($\cdot\mathcal{R}_\varepsilon$) $e\tau$-dataset to the $\mu\tau$-dataset using all $e/\mu$-symmetric background processes estimated from simulations ($Z \to \tau\tau$, top-quark production, diboson-production, $H \to \tau\tau$ and $H \to WW$). The indicated uncertainties only represent the statistical uncertainties of the simulated events.

*nonVBF NN fakes* node probability distribution, a slight disagreement between prediction and data is visible which is more pronounced for the $\mu\tau$-dataset. The $\mu\tau$-dataset is dominated by jets faking electrons. As the *fakes* node distribution is not used in any further steps of the analysis, the observed behaviour is acceptable.

The prediction by the simulated events in the VBF SR is known to overestimate the data as discussed in the beginning of Section 8.2.3 and hence the modelling is not further discussed here. But the strengths of the individual *VBF NN*s is clearly visible (Figure 8.37): in the $VBF_{Z\to\tau\tau}$ *NN* distribution, the $Z \to \tau\tau$-events are shifted to the left in contrast to the other processes. The same applies to events from top-quark and diboson production in the $VBF_{top}$ *NN*. For events of the $j \to \ell$-fakes in the $VBF_{fakes}$ *NN* distribution it is not as noticeable as for the other two.

**Modelling using the *Symmetry Method* for the $e/\mu$-symmetric background** The overall modelling of the NN distributions with the $e/\mu$-symmetric background estimated with




**(a)** *VBF NN* with $e\tau$ coefficients **(b)** *VBF NN* with $\mu\tau$ coefficients

**Figure 8.35:** Distributions of the combined signal output node probabilities of the *VBF NN*s using the combination coefficients for $e\tau$ (a) and for $\mu\tau$ (b). The $e\tau$- and the efficiency-corrected ($\cdot\mathcal{R}_\varepsilon$) $e\tau$-dataset are compared to the $\mu\tau$-dataset using all $e/\mu$-symmetric background processes estimated from simulations ($Z \to \tau\tau$, top-quark production, diboson-production, $H \to \tau\tau$ and $H \to WW$). The indicated uncertainties only represent the statistical uncertainties of the simulated events.

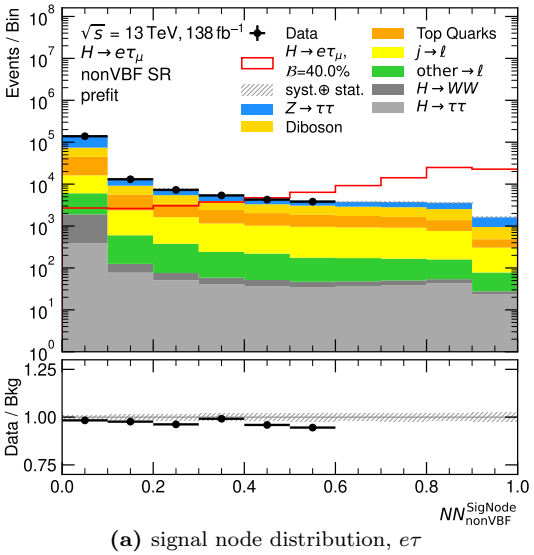
the *Symmetry Method* is very good. The opposite behaviour of the data-over-background ratio in the lower panel between the $e\tau$- and $\mu\tau$-datasets is visible which is intrinsic to the *Symmetry Method* as discussed in Sections 7.2.4 and 8.2.3. In the nonVBF SR, there is a small slope in the ratio in the lower panel of the *fakes* node probability distribution, but this is still in agreement with 1.0 within uncertainties. In addition, the same behaviour for higher values of the *fakes* node probability is visible as observed when estimating the $e/\mu$-symmetric background from MC simulations.

The modelling in the VBF SR is reasonable. In the distributions of the combined *VBF NN* only one bin shows a deviation larger than $1\sigma$ (1 standard deviation).

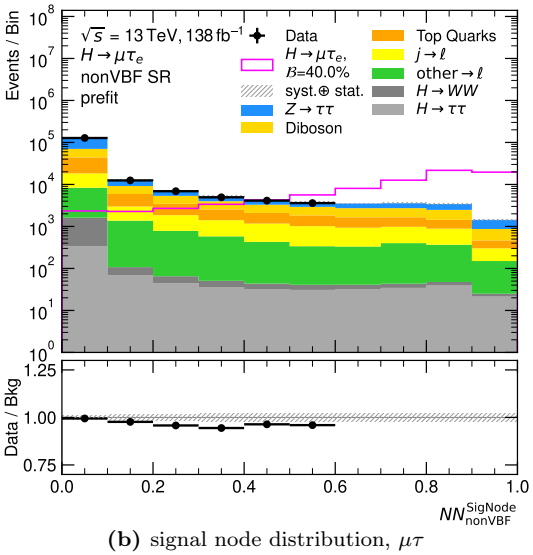
**Modelling in the SC-SR using simulations for the $e/\mu$-symmetric background**

As the signal sensitive bins are blinded, another cross check where also the modelling in these bins can be inspected, is useful. For this purpose, the output node probability distributions in the SC-SR are used. This region was already utilised in Section 7.2.2 to verify the estimation of the $j \to \ell$-fakes since it is enriched with such events. Its definition is identical with the *Basic Selection Signal Region*, only the opposite-sign electric charge requirement is inverted such that same-sign electric charges of the two leptons are required. Since the fakes are the dominant contribution in this region, it is not meaningful to use the *Symmetry Method* here to estimate the $e/\mu$-symmetric background. The respective distributions with the $e/\mu$-symmetric background estimated from simulations are shown in Figures 8.42 and 8.43.

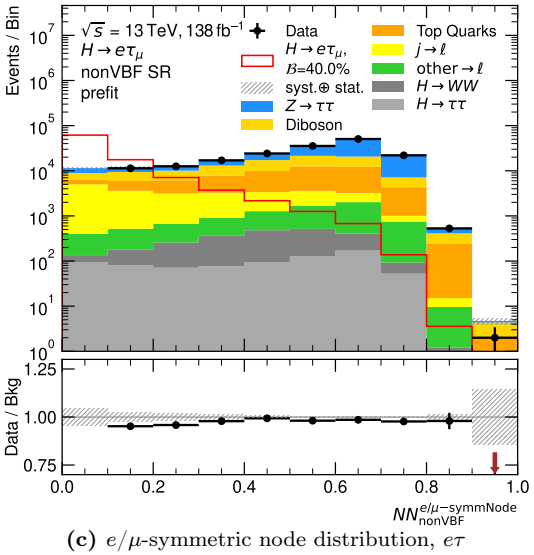
The modelling of data by the background prediction in the signal node distributions of the nonVBF SR is very good for the $e\tau$-dataset and good within uncertainties for the $\mu\tau$-dataset, although an increasing slope is visible in the data-over-background ratio of the latter. The $e/\mu$-symmetric background node distribution only has very few events in the last bins. Apart from the third to last bin in the distribution of the $e\tau$-dataset, the modelling is very good. The distributions of the *fakes*-node is very good for the $e\tau$-dataset while the data in the last bin in the distribution of the $\mu\tau$-dataset is slightly overestimated by the prediction.



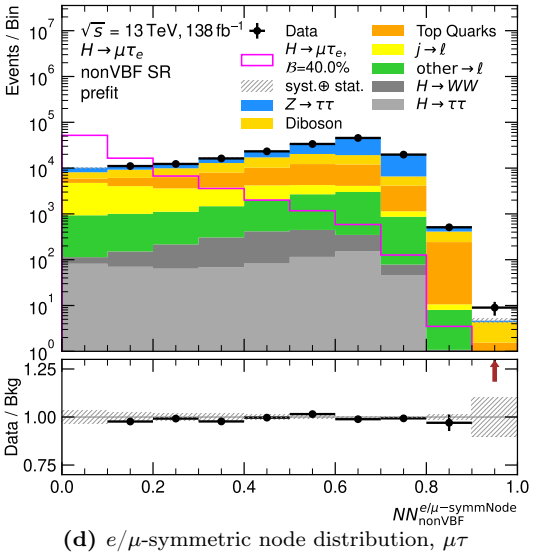

**(a)** signal node distribution, $e\tau$



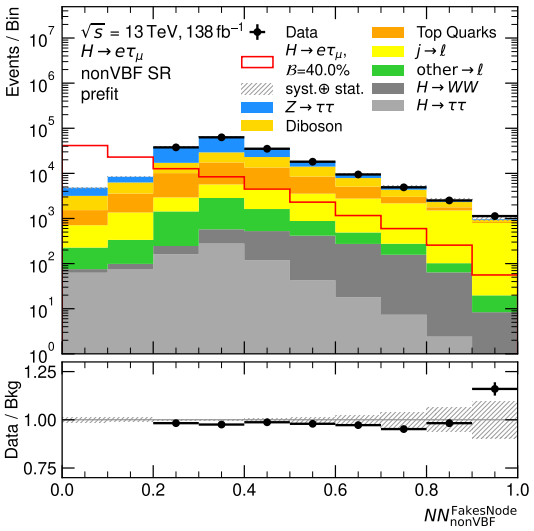

**(b)** signal node distribution, $\mu\tau$



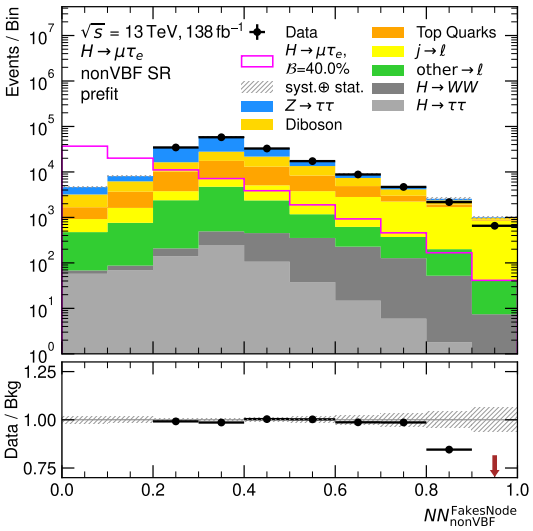

**(c)** $e/\mu$-symmetric node distribution, $e\tau$




**(d)** $e/\mu$-symmetric node distribution, $\mu\tau$

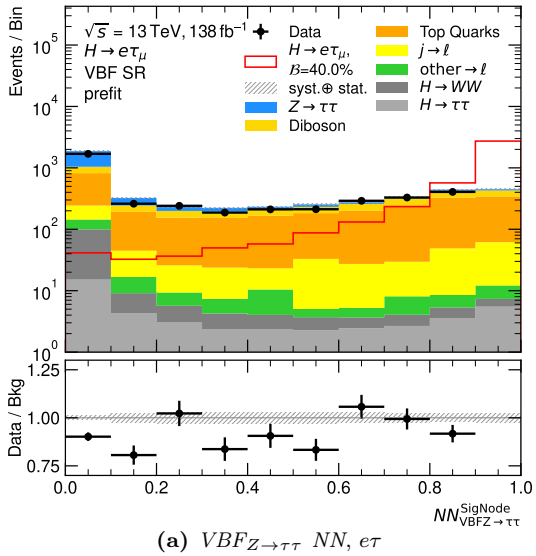



**(e)** *fakes* node distribution, $e\tau$

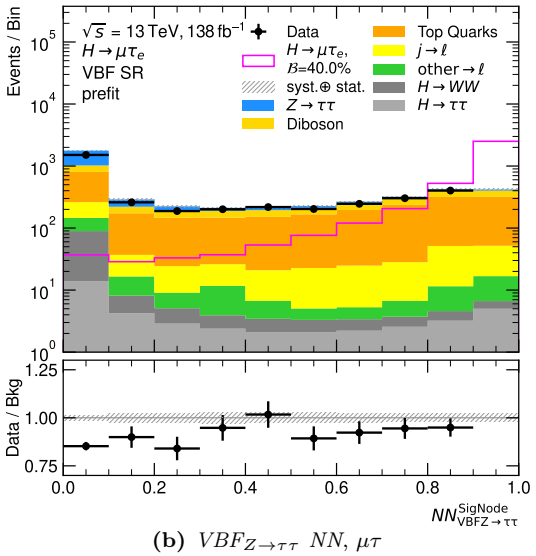



**(f)** *fakes* node distribution, $\mu\tau$

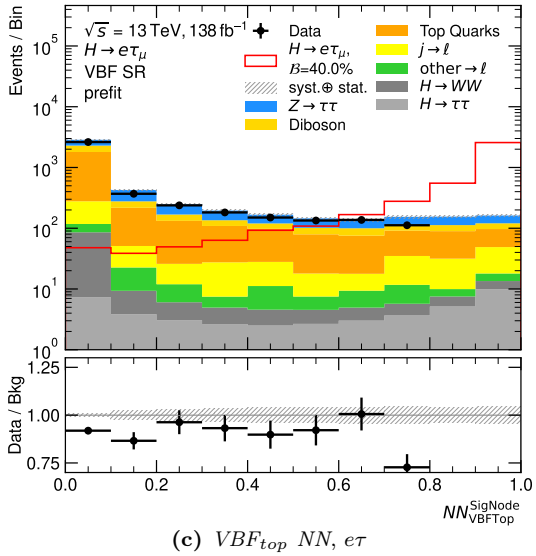
**Figure 8.36:** Distributions of the output node probabilities of the *nonVBF NN* separately for the $e\tau$- (left) and $\mu\tau$-dataset (right), obtained with the *test set*. The $e/\mu$-symmetric background is estimated from MC simulations. The uncertainty band includes the statistical uncertainties on the background predictions and the systematic uncertainties on the $j \to \ell$-fake estimate.



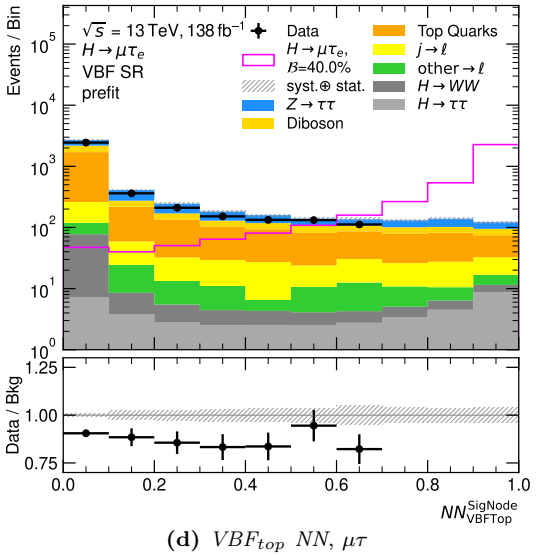

**(a)** $VBF_{Z\to\tau\tau}$ *NN*, $e\tau$



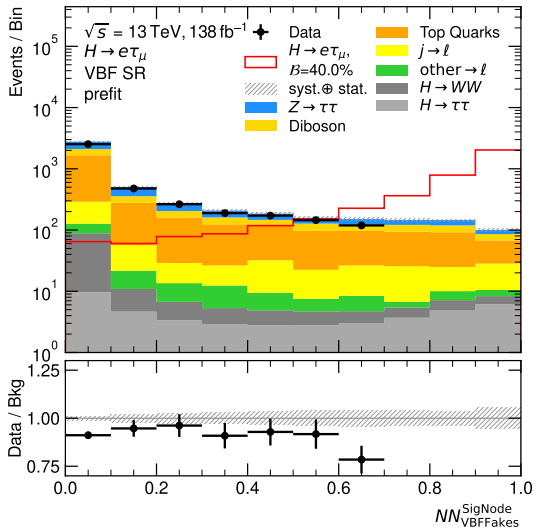

**(b)** $VBF_{Z\to\tau\tau}$ *NN*, $\mu\tau$



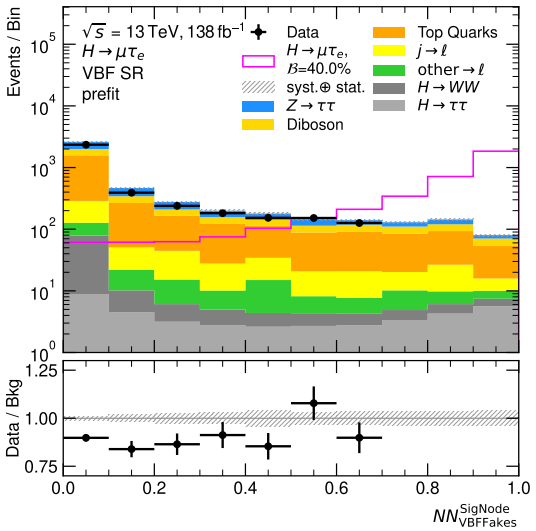

**(c)** $VBF_{top}$ *NN*, $e\tau$




**(d)** $VBF_{top}$ *NN*, $\mu\tau$

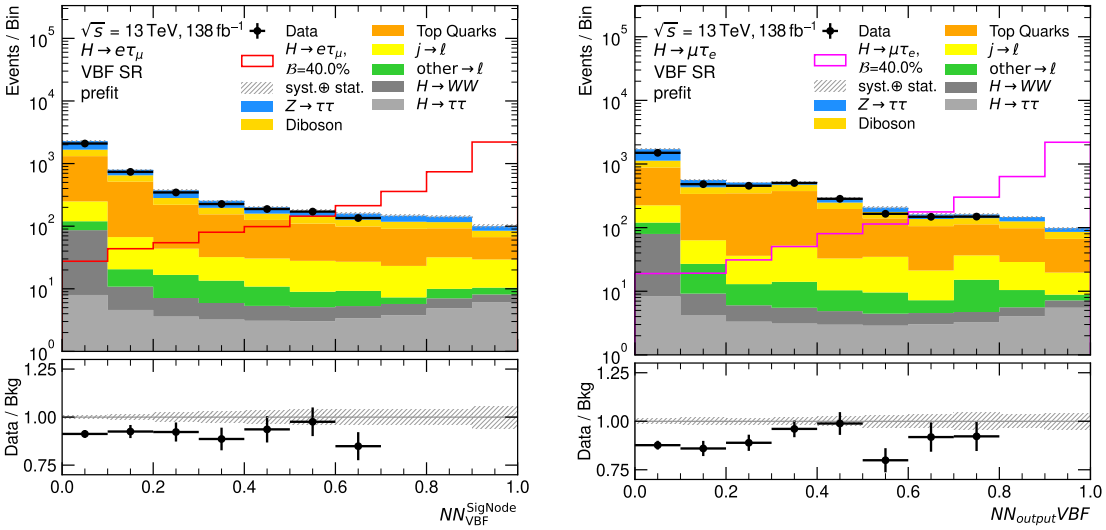



**(e)** $VBF_{fakes}$ *NN*, $e\tau$




**(f)** $VBF_{fakes}$ *NN*, $\mu\tau$

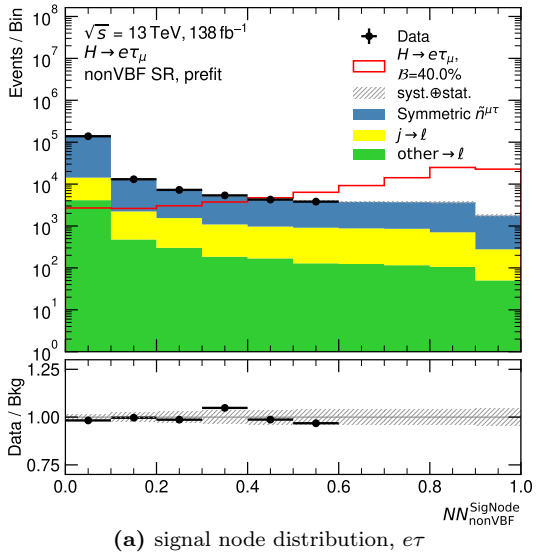
**Figure 8.37:** Distributions of the output node probabilities of the single *VBF NN*s, separately for the $e\tau$- (left) and $\mu\tau$-dataset (right), obtained with the *test set*. The $e/\mu$-symmetric background is estimated from MC simulations. The uncertainty band includes the statistical uncertainties on the background predictions and the systematic uncertainties on the $j \to \ell$-fake estimate.



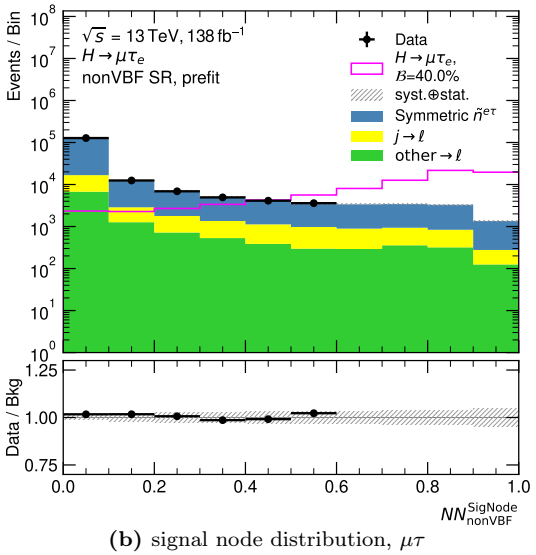

**(a)** *VBF NN*, $e\tau$ **(b)** *VBF NN*, $\mu\tau$

**Figure 8.38:** Distributions of the combined *VBF NN*s, separately for the $e\tau$- (left) and $\mu\tau$-dataset (right), obtained with the *test set*. The $e/\mu$-symmetric background is estimated from MC simulations. The uncertainty band includes the statistical uncertainties on the background predictions and the systematic uncertainties on the $j \to \ell$-fake estimate.

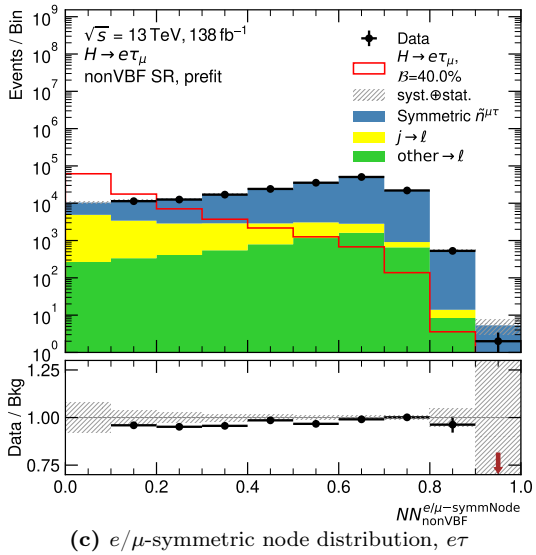
The distributions of the combined *VBF NN*s show a good description of the data by the prediction. Fluctuations are visible due to reduced statistics but apart from the third last bin in the distribution of the $\mu\tau$-dataset, agreement between prediction and data in all bins is observed. The last bin in the distribution of the $e\tau$-dataset is blinded which means that the signal over background ratio is larger than 5 % when assuming a branching ratio of 1 % for the signal. For a background yield of 7.7 events, 0.39 signal events are sufficient for the blinding of this bin.



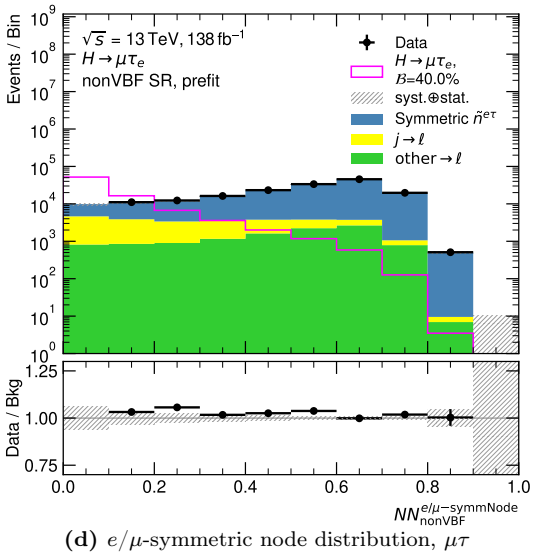

**(a)** signal node distribution, $e\tau$



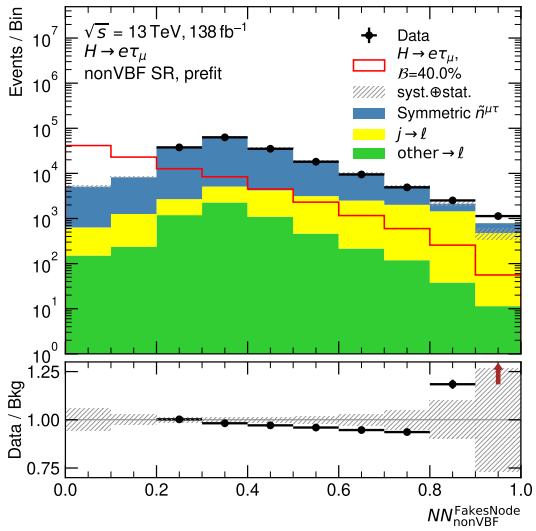

**(b)** signal node distribution, $\mu\tau$



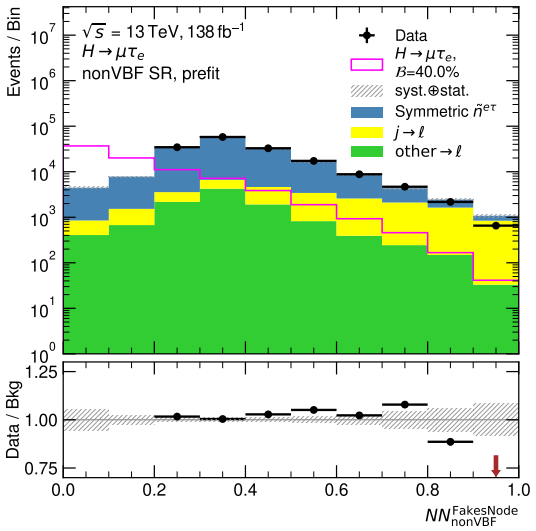

**(c)** $e/\mu$-symmetric node distribution, $e\tau$




**(d)** $e/\mu$-symmetric node distribution, $\mu\tau$

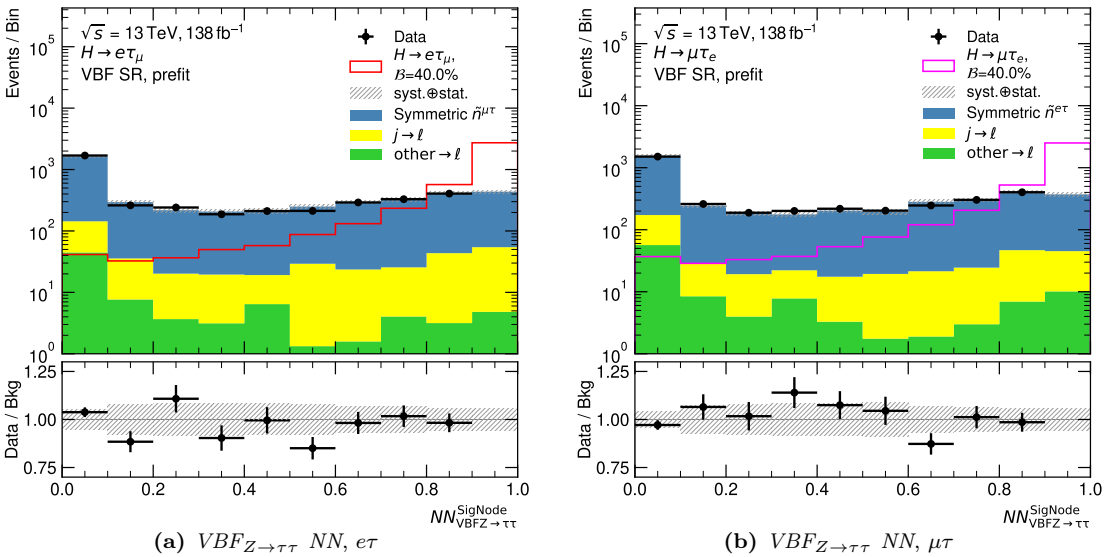



**(e)** *fakes* node distribution, $e\tau$




**(f)** *fakes* node distribution, $\mu\tau$

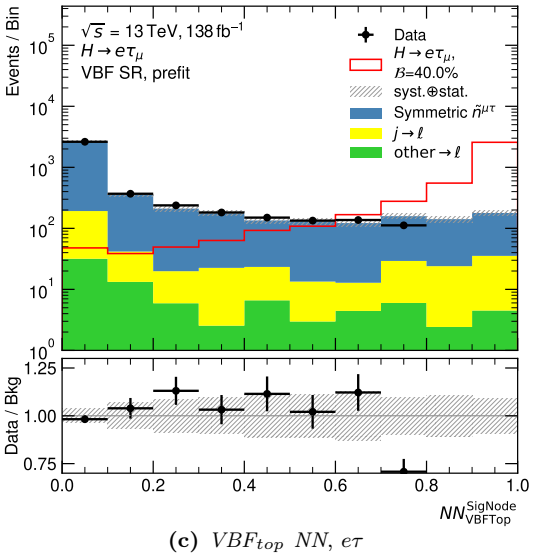
**Figure 8.39:** Distributions of the output node probabilities of the *nonVBF NN* separately for the $e\tau$- (left) and $\mu\tau$-dataset (right), obtained with the *test set*. The $e/\mu$-symmetric background is estimated with the *Symmetry Method*. The uncertainty band includes the statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature.



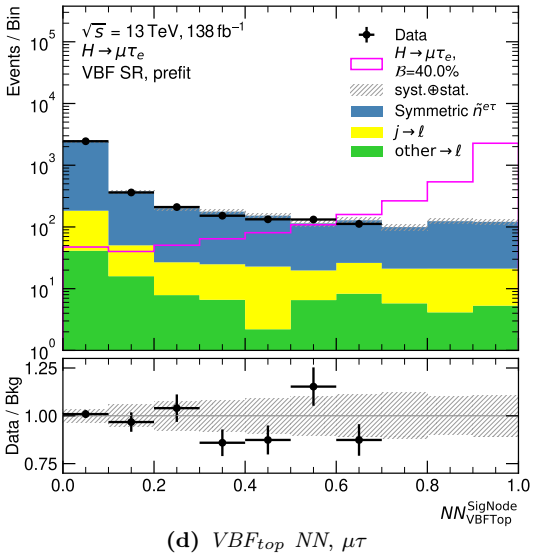

**(a)** $VBF_{Z\to\tau\tau}$ *NN*, $e\tau$



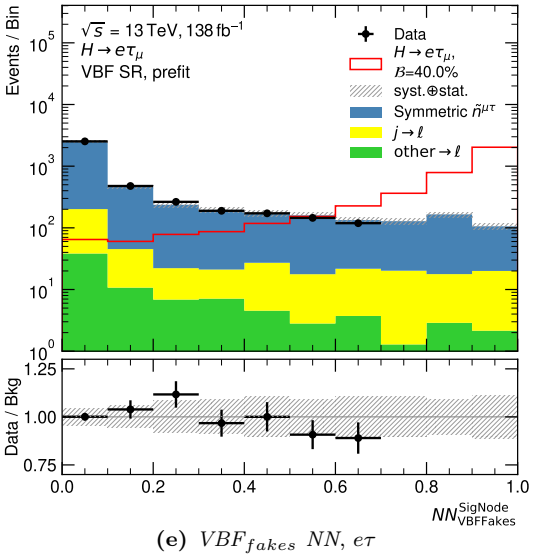

**(b)** $VBF_{Z\to\tau\tau}$ *NN*, $\mu\tau$



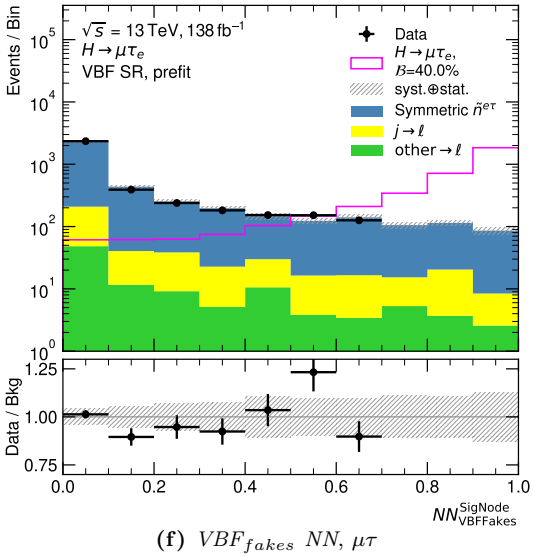

**(c)** $VBF_{top}$ *NN*, $e\tau$




**(d)** $VBF_{top}$ *NN*, $\mu\tau$

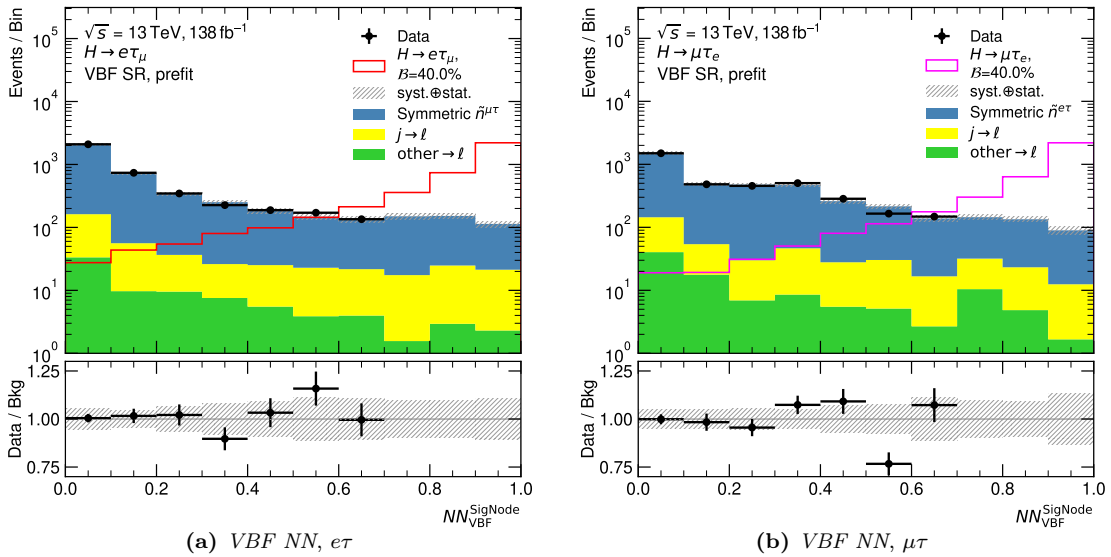



**(e)** $VBF_{fakes}$ *NN*, $e\tau$



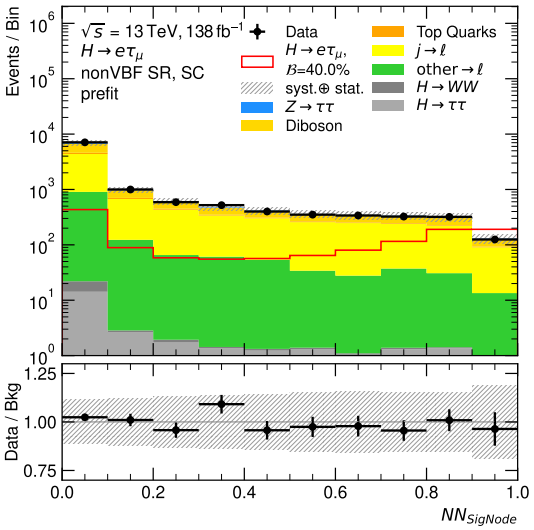

**(f)** $VBF_{fakes}$ *NN*, $\mu\tau$

**Figure 8.40:** Distributions of the output node probabilities of the single *VBF NN*s, separately for the $e\tau$- (left) and $\mu\tau$-dataset (right), obtained with the *test set*. The $e/\mu$-symmetric background is estimated with the *Symmetry Method*. The uncertainty band includes the statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature.

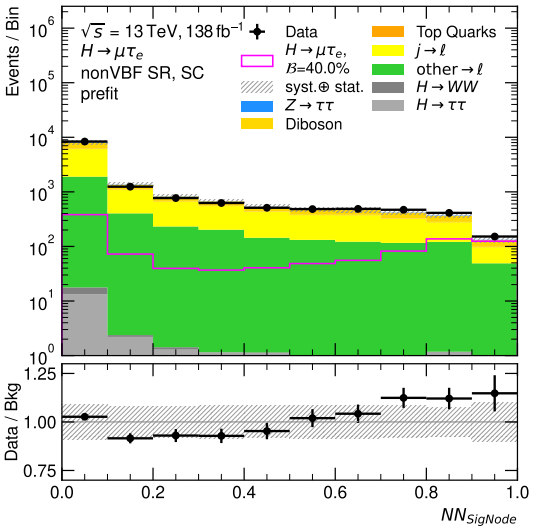



**(a)** *VBF NN*, $e\tau$ **(b)** *VBF NN*, $\mu\tau$

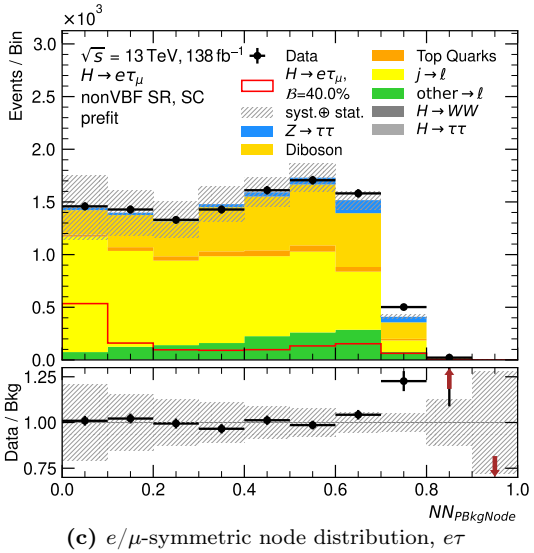
**Figure 8.41:** Distributions of the combined *VBF NN*s, separately for the $e\tau$- (left) and $\mu\tau$-dataset (right), obtained with the *test set*. The $e/\mu$-symmetric background is estimated with the *Symmetry Method*. The uncertainty band includes the statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature.



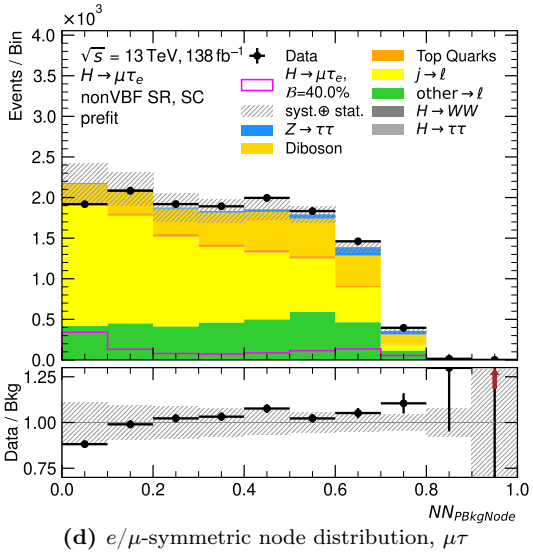

**(a)** signal node distribution, $e\tau$



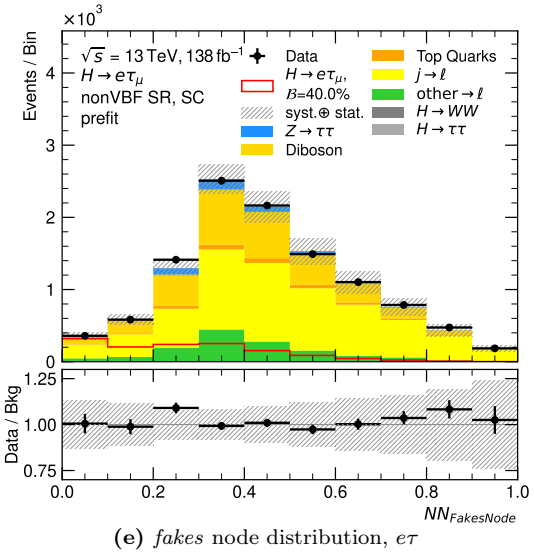

**(b)** signal node distribution, $\mu\tau$



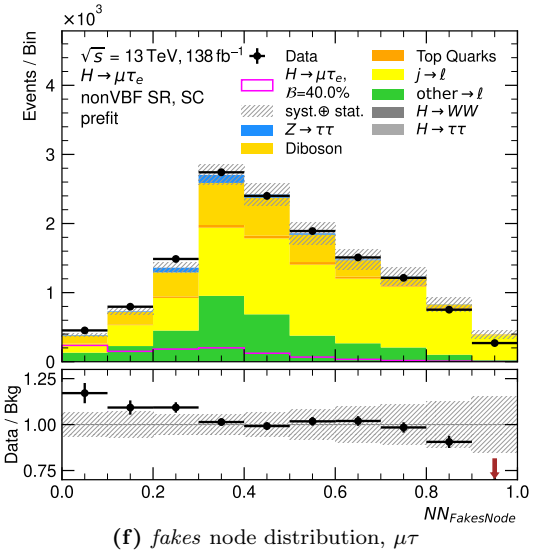

**(c)** $e/\mu$-symmetric node distribution, $e\tau$




**(d)** $e/\mu$-symmetric node distribution, $\mu\tau$

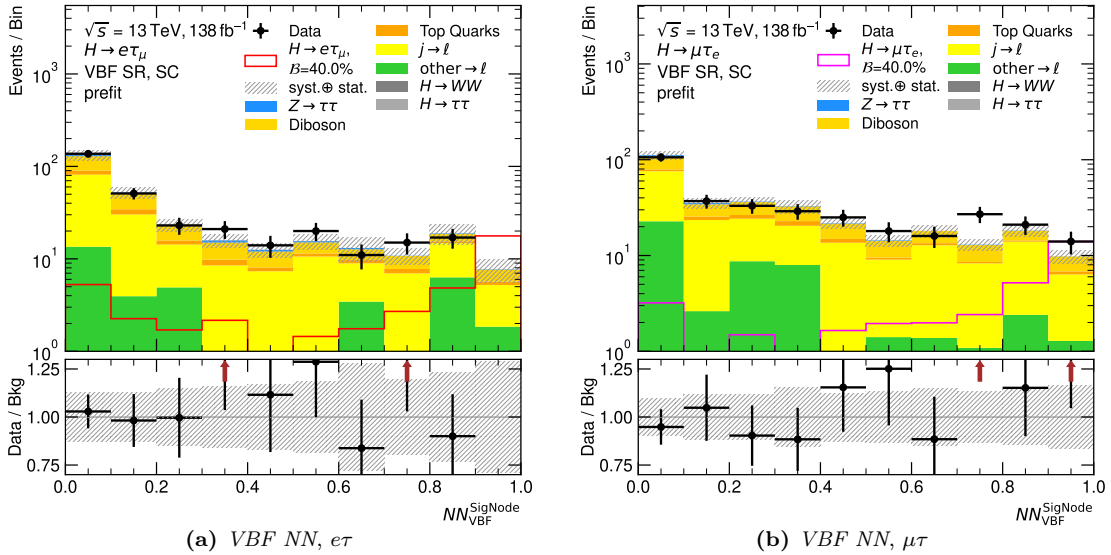



**(e)** *fakes* node distribution, $e\tau$




**(f)** *fakes* node distribution, $\mu\tau$

**Figure 8.42:** Distributions of the output node probabilities of the *nonVBF NN* separately for the $e\tau$- (left) and $\mu\tau$-dataset (right), obtained in the SC-nonVBF SR. The $e/\mu$-symmetric background is estimated from MC simulations. The uncertainty band includes the statistical uncertainties on the background predictions and the systematic uncertainties on the $j \to \ell$-fake estimate.




**(a)** *VBF NN*, $e\tau$ **(b)** *VBF NN*, $\mu\tau$

**Figure 8.43:** Distributions of the combined *VBF NN*s, separately for the $e\tau$- (left) and $\mu\tau$-dataset (right), obtained in the SC-VBF SR. The $e/\mu$-symmetric background is estimated from MC simulations. The uncertainty band includes the statistical uncertainties on the background predictions and the systematic uncertainties on the $j \to \ell$-fake estimate.

### 8.2.6 Binning of the Final Discriminants for the Statistical Analysis

The *nonVBF NN* signal node probability distribution and the distribution of the combined *VBF NN*s are used as final discriminants in the respective signal regions in the statistical analysis. When choosing the binning of the final discriminants, it must be ensured that each process contributes events to each bin. Furthermore, the sensitivity can be enhanced by the choice of the binning. In principle, the sensitivity is improved or stays the same with more bins. However, the previously mentioned requirement must be met and the statistical uncertainties on the prediction should be kept limited.

This results in the following binning strategy. A finely binned histogram of 100 equidistant bins serves as basis. Starting from the right of the distribution and proceeding to the left, these sub-bins are merged into a new bin once the following criteria are met:

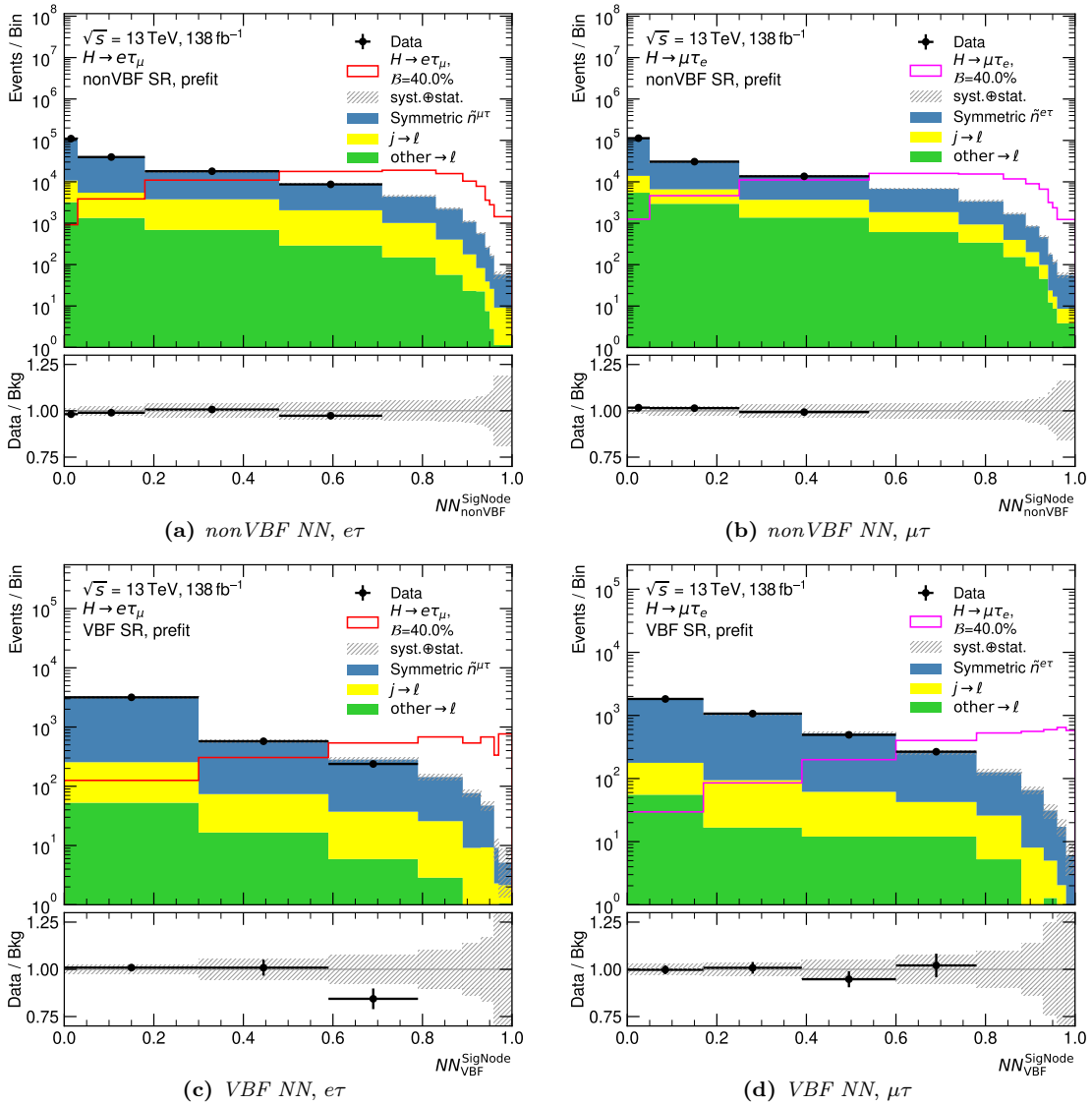
1. The yield of all single processes in the new bin must be larger than zero. The single processes are the data-part of the $e/\mu$-symmetric background, the efficiency-corrected fake contributions $\mathcal{R} \cdot j \to \ell$- and $\mathcal{R} \cdot other \to \ell$-fakes and the $j \to \ell$- and $other \to \ell$-fakes.

2. The background prediction in the new bin must be larger than in the previous bins combined.

3. The relative statistical uncertainty of all single processes in the new bin must be smaller than 100 %.

The binning procedure is performed separately for the distributions in the $e\tau$- and $\mu\tau$-dataset. For the rebinning of the distributions of the combined *VBF NN*s, the $j \to \ell$- and $other \to \ell$-fakes and the $\mathcal{R} \cdot j \to \ell$- and $\mathcal{R} \cdot other \to \ell$-fakes are combined into one process, respectively.

The resulting distributions are shown in Figure 8.44. The bins to the right, where the signal accumulates, are finer than the bins more to the left, as expected. The right-most bins are a bit wider again due to the requirement on the relative statistical uncertainty. The left-most bins, especially for the *nonVBF NN*, are thinner as they comprise the remaining events. The signal-enhanced bins are blinded whenever the signal-over-background ratio exceeds 5 % when assuming a branching ratio of 1 % for the signal. The data modelling by the background prediction using the *Symmetry Method* is very good in the remaining bins, indicating effective background estimation techniques.




**(a)** *nonVBF NN*, $e\tau$ **(b)** *nonVBF NN*, $\mu\tau$

**(c)** *VBF NN*, $e\tau$ **(d)** *VBF NN*, $\mu\tau$

**Figure 8.44:** Distributions of the *nonVBF NN* (top) and of the combined *VBF NN* (bottom), separately for the $e\tau$- (left) and $\mu\tau$-dataset (right), with the binning used in the statistical analysis. The $e/\mu$-symmetric background is estimated with the *Symmetry Method*. The uncertainty band includes the statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature.

### 8.2.7 Precursory Studies

This section presents the results of three studies which, among others, led to the results presented above. The studies were performed at different earlier stages of the analysis and hence various aspects, for instance the background estimation, the event selection, the considered uncertainties as well as the input variables, may differ. Therefore, only setups that are comparable, i.e. arose from similar stages of the analysis, are used for comparing sensitivities within one study.

#### Comparison of one multiclass and three binary NNs for the nonVBF SR

The signal node probability distribution of a multiclass NN is eventually used for the *nonVBF NN* as discussed above. However, the same strategy as finally chosen for the *VBF NN* was also tested for the *nonVBF NN* and its results are presented in the following.

Three binary NNs are trained and the signal node probabilities of all three are linearly combined to obtain one probability value per event (Eq. 8.11). The classes of the networks are the same as for the single *VBF NN*s. For each of the three binary networks an *Optuna* optimisation is performed and the best performing trial of each optimisation is chosen. The optimal values of the coefficients $c_i$ are obtained according to Eq. 8.11 in order to obtain one final discriminant.

The expected limit on the branching ratio of the signal is evaluated based on the *pre-fit Asimov* dataset. Details can be found in Section 10.3. This limit is compared to the expected limit obtained from the signal node probability distribution of a multiclass (three classes) NN. However, not the one presented above is used but one that was obtained at the same stage of the analysis as the binary NNs in order to have a fair as possible comparison. The expected limits are summarised in Table 8.10. The multiclass NN shows

| search | combined binary NNs | multiclass NN |
|---|---|---|
| $H \to e\tau_\mu$ | $0.272^{+0.105}_{-0.076}$ | $0.234^{+0.093}_{-0.065}$ |
| $H \to \mu\tau_e$ | $0.259^{+0.105}_{-0.072}$ | $0.239^{+0.094}_{-0.067}$ |

**Table 8.10:** Expected limits on the branching ratios of the LFV signals in percent obtained with the *pre-fit Asimov* dataset.

better limits for both searches, $H \to e\tau_\mu$ and $H \to \mu\tau_e$. The limit is improved by 14 and 8 %, respectively.

#### Investigating the impact of higher class weights for the background classes in the VBF SR

All NNs presented above are trained with an equal weighting of the single classes, i.e. all classes are granted the same importance. An additional parameter to be optimised was added in one hyperparameter optimisation with *Optuna*, allowing class weights larger than 1 for the background classes. This background class weight can adopt any value in $\{1, 2, 3, 5, 7, 10, 15, 20, 30, 50\}$ with respect to the signal class weight. All best performing trials favoured large values for the background class weights.

In an optimisation for the *VBF NN* with two classes, a signal class consisting of the ggF- and VBF-signal and a background class consisting of all background processes, the best performing trial used 50 for the background class weight. The signal node probability distributions obtained for this trial are shown in Figure 8.45. Two additional NNs were trained with the same hyperparameters but once the background class value is replaced by

25 and once by 1. The corresponding signal node probability distributions are shown in Figure 8.45 as well.

In the distributions with the background class weights of 50 and 25 all but the first bin are blinded. The background processes as well as the signal processes accumulate more to the left compared to the distributions with a background class weight of 1. Although the signal distributions decrease from left to right and the signal yield in the bins to the right is small, the sensitivity is considerably improved due to the absence or strong reduction of background events in these bins. The expected limits obtained with the *pre-fit Asimov* dataset (Section 10.3) are summarised in Table 8.11.

| search | background class weight | | |
|---|---|---|---|
| | 50 | 25 | 1 |
| $H \to e\tau_\mu$ | $0.212^{+0.096}_{-0.059}$ | $0.250^{+0.112}_{-0.070}$ | $0.495^{+0.195}_{-0.138}$ |
| $H \to \mu\tau_e$ | $0.303^{+0.127}_{-0.085}$ | $0.346^{+0.143}_{-0.097}$ | $0.508^{+0.202}_{-0.142}$ |

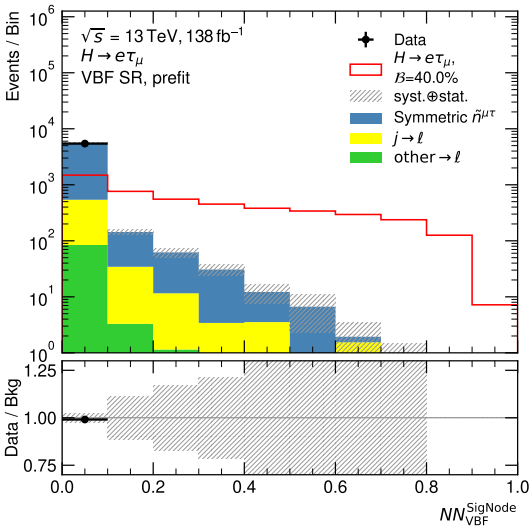
**Table 8.11:** Expected limits on the branching ratios of the LFV signals in percent obtained with the *pre-fit Asimov* dataset in the nonVBF SR.

The expected limits are improved by approximately 50 % with a background class weight of 50 compared to 1. However, the statistical analysis requires that each process contributes in each bin. Thus, if a process has no yield in one bin, it is artificially set to $10^{-6}$. When adding systematic uncertainties they also must be treated properly which is not straightforward. A correct solution to this problem is rebinning, i.e. merging bins from the right until each process sufficiently contributes. For the distribution of the $e\tau$-dataset and of the NN with a background class weight of 50 (Figure 8.45(a)) there would be one large bin from 0.2 to 1.0. This would inevitably lead to a reduced sensitivity and suggests that a similar behaviour can also be achieved by having an optimised and hence finer binning for higher values of the signal node probability distribution with a background class weight of 1. Therefore, it was decided to proceed with an equal weighting of all classes and to apply an optimised binning to the final discriminants. This was outlined in Section 8.2.6.

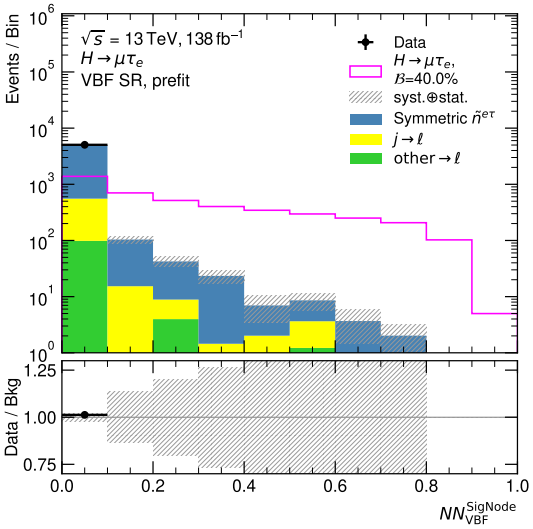
### Comparison of only low-level and high-level input variables for the nonVBF SR

In order to study the power of only using low-level variables, this section compares the expected limits obtained with the *pre-fit Asimov* dataset (see Section 10.3) of one NN trained with only low-level input variables and one NN that incorporates high-level variables. This comparison was performed at an early stage of the analysis which means that among other aspects also the set of the high-level input variables differed considerably from the set introduced above. This includes that the variables were defined in terms of $\ell_0$ and $\ell_1$ instead of $\ell_H$ and $\ell_\tau$(more details can be found in Section 7.1.3).

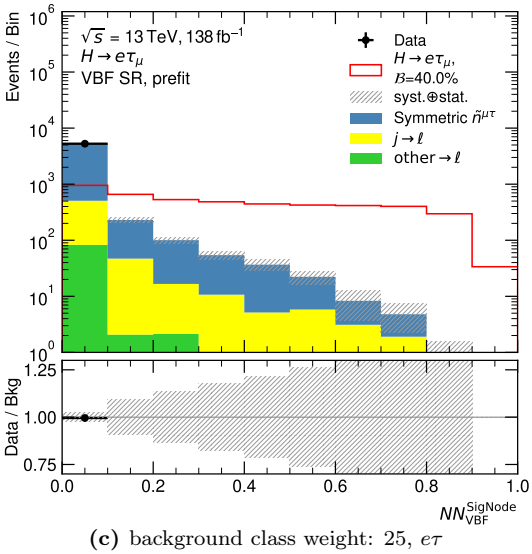
The set of low-level variables consists of 7 variables. These are the transverse momenta of the two leptons, their pseudorapidities, the absolute value of the missing transverse energy and the differences in the azimuthal angle of each of the two leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$. The set of low- and high-level variables consists of 22 variables. The architecture of the NN trained with the set of low-level variables was slightly extended compared to the NN trained with high-level variables to allow for more complex patterns potentially adopted by the network. The expected limits obtained with the *pre-fit Asimov* dataset are summarised in Table 8.12. The limits are obtained without considering any systematic or statistical uncertainties on the predictions. The sensitivity of both networks is very close, the NN trained with the set of low-level input variables performs only slightly worse. Nevertheless,



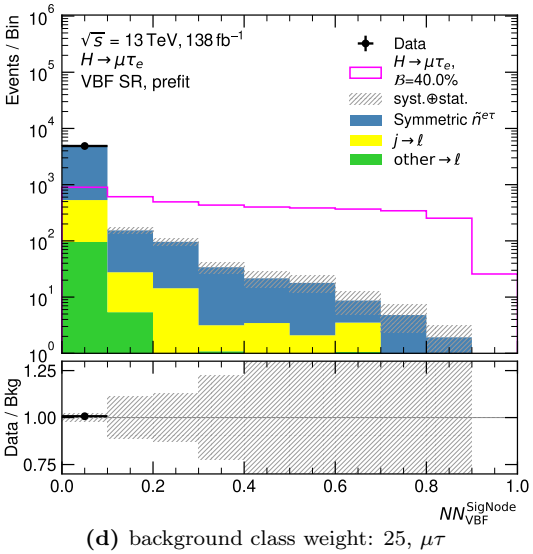

**(a)** background class weight: 50, $e\tau$



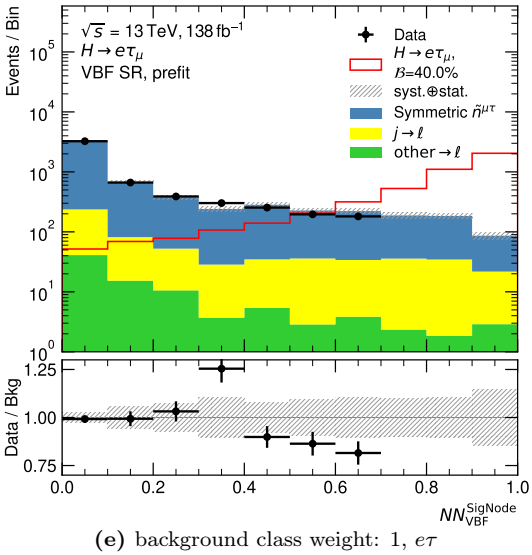

**(b)** background class weight: 50, $\mu\tau$



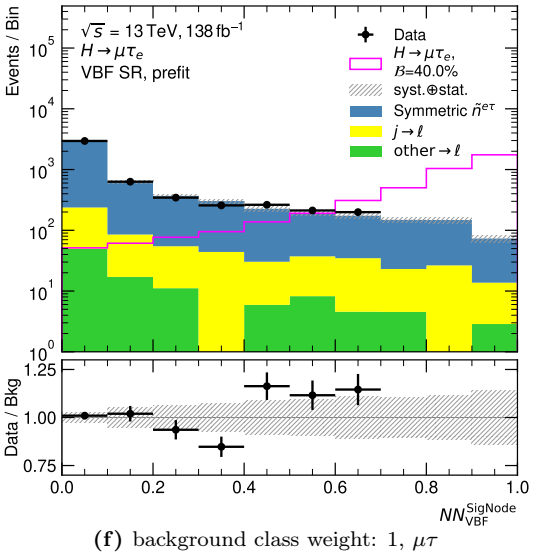

**(c)** background class weight: 25, $e\tau$




**(d)** background class weight: 25, $\mu\tau$




**(e)** background class weight: 1, $e\tau$




**(f)** background class weight: 1, $\mu\tau$

**Figure 8.45:** Distributions of the signal node probabilities of the *VBF NN*s with separate weights for the background class with respect to the signal class, separately for the $e\tau$- (left) and $\mu\tau$-dataset (right), obtained with the *test set*. The $e/\mu$-symmetric background is estimated with the *Symmetry Method*. The uncertainty band includes the statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature.

| search | set of input variables | |
|---|---|---|
| | high-level | low-level |
| $H \to e\tau_\mu$ | 0.107 | 0.113 |
| $H \to \mu\tau_e$ | 0.122 | 0.127 |

**Table 8.12:** Expected limits on the branching ratios of the LFV signals in percent obtained with the *pre-fit Asimov* dataset in the nonVBF SR.

it was decided to proceed with sets of input features that include high-level variables due to the expectation that they provide more room for improvement in further optimisations. In addition, some low-level variables like the pseudorapidities $\eta$ of the leptons show small issues in their modelling which, however, do not propagate to other high-level observables they are part of.

# Chapter 9

# Systematic Uncertainties

The different signal and background predictions come with experimental uncertainties, uncertainties of the estimation techniques and uncertainties on the theoretical predictions.

Processes estimated from MC simulations are subject to experimental uncertainties. In the $H$-LFV-analysis these are the signal processes and the *other* $\rightarrow \ell$-fakes contribution to the background. In addition, theoretical uncertainties on the signal processes are considered. They are neglected for the *other* $\rightarrow \ell$-fakes. Instead, an uncertainty on the normalisation of the *other* $\rightarrow \ell$-fakes is considered, derived in a validation region.

The $j \rightarrow \ell$-fakes and the $e/\mu$-symmetric background are mainly estimated from data and hence are neither subject to experimental nor to theoretical uncertainties. However, the methods to estimate these background components come with several uncertainty sources.

The single contributions to the experimental, theoretical and estimation technique uncertainties are described in Sections 9.1, 9.2 and 9.3, respectively. Section 9.4 discusses the statistical uncertainties on the predictions which are also treated as systematic uncertainties. The last Section 9.5 summarises preprocessing steps applied to the systematic uncertainties before the statistical analysis is performed.

An overview of the effects of the systematic uncertainties on the normalisation, i.e. yields, of the individual processes after the preprocessing steps is given in Table 9.1 for the $e\tau$-dataset in the search for $H \rightarrow e\tau$ and in Table 9.2 for the $\mu\tau$-dataset in the search for $H \rightarrow \mu\tau$ at the end of Section 9.3. The individual uncertainty sources are combined into groups. The effects are listed separately for the two signal regions and for all processes considered individually in the statistical analysis.

In order to propagate the effects of the individual uncertainty sources to the final discriminants, which are used in the statistical analysis to derive the results, each uncertainty source is once varied by $+1\sigma$ and once by $-1\sigma$ (standard deviation) and the final discriminant is re-calculated for each variation. The acceptance (normalisation) as well as the shape of the distribution can be altered. Each uncertainty source is treated as correlated between all processes it acts on and between both signal regions. Whereas the effects of different uncertainty sources are treated fully uncorrelated before the statistical analysis and thus are added in quadrature of which the square root is taken to obtain the total uncertainties.

The effects of the individual uncertainties are represented in the statistical analysis by so-called *nuisance parameter*s (NPs). Their names are listed in the following sections in *italic* and correspond to the names used in tables and figures related to the statistical analysis in Chapter 11. In addition, all nuisance parameters are listed in Appendix D.

In the following sections, numbers are quoted quantifying the effect of individual uncertainty sources on the different processes. Most often, the largest effects with respect to the individual processes and the two datasets, $e\tau$- and $\mu\tau$-datasets, are quoted. Typically, the uncertainty sources behave similarly for the $e\tau$- and $\mu\tau$-dataset. However, opposite behaviour is expected for the uncertainties related to the $j \rightarrow \ell$-fakes, the efficiency-

correction ratio $\mathcal{R}$ and the electron and muon scale factors. This is in all cases attributed to the fact that in the $e\tau$-dataset mainly muons are sub-leading with respect to $p_{\mathrm{T}}$, while in the $\mu\tau$-dataset these are mainly electrons. In addition, the reconstruction and identification efficiencies as well as the probabilities for other objects to fake leptons depend on the flavour and kinematics of the lepton.

## 9.1 Experimental Uncertainties

The experimental uncertainties of the different physics objects discussed in the following are uncertainties related to the reconstruction and identification procedures addressed in Chapter 5. They comprise uncertainties on the energy scale and resolution and on the *scale factor*s ($SF$s) which are simulation-to-data correction factors .

First, the uncertainties related to electrons and muons are discussed, followed by the uncertainties on jets and on $E_{\mathrm{T}}^{\mathrm{miss}}$. In addition, the uncertainty on the integrated luminosity is presented.

These uncertainties are applied to simulated events and thus act on the *other* $\rightarrow \ell$-fakes and on signal in the $H$-LFV analysis. All of these uncertainties are determined in separate auxiliary measurements performed centrally by the ATLAS collaboration which provides these uncertainties to all analyses.

**Electrons**

A list of electron related uncertainties with a brief description is given below. More details can be found in [180].

Uncertainties on the energy scale and resolution depend on the transverse energy and the pseudorapidity and are parametrised in bins of the latter [184].

There are different sources contributing to the scale uncertainty, e.g. limited knowledge of dead detector material, non-linearity of the cell energy measurements and modelling of the shower shape. Overall, there are 69 independent contributions corresponding to different uncertainty sources in specific decorrelated regions in pseudorapidity. For this analysis, however, the energy scale is known well enough and its uncertainties impact the analysis only very little. Thus, a simplified scheme is used which results in just one nuisance parameter for the energy scale. This is achieved by taking the square root of the individual sources added in quadrature. The resulting total uncertainty is treated as correlated over all regions in pseudorapidity which results in an overall larger uncertainty. The effect of the electron energy scale on the normalisation of the individual processes is below 1 %.

There are 9 independent contributions to the uncertainty of the energy resolution which, amongst others, come from limited knowledge of shower fluctuations, noise of the electronics and energy loss before the calorimeters. Also here the simplified scheme is used that results in just one nuisance parameter for the energy resolution. The effect of the electron energy resolution on the normalisation of the individual processes is well below 1 %.

The effects of both, energy scale and resolution, are smaller than the statistical uncertainties for the *other* $\rightarrow \ell$-fakes and signal. The names in the statistical analysis are:

- energy scale uncertainty: $e\gamma$ *scale*
- energy resolution uncertainty: $e\gamma$ *resolution*

Other sources of uncertainties for electrons are uncertainties on the different $SF$s for trigger, reconstruction, identification and isolation [185]. More details on their determination and uncertainties were discussed in Section 7.2.3. The nuisance parameter names are:

- uncertainty on trigger $SF$: *El Trigger SF*

- uncertainty on reconstruction $SF$: *El Reco Eff*
- uncertainty on identification $SF$: *El Id Eff Uncorr NP0-11*, *El Id Eff Corr NP0-15*
- uncertainty on isolation $SF$: *El Iso Eff*

For all $SF$s excluding the identification $SF$s, a simplified scheme is used which results in just one nuisance parameter per $SF$. All systematic uncertainty sources as well as the statistical uncertainties are added in quadrature and are correlated across all $p_{\mathrm{T}}$- and $\eta$-bins.

When using the simplified scheme, the single nuisance parameter for the identification $SF$ showed a relatively strong constraint when performing a fit with *pre-fit Asimov* data (Section 10.3). In such cases, the simplified scheme is not justified anymore and a more complex scheme must be used. In the $H$-LFV analysis it consists of 12 nuisance parameters (*El Id Eff Uncorr NP0-11*) representing the statistical uncertainties in different bins of $p_{\mathrm{T}}$ and $\eta$: *NP0-5* cover the barrel region in $\eta$ and *NP6-11* the endcap regions. The 6 bins in $p_{\mathrm{T}}$ correspond in increasing order to $15 \leq p_{\mathrm{T}} < 20\,\mathrm{GeV}$, $20 \leq p_{\mathrm{T}} < 25\,\mathrm{GeV}$, $25 \leq p_{\mathrm{T}} < 30\,\mathrm{GeV}$, $30 \leq p_{\mathrm{T}} < 60\,\mathrm{GeV}$, $60 \leq p_{\mathrm{T}} < 80\,\mathrm{GeV}$ and $\geq 80\,\mathrm{GeV}$. There are 16 additional nuisance parameters (*El Id Eff Corr NP0-15*) representing systematic uncertainties. A decomposition of the full covariance matrix of the systematic uncertainties is performed to obtain these 16 uncorrelated sources.

The impact of the uncertainties of the electron $SF$s is small, at most 0.7 % on the normalisation.

These electron $SF$ nuisance parameters acting on the *other* $\rightarrow \ell$-fakes and signal are correlated with the respective nuisance parameters representing the uncertainties on the electron $SF$s used in the efficiency-correction ratio (Section 9.3).

### Muons

Uncertainties related to the muon reconstruction and identification are listed below with brief descriptions. Detailed information can be found in [188]. Similar to electrons, there are uncertainties on the momentum scale and resolution and on the trigger, reconstruction and isolation *scale factor*s.

- track resolution uncertainty in the *Inner Detector*: *Muon ID*
- track resolution uncertainty in the *Muon System*: *Muon MS*
- momentum scale uncertainty: *Muon scale*
- charge-dependent momentum scale correction uncertainties: *Muon Sagitta ρ*, *Muon Sagitta res.bias*
- statistical and systematic uncertainty on trigger $SF$: *Mu Trigger SF Stat*, *Mu Trigger SF Sys*
- statistical and systematic uncertainty on reconstruction $SF$: *Mu Reco Eff Stat*, *Mu Reco Eff Sys*
- statistical and systematic uncertainty on isolation $SF$: *Mu Iso Eff Stat*, *Mu Iso Eff Sys*

Either a correction of a residual momentum scale bias between positively and negatively charged muons can be applied to data, or additional uncertainties on simulated events can be applied. The latter is done for the $H$-LFV analysis and represented by *Muon Sagitta ρ* and *Muon Sagitta res.bias*. The effect of the momentum scale, resolution and correction uncertainties in single bins of the final discriminants is at maximum 3 % and 1.5 % on the overall normalisation.

