מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

Thesis for the degree
Doctor of Philosophy

עבודת גמר (תזה) לתואר
דוקטור לפילוסופיה

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

By
Mattias Birman

מאת
מתיאס בירמן

חיפושים מכווני נתונים לפיזיקה מעבר למודל הסטנדרטי המבוססים על נתוני התנגשות באנרגיה גבוהה באמצעות שיטת הסימטריה eμ

Data-driven searches for physics beyond the Standard Model in high energy collision data using the eμ-symmetry method

Advisor:
Dr. Shikma Bressler

מנחה:
ד"ר שקמה ברסלר

July 2022

תמוז - אב ה'תשפ"ב

Acknowledgments

I am very grateful to Shikma Bressler (my advisor) from the Weizmann Institute of Science for the guidance, support, patience, and feedback she gave me throughout my Ph.D. Shikma is a person I greatly admire for many reasons, including her dedication to her research, her openness, and her desire to learn continually and to find innovative approaches which deviate from the norm. I feel gratified and honored to have been advised by her, as it gave me direction, motivation, and the tools to fulfill research that I find meaningful.

I am deeply indebted to Eilam Gross and Yosef Nir (from my advisory committee) from the Weizmann Institute of Science for their encouragement and reassuring backing. I greatly benefited from conversations with them at different stages of my Ph.D. and was touched by how approachable and attentive they are. Eilam also taught me the Elementary Particles course during my first year at the Weizmann Institute of Science. His enthusiasm when teaching piqued my interest in this field, contributing to my choice of experimental High Energy Physics (HEP) for my research. I am also thankful to Noam Tal Hod from the Weizmann Institute of Science and Yoram Rozen from the Technion (from my approval committee) for taking the time to review my research and constructive discussions.

As a member of the ATLAS collaboration at CERN, I completed a large part of my research jointly with other ATLAS members from different institutes. I want to thank in particular Markus Schumacher from the Albert-Ludwigs-Universitaet who closely followed and guided me during most of my Ph.D., as well as his student Katharina Schleicher without whom one of the main efforts presented in this thesis – the ATLAS search for Higgs Lepton Flavor Violating (LFV) decays – couldn't have been completed. I greatly enjoyed our many scientific discussions and their friendly company on the different occasions where we visited Freiburg, where they visited Rehovot, or where we met at CERN near Geneva. I am also very grateful to Valerie Lang from Albert-Ludwigs-Universitaet for her considerable help in the development of the analysis in its final stages and to Duc Bao Ta from Johannes Gutenberg University, Alexander Armstrong and Anyes Taffard from the University of California Irvine, and Kathrin Becker from the University of Warwick for their help in the development of the analysis in early stages. I also thank the whole team of the twin ATLAS Monte Carlo (MC) based analysis for the frequent and insightful joint meetings and collaborative work; and Kristin Lohwasser and Philip Sommer from the University of Sheffield and Julien Maurer from IFIN-HH for valuable discussions related to the electron efficiency measurements.

Special thanks to Benjamin Nachman from the University of California Berkeley for his interest in one of our projects and our fruitful collaboration, which led to the publication in [1].

In my early years of research, I was very fortunate to share my office with Daniel Aloni, Avital Dery, and Aielet Efrati, then more senior Ph.D. students of Yosef Nir. I warmly thank them for the long conversations we frequently held, where I benefitted from invaluable advice and help towards understanding theoretical aspects of the world of elementary particles. I am also very grateful to the other Weizmann Institute of Science students with whom I shared projects, discussed research, or enjoyed relaxing times, particularly Ta-El Coren, Federico De Vito Halevy, Anna Ivina, Gal Sela, and Ophir Turetz.

One significant aspect of my Ph.D., which is not presented here, was the development of a database and website for following the construction of the sTGC (small-strip Thin Gap Chamber) muon detectors included in the ATLAS New Small Wheel, performed towards the completion of my ATLAS qualification task. During this project, I was fortunate to work jointly with the late Daniel Lellouche from the Weizmann Institute of Science, who was inspiring and friendly. I also worked closely with David Front from the Weizmann Institute of Science, Kim Heidegger from Albert-Ludwigs-Universitaet, and the sTGC construction and development team in the Mexico building at the Weizmann Institute of Science, for which I am thankful.

Declaration

I declare that this thesis is an original report of my research, has been written by me, and has not been submitted for any previous degree. I confirm that the work submitted is my own, except for collaborative contributions, which have been indicated clearly and acknowledged below. Due references have been provided on all supporting literature and resources.

The work presented in sections 3, 4, and 5.5.3 uses data (measured or simulated) obtained from proton-proton (*pp*) collisions delivered by the Large Hadron Collider (LHC) and measured by the ATLAS detector at CERN. The "search for Higgs LFV decays" example study presented in section 3 was performed in collaboration with Katharina Schleicher from Albert-Ludwigs-Universitaet, during the development of the ATLAS full Run-2 same search presented in section 4. Completing this analysis was performed in collaboration with other ATLAS members. I only describe in detail the parts of the analysis in which I was directly implicated: the development of the analysis method, including the event classification and selection, the efficiency correction, the background estimation, and the development of the statistical model (section 4.3); the measurement of electron efficiencies (section 4.4); the fake composition study for estimating systematic uncertainties on the data-driven fake background (section 4.5.2); the fake contribution estimated from MC (section 4.5.3); validation of the different background estimates (section 4.6); a study on uncertainty preprocessing (section 4.9.1). Efforts towards estimating the data-driven fake background contribution (section 4.5.1) were mainly provided by Emanuel Dorbath and Valerie Lang from the Albert-Ludwigs-Universitaet. And the implementation of a Neural Network (NN) for enhancing the sensitivity (section 4.7), the inclusion of systematic uncertainties (section 4.8), as well as implementing the statistical analysis for the obtention of results (sections 4.9.2 and 4.9.3) is mainly thanks to Katharina Schleicher. Other contributors to this analysis are Kathrin Becker from the University of Warwick, Shikma Bressler and Ta-El Coren from the Weizmann Institute of Science, and Markus Schumacher from the Albert-Ludwigs-Universitaet.

Moreover, the final results of the ATLAS Run-2 Higgs LFV search are combined results of the analysis presented here with those of two other MC-based searches for the same signal. Although the combined results are summarized (section 4.9.3), the MC-based analyses aren't presented. In addition, these results are still unpublished at the time where this is written, although the paper is ready and going through an internal ATLAS review prior to being submitted for publication. As such, they may be subject to small changes during this process. The contributors to the MC-based analyses are Kieran Robert Amos, Xin Chen, Kwok Lam Chu, Antonio De Maria, Luca Fiorini, Luis Flores Castillo, Antonio Jesus Gomez Delegido, Julia Mariana Iturbe Ponce, Dylan Perry Kisliuk, Michaela Mlynarikova, Robert Orr, Hao Pang, Joseph Patton, Gianantonio Pezzullo, Sergi Rodriguez Bosca, Daniel Scheirich, Duc Bao Ta and Minlin Wu.

The work presented in section 5.2 is the outcome of a collaborative work published in [1]. In particular, the idea of quantifying asymmetries based on 2D histograms and using the N_σ test is thanks to Ophir Turetz from the Weizmann Institute of Science; and the NN implementation (described in section 5.2.4) was solely developed by Benjamin Nachman from the University of California Berkeley. My contributions include the development of the statistical analysis, the data preparation, and the obtention of results with the N_σ test. Other contributors to this study are Shikma Bressler, Ta-El Coren, Raphael Sebbah, and Gal Sela from the Weizmann Institute of Science.

The production of standalone MC simulated samples, described in section 5.4.1, was performed in collaboration with Ta-El Coren and Ophir Turetz.

Contents

Abbreviations

ATLAS	A Toroidal LHC ApparatuS
AUC	Area Under Curve
BR	Branching Ratio
BSM	Beyond Standard Model
CERN	European Organization for Nuclear Research
CF	Correction Factor
CL	Confidence Level
CMS	Compact Muon Solenoid
CR	Control Region
DAOD	Derived Analysis Object Data
DDP	Data-Directed Paradigm
EFT	Effective Field Theory
EM	ElectroMagnetic
FCCC	Flavor Changing Charged Current
FCNC	Flavor Changing Neutral Current
FF	Fake-Factor
FSR	Final State Radiation
ggF	gluon-gluon Fusion
HEP	High Energy Physics
ID	Inner Detector
Id	Identification
Iso	Isolation
LFC	Lepton Flavor Conservation
LFV	Lepton Flavor Violation
LHC	Large Hadron Collider
LO	Leading Order
LU	Lepton Universality
LUV	Lepton Universality Violation
MC	Monte Carlo
MFV	Minimal Flavor Violation
MS	Muon Spectrometer
NN	Neural Network
NP	New Physics
PDF	Probability Distribution Function
POI	Parameter Of Interest
pp	proton-proton
Reco	Reconstruction
ROC	Receiver Operating Characteristic
SF	Scale-Factor
SM	Standard Model
SR	Signal Region
stat	statistical
syst	systematic
VBF	Vector Boson Fusion
VEV	Vacuum Expectations Value
VLL	Vector-Like Lepton
VR	Validation Region

Abstract

In pursuit of New Physics (NP) to extend the Standard Model (SM), no effort is spared, as evidenced by the LHC and its experiments at CERN, the world's largest scientific enterprise. But with no conclusive evidence of deviations from SM predictions, novel and complementary approaches for discovery should be considered. The amount of LHC data recorded is the greatest ever accumulated and is largely unexplored. Hence the proposed data-directed paradigm (DDP); letting the data itself guide us towards its regions of interest, significantly enhancing our discovery potential. Symmetries of the SM can be utilized for this endeavor, as they provide two exclusive datasets that can be compared without requiring simulation. Focusing on the data, selections which exhibit significant asymmetry can be identified efficiently and marked for further study. An example is the approximate symmetry between electrons and muons in SM processes. This property can be exploited to probe the data for deviations from SM features such as Lepton Universality (LU) or Lepton Flavor Conservation (LFC). To this end, the e/μ-symmetry method is developed. This analysis method provides a data-driven approach to estimate SM background contributions. It involves, in particular, a technique to account for detection effects that invalidate the expected SM symmetry and a statistical analysis procedure. A direct search for the LFV decays of the Higgs boson $H \to e\tau$ and $H \to \mu\tau$ using the e/μ-symmetry method is presented, which uses the Run-2 dataset of pp collisions delivered by the LHC at $\sqrt{s} = 13$ TeV and recorded by the ATLAS detector, corresponding to an integrated luminosity of 138.42 fb^{-1}. The final results are combined results of this analysis with two other MC-based searches for the same signals. Some tension from SM predictions is observed at the level of 2.5σ in the search for $H \to \mu\tau$, while no evidence of the $H \to e\tau$ decay is found. Upper limits on the branching ratios (BR) are set at a 95% confidence level (CL): 0.230% (0.192%) for $H \to e\tau$ and 0.163% (0.182%) for $H \to \mu\tau$ when the two searches are conducted independently (simultaneously). The successful completion of this analysis is an important endorsement of the e/μ-symmetry method. However, only a specific signal is searched for in a small theoretically motivated data region. In terms of the DDP proposed, the broadest possible region is scanned for any deviation from SM predictions. Implementing such data-directed and generic searches based on symmetries of the SM is still at an initial stage. Towards laying the groundwork, a generic approach to identify asymmetries between two measurements is developed in a simplified framework. With little optimization, the sensitivity achieved is only slightly lower than that of optimal likelihood-based tests, which have full knowledge of the signal. This approach has the advantage of being extremely fast, enabling to efficiently scan large portions of the data for any hints of NP.

1 Introduction

The SM encapsulates our best understanding of the elementary particles and their interactions, including the description of three of the four known fundamental laws in the universe (electromagnetic, weak, and strong interactions, omitting gravity). Yet, we are still confronted with unresolved puzzles it does not explain, e.g., how neutrinos acquire their mass, the true nature of dark matter, or the matter-antimatter asymmetry. As such, the SM can only be regarded as a low-energy, effective incarnation of a more global theory that eludes us so far. Hence, one of the main goals of the LHC program at CERN is to discover physics Beyond the SM (BSM), which would provide hints toward uncovering the more fundamental theory that encompasses the SM. But despite hundreds of searches for BSM physics that have been conducted at the LHC, particularly in the general-purpose experiments, ATLAS and CMS, no significant deviation from the SM predictions have been found. The SM agrees with most experimental results, often with utmost precision.

Most of these searches have been conducted based on predictions from proposed theoretical extensions of the SM, which aim to resolve some of the remaining puzzles. Some popular examples are supersymmetry, models with extra dimensions, or models with more than one Higgs-like neutral boson. Such proposed theories predict the presence of signatures within the accumulated data that differ from SM-only physics. These are then used as the basis for conducting searches, where the searched-for signal is well known, and only a restricted region of the observables space (the space spanned by the observables of the measured data) is probed. In addition, these searches are generally conducted following the "blind analysis" paradigm, where the data is only looked at in the last step of the analysis after most of the efforts and time have been invested. Priority is given to well-motivated SM extensions, which predict features possibly observable within the energy scale of LHC collisions ($E \sim 1$ TeV). But given the many searches already conducted, the more popular extensions are highly constrained.

The remaining potential BSM signatures which haven't been experimentally searched for are many, significantly more numerous than those that have. In addition, the amount of accumulated data is by far the largest ever available and remains largely unexploited; only a small portion of the entire observables space has been probed. But without the theoretical guidance which has, until today, steered the choice of experimental searches, there is no particular motivation to conduct one search before another. As a result, more and more model-independent searches are being conducted, with the intention to cover more and more regions of the observables space. Still, these searches are resource-intensive tasks with limited sensitivity or probe only limited regions. To tackle this problem, we pursue an efficient method for scanning the data in wide ranges of the observables space, intending to find regions more likely to include evidence of BSM physics. Such regions, when found, would be marked for further study using traditional data analysis methods. Complementary to the "blind analysis" paradigm, where the targeted region of the observables space is selected before conducting the analysis, this proposed DDP looks upon the data itself to steer the choice of searches to be performed.

In this context, we develop the e/μ-symmetry method, a data-driven analysis method to search for BSM physics in HEP experimental data. Electrons (e) and muons (μ) are two charged leptons, each carrying a different *flavor* (generation) number. They are among the most efficiently reconstructed particles by detectors in accelerator experiments. In the SM theory, there is an approximate symmetry between processes that lead to final states including electrons and muons, which derives from LU. In addition, the accidental LFC SM symmetry forbids interactions that don't conserve lepton flavor such as $\mu \to e\gamma$ (where γ is a photon). Both LU and LFC derive from special features of the SM, meaning that they don't necessarily hold in the fundamental theory of nature[1]. As a result, finding experimental evidence that they are not conserved could have far-reaching implications. In fact, the strongest direct evidence that the SM is an incomplete theory – the discovery of neutrino masses and oscillation – entails that LFC is not preserved in nature. However, conclusive evidence from within accelerator experiments still eludes us. The e/μ-symmetry method exploits this approximate electron/muon symmetry to search for deviations from SM predictions in a data-driven way. It is sensitive to a wide range of signatures from potential BSM processes, including violations of LU or LFC. Since it doesn't rely on the simulation of background processes, it can be used to rapidly scan large portions of the observables space, efficiently uncovering hints of BSM physics hidden in the accumulated data.

The development, implementation, and application of the e/μ-symmetry method in different contexts are at the core of the research presented in this thesis. In particular, its application in a direct search for Higgs LFV decays, using LHC data collected by the ATLAS detector, is presented. The rest of this thesis is organized as follows. In section 2, the theoretical framework is described, focusing on the SM properties of LU and LFC. Motivation towards searching for evidence that they are not conserved is advanced, relying on existing experimental results and describing some examples of SM extensions that allow them to be violated. In addition, a description of the LHC and the ATLAS experiment is given and of the data analysis methods used. In section 3, the e/μ-symmetry method is described in detail, and different implementations are compared. The example of the Higgs LFV search is introduced based on partial ATLAS Run-2 simulated data. The ATLAS full Run-2 Higgs LFV analysis using the e/μ-symmetry method is presented in section 4. The complete ATLAS full Run-2 Higgs LFV analysis combines the analysis presented here with two MC-based searches for the same signal. The combined results are summarized, although the MC-based analyses aren't described. In section 5, various studies are presented that focus on implementing the e/μ-symmetry method in data-directed and generic searches. In particular, the results reported in [1] are described, and an alternative approach for symmetry restoration within the e/μ-symmetry method is considered. A summary of the research conducted is given in section 6, where conclusions are drawn.

[1] LU is not a symmetry of the SM since Higgs couplings are not flavor universal. Still, it leads to an approximate lepton symmetry which can be precisely tested

2 Scientific background

This section introduced the scientific background related to the research presented in this thesis. In section 2.1, the SM of particle physics is summarized, describing, in particular, the laws of LU and LFC. Performing searches for violations of these conservation laws is motivated in section 2.2, examples of theories extending the SM where they are violated are presented in section 2.3, and in section 2.4, current bounds found from previous searches are listed.

In addition, the LHC and the ATLAS experiment are described in section 2.5, and statistical analysis methods standardly used in searches for BSM physics are described in section 2.6.

2.1 The Standard Model, Lepton Universality, and Lepton Flavor Conservation

The SM is a theory of the elementary particles and their interactions that continually provides successful experimental predictions [2]. It is defined - in the Lagrangian formalism - by its fundamental local symmetry described in (1), a set of fermion fields, and the Higgs field.

$$G_{SM} = SU(3)_C \times SU(2)_L \times U(1)_Y \tag{1}$$

The local symmetry (1) breaks spontaneously by the vacuum expectations value (VEV) of the single Higgs scalar as follows:

$$G_{SM} \xrightarrow{SSB} SU(3)_C \times U(1)_{EM} \tag{2}$$

The SM fermions, quarks or leptons, are described in five different representations of the gauge group (1):

$$L_L(1,2)_{-1/2}, \quad E_R(1,1)_{-1}, \quad Q_L(3,2)_{+1/6}, \quad U_R(3,1)_{+2/3}, \quad D_R(3,1)_{-1/3} \tag{3}$$

where the subscripts $_{L/R}$ indicate the chirality of the fields (left- or right-handedness). Each comes in three flavors (generations) and is either an $SU(2)$ doublets (L_L and Q_L) or singlet (E_R, U_R, and D_R), depending on its chirality. We denote their $SU(2)$ components as:

$$L_{iL} = \begin{pmatrix} \nu_{iL} \\ \ell_{iL} \end{pmatrix}, \quad E_{iR} = \ell_{iR}, \quad Q_{iL} = \begin{pmatrix} u_{iL} \\ d_{iL} \end{pmatrix}, \quad U_{iR} = u_{iR}, \quad D_{iR} = d_{iR} \tag{4}$$

where $i = 1, 2, 3$ indicates the flavor.

After electroweak symmetry breaking, they are grouped into four types:

- up-type quarks: $u_{1,2,3} = u, c, t$ - (up, charm, top)
- down-type quarks: $d_{1,2,3} = d, s, b$ - (down, strange, bottom)
- charged leptons: $\ell_{1,2,3} = e, \mu, \tau$ - (electron, muon, tau)
- neutral leptons: $\nu_{1,2,3} = \nu_e, \nu_\mu, \nu_\tau$ [2] - (electron neutrino, muon neutrino, tau neutrino)

[2] We only consider the interaction eigenstates since the neutrinos are massless in the SM

The term "flavor" distinguishes between the different copies of the same gauge representation, namely fields with common quantum numbers, although they correspond to particles with distinct masses. Antiparticles share the flavor of their counterparts but with an opposite flavor charge (similarly to the electromagnetic (EM) charge). Interactions between the SM fermions are mediated by different bosons, each responsible for a specific force type. These interactions can be either "flavor universal" - with couplings proportional to the unit matrix in flavor space, "flavor diagonal" - with couplings that are diagonal in flavor space, or "flavor changing" - which can allow for a different "flavor-number" (number of particles minus number of antiparticles of a given flavor) in the initial and final states. The different fermion SM interactions are summarized in Table 1.

Interaction	Force carrier	Symbol	Fermions	Flavor
EM	Photon	A^0	u, d, ℓ	universal
Strong	Gluon	g	u, d	universal
Weak Neutral Current	Z boson	Z^0	u, d, ℓ, ν	universal
Weak Charged Current	W boson	$W^\pm$	$u\bar{d}/\ell\bar{\nu}$	changing/universal
Yukawa	Higgs	h	u, d, ℓ	diagonal

Table 1: The SM fermion interactions.

In the leptonic sector, only couplings mediated by the Higgs boson are non-universal since they are proportional to the fermion's masses. The universality of the photon and gluon couplings results from the gauge invariance in (2) and thus holds in any model extending the SM. On the other hand, the Z couplings to leptons are universal due to a special feature of the SM; all leptons of a given charge and chirality come from the same $SU(2)_L \times U(1)_Y$ representation. The only possible flavor-changing processes are mediated by the $W^\pm$ boson and are referred to as "flavor changing charged current" (FCCC) interactions. They involve up-type and down-type flavors or charged lepton and neutrino flavors. Some examples are $\mu \to e\bar{\nu}_e\nu_\mu$, $s\bar{u} \to \mu^-\bar{\nu}_\mu$, or $b \to c\bar{c}s$. Note that in the leptonic sector, the charged-current weak interaction is considered universal since the W couplings to $\tau\bar{\nu}_\tau$, $\mu\bar{\nu}_\mu$, and $e\bar{\nu}_e$ are equal. This diagonality follows from the local symmetry in (1). Note also that these FCCC processes do not violate the SM's law of LFC (no flavor changing processes). Indeed, LFC does not distinguish between lepton flavor and lepton family number (meaning that it interprets the charged lepton's flavor and its neutral counterpart's as the same). "Flavor changing neutral current" (FCNC) processes do not occur in the SM at tree level and are usually highly suppressed.

The fact that lepton flavor is always conserved is an "accidental symmetry" of the SM (meaning that it isn't a law used to define the theory, but it is preserved by all the renormalizable terms of its lagrangian). Other accidental symmetries are the conservation of the number of leptons or baryons. Equation (5) summarizes all the accidental symmetries of the SM: baryon number, electron flavor, muon flavor, and tauon flavor conservation.

$$G_{SM}^{global} = U(1)_B \times U(1)_e \times U(1)_\mu \times U(1)_\tau \tag{5}$$

A particular process will be considered LFV if it breaks at least one of the leptonic global symmetries of (5). Similarly, we denote as LU Violating (LUV) any process which is non-universal in the leptonic sector.

2.2 Hints of Lepton Flavor / Universality Violation

2.2.1 Neutrino oscillations

Violation of the lepton flavor symmetry has already been experimentally determined by the discovery of neutrino mass and oscillations [3].

In the SM, fermions masses arise after the spontaneous symmetry breaking from the Yukawa interactions mediated by the Higgs boson. In the leptonic sector, these interactions mix between doublets of the $SU(2)_L$ group $L_L = \binom{\nu_{L\ell}}{\ell_L}$ and its corresponding singlet $E_R = \ell_R$, with ℓ standing for e, μ, τ, and subscripts $_L$ $(_R)$ standing for left-handed (right-handed), describing the particle's chirality. These interaction terms are given in (6), where the i and j indices run on the possible flavors $e, \mu,$ and τ.

$$-L^{\ell}_{Yukawa} = Y^{\ell}_{ij}\bar{L}_{Li}\phi E_{Rj} + H.c. \tag{6}$$

In the mass basis, the coupling matrix Y^{ℓ} becomes diagonal after the spontaneous symmetry breaking. Furthermore, the acquired VEV of the SM Higgs field being $\langle\phi\rangle = \binom{0}{v/\sqrt{2}}$, none of these terms includes neutrinos. Therefore, the SM predicts the neutrinos to be massless. Nevertheless, many SM extensions allow for non-zero neutrino masses by explicitly breaking these symmetries (see, for example, [4–7]).

In the SM Effective Field Theory (EFT) formalism described further in (16), neutrino mass is introduced via the seesaw terms [8]:

$$-L^{dim-5}_{seesaw} = \frac{Z^{\nu}_{ij}}{\Lambda}\phi\phi L_{L_i}L_{L_j} + h.c. \tag{7}$$

These are dimension five terms (the lowest dimension of non-renormalizable terms one can add) that break lepton flavor and total lepton number. The fermionic fields included are all SM fields, but the neutrino mass terms are Majorana mass terms, and the lepton number conservation is violated by 2. These lead, upon spontaneous symmetry breaking, to neutrino masses:

$$M^{\nu}_{ij} = \frac{Z^{\nu}_{ij}\, v^2}{2\,\Lambda} \tag{8}$$

Moreover, since these terms break the lepton flavor symmetry, they also permit lepton mixing.

Neutrino masses can also be achieved by adding new fields, called sterile neutrinos, that don't couple to any of the SM force fields. They couple to both the Higgs field and a SM "active" neutrino through the following set of terms:

$$Y^{\nu}_{ij}\bar{L}_{Li}\tilde{\phi}\nu_{sj} \tag{9}$$

which give rise, after spontaneous symmetry breaking, to a Dirac mass term. In addition, a bare mass term including only the sterile neutrinos can be added, which will be a Majorana mass term. These are explicated in the following:

$$-L_{M_\nu} = M_{Dij}\bar{\nu}_{Li}\nu_{sj} + \frac{1}{2}M_{Nij}\bar{\nu}^{c}_{si}\nu_{sj} + h.c. \tag{10}$$

Whether the neutrino mass terms are Majorana or Dirac, neutrino oscillation follows from the charged current interactions with the $W^{\pm}$ field, described by:

$$-L^{\ell}_{CC} = \frac{g}{\sqrt{2}}\begin{pmatrix}\bar{e}_L & \bar{\mu}_L & \bar{\tau}_L\end{pmatrix}\gamma^{\mu}U\begin{pmatrix}\nu_1\\ \vdots \\ \nu_n\end{pmatrix}W^{+}_{\mu} \tag{11}$$

where n is the number of active+sterile neutrinos. U is the mixing matrix, which can be expressed in terms of the diagonalizing matrices of the mass matrices from (10). In this set of terms, the neutrinos are in their mass eigenstates. So the weak eigenstates ν_α produced in a weak interaction are, in general, linear combinations of the mass eigenstates ν_i:

$$|\nu_\alpha(t)\rangle = \sum_{i=1}^{n} U^*_{\alpha i} |\nu_i(t)\rangle \tag{12}$$

Thus the probability of detecting the neutrino, produced in a state ν_α, in a state ν_β is given by:

$$P_{\alpha\beta} = \left|\left\langle \nu_\beta \middle| \nu_\alpha(t) \right\rangle\right|^2 = \left| \sum_{i=1}^{n} \sum_{j=1}^{n} U^*_{\alpha i} U^*_{\beta j} \left\langle \nu_j(0) \middle| \nu_i(t) \right\rangle \right|^2 \tag{13}$$

Neutrino oscillations explicitly violate the conservation of lepton flavor. They can be detected in the charged-current interaction $\nu_\alpha(t)N' \to \ell_\beta N$ by comparing the measured flux to the total expected flux. The first evidence for such processes was detected in the Homestake gold mine using a chlorine detector [9]. In this experiment, the flux of ν_e solar neutrinos was measured to be smaller than expected. Since then, multiple experiments, either solar [10], atmospheric [11], or beam-like [12, 13], have confirmed neutrino oscillations with a very high ($> 5\sigma$) significance.

2.2.2 B decay anomalies

The summary presented in this section relies mainly on the review presented in [14].

Many experimental results from various experiments that probe LU between the first two lepton families (electron and muon) are in excellent agreement with theoretical predictions based on the SM. For example, the measured ratio $\Gamma(Z \to \mu\mu)/\Gamma(Z \to ee)$ agrees with the SM and LU with a precision within 0.3% [14–16]. Other (most precise) examples include: electroweak $W^- \to \ell^-\bar{\nu}_\ell$ (agreement within 0.8%) [14, 17, 18], pseudoscalar decays $K^- \to \ell^-\bar{\nu}_\ell$ [14, 19, 20] and $\pi^- \to \ell^-\bar{\nu}_\ell$ [14, 21] (agreement within 0.2%), τ decays $\tau^- \to \ell^-\bar{\nu}_\ell\nu_\tau$ [14, 17] (agreement within 0.14%), and quark resonance decays $J_{/\psi} \to \ell\ell$ [14] (agreement within 0.31%). Similar tests involving the third lepton family (τ) are generally less precise. This is due to the complex decay of the τ which makes its reconstruction more challenging (τ leptons are more massive – 17 and 3500 times heavier than the muon and electron – and decay close to the interaction point due to their short lifetime of 0.3 ps). Only in Z boson decays the constraint on LUV is as precise as with the first two families [15] (agreement within 0.3%). With W boson decays, for example, it is only within 3%, an order of magnitude less precise, and manifests a tension with the SM expectation at the level of 2.6σ [18]. Still, no strong indication of LUV is observed in most experimental results.

B decay anomalies are an exception, where significant tension with SM predictions has been observed. These anomalies refer to two distinct classes of measurements that involve semileptonic decays of the B hadron, each probing a different decay of b quarks to lighter quarks and leptons:

- $b \to c\ell^-\bar{\nu}_\ell$
- $b \to s\ell^+\ell^-$

In the first case ($b \to c\ell^-\bar{\nu}_\ell$), the transition involves FCCCs, which occur at tree level in the SM (mediated by virtual W bosons). The measurements discussed relate to B mesons ($q\bar{b}$ with $q = u, d, s, c$) decaying to $D^{(*)}$ mesons ($c\bar{q}$ with $q = u, d, s$ and where $(*)$ indicates if it is an excited state in terms of total angular momentum), and compare the ratios in (14) to SM predictions:

$$R_{D^{(*)}}(\ell, \ell') = \frac{\Gamma(B \to D^{(*)}\ell\bar{\nu}_\ell)}{\Gamma(B \to D^{(*)}\ell'\bar{\nu}'_\ell)} \tag{14}$$

In fact, $R_{D^{(*)}}(e, \mu)$ measurements which involve only the first two lepton families are in agreement with SM predictions and LU, such that $B \to D^{(*)}e\bar{\nu}_e$ and $B \to D^{(*)}\mu\bar{\nu}_\mu$ decays are assumed free of NP contributions and are used for the measurements of the $b \leftrightarrow c$ couplings. The tension arises in the measurements involving the third lepton family of $R_{D^{(*)}}(\tau, e \text{ or } \mu)$, where the decays involving electrons and muons are averaged in the denominator. The experimental world averages – combining results from the LHCb, Belle, and BaBar experiments – are 0.340 ± 0.030 (0.295 ± 0.014) for R_D (R_{D^*}), which exceeds the expected SM value of 0.299 ± 0.003 (0.285 ± 0.005) by 1.4σ (2.5σ). This corresponds to a combined disagreement at the level of 3.1σ [22]. Since the τ is much heavier than the electron or muon, we can speculate that B decays to the third generation are more sensitive to the presence of NP. As such, most theoretical studies which attempt to account for this discrepancy consider additions to $b \to c\tau\bar{\nu}_\tau$.

The second case ($b \to s\ell^+\ell^-$) involves FCNCs, which are forbidden at tree level in the SM but can occur at the loop level. These are, therefore, much rarer decays than in the first case. The measurements discussed relate to B mesons decaying to $K^{(*)}$ mesons ($q\bar{s}$ with $q = u, d$) and compare the ratios in (15) to SM predictions:

$$R_{K^{(*)}}(\mu, e) = \frac{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma(B\to K^{(*)}\mu^+\mu^-)}{dq^2}}{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma(B\to K^{(*)}e^+e^-)}{dq^2}} dq^2 \tag{15}$$

where q^2 is the invariant mass squared of the dilepton system integrated between $q^2_{\min}$ and $q^2_{\max}$. In this case, the only sensitive measurements involve the first two lepton families due to the rarity of the process considered and the difficulty of reconstructing τ leptons. The most precise measurements of these ratios are from the LHCb experiment, finding $0.846^{+0.044}_{-0.041}$ ($0.69^{+0.12}_{-0.09}$) for R_K (R_{K^*}) – measured in the range $1.1 < p^2 < 6.0\,\text{GeV}^2$ – in tension with the SM expected value of 1.00 ± 0.01 at the level of 3.1σ (2.5σ) [23, 24] (similar but less precise measurements from the Belle and BaBar experiments are compatible with the SM expectation). These results are more surprising than the tension found in the $R_{D^{(*)}}$ measurements since it involves the first two lepton families. Attempts to account for this discrepancy consider NP additions which modify the $b \to s$ coupling.

The B decay anomalies refer to the tension observed between experimental measurements of the $R_{D^{(*)}}$ and $R_{K^{(*)}}$ ratios with their SM predictions. Both experimental and theoretical values can be estimated precisely since the dominant uncertainties cancel in the ratios[3]. These tensions are evidence of LUV and hint at NP which modify the $b \to c\ell^-\bar{\nu}_\ell$ or $b \to s\ell^+\ell^-$ couplings. Various models extending the SM which account for these discrepancies have been proposed, for example, with the addition of a new neutral heavy boson or leptoquarks. Relevant to the research presented in this thesis, these anomalies could also be linked to LFV decays of the Higgs or other massive resonances [25–27].

[3] The experimental systematic uncertainties are effectively reduced by normalizing the two components in each ratio with reference measurements, namely $B \to D^{(*)}3\pi$ and $B \to K^{(*)}J_{/\psi}(\to \ell\ell)$ for (14) and (15) respectively.

2.3 Lepton Flavor Violation in Standard Model extensions

2.3.1 Physics beyond the Standard Model

The SM is not a complete theory of nature but rather a successful (low energy) EFT, valid up to some undetermined cut-off energy scale Λ. When searching for BSM physics, the hope is to find evidence for effects present in the theory with a finite value of Λ but disappear in the limit $\Lambda \to \infty$. Such BSM effects can potentially violate the accidental symmetries of the SM. In the EFT formalism, they are virtually introduced by adding non-renormalizable higher-order terms to the SM Lagrangian:

$$L_{EFT} = L_{SM} + \sum_i \frac{c_i^{(5)}}{\Lambda} O_i^{(5)} + \sum_i \frac{c_i^{(6)}}{\Lambda} O_i^{(6)} + \ldots \tag{16}$$

where the $O_i^{(D)}$ are operators of dimension D and the $c_i^{(D)}$ are parameters called the Wilson coefficients. The $O_i^{(D)}$ must be invariant under the fundamental symmetry of the SM but can violate any of its accidental symmetries.

Relevant to our work, off-diagonal Higgs couplings to leptons would violate the lepton flavor symmetries introduced in (5). Such effect is not observed in the SM since (accidentally) the Yukawa and mass matrices are simultaneously diagonalized. However, flavor violating Higgs couplings arise naturally in different proposed extensions of the SM and can be introduced in different ways. Some examples are described in more detail below; in the 2HDM model, this is achieved by introducing a second Higgs scalar field (see, for instance, [28, 29]); the addition of Vector-Like Leptons (VLL), which mix with the SM leptons is also possible (see, for instance, [30–34]); and they can result from adding particles with bare mass terms, e.g., in specific applications of the Minimal Flavor Violation (MFV) principle to the lepton sector (see, for instance, [35, 36]) where the bare mass terms appear via the seesaw mechanism [8].

In the research presented in this thesis, we look for the Higgs LFV decays $H \to \mu\tau$ and $H \to e\tau$, which violate the lepton flavor symmetries $U(1)_\mu \times U(1)_\tau$ and $U(1)_e \times U(1)_\tau$ respectively. These decays are FCNC processes with leptons, which are absent from the SM at any level. When adding neutrino masses (via the seesaw terms in (7)), FCNC processes with leptons occur at higher order. However, they are highly suppressed by loop factors, the GIM (or Glashow–Iliopoulos–Maiani) mechanism, and the flavor mixing coefficients and contribute to these decays only at tiny and unobservable rates ($B(H \to \mu\tau) \sim 10^{-50}$). Therefore, observing any of these decays with non-negligible rates (BR of a few per-mil) would have far-reaching implications and could hint towards NP at higher dimension terms of the effective theory. One possibility is the dimension six terms:

$$L^{dim-6} = \frac{Y'_{e,ij}}{\Lambda^2_{LFV}} (\phi^\dagger \phi)\phi \overline{L_{L_i}} E_{R_j} + h.c. \tag{17}$$

Furthermore, an observable Higgs LFV decay suggests that the non-diagonal couplings are not much smaller than the diagonal ones. This is challenging in extensions where the decay is loop suppressed, as in the case of the Minimal Supersymmetric SM (MSSM) [37]. Thus we can speculate that the searched-for decays, if found, would proceed at tree level.

2.3.2 Two Higgs doublet model

The simplest model that allows for non-negligible Higgs LFV decays is the two Higgs doublet model (2HDM). It has two distinct scalar Higgs fields, ϕ_1, and ϕ_2[4]. In general, the Yukawa couplings of leptons to both fields, shown in (18), aren't simultaneously diagonalizable, allowing for FCNCs.

$$L^{2\text{HDM}}_{Yuk,\ell} = \eta_{ij}\bar{L}_{Li}\phi_1 E_{Rj} + \hat{\xi}_{ij}\overline{L_{Li}}\phi_2 E_{Rj} + h.c. \tag{18}$$

Without loss of generality, we can work in the so-called Higgs basis where $\langle\phi_1\rangle = v/\sqrt{2}$ and $\langle\phi_2\rangle = 0$. In this basis, the η_{ij} couplings will generate fermions masses after spontaneous symmetry breaking, while the $\hat{\xi}_{ij}$ lead to FCNC at tree level:

$$L^{2\text{HDM}}_{FCNC,\ell} = \xi_{ij} h\overline{L_{Li}}E_{Rj}$$

In type III 2HDM, no prior assumptions are placed on the couplings of the two Higgs fields, and the ξ coefficients are arbitrary. Nevertheless, it is common to use the Cheng-Sher *Ansatz* [39]; the flavor-changing couplings should be of the order of the geometric mean of the Yukawa couplings of the two fermions:

$$\xi_{ij} \propto \sqrt{m_i m_j}\frac{\sqrt{2}}{v}$$

2.3.3 Vector-like leptons

The SM leptons are chiral in that left-handed and right-handed fermions transform differently under the $SU(2)_L$ group. They come in three generations, and a chiral fourth generation is excluded based on experimental measurements of the Higgs couplings [40]. On the other hand, adding VLLs is still a valid possibility, exploited in various SM extensions. As opposed to chiral fermions, right-handed and left-handed fermions of the vector-like family transform similarly under the $SU(2)_L$ group. As a result, they include Dirac mass terms – independent of the electroweak symmetry breaking – negating the need for large Higgs couplings, which conflicts with existing measurements. The addition of VLLs in various well-motivated extensions of the SM leads to natural Higgs LFV couplings, for example, in composite Higgs models [30, 31] or models with warped extra dimensions [31–33].

Many different setups, including VLLs, can be considered. VLLs can be $SU(2)$ singlets, doublets, or triplets and come in multiple types (generations). In this description, we concentrate on a specific setup following [34], inspired by composite Higgs models. In this setup, we include singlet and doublet VLLs with three generations for each (similarly to the SM leptons), and the SM charged leptons get their masses due to their mixing with the added vector-like states.

We denote the chiral SM leptons by $l^i_L = (\nu^i_L, e^i_L), e^i_R, i = 1\dots3$, and the three generations of VLLs by $L^i_{L/R} = (N^i_{L/R}, E^i_{L/R}), \tilde{E}^i_{L/R}$, transforming as $2_{-1/2}$ and 1_{-1} under the electroweak gauge group. The VLLs acquire masses via Dirac terms and Yukawa couplings with the Higgs boson:

$$L_{VLL,\text{mass}} = -M(\bar{L}C_L L + \bar{\tilde{E}}C_E\tilde{E}) - (\bar{L}_L Y\tilde{E}_R H + \bar{L}_R\tilde{Y}\tilde{E}_L H + \text{h.c.}) \tag{19}$$

[4] Although the formalism used here is different, 2HDM can also be described in the SM EFT formalism when assuming that the new degrees of freedom introduced are much heavier than the weak energy scale (see, for example, [38]). In this case, off-diagonal Higgs couplings arise from the dimension six terms described in (17).

where $C_L, C_E, Y, \tilde{Y}$ are 3×3 matrices in generation space, and M is a common mass scale of the VLLs. The chiral leptons acquire mass via mixing with the VLLs:

$$L_{VLL,\text{mix}} = M(\bar{l}_L \lambda_l L_R + \bar{\tilde{E}}_L \lambda_e e_R) + \text{h.c.} \tag{20}$$

where $\lambda_{l/e}$ are also 3×3 matrices in flavor space. Assuming the VLLs are very heavy and the elements of $C_{L/E}$ are $O(1)$, then $M \gg v$ sets the mass scale of the VLLs.

In the EFT formalism, we integrate out the new heavy particles. In this case, the effective Higgs couplings to the chiral charged leptons can be written as:

$$L^{Y_l}_{VLL,\text{eff}} = -\frac{h}{\sqrt{2}} \bar{e}_L c_{\text{eff}} e_R + \text{h.c.} \tag{21}$$

where c_{eff} is the effective Yukawa couplings matrix of the chiral leptons. It is proportional to the mixing couplings $\lambda_{l/e}$ and, at the zero derivative level, includes a term that is mediated by each of the two VLL Yukawa couplings, the Y and $\tilde{Y}$ terms in (19). This is illustrated by the diagrams in Figure 1, which were taken from [34]:

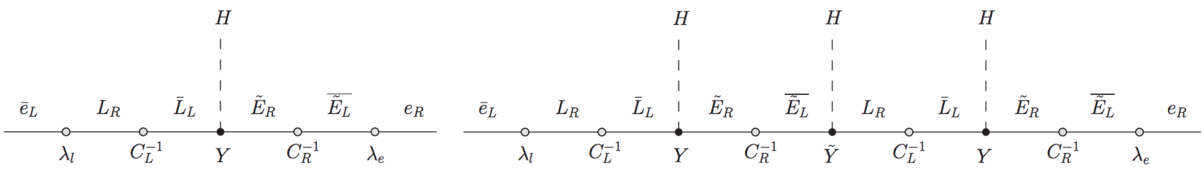


Figure 1: Diagrammatic illustration of the effective Yukawa coupling to the chiral leptons.

Explicitly, the Yukawa couplings to the chiral leptons are given by:

$$c_{\text{eff}} = \lambda_l C_L^{-1} Y C_R^{-1} \lambda_e + \frac{3v^2}{2M^2} \lambda_l C_L^{-1} \left[Y C_R^{-1} \tilde{Y} C_L^{-1} Y \right] C_R^{-1} \lambda_e + O(\lambda^4) \tag{22}$$

We see that at Leading Order (LO), they depend on the VLL Yukawa matrix Y and are flavor diagonal in an adequate basis. The contribution of $\tilde{Y}$ is suppressed by v^2/M^2. For non-zero $\tilde{Y}^{ij}$ non-diagonal elements, LFV decays of the Higgs boson can be generated:

$$\frac{\Gamma(H \to e_i^{\pm} e_j^{\pm})}{\Gamma(H \to e_j^{+} e_j^{-})_{SM}} = \frac{v^2}{2m_{e_j}^2} (|c^{ij}_{\text{eff}}|^2 + |c^{ji}_{\text{eff}}|^2) \tag{23}$$

2.3.4 Minimal Flavor Violation in the lepton sector

Another interesting model is found when imposing the MFV principle on the charged lepton sector. MFV assumes that the Yukawa structure of the SM is the only source of flavor-changing processes. In the absence of Yukawa interactions (taking the Yukawa couplings to zero), the SM has a global $[U(3)]^5$ symmetry under which each of the SM fermion representations with three flavors, $Q_{L_i}, U_{R_i}, D_{R_i}, L_{L_i}$, and E_{R_i}, can transform following $X \to V_X X$ where V_X is a 3×3 unitary matrix. Therefore, this symmetry includes what can be described as *flavor invariance*. The Yukawas break this flavor symmetry, leaving only the global SM symmetry described by (5). In MFV, the Yukawa matrices are treated as spurions (dimensionless non-dynamical fields) with transformation properties that permit recovery of this flavor invariance in the

SM. In addition, any new operator introduced must also be flavor invariant. Under these requirements, the couplings of the dimension six terms in (17) can be written as linear combinations of the Yukawa matrices. As mentioned earlier, these couplings induce FCNC at tree level.

MFV applied to the SM lepton sector does not give rise to off-diagonal Higgs couplings; we want to recover the $SU(3)_L \times SU(3)_E$ flavor symmetry[5] in the Yukawa terms $Y_e \overline{L_L} \phi E_R$. Since $\overline{L_L} \sim (\bar{3}, 1) \rightarrow \overline{L_L} V_L^\dagger$ and $E_R \sim (1, 3) \rightarrow V_R E_R$, we need $Y_e \sim (3, \bar{3}) \rightarrow V_L Y_e V_R^\dagger$. For the dimension six operator $\frac{Y_e'}{\Lambda^2}(\phi^\dagger \phi)\phi \overline{L_L} E_R$ to be flavor invariant, we impose $Y_e' \sim (3, \bar{3}) \rightarrow V_L Y_e' V_R^\dagger$. So Y_e' must be made up of odd powers of Y_e, and we simply write $Y_e' = aY_e + b(Y_e^\dagger Y_e)Y_e + O(Y_e^5)$, where a and b are constants. Since Y_e' and Y_e are simultaneously diagonalizable, no off-diagonal Higgs couplings arise here.

This is different from the case of MFV applied to the SM quark sector. With both up and down right-handed quarks, the two different Yukawa matrices Y_u and Y_d are treated as separate spurions. As a result, off-diagonal Higgs couplings arise. A similar situation can occur in the lepton sector if we introduce neutrino masses. For the case of right-handed Dirac neutrinos, the flavor symmetry is now $SU(3)_L \times SU(3)_E \times SU(3)_\nu$. There is an additional Yukawa term, $Y_\nu \overline{L_L} \widetilde{\phi} \nu_R$, two spurions: $Y_e \sim (3, \bar{3}, 1) \rightarrow V_L Y_e V_E^\dagger$ and $Y_\nu \sim (3, 1, \bar{3}) \rightarrow V_L Y_\nu V_\nu^\dagger$ and again, we impose $Y_e' \sim (3, \bar{3}, 1) \rightarrow V_L Y_e' V_E^\dagger$ for the dimension six operator to be flavor invariant. Y_e' is still expanded in odd powers of Y_e: $Y_e' = aY_e + b(Y_e^\dagger Y_e)Y_e + c(Y_\nu^\dagger Y_\nu)Y_e + O(Y^5)$, but this time the extension contains a $Y_\nu^\dagger Y_\nu$ term. Since Y_e and Y_ν are not necessarily diagonalized in the same basis, the leading contribution to flavor-changing Higgs couplings comes from the $Y_\nu^\dagger Y_\nu$ matrix. In the mass basis where Y_e is diagonal, we can express Y_e' in terms of the lepton mixing matrix U_{PMNS} coefficients and the Y_ν eigenvalues $(y_1, y_2, y_3) \propto (m_{\nu 1}, m_{\nu 2}, m_{\nu 3})$. For instance, keeping only the leading contributions and assuming standard neutrino mass hierarchy, we find:

$$y'_{\mu\tau} = c(Y_\nu^\dagger Y_\nu)_{\mu\alpha}(Y_e)_{\alpha\tau} = c(UY^2_{\nu,diag}U^\dagger)_{\mu\tau} y_\tau = c\sum_{i=1}^{3} u_{\mu i}u^*_{\tau i} y_i^2 y_\tau \sim c u_{\mu 3} u^*_{\tau 3} y_3^2 y_\tau$$

$$y'_{e\mu} \sim c u_{e3} u^*_{\mu 3} y_3^2 y_\mu$$

From the above, we have $y'_{e\mu}/y'_{\mu\tau} \sim 10^{-2}$. The current bound on $y'_{e\mu}$ results from the bound on the $\mu \rightarrow e\gamma$ decay [41]: $\sqrt{|y'_{e\mu}|^2 + |y'_{\mu e}|^2} \leq 10^{-6}$. Hence, lepton MFV with Dirac neutrinos would lead only to unobservably small rates for Higgs LFV decays, with $y'_{\mu\tau} \leq 10^{-4}$, corresponding to a BR of order 10^{-6}.

A more interesting scenario is the case of Majorana neutrinos, with masses generated by the seesaw mechanism from the dimension five non-renormalizable terms shown in (7). Here we introduce three right-handed sterile neutrinos $N_{R_i}, i = 1, 2, 3$. The flavor symmetry is now $SU(3)_L \times SU(3)_E \times SU(3)_N$. Additional terms are the Yukawa terms $Y_\nu \overline{L_L} \widetilde{\phi} N_R$ and the bare mass terms $M_N \overline{N_R^c} N_R$. There are three spurions this time: $Y_e \sim (3, \bar{3}, 1) \rightarrow V_L Y_e V_E^\dagger$, $Y_\nu \sim (3, 1, \bar{3}) \rightarrow V_L Y_\nu V_N^\dagger$ and $S_N \sim (1, 1, 6) \rightarrow V_N^* S_N V_N^\dagger$, where we define S_N by $M_N = M \cdot S_N$ in order to decouple the mass scale from the flavor structure. Again we impose $Y_e' \sim (3, \bar{3}, 1) \rightarrow V_L Y_e' V_E^\dagger$ for the dimension six operator to be flavor invariant, and we have $Y_e' = aY_e + b(Y_e^\dagger Y_e)Y_e + c(Y_\nu^\dagger Y_\nu)Y_e + O(Y^5)$. So here also, the leading contribution to flavor changing Higgs couplings comes from the $Y_\nu^\dagger Y_\nu$ matrix. The difference with the Dirac case is that the light-neutrino mass matrix m_ν and Y_ν aren't simultaneously diagonalizable due to the additional bare mass term. By diagonalizing the effective neutrino mass matrix shown in (24), we find $m_\nu = diag(m_{\nu 1}, m_{\nu 2}, m_{\nu 3}) = v^2 U Y_\nu^T M_N^{-1} Y_\nu U^\dagger$.