More details on the uncertainties related to the $SF$s can be found in Section 7.2.3. The largest effect of these uncertainties in single bins of the final discriminant is 2 % and 1.6 % on the overall normalisation.

All muon related uncertainties are smaller than the statistical uncertainties of the respective processes in almost all bins.

**Jets**

For jets, uncertainties on the energy scale and resolution as well as on the flavour composition are considered. In the case of jet energy scale (JES) and jet energy resolution (JER), eigenvector decompositions of the covariance matrices of the uncertainty sources are performed in order to reduce the number of nuisance parameters. More information can be found in [195]. In the $H$-LFV analysis 31 JES-related and 14 JER-related nuisance parameters and 4 nuisance parameters related to jet flavour are used. For the latter, separate nuisance parameters for both of the signal regions (nonVBF SR and VBF SR) are used in accordance with the partner analyses [22].

- jet energy scale:
  - *Jet effectiveNP Stat. 1-6*
  - *Jet effectiveNP mixed 1-3*
  - *Jet effectiveNP Modell. 1-4*
  - *Jet effectiveNP Detect. 1-2*
  - *Jet $\eta$-intercalibr. modelling*
  - *Jet $\eta$-intercalibr. non-closure highE*
  - *Jet $\eta$-intercalibr. non-closure neg.$\eta$*
  - *Jet $\eta$-intercalibr. non-closure pos.$\eta$*
  - *Jet $\eta$-intercalibr. non-closure total stat*
  - *Jet pileup offset $\mu$*
  - *Jet pileup offset npv*
  - *Jet pileup $pT$ term*
  - *Jet pileup $\rho$ topology*
  - *Jet punch through mc16*
  - *Jet single particle highPt*
  - *Jet b-JES response*
- jet energy resolution:
  - *Jet JER effectiveNP 1-11*
  - *Jet JER effectiveNP 12restTerm*
  - *Jet JER DataVsMC*
- jet flavour:
  - *JET Flavour Comp nonVBF*
  - *JET Flavour Comp VBF*
  - *JET Flavour Resp nonVBF*
  - *JET Flavour Resp VBF*

A few of the JES related nuisance parameters are almost fully pruned (the criteria are described in Section 9.5). The nuisance parameters contributing the most to the normalisation of the final discriminants are *Jet $\eta$-intercalibr. modelling* with up to 9 %, *Jet pileup offset $\mu$* with up to 7 %, *Jet pileup $\rho$ topology* with up to 6 % and *Jet effectiveNP Modell. 1* with up to 4 %. The quoted numbers refer to the effects in the VBF SR as the jet uncertainties are larger in this signal region, as expected.

The same applies to the JER related nuisance parameters. The nuisance parameter contributing the most on the normalisation of the single processes in the final discriminants is the *Jet JER EffectiveNP 1* with up to 6 %.

The nuisance parameters related to jet flavor have a larger effect on the final discriminants. *JET Flavour Comp* has a normalisation effect on the single processes of up to 15 % and *JET Flavour Resp* of up to 6 %.

Furthermore, uncertainties on the *scale factor*s of the (forward) jet-vertex taggers (JVT and fJVT) [192, 193] are considered. In the statistical analysis they are denoted by:

- *Jet JVT Efficiency*
- *Jet fJVT Efficiency*

Both nuisance parameters have effects on the normalisation of the single processes which are less than 1 %.

Additionally, uncertainties on the *b*-jet tagging efficiencies are considered [197, 243], which are uncertainties on the corresponding *scale factor*s, based on an eigenvector decomposition.

- uncertainties on the *b*-jet tagging efficiencies:
  - *b-tag Eigenvar. 0-2*
  - *c-tag Eigenvar. 0-2*
  - *light-tag Eigenvar. 0-3*
  - *b-tag Extrapolation*
  - *b-tag Extrapolation from c*

On one hand there are $SF$s for the efficiencies to tag *b*-jets as such and on the other hand there are scale factors on the rate of mistakenly tagging *c*- or light jets as *b*-jets. Each of these $SF$s have various uncertainty sources. The number of resulting nuisance parameters is reduced by performing an eigenvalue decomposition of the total covariance matrix of these systematic and statistical uncertainties. Two additional nuisance parameters for the extrapolation to high-$p_{\mathrm{T}}$ jets and for the extrapolation of $SF$s for *c*-jets to hadronic $\tau$-lepton jets are considered. Aside from *b-tag Eigenvar. 0*, *c-tag Eigenvar. 0*, *light-tag Eigenvar. 0* and *b-tag Extrapolation from c*, the remaining nuisance parameters are pruned. For them, the maximal normalisation effects are 0.3 %, 0.9 %, 0.8 % and 1 %, respectively.

Overall, the jet related uncertainties have larger contributions in the VBF SR than in the nonVBF SR, as expected.

**$\boldsymbol{E_{\mathrm{T}}^{\mathrm{miss}}}$**

The uncertainties on $E_{\mathrm{T}}^{\mathrm{miss}}$ [203] are represented by three nuisance parameters:

- energy scale of the soft term:
  - *met soft track scale*
- energy resolution of the soft term:
  - *met soft track res.paral.*
  - *met soft track res.perp.*

Separate uncertainties on $E_{\mathrm{T}}^{\mathrm{miss}}$ are only considered for the soft term as the other objects contributing to the calculation of $E_{\mathrm{T}}^{\mathrm{miss}}$ (hard objects) have their own systematic uncertainties which are propagated to the calculation of $E_{\mathrm{T}}^{\mathrm{miss}}$ within the analysis. The uncertainties on the soft term are obtained by varying its energy scale and resolution. For the resolution, independent variations parallel and perpendicular to the direction of the hard objects' $p_{\mathrm{T}}$ are considered.

The $E_{\mathrm{T}}^{\mathrm{miss}}$-related nuisance parameters have a moderate impact in this analysis as they result in sizeable variations of the shape of the final discriminants and show up to 6 % variation in the signal-like bins.

#### Luminosity

The integrated luminosity determination is based on the LUCID2 detector [115]. Its combined uncertainty for the Run 2 dataset is 1.7 % [113] of which the largest single uncertainty source is from the so-called calibration transfer. The corresponding nuisance parameter is called

- *Luminosity*

in the statistical analysis and considered as a normalisation uncertainty on the prediction of signal and *other* $\rightarrow \ell$-fakes.

#### Scaling Uncertainty on *other* $\rightarrow \ell$-fakes

As described in Section 7.2.2, a normalisation uncertainty on $Z \rightarrow \mu\mu$-events where one muon fakes an electron is derived in a *Validation Region*. It has a value of 16 % and is applied as an overall normalisation uncertainty to all *other* $\rightarrow \ell$-fakes in the statistical analysis and is denoted by:

- *MCfakesScalingUnc*

## 9.2 Uncertainties on the Theoretical Predictions of the Signal

In the simulation of LFV signal events, the production of the Higgs bosons is assumed to follow standard model predictions. These theoretical predictions come with various sources of uncertainties which include QCD scale uncertainties, uncertainties on the parton-density functions and on the strong coupling constant $\alpha_s$. Additional uncertainties are obtained by varying the generators used for the matrix element calculation and the parton shower and hadronisation model. All uncertainty sources apart from the pdf and $\alpha_s$ uncertainties are treated as uncorrelated between the different production processes.

By having specific requirements on the jet multiplicity (in the definition of the two *Signal Region*s and indirectly in the NN output node probability distributions), additional theoretical uncertainties are introduced.

Perturbative uncertainties, i.e. uncertainties due to higher orders in perturbation theory for the ggF signal are computed following the recommendations in [51, 244]. This results in 9 independent sources. The first two correspond to uncertainties on the cross sections as a function of the jet multiplicity, obtained by varying the renormalisation and factorisation scales for the first and the resummation scale for the second. The fourth and fith source take care of migration uncertainties between different jet multiplicities [245]. Source five and six represent uncertainties on the acceptance of ggF events in the VBF phase space when requiring two and at least three jets, respectively. Uncertainties on the shape of the transverse momentum of the Higgs boson are considered in the seventh and eighth source. The last source takes into account uncertainties due to assumptions on the top quark mass.

- QCD scale uncertainties on the ggF production mode:
  - uncertainties on the cross section: *theory ggF qcd 0*
  - resummation uncertainties: *theory ggF qcd 1*
  - uncertainties from bin migration between 0 and 1 jet and between 1 and 2 jets: *theory ggF qcd 2/3*
  - uncertainties from the VBF topology with 2/3 jets: *theory ggF qcd 4/5*
  - uncertainties on the shape of the Higgs-boson $p_{\mathrm{T}}$: *theory ggF qcd 6/7*
  - uncertainty due to assumptions on the top quark mass: *theory ggF qcd 8*

The nuisance parameter *theory ggF qcd 0* has an impact on the acceptance of the ggF-signal of up to 5 % and is larger in the nonVBF SR than in the VBF SR. The effects of *theory ggF qcd 1-3* are smaller. *theory ggF qcd 4* impacts the acceptance of up to almost 12 % in the VBF SR, while its impact in the nonVBF SR is very small, as expected. Figure 9.1 shows the effect of *theory ggF qcd 0* and *theory ggF qcd 4* on the distribution of the final discriminants. The remaining nuisance parameters influence the acceptance below 4 %.

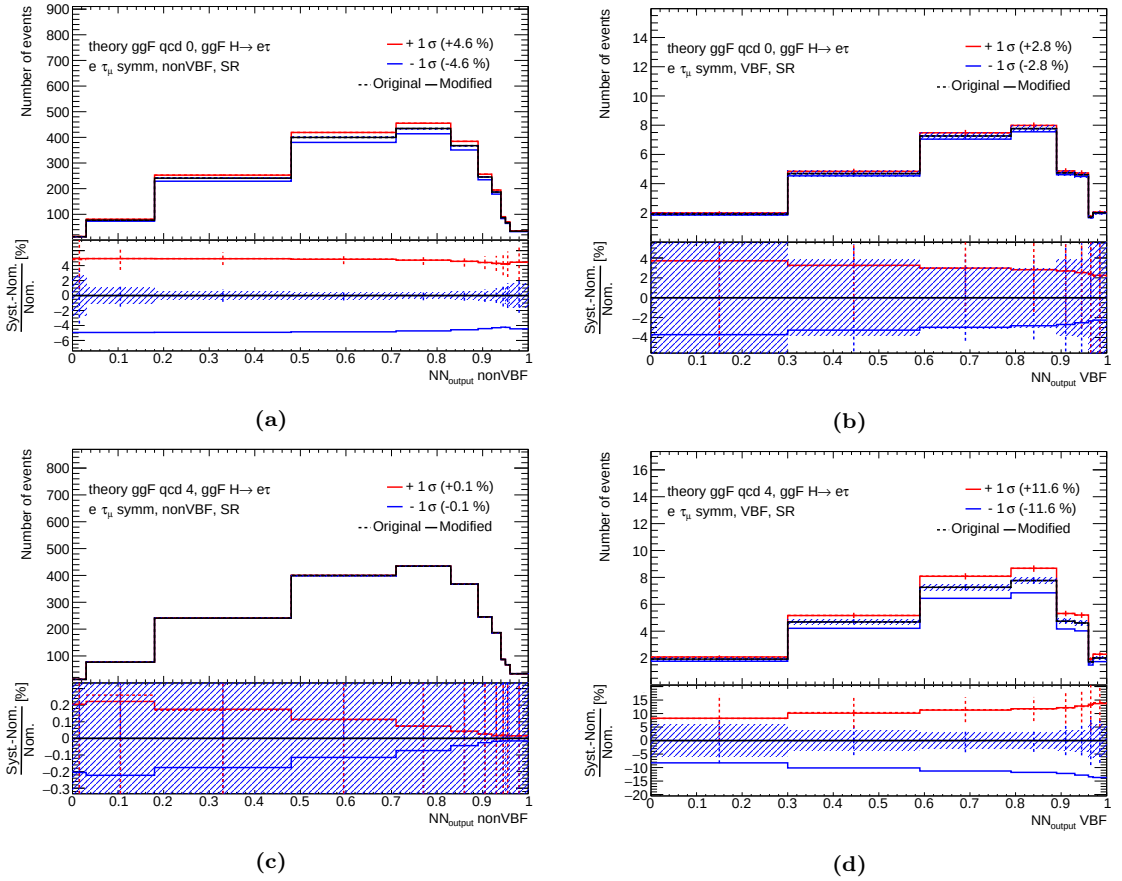



**Figure 9.1:** The effect of *theory ggF qcd 0* (a,b) and *theory ggF qcd 4* (c,d) on the distributions of the final discriminant in the nonVBF SR (a,c) and the VBF SR (b,d) for the $e\tau$-dataset. Black indicates the nominal distribution, red the up-variation and blue the down-variation. The dashed lines correspond to the original effects while the solid lines are after smoothing and symmetrisation (Section 9.5). The effect on the overall acceptance is quoted in parentheses.

For the VBF- and VH-signal, uncertainties on the cross section due to missing higher-order QCD corrections are obtained by varying the renormalisation ($\mu_R$) and factorisation ($\mu_F$) scales. The nominal values are chosen to be half the Higgs-boson mass: $\mu_R = \mu_F = m_H/2$. Both scales are varied by factors 1/2 and 2 which results in 9 variations. In each bin of the final discriminant distributions, the variation with the largest effect is chosen and a correlation over all bins is assumed (envelope). This results in one nuisance parameter per production process.

- QCD scale uncertainties on the VBF and VH production modes obtained by varying the renormalisation ($\mu_R$) and factorisation ($\mu_F$) scales:
  - *theory VBF qcd*
  - *theory VH qcd*

The impact on the acceptance of the VBF-signal is up to 2 % and on the VH-signal up to 12.8 %.

The uncertainties on the parton-density distribution functions (pdf) are obtained following the recommendations of the LHC4PDF collaboration [158]. They are represented by 30 eigenvector variations and two variations (up and down) of $\alpha_s$.

- pdf uncertainties:
  - *theory sig pdf 0-29*
- uncertainty on $\alpha_s$:
  - *theory sig* $\alpha_s$

The pdf variations are largely pruned. The largest contribution with an impact of 1 % on the acceptance of the ggF-signal in the VBF SR is from *theory sig pdf 4*. The impact of the uncertainty on $\alpha_s$ is at most 1.4 %.

Uncertainties due to the choice of the matrix element generator are estimated by comparing the prediction of the nominal generator (`Powheg-Box v2`) to the prediction of an alternative generator (`MadGraph5_aMCNLO5` [246]). In the statistical analysis, these variations are represented by the following nuisance parameters:

- matrix element uncertainties:
  - *theory ggF ME*
  - *theory VBF ME*
  - *theory VH ME*

The impact of *theory ggF ME* is visualised in Figure 9.2 since this nuisance parameter is one of the more important ones in the final statistical analysis. Sizeable effects on the shape of the final discriminant distribution are observed. The impact on the acceptance of *theory VBF ME* is up to 7.2 % in the VBF SR and a bit smaller in the nonVBF SR. The same pattern is observed for *theory VH ME* with an impact on the acceptance of up to 10 %.

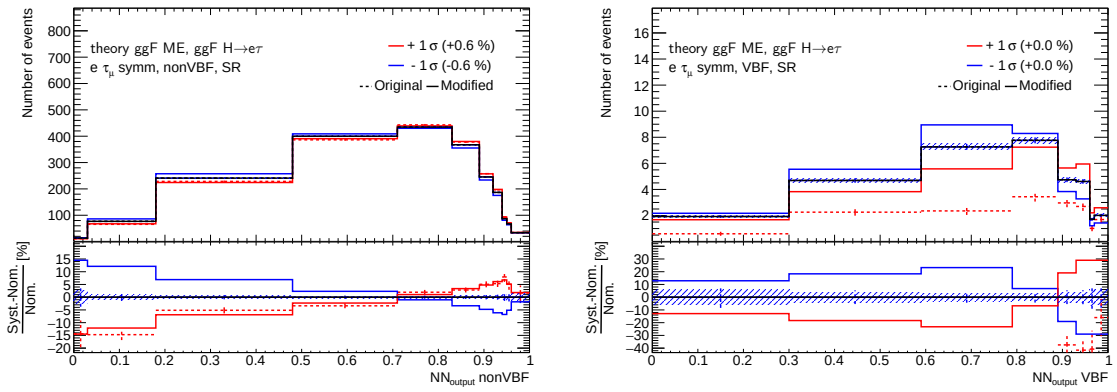



**Figure 9.2:** The effect of *theory ggF ME* on the distributions of the final discriminant in the nonVBF SR (a) and the VBF SR (b) for the $e\tau$-dataset. Black indicates the nominal distribution, red the up-variation and blue the down-variation. The dashed lines correspond to the original effects while the solid lines are after smoothing and symmetrisation (Section 9.5). The effect on the overall acceptance is quoted in parentheses. In the VBF SR only the shape effect is considered, hence the acceptance effect is quoted with 0 %.

In order to estimate uncertainties on the choice of the underlying event, parton shower and hadronisation model, the nominal algorithm (`Pythia 8.2`) is replaced by an alternative one (`Herwig 7` [247, 248]).

The following nuisance parameters describe the resulting variations:

- parton shower uncertainties:
  - *theory ggF PS*
  - *theory VBF PS*
  - *theory VH PS*

The impact of *theory ggF PS* on the normalisation of the ggF-signal is up to 10 % in the VBF SR. In the last, most signal-like, bin of the final discriminant *theory ggF PS* impacts the yield of the ggF-signal by up to 20 %. The same pattern is observed in the nonVBF SR, but with smaller effects. Stronger shape effects by *theory VBF PS* and *theory VH PS* are observed as well, with effects on the normalisation of up to 6 %.

## 9.3 Uncertainties of the Data-Driven Background Estimation Techniques

The uncertainties on the background estimation techniques consist of uncertainties on the $j \to \ell$-fake estimate and on the efficiency-correction ratio $\mathcal{R}$ used in the course of the *Symmetry Method*. Both are discussed in the following.

### Uncertainties on the $j \to \ell$-fake Estimate

Uncertainties on the $j \to \ell$-fake estimate consist of statistical and systematic uncertainties on the *Fake Factor*s (FFs) and *Correction Factor*s (CFs) which are part of the *Fake Factor Method* that is used to estimate the $j \to \ell$-fakes. A detailed discussion of the method as well as of the uncertainty sources is given in Section 7.2.2.

The *Fake Factor*s are calculated in the $Z$ + jets-*Extraction Region* (ER). The statistical uncertainties on the data yield as well as on the subtracted simulated prompt events are propagated to the FFs. The statistical uncertainty of each FF-bin is represented by a separate nuisance parameter in the statistical analysis to keep them uncorrelated a priori. Separate nuisance parameters for the electron and muon FFs are used.

- electron FF statistical uncertainties: *El. Fake FF Stat NP0-3, 8-11*
- muon FF statistical uncertainties: *Mu. Fake FF Stat NP0-2*

For the electron FFs, *NP0-3* and *NP8-11* are for small and large $\Delta\phi(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}})$, respectively. For each of the two $\Delta\phi(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}})$-regions, the *NP*s correspond to increasing $p_{\mathrm{T}}$: 15 to 20 GeV, 20 to 25 GeV, 25 to 35 GeV and $>$ 35 GeV. For the muon FFs, *NP0-2* denote three bins in $p_{\mathrm{T}}$ in increasing order: 15 to 20 GeV, 20 to 25 GeV and $>$ 25 GeV. The largest impact of the electron FF statistical uncertainties on the normalisation of the $j \to \ell$- and $\mathcal{R} \cdot j \to \ell$-fakes is from *El. Fake FF Stat NP8* with 1.5 %, followed by *El. Fake FF Stat NP11* with an impact of 1.3 %. The smallest is from *El. Fake FF Stat NP1* with 0.3 %. The $\mathcal{R} \cdot j \to \ell$-fakes are the $j \to \ell$-fake contribution from the respective other dataset ($e\tau$ or $\mu\tau$), where each event is weighted with the efficiency-correction ratio $\mathcal{R}$ (Section 7.2.1).

For the muon FFs, the number of single components which corresponds to the number of FF-bins is only 3 compared to 8 for the electron FFs, which helps to reduce the overall statistical uncertainties. Here, *Mu. Fake FF Stat NP0* contributes to the normalisation of the $j \to \ell$- or $\mathcal{R} \cdot j \to \ell$-fakes with 1.5 %, *Mu. Fake FF Stat NP1* with 0.8 % and *Mu. Fake FF Stat NP2* with 1.4 %. The shape effect of all *NP*s is entirely negligible in the VBF SR following the criterium described in Section 9.5.

The uncertainties on the normalisation of the $WZ$- and $ZZ$-background in the determination of the FFs are considered as systematic uncertainties, correlated over all FF-bins and over electron and muon FFs.

- FF systematic uncertainties:
  - *Fake WZxsec*
  - *Fake ZZxsec*

The impact on the acceptance of the two $j \to \ell$-fake contributions is up to 6.4 % for *Fake WZxsec* and 2.2 % for *Fake ZZxsec*. The shape effects are very small and again entirely negligible in the VBF SR.

The *Basic Selection*-SR and the $Z$ + jets-ER differ in the flavour composition of the jets faking leptons. This difference is accounted for by the additional application of *Correction Factor*s (CFs). They are determined from simulated events as the ratio of the FFs in the *Basic Selection*-SR over the ones in the $Z$ + jets-ER. The statistical uncertainties on the event yields in the respective regions are propagated to the CFs. They are represented by separate nuisance parameters per electron and muon CFs and separately for each CF-bin in order to keep them uncorrelated a priori.

- electron CF statistical uncertainties: *El. Fake CF Stat NP0-3*
- muon CF statistical uncertainties: *Mu. Fake CF Stat NP0-2*

The binning of the CFs for electrons is reduced compared to the FFs such that it is one-dimensional (in $p_{\mathrm{T}}$). The correspondence of the *NP* numbers and the $p_{\mathrm{T}}$-bins is the same as for the FFs. The impact on the acceptance of electron *NP*s ranges from 0.8 % (*El. Fake CF Stat NP3*) to 1.9 % (*El. Fake CF Stat NP0*). For the muon *NP*s, *Mu. Fake CF Stat NP0* has the largest impact with 1.5 %. The shape effects are pruned for all *NP*s in the VBF SR.

Systematic uncertainties on the CFs are derived by comparing their values to values obtained with an alternative MC event generator. This uncertainty is derived separately for electron and muon fakes as the mechanisms for jets to fake either of them differ.

- electron systematic uncertainties: *El. Fake CF Sys*
- muon systematic uncertainties: *Mu. Fake CF Sys*

The electron *NP* has a maximum overall impact of 13.2 % on the $j \to \ell$-fake contributions with up to 18 % for the most signal-like bin of the final discriminant. For the muon *NP* the maximum overall impact is 9.2 %. In the VBF SR, the shape effects are negligible. Figure 9.3 shows the impact of the two nuisance parameters on the distribution of the final discriminant in the nonVBF SR.

### Uncertainties on the Efficiency-Correction Ratio $\mathcal{R}$

The efficiency-correction ratio $\mathcal{R}$ is part of the estimation of the $e/\mu$-symmetric background described in Section 7.2.1. The $e/\mu$-symmetric background consists of three contributions: data, $j \to \ell$-fakes and *other* $\to \ell$-fakes. The latter two are subtracted from the former and each event is weighted with $\mathcal{R}$. The uncertainties on the $j \to \ell$- and *other* $\to \ell$-fakes, which were discussed in the previous sections, are propagated to the total $e/\mu$-symmetric background.

This section focuses on the uncertainty sources affecting $\mathcal{R}$. The denominator of the efficiency-correction ratio consists of the product of trigger, reconstruction, identification and isolation efficiencies of the electron and the muon in the event (*real event*). Whereas the numerator consists of the same product but with the kinematic properties of the electron and the muon interchanged (*switch event*).

The determination of the different efficiencies and their uncertainties are discussed in Section 7.2.3. The uncertainties of the *real* and *switch* event are treated correlated and hence are represented by the same nuisance parameters.

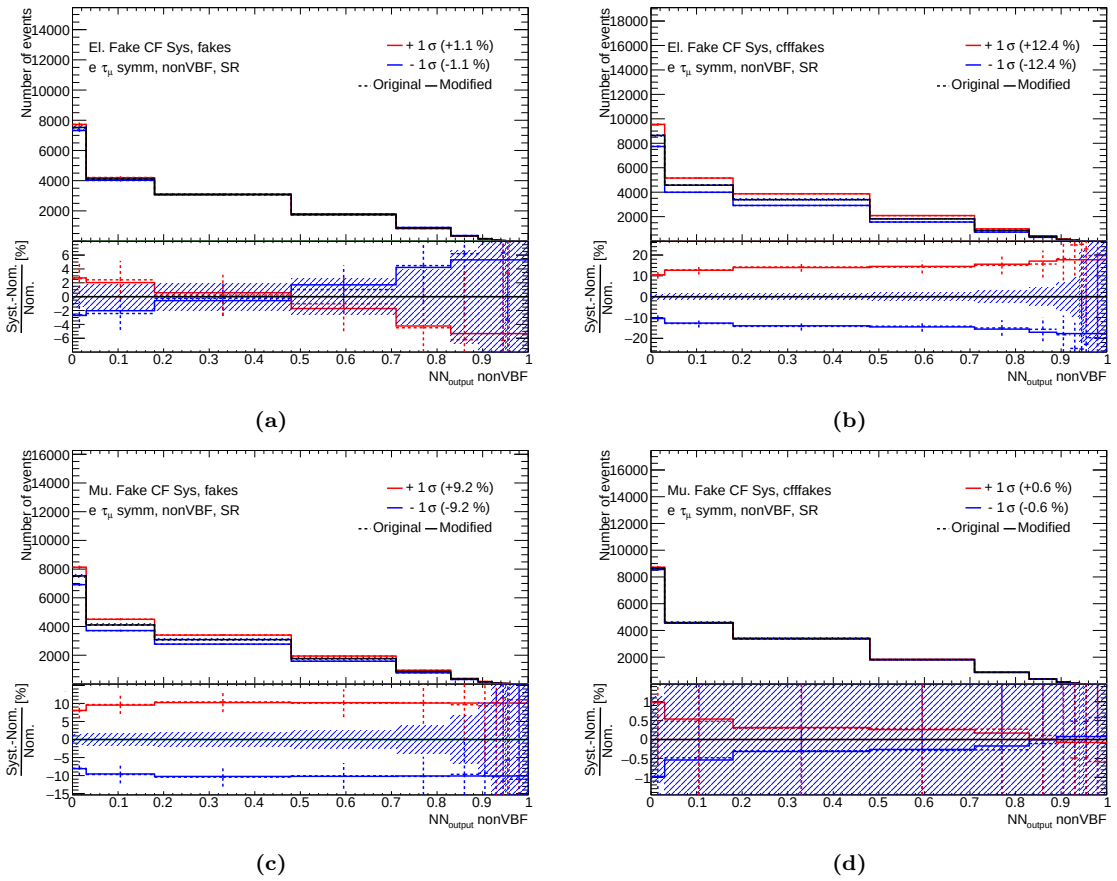



**Figure 9.3:** The effect of *El. Fake CF Sys* (a,b) and *Mu. Fake CF Sys* (c,d) on the distributions of the final discriminant for the $j \to \ell$-fakes (a,c) and the $\mathcal{R} \cdot j \to \ell$-fakes (b,d) in the nonVBF SR for the $e\tau$-dataset. Black indicates the nominal distribution, red the up-variation and blue the down-variation. The dashed lines correspond to the original effects while the solid lines are after smoothing and symmetrisation (Section 9.5). The effect on the overall acceptance is quoted in parentheses.

For each of the muon efficiencies, one nuisance parameter representing the statistical uncertainty sources and one representing the systematic sources are considered:

- muon trigger efficiency uncertainties: $\varepsilon$ *Mu Eff Trigger Stat* and *Sys*
- muon reconstruction efficiency uncertainties: $\varepsilon$ *Mu Eff Reco Stat* and *Sys*
- muon isolation efficiency uncertainties: $\varepsilon$ *Mu Eff Iso Stat* and *Sys*

Apart from $\varepsilon$ *Mu Eff Trigger Stat*, $\varepsilon$ *Mu Eff Iso Stat* and $\varepsilon$ *Mu Eff Iso Sys*, the remaining nuisance parameters are fully pruned. $\varepsilon$ *Mu Eff Trigger Stat* has the largest effect on the acceptance with up to 0.3 %.

The total (i.e. reconstruction, identification and isolation) electron offline efficiency consists of the product of the respective $SF$s and the total efficiency measured with MC-simulated events in the *Basic Selection*-SR (Eq. 7.15). These $SF$s and their uncertainties are the same as the ones described in Section 9.1. Hence, the uncertainties of the $SF$s in $\mathcal{R}$ are treated fully correlated with the $SF$s applied to signal and *other* $\to \ell$-fakes in Section 9.1. Therefore, they are represented by the same nuisance parameters:

- uncertainty on reconstruction $SF$: *El Reco Eff*
- uncertainty on identification $SF$:

– *El Id Eff Uncorr NP0-11* – *El Id Eff Corr NP0-15*

- uncertainty on isolation $SF$: *El Iso Eff*

Before correlating, *El Id Eff Uncorr NP0* has the largest impact on the acceptance, it is at most 1.6 %. Its impact on the distribution of the final discriminant before and after correlating is shown in Figure 9.4. The up- and down-variations are in opposite directions

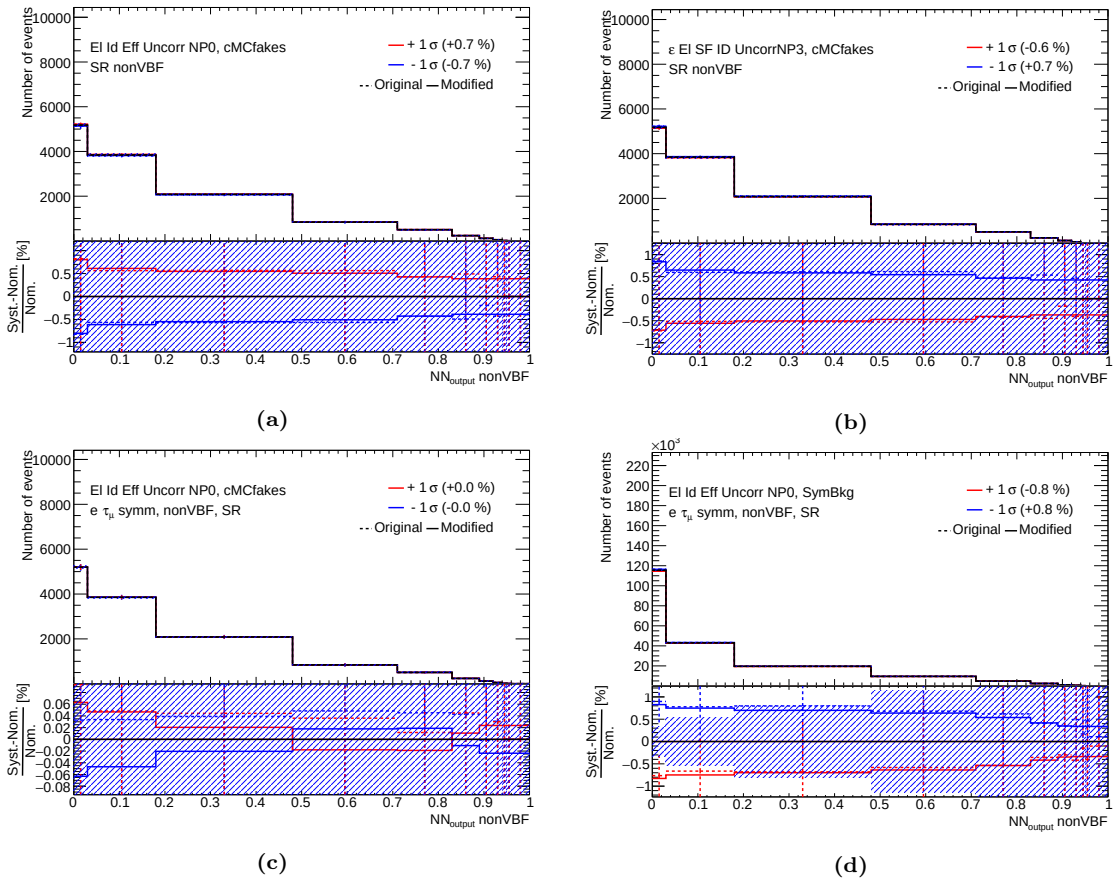



**Figure 9.4:** The effect of *El Id Eff Uncorr NP0* on the distribution of the final discriminant in the nonVBF SR and the $e\tau$-dataset for the $\mathcal{R} \cdot other \to \ell$-fakes contribution (a-c) and for the data part of the $e/\mu$-symmetric background (d). (a) shows the effect of the usual $SF$ (i.e. simulation-to-data correction factor) uncertainty, (b) for the uncertainty within the efficiency-correction ratio $\mathcal{R}$ and (c) for both combined. (d) includes only the uncertainty within $\mathcal{R}$ since no usual $SF$s are applied to data. Black indicates the nominal distribution, red the up-variation and blue the down-variation. The dashed lines correspond to the original effects while the solid lines are after smoothing and symmetrisation (Section 9.5). The effect on the overall acceptance is quoted in parentheses.

in Figures 9.4(a) and 9.4(b) and hence the combined effect (Figure 9.4(c)) is very small.

In addition, there are statistical as well as systematic uncertainties on the total efficiency measured in the *Basic Selection*-SR:

- electron offline MC efficiency uncertainties: $\varepsilon$ *El Eff MC Stat* and *Sys*

The impact of the statistical $NP$s on the acceptance of the individual processes that are weighted with $\mathcal{R}$ is at most 0.4 % and of the systematic $NP$s at most 0.9 %.

Finally, the uncertainty on the electron trigger efficiency is represented by the following nuisance parameter:

- electron trigger efficiency uncertainty: $\varepsilon$ *El Eff Trigger*

This nuisance parameter is pruned for all processes in both regions.

| | ggF | VBF | VH | $\mathcal{R}\cdot$ ggF | $\mathcal{R}\cdot$ VBF | $\mathcal{R}\cdot$ VH | $\mathcal{R}\cdot$ data | $\mathcal{R}\cdot$ $j \to \ell$-fakes | $\mathcal{R}\cdot$ *other* $\to \ell$-fakes | $j \to \ell$-fakes | *other* $\to \ell$-fakes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **nonVBF SR** | | | | | | | | | | | |
| Electrons and Muons | +1.2/−1.2 | +1.1/−1.1 | +1.0/−1.0 | +1.0/−1.0 | +0.9/−0.9 | +0.9/−1.1 | – | – | +1.3/−1.7 | – | +1.3/−1.3 |
| $E_{\mathrm{T}}^{\mathrm{miss}}$ | +0.4/−0.4 | +0.5/−0.5 | +0.3/−0.3 | +6.0/−6.0 | +5.1/−5.1 | +1.4/−1.4 | – | – | $< 0.1$ | – | +0.1/−0.1 |
| Jets | +1.0/−1.0 | +3.5/−3.5 | +1.5/−1.5 | +2.7/−2.7 | +4.8/−4.8 | +2.6/−2.6 | – | – | +0.8/−0.8 | – | +1.2/−1.2 |
| Flavour tagging | +0.3/−0.3 | +0.5/−0.5 | +1.2/−1.2 | +0.4/−0.4 | +0.5/−0.5 | +1.9/−1.9 | – | – | +0.3/−0.3 | – | +0.3/−0.3 |
| Luminosity | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | – | – | +1.7/−1.7 | – | +1.7/−1.7 |
| *other* $\to \ell$ normalisation | – | – | – | – | – | – | – | – | +16.0/−16.0 | – | +16.0/−16.0 |
| Signal theoretical unc. | +5.2/−5.2 | +4.8/−4.8 | +13.0/−13.0 | +8.3/−8.3 | +6.7/−6.7 | +12.7/−12.7 | – | – | – | – | – |
| Efficiency correction $\mathcal{R}$ | +0.2/−0.2 | +0.2/−0.2 | +0.3/−0.3 | +0.3/−0.3 | +0.3/−0.3 | +0.4/−0.4 | +1.1/−1.1 | +1.8/−1.8 | +1.1/−1.1 | – | +0.2/−0.2 |
| Electron FF stat. | – | – | – | – | – | – | – | +2.5/−2.5 | – | +0.3/−0.3 | – |
| Muon FF stat. | – | – | – | – | – | – | – | +0.6/−0.6 | – | +2.8/−2.8 | – |
| FF syst | – | – | – | – | – | – | – | +3.2/−3.5 | – | +4.9/−5.8 | – |
| Electron CF | – | – | – | – | – | – | – | +12.6/−12.6 | – | +1.2/−1.2 | – |
| Muon CF | – | – | – | – | – | – | – | +0.7/−0.7 | – | +9.4/−9.4 | – |
| **VBF SR** | | | | | | | | | | | |
| Electrons and Muons | +1.2/−1.2 | +1.0/−1.0 | +1.2/−1.2 | +1.8/−1.8 | +1.3/−1.3 | +1.4/−1.4 | – | – | +1.7/−2.0 | – | +2.2/−2.1 |
| $E_{\mathrm{T}}^{\mathrm{miss}}$ | +0.7/−0.7 | +0.3/−0.3 | +1.4/−1.4 | +5.5/−5.5 | +4.1/−4.1 | +6.4/−6.4 | – | – | +0.5/−0.5 | – | +1.3/−1.2 |
| Jets | +14.3/−14.3 | +4.7/−4.7 | +13.7/−13.7 | +20.3/−20.3 | +8.9/−8.9 | +22.9/−22.9 | – | – | +22.5/−22.6 | – | +18.9/−18.5 |
| Flavour tagging | +0.7/−0.7 | +0.5/−0.5 | +1.2/−1.2 | +0.8/−0.7 | +0.5/−0.5 | +1.4/−1.4 | – | – | +0.8/−0.8 | – | +0.8/−0.8 |
| Luminosity | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | – | – | +1.7/−1.7 | – | +1.7/−1.7 |
| *other* $\to \ell$ normalisation | – | – | – | – | – | – | – | – | +16.0/−16.0 | – | +16.0/−16.0 |
| Signal theoretical unc. | +15.3/−15.3 | +8.0/−8.0 | +12.0/−12.0 | +14.2/−14.2 | +7.4/−7.4 | +21.8/−21.8 | – | – | – | – | – |
| Efficiency correction $\mathcal{R}$ | +0.2/−0.2 | +0.3/−0.3 | +0.3/−0.3 | +0.3/−0.3 | +0.3/−0.3 | +0.5/−0.6 | +0.7/−0.8 | +1.5/−1.5 | +1.1/−1.1 | – | +0.5/−0.5 |
| Electron FF stat. | – | – | – | – | – | – | – | +2.6/−2.6 | – | +1.3/−1.3 | – |
| Muon FF stat. | – | – | – | – | – | – | – | +0.8/−0.8 | – | +2.8/−2.8 | – |
| FF syst | – | – | – | – | – | – | – | +3.8/−4.2 | – | +5.8/−6.5 | – |
| Electron CF | – | – | – | – | – | – | – | +13.4/−13.4 | – | +4.9/−4.9 | – |
| Muon CF | – | – | – | – | – | – | – | +1.3/−1.3 | – | +7.8/−7.8 | – |

**Table 9.1:** Relative impact of groups of uncertainties in percent on the normalisation of the individual signal and background processes (before the statistical analysis) in the search for $H \to e\tau$ ($e\tau$-dataset), separate for the up- and down-variations. The individual uncertainty sources of one group are added in quadrature and the square root is taken. The upper part shows the impact in the nonVBF SR, the lower part in the VBF SR. The signal is split in the different production processes (ggF, VBF, VH) and the signal which is wrongly classified into the $\mu\tau$-dataset must be considered, corrected with the efficiency ratio $\mathcal{R}$ (see Section 10.4). The "Efficiency correction $\mathcal{R}$" group includes the "usual" electron identification *SF* uncertainties in addition to the ones applied within $\mathcal{R}$. The abbreviations FF and CF stand for "fake factor" and "correction factor", respectively, which are part of the $j \to \ell$-fake estimate.

| | ggF | VBF | VH | $\mathcal{R}\cdot$ ggF | $\mathcal{R}\cdot$ VBF | $\mathcal{R}\cdot$ VH | $\mathcal{R}\cdot$ data | $\mathcal{R}\cdot$ $j \to \ell$-fakes | $\mathcal{R}\cdot$ *other* $\to \ell$-fakes | $j \to \ell$-fakes | *other* $\to \ell$-fakes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **nonVBF SR** | | | | | | | | | | | |
| Electrons and Muons | +0.9/−1.1 | +0.8/−1.0 | +0.8/−0.9 | +1.0/−1.0 | +0.9/−1.0 | +0.8/−0.9 | – | – | +1.3/−1.3 | – | +1.1/−1.3 |
| $E_{\mathrm{T}}^{\mathrm{miss}}$ | +0.4/−0.4 | +0.5/−0.5 | +0.1/−0.1 | +5.6/−5.6 | +4.9/−4.9 | +1.1/−1.1 | – | – | $< 0.1$ | – | $< 0.1$ |
| Jets | +0.8/−0.8 | +3.2/−3.2 | +1.1/−1.1 | +2.3/−2.3 | +5.2/−5.2 | +1.9/−1.9 | – | – | +0.9/−0.8 | – | +0.6/−0.6 |
| Flavour tagging | +0.3/−0.3 | +0.5/−0.5 | +1.2/−1.2 | +0.4/−0.4 | +0.5/−0.5 | +1.9/−1.9 | – | – | +0.3/−0.3 | – | +0.3/−0.3 |
| Luminosity | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | – | – | +1.7/−1.7 | – | +1.7/−1.7 |
| *other* $\to \ell$ normalisation | – | – | – | – | – | – | – | – | +16.0/−16.0 | – | +16.0/−16.0 |
| Signal theoretical unc. | +5.3/−5.3 | +7.1/−7.1 | +13.0/−13.0 | +8.1/−8.1 | +6.0/−6.0 | +13.0/−13.0 | – | – | – | – | – |
| Efficiency correction $\mathcal{R}$ | +1.0/−1.0 | +0.9/−0.9 | +0.8/−0.8 | +0.2/−0.2 | +0.3/−0.3 | +0.4/−0.4 | +0.9/−0.8 | +1.3/−1.3 | +0.9/−0.9 | – | +1.1/−1.1 |
| Electron FF stat. | – | – | – | – | – | – | – | +0.5/−0.5 | – | +2.5/−2.5 | – |
| Muon FF stat. | – | – | – | – | – | – | – | +2.8/−2.8 | – | +0.8/−0.8 | – |
| FF syst | – | – | – | – | – | – | – | +5.0/−5.9 | – | +3.7/−4.1 | – |
| Electron CF | – | – | – | – | – | – | – | +2.1/−2.1 | – | +13.1/−13.1 | – |
| Muon CF | – | – | – | – | – | – | – | +8.4/−8.4 | – | +1.1/−1.1 | – |
| **VBF SR** | | | | | | | | | | | |
| Electrons and Muons | +0.9/−0.9 | +0.7/−0.8 | +0.9/−1.1 | +1.5/−1.5 | +1.3/−1.4 | +1.2/−1.2 | – | – | +2.5/−2.5 | – | +1.7/−1.9 |
| $E_{\mathrm{T}}^{\mathrm{miss}}$ | +0.6/−0.6 | +0.4/−0.4 | +2.4/−2.4 | +3.9/−3.9 | +4.3/−4.3 | +4.9/−4.9 | – | – | +0.9/−0.9 | – | +0.7/−0.7 |
| Jets | +12.6/−12.6 | +4.9/−4.9 | +16.7/−16.7 | +17.9/−17.9 | +10.1/−10.1 | +18.6/−18.6 | – | – | +20.2/−20.2 | – | +23.0/−22.9 |
| Flavour tagging | +0.7/−0.7 | +0.5/−0.5 | +1.4/−1.4 | +0.8/−0.7 | +0.5/−0.5 | +1.6/−1.6 | – | – | +0.9/−0.9 | – | +0.8/−0.8 |
| Luminosity | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | +1.7/−1.7 | – | – | +1.7/−1.7 | – | +1.7/−1.7 |
| *other* $\to \ell$ normalisation | – | – | – | – | – | – | – | – | +16.0/−16.0 | – | +16.0/−16.0 |
| Signal theoretical unc. | +17.0/−17.0 | +5.8/−5.8 | +16.0/−16.0 | +15.1/−15.1 | +2.8/−2.8 | +34.1/−34.1 | – | – | – | – | – |
| Efficiency correction $\mathcal{R}$ | +0.8/−0.8 | +0.7/−0.7 | +0.9/−0.9 | +0.3/−0.3 | +0.3/−0.3 | +0.3/−0.3 | +0.4/−0.4 | +0.7/−0.7 | +0.8/−0.8 | – | +0.7/−0.7 |
| Electron FF stat. | – | – | – | – | – | – | – | +1.6/−1.6 | – | +2.4/−2.4 | – |
| Muon FF stat. | – | – | – | – | – | – | – | +2.8/−2.8 | – | +1.7/−1.7 | – |
| FF syst | – | – | – | – | – | – | – | +6.0/−6.7 | – | +5.3/−5.7 | – |
| Electron CF | – | – | – | – | – | – | – | +6.2/−6.2 | – | +12.9/−12.9 | – |
| Muon CF | – | – | – | – | – | – | – | +6.4/−6.4 | – | +2.1/−2.1 | – |

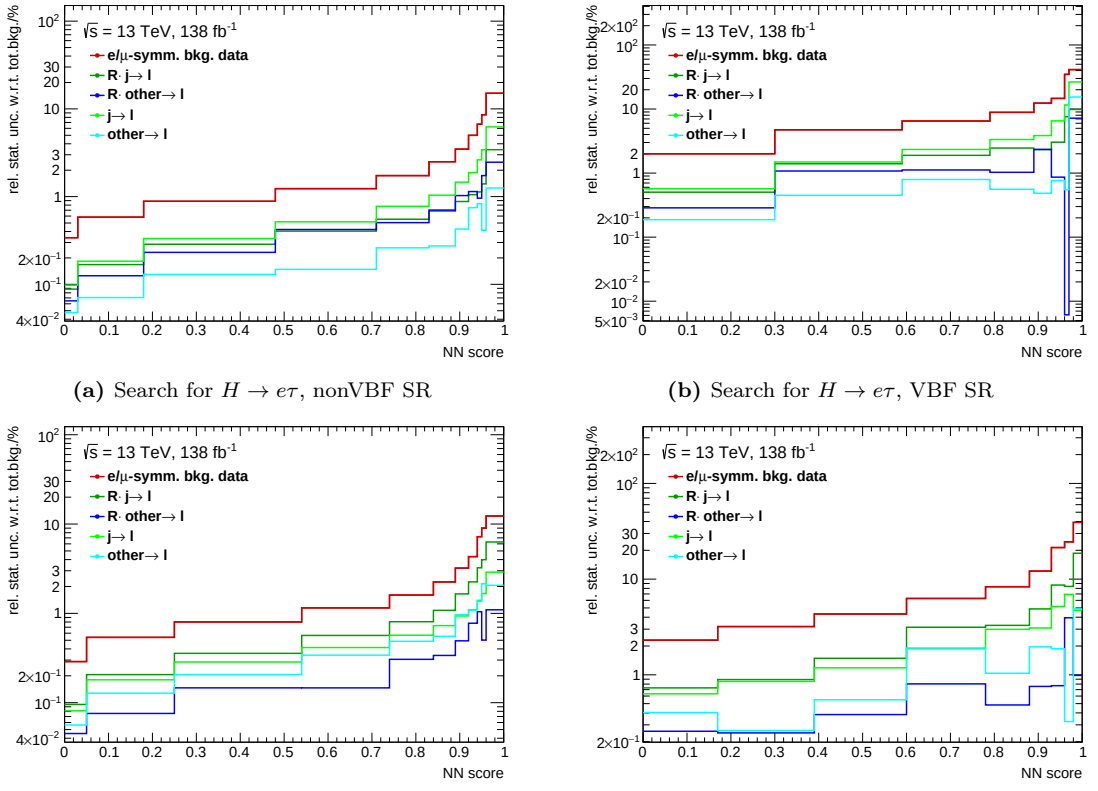
**Table 9.2:** Relative impact of groups of uncertainties in percent on the normalisation of the individual signal and background processes (before the statistical analysis) in the search for $H \to \mu\tau$ ($\mu\tau$-dataset), separate for the up- and down-variations. The individual uncertainty sources of one group are added in quadrature and the square root is taken. The upper part shows the impact in the nonVBF SR, the lower part in the VBF SR. The signal is split in the different production processes (ggF, VBF, VH) and the signal which is wrongly classified into the $e\tau$-dataset must be considered, corrected with the efficiency ratio $\mathcal{R}$ (see Section 10.4). The "Efficiency correction $\mathcal{R}$" group includes the "usual" electron identification $SF$ uncertainties in addition to the ones applied within the $\mathcal{R}$. The abbreviations FF and CF stand for "fake factor" and "correction factor", respectively, which are part of the $j \to \ell$-fake estimate.

## 9.4 Statistical Uncertainties on the Background Prediction

Another source of uncertainties are the statistical uncertainties on the background prediction. Although of statistical nature, they are treated as systematic uncertainties. The background yield $b_i$ in each bin $i$, is the sum of the weights $w_j$ of the single events while its statistical uncertainty $\sigma_{b_i}$ is the square root of the sum of the squared weights:

$$b_i = \sum_{j=1}^{N_{e_j}} w_j, \qquad \sigma_{b_i} = \sqrt{\sum_{j=1}^{N_{e_j}} w_j^2}\,. \tag{9.1}$$

The event weights $w_j$ are the product of several weights which for example arise from MC simulations, scale factors, the efficiency-correction ratio $\mathcal{R}$ or the fake factors $FF$. The relative statistical uncertainties of each background process with respect to the total background prediction in percent in each bin of the final discriminants is shown in Figure 9.5. The figures clearly show that the data part of the $e/\mu$-symmetric background has the largest




**(a)** Search for $H \to e\tau$, nonVBF SR

**(b)** Search for $H \to e\tau$, VBF SR

**(c)** Search for $H \to \mu\tau$, nonVBF SR

**(d)** Search for $H \to \mu\tau$, VBF SR

**Figure 9.5:** Relative statistical uncertainties of the individual background processes with respect to the total background in percent. Given for the $e\tau$-dataset (top) and the $\mu\tau$-dataset (bottom) in the bins of the distributions of the final discriminants in the nonVBF SR (left) and the VBF SR (right).

statistical uncertainties with respect to the total background prediction of up to $\sim 15\,\%$ in the nonVBF SR and up to $\sim 40\,\%$ in the VBF SR. For this contribution the event weights are close to 1. Overall the statistical uncertainties of the data part of the $e/\mu$-symmetric background are larger in the $e\tau$-dataset (i.e. in the search for $H \to e\tau$) compared to the $\mu\tau$-dataset. The data part of the $e/\mu$-symmetric background of the $e\tau$-dataset corresponds to the data of the $\mu\tau$-dataset which has a lower yield than the data of the $e\tau$-dataset and hence larger statistical uncertainties.

The second and third largest contribution is from both $j \to \ell$-fake contributions. In

the search for $H \to e\tau$, the $j \to \ell$-fakes have higher statistical uncertainties than the $\mathcal{R} \cdot j \to \ell$-fakes while it is vice versa in the search for $H \to \mu\tau$ since, apart from $\mathcal{R}$, they are the same, respectively.

The contribution with the lowest statistical uncertainties is from the *other* $\to \ell$-fakes which are estimated from MC simulations, where typically more events are generated than expected for this integrated luminosity. Hence the event weights are considerably smaller than 1 on average. Again, the size of the statistical uncertainties of the *other* $\to \ell$- and the $\mathcal{R} \cdot$ *other* $\to \ell$-fakes relative to each other is opposite in the search for $H \to e\tau$ and $H \to \mu\tau$ as they are, apart from the factor $\mathcal{R}$, the same, respectively.

## 9.5 Preprocessing of the Uncertainties for the Statistical Analysis

The large number of uncertainty sources results in a high number of parameters (NPs) in the statistical analysis. In order to simplify and stabilise the fitting procedure, uncertainty sources that only contribute marginally are removed (pruning). Furthermore, a symmetric behaviour of the $+1\sigma$- and the $-1\sigma$-variations is enforced (symmetrisation) and strong changes of the effect of the uncertainty from one bin to the next in the distribution of the final discriminant due to statistical fluctuations are smoothed (smoothing).

The three procedures are implemented in the software package TRExFitter [249] which is used to perform the statistical analysis. Details on the three procedures are given in the following. Each of the procedures is performed separately for each signal region (nonVBF SR and VBF SR) and each background contribution.

### Pruning

In order to reduce the number of nuisance parameters in the statistical analysis which helps with stabilising the fitting procedure, nuisance parameters which have a negligible effect on the final discriminant are omitted (pruned). The normalisation or acceptance effect and the shape effect of the NPs are treated independently.

The acceptance component is pruned once its effect on the overall yield of the respective sample is below 0.1 %. Both variations, the $+1\sigma$- and the $-1\sigma$-variation, must be below this threshold.

The shape effect of a nuisance parameter is pruned if there is no bin in the distribution of the final discriminant for which either the $+1\sigma$- or the $-1\sigma$-variation is above 0.1 % with respect to nominal.

An overview of all nuisance parameters and whether the acceptance effect, the shape effect, both effects or neither is pruned is given in Appendix D.

### Symmetrisation

If only the the $+1\sigma$- or only the $-1\sigma$-variation is available for a specific uncertainty source, the existent variation is mirrored with respect to nominal.

Kinematic systematics are systematics which impact the four-momentum of the objects in the event and not only the event weight. This comprises all uncertainties related to energy scale and resolution. These kinematic systematics often show an asymmetric behaviour of the $+1\sigma$- and $-1\sigma$-variations with respect to nominal. There are also cases where in some bins of the final discriminant both variations have effects in the same direction with respect to nominal. Hence, all kinematic systematics are symmetrised which is done bin-by-bin. To do so, the mean value of the absolute values of both variations is calculated (per bin). If both variations are in the same direction with respect to nominal, the one with the smaller absolute value is assigned the opposite sign and the absolute values of both are set to the

mean value. If both variations are in opposite directions with respect to nominal, the signs are kept but the absolute values are set to the mean value.

The same procedure is applied to the uncertainties related to the electron identification $SF$s in accordance with the partner analyses [22]. No symmetrisation is applied to all other uncertainties.

**Smoothing**

Statistical fluctuations can lead to a discontinuous behaviour of the variations over the course of the individual bins. Smoothing is applied in order to flatten these effects. If the overall relative statistical uncertainty of the respective process is larger than 5 %, all bins for the variation are merged which results in a fully flat behaviour over all bins and no shape-effect is left. Otherwise, a rebinning in the distribution of the relative statistical variations is performed until at most three local extrema are left and the relative statistical uncertainty in each bin is below 5 %.

# Chapter 10

# Statistical Analysis

The final results are obtained using a statistical analysis which is based on binned maximum likelihood fits. The prediction is fitted to the data in the distributions of the neural network probabilities in order to extract the parameters of interest. In this analysis, the parameters of interest are the signal strengths of the two LFV-signals; or, more generally, the difference of their branching ratios. The likelihood function, taking into account statistical and systematic uncertainties of the prediction, is discussed in Section 10.1. In order to quantify the significance of a potential signal and to set upper limits on its branching ratio, hypotheses tests are performed. This is detailed in Section 10.2. The measures and tools used to investigate the results of the fit are discussed in Section 10.3. The final Section 10.4 summarises the fit setup used to extract information about the two LFV-signals, $H \to e\tau$ and $H \to \mu\tau$ and discusses the different options for interpretation of the parameters of interest.

The technical realisation of the statistical analysis is performed with the `TRExFitter` software package [249], which makes use of `HistFactory` [250] including `RooStats` [251] and `RooFit` [252].

The following sections are based on Refs. [216, 250, 253, 254].

## 10.1 Likelihood

The probability to observe $n_i$ events in bin $i$ of one of the final discriminants, which are the neural network probability distributions in this analysis, is described by a Poisson distribution. The expected number of events in bin $i$, potentially consisting of background ($b_i$) and signal ($s_i$) events, can be expressed by $\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})$. The predictions $b_i$ and $s_i$ are obtained from the background estimation discussed in Section 7.2 and from MC simulations, respectively. The parameter $\mu$ is the signal strength. If it is 0 only background events contribute to this bin, if it is 1 the signal contribution equals the predicted one. Hence, this parameter is the so-called *parameter of interest* (POI) in the analysis. The symbol $\boldsymbol{\theta}$ represents a set of *nuisance parameter*s (NPs) $\theta_p$ affecting the number of expected signal and background events but whose measurement is not of primary interest. Then, the Poisson probability for bin $i$ becomes:

$$P_{\mathrm{P}}(n_i|\mu,\boldsymbol{\theta}) = \frac{(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))^{n_i}\, e^{-(\mu s_i(\boldsymbol{\theta})+b_i(\boldsymbol{\theta}))}}{n_i!}\,. \tag{10.1}$$

After a measurement is performed, the number of observed events in the single bins is fixed and the parameters of the expectation are fitted to the data with the aim to extract the signal strength $\mu$. This is realised via the likelihood function $\mathcal{L}(\mu,\boldsymbol{\theta}|n_i)$ which is a function of the parameters $\mu$ and $\boldsymbol{\theta}$ rather than the data. The likelihood function for the full distribution with observed events $\boldsymbol{n} = (n_1, ..., n_N)$ in $N$ statistically independent bins

is constructed as the product over all bins:

$$\mathcal{L}(\mu, \boldsymbol{\theta}|\boldsymbol{n}) = \prod_{i=1}^{N} \frac{(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))^{n_i}\, e^{-(\mu s_i(\boldsymbol{\theta})+b_i(\boldsymbol{\theta}))}}{n_i!}\,. \tag{10.2}$$

The parameter values $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ that maximise $\mathcal{L}(\mu, \boldsymbol{\theta}|\boldsymbol{n})$ are the maximum likelihood estimators and represent the best-fit values. The minimisation of $-\ln\mathcal{L}$ is performed with MINUIT [255]. In order to combine the distributions of several signal regions, equivalent terms for the additional distributions are multiplied in Eq. 10.2.

The covariance matrix $V_{ij}$ of the parameters $(\hat{\mu}, \hat{\boldsymbol{\theta}})$ is estimated in the large sample limit by inverting the second derivative of the negative logarithm of the likelihood at the minimum [253]:

$$\left(\widehat{V^{-1}}\right)_{ij} = -\frac{\partial^2 \ln\mathcal{L}}{\partial\theta_i \partial\theta_j}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\,, \tag{10.3}$$

and is used to obtain the correlation matrix of the nuisance parameters. The covariance matrix results in symmetric uncertainties on the parameters. However, the logarithm of the likelihood function around the minimum is not necessarily parabolic. Hence, the MINOS algorithm of MINUIT is run for all parameters after the minimisation to obtain asymmetric uncertainties by scanning the likelihood.

### 10.1.1 Incorporating Systematic Uncertainties

Different systematic uncertainty sources affecting the background and signal prediction were discussed in Chapter 9. They are parametrised by the NPs $\boldsymbol{\theta} = (\theta_1, ..., \theta_{N_\theta})$ for $N_\theta$ systematic uncertainty sources. The statistical uncertainties on the background prediction are excluded from this; their description follows in the next section. Typically, the systematic uncertainties are determined in independent auxiliary measurements, which results in additional information about them. In order to incorporate this additional information into the statistical model, the likelihood function is adapted. For each NP $\theta_p$, a Gaussian constraint term is multiplied, assuming that the probability density function of the auxiliary measurement $\theta'$ is Gaussian:

$$\mathcal{L}_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\boldsymbol{\theta'}, \boldsymbol{\sigma}) = \prod_{p=1}^{N_\theta} \frac{1}{\sqrt{2\pi}\sigma_p}\, e^{-\frac{1}{2}\left(\frac{\theta_p - \theta'_p}{\sigma_p}\right)^2}\,. \tag{10.4}$$

Each factor of the product is maximised for $\theta_p = \theta'_p$ with $\theta'_p$ reproducing the nominal value of the prediction and $\sigma_p$ being the absolute value of the $\pm1\sigma$-variations of the systematic uncertainties determined in the auxiliary measurements. The further the prediction is pulled away from its nominal values with $\theta_p = \theta'_p$, the more the additional likelihood term decreases, which is steered by $\sigma_p$. However, this can nevertheless be beneficial for the maximisation of the overall likelihood as these pulls can help to adapt the prediction such that it describes the data better. Since there is no dependence of the additional likelihood terms on the individual bins, the systematic uncertainty sources are correlated over all bins, each.

The effect of the systematic uncertainties is split into an acceptance or normalisation component $\eta$ and a shape component $\sigma_i$ in bin $i$ with $\eta' = 1$ and $\sigma'_i$ reproducing the nominal prediction and $\eta^{\pm}$ and $\sigma_i^{\pm}$ representing the $\pm1\sigma$-variations. An interpolation is performed to obtain a continuous parametrisation between the $\pm1\sigma$-variations and beyond. For the acceptance (normalisation) effect, an exponential interpolation is used while a linear

interpolation is used for the shape effect:

$$\eta\left(\theta_p\right) = \begin{cases} \left(\eta^+\right)^{\theta_p} & \theta_p \geq 0 \\ \left(\eta^-\right)^{-\theta_p} & \theta_p < 0 \end{cases} \tag{10.5}$$

$$\sigma_i\left(\theta_p\right) = \begin{cases} \sigma_i' + \theta_p\left(\sigma_i^+ - \sigma_i'\right) & \theta_p \geq 0 \\ \sigma_i' - \theta_p\left(\sigma_i^- - \sigma_i'\right) & \theta_p < 0 \end{cases} \tag{10.6}$$

The exponential interpolation ensures that $\eta\left(\theta_p\right)$ is always positive. For $\theta_p = 0$, the nominal case with $\eta = \eta' = 1$ or $\sigma_i = \sigma_i'$ is obtained. Whereas $\theta_p = \pm 1$ results in the $\pm 1\sigma$-variations $\eta^\pm$ or $\sigma_i^\pm$. This results in $\theta_p' = 0$ and $\sigma_p = 1$ and hence Eq. 10.4 becomes:

$$\mathcal{L}_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\boldsymbol{\theta'} = 0, \boldsymbol{\sigma} = 1) = \prod_{p=1}^{N_\theta} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(\theta_p)^2}\,. \tag{10.7}$$

### 10.1.2 Incorporating Statistical Uncertainties on the Prediction

Other sources of uncertainties are the statistical uncertainties on the background prediction while the ones on the signal prediction are assumed to be negligible. They are parametrised by one nuisance parameter $\gamma_i$ per statistical independent bin $i$ reflecting that the true rate may differ from the predictied $b_i$. The true rates are obtained by fitting the prediction, i.e. the $\gamma_i$ to the data. In principle, a nuisance parameter per process and bin could be introduced for statistically independent processes [256]. The statistical uncertainties of the individual processes are shown in Figure 9.5. However, this introduces a large number of those parameters, 95 in the search for $H \to e\tau$ and 100 in the search for $H \to \mu\tau$. Therefore, the total background prediction is treated in a combined way as one process or sample instead, which reduces the number of parameters to 19 and 20, respectively.

The background prediction $b_i$ in each bin $i$ consists of the sum of weights $w_j$ of the single $N_{e_i}$ events in this bin: $b_i = \sum_{j=1}^{N_{e_i}} w_j$. The statistical uncertainty $\sigma_{b_i}$ on $b_i$ is: $\sigma_{b_i} = \sum_{j=1}^{N_{e_i}} w_j^2$. This is incorporated into the likelihood function by the additional term $\mathcal{L}_{\boldsymbol{\gamma}}$. Assuming an auxiliary measurement of the true rate with the same relative statistical uncertainty $\sigma_{\mathrm{rel},i}$ as of $b_i$, which results in the measurement of $m_i$ unweighted (i.e. all $w_j = 1$) events, allows for a Poissonian constraint term of the form [250]

$$P_{\mathrm{P}}(m_i|\gamma_i m_i) = \frac{(\gamma_i m_i)^{m_i}\, e^{-\gamma_i m_i}}{m_i!}\,. \tag{10.8}$$

Here, $m_i$ fluctuates around the true rate $\gamma_i m_i$ and equals the true rate for $\gamma_i' = 1$. From the relative statistical uncertainty $\sigma_{\mathrm{rel},i} = \sigma_{b_i}/b_i$ and the requirement that it has to equal $\sigma_{m_i}/m_i = 1/\sqrt{m_i}$ (following Poisson statistics), the following relation for $m_i$ can be derived:

$$m_i = \frac{b_i^2}{\sigma_{b_i}^2}\,. \tag{10.9}$$

However, as $m_i$ is not necessarily an integer, the gamma distribution is used instead and the following additional term is added to the likelihood function:

$$\mathcal{L}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}|\boldsymbol{m}) = \prod_{i=1}^{N} \frac{m_i^{m_i}\gamma_i^{m_i-1}e^{-\gamma_i m_i}}{\Gamma(m_i - 1)}\,. \tag{10.10}$$

Here, $\Gamma$ is the gamma function. Additionally, the prediction $b_i$ in Eq. 10.2 must be multiplied with the nuisance parameter $\gamma_i$.

## 10.2 Hypothesis Testing

In order to quantify the significance of a new signal or to set upper limits on its cross section times branching ratio ($\mu$), hypothesis tests are performed. This involves formulating a null hypothesis $H_0$ and an alternative hypothesis $H_1$. For example, $H_0$ could state that the data is described by known processes, i.e. background processes only (corresponding to $\mu = 0$) while $H_1$ states that in addition a new signal ($\mu > 0$) is required to describe the data. A metric to quantify the (non-)compatibility of the data with the hypotheses is constructed which is a function of the data and called *test statistic* $t$. The probability density functions (pdf) of this test statistic for each of the hypotheses must be known to evaluate the test statistic. Then, a critical region in $t$ is defined. Assuming that the pdf of $t$ under $H_1$, $f(t|H_1)$, has larger values of $t$ compared to the pdf under $H_0$, $f(t|H_0)$, the value of $t$ defining the critical region, $t_{\text{cut}}$, is defined by:

$$\alpha = \int_{t_{\text{cut}}}^{\infty} f(t|H_0)dt\,. \tag{10.11}$$

If the observed value of $t$ is larger than $t_{\text{cut}}$, the null hypothesis $H_0$ is rejected in favour of the alternative hypothesis $H_1$. The significance level $\alpha$ must be defined beforehand. It represents the probability to reject $H_0$ although it is true. Similarly, the probability $\beta$ to reject $H_1$ (not reject $H_0$) although $H_1$ is true is defined by:

$$\beta = \int_{-\infty}^{t_{\text{cut}}} f(t|H_1)dt\,. \tag{10.12}$$

The power of the test statistic is given by $1 - \beta$. It is the probability to not reject $H_1$ when $H_1$ is true. That means that the test statistic $t$ should be chosen such that $1 - \beta$ is maximised for a given $\alpha$. Following the Neyman-Pearson Lemma [257], this is achieved by the ratio of the likelihoods of the two hypotheses. However, this is only true for simple hypotheses which do not depend on nuisance parameters. Nonetheless, a test statistic where the hypothesis of a fixed value $\mu$ is tested, is formulated as

$$t_\mu = -2\ln\frac{\mathcal{L}\left(\mu, \hat{\hat{\boldsymbol{\theta}}}, \hat{\hat{\boldsymbol{\gamma}}}\right)}{\mathcal{L}\left(\hat{\mu}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\gamma}}\right)}\,, \tag{10.13}$$

where the pdf of $t_\mu$ is known from Wilks theorem [258] and $\mathcal{L}$ is the likelihood function discussed in Section 10.1. In the denominator, $\hat{\mu}$ and $\hat{\theta}$ are the unconditional maximum likelihood estimators (MLEs), while $\hat{\hat{\theta}}$ in the numerator is the conditional MLE for a specific fixed $\mu$. The numerator represents the null hypothesis $H_0$ which can either be a background plus signal hypothesis with signal strength $\mu$ or a background-only hypothesis for $\mu = 0$. In case of perfect agreement of the data (denominator) with the hypothesis (numerator), the ratio equals 1, i.e. $t_\mu = 0$. The further data is away from the hypothesis, the smaller the ratio gets which is equivalent to increasing values of $t_\mu$. The value of $t_\mu$ observed in data is denoted by $t_{\mu,\text{obs}}$. The probability to observe data that is equally or less compatible with $H_0$ than the current observation is represented by the $p$-value which is calculated as follows:

$$p = \int_{t_{\mu,\text{obs}}}^{\infty} f(t_\mu|\mu)dt_\mu\,, \tag{10.14}$$

with $f(t_\mu|\mu)$ being the pdf of $t_\mu$ under the null hypothesis $H_0$ which is defined by the value of $\mu$. If $p < \alpha$, the null hypothesis is rejected. The $p$-value can be translated into a

significance $Z$ via

$$Z = \Phi^{-1}(1-p) \tag{10.15}$$

where $\Phi$ is the cumulative distribution of a standard Gaussian distribution. In order to obtain expected sensitivities, so-called *Asimov datasets* are constructed which is described in Section 10.3.

In addition, $t_\mu$ as defined in Eq. 10.13 is known to follow a $\chi^2$-distribution for one degree of freedom in the large sample limit [216, 258].

### 10.2.1 Discovery of a New Signal

For the discovery of a new signal, i.e. rejection of the background-only hypothesis for which $\mu = 0$, the following slightly modified test statistic $q_0$ is used:

$$q_0 = \begin{cases} -2\ln\left(\frac{\mathcal{L}\left(0,\hat{\hat{\boldsymbol{\theta}}},\hat{\hat{\gamma}}\right)}{\mathcal{L}\left(\hat{\mu},\hat{\boldsymbol{\theta}},\hat{\gamma}\right)}\right) & \text{for } \hat{\mu} \geq 0 \\ +2\ln\left(\frac{\mathcal{L}\left(0,\hat{\hat{\boldsymbol{\theta}}},\hat{\hat{\gamma}}\right)}{\mathcal{L}\left(\hat{\mu},\hat{\boldsymbol{\theta}},\hat{\gamma}\right)}\right) & \text{for } \hat{\mu} < 0 \end{cases} \tag{10.16}$$

For $\hat{\mu} > 0$, $q_0$ is positive and for $\hat{\mu} < 0$ negative. The best-fit signal strength $\hat{\mu}$ can become negative due to statistical fluctuations.

In order to calculate the $p$-value, the pdf of $q_0$ must be known. It can either be obtained by sampling its distribution with the Monte Carlo technique or by using an approximate analytical form. In this work the latter is used as the large sample limit is reached which allows for making use of results obtained by Wald [259] and Wilks [258]: the pdf of $q_0$ is $\phi(\sqrt{q_0})$ for $\hat{\mu} \geq 0$, as outlined in [216], and $\phi(-\sqrt{-q_0})$ for $\hat{\mu} < 0$ [254] with $\phi$ being the Gaussian pdf. And hence, with Eq. 10.15, the significance $Z$ can be obtained by $Z = +\sqrt{+q_0}$ for $\hat{\mu} \geq 0$ and by $Z = -\sqrt{-q_0}$ for $\hat{\mu} < 0$.

By convention in the particle physics community, a new signal is discovered if the background-only hypothesis can be rejected with at least $5\sigma$ ($p \leq \alpha = 2.87 \times 10^{-7}$). Evidence for a new signal is found if the background-only hypothesis can be rejected with $3\sigma$.

### 10.2.2 Limit Setting

In the case of limit setting, the background plus signal hypothesis with a fixed value $\mu$ for the signal strength serves as $H_0$ whereas the background only hypothesis serves as $H_1$. The modified test statistic $\tilde{q}_\mu$ used to extract upper limits on the signal strength $\mu$ is:

$$\tilde{q}_\mu = \begin{cases} -2\ln\left(\frac{\mathcal{L}\left(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu),\hat{\hat{\gamma}}(\mu)\right)}{\mathcal{L}\left(0,\hat{\hat{\boldsymbol{\theta}}}(0),\hat{\hat{\gamma}}(0)\right)}\right) & \text{for } \hat{\mu} < 0 \\ -2\ln\left(\frac{\mathcal{L}\left(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu),\hat{\hat{\gamma}}(\mu)\right)}{\mathcal{L}\left(\hat{\mu},\hat{\boldsymbol{\theta}},\hat{\gamma}\right)}\right) & \text{for } 0 \leq \hat{\mu} \leq \mu \\ 0 & \text{for } \hat{\mu} > \mu \end{cases} \tag{10.17}$$

The value of $\mu$ in the denominator of the first case is set to 0 when only allowing for $\mu > 0$. The value of the test statistic observed in data is denoted by $\tilde{q}_{\mu,\text{obs}}$. The pdfs of $\tilde{q}_\mu$ under the assumption of $\mu$ and of a different signal strength $\mu'$ are given in [216].

In order to exclude null hypotheses $H_0$ with signal strengths $\mu$, the $\text{CL}_s$-method [260]

is used with the $\mathrm{CL}_s$-value defined as:

$$\mathrm{CL}_s = \frac{p_\mu}{1 - p_{\mu=0}} \,. \tag{10.18}$$

The value $p_\mu$ is obtained according to Eq. 10.14 for a signal hypothesis with fixed $\mu$ and by replacing $t_\mu$ with $\tilde{q}_\mu$. Whereas $p_{\mu=0}$ ($\mu = 0$ indidcating the background-only hypothesis) is defined as:

$$p_{\mu=0} = \int_{-\infty}^{\tilde{q}_{\mu,\mathrm{obs}}} f\left(\tilde{q}_\mu|0\right) d\tilde{q}_\mu \,. \tag{10.19}$$

The null hypothesis with signal strength $\mu$ is excluded at a confidence level (CL) of $1 - \alpha$ if $\mathrm{CL}_s < \alpha$. Typically, $\alpha = 0.05$ is used. The upper limit on the signal strength $\mu$ is defined as the largest value of $\mu$ that is not excluded. This value can, for example, be found by iteratively testing different values of $\mu$.

The reason for using $\mathrm{CL}_s$ instead of only $p_\mu$ is that hypotheses for which the sensitivity is very small[1], which is the case if signal is much smaller than the expected background, are excluded with a probability of almost $\alpha$ [260]. Therefore, the $p_\mu$-value of the measurement is basically penalised by the power $\beta = 1 - p_{\mu=0}$ (Eq. 10.12) of the test statistic. The closer the power is to 1, i.e. the better the sensitivity is, the smaller is the penalty and $\mathrm{CL}_s \approx p_\mu$. Whereas if the power of the test is low which results in an increased $\mathrm{CL}_s$-value (compared to $p_\mu$), the rejection of the background plus signal hypothesis is prevented. Hence, the upper limits obtained with the $\mathrm{CL}_s$-method are more conservative compared to the limits obtained from $p_\mu$ alone.

## 10.3 Validation of the Fit Results

Several measures and tools that are used to validate and present the fit results in Chapter 11 are introduced in the following.

### Pulls and Constraints

The minimisation of the negative logarithm of the likelihood can on one hand pull the best-fit value of a nuisance parameter $\theta$ away from its nominal pre-fit value and on the other hand constraint it, i.e. reduce the size of its uncertainty due to additional knowledge from the data.

The *pull* of a nuisance parameter $\theta$ is defined by the relative change of its best-fit value $\hat{\theta}$ with respect to its nomininal pre-fit value $\theta'$:

$$\mathrm{pull} = \frac{\hat{\theta} - \theta'}{\sigma_\theta} \,, \tag{10.20}$$

with $\sigma_\theta$ denoting its pre-fit uncertainty assumed in an auxiliary measurement (see Eq. 10.4).

The constraint of a nuisance parameter denotes the reduction in size of its associated uncertainty from the fit, $\hat{\sigma}_{\hat{\theta}}$ as extracted with MINOS (see Section 10.1), compared to the pre-fit value $\sigma_\theta$:

$$\mathrm{constraint} = \frac{\hat{\sigma}_{\hat{\theta}}}{\sigma_\theta} \,. \tag{10.21}$$

[1] The sensitivity is small if the pdfs of $H_0$ and $H_1$ strongly overlap.

### Ranking of Nuisance Parameters

The single nuisance parameters (NPs) can have differently strong impacts on the POI, i.e. the signal strength $\mu$. Therefore, a ranking of the NPs with respect to their impact on $\mu$ can be constructed in order to evaluate their importance on the final results. To obtain the impact of one NP on $\mu$, the value of this NP is fixed to e.g. its $-1\sigma$-variation. Then the full fit is repeated resulting in a new value of $\mu$. The difference of the nominal value of $\mu$ and its new value is the impact of the corresponding NP. This is repeated for the $+1\sigma$-variation of the NP and for all other NPs.

Pre-fit as well as post-fit impacts of the NPs can be derived by either using the pre-fit or the post-fit uncertainty on the NPs, respectively.

Finally, the ranking of the NPs is obtained by sorting the impacts obtained from the post-fit $+1\sigma$-variation in decreasing order.

### Breakdown of Uncertainties

The breakdown of the uncertainties gives an overview of the impact of NPs on the uncertainty of the signal strength $\mu$, $\sigma_\mu$. Typically, several NPs of similar origin (e.g. all uncertainties related to electrons) are combined into groups. The impact of one group on $\sigma_\mu$ is obtained as follows: a fit is performed where all NPs of this group are fixed to their best-fit values of the nominal fit. The resulting new $\sigma_\mu^{\text{fixed}}$ is subtracted in quadrature from the nominal $\sigma_\mu$ which results in the impact of this group of NPs on $\sigma_\mu$, denoted by $\sigma_\mu^{\text{group}}$:

$$\sigma_\mu^{\text{group}} = \sqrt{\sigma_\mu^2 - \left(\sigma_\mu^{\text{fixed}}\right)^2}\,. \tag{10.22}$$

This is performed for any desired group of NPs. The quadratic sum of the single $\sigma_\mu^{\text{group}}$ does not necessarily add up to $\sigma_\mu$ due to correlations between the NPs. Owing to the usage of MINOS, the up and down value of $\sigma_\mu$ can differ and hence also the up and down impact of one group can be different.

The impact on $\sigma_\mu$ due to data statistical uncertainties only (without any uncertainties on the prediction), $\sigma_\mu^{\text{stat}}$, is obtained by performing a fit where all NPs are set to their best-fit values of the nominal fit while only the signal strength parameter $\mu$ is free. By subtracting $\sigma_\mu^{\text{stat}}$ in quadrature from $\sigma_\mu$, following Eq. 10.22, the impact of all NPs on $\sigma_\mu$ is obtained, denoted by $\sigma_\mu^{\text{syst}}$: $\sigma_\mu^{\text{syst}} = \sqrt{\sigma_\mu^2 - \left(\sigma_\mu^{\text{stat}}\right)^2}$.

### Asimov Dataset

The $H$-LFV analysis is developed and optimised while keeping signal-sensitive bins in the distributions of the different observables blinded (see Chapter 3). However, a measure of the sensitivity of the analysis is needed for the optimisation process. Furthermore, the statistical model must be examined for its general soundness before performing the fit with the observed data. Both is achieved by performing the full statistical analysis, i.e. obtaining the maximum likelihood estimators, deriving the significance for the potentially existing new signal and deriving the upper limit on its branching ratio, by making use of the *Asimov dataset* [216].

The Asimov dataset is an artificially created dataset. It is constructed as the sum of all expected background processes and the signal, assuming a certain value $\mu'$ for the signal strength.

Thus, when performing the maximum likelihood fit on the Asimov dataset, the best-fit values (MLEs) of the signal strength $\mu$ and of all nuisance parameters result in their pre-fit values: $\hat{\mu} = \mu'$, $\hat{\gamma}_i = \gamma_i'$ and $\hat{\theta}_p = \theta_p'$, but $\hat{\sigma}_p \neq \sigma_p$ (Eq. 10.4) can occur. This means no pulls of the nuisance parameters are observed, whereas constraints can appear. If the fit

strongly constrains a nuisance parameter, it indicates that the analysis is sensitive to it and can reduce the related uncertainty compared to the value provided by the auxiliary measurement. This typically also leads to noticeable pulls once the fit is performed with observed data. Stronger constraints in a fit with Asimov data hint to an overestimated or overly simplified description of the uncertainty and as such should be investigated.

Two different versions of the Asimov dataset are used in this analysis. In order to get measures for the expected sensitivity while optimising the analysis, the *pre-fit Asimov* dataset is used. When quoting the expected limit and significance in conjunction with the observed results, the *post-fit Asimov* dataset is used. Each of them is constructed as follows.

***Pre-fit Asimov* dataset** The *pre-fit Asimov* dataset is constructed from the sum of all expected background processes as they are predicted before the maximum likelihood fit to observed data is performed (pre-fit) and optionally signal. The pre-fit expected significance is derived assuming $\mu' = 1$ for the signal strength. This corresponds to a branching ratio of the LFV-signal of 1 % if not stated otherwise and therefore to the background plus signal hypothesis. The pre-fit expected limit is derived using $\mu' = 0$, resulting in a background-only dataset and therefore the background-only hypothesis.

***Post-fit Asimov* dataset** The *post-fit Asimov* dataset is constructed from the sum of all considered background processes, setting the NPs to their post-fit values obtained from the fit to the observed data, and optionally signal. Using the post-fit values for the NPs when constructing the Asimov dataset allows for taking into account changes in the prediction due to the fit to observed data. The post-fit expected significance and limit are derived assuming $\mu' = 1$ and $\mu' = 0$, respectively, as above.

In the process of investigating the soundness of the statistical model, also a *sideband* and a *mixed dataset* were used.

The sideband dataset only uses the bins of the final discriminants which are not sensitive, i.e. not blinded. Hence, the observed data can be used in these bins and a more realistic picture of the behaviour of the individual nuisance parameters with regard to their pulls and constraints can be obtained.

The mixed dataset combines the sideband dataset with a post-fit Asimov dataset in the sensitive bins. This Asimov dataset is built with the NPs set to the post-fit values obtained from a background-only fit to observed data in the sideband bins. With this setup a more realistic estimation of the expected sensitivity can be obtained as changes in the background prediction from the background-only fit are taken into account, without using observed data in the sensitive bins.

## 10.4 Fit Setup

The statistical analysis as introduced above is performed separately with the $e\tau$- and the $\mu\tau$-dataset in order to search for the $H \to e\tau$- and the $H \to \mu\tau$-signal, respectively. The distributions of the neural network probabilities serve as final discriminants which are used in the maximum likelihood fits. The fit is performed simultaneously in both signal regions, the nonVBF SR and the VBF SR, with the respective neural network probablity distributions.

The parameter of interest (POI), an unconstrained overall normalisation parameter, is the signal strength $\mu^{e\tau}$ or $\mu^{\mu\tau}$. It scales the respective signal template, $s_{H\to e\tau}$ or $s_{H\to\mu\tau}$. The signal templates are obtained from MC simulations assuming a production cross section of the Higgs boson as predicted in the standard model and a branching ratio of 1 % for

each of the LFV-signals. Hence, $\mu = 1$ corresponds to a branching ratio of 1 % and e.g. $\mu = 0.1$ to a branching ratio of 0.1 %. A short discussion of the interpretation of $\mu^{e\tau}$ or $\mu^{\mu\tau}$ is given in the next paragraph.

The main term of the binned likelihood function (Eq. 10.2), without the additional terms for systematic and statistical uncertainties of the prediction, is given in the following for the search for $H \to \mu\tau_e$. This likelihood is used to perform a maximum likelihood fit, to derive significances with which the background-only hypothesis can potentially be rejected and to set upper limits on their branching ratios.

$$\mathcal{L}(\mu, \boldsymbol{\theta}, \boldsymbol{\gamma} | \boldsymbol{n}^{\mu\tau}) = \prod_{i=1}^{N} P_{\mathcal{P}} \left( n_i^{\mu\tau} | \gamma_i b_i^{\mu\tau}(\boldsymbol{\theta}) + \mu^{\mu\tau} \cdot s_{H\to\mu\tau,i}^{\mu\tau}(\boldsymbol{\theta}) \right) \tag{10.23}$$

$$= \prod_{i=1}^{N} P_{\mathcal{P}} \left( n_i^{\mu\tau} | \gamma_i \left( \tilde{n}_i^{e\tau}(\boldsymbol{\theta}) + f_{j\to\ell,i}^{\mu\tau}(\boldsymbol{\theta}) + f_{other\to\ell,i}^{\mu\tau}(\boldsymbol{\theta}) \right) + \mu^{\mu\tau} \cdot s_{H\to\mu\tau,i}^{\mu\tau}(\boldsymbol{\theta}) \right) \tag{10.24}$$

with

$$\tilde{n}_i^{e\tau} = \mathcal{R}^{e\tau} \cdot \left( n_i^{e\tau} - f_{j\to\ell,i}^{e\tau}(\boldsymbol{\theta}) - f_{other\to\ell,i}^{e\tau}(\boldsymbol{\theta}) - \mu^{\mu\tau} \cdot s_{H\to\mu\tau,i}^{e\tau}(\boldsymbol{\theta}) \right) . \tag{10.25}$$

The likelihood for the search for $H \to e\tau_\mu$ can be formulated eqivalently by exchanging $\mu\tau$ with $e\tau$ and vice versa.

The product is over the bins $i$ of the final discriminants in both signal regions. The binning of the final discriminants is discussed in Section 8.2.6. The Poisson pdf is denoted by $P_{\mathcal{P}}$. The number of detected events is denoted by the vector $\boldsymbol{n}$ where the components $n_i$ represent the number of detected events in bin $i$. The superscript $\mu\tau$ denotes the $\mu\tau$-dataset, $e\tau$ the $e\tau$-dataset. The nuisance parameters $\boldsymbol{\gamma}$ describe the statistical uncertainties on the background prediction $\boldsymbol{b}$ with components $\gamma_i$ and $b_i$ for bin $i$, respectively. Each component $\theta_p$ of the vector of nuisance parameters $\boldsymbol{\theta}$ describes one source of systematic uncertainty. The background prediction is split in its individual contributions as given in Eqs. 10.24 and 10.25. There are the $e/\mu$-symmetric background $\tilde{n}^{e\tau}$ (Eq. 10.25), the $j \to \ell$- and the $other \to \ell$-fakes.

The $e/\mu$-symmetric background consists of the data of the respective other dataset, here of the $e\tau$-dataset, and of the $j \to \ell$- and the $other \to \ell$-fakes of the $e\tau$-dataset, subtracted and corrected with the efficiency ratio $\mathcal{R}$. In addition, the part of the $H \to \mu\tau$-signal that is wrongly classified, namely that is assigned to belong to the $e\tau$-dataset, must be subtracted and also corrected with $\mathcal{R}$, since it introduces an asymmetry between the two datasets in case the $H \to \mu\tau$-signal exists. Its shape in the final discriminants differs from the correctly classified signal and hence only has a minor impact on the sensitivity. It is also scaled by the signal strength parameter $\mu^{\mu\tau}$. More details on the dataset classification, i.e. lepton assignment can be found in Section 7.1.3. In fact, the correction is applied on an event-by-event basis. More details can be found in Section 7.2.3 and Eq. 7.17.

The estimation of the $j \to \ell$-fakes is discussed in Section 7.2.2 and a detailed expression for $f_{j\to\ell}$ can be found in Eq. 7.12. The same section also includes details on the estimation of the $other \to \ell$-fakes.

### Interpretation of the Signal Strength $\mu$

As detailed in Chapter 3, the *Symmetry Method* is only sensitive to the difference in branching ratios of the $H \to \mu\tau$- and the $H \to e\tau$-signal, meaning that the signal strength $\mu^{\mu\tau}$ ($\mu^{e\tau}$) in the search for $H \to \mu\tau$ ($H \to e\tau$) does not necessarily represent the branching

ratio $\mathcal{B}(H \to \mu\tau)$ ($\mathcal{B}(H \to e\tau)$) in percent, but instead a difference $\Delta^{\mu\tau}$ ($\Delta^{e\tau}$):

$$\Delta^{\mu\tau} = \mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)\,. \tag{10.26}$$

And equivalently, the difference $\Delta^{e\tau}$ is:

$$\Delta^{e\tau} = \mathcal{B}(H \to e\tau) - \mathcal{B}(H \to \mu\tau)\,. \tag{10.27}$$

Hence, the absolute values of both are equal:

$$\Delta^{\mu\tau} = -\Delta^{e\tau}\,. \tag{10.28}$$

That means, in the most general case, the signal strengths $\mu^{\mu\tau}$ and $\mu^{e\tau}$ are parameters for the differences. If it is assumed, however, that $\mathcal{B}(H \to e\tau) = 0$ in the search for $H \to \mu\tau$ and $\mathcal{B}(H \to \mu\tau) = 0$ in the independent search for $H \to e\tau$, the signal strengths $\mu^{\mu\tau}$ and $\mu^{e\tau}$ are direct measurements of the branching ratios. The limit on the branching ratio of the decay $\mu \to e\gamma$ allows for deriving a bound on the product of the two off-diagonal Yukawa couplings $Y_{e\tau}$ and $Y_{\mu\tau}$, as discussed in Section 1.4.2, suggesting that not both LFV signals can be "large" simultaneously. The largest value both branching ratios, $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$, can take on simultaneously is $8.28 \times 10^{-3}\,\%$. In this case no signal could be detected with the *Symmetry Method*. As soon as one of the two signals is greater than this value, the other must be smaller, resulting in an asymmetry between both and hence the possibility to detect this difference with the *Symmetry Method*. Further discussions of the relations of both measurements of $\Delta^{\mu\tau}$ and $\Delta^{e\tau}$ are given in Sections 11.1.3 and 11.1.4.

# Chapter 11

# Results

The results obtained in the search for lepton-flavour violating decays of the Higgs boson, $H \to e\tau_\mu$ and $H \to \mu\tau_e$, are presented and discussed in the following. Section 11.1 focuses on the results derived with the *Symmetry Method* which are the main focus of this work.

Combinations of the results obtained with the *Symmetry Method* and of results derived by the partner-analyses are discussed in Sections 11.2 and 11.3. The estimation of all backgrounds apart from fakes of the two partner-analyses relies on templates from MC simulations where the normalisation of the main backgrounds is obtained from data. One analysis searches for $H$-LFV decays by exploiting hadronic $\tau$-lepton decays (called MC-based lephad) while the other exploits leptonic $\tau$-lepton decays (called MC-based leplep) considering the same final state as in this work.

Section 11.2 presents the combination with the MC-based lephad analysis, Section 11.3 the combination of all three analyses by using one of the leplep analyses for each of the two signal regions.

## 11.1 Results with the Symmetry Method

In order to obtain the signal strengths $\mu^{e\tau}$ and $\mu^{\mu\tau}$ in the searches for $H \to e\tau_\mu$ and $H \to \mu\tau_e$, respectively, and to set upper limits, a statistical analysis as described in Chapter 10 is performed.

First, a maximum likelihood fit is performed using the likelihood function in Eq. 10.24, thereby obtaining the best-fit values or maximum likelihood estimators (MLEs) of the signal strength and the nuisance parameters. The signal strength is defined such that a values of 1 corresponds to a branching ratio or a branching ratio difference of 1 %. The significance of the observed signal strength values and the upper limit on the signal strength at 95 % CL using the $\mathrm{CL}_s$-method is calculated. This is discussed in Section 11.1.1.

Section 11.1.2 quantifies the compatibility of the results between the two signal regions, Section 11.1.3 discusses the relation between $\mu^{e\tau}$ and $\mu^{\mu\tau}$ and Section 11.1.4 quantifies the compatibility of the results among the two searches.

### 11.1.1 Results

The analysis was developed without considering data in sensitive bins (blinding). Bins are considered as sensitive bins whenever the signal-over-background ratio exceeds 5 % assuming a branching ratio of 1 % for the signal. Expected significances and upper limits were derived using the *pre-fit Asimov* dataset discussed in Section 10.3 in order to serve as measures for optimising the analysis. The fit model was validated using a *mixed* dataset, introduced in Section 10.3 as well.

The following presents and discusses the observed results obtained from performing the fit with data in all bins of the final discriminants after all development and optimisation was finalised. In addition, expected limits and significances are quoted obtained from a fit with the *post-fit Asimov* dataset where the nuisance parameters are fixed to their post-fit values obtained from the fit to data as detailed in Section 10.3. The final results are obtained from a combined fit in both signal regions (nonVBF SR and VBF SR), in the following denoted by "combined", for one particular search channel, either $H \to e\tau_\mu$ or $H \to \mu\tau_e$. In addition, results are derived by performing the fit in each of the signal regions separately, in the following simply denoted by nonVBF SR and VBF SR.

The best-fit values of the signal strengths $\mu^{e\tau}$ and $\mu^{\mu\tau}$ are summarised in Table 11.1 where also the corresponding significances $Z$ with which the background-only hypothesis can be rejected and the upper limits at 95 % CL on the signal strengths are given, denoted by $\mu^{e\tau}_{95}$ and $\mu^{\mu\tau}_{95}$. The expected significance for a signal strength of 0.1 which is equivalent to a branching ratio of 0.1 % and the expected upper limits are given in brackets.

| | combined | nonVBF SR | VBF SR |
|---|---|---|---|
| $\hat{\mu}^{e\tau}$ | $-0.330^{+0.105}_{-0.110}$ | $-0.375^{+0.114}_{-0.121}$ | $-0.090^{+0.222}_{-0.253}$ |
| $Z/\sigma$ | −3.253 (1.010) | −3.399 (0.879) | −0.391 (0.507) |
| $\mu^{e\tau}_{95}$ | 0.077 $(0.190^{+0.077}_{-0.053})$ | 0.083 $(0.217^{+0.088}_{-0.061})$ | 0.361 $(0.400^{+0.168}_{-0.112})$ |
| $\hat{\mu}^{\mu\tau}$ | $0.249^{+0.103}_{-0.100}$ | $0.299^{+0.114}_{-0.111}$ | $0.060^{+0.217}_{-0.195}$ |
| $Z/\sigma$ | 2.502 (1.042) | 2.693 (0.904) | 0.304 (0.523) |
| $\mu^{\mu\tau}_{95}$ | 0.422 $(0.191^{+0.076}_{-0.053})$ | 0.490 $(0.222^{+0.089}_{-0.062})$ | 0.476 $(0.386^{+0.163}_{-0.108})$ |

**Table 11.1:** The best-fit values $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$ of the signal strengths, the corresponding significances $Z$ and the upper limits on the signal strengths at 95 % CL, denoted by $\mu^{e\tau}_{95}$ and $\mu^{\mu\tau}_{95}$, for the combined fit and the fits solely performed in the individual signal regions. The expected significance for a signal strength of 0.1 and the expected upper limits, obtained with the *post-fit Asimov* dataset, are given in brackets. The upper half of the table summarises the results obtained with the $e\tau$-dataset in the search for $H \to e\tau$. The lower half shows the results obtained with the $\mu\tau$-dataset in the search for $H \to \mu\tau$.

An upwards fluctuation in data in the $\mu\tau$-dataset (i.e. in the search for $H \to \mu\tau$), corresponding to a significance of $2.50\,\sigma$ and a best-fit value $\hat{\mu}^{\mu\tau} = 0.25 \pm 0.10$ is observed in the combined fit. This translates to an observed downwards fluctuation in the $e\tau$-dataset with a significance of $-3.25\,\sigma$ and a best-fit value $\hat{\mu}^{e\tau} = -0.33 \pm 0.11$. These results are either caused by an actual upwards fluctuation in data of the $\mu\tau$-dataset, an actual downwards fluctuation in data of the $e\tau$-dataset or by both since with the *Symmetry Method* $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$ are directly correlated through the $e/\mu$-symmetric background estimate. This implicates that the *Symmetry Method* is only sensitive to the difference of the two signal strengths. A detailed investigation of these results, in particular with regard to the relation and the compatibility of $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$ is performed in Sections 11.1.3 and 11.1.4. Due to the negative $\hat{\mu}^{e\tau}$, the observed upper limit on the signal strength $\mu^{e\tau}$ is smaller than the expected. And due to the positive $\hat{\mu}^{\mu\tau}$ the observed upper limit on the signal strength $\mu^{\mu\tau}$ is larger than the expected. When assuming that the branching ratio $\mathcal{B}$ of $H \to \mu\tau$ is zero, an upper limit on the branching ratio of $H \to e\tau$ can be set:

$$\mathcal{B}(H \to e\tau) < 0.077\,\%\,(0.190^{+0.077}_{-0.053}\,\%)\,, \tag{11.1}$$

where the expected limit is given in brackets. When, in contrast, assuming that the branching ratio of $H \to e\tau$ is zero, an upper limit on the branching ratio of $H \to \mu\tau$ can

be set:

$$\mathcal{B}(H \to \mu\tau) < 0.422\,\% \,(0.191^{+0.076}_{-0.053}\,\%)\,. \tag{11.2}$$

The observed fluctuations are driven by the nonVBF SR where the absolute values of the signal strength deviate further from 0 than in the combined fit. Small deviations of $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$ are also observed in the VBF SR. However, within uncertainties they are compatible with 0 which is also reflected in the small significances.

The post-fit distributions of the final discriminants are shown in Figure 11.1 and the post-fit yields are listed in Table 11.2. The distributions show the opposite behaviour of the observed fluctuations in data. The up- or down-fluctuation in one dataset results in an down- or up-fluctuation in the other and hence in a positive best-fit signal contribution in one dataset and a negative best-fit signal contribution in the other. Compared to the pre-fit distributions (Figure 8.44) and the pre-fit yields (Tables 8.1 and 8.2), the uncertainties on the prediction are reduced due to constraints of the NPs and due to correlations between the NPs and the processes. The constraints as well as the correlations of the NPs are discussed below. The post-fit prediction describes the data well over the entire range of the distributions.

### Influence of Individual Uncertainty Sources

The breakdown of the uncertainties on the signal strength $\mu$, i.e. the impact of groups of uncertainty sources on the uncertainty of $\mu$, is summarised in Table 11.3. Details of its calculation can be found in Section 10.3. The pulls and constraints (defined in Section 10.3, likewise) of the 25 nuisance parameters (NPs) with the highest impact on the value of the signal strength $\mu$ are shown in Figure 11.2 for the combined fit and in Figure 11.3 for the fit in the VBF SR only. The corresponding figure for the fit performed solely in the nonVBF SR shows similar features to the combined fit (shown in Appendix E). Figures showing the pulls and constraints of all NPs obtained in the combined fit and in the fits performed solely in the single signal regions are given in Appendix E as well.

The total uncertainty on $\mu$ is smaller in the search for $H \to \mu\tau$ than in the search for $H \to e\tau$ (Table 11.3). In contrast, the uncertainty due to the data sample size (i.e. due to the amount of collected and selected data) is larger in $H \to \mu\tau$ which reflects the lower data yield observed in the $\mu\tau$-dataset (see Table 11.2). This automatically leads to a larger uncertainty due to the background sample size in $H \to e\tau$ compared to $H \to \mu\tau$. The uncertainty due to the background sample size represents the statistical uncertainties on the total background prediction. It is dominated by the data sample size of the "other" dataset[1] (as shown in Figure 9.5) which is the basis for the $e/\mu$-symmetric background estimate. Hence, the data sample size uncertainty and the background sample size uncertainty are correlated among the two searches.

**Statistical Uncertainties on the Background Prediction** The contribution of the background sample size to the total uncertainty is larger than the contribution of all other systematic uncertainties and hence, together with the uncertainty due to the data sample size, the analysis is clearly dominated by statistical uncertainties. This is particularly the case in the VBF SR where the yields are lower and the systematic uncertainties play an even smaller role. However, the latter could be amplified by the fact that the shape effect of more NPs are dropped in the VBF SR due to being flat after the smoothing procedure (Section 9.5), which is again provoked by low yields. The importance of the statistical

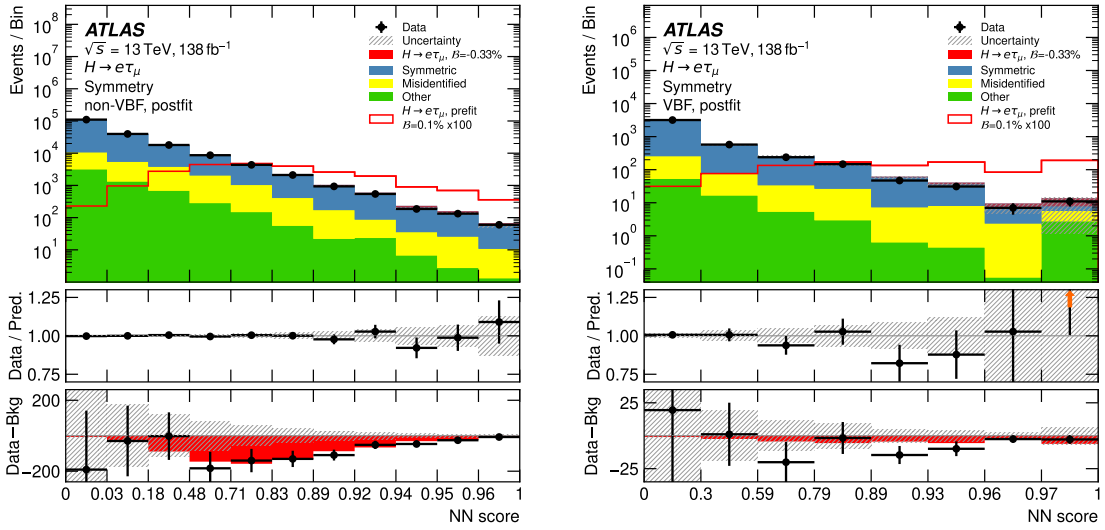
[1] In the search for $H \to e\tau$, for example, the $e\tau$-dataset is used while the "other" dataset, in this case the $\mu\tau$-dataset, is used for the estimation of the $e/\mu$-symmetric background.

| | Process | Yields | | | |
|---|---|---|---|---|---|
| | | $e\tau$ | | $\mu\tau$ | |
| | | nonVBF SR | VBF SR | nonVBF SR | VBF SR |
| Signal | ggF$H \to \ell\tau$ | $-720 \pm 230$ | $-11 \pm 4$ | $480 \pm 190$ | $8 \pm 4$ |
| | VBF$H \to \ell\tau$ | $-38 \pm 12$ | $-20 \pm 7$ | $26 \pm 11$ | $14 \pm 6$ |
| | VH$H \to \ell\tau$ | $-31 \pm 11$ | $-0.55 \pm 0.21$ | $21 \pm 9$ | $0.33 \pm 0.16$ |
| | $\mathcal{R} \cdot$ ggF$H \to \ell\tau$ | $48 \pm 17$ | $1.30 \pm 0.50$ | $-35 \pm 15$ | $-1.00 \pm 0.50$ |
| | $\mathcal{R} \cdot$ VBF$H \to \ell\tau$ | $3.90 \pm 1.30$ | $1.70 \pm 0.60$ | $-2.80 \pm 1.20$ | $-1.20 \pm 0.50$ |
| | $\mathcal{R} \cdot$ VH$H \to \ell\tau$ | $5.70 \pm 2$ | $0.14 \pm 0.06$ | $-4.20 \pm 1.80$ | $-0.10 \pm 0.05$ |
| Fake Contributions | $j \to \ell$ | $17\,900 \pm 1\,800$ | $330 \pm 40$ | $17\,200 \pm 1\,700$ | $320 \pm 40$ |
| | $other \to \ell$ | $5\,700 \pm 800$ | $81 \pm 19$ | $11\,500 \pm 1\,600$ | $106 \pm 28$ |
| Symmetric Background | $R \cdot$ Data | $193\,700 \pm 1\,200$ | $4\,300 \pm 50$ | $162\,900 \pm 900$ | $3\,890 \pm 50$ |
| | $R \cdot j \to \ell$ | $-19\,000 \pm 2\,000$ | $-340 \pm 40$ | $-15\,100 \pm 1\,500$ | $-300 \pm 34$ |
| | $R \cdot other \to \ell$ | $-12\,600 \pm 1\,800$ | $-112 \pm 30$ | $-5\,400 \pm 700$ | $-2 \pm 17$ |
| Total Background | | $185\,700 \pm 500$ | $4\,260 \pm 50$ | $171\,000 \pm 500$ | $3\,950 \pm 50$ |
| | Data | 184 887 | 4 230 | 171 675 | 3 967 |

**Table 11.2:** Observed event yields in the nonVBF SRs and the VBF SRs obtained from the combined fit in the search for $H \to e\tau$ (denoted by $e\tau$) and in the search for $H \to \mu\tau$ (denoted by $\mu\tau$), respectively. All processes denoted with an $\mathcal{R}$ in front are originally of the respective other dataset, corrected with the efficiency ratio $\mathcal{R}$. The uncertainties include statistical and systematic uncertainties. The uncertainty on the total background prediction is smaller than the quadratic sum of the uncertainties of the single contributions due to correlations between them.

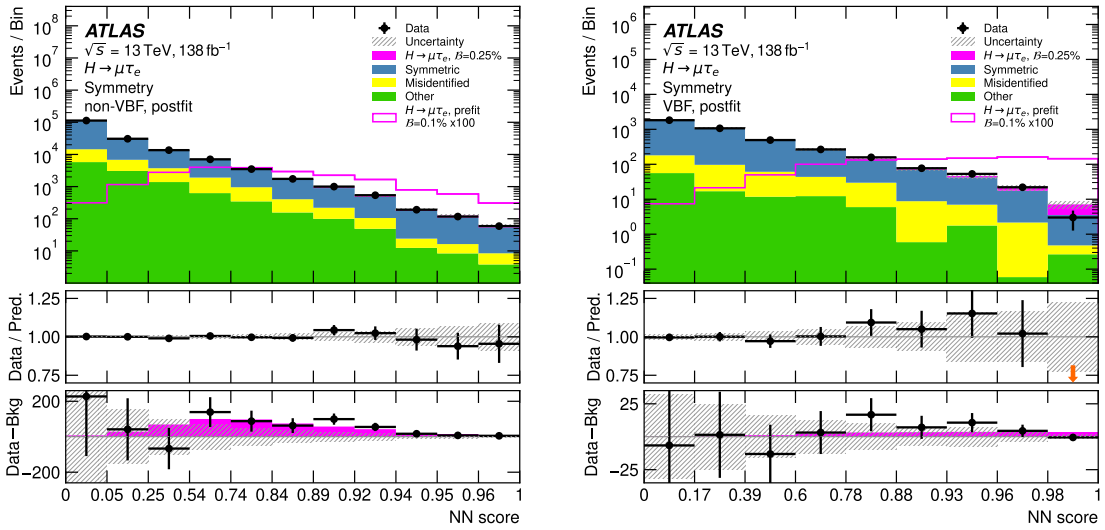
| **Group** | **Uncertainty on observed $\mu$ by NP group** | | | | | |
|---|---|---|---|---|---|---|
| | combined | | nonVBF SR | | VBF SR | |
| | $H \to e\tau$ | $H \to \mu\tau$ | $H \to e\tau$ | $H \to \mu\tau$ | $H \to e\tau$ | $H \to \mu\tau$ |
| Total uncertainty | $+0.105 \,/\, -0.110$ | $+0.103 \,/\, -0.100$ | $+0.114 \,/\, -0.121$ | $+0.114 \,/\, -0.111$ | $+0.222 \,/\, -0.253$ | $+0.217 \,/\, -0.195$ |
| Data sample size | $+0.047 \,/\, -0.046$ | $+0.056 \,/\, -0.055$ | $+0.051 \,/\, -0.051$ | $+0.059 \,/\, -0.058$ | $+0.124 \,/\, -0.108$ | $+0.162 \,/\, -0.135$ |
| Syst. unc. + Bkg. sample size | $+0.093 \,/\, -0.100$ | $+0.087 \,/\, -0.084$ | $+0.102 \,/\, -0.109$ | $+0.098 \,/\, -0.094$ | $+0.185 \,/\, -0.229$ | $+0.145 \,/\, -0.140$ |
| Background sample size | $+0.080 \,/\, -0.085$ | $+0.068 \,/\, -0.068$ | $+0.082 \,/\, -0.086$ | $+0.073 \,/\, -0.074$ | $+0.183 \,/\, -0.226$ | $+0.140 \,/\, -0.139$ |
| Systematic uncertainties | $+0.065 \,/\, -0.073$ | $+0.062 \,/\, -0.058$ | $+0.071 \,/\, -0.079$ | $+0.073 \,/\, -0.069$ | $+0.042 \,/\, -0.078$ | $+0.046 \,/\, -0.021$ |
| Electron FF stat. | $+0.014 \,/\, -0.014$ | $+0.014 \,/\, -0.014$ | $+0.017 \,/\, -0.015$ | $+0.016 \,/\, -0.017$ | $+0.004 \,/\, -0.003$ | $+0.002 \,/\, -0.002$ |
| Muon FF stat. | $+0.003 \,/\, -0.003$ | $+0.003 \,/\, -0.003$ | $+0.007 \,/\, < 0.001$ | $+0.004 \,/\, -0.003$ | $+0.006 \,/\, -0.009$ | $+0.004 \,/\, -0.003$ |
| FF syst. | $+0.005 \,/\, -0.006$ | $+0.005 \,/\, -0.004$ | $+0.003 \,/\, -0.005$ | $+0.005 \,/\, -0.003$ | $+0.011 \,/\, -0.015$ | $+0.006 \,/\, -0.006$ |
| Electron CF | $+0.032 \,/\, -0.035$ | $+0.034 \,/\, -0.032$ | $+0.039 \,/\, -0.043$ | $+0.042 \,/\, -0.039$ | $< 0.001 \,/\, -0.017$ | $+0.005 \,/\, -0.006$ |
| Muon CF | $+0.002 \,/\, -0.003$ | $+0.002 \,/\, -0.003$ | $+0.005 \,/\, -0.002$ | $+0.004 \,/\, -0.004$ | $+0.016 \,/\, -0.029$ | $+0.011 \,/\, -0.012$ |
| Efficiency correction $\mathcal{R}$ | $+0.020 \,/\, -0.023$ | $+0.020 \,/\, -0.018$ | $+0.023 \,/\, -0.027$ | $+0.025 \,/\, -0.022$ | $< 0.001 \,/\, -0.014$ | $+0.005 \,/\, -0.005$ |
| Signal theoretical unc. | $+0.015 \,/\, -0.028$ | $+0.025 \,/\, -0.012$ | $+0.017 \,/\, -0.033$ | $+0.031 \,/\, -0.014$ | $+0.022 \,/\, -0.056$ | $+0.037 \,/\, -0.015$ |
| *other* $\to \ell$ normalisation | $+0.006 \,/\, -0.009$ | $+0.005 \,/\, -0.005$ | $+0.007 \,/\, -0.005$ | $+0.007 \,/\, -0.005$ | $+0.006 \,/\, -0.009$ | $+0.001 \,/\, -0.002$ |
| Jets | $+0.026 \,/\, -0.034$ | $+0.018 \,/\, -0.016$ | $+0.030 \,/\, -0.038$ | $+0.024 \,/\, -0.021$ | $+0.021 \,/\, -0.046$ | $+0.020 \,/\, -0.011$ |
| $E_{\mathrm{T}}^{\mathrm{miss}}$ | $+0.008 \,/\, -0.005$ | $+0.009 \,/\, -0.007$ | $+0.008 \,/\, -0.006$ | $+0.012 \,/\, -0.008$ | $+0.002 \,/\, -0.013$ | $+0.011 \,/\, -0.002$ |
| Electrons and Muons | $+0.005 \,/\, -0.007$ | $+0.006 \,/\, -0.004$ | $+0.005 \,/\, -0.009$ | $+0.007 \,/\, -0.005$ | $+0.004 \,/\, -0.002$ | $+0.003 \,/\, -0.001$ |
| Flavour tagging | $< 0.001 \,/\, -0.001$ | $+0.001 \,/\, < 0.001$ | $+0.006 \,/\, < 0.001$ | $+0.001 \,/\, -0.000$ | $+0.003 \,/\, < 0.001$ | $+0.000 \,/\, -0.001$ |
| Luminosity | $+0.003 \,/\, -0.007$ | $+0.005 \,/\, -0.002$ | $+0.002 \,/\, -0.009$ | $+0.006 \,/\, -0.002$ | $+0.004 \,/\, -0.003$ | $+0.005 \,/\, -0.002$ |

**Table 11.3:** Breakdown of the uncertainties on the signal strength $\mu$. Each group in the row just above a horizontal line is decomposed in sub-groups listed below this line. The square root of the quadratic sum of all single contributions does not add up to the total uncertainty due to correlations between the single groups. The abbreviations *stat.* and *syst.* stand for *statistical* and *systematic*, *unc.* stands for *uncertainty* and *bkg.* for *background.*




**(a)** Search for $H \to e\tau$, nonVBF SR

**(b)** Search for $H \to e\tau$, VBF SR



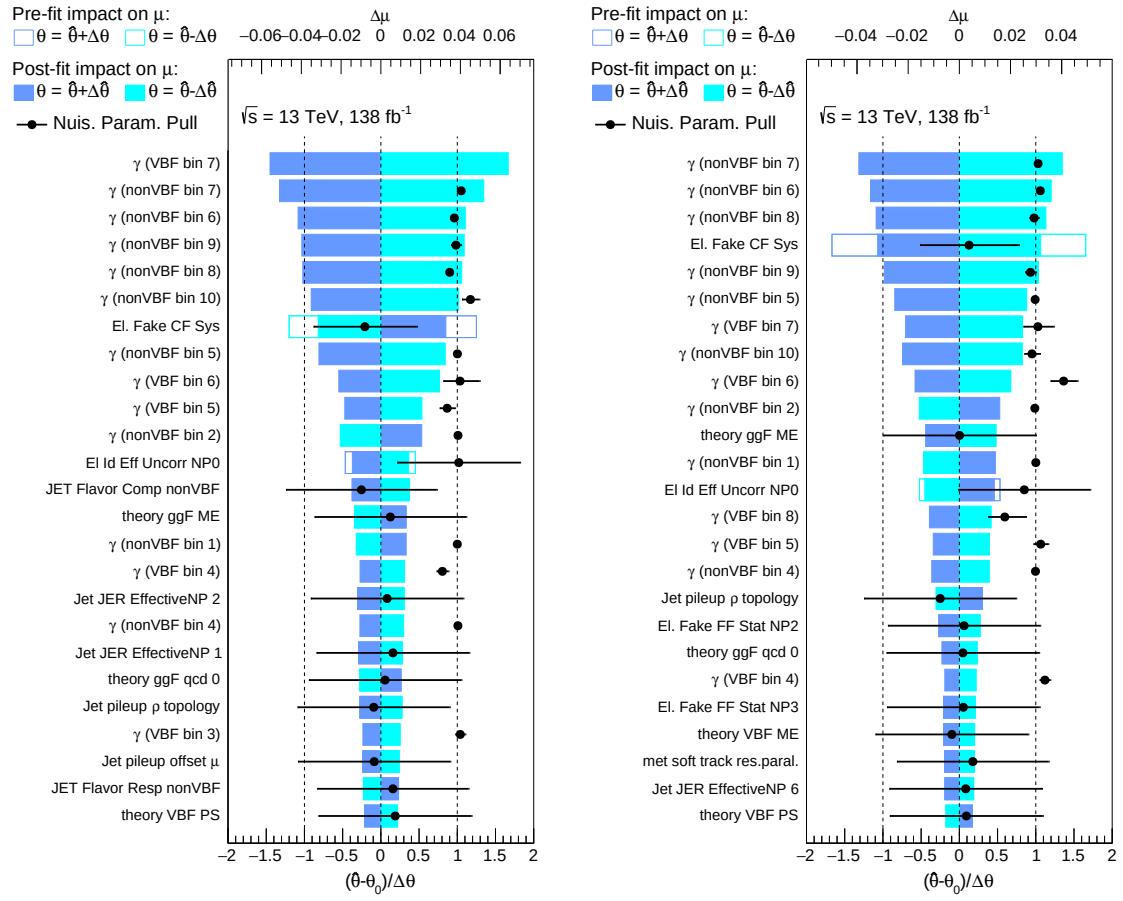

**(c)** Search for $H \to \mu\tau$, nonVBF SR

**(d)** Search for $H \to \mu\tau$, VBF SR

**Figure 11.1:** Post-fit distributions of the final discriminants (neural network output node probability distributions) for the $e\tau$- (top) and $\mu\tau$-dataset (bottom) in the nonVBF SR (left) and in the VBF SR (right), obtained from the combined fit [22]. The binning is equivalent to the one described in Section 8.2.6 but visualised equidistantly to enhance the visibility of the narrow bins. The hashed band indicates the post-fit statistical and systematic uncertainties on the prediction. In addition to the stacked signal corresponding to the best-fit prediction, the signal is overlaid (line) assuming a branching ratio of 0.1 % and scaled by 100. The central panel shows the ratio of data over the prediction (including signal) while the lower panel shows the difference of data and the background prediction.

uncertainties on the background prediction in this analysis is also reflected in the ranking of the NPs (Figures 11.2 and 11.3), where several NPs reflecting these uncertainties are ranked highest. These NPs are denoted by $\gamma(<region>$ bin $<bin\ number>)$ where the bin numbering starts at the left most bin of the final discriminants with 0 and ends at 10 in the nonVBF SR and at 7 (8) for $e\tau$ ($\mu\tau$) in the VBF SR. Some of them experience stronger pulls as they are defined bin-by-bin and thus allow for adjustments of the background prediction for each bin separately.

**Systematic Uncertainties on the Prediction** The largest impact within the systematic uncertainties in the combined fit and the fit in the nonVBF SR comes from the uncertainties on the electron *Correction Factor*s (CFs), followed by uncertainties related to jets, the

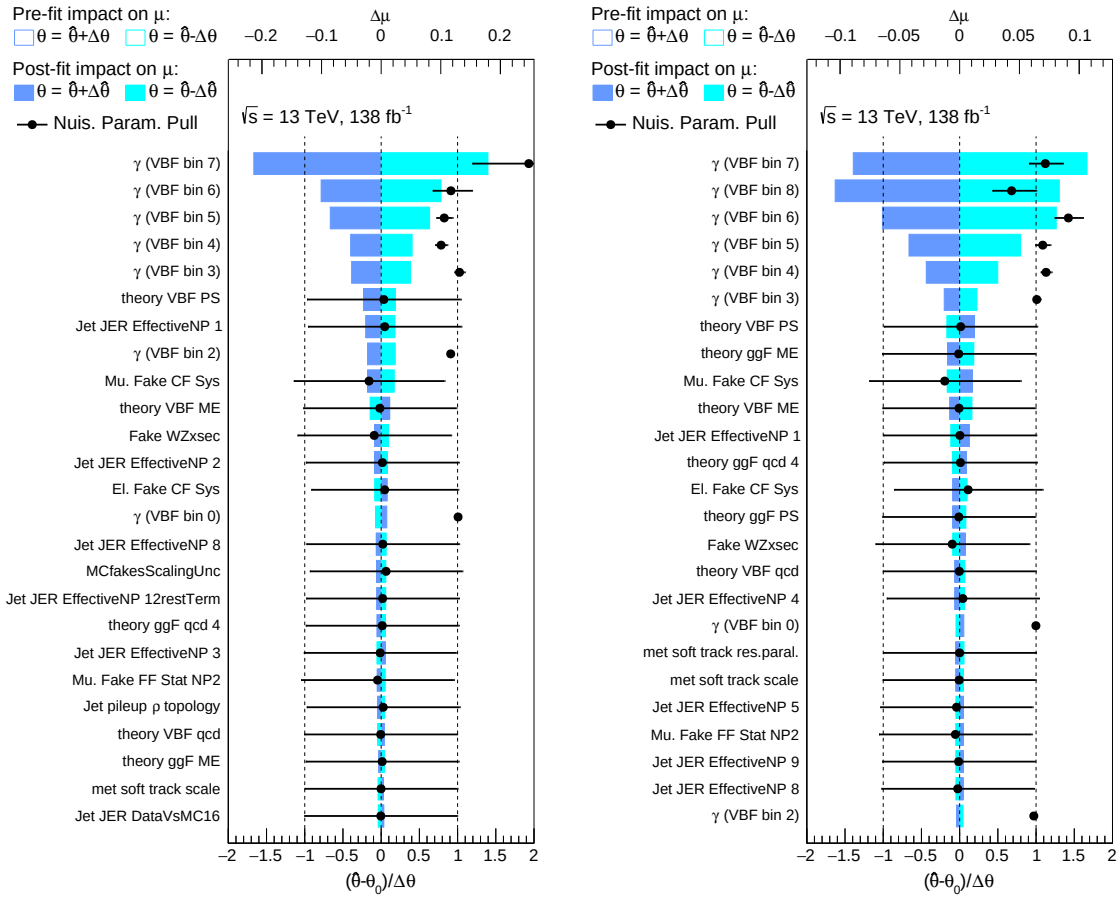



**(a)** Search for $H \to e\tau$ **(b)** Search for $H \to \mu\tau$

**Figure 11.2:** The pulls and constraints (black dots and bars and lower horizontal axis) of the 25 highest ranked nuisance parameters obtained from a combined maximum likelihood fit in both signal regions in the $e\tau$-dataset (a) and the $\mu\tau$-dataset (b). The nominal value of the $\gamma$-parameters is at 1 and deviations from it indicate a pull. For all other NPs the nominal value is at 0 and the pre-fit $\pm 1\sigma$-variations correspond to $\pm 1$. The value of the highest ranked NP in $e\tau$, the $\gamma$-parameter of the last bin of the *VBF NN* distribution, is outside the range and is: $2.66^{+0.58}_{-0.52}$. The post-fit impact on the signal strength $\mu$ (filled coloured bars and the upper horizontal axis) is the basis of the ranking (see Section 10.3). In addition, the pre-fit impacted is shown (empty coloured bars).

efficiency-correction ratio $\mathcal{R}$ and the signal theory uncertainties. For the fit in the VBF SR, the efficiency-correction ratio uncertainties and the electron CF uncertainties have a comparably minor impact. The electron identification $SF$ uncertainties which are correlated between the usual $SF$s applied to all simulated events and the $SF$s used in the efficiency-correction ratio $\mathcal{R}$, are considered in the group "Efficiency correction $\mathcal{R}$" and not in the group "Electrons and Muons". The observations made are also reflected in the NP ranking which gives a more precise insight into which NP(s) of the single groups drive(s) their impact. In the combined fit, apart from the $\gamma$-parameters, the NP *El. Fake CF Sys* describing the systematic uncertainties of the electron CFs which are part of the $j \to \ell$-fake estimate, is ranked highest, followed by *El Id Eff Uncorr NP0* which represents the statistical uncertainty on the electron identification $SF$ in the bin $15\,\text{GeV} \leq p_\text{T} < 20\,\text{GeV}$ and $0 < |\eta| < 1.37$.

The NP *El. Fake CF Sys* is constrained which is also reflected by its reduced post-fit impact compared to its pre-fit impact. The *Mu. Fake CF Sys* NP, however, is not among the highest ranked NPs although its overall pre-fit impact on the $j \to \ell$-fakes distribution of the final discriminant is up to 9.2 %. This is smaller compared to the *El. Fake CF Sys* with 13.2 % and also the shapes are different such that the impact of *Mu. Fake CF Sys* in the sensitive bins is smaller. In addition, a sizeable correlation of around 50 % between




**(a)** Search for $H \to e\tau$ **(b)** Search for $H \to \mu\tau$

**Figure 11.3:** The pulls and constraints (black dots and bars and lower horizontal axis) of the 25 highest ranked nuisance parameters obtained from a maximum likelihood fit in the VBF SR in the $e\tau$-dataset (a) and the $\mu\tau$-dataset (b). The nominal value of the $\gamma$-parameters is at 1 and deviations from it indicate a pull. For all other NPs the nominal value is at 0 and the pre-fit $\pm 1\sigma$-variations correspond to $\pm 1$. The post-fit impact on the signal strength $\mu$ (filled coloured bars and the upper horizontal axis) is the basis of the ranking (see Section 10.3). In addition, the pre-fit impacted is shown (empty coloured bars).

the two NPs is determined in the combined fit. Hence, it seems that *El. Fake CF Sys* partially absorbs effects of *Mu. Fake CF Sys* which results in a reduced importance of the latter. The correlation matrices of the NPs obtained from the combined fit are shown in Figures 11.4 and 11.5 in the search for $H \to e\tau$ and for $H \to \mu\tau$, respectively. Figures showing the correlation matrices obtained from fits in the individual signal regions are given in Appendix E. Furthermore, the NP *El. Fake CF Sys* has a higher rank in the search for $H \to \mu\tau$ than in the search for $H \to e\tau$ but the impact on $\mu$ of $\sim 0.032$ is approximately the same in both searches. The differences in the ranking are rather a result of the lower statistical uncertainties on the background prediction in $H \to \mu\tau$ which are therefore ranked lower. Thus *El. Fake CF Sys* is automatically ranked higher.

The NP *El Id Eff Uncorr NP0* is only slightly constrained but shows a pull of up to $+1\sigma$. Its pre-fit impact on the distributions of the single processes in the final discriminant is not that large, however, due to light tensions between its effect on the "real" and "switch" efficiencies the fit is sensitive to it. After correlating this uncertainty source between the "real" and "switch" efficiencies, it showed a slightly stronger constraint and pull which was investigated using the *mixed* dataset.

Overall, the impact of the systematic uncertainties on the total uncertainty is similar in both searches. However, the uncertainties related to jets are larger in the search for $H \to e\tau$

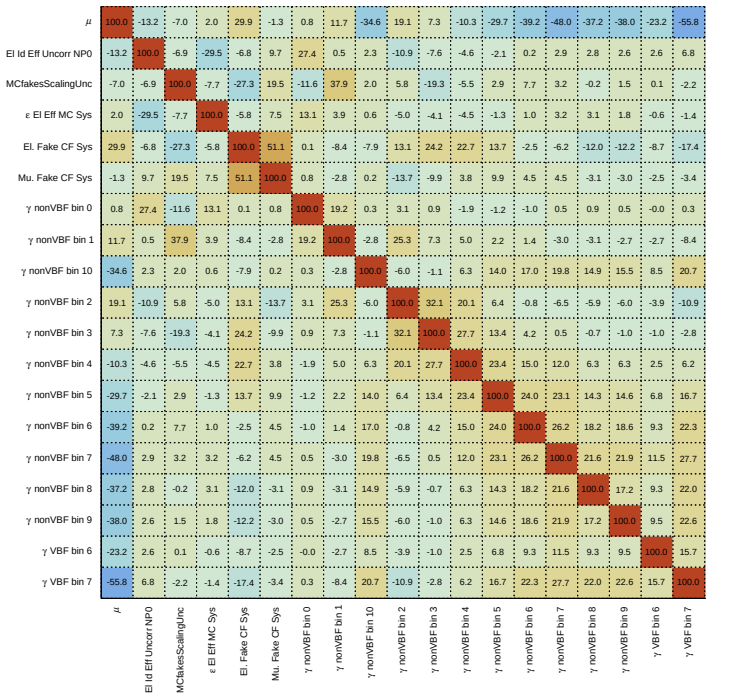



**Figure 11.4:** The correlations in percent of all nuisance parameters that have at least one correlation with an absolute value larger than 20 %, obtained from a combined maximum likelihood fit in both signal regions with the $e\tau$-dataset.

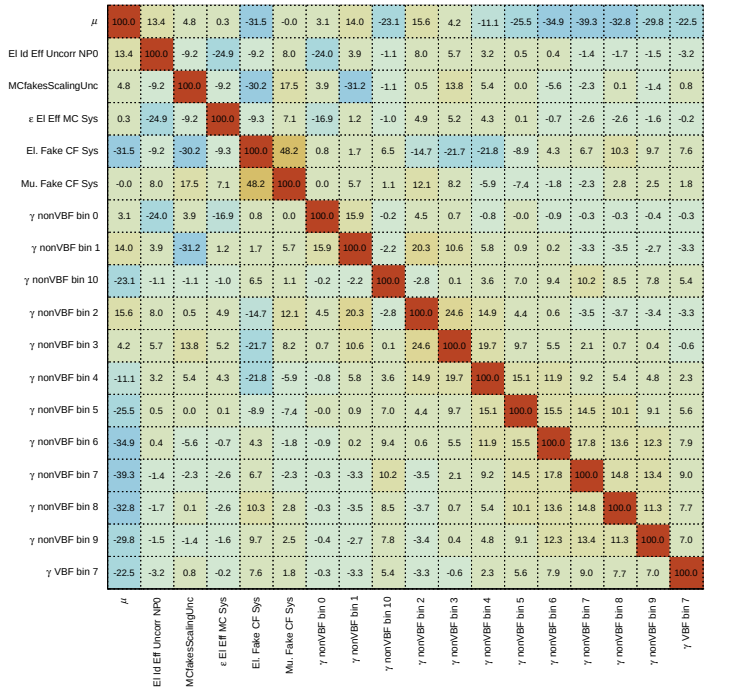



**Figure 11.5:** The correlations in percent of all nuisance parameters that have at least one correlation with an absolute value larger than 20 %, obtained from a combined maximum likelihood fit in both signal regions with the $\mu\tau$-dataset.

than in the search for $H \to \mu\tau$. This is, again, also reflected in the ranking of the NPs of the combined fit. In $H \to e\tau$, the first jet-related NP, *Jet Flavor Comp nonVBF*, is at rank 13 while the first jet-related NP, *Jet pileup $\rho$ topology*, in $H \to \mu\tau$ is at rank 17 (this NP is at rank 21 in $H \to e\tau$). In addition, *Jet JER EffectiveNP 2* and *Jet JER EffectiveNP 1* are also ranked higher in $H \to e\tau$ but are not among the first 25 in $H \to \mu\tau$. Although the NP *Jet pileup $\rho$ topology* has indeed different ranks for $H \to e\tau$ and $H \to \mu\tau$, the absolute value of its $\Delta\mu$ of $\sim 0.01$ is approximately the same in both searches. This is not the case for the other mentioned NPs. For them, the value of $\Delta\mu$ is larger by a factor 5 to 10 for $H \to e\tau$. When comparing their effects on the distributions of the final discriminant of the individual processes, it is found that it is slightly larger in the right-most, i.e. signal-like bins in $H \to e\tau_\mu$. This is consistently the case for all processes.

**Theoretical Uncertainties on the Signal Prediction** The impact of the signal theory uncertainties on the up- and down-uncertainty of $\mu$ is asymmetric. For $H \to e\tau$, the down-uncertainty is larger, for $H \to \mu\tau$ the up-uncertainty is larger. This is more pronounced for the fit in the VBF SR. The largest impact on the value of $\mu$ and hence the highest ranked single NPs of this group are *theory ggF ME* and *theory ggF qcd 0* in the combined (and also in the nonVBF SR) fit while it is *theory VBF PS* and *theory VBF ME/ theory ggF ME* in the fit in the VBF SR for $H \to e\tau/H \to \mu\tau$. Their up- and down-variations with respect to nominal are perfectly symmetric as only one variation is available which is mirrored to obtain the other variation. Hence, the opposite asymmetric behaviour in the searches for $H \to e\tau$ and $H \to \mu\tau$ probably results from the opposite signs of the respctive best-fit values $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$.

**Correlations of Nuisance Parameters** Other nuisance parameters with sizeable correlations but which were not discussed yet as they do not have a large impact on $\mu$ are *MCfakesScalingUnc* and *$\varepsilon$ El Eff MC Sys.* The former shows a correlation of 38 % in $e\tau$ and of $-31$ % in $\mu\tau$ with the statistical uncertainties on the background prediction of the second bin of the final discriminant in the nonVBF SR. This correlation is reasonable as the bins to the left have large background contributions and hence serve as control regions to constrain the background predictions. Due to the same reason other NPs have sizeable correlations with the statistical uncertainties on the background prediction of the left bins; these are *El Id Eff Uncorr NP0* and *El. Fake CF Sys.* Furthermore, there are also correlations of up to $\sim 30$ % between the statistical uncertainties of the different bins. The opposite sign of the correlation of *MCfakesScalingUnc* and the statistical uncertainty in the two searches originates from the fact that this uncertainty is dominated by the $Z \to \mu\mu$-contribution to the *other* $\to \ell$-fakes which is more abundant in the $\mu\tau$-dataset which enters the background estimate with a positive sign in the search for $H \to \mu\tau$ and with a negative sign as part of the $e/\mu$-symmetric background in the search for $H \to e\tau$. The NP *$\varepsilon$ El Eff MC Sys* which is an uncertainty on the efficiency correction ratio $\mathcal{R}$ shows a correlation of $-30$ % with *El Id Eff Uncorr NP0.* Also the latter is dominated by uncertainties on $\mathcal{R}$.

### 11.1.2 Validation of the Compatibility among Signal Regions

A hypothesis test is performed in order to quantify the compatibility of the best-fit values of the signal strengths in the two signal regions. The default fit setup with one *parameter of interest* (POI) for both signal regions serves as null hypothesis, claiming that the same signal strength in both regions can be assumed. A fit setup with two POIs, one for each of the two signal regions, serves as alternative hypothesis. In both fit setups a combined fit is performed in both signal regions simultaneously.

The resulting best-fit values for the two POIs of the alternative hypothesis are summarised in Table 11.4. They are very close to the values obtained when performing

independent fits in each of the signal regions (Table 11.1).

| | nonVBF SR | VBF SR |
|---|---|---|
| $\hat{\mu}^{e\tau}$ | $-0.376^{+0.114}_{-0.120}$ | $-0.084^{+0.219}_{-0.249}$ |
| $\hat{\mu}^{\mu\tau}$ | $0.298^{+0.114}_{-0.111}$ | $0.055^{+0.214}_{-0.192}$ |

**Table 11.4:** The observed best-fit values for the signal strength $\mu$ from combined 2-POI fits (1 POI per signal region) in the search for $H \to e\tau$ and $H \to \mu\tau$.

The test statistic $t_{\text{SRs}}$ to quantify the compatibility among the regions is again a likelihood ratio:

$$t_{\text{SRs}} = -2 \ln \frac{\mathcal{L}_{\text{1POI}}}{\mathcal{L}_{\text{2POI}}} = 2 \left( \ln \mathcal{L}_{\text{2POI}} - \ln \mathcal{L}_{\text{1POI}} \right) , \tag{11.3}$$

with the null hypothesis in the numerator and the alternative hypothesis in the denominator. As noted in Section 10.2, this test statistic follows a $\chi^2$-distribution with 1 degree of freedom corresponding to the difference in the number of POIs $(2 - 1)$. The corresponding negative log-likelihood values and the resulting $p$-value and significance are summarised in Table 11.5. For the significance, a two-sided fluctuation of a Gaussian-distributed variable is assumed. The null hypothesis cannot be rejected and hence the signal regions show a compatible behaviour.

| | | $-\ln \mathcal{L}$ | $p$ | $Z/\sigma$ |
|---|---|---|---|---|
| $H \to e\tau$ | 1 POI | -2182526.009223 | 0.281 | 1.078 |
| | 2 POI | -2182526.590176 | | |
| $H \to \mu\tau$ | 1 POI | -2029627.347605 | 0.309 | 1.017 |
| | 2 POI | -2029627.865049 | | |

**Table 11.5:** The negative log-likelihood values of the 1-POI and 2-POI combined fits in the searches for $H \to e\tau$ and $H \to \mu\tau$, respectibely and the resulting measures for compatibility, $p$-values and significance in terms of standard deviations $\sigma$ (assuming a two-sided fluctuation of a Gaussian-distributed variable).

### 11.1.3 Investigation and Visualisation of the Relation of $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$

The *Symmetry Method* is eventually only sensitive to the difference of branching ratios of the two LFV signals as discussed in Chapter 3 and Section 10.4. In order to investigate and visualise the relation of the two signals, the presence of a $H \to \mu\tau_e$ signal is considered in the fit for the search for $H \to e\tau_\mu$. A contribution of $H \to \mu\tau_e$ with a fixed signal strength $\mu^{\mu\tau}_{\text{fixed}}$ is included in the likelihood (Eq. 10.24) and the behaviour of $\hat{\mu}^{e\tau}$ is investigated. The modified likelihood for the search for $H \to e\tau$ is given in the following where the new additional terms are indicated in red and are underlined.

$$\mathcal{L}(\mu, \boldsymbol{\theta}, \gamma_i | \boldsymbol{n}^{e\tau}) =$$
$$\prod_{i=1}^{N} P_{\mathcal{P}} \left( n_i^{e\tau} | \gamma_i \left( \tilde{n}_i^{\mu\tau}(\boldsymbol{\theta}) + f^{e\tau}_{j\to\ell,i}(\boldsymbol{\theta}) + f^{e\tau}_{other\to\ell,i}(\boldsymbol{\theta}) \right) \underline{+\mu^{\mu\tau}_{\text{fixed}} \cdot s^{e\tau}_{H\to\mu\tau,i}(\boldsymbol{\theta})} + \mu^{e\tau} \cdot s^{e\tau}_{H\to e\tau,i}(\boldsymbol{\theta}) \right) \tag{11.4}$$

with

$$\tilde{n}_i^{\mu\tau} = \mathcal{R}^{\mu\tau} \cdot \left( n_i^{\mu\tau} - f_{j\to\ell,i}^{\mu\tau}(\boldsymbol{\theta}) - f_{other\to\ell,i}^{\mu\tau}(\boldsymbol{\theta}) - \mu_{\text{fixed}}^{\mu\tau} \cdot s_{H\to\mu\tau,i}^{\mu\tau} - \mu^{e\tau} \cdot s_{H\to e\tau,i}^{\mu\tau}(\boldsymbol{\theta}) \right) . \tag{11.5}$$

The $H \to \mu\tau$ signal is treated as part of the background prediction since a possible $H \to \mu\tau$ signal or up-fluctuation in the $\mu\tau$-data propagates to the $e/\mu$-symmetric background estimate used in the search for $H \to e\tau$. The contribution of $\mu_{\text{fixed}}^{\mu\tau} s_{H\to\mu\tau}^{\mu\tau}$, which is subtracted, is larger than the contribution $\mu_{\text{fixed}}^{\mu\tau} s_{H\to\mu\tau}^{e\tau}$, which is added, and hence the new overall background prediction is reduced when including the $H \to \mu\tau$ signal and assuming $\mu_{\text{fixed}}^{\mu\tau} > 0$. The reduced background prediction, however, causes issues with low background yields in the last bin of the final discriminant in the VBF SR. Therefore, the last two bins are merged for this study.

First, the signal strength $\mu_{\text{fixed}}^{\mu\tau}$ is set to 0.25 which is the best-fit value obtained from the original combined fit (Table 11.1). The obtained best-fit signal strength $\hat{\mu}^{e\tau}$ for $\mu_{\text{fixed}}^{\mu\tau} = 0.25$ is summarised in Table 11.6. This study was performed before the signal parton shower and the signal matrix element uncertainties were included in the statistical analysis. Hence, the results of the original fit without these uncertainties are additionally quoted in the table. The last two bins in the VBF SR were not merged in this fit. These results are denoted by $\mu_{\text{fixed}}^{\mu\tau} = 0$ since not considering the $H \to \mu\tau$ signal in the likelihood equals the assumption that this signal is zero. The changes compared to Table 11.1 where the additional signal uncertainties are included are mininmal. The best-fit signal strength $\hat{\mu}^{e\tau}$

| | combined | nonVBF SR | VBF SR |
|---|---|---|---|
| $\hat{\mu}^{e\tau}$ if $\mu_{\text{fixed}}^{\mu\tau} = 0$ | $-0.329^{+0.104}_{-0.108}$ | $-0.377^{+0.114}_{-0.119}$ | $-0.089^{+0.221}_{-0.247}$ |
| $\hat{\mu}^{e\tau}$ if $\mu_{\text{fixed}}^{\mu\tau} = 0.25$ | $-0.081^{+0.102}_{-0.106}$ | $-0.121^{+0.113}_{-0.119}$ | $0.090^{+0.218}_{-0.236}$ |

**Table 11.6:** The observed results obtained in the search for $H \to e\tau_\mu$ when considering a $H \to \mu\tau$-signal with a branching ratio of 0.25 % (best-fit value). The upper row summarises the results obtained with the original fit where no $H \to \mu\tau$-signal is considered. The last two bins of the final discriminant in the VBF SR are merged when $\mu_{\text{fixed}}^{\mu\tau} = 0.25$ to circumvent issues with low yields.

for $\mu_{\text{fixed}}^{\mu\tau} = 0.25$ is now compatible with 0 for the combined fit and the fits in both regions separately. The change in signal strength with respect to the original fit corresponds to the expected difference of 0.25 in the nonVBF SR. This is not exactly the case in the VBF SR, and therefore also not for the combined fit, which is expected due to the change in binning.

In a next step, the fit in the search for $H \to e\tau$ is repeated 10 times for different values of $\mu_{\text{fixed}}^{\mu\tau}$. It is scanned from 0.0 to 0.50 in steps of 0.05. Each time, the corresponding best-fit value $\hat{\mu}^{e\tau}$ is obtained and its dependence on $\mu_{\text{fixed}}^{\mu\tau}$ is shown in Figure 11.6. The scan is performed with fits in the nonVBF SR only due to low yields and eventually negative background predictions in the VBF SR for higher values of $\mu_{\text{fixed}}^{\mu\tau}$. A clear linear dependence of the two signal strengths is visible which indicates a direct correlation between them. The orange cross indicates the best-fit value $\hat{\mu}^{e\tau}$ in the nonVBF SR when assuming $\mu_{\text{fixed}}^{\mu\tau} = 0$ which corresponds to the original fit. The red cross at $\mu_{\text{fixed}}^{\mu\tau} = 0.3$ corresponds to the best-fit value of $\hat{\mu}^{\mu\tau}$ in the nonVBF SR of 0.299 and results in $\hat{\mu}^{e\tau} = -0.070^{+0.113}_{-0.118}$ which is compatible with zero.

This study clearly shows the direct connection between $\mu^{e\tau}$ and $\mu^{\mu\tau}$. However, the conclusion that the upwards fluctuation observed in the $\mu\tau$-dataset actually originates from a $H \to \mu\tau$ signal while no $H \to e\tau$ signal is observed, as perhaps suggested by the previous elaborations, would be wrong. On one hand, the significance for the upwards fluctuation (Table 11.1) is not large enough to claim evidence for a new signal and on the other hand

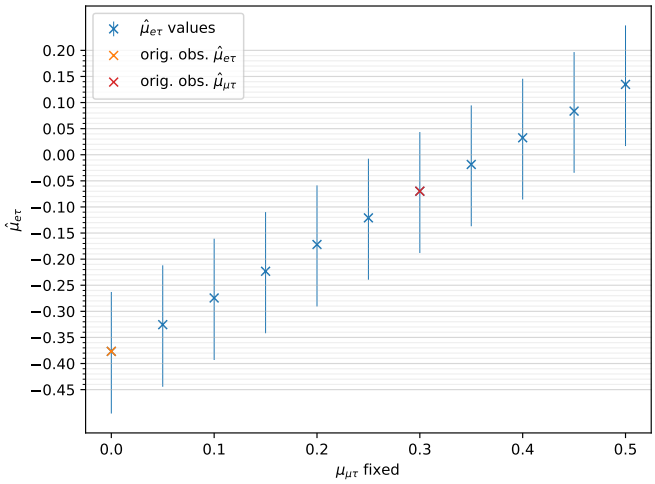



**Figure 11.6:** Dependence of the best-fit values $\hat{\mu}^{e\tau}$ in the search for $H \to e\tau$ as a function of different fixed $\mu^{\mu\tau}_{\text{fixed}}$ in the nonVBF SR where the $H \to \mu\tau$ signal is considered as part of the background estimate. The orange cross indicates the best-fit value $\hat{\mu}^{e\tau}$ when assuming $\mu^{\mu\tau}_{\text{fixed}} = 0$. The red cross at $\mu^{\mu\tau}_{\text{fixed}} = 0.3$ corresponds to the best-fit value of $\hat{\mu}^{\mu\tau}$ of 0.299.

one cannot judge whether the excess observed in the $\mu\tau$-dataset and the deficit observed in the $e\tau$-dataset originate solely from an up-fluctuation in $\mu\tau$, solely from a down-fluctuation in $e\tau$ or from both.