[5] This is the part of the global $[U(3)]^5$ symmetry that is broken by the leptonic Yukawa terms.

$$M_\nu = \begin{pmatrix} 0 & Y_\nu \frac{v}{\sqrt{2}} \\ Y_\nu^T \frac{v}{\sqrt{2}} & M_N \end{pmatrix} \tag{24}$$

So $Y_\nu = i\frac{\sqrt{M_N}}{v} R\sqrt{m_\nu} U^\dagger$, where R is a general orthogonal matrix. Keeping only the leading contributions, this leads to the flavor-changing couplings:

$$y_{ij} \propto (Y_\nu^\dagger Y_\nu)_{ij} y_j = \frac{1}{v^2}(U\sqrt{m_\nu} R^\dagger M_N R \sqrt{m_\nu} U^\dagger)_{ij} y_j \tag{25}$$

Concretely, R and M_N are completely unknown, while there is only partial information on m_ν. If we assume that the heavy sterile neutrinos are degenerate and that R is real, then (25) simplifies to $y_{ij} \propto (U m_\nu U^\dagger)_{ij} y_j$. So we obtain the same relation $y'_{e\mu}/y'_{\mu\tau} \sim 10^{-2}$ as in the Dirac case. To conclude, in order to allow for significant flavor-changing Higgs couplings, R and M_N must play a non-trivial role, and the relations they must fulfill have no special motivation.

2.4 Previous searches and current bounds

The best current upper limit at 95% CL on the BR of various LFV processes with the Higgs and Z bosons are the following:

- $B(h \to \tau\mu) \leq 0.15\%$, set by CMS using 137 fb^{-1} of 13 TeV data [42]. The best corresponding limit from ATLAS searches is $B(h \to \tau\mu) \leq 0.28\%$, using 36.1 fb^{-1} of 13 TeV data [43].
- $B(h \to \tau e) \leq 0.22\%$, also set by CMS using 137 fb^{-1} of 13 TeV data [42]. The best corresponding limit from ATLAS searches is $B(h \to \tau e) \leq 0.47\%$, using 36.1 fb^{-1} of 8 TeV data [43].
- $B(h \to \mu e) \sim 10^{-8}$, set by indirect constraints derived from the results of the search for $\mu \to e\gamma$ conducted by MEG [41].
- $B(Z \to \tau\mu) \leq 7.2 \times 10^{-6}$, set by ATLAS combining 20.3 fb^{-1} of 8 TeV data and 139 fb^{-1} of 13 TeV data [44].
- $B(Z \to \tau e) \leq 7.0 \times 10^{-6}$, set by ATLAS combining 20.3 fb^{-1} of 8 TeV data and 139 fb^{-1} of 13 TeV data [44].
- $B(Z \to \mu e) \leq 2.62 \times 10^{-7}$, set by ATLAS using 139 fb^{-1} of 13 TeV data [45]. The best corresponding limit set by CMS searches is $B(Z \to \mu e) \leq 7.3 \times 10^{-7}$, using 20.3 fb^{-1} of 8 TeV data [46].

No significant deviation from the SM has been found in these different flavor leptonic final states.

2.5 The LHC and the ATLAS detector

2.5.1 Overview

The LHC is the world's largest and most powerful particle accelerator and collider [47], located at the headquarters of CERN in the vicinity of Geneva, Switzerland. It is housed in a circular tunnel of 27 km in circumference, dug 50 to 175 m deep underground, across the Franco-Swiss border. It collides bunches of protons with a targeted center of mass energy of 14 TeV. Its first data-taking period (2010-2013) attained the center-of-mass energy of 7 TeV (8 TeV from 2012) and recorded data equivalent to about 5.5 fb^{-1}

(22.8 fb^{-1}) of integrated luminosity. During its second run (2015 - 2016), it delivered collisions at energy 13 TeV and about 140 fb^{-1} of integrated luminosity.

The ATLAS experiment [48] is one of the two general-purpose detectors built in an attempt to exploit the discovery potential of the LHC. It was designed to investigate a wide range of physics processes, including SM precision measurements and searches for physics beyond the SM. Particles produced in the LHC collisions emerge from the center of the detector in all directions, and the different signatures they leave are recorded. ATLAS comprises a series of concentric subsystems, each sensitive to different types of particles produced in the collisions. The Inner Detector (ID) is closest to the interaction point and measures the trajectories of charged particles [49, 50]. The calorimeters [51, 52] surround the ID; first, the EM calorimeter, which detects and stops electrons and photons, then the hadron calorimeter, which detects and stops hadrons. In the outermost layer is the Muon Spectrometer (MS), which detects the muons penetrating through the calorimeters [53].

2.5.2 ATLAS subdetectors

In this section, we briefly describe of the ATLAS subsystems most relevant to the research presented in this thesis. The focus is given to detector components that play a role in electron and muon detection in the pseudorapidity range $|\eta| < 2.5$: the ID, the EM calorimeters, and the MS.

The inner detector The ID is the closest subdetector to the interaction point. It has the following tasks; reconstruct the tracks and vertices of each event with high precision and at high efficiency; contribute to charged particle recognition and identification of secondary vertices from short-lived particle decays; permit precise measurements of the momenta of charged particles. These tasks are fulfilled by combining high-resolution detectors at inner radii with continuous tracking elements at outer radii, all contained in a solenoidal magnet with a central field of 2T.

Semiconductor tracking (SCT) detectors are used in the inner radii. The highest granularity around the vertex region is achieved using at least three layers of semiconductor pixel detectors. Four additional layers of silicon microstrips complete the measurements. The outer radii consist of straw tube tracker (TRT) systems which provide many tracking points (typically 36 per track), permitting continuous track-following with much less material and at lower costs.

The electromagnetic calorimeter The main task of the EM calorimeter is an accurate measurement of the energy and track position of electrons and photons. It also provides electron and photon triggers. The requirements include fast response, high granularity, and sufficient radiation length.

It is a sampling calorimeter with liquid argon as the active medium and lead plates as the absorber. The lead plates are accordion-shaped, providing full ϕ coverage and symmetry without azimuthal cracks. Readout electrodes are installed between the lead plates, and the remaining space is filled with liquid argon. The electrons and photons escaping the ID deposit all their energy through the production of EM showers. From the sampled energy of the shower, the particle's energy can be inferred.

The muon spectrometers The MS is the outermost detector system of ATLAS. It is designed to measure high-p_T muons with a high precision independent of the ID. The spectrometer also provides an independent muon trigger.

Two types of trigger detectors are used, Resistive Plate Chambers (RPC) in the barrel region and Thin Gap Chambers (TGC) in the endcap region. In addition, precision tracking and momentum measurement are provided by Monitored Drift Tube (MDT) chambers and, in the innermost layer of the endcap, by Cathode Strip Chambers (CSC). Measurement of the muon p_T requires at least three hits and a high magnetic field to bend their trajectories. Therefore, three layers of MDT chambers or CSCs usually combined with RPCs or TGCs are used. The magnet system of the MS includes three air-core superconducting systems located in the barrel and both endcap regions. Each consists of eight coils positioned symmetrically around the beam axis. The bending power ranges from 1.5-7.5 T, depending on the region.

2.5.3 ATLAS coordinate system and kinematic variables

A right-handed Cartesian coordinate system whose origin coincides with the nominal interaction point is defined: the positive x-axis points towards the center of the LHC ring while the y-axis points upwards. The z-axis lies along the beam pipe. Spherical coordinates are also used: the azimuthal angle ϕ is defined on the transverse plane identified by the x- and y-axes and by convention ranges between $-\pi$ and π. The polar angle θ is measured from the positive direction of the z-axis, but it is conveniently expressed in terms of the pseudorapidity $\eta = -\ln(\tan(\theta/2))$. For a massless particle, the pseudorapidity η is equal to the rapidity:

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right) \tag{26}$$

where E is the particle energy and p_z its longitudinal momentum.

The distance between two particles in the $\eta - \phi$ plane is measured by the variable:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{27}$$

A particle's transverse momentum is the projection of the particle's momentum vector on the traverse plane:

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} \tag{28}$$

Since the momentum of the incoming partons is, at first approximation, only directed along the z-axis, the transverse momenta of all the outgoing particles should sum to zero unless some particles remain undetected, such as neutrinos. Therefore, the missing transverse energy (or missing transverse momentum) is defined as:

$$E_{\mathrm{T}}^{\mathrm{miss}} = |\vec{p}_{\mathrm{T}}^{\mathrm{miss}}| = -\sum_i \vec{p}_{\mathrm{T},i} \tag{29}$$

where i loops on all the visible particles in the event.

2.6 Statistical analysis methods

In this section, we present statistical analysis methods standardly implemented in searches for BSM physics from HEP collision data [54].

2.6.1 Binned likelihood-ratio test statistic

BSM searches are usually cast into statistical hypothesis testing, where the compatibility of the data is evaluated w.r.t. two statistical hypotheses, the null hypothesis H_0 and the alternative hypothesis H_1. In the context of BSM searches, H_0 is the background-only hypothesis, where only SM processes contribute to the data, while H_1 is the signal hypothesis, where the searched-for signal is expected in addition to the SM background. The signal discovery is quantified by computing the p-value p, representing the probability at which the data is compatible with H_0. This is often translated in terms of significance $Z = \Phi^{-1}(1-p)$, where Φ is the standard Gaussian cumulative distribution. Discovery is claimed when H_0 is excluded by $Z \geq 5\sigma$ significance, or equivalently $p \leq 2.87 \times 10^{-7}$. In the absence of any significant signal found in the data, an upper limit on the signal's BR is set. In this case, H_1 is excluded by $p = 0.05$, corresponding to a 95% CL.

The compatibility of the data with a given hypothesis is computed based on histograms of the final discriminant (a selected observable which enhances the signal to background separation) using a binned likelihood-ratio test statistic. In each bin i, the measured data n_i is compared to the predicted expected value $E[n_i] = b_i(\boldsymbol{\theta}) + \mu \cdot s_i(\boldsymbol{\theta})$, where b_i (s_i) is the expected number of background (signal) entries in this bin, μ is the signal strength parameter which, common to all bins, regulates the signal's BR, and $\boldsymbol{\theta}$ is a set of nuisance parameters associated to the uncertainties in the predictions. The likelihood function $L(\mu, \boldsymbol{\theta})$ is then a bin-by-bin product of terms comparing the data to the prediction Poisson$(n_i|b_i(\boldsymbol{\theta}) + \mu \cdot s_i(\boldsymbol{\theta})$, of (optional) similar terms from subsidiary measurements of the nuisance parameters in Control Regions (CR) and of constraint terms penalizing variations of the nuisance parameters from their estimated nominal value (log-normal for normalizations or Gaussian for shape systematics). The signal strength μ and the nuisance parameters are determined by finding the best fit, which maximizes $L(\mu, \boldsymbol{\theta})$.

The significance and the upper limit on the signal's BR are derived using the test statistics q_0 and $\tilde{q}_\mu$, respectively, together with their asymptotic distributions described in [54]. Their definitions are shown in (30) and (31):

$$q_0 = \begin{cases} -2\ln\frac{L\left(\mu=0,\hat{\hat{\boldsymbol{\theta}}}\right)}{L\left(\hat{\mu},\hat{\boldsymbol{\theta}}\right)} & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0 \end{cases} \tag{30}$$

$$\tilde{q}_\mu = \begin{cases} -2\ln\frac{L\left(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu)\right)}{L\left(0,\hat{\hat{\boldsymbol{\theta}}}(0)\right)} & \hat{\mu} < 0, \\ -2\ln\frac{L\left(\mu,\hat{\hat{\boldsymbol{\theta}}}(\mu)\right)}{L\left(\hat{\mu},\hat{\boldsymbol{\theta}}\right)} & 0 \leq \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu \end{cases} \tag{31}$$

The term $L(\hat{\mu}, \hat{\boldsymbol{\theta}})$ is the likelihood evaluated at its global maximum, where $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ denote the values of the parameters at this maximum. And $\hat{\hat{\boldsymbol{\theta}}}$ represents the value of $\boldsymbol{\theta}$ that maximizes the likelihood for a certain fixed value of μ.

For a given test statistic t_μ, the p-value that measures the compatibility between the data and a given hypothesis – which assumes $\mu = \mu'$ – is provided by:

$$p_{\mu'} = \int_{t_{\mu,\mathrm{obs}}}^{\infty} f(t_\mu|\mu')\mathrm{d}t_\mu \tag{32}$$

where $t_{\mu,\text{obs}}$ is the value of t_μ observed in the data, and $f(t_\mu|\mu')$ is the Probability Distribution Function (PDF) of t_μ under the assumption μ'. For the special case $\mu = \mu'$, $f(t_\mu|\mu)$ is well approximated by the *half*-χ^2 distribution with 1 degree of freedom in the large sample approximation (see [54]). In this case, the significance Z_0 is obtained from (30) simply via $Z_0 = \sqrt{q_0}$, and the upper limit on the signal's BR is derived at 95% CL from (31) using:

$$\mu_{\text{up}} = \hat{\mu} + \sigma\Phi^{-1}(1 - 0.05) = \hat{\mu} + 1.64\sigma \tag{33}$$

where σ represents the standard deviation of $\hat{\mu}$ measured from the covariance matrix of the best fit maximizing the likelihood function.

Usually, the upper limits reported are modified versions of the one computed by (33), using the CL$_s$ method [55] which instead of p_μ uses:

$$p'_\mu = \frac{p_\mu}{1 - p_b} \tag{34}$$

where p_b is the p-value derived from the same test statistic under the background-only hypothesis:

$$p_b = 1 - \int_{\tilde{q}_{\mu,\text{obs}}}^{\infty} f(\tilde{q}_\mu|0)\mathrm{d}\tilde{q}_\mu \tag{35}$$

The CL$_s$ upper limit on μ at 95% CL is obtained by solving (34) for $p'_\mu = 0.05$.

Finally, *expected* significances or upper limits can be derived, which quantify the analysis' sensitivity prior to looking at the data in the sensitive Signal Regions (SR). For this, we replace the data in the likelihood-ratio test statistic with the so-called *Asimov dataset*, constructed as the sum of the combined background prediction and of the signal with a fixed signal strength μ ($\mu = 1$ for the expected significance and $\mu = 0$ for the expected upper limit).

2.6.2 Evaluation of a fit's performance

In the following, we describe some standard methods and tools used to evaluate a binned likelihood fit's performance. The fit is used to find the values of the signal strength parameter μ and the nuisance parameters $\boldsymbol{\theta}$, which maximize the likelihood function.

Fit inputs The fit may be performed with different types of inputs (corresponding to the left side of the Poisson terms in the likelihood function)

- **Asimov dataset:** constructed as the sum of the combined background prediction and the signal with a fixed signal strength μ. This dataset is used for validating the fit's performance, as well as for the obtention of the expected significance (with $\mu = 1$) and expected limit (with $\mu = 0$) of the search
- **Sideband/Mixed dataset:** includes the data only in the bins where the expected signal fraction (w.r.t. the background prediction) is smaller than 5% – the non-blinded bins – measured for a signal BR of 1%. Other bins – the blinded bins – are discarded in the sideband dataset, while in the mixed dataset, they include Asimov dataset yields but modified with post-fit background yields from a background-only fit in the non-blinded bins. These datasets are also used for validating the fit's performance, taking into account some of the possible post-fit values of the nuisance parameters

- **Unblinded dataset:** includes the data in all the bins, without any blinding restrictions. This dataset is used for the obtention of the actual results, the observed significance, and the observed limit of the search

Parameter pulls, constraints, and correlations After a fit, the parameter pulls are the adjustments on each parameter's value towards better prediction to data agreement (for the Asimov dataset, no pulls are expected by construction). In general, pulls should be within 1σ of the parameter's pre-fit uncertainties. The post-fit parameter uncertainties are derived from the covariance matrix of the parameters, which includes correlation coefficients. These correlations are built by the fit, even if the correlated parameters are from completely independent sources. One should ensure that the correlations are sensible, e.g., if two parameters are found largely correlated, then it is expected to some degree. Correlations may lead to constraints, meaning the reduction of the parameters' uncertainties w.r.t. the pre-fit values. Considerable constraints should be investigated.

Ranking of nuisance parameters A ranking of the nuisance parameters by their impact on the signal strength μ can be constructed. To obtain the impact of one nuisance parameter on μ, the value of this nuisance parameter is fixed, e.g., to its -1σ-variation, the whole fit is repeated, and the resulting value of μ is saved. The impact is the value of μ from the nominal fit minus the new value of μ. This is repeated for the $+1\sigma$-variation of the nuisance parameter and all other nuisance parameters. Finally, a ranking can be made.

Breakdown of Uncertainties The breakdown of the uncertainties, i.e., the impact of a group of nuisance parameters on the uncertainty of the signal strength μ, is obtained by the following procedure: a fit is performed where the nuisance parameters of one group are fixed to their best-fit value from the nominal fit. The resulting uncertainty on μ is subtracted in quadrature from the uncertainty on μ from the nominal fit. The result is the impact of this group of nuisance parameters on μ.

3 The e/μ-symmetry method

The electron/muon (e/μ) symmetry method is a data-driven analysis method to search for BSM physics in HEP experimental data. It was first proposed in [56] and implemented, in a simplified form, in the ATLAS Run-1 search for Higgs LFV decays [57]. The results described in this thesis mainly center around the development, implementation, and application of this analysis method in different contexts.

The e/μ-symmetry method is described in section 3.1; the example of the Higgs LFV search is detailed in section 3.2; the efficiency correction – which accounts for asymmetries induced by different detection efficiencies – is presented in section 3.3; the implementation of the method in a statistical analysis is discussed in section 3.4; section 3.5 discusses possible implementations and generalizations of the method.

3.1 Description

The e/μ-symmetry method is based on the premise that the kinematic properties of the SM background are, to a good approximation, symmetric under the exchange of electrons and muons ($e \leftrightarrow \mu$). This electron-muon symmetry, which derives from LU, is not an exact symmetry. It is invalidated by phase-space effects and Higgs Yukawa couplings due to the mass difference between electrons and muons. But the Yukawa couplings are very small, and at the energy achieved in the LHC's collisions, the differences due to phase-space effects are negligible, rendering the e/μ-symmetry method applicable.

Based on this assumption, any two mutually exclusive datasets, one comprising of a number of electrons in its final state and the other of the same number of muons instead, are expected to be statistically consistent – if they consist solely of contributions from SM processes. In this case, the distribution of any observable (kinematic property or combination of kinematic properties of the data) from these two datasets should be equivalent up to statistical uncertainties. In the e/μ-symmetry method, these two datasets are compared, and any observed asymmetry is interpreted as a sign of NP. In terms of one-sided searches for BSM physics, one dataset is probed for excess above the other, which serves as the SM background estimate (the roles of the two datasets can be switched).

The main advantage of this analysis method is that it provides a data-driven background estimation method, evading the need to use MC samples that are computationally heavy, include systematic uncertainties, and need to be validated at length before being used in an analysis. As such, the e/μ-symmetry method can be applied more efficiently, and a wide range of final states can be probed rapidly. Indeed, the number of different final states, including electrons or muons in association with any additional object(s) and any combination of charges, is vast, each potentially sensitive to different BSM manifestations. Some examples are ee vs. $\mu\mu$, $e^+\mu^-$ vs. $e^-\mu^+$, e+jet vs. μ+jet, . . . In addition, any other symmetry of the SM – exact or approximate – could be similarly probed. The main disadvantages of this method are that the background estimate is limited by the statistics in the recorded data; and that it is only sensitive to eventual BSM processes that contribute asymmetrically to the two datasets considered. Indeed, if one or different BSM processes would lead to symmetric contributions, they wouldn't be detectable.

An important aspect to consider in the e/μ-symmetry method is that detection effects invalidate the expected SM symmetry. Electrons and muons are different objects detected by different detectors and technologies. As a result, some asymmetry between the datasets, including electrons vs. muons, is induced when recording the events. These effects must be identified and corrected to restore the expected SM symmetry. Among these effects, we identified two dominant ones:

- Contribution of events containing misidentified and non-prompt leptons, namely fake events. These fake leptons originate mainly from misidentified jets or leptonic decays within jets, and the rates of faking electrons and muons are different.
- Difference in detection efficiencies of electrons and muons, which vary differently with the kinematic properties of each lepton.

Other effects include resolution of the kinematic parameters of each lepton, correct determination of the charge, and others. But so far, these have not been taken into account as their impact is smaller than the overall uncertainties (the combination of the systematic uncertainties deriving from the corrections applied to the two main effects listed above and of the statistical uncertainties).

3.2 The search for Higgs Lepton Flavor Violating decays example

Searching for Higgs LFV decays is well-motivated: LFV occurs in nature (section 2.2), and such decays are predicted in various SM extensions (section 2.3). The e/μ-symmetry method was first implemented in the ATLAS Run-1 Higgs LFV analysis [57], and this same search with the ATLAS Run-2 data is presented in section 4. In the following, we describe the application of the e/μ-symmetry method to these searches.

More specifically, the e/μ-symmetry method is described when the τ lepton further decays to the lepton of the *other* flavor in the Higgs LFV decays considered:

- $H \to \mu\tau \to \mu e + 2\nu$ ($H \to \mu\tau_e$)
- $H \to e\tau \to e\mu + 2\nu$ ($H \to e\tau_\mu$)

In principle, the e/μ-symmetry method could also be applied to the same search where the τ decays hadronically or to the same-flavor lepton, but this is not considered in the scope of the research presented here and will therefore not be addressed. The two decay modes $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ are illustrated in Figure 2, where ℓ/ℓ' is used to denote electrons and muons.

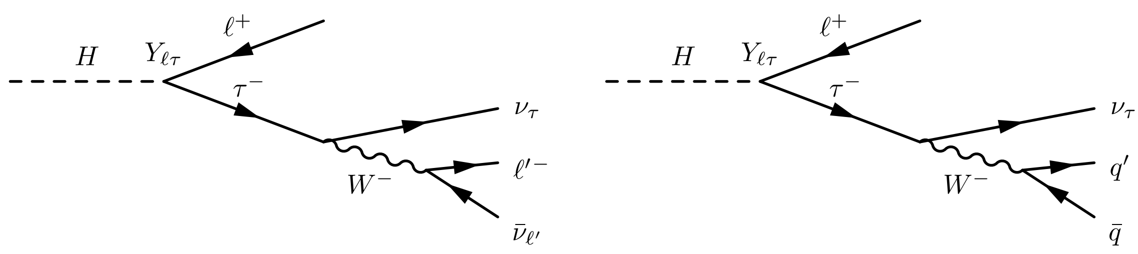


Figure 2: LFV decay schemes of the Higgs boson for the $\ell\tau_{\ell'}$ (left) and $\ell\tau_{\text{had}}$ (right) final states. The off-diagonal Yukawa coupling term is indicated by the $Y_{\ell\tau}$ symbol.

In this context, the e/μ-symmetry method provides sensitivity to the difference between the two BRs $|B(H \to \mu\tau_e) - B(H \to e\tau_\mu)|$, and we perform each of the two searches independently, assuming that the BR of the other decay is zero. This is motivated by the experimental limit on $B(\mu \to e\gamma)$ [41], which indirectly limits the product $B(H \to \mu\tau_e) * B(H \to e\tau_\mu) < 10^{-8}$, ensuring that at least one of the two BRs is currently undetectable. For simplicity, we describe the method assuming an $H \to \mu\tau_e$ signal. The same procedure, but with e and μ exchanged, is valid under the $H \to e\tau_\mu$ assumption.

Events selected must contain two opposite-sign leptons, one electron and one muon. The leading lepton (the lepton with the larger transverse momentum) is indicated by ℓ_0 and the subleading lepton by ℓ_1. These dilepton events are then divided into two mutually exclusive datasets:

- μe dataset: ℓ_0 is the muon and ℓ_1 is the electron ($p_{\mathrm{T}}^{\mu} > p_{\mathrm{T}}^{e}$)
- $e\mu$ dataset: ℓ_0 is the electron and ℓ_1 is the muon ($p_{\mathrm{T}}^{e} > p_{\mathrm{T}}^{\mu}$)

Based on the expected SM $e \leftrightarrow \mu$ symmetry assumption, the SM background is split equally between the two datasets. The $H \to \mu\tau_e$ signal, however, is present only in the μe dataset because the p_{T} spectrum of electrons from $H \to \mu\tau_e$ decays is softer than the muon p_{T} spectrum. Indeed, the energy of the τ lepton is divided between the electron and two neutrinos it decays to. Therefore, we look for an excess of events in the μe dataset compared to the $e\mu$ dataset, which estimates the SM background contributions.

In particular, one can look at histograms of the reconstructed Higgs mass (determined from the properties of the two leptons and other objects in each event) and search for an excess around the known Higgs mass value at 125 GeV. This is illustrated in Figure 3.

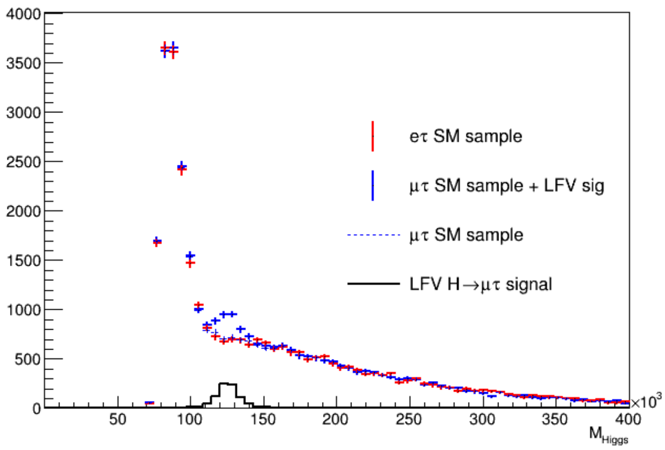


Figure 3: Illustration of the $e\mu$-symmetry method showing how the $H \to \mu\tau$ LFV signal can be discovered by comparing data yields in the $e\tau$ and $\mu\tau$ channels. Based on toy simulated data described in section 5.4.1.

One method to estimate the Higgs mass from its decay products is using the collinear approximation. The so-called collinear mass m_{coll} is based on the assumption that the decay products of the Higgs are back-to-back, that the decay products of the τ-lepton go in the same direction as the τ-lepton and that all missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ (see (29)) originates from the neutrinos of the τ-decay. Its definition is given by:

$$m_{\mathrm{coll}} = \sqrt{2p_{\mathrm{T}}^{\ell_0}\left(p_{\mathrm{T}}^{\ell_1} + E_{\mathrm{T}}^{\mathrm{miss}}\right)\left(\cosh\Delta\eta(\ell_0,\ell_1) - \cos\Delta\phi(\ell_0,\ell_1)\right)} \tag{36}$$

where the transverse momentum p_T, the pseudorapidity η, and the azimuthal angle ϕ of the two leptons are described in section 2.5.3.

Preliminary study definitions Two distinct efficiency correction methods for the e/μ-symmetry method have been developed, and they are described and compared in sections 3.3 and 3.4. The first is based on the implementation used in the Run-1 Higgs LFV analysis, but the second one is implemented in the full Run-2 Higgs LFV analysis, described in detail in section 4.

In the following sections, we illustrate the first method based on a preliminary study of the Run-2 Higgs LFV search using MC simulated data. The details and definitions given below are used in this study:

The samples used are MC simulated samples which correspond to the partial Run-2 dataset of 2015+2016 LHC data with center of mass energy 13 TeV, corresponding to a total integrated luminosity of 36.1 fb^{-1}. They are produced with the ATLAS simulation infrastructure [58] to reproduce the ATLAS detector response. The signal Higgs LFV decays are modeled using EvtGen 2.0 [59]. The production modes considered are gluon-gluon Fusion (ggF) and Vector Boson Fusion (VBF), produced with Powheg [60] generator interfaced with Pythia8 [61], and $WH + ZH$ produced with Pythia8 generator. The main SM processes that contribute to the background and are considered are $Z \to \tau\tau$, $t\bar{t}$, single top, WW, WZ, ZZ, $H \to \tau\tau$ and $H \to WW$, jointly referred to as the *SM MC sample*. Figure 4 shows diagrams illustrating some of these contributions.

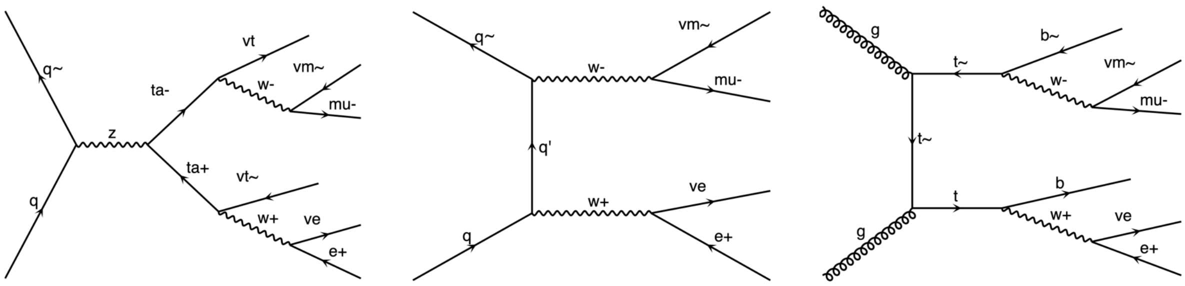


Figure 4: Diagrams illustrating some of the main SM background contributions to the Higgs LFV $\ell\tau_{\ell'}$ searches, namely $Z \to \tau\tau$, WW and $t\bar{t}$ from left to right.

Events selected must contain precisely two opposite-sign leptons of different flavors, an electron and a muon. The leptons are required to have $p_T > 15$ GeV, $|\eta| < 2.47$, and to pass a medium Identification (Id) and a gradient Isolation (Iso) criteria (see section 4.2 for details on lepton reconstruction in ATLAS). Furthermore, electrons from the transition region between the barrel and endcap calorimeters are excluded ($1.32 < |\eta| < 1.52$).

The trigger selection is based on a combination of single-electron and single-muon trigger chains, as detailed in Table 2. Priority is given to single-electron triggers, which are used in all events where the electron has p_T above the applicable threshold + 1 GeV. If the electron's p_T is below the threshold, single-muon triggers are used if the muon has p_T above the relevant threshold + 1 GeV. A further matching between the triggered lepton and the trigger track is required.

We define two event selections. The symmetric baseline selection is a loose selection that includes the Higgs LFV signal and where we can expect an $e\mu/\mu e$ SM symmetry. It contains the following selection cuts:

Trigger menu	Data period	Thresholds [GeV]
Single electron	2015	24 + 60 + 120
	2016	26 + 60 + 140
Single muon	2015	20 + 50
	2016	26 + 50

Table 2: Trigger items (more details on the triggers used are given in Table 7).

- Minimum p_{T} and symmetric trigger coverage:
  - for 2015 data: ($p_{\mathrm{T}}^{\ell_1} > 15$ GeV and $p_{\mathrm{T}}^{\ell_0} > 25$ GeV) or ($p_{\mathrm{T}}^{\ell_0/\ell_1} > 21$ GeV)
  - for 2016 data: ($p_{\mathrm{T}}^{\ell_1} > 15$ GeV and $p_{\mathrm{T}}^{\ell_0} > 28$ GeV) or ($p_{\mathrm{T}}^{\ell_0/\ell_1} > 27$ GeV)
- $30 < m_{\ell\ell} < 150$ GeV (invariant mass of the two leptons)
- No b-jets in the event (b-tagged jets with $p_{\mathrm{T}} > 25$ GeV)

The SR is a selection included in the baseline selection, constructed to enhance the signal over background ratio. At this stage, the Run-1 SR definitions are used without being optimized for the Run-2 data. The additional selection cuts are:

$p_{\mathrm{T}}^{\ell_0}$	$\geq$ 35GeV
$\Delta\phi(\ell_1, E_{\mathrm{T}}^{\mathrm{miss}})$	≤ 0.7
$\Delta\phi(\ell_0, \ell_1)$	≥ 2.3
$\Delta\phi(\ell_0, E_{\mathrm{T}}^{\mathrm{miss}})$	≥ 2.5
$\Delta p_{\mathrm{T}}(\ell_0, \ell_1)$	$\geq$ 7GeV

Table 3: SR selection cuts applied in addition to the symmetric baseline selection.

3.3 Efficiency correction

3.3.1 Main concept

As detailed earlier, the approximate e/μ symmetry in the SM is invalidated by detector effects. Thus, to apply the method, the symmetry needs to be restored. The correction to the difference in fake rates is done by estimating the contribution of events, including fake leptons (fake events), and taking this contribution into account. Estimating fake contributions is not specific to the e/μ-symmetry method since it is also required in more standard analysis methods which rely on MC samples for the background estimation. As such, it is not described here, although an example is detailed in section 4.5. On the other hand, the efficiency correction is specific to this analysis method.

The central concept of the efficiency correction is as follows. Say we compare two datasets we label as the e-dataset and the μ-dataset. Since we discuss restoring the expected SM symmetry, we currently assume that only SM processes occur. Therefore, we expect that at interaction point, the number of events contributing to each of these two datasets is equal:

$$N_0 = N_0^e = N_0^\mu \tag{37}$$

(Here, e and μ are only used to label the two datasets that we compare, which can include any switched number of electrons or muons and possibly other objects.)

We lose some of these events in different fractions from each dataset at detection level. The relation between N_0 and the number of detected events in each dataset is given through the efficiencies. Naively:

$$N_0 = \frac{N^e}{\epsilon^e} = \frac{N^\mu}{\epsilon^\mu} \tag{38}$$

So, if we apply the inverse of the efficiencies to our measured events, we scale the two datasets back to the number of events at interaction point, restoring the expected symmetry.

Another method to restore the symmetry is to apply an efficiency-*ratio* correction. Instead of scaling both datasets back to N_0, we can restore the symmetry by using the *ratio* of ϵ^μ and ϵ^e to scale the e-dataset to match with the μ-dataset (or vice-versa by switching e and μ):

$$N^\mu = \frac{\epsilon^\mu}{\epsilon^e} N^e = R^\epsilon \cdot N^e := \tilde{N}^e_{\mathrm{sym}} \tag{39}$$

We introduced the efficiency-ratio Correction Factor (CF) R^ϵ, which is the ratio of the efficiencies of the μ over e-datasets (in this example). This approach is usually preferred when using one dataset for background estimation (here the e-dataset), leaving the dataset tested for NP as raw uncorrected data (here the μ-dataset). Of course, the roles of the two datasets can be switched. $R^\epsilon \cdot N^e$ is the *SM symmetric background* contribution within the μ-dataset. This is estimated from the e-dataset; thus, we label it $\tilde{N}^e_{\mathrm{sym}}$.

The description in (38) and (39) is naive since it assumes that the efficiencies ϵ^e and ϵ^μ are constant for all events in their respective datasets. But this assumption is generally wrong since the efficiencies depend on the kinematic parameters of the objects in the events considered. In the following, we describe two distinct approaches considered to account for these variations, along with and their respective statistical model implementations.

3.3.2 Sub-region correction

The first approach considered for the efficiency correction is to divide the observables space (the space spanned by all the observables of the events in the datasets considered) into sub-regions where the efficiency ratio introduced in (39) is approximately constant. The primary motivation for using this approach is in the implementation and treatment of the statistical analysis when testing for BSM discovery. In this case, a single efficiency-correcting parameter can be assigned to each sub-region in the statistical model, and the symmetric background component is left as a fit parameter constrained by the *uncorrected* measurements of the two datasets (see section 3.4.1 for more details).

A version of this approach was used in the ATLAS Run-1 Higgs LFV analysis [57], and some further developments were considered in the research presented here. But in the end, this is not the approach implemented in the ATLAS Run-2 Higgs LFV analysis presented in section 4. Therefore it is described in a non-optimized form and only some preliminary results are shown based on the 2015+16 SM MC sample (see sample and selection definitions in section 3.2).

In the context of the Higgs LFV search, the two datasets compared are the $e\mu$ and μe datasets. Assuming only symmetric contributions, (39) is rewritten as:

$$N^{\mu e} = \frac{\epsilon^{\mu e}}{\epsilon^{e\mu}} N^{e\mu} = R^{\epsilon} \cdot N^{e\mu} \tag{40}$$

Although both the $e\mu$ and μe datasets include one electron and one muon, their efficiencies can differ. Therefore their distributions from SM processes are not symmetric after detection, as seen in Figure 5. Indeed, these datasets differ by the p_{T} ordering of the two leptons, and the lepton's p_{T} is, in general, the most sensitive parameter which affects the efficiency values. Typically, lepton detection efficiencies are very low for low p_{T} leptons, increasing when the p_{T} increases along a *turn-on* curve until they reach a certain *plateau* value. An example of this p_{T} efficiency dependency can be seen in Figure 12. In the region where both leptons' p_{T} is in the plateau region of their respective efficiencies, the efficiency ratio is assumed constant. The fact that the turn-on curves are different between electrons and muons leads to the need to define several sub-regions in the low to medium p_{T} range. In the general case, other efficiency dependencies could also affect the symmetry.

In the Run-1 Higgs LFV search, the event selection included a cut on the leading lepton's p_{T} above the p_{T} value where the efficiency plateaus start, $p_{\mathrm{T}}^{\ell_0} > 35$ GeV. As a result, the various sub-regions of constant efficiency ratio were defined solely based on the subleading lepton's p_{T}[6]. But using the same strategy with the Run-2 data results in non-negligible residual asymmetry, as seen in Figure 6. In particular, the ratio of $e\mu$ over μe $p_{\mathrm{T}}^{\ell_0}$ distributions presents a slope increasing from 0.9 to 1.2 in the considered range.

To correct this residual asymmetry, we include a $p_{\mathrm{T}}^{\ell_0}$ dependency in the efficiency correction as follows. We derive a 2D efficiency-ratio map as a function of $p_{\mathrm{T}}^{\ell_0}$ and $p_{\mathrm{T}}^{\ell_1}$ by dividing the equivalent 2D distributions of the $e\mu$ over μe MC datasets[7] in the baseline selection. In order to avoid bins with low statistics and extreme values, a simple bin-merging algorithm is applied. The resulting efficiency-ratio map obtained is shown in Figure 7 for the combined SM sample. The next step is to split the efficiency-ratio map into a finite number of channels (the sub-regions of approximately constant efficiency ratio). This is done based on a specific approach, which could be improved or replaced; still assuming a search for $H \to \mu\tau_e$, we draw the 1D histogram of the efficiency-ratio values obtained from the map in Figure 7 and from the $e\mu$ events which pass the SR selection. The resulting R^{ϵ} histogram is shown in Figure 8(a). The channels are then defined by merging bins of this R^{ϵ} histogram. The specific choice is determined from parameters such as the statistics per channel or the variance of the R_c^{ϵ} per channel; the optimal considerations have yet to be determined. Here, four channels have been selected and are shown in Figure 8(b). The measured average R_c^{ϵ} value per channel is used as the nominal value for the efficiency-CFs, and the standard deviation is used as a constraint (see Table 4). For validation, the $e\mu$ and μe kinematic distributions are compared in each channel after applying the relevant mean R_c^{ϵ} values as CFs to the μe dataset. The resulting SM MC sample's collinear mass distributions per channel in the SR selection are shown in Figure 9. We find that the symmetry is indeed restored, and the agreement in most bins is within statistical uncertainties. We observe an offset of 10% in channel 4, which exhibits a significant downward fluctuation in the μe

[6] In the Run-1 search, the SR was further split in two – whether at least one central high-p_{T} jet was reconstructed in the event or not. But in the Run-2 search presented in this thesis, this additional efficiency dependency was no longer observed and dropped; therefore, we don't address it here.

[7] This is opposite to the description in (40), where the efficiency ratio is from the μe over $e\mu$ datasets when searching for the $H \to \mu\tau_e$ signal. This choice depends on the statistical model implementation; see description in section 3.4.1.

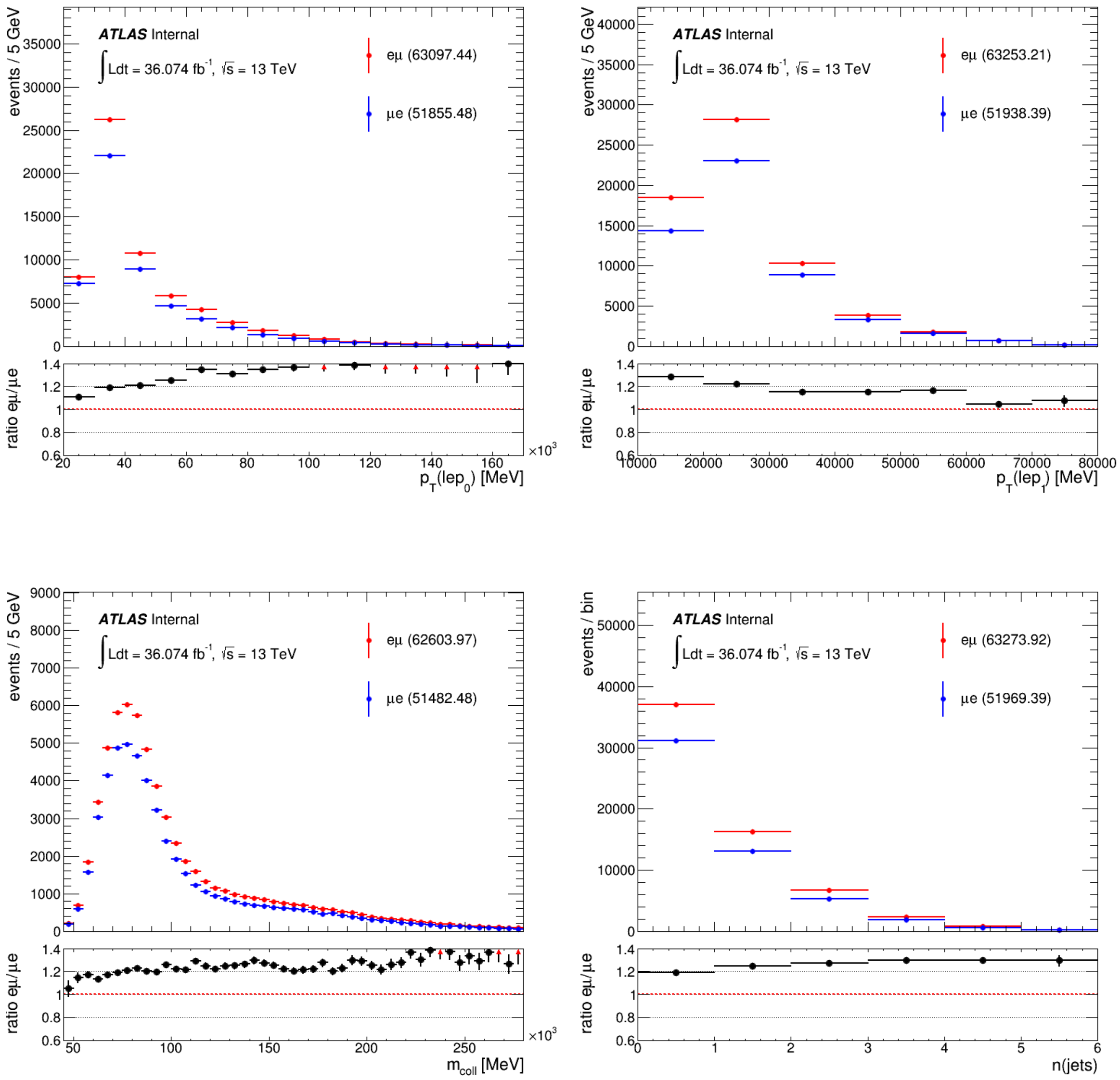


Figure 5: SM MC sample distributions in the $e\mu$ and μe datasets for events passing the symmetric baseline selection. Only statistical uncertainties are shown. No correction is applied.

dataset around 130 - 140 GeV, which we attribute to a statistical fluctuation. This channel suffers from low statistics and a wider spread of the R^{ϵ} values compared to the other three but includes the most important signal contribution.

One significant advantage of the method used here for defining the different sub-regions of approximately constant efficiency ratio is that it transforms an N-dimensional problem, where N is the number of different parameterizations of the efficiency-CFs, into a 1D problem, by using the statistics of the dataset that serves to estimate the SM background. In addition, the obtained standard deviation of the R_c^{ϵ} can be used to constrain the efficiency-CFs in the fit model (In the Run-1 implementation, these parameters were

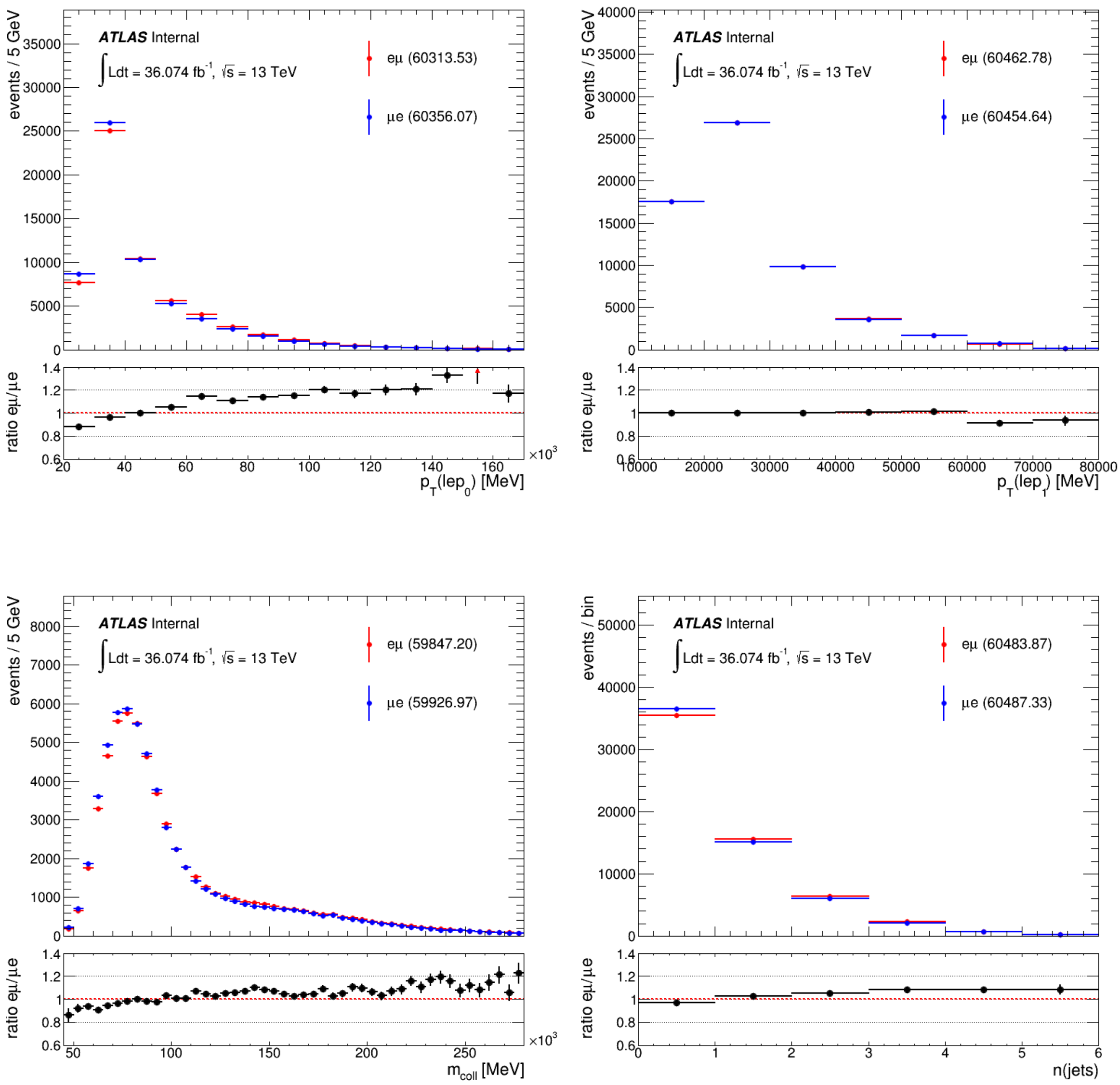


Figure 6: SM MC sample distributions in the $e\mu$ and μe datasets for events passing the symmetric baseline selection. Only statistical uncertainties are shown. Efficiency correction based on the Run-1 method is applied to the μe dataset. The CFs were measured based on the $p_T^{\ell_1}$ distributions and found to be 1.24 / 1.18 / 1.12 for events where $p_T^{\ell_1} < 20$ GeV / $20 < p_T^{\ell_1} < 30$ GeV / $p_T^{\ell_1} > 30$ GeV.

left unconstrained in the fit). Still, the need to split the SR into sub-regions of approximately constant efficiency ratio is a significant caveat because of two opposing features: this splitting comes at the cost of reduced statistics in each sub-region, increasing, in particular, the uncertainty of the estimation of the SM symmetric background; and the constant efficiency-ratio value which is used is only an approximation in each sub-region, which can only be improved by increasing the number of sub-regions. Because of these caveats, another approach for the efficiency correction was developed, bypassing the need for this splitting into sub-regions of constant efficiency ratio.