### 11.1.4 Compatibility of the Results of the Searches for $H \to e\tau_\mu$ and $H \to \mu\tau_e$ and Measurement of $\Delta = \mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$

The linear dependence of $\mu^{e\tau}$ and $\mu^{\mu\tau}$ discussed in the previous section illustrates that both parameters essentially describe the same observable (apart from its sign), which is the difference of both branching ratios, $\Delta$. In the search for $H \to e\tau$ the difference $\Delta^{e\tau} = \mathcal{B}(H \to e\tau) - \mathcal{B}(H \to \mu\tau)$ is measured, in the search for $H \to \mu\tau$ the difference $\Delta^{\mu\tau} = \mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ is measured for which $\Delta^{e\tau} = -\Delta^{\mu\tau}$ should hold. Therefore, the absolute values of the best-fit signal strengths $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$ should be compatible with each other. This is equivalent to requiring that the sum of $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$ is compatible with 0 which allows to formulate the following $\chi^2$ test in order to quantify the compatibility.

$$\chi^2 = \frac{((\hat{\mu}^{e\tau} + \hat{\mu}^{\mu\tau}) - 0)^2}{e^2} \tag{11.6}$$

with

$$e^2 = e^2_{\hat{\mu}^{e\tau}} + e^2_{\hat{\mu}^{\mu\tau}} + 2e_{\hat{\mu}^{e\tau}} e_{\hat{\mu}^{\mu\tau}} \rho\,. \tag{11.7}$$

Here, $e_{\hat{\mu}^{e\tau}}$ ($e_{\hat{\mu}^{\mu\tau}}$) is the mean value of the up and down uncertainty on $\hat{\mu}^{e\tau}$ ($\hat{\mu}^{\mu\tau}$) and $\rho$ the correlation coefficient between $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$. The values of all variables except $\rho$ are known and given in Table 11.1. In order to obtain the value of $\rho$, a bootstrap [261, 262] analysis is performed. With the bootstrap technique, statistical correlations between measurements can be determined by generating pseudo-experiments in a coherent way. The code documented in [262] is utilised to conduct the analysis. 500 replicas of the data samples as well as of the data part of the $e/\mu$-symmetric background estimates are created. It is ensured that the $e\tau(\mu\tau)$-data used as data in the search for $H \to e\tau(H \to \mu\tau)$ and as part of the $e/\mu$-symmetric background estimate in the search for $H \to \mu\tau(H \to e\tau)$ is fluctuated coherently. Furthermore, the systematic uncertainties on the data part of the

$e/\mu$-symmetric background estimates are varied accordingly and the statistical uncertainties on the total background prediction are recalculated for each replica. The combined fit in both signal regions and for both searches, $H \to e\tau$ and $H \to \mu\tau$, including all uncertainties is performed for each replica dataset and the respective signal strengths $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$ are obtained. The results are visualised in Figure 11.7. The absolute value of the best-fit signal

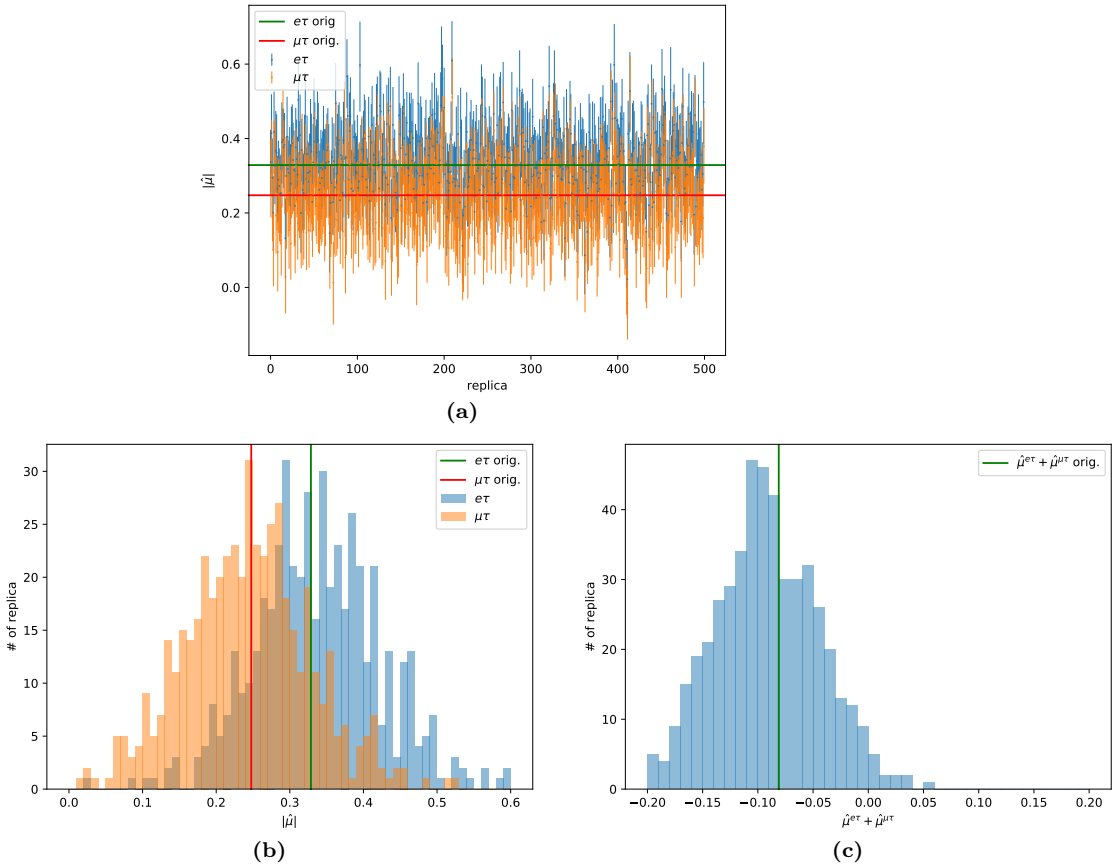



**Figure 11.7:** The absolute values of the best-fit signal strengths $\hat{\mu}$ for each of the 500 bootstrap replicas (a), the number of the replica as a function of the absolute value of the best-fit signal strength in bins of width 0.01 (b) and the number of replicas as a function of the sum of the two best-fit signal strengths (c). In (a) and (b) blue indicates the results obtained from fits in the search for $H \to e\tau$ and orange the results obtained from fits in the search for $H \to \mu\tau$. The lines indicate the results of the original fits.

strengths, $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$, are shown for each replica in Figure 11.7(a) while Figure 11.7(b) shows the number of replicas against the absolute value of the best-fit signal strength in bins of width 0.01. Both figures indicate that the mean of $|\hat{\mu}^{e\tau}|$ is larger than the mean of $|\hat{\mu}^{\mu\tau}|$. Figure 11.7(c) shows the number of replicas against the sum of the best-fit signal strengths, $\hat{\mu}^{e\tau} + \hat{\mu}^{\mu\tau}$, in bins of width 0.01. Concordant to the previous observation, the mean of the sum is shifted to negative values. Possible reasons for these small differences in absolute values of the signal strengths could be the different binning of the final discriminants and the different coefficients used in the combined *VBF NN* distribution (Section 8.2.5) as well as the different statistics of the background predictions in the $e\tau$- and $\mu\tau$-dataset and thus, in addition, different impacts of the uncertainties.

The correlation coefficient $\rho$ is calculated from the results of the $N_{\text{rep}} = 500$ replicas

using the usual equations for the covariance (cov):

$$\mathrm{cov}(\hat{\mu}^{e\tau}, \hat{\mu}^{\mu\tau}) = \frac{1}{N_{\mathrm{rep}}} \sum_{i=1}^{N_{\mathrm{rep}}} \left(\hat{\mu}_i^{e\tau} - \overline{\hat{\mu}}^{e\tau}\right) \left(\hat{\mu}_i^{\mu\tau} - \overline{\hat{\mu}}^{\mu\tau}\right) , \tag{11.8}$$

$$\sigma_{\hat{\mu}^{x\tau}} = \sqrt{\mathrm{cov}\left(\hat{\mu}^{x\tau}, \hat{\mu}^{x\tau}\right)} , \tag{11.9}$$

$$\rho = \mathrm{cov}\left(\hat{\mu}^{e\tau}, \hat{\mu}^{\mu\tau}\right) / \left(\sigma_{\hat{\mu}^{e\tau}} \sigma_{\hat{\mu}^{\mu\tau}}\right) . \tag{11.10}$$

Here, $x \in \{e, \mu\}$ and $\overline{\hat{\mu}}^{x\tau}$ denotes the mean value. The correlation coefficient is calculated to be $\rho = -0.80$ which allows for calculating the $\chi^2$-value using Eqs. 11.6 and 11.7 and the results in Table 11.1. The obtained $\chi^2$-values for the results of the combined fit and of the fits in only one of the signal regions are summarised in Table 11.7 using the same value for $\rho$ in all three columns. Additionally, the corresponding $p$-value and singificance $Z$ are listed. Both measurements of the signal strength and therefore of the difference in

| | combined | nonVBF SR | VBF SR |
|---|---|---|---|
| $\chi^2$-value | 1.528 | 1.114 | 0.045 |
| $p$-value | 0.216 | 0.291 | 0.832 |
| $Z/\sigma$ | 1.236 | 1.056 | 0.212 |

**Table 11.7:** The $\chi^2$-values with the corresponding $p$-values and singificances $Z$ obtained to quantify the compatibility of the best-fit signal strengths $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$.

branching ratios of $H \to e\tau$ and $H \to \mu\tau$, $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$, agree within $1.2\,\sigma$.

The agreement of both measurements improves further if only statistical uncertainties on the prediction and no systematic uncertainties are considered. In this case a correlation coefficient of $\rho = -0.86$ is obtained and the measurements of both signal strengths agree within $0.3\,\sigma$. The corresponding best-fit signal strengths, $\chi^2$-value, $p$-value and significance for the combined fit are summarised in Table 11.8. However, the values of the best-fit signal strengths themselves are not relevant as they do not have a reasonable physical meaning without considering all relevant uncertainties.

| $\hat{\mu}^{e\tau}$ | $\hat{\mu}^{\mu\tau}$ | $\chi^2$-value | $p$-value | $Z/\sigma$ |
|---|---|---|---|---|
| $-0.336^{+0.084}_{-0.084}$ | $0.322^{+0.082}_{-0.082}$ | 0.105 | 0.746 | 0.324 |

**Table 11.8:** The best-fit signal strengths $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$ obtained from a combined fit in both signal regions neglecting systematic uncertainties and considering only statistical uncertainties on the background prediction. In addition, the $\chi^2$-value with the corresponding $p$-value and singificance $Z$ obtained to quantify the compatibility of the best-fit signal strengths.

Since both measurements of the signal strength and hence of the branching ratio difference agree with each other, the measurement with the smaller expected uncertainty is chosen as final result of the analysis. The best-fit signal strengths obtained with the *pre-fit Asimov* dataset (see Section 10.3), assuming 1 % branching ratio for the signal templates, is $\hat{\mu}^{e\tau} = 1.000^{+0.148}_{-0.135}$ and $\hat{\mu}^{\mu\tau} = 1.000^{+0.139}_{-0.128}$. Hence, smaller uncertainties are expected from the fit in the $\mu\tau$-dataset and thus the results obtained in the search for $H \to \mu\tau$ are chosen as the main results.

Therefore, a best-fit value of the branching ratio difference

$$\hat{\Delta}^{\mu\tau} = \mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau) = 0.249^{+0.103}_{-0.100}\,\% \tag{11.11}$$

is obtained, suggesting an upward fluctuation of the $\mu\tau$-data compared to the $e\tau$-data; although this fluctuation corresponding to $2.5\,\sigma$ is not highly significant. An upper limit on the branching ratio difference of

$$\Delta_{95}^{\mu\tau} = 0.422\,\%\,(0.191^{+0.076}_{-0.053}\,\%) \tag{11.12}$$

is derived where the expected upper limit is given in brackets. The observed as well as the expected upper limit are visualised in Figure 11.8, assuming that the limit is on the absolute value of the branching ratio difference. The observed limit is just above the $2\,\sigma$-band of the

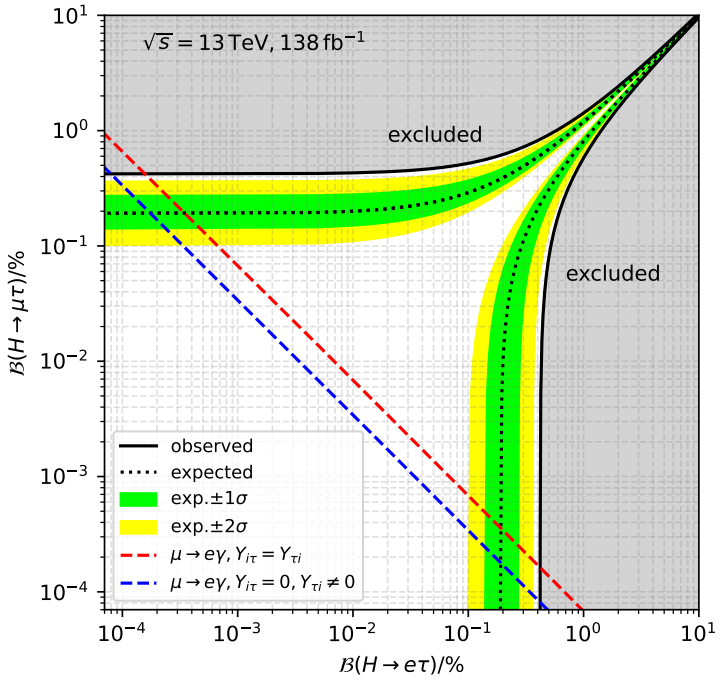



**Figure 11.8:** The observed and expected upper limit on the difference $\Delta^{\mu\tau}$ (black solid and black dotted lines, respectively), assuming the limit is on the absolute value of the branching ratio difference. The red and blue dashed lines indicates the bound derived from the limit on $\mathcal{B}(\mu \to e\gamma)$ using different assumptions on the $Y_{i\tau}$; many thanks to [263] for the support.

expected upper limit as anticipated from the observed fluctuation.

In addition, a bound derived from the limit on the branching ratio of the decay $\mu \to e\gamma$ is drawn. The indirect limits on the LFV decays of the Higgs boson are discussed in Section 1.4.2 where Table 1.5 includes the bound on the combination $Y_{e\tau}Y_{\tau\mu}$: $\sqrt[4]{|Y_{e\tau}Y_{\tau\mu}|^2 + |Y_{\tau e}Y_{\mu\tau}|^2} < 2.2 \times 10^{-4}$, with $Y_{ij}$ being the off-diagonal Yukawa couplings. Eq. 1.54 allows for translating this bound into a simultaneous bound on the branching ratios of $H \to e\tau$ and $H \to \mu\tau$ which is displayed in Figure 11.8. The details are given in Appendix F. While deriving the bound on the branching ratios an assumption on the relation of $Y_{i\tau}$ and $Y_{\tau i}$ is required. For the red curve in Figure 11.8, it is assumed that they are equal: $Y_{i\tau} = Y_{\tau i}$. Whereas for the blue curve it is assumed that one of them is zero while the other is not: $Y_{i\tau} = 0, Y_{\tau i} \neq 0$ or $Y_{i\tau} \neq 0, Y_{\tau i} = 0$. The bounds show that if the branching ratio of $H \to \mu\tau$ were equal to the limit of 0.42 %, the branching ratio of $H \to e\tau$ would only be allowed to have a maximum value of $1.6 \times 10^{-4}$ % ($8.1 \times 10^{-5}$ %) using the red (blue) bound. If both branching ratios had the same size, in which case no signal could be detected with the *Symmetry Method*, they would at most be $8.28 \times 10^{-3}$ % ($5.86 \times 10^{-3}$ %) which is below the current sensitivity. This means that if there is a LFV signal in the order of magnitude of the current sensitivity it will be detected with the *Symmetry Method*.

## 11.2 Combination with the Hadronic Channel

This section briefly discusses the combination with the partner analysis exploiting the hadronic $\tau$-lepton decays in the search for LFV decays of the Higgs boson. Two LFV-decay channels are considered: on one hand one electron and one hadronically decaying $\tau$-lepton in the final state ($e\tau_{\text{had}}$) and on the other hand one muon and one hadronically decaying $\tau$-lepton in the final state ($\mu\tau_{\text{had}}$). The estimation of all background processes apart from fakes is based on templates from MC simulations, thus this analysis is denoted as "MC-based lephad" in the following. The fakes are events passing the selection where a jet mimics the visible part of a hadronically decaying $\tau$-lepton ($\tau_{\text{had-vis}}$). They are estimated with an adapted version of the *Fake Factor Method*. The largest background contribution is from $Z \to \tau\tau$-events and is normalised from the most background-like bins of the final discriminants. The MC-based lephad analysis also makes use of two signal regions, the nonVBF SR and the VBF SR. In each of the regions and separately for the $e\tau_{\text{had}}$ and the $\mu\tau_{\text{had}}$ decay channels, two or three *Boosted Decision Tree*s (BDTs) are trained, each focusing on the separation of one background class and signal. For each region and channel, the single BDTs are combined into one final discriminant, similar to the method described for the VBF SR of the analysis described in this thesis (Section 8.2.5). More details of the MC-based lephad analysis can be found in [22].

The post-fit distributions of the final discriminants (the combined BDT scores) are shown in Figure 11.9. The fit performed to obtain these distributions is a 2-POI fit where the two signal strengths $\mu^{e\tau}$ and $\mu^{\mu\tau}$ (the 2 POIs) are extracted simultaneously by performing the fit simultaneously in the $e\tau_{\text{had}}$- and the $\mu\tau_{\text{had}}$-channel. In addition, the fit combined the MC-based lephad and the MC-based leplep analyses. More information can be found in [22].

In order to combine the *Symmetry Method*-based leplep analysis, in short referred to as "symmetry-based leplep", with the MC-based lephad analysis, separate 1-POI fits are performed in the search for $H \to e\tau$ and for $H \to \mu\tau$, respectively. A simultaneous 2-POI fit is not feasible for the symmetry-based leplep analysis due to the direct correlation of the data part of the $e/\mu$-symmetric background in one search and the data in the other search.

The combination requires the assumption that the branching ratio of the respective other signal is zero. This allows interpreting the signal strengths $\mu$ as branching ratios in percent. Then, a simultaneous fit in all four regions, symmetry-based leplep nonVBF SR and VBF SR and MC-based lephad nonVBF SR and VBF SR, is performed; once in the search for $H \to e\tau$ and once in the search for $H \to \mu\tau$. The results of the best-fit signal strengths, significances and upper limits are summarised in Table 11.9. In addition, the limits are visualised in Figure 11.10. The breakdown of the uncertainties on the signal strength $\mu$, i.e. the impact of groups of uncertainty sources on the uncertainty of $\mu$, is summarised in Table 11.10 and the ranking of the 25 highest ranked NPs is shown in Figure 11.11.

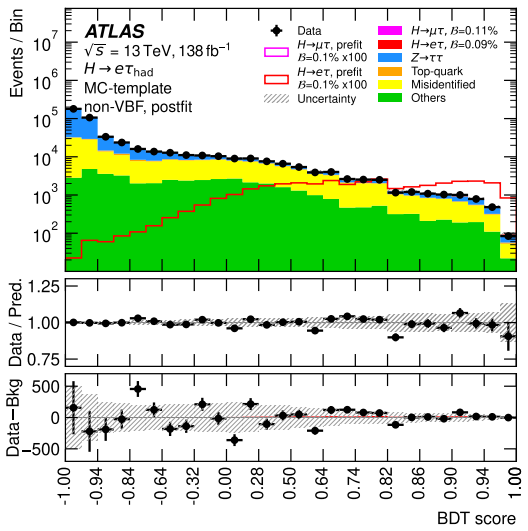
As discussed before, $\hat{\mu}^{e\tau}$ of the symmetry-based leplep analysis in the search for $H \to e\tau$ is $-0.33^{+0.10}_{-0.11}$ while $\hat{\mu}^{e\tau}$ of the MC-based analysis is $0.05^{+0.09}_{-0.09}$ which results in a combined best-fit value $\hat{\mu}^{e\tau}$ of $-0.12^{+0.07}_{-0.07}$ with a significance of $-1.84\,\sigma$ (under the assumption that $\mathcal{B}(H \to \mu\tau) = 0$). Therefore, the observed upper limit on the signal strength of 0.07 is below the expected upper limit of $0.13^{+0.05}_{-0.04}$ but still within the $2\,\sigma$-band of the expected upper limit.

In the search for $H \to \mu\tau$ the best-fit values of $\mu^{\mu\tau}$ of both analyses are larger than 0 which results in a combined $\hat{\mu}^{\mu\tau}$ of $0.13^{+0.05}_{-0.05}$ with a significance of $2.88\,\sigma$ (under the assumption that $\mathcal{B}(H \to e\tau) = 0$). Therefore, the observed upper limit of 0.22 is above the expected upper limit of $0.09^{+0.04}_{-0.03}$ exceeding the $2\,\sigma$-band.

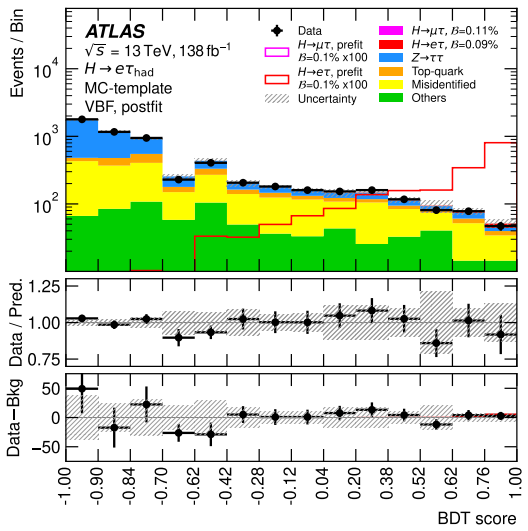
The combination of both analyses is still dominated by the uncertainties due to the

| | combined | Symmetry-based Leplep | | MC-based Lephad | |
|---|---|---|---|---|---|
| | | nonVBF SR | VBF SR | nonVBF SR | VBF SR |
| $\hat{\mu}^{e\tau}$ | $-0.124^{+0.068}_{-0.069}$ | $-0.375^{+0.114}_{-0.121}$ | $-0.090^{+0.222}_{-0.253}$ | $0.034^{+0.123}_{-0.124}$ | $0.066^{+0.129}_{-0.127}$ |
| $Z/\sigma$ | $-1.842$ (9.202) | $-3.399$ (0.879) | $-0.391$ (0.507) | 0.281 (6.279) | 0.517 (4.614) |
| $\mu_{95}^{e\tau}$ | 0.074 ($0.128^{+0.051}_{-0.036}$) | 0.083 ($0.217^{+0.088}_{-0.061}$) | 0.361 ($0.400^{+0.168}_{-0.112}$) | 0.267 ($0.244^{+0.097}_{-0.068}$) | 0.309 ($0.260^{+0.111}_{-0.073}$) |
| $\hat{\mu}^{\mu\tau}$ | $0.134^{+0.049}_{-0.047}$ | $0.299^{+0.114}_{-0.111}$ | $0.060^{+0.217}_{-0.195}$ | $0.120^{+0.066}_{-0.063}$ | $0.049^{+0.099}_{-0.092}$ |
| $Z/\sigma$ | 2.877 (12.004) | 2.693 (0.904) | 0.304 (0.523) | 1.921 (9.295) | 0.520 (6.399) |
| $\mu_{95}^{\mu\tau}$ | 0.216 ($0.091^{+0.037}_{-0.026}$) | 0.490 ($0.222^{+0.089}_{-0.062}$) | 0.476 ($0.386^{+0.163}_{-0.108}$) | 0.231 ($0.126^{+0.052}_{-0.035}$) | 0.237 ($0.199^{+0.087}_{-0.056}$) |

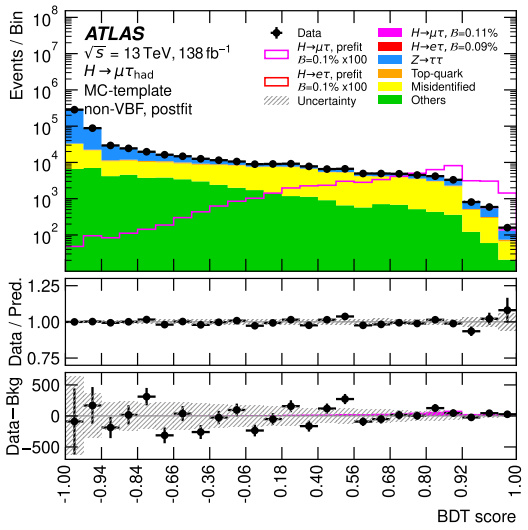
**Table 11.9:** The best-fit values $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$ of the signal strengths, the corresponding significances $Z$ and the upper limits $\mu_{95}^{e\tau}$ and $\mu_{95}^{\mu\tau}$ on the signal strengths at 95 % for the combination with the MC-based lephad analysis and the fits solely performed in the single signal regions. The expected significance for a signal strength of 1.0 and the expected upper limits, obtained with the *post-fit Asimov* dataset, are given in brackets. The upper half of the table summarises the results obtained with the $e\tau$-dataset in the search for $H \to e\tau$ and the lower half the results obtained with the $\mu\tau$-dataset in the search for $H \to \mu\tau$.



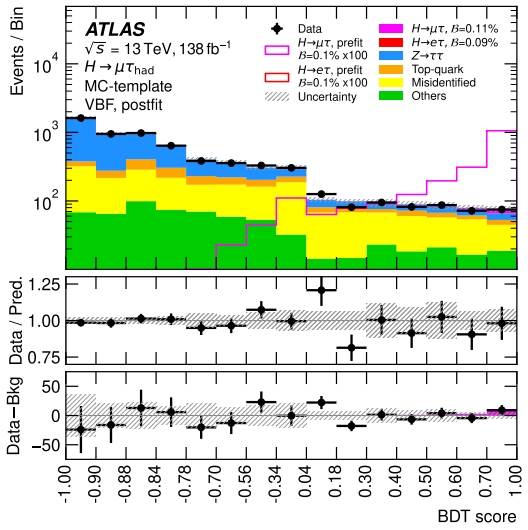

**(a)** Search for $H \to e\tau$, nonVBF SR




**(b)** Search for $H \to e\tau$, VBF SR



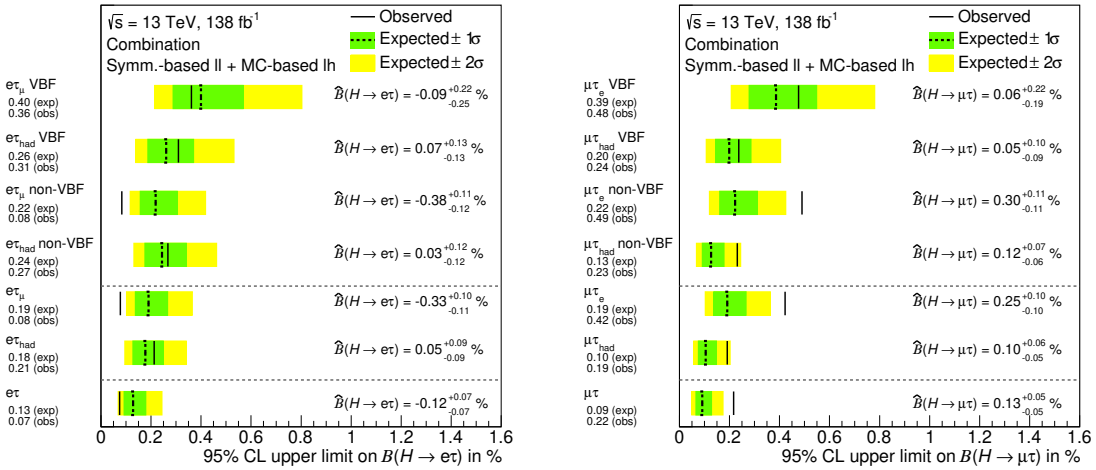

**(c)** Search for $H \to \mu\tau$, nonVBF SR




**(d)** Search for $H \to \mu\tau$, VBF SR

**Figure 11.9:** Post-fit distributions of the final discriminants (BDT score distributions) in the lephad final state in the search for $H \to e\tau$ (top) and for $H \to \mu\tau$ (bottom) in the nonVBF SR (left) and in the VBF SR, obtained from a combined fit in both regions which extracts both signal strengths simultaneously. The binning is visualised equidistantly to enhance visibility of the narrow bins. The hashed band indicates the post-fit statistical and systematic uncertainties on the prediction. In addition to the stacked signal corresponding to the best-fit prediction, the signal is overlaid (line) assuming a branching ratio of 0.1 % and scaled by 100. The central panel shows the ratio of data over the prediction (including signal) while the lower panel shows the difference of data and the background prediction [22].

background sample size although they play a reduced role in MC-based analyses if the number of generated events is considerably larger than the yield corresponding to the integrated luminosity. However, in the search for $H \to e\tau$ also $\gamma$-parameters (representing the statistical uncertainties) related to the MC-based lephad search are ranked high (Figure 11.11(a)). This is also the case in the search for $H \to \mu\tau$, but less pronounced. The breakdown of the uncertainties (Table 11.10) also shows that the impact on the uncertainty of $\mu$ of the uncertainties due to the lephad background sample size is relatively large, however still smaller than the impact of the corresponding leplep uncertainties. The uncertainties related to the fake-estimate of the lephad analysis have the highest impact on the uncertainty of $\mu$ among all systematic uncertainties, followed by the uncertainties related to the fake-estimate of the leplep analysis and the theoretical uncertainties on



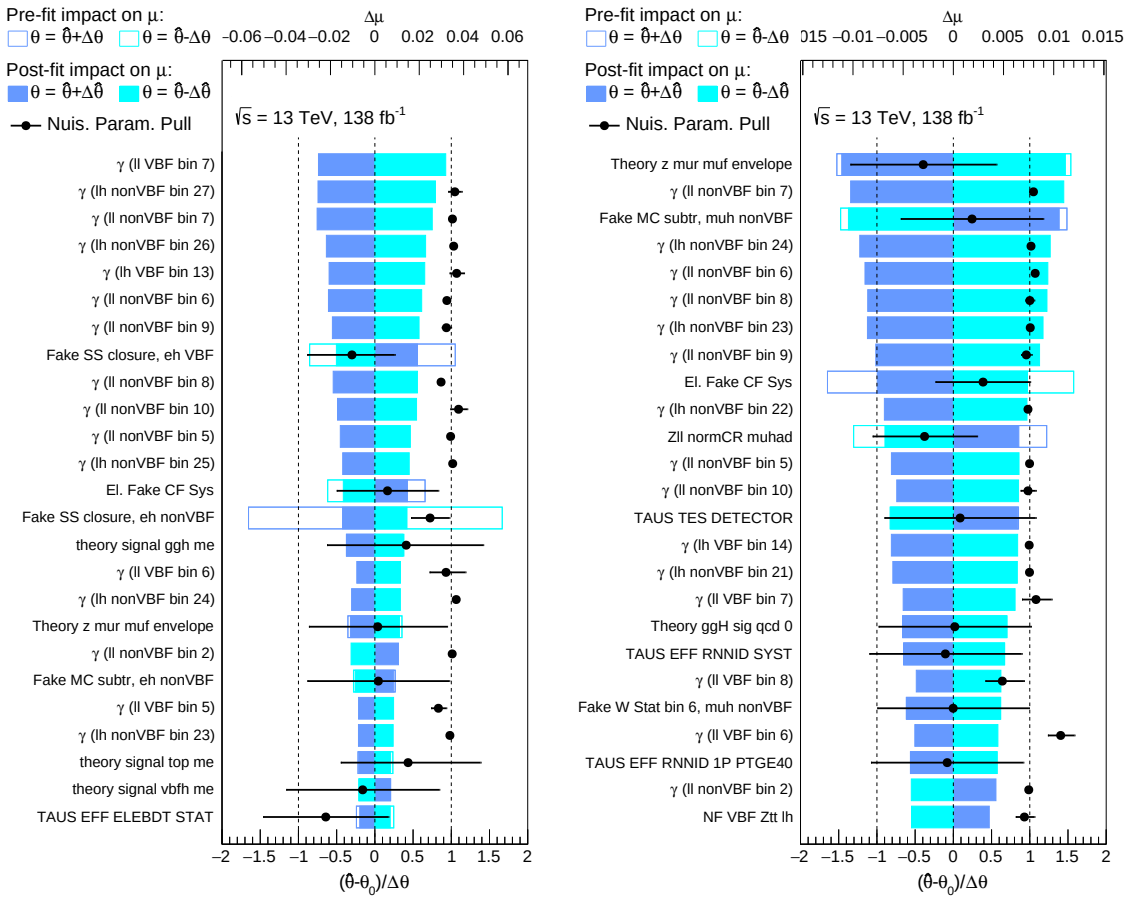

**(a)** Search for $H \to e\tau$ **(b)** Search for $H \to \mu\tau$

**Figure 11.10:** Upper limits at 95 % CL and best-fit values of the signal strengths in the combination with the lephad analysis in the search for $H \to e\tau$ (left) and for $H \to \mu\tau$ (right). The observed limits are indicated by a solid line, the expected by a dashed line. The results are obtained while assuming that the branching ratio of the respective other signal is zero and hence the signal strength $\mu$ can be interpreted as the branching ratio in %.

| Group | Uncertainty on observed $\mu$ by NP group | |
|---|---|---|
| | $H \to e\tau$ | $H \to \mu\tau$ |
| Total uncertainty | $+0.068\,/\,-0.069$ | $+0.049\,/\,-0.047$ |
| Data sample size | $+0.030\,/\,-0.030$ | $+0.027\,/\,-0.027$ |
| Syst. unc. + Bkg. sample size | $+0.061\,/\,-0.062$ | $+0.041\,/\,-0.039$ |
| Background sample size lephad | $+0.039\,/\,-0.040$ | $+0.017\,/\,-0.017$ |
| Background sample size leplep | $+0.045\,/\,-0.047$ | $+0.021\,/\,-0.020$ |
| Fakes leplep | $+0.020\,/\,-0.022$ | $+0.012\,/\,-0.011$ |
| Fakes lephad | $+0.026\,/\,-0.027$ | $+0.014\,/\,-0.014$ |
| Efficiency correction $\mathcal{R}$ | $+0.003\,/\,-0.003$ | $+0.001\,/\,-0.001$ |
| *other* $\to \ell$ normalisation leplep | $+0.001\,/\,-0.002$ | $+0.001\,/\,-0.001$ |
| $Z \to \ell\ell$ normalisation lephad | – | $+0.006\,/\,-0.007$ |
| Jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ | $+0.015\,/\,-0.014$ | $+0.010\,/\,-0.009$ |
| Hadronic $\tau$-leptons | $+0.011\,/\,-0.012$ | $+0.011\,/\,-0.009$ |
| Electrons and Muons | $+0.007\,/\,-0.008$ | $+0.005\,/\,-0.004$ |
| Flavour tagging | $+0.001\,/\,< 0.001$ | $+0.002\,/\,-0.002$ |
| Signal theoretical unc. | $+0.016\,/\,-0.018$ | $+0.011\,/\,-0.007$ |
| $Z \to \tau\tau$ theoretical unc. | $+0.011\,/\,-0.012$ | $+0.012\,/\,-0.012$ |
| Top-quark prod. theoretical unc. | $+0.009\,/\,-0.009$ | $+0.005\,/\,-0.005$ |
| Diboson prod. theoretical unc. | $+0.000\,/\,-0.000$ | $+0.002\,/\,-0.002$ |
| Luminosity | $+0.002\,/\,-0.001$ | $+0.005\,/\,-0.003$ |

**Table 11.10:** Breakdown of the uncertainties on the signal strength $\mu$ in the combination with the MC-based lephad analysis. Each group in the row just above a horizontal line is decomposed in sub-groups listed below this line. The square root of the quadratic sum of all single contributions does not add up to the total uncertainty due to correlations between the single groups. The abbreviation *syst.* stands for *systematic*, *unc.* stands for *uncertainty*, *bkg.* for *background* and *prod.* for *production*.

the signal prediction. Other important systematic uncertainties stem from jets and $E_{\mathrm{T}}^{\mathrm{miss}}$, hadronically decaying $\tau$-leptons and theoretical predictions of the $Z \to \tau\tau$-estimate which are mainly driven by or exclusive to the MC-based lephad analysis.




**(a)** Search for $H \to e\tau$ **(b)** Search for $H \to \mu\tau$

**Figure 11.11:** The pulls and constraints (black dots and bars and lower horizontal axis) of the 25 highest ranked nuisance parameters obtained from the combination with the MC-based lephad analysis in the $e\tau$-dataset (a) and the $\mu\tau$-dataset (b). The nominal value of the $\gamma$-parameters is at 1 and deviations from it indicate a pull. For all other NPs the nominal value is at 0 and the pre-fit $\pm 1\sigma$-variations correspond to $\pm 1$. The value of the highest ranked NP in $e\tau$, the $\gamma$-parameter of the last bin of the *VBF NN* distribution of the symmetry-based leplep analysis, is outside the range and is: $2.07^{+0.53}_{-0.46}$. In the names of the $\gamma$-parameters, "ll" denotes leplep and "lh" lephad. The post-fit impact on the signal strength $\mu$ (filled coloured bars and the upper horizontal axis) is the basis of the ranking (see Section 10.3). In addition, the pre-fit impacted is shown (empty coloured bars).

In order to quantify the compatibility of the best-fit values for the signal strengths among the four signal regions, a hypothesis test as described for the *Symmetry-Method* in Section 11.1.2 is performed. Again, the default fit setup with 1 POI serves as null hypothesis, claiming that the same signal strength in all regions can be assumed. While a fit setup with four POIs, one for each of the regions, serves as alternative hypothesis. The significance obtained from the test statistic (Eq. 11.3) using three (4-1) degrees of freedom for the $\chi^2$-distribution is $2.1\,\sigma$ in the search for $H \to e\tau$ and $0.9\,\sigma$ in the search for $H \to \mu\tau$ which means that the two fit setups agree within $2.1\,\sigma$ and $0.9\,\sigma$, respectively and hence the null hypothesis cannot be rejected.

## 11.3 Combination with both MC-based Analyses

In addition to a combination with the MC-based lephad analysis, a combination of all three analyses is performed: symmetry-based leplep, MC-based leplep and MC-based lephad. The two leplep analyses, however, are statistically correlated. Hence, one of them is used in the nonVBF SR and the other in the statistically independent VBF SR. The definition of the two signal regions is the same in both analyses, although the *Basic Selection* differs slightly due to different challenges of the different background estimation techniques. The choice of which leplep analysis to use for which region is made based on the expected sensitivity of each analysis in each region. The symmetry-based analysis showed a better expected sensitivity in the VBF SR while the MC-based analysis showed a better expected sensitivity in the nonVBF SR for both searches.

The estimation of all background processes apart from *fakes* in the MC-based leplep analysis is based on templates from MC simulations, similar to the MC-based lephad analysis. The largest background contributions from $Z \to \tau\tau$ and top-quark production are normalised in dedicated control regions which are enriched with events of the respective process. The *fakes* are estimated in a data-driven way assuming that the ratio of the number of events with an electron and a muon of opposite sign electric charges over the number of events with same sign electric charges is approximately the same when applying the default isolation and identification criteria and when inverting the isolation and/or the identification criteria of the sub-leading lepton. More information can be found in [22].

For the combination of all three analyses, two separate 1-POI fits in the search for $H \to e\tau$ and for $H \to \mu\tau$ are performed. For each of the two fits it is assumed that the branching ratio of the respective other signal is 0 which allows interpreting the signal strengths $\mu$ (the POIs) as branching ratios in percent of the LFV-signals. Each of the two fits is performed simultaneously in the four signal regions, MC-based leplep nonVBF SR, symmetry-based leplep VBF SR and MC-based lephad nonVBF SR and VBF SR, and in the nonVBF $Z \to \tau\tau$- and top-quark production control regions defined by the MC-based leplep analysis. The former control region is used to normalise the $Z \to \tau\tau$-contribution in the MC-based leplep analysis while this contribution in the MC-based lephad analysis is normalised from the most background-like bins of the final discriminants in the lephad signal regions. The latter control region is used to normalise the background contribution from top-quark production in the MC-based leplep analysis as well as in the nonVBF SR of the MC-based lephad analysis where top-quark production is only a minor contribution.

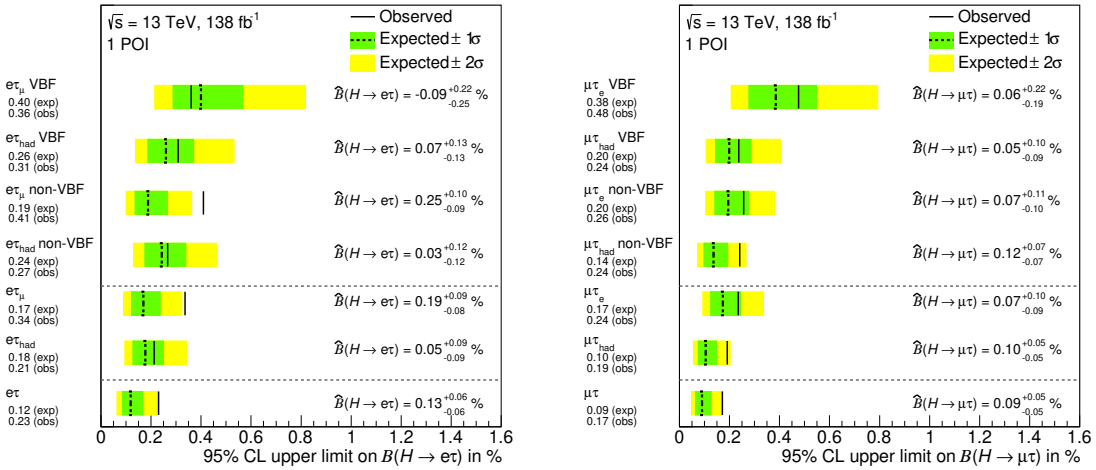
The results of the best-fit signal strengths, significances and upper limits are summarised in Table 11.11. In addition, the upper limits are visualised in Figure 11.12. The breakdown of the uncertainties on the signal strength $\mu$ in groups is summarised in Table 11.12 and the ranking of the 25 highest ranked NPs is shown in Figure 11.13.

Compared to the previous combination of only the symmetry-based leplep and the MC-based lephad analyses, the best-fit signal strength $\hat{\mu}$ in the nonVBF SR of the leplep analysis is now provided by the MC-based leplep analysis and is larger than 0 in both searches. In $H \to e\tau$, where it is $0.25^{+0.10}_{-0.09}$, a combined best-fit signal strength $\hat{\mu}^{e\tau}$ of $0.13^{+0.06}_{-0.06}$ is obtained with a significance of $2.1\,\sigma$. Hence, the observed upper limit of 0.23 is above the expected upper limit of $0.12^{+0.05}_{-0.03}$ just beyond the $2\,\sigma$-band. In $H \to \mu\tau$, where $\hat{\mu}^{\mu\tau}$ in the nonVBF SR of the leplep analysis is $0.07^{+0.11}_{-0.10}$, a combined best-fit signal strength of $0.09^{+0.05}_{-0.05}$ is obtained with a significance of $2.0\,\sigma$. Therefore, the observed upper limit of 0.17 is above the expected upper limit of $0.09^{+0.04}_{-0.03}$ slightly below the edge of the $2\,\sigma$-band.

The uncertainties due to the background sample size also play an important role in this combination, although their impact is reduced. In $H \to e\tau$ the impact is considerably larger than in $H \to \mu\tau$. The largest contribution among the systematic uncertainties is from the uncertainties related to the fake-estimate of the MC-based leplep analysis, followed

| | combined | MC-based Leplep | Symmetry-based Leplep | MC-based Lephad | |
|---|---|---|---|---|---|
| | | nonVBF SR | VBF SR | nonVBF SR | VBF SR |
| $\hat{\mu}^{e\tau}$ | $0.128^{+0.061}_{-0.060}$ | $0.247^{+0.098}_{-0.094}$ | $-0.090^{+0.222}_{-0.253}$ | $0.034^{+0.123}_{-0.124}$ | $0.066^{+0.129}_{-0.127}$ |
| $Z/\sigma$ | 2.135 (10.557) | 2.634 (8.007) | −0.391 (0.507) | 0.286 (6.295) | 0.517 (4.614) |
| limit $\mu^{e\tau}$ | 0.230 $(0.119^{+0.047}_{-0.033})$ | 0.410 $(0.188^{+0.077}_{-0.053})$ | 0.361 $(0.400^{+0.168}_{-0.112})$ | 0.266 $(0.243^{+0.097}_{-0.068})$ | 0.309 $(0.260^{+0.111}_{-0.073})$ |
| $\hat{\mu}^{\mu\tau}$ | $0.091^{+0.048}_{-0.046}$ | $0.067^{+0.106}_{-0.100}$ | $0.060^{+0.217}_{-0.195}$ | $0.119^{+0.072}_{-0.068}$ | $0.049^{+0.099}_{-0.092}$ |
| $Z/\sigma$ | 1.985 (12.048) | 0.659 (7.585) | 0.304 (0.523) | 1.766 (8.953) | 0.520 (6.399) |
| limit $\mu^{\mu\tau}$ | 0.171 $(0.090^{+0.037}_{-0.025})$ | 0.256 $(0.195^{+0.081}_{-0.055})$ | 0.476 $(0.386^{+0.163}_{-0.108})$ | 0.241 $(0.136^{+0.057}_{-0.038})$ | 0.237 $(0.199^{+0.087}_{-0.056})$ |

**Table 11.11:** The best-fit values $\hat{\mu}^{e\tau}$ and $\hat{\mu}^{\mu\tau}$ of the signal strengths, the corresponding significances $Z$ and the upper limits on the signal strengths for the combination of all three analyses and the fits solely performed in the single signal regions. The expected significance for a signal strength of 1.0 and the expected upper limits, obtained with the *post-fit Asimov* dataset, are given in brackets. The upper half of the table summarises the results obtained with the $e\tau$-dataset in the search for $H \to e\tau$ and the lower half the results obtained with the $\mu\tau$-dataset in the search for $H \to \mu\tau$.




**(a)** Search for $H \to e\tau$ **(b)** Search for $H \to \mu\tau$

**Figure 11.12:** Upper limits at 95 % CL and best-fit values of the signal strengths in the combination of all three analyses in the search for $H \to e\tau$ (left) and for $H \to \mu\tau$ (right). The observed limits are indicated by a solid line, the expected by a dashed line. The results are obtained while assuming that the branching ratio of the respective other signal is zero and hence the signal strength $\mu$ can be interpreted as the branching ratio in %.

by the uncertainties related to the fake-estimate of the the MC-based lephad analysis. Other important systematic uncertainty sources are, as in the previous combination, the theoretical uncertainties on the signal and on the $Z \to \tau\tau$-prediction, from jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ and from hadronically decaying $\tau$-leptons. All uncertainty sources related solely to the *Symmetry Method* which are uncertainties on the *fake*-estimates and uncertainties on the efficiency-correction ratio $\mathcal{R}$ are combined in one group called "Symmetry-based bkg. estimate". The NPs related to the electron identification scale factors which are also part of $\mathcal{R}$, however, are assigned to the group called "Electrons and Muons". Overall, this combination is as expected dominated by the MC-based analyses. This combination is also documented in [22].

In order to quantify the compatibility of the best-fit values for the signal strengths among the four signal regions equivalently to the combination with only the MC-based lephad analysis, a hypothesis test as described for the *Symmetry-Method* in Section 11.1.2 is performed. Again, the default fit setup with 1 POI serves as null hypothesis, claiming that the same signal strength in all regions can be assumed. While a fit setup with four POIs, one for each of the signal regions, serves as alternative hypothesis. The significance obtained from the test statistic (Eq. 11.3) and using three (4-1) degrees of freedom for the $\chi^2$-distribution is $1.0\,\sigma$ in the search for $H \to e\tau$ and $0.13\,\sigma$ in the search for $H \to \mu\tau$ which means that the two fit setups agree within $1.0\,\sigma$ and $0.13\,\sigma$, respectively and hence the null hypothesis cannot be rejected.

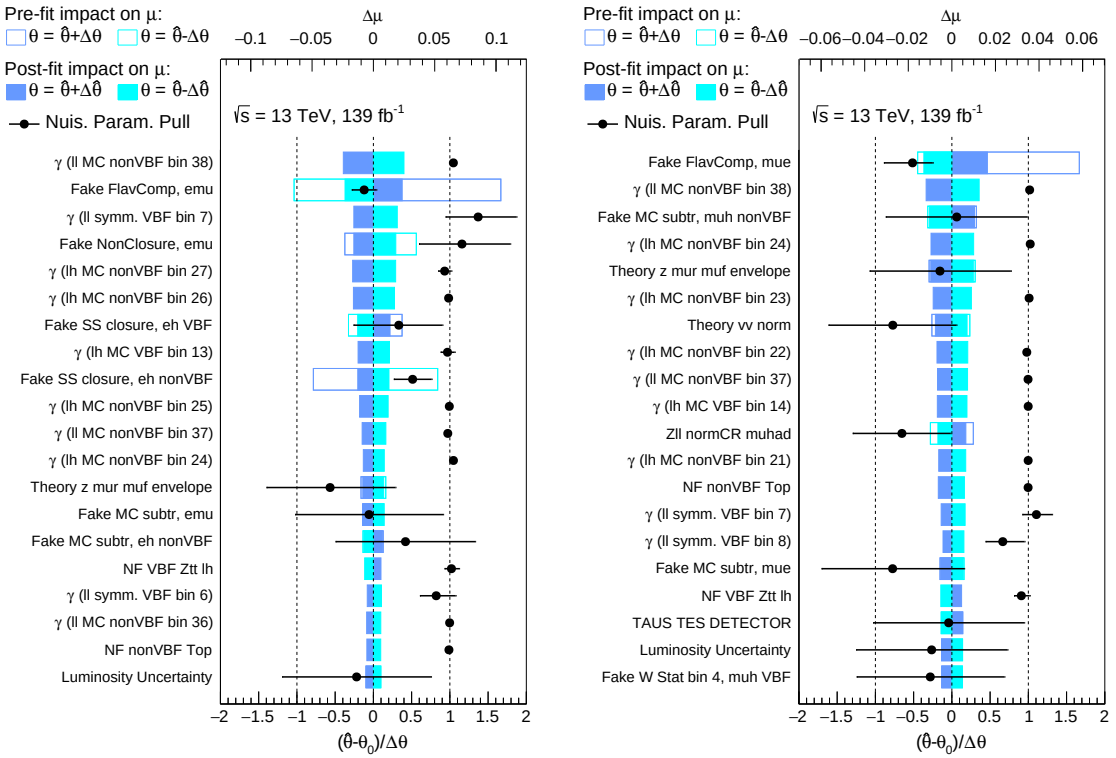
Compared to the combination of only the symmetry-based leplep and the MC-based lephad analyses, the upper limit on $\mu^{e\tau}$ in the search for $H \to e\tau$ is larger in the combination of all three analyses which can be explained by $\hat{\mu}^{e\tau}$ being negative in the former and positive in the latter combination. The expected limit, however, is lower in the combination of all three analyses. In the search for $H \to \mu\tau$, the observed upper limit in the combination of all three analyses is lower while the expected upper limit is almost the same as in the combination of only the symmetry-based leplep and the MC-based lephad analyses.

In addition to the 1-POI fits of the combination of all three analyses, another combination is documented in [22]. This combination solely considers the two MC-based methods

| Group | Uncertainty on observed $\mu$ by NP group | |
|---|---|---|
| | $H \to e\tau$ | $H \to \mu\tau$ |
| Total uncertainty | $+0.061 \,/\, -0.060$ | $+0.048 \,/\, -0.046$ |
| Data sample size | $+0.029 \,/\, -0.029$ | $+0.027 \,/\, -0.026$ |
| Syst. unc. + Bkg. sample size | $+0.054 \,/\, -0.052$ | $+0.040 \,/\, -0.038$ |
| Background sample size MC-based ll+lh | $+0.037 \,/\, -0.037$ | $+0.022 \,/\, -0.022$ |
| Background sample size symm.-based ll | $+0.019 \,/\, -0.020$ | $+0.008 \,/\, -0.009$ |
| Fakes leplep MC-based | $+0.029 \,/\, -0.028$ | $+0.017 \,/\, -0.015$ |
| Fakes lephad | $+0.020 \,/\, -0.021$ | $+0.015 \,/\, -0.014$ |
| Symmetry-based bkg. estimate | $+0.002 \,/\, -0.002$ | $+0.001 \,/\, -0.001$ |
| Signal theoretical unc. | $+0.011 \,/\, -0.006$ | $+0.009 \,/\, -0.005$ |
| $Z \to \tau\tau$ theoretical unc. | $+0.009 \,/\, -0.010$ | $+0.012 \,/\, -0.012$ |
| Top-quark prod. theoretical unc. | $+0.003 \,/\, -0.003$ | $+0.003 \,/\, -0.003$ |
| Diboson prod. theoretical unc. | $+0.004 \,/\, -0.004$ | $+0.007 \,/\, -0.007$ |
| $Z \to \ell\ell$ normalisation MC-based | $+0.002 \,/\, -0.001$ | $+0.006 \,/\, -0.007$ |
| Jets + $E_{\mathrm{T}}^{\mathrm{miss}}$ | $+0.011 \,/\, -0.010$ | $+0.011 \,/\, -0.010$ |
| Hadronic $\tau$-leptons | $+0.009 \,/\, -0.008$ | $+0.010 \,/\, -0.009$ |
| Electrons and Muons | $+0.003 \,/\, -0.001$ | $+0.005 \,/\, -0.005$ |
| Flavour tagging | $+0.006 \,/\, -0.006$ | $+0.004 \,/\, -0.003$ |
| Luminosity | $+0.007 \,/\, -0.005$ | $+0.006 \,/\, -0.004$ |

**Table 11.12:** Breakdown of the uncertainties on the signal strength $\mu$ in the combination with both MC-based analyses. Each group in the row just above a horizontal line is decomposed into sub-groups listed below the respective line. The square root of the quadratic sum of all single contributions does not add up to the total uncertainty due to correlations between the single groups. The abbreviations *syst.* stands for *systematic*, *unc.* stands for *uncertainty*, *bkg.* for *background*, *prod.* for *production* and *ll* and *lh* for *leplep* and *lephad*, respectively.

which allows to perform a 2-POI fit where both signal strengths, $\mu^{e\tau}$ and $\mu^{\mu\tau}$, are fitted simultaneously and hence no assumptions on the branching ratio of the respective other signal must be made. This combination observes best-fit signal strengths of $\hat{\mu}^{e\tau} = 0.095^{+0.059}_{-0.059}$ and $\hat{\mu}^{\mu\tau} = 0.108^{+0.046}_{-0.045}$ with significances of $1.6\,\sigma$ and $2.4\,\sigma$, respectively. The observed upper limit on $\mu^{e\tau}$ of 0.195 is lower than in the combination of all three analyses while the expected upper limit of $0.115^{+0.046}_{-0.032}$ is slightly lower as well. Whereas the observed upper limit on $\mu^{\mu\tau}$ of 0.184 is higher than in the combination of all three analyses while the expected upper limit of $0.088^{+0.035}_{-0.025}$ is slightly lower than in the combination of all three analyses.




**(a)** Search for $H \to e\tau$ **(b)** Search for $H \to \mu\tau$

**Figure 11.13:** The pulls and constraints (black dots and bars and lower horizontal axis) of the 25 highest ranked nuisance parameters obtained from the combination with both MC-based analyses in the $e\tau$-dataset (a) and the $\mu\tau$-dataset (b). The nominal value of the $\gamma$-parameters is at 1 and deviations from it indicate a pull. For all other NPs the nominal value is at 0 and the pre-fit $\pm 1\sigma$-variations correspond to $\pm 1$. In the names of the $\gamma$-parameters, "ll" denotes leplep, "lh" lephad, "symm." symmetry-based and "MC" MC-based. The post-fit impact on the signal strength $\mu$ (filled coloured bars and the upper horizontal axis) is the basis of the ranking (see Section 10.3). In addition, the pre-fit impacted is shown (empty coloured bars).

## 11.4 Compatibility of both Leplep Analyses

The compatibility of the results of the two leplep analyses is investigated in reference [22]. As discussed in Section 11.1.4, the final result of the symmetry-based leplep analysis is the best-fit signal strength $\hat{\mu}^{\mu\tau}$ interpreted as the difference in branching ratios $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ as well as the upper limit on this difference. The MC-based analysis, however, measures absolute branching ratios. For the comparison with the symmetry-based results, the difference of these absolute branching ratios, obtained from a simultaneous fit of $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$ in the MC-based leplep analysis is calculated. The fit also provides the correlation between the best-fit values of both branching ratios which is taken into account when calculating the difference.

Both analyses are highly correlated as they are based on the same data events to a large extent. In addition, the same simulated signal samples are used. Hence, all uncertainties on the signal are treated correlated between both analyses while all other uncertainty sources are treated uncorrelated due to the different background estimation techniques. In order to avoid double-counting of the correlated uncertainties, they are fixed to their post-fit nominal values in the MC-based analysis and the fit is repeated to obtain the uncertainty on the branching ratios only originating from the uncorrelated sources. The full uncertainties are kept for the results of the symmetry-based analysis. The correlation of the data due to the overlap of selected events is not explicitly taken into account. The values of the branching ratio difference $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ measured by the two analyses and their uncertainties are shown in Figure 11.14 for individual fits in each of

the two signal regions and for a combined fit in both signal regions. The full as well as only the uncorrelated uncertainties are indicated for the MC-based analysis. The measured branching ratio differences of the two analyses agree within $2\,\sigma$ taking into account the results of both signal regions individually.

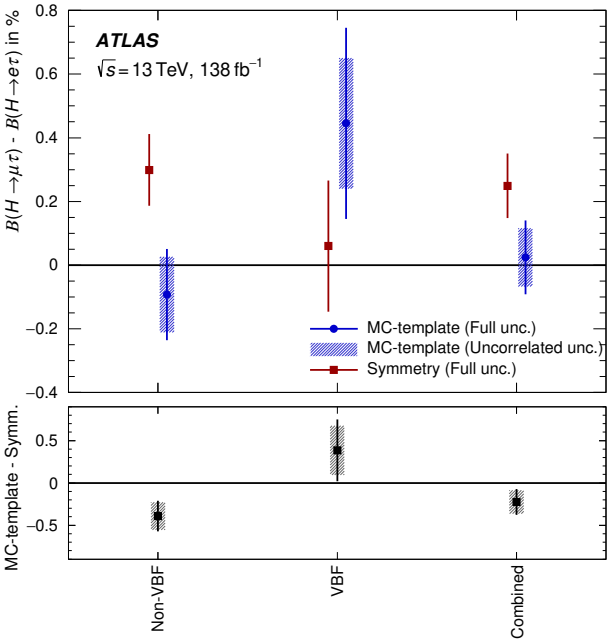



**Figure 11.14:** Best-fit values of the branching ratio difference $\mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ in percent measured by the symmetry-based and by the MC-based leplep analyses, from [22]. The lower panel shows the difference of the branching ratio differences measured by the two analyses. The uncertainty bars indicate the uncorrelated uncertainties only, while the lines indicate the full uncertainties.

# Chapter 12

# Conclusion and Outlook

In this thesis a direct search for lepton-flavour violating decays of the Higgs boson using data from proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ delivered by the LHC and recorded by the ATLAS detector has been performed. The dataset corresponds to an integrated luminosity of $138.4\,\text{fb}^{-1}$ collected during Run 2 of the LHC which lasted from 2015 to 2018. The search in this thesis has been focused on the decays $H \to e\tau$ and $H \to \mu\tau$ with leptonic decays of the $\tau$-lepton leading to an electron, a muon and two neutrinos in the final state. The signal over background ratio is small and dedicated analysis techniques were required to improve the sensitivity. On one hand two signal regions (VBF SR and nonVBF SR) targeting different production processes of the Higgs boson have been defined, thereby allowing for exploiting the characteristic topology of the VBF production. On the other hand, machine learning techniques have been deployed. Individual neural networks have been trained in each of the two signal regions aiming at classifying each collision event to be background or signal. The resulting probability distributions have been used as final discriminants in the statistical analysis. The development and training of the NNs required dedicated optimisations of the NN strategy, hyperparameters and input variables. The optimisation of the networks and other aspects of the analysis has been performed by a direct interplay with the statistical analysis in order to obtain a measure for the sensitivity changes. The statistical model has undergone substantial improvements compared to the previous publication that included the *Symmetry Method*.

The *Symmetry Method* [21] has been used for background estimation. It exploits the fact that SM processes are symmetric with respect to an exchange of electrons with muons and vice versa at the high energies prevalent at the LHC.

The difference in branching ratios $\Delta^{\mu\tau} = \mathcal{B}(H \to \mu\tau) - \mathcal{B}(H \to e\tau)$ has been measured using the $\mu\tau$-dataset as

$$\hat{\Delta}^{\mu\tau} = (0.25 \pm 0.10)\,\%\,, \tag{12.1}$$

which indicates that the data favours a larger branching ratio for the decay $H \to \mu\tau$ than for the decay $H \to e\tau$. However, the significance of this measurement of $2.5\,\sigma$ is not highly significant and hence does not allow for rejecting the background-only hypothesis. It is not possible to judge whether the observed fluctuation originates from an up-fluctuation in the $\mu\tau$-dataset, from a down-fluctuation in the $e\tau$-dataset or from both. An upper limit on $\Delta^{\mu\tau}$ at 95 % confidence level (CL) is set:

$$\Delta^{\mu\tau} < 0.42\,\%\,, \tag{12.2}$$

where the expected limit is $\Delta^{\mu\tau} < 0.19^{+0.08}_{-0.05}\,\%$. Thus, the observed limit is outside the upper edge of the $2\,\sigma$-band. When assuming that the $H \to e\tau$-signal is non-existent, the upper limit on the difference can be interpreted as an upper limit on the branching ratio of

the $H \to \mu\tau$-signal:

$$\mathcal{B}(H \to \mu\tau) < 0.42\,\%\,(0.19^{+0.08}_{-0.05})\,, \tag{12.3}$$

with the expected limit given in brackets. When, in contrast, a non-existent $H \to \mu\tau$-signal is assumed, an upper limit at 95 % CL on the branching ratio of $H \to e\tau$ can be derived from analysing the $e\tau$-dataset:

$$\mathcal{B}(H \to e\tau) < 0.08\,\%\,(0.19^{+0.08}_{-0.05})\,. \tag{12.4}$$

The observed limit is just outside the lower edge of the $2\,\sigma$-band of the expected limit. This is a consequence of the *Symmetry Method* which is sensitive to the difference in branching ratios but not to the absolute values and hence the up-fluctuation observed in the $\mu\tau$-dataset propagates to the $e\tau$-dataset as a down-fluctuation.

This analysis is clearly dominated by statistical uncertainties since it relies heavily on data for estimating the background contributions and therefore is limited by the size of the dataset.

Furthermore, combinations with two partner analyses which also search for LFV decays of the Higgs boson have been performed. One partner analysis has investigated events where the $\tau$-lepton decays hadronically (called MC-based lephad), the other events where the $\tau$-lepton decays leptonically (called MC-based leplep) [22]. Both have used a different technique for background estimation compared to the analysis discussed in this thesis (denoted as symmetry-based leplep). Two combinations have been performed: 1. symmetry-based leplep and MC-based lephad and 2. VBF SR of symmetry-based leplep, nonVBF SR of MC-based leplep and MC-based lephad. The resulting upper limits are summarised in Table 12.1. In order to perform these combinations, it must be assumed that the branching ratio of one of the potential LFV signals is zero. In addition, the MC-based analyses have performed a combination where the branching ratios of both signals are extracted simultaneously and hence no assumption on the branching ratio of the other signal is needed. These results (indicated by Comb. 3) as well as the latest direct limits by the CMS collaboration, obtained with the full Run 2 dataset, are also listed in Table 12.1. The

| | Comb. 1 | Comb. 2 | Comb. 3 [22] | CMS [20] |
|---|---|---|---|---|
| $\mathcal{B}_{95}(H \to e\tau)/\%$ | 0.07 $(0.13^{+0.05}_{-0.04})$ | 0.23 $(0.12^{+0.05}_{-0.03})$ | 0.20 (0.12) | 0.22 (0.16) |
| $\mathcal{B}_{95}(H \to \mu\tau)/\%$ | 0.22 $(0.09^{+0.04}_{-0.03})$ | 0.17 $(0.09^{+0.04}_{-0.03})$ | 0.18 (0.09) | 0.15 (0.15) |

**Table 12.1:** The observed (expected) upper limits on the branching ratios of the decays $H \to e\tau$ and $H \to \mu\tau$ at 95 % CL of the combination (Comb.) of the symmetry-based leplep and the MC-based lephad analysis (1) of the symmetry-based leplep, the MC-based leplep and the MC-based lephad analysis (2), both MC-based analyses (3) and the latest direct limits by the CMS collaboration. All limits are derived with the full Run 2 dataset. In the derivation of all limits apart from combination 3 it is assumed that the respective other LFV signal is zero.

expected upper limits of all three combinations by the ATLAS collaboration are comparable to each other but lower by $\sim 20\,\%\,(40\,\%)$ for $H \to e\tau\,(H \to \mu\tau)$, i.e. more sensitive, compared to the CMS collaboration. All observed limits by the ATLAS collaboration, except the one of Combination 1 on $\mathcal{B}(H \to e\tau)$, are larger than the expected ones. This indicates the observation of up-fluctuations in data of which, however, none is statistically highly significant. The small observed limit of Combination 1 on $\mathcal{B}(H \to e\tau)$ is driven by the up- and/or down-fluctuation in the $\mu\tau$- and/or $e\tau$-dataset observed by the symmetry-based analysis.

# Outlook

Using the *Symmetry Method* for the search for LFV decays of the Higgs boson is an innovative approach with a reduced dependence on simulated events. Consequently, the available amount of data plays a considerable role regarding the sensitivity of the analysis.

However, not only the size of the dataset contributes to the statistical uncertainties but also the limited amount of simulated events currently used to estimate background contributions where other objects than jets are mis-identified as an electron or a muon. In addition, all systematic uncertainties related to these simulated events as well as uncertainties on their normalisation must be considered. Hence, it will be beneficial if a method is found with the ability to estimate these contributions in a data-driven way as well. At the same time, seeking methods to reduce the systematic uncertainties on the data-driven estimation of the background contribution where jets are mis-identified as leptons would be useful.

Furthermore, it could be beneficial to further improve the statistical model such that it is less reliant on the subtraction of various contributions which currently enhances the statistical uncertainties in the sensitive bins of the final discriminant. In addition, related to the previous point, it would be advantageous to be able to measure the difference in branching ratios directly by including both signal contributions at once which implies incorporating both datasets, the $e\tau$- and $\mu\tau$-dataset, simultaneously as data and as background contribution.

Finally, the analysis will evidently benefit from an enhanced dataset. In order to estimate the sensitivity potentially reachable in the future with the ATLAS detector, an extrapolation of the analysis presented in this thesis to the high-luminosity LHC (HL-LHC) scenario is performed assuming an integrated luminosity of $3\,000\,\text{fb}^{-1}$ at a center-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$ [23]. In the process, more precise theory predictions as well as expected reductions of systematic uncertainties due to detector upgrades are considered which, however, play a minor role in this analysis. Expected limits as well as expected significances assuming branching ratios of 0.1 % for the LFV signals are derived in the searches for $H \to e\tau$ and $H \to \mu\tau$ based on the *Symmetry Method*, thereby assuming that the respective other signal is zero. The obtained expected limits are visualised in Figure 12.1. The results are compared to the respective expected results obtained with the Run 2 dataset. Overall, improvements in the sensitivity by a factor 3.7 can be expected. This results in an expected upper limit at 95 % CL on each of the branching ratios of $0.05^{+0.02}_{-0.01}$ % for the HL-LHC scenario. Such a degree of sensitivity should be sufficient to judge whether the observed hint for a difference in branching ratios of the two LFV decays of that size and thus for the existence of the $H \to \mu\tau$-signal hardens or is attributed to statistical fluctuations.

The sensitivity in the VBF SR profits more from the larger dataset compared to the nonVBF SR as it is stronger limited by statistical uncertainties due to the lower yield in this region. The analysis is still dominated by statistical uncertainties when based on a dataset of $3\,000\,\text{fb}^{-1}$ as systematic uncertainties are also expected to be reduced. Nevertheless, the impact of the systematic uncertainties with respect to the overall uncertainty will be increased.

The search for LFV decays of the Higgs boson will evidently benefit from a larger dataset and remain an exciting topic in the search for physics beyond the standard model.

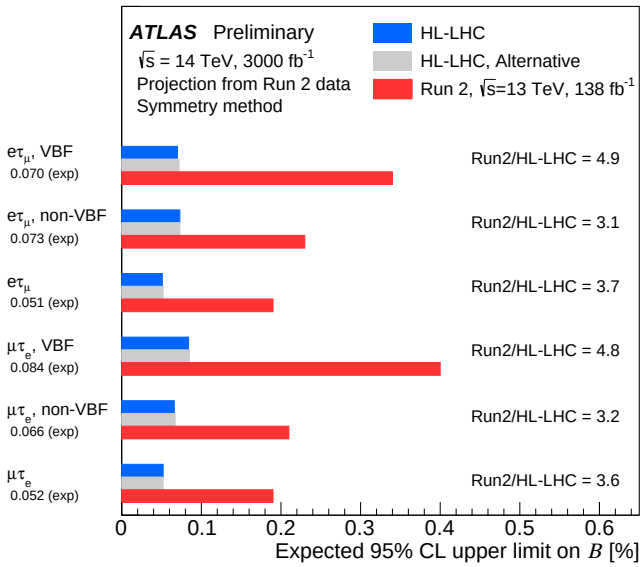



**Figure 12.1:** Expected upper limits on the branching ratios $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$ at 95 % CL for the Run 2 (red) and the HL-LHC (blue, gray) scenario [23]. The respective other LFV signal is assumed to be zero when deriving the limits. For the HL-LHC scenario two different assumptions on the scaling of the statistical uncertainties of simulated samples are made: either they are scaled with luminosity (gray) or are set to 0 (blue).

# Bibliography


[1] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (9 1964) 321 (cit. on pp. 1, 12).

[2] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (16 1964) 508 (cit. on pp. 1, 12).

[3] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (20 1964) 585 (cit. on pp. 1, 12).

[4] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** (2012) 1, ISSN: 0370-2693 (cit. on pp. 1, 18).

[5] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B **716** (2012) 30, ISSN: 0370-2693 (cit. on pp. 1, 18).

[6] R. Davis Jr., D. S. Harmer, and K. C. Hoffman, *Search for neutrinos from the sun*, Phys. Rev. Lett. **20** (1968) 1205 (cit. on pp. 1, 22).

[7] Y. Fukuda et al., *Evidence for Oscillation of Atmospheric Neutrinos*, Phys. Rev. Lett. **81** (8 1998) 1562 (cit. on pp. 1, 22).

[8] Q. R. Ahmad et al., *Measurement of the Rate of $\nu_e + d \to p + p + e^-$ Interactions Produced by $^8B$ Solar Neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87** (7 2001) 071301 (cit. on pp. 1, 22).

[9] J. D. Bjorken and S. Weinberg, *Mechanism for Nonconservation of Muon Number*, Phys. Rev. Lett. **38** (12 1977) 622 (cit. on pp. 1, 23).

[10] J. L. Diaz-Cruz and J. J. Toscano, *Lepton flavor violating decays of Higgs bosons beyond the standard model*, Phys. Rev. D **62** (2000) 116005 (cit. on pp. 1, 23, 24).

[11] G. C. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rept. **516** (2012) 1 (cit. on pp. 1, 23, 25).

[12] K. Agashe and R. Contino, *Composite Higgs-Mediated FCNC*, Phys. Rev. D **80** (2009) 075016 (cit. on pp. 1, 23).

[13] A. Azatov, M. Toharia, and L. Zhu, *Higgs Mediated FCNC's in Warped Extra Dimensions*, Phys. Rev. D **80** (2009) 035016 (cit. on pp. 1, 23).

[14] A. Brignole and A. Rossi, *Anatomy and phenomenology of mu-tau lepton flavor violation in the MSSM*, Nucl. Phys. B **701** (2004) 3 (cit. on pp. 1, 23).