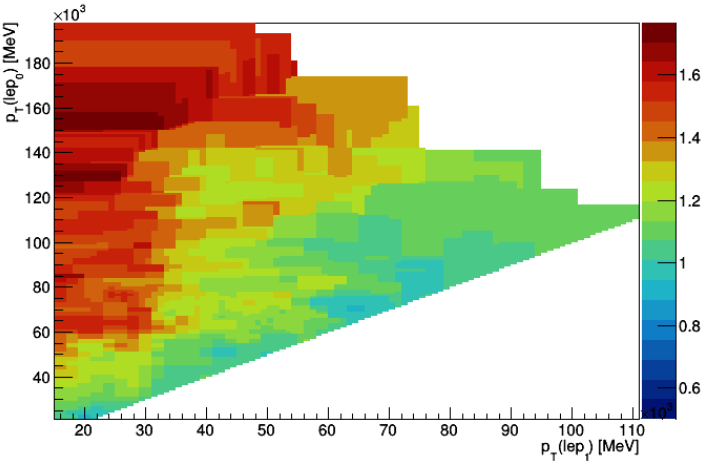
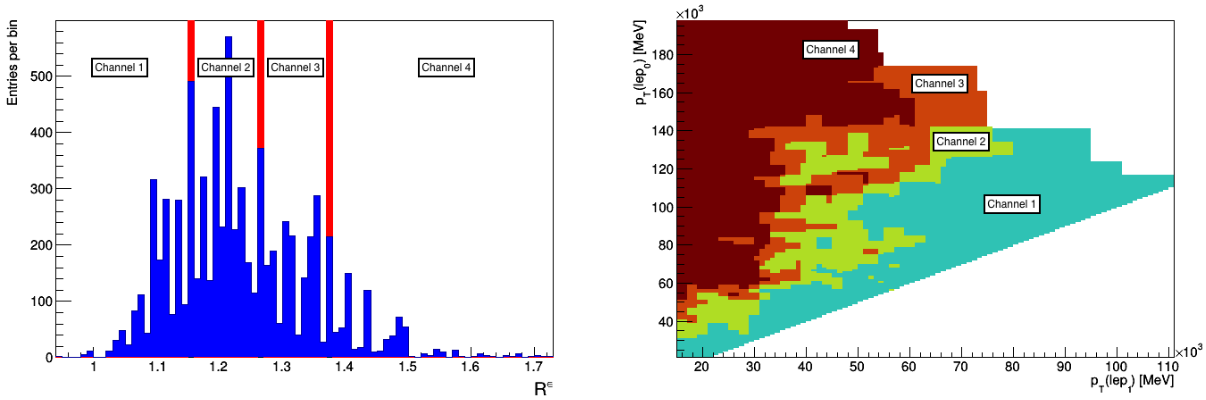


Figure 7: Efficiency-ratio map derived from the ratio of the $e\mu$ and μe SM MC samples in the symmetric baseline selection.



(a)

(b)

Figure 8: (a) Efficiency-ratio distribution from the $e\mu$ events in the Run-1-based SR. The red lines delimit the different channels in the statistical model. (b) Occupancy of each channel in the $p_T^{\ell_0}$ vs. $p_T^{\ell_1}$ space.

3.3.3 Single event correction

The second approach considered for implementing the efficiency correction is an event-by-event correction method. As discussed above, it bypasses the need for splitting the SR into sub-regions of constant efficiency ratio, which is a substantial advantage (more precise efficiency correction and more statistics for estimating the SM symmetric background contribution). In this implementation, the *other* dataset (the one which is not probed for NP) is directly used – after efficiency correction – as the SM symmetric background estimate.

In this case, we obtain the SM symmetric background $\tilde{N}^e_{\rm sym}$ (defined in (39)) by applying efficiency CFs as

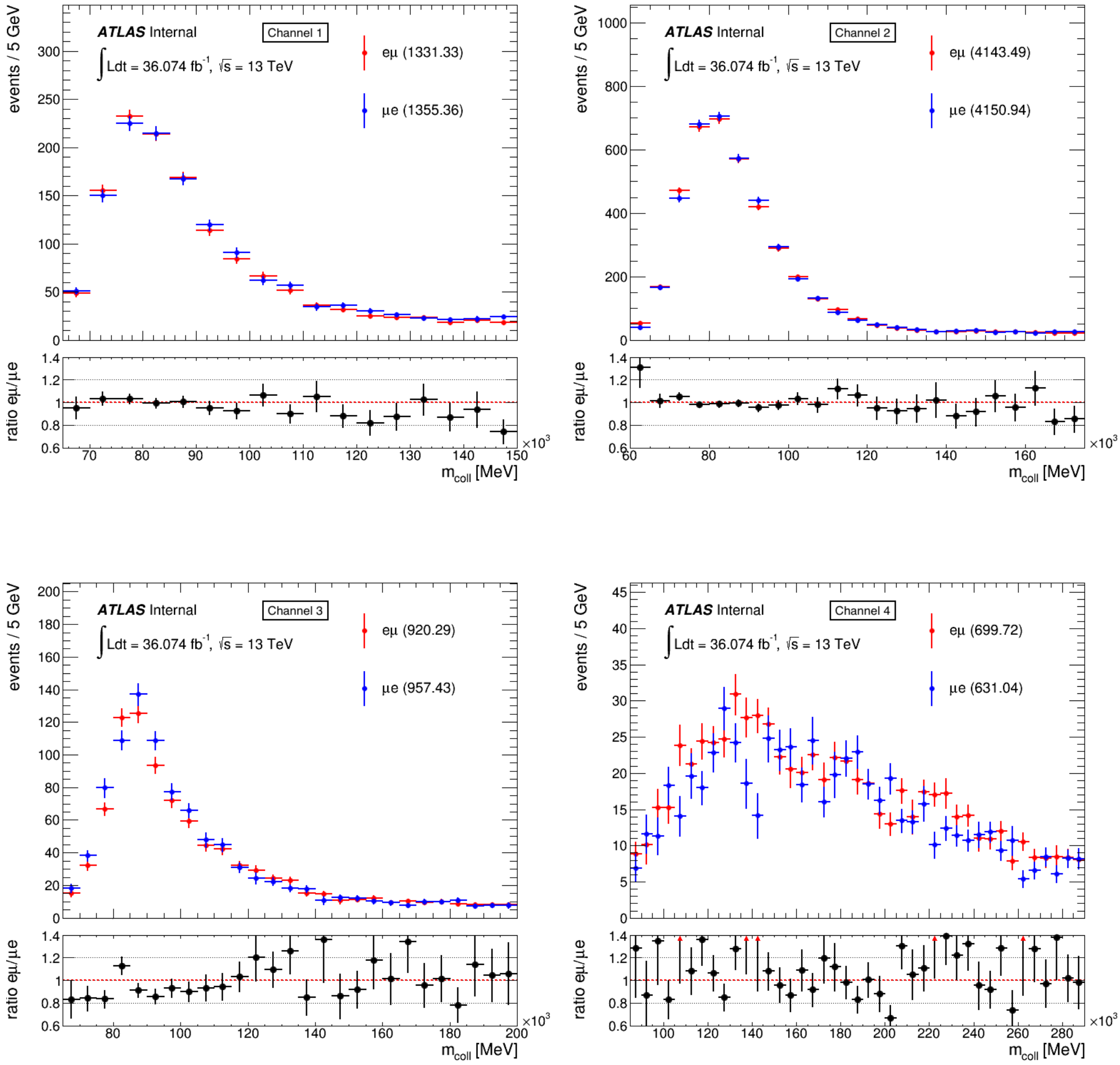


Figure 9: SM MC sample's collinear mass $e\mu$ and μe distributions in the Run-1-based SR selection, per channel of the statistical model. Only statistical uncertainties are shown. Efficiency correction using the mean R_c^ϵ values per channel is applied to the μe dataset.

weights to the e-dataset events:

$$\tilde{N}^e_{\text{sym}} = \sum_{i=1}^{N^e} \frac{\epsilon_i^\mu}{\epsilon_i^e} w_i = \sum_{i=1}^{N^e} R_i^\epsilon \cdot w_i \tag{41}$$

where $w_i = 1$ for data events, and $w_i = (\text{MC weight})_i$ for MC events.

We considered two distinct approaches to obtain the efficiency CFs:

- From the ratio of e- and μ-dataset n-dimensional distributions

- From the ratio of e- and μ-efficiencies estimated per event

The first method to obtain the efficiency CFs is illustrated based on the Higgs LFV study described in the previous section. The CFs are taken from the 2D efficiency-ratio map shown in Figure 7, derived from the ratio of $e\mu$ over μe distributions in the baseline selection. Applying them to the μe events in the Run-1-based optimized selection successfully restores the $e\mu$ vs. μe symmetry in this selection – within statistical uncertainties in most bins – as seen in Figure 10.

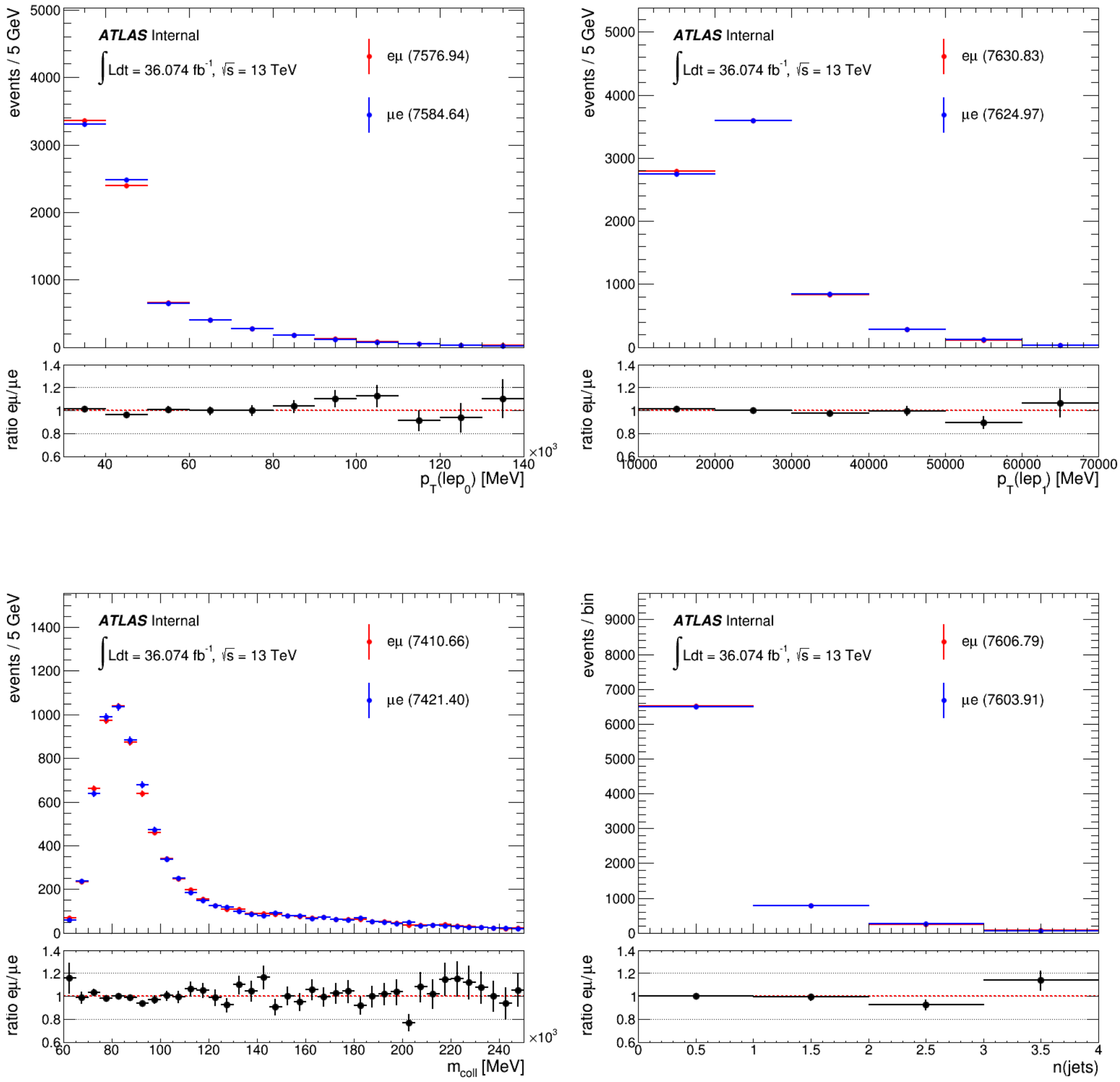


Figure 10: SM MC sample distributions in the $e\mu$ and μe datasets for events passing the Run-1-based SR. Only statistical uncertainties are shown. The efficiency correction using CFs from the efficiency-ratio map derived in the symmetric baseline selection is applied event-by-event.

This method is straightforward and gives good results when applied to MC-simulated samples limited to

SM processes, but applying this to real data is challenging. Indeed the presence of fake contributions – which can only be estimated up to some rather large uncertainty – could lead to biasing the values obtained. A possible approach is to rely on the ATLAS-provided Scale-Factors (SF), which correct MC mismodelings, to translate the R_ϵ values we derive from MC into CFs applicable to the data. Further developments are presented in section 5.5.

The second method to obtain the efficiency CFs was implemented in the ATLAS Run-2 Higgs LFV analysis, presented in section 4. The i^{th} e-dataset event's CF is determined from a ratio of efficiencies $R_i^\epsilon = \epsilon_i^\mu / \epsilon_i^e$. The ϵ_i^μ efficiency is determined by considering the i^{th} e-dataset's *mirror event*. It has the same kinematic properties as the i^{th} e-dataset event, but with switched electrons and muons. This is similar to when applying SFs to MC samples in order to reach better data vs. MC agreement. Indeed, the event-by-event SFs are, in fact, $SF_i = \epsilon_i^{\text{data}} / \epsilon_i^{\text{MC}}$.

The efficiency values ϵ_i^e and ϵ_i^μ are measured using estimates of the different component efficiencies from which they are composed. These vary depending on the final state composition of the compared datasets but can be summarized in three main categories:

- trigger efficiencies
- electron detection efficiencies
- muon detection efficiencies

As further detailed in section 4.3.4, lepton (electron or muon) detection efficiencies are a product of sub-efficiencies of the different constraints applied to them during their reconstruction, namely Reconstruction (Reco), Id, and Iso efficiencies. As for trigger efficiencies, they depend on the triggers used in the analysis. Each of these components can have different dependencies on the leptons' kinematic variables, which enables precise symmetry restoration.

In the Higgs LFV example, assuming a search for the $H \to \mu\tau_e$ signal, the efficiency-ratio CFs applied to the $e\mu$ events to estimate the SM symmetric background contribution can be written as:

$$R_i^\epsilon(e\mu \text{ event}) = \frac{\epsilon_i^{\mu e}}{\epsilon_i^{e\mu}} = \frac{\epsilon_{trig}^{\mu e}(\{k_i\}_e, \{k_i\}_\mu) \cdot \epsilon_\mu(\{k_i\}_e) \cdot \epsilon_e(\{k_i\}_\mu)}{\epsilon_{trig}^{e\mu}(\{k_i\}_e, \{k_i\}_\mu) \cdot \epsilon_e(\{k_i\}_e) \cdot \epsilon_\mu(\{k_i\}_\mu)} \tag{42}$$

where $\{k_i\}_\ell$ represents the kinematic properties of lepton ℓ.

3.4 Statistical analysis implementation

3.4.1 Sub-region correction

In this section, we describe the implementation of the statistical analysis when searching for a BSM signal in a μ-dataset by comparing it to an e-dataset for the case of the efficiency correction per sub-region of constant efficiency ratio. The roles of the e and μ-datasets can be switched by exchanging $e \leftrightarrow \mu$. We also show some preliminary results based on the Higgs LFV SM MC sample study, described in section 3.3.2. As a reminder, this implementation was used in the ATLAS Run-1 Higgs LFV search [57]. Still, due to the disadvantages of splitting the SR into sub-regions and the approximate efficiency correction, a different implementation was used in the ATLAS Run-2 Higgs LFV search, described in the following (sections 3.4.2 and 4.3.6).

Here, we split the SR into different sub-regions – or channels c – with a single efficiency-ratio correction parameter r_c^ϵ per sub-region. These parameters can either be constrained by exteriorly provided measurements – $\text{Gaus}(r_c^\epsilon \mid R_c^\epsilon, \sigma_{R_c^\epsilon})$ – as is the case in the Higgs LFV SM MC sample study, or left unconstrained if no measurement is available (as in the Run-1 Higgs LFV search).

As discussed in section 2.6, the statistical analysis uses a binned likelihood function applied to 1D histograms of an observable with a good signal over background separation power, the final discriminant. For example, in the Higgs LFV SM MC sample study, we chose the collinear mass distribution described in (36), using the histograms shown in Figure 10 per channel. We introduce a SM symmetric background fit parameter $b_{c,b}^{\text{sym}}$ for each histogram bin b of each channel c. The $b_{c,b}^{\text{sym}}$ are simultaneously Poisson constrained by the two data measurements $N_{c,b}^e$ and $N_{c,b}^\mu$ in each bin of the e and μ-datasets respectively, but multiplied by the r_c^ϵ for the e-dataset to account for the difference in efficiencies. We also include the searched for signal $S_{c,b}$ per bin for the μ-dataset (generally provided by MC simulation of BSM processes), multiplied by a single signal strength parameter μ^s, which is the Parameter Of Interest (POI), common to all channels and all bins. Denoting by C the number of channels c, and B_c the number of bins in the histograms of each channel, the likelihood function is given by:

$$L(\mu^s, \{r_c^\epsilon\}, \{b_{c,b}^{\text{sym}}\}) = \prod_{c=1}^{C} \prod_{b=1}^{B_c} \text{Pois}(\ N_{c,b}^\mu \mid b_{c,b}^{\text{sym}} + \mu^s \cdot S_{c,b}\ ) \times \text{Pois}(\ N_{c,b}^e \mid r_c^\epsilon \cdot b_{c,b}^{\text{sym}}\ ) \times \text{Gaus}(r_c^\epsilon \mid R_c^\epsilon, \sigma_{R_c^\epsilon}) \tag{43}$$

We omitted from this description the treatment of signal uncertainties since this depends on how the signal is provided and is not unique to the e/μ-symmetry method.

In terms of fake contributions briefly discussed in section 3.1, denoting them as $F_{c,b}^e$ and $F_{c,b}^\mu$ for each bins of the e and μ-datasets, respectively, we include them in the overall background estimate as follows (omitting now, in addition, the constraint term on the r_c^ϵ and the treatment of fake uncertainties):

$$L(\mu^s, \{r_c^\epsilon\}, \{b_{c,b}^{\text{sym}}\}) = \prod_{c=1}^{C} \prod_{b=1}^{B_c} \text{Pois}(\ N_{c,b}^\mu \mid b_{c,b}^{\text{sym}} + F_{c,b}^\mu + \mu^s \cdot S_{c,b}\ ) \times \text{Pois}(\ N_{c,b}^e \mid r_c^\epsilon \cdot b_{c,b}^{\text{sym}} + F_{c,b}^e\ ) \tag{44}$$

For example, we implement the statistical model described by (43) in the Higgs LFV SM MC study, with the channels and efficiency-ratio constraints derived in section 3.3.2. The $H \to \mu\tau_e$ MC simulated signal described in section 3.2 is used for the signal S, normalized to 1% BR. Instead of the N^e data, we input the contribution from the $e\mu$ SM MC dataset. Instead of the N^μ data, the sum of contributions from the μe SM MC dataset and the $H \to \mu\tau_e$ signal also normalized to 1% BR. Distributions from the best-fit that minimizes the negative log-likelihood ratio are displayed in Figure 11 (all channels combined); the post-fit prediction is compatible with the input data. The fitted efficiency-ratio CFs per channel are consistent with the pre-fit values, as shown in Table 4. The signal strength found is $\mu^s_{\text{best-fit}} = 0.74 \pm 0.20$, which is less than 2σ away from the input value of $\mu^s_{\text{input}} = 1$. We explain this discrepancy from the presence of the downward fluctuation in the ratio of μe over $e\mu$ events around 130-140 GeV, observed mainly in channel 4, which results in an underestimation of the signal contribution. The expected significance for the discovery of the $H \to \mu\tau_e$ signal found in this study is 3.1σ. This is with a partial (2015+16) dataset of the Run-2 data but doesn't consider fake contributions, which have large uncertainties. Also, no optimization of the SR selection was done.

We also implemented the same experience, but this time using unconstrained nuisance parameters in place of the R_c^ϵ, to test the advantage of using constrained versus unconstrained parameters. Due to the additional freedom when removing the constraints, the resulting signal strength found is more consistent with the input value, $\mu^s_{\text{best-fit}} = 0.83 \pm 0.20$, but the expected significance measured at 2.80σ is lower by $\sim 10\%$.

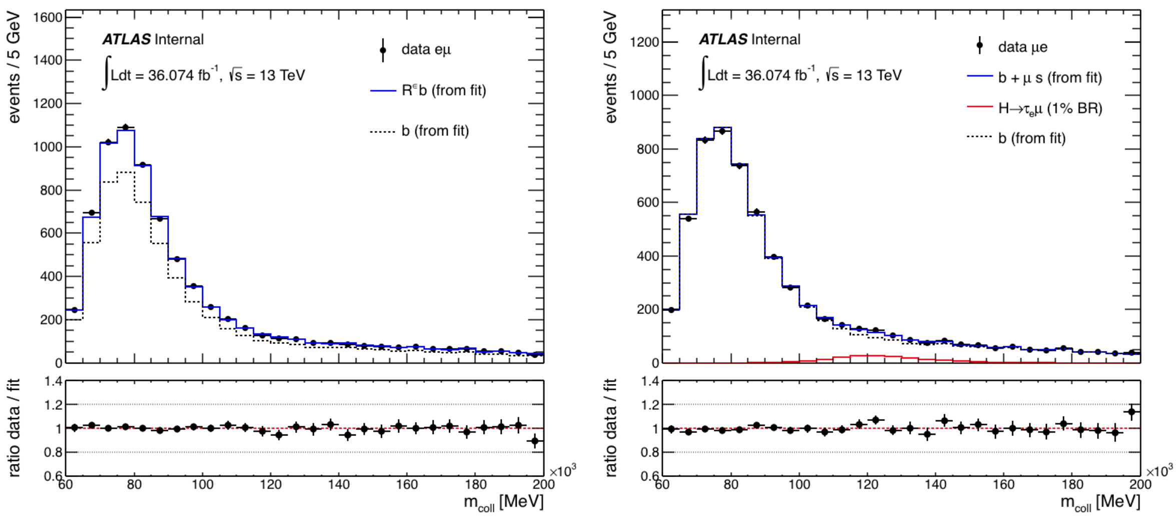


Figure 11: SM MC sample m_{coll} distributions in the $e\mu$ (left) and μe (right) datasets for events passing the Run-1-based SR, all channels combined. Comparing the input data (black dots) to the best-fit expectation values (blue). The fitted symmetric background (b - dotted line), common to both datasets, is also shown. In the μe dataset, the $H \to \mu\tau_e$ signal contribution normalized to 1% BR is shown (red).

Channel #	R^{ϵ}_{c} - prefit	R^{ϵ}_{c} - postfit
1	1.12 ± 0.03	1.12 ± 0.02
2	1.22 ± 0.03	1.24 ± 0.02
3	1.33 ± 0.03	1.32 ± 0.03
4	1.45 ± 0.06	1.47 ± 0.04

Table 4: Efficiency-ratio values per channel.

3.4.2 Single event correction

When the efficiency correction is applied event-by-event, as described in section 3.3.3, the likelihood function used is different mainly because we use the other dataset (that is not probed for BSM physics) directly to estimate the SM symmetric background. When searching for a signal in the μ-dataset, the SM symmetric background is obtained from $\tilde{N}^{e}_{\text{sym}}$ described in (41). In this case, there is no splitting of the SR into sub-regions, and the likelihood function is simply defined by:

$$\begin{aligned} L(\mu^s) &= \textstyle\prod_{b=1}^{B} \text{Pois}\left(N^{\mu}_{b} \,\middle|\, \tilde{N}^{e}_{\text{sym},b} + \mu^s \cdot S_b\right) \\ &= \textstyle\prod_{b=1}^{B} \text{Pois}\left(N^{\mu}_{b} \,\middle|\, \sum_{i=1}^{N^{e}_{b}} R^{\epsilon}_{i} + \mu^s \cdot S_b\right) \end{aligned} \tag{45}$$

where B is the number of bins in the histograms considered, N^{e}_{b} (N^{μ}_{b}) is the number of events in bin b from the e-dataset (μ-dataset), and S_b and μ^s are the signal contribution in bin b and the signal strength already introduced in (43). The treatment of uncertainties on the efficiency-ratio CFs, not shown here, is done in the same manner as when applying SFs in a standard MC-based analysis. For this, we derive the upward and downward fluctuated $\tilde{N}^{e}_{\text{sym}}$ histograms – by inputting the $R^{\epsilon}_{i} \pm 1\sigma$ values instead of the nominal ones – which are also input to the fit. A parameter per bin that multiplies the symmetric background contribution allows it to fluctuate within a Gaussian constraint delimited by these upward and downward fluctuated values in the same bin. This enables conveniently accounting for various sources of uncertainties – statistical or systematic – which depend on how the R^{ϵ}_{i} values are determined.

The inclusion of fake contributions in this statistical model must be done with more care. Assuming the presence of fake events in the data, but still with no BSM processes, (39) becomes:

$$N^{\mu} - F^{\mu} = R^{\epsilon}(N^{e} - F^{e}) := \tilde{N}^{e}_{\text{sym}} \tag{46}$$

As can be seen, the fake contribution of the e-dataset F^{e} needs to be subtracted from the e-data prior to applying the efficiency correction to estimate the SM symmetric background. Meaning that F^{e} is also efficiency-corrected, using the same CFs as if it were a SM event. On the other hand, the μ-dataset fake contribution F^{μ} is left uncorrected. Implementing this in the likelihood function is done as follows:

$$\begin{aligned} L(\mu^{s}) &= \textstyle\prod_{b=1}^{B} \text{Pois}\left(N^{\mu}_{b} \;\middle|\; \tilde{N}^{e}_{\text{sym},b} + F^{\mu}_{b} + \mu^{s} \cdot S_{b}\right) \\ &= \textstyle\prod_{b=1}^{B} \text{Pois}\left(N^{\mu}_{b} \;\middle|\; \sum_{i=1}^{N^{e}_{b}} R^{\epsilon}_{i} - \sum_{i=1}^{F^{e}_{b}} R^{\epsilon}_{i} + F^{\mu}_{b} + \mu^{s} \cdot S_{b}\right) \end{aligned} \tag{47}$$

As previously discussed, this statistical model is implemented in the ATLAS Run-2 Higgs LFV search described in section 4, such that a detailed implementation example is discussed within.

3.5 Discussion and outlook

In this section, we presented the e/μ-symmetry method, an analysis method that exploits the approximate SM $e \leftrightarrow \mu$ symmetry to provide a data-driven background estimate. It was illustrated based on the example of the Higgs LFV search, and its implementation in the ATLAS full Run-2 Higgs LFV analysis – the topic of section 4 – is a major component of the research presented in this thesis. In this context, some details were given towards its implementation in a statistical analysis which searches for a BSM signal *priorly anticipated*, as is standard in the ATLAS collaboration and searches in HEP in general. This falls in the category of *blind* analyses, where a specific signal is searched for within a predefined region of the collected data's observables space; and where the actual data included in this sensitive region is only looked at after all of the analysis is set up.

But one of the main advantages of the e/μ-symmetry method is that it provides a data-driven way to model the SM background. Once the expected SM symmetry is under control, any asymmetry found in the data could be a sign of NP. This enables implementing generic searches in large portions of the collected data's observables space, significantly enhancing the potential for discovering BSM physics simply by better exploiting the already collected data. Such generic analyses fall in the category of the DDP proposed by us in [1, 62], where we advocate permitting the data to lead us towards the interesting regions within its observables space. In section 5, more considerations and motivation towards applying such symmetry-based, data-directed, and generic searches are discussed, as well as some progress towards possible implementations. In particular, we show that the potential for discovery, even when searching for a generic signal, can be significant.

4 ATLAS Run-2 search for Higgs Lepton Flavor Violating decays

This section presents the Symmetry-based search for Higgs LFV decays to $e\tau$ and $\mu\tau$ using the full Run-2 data from pp collisions at $\sqrt{s} = 13$ TeV provided by the LHC and collected by the ATLAS detector. As described in section 3.2, two searches are performed, $H \to e\tau \to e\mu 2\nu$ and $H \to \mu\tau \to \mu e 2\nu$, where the SM background is estimated in a data-driven way using the e/μ-symmetry method.

Moreover, the complete ATLAS Run-2 Higgs LFV search comprises three separate analyses:

- MC-based searches for $H \to e\tau_{\text{had}}$ and $H \to \mu\tau_{\text{had}}$
- MC-based searches for $H \to e\tau_{\mu}$ and $H \to \mu\tau_{e}$
- Symmetry-based searches for $H \to e\tau_{\mu}$ and $H \to \mu\tau_{e}$

The MC-based searches, in contrast, use MC simulation to estimate the SM background. Although the combined results are summarized, the MC-based analyses aren't presented. In addition, the results reported here are still unpublished at the time where this is written, although the paper is ready and going through an internal ATLAS review prior to being submitted for publication. As such, they may be subject to small changes during this process.

Section 4.1 describes the samples (measured and simulated) used in the analysis, and section 4.2 the methods and definitions to reconstruct the different objects in each event; section 4.3 gives an overview of the strategy implemented; section 4.4 presents measurements of electron efficiencies used in the efficiency correction and section 4.5 the fake background estimate; in section 4.6 the different steps performed to validate the background estimation procedure are detailed; a NN implementation to optimize the signal to background separation is described in section 4.7; section 4.8 lists the systematic uncertainties considered. The statistical analysis implementation and the obtained results are presented in section 4.9, and in section 4.10, conclusions are drawn.

4.1 Collision data and simulated samples

The dataset used for these searches consists of the LHC data recorded by the ATLAS experiment at a center of mass energy of 13 TeV from 2015-2018. The total integrated luminosity of analyzed data corresponds to 138.42 fb^{-1} after the application of data quality requirements [63]. Events used in this analysis were triggered by a combination of single-electron, single-muon, or electron-muon triggers, as detailed in Table 7 [64–67].

Although this analysis is data-driven, we still use MC simulation to model the signal processes, measure electron detection efficiencies (section 4.4), fine-tune the fake background estimation (sections 4.5.1 and 4.5.2), and validate the different stages of the analysis (section 4.6). We do not perform MC studies usually implemented in standard analyses, such as normalizing specific samples for better data to MC agreement, since the SM background estimate is based on data in our analysis.

All MC samples are processed through the full ATLAS detector simulation [58] based on Geant4 [68], and the same event reconstruction algorithms are applied as with the data. Simulated events are reconstructed with pileup events overlaid following the expected pileup profile. In addition, all simulated events were weighted by SFs to correct for differences in object selection efficiencies in MC compared to data, as recommended by the respective combined performance groups.

For the Higgs LFV signal samples ($H \to \ell\tau$) and the Higgs background samples ($H \to \tau\tau$ and $H \to WW$), we consider the dominant Higgs boson production modes at the LHC: ggF, VBF, and the associated production modes WH and ZH. Others are neglected. The Higgs mass is set to $m_H = 125$ GeV [69], and the cross-sections are fixed to the SM predictions [70]. The generator used for the Higgs' production and decay is Powheg-Box v2 [60], while Pythia8 [61] is used for parton shower, hadronization, and other decays.

SM background processes that contribute to our SR include $Z/\gamma^* \to \tau\tau$, $t\bar{t}$, single top-quark, and diboson (WW, WZ, and ZZ), together with the Higgs background samples previously described. The combination of these samples is referred to as the *SM MC sample*. Smaller contributions result from $Z/\gamma^* \to ee$ and $Z/\gamma^* \to \mu\mu$. For fake background studies or electron MC efficiency measurements, W+jets, $V + \gamma$, and ttV productions are also considered.

Z + jets and W+ jets production are simulated with Powheg-Box v2 + Pythia8 or Sherpa 2.2.1. The latter is generally considered the standard, but Powheg-Box v2 + Pythia8 has been found to provide more statistics in our selection, critically needed in some low statistics regions. It is therefore used as the default for $Z \to \mu\mu$ in the main analysis and Z + jets and W+ jets in the fake composition estimate.

$t\bar{t}$ and single top-quark production are simulated using Powheg-Box v2 + Pythia8, diboson using Sherpa2.2.1, $V + \gamma$ with Sherpa 2.2.8, and ttV (only relevant for the Z + jets-CR) with aMC@NLO + Pythia8.

We use different types of data formats in our analysis. These are input files provided within the ATLAS collaboration, including events with different types of final states and applied selection rules. The primary input type is DAOD_HIGG4D1. To measure electron MC efficiencies, using only $Z/\gamma^* \to \tau\tau$, diboson, $t\bar{t}$, and single top-quark MC samples (the main SM contributions to our SR), the data format DAOD_TOPQ1 is used. To validate the electron MC efficiency measurement method – based only on $Z/\gamma^* \to ee$ MC samples – the data type used is DAOD_EGAM1.

4.2 Object reconstruction

In this section, the objects used in this analysis are briefly described. Selected events require one electron and one muon and a veto of hadronically decaying τ-leptons. For this analysis, it is essential that electrons and muons are symmetric and can originate from τ-lepton decays. Therefore, deviations from the standard ATLAS treatment of those objects are present in this analysis, which are indicated below.

Primary vertices require at least two associated tracks, each with transverse momenta $p_{\mathrm{T}} > 500$ MeV. If more than one primary vertex is reconstructed in the event, the one with the largest $\sum p_{\mathrm{T}}^2$ is chosen as the hard-scatter primary vertex. It is subsequently used to calculate the physics objects in the event.

Leptons are reconstructed with different levels of constraints depending on specific uses in the analysis. *Baseline* leptons are defined to perform the overlap removal, calculate the missing transverse momentum (see further in this section and in [80]), and veto additional leptons. Additional more stringent criteria are applied to the *signal* or *selected* leptons used in the reconstruction of Higgs decay properties to ensure higher background rejection and symmetry between electrons and muons.

Electrons are selected from clusters of energy deposits in the EM calorimeter that match a track reconstructed in the ID. They are identified using likelihood Id criteria [71]. Baseline electrons are required to pass *Loose and BLayer LH* (Likelihood-Hypotheses based) Id, which provides efficiency of about 93%. A cut on the quality requirement *BADCLUSELECTRON* is imposed to reject electrons with problematic calorimeter measurements. Baseline electrons are required to have $p_T > 15$ GeV and $|\eta| < 2.47$ and electrons inside the calorimeter crack region defined as $1.37 < |\eta| < 1.52$ are vetoed. The baseline electrons are used for the overlap removal described below. The remaining baseline electrons are then used for the E_T^{miss} calculation and the third lepton veto.

Selected electrons are further required to pass the *Medium LH* Id, which gives an average efficiency of 88% for typical electroweak processes, and the *Gradient* Iso criteria. The Gradient Iso criteria gives an efficiency that increases with the electron's p_T, designed to give 90% at p_T = 25 GeV and 99% at p_T = 60 GeV, uniform in η [71]. 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 $|z_0 \sin\theta| < 0.5$ mm. The significance of the transverse impact parameter calculated with respect to the beam-line is required to be $|d_0|/\sigma_{d_0} < 10$. The latter is looser than standard impact parameter requirements for electrons, allowing higher efficiency to select electrons originating from τ decays.

Muons are reconstructed by combining ID and MS tracks with consistent trajectories and curvatures. The combination is performed through an overall fit using the hits of the ID track, the energy loss in the calorimeter, and the hits of the tracks in the MS. These muons are called combined muons [72], and only those muons are used in this analysis. Based on the quality of the Reco and Id, muon candidates are defined as *Loose*, *Medium*, and *Tight*. Baseline muons are required to pass *Loose*, which provides an efficiency of about 98% relatively uniform in p_T, and to have $p_T > 10$ GeV and $|\eta| < 2.47$. As for the electrons, baseline muons are used for the overlap removal. The remaining baseline muons are then used for the E_T^{miss} calculation and the third lepton veto.

Selected muons are further required to pass *Medium* Id and the *FCTightTrackOnly* Iso criteria. The same impact parameter requirements as for electrons are applied: $|z_0 \sin\theta| < 0.5$ mm and $|d_0|/\sigma_{d_0} < 10$, where the latter is again loosened compared to standard muon impact parameter requirements to be consistent with the hadronic decay of a τ-lepton. For symmetry reasons, only muons with $p_T > 15$ GeV are considered, and a veto for the calorimeter crack of $1.37 < |\eta| < 1.52$ is also applied to the muons. The calorimeter crack veto for muons is specific to this analysis.

Hadronic τ (τ_{had}) decays are composed of a neutrino and a set of visible decay products, typically one or three charged pions and up to two neutral pions. The reconstruction of the visible decay products is seeded by jets with the $anti-k_t$ algorithm [73] applied to calibrated topo clusters with a distance parameter of $R = 0.4$ [74]. Jets seeding $\tau_{had\text{-}vis}$ candidates are required to have $p_T > 10$ GeV and $|\eta| < 2.5$. To separate visible decay products of hadronically decaying τ leptons, $\tau_{had\text{-}vis}$, from quark and gluon-initiated jets, isolation criteria based on a Recurrent NN (RNN) are constructed [75]. A boosted decision tree (*eBDT*) is constructed to reject electrons faking $\tau_{had\text{-}vis}$ candidates. Baseline taus are required to have 1 or 3 associated tracks, $p_T > 20$ GeV and $|\eta| < 2.47$, excluding the crack region of $1.37 < |\eta| < 1.52$. The *Medium* RNN τ Id and the *Loose electron Boosted Decision Tree (eBDT)* working points are applied as well as a muon veto criterion called *MuonORL*. Baseline taus are used for the overlap removal. The remaining baseline taus are then used for the E_T^{miss} calculation and the hadronic τ veto.

Jets are reconstructed from particle flow objects using the $anti-k_t$ algorithm with a radial distance parameter of $R = 0.4$ [76]. Energy clusters are measured at the EM scale and calibrated to the jet energy scale (JES) using the Global Sequential Calibration (GSC) [77]. Baseline jets are required to have $p_T > 20$ GeV, $|\eta| < 4.5$, and pass the *LooseBad* cleaning working point. These jets are used for overlap removal. For jets surviving the overlap removal, additional requirements on (f)JVT ((forward) Jet Vertex Tagger) are imposed to suppress jets originating from pileup [78]. The *Medium* JVT working point is used. Since the fJVT calibration is not yet available for PFlow (Particle Flow) jet, the fJVT decision and the corresponding CFs are taken from the matched EM jets. In order to identify and reject jets originating from b-hadrons, the DL1r tagger is used with an 85% working point [79].

Since the reconstructed objects are not mutually exclusive, an overlap removal procedure is applied to the baseline objects. The overlap procedure is summarized in Table 5.

Step	Reject	Against	Criteria
1	electron	electron	shared track, $p_T^1 < p_T^2$
2	$\tau_{\text{had-vis}}$	electron	$\Delta R < 0.2$
3	$\tau_{\text{had-vis}}$	muon	$\Delta R < 0.2$
4	muon	electron	is-calo muon and shared ID track
5	electron	muon	shared ID track
6	jet	electron	$\Delta R < 0.2$
7	electron	jet	$\Delta R < 0.4$
8	jet	muon	Number of tracks < 3 and ghost-associated or $\Delta R < 0.2$
9	muon	jet	$\Delta R < 0.4$
10	jet	$\tau_{\text{had-vis}}$	$\Delta R < 0.2$

Table 5: Overlap Removal selection. The selection criteria are applied sequentially as ordered in this table.

The missing transverse energy [80] is an estimate of the imbalance in the transverse momentum in the detector, and it is computed as the missing transverse vector $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, with its magnitude – the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ [81]. The $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ vector is calculated as the negative vector sum of the transverse momenta of electrons, muons, taus, and jets. Tracks not associated with any reconstructed object are considered in the calculation as soft terms. The electrons, muons, and $\tau_{\text{had-vis}}$ objects passing the baseline selection described in the previous sections are used for the missing transverse energy reconstruction together with the entire jet collection. The *Tight* criterion of the official ATLAS Missing Transverse Energy Tool has been chosen. This criterion ensures that the transverse momentum of the forward jets (with $|\eta| > 2.5$) used for the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation is above 30 GeV.

4.3 Analysis strategy

4.3.1 Overview

Similarly to the description given in section 3.2, we discuss two searches for Higgs LFV decays using the e/μ-symmetry method:

- $H \to \mu\tau \to \mu e + 2\nu$ ($H \to \mu\tau_e$)
- $H \to e\tau \to e\mu + 2\nu$ ($H \to e\tau_\mu$)

The e/μ-symmetry method provides sensitivity to the difference between the two BRs $|B(H \to \mu\tau_e) - B(H \to e\tau_\mu)|$, and we perform each of the two searches independently, assuming that the BR of the other decay is zero.

Events selected must contain precisely two opposite-sign leptons, one an electron and the other a muon. These events are then divided into two mutually exclusive datasets, one where the searched-for signal is expected and the other serving as a SM background estimate. Details on the event classification and selection are given in section 4.3.2 and section 4.3.3, respectively.

The obtention of background predictions includes several components. The SM contributions are estimated using the data-driven e/μ-symmetry method, using the single event efficiency correction (described in section 3.3.3) to account for efficiency differences between electrons vs. muons. Further details on the implementation of the efficiency correction method are given in section 4.3.4, and the measurements of electron efficiencies used in this correction are detailed in section 4.4. The fake or misidentified background is mainly estimated using a different data-driven method, although some minor contributions are estimated directly from MC. This is detailed in section 4.5. The combination of these various components which constitute the overall background prediction is described in section 4.3.5.

Validation of the combined background prediction is done in several steps. The efficiency correction is validated based on SM MC samples, where we verify that the expected symmetry is restored, as detailed in section 4.6.1. The fake background estimate is validated by comparing MC predictions vs. data in a dedicated same-sign CR, dominated by fake events, as well as in the symmetric baseline selection – a selection that includes the expected signal contribution but is loose enough that it is dominated by background contributions. This is shown in section 4.6.2. Finally, the combined background estimate is validated by inspecting the obtained prediction vs. data in the baseline selection, as is shown in section 4.6.3.

The strategy to achieve high sensitivity in this analysis is, on the one hand, a relatively loose selection to keep high statistics and, on the other hand, constructing an observable that separates signal and background as best as possible. The distribution of this observable is then used as the final discriminant in the fit. To construct such an observable, a deep NN is trained. Further details on the NN implementation are given in section 4.7. As a result, no optimized SR selection is defined, and all of the events which pass the baseline selection are used for the training of the NN and the obtention of results in the statistical analysis. On the other hand, selected events are further divided into two categories: VBF and non-VBF. The VBF category is designed to enhance the sensitivity to the VBF Higgs production mode (see definitions of the SRs in section 4.3.3).

The statistical analysis – used to extract the results – is performed separately for the search for $H \to e\tau_\mu$ and $H \to \mu\tau_e$. It uses a binned likelihood function based on histograms of the NN outputs. A simultaneous fit that combines the non-VBF and VBF categories is done. Details on the implementation are given in section 4.3.6. A discussion of the uncertainties considered and their treatment is given in sections 4.8 and 4.9.1. The results obtained are described in section 4.9.2 for this analysis, while final results in combination with the MC-based analyses are presented in 4.9.3.

4.3.2 Event classification

Event classification refers to assigning an event with precisely two opposite-sign leptons, one an electron and the other a muon, to either of the two mutually exclusive datasets, which we compare when applying the e/μ-symmetry method.

As described in section 3.2, we expect that in the case of the $H \to \mu\tau$ signal, the p_T spectrum of electrons originating from τ decays is softer than the muon p_T spectrum. Based on this assumption, the event classification is implemented following the p_T-ordering of the two leptons, and we define the $e\mu$ and μe datasets:

- $\boldsymbol{\mu e}$ **dataset**: ℓ_0 is the muon, and ℓ_1 is the electron ($p_\mathrm{T}^\mu > p_\mathrm{T}^e$)
- $\boldsymbol{e\mu}$ **dataset**: ℓ_0 is the electron, and ℓ_1 is the muon ($p_\mathrm{T}^e > p_\mathrm{T}^\mu$)

where the leading lepton (the lepton with the larger transverse momentum) is indicated by ℓ_0 and the subleading lepton by ℓ_1. As previously discussed, this method was used in the Run-1 Higgs LFV search also based on the e/μ-symmetry method [57].

But this assumption is only valid in the rest frame of the decaying Higgs boson, which can be different than the laboratory frame due to the eventual boosting of the Higgs. This leads to a non-negligible contribution of wrongly classified signal events. These misclassified signal events negatively impact the sensitivity of this analysis, not only because they are lost to our SR but also contribute instead to the background model. This effect is found to be even worse for VBF signal events as compared to ggF signal events, due to the Higgs being more boosted in average in the VBF case. We define the event classification *accuracy* as the percentage of signal events correctly assigned.

The Higgs rest frame cannot be fully known due to missing information on the z-component of the missing energy. Still, it can be estimated by adding the 4-vectors of the two leptons and the transverse component of the missing energy and by constraining the resulting Higgs mass to 125 GeV. Event classification based on the p_T-ordering of the two leptons, this time boosted to this estimated Higgs rest frame, leads to significantly improved accuracy. It was found that this can be improved even further when evaluating the pseudorapidity component of the missing energy with the pseudorapidity of the system of the two leptons.

Accuracies for the different signal samples used in our analysis, showing the improvement achieved with this refined event classification method, are shown in Table 6. The improved method is the one implemented in this analysis.

	(ggF) $H \to \mu\tau$	(ggF) $H \to e\tau$	(VBF) $H \to \mu\tau$	(VBF) $H \to e\tau$
laboratory frame	85.1	88.7	71.1	78.0
estimated Higgs frame	92.8	93.6	90.5	91.3

Table 6: Event classification accuracy in % for the different signal samples considered, based on the p_T-ordering of the two leptons either in the laboratory frame, or boosted to the estimated Higgs rest frame.

An ideal event classification method means that one can perfectly assign, within a signal event, which lepton, ℓ_H, originates directly from the Higgs decay and which lepton, ℓ_τ, from the secondary τ decay. Therefore we refer to the two mutually exclusive datasets obtained by applying this improved event classification method using the following notations:

- $\boldsymbol{\mu\tau_e}$ **dataset**: ℓ_H is the muon and ℓ_τ is the electron ($p_\mathrm{T}^\mu > p_\mathrm{T}^e$ in the estimated Higgs rest frame)
- $\boldsymbol{e\tau_\mu}$ **dataset**: ℓ_H is the electron and ℓ_τ is the muon ($p_\mathrm{T}^e > p_\mathrm{T}^\mu$ in the estimated Higgs rest frame)

As before, the SM background is split equally between the two datasets. We sometimes refer to these two datasets simply as the $\mu\tau$ and $e\tau$ datasets, implying that the τ further decays to the opposite flavored lepton.

4.3.3 Selections

In this analysis, different event selections for distinct tasks are considered.

The symmetric baseline selection – or baseline selection – is a loose selection that includes the Higgs LFV signal and preserves the $e\mu/\mu e$ SM symmetry. It contains events with one electron and one muon with opposite signs in the final state. The trigger logic to select those events is an `OR` between single-lepton and dilepton triggers. The triggers considered are summarized in Table 7. Symmetric p_T thresholds for electron and muon triggers are chosen to ensure a symmetric selection between electrons and muons and to not artificially violate the symmetry assumption.

Trigger Menu	**Data Period**	**Trigger Chain Name**	**recommended p_T thresholds**	**used p_T thresholds**

Single Electron	2015	`HLT_e24_lhmedium_L1EM20VH`
		`HLT_e60_lhmedium`
		`HLT_e120_lhloose`
			25 GeV	25 GeV

	2016-18	`HLT_e26_lhtight_nod0_ivarloose`
		`HLT_e60_lhmedium_nod0`
		`HLT_e140_lhloose_nod0`
			27 GeV	27 GeV

Single Muon	2015	`HLT_mu20_iloose_L1MU15` `HLT_mu50`	21 GeV	25 GeV
	2016-18	`HLT_mu26_ivarmedium` `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: Triggers used in this analysis. An `OR` between all of them is required. Symmetric p_T thresholds are used instead of the recommended ones to avoid introducing an artificial asymmetry between electrons and muons.

Further event selection criteria applied are:

- τ_{had} veto (to combine results with the $H \rightarrow e\tau_{had}$ and $H \rightarrow \mu\tau_{had}$ searches)
- b-jet veto ($t\bar{t}$ background reduction)
- $p_T^{\ell_0} > 35$ GeV and $30 < m_{\ell\ell} < 150$ GeV (background reduction)

Events that pass the baseline selection are further divided into two categories: the VBF and non-VBF selections. The VBF category is designed to enhance the sensitivity to the VBF production mode. The VBF selection includes events with at least two jets, which pass further dedicated requirements applied to the jet kinematics and topology of the two jets j_0 and j_1 (denoting the leading and subleading jet ordered in p_T, respectively):

- $p_{\mathrm{T}}^{j_0} > 40$ GeV and $p_{\mathrm{T}}^{j_1} > 30$ GeV
- $|\Delta\eta_{jj}| > 3$
- $m_{jj} > 400$ GeV

The non-VBF category contains events failing the VBF selection.

The *same-sign* selection has the same requirements as the baseline selection, except that the leptons are required to have charges of the same sign. This CR, mainly populated by fake events, is used for validating the fake background estimation.

The Z+jets CR is a selection also mainly populated by fake events used in the fake background estimate. It includes events with three baseline leptons: two leptons of the same flavor and opposite sign, *tagged* to the Z boson decay, and a third *probe* lepton which is often a fake lepton. We further require that the tagged pair's invariant mass is between 80 and 100 GeV, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is under 60 GeV, and the invariant mass of the system of the probe lepton with the $E_{\mathrm{T}}^{\mathrm{miss}}$ is under 40 GeV.

4.3.4 Efficiency correction

One of the main focuses of this analysis, based on the e/μ-symmetry method, is to restore the expected symmetry between SM events which contribute to the $e\tau$ and $\mu\tau$ datasets. Asymmetries between these two datasets arise after detection due to different lepton trigger and lepton detection efficiencies. The efficiency correction is applied event-by-event, similar to the description in section 3.3.3. For this, we estimate each event's overall efficiency by measuring the various efficiency components affecting the detection of the electron and the muon in the event.

We emphasize that the efficiency correction only attempts to restore the symmetry between *real* SM contributions to the $e\tau$ and $\mu\tau$ datasets. Another source of asymmetry, the contribution of events containing fake or non-prompt leptons, are accounted for with different methods, as detailed in section 4.3.5.

Event efficiency

We select events with precisely one electron and one muon of opposite sign. The event efficiency is a product of the efficiencies to detect the electron and the muon in the event and to trigger the event:

$$\epsilon_{\mathrm{tot}} = \epsilon_{\mathrm{trig}} \cdot \epsilon_e \cdot \epsilon_\mu \tag{48}$$

Furthermore, the lepton efficiencies have different components for the different steps of the reconstruction algorithms: Reco, Id, and Iso [71, 82].

$$\epsilon_{\ell\in\{e,\mu\}} = \epsilon_\ell^{\mathrm{reco}} \cdot \epsilon_\ell^{\mathrm{Id}} \cdot \epsilon_\ell^{\mathrm{Iso}} \tag{49}$$

Trigger efficiencies are provided by the ATLAS electron and muon trigger working groups [64–67]. We use an inclusive combination of single-electron, single-muon, or di-lepton trigger chains, as detailed in section 4.3.3. The global trigger efficiency of the event is calculated as a function of the kinematics of the two leptons in the event, the relevant single-lepton trigger efficiency values, and combinatoric formulas provided in the ATLAS TrigGlobalEfficiencyCorrection tool.

The single-electron trigger efficiencies are parameterized in terms of the electron's p_{T} and pseudorapidity η. They rise with the electron's p_{T} up to around 80 GeV before reaching a plateau of over 95% efficiency. The efficiencies significantly vary between the barrel and endcap regions of the detector, and separate efficiency maps are provided for the different data-taking years.

The single-muon trigger efficiencies are lower than that of the single-electron trigger, especially in the barrel region of the detector (~60-80%). They are parameterized in terms of the muon's pseudorapidity and azimuthal angle. There is no p_{T} dependence, but the efficiencies are only valid above the trigger's p_{T} threshold value. Separate efficiency maps are provided for the barrel and endcap regions and different data-taking periods.

The dilepton triggers efficiencies are assumed to factorize into the two single-lepton triggers efficiencies, and the properties are described by the behavior of the combined efficiencies.

Muon efficiencies are the product of the muon's Reco, Id and Iso efficiencies [82]. Joint Reco and Id efficiency maps are provided by the ATLAS muon working groups, parameterized by the muon's η and ϕ, and a p_{T} dependence is given to the systematic uncertainty of the efficiency values. Again separate maps are provided for different data-taking periods, but the variations are much smaller than in the case of the single-muon trigger efficiencies.

Muon Iso efficiencies are also provided and depend mainly on the muon's p_{T}. An additional parameter, the distance to the closest jet, is used to further parameterize the systematic uncertainties on the muon Iso SFs.

Electron efficiencies are also a combination of the electron's Reco, Id and Iso efficiencies [71]. These efficiencies were found by the EGamma performance group to depend strongly on the kinematic selection and not only on the properties of the electrons. Therefore, they are not provided in the form of efficiency maps. On the other hand, the SFs - which account for the difference in electron efficiencies between MC and data measured by the performance group - were found to be independent of the selection and are provided.

The recommended procedure to derive electron efficiencies valid in our SR is the following:

- Estimate first the MC electron efficiency from simulated samples; count the number of electrons in the SR which pass the Reco, Id, and Iso criteria. Divide this number by the total number of electrons generated:
$$\epsilon_X^{MC} = \frac{N_{\text{electrons passing cuts X}}}{N_{\text{all electrons}}}, X \subset \{\text{Reco,Id,Iso}\} \tag{50}$$
- Estimate the data electron efficiency by multiplying the measured MC efficiency by the relevant electron SFs:
$$\epsilon_X^{Data} = SF_X \cdot \epsilon_X^{MC} \tag{51}$$

The electron candidates to consider in the first step must be carefully selected based on their *truth* information. In particular, the origin of the electron needs to be matched to the relevant process that was simulated, e.g., a Z boson in the case of a $Z \to ee$ process. The electrons considered are:

- Prompt electrons with matched origin.
- Electrons that underwent bremsstrahlung conversion originating from a prompt electron with matched origin.
- Electrons originating from a Final State Radiation (FSR) photon that converted.

The two latter types are jointly referred to as *background electrons*. The measurement of the electron MC efficiencies is presented in section 4.4.

Efficiency correction strategy

As a reminder, when searching for an $H \to \mu\tau_e$ signal within the $\mu\tau$ dataset, we use the $e\tau$ dataset after efficiency correction to estimate the SM background contributions (in a search for the $H \to e\tau_\mu$ signal, the roles of the two datasets are reversed). The efficiency correction is applied following the single event correction formalism developed in section 3.3.3, where we use the event-by-event efficiency-ratio CF introduced in (42) to the events of the $e\tau$ dataset.

Due to the different efficiency dependencies, the efficiency-ratio CF is relative to whether the event belongs to the $e\mu$ or μe dataset, which differ by the p_{T} ordering of the two leptons in the *laboratory* frame as described in section 4.3.2. In terms of the $e\tau_\mu$ and $\mu\tau_e$ datasets (which differ by the p_{T} ordering of the two leptons in the *estimated Higgs* frame as described in section 4.3.2), the procedure is almost the same. The efficiency correction is still applied only to the dataset used as the background estimate, but the efficiency-CF is still relative to whether the event belongs to the $e\mu$ or μe dataset. For example, if the $e\tau_\mu$ is the basis for the background estimate, the efficiency-CF for the few μe events which contribute to it will be $R_i^\epsilon(\mu e \text{ event})$.

We validate the efficiency correction based on the SM MC simulated samples, as described in section 4.6.1.

4.3.5 Background estimation

Several background processes result in signatures similar to the one expected for the $H \to e\tau$ or $H \to \mu\tau$ LFV signal. They can be distinguished into two broad categories:

- *fake* backgrounds
- *real* SM backgrounds

Events of the first category result from particles or objects misidentified as leptons – electrons or muons – which pass the lepton selection criteria, therefore allowing these events to be selected. We refer to these leptons as fake leptons and the events selected, including them as fake events. The estimation of the fake background contribution is separated in two depending on the source of the fake leptons; a data-driven method called the Fake-Factor (FF) method for the main contribution (jets faking leptons) and directly from MC for other sources. We denote as FF-fakes (F_{FF}) the fakes estimated with the FF method, and as MC-fakes (F_{MC}) the fakes estimated from MC. The fake estimate is presented in section 4.5. It is validated in the same-sign selection – a fake enriched CR defined in section 4.3.3 – as shown in section 4.6.2.

Events of the second category result from SM processes, producing the same numbers and types of particles as the signal process. They include $Z \to \tau\tau$, diboson (WW, WZ, ZZ), top-quark production ($t\bar{t}$ and single top), and SM processes involving the Higgs boson: $H \to WW$ and $H \to \tau\tau$. As detailed in section 3.1, these SM processes are symmetric to the exchange $e \leftrightarrow \mu$, and their contributions are evaluated using the e/μ-symmetry method. In particular, the efficiency correction described in section 4.3.4 is applied to the data samples from which the fake contribution is subtracted. We denote the resulting estimate as the *SM symmetric background*. Based on the formalism developed in (46), it is written as:

$$\begin{cases} \tilde{N}^{e\tau}_{\text{sym}} = R^{\epsilon} \cdot (N^{e\tau} - F^{e\tau}_{\text{FF}} - F^{e\tau}_{\text{MC}}) & (H \to \mu\tau \text{ search}) \\ \tilde{N}^{\mu\tau}_{\text{sym}} = R^{\epsilon} \cdot (N^{\mu\tau} - F^{\mu\tau}_{\text{FF}} - F^{\mu\tau}_{\text{MC}}) & (H \to e\tau \text{ search}) \end{cases} \tag{52}$$

where $N^{e\tau}$ ($N^{\mu\tau}$) is the data, $F^{e\tau}_{\text{FF}}$ ($F^{\mu\tau}_{\text{FF}}$) is the FF-fakes, and $F^{e\tau}_{\text{MC}}$ ($F^{\mu\tau}_{\text{MC}}$) is the MC-fakes – in the $e\tau$ ($\mu\tau$) dataset. The efficiency-ratio CF R^{ϵ} is applied event-by-event and differs whether the event is an $e\mu$ or μe event, as described in section 4.3.4. The combined background prediction is then written as:

$$\begin{cases} b^{\mu\tau} = \tilde{N}^{e\tau}_{\text{sym}} + F^{\mu\tau}_{\text{FF}} + F^{\mu\tau}_{\text{MC}} & (H \to \mu\tau \text{ search}) \\ b^{e\tau} = \tilde{N}^{\mu\tau}_{\text{sym}} + F^{e\tau}_{\text{FF}} + F^{e\tau}_{\text{MC}} & (H \to e\tau \text{ search}) \end{cases} \tag{53}$$

Validation of the combined background estimate is shown in section 4.6.3.

4.3.6 Statistical analysis

The statistical analysis is used to extract the results. It is performed separately for the search for $H \to e\tau_{\mu}$ and $H \to \mu\tau_e$, under the assumption that the other LFV decay's BR is zero. As previously discussed, the final discriminant used in the analysis is the output histogram of the NN scores, implemented to enhance the signal over background separation (see section 4.7).

The statistical analysis uses a binned likelihood function, constructed as a product of Poisson probability terms – one for each bin in the distribution of the NN outputs. A simultaneous fit that combines the non-VBF and VBF categories is done. Two measures of interest are derived: a) the significance of a signal, which quantifies the disagreement with the background-only hypothesis, and b) an upper limit on the cross-section times BR of the signal process, which quantifies the maximum signal strength consistent with the data.

The likelihood function for the $H \to \mu\tau_e$ search, based on the formalism introduced in (47), is written as:

$$L(\mu^s) = \prod_{b=1}^{B} \text{Pois}\left(N^{\mu\tau}_b \,\middle|\, \tilde{N}^{e\tau}_{\text{sym},b} + F^{\mu\tau}_{\text{FF},b} + F^{\mu\tau}_{\text{MC},b} + \mu^s \cdot \left(S^{\mu\tau}_b - \tilde{S}^{e\tau}_b\right)\right) \tag{54}$$

where B is the total number of bins in the histograms considered, $N^{\mu\tau}$ is the $\mu\tau$ data, $\tilde{N}^{e\tau}_{\text{sym}}$ is the SM symmetric background (52), F_{FF} and F_{MC} are the FF-fakes and MC-fakes, $S^{\mu\tau}_b$ is the searched-for $H \to \mu\tau_e$ signal, and μ^s is the signal strength parameter. We also include here the signal contamination $\tilde{S}^{e\tau}$: the $H \to \mu\tau_e$ events, which are wrongly classified to the $e\tau$ dataset (see section 4.3.2). Due to symmetry considerations, this contribution needs to be included after efficiency correction, similarly to the $e\tau$ fake contributions. Likewise, for the $H \rightarrow e\tau_\mu$ search, the likelihood function is:

$$L(\mu^s) = \prod_{b=1}^{B} \text{Pois}\left(N_b^{e\tau} \,\middle|\, \tilde{N}_{\text{sym},b}^{\mu\tau} + F_{\text{FF},b}^{e\tau} + F_{\text{MC},b}^{e\tau} + \mu^s \cdot \left(S_b^{e\tau} - \tilde{S}_b^{\mu\tau}\right)\right) \tag{55}$$

Not shown in the above expressions, systematic uncertainties – expressed as a set of nuisance parameters – are considered as additional Gaussian constraint terms (limited by the upward and downward fluctuated distributions per nuisance parameter and per sample) in the likelihood function. All considered systematic uncertainties are summarized in section 4.8. In addition, statistical uncertainties of the total background expectations are considered bin-by-bin via additional Poisson constraint terms in the likelihood. The corresponding nuisance parameters – the γ-parameters – are parameterized to have a mean value of 1 and a variance corresponding to the relative statistical uncertainty.

The significance and the upper limit on the BR are derived using the test statistics q_0 and $\tilde{q}_\mu$ respectively, which are described in section 2.6.

4.4 Electron efficiency measurements

As discussed in section 4.3.4, the electron efficiencies – used in the efficiency correction – depend strongly on the kinematic selection and not only on the properties of the electrons. On the other hand, the SFs are independent of the selection. As per the recommendation, we measure electron efficiencies valid in our selection from MC and translate them to data efficiencies via the provided SFs.

In the following, we present the measurements of the electron MC efficiencies valid in our selection. In particular, $Z \rightarrow ee$ samples aren't included in these measurements since they don't contribute to our SR. In contrast, they are the dominant contribution to electron efficiencies and SFs derived by the EGamma performance group. We describe the method used in section 4.4.1 and validate it based on $Z \rightarrow ee$ samples in section 4.4.2. Results and strategies implemented to boost the MC statistics are shown in section 4.4.3, and the derivation of uncertainties is described in section 4.4.4. A study comparing efficiencies measured for leading vs. subleading electrons is presented in section 4.4.5. In section 4.4.6, we compare the efficiencies obtained for our selection to those obtained from the $Z \rightarrow ee$ samples. Finally, efficiencies measured when tighter constraints on the electrons are applied are presented in section 4.4.7.

Validation of the measured electron efficiencies and the efficiency correction (described in section 4.3.4) is shown in section 4.6.1 based on the SM MC sample and in section 4.6.3 based on the data.

4.4.1 Description

Samples
One of the technical challenges in measuring electron efficiencies from MC samples is to circumvent any bias induced by preselection constraints applied to the input data files. For instance, the HIGG4D1 DAODs which we use in our analysis only include events with at least two loosely identified leptons. Therefore it isn't possible to provide unbiased electron efficiency measurements from them.

The next obvious candidates are the EGAM1 DAODs, which are used by the performance group to measure electron efficiencies and provide electron SFs. Indeed, EGAM1 preselection requires at least two loosely identified electrons, or one electron and one photon. Therefore these can be used to obtain unbiased electron efficiency measurements. On the other hand, the EGAM1 DAODs do not include all the SM processes which contribute to our SR, nor events with one electron and one muon which are relevant to us. Therefore it isn't possible to provide electron efficiency measurements in a selection close to our analysis selection from them.

Ultimately, we derive the electron MC efficiencies using the TOPQ1 DAODs. With a preselection requiring at least one loose lepton with $p_{\mathrm{T}} > 20$ GeV, and all the relevant SM processes available, these are adequate for measuring unbiased electron efficiencies in a selection close to our analysis selection.

The SM processes we consider to measure the electron MC efficiencies are Top, $Z \to \tau\tau$, and Diboson. These are the dominant processes with *real* SM contributions to our SR. We refer to their joint contribution as the *SM sample*. In particular, we don't include $Z \to \ell\ell$ samples for this measurement since it isn't a *real* SM background in our analysis (given that we only include events with one electron and one muon).

Electron candidates

Electron candidates used in the electron MC efficiency measurements are truth-matched electrons with $p_{\mathrm{T}} > 15$ GeV and $|\eta| < 2.47$, measured away from the crack-region ($1.37 < |\eta| < 1.52$). Truth-matching refers to looking at a reconstructed particle's information within the simulated event generation origin, such as which type of particle it corresponds to or from which mother particle it decayed. This information is only available in MC simulated samples. Truth-matching is applied here to select only prompt electrons and is done using the ATLAS IFFTruthClassifier tool. We accept the following electron categories:

- Prompt electrons with matched origin
- Background electrons or photons which originate from a prompt electron that underwent electron-photon conversion and with matched origin
- Background electrons that originate from FSR photons

Electron candidates can also be found in the PhotonContainer, meaning they were reconstructed as photons but are true electrons. Including electrons reconstructed as photons means that the denominator in (50) is filled with all the reconstructed clusters found in the EMCal. Ideally, we would include *all* real electrons produced in the simulated collision in the denominator, but we are missing electrons whose cluster was not reconstructed. Still the cluster reconstruction efficiency is > 99% for electrons with $p_{\mathrm{T}} > 15$ GeV (see Figure 12), so it is only a small effect. These electron candidates are used to measure the electron Reco efficiency. If there is more than one electron candidate in a given event, we apply a simple Overlap Removal among them based on the $|\Delta R| < 0.2$ criteria. We then prioritize reconstructed electrons over reconstructed photons, then the highest p_{T}.

Event selection

Events considered for measuring electron efficiencies contain either two electron candidates, or one electron candidate and one muon, with $p_{\mathrm{T}} > 15$ GeV, $|\eta| < 2.47$. Selection constraints are applied to approach the definitions of our baseline selection discussed in section 4.3.3:

- opposite sign leptons away from the crack-region ($1.37 < |\eta| < 1.52$)
- Symmetric trigger coverage (same p_{T} thresholds for electrons and muons) (no trigger matching)
- τ_{had} veto

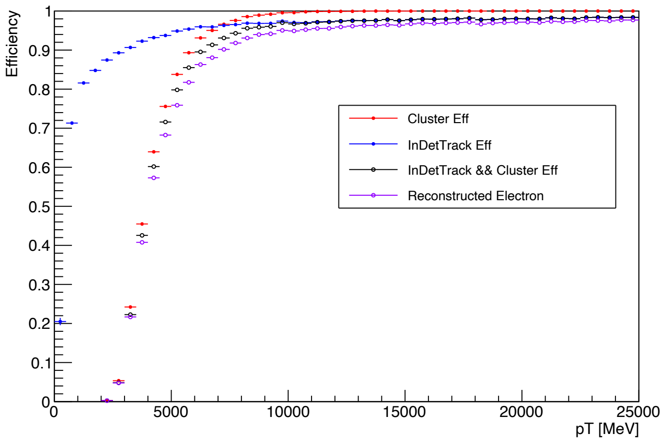


Figure 12: Electron cluster, track, and reconstruction efficiencies. This plot was provided by the ATLAS EGamma working group.

- b-jet veto
- $30 < m_{\ell\ell} < 150$ GeV

Reconstructed muons are required to be truth-matched (using the IFFTruthClassifier) to prompt muons. Electrons are truth-matched, as discussed above.

Tag&probe

We use a *tag&probe* method to obtain an unbiased sample of electrons (the *probes*). Indeed, due to the requirements applied in the input files to include events, we know that every event contains at least one loose lepton with $p_T > 20$ GeV, so the overall sample of electrons available is biased. The tag&probe method is used to identify and reject the lepton which originated in its event selection at the time of the creation of the input file – the *tag* lepton. The second lepton – the *probe* lepton – is then unbiased by this original selection. The *tags* are baseline leptons with $p_T > 20$ GeV. In the case of μe events, the muon is a tag candidate, and the electron is a probe candidate, and in the case of ee events, both electrons are tag and probe candidates in turn. If a tag candidate satisfies the tag requirements, its paired probe electron will be used to measure electron efficiencies.

Electron working points

All electron probes are included in the denominator of (50). To be included in its numerator, they need to satisfy the constraints associated with the following working points, which are the constraints used to define the signal electrons used in the analysis:

- Reco: Reconstructed as an electron and passes TrackQuality criteria as defined by the EGamma performance group (enough sensor hits, not too many dead sensors)
- Baseline: Passes the electron Loose Id requirements, standard overlap removal (with all objects in the event), and relaxed impact-parameter cuts (see section 4.2)
- Id: Passes the electron Medium Id requirements

- Iso: Passes the electron Gradient Iso requirements

Parameterization
We derive 2D electron MC efficiency maps as a function of the electron's p_{T} and η. The binning used is the same as the binning for the electron SFs. We investigate separate maps for the following:

- Different data-taking years
- Electron or muon tags
- Leading or subleading electron probes

Due to limited statistics, this split parameterization comes with large uncertainties. We discuss below some merging strategies that we considered and used to bypass this limitation.

Uncertainties
The statistical uncertainties of the separate numerator and denominator of the efficiency measurement are simply the statistical uncertainties of the weighted count of probe electrons included in each. But the uncertainty of the numerator and the denominator are correlated. We estimate the statistical uncertainty on an efficiency value ϵ measured with N_{den} electrons included in its denominator by $\sqrt{\epsilon(1-\epsilon)/N_{\mathrm{den}}}$, based on the binomial approximation.

We assign systematic uncertainties to the electron efficiency measurements from two different sources: the tag flavor (electron or muon) and the kinematic selection. We then combine them to provide one systematic uncertainty to be used in the fit. Further details are given in section 4.4.4.

4.4.2 Validation in $Z \to ee$

We validate our measurements of the electron MC efficiencies by comparing them to measurements from the EGamma ATLAS performance group. The measurements performed by EGamma are described in [71]. For this measurement, we derive efficiencies from the same input used by EGamma, the DAOD_EGAM1 $Z/\gamma^* \to ee$ samples, and match their selection. The main differences with the selection used to measure efficiencies for our analysis are:

- Use only $Z \to ee$ samples
- Use only ee events with $75 < M_{\ell\ell} < 105$ GeV
- Use standard impact parameter cuts on the electrons
- Tag electrons have p_{T}>25 GeV (instead of 20 GeV for our efficiencies)

Plots showing the comparison of our measurements to reference values obtained from EGamma are shown in appendix A.1. The values obtained agree within uncertainties.

4.4.3 Results and merging schemes

In this section, we present the electron MC efficiencies used in the analysis for the efficiency correction, measured from the combined SM sample in the baseline selection.

As previously discussed, MC statistics are limited for these measurements, mainly due to the exclusion of $Z \to ee$ samples which do not contribute to our SR. Towards obtaining measurements with sufficient statistics, we consider different merging schemes of the separate efficiency maps:

- Use $|\eta_{el}|$ instead of η_{el}
- Release the $p_{\mathrm{T}}(\ell_0) > 35$ GeV cut
- Combining data-taking years
- Electron and muon tags
- Leading and subleading electron probes

A merging scheme is considered potentially adequate at first if the efficiency values of the to-be merged maps are more or less consistent up to statistical uncertainties. But in the end, the determining criteria is how well the symmetry is restored when using the merged efficiencies in the efficiency correction of the $e\mu$ and μe datasets.

The most apparent merging scheme we consider is the use of $|\eta_{el}|$ instead of η_{el} (Figure 13). Indeed the measured efficiency values found in a specific η_{el} bin are almost always consistent with those found in the corresponding $-\eta_{el}$ bin due to the geometrical symmetry of the detector.

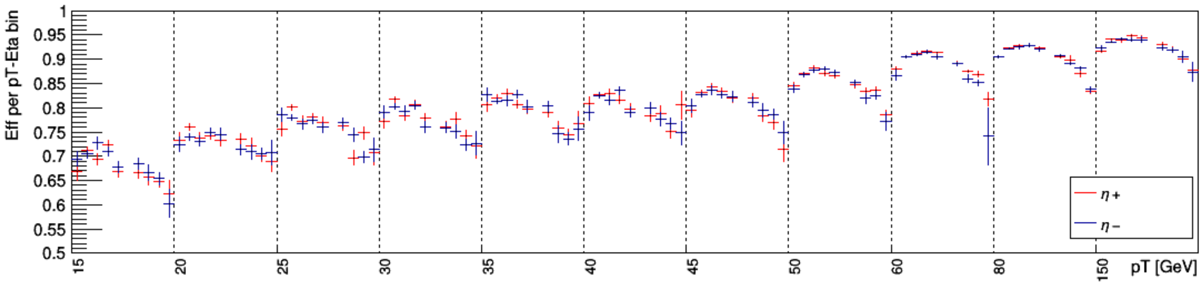


Figure 13: Measured electron MC efficiencies per p_{T} and η bin. The η bin efficiencies are displayed within each p_{T} bin ($0 < \eta < 2.47$), positive η bins in red and their respective negative η bins in blue. The efficiencies here use combined data-taking years, electron and muon tags, leading and subleading probes, and no $p_{\mathrm{T}}(\ell_0) > 35$ GeV cut.

Releasing the $p_{\mathrm{T}}(\ell_0) > 35$ GeV cut (Figure 14) leads to efficiency values with lower statistical uncertainties and which are consistent with the efficiency values which are measured with this cut. Therefore this cut is dropped when measuring the electron MC efficiencies.

Regarding the different data-taking years (Figure 15), we observe that efficiency values for 2017 and 2018 are consistent, while those for 2015+16 are usually higher. This is expected since the electron Id efficiencies were found to decrease with higher pileup, as seen in Figure 16 of [71]. Still, due to low statistics, we find that merging the efficiency maps per year instead of using separate ones leads to a better-restored symmetry after efficiency correction.

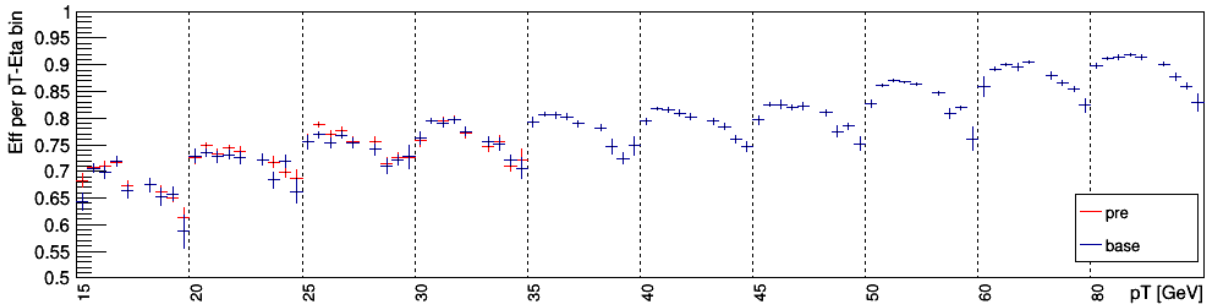


Figure 14: Measured electron MC efficiencies per p_{T} and η bin, with (*base*) and without (*pre*) the $p_{\mathrm{T}}(\ell_0) > 35$ GeV cut. The efficiencies here use combined data-taking years, electron and muon tags, leading and subleading probes, and $|\eta_{el}|$ bins.

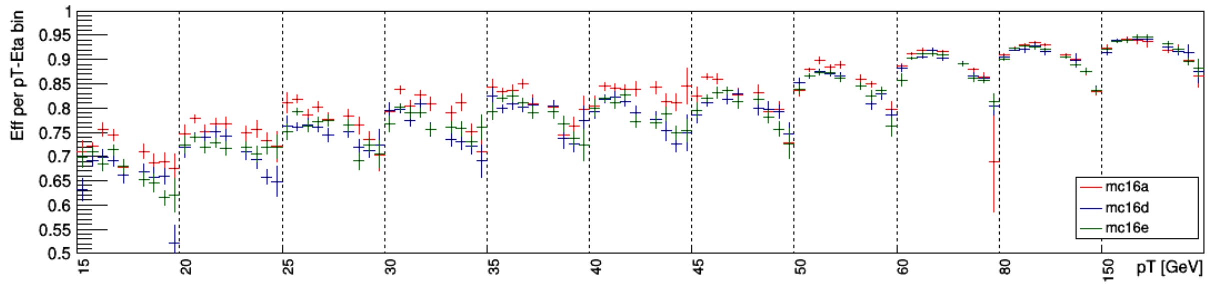


Figure 15: Measured electron MC efficiencies per data-taking year and per p_{T} and η bin. The labels mc16a, mc16d, and mc16e are the MC campaigns corresponding to data-taking years 2015+16, 2017, and 2018 respectively. The efficiencies here use combined electron and muon tags, leading and subleading probes, $|\eta_{el}|$ bins, and no $p_{\mathrm{T}}(\ell_0) > 35$ GeV cut.

For the tag lepton's flavor, in principle, we should use muons since our SR contains events with one electron and one muon. But we expect that within our SM sample (which excludes $Z \to ee$ samples), $e\mu$ events and ee events have similar topology. Therefore we consider including also ee events for the electron efficiency calculation, meaning using electron tags, which would lead to significantly increased statistics. In Figure 16, we compare efficiencies obtained from events with muon tags versus electron tags. A slight tendency for higher values in the case of electron-tagged efficiencies is observed, especially for p_{T} bins higher than 50 GeV. However, including also electron-tagged events in the efficiency measurement leads to improved restored symmetry. Therefore, electron and muon tagged events are combined for the electron MC efficiency measurement.

Finally, in Figure 17, we compare efficiency values measured from electron probes whether the electron is the leading vs. subleading lepton in the event. We find that leading electron efficiencies are systematically higher than subleading electron efficiencies (within the same p_{T} and η bin), with differences going up to 5% in efficiency for some bins. Using *separate* maps for leading and subleading electron probes leads to improved restored symmetry; hence the MC efficiencies, in this case, are kept separate.

In the end, we chose to use efficiency maps parameterized as a function of the electron's p_{T} and $|\eta|$, corresponding to the full 2015-18 data-taking period, using both electron and muon tags and without the $p_{\mathrm{T}}(\ell_0) > 35$ GeV cut, but separate for leading and subleading electrons. These are the efficiencies shown in Figure 17, keeping in mind that for the leading electron efficiency maps, only $p_{\mathrm{T}} > 35$ GeV bins are

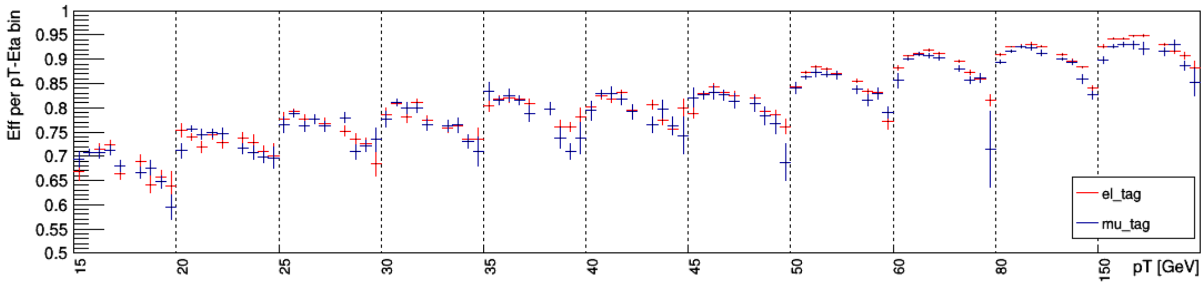


Figure 16: Measured electron MC efficiencies per tag lepton's flavor and per p_{T} and η bin. The efficiencies here use combined data-taking years, leading and subleading probes, and $|\eta_{el}|$ bins, and no $p_{\mathrm{T}}(\ell_0) > 35$ GeV cut.

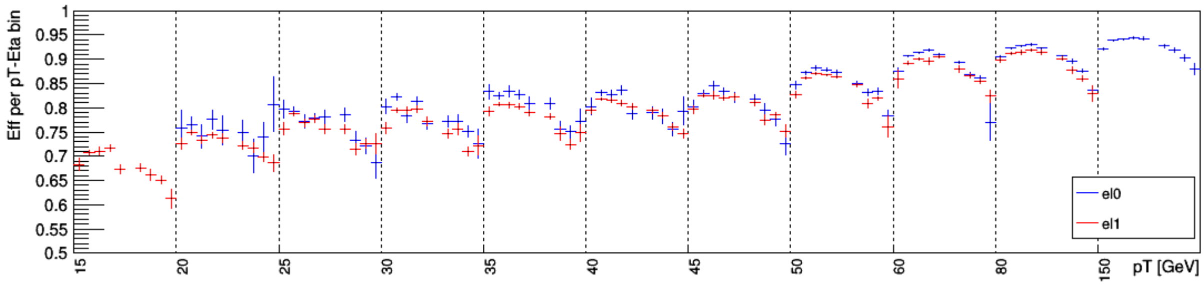


Figure 17: Measured electron MC efficiencies per leading (blue) versus subleading (red) electrons and per p_{T} and η bin. The efficiencies here use combined data-taking years, electron and muon tags, $|\eta_{el}|$ bins, and no $p_{\mathrm{T}}(\ell_0) > 35$ GeV cut. The last p_{T} bin in the subleading electron map is $p_{\mathrm{T}} > 80$ GeV, and in the leading electron map, $p_{\mathrm{T}} > 150$ GeV.

relevant. Using this scheme, we show in Figure 18 the obtained electron MC efficiencies per electron working point projected on the p_{T} and η axes.

4.4.4 Uncertainties

To estimate systematic uncertainties in the electron MC efficiency measurements, we vary the flavor of the tag lepton (electron or muon) and the kinematic selection (applying or not the $p_{\mathrm{T}}(\ell_0) > 35$ GeV cut). This approximately mirrors systematic uncertainties derived by the EGamma group obtained from varying the tag Id criterion for the first and the dilepton mass window range for the latter.

In the case of the tag's flavor, the difference between efficiencies measured from electron tagged and muon tagged events is shown in Figure 16. The nominal values use combined electron and muon tags, and the systematic uncertainty per bin is set as the largest difference between |nominal - electron tag| and |nominal - muon tag| efficiencies. This uncertainty is set symmetrically as both upward and downward variations.

In the case of the kinematic selection, the difference between efficiencies measured with or without applying the $p_{\mathrm{T}}(\ell_0) > 35$ GeV cut is shown in Figure 14. The nominal values are measured without the cut (in order to boost statistics), although in our base selection, this cut is applied. The systematic uncertainty per bin is set as the absolute difference between the two measurements, set symmetrically as both upward and downward variations. This only affects efficiency measurements of subleading electrons with $p_{\mathrm{T}} < 35$ GeV.

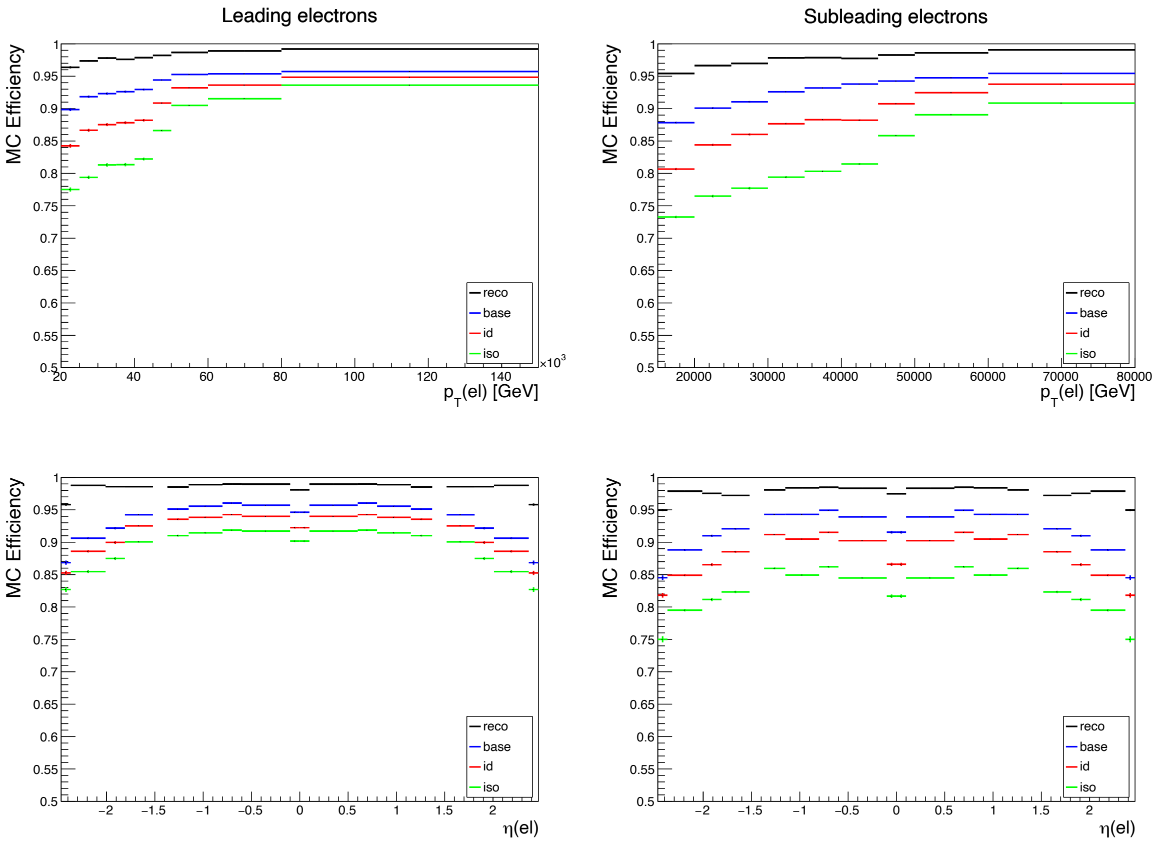


Figure 18: Electron MC efficiencies per electron working point projected on the p_{T} and η axes. Separate plots are shown for leading and subleading electrons. Using combined data-taking years, electron and muon tags, and $|\eta_{el}|$ bins. The last p_{T} bins (overflow) are not shown.

In Figure 19, we show the resulting systematic bands per bin with respect to the nominal values for the leading and subleading electron MC efficiency maps and the tag's flavor, selection, and combined systematics.

4.4.5 Leading vs. subleading electrons

In an attempt to understand better the origin of the differences in efficiency values measured from leading vs. subleading electrons, shown in Figure 17, we derived a similar comparison when using the DAOD_EGAM1 $Z/\gamma^* \to ee$ samples used in the efficiency measurement validation described in section 4.4.2.

As is shown in Figure 20, here as well, the values can differ significantly whether the electron probe is the leading or subleading electron in the event. We find differences mainly in the $p_{\mathrm{T}} < 40$ GeV range, where the leading electron efficiency values can be higher by up to 5% efficiency. We also compared leading and subleading efficiencies per working point (Reco, Id, and Iso) – the individual plots are shown in appendix A.2. We find that this difference originates mainly from the Iso constraint but is also observed when requiring Id.

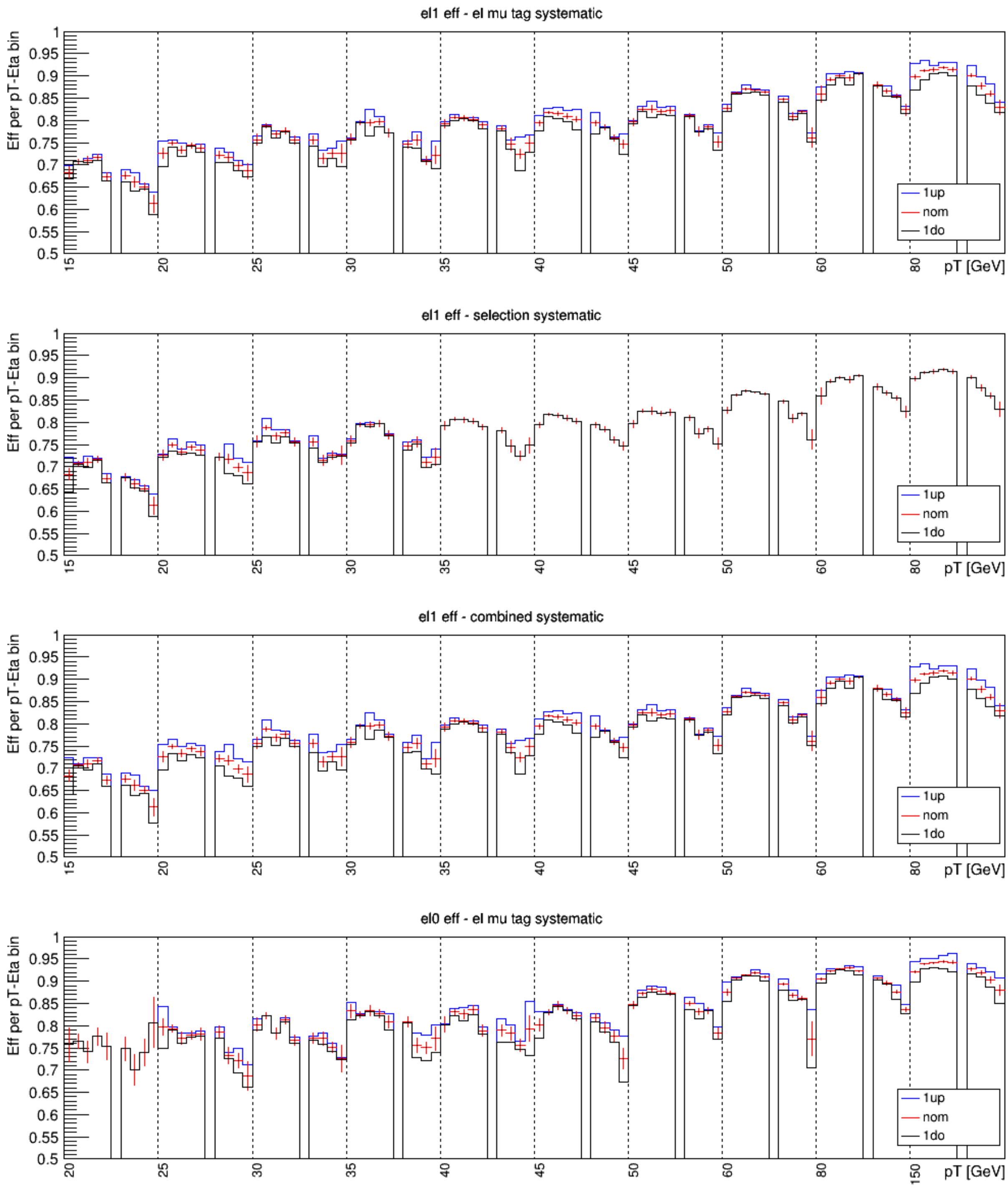


Figure 19: Electron MC efficiency measurements used in the analysis. Nominal values and statistical errors are in red, and upward and downward systematic variations are in blue and black. The top three plots are for the subleading electron efficiencies, showing the tag's flavor, selection, and combined systematic variations. The last plot is for the leading electron efficiencies, showing the tag's flavor systematic variations only since these efficiencies are not affected by the selection systematic. For this last map, only the bins with $p_{\mathrm{T}} > 35$ GeV are used in our analysis.

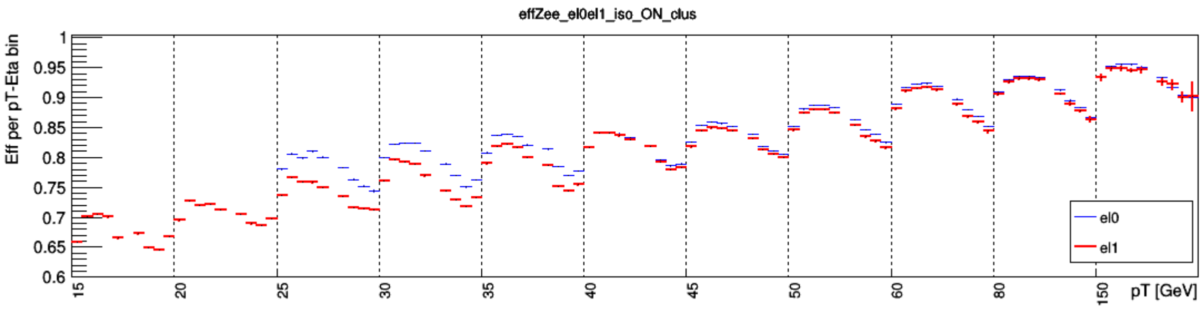
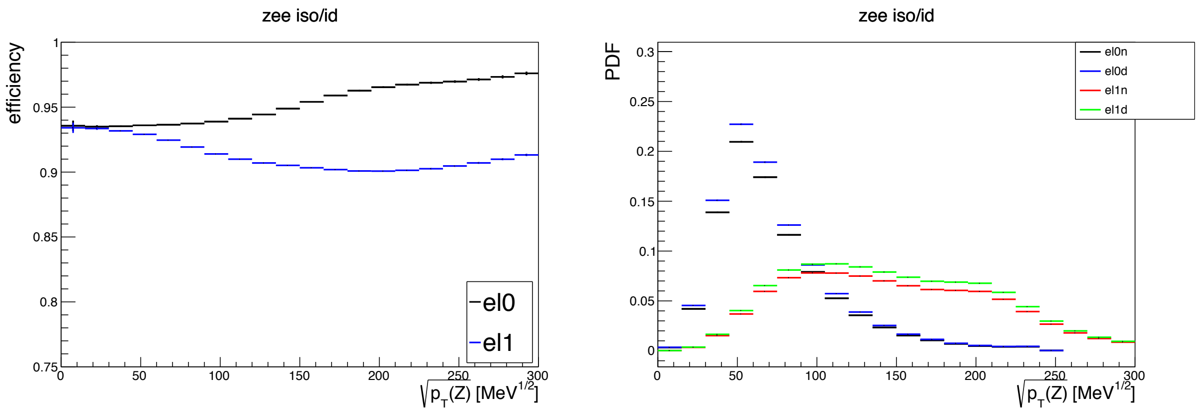


Figure 20: Measured electron MC efficiencies using the reference $Z \to ee$ samples per leading (blue) versus subleading (red) electrons and per p_{T} and η bin. The efficiencies here use combined data-taking years and $|\eta_{el}|$ bins.

To further investigate, we looked at the $Z \to ee$ Iso efficiencies as a function of the kinematics of the dielectron (probe + tag) system. In Figure 21(a), we show the comparison of leading vs. subleading efficiencies measured as a function of the p_{T} of the Z boson (measured from the two electrons it decayed to). We find that when $p_{\mathrm{T}}(Z) = 0$, the efficiencies agree, but the difference observed increases the larger $p_{\mathrm{T}}(Z)$ is. This hints at some correlation with the hadronic activity within the underlying event.

Another check performed was to look at dielectron distributions when the electron probe – either the leading or the subleading electron – is constrained to a specific p_{T} and η bin (constraining one electron fully constrains the dielectron system since they both decay from the Z boson). An example of $p_{\mathrm{T}}(Z)$ distributions in the $25 < p_{\mathrm{T}} < 35$ GeV and $0.1 < \eta < 0.6$ bin is shown in Figure 21(b). We find that the underlying distributions of dielectron kinematics are very different whether the probe is the leading or subleading electron, which leads to the differences in efficiency values observed.

More studies are needed to fully understand these effects; for example, one could look at efficiencies as a function of the number of jets near the electrons or the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft terms. Still, this study shows that the difference in leading vs. subleading electron efficiencies is a genuine effect, which justifies using separate efficiency maps when applying the efficiency correction.



(a)

(b)

Figure 21: (a) Leading vs. subleading $Z \to ee$ Iso MC efficiencies measured as a function of the dielectron (probe + tag) p_{T}. (b) Normalized dielectron p_{T} distributions from $Z \to ee$ when the leading vs. subleading probe electron is constrained to the $25 < p_{\mathrm{T}} < 35$ GeV and $0.1 < \eta < 0.6$ bin. In the labels, n stands for numerator (Iso electrons) and d for denominator (Id electrons).

4.4.6 Comparison to $Z \rightarrow ee$

As detailed in section 4.3.4, the electron efficiencies are not provided by EGamma since they were found to depend strongly on the kinematic selection and not only on the properties of the electrons. In Figure 22, we compare the electron MC efficiencies measured for our analysis to those measured with the $Z \rightarrow ee$ samples in order to show the differences observed. We find that, indeed, the efficiency values can differ significantly, with a difference of up to 4-5% in specific bins, mainly in the $p_T < 25$ GeV and $p_T > 45$ GeV ranges. The same comparison for the individual working points is shown in appendix A.3.

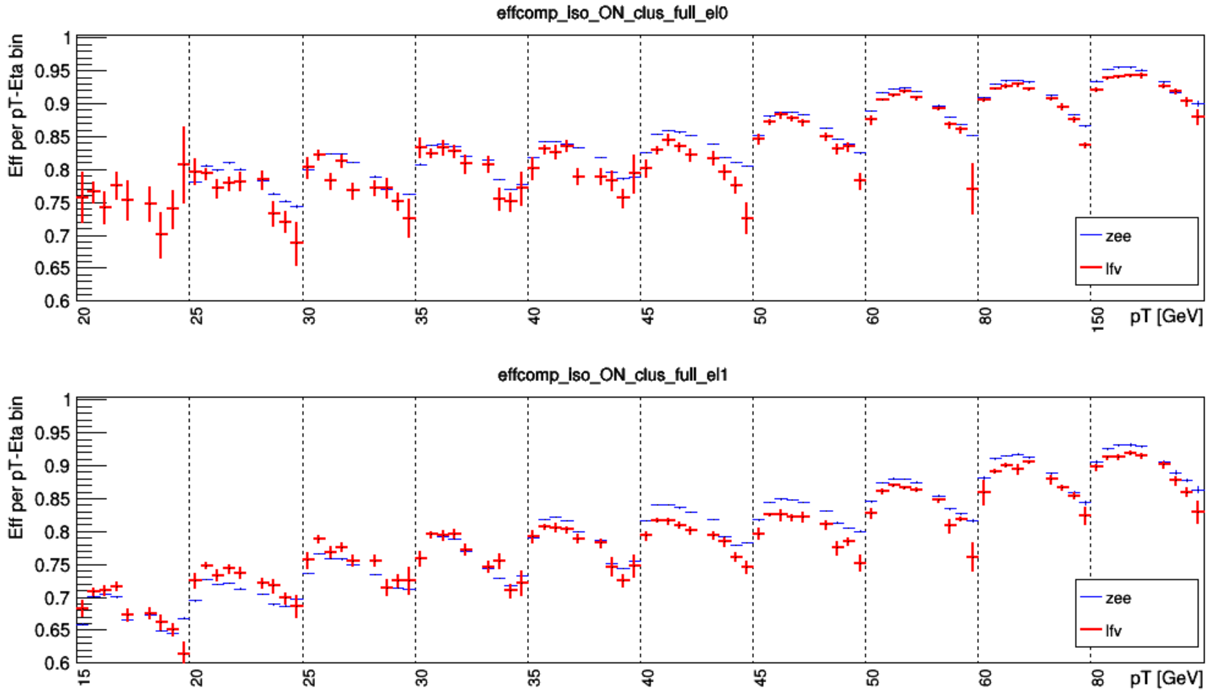


Figure 22: Comparison of LFV vs. Zee electron combined efficiencies for leading and subleading electron probes.