[15] G. Perez and L. Randall, *Natural Neutrino Masses and Mixings from Warped Geometry*, JHEP **01** (2009) 077 (cit. on pp. 1, 23).

[16] C. D. Froggatt and H. B. Nielsen, *Hierarchy of quark masses, Cabibbo angles and CP violation*, Nucl. Phys. B **147** (1979) 277 (cit. on pp. 1, 23).

[17] G. Blankenburg, J. Ellis, and G. Isidori, *Flavour-Changing Decays of a 125 GeV Higgs-like Particle*, Phys. Lett. B **712** (2012) 386 (cit. on pp. 1, 2, 27, 28).

[18] R. Harnik, J. Kopp, and J. Zupan, *Flavor Violating Higgs Decays*, JHEP **03** (2013) 026 (cit. on pp. 1, 2, 24, 27, 28, 331).

[19] ATLAS Collaboration, *Searches for lepton-flavour-violating decays of the Higgs boson in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, Physics Letters B **800** (2020) 135069 (cit. on pp. 2, 3, 29, 46, 100).

[20] CMS Collaboration, *Search for lepton-flavor violating decays of the Higgs boson in the $\mu\tau$ and $e\tau$ final states in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. D **104** (2021) 032013 (cit. on pp. 2, 29, 258).

[21] S. Bressler, A. Dery, and A. Efrati, *Asymmetric lepton-flavor violating Higgs boson decays*, Physical Review D **90** (2014) (cit. on pp. 2, 43, 257).

[22] ATLAS Collaboration, *Searches for lepton-flavour-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$ in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, (2023), arXiv: 2302.05225 [hep-ex] (cit. on pp. 2, 3, 152, 204, 217, 234, 245, 247, 250, 252, 254, 255, 258).

[23] ATLAS Collaboration, *Searches of lepton-flavour-violating decays of the Higgs bosons with the ATLAS detector at the HL-LHC*, CERN, 2022, URL: https://cds.cern.ch/record/2841245 (cit. on pp. 3, 259, 260).

[24] ATLAS Collaboration, *The ATLAS Tau Trigger in Run 2*, CERN, 2017, URL: https://cds.cern.ch/record/2274201 (cit. on p. 3).

[25] R. P. Feynman, *Space-Time Approach to Quantum Electrodynamics*, Phys. Rev. **76** (6 1949) 769 (cit. on pp. 5, 49).

[26] D. J. Griffiths, *Introduction to elementary particles; 2nd rev. version*, Physics textbook, Wiley, 2008, ISBN: 978-3-527-40601-2 (cit. on p. 5).

[27] F. Halzen and A. D. Martin, *Quarks & Leptons: An introductory course in modern particle physics*, John Wiley & Sons, 1984, ISBN: 0-471-88741-2 (cit. on pp. 5, 9).

[28] P. Schmüser, *Feynman-Graphen und Eichtheorien für Experimentalphysiker*, Springer-Verlag, 1988, ISBN: 3-540-18797-9 (cit. on p. 5).

[29] S. Dittmaier and M. Schumacher, *The Higgs Boson in the Standard Model - From LEP to LHC: Expectations, Searches, and Discovery of a Candidate*, Prog. Part. Nucl. Phys. **70** (2013) 1 (cit. on pp. 5, 14).

[30] R. L. Workman et al., *Review of Particle Physics*, PTEP **2022** (2022) 083C01 (cit. on pp. 6–8, 13, 18, 19, 21).

[31] E. Noether, *Invariante Variationsprobleme*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse (1918) 235, URL: http://eudml.org/doc/59024 (cit. on p. 8).

[32] M. Gell-Mann, *Symmetries of Baryons and Mesons*, Phys. Rev. **125** (3 1962) 1067 (cit. on p. 8).

[33] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Phys. Rev. Lett. **30** (26 1973) 1343 (cit. on p. 9).

[34] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (26 1973) 1346 (cit. on p. 9).

[35] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** (1961) 579, ISSN: 0029-5582 (cit. on p. 9).

[36] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (21 1967) 1264 (cit. on p. 9).

[37] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519** (1968) 367 (cit. on p. 9).

[38] T. Nakano and K. Nishijima, *Charge Independence for V-particles*, Progress of Theoretical Physics **10** (1953) 581, ISSN: 0033-068X, eprint: https://academic.oup.com/ptp/article-pdf/10/5/581/5364926/10-5-581.pdf (cit. on p. 10).

[39] M. Gell-Mann, *The interpretation of the new particles as displaced charge multiplets*, Il Nuovo Cimento **4** (1956) 848 (cit. on p. 10).

[40] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (12 1963) 531 (cit. on p. 10).

[41] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49** (1973) 652, ISSN: 0033-068X (cit. on p. 10).

[42] UA1 Collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at* $\sqrt{s} = 540$ *GeV*, Phys. Lett. B **122** (1983) 103 (cit. on p. 11).

[43] UA2 Collaboration, *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN pp Collider*, Phys. Lett. B **122** (1983) 476 (cit. on p. 11).

[44] UA1 Collaboration, *Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider*, Phys. Lett. B **126** (1983) 398 (cit. on p. 11).

[45] Y. Nambu, *Quasi-Particles and Gauge Invariance in the Theory of Superconductivity*, Phys. Rev. **117** (3 1960) 648 (cit. on p. 13).

[46] J. Goldstone, *Field Theories with Superconductor Solutions*, Nuovo Cim. **19** (1961) 154 (cit. on p. 13).

[47] S. Weinberg, *Physical Processes in a Convergent Theory of the Weak and Electromagnetic Interactions*, Phys. Rev. Lett. **27** (24 1971) 1688 (cit. on p. 13).

[48] S. Weinberg, *General Theory of Broken Local Symmetries*, Phys. Rev. D **7** (4 1973) 1068 (cit. on p. 13).

[49] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the Unified Model of Elementary Particles*, Progress of Theoretical Physics **28** (1962) 870, ISSN: 0033-068X (cit. on p. 15).

[50] ATLAS Collaboration, *Standard Model Summary Plots February 2022*, CERN, 2022, URL: http://cds.cern.ch/record/2804061 (cit. on p. 16).

[51] LHC Higgs Cross Section Working Group, *CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*, en, 2017 (cit. on pp. 16–18, 206).

[52] *High precision determination of the gluon fusion Higgs boson cross-section at the LHC*, JHEP **05** (2016) 058 (cit. on p. 16).

[53] U. Aglietti, R. Bonciani, G. Degrassi, and A. Vicini, *Two-loop light fermion contribution to Higgs production and decays*, Physics Letters B **595** (2004) 432, ISSN: 0370-2693 (cit. on p. 16).

[54] S. Actis, G. Passarino, C. Sturm, and S. Uccirati, *NLO electroweak corrections to Higgs boson production at hadron colliders*, Physics Letters B **670** (2008) 12, ISSN: 0370-2693 (cit. on p. 16).

[55] S. Actis, G. Passarino, C. Sturm, and S. Uccirati, *NNLO computational techniques: The cases* $H \to \gamma\gamma$ *and* $H \to gg$, Nuclear Physics B **811** (2009) 182, ISSN: 0550-3213 (cit. on p. 16).

[56] C. Anastasiou, R. Boughezal, and F. Petriello, *Mixed QCD-electroweak corrections to Higgs boson production in gluon fusion*, Journal of High Energy Physics **2009** (2009) 003 (cit. on p. 16).

[57] P. Bolzoni, F. Maltoni, S.-O. Moch, and M. Zaro, *Higgs Boson Production via Vector-Boson Fusion at Next-to-Next-to-Leading Order in QCD*, Physical Review Letters **105** (2010) (cit. on pp. 17, 53).

[58] M. Ciccolini, A. Denner, and S. Dittmaier, *Strong and Electroweak Corrections to the Production of a Higgs boson + 2 Jets Weak Interactions at the Large Hadron Collider*, Phys. Rev. Lett. **99** (16 2007) 161803 (cit. on p. 17).

[59] M. Ciccolini, D. A., and D. S., *Electroweak and QCD corrections to Higgs production via vector-boson fusion at the CERN LHC*, Physical Review D **77** (2008) (cit. on pp. 17, 53).

[60] O. Brein, A. Djouadi, and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, Physics Letters B **579** (2004) 149 (cit. on pp. 17, 53).

[61] A. Denner, S. Dittmaier, S. Kallweit, and A. Mück, *Electroweak corrections to Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with Hawk*, Journal of High Energy Physics **2012** (2012) (cit. on pp. 17, 53).

[62] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak, and T. J. E. Zirke, *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, Journal of High Energy Physics **2013** (2013) (cit. on pp. 17, 53).

[63] S. Dawson, L. H. Orr, L. Reina, and D. Wackeroth, *Associated top quark Higgs boson production at the LHC*, Phys. Rev. D **67** (2003) 071503 (cit. on p. 17).

[64] S. Dawson, C. Jackson, L. H. Orr, L. Reina, and D. Wackeroth, *Associated Higgs boson production with top quarks at the CERN Large Hadron Collider: NLO QCD corrections*, Phys. Rev. D **68** (3 2003) 034022 (cit. on p. 17).

[65] W. Beenakker et al., *Higgs Radiation Off Top Quarks at the Tevatron and the LHC*, Phys. Rev. Lett. **87** (20 2001) 201805 (cit. on p. 17).

[66] W. Beenakker et al., *NLO QCD corrections to ttH production in hadron collisions*, Nuclear Physics B **653** (2003) 151, ISSN: 0550-3213 (cit. on p. 17).

[67] Y. Zhang, W.-G. Ma, R.-Y. Zhang, C. Chen, and L. Guo, *QCD NLO and EW NLO corrections to tH production with top quark decays at hadron collider*, Physics Letters B **738** (2014) 1, ISSN: 0370-2693 (cit. on p. 17).

[68] S. Frixione, V. Hirschi, D. Pagani, H.-S. Shao, and M. Zaro, *Weak corrections to Higgs hadroproduction in association with a top-quark pair*, Journal of High Energy Physics **2014** (2014) (cit. on p. 17).

[69] S. Frixione, V. Hirschi, D. Pagani, H. .-. Shao, and M. Zaro, *Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons*, JHEP **06** (2015) 184 (cit. on p. 17).

[70] L. D. Landau, *On the angular momentum of a system of two photons*, Dokl. Akad. Nauk SSSR **60** (1948) 207 (cit. on p. 19).

[71] C. N. Yang, *Selection Rules for the Dematerialization of a Particle into Two Photons*, Phys. Rev. **77** (2 1950) 242 (cit. on p. 19).

[72] ATLAS Collaboration, *Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*, The European Physical Journal C **75** (2015) 1 (cit. on p. 19).

[73] CMS Collaboration, *A measurement of the Higgs boson mass in the diphoton decay channel*, Physics Letters B **805** (2020) 135425 (cit. on p. 19).

[74] ATLAS Collaboration, *Measurement of the Higgs boson mass in the $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$ channels with $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detecto*, Physics Letters B **784** (2018) 345, ISSN: 0370-2693 (cit. on p. 19).

[75] L. J. Dixon and Y. Li, *Bounding the Higgs Boson Width through Interferometry*, Phys. Rev. Lett. **111** (11 2013) 111802 (cit. on p. 19).

[76] F. Caola and K. Melnikov, *Constraining the Higgs boson width with ZZ production at the LHC*, Phys. Rev. D **88** (5 2013) 054024 (cit. on p. 19).

[77] J. M. Campbell, R. K. Ellis, and C. Williams, *Bounding the Higgs width at the LHC using full analytic results for $gg \to e^-e^+\mu^-\mu^+$*, Journal of High Energy Physics **2014** (2014) (cit. on p. 19).

[78] ATLAS Collaboration, *Evidence of off-shell Higgs boson production and constraints on the total width of the Higgs boson in the $ZZ \to 4\ell$ and $ZZ \to 2\ell 2\nu$ decay channels with the ATLAS detector*, CERN, 2022, URL: `https://cds.cern.ch/record/2842520` (cit. on p. 19).

[79] CMS Collaboration, *Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production*, Nature Phys. **18** (2022) 1329 (cit. on p. 19).

[80] ATLAS Collaboration, *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, Nature **607** (2022) 52, [Erratum: Nature 612, E24 (2022)] (cit. on pp. 20, 21).

[81] CMS Collaboration, *A portrait of the Higgs boson by the CMS experiment ten years after the discovery*, Nature **607** (2022) 60 (cit. on p. 20).

[82] CMS Collaboration, *Evidence for Higgs boson decay to a pair of muons*, Journal of High Energy Physics **2021** (2021) 1 (cit. on p. 20).

[83] C. T. Potter et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*, en, 2013 (cit. on pp. 20, 44).

[84] L. Calibbi and G. Signorelli, *Charged Lepton Flavour Violation: An Experimental and Theoretical Introduction*, Riv. Nuovo Cim. **41** (2018) 71 (cit. on p. 22).

[85] A. de Gouvêa and P. Vogel, *Lepton flavor and number conservation, and physics beyond the standard model*, Progress in Particle and Nuclear Physics **71** (2013) 75 (cit. on p. 22).

[86] S. M. Bilenky, S. T. Petcov, and B. Pontecorvo, *Lepton Mixing, $\mu \to e\gamma$ Decay and Neutrino Oscillations*, Phys. Lett. B **67** (1977) 309 (cit. on p. 22).

[87] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. D **2** (7 1970) 1285 (cit. on p. 22).

[88] T. Appelquist and J. Carazzone, *Infrared singularities and massive fields*, Phys. Rev. D **11** (10 1975) 2856 (cit. on p. 23).

[89] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, *Dimension-Six Terms in the Standard Model Lagrangian*, JHEP **10** (2010) 085 (cit. on p. 24).

[90] A. Goudelis, O. Lebedev, and J.-h. Park, *Higgs-induced lepton flavor violation*, Phys. Lett. B **707** (2012) 369 (cit. on p. 24).

[91] A. Vicente, *Higgs Lepton Flavor Violating Decays in Two Higgs Doublet Models*, Frontiers in Physics **7** (2019), ISSN: 2296-424X (cit. on pp. 25, 26).

[92] J. Bernon, J. F. Gunion, H. E. Haber, Y. Jiang, and S. Kraml, *Scrutinizing the alignment limit in two-Higgs-doublet models: $m_h$=125 GeV*, Phys. Rev. D **92** (2015) 075004 (cit. on p. 26).

[93] J. F. Gunion and H. E. Haber, *The CP conserving two Higgs doublet model: The Approach to the decoupling limit*, Phys. Rev. D **67** (2003) 075019 (cit. on p. 26).

[94] A. Baldini et al., *Search for the lepton flavour violating decay $mu^+ \to e^+\gamma$ with the full dataset of the MEG experiment*, The European Physical Journal C **76** (2016) 434 (cit. on p. 27).

[95] B. Aubert et al., *Searches for lepton flavor violation in the decays $\tau^\pm \to e^\pm\gamma$ and $\tau^\pm \to \mu^\pm\gamma$*, Physical review letters **104** (2010) 021802 (cit. on p. 27).

[96] J. Adam et al., *New limit on the lepton-flavour violating decay $\mu^+ \to e^+\gamma$*, Phys. Rev. Lett. **107** (2011) 171801 (cit. on p. 28).

[97] ATLAS Collaboration, *Search for the Higgs boson decays $H \to ee$ and $H \to e\mu$ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **801** (2020) 135148 (cit. on p. 29).

[98] CMS Collaboration, *Search for lepton flavour violating decays of the Higgs boson to $e\tau$ and $e\mu$ in proton–proton collisions at $\sqrt{s} = 8$ TeV*, Phys. Lett. B **763** (2016) 472 (cit. on p. 29).

[99] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) S08001 (cit. on pp. 31, 32).

[100] Education, Communications and Outreach Group, *LHC the guide*, CERN-Brochure-2017-002-Eng, 2017, URL: `https://home.cern/sites/default/files/2018-07/CERN-Brochure-2017-002-Eng.pdf` (cit. on p. 31).

[101] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, Journal of Instrumentation **3** (2008) S08002 (cit. on p. 31).

[102] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** (2008) S08003. 437 p (cit. on pp. 31, 33, 34, 37, 38).

[103] CMS Collaboration, *The CMS experiment at the CERN LHC*, Journal of Instrumentation **3** (2008) S08004 (cit. on p. 31).

[104] The LHCb Collaboration, *The LHCb Detector at the LHC*, Journal of Instrumentation **3** (2008) S08005 (cit. on p. 31).

[105] E. Mobs, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, 2016, URL: `https://cds.cern.ch/record/2197559` (cit. on p. 32).

[106] J. Wenninger, *Operation and Configuration of the LHC in Run 2*, 2019, URL: `https://cds.cern.ch/record/2668326` (cit. on p. 32).

[107] J. Pequenao and P. Schaffner, *How ATLAS detects particles: diagram of particle paths in the detector*, 2013, URL: `https://cds.cern.ch/record/1505342` (cit. on p. 33).

[108] ATLAS Collaboration, *Study of the material of the ATLAS inner detector for Run 2 of the LHC*, JINST **12** (2017) P12009 (cit. on p. 35).

[109] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013, ATLAS-TDR-19, 2010, URL: `https://cds.cern.ch/record/1291633` (cit. on p. 35).

[110] B. Mindur and A. Collaboration, *ATLAS Transition Radiation Tracker (TRT): Straw tubes for tracking and particle identification at the Large Hadron Collider*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **845** (2017) 257, Proceedings of the Vienna Conference on Instrumentation 2016, ISSN: 0168-9002 (cit. on p. 36).

[111] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. C **77** (2017) 317 (cit. on p. 39).

[112] ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2*, Journal of Instrumentation **15** (2020) P10004 (cit. on p. 39).

[113] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019, URL: `http://cds.cern.ch/record/2677054` (cit. on pp. 40, 206).

[114] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, 12, 2016 653 (cit. on p. 40).

[115] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, Journal of Instrumentation **13** (2018) P07017 (cit. on pp. 40, 206).

[116] S. van der Meer, *Calibration of the effective beam height in the ISR*, tech. rep., CERN, 1968, URL: `https://cds.cern.ch/record/296752` (cit. on p. 40).

[117] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*, URL: `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2` (visited on 06/12/2022) (cit. on pp. 41, 51).

[118] W. Altmannshofer, J. Brod, and M. Schmaltz, *Experimental constraints on the coupling of the Higgs boson to electrons*, Journal of High Energy Physics **2015** (2015) (cit. on p. 44).

[119] ATLAS Collaboration, *Search for lepton-flavour-violating $H \to \mu\tau$ decays of the Higgs boson with the ATLAS detector*, Journal of High Energy Physics **2015** (2015) (cit. on p. 46).

[120] ATLAS Collaboration, *Search for lepton-flavour-violating decays of the Higgs and Z bosons with the ATLAS detector*, The European Physical Journal C **77** (2017) (cit. on pp. 46, 100).

[121] G. Sela, "A data-directed search for lepton-flavor violating decays of the Higgs in the hadronic tau decay channel", MA thesis: Weizmann Institute of Science, 2022 (cit. on p. 46).

[122] J. Rojo, *What hides inside a proton? From heavy quarks and photons to leptons and Higgs bosons*, Seminar of the RTG 2044 - Mass and Symmetries after the Discovery of the Higgs Particle at the LHC, 2020 (cit. on p. 47).

[123] M. L. Mangano and T. J. Stelzer, *Tools For The Simulation Of Hard Hadronic Collisions*, Annual Review of Nuclear and Particle Science **55** (2005) 555 (cit. on p. 47).

[124] P. Skands, "Introduction to QCD", *Searching for New Physics at Small and Large Scales*, WORLD SCIENTIFIC, 2013 (cit. on p. 47).

[125] A. Buckley et al., *General-purpose event generators for LHC physics*, Phys. Rept. **504** (2011) 145 (cit. on pp. 47, 49).

[126] R. Ellis, W. Stirling, and B. Webber, *QCD and Collider Physics*, Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, Cambridge University Press, 2003, ISBN: 9780521545891 (cit. on p. 47).

[127] J. Feltesse, *Introduction to Parton Distribution Functions*, Scholarpedia **5** (2010) 10160, revision #186761 (cit. on p. 47).

[128] J. Collins, *Parton distribution functions (definition)*, Scholarpedia **7** (2012) 10851, revision #128140 (cit. on p. 47).

[129] G. Altarelli, *QCD evolution equations for parton densities*, Scholarpedia **4** (2009) 7124, revision #91681 (cit. on p. 47).

[130] J. C. Collins and D. E. Soper, *The Theorems of Perturbative QCD*, Annual Review of Nuclear and Particle Science **37** (1987) 383 (cit. on p. 47).

[131] R. D. Ball et al., *Parton distributions from high-precision collider data*, Eur. Phys. J. C **77** (2017) 663 (cit. on p. 48).

[132] V. N. Gribov and L. N. Lipatov, *Deep inelastic electron scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438, URL: `https://cds.cern.ch/record/427157` (cit. on p. 48).

[133] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nuclear Physics B **126** (1977) 298 (cit. on p. 48).

[134] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, Sov. Phys. JETP **46** (1977) 641 (cit. on p. 48).

[135] L. Del Debbio, *Parton distributions in the LHC era*, EPJ Web of Conferences **175** (2018) 01006 (cit. on p. 48).

[136] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Phys. Rev. Lett. **30** (26 1973) 1343 (cit. on p. 49).

[137] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (26 1973) 1346 (cit. on p. 49).

[138] T. Gleisberg et al., *Event generation with SHERPA 1.1*, Journal of High Energy Physics **2009** (2009) 007 (cit. on p. 50).

[139] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, *Parton fragmentation and string dynamics*, Physics Reports **97** (1983) 31, ISSN: 0370-1573 (cit. on p. 49).

[140] B. Andersson, *The Lund model*, vol. 7, Cambridge University Press, 2005, ISBN: 978-0-521-01734-3, 978-0-521-42094-5, 978-0-511-88149-7 (cit. on p. 49).

[141] B. Webber, *A QCD model for jet fragmentation including soft gluon interference*, Nuclear Physics B **238** (1984) 492, ISSN: 0550-3213 (cit. on p. 49).

[142] J.-C. Winter, F. Krauss, and G. Soff, *A Modified cluster hadronization model*, Eur. Phys. J. C **36** (2004) 381 (cit. on p. 49).

[143] S. Catani, F. Krauss, B. R. Webber, and R. Kuhn, *QCD Matrix Elements + Parton Showers*, Journal of High Energy Physics **2001** (2001) 063 (cit. on p. 50).

[144] M. L. Mangano, M. Moretti, and R. Pittau, *Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b}$ + n jets as a case study*, Nucl. Phys. B **632** (2002) 343 (cit. on p. 50).

[145] Z. Marshall, *Simulation of Pile-up in the ATLAS Experiment*, ATL-SOFT-PROC-2013-030, CERN, 2013, URL: `https://cds.cern.ch/record/1616394` (cit. on p. 51).

[146] S. Agostinelli et al., *GEANT4–a simulation toolkit. GEANT4. A Simulation toolkit*, Nucl. Instrum. Methods Phys. Res., A **506** (2002) 250 (cit. on p. 51).

[147] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040 (cit. on pp. 53, 54, 56, 57).

[148] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070 (cit. on pp. 53, 54, 56, 57).

[149] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043 (cit. on pp. 53, 54, 56, 57).

[150] J. M. Campbell et al., *NLO Higgs Boson Production Plus One and Two Jets Using the POWHEG BOX, MadGraph4 and MCFM*, JHEP **07** (2012) 092 (cit. on p. 53).

[151] K. Hamilton, P. Nason, E. Re, and G. Zanderighi, *NNLOPS simulation of Higgs boson production*, Journal of High Energy Physics **2013** (2013) (cit. on p. 53).

[152] K. Hamilton, P. Nason, and G. Zanderighi, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, 2015 (cit. on p. 53).

[153] C. Anastasiou, C. Duhr, F. Dulat, F. Herzog, and B. Mistlberger, *Higgs Boson Gluon-Fusion Production in QCD at Three Loops*, Phys. Rev. Lett. **114** (21 2015) 212001 (cit. on p. 53).

[154] S. Actis, G. Passarino, C. Sturm, and S. Uccirati, *NLO electroweak corrections to Higgs boson production at hadron colliders*, Physics Letters B **670** (2008) 12 (cit. on p. 53).

[155] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, JHEP **02** (2010) 037 (cit. on p. 53).

[156] K. Hamilton, P. Nason, and G. Zanderighi, *MINLO: multi-scale improved NLO*, Journal of High Energy Physics **2012** (2012) (cit. on p. 53).

[157] G. Luisoni, P. Nason, C. Oleari, and F. Tramontano, *HW ±/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO*, Journal of High Energy Physics **2013** (2013) (cit. on p. 53).

[158] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, Journal of Physics G: Nuclear and Particle Physics **43** (2016) 023001 (cit. on pp. 53, 208).

[159] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Computer Physics Communications **191** (2015) 159 (cit. on pp. 53, 56, 57).

[160] ATLAS Collaboration, *Measurement of the $Z/\gamma^*$ boson transverse momentum distribution in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, JHEP **09** (2014) 145 (cit. on p. 53).

[161] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012 (cit. on p. 53).

[162] E. Bothmann et al., *Event Generation with Sherpa 2.2*, SciPost Phys. **7** (3 2019) 34 (cit. on p. 54).

[163] F. Krauss, R. Kuhn, and G. Soff, *AMEGIC++ 1.0: A Matrix element generator in C++*, JHEP **02** (2002) 044 (cit. on p. 54).

[164] C. Duhr, S. Hoeche, and F. Maltoni, *Color-dressed recursive relations for multi-parton amplitudes*, JHEP **08** (2006) 062 (cit. on p. 54).

[165] F. Cascioli, P. Maierhofer, and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601 (cit. on p. 54).

[166] R. D. Ball et al., *Parton distributions for the LHC run II*, Journal of High Energy Physics **2015** (2015) (cit. on pp. 54, 56, 57).

[167] S. Hoeche, F. Krauss, M. Schonherr, and F. Siegert, *A critical appraisal of NLO+PS matching methods*, JHEP **09** (2012) 049 (cit. on p. 54).

[168] Y. Li and F. Petriello, *Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code*, Physical Review D **86** (2012) (cit. on pp. 54, 55).

[169] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO vector-boson production matched with shower in POWHEG*, JHEP **07** (2008) 060 (cit. on p. 54).

[170] D. Lange, *The EvtGen particle decay simulation package*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462** (2001) 152 (cit. on pp. 54, 56, 57).

[171] E. Barberio and Z. Was, *PHOTOS: A Universal Monte Carlo for QED radiative corrections. Version 2.0*, Comput. Phys. Commun. **79** (1994) 291 (cit. on p. 55).

[172] Z. Was, P. Golonka, and G. Nanava, *PHOTOS Monte Carlo and its theoretical accuracy*, Nucl. Phys. B Proc. Suppl. **181-182** (2008) 269, ed. by C. Bini and G. Venanzoni (cit. on p. 55).

[173] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, CERN, 2014, URL: https://cds.cern.ch/record/1966419 (cit. on p. 56).

[174] T Cornelissen and M Elsing and I Gavrilenko and W Liebig and E Moyse and A Salzburger, *The new ATLAS track reconstruction (NEWT)*, Journal of Physics: Conference Series **119** (2008) 032014 (cit. on p. 62).

[175] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, The European Physical Journal C **77** (2017), ISSN: 1434-6052 (cit. on p. 62).

[176] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **262** (1987) 444, ISSN: 0168-9002 (cit. on p. 62).

[177] ATLAS Collaboration, *Tracking CP Recommendations for Early 2018 Analyses*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TrackingCPRecsEarly2018%5C#Track_Selection (visited on 03/07/2022) (cit. on p. 62).

[178] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, The European Physical Journal C **77** (2017), ISSN: 1434-6052 (cit. on p. 62).

[179] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-026, CERN, 2015, URL: https://cds.cern.ch/record/2037717 (cit. on p. 63).

[180] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, Journal of Instrumentation **14** (2019) P12006, ISSN: 1748-0221 (cit. on pp. 63–67, 124, 202).

[181] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, The European Physical Journal C **77** (2017), ISSN: 1434-6052 (cit. on p. 63).

[182] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, ATL-COM-LARG-2008-003, CERN, 2008, URL: https://cds.cern.ch/record/1099735 (cit. on p. 63).

[183] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047, CERN, 2012, URL: https://cds.cern.ch/record/1449796 (cit. on p. 64).

[184] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data*, Journal of Instrumentation **14** (2019) P03017, ISSN: 1748-0221 (cit. on pp. 64, 202).

[185] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s} = 13\,TeV$, The European Physical Journal C **79** (2019), ISSN: 1434-6052 (cit. on pp. 65, 66, 202).

[186] Cacciari, Matteo and Salam, Gavin P., *Pileup subtraction using jet areas*, Physics Letters B **659** (2008) 119, ISSN: 0370-2693 (cit. on p. 66).

[187] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s} = 13\ TeV$, Eur. Phys. J. C **81** (2021) 578 (cit. on pp. 67–70, 123).

[188] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s} = 13\,TeV$, The European Physical Journal C **76** (2016), ISSN: 1434-6052 (cit. on pp. 67, 68, 120, 203).

[189] ATLAS Collaboration, *Lepton Isolation Working Point Performance*, URL: `https://indico.cern.ch/event/854783/contributions/3641582/attachments/1944964/3228483/booklet_isoEff.pdf` (visited on 04/20/2022) (cit. on p. 71).

[190] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, The European Physical Journal C **77** (2017) (cit. on p. 71).

[191] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-$k_t$ jet clustering algorithm*, Journal of High Energy Physics **2008** (2008) 063 (cit. on p. 71).

[192] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp-collisions at* $\sqrt{s} = 8\,TeV$ *using the ATLAS detector*, The European Physical Journal C **76** (2016) (cit. on pp. 72, 205).

[193] ATLAS Collaboration, *Forward jet vertex tagging using the particle flow algorithm*, ATL-PHYS-PUB-2019-026, CERN, 2019, URL: `https://cds.cern.ch/record/2683100` (cit. on pp. 72, 205).

[194] ATLAS Collaboration, *JVT Calibration*, URL: `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JVTCalibration` (visited on 05/01/2022) (cit. on p. 72).

[195] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at* $\sqrt{s} = 13\ TeV$ *with the ATLAS detector*, Eur. Phys. J. C **81** (2021) 689 (cit. on pp. 72, 204).

[196] M. Guth, "*Search for ttH(bb) Production in the Lepton + Jets Channel and Quark Flavour Tagging with Deep Learning at the ATLAS Experiment*", PhD thesis: Albert-Ludwigs-Universität Freiburg, 2021, URL: `https://doi.org/10.6094/UNIFR/194834` (cit. on p. 73).

[197] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with* $t\bar{t}$ *events in pp collisions at* $\sqrt{s} = 13\ TeV$, The European Physical Journal C **79** (2019), ISSN: 1434-6052 (cit. on pp. 73, 205).

[198] P. Zyla et al., *Review of Particle Physics*, PTEP **2020** (2020) 083C01, and 2021 update (cit. on p. 73).

[199] ATLAS Collaboration, *Expected performance of the 2019 ATLAS b-taggers*, URL: `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-005/` (visited on 04/27/2022) (cit. on p. 73).

[200] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at* $\sqrt{s} = 13$ TeV, ATLAS-CONF-2017-029, CERN, 2017, URL: https://cds.cern.ch/record/2261772 (cit. on p. 74).

[201] T. Barillari et al., *Local Hadronic Calibration*, ATL-LARG-PUB-2009-001-2, ATL-COM-LARG-2008-006, ATL-LARG-PUB-2009-001, CERN, 2008, URL: https://cds.cern.ch/record/1112035 (cit. on p. 74).

[202] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, ATL-PHYS-PUB-2019-033, CERN, 2019, URL: https://cds.cern.ch/record/2688062 (cit. on p. 74).

[203] ATLAS Collaboration, $E_T^{miss}$ *performance in the ATLAS detector using 2015-2016 LHC p-p collisions*, ATLAS-CONF-2018-023, CERN, 2018, URL: https://cds.cern.ch/record/2625233 (cit. on pp. 75, 76, 205).

[204] ATLAS Collaboration, *Overlap removal tools*, URL: https://gitlab.cern.ch/atlas/athena/blob/21.2/PhysicsAnalysis/AnalysisCommon/AssociationUtils/README.rst (visited on 04/29/2022) (cit. on pp. 76, 77).

[205] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed., Wiley, 2001, ISBN: 978-0-471-05669-0 (cit. on p. 79).

[206] M. A. Nielsen, *Neural Networks and Deep Learning*, Determination Press, 2015 (cit. on pp. 80–82, 85, 87).

[207] K. Fukushima, *Visual Feature Extraction by a Multilayered Network of Analog Threshold Elements*, IEEE Transactions on Systems Science and Cybernetics **5** (1969) 322 (cit. on p. 81).

[208] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models", *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 2013 (cit. on p. 81).

[209] A. Cauchy, *Méthode générale pour la résolution des systèmes d'équations simultanées*, C. R. Acad. Sci. Paris **25:536–538** (1847) (cit. on p. 83).

[210] J. Hadamard, *Mémoire sur le problème d'analyse relatif à Véquilibre des plaques élastiques encastrées*, présentés à l'Académie des Sciences de l'Institut de France (1908) (cit. on p. 83).

[211] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning Representations by Back-propagating Errors*, Nature **323** (1986) 533 (cit. on p. 85).

[212] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation", *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*, MIT Press, 1986 318 (cit. on p. 85).

[213] Kingma, Diederik P. and Ba, Jimmy, *Adam: A Method for Stochastic Optimization*, 2014 (cit. on pp. 86, 87).

[214] P. Mehta et al., *A high-bias, low-variance introduction to Machine Learning for physicists*, Physics Reports **810** (2019) 1 (cit. on p. 87).

[215] J. P. Egan, "Signal detection theory and ROC analysis", 1975 (cit. on p. 88).

[216] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, The European Physical Journal C **71** (2011) (cit. on pp. 89, 219, 223, 225).

[217] C. Molnar, *Interpretable Machine Learning, A Guide for Making Black Box Models Explainable*, 2nd ed., 2022, URL: https://christophm.github.io/interpretable-ml-book (cit. on p. 92).

[218] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework", *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019 (cit. on pp. 93, 94, 172).

[219] S. Falkner, A. Klein, and F. Hutter, "BOHB: Robust and Efficient Hyperparameter Optimization at Scale", *Proceedings of the 35th International Conference on Machine Learning*, ed. by J. Dy and A. Krause, vol. 80, Proceedings of Machine Learning Research, PMLR, 2018 1437, URL: https://proceedings.mlr.press/v80/falkner18a.html (cit. on p. 93).

[220] J. Bergstra and Y. Bengio, *Random Search for Hyper-Parameter Optimization*, Journal of Machine Learning Research **13** (2012) 281, URL: http://jmlr.org/papers/v13/bergstra12a.html (cit. on p. 93).

[221] J. Močkus, "On bayesian methods for seeking the extremum", *Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974*, ed. by G. I. Marchuk, Springer Berlin Heidelberg, 1975 400, ISBN: 978-3-540-37497-8 (cit. on p. 94).

[222] D. Jones, M. Schonlau, and W. Welch, *Efficient Global Optimization of Expensive Black-Box Functions*, Journal of Global Optimization **13** (1998) 455 (cit. on p. 94).

[223] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for Hyper-Parameter Optimization", *Advances in Neural Information Processing Systems*, ed. by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, vol. 24, Curran Associates, Inc., 2011, URL: https://proceedings.neurips.cc/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf (cit. on p. 94).

[224] J. Bergstra, D. Yamins, and D. Cox, "Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures", *Proceedings of the 30th International Conference on Machine Learning*, ed. by S. Dasgupta and D. McAllester, vol. 28, Proceedings of Machine Learning Research 1, PMLR, 2013 115, URL: https://proceedings.mlr.press/v28/bergstra13.html (cit. on p. 94).

[225] Y. Ozaki, Y. Tanigaki, S. Watanabe, and M. Onishi, "Multiobjective Tree-Structured Parzen Estimator for Computationally Expensive Optimization Problems", *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, GECCO '20, Association for Computing Machinery, 2020 533, ISBN: 9781450371285 (cit. on p. 94).

[226] Optuna Contributors, *Optuna Documentation*, URL: https://optuna.readthedocs.io/en/stable/reference/generated/optuna.pruners.MedianPruner.html (visited on 07/11/2022) (cit. on p. 94).

[227] ATLAS Collaboration, *TrigGlobalEfficiencyCorrectionTool*, URL: https://gitlab.cern.ch/atlas/athena/-/blob/21.2/Trigger/TrigAnalysis/TrigGlobalEfficiencyCorrection/doc/formulas.pdf (visited on 07/20/2022) (cit. on p. 97).

[228] ATLAS Collaboration, *Bad Batman Event Event Cleaning*, URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HowToCleanJets2017#IsBadBatMan_Event_Flag_and_EMEC (visited on 07/20/2022) (cit. on p. 97).

[229] V. Lang, private communication, 2022 (cit. on pp. 106, 108).

[230] ATLAS Collaboration, *IFFTruthClassifier*, URL: https://gitlab.cern.ch/ATLAS-IFF/IFFTruthClassifier (visited on 08/08/2022) (cit. on pp. 107, 115, 126).

[231] E. J. Dorbath, "Search for lepton-flavour violation in $pp \to X \to \tau e(\tau\mu)$ decays with the ATLAS experiment", MA thesis: Albert-Ludwigs-Universität Freiburg, 2020 (cit. on p. 107).

[232] A. Collaboration, *Multi-Boson Simulation for 13 TeV ATLAS Analyses*, CERN, 2017, URL: https://cds.cern.ch/record/2261933 (cit. on p. 109).

[233] ATLAS Collaboration, *Search for lepton-flavour violating decays of the Higgs boson at $\sqrt{s} = 13$ TeV using data-driven background estimates*, CERN, 2020, URL: https://cds.cern.ch/record/2743987 (cit. on pp. 109, 111).

[234] M. Birman, private communication, 2022 (cit. on pp. 110, 127).

[235] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, JINST **15** (2020) P09015 (cit. on pp. 120, 122).

[236] A. Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, Eur. Phys. J. C **80** (2020) 47 (cit. on p. 123).

[237] T. Hryn'ova, F. Monticelli, M. Spina, B. Safarzadeh Samani, and D. M. Koeck, *Single electron trigger 2018 performance plots*, CERN, 2019, URL: https://cds.cern.ch/record/2668252 (cit. on pp. 124, 125).

[238] R. K. Ellis, I. Hinchliffe, M. Soldate, and J. J. Van Der Bij, *Higgs decay to $\tau^+\tau^-$: A possible signature of intermediate mass Higgs bosons at high energy hadron colliders*, Nuclear Physics B **297** (1988) 221, ISSN: 0550-3213 (cit. on p. 161).

[239] A. Elagin, P. Murat, A. Pranko, and A. Safonov, *A new mass reconstruction technique for resonances decaying to $\tau\tau$*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **654** (2011) 481, ISSN: 0168-9002 (cit. on p. 161).

[240] S. Davidson, S. Lacroix, and P. Verdier, *LHC sensitivity to lepton flavour violating Z boson decays*, JHEP **09** (2012) 092 (cit. on p. 164).

[241] J. Erdmann, T. Kallage, K. Kröninger, and O. Nackenhorst, *From the bottom to the top—reconstruction of $t\bar{t}$ events with deep learning*, JINST **14** (2019) P11015 (cit. on p. 172).

[242] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks", *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, ed. by Y. W. Teh and M. Titterington, vol. 9, Proceedings of Machine Learning Research, PMLR, 2010 249, URL: https://proceedings.mlr.press/v9/glorot10a.html (cit. on p. 172).

[243] *Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, Journal of High Energy Physics **2018** (2018) (cit. on p. 205).

[244] I. W. Stewart and F. J. Tackmann, *Theory uncertainties for Higgs mass and other searches using jet bins*, Physical Review D **85** (2012) (cit. on p. 206).

[245] S. Gangal and F. J. Tackmann, *NLO Uncertainties in Higgs + 2 jets from Gluon Fusion*, Physical Review D **87** (2013) (cit. on p. 206).

[246] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, Journal of High Energy Physics **2014** (2014) (cit. on p. 208).

[247] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, The European Physical Journal C **76** (2016) (cit. on p. 208).

[248] M. Bähr et al., *Herwig++ physics and manual*, The European Physical Journal C **58** (2008) 639 (cit. on p. 208).

[249] ATLAS Collaboration, *TRExFitter*, URL: https://gitlab.cern.ch/TRExStats/TRExFitter (visited on 10/24/2022) (cit. on pp. 216, 219).

[250] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, New York U., 2012, URL: https://cds.cern.ch/record/1456844 (cit. on pp. 219, 221).

[251] L. Moneta et al., *The RooStats Project*, PoS **ACAT2010** (2010) 057, ed. by T. Speer et al. (cit. on p. 219).

[252] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, ed. by L. Lyons and M. Karagoz (cit. on p. 219).

[253] G. Cowan, *Statistical data analysis*, 1998, ISBN: 978-0-19-850156-5 (cit. on pp. 219, 220).

[254] ATLAS Collaboration, *Statistical Methods for the LHC*, URL: https://atlas-stats-doc-dev.web.cern.ch/atlas-stats-doc-dev/recommendations/rec_stat_computations/ (visited on 11/03/2022) (cit. on pp. 219, 223).

[255] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, Comput. Phys. Commun. **10** (1975) 343 (cit. on p. 220).

[256] R. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Computer Physics Communications **77** (1993) 219, ISSN: 0010-4655 (cit. on p. 221).

[257] J. Neyman and E. S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character **231** (1933) 289, ISSN: 02643952, URL: https://doi.org/10.1098/rsta.1933.0009 (cit. on p. 222).

[258] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, The Annals of Mathematical Statistics **9** (1938) 60 (cit. on pp. 222, 223).

[259] A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, Transactions of the American Mathematical Society **54** (1943) 426, ISSN: 00029947, URL: http://www.jstor.org/stable/1990256 (cit. on p. 223).

[260] ATLAS Statistics Forum, *The $CL_s$ method: information for conference speakers*, URL: https://twiki.cern.ch/twiki/pub/Sandbox/NotesOnStatistics/CLsInfo.pdf (visited on 11/03/2022) (cit. on pp. 223, 224).

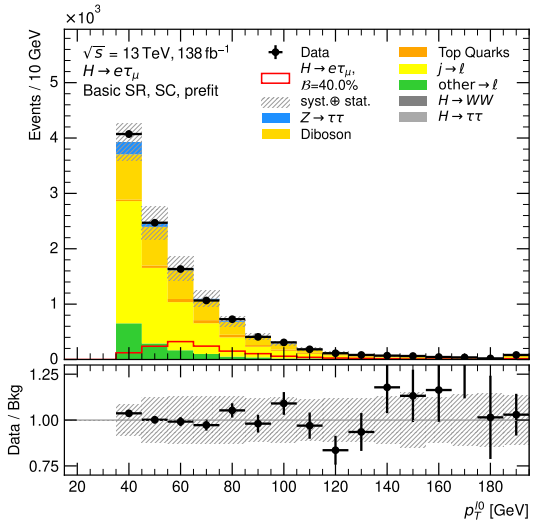
[261] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, 1994, ISBN: 9780412042317 (cit. on p. 241).

[262] ATLAS Collaboration, *Evaluating statistical uncertainties and correlations using the bootstrap method*, ATL-PHYS-PUB-2021-011, 2021, URL: https://cds.cern.ch/record/2759945 (cit. on p. 241).

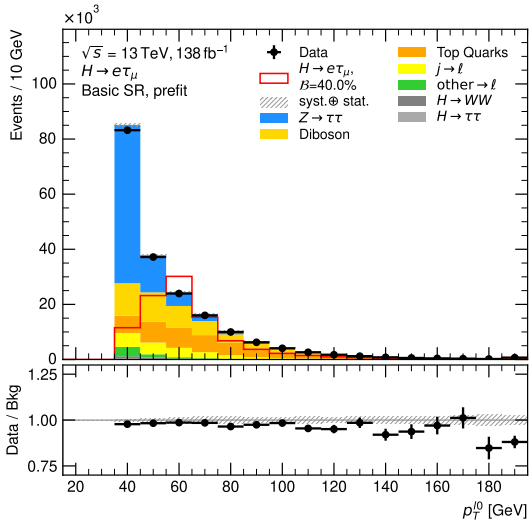
[263] A. Dery, private communication, 2023 (cit. on pp. 244, 332).

# Appendix A

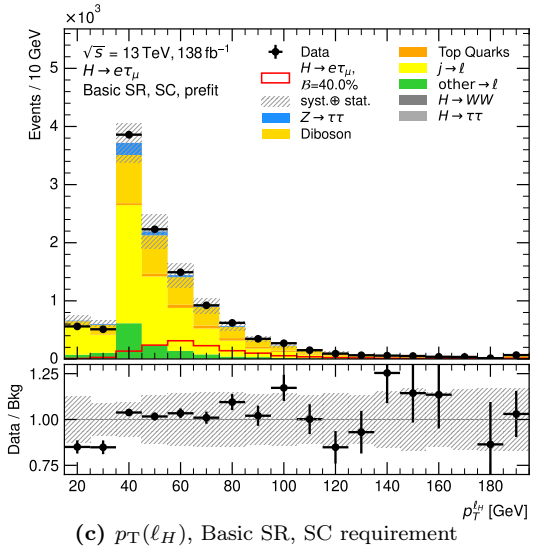
# Validation of the *Fake*-Lepton Background Estimate



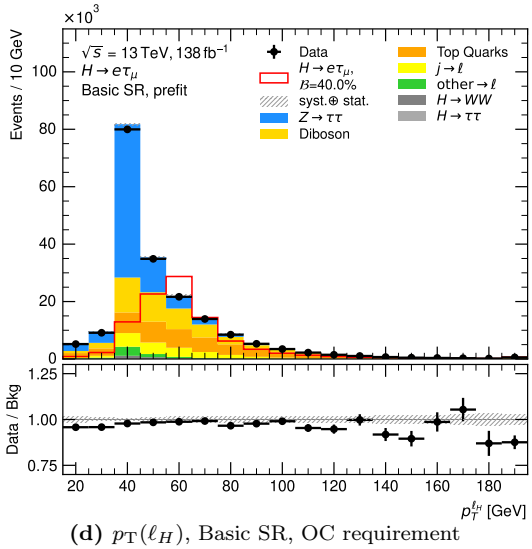

**(a)** $p_{\mathrm{T}}(\ell_0)$, Basic SR, SC requirement



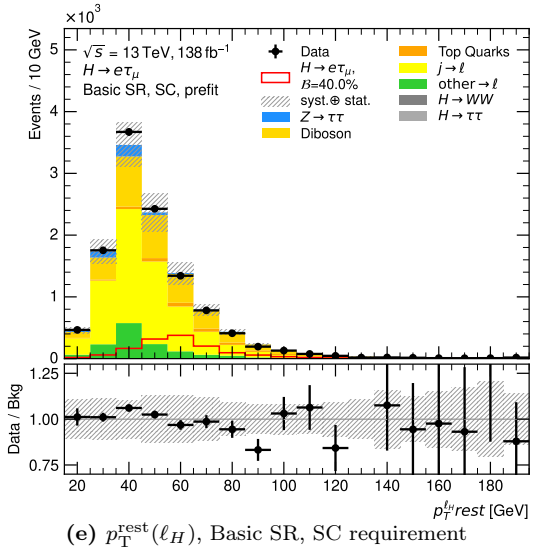

**(b)** $p_{\mathrm{T}}(\ell_0)$, Basic SR, OC requirement



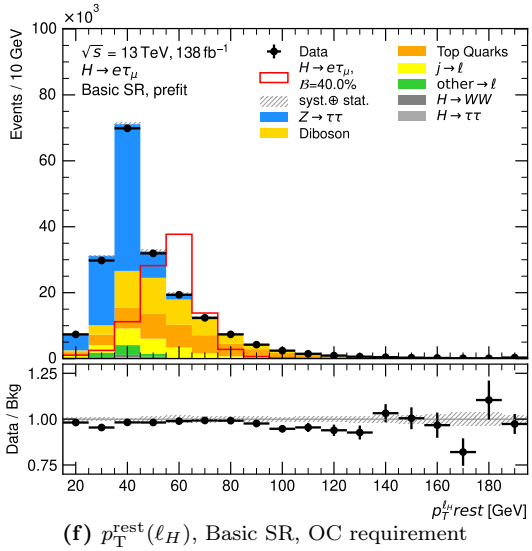

**(c)** $p_{\mathrm{T}}(\ell_H)$, Basic SR, SC requirement




**(d)** $p_{\mathrm{T}}(\ell_H)$, Basic SR, OC requirement

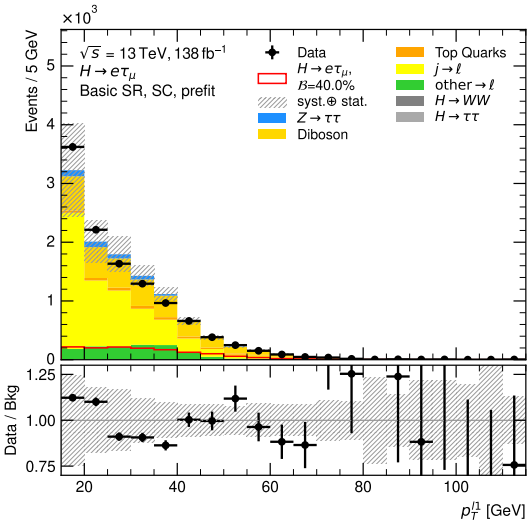



**(e)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_H)$, Basic SR, SC requirement

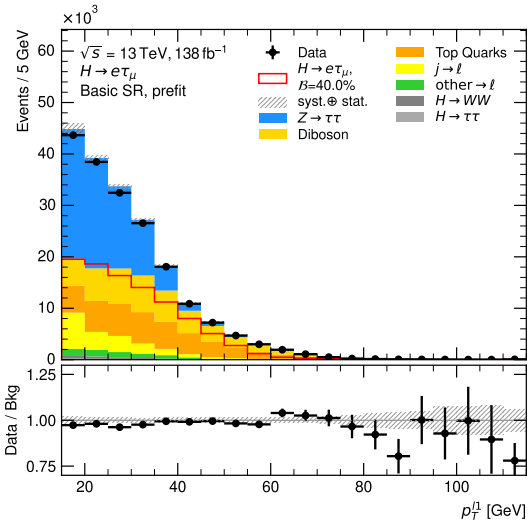



**(f)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_H)$, Basic SR, OC requirement

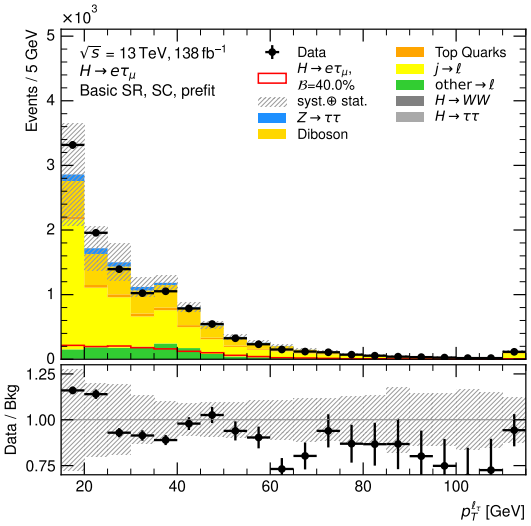
**Figure A.1:** Distributions in the $e\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate.



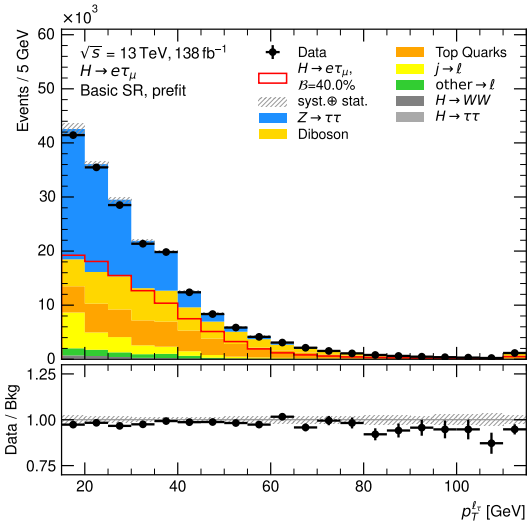

**(a)** $p_{\mathrm{T}}(\ell_1)$, Basic SR, SC requirement



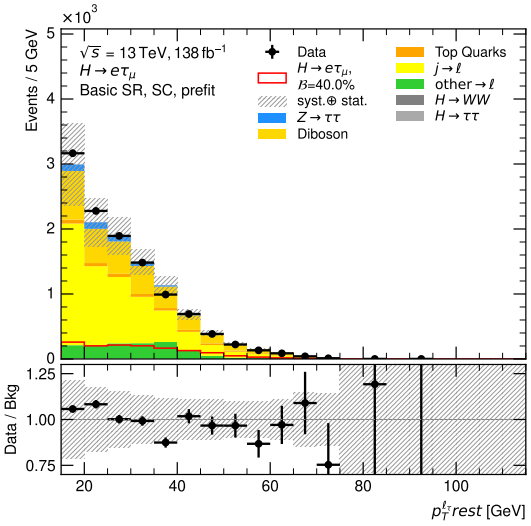

**(b)** $p_{\mathrm{T}}(\ell_1)$, Basic SR, OC requirement



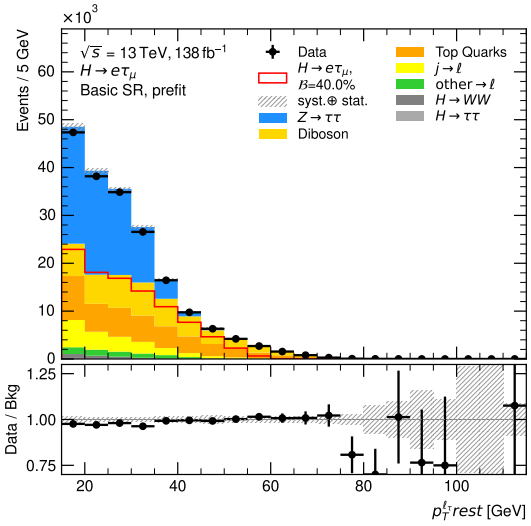

**(c)** $p_{\mathrm{T}}(\ell_\tau)$, Basic SR, SC requirement




**(d)** $p_{\mathrm{T}}(\ell_\tau)$, Basic SR, OC requirement

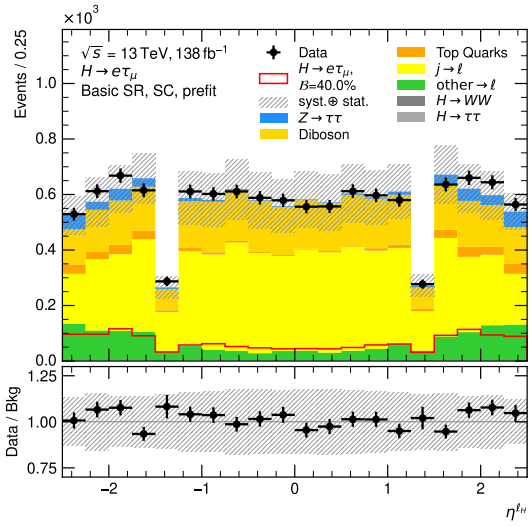



**(e)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_\tau)$, Basic SR, SC requirement

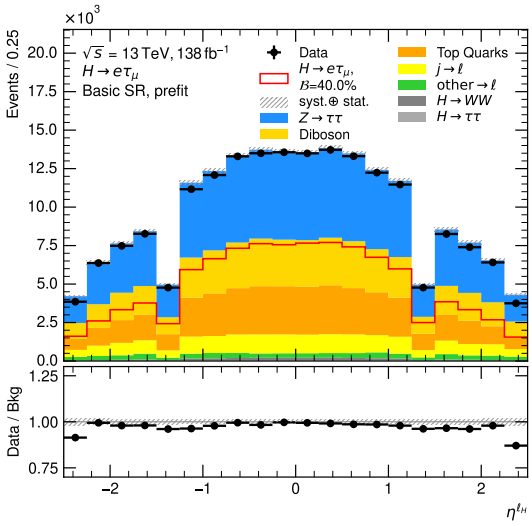



**(f)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_\tau)$, Basic SR, OC requirement

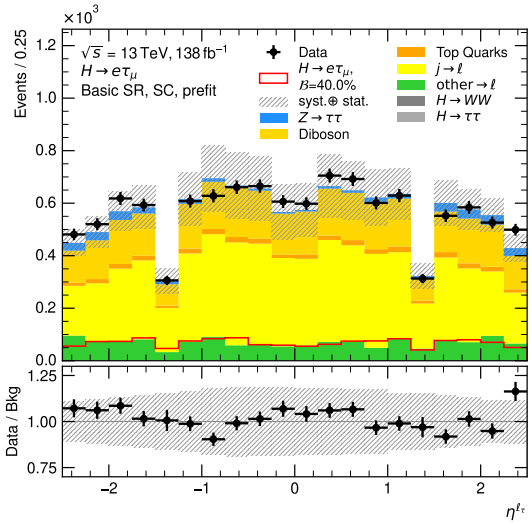
**Figure A.2:** Distributions in the $e\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate.



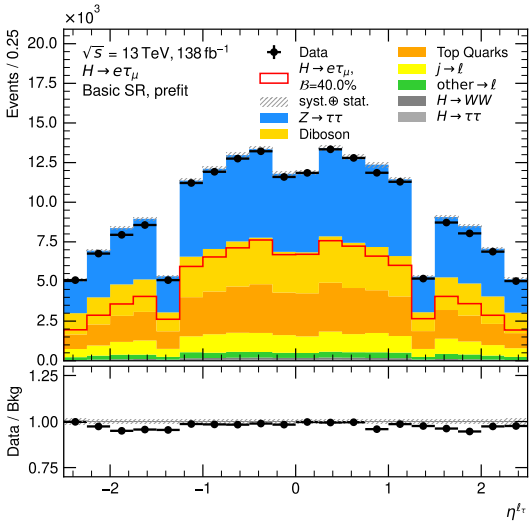

**(a)** $\eta(\ell_H)$, Basic SR, SC requirement



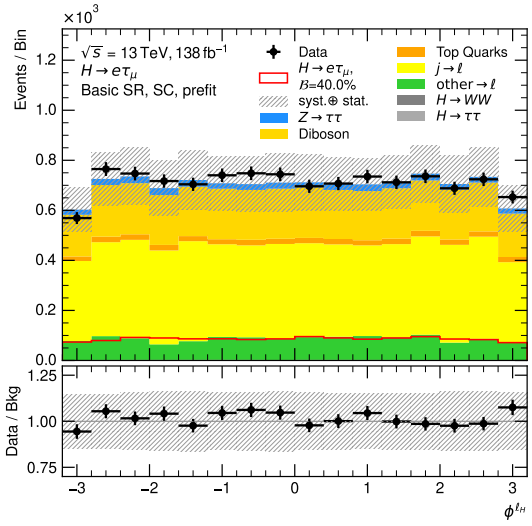

**(b)** $\eta(\ell_H)$, Basic SR, OC requirement



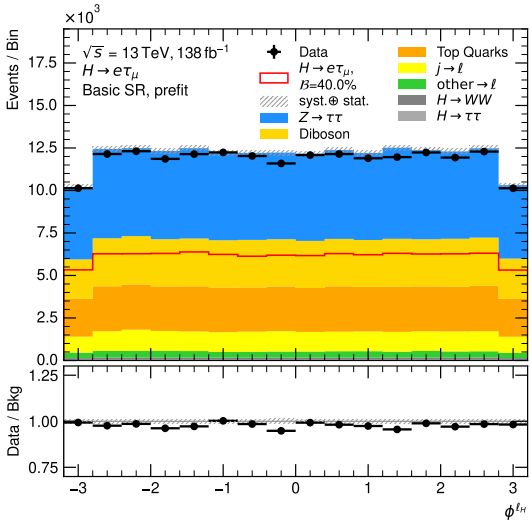

**(c)** $\eta(\ell_\tau)$, Basic SR, SC requirement




**(d)** $\eta(\ell_\tau)$, Basic SR, OC requirement

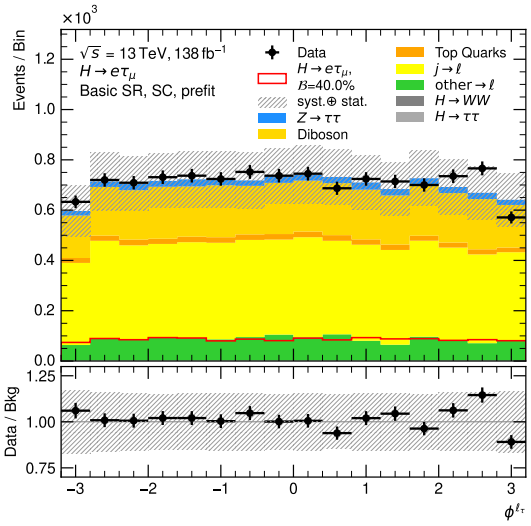



**(e)** $\phi(\ell_H)$, Basic SR, SC requirement

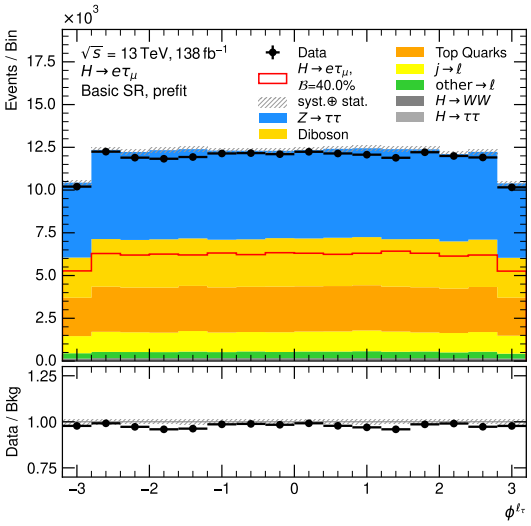



**(f)** $\phi(\ell_H)$, Basic SR, OC requirement

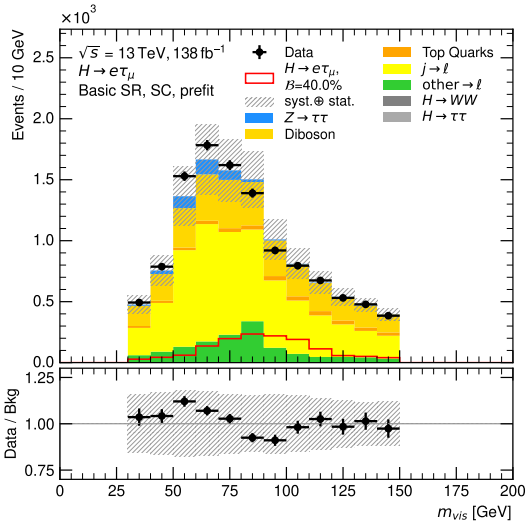
**Figure A.3:** Distributions in the $e\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate.



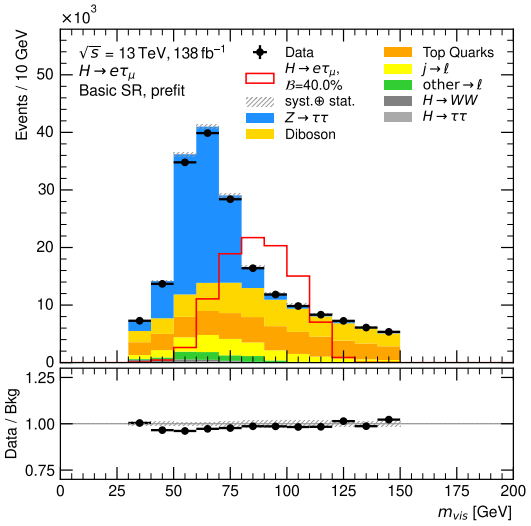

**(a)** $\phi(\ell_\tau)$, Basic SR, SC requirement



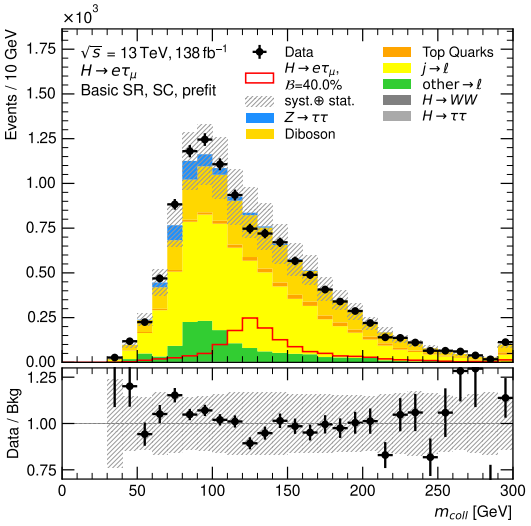

**(b)** $\phi(\ell_\tau)$, Basic SR, OC requirement



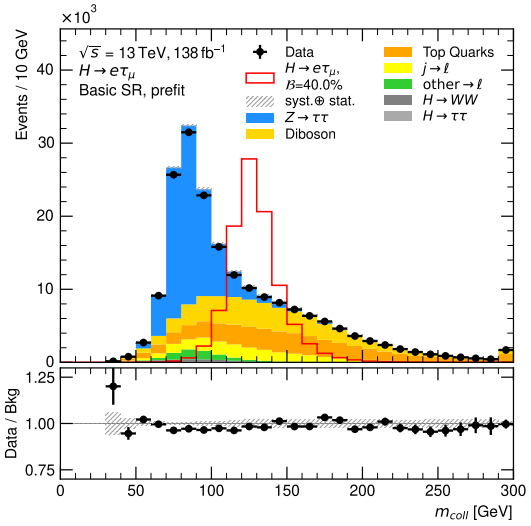

**(c)** $m_{\ell\ell}$, Basic SR, SC requirement




**(d)** $m_{\ell\ell}$, Basic SR, OC requirement

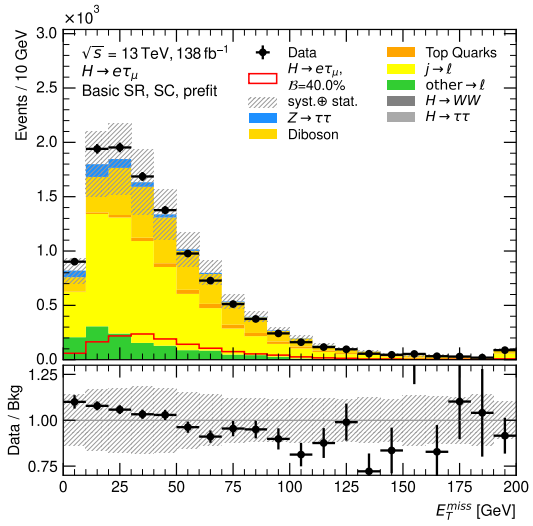



**(e)** $m_{\text{coll}}$, Basic SR, SC requirement

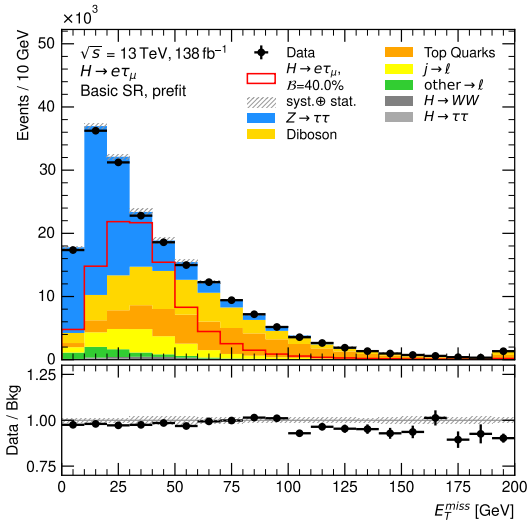



**(f)** $m_{\text{coll}}$, Basic SR, OC requirement

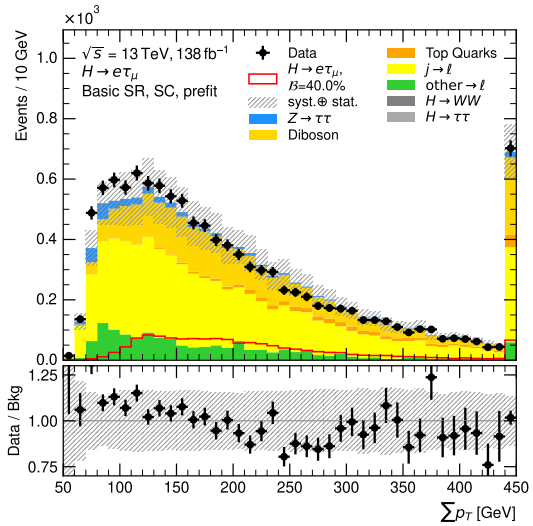
**Figure A.4:** Distributions in the $e\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate.



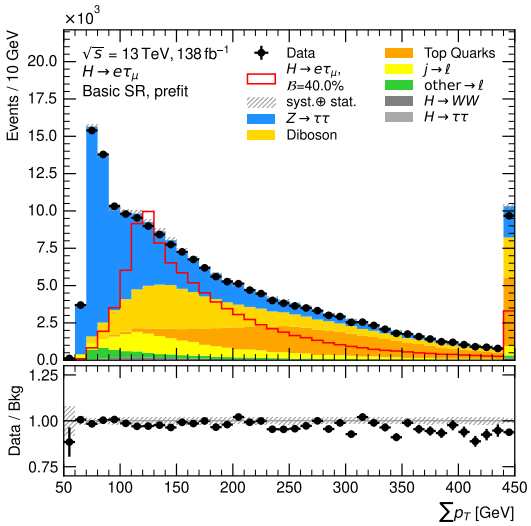

**(a)** $E_{\mathrm{T}}^{\mathrm{miss}}$, Basic SR, SC requirement



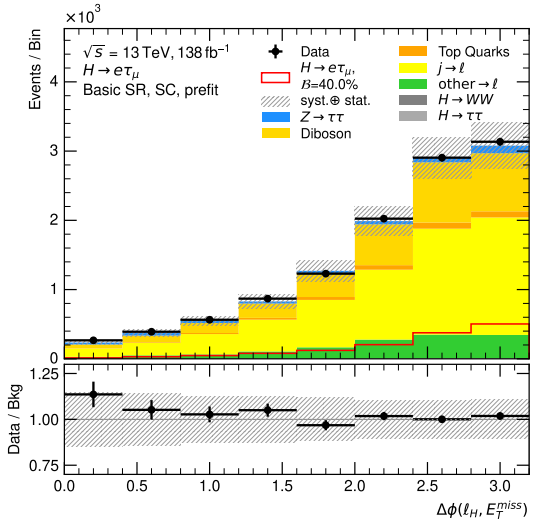

**(b)** $E_{\mathrm{T}}^{\mathrm{miss}}$, Basic SR, OC requirement



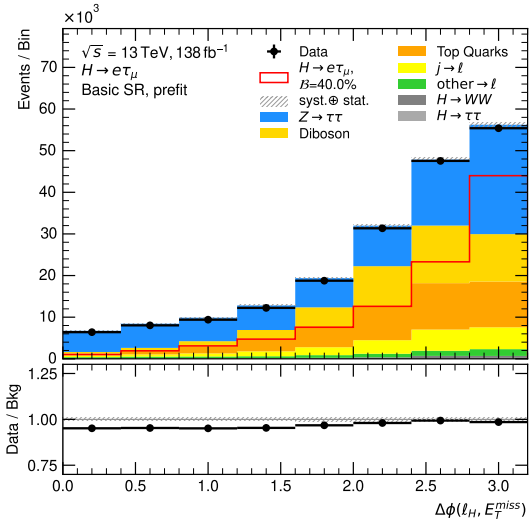

**(c)** $\sum p_{\mathrm{T}}$, Basic SR, SC requirement




**(d)** $\sum p_{\mathrm{T}}$, Basic SR, OC requirement

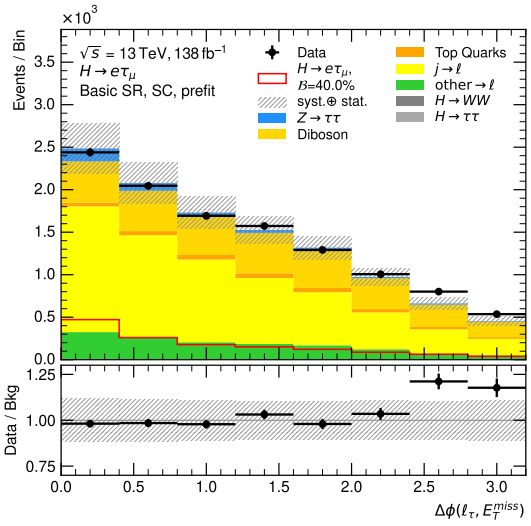



**(e)** $\Delta\phi(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, SC requirement

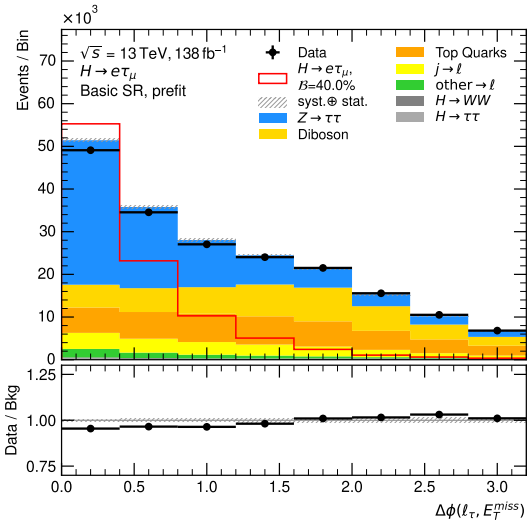



**(f)** $\Delta\phi(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, OC requirement

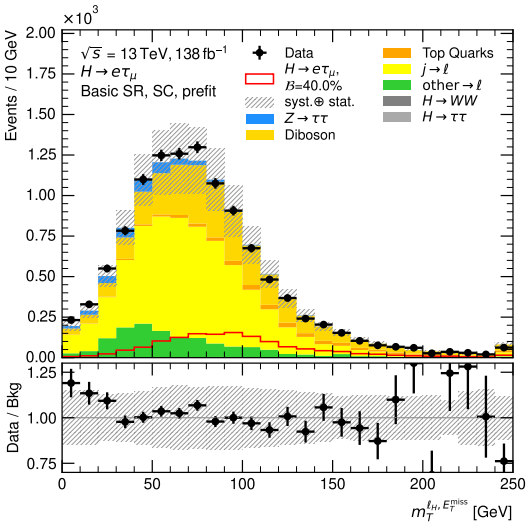
**Figure A.5:** Distributions in the $e\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate.