4.4.7 Alternative electron constraints

During the development of this analysis, we considered applying additional selection cuts, which our partner MC-based analysis used to suppress, particularly the $Z \rightarrow \mu\mu$ background. This $Z \rightarrow \mu\mu$ contributes to our SR in the $\mu\tau_e$ dataset due to low-energy muons misidentified as electrons (see details in section 4.5.3). The tighter requirements considered, applied to the electrons, are the following:

- $d0\mathrm{sig}(e) < 5$
- $0.2 < p_T^{\mathrm{track}}/p_T^{\mathrm{cluster}}(e) < 1.25$

Applying these cuts significantly affects the electron efficiencies. Therefore dedicated measurements were performed, and the symmetry after efficiency correction was inspected. These studies are recorded in appendix A.4.

In the end, due to the small impact on the sensitivity of the analysis, these additional requirements weren't implemented. Still, the fact that the efficiency correction successfully restores the symmetry in significantly different selections is a further endorsement of the e/μ-symmetry method.

4.5 Fake background

The fake lepton background is the contribution of events where one or two of the leptons reconstructed in the event is a *fake* lepton. Fake leptons are usually jets misidentified as leptons but can come from other sources as well. Since the probability of faking an electron or a muon is different and varies with the kinematic properties of the lepton, we don't expect the fake background to contribute evenly to the $e\tau$ and $\mu\tau$ datasets. Therefore estimating precisely the fake background is key to restoring the expected symmetry, one of the main assumptions in this analysis.

To estimate the fake background contribution originating from jets faking leptons, we use a data-driven technique called the FF method, standardly used in ATLAS analyses – for example in the $H \to WW^* \to \ell\nu\ell\nu$ production cross-section measurement [83]. This technique measures certain properties from a dedicated fake-enriched data selection, in our case, the Z+jets CR (defined in section 4.3.3), which are extrapolated to the baseline or same-sign selections in order to estimate their fake background contribution, as described in section 4.5.1. It is more accurate than using the fake background estimate obtained from simulated MC samples, for instance, W+jets samples since fake estimation in MC doesn't reflect precisely the actual fake distributions.

Still, MC samples are used to estimate the dominant uncertainty on the jet fakes estimate, the fake flavor composition uncertainty, which originates from the difference in the relative abundance of fakes originating from various sources. This is described in section 4.5.2.

Other sources of fake leptons are estimated separately and directly from MC simulation due to being present in the baseline selection and same-sign CR but absent from the Z+jets CR:

- Electron fakes from photon conversions
- Electron fakes from misidentified muons or prompt muon conversions
- Electron and muon fakes from hadronic τ decays

This is described in section 4.5.3. We jointly refer to the combined contribution of the fake background estimated from MC as the MC-fakes and the more dominant contribution of jet fakes estimated with the data-driven FF method as the FF-fakes.

The fake contributions are estimated separately in the baseline selection and in the same-sign CR, which is used for validation and the $e\tau$ and $\mu\tau$ datasets. Their validation is done by comparing MC + fake prediction to the data and is presented in section 4.6.2.

4.5.1 Jet fakes estimate

The FF is a transfer factor measured from data in the Z+jets CR. It is calculated as the ratio of the number of so-called Id versus anti-Id leptons. The Id-leptons are the signal leptons that pass all the selection criteria used in the analysis. In contrast, anti-Id leptons are leptons that fail to pass some of these selection criteria. The method to estimate the contribution of fake electrons and fake muons is the same, except for the requirements defining the anti-Id leptons:

- Id electrons and muons pass medium Id and Iso criteria
- anti-Id electrons pass loose Id but not medium Id or not Iso criteria
- anti-Id muons pass medium Id but not Iso criteria

The assumption is that for any given source of lepton fake, the number of Id versus anti-Id fakes is proportional and independent of other selection criteria. The choice of the Z+jets CR to measure the FF is to use a selection kinematically close to our SR and containing a high number and purity of fakes. The FF is then applied in the baseline or same-sign region event by event as a SF to all the data events that include an anti-Id lepton, constituting the fake background estimate.

The FF is measured from data in the Z+jets CR. As described in section 4.3.3, it consists of events with three leptons, two leptons are tagged to the Z and the third one, the *probe* lepton, has a high probability of being a fake. Although the fake purity is high in this CR, some of the Id and anti-Id leptons found can actually be prompt leptons or fake leptons from the MC-fakes category. To correct this, we subtract these contributions as obtained from MC simulation, which contaminate both the Id and anti-Id fake lepton samples. Similarly, when applying the FF in either the baseline or same-sign regions, both real and MC-fake contributions are subtracted. In Figure 23, we show the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $m_{\mathrm{T}}(\ell_{\mathrm{probe}})$ distributions of the probe leptons in a loosened Z+jets selection (without applying the $E_{\mathrm{T}}^{\mathrm{miss}} < 60$ GeV and $m_{\mathrm{T}}(\ell_{\mathrm{probe}}) < 40$ GeV cuts). The gap between the data and the MC prediction illustrates the jet faking leptons contribution in this selection.

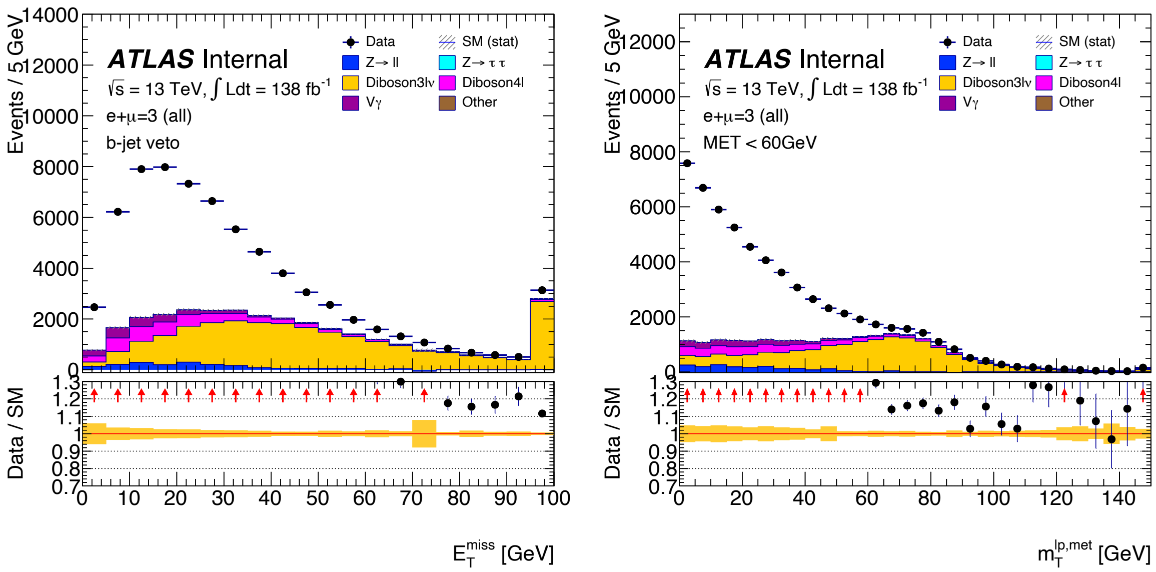


Figure 23: Distributions of the missing transverse energy (left) and of the transverse mass distribution of the probe lepton (right) before the application of the respective requirements for the Z + jets-CR ($E_{\mathrm{T}}^{\mathrm{miss}} < 60$ GeV and $m_{\mathrm{T}}(\ell_{\mathrm{probe}}) < 40$ GeV).

FFs are extracted from the $Z+$jets-CR by subtracting the MC-based backgrounds from the data and then building the ratio of the Id leptons and the anti-Id leptons selections:

$$FF = \frac{N_{\text{id}}^{\text{data}} - N_{\text{id}}^{\text{MC-based bkgs}}}{N_{\text{anti-Id}}^{\text{data}} - N_{\text{anti-Id}}^{\text{MC-based bkgs}}}. \tag{56}$$

FFs are calculated for electrons and muons separately. They are binned in p_{T} and, in the case of electrons, in $\Delta\phi(\ell_p, E_{\text{T}}^{\text{miss}})$. Binning in $|\eta|$ was also investigated for both electrons and muons but not found to be important.

Figure 24 displays the FFs determined from the $Z+$jets-CR for electrons and muons separately. Electron FFs are between 0.1 and 0.4, with smaller FFs for small $\Delta\phi(\ell_p, E_{\text{T}}^{\text{miss}})$ and larger for larger $\Delta\phi(\ell_p, E_{\text{T}}^{\text{miss}})$. Muon FFs are between 0.3 and about 0.6, with a large statistical uncertainty for the highest p_{T} and FF value.

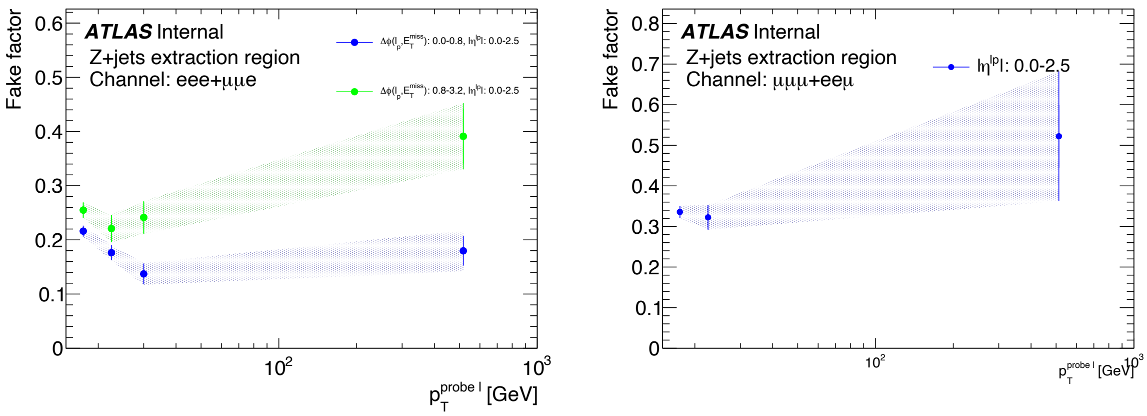


Figure 24: FFs for electrons (left) and muons (right) determined in the $Z+$jets-CR. The error bars indicate the statistical uncertainties from data and MC-based backgrounds. The dotted band summarizes the total uncertainty, including the systematic uncertainty from the subtraction of the MC-based backgrounds.

The systematic uncertainties on the FFs result from the MC subtraction, dominated by theoretical uncertainties on ZZ and WH cross-sections (7.1% and 6%, respectively). Still, statistical uncertainties are larger. The measured FF values and their uncertainties are summarized in Table 8.

A closure test in the $Z+$jets CR is done, where the FFs are applied to the anti-Id leptons in this selection. Figure 25 shows that the prediction agrees with the data within uncertainties in most bins. The remaining small mismodeling is in individual bins or very close to the systematic uncertainty and thus considered acceptable. The same procedure is applied to the anti-Id leptons in the baseline or same-sign regions in order to estimate the jets faking leptons contributions in each selection, respectively – with the exception that additional CFs are applied to account for differences in fake composition, as described in section 4.5.2. The results are shown in section 4.6.2 when presenting the validation of the overall fake background estimates.

Electron FFs		
	$\Delta\phi(\ell_p, E_{\mathrm{T}}^{\mathrm{miss}}) < 0.8$	dphilmet > 0.8
15<p_{T}<20GeV	0.2161 ± 0.0101 (stat) ± 0.0011 (syst)	0.255 ± 0.014 (stat) ± 0.0023 (syst)
20<p_{T}<25GeV	0.1763 ± 0.0141 (stat) ± 0.0027 (syst)	0.2209 ± 0.0248 (stat) ± 0.0055 (syst)
25<p_{T}<35GeV	0.1371 ± 0.0193 (stat) ± 0.0061 (syst)	0.2414 ± 0.0285 (stat) ± 0.0111 (syst)
p_{T}>35GeV	0.1797 ± 0.0272 (stat) ± 0.0259 (syst)	0.3912 ± 0.0504 (stat) ± 0.0346 (syst)
Muon FFs		
15<p_{T}<20GeV	0.3358 ± 0.0142 (stat) ± 0.0055 (syst)	
20<p_{T}<25GeV	0.3224 ± 0.0274 (stat) ± 0.0136 (syst)	
p_{T}>25GeV	0.5222 ± 0.0773 (stat) ± 0.14 (syst)	

Table 8: FFs and associated uncertainties, measured from the Z+jets CR.

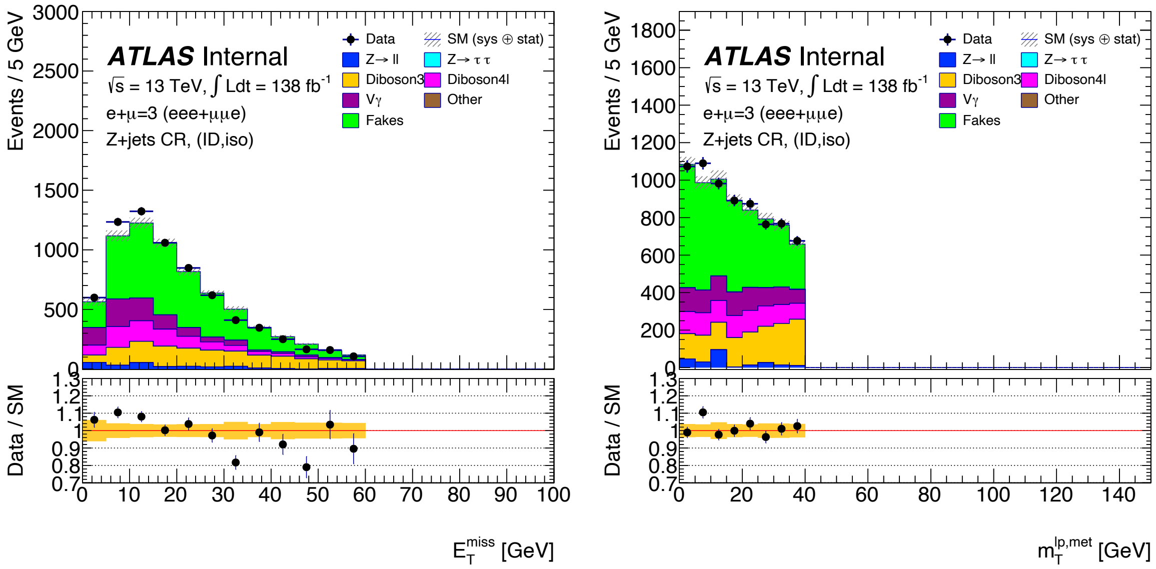


Figure 25: $E_{\mathrm{T}}^{\mathrm{miss}}$ (left) and m_{T} (ℓ_{probe}) (right) distributions including the electron fake estimate in the Z + jets CR, using the nominal FFs, but including systematic uncertainties from the fake background estimate.

4.5.2 Jet fakes flavor uncertainty

The FFs are measured in the Z+jets CR and applied to the baseline selection or the same-sign CR used for validation. But fake leptons can originate from a number of different sources whose relative abundance will vary between the different selections. If these fake sources have distinct chances of passing the Id- and anti-Id selections, it leads to a discrepancy in their ratio from one region to the other.

In order to quantify this discrepancy, a flavor composition study of the fake leptons on MC is performed. The classification of the leptons into different flavor categories is done using the IFFTruthClassifier tool. The MC samples considered are the Z+jets, W+jets, $Z \to \tau\tau$, $V + \gamma$, Top, Diboson, and Higgs. For the first three, we compare results from samples with two distinct generators, PowhegPythia and Sherpa.

The FF method gives an estimate of the jet faking leptons contributions which originate from the following distinct sources:

- light jets
- charm jets
- bottom jets
- *unknown* (electrons or muons with incomplete truth information).

Electron fake flavor composition

Table 9 summarizes per region the flavor fractions for Id and anti-Id electrons and the electron FFs per flavor component. The *unknown* fractions being always under 2%, we don't include this category. The Id and anti-Id electrons are mainly light jets, followed by charm jets in the baseline region, mainly light jets in the same-sign region. In the Z+jets region, bottom flavor is also significant. The flavor fractions reasonably agree between the two generators, but with more bottom in Sherpa and more charm in PowhegPythia.

The electron flavor FFs per region are similar, with the largest fluctuations for light flavor (smaller in same-sign, larger in base) and bottom flavor (larger in Z+jets). Values between both generators can vary, generally lower with Sherpa, but the trends per region are similar.

Region	LightFlavorDecay	CHadronDecay	BHadronDecay
	Id Electron Flavor Fractions		
Baseline	68.9 ± 1.3 (65.4 ± 2.8)	22.8 ± 1.2 (16.7 ± 2.0)	7.1 ± 0.5 (17.0 ± 1.4)
Same-Sign	74.6 ± 1.7 (66.5 ± 3.6)	3.5 ± 0.7 (4.6 ± 1.0)	18.2 ± 1.4 (27.8 ± 3.0)
Z+jets	56.5 ± 0.4 (57.7 ± 1.9)	8.0 ± 0.2 (6.3 ± 0.7)	34.3 ± 0.4 (34.8 ± 1.6)
	Anti-Id Electron Flavor Fractions		
Baseline	61.8 ± 0.6 (66.0 ± 0.9)	30.7 ± 0.5 (19.5 ± 0.7)	7.1 ± 0.2 (13.9 ± 0.4)
Same-Sign	79.5 ± 0.5 (71.0 ± 0.9)	3.5 ± 0.3 (4.7 ± 0.4)	15.6 ± 0.4 (23.5 ± 0.7)
Z+jets	61.6 ± 0.2 (61.0 ± 0.7)	10.5 ± 0.1 (8.0 ± 0.3)	27.5 ± 0.2 (30.5 ± 0.6)
	Electron Flavor FFs		
Baseline	0.188 ± 0.007 (0.112 ± 0.013)	0.125 ± 0.008 (0.096 ± 0.011)	0.168 ± 0.014 (0.138 ± 0.006)
Same-Sign	0.129 ± 0.007 (0.098 ± 0.015)	0.137 ± 0.031 (0.103 ± 0.022)	0.160 ± 0.013 (0.124 ± 0.007)
Z+jets	0.164 ± 0.002 (0.136 ± 0.010)	0.136 ± 0.004 (0.112 ± 0.014)	0.222 ± 0.003 (0.164 ± 0.007)

Table 9: Electron flavor fractions and flavor FFs per region, integrated in p_{T} and η. Results measured from PowhegPythia (Sherpa) samples are displayed in each cell. Only statistical uncertainties are considered.

Muon fake flavor composition

Table 10 summarizes per region the flavor fractions for Id and anti-Id muons, as well as the muon

FFs per flavor component. The Id and anti-Id muon fakes are mainly charm flavor in the baseline region and mainly charm and bottom flavor in the same-sign and Z+jets regions. For anti-Id muons, light is also significant in same-sign. The flavor fractions are similar between the two generators, with again more bottom in Sherpa.

The muon flavor FFs are similar between regions and generators for the charm and bottom flavors. While larger differences are found for light flavor and *unknown*, their fractions are relatively small except for same-sign anti-Id light.

Region	LightFlavorDecay	CHadronDecay	BHadronDecay	Unknown
		Id Muon Flavor Fractions		
Baseline	9.1 ± 0.8 (4.1 ± 2.2)	76.7 ± 1.1 (76.5 ± 2.4)	6.1 ± 0.5 (14.2 ± 0.9)	8.1 ± 0.7 (5.1 ± 1.6)
Same-Sign	10.6 ± 2.4 (8.7 ± 4.2)	28.8 ± 3.3 (31.5 ± 3.2)	52.0 ± 3.5 (50.6 ± 3.7)	8.7 ± 2.3 (9.2 ± 3.3)
Z+jets	8.2 ± 0.3 (3.6 ± 3.2)	35.0 ± 0.6 (34.7 ± 2.5)	50.3 ± 0.6 (57.4 ± 2.9)	6.4 ± 0.3 (4.3 ± 1.4)
		Anti-Id Muon Flavor Fractions		
Baseline	8.5 ± 0.5 (9.4 ± 1.4)	78.1 ± 0.7 (64.4 ± 1.4)	11.4 ± 0.4 (24.8 ± 0.8)	2.0 ± 0.2 (1.4 ± 0.8)
Same-Sign	23.5 ± 1.7 (14.3 ± 2.6)	18.2 ± 1.3 (18.8 ± 1.4)	51.9 ± 1.8 (62.1 ± 2.4)	6.4 ± 1.0 (4.8 ± 1.5)
Z+jets	7.9 ± 0.2 (8.6 ± 1.3)	27.6 ± 0.3 (25.0 ± 1.1)	57.3 ± 0.4 (60.1 ± 1.4)	7.3 ± 0.2 (6.3 ± 1.1)
		Muon Flavor FFs		
Baseline	0.437 ± 0.049 (0.157 ± 0.089)	0.401 ± 0.016 (0.424 ± 0.026)	0.218 ± 0.019 (0.205 ± 0.011)	1.656 ± 0.240 (1.321 ± 0.892)
Same-Sign	0.103 ± 0.027 (0.158 ± 0.089)	0.360 ± 0.059 (0.434 ± 0.060)	0.228 ± 0.021 (0.211 ± 0.012)	0.311 ± 0.104 (0.502 ± 0.253)
Z+jets	0.361 ± 0.016 (0.108 ± 0.100)	0.439 ± 0.011 (0.354 ± 0.035)	0.303 ± 0.006 (0.245 ± 0.013)	0.305 ± 0.015 (0.174 ± 0.066)

Table 10: Muon flavor fractions and flavor FFs per region, integrated in p_{T} and η. Results measured from PowhegPythia (Sherpa) samples are displayed in each cell. Only statistical uncertainties are considered.

Correction factors

This section introduces CFs, which account for the difference in flavor composition between the Z+jets CR on one side and the baseline or same-sign selections on the other when applying the data-driven FFs. These CFs, evaluated in MC, are defined as the ratio between the FFs in the baseline or same-sign regions, divided by the FFs in the Z+jets region, $FF^{\mathrm{MC}}_{base/SS}/FF^{\mathrm{MC}}_{Z\text{+jets}}$. In data, the application of the FFs is then adjusted as:

$$FF_{base/SS} = FF^{\mathrm{data}}_{Z\text{+jets}} \cdot \frac{FF^{\mathrm{MC}}_{base/SS}}{FF^{\mathrm{MC}}_{Z\text{+jets}}} \tag{57}$$

where $FF^{\mathrm{data}}_{Z\text{+jets}}$ are the FFs obtained from data described in (56).

Although we don't expect the FFs evaluated in MC to be fully reliable, we still expect, due to semi-cancellations, that this ratio can help translate the data-driven FFs measured from the Z+jets region to FFs valid within the region where they are applied. By comparing results obtained with different MC generators, we derive an uncertainty on the data-driven fake background due to the fake flavor composition. This systematic uncertainty reflects the difference in the fake composition between the $Z +$jets CR where the FFs are determined and the baseline or same-sign selections where they are applied. Since the source of the fakes differs for electron and muon fakes, this systematic uncertainty is considered uncorrelated between electron and muon fakes. The determination of the CFs and the associated flavor composition uncertainty is done separately in the baseline and same-sign selections.

In Figure 26, we compare the FFs in the Z+jets region measured from data to those measured in MC with Sherpa and PowhegPythia samples. The data FFs are generally higher than those measured in MC. The MC FFs measured with PowhegPythia samples agree better with the data than those measured with Sherpa samples.

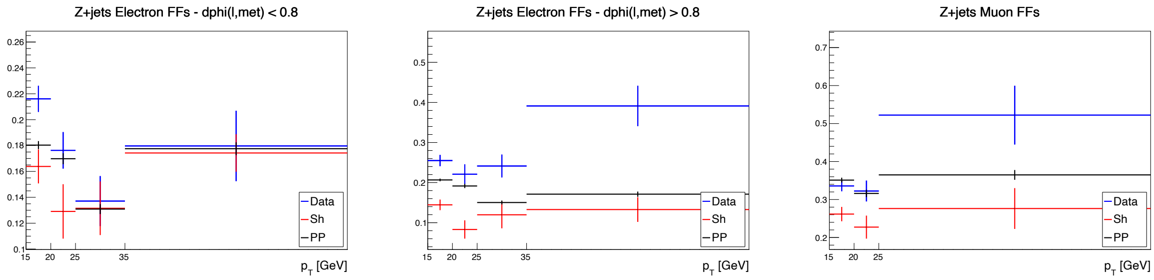


Figure 26: Comparison of electron and muon Z+jets FFs, measured from data, from MC using Sherpa samples, and from MC using PowhegPythia samples.

Figure 27 shows the FFs and CFs in the different regions, measured with the different MC generators. Results with PowhegPythia have smaller statistical uncertainties and agree better with data in the Z+jets region. Therefore, we use the CFs measured with PowhegPythia as our nominal values and derive systematic uncertainties by comparing them to those measured with Sherpa.

Table 11 summarizes the CFs and their uncertainties that we apply to the data-driven FFs, in order to account for the differences in flavor composition of the fakes between the different regions.

	Baseline Region	Same-Sign Region
	Electron CFs	
15<pT<20GeV	0.788 ± 0.037(stat) ± 0.050(syst)	0.799 ± 0.054(stat) ± 0.168(syst)
20<pT<25GeV	0.879 ± 0.064(stat) ± 0.080(syst)	0.704 ± 0.076(stat) ± 0.214(syst)
25<pT<35GeV	1.153 ± 0.089(stat) ± 0.449(syst)	0.724 ± 0.092(stat) ± 0.154(syst)
pT>35GeV	1.301 ± 0.086(stat) ± 0.472(syst)	0.823 ± 0.086(stat) ± 0.134(syst)
	Muon CFs	
15<pT<20GeV	0.962 ± 0.048(stat) ± 0.160(syst)	0.621 ± 0.067(stat) ± 0.294(syst)
20<pT<25GeV	1.250 ± 0.096(stat) ± 0.389(syst)	0.662 ± 0.116(stat) ± 0.319(syst)
pT>25GeV	1.888 ± 0.147(stat) ± 0.089(syst)	0.793 ± 0.116(stat) ± 0.644(syst)

Table 11: CFs and corresponding uncertainties per kinematic bin. The nominal values and statistical errors are derived with PowhegPythia samples, while the systematic uncertainty is evaluated by comparing with Sherpa samples.

Table 12 summarizes the data-driven fake yields and their uncertainties from different sources in the baseline and same-sign selections.

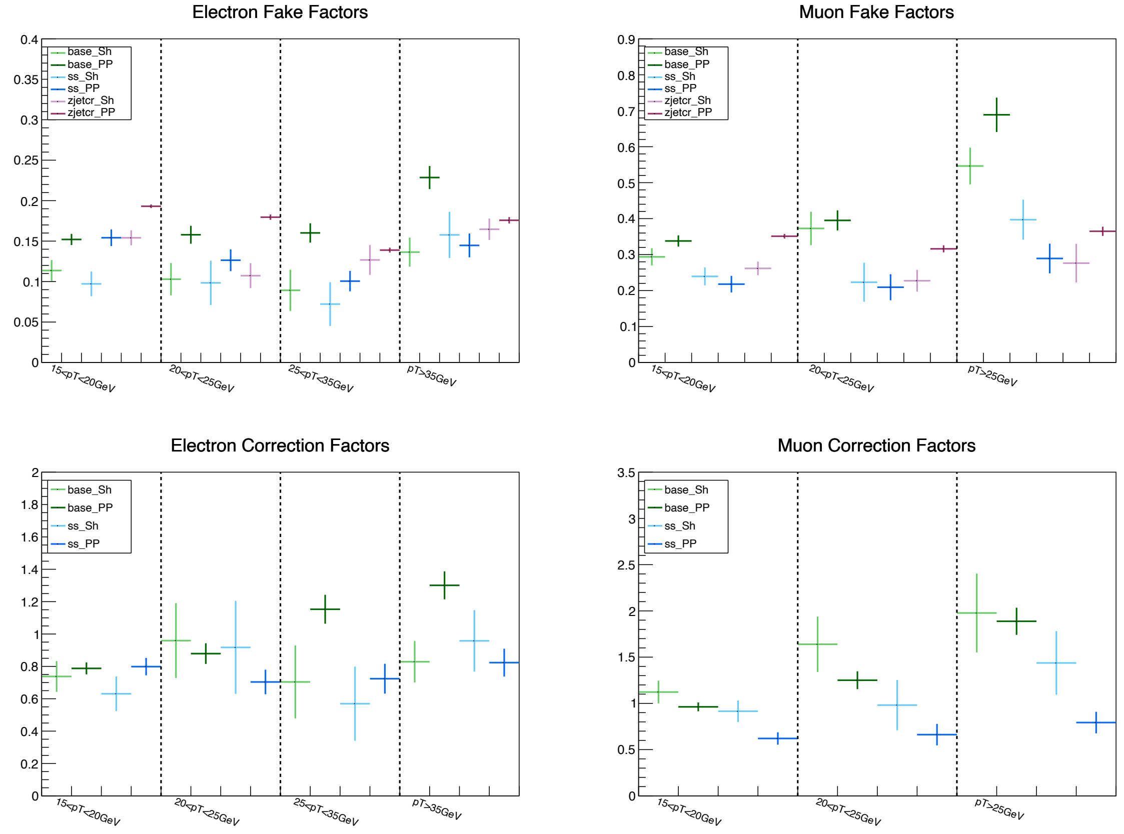


Figure 27: Comparison of electron and muon FFs and CFs in the different regions (base, same-sign and Z+jets), and measured with the different MC generators (Sherpa and PowhegPythia).

selection	nominal	stat	ffelstat	ffmustat	ffwzsys	ffzzsys	cfelstat	cfmustat	cfelsys	cfmusys	sys-combined
base-$e\tau$	18085.1	0.89	0.52	4.58	7.99	3.04	0.32	3.59	1.47	9.21	14.14
base-$\mu\tau$	16761.5	0.72	6.24	1.03	5.67	2.08	4.31	0.63	12.70	0.99	16.12
ss-$e\tau$	5896.5	0.80	4.19	3.65	0.91	0.37	3.68	6.25	6.74	26.83	29.16
ss-$\mu\tau$	6813.8	0.63	7.08	0.94	0.61	0.25	7.53	1.16	18.80	5.62	22.28

Table 12: FF fake yields (nominal) and relative uncertainties in [%] induced from statistical uncertainties (stat) or the different systematic uncertainty sources determined: the FF statistical errors (ffelstat and ffmustat), the WZ and ZZ cross-section errors (ffwzsys and ffzzsys), the flavor composition CFs statistical error (cfelstat and cfmustat) and systematic uncertainty (cfelsys and cfmusys) and the combined systematic uncertainty (sys-combined). These are shown separately in the $e\tau$ and $\mu\tau$ datasets and the baseline and same-sign selections.

4.5.3 Other fakes estimate

The background from hadronically decaying τ-leptons misidentified as leptons ($\tau_{\text{had}} \to \ell$), muons misidentified as electrons ($\mu \to e$), and photons misidentified as electrons ($\gamma \to e$) is estimated from MC simulations. It corresponds to the remaining part of the fake backgrounds, not covered by the data-driven FF method, discussed in section 4.5.1.

This MC-based fake estimate is done separately in the baseline and same-sign selections, based on the following simulated processes: $Z \to \tau\tau$, $t\bar{t}$, single-top, diboson, $Z \to \mu\mu$, $Z \to ee$, and $V\gamma$. Contributions from $H \to \tau\tau$ and $H \to WW$ are neglected. Fake leptons in MC simulation are selected employing a truth matching based on the IFFTruthClassifier. If at least one lepton is classified in one of the MC-fakes categories, then the event is included in the MC-fakes estimate.

The MC-fake yields in the baseline selection are summarized in Table 13. The largest contribution, when considering both $e\tau$ and $\mu\tau$ final states, comes from $Z \to \mu\mu$, for which one muon – typically the one with the lower transverse momentum – is misidentified as an electron. This process mainly contributes to the $\mu\tau$-dataset where the electron is sub-leading in most events, but not in all due to the chosen lepton-assignment strategy described in section 4.3.2. Only the $\mu\tau$-dataset is used to study the $Z \to \mu\mu$-background in the following.

fake source	$e\tau$	$\mu\tau$
$\mu \to e$	1401.9 ± 25	6855.7 ± 43.7
$\tau_{\text{had}} \to \ell$	3658.4 ± 54.2	2057.5 ± 43.3
$\gamma \to e$	811.6 ± 33.1	2239.5 ± 59.3

Table 13: MC-fake yields in the baseline selection, shown separately per fake source and for the $e\tau$- and the $\mu\tau$-datasets.

In order to check if the $\mu \to e$-fake rate is well modeled by the MC predictions, a dedicated validation region (VR) is defined. High purity of $Z \to \mu\mu$-events is desired such that the expected yield can be compared to the corresponding data yield. To achieve this, in addition to the baseline selection requirements, the following criteria are applied:

- $35\,\text{GeV} < p_{\text{T}}(\ell 0) < 45\,\text{GeV}$
- $75\,\text{GeV} < m_{\ell\ell} < 100\,\text{GeV}$
- $1.25 <$ trk-p_{T}/cluster-$p_{\text{T}} < 3$ for electrons
- $m_{\text{coll}} < 115\,\text{GeV}$

The $Z \to \mu\mu$ contribution in this VR amounts to around 70% of the events. An overshoot of the prediction compared to data is visible, shown in Figure 28. Therefore, a corresponding uncertainty of 16% on the $Z \to \mu\mu$-normalization in MC is defined and applied in full size – to be conservative – to the total MC-fakes contribution and not only to its $Z \to \mu\mu$-part.

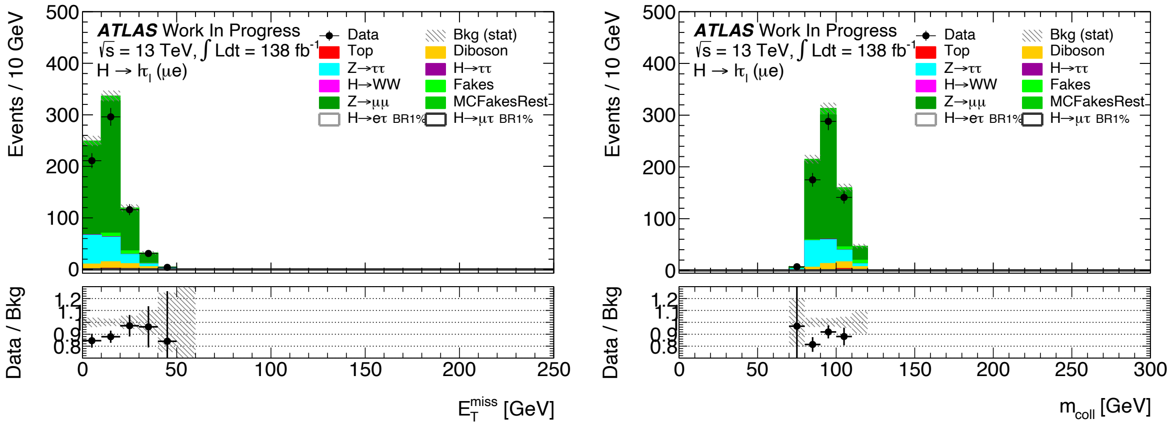


Figure 28: $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{coll} distributions in the $Z \to \mu\mu$-VR in the $\mu\tau$-dataset. 'MCfakesRest' in the legend is MCfakes without $Z \to \mu\mu$ which is shown separately, while 'Fakes' are the data-driven jet fakes. Only statistical uncertainties are displayed.

4.6 Validation of the background estimates

In this section, we present the different steps to validate the estimates of the various background contributions, as well as their combination. Validation is performed by inspecting various kinematic distributions of the selected events and comparing the agreement between different datasets depending on the validation step considered.

The efficiency correction is validated based on the SM MC samples by inspecting the resulting $e\tau$ vs. $\mu\tau$ symmetry, shown in section 4.6.1. The fake estimate is validated by comparing MC background + fake predictions to the data, separately for the $e\tau$ and $\mu\tau$ datasets and in the same-sign and baseline selections, shown in section 4.6.2. The combined background prediction is validated by inspecting the achieved $e\tau$ vs. $\mu\tau$ symmetry in the data, as is shown in section 4.6.3.

4.6.1 Efficiency correction

We start by validating the efficiency correction method described in section 4.3.4 (together with the measured electron MC efficiencies presented in section 4.4) based on the SM MC simulated samples, with events passing the baseline selection. Due to the different efficiency dependencies, the efficiency-ratio CF is relative to whether the event belongs to the $e\mu$ or μe dataset, which differ by the p_{T} ordering of the two leptons in the *laboratory* frame (as described in section 4.3.2). Therefore we start by comparing the corrected $e\mu$ and μe datasets. We first apply standard (MC to data) SFs to both datasets in order to scale them to what is expected in data. The $e\mu$ dataset is then corrected using *data* efficiencies. As per (51), the measured electron MC efficiencies are converted to electron data efficiencies by multiplying them with the relevant standard (MC to data) electron SFs.

We inspect the level of the agreement after efficiency correction for various kinematic distributions, as shown in Figure 29. Any particular trend in the ratio of corrected $e\mu$ over μe events as a function of a distribution could hint at a residual asymmetry which should be investigated. Applying the efficiency correction successfully recovers the expected symmetry between the $e\mu$ and μe datasets, which agree within statistical uncertainties in most bins.

In terms of the $e\tau_\mu$ and $\mu\tau_e$ datasets (which differ by the p_T ordering of the two leptons in the *estimated Higgs* frame as described in section 4.3.2), the procedure is almost the same. The efficiency correction is still applied only to the dataset used as the background estimate. However, the efficiency-CF is still relative to whether the event belongs to the $e\mu$ or μe dataset. For example, if the $e\tau_\mu$ is the basis for the background estimate, the efficiency-CF for the few μe events which contribute to it will be $R_i^\epsilon(\mu e \text{ event})$.

We proceed with the same comparison, but with the corrected $e\tau$ vs. $\mu\tau$ events instead of corrected $e\mu$ vsμe. This is shown separately for the non-VBF category in Figure 30 and the VBF category in Figure 31. Here as well, applying the efficiency correction restores the expected SM symmetry in simulated events.

The same validation but with tighter constraints applied to the electrons (as described in section 4.4.7) is presented in appendix A.4. Although the electron efficiencies are significantly impacted by these tighter constraints, the expected SM symmetry is successfully restored in this case as well.

4.6.2 Fake background

The overall fake background estimate is validated in the baseline and same-sign regions by comparing MC background + fake prediction to the data. In the baseline region, we assume that the selection is wide enough such that the expected signal wouldn't impact the MC to data agreement if present. In the same-sign region, a fake purity – calculated based on the predictions – of 64.9% in $e\tau$ and 75.9% in $\mu\tau$ – is reached.

Similar to the $e\mu$-symmetry validation in MC shown in section 4.6.1, this validation step is intermediate and isn't decisive in our analysis since we estimate the SM contribution to our SR from the data directly. It is important, nonetheless, since when validating the $e\mu$-symmetry in data, we can't decouple features originating from the efficiency correction and the fake estimate. Still, good MC vs. data agreement isn't critical for us as it is in more conventional analyses where the background is estimated from MC and where additional studies are performed to achieve this, such as normalizing specific SM contributions based on MC vs. data comparison in dedicated CRs.

The overall fake background estimate is composed of the FF-fake and MC-fake contributions. The FF-fake contribution is obtained by multiplying the anti-id events in the data minus the prompt anti-id MC events by the relevant FFs and CFs, as described in section 4.5.

Different kinematic distributions of the $e\tau$ and $\mu\tau$ datasets are shown in the baseline (Figures 32 and 33) and same-sign (Figures 34 and 35) selections. The average level of disagreement is around 3-4% in the baseline selection, and we expect it to be covered by MC systematics (only systematics on the FF-fake estimate are shown). In the same-sign selection, the modeling is good, and the prediction agrees with data within the statistical uncertainties and the uncertainties on the FF-fake estimate depicted by the uncertainty band.

4.6.3 Combined background prediction

The combined background is the sum of the SM symmetric background defined in (52) and the fake estimates. It is described in (53).

The comparison of the combined background predictions to the data in the $\mu\tau$ dataset is shown separately for the non-VBF category in Figure 36 and the VBF category in Figure 37. In the plots, only statistical uncertainties and systematic uncertainties on the FF-fakes (the dominant ones) are included. Both signals are shown, normalized to 1% BR, and scaled by 10. The $H \to e\tau_\mu$ signal corresponds to the "contamination" of wrongly classified signal events, as described in section 4.3.2, and its contribution is very small. Only plots in the $\mu\tau$ dataset are shown since, due to the symmetry, the background predictions to data agreement is similar – albeit inverted – in the $e\tau$ dataset, and the fake estimates from both datasets are included in the background predictions. In general, the modeling of the data by the predictions is good – within systematic uncertainties. Disagreements in isolated bins are found, but no general trends are observed.

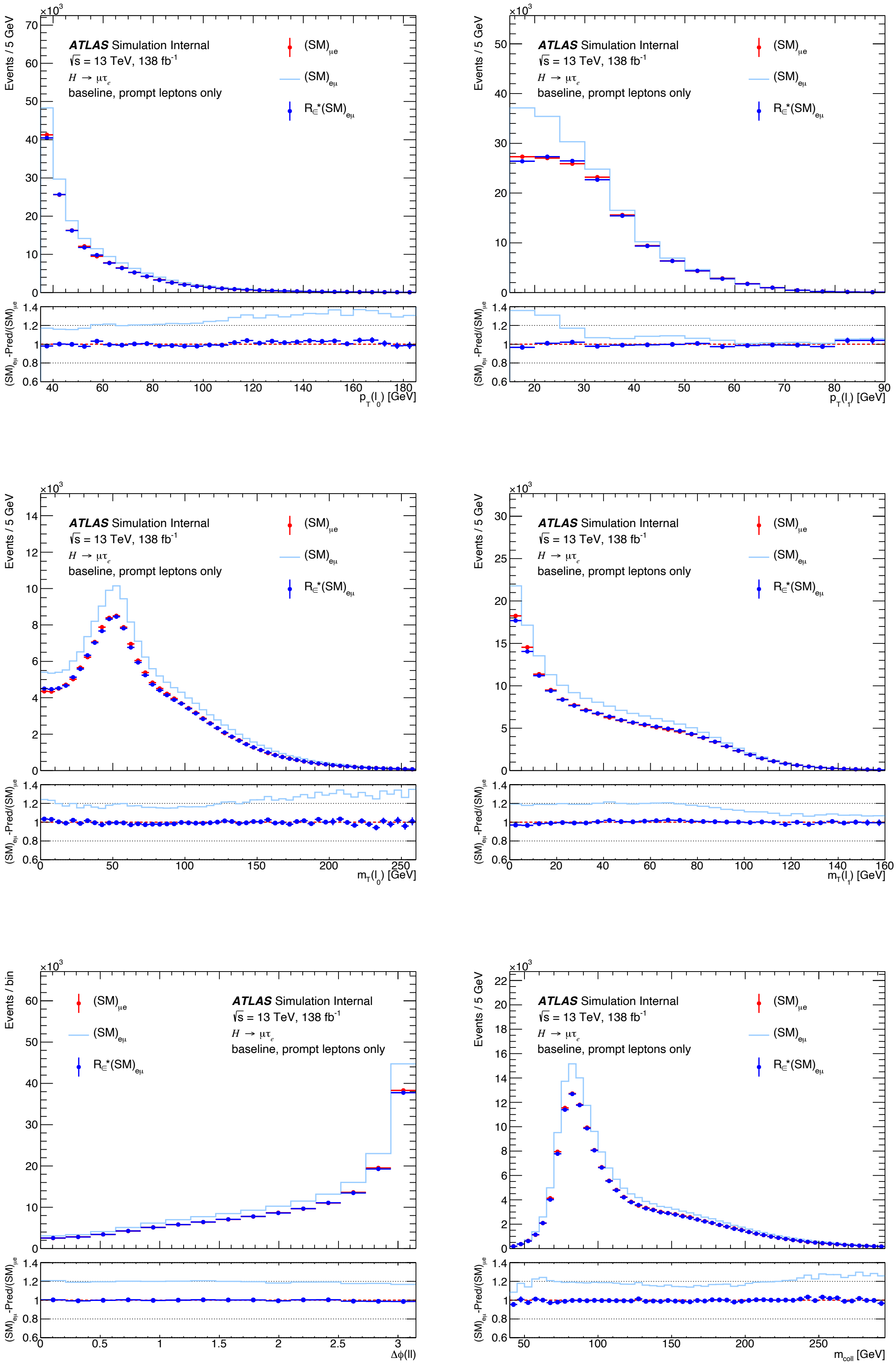


Figure 29: Comparison of efficiency-corrected $e\mu$ (dark blue) over μe (red) SM MC sample datasets in the baseline selection for the $p_{\mathrm{T}}(\ell_0)$, $p_{\mathrm{T}}(\ell_1)$, $m_{\mathrm{T}}(\ell_0, E_T^{\mathrm{miss}})$, $m_{\mathrm{T}}(\ell_1, E_T^{\mathrm{miss}})$, $\Delta\phi(\ell\ell)$ and m_{coll} distributions. The uncorrected $e\mu$ dataset is also displayed (light blue).

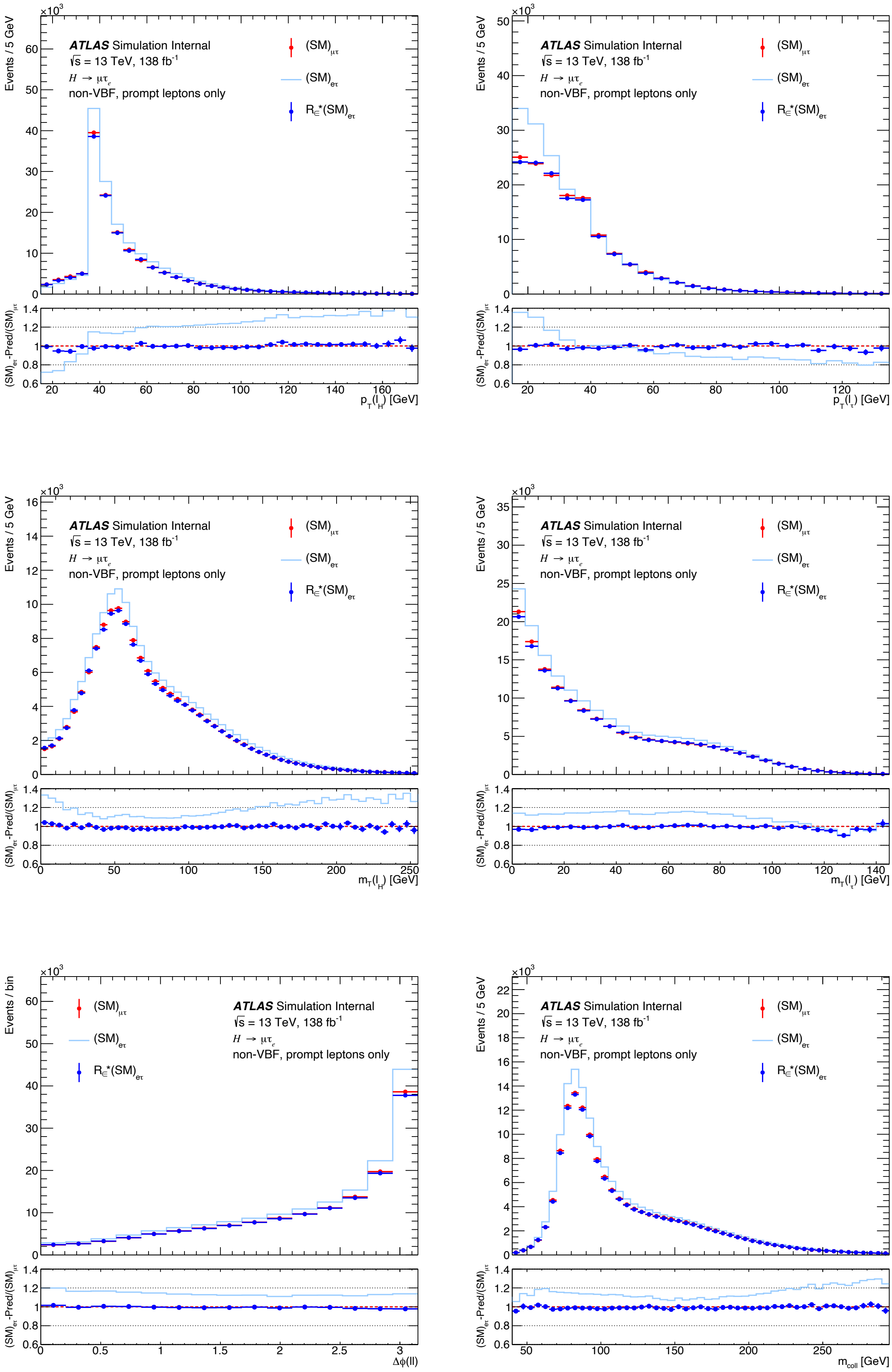


Figure 30: Comparison of efficiency-corrected $e\tau$ (dark blue) over $\mu\tau$ (red) SM MC sample datasets in the non-VBF selection for the $p_{\mathrm{T}}(\ell_H)$, $p_{\mathrm{T}}(\ell_\tau)$, $m_{\mathrm{T}}(\ell_H, E_T^{\mathrm{miss}})$, $m_{\mathrm{T}}(\ell_\tau, E_T^{\mathrm{miss}})$, $\Delta\phi(\ell\ell)$ and m_{coll} distributions. The uncorrected $e\tau$ dataset is also displayed (light blue).