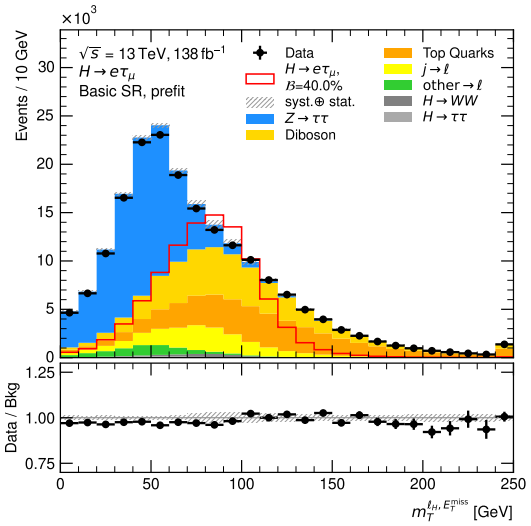

**(a)** $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, SC requirement



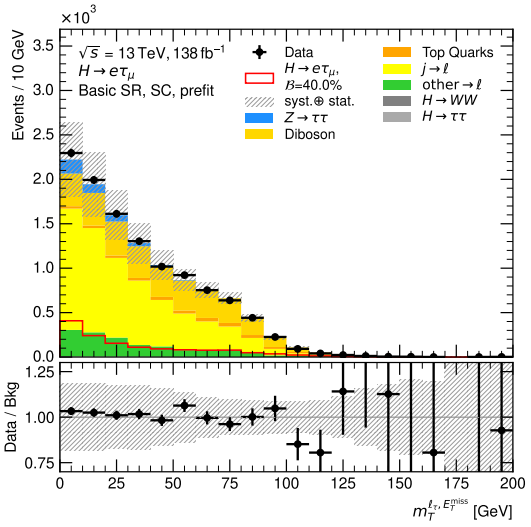

**(b)** $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, OC requirement



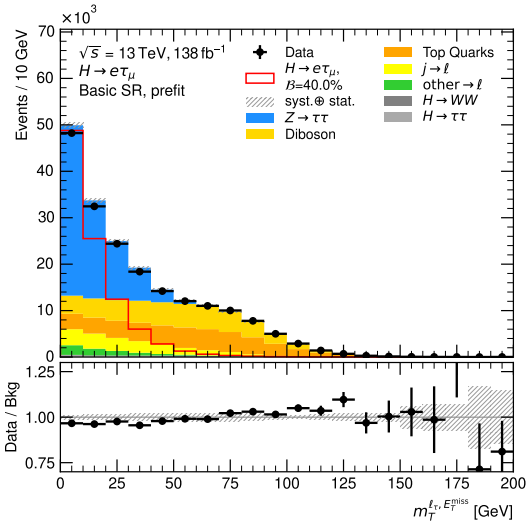

**(c)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, SC requirement




**(d)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, OC requirement

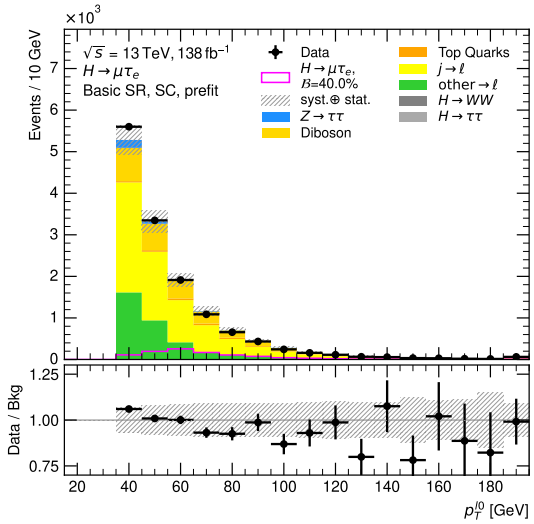



**(e)** $m_T(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, SC requirement

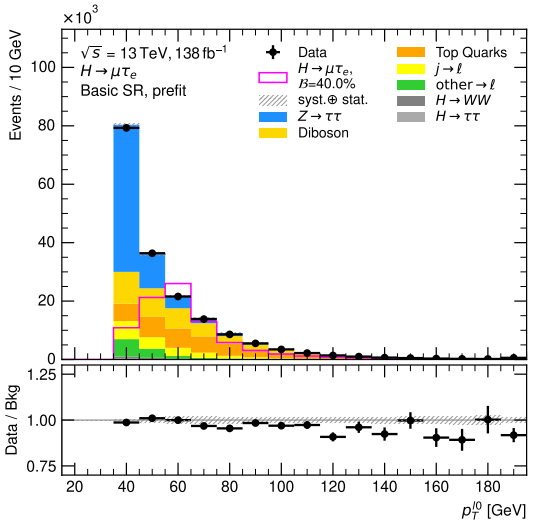



**(f)** $m_T(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, OC requirement

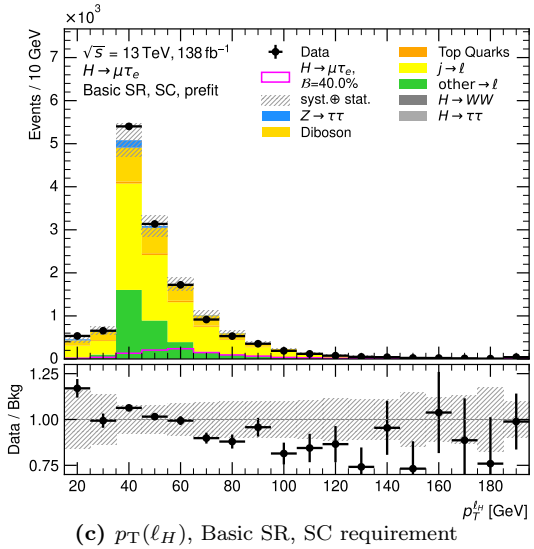
**Figure A.6:** Distributions in the $e\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate.



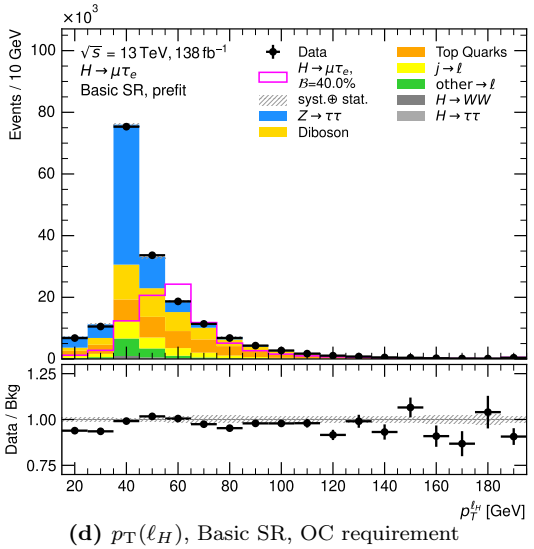

**(a)** $p_{\mathrm{T}}(\ell_0)$, Basic SR, SC requirement



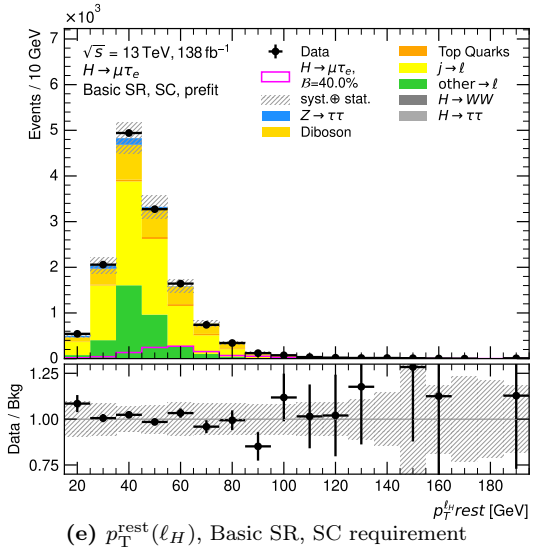

**(b)** $p_{\mathrm{T}}(\ell_0)$, Basic SR, OC requirement



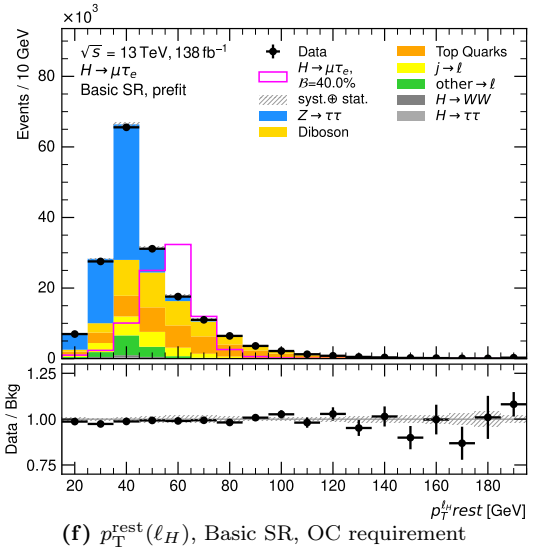

**(c)** $p_{\mathrm{T}}(\ell_H)$, Basic SR, SC requirement




**(d)** $p_{\mathrm{T}}(\ell_H)$, Basic SR, OC requirement

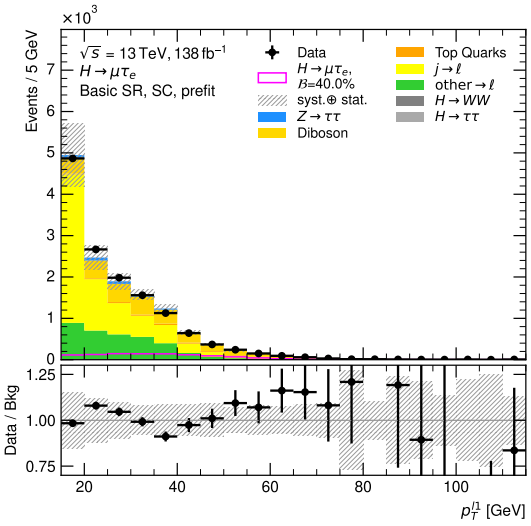



**(e)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_H)$, Basic SR, SC requirement

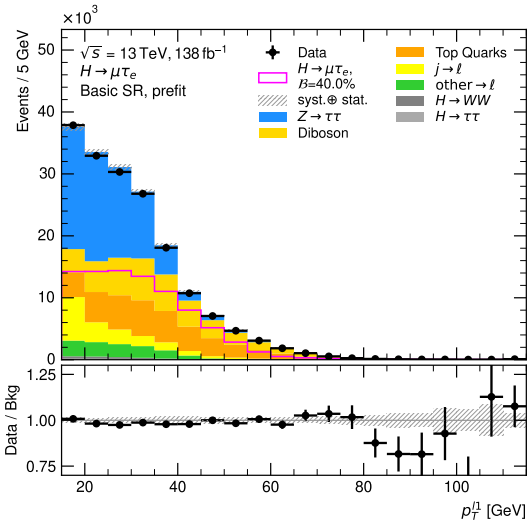



**(f)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_H)$, Basic SR, OC requirement

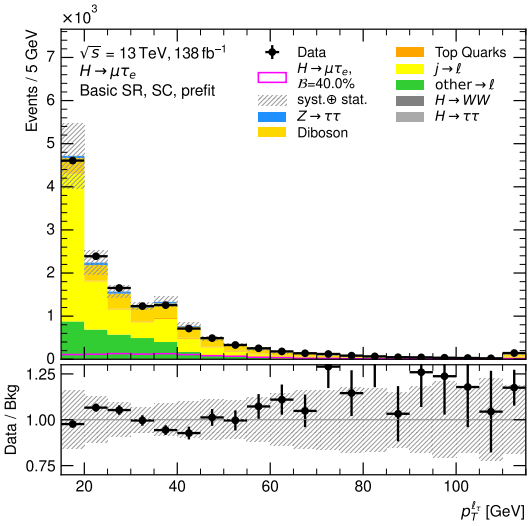
**Figure A.7:** Distributions in the $\mu\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate.



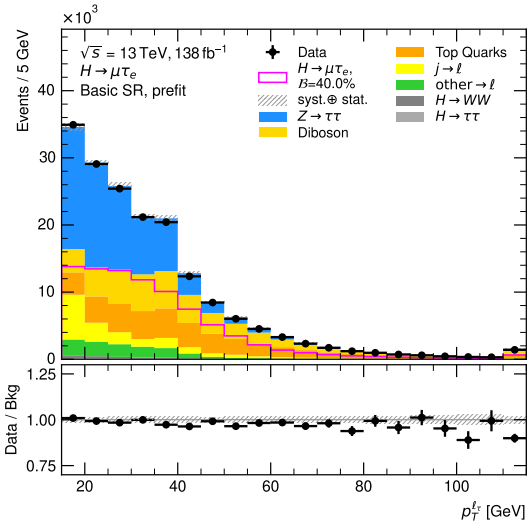

**(a)** $p_{\mathrm{T}}(\ell_1)$, Basic SR, SC requirement



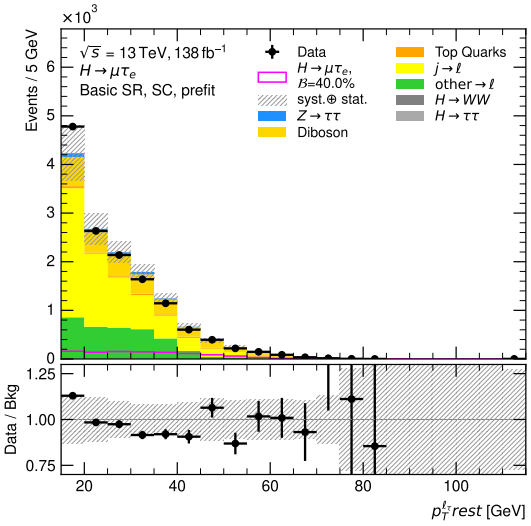

**(b)** $p_{\mathrm{T}}(\ell_1)$, Basic SR, OC requirement



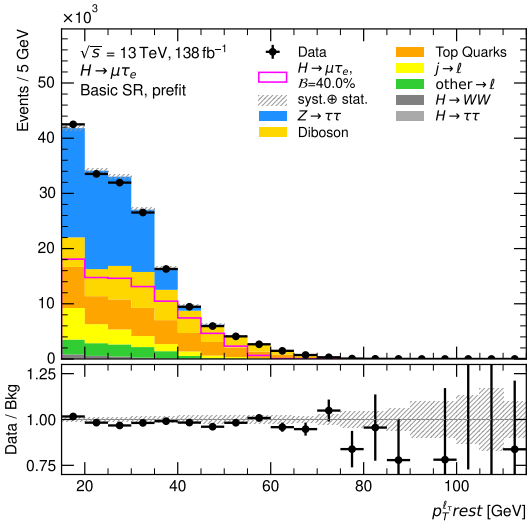

**(c)** $p_{\mathrm{T}}(\ell_\tau)$, Basic SR, SC requirement




**(d)** $p_{\mathrm{T}}(\ell_\tau)$, Basic SR, OC requirement

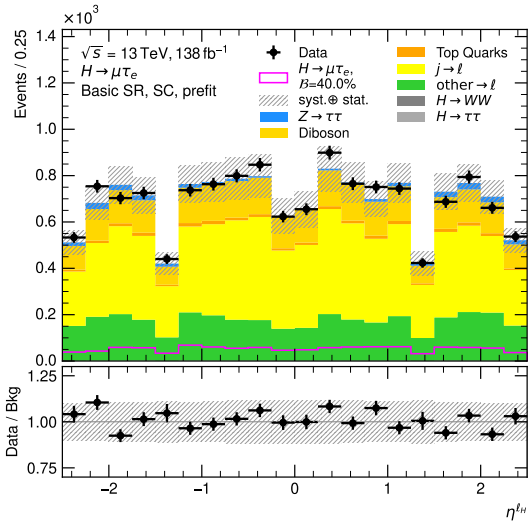



**(e)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_\tau)$, Basic SR, SC requirement

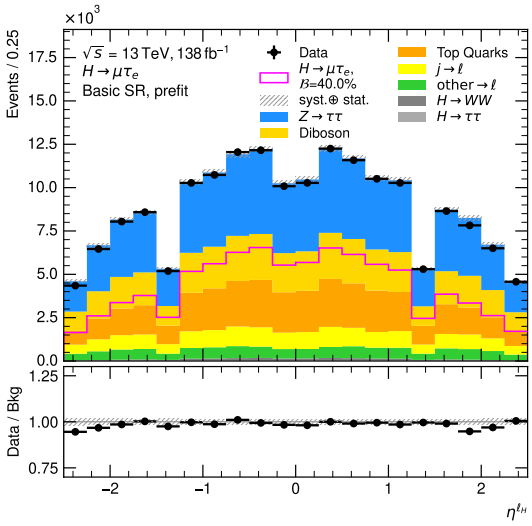



**(f)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_\tau)$, Basic SR, OC requirement

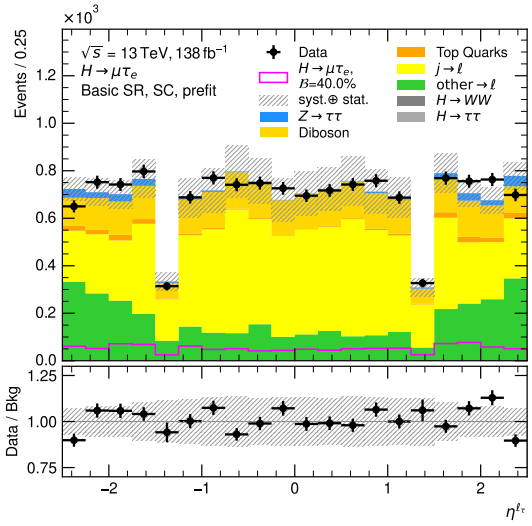
**Figure A.8:** Distributions in the $\mu\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate.



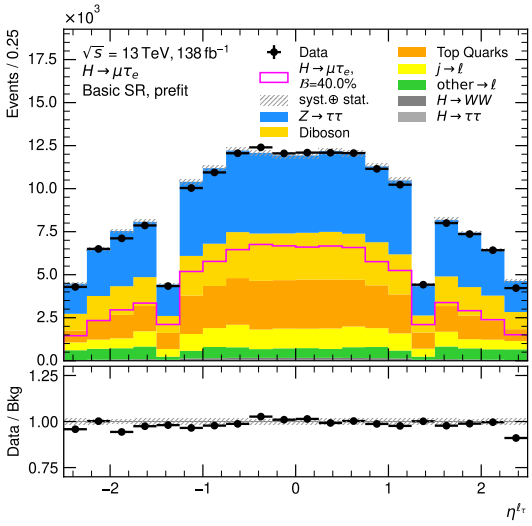

**(a)** $\eta(\ell_H)$, Basic SR, SC requirement



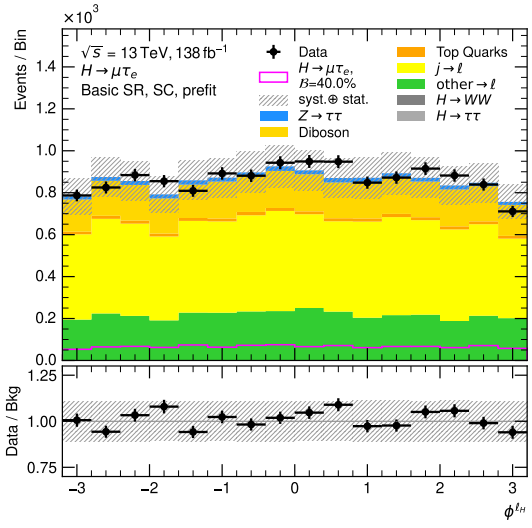

**(b)** $\eta(\ell_H)$, Basic SR, OC requirement



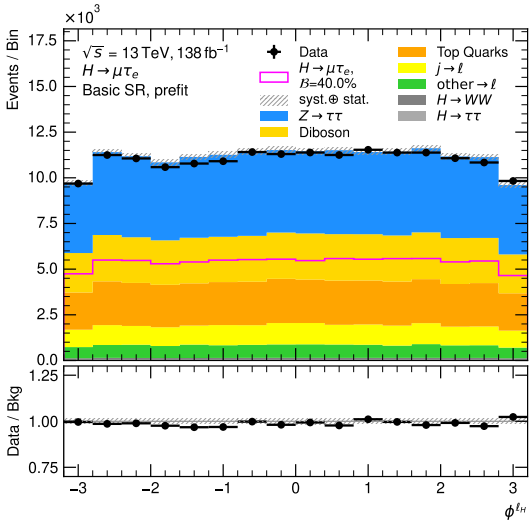

**(c)** $\eta(\ell_\tau)$, Basic SR, SC requirement




**(d)** $\eta(\ell_\tau)$, Basic SR, OC requirement

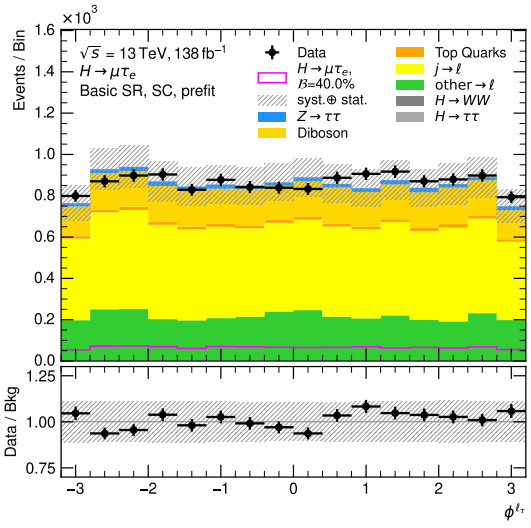



**(e)** $\phi(\ell_H)$, Basic SR, SC requirement

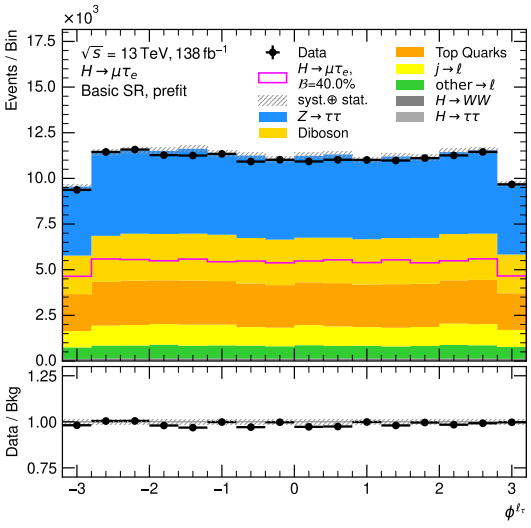



**(f)** $\phi(\ell_H)$, Basic SR, OC requirement

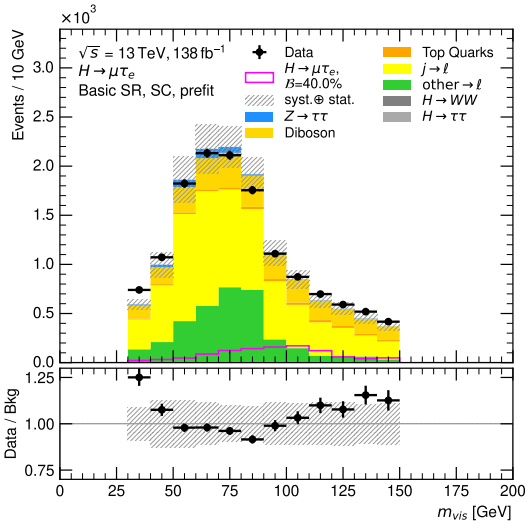
**Figure A.9:** Distributions in the $\mu\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate.



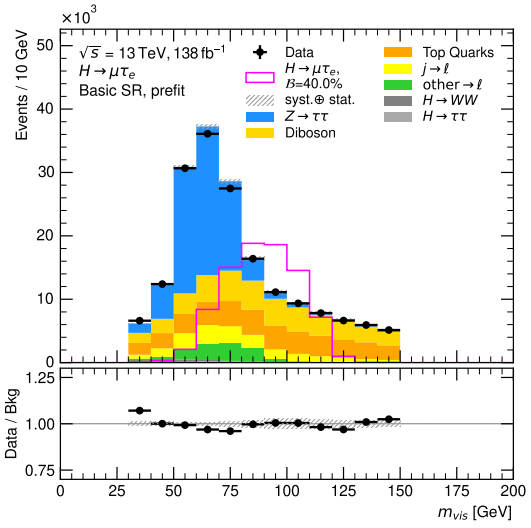

**(a)** $\phi(\ell_\tau)$, Basic SR, SC requirement



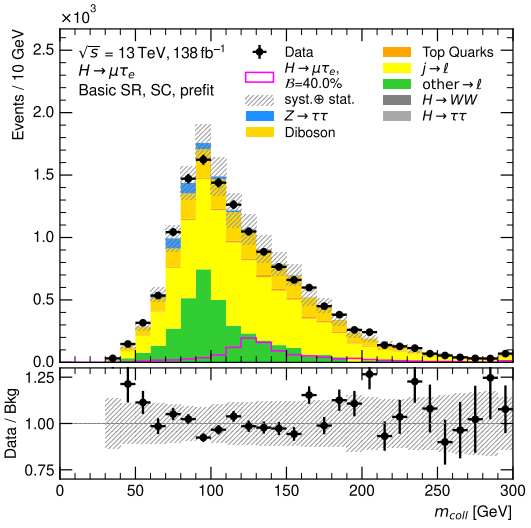

**(b)** $\phi(\ell_\tau)$, Basic SR, OC requirement



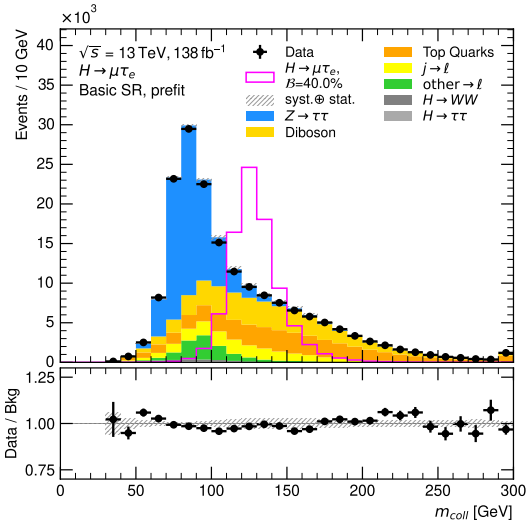

**(c)** $m_{\ell\ell}$, Basic SR, SC requirement




**(d)** $m_{\ell\ell}$, Basic SR, OC requirement

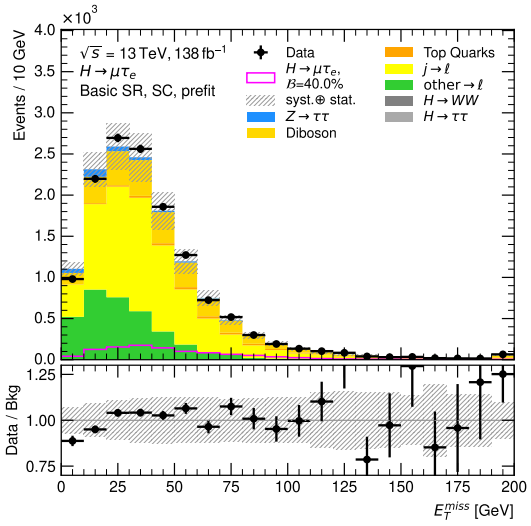



**(e)** $m_{\text{coll}}$, Basic SR, SC requirement

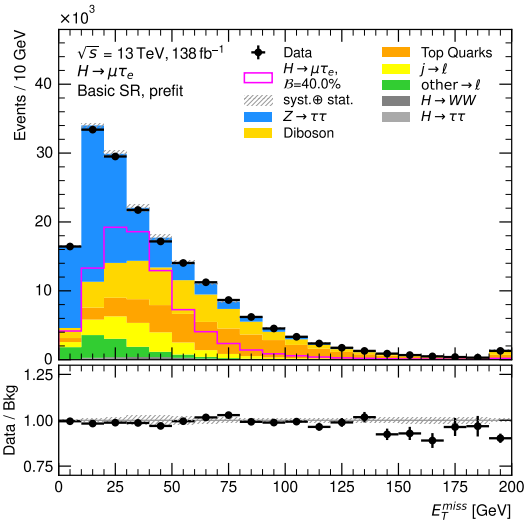



**(f)** $m_{\text{coll}}$, Basic SR, OC requirement

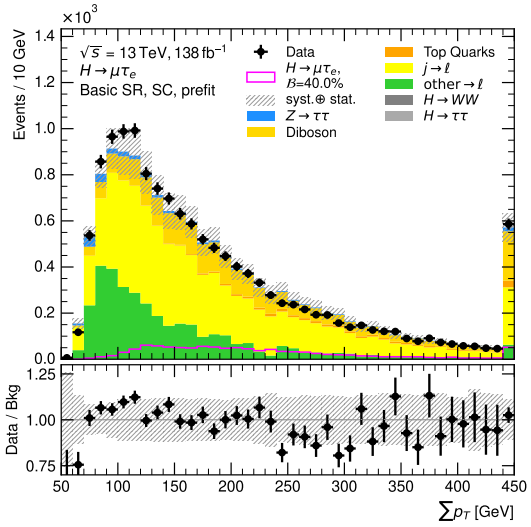
**Figure A.10:** Distributions in the $\mu\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate.



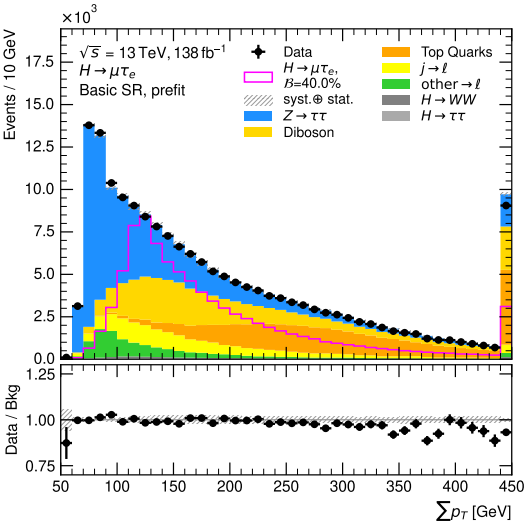

**(a)** $E_{\mathrm{T}}^{\mathrm{miss}}$, Basic SR, SC requirement



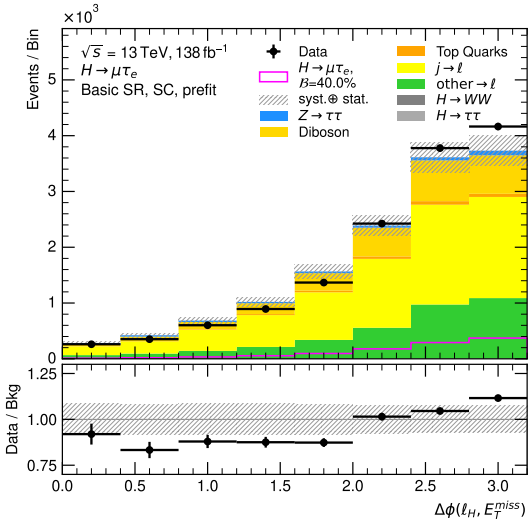

**(b)** $E_{\mathrm{T}}^{\mathrm{miss}}$, Basic SR, OC requirement



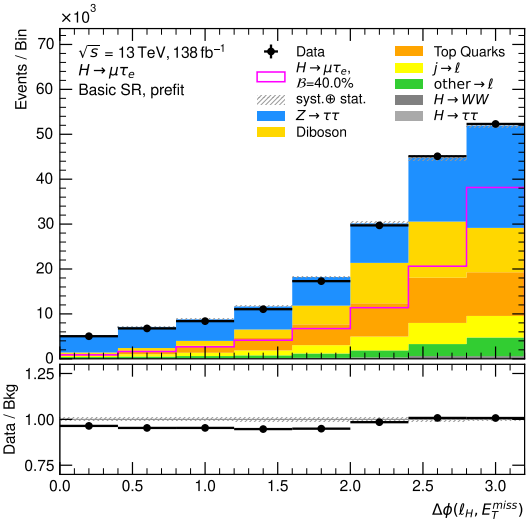

**(c)** $\sum p_{\mathrm{T}}$, Basic SR, SC requirement




**(d)** $\sum p_{\mathrm{T}}$, Basic SR, OC requirement

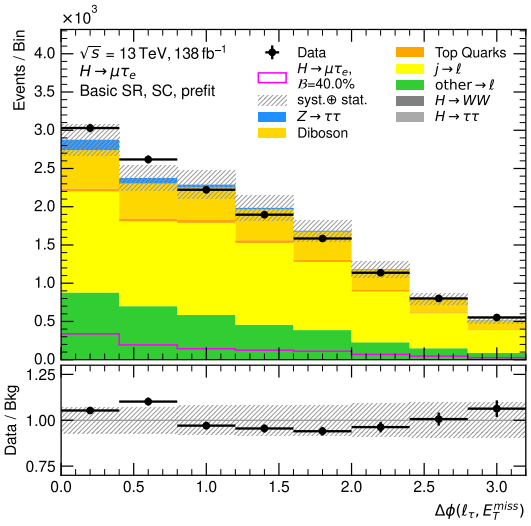



**(e)** $\Delta\phi(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, SC requirement

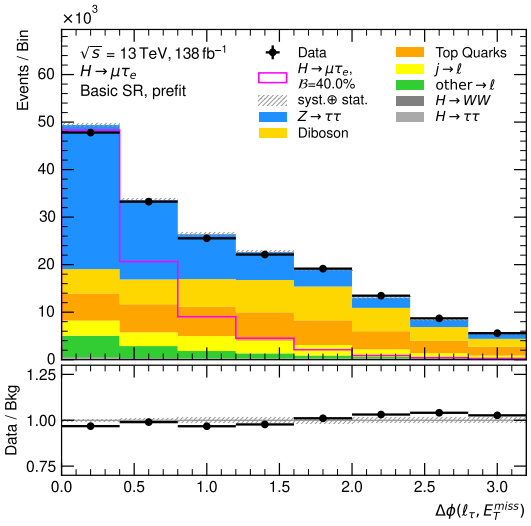



**(f)** $\Delta\phi(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, OC requirement

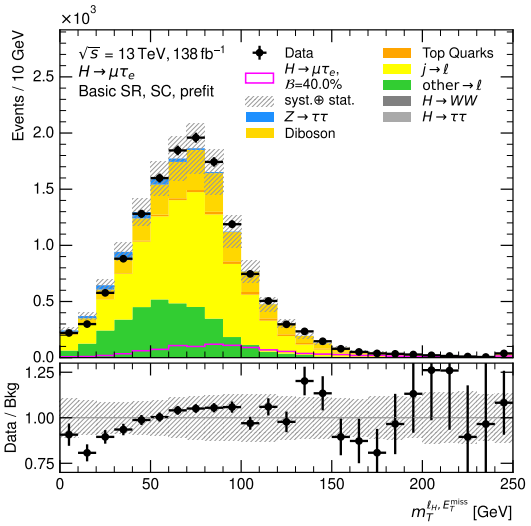
**Figure A.11:** Distributions in the $\mu\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate.



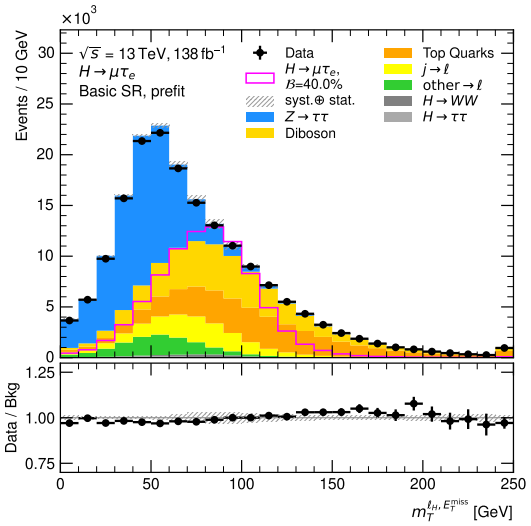

**(a)** $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, SC requirement



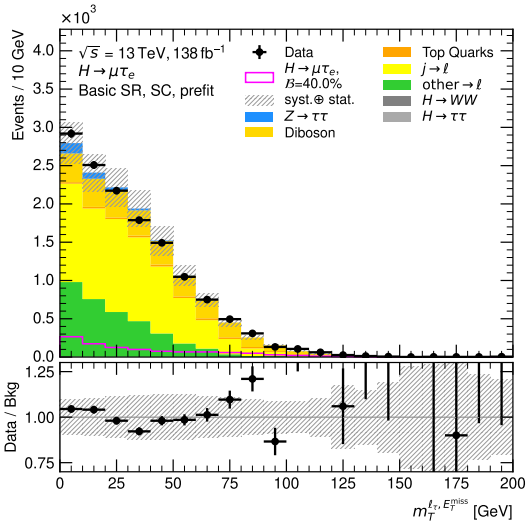

**(b)** $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, OC requirement



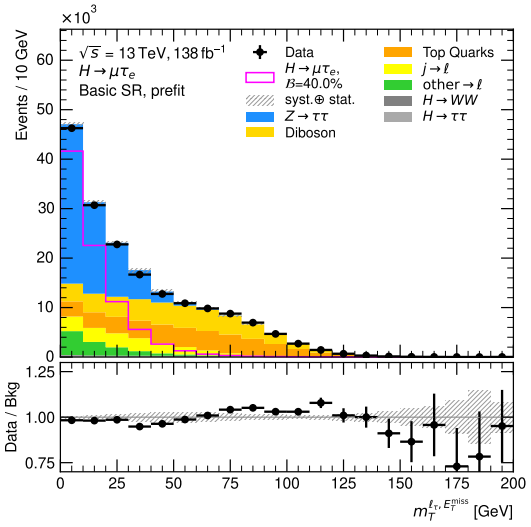

**(c)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, SC requirement




**(d)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, OC requirement




**(e)** $m_T(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, SC requirement

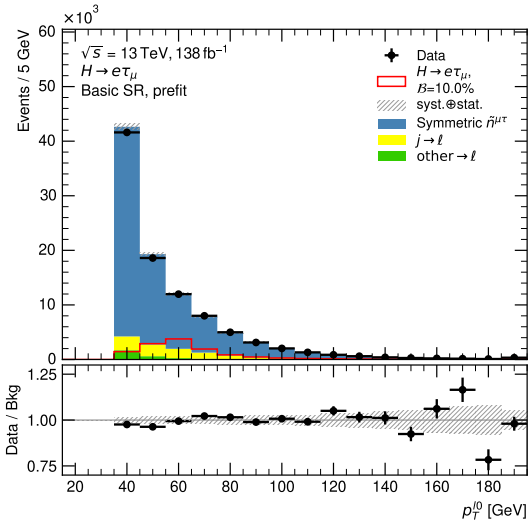



**(f)** $m_T(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, OC requirement

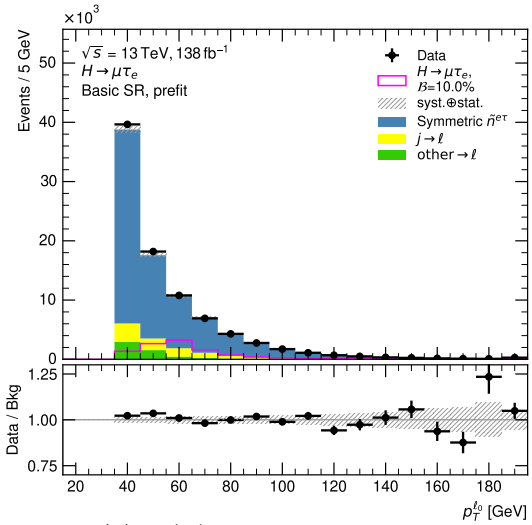
**Figure A.12:** Distributions in the $\mu\tau$-dataset after the *Basic SR Selection* with the same-sign electric charge (SC) requirement (left) and the opposite-sign electric charge (OC) requirement (right). The $e/\mu$-symmetric background components are estimated from MC-simulations, the $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate.

# Appendix B

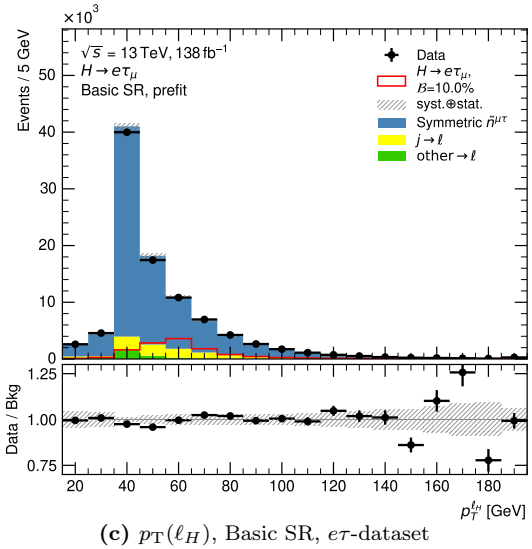
# Validation of the $e/\mu$-Symmetric Background Estimate



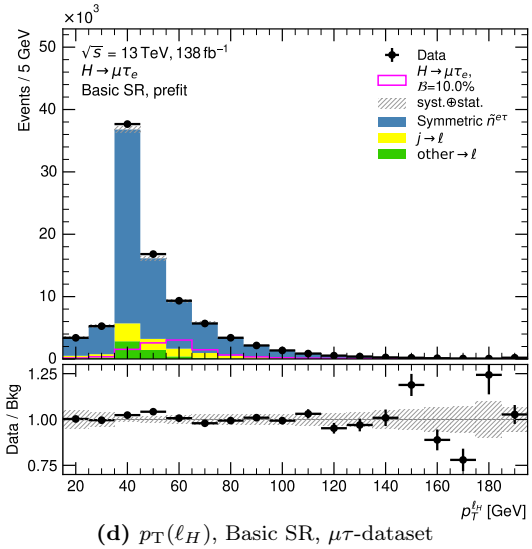

**(a)** $p_{\mathrm{T}}(\ell_0)$, Basic SR, $e\tau$-dataset



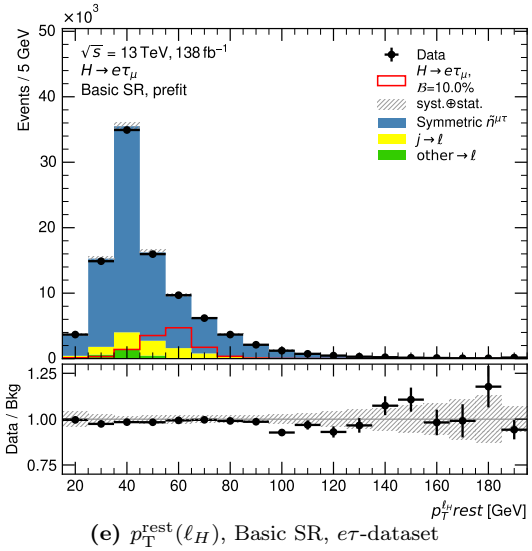

**(b)** $p_{\mathrm{T}}(\ell_0)$, Basic SR, $\mu\tau$-dataset



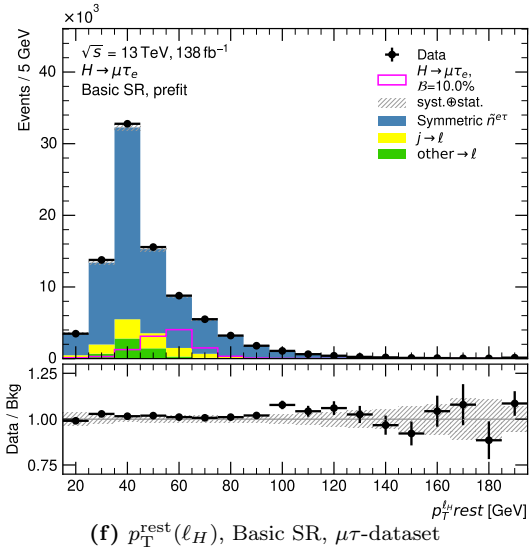

**(c)** $p_{\mathrm{T}}(\ell_H)$, Basic SR, $e\tau$-dataset




**(d)** $p_{\mathrm{T}}(\ell_H)$, Basic SR, $\mu\tau$-dataset

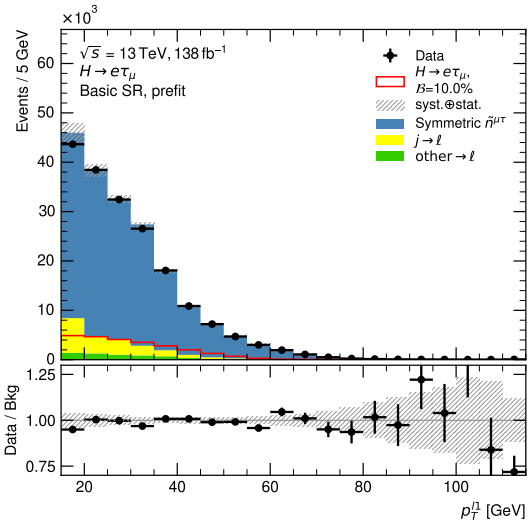



**(e)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_H)$, Basic SR, $e\tau$-dataset

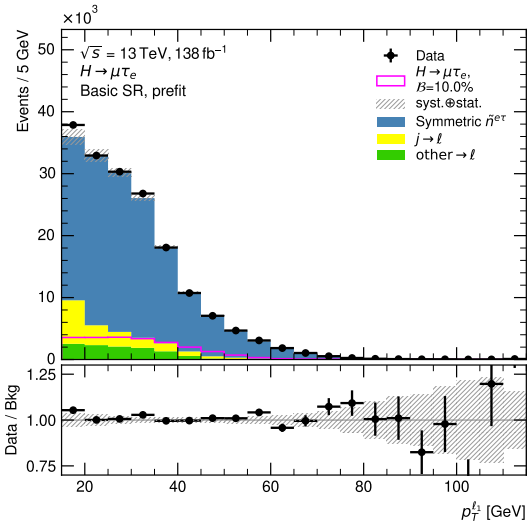



**(f)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_H)$, Basic SR, $\mu\tau$-dataset

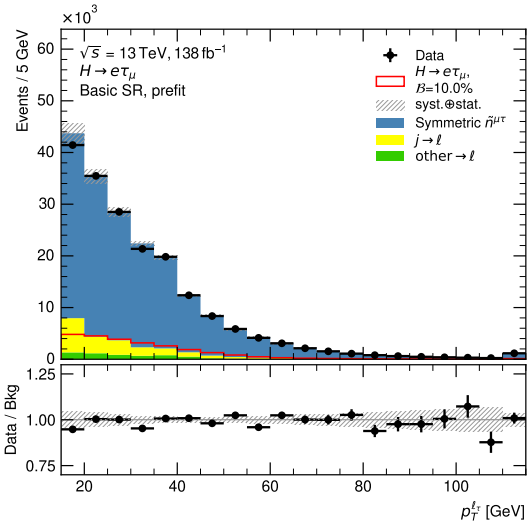
**Figure B.1:** Kinematic distributions with the $e/\mu$-symmetric background contribution estimated with the *Symmetry Method* for the $e\tau$-dataset (left) and the $\mu\tau$-dataset (right) after the *Basic Selection* SR. The $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical and full systematic uncertainties on the background predictions added in quadrature. The signal prediction assuming a branching ratio of 10 % is overlaid.



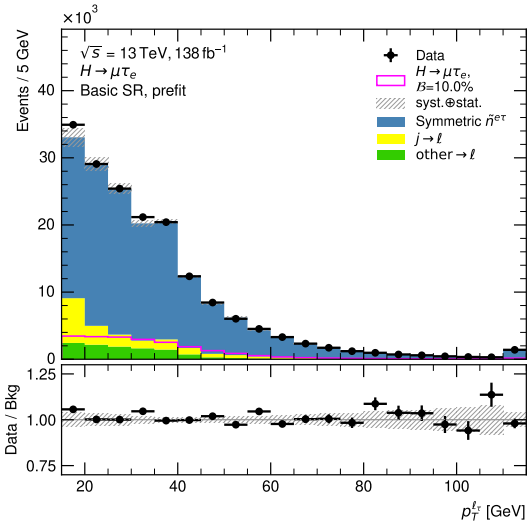

**(a)** $p_{\mathrm{T}}(\ell_1)$, Basic SR, $e\tau$-dataset



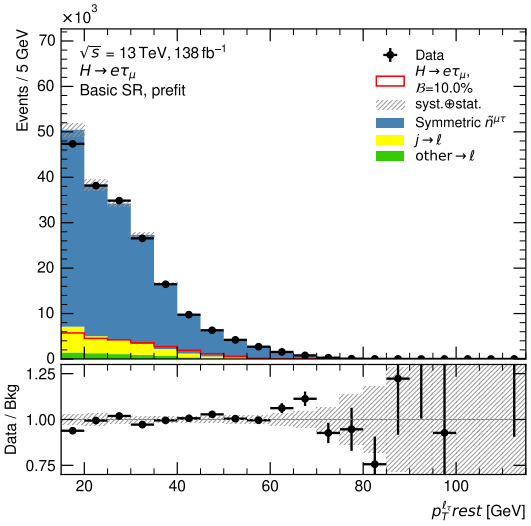

**(b)** $p_{\mathrm{T}}(\ell_1)$, Basic SR, $\mu\tau$-dataset



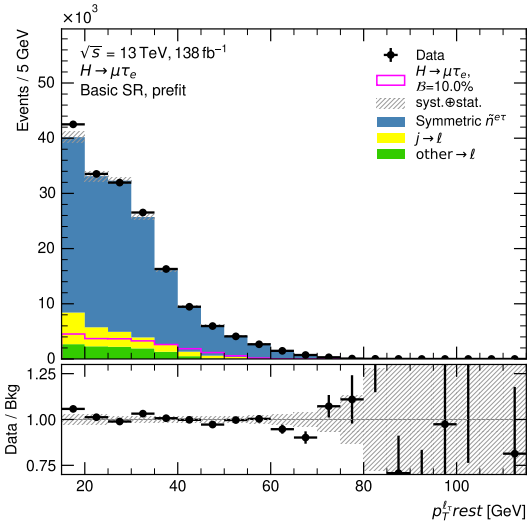

**(c)** $p_{\mathrm{T}}(\ell_\tau)$, Basic SR, $e\tau$-dataset




**(d)** $p_{\mathrm{T}}(\ell_\tau)$, Basic SR, $\mu\tau$-dataset

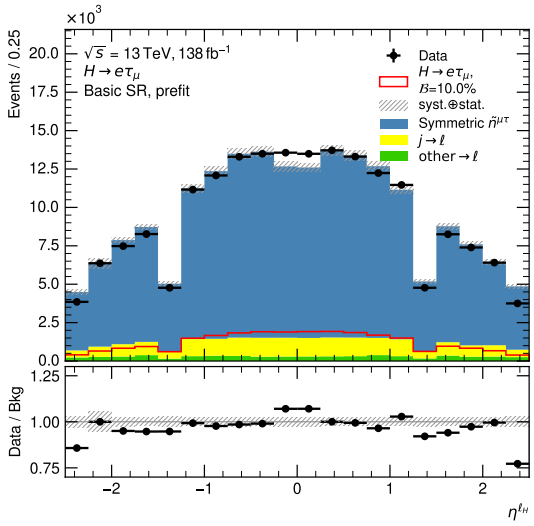



**(e)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_\tau)$, Basic SR, $e\tau$-dataset

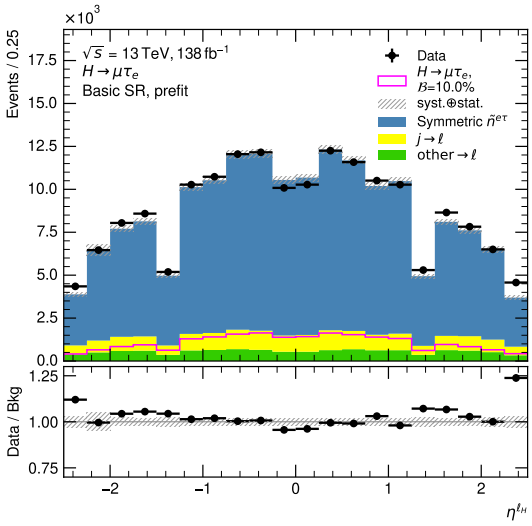



**(f)** $p_{\mathrm{T}}^{\mathrm{rest}}(\ell_\tau)$, Basic SR, $\mu\tau$-dataset

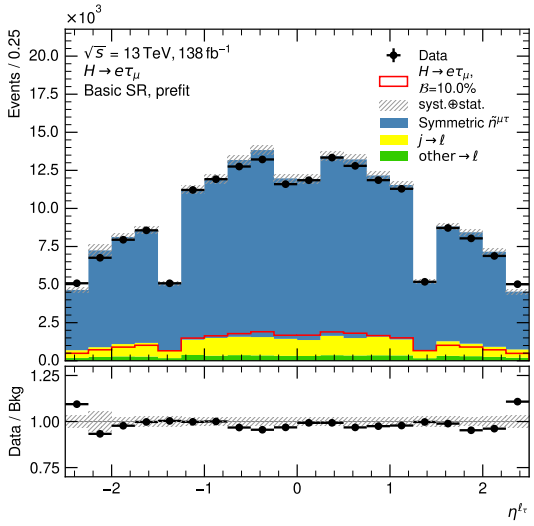
**Figure B.2:** Kinematic distributions with the $e/\mu$-symmetric background contribution estimated with the *Symmetry Method* for the $e\tau$-dataset (left) and the $\mu\tau$-dataset (right) after the *Basic Selection* SR. The $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical and full systematic uncertainties on the background predictions added in quadrature. The signal prediction assuming a branching ratio of 10 % is overlaid.



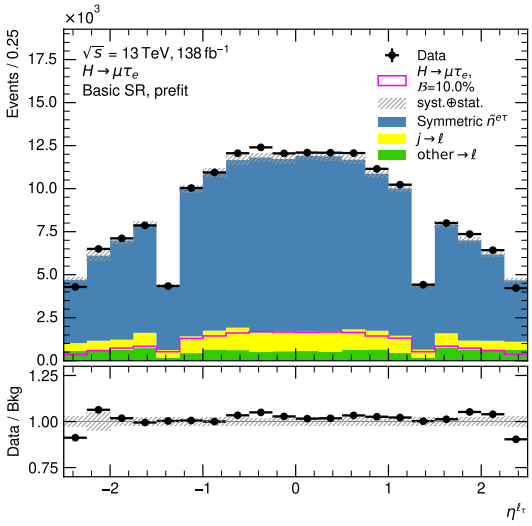

**(a)** $\eta(\ell_H)$, Basic SR, $e\tau$-dataset



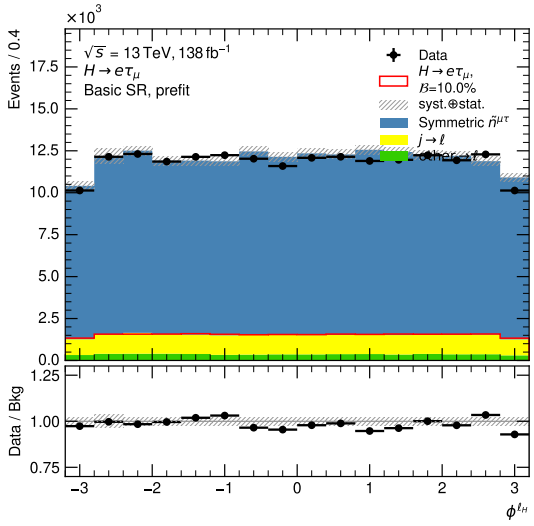

**(b)** $\eta(\ell_H)$, Basic SR, $\mu\tau$-dataset



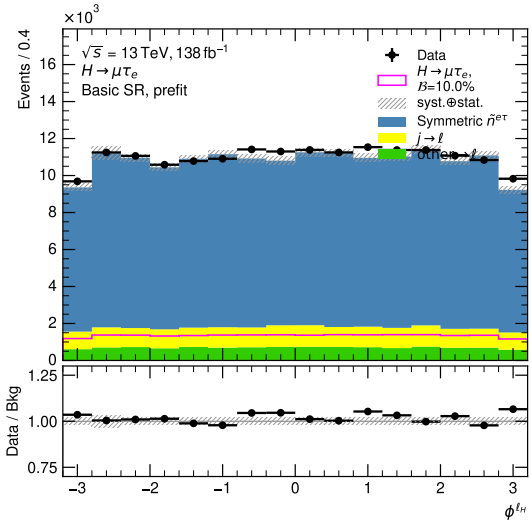

**(c)** $\eta(\ell_\tau)$, Basic SR, $e\tau$-dataset




**(d)** $\eta(\ell_\tau)$, Basic SR, $\mu\tau$-dataset

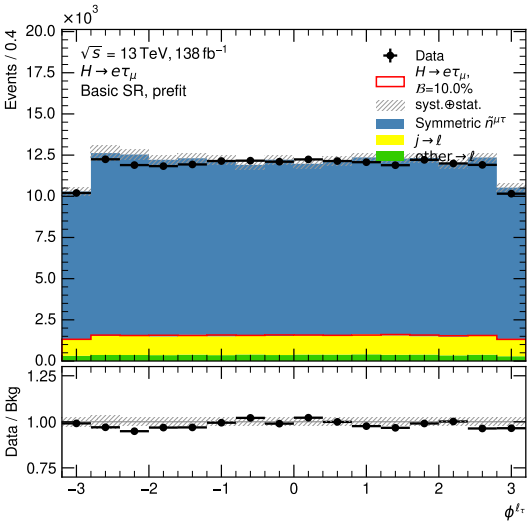



**(e)** $\phi(\ell_H)$, Basic SR, $e\tau$-dataset

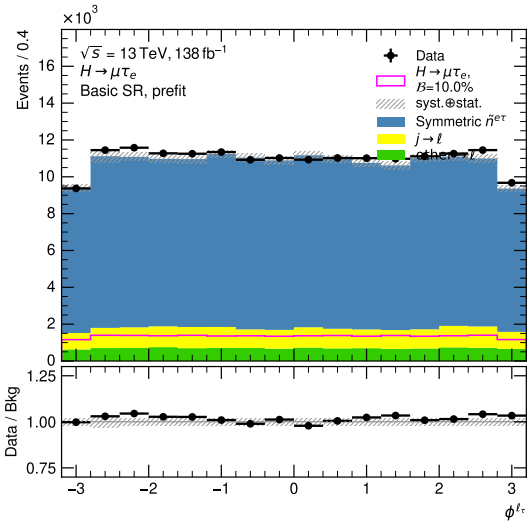



**(f)** $\phi(\ell_H)$, Basic SR, $\mu\tau$-dataset

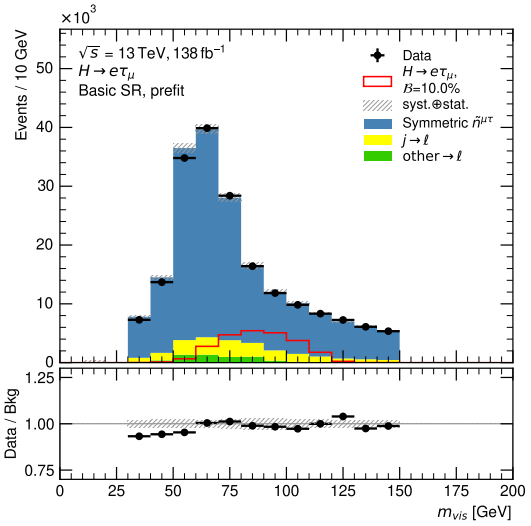
**Figure B.3:** Kinematic distributions with the $e/\mu$-symmetric background contribution estimated with the *Symmetry Method* for the $e\tau$-dataset (left) and the $\mu\tau$-dataset (right) after the *Basic Selection* SR. The $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical and full systematic uncertainties on the background predictions added in quadrature. The signal prediction assuming a branching ratio of 10 % is overlaid.



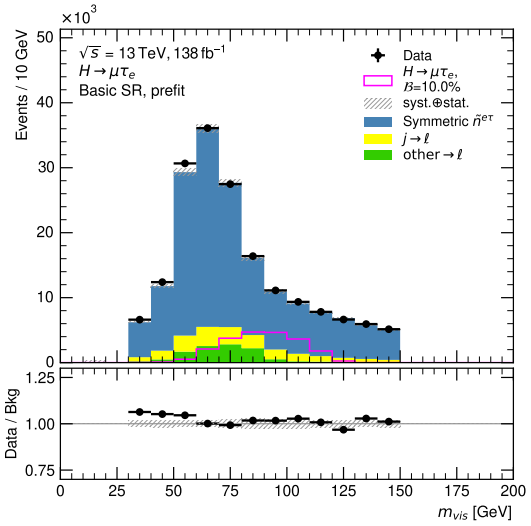

**(a)** $\phi(\ell_\tau)$, Basic SR, $e\tau$-dataset



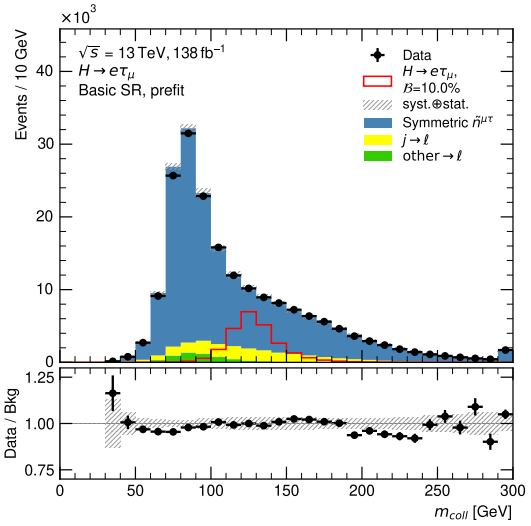

**(b)** $\phi(\ell_\tau)$, Basic SR, $\mu\tau$-dataset



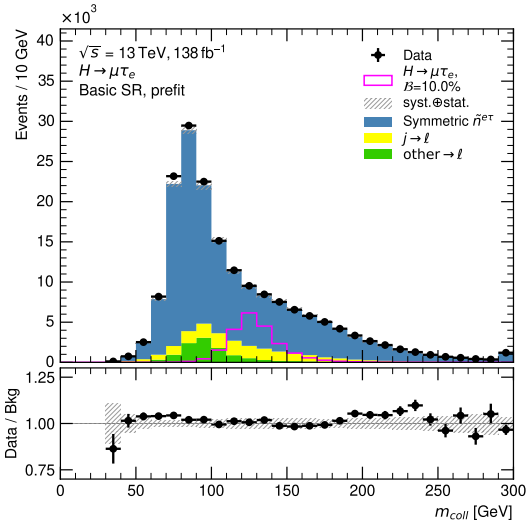

**(c)** $m_{\ell\ell}$, Basic SR, $e\tau$-dataset




**(d)** $m_{\ell\ell}$, Basic SR, $\mu\tau$-dataset

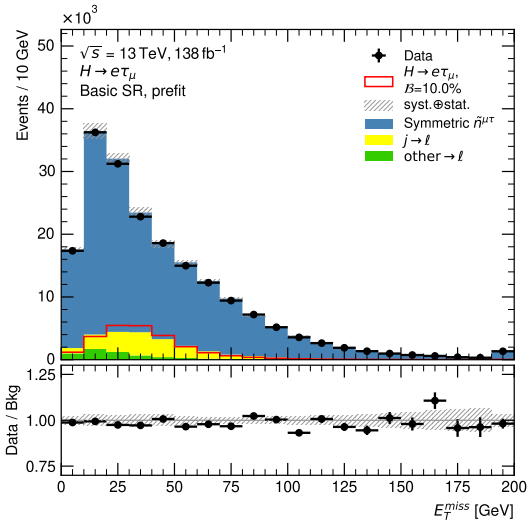



**(e)** $m_{\text{coll}}$, Basic SR, $e\tau$-dataset

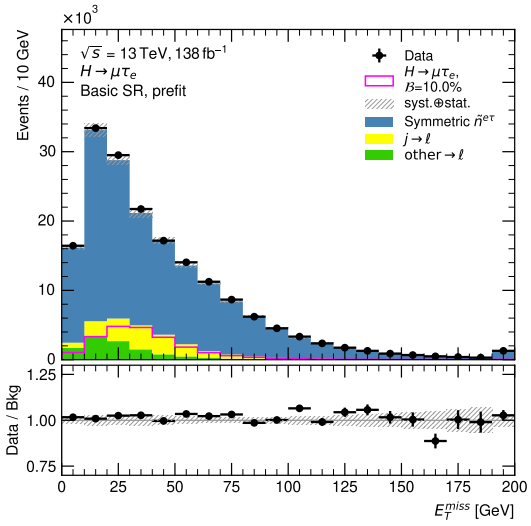



**(f)** $m_{\text{coll}}$, Basic SR, $\mu\tau$-dataset

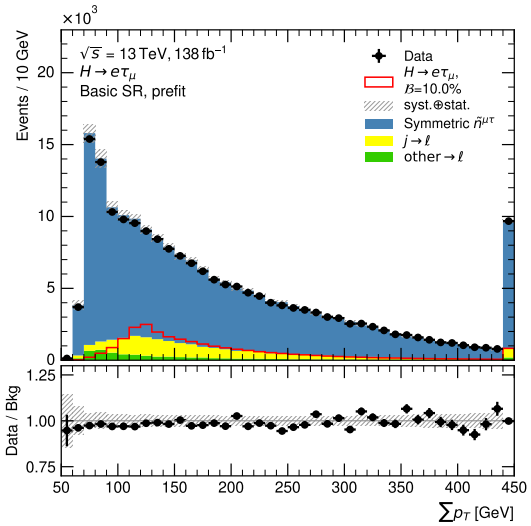
**Figure B.4:** Kinematic distributions with the $e/\mu$-symmetric background contribution estimated with the *Symmetry Method* for the $e\tau$-dataset (left) and the $\mu\tau$-dataset (right) after the *Basic Selection* SR. The $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical and full systematic uncertainties on the background predictions added in quadrature. The signal prediction assuming a branching ratio of 10 % is overlaid.



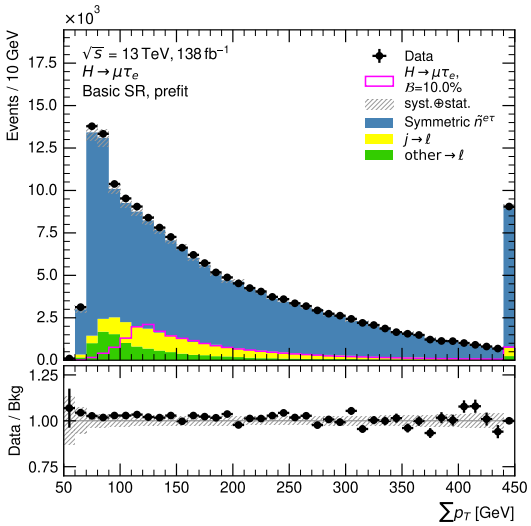

**(a)** $E_{\mathrm{T}}^{\mathrm{miss}}$, Basic SR, $e\tau$-dataset



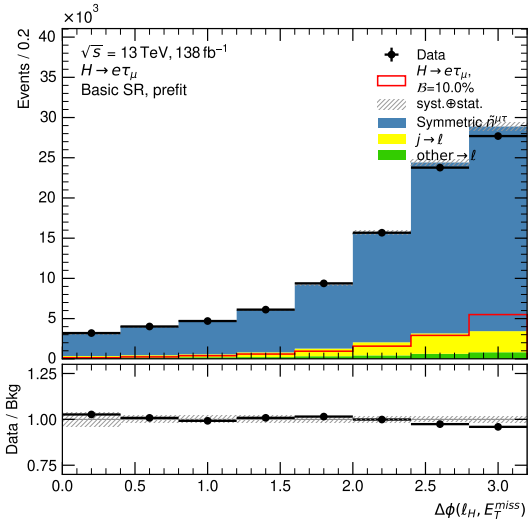

**(b)** $E_{\mathrm{T}}^{\mathrm{miss}}$, Basic SR, $\mu\tau$-dataset



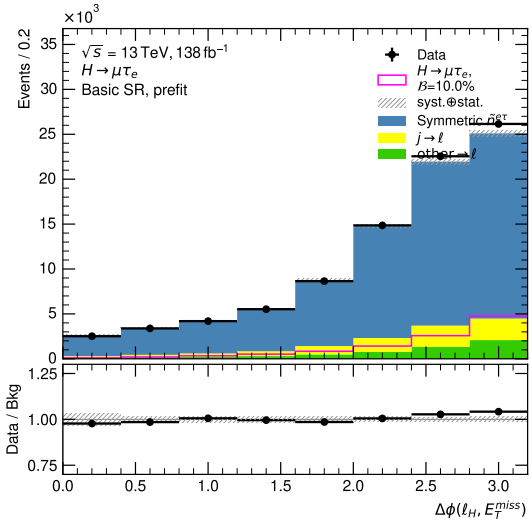

**(c)** $\sum p_{\mathrm{T}}$, Basic SR, $e\tau$-dataset




**(d)** $\sum p_{\mathrm{T}}$, Basic SR, $\mu\tau$-dataset

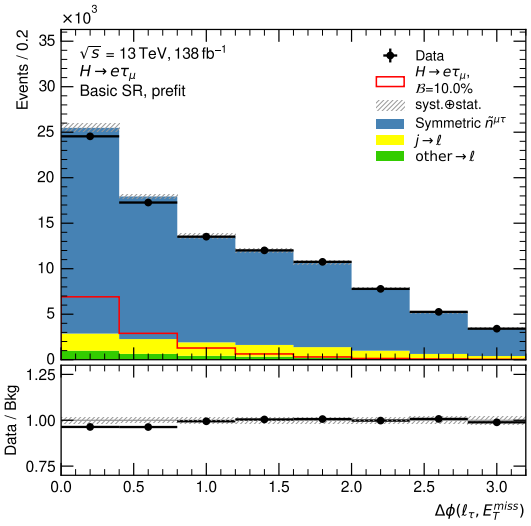



**(e)** $\Delta\phi(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, $e\tau$-dataset

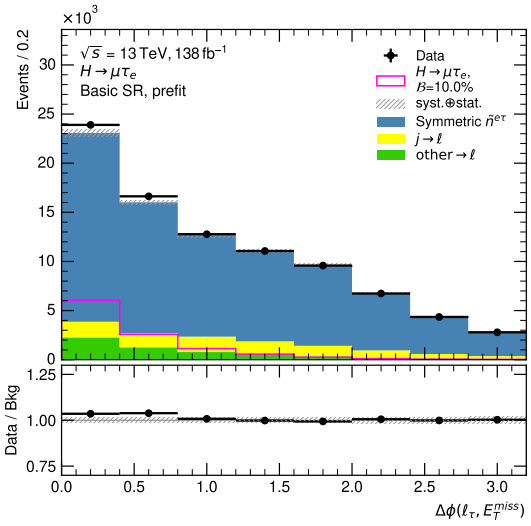



**(f)** $\Delta\phi(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, $\mu\tau$-dataset

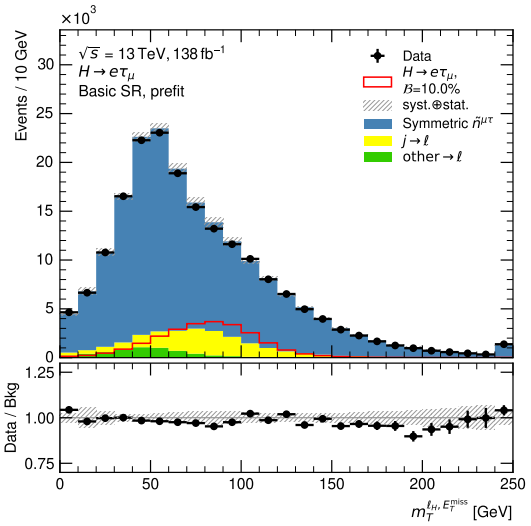
**Figure B.5:** Kinematic distributions with the $e/\mu$-symmetric background contribution estimated with the *Symmetry Method* for the $e\tau$-dataset (left) and the $\mu\tau$-dataset (right) after the *Basic Selection* SR. The $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical and full systematic uncertainties on the background predictions added in quadrature. The signal prediction assuming a branching ratio of 10 % is overlaid.



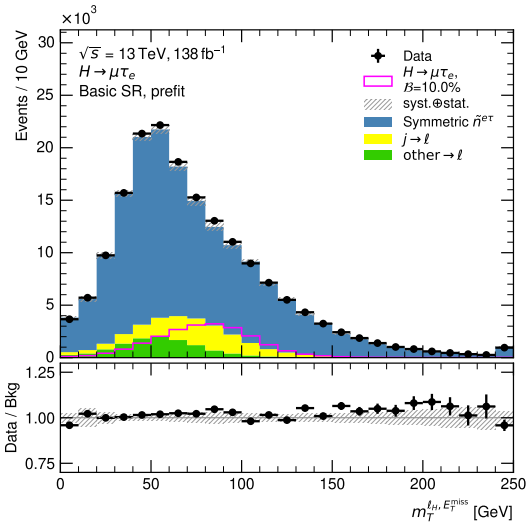

**(a)** $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, $e\tau$-dataset



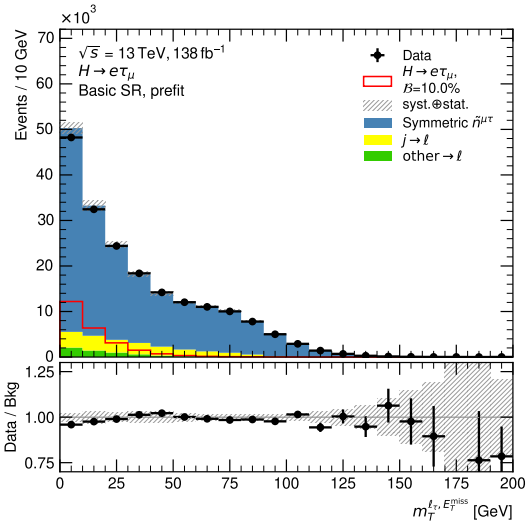

**(b)** $\Delta\phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, $\mu\tau$-dataset



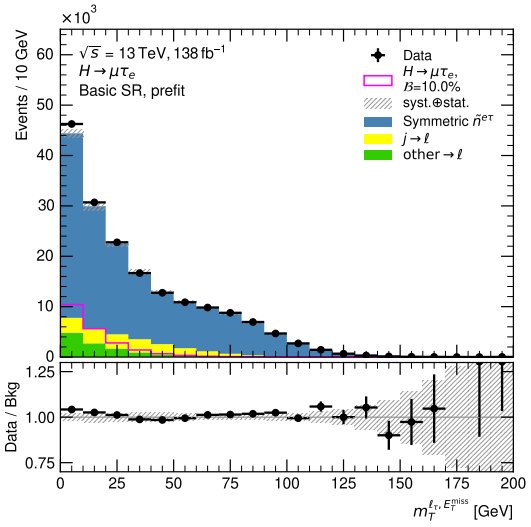

**(c)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, $e\tau$-dataset




**(d)** $m_T(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, $\mu\tau$-dataset




**(e)** $m_T(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, $e\tau$-dataset




**(f)** $m_T(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, Basic SR, $\mu\tau$-dataset

**Figure B.6:** Kinematic distributions with the $e/\mu$-symmetric background contribution estimated with the *Symmetry Method* for the $e\tau$-dataset (left) and the $\mu\tau$-dataset (right) after the *Basic Selection* SR. The $j \to \ell$-fakes are estimated with the *Fake Factor Method* and the remaining other $\to \ell$-fakes are estimated from MC-simulations. The uncertainty band includes statistical and full systematic uncertainties on the background predictions added in quadrature. The signal prediction assuming a branching ratio of 10 % is overlaid.

# Appendix C

# Input Variables to the NNs for the $\mu\tau$-Dataset

The same distributions of the input variables to the *nonVBF NN* and to the *VBF NN*s as shown in Section 8.2.3 but for the $\mu\tau$-dataset instead of the $e\tau$-dataset are given below. The right column in all figures is the same as in Section 8.2.3 since the $e\tau$- and $\mu\tau$-dataset are summed up for these figures. The signal over background ratio in the middle column often shows the opposite behaviour to what is observed for the $e\tau$-dataset in Section 8.2.3 which is intrinsic to the *Symmetry Method*. If the data in the $e\tau$-dataset shows an upwards fluctuation, this propagates to the symmetric background estimate $\tilde{n}^{e\tau}$ for the $\mu\tau$-dataset and leads to an upwards fluctuation of the prediction instead.

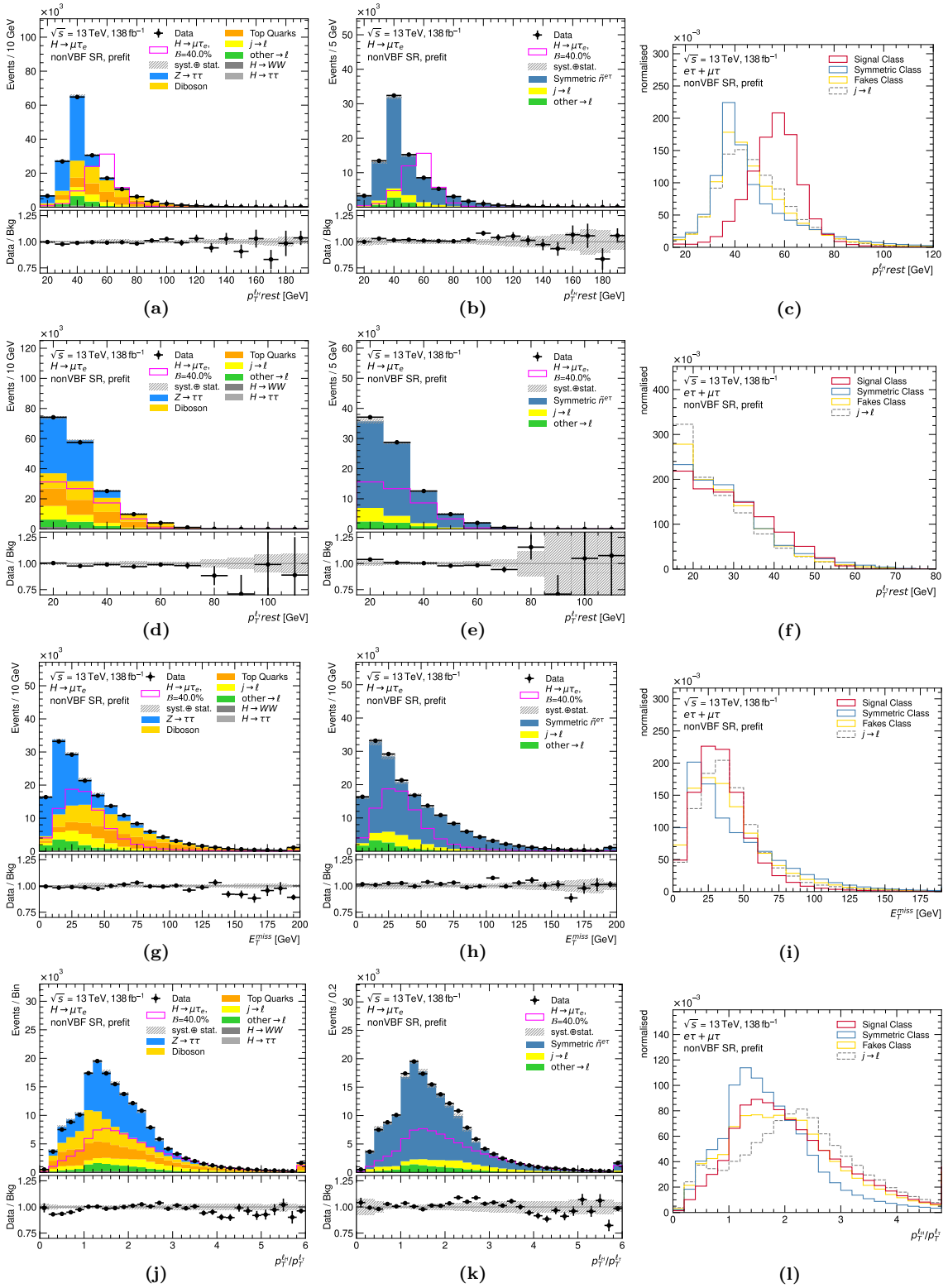



**Figure C.1:** Distributions of input variables to the *nonVBF NN*. The left and middle columns compare the prediction to the data in the $\mu\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datsets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

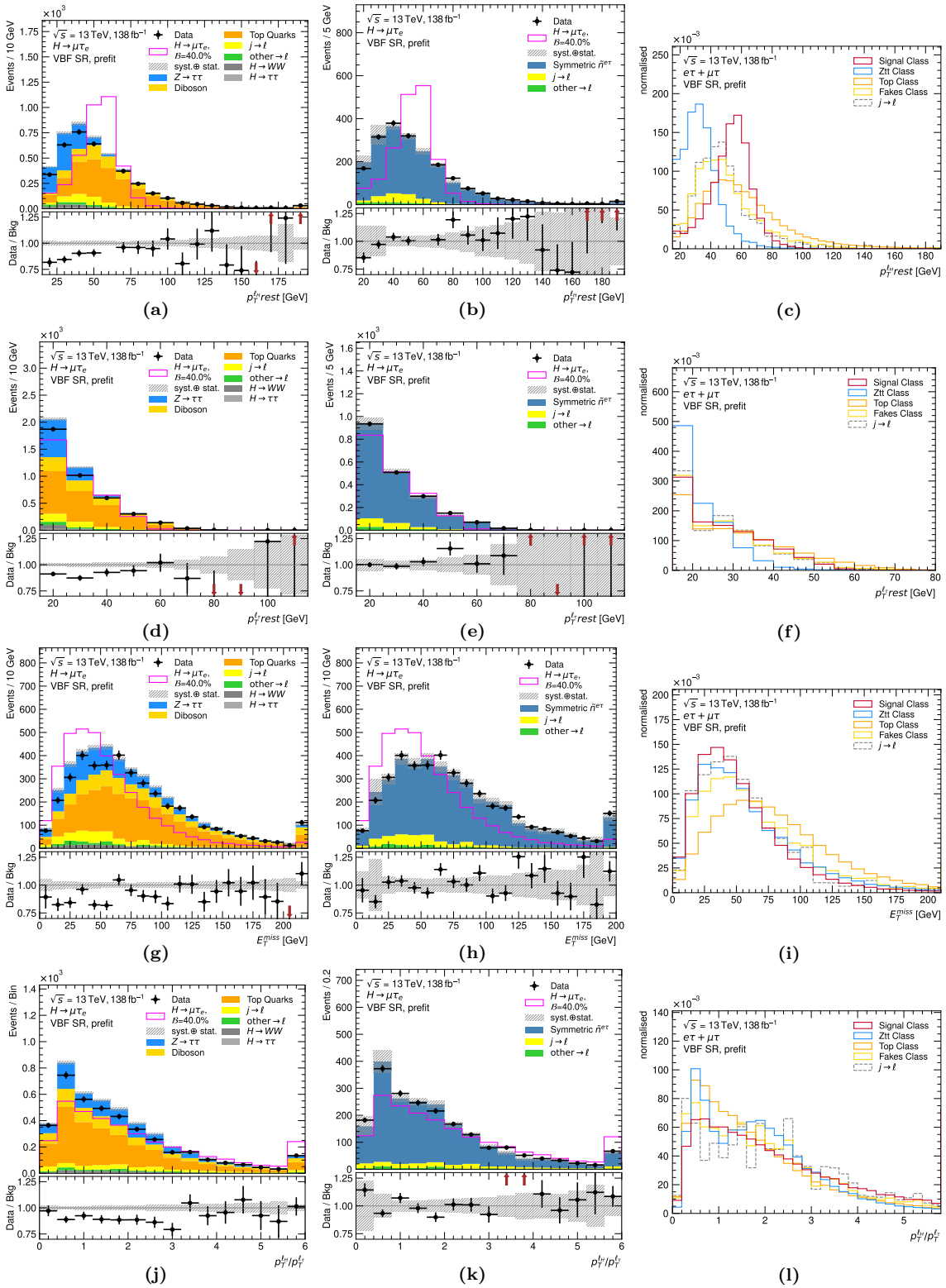



**Figure C.2:** Distributions of input variables to the *VBF NN*s. The left and middle columns compare the prediction to the data in the $\mu\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datsets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

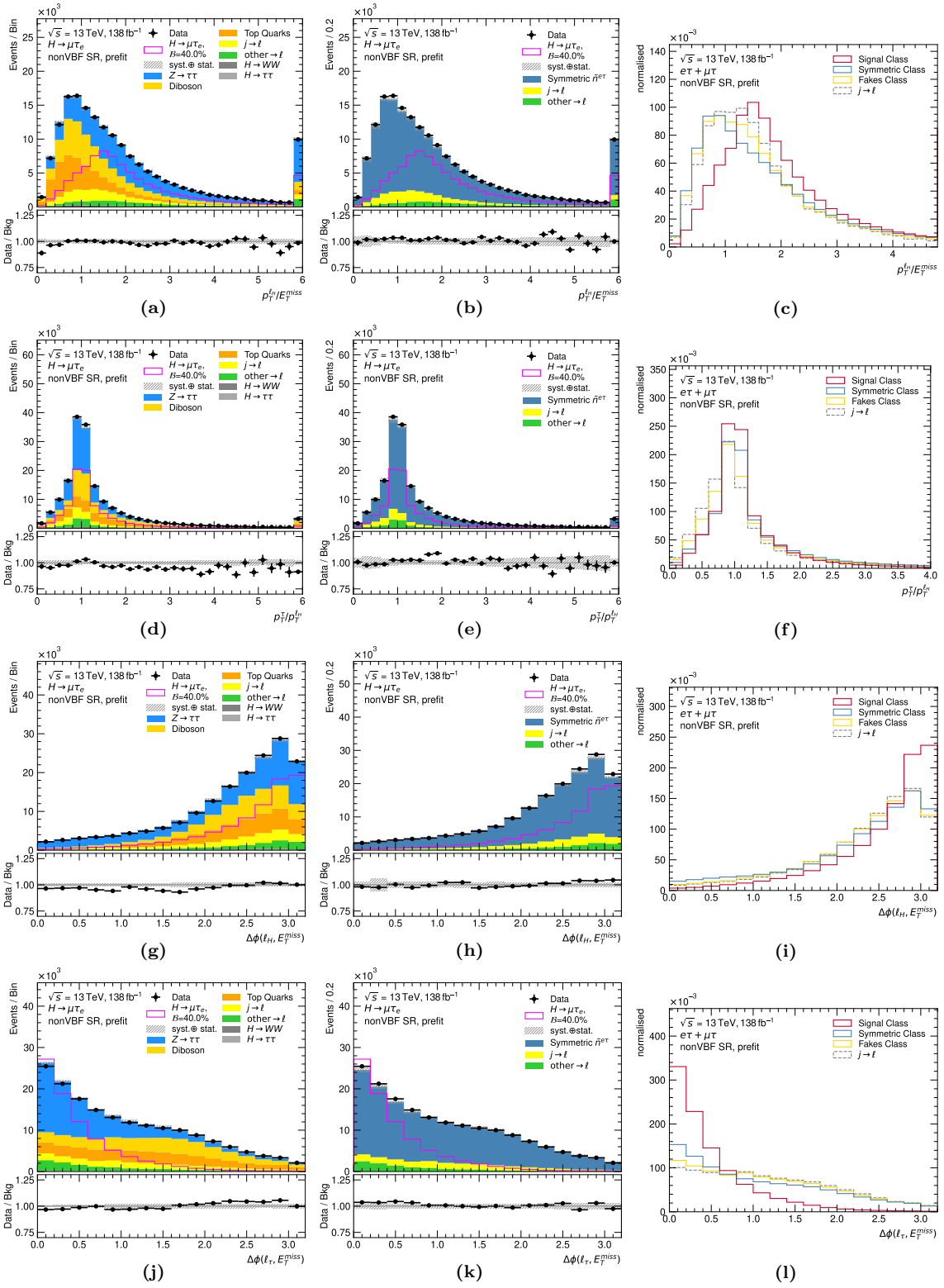



**Figure C.3:** Distributions of input variables to the *nonVBF NN*. The left and middle columns compare the prediction to the data in the $\mu\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datsets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

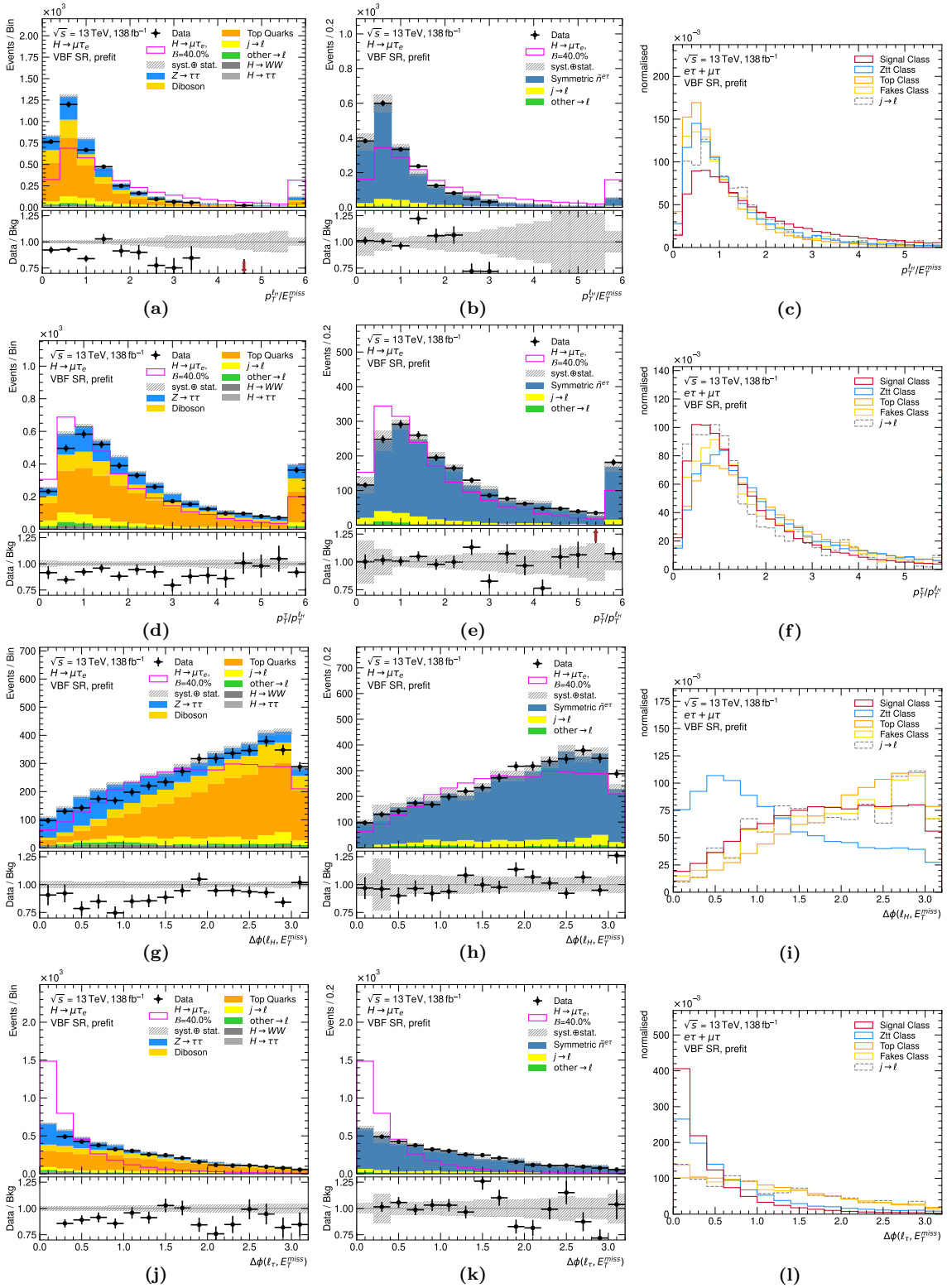



**Figure C.4:** Distributions of input variables to the *VBF NN*s. The left and middle columns compare the prediction to the data in the $\mu\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datsets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

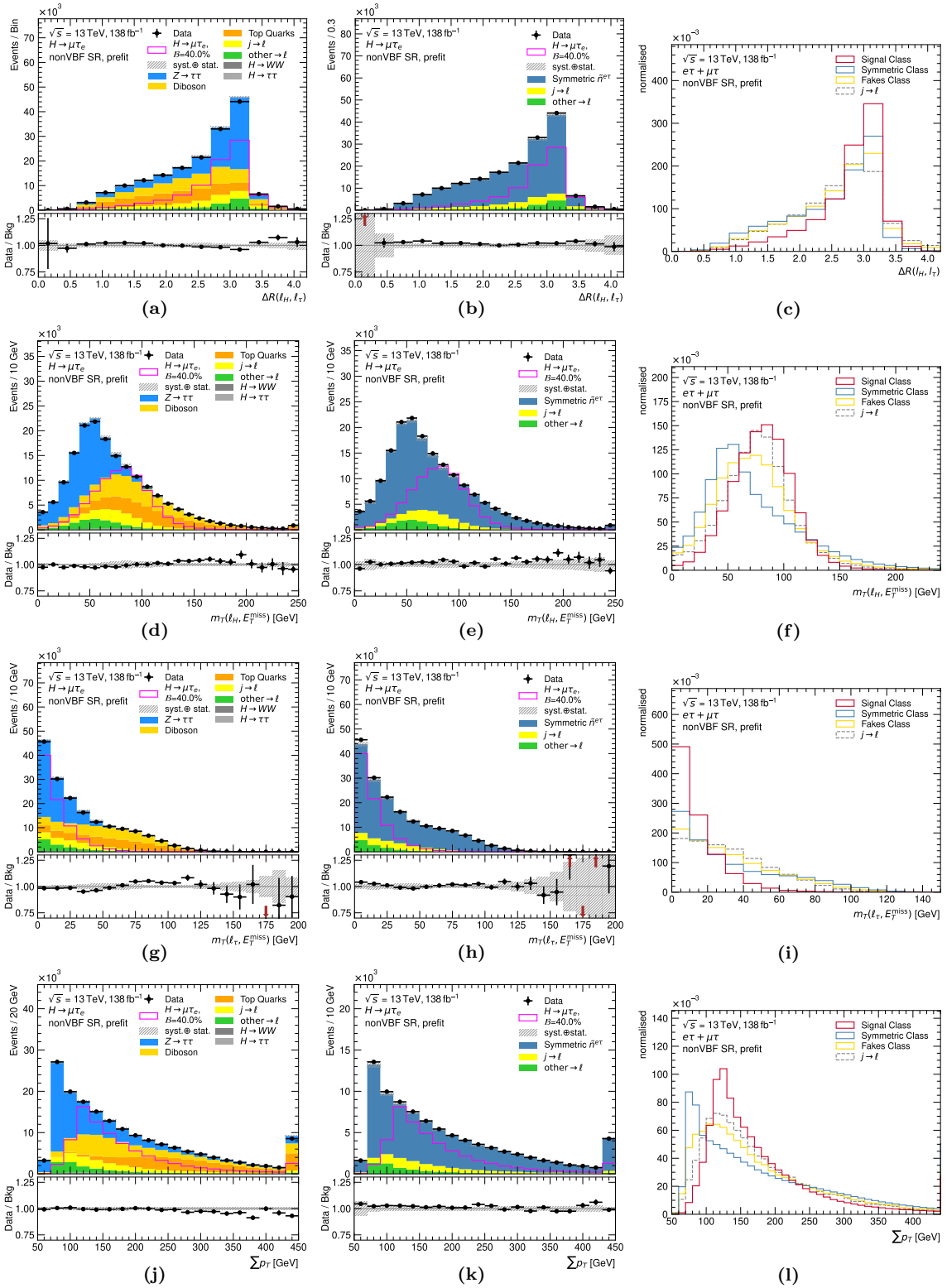



**Figure C.5:** Distributions of input variables to the *nonVBF NN*. The left and middle columns compare the prediction to the data in the $\mu\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datsets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

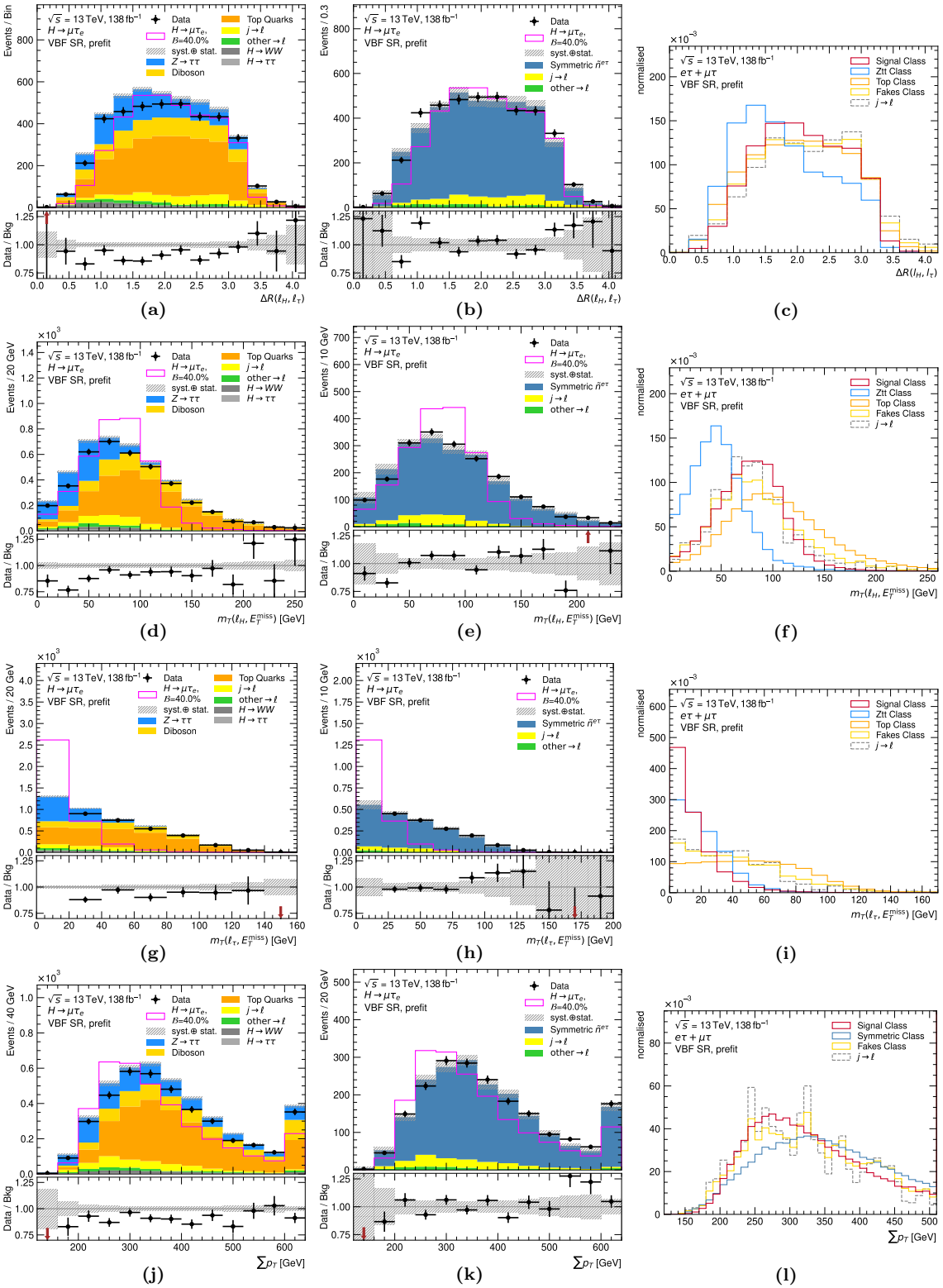



**Figure C.6:** Distributions of input variables to the *VBF NN*s. The left and middle columns compare the prediction to the data in the $\mu\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datsets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

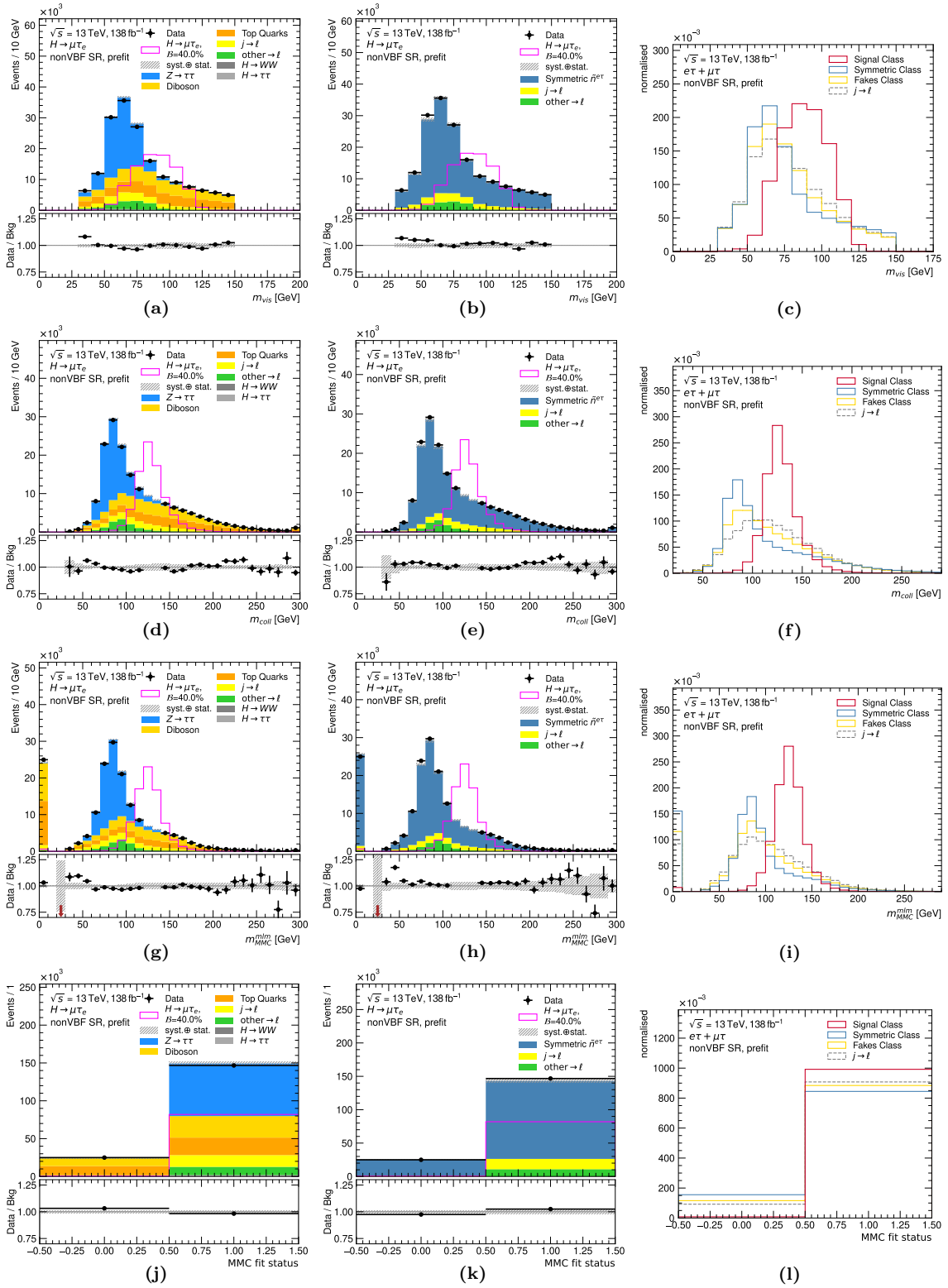



**Figure C.7:** Distributions of input variables to the *nonVBF NN*. The left and middle columns compare the prediction to the data in the $\mu\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datsets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

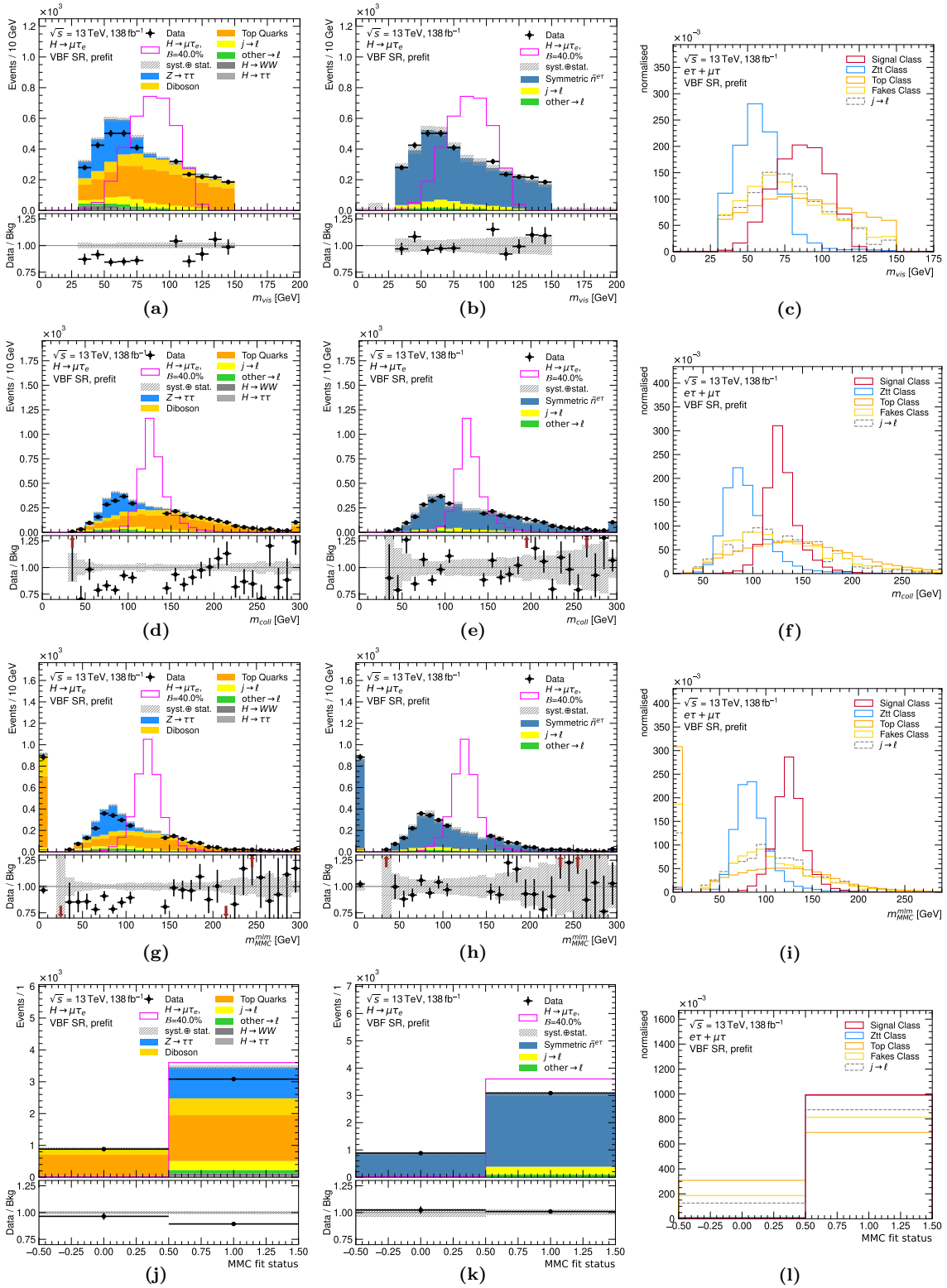



**Figure C.8:** Distributions of input variables to the *VBF NN*s. The left and middle columns compare the prediction to the data in the $\mu\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datsets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

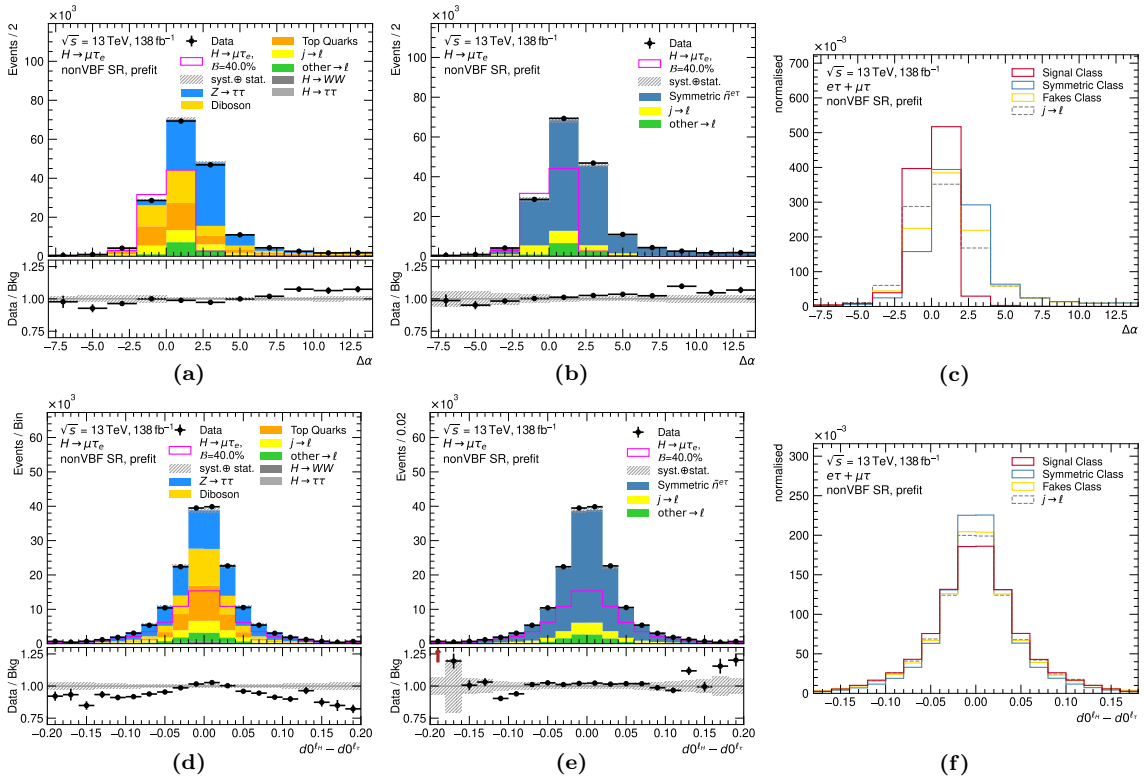



**Figure C.9:** Distributions of input variables to the *nonVBF NN*. The left and middle columns compare the prediction to the data in the $\mu\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datsets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

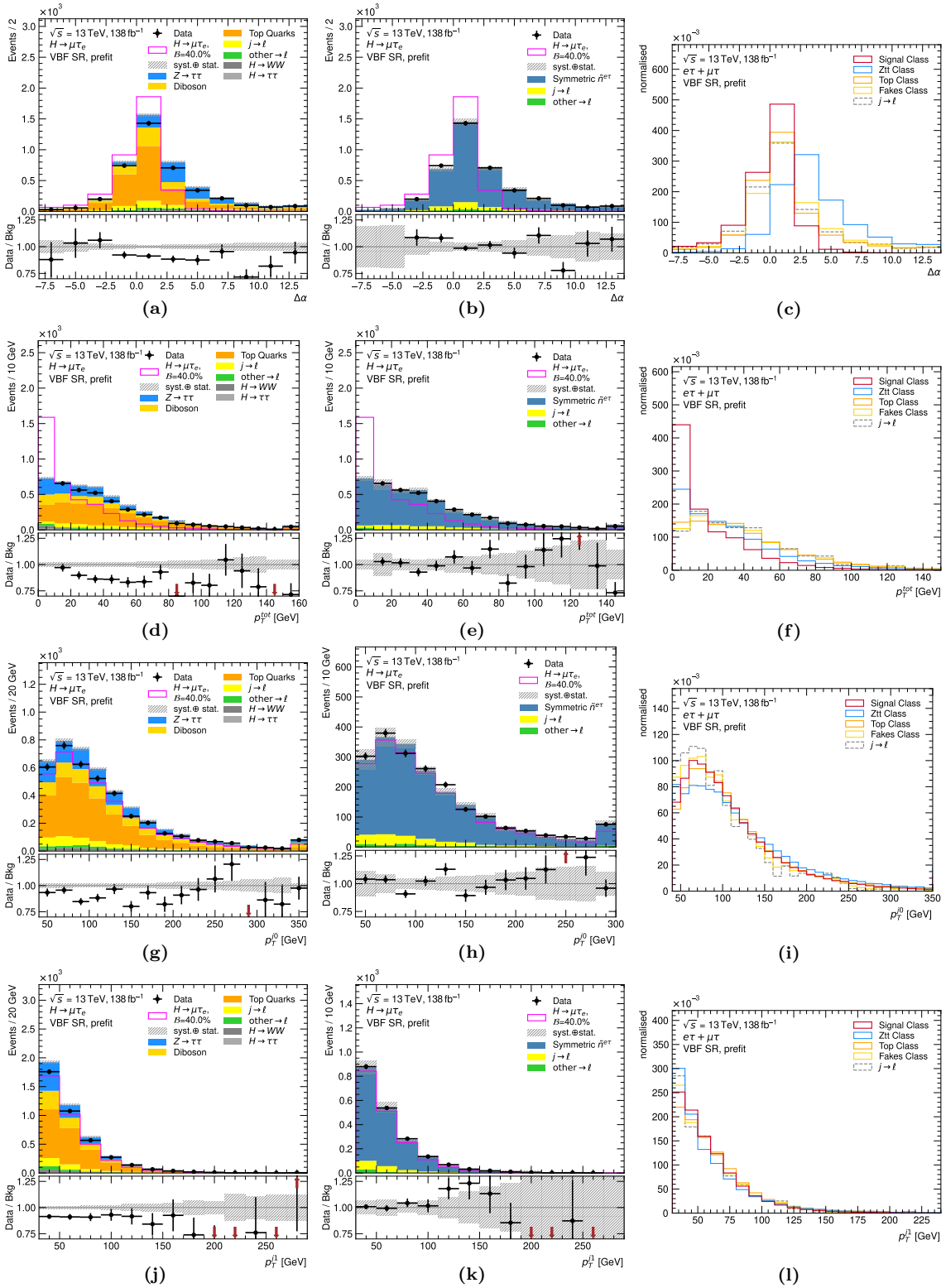



**Figure C.10:** Distributions of input variables to the *VBF NN*s. The left and middle columns compare the prediction to the data in the $\mu\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datsets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

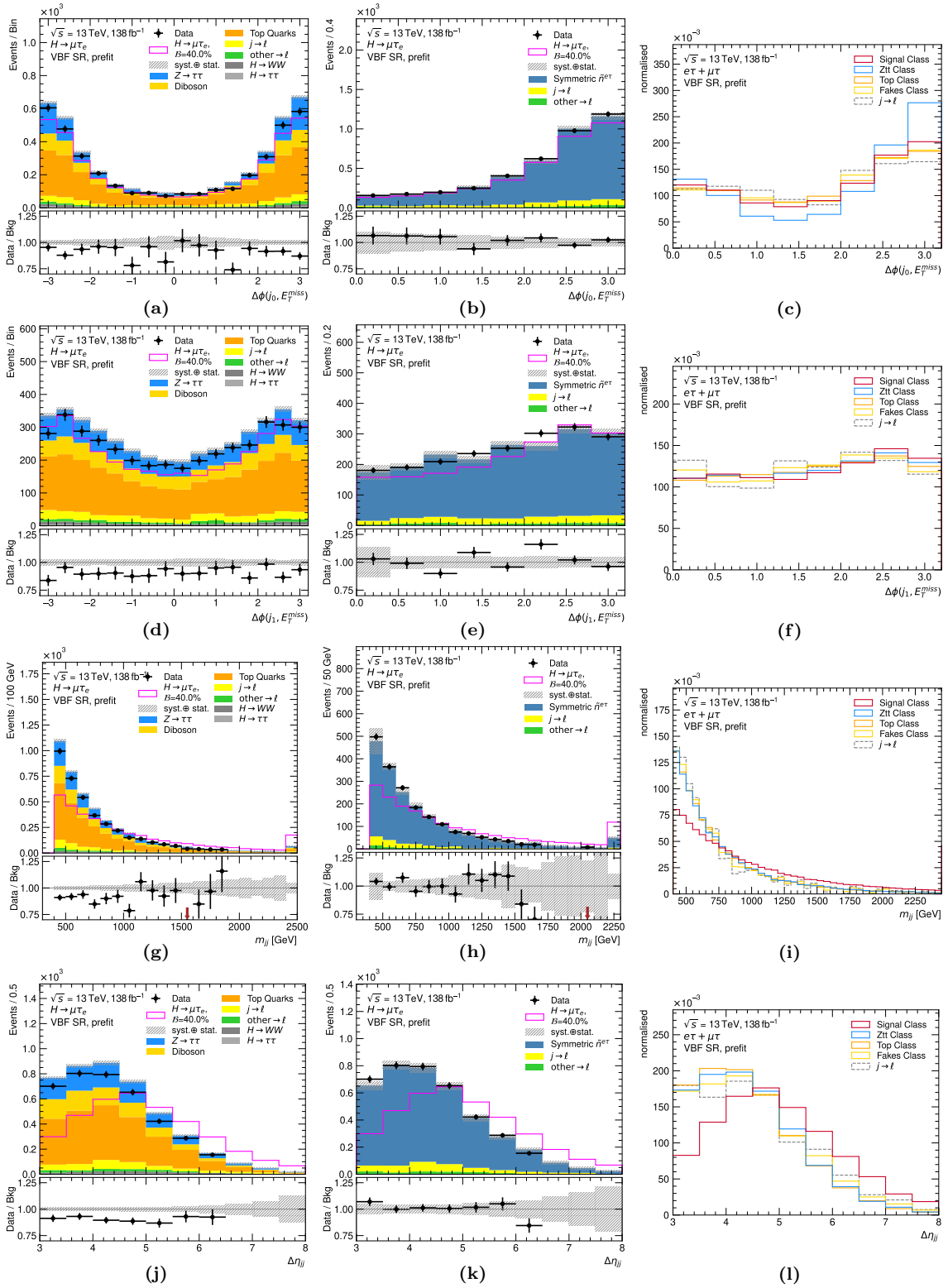



**Figure C.11:** Distributions of input variables to the *VBF NN*s. The left and middle columns compare the prediction to the data in the $\mu\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datsets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

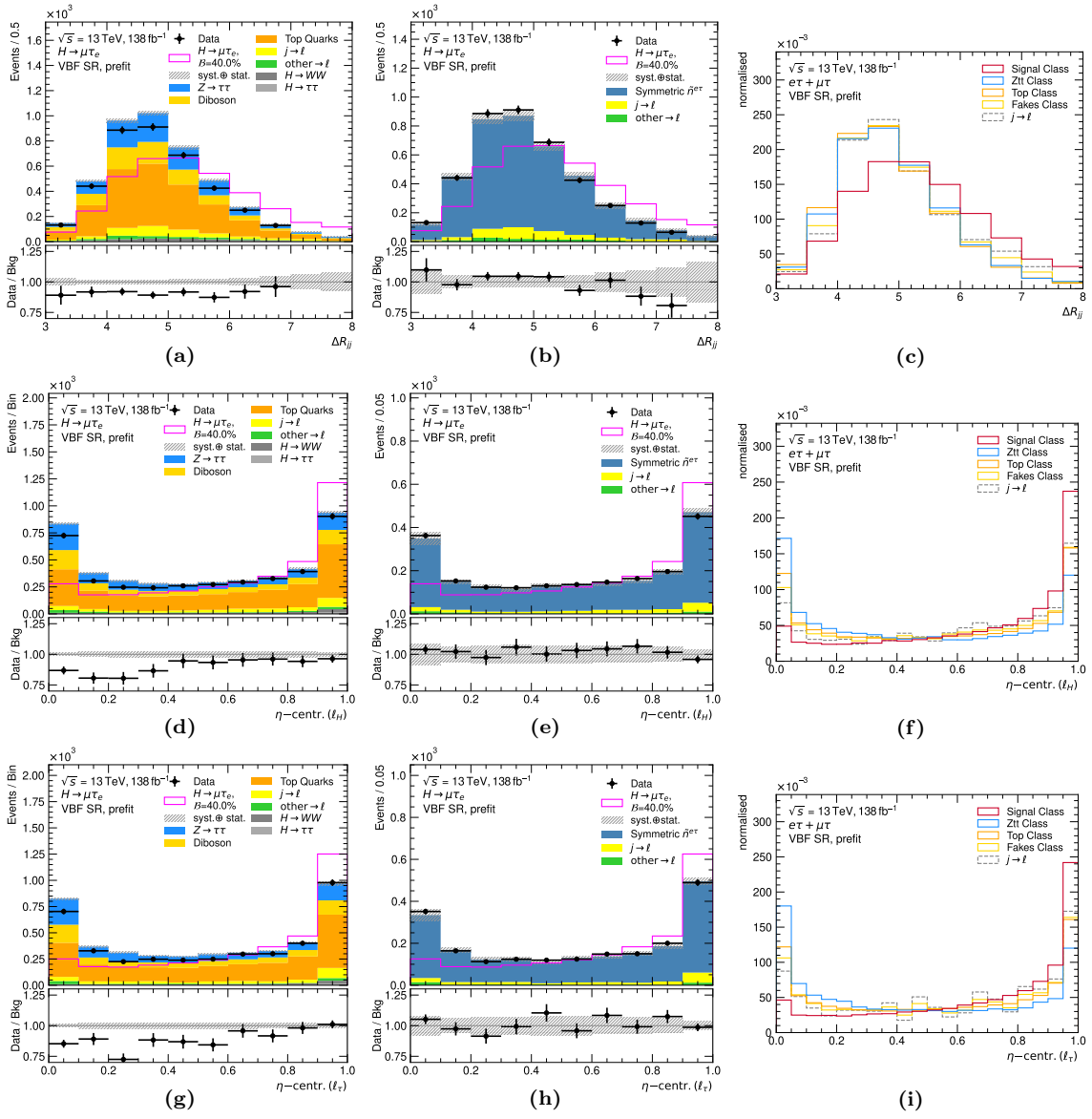



**Figure C.12:** Distributions of input variables to the *VBF NN*s. The left and middle columns compare the prediction to the data in the $\mu\tau$-dataset and show the ratio of data over the background prediction in the lower panel. The LFV signal prediction assuming a branching ratio of 40 % is overlaid. The $e/\mu$-symmetric background is once estimated from MC simulations (left) and once with the *Symmetry Method* (middle). The uncertainty band in the left column comprises statistical uncertainties on the background prediction and systematic uncertainties on the $j \to \ell$-fake estimate added in quadrature. The uncertainty band in the middle column comprises statistical uncertainties and the full set of systematic uncertainties on the background predictions added in quadrature. The right column shows normalised distributions for the sum of the $e\tau$- and $\mu\tau$-datsets, separately for each classes used in the NN training. In addition, the distribution of the full $j \to \ell$ fake estimate is shown for comparison.

# Appendix D

# List of Nuisance Parameters

## Search for $H \to e\tau$

| | nonVBF SR | | | | | | | | | | | VBF SR | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nuisance Parameter | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot$ *other* $\to \ell$-fakes | $j \to \ell$-fakes | *other* $\to \ell$-fakes | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot$ *other* $\to \ell$-fakes | $j \to \ell$-fakes | *other* $\to \ell$-fakes |
| $e\gamma$ scale | ◊ | ★ | ◊ | ★ | ★ | ★ | – | – | ★ | – | ★ | ★ | ◊ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| $e\gamma$ resolution | ◊ | ◊ | ◊ | ★ | ○ | ○ | – | – | ◊ | – | ○ | ◊ | ◊ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| El Trigger SF | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | ○ |
| El Reco Eff | ○ | ○ | ○ | ★ | ★ | ○ | ★ | ★ | ○ | – | ○ | ○ | ○ | ○ | ○ | ★ | ○ | ○ | ○ | ○ | – | ○ |
| El Id Eff Uncorr NP0 | × | × | × | ◊ | ◊ | × | ★ | ★ | × | – | × | × | × | × | ◊ | ◊ | × | ★ | ○ | × | – | ○ |
| El Id Eff Uncorr NP1 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP2 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP3 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP4 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP5 | × | ◊ | ★ | × | × | × | × | × | × | – | × | ○ | ○ | ○ | × | ○ | × | × | × | × | – | × |
| El Id Eff Uncorr NP6 | × | × | × | × | × | × | ★ | ○ | × | – | × | × | × | × | × | × | × | ○ | ○ | × | – | × |
| El Id Eff Uncorr NP7 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP8 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP9 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP10 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP11 | × | × | × | ◊ | ★ | ★ | × | × | ★ | – | × | × | × | × | ○ | ★ | ○ | × | × | ○ | – | ○ |
| El Id Eff Corr NP0 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP1 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP2 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP3 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP4 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP5 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP6 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP7 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP8 | × | × | × | × | × | × | × | ○ | × | – | × | × | × | × | × | × | × | × | ○ | × | – | × |
| El Id Eff Corr NP9 | × | × | × | × | × | × | ○ | ○ | × | – | × | × | × | × | × | × | × | × | ○ | × | – | × |
| El Id Eff Corr NP10 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP11 | × | ○ | ○ | × | ○ | ○ | × | × | × | – | × | ○ | ○ | ○ | ○ | ○ | ○ | × | × | ○ | – | ○ |
| El Id Eff Corr NP12 | × | × | × | ○ | ○ | ○ | × | × | ○ | – | ○ | × | × | × | ○ | × | ○ | × | × | ○ | – | ○ |
| El Id Eff Corr NP13 | × | × | × | ◊ | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP14 | ○ | ○ | ○ | ★ | ◊ | × | ★ | ○ | ○ | – | ○ | × | ◊ | × | ◊ | × | × | ○ | ○ | × | – | × |
| El Id Eff Corr NP15 | × | × | × | ◊ | ◊ | × | ○ | ○ | × | – | × | × | × | × | × | × | × | ○ | ○ | × | – | ○ |
| El Iso Eff | × | × | ○ | × | × | × | × | ○ | × | – | × | × | ○ | ○ | ○ | ○ | ○ | × | × | ○ | – | ○ |

**Table D.1:** List of nuisance parameters used in the statistical analysis in the search for $H \to e\tau$ and whether both shape and normalisation effect (★), only the shape effect (◊), only the normalisation effect (○) or neither (×) are considered for the individual processes in the two signal regions. If shape and/or normalisation effect are not considered, they were pruned following the criteria described in Section 9.5. The dash (–) indicates that the nuisance parameter does not act on the respective process.

| | nonVBF SR | | | | | | | | | | | VBF SR | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nuisance Parameter | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot other \to \ell$-fakes | $j \to \ell$-fakes | $other \to \ell$-fakes | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot other \to \ell$-fakes | $j \to \ell$-fakes | $other \to \ell$-fakes |
| Muon ID | ◊ | ◊ | ◊ | ◊ | ◊ | ★ | – | – | ◊ | – | ◊ | ★ | ◊ | ○ | ○ | ◊ | ○ | – | – | ○ | – | ○ |
| Muon MS | ◊ | × | ◊ | ◊ | ◊ | ★ | – | – | ◊ | – | ◊ | ★ | ◊ | ○ | ○ | ◊ | ○ | – | – | ○ | – | ○ |
| Muon scale | ◊ | ◊ | ◊ | ★ | ★ | ★ | – | – | ★ | – | ◊ | ◊ | ◊ | × | ○ | ★ | ○ | – | – | ○ | – | × |
| Muon Sagitta $\rho$ | ◊ | × | × | ◊ | ◊ | × | – | – | ◊ | – | × | ◊ | ◊ | × | × | ◊ | × | – | – | ○ | – | ○ |
| Muon Sagitta res.bias | ◊ | ◊ | × | × | ◊ | × | – | – | ◊ | – | × | ◊ | ◊ | × | × | ◊ | × | – | – | ○ | – | ○ |
| Mu Trigger $SF$ Stat | × | × | × | × | × | × | – | – | ○ | – | × | × | × | × | × | × | × | – | – | × | – | × |
| Mu Trigger $SF$ Sys | × | ○ | ○ | ★ | ★ | ★ | – | – | ★ | – | ★ | ○ | ○ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Mu Reco Eff Stat | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| Mu yeco Eff Sys | ○ | ○ | ○ | ○ | ○ | ○ | – | – | ○ | – | ○ | ○ | ○ | ○ | ○ | ○ | ○ | – | – | ○ | – | ○ |
| Mu Iso Eff Stat | × | × | × | × | ○ | ○ | – | – | × | – | × | × | × | × | ○ | ○ | ○ | – | – | ○ | – | × |
| Mu Iso Eff Sys | ★ | ★ | ★ | ★ | ★ | ○ | – | – | ○ | – | ★ | ○ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet effectiveNP Stat. 1 | × | ◊ | ◊ | ◊ | ◊ | × | – | – | × | – | × | ◊ | × | × | × | ◊ | ○ | – | – | × | – | × |
| Jet effectiveNP Stat. 2 | ◊ | ◊ | ◊ | ★ | ◊ | ◊ | – | – | ◊ | – | ◊ | ★ | ◊ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet effectiveNP Stat. 3 | × | × | ◊ | × | ◊ | ◊ | – | – | × | – | × | ◊ | × | ○ | × | ◊ | × | – | – | × | – | ○ |
| Jet effectiveNP Stat. 4 | × | ◊ | ◊ | ◊ | ◊ | × | – | – | ◊ | – | × | ◊ | ◊ | × | ◊ | ★ | × | – | – | × | – | ○ |
| Jet effectiveNP Stat. 5 | ◊ | × | ◊ | ◊ | ◊ | ◊ | – | – | ◊ | – | × | ★ | ◊ | ○ | × | ★ | ○ | – | – | × | – | × |
| Jet effectiveNP Stat. 6 | × | ◊ | ◊ | × | ◊ | × | – | – | × | – | ◊ | ◊ | ◊ | ○ | ★ | ★ | ○ | – | – | × | – | × |
| Jet effectiveNP mixed 1 | ◊ | ◊ | ◊ | ◊ | ◊ | × | – | – | ◊ | – | ◊ | ★ | ◊ | × | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet effectiveNP mixed 2 | × | ★ | ◊ | ○ | ◊ | ◊ | – | – | × | – | × | ★ | ★ | × | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet effectiveNP mixed 3 | ◊ | × | ◊ | ◊ | ◊ | ◊ | – | – | ◊ | – | ◊ | ★ | ◊ | × | ○ | ★ | ○ | – | – | ○ | – | × |
| Jet effectiveNP Modell. 1 | ★ | ★ | ★ | ★ | ★ | ★ | – | – | ◊ | – | ○ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet effectiveNP Modell. 2 | ◊ | ◊ | ◊ | ◊ | ◊ | ◊ | – | – | ◊ | – | × | ★ | ◊ | × | ◊ | ★ | ○ | – | – | × | – | ○ |
| Jet effectiveNP Modell. 3 | × | ◊ | ★ | ◊ | ◊ | × | – | – | ◊ | – | × | ◊ | ◊ | × | × | ★ | ○ | – | – | ○ | – | ○ |
| Jet effectiveNP Modell. 4 | × | ◊ | ◊ | ◊ | ◊ | × | – | – | × | – | ◊ | ◊ | ◊ | ○ | ◊ | ★ | ○ | – | – | × | – | × |
| Jet effectiveNP Detect. 1 | × | ◊ | ◊ | ★ | ◊ | × | – | – | ◊ | – | ◊ | ★ | ◊ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet effectiveNP Detect. 2 | × | ◊ | × | × | ◊ | × | – | – | × | – | × | ◊ | ◊ | × | ◊ | ◊ | ○ | – | – | × | – | × |
| Jet $\eta$-intercalibr. modelling | ★ | ★ | ★ | ★ | ★ | ★ | – | – | ★ | – | ★ | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet $\eta$-intercalibr. non-closure highE | × | × | × | × | × | × | – | – | × | – | × | ◊ | × | × | × | ◊ | × | – | – | × | – | × |
| Jet $\eta$-intercalibr. non-closure neg.$\eta$ | ◊ | ◊ | × | × | × | ◊ | – | – | ◊ | – | × | ◊ | ◊ | × | ◊ | ★ | × | – | – | ○ | – | × |
| Jet $\eta$-intercalibr. non-closure pos.$\eta$ | × | ◊ | × | × | ◊ | ★ | – | – | ◊ | – | × | ◊ | × | ○ | ★ | ◊ | ○ | – | – | × | – | × |
| Jet $\eta$-intercalibr. non-closure total stat | ◊ | ★ | ◊ | ★ | ★ | ◊ | – | – | ◊ | – | × | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet $\eta$-intercalibr. non-closure 2018data | × | ◊ | ◊ | ◊ | ◊ | ◊ | – | – | ◊ | – | × | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet pileup offset $\mu$ | ★ | ★ | ★ | ★ | ★ | ★ | – | – | ★ | – | ○ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet pileup offset npv | ★ | ★ | ★ | ★ | ★ | ★ | – | – | ◊ | – | ○ | ★ | ★ | ○ | ○ | ★ | × | – | – | ○ | – | ○ |
| Jet pileup pT term | × | ◊ | ◊ | ◊ | ○ | ○ | – | – | ◊ | – | ★ | ★ | ◊ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet pileup $\rho$ topology | ★ | ★ | ★ | ★ | ★ | × | – | – | ★ | – | ★ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet punch through mc16 | × | × | × | × | ◊ | × | – | – | × | – | × | × | × | × | × | ◊ | × | – | – | × | – | × |
| Jet single particle highPt | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | ◊ | × | – | – | × | – | × |
| Jet b-JES response | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | ◊ | × | – | – | × | – | × |
| Jet JER effectiveNP 1 | ★ | ★ | ★ | ★ | ★ | ★ | – | – | ○ | – | ○ | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 2 | ◊ | ★ | ★ | ★ | ★ | ★ | – | – | ◊ | – | ○ | ★ | ◊ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 3 | ★ | ★ | ★ | ★ | ★ | ★ | – | – | ★ | – | ○ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 4 | ★ | ◊ | ★ | ★ | ★ | ○ | – | – | ★ | – | ○ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 5 | ◊ | ◊ | ◊ | ◊ | ◊ | ◊ | – | – | ○ | – | ○ | ★ | ◊ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 6 | ◊ | ★ | ◊ | ★ | ★ | ★ | – | – | ★ | – | × | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 7 | ◊ | ◊ | ◊ | ◊ | ★ | × | – | – | ◊ | – | × | ★ | ◊ | ○ | ○ | ★ | ○ | – | – | ○ | – | × |
| Jet JER effectiveNP 8 | ◊ | ◊ | ◊ | ◊ | ★ | ○ | – | – | ★ | – | × | ◊ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 9 | ◊ | ★ | ★ | ◊ | ★ | ★ | – | – | ◊ | – | × | ◊ | ★ | ○ | × | ◊ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 10 | ◊ | ◊ | ◊ | ★ | ◊ | ◊ | – | – | ★ | – | ○ | ◊ | ★ | ○ | ○ | ◊ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 11 | ◊ | ◊ | ◊ | × | ◊ | ★ | – | – | × | – | × | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 12restTerm | ◊ | ◊ | ★ | ◊ | ★ | ○ | – | – | ◊ | – | ○ | ◊ | ◊ | ○ | ○ | ◊ | ○ | – | – | ○ | – | ○ |
| Jet JER DataVsMC | ◊ | ◊ | ◊ | ○ | ★ | ○ | – | – | ◊ | – | × | ★ | ◊ | py | ○ | ★ | ○ | – | – | ○ | – | ○ |

**Table D.2:** List of nuisance parameters used in the statistical analysis in the search for $H \to e\tau$ and whether both shape and normalisation effect (★), only the shape effect (◊), only the normalisation effect (○) or neither (×) are considered for the individual processes in the two signal regions. If shape and/or normalisation effect are not considered, they were pruned following the criteria described in Section 9.5. The dash (–) indicates that the nuisance parameter does not act on the respective process.

| | nonVBF SR | | | | | | | | | | | VBF SR | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nuisance Parameter | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot$ *other* $\to \ell$-fakes | $j \to \ell$-fakes | *other* $\to \ell$-fakes | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot$ *other* $\to \ell$-fakes | $j \to \ell$-fakes | *other* $\to \ell$-fakes |
| JET Flavour Comp non-VBF | ★ | ★ | ★ | ★ | ★ | ○ | – | – | ★ | – | ○ | – | – | – | – | – | – | – | – | – | – | – |
| JET Flavour Comp VBF | – | – | – | – | – | – | – | – | – | – | – | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| JET Flavour Resp non-VBF | ★ | ★ | ★ | ★ | ★ | × | – | – | ★ | – | ○ | – | – | – | – | – | – | – | – | – | – | – |
| JET Flavour Resp VBF | – | – | – | – | – | – | – | – | – | – | – | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JVT Efficiency | ◊ | ★ | ★ | ★ | ★ | ★ | – | – | ★ | – | ○ | ★ | ◊ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet fJVT Efficienc | ◊ | ★ | × | ★ | ★ | × | – | – | × | – | × | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| b-tag Eigenvar. 0 | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | ○ | – | – | ○ | – | ○ |
| b-tag Eigenvar. 1 | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| b-tag Eigenvar. 2 | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| c-tag Eigenvar. 0 | ○ | ○ | ★ | ○ | ★ | ○ | – | – | ○ | – | ○ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| c-tag Eigenvar. 1 | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| c-tag Eigenvar. 2 | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| light-tag Eigenvar. 0 | ★ | ○ | ★ | ★ | ○ | ★ | – | – | ○ | – | ○ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| light-tag Eigenvar. 1 | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| light-tag Eigenvar. 2 | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| light-tag Eigenvar. 3 | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| b-tag Extrapolation | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| b-tag Extrapolation from c | × | × | ★ | × | × | ★ | – | – | × | – | × | × | × | ○ | × | × | ○ | – | – | × | – | × |
| met soft track scale | ★ | ★ | ★ | ★ | ★ | ★ | – | – | ◊ | – | × | ★ | ★ | ○ | ★ | ★ | ○ | – | – | × | – | ○ |
| met soft track res.paral. | ★ | ★ | ★ | ★ | ★ | ★ | – | – | ◊ | – | ○ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| met soft track res.perp. | ★ | ★ | ★ | ★ | ★ | ★ | – | – | ◊ | – | × | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Luminosity | ○ | ○ | ○ | ○ | ○ | ○ | – | – | ○ | – | ○ | ○ | ○ | ○ | ○ | ○ | ○ | – | – | ○ | – | ○ |
| MCfakesScalingUnc | – | – | – | – | – | – | – | – | ○ | – | ○ | – | – | – | – | – | – | – | – | ○ | – | ○ |
| theory ggF qcd 0 | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory ggF qcd 1 | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory ggF qcd 2 | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory ggF qcd 3 | ◊ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ○ | – | – | – | – | – | – | – |
| theory ggF qcd 4 | ◊ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory ggF qcd 5 | ◊ | – | – | ◊ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory ggF qcd 6 | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory ggF qcd 7 | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ○ | – | – | – | – | – | – | – |
| theory ggF qcd 8 | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ○ | – | – | – | – | – | – | – |
| theory VBF qcd | – | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – |
| theory VH qcd | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – |
| theory sig pdf 0 | × | × | ◊ | × | × | ◊ | – | – | – | – | – | × | ★ | ○ | ○ | ◊ | ○ | – | – | – | – | – |
| theory sig pdf 1 | ★ | ◊ | × | ◊ | ★ | ★ | – | – | – | – | – | × | ★ | ○ | ○ | ★ | ○ | – | – | – | – | – |
| theory sig pdf 2 | × | ◊ | × | × | × | × | – | – | – | – | – | × | × | × | × | ★ | × | – | – | – | – | – |
| theory sig pdf 3 | ★ | ★ | ★ | ◊ | ◊ | ○ | – | – | – | – | – | ○ | ◊ | ○ | ★ | ◊ | ○ | – | – | – | – | – |
| theory sig pdf 4 | ★ | ◊ | ◊ | ★ | ◊ | ◊ | – | – | – | – | – | ★ | ○ | × | ○ | ★ | ○ | – | – | – | – | – |
| theory sig pdf 5 | ◊ | × | ◊ | ◊ | ◊ | × | – | – | – | – | – | ○ | ★ | ○ | ○ | ◊ | ○ | – | – | – | – | – |
| theory sig pdf 6 | × | × | × | × | × | × | – | – | – | – | – | × | ◊ | × | × | ★ | × | – | – | – | – | – |
| theory sig pdf 7 | × | ◊ | × | × | ◊ | × | – | – | – | – | – | × | × | × | × | ◊ | × | – | – | – | – | – |
| theory sig pdf 8 | × | ◊ | × | × | ◊ | × | – | – | – | – | – | ○ | × | ○ | ○ | × | ○ | – | – | – | – | – |
| theory sig pdf 9 | × | × | × | × | ◊ | × | – | – | – | – | – | × | × | × | × | × | × | – | – | – | – | – |
| theory sig pdf 10 | ◊ | × | × | ◊ | × | × | – | – | – | – | – | ○ | ○ | × | ○ | ★ | × | – | – | – | – | – |
| theory sig pdf 11 | × | ◊ | × | × | ◊ | × | – | – | – | – | – | ○ | × | × | ◊ | ◊ | × | – | – | – | – | – |
| theory sig pdf 12 | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | ★ | × | – | – | – | – | – |
| theory sig pdf 13 | × | × | × | × | × | × | – | – | – | – | – | × | × | ○ | × | × | ○ | – | – | – | – | – |
| theory sig pdf 14 | × | × | × | × | × | × | – | – | – | – | – | × | ○ | ○ | × | ◊ | ○ | – | – | – | – | – |
| theory sig pdf 15 | × | × | × | × | ◊ | × | – | – | – | – | – | ○ | × | ○ | × | ◊ | ○ | – | – | – | – | – |
| theory sig pdf 16 | ★ | ◊ | × | ★ | ◊ | × | – | – | – | – | – | ○ | ○ | ○ | ○ | ○ | ○ | – | – | – | – | – |

**Table D.3:** List of nuisance parameters used in the statistical analysis in the search for $H \to e\tau$ and whether both shape and normalisation effect (★), only the shape effect (◊), only the normalisation effect (○) or neither (×) are considered for the individual processes in the two signal regions. If shape and/or normalisation effect are not considered, they were pruned following the criteria described in Section 9.5. The dash (–) indicates that the nuisance parameter does not act on the respective process.

| | nonVBF SR | | | | | | | | | | | VBF SR | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nuisance Parameter | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot other \to \ell$-fakes | $j \to \ell$-fakes | $other \to \ell$-fakes | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot other \to \ell$-fakes | $j \to \ell$-fakes | $other \to \ell$-fakes |
| theory sig pdf 17 | × | × | × | × | ◊ | × | – | – | – | – | – | × | × | × | × | ◊ | × | – | – | – | – | – |
| theory sig pdf 18 | ★ | ◊ | ◊ | ◊ | ◊ | × | – | – | – | – | – | ○ | × | ○ | ○ | × | ○ | – | – | – | – | – |
| theory sig pdf 19 | × | × | ★ | × | × | ◊ | – | – | – | – | – | × | × | ○ | × | × | ○ | – | – | – | – | – |
| theory sig pdf 20 | × | × | × | × | ◊ | × | – | – | – | – | – | × | ◊ | × | × | ◊ | ○ | – | – | – | – | – |
| theory sig pdf 21 | ◊ | × | × | × | ◊ | × | – | – | – | – | – | ○ | × | ○ | ○ | ★ | × | – | – | – | – | – |
| theory sig pdf 22 | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – | – | – | – |
| theory sig pdf 23 | × | ◊ | × | × | × | × | – | – | – | – | – | × | × | × | × | ★ | × | – | – | – | – | – |
| theory sig pdf 24 | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – | – | – | – |
| theory sig pdf 25 | × | ◊ | × | × | ◊ | × | – | – | – | – | – | × | × | ○ | × | × | ○ | – | – | – | – | – |
| theory sig pdf 26 | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | ◊ | × | – | – | – | – | – |
| theory sig pdf 27 | ★ | × | × | × | × | × | – | – | – | – | – | × | × | ○ | ○ | × | ○ | – | – | – | – | – |
| theory sig pdf 28 | × | ◊ | × | × | ◊ | × | – | – | – | – | – | × | × | × | × | × | × | – | – | – | – | – |
| theory sig pdf 29 | × | × | × | × | ◊ | × | – | – | – | – | – | × | × | × | × | × | × | – | – | – | – | – |
| theory sig $\alpha_s$ | ★ | ◊ | ◊ | ★ | × | ◊ | – | – | – | – | – | ★ | ★ | ○ | ○ | ★ | ○ | – | – | – | – | – |
| theory ggF ME | ★ | – | – | ★ | – | – | – | – | – | – | – | ◊ | – | – | × | – | – | – | – | – | – | – |
| theory VBF ME | – | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – |
| theory VH ME | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – | ○ | – | – | ○ | – | – | – | – | – |
| theory ggF PS | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ○ | – | – | – | – | – | – | – |
| theory VBF PS | – | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – |
| theory VH PS | – | – | ★ | – | – | ○ | – | – | – | – | – | – | – | ○ | – | – | ○ | – | – | – | – | – |
| El. Fake FF Stat NP0 | – | – | – | – | – | – | – | ★ | – | × | – | – | – | – | – | – | – | – | ○ | – | × | – |
| El. Fake FF Stat NP1 | – | – | – | – | – | – | – | ★ | – | × | – | – | – | – | – | – | – | – | ○ | – | × | – |
| El. Fake FF Stat NP2 | – | – | – | – | – | – | – | ★ | – | × | – | – | – | – | – | – | – | – | ○ | – | × | – |
| El. Fake FF Stat NP3 | – | – | – | – | – | – | – | ★ | – | × | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake FF Stat NP8 | – | – | – | – | – | – | – | ★ | – | × | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake FF Stat NP9 | – | – | – | – | – | – | – | ★ | – | × | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake FF Stat NP10 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake FF Stat NP11 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Mu. Fake FF Stat NP0 | – | – | – | – | – | – | – | × | – | ★ | – | – | – | – | – | – | – | – | × | – | ○ | – |
| Mu. Fake FF Stat NP1 | – | – | – | – | – | – | – | × | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Mu. Fake FF Stat NP2 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Fake WZxsec | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Fake ZZxsec | – | – | – | – | – | – | – | ○ | – | ○ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake CF Stat NP0 | – | – | – | – | – | – | – | ★ | – | × | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake CF Stat NP1 | – | – | – | – | – | – | – | ★ | – | × | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake CF Stat NP2 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake CF Stat NP3 | – | – | – | – | – | – | – | ★ | – | ◊ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Mu. Fake CF Stat NP0 | – | – | – | – | – | – | – | × | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Mu. Fake CF Stat NP1 | – | – | – | – | – | – | – | × | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Mu. Fake CF Stat NP2 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake CF Sys | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Mu. Fake CF Sys | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| $\varepsilon$ Mu Eff Trigger Stat | – | – | – | × | × | × | ○ | ○ | ○ | – | – | – | – | – | × | × | × | ○ | ○ | ○ | – | – |
| $\varepsilon$ Mu Eff Trigger Sys | – | – | – | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – |
| $\varepsilon$ Mu Eff Reco Stat | – | – | – | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – |
| $\varepsilon$ Mu Eff Reco Sys | – | – | – | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – |
| $\varepsilon$ Mu Eff Iso Stat | – | – | – | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | ○ | – | – |
| $\varepsilon$ Mu Eff Iso Sys | – | – | – | ◊ | × | × | ◊ | ◊ | × | – | – | – | – | – | × | × | × | × | × | ○ | – | – |
| $\varepsilon$ El Eff MC Stat | – | – | – | × | ◊ | ○ | ◊ | ★ | ★ | – | – | – | – | – | ◊ | × | × | ○ | ○ | ○ | – | – |
| $\varepsilon$ El Eff MC Sys | – | – | – | ★ | ★ | ○ | ★ | ★ | ★ | – | – | – | – | – | ◊ | ◊ | ○ | ○ | ○ | ○ | – | – |
| $\varepsilon$ El Eff Trigger | – | – | – | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – |

**Table D.4:** List of nuisance parameters used in the statistical analysis in the search for $H \to e\tau$ and whether both shape and normalisation effect (★), only the shape effect (◊), only the normalisation effect (○) or neither (×) are considered for the individual processes in the two signal regions. If shape and/or normalisation effect are not considered, they were pruned following the criteria described in Section 9.5. The dash (–) indicates that the nuisance parameter does not act on the respective process.