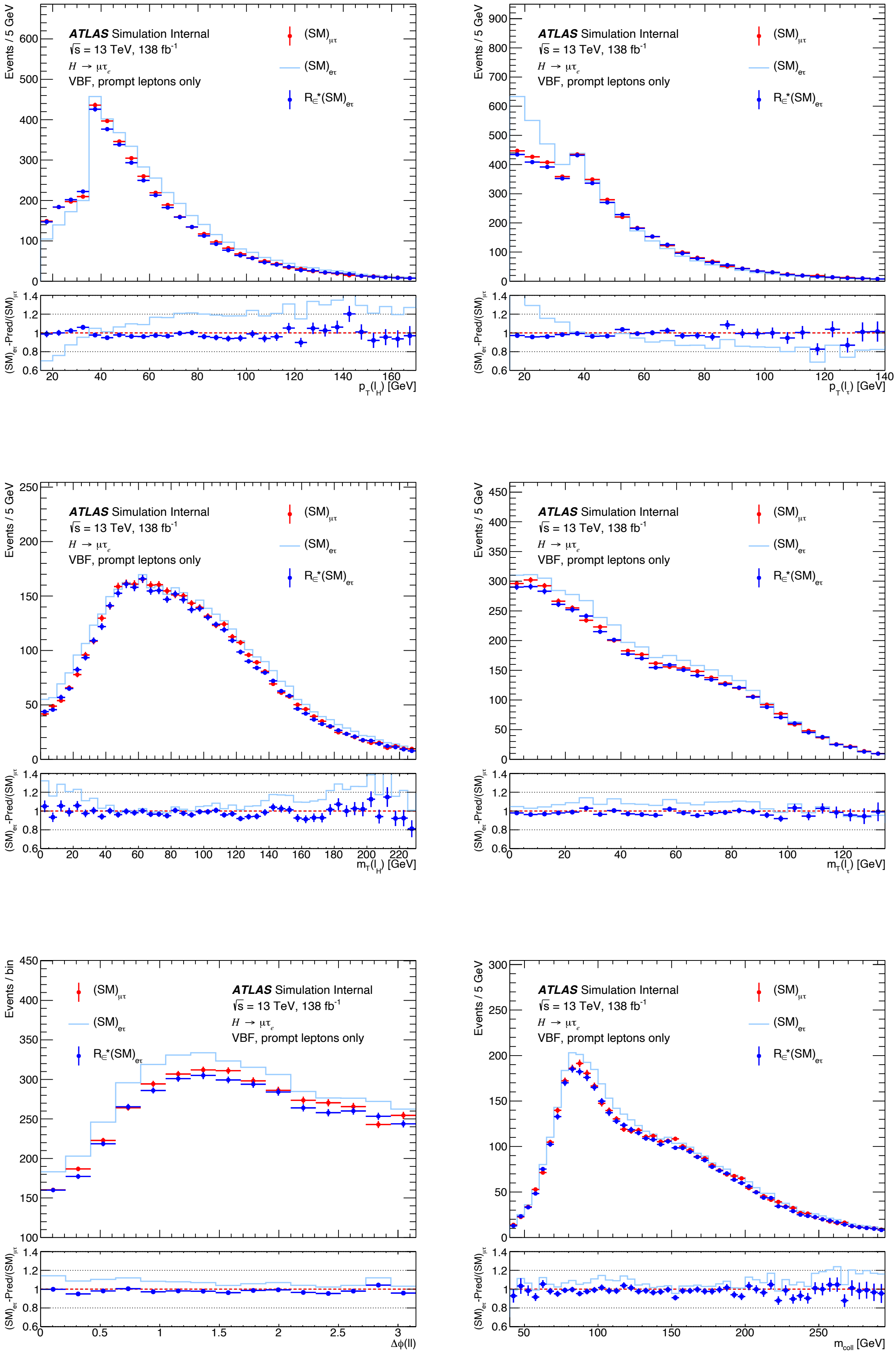


Figure 31: Comparison of efficiency-corrected $e\tau$ (dark blue) over $\mu\tau$ (red) SM MC sample datasets in the VBF selection for the $p_{\mathrm{T}}(\ell_H)$, $p_{\mathrm{T}}(\ell_\tau)$, $m_{\mathrm{T}}(\ell_H, E_T^{\mathrm{miss}})$, $m_{\mathrm{T}}(\ell_\tau, E_T^{\mathrm{miss}})$, $\Delta\phi(\ell\ell)$ and m_{coll} distributions. The uncorrected $e\tau$ dataset is also displayed (light blue).

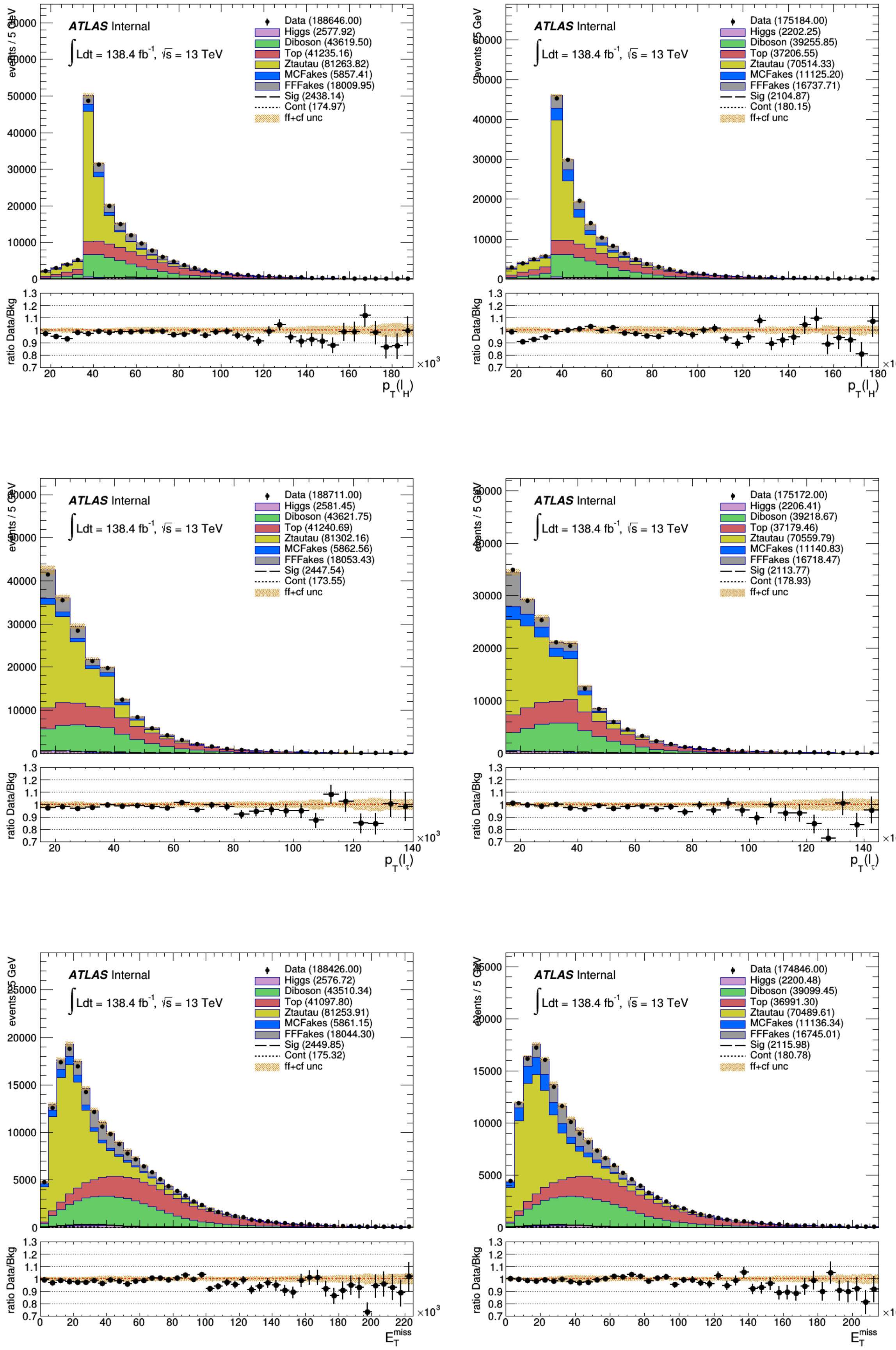


Figure 32: MC vs. data agreement in the baseline selection for different kinematic distributions in the $e\tau$ (left) and $\mu\tau$ (right) datasets. The FF-fakes are in gray and the MC-fakes in blue. Only the systematic uncertainty associated with the FF estimate is shown.

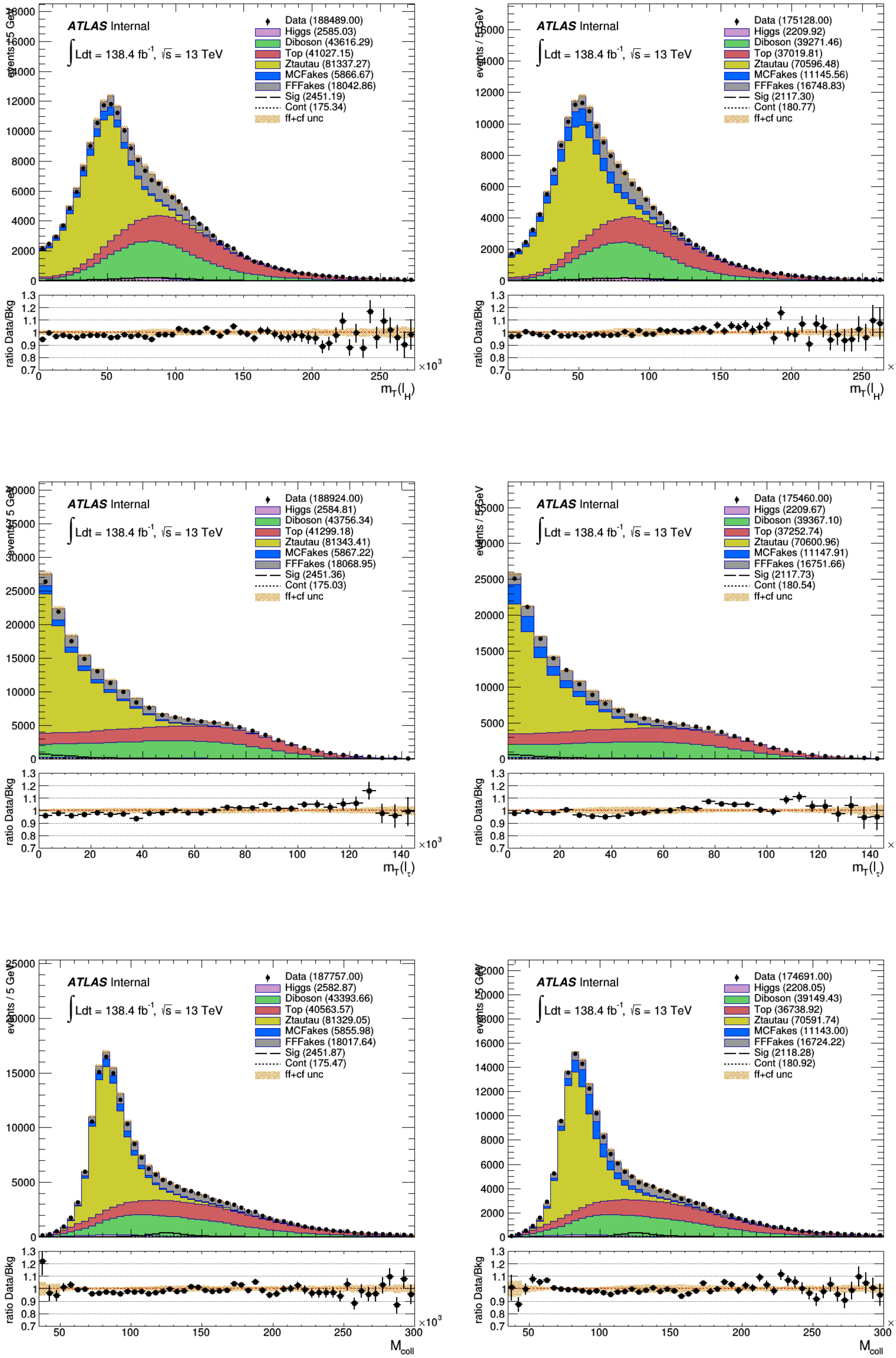


Figure 33: MC vs. data agreement in the baseline selection for more kinematic distributions in the $e\tau$ (left) and $\mu\tau$ (right) datasets. The FF-fakes are in gray and the MC-fakes in blue. Only the systematic uncertainty associated with the FF estimate is shown.

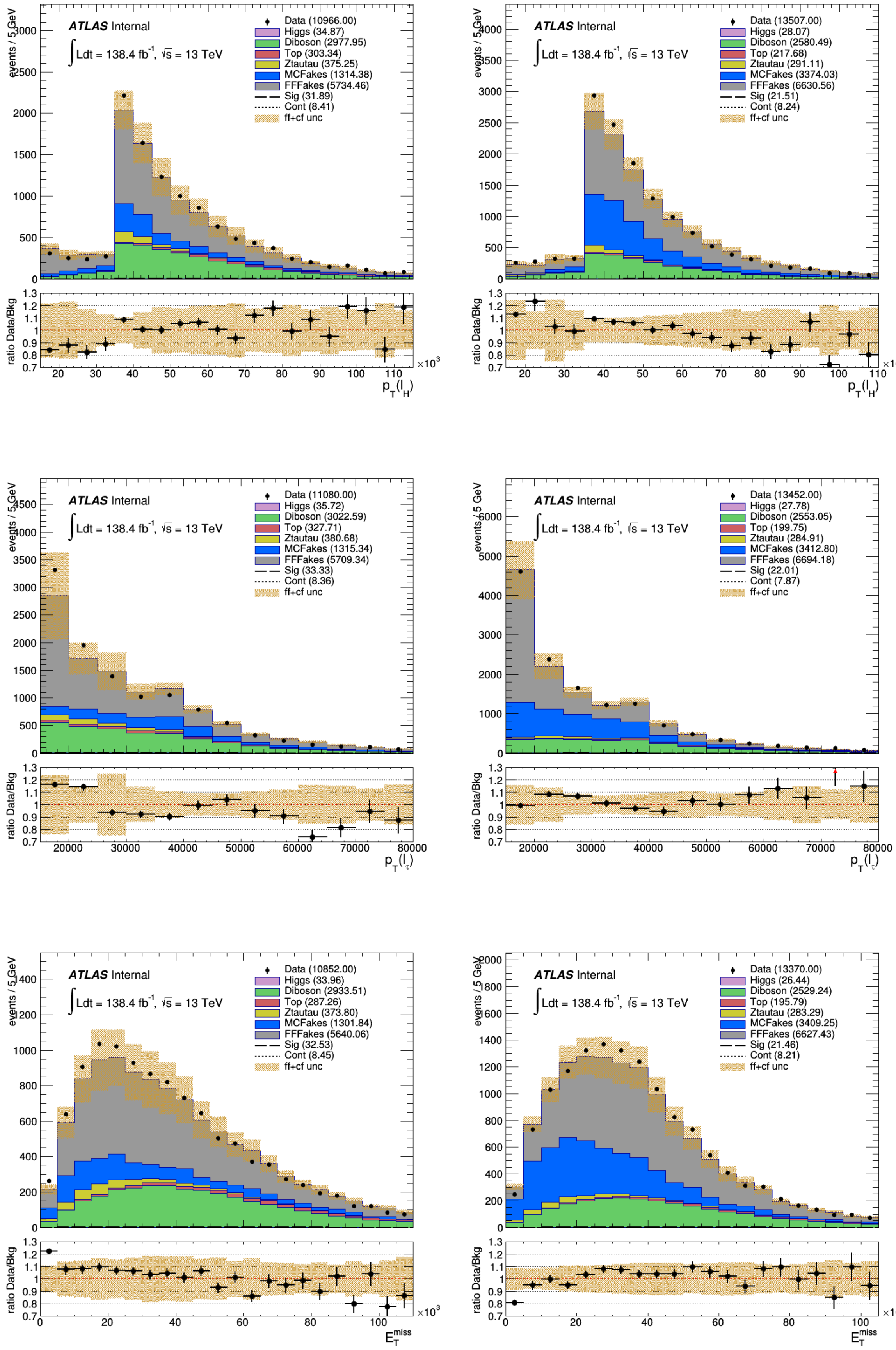


Figure 34: MC vs. data agreement in the same-sign selection for different kinematic distributions in the $e\tau$ (left) and $\mu\tau$ (right) datasets. The FF-fakes are in gray and the MC-fakes in blue. Only the systematic uncertainty associated with the FF estimate is shown.

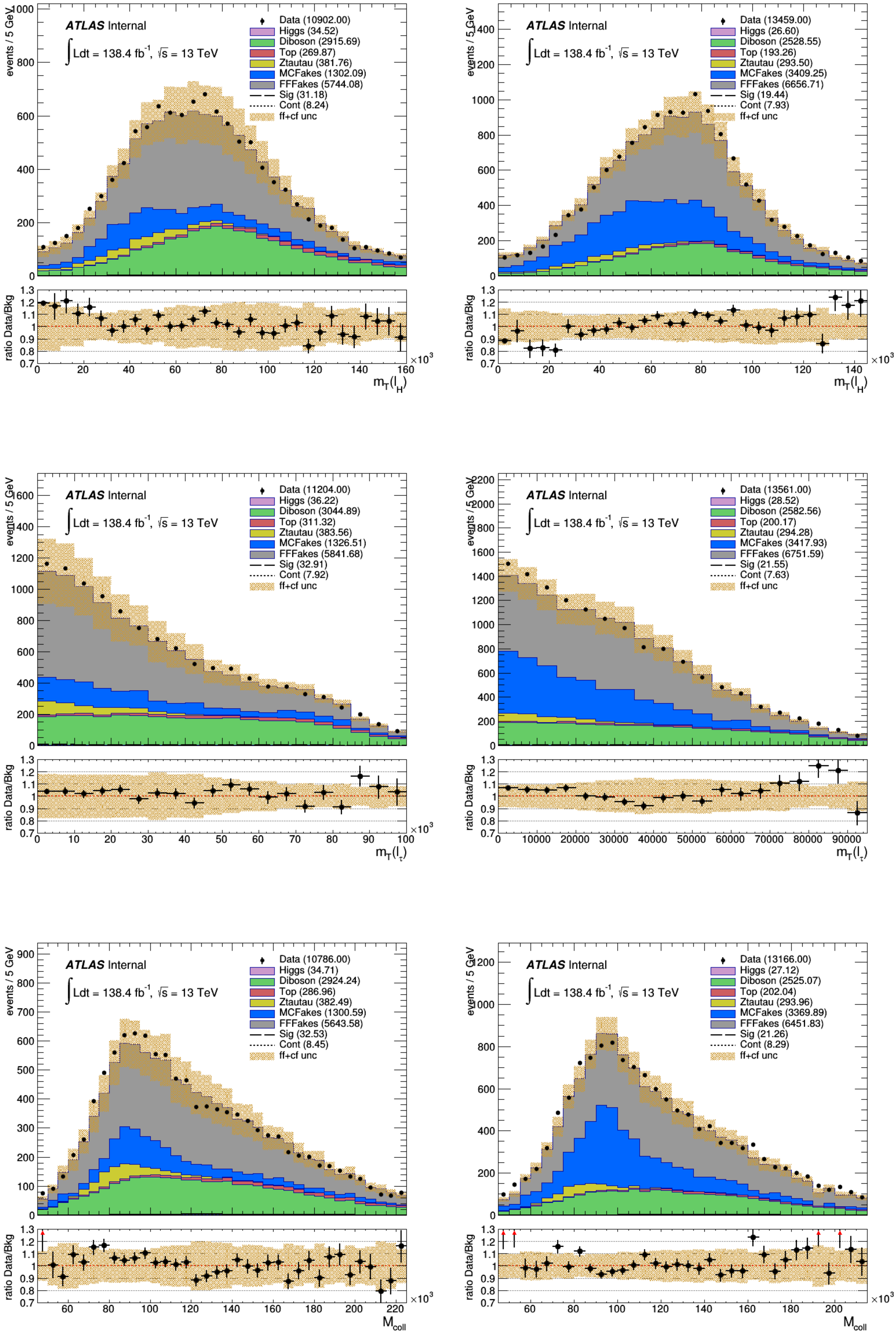


Figure 35: MC vs. data agreement in the same-sign selection for more kinematic distributions in the $e\tau$ (left) and $\mu\tau$ (right) datasets. The FF-fakes are in gray and the MC-fakes in blue. Only the systematic uncertainty associated with the FF estimate is shown.

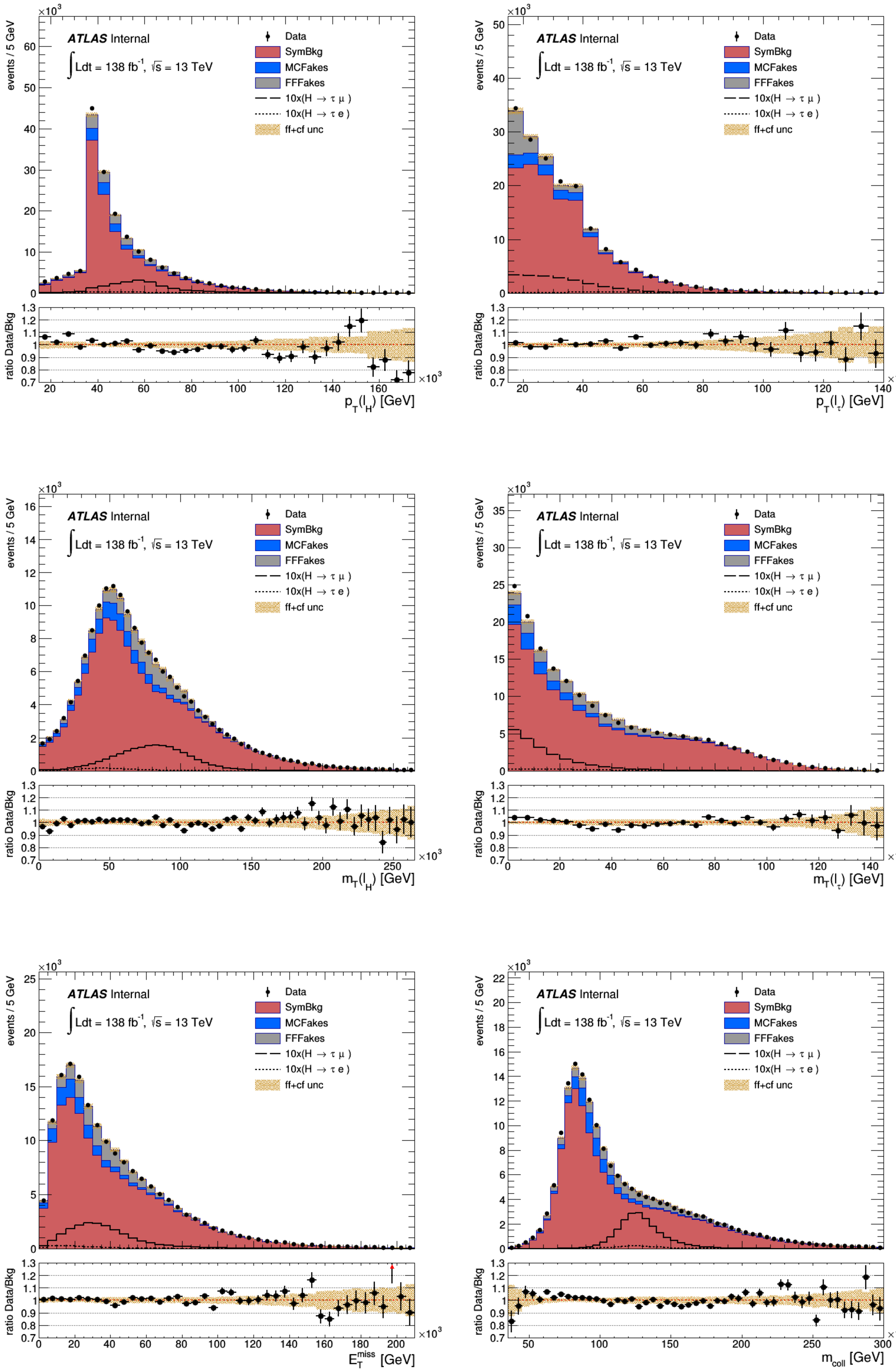


Figure 36: Comparison of the background predictions to the data in the $\mu\tau$ datasets in the non-VBF selection for the $p_{\mathrm{T}}(\ell_H)$, $p_{\mathrm{T}}(\ell_\tau)$, $m_{\mathrm{T}}(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, $m_{\mathrm{T}}(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, $E_{\mathrm{T}}^{\mathrm{miss}}$, and m_{coll} distributions. Statistical and FF-fakes systematic uncertainties on the background prediction are included in the uncertainty band.

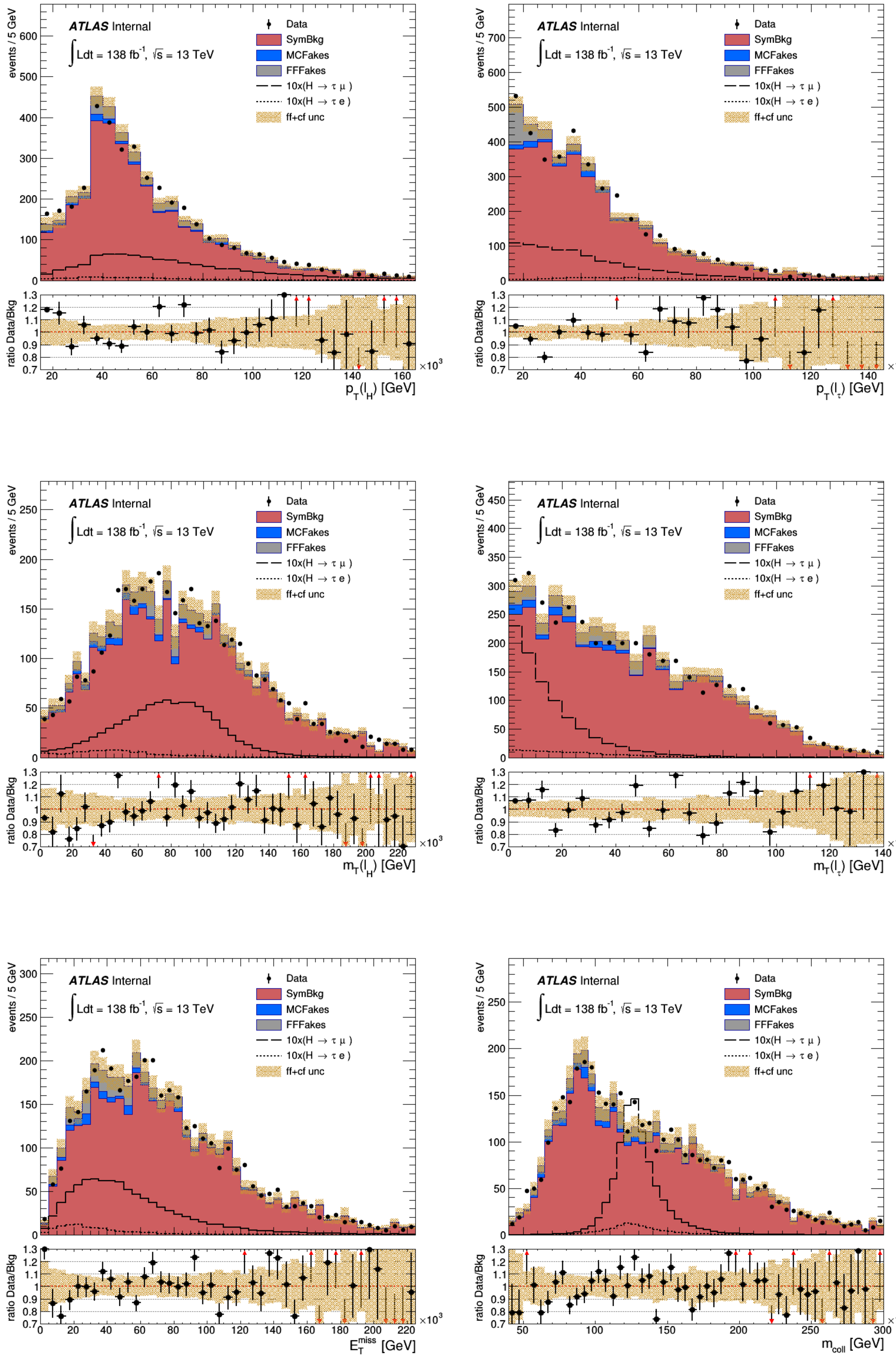


Figure 37: Comparison of the background predictions to the data in the $\mu\tau$ datasets in the VBF selection for the $p_{\mathrm{T}}(\ell_H)$, $p_{\mathrm{T}}(\ell_\tau)$, $m_{\mathrm{T}}(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, $m_{\mathrm{T}}(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, $E_{\mathrm{T}}^{\mathrm{miss}}$, and m_{coll} distributions. Statistical and FF-fakes systematic uncertainties on the background prediction are included in the uncertainty band.

4.7 Enhancing Sensitivity via Neural Network

The strategy to achieve high sensitivity in this analysis is, on the one hand, a relatively loose selection (baseline region) to keep high statistics and, on the other hand, constructing an observable that separates signal and background as best as possible. The distribution of this observable is then used as the final discriminant in the statistical analysis. To construct such an observable, fully connected deep NNs are trained.

The NNs are trained to separate the signal from the various background contributions. Different NNs are used for the non-VBF and the VBF categories to exploit the particular VBF topology, including additional kinematic jet-related information. The chosen architecture takes into account the limited training sample size. The $e\tau$ and the $\mu\tau$ datasets are combined in order to enhance the number of events in the training. In addition, the VBF cut on the jet invariant mass (m_{jj}) is lowered to 300 GeV to increase the size of the training set in this region. The NNs are trained via supervised learning using Keras (v2.2) [84], with Tensorflow (v1.12) [85] backend. The hyper-parameters and the input variables are optimized based on the Asimov discovery significance [54], carried out using the Optuna framework [86], which also allows using the *L2 weight regularization* to prevent overtraining.

For the non-VBF region, a single multiclass classifier with three output classes is trained. The three classes correspond to signal, SM symmetric background, and fake background. The MC-fakes contribution is added to the SM symmetric background class since the distributions of its main processes, $Z \to \mu\mu$, $Z \to \tau\tau$, and $V\gamma$, are more similar to the distributions of the SM symmetric background processes than to the ones originating from jets faking leptons. For the VBF region, three single binary classification networks are trained. The resulting output distributions are linearly combined in a single one – where the weights are also optimized using the Asimov discovery significance – to obtain one discriminant. The three networks are trained to separate signal events from (i) MC-fakes, $Z \to \tau\tau$ and $H \to \tau\tau$, (ii) $t\bar{t}$, single-top, diboson, and $H \to WW$ and (iii) FF-fakes.

The NNs are trained with several input variables, including the four-momenta of the analyses objects and other derived observables, like the invariant masses, angular separations, and the $E_{\mathrm{T}}^{\mathrm{miss}}$. In the VBF region, jet-related variables are used in addition. The list of variables is optimized for each category by removing the lowest-ranked variables with a marginal contribution to the sensitivity. The visible Higgs mass (m_{vis}), the collinear mass (m_{coll}), and the Higgs mass obtained with the Missing Mass Calculator technique (m_{MMC}) [87] show the highest separation power together with transverse masses obtained from the $E_{\mathrm{T}}^{\mathrm{miss}}$ and each of the leptons ($m_{\mathrm{T}}(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}}), m_{\mathrm{T}}(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$). Under the collinear approximation, the azimuthal angular difference between the prompt lepton decaying from the Higgs (from the τ) and the $E_{\mathrm{T}}^{\mathrm{miss}}$– $\Delta\Phi(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})(\Delta\Phi(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}}))$ – is expected to be large (small) for signal events. Other angular differences are also included: $\Delta R(\ell_H, \ell_\tau), \Delta\eta(\ell_H, \ell_\tau), \Delta\Phi(\ell_H, \ell_\tau)$. The angular differences $\Delta\Phi$ involving ℓ_H are evaluated in the approximate Higgs boson rest frame. Two vertex variables are also included as they provide powerful discrimination power: the difference in the transverse impact parameter (d0) between the leptons $\Delta d0(\ell_0, \ell_1)$ and the d0 significance of the τ lepton $\sigma_{d0}^{\ell_\tau}$. The $\Delta\alpha$ discriminant [88] is expected to be close to zero if the decay products of the τ are collinear and the transverse momentum of the Higgs boson can be neglected, while for the background events, this value deviates from zero. The η-centrality, defined as

$$\exp(-4/(\eta(j_0) - \eta(j_1))^2 \cdot (\eta(\ell) - 0.5(\eta(j_0) - \eta(j_1)))^2) \tag{58}$$

is included as well in the VBF category.

The binning of the NN output distributions is chosen pre-fit such that each individual background contribution has yields larger than zero and relative statistical uncertainties smaller than 100% in each bin and such that the overall background estimate has decreasing yields in bins of increasing NN score.

The resulting NN discriminant distributions are shown in Figure 38, as obtained from the fit described in section 4.9.2, separately for the $H \to e\tau_\mu$ and $H \to \mu\tau_e$ signal searches and in the non-VBF and VBF regions. Full uncertainties (statistical and systematic) corresponding to the pre-fit values are shown. The signal distributions are normalized to 0.1% BR and scaled by 100. Two distinct prediction vs. data comparisons are shown in the plots, the data over prediction (background + signal times signal strength) ratio and background-subtracted data.

4.8 Systematic uncertainties

Systematic uncertainties affect the yields of the main fit observable, the NN output distribution. These can be categorized into two groups: experimental uncertainties and theoretical uncertainties for the signal. Theoretical uncertainties for the background processes are not considered since the background estimate is mainly data-driven, except for the $Z \to \mu\mu$ normalization uncertainty determined in section 4.5.3 from a dedicated CR and variations of the WZ and ZZ cross-sections in the FF-fake estimate (subtraction of real MC contributions).

Experimental uncertainties include those originating from the trigger, Reco, Id, and Iso efficiencies of the final-state particles. These include leptons [89–91], jets [77, 92, 93] and $E_{\mathrm{T}}^{\mathrm{miss}}$ [80]. Their energy scale and resolution uncertainties are taken into account as well. Experimental uncertainties generally affect the shape of the NN output distributions, the background yields, and the signal cross-section through their effects on the acceptance and the migration between different categories. Uncertainties of the luminosity measurement [94] are also included.

The uncertainties associated with the FF-fake contribution consist of statistical and systematic components. The statistical component is estimated for each FF bin separately. It consists of the statistical uncertainties from the data, and the subtracted real MC yields within the Z+jets CR propagated to the FFs. The systematic component corresponds to the uncertainty from the subtraction of MC real events contributing to the $Z +$jets CR, where the subtraction is varied by the theory cross-section uncertainties of the two dominant contributions separately, the WZ and ZZ processes. Uncertainties associated with the CFs – which account for different relative (jet) fake source abundances between the baseline and $Z +$jets regions – include the statistical uncertainties of the MC events in both regions and the flavor composition uncertainty derived from the comparison of two MC generators – as described in section 4.5.2. Table 12 summarizes the pre-fit impact of each type of uncertainty on the FF-fakes estimate. In the statistical model (described in section 4.3.6), the statistical sources are treated as uncorrelated between FF or CF bins, the systematic sources as correlated between bins, and the MC subtraction uncertainties as correlated between bins and also between electron and muon FF-fakes contributions.

The uncertainties related to the MC-fakes contribution consist of the statistical and experimental (systematic) uncertainties associated with the available MC simulated events. An additional systematic uncertainty on the $Z \to \mu\mu$ normalization is included, as described in section 4.5.3.

The SM symmetric background estimate consists of data, FF-fakes, and MC-fakes. The latter two are subtracted from the data, and all three components are weighted with efficiency corrections event-by-event, as described in section 4.6.3. Uncertainties on the FF-fakes and MC-fakes described above are propagated

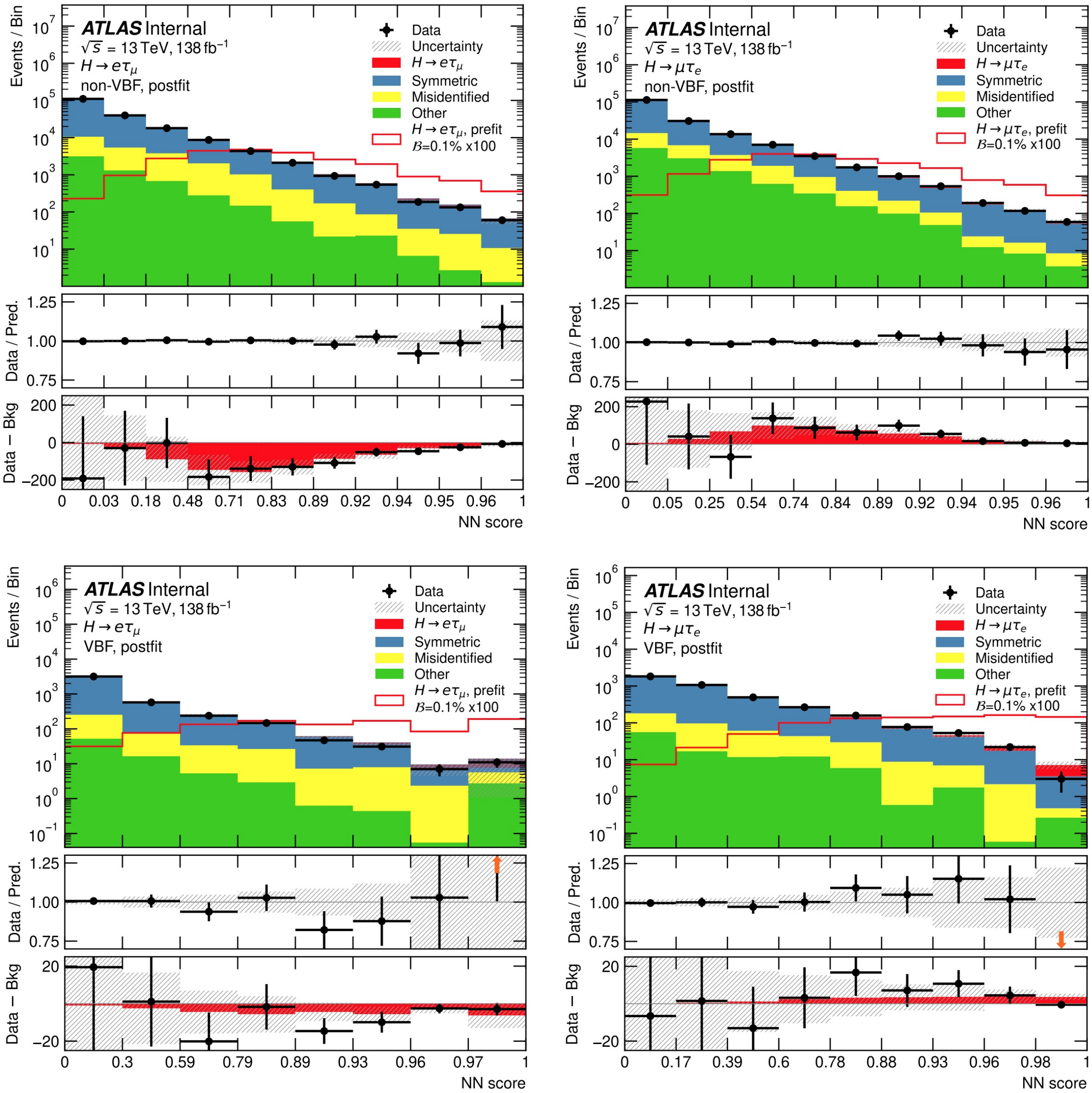


Figure 38: Postfit NN score distributions of the $H \to e\tau_\mu$ (left) and $H \to \mu\tau_e$ (right) signal in the non-VBF (top) and VBF (bottom) categories. The uncertainties (hatched band), combining the statistical, experimental, and theoretical contributions, correspond to the pre-fit values.

to the SM symmetric background. For the efficiency correction, experimental uncertainties related to electron and muon efficiencies are included in both the numerator and denominator of the efficiency-ratio CF shown in (42) and treated as correlated. Efficiency uncertainties provided by the ATLAS performance groups are used for muon and trigger efficiencies and electron, muon, and trigger SFs. For electron efficiency uncertainties, those determined in section 4.4.4 are used. SF uncertainties both included in the efficiency-ratio CF and applied to MC samples are correlated when relevant.

For the signals, the Higgs boson production cross-section uncertainties are obtained following the recommendations of the LHC Higgs Cross-Section Working Group [95]. Effects on the signal expectations are treated as uncorrelated between each production mode. Theoretical uncertainties affecting the ggF signal originate from nine sources [70]. Two of these account for the yield uncertainties and are evaluated by an overall variation of all the relevant scales, and are correlated across all the bins of the NN output distributions [96]. Two sources account for migration uncertainties of zero to one jet and one to at least two jets in the event [96–98]. Two sources account for the Higgs boson p_{T} shape uncertainties and one for the treatment of the top-quark mass in the loop corrections. Finally, two sources account for the acceptance uncertainties of ggF production in the VBF region by selecting exactly two and at least three jets, respectively [99, 100]. For VBF, WH, and ZH production cross-sections, the uncertainties due to missing higher-order QCD (Quantum ChromoDynamics) corrections are estimated by varying the factorization and renormalization scales up and down by factors of two around the nominal. For all production modes, uncertainties are estimated for *parton distribution functions* and α_{s} – the choice of the parton shower and the hadronization model – and missing higher orders in the matrix element calculation. *Parton distribution functions* and α_{s} uncertainties are estimated using the PDF4LHC15nlo set of eigenvectors [101].

4.9 Statistical analysis and results

In this section, we present the implementation of the statistical analysis and the obtained results for the $H \to e\tau_\mu$ and $H \to \mu\tau_e$ searches using the e/μ-symmetry method.

The pre-fit processing of the nuisance parameters is described in section 4.9.1; the obtained results are described in section 4.9.2 for this analysis, while final results in combination with MC-based analyses are presented in section 4.9.3.

4.9.1 Uncertainty preprocessing

There are several steps of preprocessing of the nuisance parameters before the statistical analysis to promote a stable fit procedure: pruning, symmetrization and smoothing. Each of these steps can be performed with various algorithms and options, which affect the analysis results differently. A dedicated study to which the author contributed, comparing fit results for different configurations of the symmetrization and smoothing steps, is presented in appendix B. As a result, the algorithms applied in the final analysis fits are described in the following:

Pruning Nuisance parameters are pruned, i.e., dropped pre-fit if their effects are small to stabilize the fit. The shape and normalization effects of the nuisance parameters are gauged separately. For the shape effect, it is checked if the systematic variation has at least one bin with a difference to the nominal value of at least 0.1%. If not, the shape effect of the corresponding nuisance parameter is pruned. The normalization effect, i.e., the overall effect, of a nuisance parameter is pruned if the impact of both up and down variations is below 0.1%. The pruning procedure is performed for each process and region separately.

Symmetrization In cases where only the up or down variation of a systematic uncertainty is available (one-sided variation), the existent variation is mirrored. For kinematic systematics (which can affect which bin of the NN output an event falls), the up and down variations are symmetrized. The mean of their absolute values is calculated and is set as the new value for the up and the down variation while the signs are kept. Suppose the up and down variation of the kinematic systematic uncertainties are in the same direction compared to nominal in one bin. In that case, the variation with the smaller absolute value gets the opposite sign assigned. In addition, the mean of the absolute values is set as a new value on both sides. Symmetrization is not applied in other cases.

Smoothing Smoothing of the systematic variations is applied to flatten fluctuations due to low statistics. The procedure uses a parabolic smoothing algorithm for noise reduction. This is applied prior to symmetrization for one-sided variations and afterward otherwise.

To illustrate the nuisance parameter preprocessing, a selected number of so-called *envelope* plots is shown in Figure 39 for the $H \to e\tau_\mu$ search and in Figure 40 for the $H \to \mu\tau_e$ search. In each figure, for a given sample's NN distribution, its relative upward and downward fluctuated versions are shown – before and after the preprocessing is applied. Also shown on the plots are the nuisance parameter's ranking and its impact on the signal strength's fitted value which corresponds to the Asimov fit (see definitions in section 2.6). In total, almost 2000 such distributions of systematic uncertainties (number of nuisance parameters times samples they are applied to) per search are entering the statistical analysis.

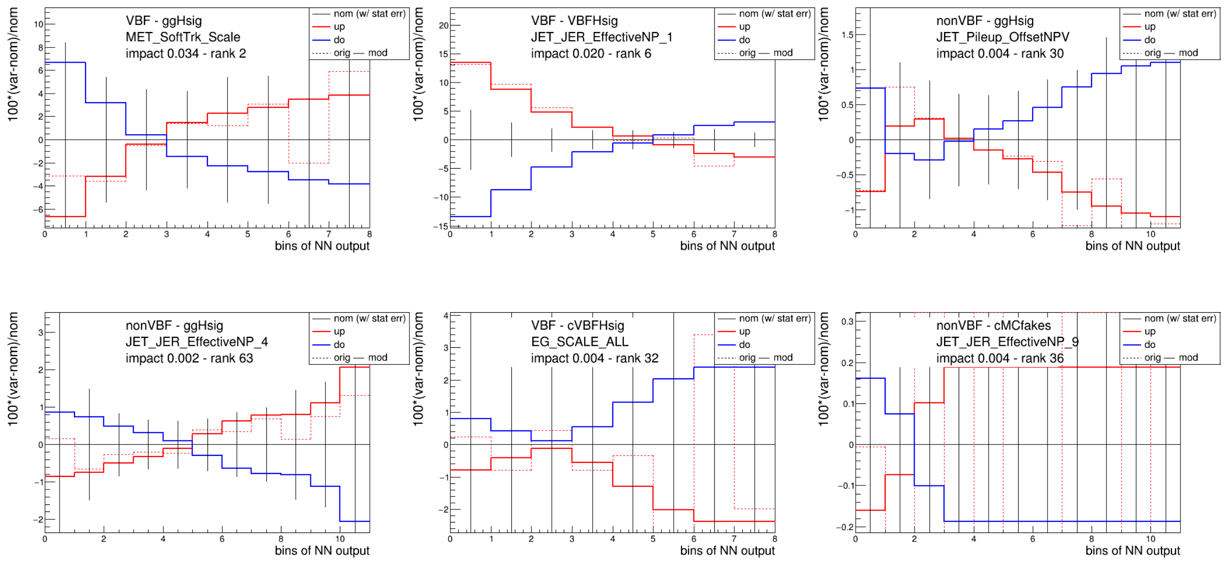


Figure 39: Selected envelope plots for the $H \to e\tau$ search.

4.9.2 Standalone results

In this section, we present the observed and expected result for the two searches, $H \to e\tau_\mu$ and $H \to \mu\tau_e$ (using the unblinded datasets). Each search is performed independently, assuming that the BR of the other signal is zero. The results are obtained by fitting the statistical model described in section 4.3.6 to the histograms of the NN output distributions. A description of the obtention of the results, the different inputs used, and the methods implemented to evaluate the fit performance is given in section 2.6.

The post-fit NN output distributions are shown in Figure 38. The measured best-fit values of the signal strengths are:

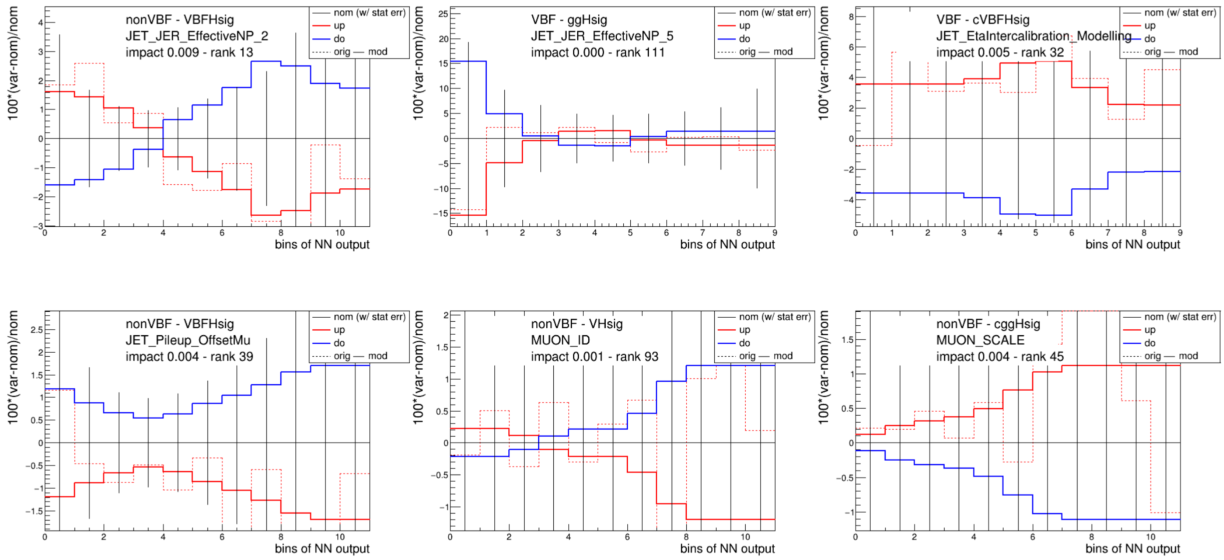


Figure 40: Selected envelope plots for the $H \to \mu\tau$ search.