# Search for $H \to \mu\tau$

| | nonVBF SR | | | | | | | | | | | VBF SR | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nuisance Parameter | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot other \to \ell$-fakes | $j \to \ell$-fakes | $other \to \ell$-fakes | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot other \to \ell$-fakes | $j \to \ell$-fakes | $other \to \ell$-fakes |
| $e\gamma$ scale | ★ | ★ | ★ | ★ | ★ | ○ | – | – | ★ | – | ★ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| $e\gamma$ resolution | ◊ | ◊ | ★ | ◊ | ★ | ◊ | – | – | ○ | – | ◊ | ◊ | ◊ | × | ○ | ◊ | × | – | – | ○ | – | ○ |
| El Trigger SF | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | ○ | – | × |
| El Reco Eff | ★ | ★ | ○ | ★ | ○ | ○ | ○ | ○ | ○ | – | ○ | ○ | ○ | ○ | ○ | ○ | ○ | × | ○ | ○ | – | ○ |
| El Id Eff Uncorr NP0 | ★ | ★ | ★ | ◊ | ◊ | ○ | ★ | ★ | ★ | – | ★ | ★ | ★ | ○ | ◊ | ◊ | ○ | ○ | ○ | ○ | – | ○ |
| El Id Eff Uncorr NP1 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP2 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP3 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP4 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP5 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | ○ | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP6 | ○ | ○ | ○ | × | × | × | ★ | ○ | ○ | – | ★ | × | × | ○ | × | × | × | × | × | ○ | – | ○ |
| El Id Eff Uncorr NP7 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP8 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP9 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP10 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Uncorr NP11 | ★ | ★ | ★ | ◊ | ★ | × | × | × | × | – | ★ | ★ | ★ | ○ | ★ | ★ | ○ | × | × | ○ | – | ○ |
| El Id Eff Corr NP0 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP1 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP2 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP3 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP4 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP5 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP6 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP7 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | × | – | × |
| El Id Eff Corr NP8 | × | × | × | × | × | × | × | ○ | × | – | × | × | × | × | × | × | × | ○ | × | – | – | × |
| El Id Eff Corr NP9 | × | × | × | × | × | × | × | ○ | × | – | × | × | × | × | × | × | × | ○ | × | – | – | × |
| El Id Eff Corr NP10 | × | × | × | × | × | × | × | × | × | – | × | × | × | × | × | × | × | × | × | – | – | × |
| El Id Eff Corr NP11 | ○ | ○ | ○ | × | ○ | ○ | × | × | ○ | – | ○ | ○ | ○ | ○ | ○ | ○ | × | × | ○ | – | – | ○ |
| El Id Eff Corr NP12 | ○ | ○ | ○ | ○ | ○ | ○ | × | × | ○ | – | ○ | ○ | ○ | ○ | – | ○ | × | × | ○ | – | – | ○ |
| El Id Eff Corr NP13 | ○ | ○ | ○ | × | × | × | × | × | ○ | – | ○ | ○ | ○ | × | × | × | × | × | × | – | – | × |
| El Id Eff Corr NP14 | ◊ | ◊ | × | ○ | ◊ | × | ★ | ★ | ○ | – | ★ | × | × | × | × | ◊ | × | ○ | ○ | ○ | – | × |
| El Id Eff Corr NP15 | ○ | ○ | ○ | ◊ | × | × | ○ | ○ | ○ | – | ○ | ○ | ○ | ○ | × | × | × | ○ | × | ○ | – | ○ |
| El Iso Eff | ○ | ○ | ○ | × | × | ○ | × | × | ○ | – | ○ | ○ | ○ | ○ | × | ○ | ○ | × | × | ○ | – | ○ |

**Table D.5:** List of nuisance parameters used in the statistical analysis in the search for $H \to \mu\tau$ and whether both shape and normalisation effect (★), only the shape effect (◊), only the normalisation effect (○) or neither (×) are considered for the individual processes in the two signal regions. If shape and/or normalisation effect are not considered, they were pruned following the criteria described in Section 9.5. The dash (–) indicates that the nuisance parameter does not act on the respective process.

| | nonVBF SR | | | | | | | | | | | VBF SR | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nuisance Parameter | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot$ *other* $\to \ell$-fakes | $j \to \ell$-fakes | *other* $\to \ell$-fakes | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot$ *other* $\to \ell$-fakes | $j \to \ell$-fakes | *other* $\to \ell$-fakes |
| Muon ID | ◊ | ◊ | ◊ | ◊ | ★ | ★ | – | – | ◊ | – | ◊ | ★ | ★ | ○ | ◊ | ★ | ○ | – | – | ○ | – | ○ |
| Muon MS | ◊ | ◊ | ◊ | ◊ | ◊ | ★ | – | – | ○ | – | ◊ | ★ | ◊ | ○ | ◊ | ◊ | ○ | – | – | × | – | × |
| Muon scale | ◊ | ◊ | ◊ | ★ | ○ | ★ | – | – | ◊ | – | ★ | ★ | ◊ | ○ | ★ | ★ | × | – | – | × | – | ○ |
| Muon Sagitta $\rho$ | ◊ | × | × | × | × | ◊ | – | – | × | – | ◊ | ◊ | ◊ | × | × | ◊ | × | – | – | ○ | – | ○ |
| Muon Sagitta res.bias | ◊ | × | × | × | × | × | – | – | × | – | ◊ | ◊ | ◊ | × | × | × | × | – | – | ○ | – | ○ |
| Mu Trigger $SF$ Stat | × | × | × | × | × | × | – | – | ○ | – | × | × | × | × | × | × | × | – | – | × | – | × |
| Mu Trigger $SF$ Sys | ○ | ★ | ★ | ★ | ★ | ★ | – | – | ★ | – | ★ | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Mu Reco Eff Stat | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| Mu Reco Eff Sys | ○ | ○ | ○ | ○ | ○ | ○ | – | – | ○ | – | ○ | ○ | ○ | ○ | ○ | ○ | ○ | – | – | ○ | – | ○ |
| Mu Iso Eff Stat | ○ | ○ | ○ | × | ○ | × | – | – | × | – | × | ○ | ○ | ○ | ○ | ○ | ○ | – | – | × | – | ○ |
| Mu Iso Eff Sys | ★ | ★ | ★ | ★ | ★ | ★ | – | – | ★ | – | ○ | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet effectiveNP Stat. 1 | ◊ | × | ◊ | ◊ | ◊ | × | – | – | ◊ | – | × | ◊ | ◊ | ○ | ★ | ◊ | × | – | – | × | – | × |
| Jet effectiveNP Stat. 2 | ◊ | ◊ | × | ★ | ○ | ◊ | – | – | ◊ | – | ◊ | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet effectiveNP Stat. 3 | × | ◊ | ◊ | × | ◊ | × | – | – | × | – | × | ◊ | ◊ | × | × | ★ | × | – | – | ○ | – | × |
| Jet effectiveNP Stat. 4 | × | ◊ | ◊ | ◊ | × | × | – | – | ◊ | – | ◊ | ★ | ◊ | ○ | ◊ | ★ | × | – | – | × | – | ○ |
| Jet effectiveNP Stat. 5 | × | ◊ | ◊ | ◊ | ◊ | × | – | – | × | – | ◊ | ◊ | ◊ | × | × | ★ | ○ | – | – | × | – | × |
| Jet effectiveNP Stat. 6 | × | ◊ | ◊ | ◊ | × | ◊ | – | – | ◊ | – | ◊ | ◊ | ◊ | ○ | ○ | ◊ | × | – | – | × | – | ○ |
| Jet effectiveNP mixed 1 | ◊ | × | ◊ | ◊ | ◊ | × | – | – | ◊ | – | ◊ | ★ | ◊ | ○ | ★ | ◊ | ○ | – | – | ○ | – | ○ |
| Jet effectiveNP mixed 2 | × | ◊ | ◊ | ◊ | ★ | × | – | – | ◊ | – | ◊ | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet effectiveNP mixed 3 | ◊ | ◊ | × | ◊ | ★ | ◊ | – | – | ◊ | – | ◊ | ★ | ◊ | ○ | ◊ | × | × | – | – | × | – | × |
| Jet effectiveNP Modell. 1 | ★ | ★ | ◊ | ◊ | ★ | ★ | – | – | ◊ | – | ◊ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet effectiveNP Modell. 2 | × | × | ◊ | ★ | × | ◊ | – | – | ◊ | – | ◊ | ★ | ◊ | ○ | × | ◊ | ○ | – | – | ○ | – | ○ |
| Jet effectiveNP Modell. 3 | ◊ | ◊ | ◊ | ★ | ◊ | ★ | – | – | ◊ | – | ◊ | ◊ | ◊ | × | ○ | ★ | ○ | – | – | × | – | ○ |
| Jet effectiveNP Modell. 4 | × | ◊ | ◊ | ◊ | × | × | – | – | ◊ | – | × | ◊ | ◊ | ○ | ★ | ◊ | × | – | – | × | – | ○ |
| Jet effectiveNP Detect. 1 | × | ◊ | × | ◊ | × | ◊ | – | – | ◊ | – | ◊ | ★ | ★ | × | ◊ | ★ | × | – | – | × | – | ○ |
| Jet effectiveNP Detect. 2 | × | × | ◊ | ◊ | × | ◊ | – | – | × | – | × | ◊ | ◊ | × | ◊ | ◊ | × | – | – | × | – | × |
| Jet $\eta$-intercalibr. modelling | ★ | ★ | ★ | ★ | ★ | ○ | – | – | × | – | ◊ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet $\eta$-intercalibr. non-closure highE | × | × | ◊ | × | × | × | – | – | × | – | × | × | × | × | × | ◊ | × | – | – | × | – | × |
| Jet $\eta$-intercalibr. non-closure neg.$\eta$ | × | ◊ | ◊ | ◊ | × | ◊ | – | – | × | – | × | ◊ | ◊ | ○ | × | ◊ | × | – | – | × | – | × |
| Jet $\eta$-intercalibr. non-closure pos.$\eta$ | × | ◊ | ◊ | ◊ | ◊ | ◊ | – | – | × | – | × | ★ | ◊ | × | ★ | ★ | × | – | – | × | – | × |
| Jet $\eta$-intercalibr. non-closure total stat | ◊ | ★ | ◊ | ◊ | ★ | ◊ | – | – | ★ | – | ◊ | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet $\eta$-intercalibr. non-closure 2018data | ◊ | ◊ | ◊ | × | ◊ | ◊ | – | – | ◊ | – | ◊ | ★ | ◊ | ○ | ○ | ★ | ○ | – | – | × | – | ○ |
| Jet pileup offset $\mu$ | ★ | ★ | ◊ | ★ | ★ | ★ | – | – | × | – | ★ | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet pileup offset npv | ★ | ★ | ◊ | ★ | ○ | ★ | – | – | ★ | – | ★ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet pileup pT term | × | ◊ | ◊ | ◊ | ◊ | ○ | – | – | ★ | – | × | ★ | ★ | × | ★ | ★ | × | – | – | ○ | – | ○ |
| Jet pileup $\rho$ topology | ★ | ★ | ★ | ★ | ★ | ★ | – | – | ◊ | – | ★ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet punch through mc16 | × | × | ◊ | × | × | × | – | – | × | – | × | × | ◊ | × | × | × | × | – | – | × | – | × |
| Jet single particle highPt | × | × | ◊ | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| Jet b-JES response | × | × | ◊ | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| Jet JER effectiveNP 1 | ★ | ★ | ★ | ★ | ★ | ★ | – | – | ○ | – | ★ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 2 | ◊ | ★ | ★ | ★ | ★ | ○ | – | – | ◊ | – | ★ | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 3 | ★ | ◊ | ★ | ★ | ★ | ◊ | – | – | × | – | ★ | ★ | ◊ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 4 | ★ | ◊ | ★ | ★ | ◊ | ★ | – | – | ○ | – | ◊ | ◊ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 5 | ◊ | ◊ | ★ | ★ | ★ | ★ | – | – | ○ | – | ○ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 6 | ◊ | ◊ | ★ | ◊ | ★ | ★ | – | – | × | – | ★ | ◊ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 7 | ◊ | ◊ | ◊ | ◊ | ◊ | ★ | – | – | ○ | – | ◊ | ★ | ◊ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 8 | ◊ | ◊ | ◊ | ★ | ★ | ★ | – | – | × | – | ★ | ★ | ◊ | × | ○ | ◊ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 9 | ◊ | ◊ | ◊ | ◊ | ◊ | ★ | – | – | × | – | ◊ | ◊ | ◊ | × | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 10 | ◊ | ◊ | ◊ | ★ | ◊ | ◊ | – | – | ★ | – | ★ | ★ | ◊ | ○ | ○ | ◊ | ○ | – | – | ○ | – | ○ |
| Jet JER effectiveNP 11 | ◊ | ◊ | ◊ | ★ | ◊ | ★ | – | – | × | – | ○ | ★ | ◊ | ○ | ○ | ★ | ○ | – | – | × | – | ○ |
| Jet JER effectiveNP 12restTerm | ◊ | ◊ | ◊ | ★ | ★ | ○ | – | – | ★ | – | ◊ | ◊ | ◊ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JER DataVsMC | ◊ | ◊ | ★ | ◊ | ★ | ★ | – | – | ○ | – | ◊ | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |

**Table D.6:** List of nuisance parameters used in the statistical analysis in the search for $H \to \mu\tau$ and whether both shape and normalisation effect (★), only the shape effect (◊), only the normalisation effect (○) or neither (×) are considered for the individual processes in the two signal regions. If shape and/or normalisation effect are not considered, they were pruned following the criteria described in Section 9.5. The dash (–) indicates that the nuisance parameter does not act on the respective process.

| Nuisance Parameter | nonVBF SR | | | | | | | | | | | VBF SR | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot$ *other* $\to \ell$-fakes | $j \to \ell$-fakes | *other* $\to \ell$-fakes | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot$ *other* $\to \ell$-fakes | $j \to \ell$-fakes | *other* $\to \ell$-fakes |
| JET Flavour Comp non-VBF | ★ | ★ | ★ | ★ | ○ | ★ | – | – | ◊ | – | ★ | – | – | – | – | – | – | – | – | – | – | – |
| JET Flavour Comp VBF | – | – | – | – | – | – | – | – | – | – | – | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| JET Flavour Resp non-VBF | ★ | ★ | ★ | ★ | ★ | ◊ | – | – | ★ | – | ◊ | – | – | – | – | – | – | – | – | – | – | – |
| JET Flavour Resp VBF | – | – | – | – | – | – | – | – | – | – | – | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Jet JVT Efficiency | ◊ | ★ | ★ | ★ | ★ | ★ | – | – | ★ | – | ★ | ◊ | ◊ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| Jet fJVT Efficienc | ◊ | ★ | ◊ | ★ | ★ | × | – | – | × | – | × | ★ | ★ | ○ | ★ | ★ | ○ | – | – | ○ | – | ○ |
| b-tag Eigenvar. 0 | × | × | ○ | × | × | × | – | – | × | – | × | ★ | × | ○ | × | × | ○ | – | – | ○ | – | ○ |
| b-tag Eigenvar. 1 | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| b-tag Eigenvar. 2 | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| c-tag Eigenvar. 0 | ○ | ○ | ★ | ○ | ○ | ○ | – | – | ★ | – | ○ | ○ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| c-tag Eigenvar. 1 | × | × | ◊ | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| c-tag Eigenvar. 2 | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| light-tag Eigenvar. 0 | ★ | ○ | ★ | ★ | ○ | ★ | – | – | ○ | – | ○ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| light-tag Eigenvar. 1 | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| light-tag Eigenvar. 2 | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| light-tag Eigenvar. 3 | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| b-tag Extrapolation | × | × | × | × | × | × | – | – | × | – | × | × | × | × | × | × | × | – | – | × | – | × |
| b-tag Extrapolation from c | × | × | ★ | × | × | ★ | – | – | × | – | × | × | × | ○ | × | × | ○ | – | – | × | – | × |
| met soft track scale | ★ | ★ | ◊ | ★ | ★ | ★ | – | – | ◊ | – | ◊ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| met soft track res.paral. | ★ | ★ | ★ | ★ | ★ | ★ | – | – | ◊ | – | ◊ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| met soft track res.perp. | ★ | ★ | ◊ | ★ | ★ | ○ | – | – | × | – | ◊ | ★ | ★ | ○ | ○ | ★ | ○ | – | – | ○ | – | ○ |
| Luminosity | ○ | ○ | ○ | ○ | ○ | ○ | – | – | ○ | – | ○ | ○ | ○ | ○ | ○ | ○ | ○ | – | – | ○ | – | ○ |
| MCfakesScalingUnc | – | – | – | – | – | – | – | – | ○ | – | ○ | – | – | – | – | – | – | – | – | ○ | – | ○ |
| theory ggF qcd 0 | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory ggF qcd 1 | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory ggF qcd 2 | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory ggF qcd 3 | ◊ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory ggF qcd 4 | ◊ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory ggF qcd 5 | ◊ | – | – | ◊ | – | – | – | – | – | – | – | ★ | – | – | ○ | – | – | – | – | – | – | – |
| theory ggF qcd 6 | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory ggF qcd 7 | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory ggF qcd 8 | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory VBF qcd | – | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – |
| theory VH qcd | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – | ○ | – | – | ○ | – | – | – | – | – |
| theory sig pdf 0 | × | × | × | × | ◊ | × | – | – | – | – | – | × | ★ | ○ | ○ | ★ | ○ | – | – | – | – | – |
| theory sig pdf 1 | ★ | ◊ | ◊ | ◊ | ◊ | ◊ | – | – | – | – | – | ◊ | ○ | ○ | ○ | ★ | ○ | – | – | – | – | – |
| theory sig pdf 2 | × | ◊ | × | × | ◊ | × | – | – | – | – | – | × | × | × | × | × | × | – | – | – | – | – |
| theory sig pdf 3 | ★ | ★ | ★ | ◊ | ◊ | × | – | – | – | – | – | ★ | ◊ | ○ | ★ | ◊ | ○ | – | – | – | – | – |
| theory sig pdf 4 | ★ | ◊ | ◊ | ◊ | × | × | – | – | – | – | – | ★ | ○ | × | ★ | ○ | ○ | – | – | – | – | – |
| theory sig pdf 5 | ◊ | × | ◊ | ◊ | ◊ | × | – | – | – | – | – | ★ | ★ | ○ | ★ | ★ | ○ | – | – | – | – | – |
| theory sig pdf 6 | × | × | × | × | × | × | – | – | – | – | – | × | ◊ | × | × | ◊ | × | – | – | – | – | – |
| theory sig pdf 7 | × | ◊ | × | × | × | × | – | – | – | – | – | × | × | × | ★ | × | × | – | – | – | – | – |
| theory sig pdf 8 | × | ◊ | × | × | × | × | – | – | – | – | – | ○ | × | ○ | ○ | × | ○ | – | – | – | – | – |
| theory sig pdf 9 | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – | – | – | – |
| theory sig pdf 10 | ◊ | × | ◊ | × | × | × | – | – | – | – | – | ○ | ○ | × | ★ | ○ | ○ | – | – | – | – | – |
| theory sig pdf 11 | × | ◊ | × | × | × | × | – | – | – | – | – | × | ◊ | × | × | × | × | – | – | – | – | – |
| theory sig pdf 12 | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – | – | – | – |
| theory sig pdf 13 | × | × | × | × | × | × | – | – | – | – | – | × | × | ○ | × | × | ○ | – | – | – | – | – |
| theory sig pdf 14 | × | × | × | × | × | × | – | – | – | – | – | × | ○ | ○ | × | ★ | ○ | – | – | – | – | – |
| theory sig pdf 15 | × | ◊ | × | × | ◊ | × | – | – | – | – | – | ○ | × | ○ | × | ◊ | × | – | – | – | – | – |
| theory sig pdf 16 | ◊ | ◊ | × | ★ | × | × | – | – | – | – | – | ○ | ○ | ○ | ○ | ○ | ○ | – | – | – | – | – |

**Table D.7:** List of nuisance parameters used in the statistical analysis in the search for $H \to \mu\tau$ and whether both shape and normalisation effect (★), only the shape effect (◊), only the normalisation effect (○) or neither (×) are considered for the individual processes in the two signal regions. If shape and/or normalisation effect are not considered, they were pruned following the criteria described in Section 9.5. The dash (–) indicates that the nuisance parameter does not act on the respective process.

| | nonVBF SR | | | | | | | | | | | VBF SR | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nuisance Parameter | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot other \to \ell$-fakes | $j \to \ell$-fakes | $other \to \ell$-fakes | ggF $H \to e\tau$ | VBF $H \to e\tau$ | VH $H \to e\tau$ | $\mathcal{R}\cdot$ ggF $H \to e\tau$ | $\mathcal{R}\cdot$ VBF $H \to e\tau$ | $\mathcal{R}\cdot$ VH $H \to e\tau$ | $e/\mu$-symm. bkg. data | $\mathcal{R}\cdot j \to \ell$-fakes | $\mathcal{R}\cdot other \to \ell$-fakes | $j \to \ell$-fakes | $other \to \ell$-fakes |
| theory sig pdf 17 | × | ◊ | × | × | ◊ | × | – | – | – | – | – | × | × | × | ◊ | ◊ | ○ | – | – | – | – | – |
| theory sig pdf 18 | ★ | ◊ | ◊ | ◊ | × | × | – | – | – | – | – | ○ | × | ○ | ○ | × | ○ | – | – | – | – | – |
| theory sig pdf 19 | × | × | ★ | × | × | × | – | – | – | – | – | × | × | ○ | × | × | × | – | – | – | – | – |
| theory sig pdf 20 | × | × | × | × | × | × | – | – | – | – | – | × | × | × | ◊ | × | × | – | – | – | – | – |
| theory sig pdf 21 | ◊ | × | × | ◊ | × | × | – | – | – | – | – | ○ | × | × | ★ | × | × | – | – | – | – | – |
| theory sig pdf 22 | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – | – | – | – |
| theory sig pdf 23 | × | ◊ | × | × | ◊ | × | – | – | – | – | – | × | × | × | × | × | × | – | – | – | – | – |
| theory sig pdf 24 | × | × | × | × | × | × | – | – | – | – | – | × | ◊ | × | × | × | × | – | – | – | – | – |
| theory sig pdf 25 | × | × | × | × | ◊ | × | – | – | – | – | – | × | × | ○ | × | × | ○ | – | – | – | – | – |
| theory sig pdf 26 | × | × | ◊ | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – | – | – | – |
| theory sig pdf 27 | ★ | × | × | × | × | × | – | – | – | – | – | ★ | × | ○ | ★ | × | ○ | – | – | – | – | – |
| theory sig pdf 28 | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – | – | – | – |
| theory sig pdf 29 | × | × | × | × | ◊ | × | – | – | – | – | – | × | ◊ | × | × | × | × | – | – | – | – | – |
| theory sig $\alpha_s$ | ★ | ◊ | ◊ | ★ | × | ◊ | – | – | – | – | – | ★ | ★ | ○ | ★ | ★ | ○ | – | – | – | – | – |
| theory ggF ME | ★ | – | – | ★ | – | – | – | – | – | – | – | ◊ | – | – | ◊ | – | – | – | – | – | – | – |
| theory VBF ME | – | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – |
| theory VH ME | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – | ○ | – | – | ○ | – | – | – | – | – |
| theory ggF PS | ★ | – | – | ◊ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – |
| theory VBF PS | – | ★ | – | – | ★ | – | – | – | – | – | – | – | ★ | – | – | ★ | – | – | – | – | – | – |
| theory VH PS | – | – | ★ | – | – | ★ | – | – | – | – | – | – | – | ○ | – | – | ○ | – | – | – | – | – |
| El. Fake FF Stat NP0 | – | – | – | – | – | – | – | × | – | ★ | – | – | – | – | – | – | – | – | × | – | ○ | – |
| El. Fake FF Stat NP1 | – | – | – | – | – | – | – | × | – | ★ | – | – | – | – | – | – | – | – | × | – | ○ | – |
| El. Fake FF Stat NP2 | – | – | – | – | – | – | – | × | – | ★ | – | – | – | – | – | – | – | – | × | – | ○ | – |
| El. Fake FF Stat NP3 | – | – | – | – | – | – | – | × | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake FF Stat NP8 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake FF Stat NP9 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake FF Stat NP10 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake FF Stat NP11 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Mu. Fake FF Stat NP0 | – | – | – | – | – | – | – | ★ | – | × | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Mu. Fake FF Stat NP1 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Mu. Fake FF Stat NP2 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Fake WZxsec | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Fake ZZxsec | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake CF Stat NP0 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake CF Stat NP1 | – | – | – | – | – | – | – | × | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake CF Stat NP2 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake CF Stat NP3 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Mu. Fake CF Stat NP0 | – | – | – | – | – | – | – | ★ | – | × | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Mu. Fake CF Stat NP1 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Mu. Fake CF Stat NP2 | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| El. Fake CF Sys | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| Mu. Fake CF Sys | – | – | – | – | – | – | – | ★ | – | ★ | – | – | – | – | – | – | – | – | ○ | – | ○ | – |
| $\varepsilon$ Mu Eff Trigger Stat | – | – | – | × | × | × | ○ | ○ | ○ | – | – | – | – | – | × | × | × | × | ○ | × | – | – |
| $\varepsilon$ Mu Eff Trigger Sys | – | – | – | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – |
| $\varepsilon$ Mu Eff Reco Stat | – | – | – | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – |
| $\varepsilon$ Mu Eff Reco Sys | – | – | – | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – |
| $\varepsilon$ Mu Eff Iso Stat | – | – | – | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – |
| $\varepsilon$ Mu Eff Iso Sys | – | – | – | × | × | × | ◊ | ◊ | × | – | – | – | – | – | × | × | × | × | × | × | – | – |
| $\varepsilon$ El Eff MC Stat | – | – | – | × | ◊ | × | ◊ | ★ | ◊ | – | – | – | – | – | × | × | × | ○ | ○ | × | – | – |
| $\varepsilon$ El Eff MC Sys | – | – | – | ◊ | ◊ | ○ | ★ | ★ | ○ | – | – | – | – | – | ★ | ◊ | × | ○ | ○ | × | – | – |
| $\varepsilon$ El Eff Trigger | – | – | – | × | × | × | × | × | × | – | – | – | – | – | × | × | × | × | × | × | – | – |

**Table D.8:** List of nuisance parameters used in the statistical analysis in the search for $H \to \mu\tau$ and whether both shape and normalisation effect (★), only the shape effect (◊), only the normalisation effect (○) or neither (×) are considered for the individual processes in the two signal regions. If shape and/or normalisation effect are not considered, they were pruned following the criteria described in Section 9.5. The dash (–) indicates that the nuisance parameter does not act on the respective process.

# Appendix E

# Supplementary Material for the Results

This appendix holds all supplementary material for the the discussion of the results obtained with the *Symmetry Method* in Chapter 11.

## Combined Fit in the nonVBF SR and the VBF SR

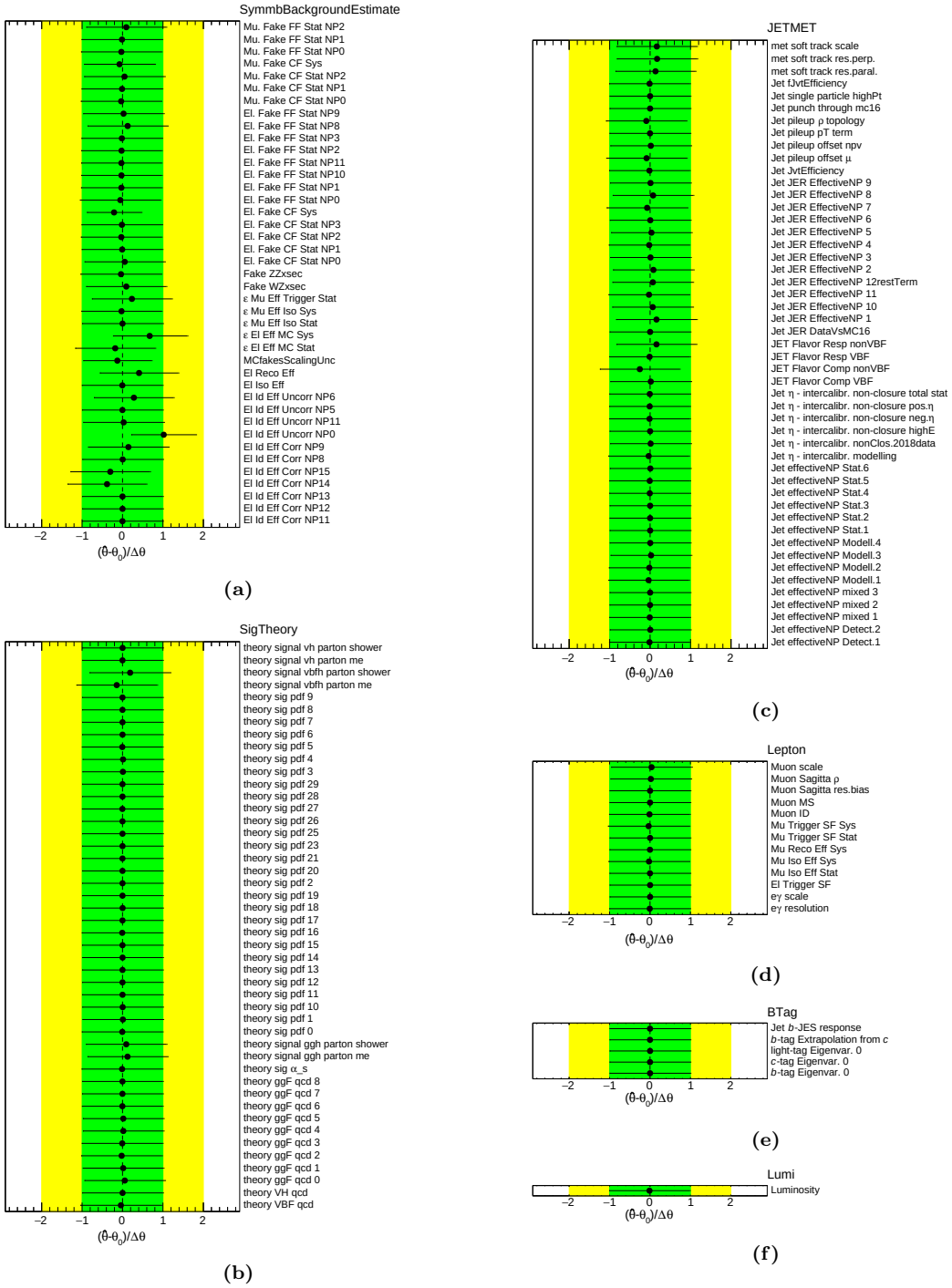



**Figure E.1:** The pulls and constraints of all nuisance parameters obtained from a combined maximum likelihood fit in both signal regions with the $e\tau$-dataset.

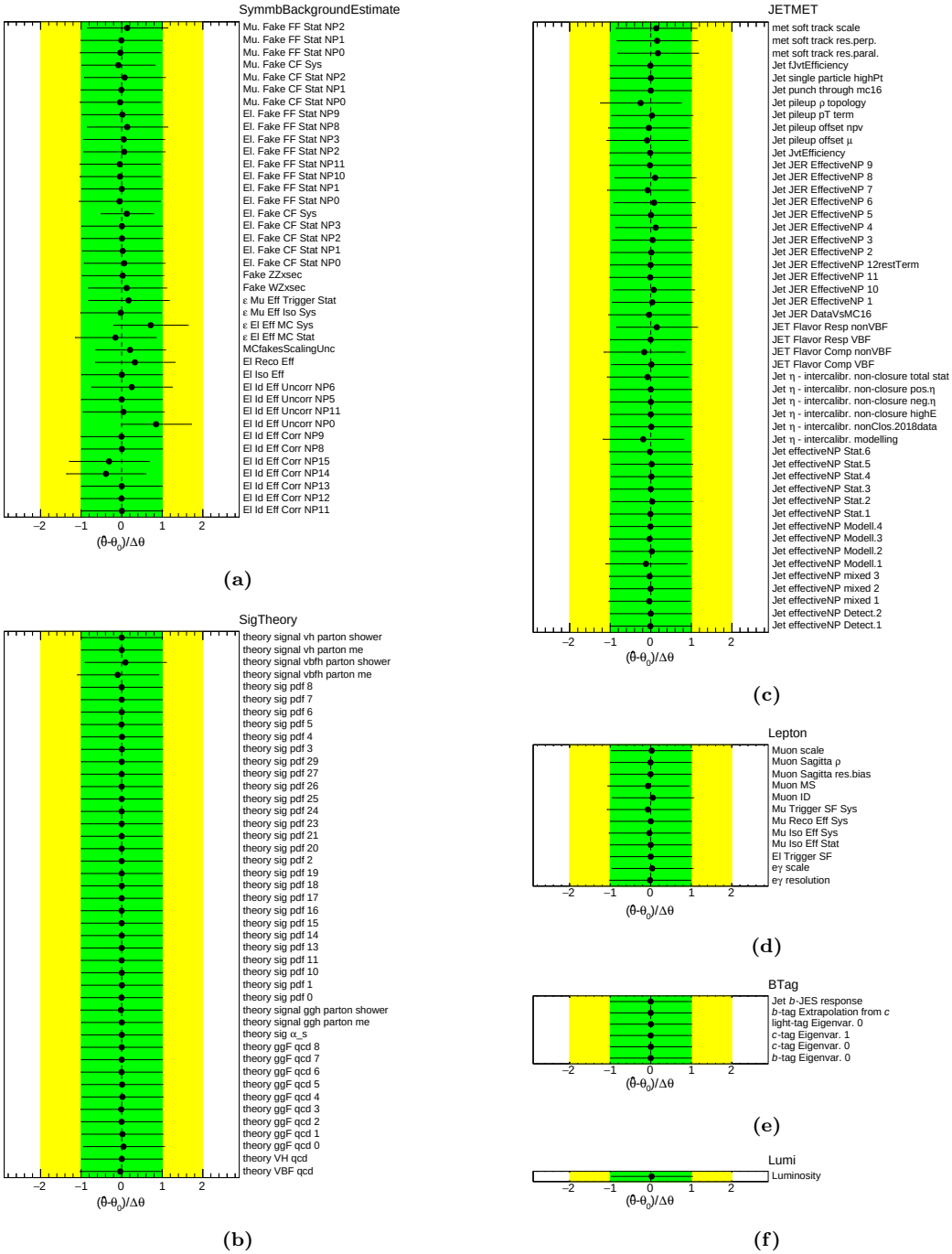
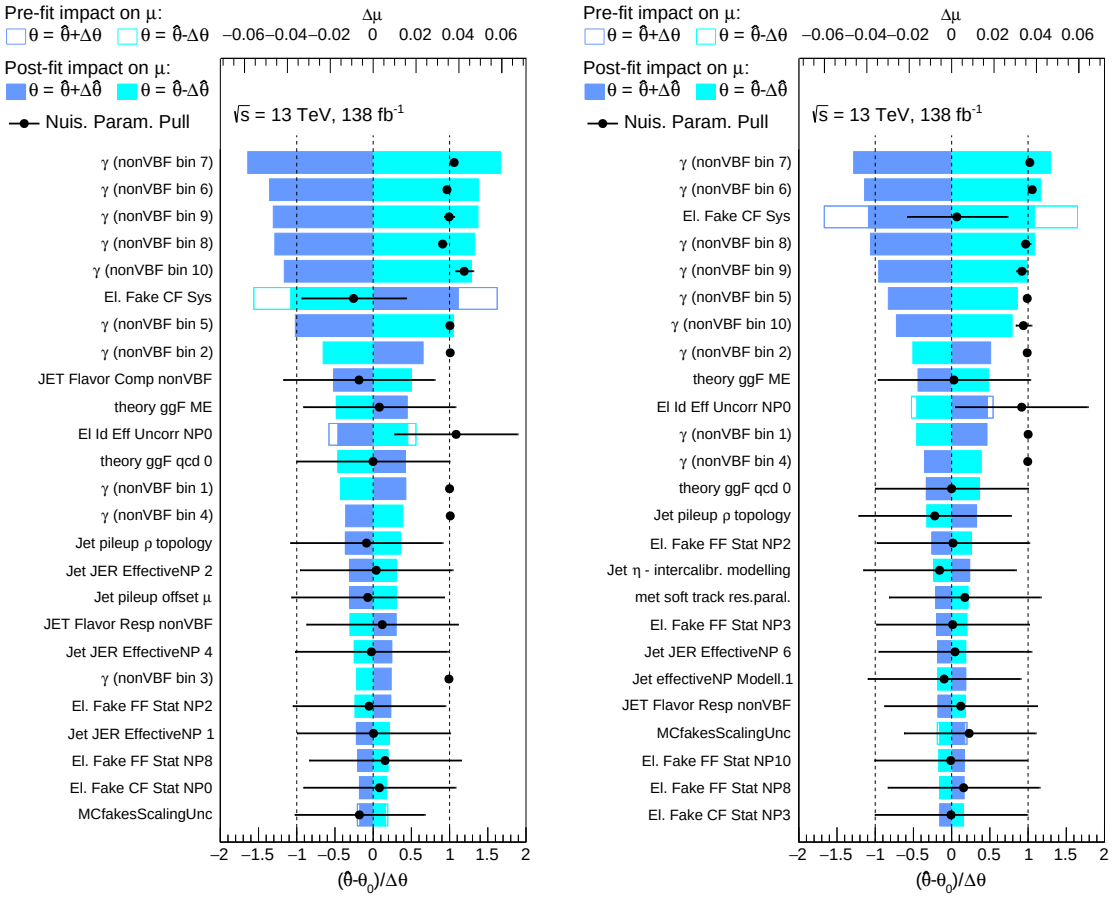



**Figure E.2:** The pulls and constraints of all nuisance parameters obtained from a combined maximum likelihood fit in both signal regions with the $\mu\tau$-dataset.

## Fit in the nonVBF SR




**(a)** Search for $H \to e\tau$ **(b)** Search for $H \to \mu\tau$

**Figure E.3:** The pulls and constraints (black dots and bars and lower horizontal axis) of the 25 highest ranked nuisance parameters obtained from a maximum likelihood fit in the nonVBF SR in the $e\tau$-dataset (a) and the $\mu\tau$-dataset (b). The nominal value of the $\gamma$-parameters is at 1 and deviations from it indicate a pull. For all other NPs the nominal value is at 0 and the pre-fit $\pm 1\sigma$-variations correspond to $\pm 1$. The post-fit impact on the signal strength $\mu$ (filled coloured bars and the upper horizontal axis) is the basis of the ranking (see Section 10.3). In addition, the pre-fit impacted is shown (empty coloured bars).

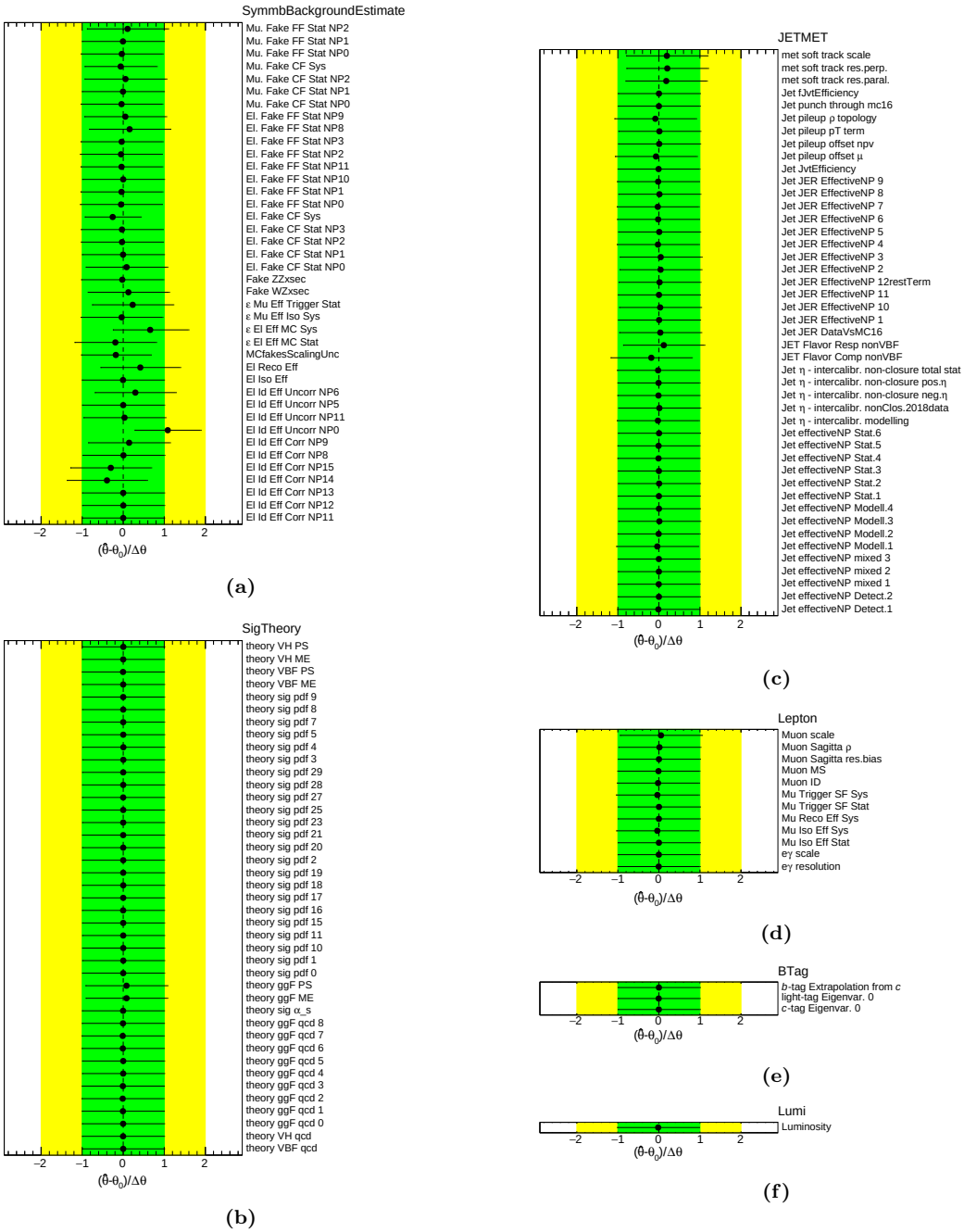



**Figure E.4:** The pulls and constraints of all nuisance parameters obtained from a maximum likelihood fit in the nonVBF SR with the $e\tau$-dataset.

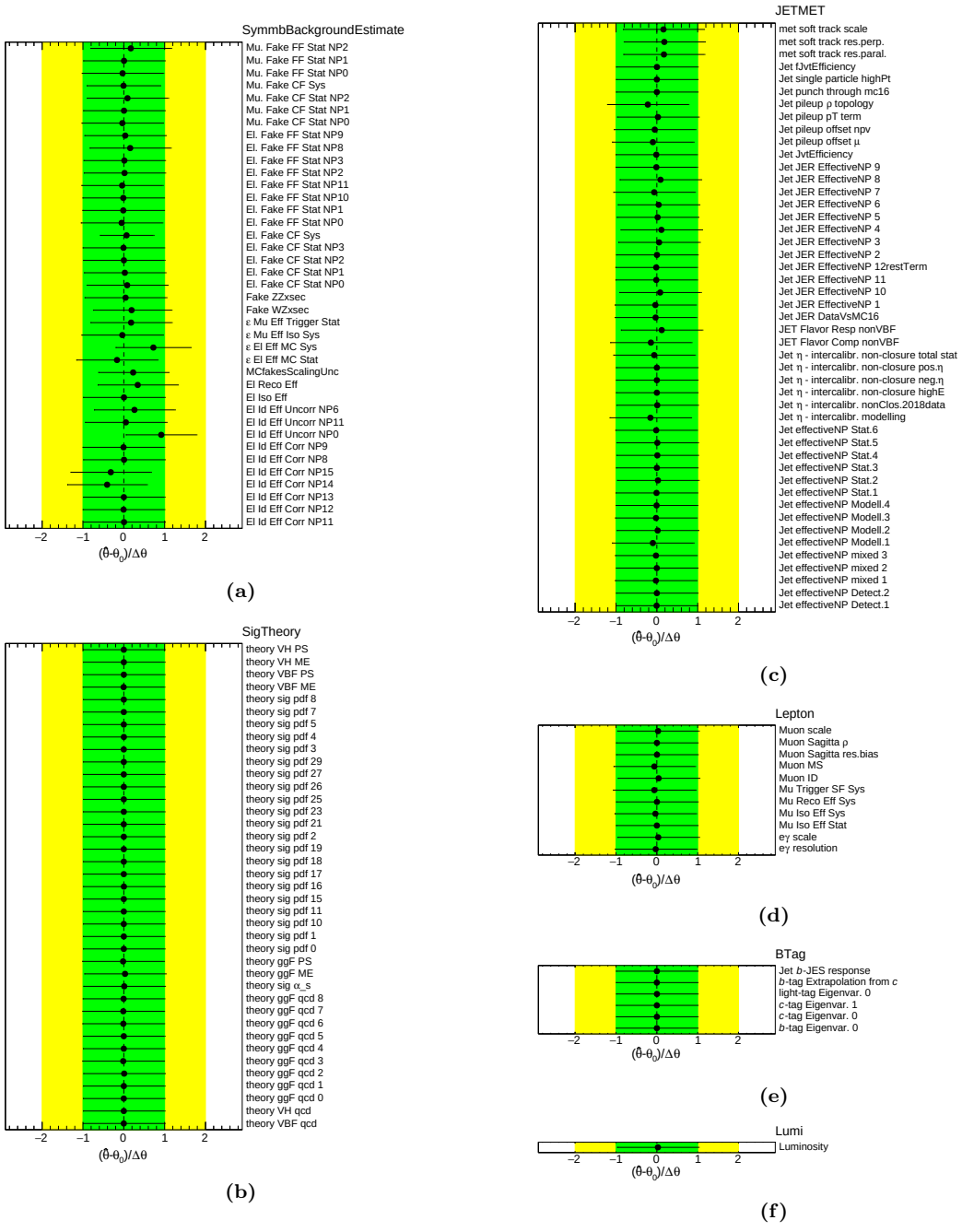



**Figure E.5:** The pulls and constraints of all nuisance parameters obtained from a maximum likelihood fit in the nonVBF SR with the $\mu\tau$-dataset.

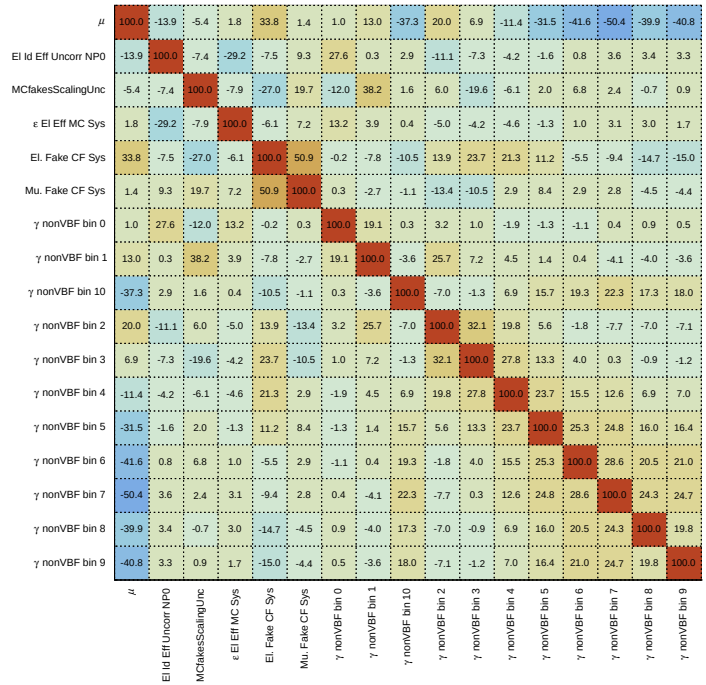



**Figure E.6:** The correlations in percent of all nuisance parameters that have at least one correlation with an absolute value larger than 20 %, obtained from a maximum likelihood fit in the nonVBF SR with the $e\tau$-dataset.

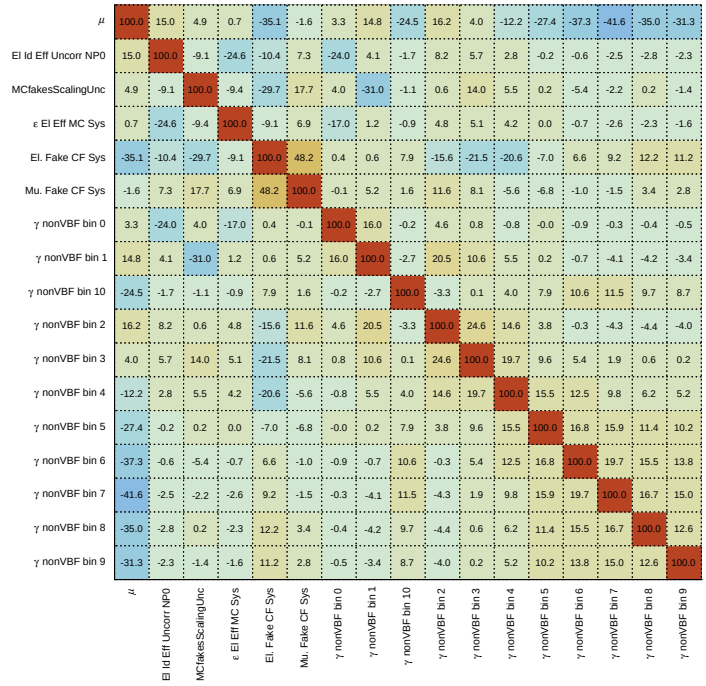



**Figure E.7:** The correlations in percent of all nuisance parameters that have at least one correlation with an absolute value larger than 20 %, obtained from a maximum likelihood fit in the nonVBF SR with the $\mu\tau$-dataset.

## Fit in the VBF SR

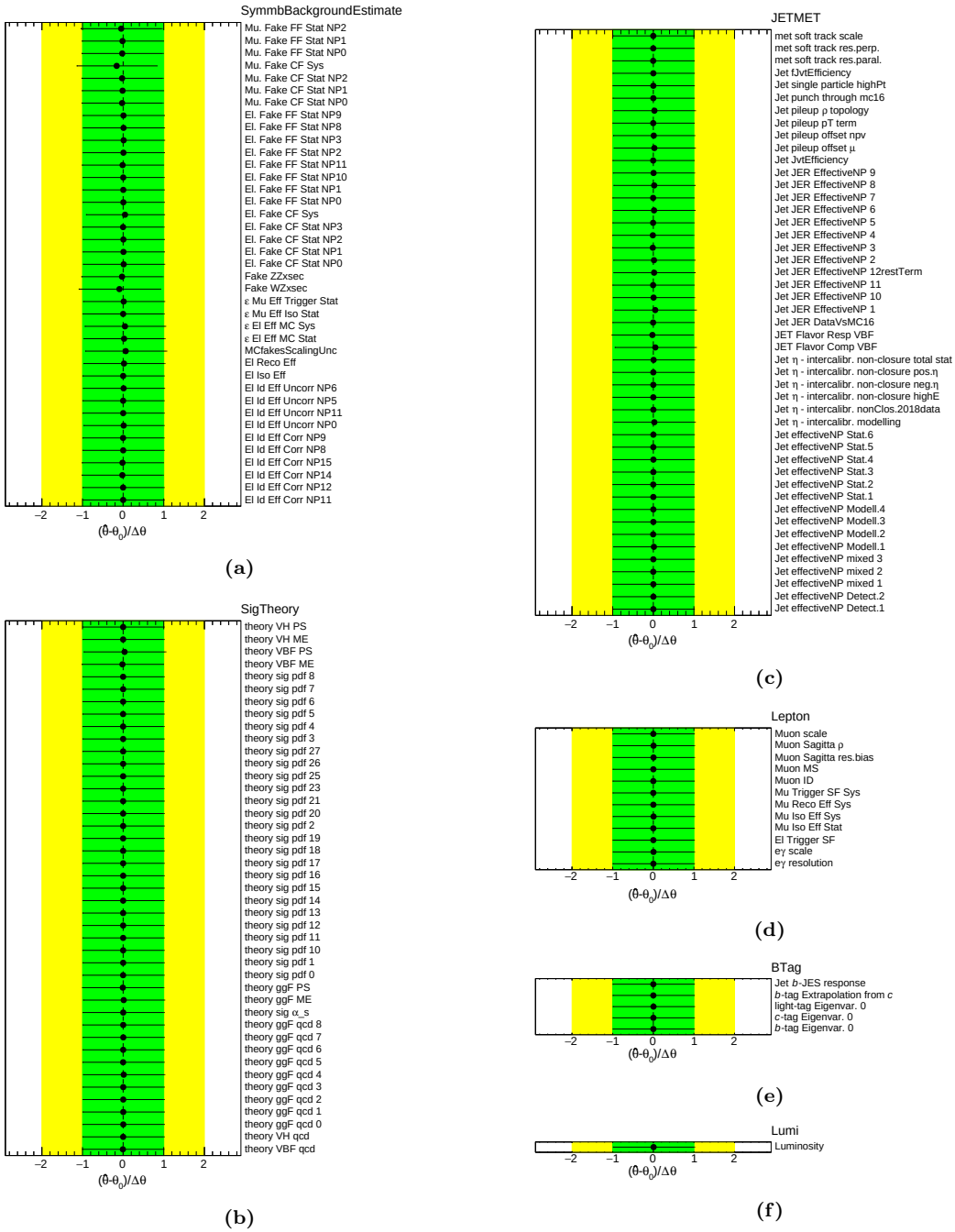
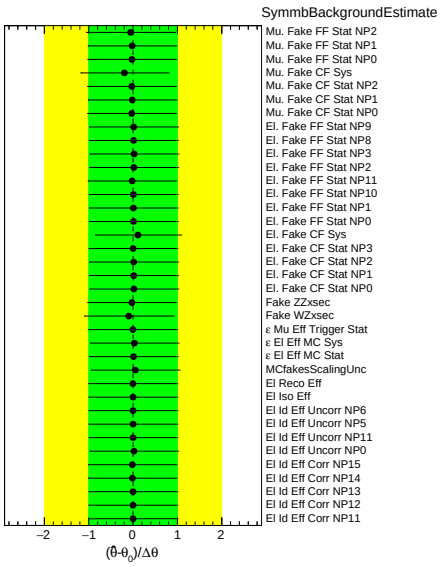
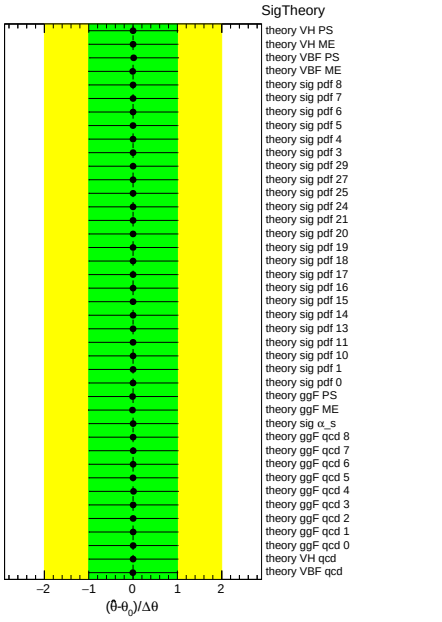


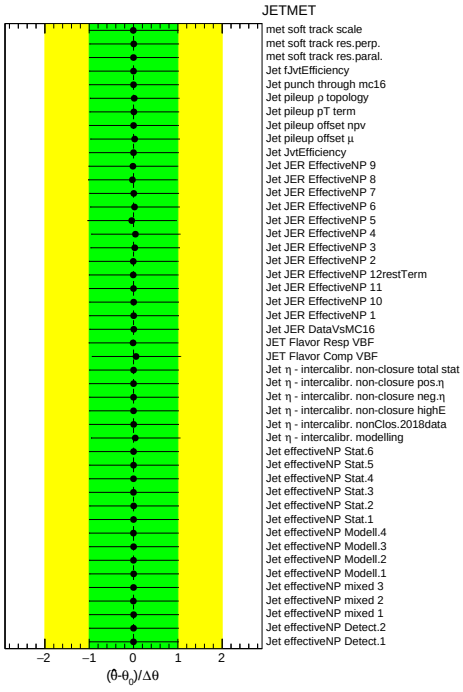

**Figure E.8:** The pulls and constraints of all nuisance parameters obtained from a maximum likelihood fit in the VBF SR with the $e\tau$-dataset.



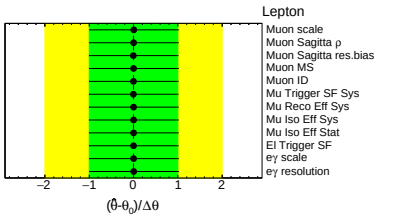

**(a)**



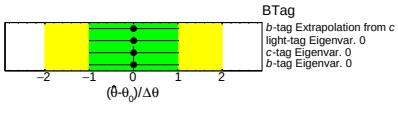

**(b)**



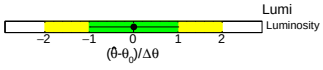

**(c)**




**(d)**




**(e)**




**(f)**

**Figure E.9:** The pulls and constraints of all nuisance parameters obtained from a maximum likelihood fit in the VBF SR with the $\mu\tau$-dataset.

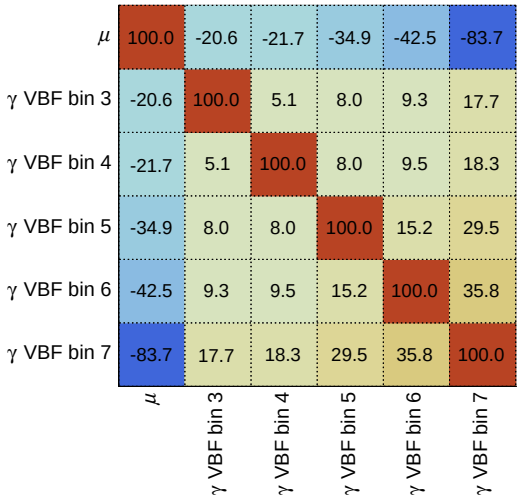



**Figure E.10:** The correlations in percent of all nuisance parameters that have at least one correlation with an absolute value larger than 20 %, obtained from a maximum likelihood fit in the VBF SR with the $e\tau$-dataset.

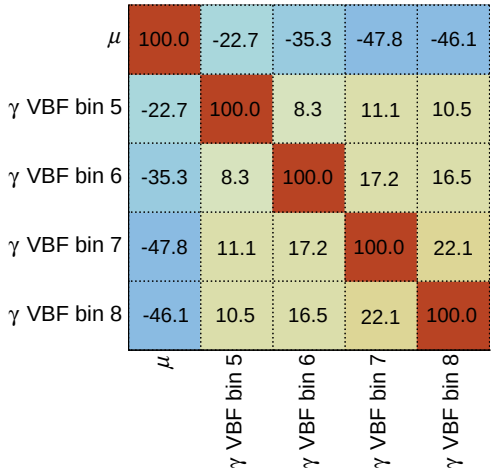



**Figure E.11:** The correlations in percent of all nuisance parameters that have at least one correlation with an absolute value larger than 20 %, obtained from a maximum likelihood fit in the VBF SR with the $\mu\tau$-dataset.

# Appendix F

# Simultaneous Bound on $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$ from Upper Limit on $\mathcal{B}(\mu \to e\gamma)$

The upper limit on the branching ratio $\mathcal{B}(\mu \to e\gamma)$ can be used to derive an upper bound on the combination $Y_{e\tau}Y_{\tau\mu}$ where the $Y_{ij}$ are off-diagonal Yukawa couplings as discussed in Section 1.4.2. The bound on the combination of the couplings can be translated into a bound on the branching ratios $\mathcal{B}(H \to e\tau)$ and $\mathcal{B}(H \to \mu\tau)$ which allows for visualising this bound in Figure 11.8 where the two branching ratios are on either axis. To do so, Eq. 1.54 [18] is used which is repeated here:

$$\Gamma(H \to i\tau) = \frac{m_H}{8\pi}\left(|Y_{i\tau}|^2 + |Y_{\tau i}|^2\right), \tag{F.1}$$

$$\mathcal{B}(H \to i\tau) = \frac{\Gamma(H \to i\tau)}{\Gamma_{\mathrm{SM}}(H) + \Gamma(H \to i\tau)}, \tag{F.2}$$

where $i$ either denotes an electron $e$ or a muon $\mu$. Assuming that $Y_{i\tau} = Y_{\tau i}$, Eq. F.1 becomes

$$\Gamma(H \to i\tau) = \frac{m_H}{4\pi}|Y_{i\tau}|^2, \tag{F.3}$$

and Eq. F.2 can be rearranged as

$$\Gamma(H \to i\tau) = \frac{\mathcal{B}(H \to i\tau)}{1 - \mathcal{B}(H \to i\tau)}\Gamma_{\mathrm{SM}}(H), \tag{F.4}$$

such that the following equation can be obtained:

$$|Y_{e\tau}|^2|Y_{\mu\tau}|^2 = \frac{4^2\pi^2\Gamma^2_{\mathrm{SM}}(H)}{m_H^2} \cdot \frac{\mathcal{B}(H \to e\tau)}{1 - \mathcal{B}(H \to e\tau)} \cdot \frac{\mathcal{B}(H \to \mu\tau)}{1 - \mathcal{B}(H \to \mu\tau)} \tag{F.5}$$

$$\implies \mathcal{B}(H \to \mu\tau) = \frac{(m_H/4\pi)^2 \cdot |Y_{e\tau}|^2|Y_{\mu\tau}|^2 \cdot (1 - \mathcal{B}(H \to e\tau))}{\Gamma^2_{\mathrm{SM}}(H) \cdot \mathcal{B}(H \to e\tau) + (m_H/4\pi)^2 \cdot |Y_{e\tau}|^2|Y_{\mu\tau}|^2 \cdot (1 - \mathcal{B}(H \to e\tau))}. \tag{F.6}$$

In principle, the partial decay width of the respective other LFV decay must also be considered in the denominator of Eq. F.2 but is neglected here. The bound $\sqrt[4]{|Y_{e\tau}Y_{\tau\mu}|^2 + |Y_{\tau e}Y_{\mu\tau}|^2} <$

$2.2 \times 10^{-4} = b$ (Table 1.5) can also be simplified by assuming that $Y_{i\tau} = Y_{\tau i}$:

$$\sqrt[4]{2|Y_{e\tau}Y_{\mu\tau}|^2} < b \tag{F.7}$$

$$\implies |Y_{e\tau}|^2|Y_{\mu\tau}|^2 < \frac{1}{2}b^4\,. \tag{F.8}$$

Inserting Eq. F.8 in Eq. F.6 leads to a term for the red curve in Figure 11.8 which indicates the bound from the limit on $\mathcal{B}(\mu \to e\gamma)$. Within the range displayed in Figure 11.8, there is no visible difference in the curve whether $\Gamma(H \to i\tau)$ is considered in the denominator of Eq. F.2 or not.

Instead of assuming $Y_{i\tau} = Y_{\tau i}$ it could also be assumed that one of both is zero while the other is not, e.g. $Y_{\tau i} = 0$ and $Y_{i\tau} \neq 0$. Then, the $4\pi$ in Eqs. F.3 and F.6 are replaced by $8\pi$ and the factor $1/2$ in Eq. F.8 vanishes. This leads to the blue curve in Figure 11.8.

Many thanks to [263] for support on this topic.

# Acknowledgements

I would like to thank everyone who has helped and supported me throughout the course of this thesis. First of all, I would like to thank Markus for the opportunity to work on my PhD project in his group, for providing an attractive topic centred around a novel concept and for his generous supervision and support. Furthermore, I am grateful for introducing me to Israel.

My gratitude also goes to my direct supervisors Duc, Kathrin and Valerie, who accompanied me in different phases of my thesis. Thanks to Duc for introducing me to the inner workings of the ATLAS collaboration and for his patience with my initially very limited programming experience. Thanks to Kathrin for many fruitful discussions which extended also beyond our time together in Freiburg. Thanks to Valerie for advocating for the *Symmetry Method* and for her invaluable contributions towards bringing the analysis to publication.

I am grateful to my colleagues in Freiburg for their companionship in the office, and for the numerous discussions and support provided, in particular to Dirk, Benjamin, Felix and Alena. Special thanks to Chris for reliably solving all non-scientific problems. Thanks to the computing team who always guaranteed smooth operation of our compute cluster and for rescuing data that was thought to be lost.

I would also like to thank the entire LFV analysis team without whom this thesis would not have been possible. On the one hand, this is the *Symmetry Method*-team, including Shikma, Mattias, Valerie and Markus. On the other hand, these are the people from our partner analyses, including Sergi, Antonio, Michaela, Kieran, Luca, Julia, Michael and Minlin.

I want to thank Serhat and Fabian for making the work on the tau-trigger efficiencies a bit more fun.

A big thanks to Markus, Valerie and Stefan for proof-reading my thesis.

Many thanks to Sophie, who has accompanied me since the beginning of our physics studies and to Timo for sharing the ups and downs of working in the ATLAS collaboration and the time at CERN.

I am thankful beyond words for Stefan, for everything he has done for me, for his support and his patience.

I would like to express my deep gratitude to my parents without whom this entire endeavour would not have been possible.