- $\mu^s_{e\tau} = -0.349^{+0.104}_{-0.108}$ ($H \to e\tau_\mu$ search)
- $\mu^s_{\mu\tau} = 0.248^{+0.101}_{-0.099}$ ($H \to \mu\tau_e$ search)

Since this analysis is performed using the e/μ-symmetry method, which searches for an $H \to \mu\tau$ signal in the $\mu\tau$ dataset by comparing it to the $e\tau$ dataset (and vice-versa for an $H \to e\tau_\mu$ signal), it is expected that the measured signal strengths are opposed in sign. Both obtained values point to an asymmetry between the two datasets in favor of the $\mu\tau$ dataset. This is seen in Figure 38, particularly from the “data - background” panels, where the signal multiplied by the measured signal strength (shown in red) is compatible with the observed disagreement, especially in the non-VBF region. Furthermore, the data agrees very well with the post-fit prediction (background + signal times signal strength) in all regions, as is shown in the “data / prediction” panels.

Scrutinizing the behavior of the various nuisance parameters included in the fit is essential in evaluating its performance. This is done first with the Asimov and mixed datasets and needs to be validated before the unblinded datasets are used for the obtention of results. This study is summarized here using the unblinded datasets, while the ones based on the Asimov and mixed datasets are shown in appendix C.

The nuisance parameters are separated into groups: “JETMET” related to jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainties, “BTag” to b-tagging uncertainties, “Lepton” to experimental electron and muon uncertainties on the (MC to data) SFs, “Lumi” to the luminosity uncertainty, “SigTheory” to the theoretical uncertainties on the signal, “SymmbBackgroundEstimate” to fake and efficiency correction uncertainties, and “Gamma” to the γ parameters which describe the combined background statistical uncertainties. The pulls and constraints on all the nuisance parameters – relative to their pre-fit nominal and uncertainty values – are shown, per group, in Figure 41 and Figure 42 for the $H \to e\tau_\mu$ and $H \to \mu\tau_e$ searches, respectively. All of the pulls of the systematic nuisance parameters are lower or very close to 1σ, and the only noticeable constraint – which appears if the spread of the black line is narrower than the green stripe – is for “El. Fake CF sys”. This nuisance parameter accounts for the flavor uncertainty on the electron fakes in the FF-fake background (see section 4.5.2). The pulls on the γ parameters can be larger, especially in the last NN bins, due to the discrepancy observed between the data and the background prediction (labels “cat inc” and “cat vbf” correspond to bins in the non-VBF and VBF regions respectively).

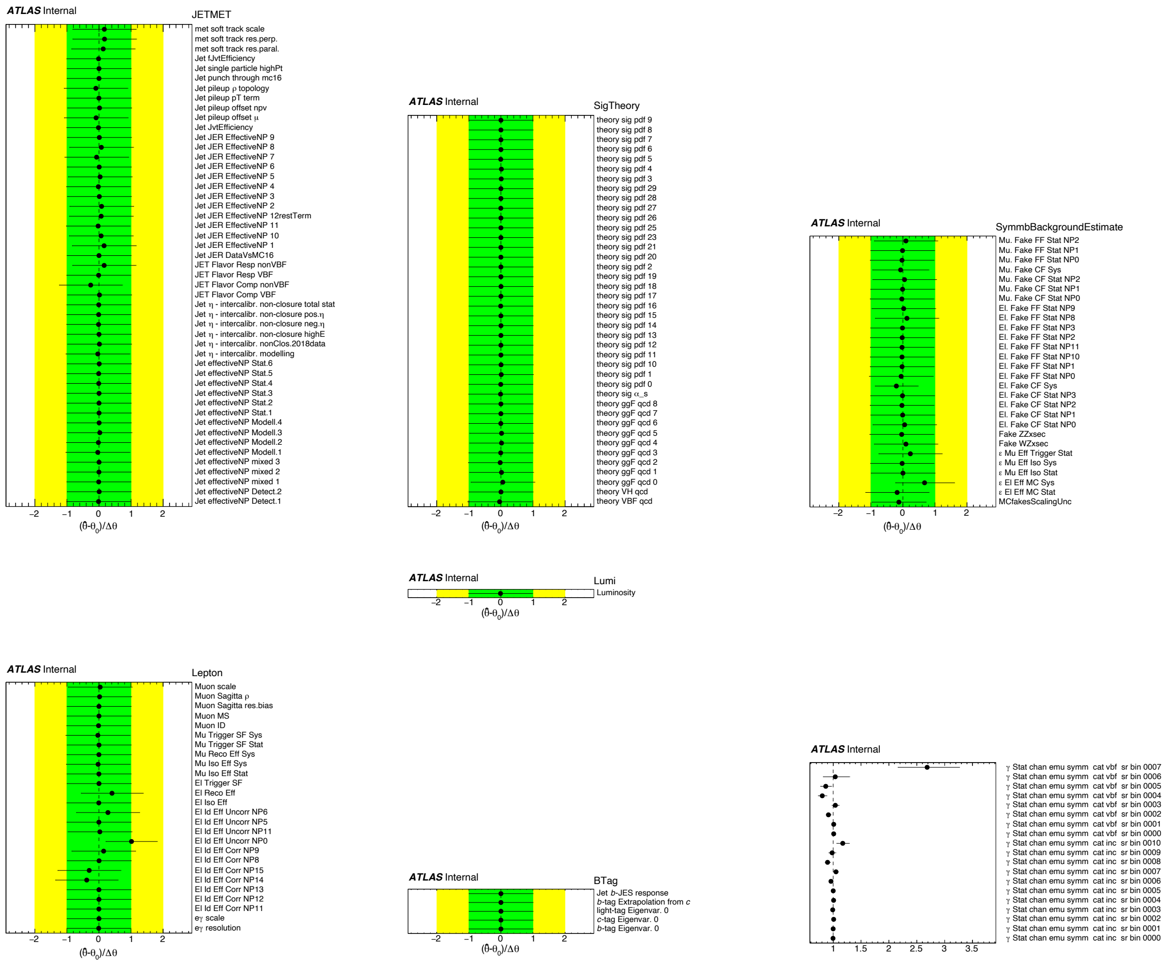


Figure 41: Pulls and constraints of the nuisance parameters in the search for $H \to e\tau_\mu$. The dots indicate the pulls, and the lines represent the constraints. The green band indicates the 1σ band relative to the pre-fit values.

The correlations found between the different nuisance parameters are displayed in Figure 43 for both $H \to e\tau_\mu$ and $H \to \mu\tau_e$. The first column/row of the correlation matrices shows the impact of each nuisance parameter on the signal strength μ^s, which is discussed below (see the ranking of nuisance parameters). The correlations found are reasonable and similar between both searches. The largest correlation is between the “El. Fake CF sys” and “Mu. Fake CF sys” parameters, which both relate to the flavor uncertainty which dominates the fake estimates.

The ranking of the nuisance parameters with respect to their impact on the signal strength μ^s is shown in Figure 44. In addition, the ranking plots also show the pulls and constraints of the nuisance parameters (the black dots with horizontal lines together with the x-axis at the bottom) discussed previously. We find that the nuisance parameters with the largest impact are the γ-parameters – which correspond to the statistical uncertainties of the combined background – especially in the bins of higher NN scores. Since the background estimation is data-driven, this is expected. The impact of the flavor uncertainty on the fake estimate is also significant since it is the largest single uncertainty in the analysis.

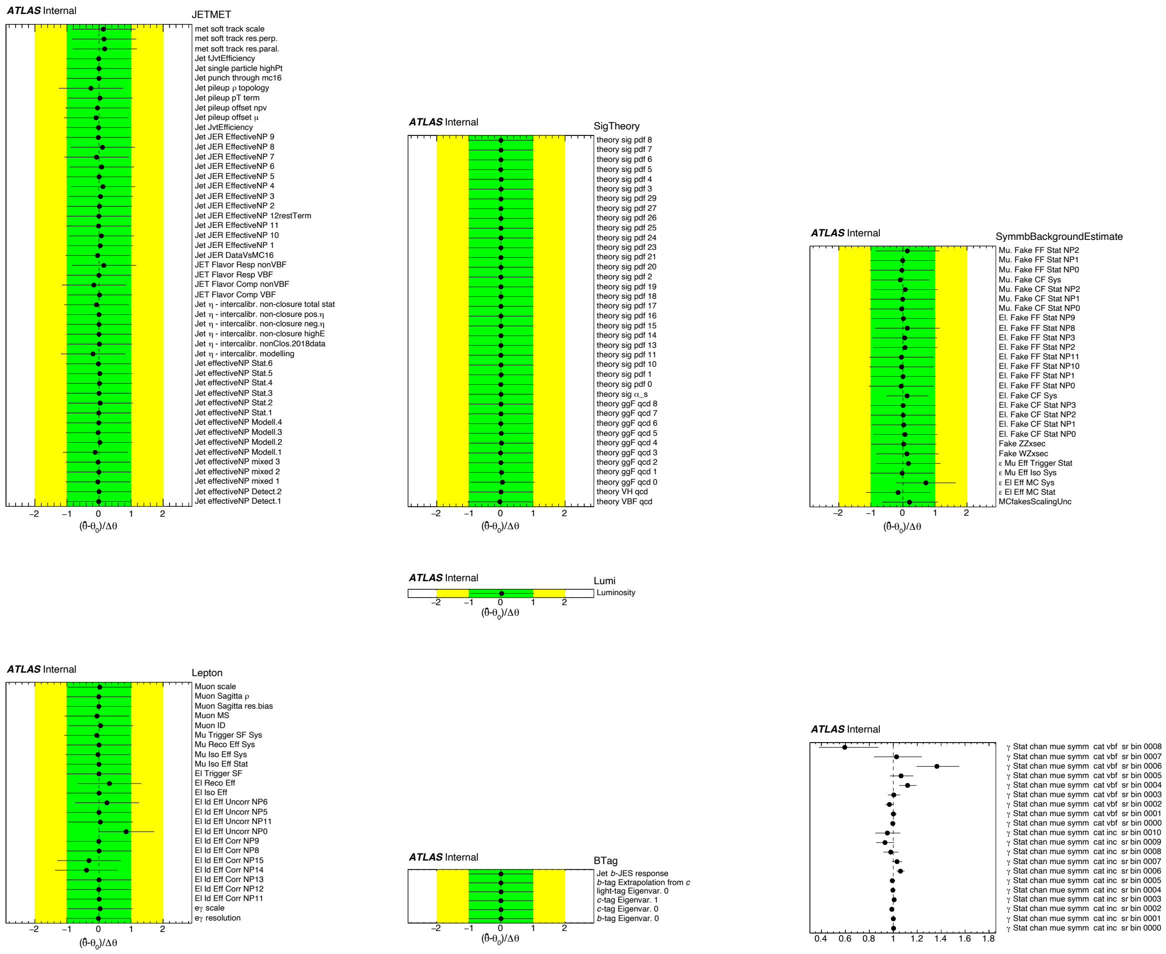


Figure 42: Pulls and constraints of the nuisance parameters in the search for $H \to \mu\tau_e$. The dots indicate the pulls, and the lines represent the constraints. The green band indicates the 1σ band relative to the pre-fit values.

The impact on the signal strength per uncertainty group is shown in Table 14 and Table 15 for the $H \to e\tau_\mu$ and $H \to \mu\tau_e$ searches, respectively. Here also, we see that the largest impact is from the γ-parameters, followed by the group combining uncertainties on the fakes and efficiency correction. With more accumulated data, we expect that both the Gamma and data uncertainties will be reduced, significantly enhancing the sensitivity of the analysis.

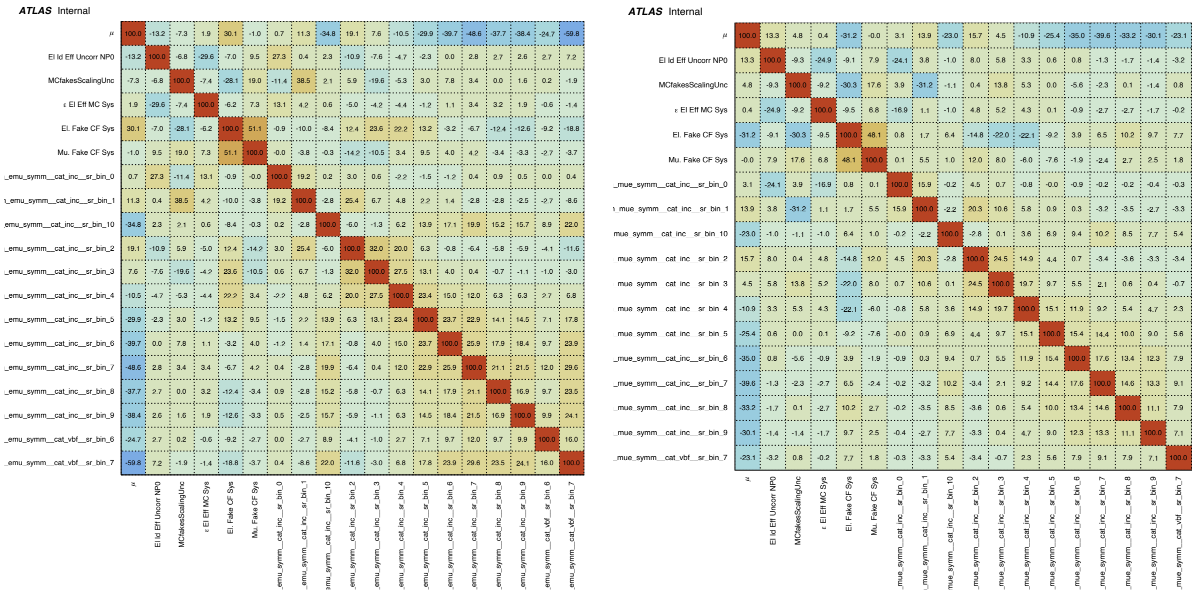


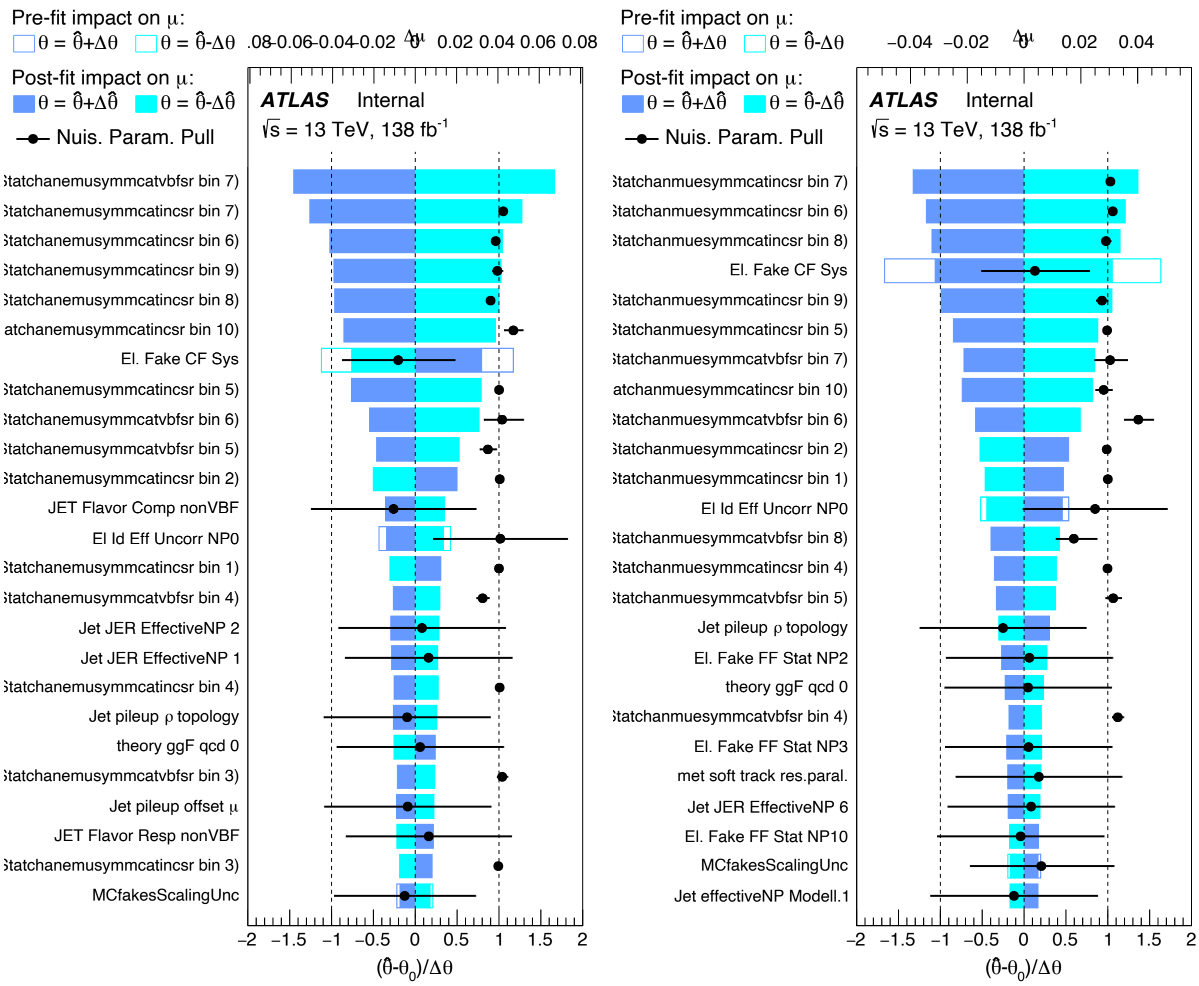
Figure 43: Correlations of the nuisance parameters in the search for $H \to e\tau_\mu$ (left) and $H \to \mu\tau_e$ (right). Only nuisance parameters with correlations above/below $+/-20\%$ are shown.

Group	Impact on unc. of μ	
Full unc.	+0.104	−0.108
Data unc.	+0.047	−0.046
Prediction unc.	+0.093	−0.098
Gammas	+0.081	−0.086
BTag	+0.000	−0.001
JETMET	+0.027	−0.034
Lepton	+0.017	−0.021
Lumi	+0.003	−0.007
SigTheory	+0.007	−0.017
Fakes + Eff. Corr.	+0.053	−0.058

Table 14: Impact of the different uncertainty groups on the uncertainty of the signal strength μ in the search for $H \to e\tau_\mu$.

Group	Impact on unc. of μ	
Full unc.	+0.101	−0.099
Data unc.	+0.056	−0.055
Prediction unc.	+0.084	−0.083
Gammas	+0.067	−0.068
BTag	+0.001	−0.000
JETMET	+0.020	−0.018
Lepton	+0.018	−0.016
Lumi	+0.005	−0.002
SigTheory	+0.013	−0.004
Fakes + Eff. Corr.	+0.053	−0.051

Table 15: Impact of the different uncertainty groups on the uncertainty of the signal strength μ in the search for $H \to \mu\tau_e$.



(a) NP Ranking $e\tau$

(b) NP Ranking $\mu\tau$

Figure 44: The ranking of the nuisance parameters in the search for $H \to e\tau_\mu$ (left) and $H \to \mu\tau_e$ (right). The postfit impact is indicated by the filled bars, while the pre-fit impact is indicated by the empty bars. In addition, the pulls are indicated by the black dots, and the black lines show the constraints.

The observed best fit signal strength (μ^s), significance, and upper limit for the $H \to e\tau_\mu$ and $H \to \mu\tau_e$ searches are listed in Table 16. The corresponding expected results obtained using the Asimov dataset – constructed from the post-fit nuisance parameter values – are shown in brackets. Results from the separate non-VBF and VBF categories are also displayed.

	non-VBF+VBF	non-VBF	VBF
$e\tau$ signif.	−3.240 (6.814)	−3.397 (5.888)	−0.387 (4.099)
$e\tau$ limit (%)	$0.077\ (0.189^{+0.075}_{-0.053})$	$0.082\ (0.216^{+0.086}_{-0.060})$	$0.356\ (0.393^{+0.158}_{-0.110})$
$\mu^s_{e\tau}$	$-0.329^{+0.104}_{-0.108}$	$-0.377^{+0.114}_{-0.119}$	$-0.089^{+0.221}_{-0.247}$
$\mu\tau$ signif.	2.498 (8.217)	2.694 (7.405)	0.301 (3.992)
$\mu\tau$ limit (%)	$0.415\ (0.189^{+0.074}_{-0.053})$	$0.485\ (0.221^{+0.087}_{-0.062})$	$0.466\ (0.381^{+0.157}_{-0.106})$
$\mu^s_{\mu\tau}$	$0.248^{+0.101}_{-0.099}$	$0.299^{+0.112}_{-0.110}$	$0.060^{+0.214}_{-0.194}$

Table 16: The observed results. The expected sensitivities are given in brackets. The expected significance is obtained when assuming a BR of 1%.

The asymmetry favoring the $\mu\tau$ over the $e\tau$ dataset discussed above corresponds to a 2.5σ excess in the $H \to \mu\tau_e$ search and -3.2σ excess in the $H \to e\tau_\mu$ search. A significance of 3σ is roughly interpreted as a 1/1000 chance that the excess results from a fluctuation in the data from the background-only hypotheses. This asymmetry is mainly found in the non-VBF region. The signal strengths μ^s measured in the VBF region – although in the same direction (sign) as those of non-VBF – are compatible with the background-only hypotheses. This asymmetry also has implications for the observed limits; the observed limit of the non-VBF and VBF regions combined is $\sim 2\sigma$ below the expected in $H \to e\tau_\mu$ and $\sim 3\sigma$ above the expected in $H \to \mu\tau_e$.

4.9.3 Combined results

As previously discussed at the beginning of section 4, the full ATLAS Run-2 Higgs LFV search includes, in addition, two other analyses which rely on MC simulation for background estimation; MC-based $\ell\tau_{\ell'}$ searching for the same signals as the analysis presented here ($H \to e\tau_\mu$ and $H \to \mu\tau_e$) and MC-based $\ell\tau_{\mathrm{had}}$ searching for the same signals but when the τ decays hadronically ($H \to e\tau_{\mathrm{had}}$ and $H \to \mu\tau_{\mathrm{had}}$). The two final states, $\ell\tau_{\ell'}$ and $\ell\tau_{\mathrm{had}}$, are combined to measure the final analysis results. In the following, we refer to the analysis presented in this thesis as the Symmetry-based $\ell\tau_{\ell'}$ analysis.

The best fit signal strengths measured with the combined MC-based analyses are $\mu^s_{e\tau} = 0.11 \pm 0.06$ and $\mu^s_{\mu\tau} = 0.12 \pm 0.05$. This means that the MC-based analyses measure an excess in the data w.r.t. the background prediction in both the $e\tau$ and $\mu\tau$ datasets. The Symmetry-based search measures $\mu^s_{e\tau}$ with an opposite sign, but in contrast with the MC-based analysis, it makes use of the $\mu\tau$ dataset to estimate the background; therefore, it measures the difference between the two signal strengths. On the other hand, the $\mu^s_{\mu\tau}$ measured values are comparable; although the MC-based value is less than half of the one measured with the Symmetry-based search, due to smaller uncertainties, the corresponding significances are similar. Compatibility tests were performed to verify that the results from the Symmetry-based and MC-based $\ell\tau_{\ell'}$ analyses are consistent, which consider the difference between the two signal strengths measured. A compatibility within 2.3σ was found between the two methods.

The results obtained when combining the Symmetry-based $\ell\tau_{\ell'}$ and MC-based $\ell\tau_{\rm had}$ analyses are shown in Table 17. The combined measured signal-strength values are decreased, especially in the $H \to e\tau$ search, since the $\mu^s_{e\tau}$ value found in the MC-based search is opposite in sign. But in the $H \to \mu\tau$ search, the significance is increased to almost 2.9σ due to the uncertainty reduction. We see from the expected results that the combination improves the sensitivity of the analysis by up to 60%.

	$H \to e\tau$	$H \to \mu\tau$
μ^s	$-0.118^{+0.066}_{-0.067}$	$0.133^{+0.048}_{-0.047}$
signif.	−1.796 (9.560)	2.864 (12.586)
limit (%)	0.073 ($0.128^{+0.051}_{-0.036}$)	0.214 ($0.091^{+0.037}_{-0.026}$)

Table 17: Observed results of the combination with the MC-based $\ell\tau_{\rm had}$ analysis. The expected sensitivities are given in brackets. The expected significance is obtained when assuming a BR of 1%.

The results presented above are relative to the Symmetry-based searches described in this thesis. The final results, which will figure in the paper once published, are based on a different combination of the distinct analyses performed. Indeed, various combinations can be considered. The choice was made prior to the data unblinding, by comparing the expected sensitivities achieved with the different options. Two sets of results with different physical implications are derived:

- 1 POI fit: Independent searches for $H \to e\tau$ and $H \to \mu\tau$ (assuming that the BR of the other signal is zero) combining the MC-based $\ell\tau_{\ell'}$ in the non-VBF SR, the Symmetry-based $\ell\tau_{\ell'}$ in the VBF SR, and the MC-based $\ell\tau_{\rm had}$ in the non-VBF and VBF SRs
- 2 POI fit: Simultaneous search for $H \to e\tau$ and $H \to \mu\tau$ combining the MC-based $\ell\tau_{\ell'}$ and $\ell\tau_{\rm had}$ in the non-VBF and VBF SRs

In the independent searches (1 POI fit), for the $\ell\tau_{\ell'}$ channel, the MC-based analysis has better expected sensitivity than the Symmetry-based in the non-VBF SR but worse in the VBF SR. This is mainly driven by the amount of MC statistics available in each SR in comparison to the data statistics. For the $\ell\tau_{\rm had}$ channel, only an MC-based analysis was performed. The simultaneous search (2 POI fit) is a single measurement of both the $H \to e\tau$ and $H \to \mu\tau$ searches. This can be implemented in the MC-based analyses since the background predictions derived for each signal are independent of the data and one another. This is not the case in the Symmetry-based searches, where only independent measurements can be performed.

Table 18 lists the observed best fit signal strength and upper limit for the $H \to e\tau$ and $H \to \mu\tau$ searches and for the 1 POI and 2 POI combinations. Based on the simultaneous fit, the $H \to \mu\tau$ search shows a 2.5σ excess, in line with the independent fit, while no significant excess was observed in the $H \to e\tau$ search, not supporting the hint in the independent fit.

4.10 Conclusion

Two direct searches for Higgs LFV decays, $H \to e\tau_\mu$ and $H \to \mu\tau_e$, are presented based on data collected by the ATLAS experiment at $\sqrt{s}$ = 13 TeV, corresponding to an integrated luminosity of 138.42 fb^{-1}.

The searches are conducted using the e/μ-symmetry method by comparing the data in the $e\tau_\mu$ and $\mu\tau_e$ datasets – both include events with one electron and one muon in the final state but differ by the $p_{\rm T}$ ordering

	1 POI fit		2 POI fit	
	$H \to e\tau$	$H \to \mu\tau$	$H \to e\tau$	$H \to \mu\tau$
μ^s	$0.129^{+0.061}_{-0.060}$	$0.084^{+0.047}_{-0.046}$	$\mu_{e\tau} = 0.094^{+0.059}_{-0.058}$	$\mu_{\mu\tau} = 0.107^{+0.045}_{-0.044}$
limit (%)	$0.230\ (0.118^{+0.047}_{-0.033})$	$0.163\ (0.089^{+0.036}_{-0.025})$	$0.192\ (0.114^{+0.046}_{-0.032})$	$0.182\ (0.087^{+0.035}_{-0.024})$

Table 18: Observed best fit signal strength and upper limit obtained from the independent (1 POI fit) and simultaneous (2 POI fit) analyses, which combine results from the $\ell\tau_{\ell}$ and $\ell\tau_{\text{had}}$ searches as described in the text. The expected sensitivities are given in brackets.

of the two leptons in the estimated Higgs rest frame. Significant efforts were made to correctly account for detector-induced asymmetries between the two datasets: applying the efficiency correction and estimating the fake background contribution. The results are extracted in a statistical analysis based on the output of a NN, trained to identify the LFV signal.

The observed best fit values of the signal strengths are $\mu_{e\tau} = -0.329^{+0.104}_{-0.108}$ and $\mu_{\mu\tau} = 0.248^{+0.101}_{-0.099}$ for the $H \to e\tau_{\mu}$ and $H \to \mu\tau_{e}$ signals, respectively, when assuming that the other signal is zero. This corresponds to an excess of -3.2σ and 2.5σ. Both values point to an asymmetry between the two datasets in favor of the $\mu\tau$ dataset. The expected significances obtained when assuming a BR of 1% on the signals are 6.814 and 8.217. Additionally, observed (expected) limits on the BRs are set at 95% CL using the CL$_s$ method: 0.077% ($0.189^{+0.075}_{-0.053}$%) for $H \to e\tau$ and 0.415 ($0.189^{+0.074}_{-0.053}$%) for $H \to \mu\tau$. The observed limits are affected by the asymmetry present between the $e\tau$ and $\mu\tau$ datasets.

Combination with the MC-based $\ell\tau_{\text{had}}$ search reduces the significance in the $H \to e\tau$ search to -1.8σ and increases it in the $H \to \mu\tau$ search to 2.9σ. The observed (expected) upper limits set are 0.073% ($0.128^{+0.051}_{-0.036}$%) for $H \to e\tau$ and 0.214% ($0.091^{+0.037}_{-0.026}$%) for $H \to \mu\tau$. The combination improves the expected sensitivity of the analysis by up to 60%.

The final results derived for this analysis include independent $H \to e\tau$ and $H \to \mu\tau$ searches combining the Symmetry and MC-based analyses in the $\ell\tau_{\ell'}$ channel and the MC-based analysis in the $\ell\tau_{\text{had}}$ channel; as well as a simultaneous $H \to e\tau$ and $H \to \mu\tau$ search combining the $\ell\tau_{\ell'}$ and $\ell\tau_{\text{had}}$ MC-based analyses. Based on the simultaneous fit, the $H \to \mu\tau$ search shows a 2.5σ excess in line with the independent fit, while no significant excess was observed in the $H \to e\tau$ search, not supporting the hint in the independent fit. The observed (expected) upper limits set with the independent searches are 0.230% ($0.118^{+0.047}_{-0.033}$%) for $H \to e\tau$ and 0.163% ($0.089^{+0.036}_{-0.025}$%) for $H \to \mu\tau$. The observed (expected) upper limits set with the simultaneous search are 0.192% ($0.114^{+0.046}_{-0.032}$%) for $H \to e\tau$ and 0.182% ($0.087^{+0.035}_{-0.024}$%) for $H \to \mu\tau$.

5 Towards symmetry-based data-directed searches

In this section, we present different studies that aim to lay the groundwork for implementing data-directed and generic searches for BSM physics, using the e/μ-symmetry method or similarly exploiting SM symmetries in general. The motivation to conduct such searches is summarized in section 5.1. A simple approach to identify asymmetries between two measurements, which exhibits good sensitivity and can easily be generalized, is presented in sections 5.2 and 5.3. This approach is illustrated in the example of the search for Higgs LFV decays, which is constructed using standalone simulated data produced with dedicated software. The procedure employed to generate this data is described in section 5.4, along with an independent study that probes the validity of generalizing the e/μ-symmetry method to datasets with various final states. Finally, section 5.5 presents an alternative approach for symmetry restoration, which can be efficiently implemented to account for any asymmetry accurately modeled in the MC samples.

5.1 Motivation

We propose extending the discovery potential of the LHC with a DDP. Similar to [102–104], its principal objective is to efficiently scan large portions of the observables space for hints of NP, but unlike [102–104], without using any MC simulation. We look directly at the data in an attempt to identify regions in the observables space that exhibit deviations from a theoretically well-established property of the SM. Such regions should be considered as data-directed BSM hypotheses, as opposed to theoretically-motivated ones, and could be studied using traditional data-analysis methods. As detailed in [62], a search in the DDP can be implemented with two key ingredients: a) a theoretically well-established property of the SM and b) an efficient algorithm to search for deviations from this property.

In this section, we show that any SM symmetry can be exploited in such a data-directed search. Symmetries can be used to split the data into two mutually exclusive datasets, which should only differ by statistical fluctuations. By comparing them, we become sensitive to any potential BSM process which breaks this symmetry. An example is the e/μ-symmetry method described in section 3, which enables comparing any two datasets of events with the same number of electrons in one, and muons in the other, expected symmetric from the approximate SM $e \leftrightarrow \mu$ symmetry. In the general case, systematic detector effects could also affect the symmetry, such as the different electron vs. muon efficiencies. These need to be accounted for – e.g., applying an efficiency correction as in section 4.3.4. Still, such correction methods exist in principle. So as a first step, we do not consider them in the proof-of-principle study presented in section 5.2. In an experimental realization of the symmetry-based DDP search, such systematic effects must be taken into account.

The concept of exploiting symmetries of the SM for data-driven BSM searches was previously proposed in [56] and [105]. It is also implemented – based on the e/μ-symmetry method– in the ATLAS Higgs LFV search presented in section 4 and in [57] and in the search for an asymmetry between $e^+\mu^-$ and $e^-\mu^+$ events [106]. However, the implementation of these searches still follows the blind-analysis paradigm where only a specific signal is searched for in a small theoretically-motivated subset of the observables space.

In terms of the DDP proposed here, no specific signal is searched for. Instead, the two SM symmetric datasets are compared in full in many different sub-regions (corresponding to exclusive selections of the data). Any significant deviation observed is a potential sign for NP to be further investigated. Thus, sensitivity to many more possible BSM processes and scenarios is enabled. As such, the $e\mu/\mu e$ comparison implemented in the Higgs LFV search discussed above becomes a general test for lepton flavor universality in the final state containing one electron and one muon of opposite charge. Similarly, different final states including a number of electrons, muons, and other objects, can be probed (ee vs. $\mu\mu$, e+jet vs. μ+jet, etc.), each potentially sensitive to different BSM manifestations. In this context, the recent hints for non-universality in the R_K measurements from LHCb [23] are, in fact, hints of an asymmetry between the ee and $\mu\mu$ datasets in the decay of b hadrons to a Kaon and two same-flavor leptons. Likewise, the comparison of $e^-\mu^+$ to $e^+\mu^-$ in [105] and [106] is a test of CP (Charge-Parity) symmetry in the lepton sector. Other symmetries could be used in similar implementations, such as forward-backward or time-reversal symmetries.

Given the large number of symmetries in the SM which can be violated in BSM scenarios, the potential benefits of implementing such symmetry-based generic searches are significant. However, interpreting the results must be done with care. Indeed, a data-directed search will naturally be tuned to identify regions including statistical fluctuations or other measurement effects which could induce asymmetries. If a detected signal originates from a statistical fluctuation, it will disappear with more collected data. If it originates from a detector or other systematic effect correctly modeled in MC simulations, then it can be ruled out. Any residual asymmetry can be considered a data-directed BSM hypothesis to be inspected using standard analysis techniques. In this manner, the risk of claiming a false discovery should not be higher than when implementing hundreds of searches in the blind paradigm since the trial factor is high in both cases [107].

5.2 Identifying asymmetries between two measurements

This section aims to draw attention to the potential for discovering BSM physics when implementing searches in the DDP, particularly data-directed searches based on symmetries of the SM. In this context, we lay the groundwork for a generic method to compare two datasets and quantify the level of any discrepancy between them, if present. As previously discussed, we do not address here the treatment of eventual systematic effects that can deteriorate the expected SM symmetry between the two datasets. Nonetheless, as shown in section 4, [57], and [106], in analyses that were based on symmetry considerations, such effects can be accounted for. The results presented in this section also figure in [1].

Since the goal is to quickly scan multiple sub-regions of the observables space in a large number of final states, a fast method for identifying asymmetries is needed. We develop this method based on a simplified framework using MC simulated data. Different test statistics can be used to compare the two datasets (e.g., Kolmogorov-Smirnov [108], student t-test [109]). In the implementation proposed in this section, the datasets are represented by 2D histograms[8] of predetermined properties of the data and compared using a simple test statistic, the N_σ test, described below. Since the method is fast, multiple 2D histograms of all the existing properties and their combinations can be compared efficiently. We leave to future work the generalization of this study for a more comprehensive and optimized implementation.

When working with histograms, there is no a priori way to choose the bins, which is particularly challenging in many dimensions. One solution to this challenge is to make use of machine learning. Starting from [110, 111] based on [112], there have been a variety of proposals to perform anomaly detection with machine learning by comparing two datasets [110, 111, 113–121] (see [122–126] for recent reviews). Complementary to the binned DDP (henceforth, simply "the DDP"), we demonstrate that asymmetries can also be identified using weakly supervised NNs, similar to the approach in [112]. Nevertheless, for now, such methods require training at least one NN for each event selection. This is time-consuming and

[8] The generalization of the proposed analysis approach to n-dimensional histograms is straightforward.

restricts the number of selections that can be tested, which could be limiting in the context of the DDP, depending on the available computational resources.

The sensitivity of the proposed DDP search is compared to that of two likelihood-based test statistics. While both assume exact knowledge of the signal shape, one represents an ideal search in which also the distribution of the symmetric background components are precisely known, and the other represents the expected sensitivity of a traditional blind analysis search employing a symmetry-based background estimation, such as the Higgs LFV analysis described in section 4. According to the Neyman-Pearson lemma [127], these are the most sensitive tests for the respective scenarios they consider.

This section is organized as follows. In section 5.2.1, we describe some of the statistical properties of the DDP symmetry search. The simulated data used for our numerical studies are presented in section 5.2.2. Results for the DDP are given in section 5.2.3, and a complementary approach using NNs is discussed in section 5.2.4. We end with conclusions and outlook in section 5.2.5.

5.2.1 Quantifying asymmetries

Given two datasets, our goal is to determine the probability that they are *asymmetric*, as opposed to originating from the same underlying distribution. The latter represents the null hypothesis, where both measurements are indeed symmetric as expected from the symmetry property of the SM considered. In the context of the symmetry-based DDP proposed here, and unlike other statistical tests commonly used in BSM searches, no signal assumptions are made. The test is intended to output the probability at which the background-only hypothesis is rejected.

In order to rapidly scan many selections and final states, the method used to quantify the asymmetry between two datasets should be efficient. This can be achieved if we ensure that the results obtained are independent of the properties of the underlying symmetric background component. Indeed, one of the most time-consuming tasks for implementing a statistical test to reject a hypothesis is determining the test statistic's PDF under said hypotheses. But if this PDF is constant and known, we avoid the need to derive it for each different dataset tested.

The generic N_σ test statistic considered is given in (59). A and B are two n-dimensional matrices representing the two tested datasets projected into histograms of n properties of the measurements. They each have M bins in total, the A_i and B_i are their respective number of entries in bin i, and the σ_{Ai} and σ_{Bi} are their respective standard errors:

$$N_\sigma(B, A) = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} \frac{B_i - A_i}{\sqrt{\sigma_{Ai}^2 + \sigma_{Bi}^2}} \tag{59}$$

In this formalism, we search for a signal in B by comparing it to the reference measurement A, but their roles are exchangeable. When A and B are two (Poisson-distributed) measurements, (59) simplifies to:

$$N_\sigma(B, A) = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} \frac{B_i - A_i}{\sqrt{A_i + B_i}} \tag{60}$$

It can be shown that in the limits of the normal approximation, applicable here provided there are enough statistics in each bin of the two matrices, the symmetry-case PDF of the N_σ test is well approximated by a standard Gaussian. This satisfies the condition that the test should be independent of the underlying symmetric component, ensuring its efficiency. We confirmed that this approximation is valid when ensuring at least 25 entries per bin. For scenarios with lower statistics, the background-only PDF can be distorted from the normal distribution. Nevertheless, large N_σ values still correspond to asymmetries. Further studies on the validity in low statistics are discussed in section 5.3.2.

The performance of the N_σ test is compared to that of two distinct likelihood-based test statistics, which are built on the test statistic for the discovery of a positive signal described in section 2.6 and rely on the full knowledge of the signal shape that is being searched for:

- q_0^{L1} assumes that the underlying symmetric component is perfectly known. This is equivalent to the ideal analysis case in which the signal and background distributions are perfectly known (no uncertainties)
- q_0^{L2} uses no a priori knowledge of the underlying symmetric distribution and estimates it from the two measurements as part of the fitting procedure. This represents the case where symmetry is the only available information

Since we aim to compare the sensitivity to detect asymmetries using the N_σ test relative to the likelihood-based tests, statistical uncertainties on the signal are not included in this study. The likelihood functions for each scenario are shown below, where S is the shape of the signal considered, B is the tested dataset, T is the true distribution of the symmetric background, and A is a measurement of T. The parameter μ represents the signal strength, and $b = \{b_i\}$ are the background parameters (one per bin of the compared matrices):

$$L1_\mu(B, T, S) = \text{Poisson}(B \mid T + \mu S) \tag{61}$$

$$L2_\mu(B, A, S; b) = \text{Poisson}(B \mid b + \mu S) \cdot \text{Poisson}(A \mid b) \tag{62}$$

The formalism used, which permits a comparison with the N_σ test, is shown in (63) and (64), where L_μ is the likelihood function (either $L1_\mu$ or $L2_\mu$), λ_μ is the profile likelihood ratio, $\hat{\mu}$ and $\hat{b}$ are the maximum likelihood estimators of μ and the b_i parameters, and $\hat{\hat{b}}$ is the maximum likelihood estimator of the b_i when μ is fixed.

$$\lambda_\mu(B, A, S) = \frac{L_\mu(B, A, S; \hat{\hat{b}})}{L_{\hat{\mu}}(B, A, S; \hat{b})} \tag{63}$$

$$q_0(B, A, S) = \begin{cases} -2\ln\lambda_0(B, A, S) & , \hat{\mu} \geq 0 \\ +2\ln\lambda_0(B, A, S) & , \hat{\mu} < 0 \end{cases} \tag{64}$$

When performing a test for discovery, we compare the test's score to the background-only PDF to obtain a p-value (p), which measures the level at which the background hypothesis can be rejected. We then translate this p-value into an equivalent significance $Z = \Phi^{-1}(1-p)$, where Φ^{-1} is the quantile of the standard Gaussian. A significance of 5 is commonly considered an appropriate level to constitute a discovery, corresponding to $p \approx 2.87 \times 10^{-7}$. For the case of the N_σ test, the background-only PDF is itself a standard Gaussian. Therefore the score obtained is directly a measure of the obtained significance Z, bypassing the need to compute the p-value:

$$Z = N_\sigma(B, A) \tag{65}$$

Similarly, regarding the q_0 test, we know from [54] that:

$$Z = \sqrt{q_0}(B, A, S) \tag{66}$$

So the $\sqrt{q_0}$ background-only PDF is again a standard Gaussian[9]. Therefore, in the following, we directly compare the N_σ and $\sqrt{q_0}$ significance values.

5.2.2 Data preparation

The symmetry-based DDP is demonstrated in a practical example, the search for Higgs LFV decay $H \to \tau\mu$ where the τ further decays to an electron, based on the example discussed in section 3.2. The SM processes considered which contribute to the symmetric background include Drell-Yan, diboson, Wt, $t\bar{t}$, and SM Higgs ($H \to WW/\tau\tau$). For each of these processes, a dataset equivalent to 40 fb^{-1} of pp collisions at $\sqrt{s}$ = 13 TeV was generated using MadGraph 2.6.4 [128] and Pythia 8.2 [61]. The response of the ATLAS detector was emulated using Delphes 3 [129]. The signal processes considered are ggF and VBF Higgs production mechanisms. The generation of these MC simulated samples is described in detail in section 5.4.1. The SM events are used to construct an $e\mu$ symmetric template matrix T – representing the SM background underlying distributions from which symmetric datasets will be drawn (see description of this process further below). The Higgs LFV signal events are used to construct a normalized signal template matrix S. This is done by projecting the simulated measured events on a 28×28 2D histogram with two selected event properties:

- x-axis: collinear mass (defined in (36)), 5 GeV bins from 30-170 GeV
- y-axis: leading lepton p_{T}, 5 GeV bins from 10-150 GeV

To demonstrate the concept and allow quantitative comparisons to the performance of the likelihood-based tests, we avoided bins with low statistics by adding a flat 25 entries to each bin in T. The resulting T and S templates are shown in Figure 45. Other background and signals considered are flat T background distributions (with either 100 or 10^4 entries per bin) and rectangle, 2D Gaussian, and slope-like signal S templates described in section 5.3.1.

Given a background template T – which represents the underlying symmetric distribution – and a signal template S – which can be injected with different levels of signal strength – the procedure to generate the datasets used to qualify the different tests is as follows. From T, we Poisson draw N pairs of (A, B) background-only measurements, which are symmetric up to statistical fluctuations. The background + signal measurements B^s are obtained by injecting some signal into the B datasets. We inject the signal with a signal strength μ_{inj}, determined such that a q_0 test for discovery (q_0^{L1} or q_0^{L2}) outputs a given significance Z_{inj} when testing $B^s = B + \mu_{\mathrm{inj}} S$ against B:

$$\sqrt{q_0}(B + \mu_{\mathrm{inj}} S, B, S) = Z_{\mathrm{inj}} \tag{67}$$

Since S is normalized, μ_{inj} is the number of signal events added to the B dataset.

Explicitly, for the q_0^{L1} and q_0^{L2} cases, it is found by solving (68) and (69), respectively:

$$2\left(-\mu_{\mathrm{inj1}} + \sum_{i=1}^{M}\left[(B_i + \mu_{\mathrm{inj1}} S_i)\ln\left(1 + \mu_{\mathrm{inj1}}\frac{S_i}{B_i}\right)\right]\right) = Z^2_{\mathrm{inj1}} \tag{68}$$

[9] This can also be shown in the more common single-sided formalism presented in [54], where the background-only PDF of q_0 in the asymptotic limit is given by $\frac{1}{2}(\delta(0) + \chi^2_1)$, where χ^2_1 is the χ^2 distribution with one degree of freedom. Thus the PDF of $\sqrt{q_0}$ is $\frac{1}{2}(\delta(0) + \chi_1)$, and the χ_1 distribution is the *half-normal* distribution.

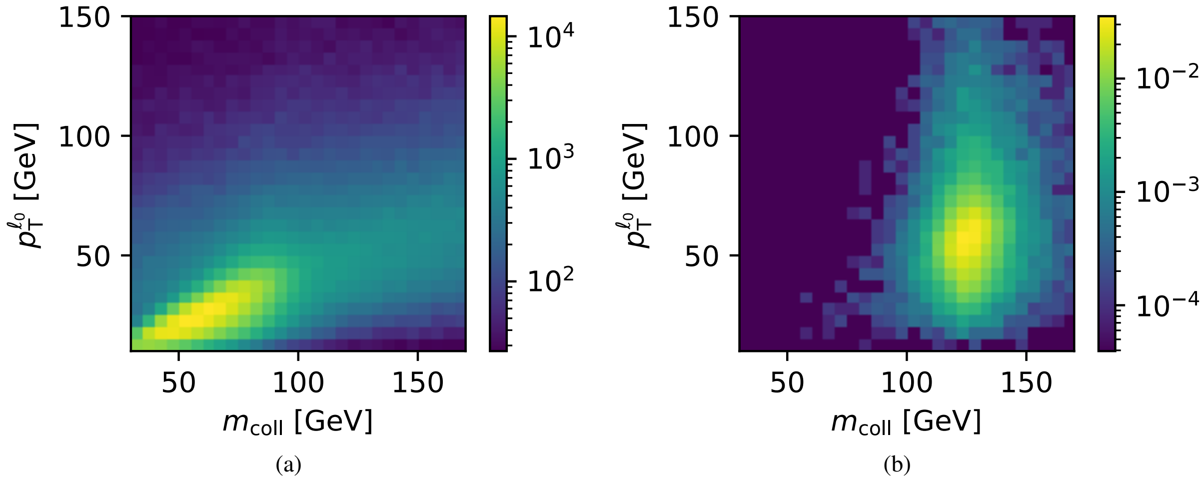


Figure 45: The e/μ background template matrix T (left) and the Higgs LFV signal template matrix S (right). The x, y, and z axes are the collinear mass, leading lepton p_T, and the number of entries per bin, respectively (S is normalized).

$$2\sum_{i=1}^{M}\left[(B_i+\mu_{\text{inj2}}S_i)\ln\left(1+\mu_{\text{inj2}}\frac{S_i}{2B_i+\mu_{\text{inj2}}S_i}\right)-B_i\ln\left(1+\mu_{\text{inj2}}\frac{S_i}{2B_i}\right)\right]=Z_{\text{inj2}}^2 \tag{69}$$

For each separate experiment considered and detailed below, the number of A, B, and $B^s = B + \mu_{\text{inj}}S$ matrices we generate is $N = 20\text{K}$. For the N_σ and q_0^{L2} tests, the PDFs of the symmetric case (background-only) are obtained by comparing the B and A pairs, and the PDFs of the asymmetric case (signal+background) by comparing the B^s and A pairs. The same applies to the q_0^{L1} test when the A matrices are replaced by the template T.

5.2.3 Results

Focusing on the Higgs LFV example, using the signal (S) and background (T) templates shown in Figure 45, we apply an injected signal strength μ_{inj} which corresponds to the 5σ significance of the ideal q_0^{L1} test. To give an impression, when applied to T, this corresponds to a signal fraction of 0.2% or in a 6×6 window centered around the signal of 2.8%. Figure 46 compares Z PDFs obtained with the q_0^{L1}, q_0^{L2}, and N_σ tests. As expected, the symmetric-case PDFs of all tests are consistent with standard Gaussian distributions. We observe that the background + signal (asymmetric-case) PDFs are consistent with Gaussians with variance 1 ± 0.05 (for all examples considered), centered around the resulting average significance Z_{avg} of the relevant test. The Z_{avg} of each test can be directly estimated using the *Asimov* dataset described in section 2.6, setting $A = T$ and $B^s = T + \mu_{\text{inj}}S$. The resulting significance with the q_0^{L1} test is predictably $Z_{\text{avg}} = 5.0 \approx Z_{\text{inj}}$. With $Z_{\text{avg}} = 3.53$, q_0^{L2} is less sensitive than q_0^{L1} since it does not use an a priori knowledge of the background but estimates it from the two measurements as part of the fitting procedure. Since the N_σ test is averaged on all the bins, and most of them only include background contributions, the resulting average significance $Z_{\text{avg}} = 1.48$ is significantly lower than the separation power measured with the q_0^{L2} test.

In general, applying the N_σ test in a sub-region of the datasets can be much more efficient. Even though the signal's shape and location are not known in a generic test, since the calculation of N_σ is fast, one

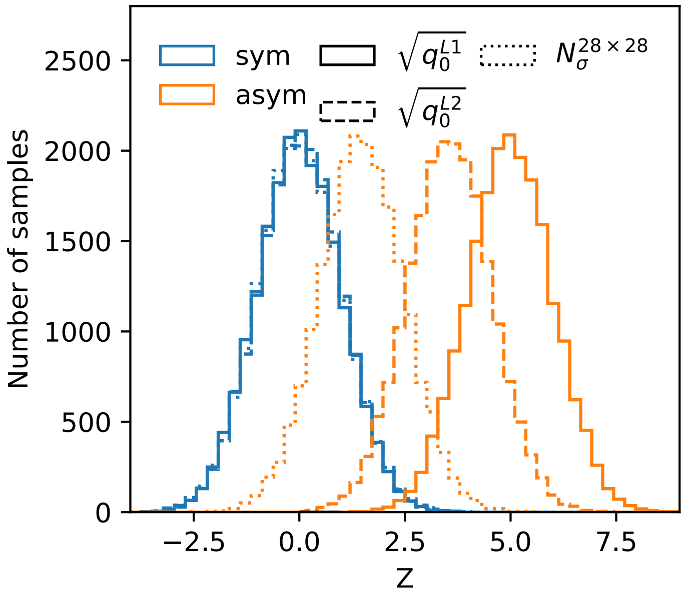


Figure 46: Significance PDFs comparing results of the N_σ, q_0^{L1}, and q_0^{L2} tests for the Higgs LFV example, with injected signal strength corresponding to 5σ of q_0^{L1}.

could test multiple bin subsets[10]or develop an algorithm to optimize this selection. In Figure 47, we show N_σ scores with the Asimov data obtained when the test is performed on square windows of different sizes centered around the location of the signal. The N_σ sensitivity increases when the window encapsulates the SR more precisely, reaching up to $Z_{\text{avg,max}} = 2.74$ with the 6×6 bins window. Thus, for this example, the sensitivity achieved is only slightly worse than the one achieved with the q_0^{L2} test, which exploits a full knowledge of the signal shape. The N_σ results presented hereafter are for the best-suited window (6×6 bins for all examples considered, unless specified otherwise).

In Figure 48, we show the Receiver Operating Characteristic (ROC) curves obtained from the PDFs of the different tests. The Area-Under-Curve (AUC) measured is approximately 1.0 for the q_0^{L1} test and 0.994 for the q_0^{L2} test. With an AUC of 0.973, the N_σ test is only 2.6% less sensitive than the q_0^{L1} test, and 2.0% less sensitive than the q_0^{L2} test. Finally, in Figure 49, we show Z_{avg} per test (estimated from the Asimov data), for increasing injected signal strength. Using the N_σ test statistic, the symmetric case (background only) can be separated from the asymmetric case at the level of 2σ if the signal that would have been measured assuming an ideal analysis (q_0^{L1}) is at the level of 3.5σ. This should also be compared to the 2.5σ separation that would have been obtained in the same case using the profile likelihood ratio test statistic that uses the two datasets to estimate the symmetric background and full knowledge of the signal shape (q_0^{L2}).

For clarity, we also consider a *flat* background template T with 10^4 entries in each bin, and a *flat rectangle* signal template S of size 6×6 bins, located at the center of T. Since the q_0^{L1} and q_0^{L2} are independent of the background and signal shapes, and only depend on the injected signal strength, their symmetry- and asymmetry-case PDF will remain unchanged. The PDF associated with the N_σ in the asymmetric case

[10] There is a trial factor for performing multiple tests, but as stated earlier, the goal is to identify interesting regions and not to compute a precise global p-value. That could be done with k-folding or other divide-and-test schemes, which we leave for future work to explore.

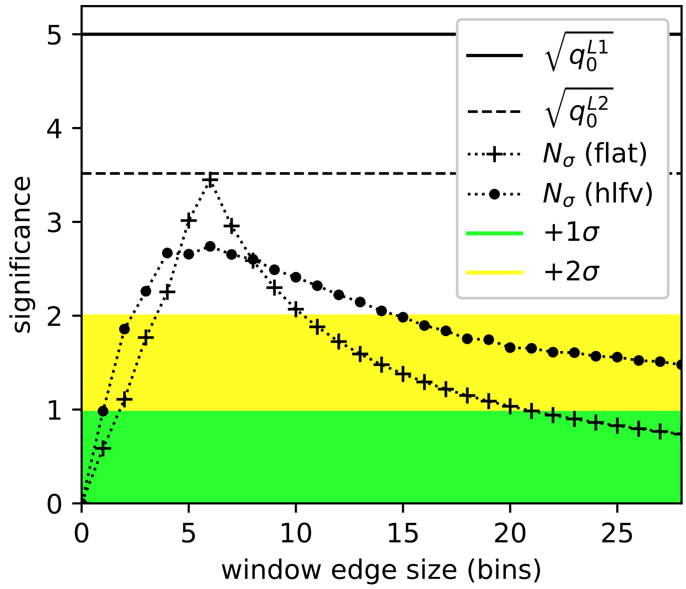


Figure 47: Significance measured from the Asimov data, with the N_σ test applied to increasing window sizes, and compared to the q_0^{L1} and q_0^{L2} significance. Results for the Higgs LFV example and the ideal (flat) scenario are shown, with injected signal strength corresponding to 5σ of q_0^{L1}. The green and yellow bands correspond to the 1σ and 2σ deviations from the symmetry (no signal) assumption, respectively.

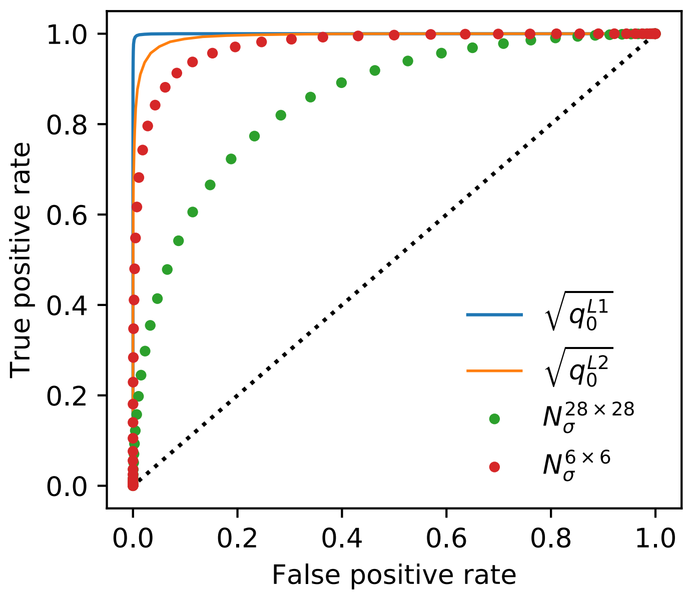


Figure 48: ROC curves comparing results of the N_σ, q_0^{L1}, and q_0^{L2} tests for the Higgs LFV example, with injected signal strength corresponding to 5σ of q_0^{L1}.

will change. As shown in Figures 47 and 49, in this simplified case, the N_σ sensitivity matches exactly the sensitivity of the q_0^{L2} test. This suggests that the loss of sensitivity of the generic N_σ test, compared to q_0^{L2}, is mainly due to shape variations of the background and the signal (in the optimal sub-region that is tested). But even in a realistic scenario like the Higgs LFV example, the sensitivity loss is reasonable (from $Z_{\text{avg}} = 3.53$ to 2.74), and the power achieved to identify regions with asymmetry, even though the N_σ test is generic, is significant. In terms of the ability to identify asymmetries, similar performance was obtained for all the other shapes of signal and background considered. Additional studies characterizing the N_σ test statistic are presented in section 5.3.

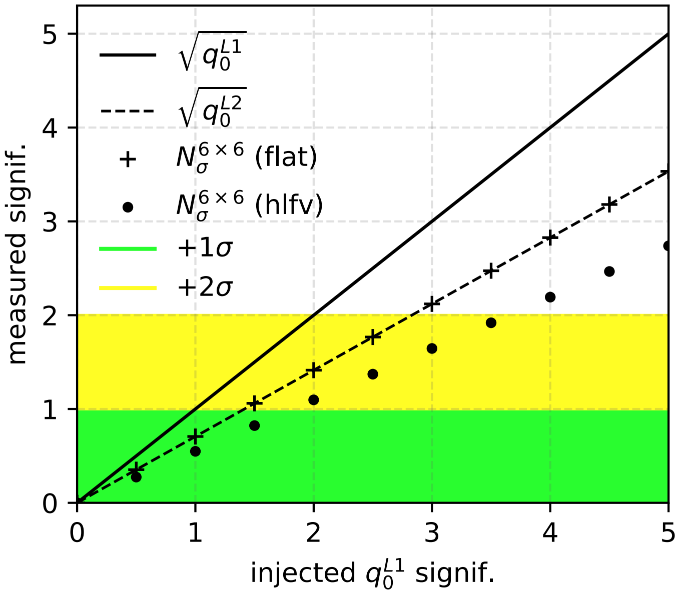


Figure 49: Significance measured from the Asimov data for increasing injected signal, comparing results of the N_σ, q_0^{L1}, and q_0^{L2} tests. Results for the Higgs LFV example and the ideal (flat) scenario are shown. The green and yellow bands correspond to the 1σ and 2σ deviations from the symmetry (no signal) assumption, respectively.

5.2.4 Neural Network approach

Machine learning-based anomaly detection methods constructed by comparing two datasets are categorized as weakly- or semi-supervised learning because both datasets are mostly background, and one will have more signal than the other. The dataset with the most signal potential is given a noisy label of one, and the other dataset is given a label of zero. A classifier trained to distinguish the two datasets can then automatically identify subtle differences between the datasets without explicitly setting up bins. Existing proposals construct the datasets from SR / sideband regions [110, 111, 116], from data versus simulation [113, 114, 121, 130], as well as other approaches [115, 117–120]. We propose to extend this methodology to symmetries.

The combination of machine learning and symmetry has received significant attention. For a given symmetry, one can construct machine learning methods that are invariant or covariant (in machine learning, this is called *equivariant*) under the action of that symmetry. For example, recent proposals have shown how to construct Lorentz covariant NNs [131–133]. Symmetries can also be used to build a learned representation of a dataset [134]. There have also been proposals to use machine learning methods to discover symmetries automatically in datasets [135–137]. In the context of BSM searches, [138, 139] recently described how to use a weakly supervised-like approach to test if a given symmetry is broken by applying the transformation to the input data. Our approach also starts by positing a symmetry, but we do not apply the symmetry transformation to each data point. Instead, we have two datasets that should be statistically identical in the presence of symmetry but could be different when BSM is present.

In the following, we demonstrate the concept of identifying asymmetries using a weakly supervised approach. Considering the $e\mu$ symmetry example discussed above, one of the datasets is the $e\mu$ dataset, and the other is the μe dataset. The same two-dimensional space described earlier is used for illustration; extending to higher dimensions is technically straightforward. A deep NN with three hidden layers and 50 nodes per layer is used for the classifier. Rectified Linear Units (ReLU) are used for all intermediate layers, and the output is passed through a sigmoid function. This network is implemented using Keras [84] and Tensorflow [85] using Adam [140] for optimization. We train for 20 epochs with a batch size of 200. None of these parameters were optimized. Figure 50 shows the NN's symmetry/asymmetry separation power as a function of the signal fraction injected into the μe dataset. The background-only band is computed via bootstrapping [141]. For each bootstrap, two datasets are created by drawing from the $e\mu$ and μe events with replacement. By mixing the two datasets, any asymmetry is removed.

There is no unique way to quantify the NN performance. An optimal test statistic by the Neyman-Pearson Lemma [127] is monotonically related to the likelihood ratio. [113, 114, 121, 142] show how to modify the loss function so that the average loss approximates the (log) likelihood ratio. Here, we find that in practice, the maximum NN score using the standard binary cross-entropy loss function is an effective statistic, which goes from 0.5 in the case of no signal and increases as more signal is injected. The background-only band in Figure 50 is computed via bootstrapping. Where the blue line and green/yellow bands cross indicate the approximate $1\sigma/2\sigma$ exclusion. The NN is able to automatically identify the presence of BSM for signal fractions that are a few per mil, corresponding to around 5σ significance calculated with the ideal q_0^{L1} test. Future explorations of this idea will understand the best way to set up the training, what statistics are most effective, and how to best extend to higher dimensions.

5.2.5 Discussion

With limited resources at hand and yet no conclusive indication of BSM physics found, we must try novel and complementary avenues for discovery. To overcome the limitations stemming from adapting the blind-analysis strategy, we propose developing the DDP. Similarly to [102–104, 143, 144], yet without relying on MC simulations, its principal objective is to allow scanning of as many regions of the observables space as possible and direct dedicated analyses towards the ones in which the data itself exhibits deviation from some fundamental and theoretically well-established property of the SM. Relative to regions in which the data agrees well with the SM predictions, the ones that exhibit deviations are promising for further investigations into BSM physics.

We propose developing the DDP based on symmetries of the SM and demonstrate its potential sensitivity using, as an example, the e/μ symmetry. Symmetries allow splitting the data into two mutually exclusive datasets, which, under the symmetry assumption, differ only by statistical fluctuations. Thus, asymmetry observed between the two datasets in any observable and at any sub-selection is potentially interesting and should be considered for further study.

While different algorithms can be developed to identify asymmetries, even the most simple one developed, the N_σ test statistic, already provides good sensitivity. It is compared to the sensitivity obtained with two likelihood-based test statistics; the first, q_0^{L1}, represents an ideal analysis in which both the signal and the symmetric contribution from the SM processes is perfectly known. The second, q_0^{L2}, represents the expected

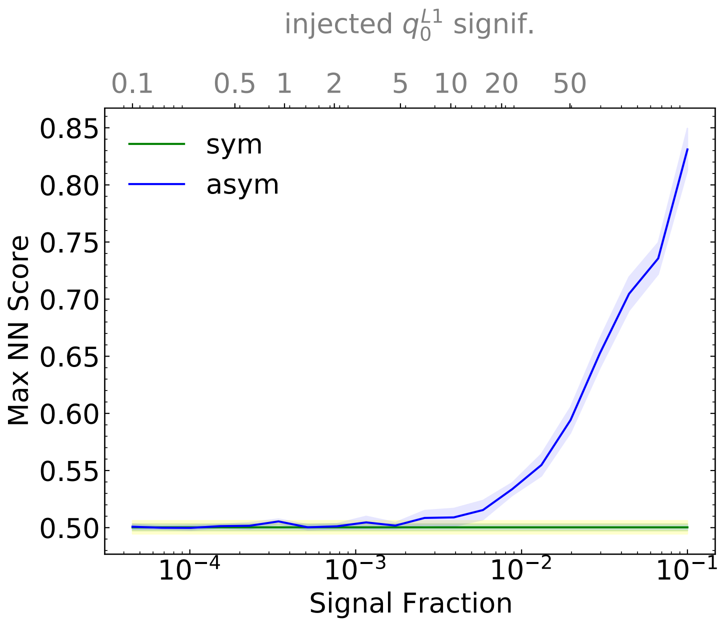


Figure 50: The maximum NN score from training a classifier to distinguish the $e\mu$ from μe datasets with (asym) and without (sym) a BSM contribution. The green (yellow) and blue bands represent (twice) the standard deviation over 10 bootstrap datasets. The separation power is shown as a function of the injected signal fraction (bottom scale) and the corresponding significance calculated with the ideal q_0^{L1} test. Note that these results are not directly comparable to the binned DDP because it is not possible to ignore signal statistical uncertainties.

sensitivity of a traditional blind analysis search for a predefined signal that employs a symmetry-based background estimation, such as the Higgs LFV search described in section 4.

Compared to the sensitivity obtained in an ideal analysis, the separation power between the symmetric case and an asymmetry at the level of 5σ is less than 3% lower in terms of the area under the ROC curve, and a separation at the level of 2σ is achieved for 3.5σ signal injected. Compared to traditional symmetry-based analysis, the separation power between the symmetric case and an asymmetry at the level of 3.5σ is less than 2% lower in terms of the area under the ROC curve, and a separation at the level of 2σ achieved using the N_σ test is only slightly degraded relative to the 2.5σ obtained with the q_0^{L2} test. The results quoted are when applying the N_σ test in the best-suited window for the examples considered. The ability to find this optimal window demonstrates the strength of the DDP. Since the test is rapid, a large number of n-dimensional histograms and windows within can be tested efficiently. This could permit scanning the data systematically in search of asymmetries.

We have shown that weakly-supervised NNs can also be used to identify asymmetries between two datasets. This paves the way towards NN-based DDP.

We emphasize that traditional blind analyses are expected to be the most sensitive for any predefined signal. Nonetheless, it is impossible to conduct a dedicated search in any possible final state and at any possible event selection. Moreover, not all potential signals can be thought of. Thus, the DDP could significantly expand our discovery reach.

5.3 Characterization of the N_σ test statistic

This section presents additional studies characterizing the N_σ test statistic defined in (59) and used in section 5.2.3 to quantify asymmetries between two measurements. The study's framework, data analysis methods (including benchmark tests), data preparation, and obtention of results are already described in sections 5.2.1-5.2.3.

5.3.1 Varying backgrounds and signals

As discussed in section 5.2.3, the N_σ sensitivity for a given signal can vary greatly on how well or poorly the signal is encapsulated in the sub-region tested (see Figure 47). To further emphasize this property, we consider the case of a flat background template T with 10^4 entries in each bin (28×28) and a signal template S, which is only present in a single bin and injected with a signal strength corresponding to 5σ of q_0^{L1}. In Figure 51, we compare the ROC curves obtained from different tests: q_0^{L1}, q_0^{L2}, N_σ applied to all the bins, and N_σ applied only to the bin where the signal is present.

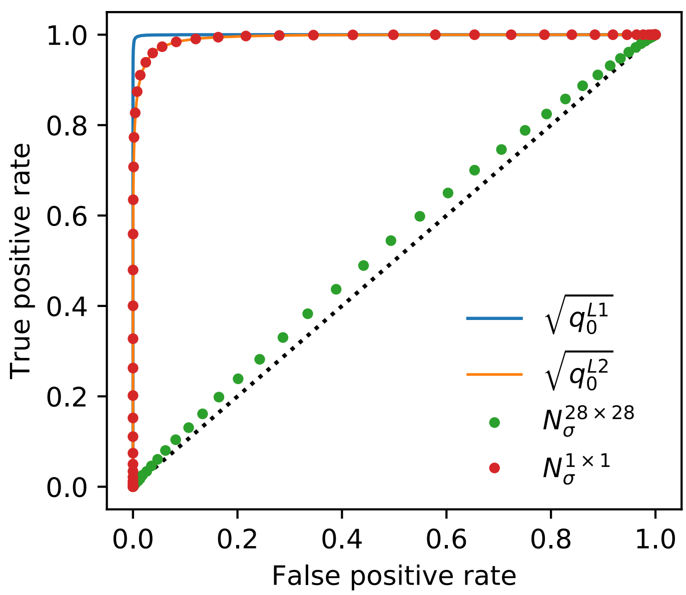


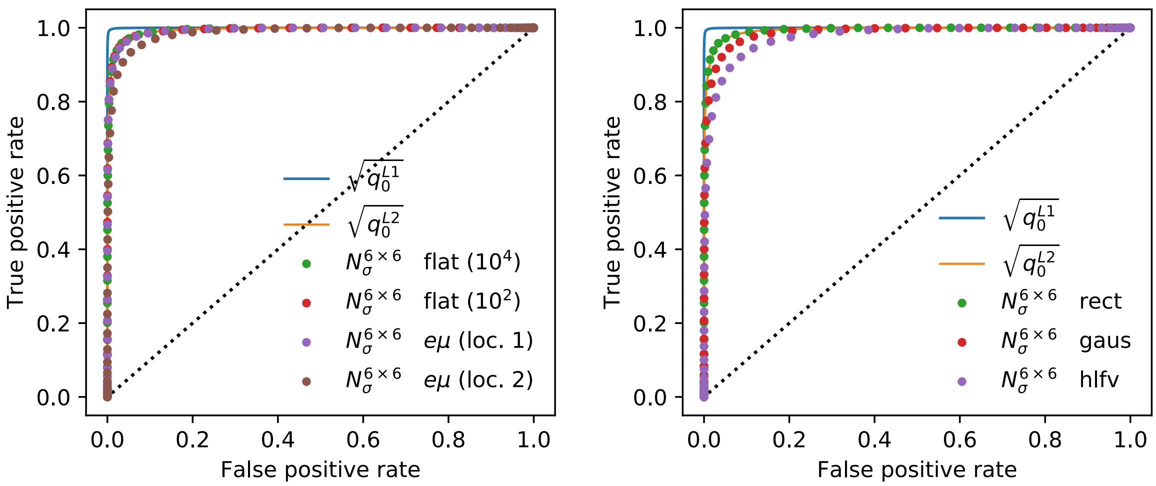
Figure 51: ROC curves comparing results of the N_σ, q_0^{L1}, and q_0^{L2} tests for the example of a flat background (10^4 entries/bin) and a single-bin signal with injected signal strength corresponding to 5σ of q_0^{L1}. Results when N_σ is applied to all bins, or the single bin with signal, are compared.

The q_0^{L1} and q_0^{L2} sensitivities do not depend on the properties of the background or signal templates considered, since the signal is injected with a signal strength corresponding to a fixed significance of q_0^{L1} or q_0^{L2}. So their respective ROC curves in Figure 51 are the same as those shown, for example, in Figure 48. When the N_σ test is applied to all the bins, it has no sensitivity to detect this (single bin) signal; its measured average significance is $Z_{\text{avg}} \approx 0.1$. But when it is applied only to the bin that includes the signal, its sensitivity ($Z_{\text{avg}} \approx 3.5$) is equivalent to that of q_0^{L2} – which uses the same background assumptions but full knowledge of the signal. This one-bin example is analogous to a counting experiment. In this case, the N_σ and q_0^{L2} sensitivities are equivalent. This was already observed in section 5.2.3 for the similar case of a flat background and rectangle signal, where the N_σ test was only applied to the bins, that include signal contributions.

We also showed in section 5.2.3 that in the case of the Higgs LFV example, even in an optimized sub-region, the N_σ sensitivity is degraded compared to that of q_0^{L2} (see Figure 47). We attribute this to variations in the background or signal. To further characterize the test's performance, we consider different background templates and signal templates.

In Figure 52(a), we compare ROC curves obtained using the rectangle 6×6 signal with variable background templates: *flat* with 100 entries per bin, *flat* with 10^4 entries per bin, the $e\mu$ background from the Higgs LFV example (see Figure 45) with signal centered around coordinates (130, 110) (location 1), and the $e\mu$ background with signal centered around coordinates (60, 40) (location 2). In all cases, the injected signal amounts to 5σ of q_0^{L1}. We find that the N_σ test performs equally well ($Z_{\text{avg}} \approx 3.5$) for a flat background with 100 or 10^4 entries per bin. It is slightly penalized for the $e\mu$ background, especially at location 2 ($Z_{\text{avg}} \approx 3.1$), due to larger differences in the background statistics from bin to bin.

In Figure 52(b), we compare ROC curves obtained using the usual flat background (10^4 entries per bin) but with different signal templates: *rectangular* (size 6×6), *gaussian* (std 2 bins), and Higgs LFV (see Figure 45). In all cases, the injected signal amounts to 5σ of q_0^{L1}. We find that, compared to the ideal rectangle signal case ($Z_{\text{avg}} \approx 3.5$), the N_σ sensitivity is slightly penalized for the gaussian signal ($Z_{\text{avg}} \approx 3.2$), and even more for the Higgs LFV signal ($Z_{\text{avg}} \approx 2.8$) since it is spread out on more bins, but the obtained sensitivity is still close to that of q_0^{L2}.



(a) (b)

Figure 52: ROC curves obtained with variable background (a) or signal (b) templates. Otherwise, using the flat 10^4 background template and rectangle 6×6 signal template. The injected signal strength corresponds to 5σ of q_0^{L1} for all cases.

All the signals considered up to now are bump-like signals, illustrating searches for resonances. But other types of BSM signatures can be considered, for example, in non-resonant or indirect searches (see, e.g., [145, 146]). We also consider a slope-like signal. It is uniform along the y-axis and linearly rising along the x-axis (still normalized to unity). In Figure 53, we show N_σ scores with the Asimov data obtained when the test is performed on windows of size $x \times 28$, which start from the right-most edge of the matrix and where x is varied from 0-28. With an injected signal corresponding to 5σ of q_0^{L1}, the best N_σ score for the case of the flat background (10^4 entries per bin) is $Z_{\text{avg,max}} = 3.33$ with a window of size 18×28, and for the case of the $e\mu$ background $Z_{\text{avg,max}} = 2.95$ with a window of size 17×28. These scores are again close to that of q_0^{L2} ($Z_{\text{avg}} = 3.53$).

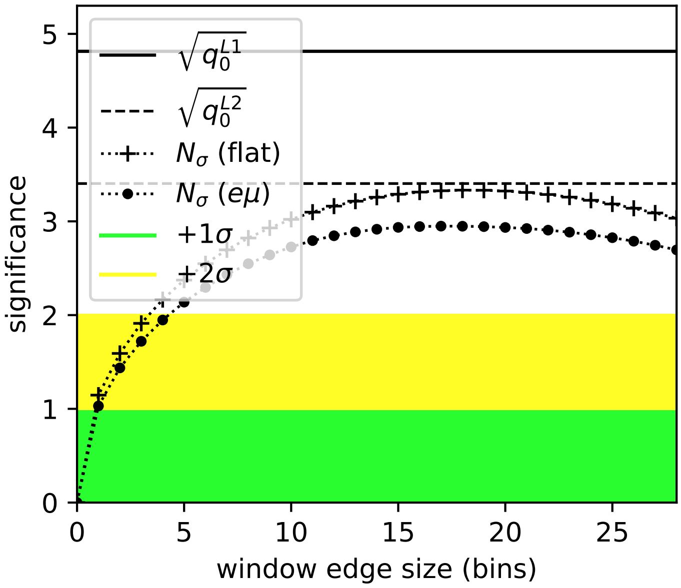


Figure 53: Significance measured from the Asimov data, with the N_σ test applied to windows of size $x \times 28$ with increasing x values, and compared to the q_0^{L1} and q_0^{L2} significance. The injected signal is the slope-like signal described in the main text. Results for the flat and $e\mu$ backgrounds are shown, with injected signal strength corresponding to 5σ of q_0^{L1}. The green and yellow bands correspond to the 1σ and 2σ deviations from the symmetry (no signal) assumption.

Table 19 summarizes the results comparing the q_0^{L1}, q_0^{L2}, and N_σ sensitivities obtained from the Asimov datasets for the various background and signal templates considered. The signals are injected with a significance corresponding to 5σ of q_0^{L1} (q_0^{L2}) in the first half (second half) of the table. For the $e\mu$ background with a rectangle or Gaussian signals, results for both signal locations already considered above are listed (loc1: (130, 110), loc2: (60, 40)). We also indicate, for each case, the corresponding μ_{inj}^s – which is the total number of injected signal events – and the signal fraction w.r.t. the number of background events. Both the N_σ scores and the signal fractions are measured in the optimal window in which the N_σ score is maximized (6×6 for all cases except for the slope signal where it is 18×28 (17×28) if the background is flat ($e\mu$)).

background	signal	μ^s_{inj}	signal fraction	$Z(q_0^{L1})$	$Z(q_0^{L2})$	$Z_{\text{max}}(N_\sigma)$
flat (10^4)	rect	3004.2	0.83	5.00	3.53	3.53
	gaus	3550.5	0.75	5.00	3.53	3.16
	hlfv	4281.1	0.66	5.00	3.53	2.80
	slope	12018.3	0.19	5.00	3.53	3.33
flat (10^2)	rect	304.1	8.45	5.00	3.51	3.51
	gaus	360.0	7.58	5.00	3.51	3.15
	hlfv	434.2	6.71	5.00	3.51	2.79
	slope	1206.0	1.88	5.00	3.53	3.33
$e\mu$	rect (loc 1)	268.1	8.84	5.00	3.51	3.47
	rect (loc 2)	877.2	0.68	5.00	3.52	3.10
	gaus (loc 1)	315.8	7.89	5.00	3.50	3.11
	gaus (loc 2)	1069.5	0.63	5.00	3.52	2.85
	hlfv	1011.7	2.76	5.00	3.52	2.74
	slope	1195.7	0.23	5.00	3.52	2.95
flat (10^4)	rect	4255.1	1.18	7.08	5.00	5.00
	gaus	5029.9	1.06	7.08	5.00	4.48
	hlfv	6065.4	0.94	7.08	5.00	3.96
	slope	17003.9	0.26	7.07	5.00	4.71
flat (10^2)	rect	436.8	12.13	7.14	5.00	5.00
	gaus	517.9	10.90	7.15	5.00	4.48
	hlfv	625.1	9.66	7.15	5.00	3.97
	slope	1713.0	2.67	7.09	5.00	4.72
$e\mu$	rect (loc 1)	386.4	12.74	7.16	5.00	4.95
	rect (loc 2)	1252.4	0.98	7.11	5.00	4.41
	gaus (loc 1)	455.9	11.39	7.16	5.00	4.44
	gaus (loc 2)	1524.6	0.90	7.11	5.00	4.04
	hlfv	1440.4	3.92	7.10	5.00	3.89
	slope	1704.3	0.33	7.11	5.00	4.19

Table 19: Comparison of q_0^{L1}, q_0^{L2}, and N_σ significances obtained from the Asimov datasets for the various background and signal templates considered. The signals are injected with a significance corresponding to 5σ of q_0^{L1} (q_0^{L2}) in the first half (second half) of the table. The injected μ^s_{inj} and signal fractions are also listed (see details in the main body's text).

Although the injected number of signal events and signal fractions can vary significantly from one case to the other, the significance obtained with the generic N_σ test is consistently between 75-100% of the signal-specific q_0^{L2} significance for all examples considered. We observe some small variations of the q_0^{L2} (q_0^{L1}) score for a fixed q_0^{L1} (q_0^{L2}) injected significance. We attribute this to the approximate validity of the asymptotic assumption used to measure these significances depending on the statistics available. Since this effect is small for all examples considered, we neglected it in all the descriptions above.

5.3.2 Validity in low statistics

All the examples considered up to now were constructed with at least 25 entries in each bin of the background template T in order to avoid regions with low statistics. For the case of very low statistics, the PDF of the symmetric scenario (where the two compared matrices are indeed drawn from the same underlying distribution) is no longer consistent with the standard Gaussian distribution. We illustrate this in Figure 54(a) by comparing PDFs of N_σ scores obtained from 50K pairs of background-only matrices, where the background templates T considered are flat backgrounds with 0.1, 1, and 10 entries per bin. In this example, the N_σ test is applied in windows of size 6×6. When T has at least 10 entries per bin, the $N_\sigma^{6\times6}$ PDF is consistent with the standard Gaussian (the dotted black line in the plot). But for lower entries per bin in T, the PDF is narrower, albeit still Gaussian-like.

We attribute the distortion of the N_σ PDF from the standard Gaussian (at least in part) to the discreetness of the Poisson statistics, which govern the yields in the compared matrices. As such, the statistics in T are not the only parameters that affect the shape of the N_σ PDF; the size of the tested window should also be considered. Indeed, the more bins are tested together, the more combinatorics and the more the different N_σ scores obtained are continuous-like. In Figure 54(b), we compare N_σ PDFs applied to windows of various sizes (1×1, 3×3, and 4×4) when T has only 2 entries per bin. We see that in the single-bin case (1×1), the obtained PDF is narrow and doesn't resemble a Gaussian due to the discreetness of the obtained scores. Whereas for 3×3 or 4×4 windows, the N_σ PDF is consistent with the standard Gaussian.

In Figure 54(c) (Figure 54(d)), we compare N_σ PDFs when the tested windows are of size 2×2 (1×1) for flat background templates T with variable entries per bin. In this case, the N_σ PDF is consistent with the standard Gaussian for a minimum of 3 (15) entries per bin of T.

We summarize the minimal empirical conditions for the validity of the standard Gaussian approximation of the N_σ PDF in the symmetric (background-only) case:

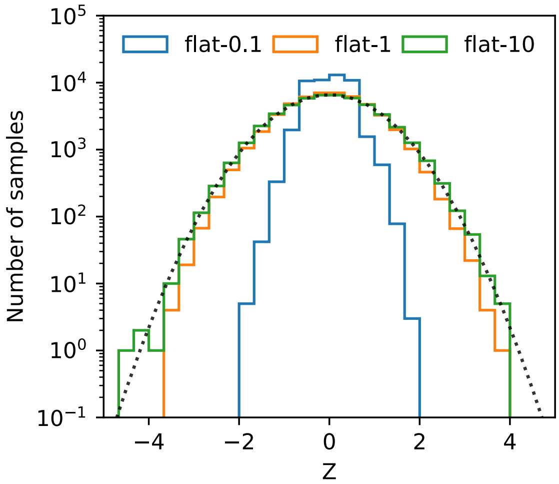
- For a single-bin N_σ test, the bins compared should include at least 15 entries
- For a 2×2 N_σ test, the bins compared should include at least 3 entries
- For a 3×3 N_σ test or larger, the bins compared should include at least 2 entries

When confronted with a scenario that doesn't satisfy these conditions, one can try rebinning the compared matrices to include more events in each bin or determining the expected symmetric-case PDF in more detail.

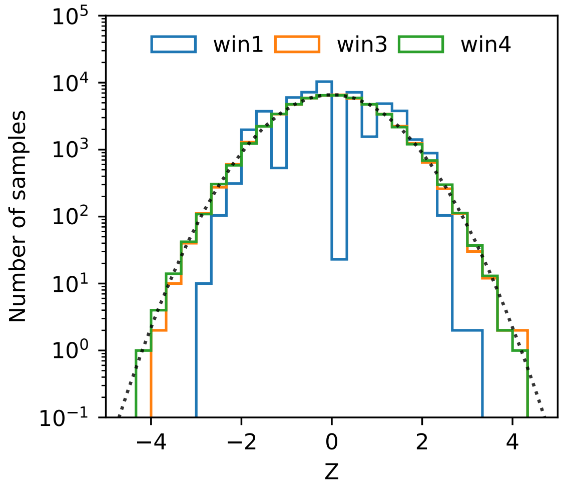
Still, whatever the scenario, a larger N_σ score indicates a larger asymmetry between the two measurements considered. Furthermore, since the symmetric-case PDF becomes narrower at low statistics, the obtained score under the (invalid) normal approximation will underestimate the actual significance of the discrepancy, which is conservative. Since our goal is to identify regions that exhibit asymmetries – to be studied using more traditional data-analysis methods – this is enough in most applications to be considered.

5.3.3 Known background underlying distribution

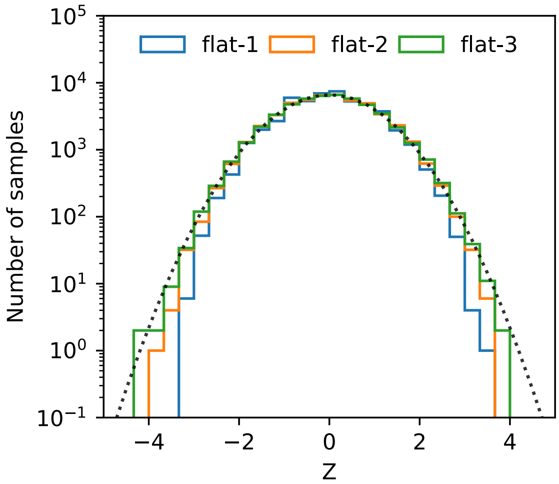
In the previous discussions, we compared results obtained with the generic N_σ test to those obtained from two distinct signal-specific likelihood-based tests, q_0^{L1}, which knows the exact underlying background distribution, and q_0^{L2}, which – similarly to N_σ – makes no prior assumptions on the underlying background distribution. For completeness, one can also consider a generic test similar to N_σ but which compares a



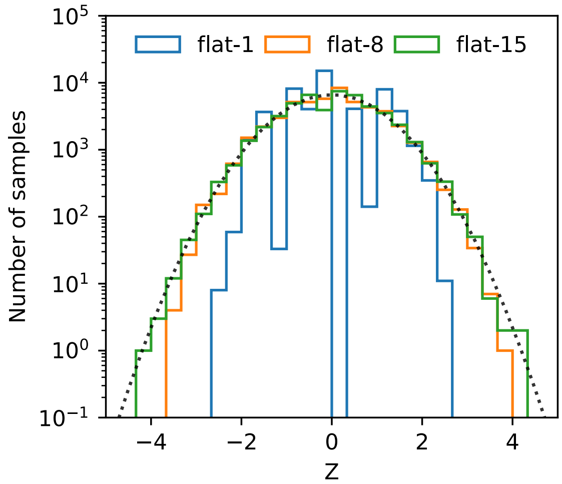
(a) $N_\sigma^{6\times6}$ - variable flat bkgs



(b) N_σ variable windows - flat bkg (2 entries/bin)



(c) $N_\sigma^{2\times2}$ - variable flat bkgs



(d) $N_\sigma^{1\times1}$ - variable flat bkgs

Figure 54: N_σ PDFs obtained from 50K pairs of background-only matrices. The background template T is always flat with variable entries per bin in (a), (c), and (d), and with 2 entries per bin in (b). The N_σ test is applied to windows of variable sizes as described in the captions of (a), (c), and (d), and in the legend of (b). The PDFs are compared to the standard Gaussian distribution, displayed as the dotted black line in the plots.

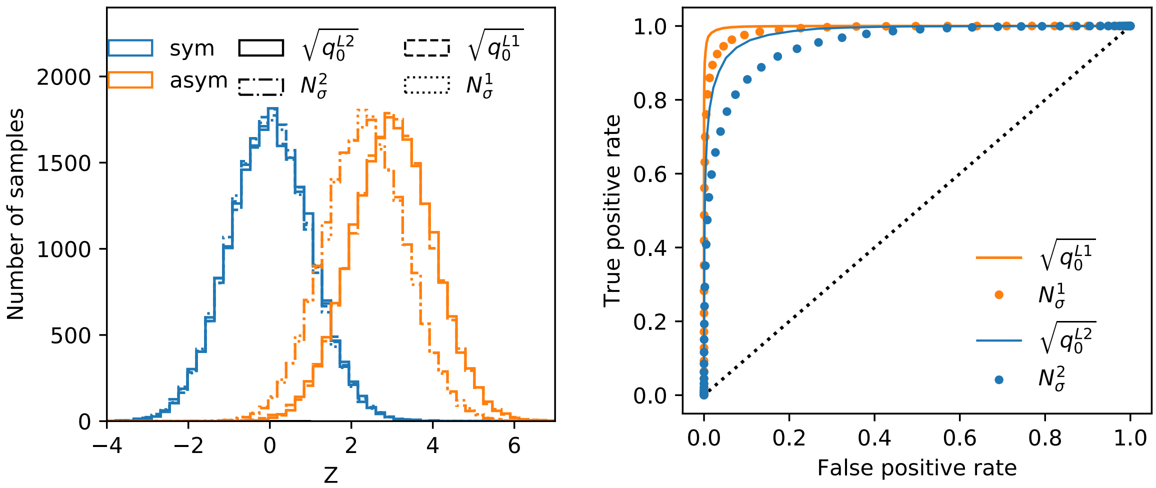
measurement B to a known background distribution T. We denote this test as N_σ^1 and define it in (70); and in the following, we denote the N_σ test between two measurements defined in (60) as N_σ^2.

$$N_\sigma^1(B,T) = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} \frac{B_i - T_i}{\sqrt{T_i}} \tag{70}$$

Although such a test wouldn't be applicable in the comparison of two measurements, which is our goal, and although methods already exist for comparing a measurement to a known background distribution, the main motivation to discuss this test is that the results obtained with N_σ^1 compared to q_0^{L1} are equivalent to the results obtained with N_σ^2 compared to q_0^{L2}. This means that all the results presented up to now are also applicable to this scenario when switching N_σ^2 and q_0^{L2} by N_σ^1 and q_0^{L1}.

This is illustrated in Figure 55(a) in the case of the Higgs LFV example, where we compare the PDFs of (N_σ^1, q_0^{L1}) for an injected signal corresponding to 3σ of q_0^{L1} and of (N_σ^2, q_0^{L2}) for an injected signal corresponding to 3σ of q_0^{L2}. For both the symmetric and asymmetric cases, the obtained N_σ PDFs are in agreement. These are obtained when applying the N_σ^1 and N_σ^2 tests in the same optimal 6×6 window centered around the signal. We also show that the q_0^{L1} and q_0^{L2} PDFs agree as well.

To further clarify the relationship between the different tests considered, we compare in Figure 55(b) the ROC curves per test for the Higgs LFV example, obtained with an injected signal corresponding to 3σ significance of q_0^{L2} for all tests. As expected, q_0^{L1} ($Z_{\text{avg}} \approx 4.2$) is more sensitive than q_0^{L2} ($Z_{\text{avg}} \approx 3.0$). Similarly, N_σ^1 ($Z_{\text{avg}} \approx 3.3$) is more sensitive than N_σ^2 ($Z_{\text{avg}} \approx 2.3$).



(a)

(b)

Figure 55: Comparison of results with (N_σ^1, q_0^{L1}) vs. (N_σ^2, q_0^{L2}) for the Higgs LFV example. In (a), PDFs are shown when $Z_{\text{inj}} = 3\sigma$ of q_0^{L1} for (N_σ^1, q_0^{L1}) and $Z_{\text{inj}} = 3\sigma$ of q_0^{L2} for (N_σ^2, q_0^{L2}). In (b), ROC curves are shown when $Z_{\text{inj}} = 3\sigma$ of q_0^{L2} for all tests. The N_σ^1 and N_σ^2 scores are obtained in the optimal 6×6 window centered around the signal

5.4 Efficiency correction in standalone MC samples

This section describes efforts towards generating MC simulated samples of pp collisions at $\sqrt{s} = 13$ TeV. The goal is to provide test data for proof-of-concept and the development of generic data-directed searches using the e/μ-symmetry method in the context of the proposed symmetry-based DDP. Events were generated using MadGraph5_aMC@NLO 2.6.4 [128] and Pythia 8.2 [61]. Delphes 3 [129] was used to simulate the effects and measurements of the ATLAS detector. The generation procedure implemented is detailed in section 5.4.1.

Using this data, we probe the expected e/μ symmetry from SM processes by comparing the $e\mu$ vs. μe or ee vs. $\mu\mu$ datasets. This comparison, before and after the data is processed through the detector simulation, is presented in section 5.4.2 and section 5.4.3, respectively. In the case of the reconstructed data, a correction for the efficiencies introduced in Delphes is applied to restore the symmetry. In section 5.4.4, we summarize the results.

5.4.1 Samples and selection

The different SM processes simulated, which are the main background contributions to dilepton final states, are $Z/\gamma^* \to \ell\ell$, $Z/\gamma^* \to \tau\tau$, WW, Wt, $t\bar{t}$, (ggF + VBF)$H \to WW$, and (ggF + VBF)$H \to \tau\tau$. We also produced signal samples for LFV decays of the Higgs (ggF + VBF) and Z bosons to τe and $\tau\mu$ final states. The SM background and Higgs LFV signal templates used in the study presented in section 5.2 were derived using these simulated samples (see Figure 45).

All the samples were generated at LO, with two leptons per event (electron, muon, or tau) and up to two hard jets, except for VBF samples which have exactly two. We worked in the five-flavor scheme, used the NNPDF3.0 PDF set [147], and applied the MLM scheme [148] with kT jet matching following the recommended settings. MadSpin [149] was used to decay the W boson and the top quark, while the τ decays (left unspecified) were handled by Pythia. In Delphes, we adjusted the default ATLAS card such that the electron and muon efficiencies resembled those in ATLAS. In Table 20, we show the different cross-sections and BRs used to estimate the cross-sections of the generated processes. Table 21 compares the calculated cross-sections per generated sample to those estimated at LO by MadGraph.

Sample	Value	Source
	Cross-Sections [pb]	
tt	826.4	https://cds.cern.ch/record/2686255
Wt	94	https://link.springer.com/article/10.1007%2FJHEP01%282018%29063
WW	115.3	http://inspirehep.net/record/1469339/
(gg)H	43.62	CERNYellowReportPageAt1314TeV2014
(vbf)H	3.727	CERNYellowReportPageAt1314TeV2014
$Z/\gamma^* \to ee$	2087	mg5_aMC@NLO
$Z/\gamma^* \to \mu\mu$	2087	mg5_aMC@NLO
$Z/\gamma^* \to \tau\tau$	2101	mg5_aMC@NLO
Z	51450	mg5_aMC@NLO
	BRs	
$W \to e$	0.1071	pdg
$W \to \mu$	0.1063	pdg
$W \to \tau$	0.1138	pdg
$t \to q$	0.665	pdg
$t \to \ell$	0.111666667	$(1 - B(t \to q))/3$
$H \to \tau\tau$	0.06192	CERNYellowReportPageBR
$H \to WW$	0.2219	CERNYellowReportPageBR
$Z \to \tau\ell$	0.00001	fixed
$H \to \tau\ell$	0.01	fixed

Table 20: Cross-sections and BRs used to estimate the luminosity generated per simulated sample.

Sample	LO XSec [pb]	XSec [pb]	kFactor
$WW \to ee$	1.037	1.3225	1.276
$WW \to e^-\mu^+$	1.039	1.3127	1.263
$WW \to e^+\mu^-$	1.039	1.3127	1.263
$WW \to \mu\mu$	1.040	1.3029	1.252
$WW \to \tau^-\ell^+$	2.074	2.8001	1.350
$WW \to \tau^+\ell^-$	2.079	2.8001	1.347
$WW \to \tau\tau$	1.034	1.4932	1.444
$Wt \to ee$	0.703	1.1242	1.600
$Wt \to e^-\mu^+$	0.702	1.1200	1.596
$Wt \to e^+\mu^-$	0.701	1.1200	1.597
$Wt \to \mu\mu$	0.701	1.1158	1.593
$Wt \to \tau^-\ell^+$	1.400	2.3145	1.654
$Wt \to \tau^+\ell^-$	1.401	2.3145	1.652
$Wt \to \tau\tau$	0.699	1.1945	1.708
$Z/\gamma^* \to ee$	1057.962	2087.0	1.973
$Z/\gamma^* \to \mu\mu$	1057.993	2087.0	1.973
$Z \to \tau e$	0.241	0.5145	2.138
$Z \to \tau\mu$	0.240	0.5145	2.140
$Z/\gamma^* \to \tau\tau$	1050.212	2101.0	2.001
$tt \to ee$	6.188	10.3047	1.665
$tt \to e^-\mu^+$	6.188	10.3047	1.665
$tt \to e^+\mu^-$	6.188	10.3047	1.665
$tt \to \mu\mu$	6.192	10.3047	1.664
$tt \to \tau^-\ell^+$	12.369	20.6095	1.666
$tt \to \tau^+\ell^-$	12.371	20.6095	1.666
$tt \to \tau\tau$	6.176	10.3047	1.668
(gg) $HWW \to \ell\ell$	0.113	0.4408	3.886
(gg) $H \to \tau e$	0.132	0.4362	3.309
(gg) $H \to \tau\mu$	0.132	0.4362	3.306
(gg) $H \to \tau\tau$	0.708	2.7010	3.817
(vbf)$HWW \to \ell\ell$	0.025	0.0377	1.522
(vbf)$H \to \tau e$	0.029	0.0373	1.293
(vbf)$H \to \tau\mu$	0.029	0.0373	1.293

Table 21: List of all the simulated samples and their estimated cross-sections.

For the studies described below, we use events from the simulated SM processes corresponding to 20 fb^{-1} of integrated luminosity while ensuring that the mc weights per event are equal to unity.

We consider two sets of data, the *truth* data, which consists of the generated events after showering (initial- and final-state parton and lepton radiation), and the *reconstructed* data, which consists of the showered events after being processed through the detector simulation. The different detection modules implemented in Delphes which affect our study the most are:

- Lepton track Reco
- Lepton Reco
- Lepton Iso
- Lepton transverse momentum smearing

The lepton and lepton track Reco are implemented through efficiency maps binned in the lepton's p_T and η to simulate ATLAS measurements. The lepton Iso constraint is implemented by summing the p_T of the objects surrounding the lepton track within a cone and dividing the result with the lepton's p_T. If this ratio is above a certain threshold, the lepton is non-isolated. The p_T smearing is to simulate the detector's resolution.

For both truth and reconstructed datasets, we select events containing two opposite-sign leptons with $p_T > 10$ GeV and $|\eta| < 2.5$. The selected events are further split into the ee, $e\mu$, μe or $\mu\mu$ datasets depending on the flavor of the selected leptons and their p_T ordering.

5.4.2 Symmetry in truth data

As previously described, truth data refers to the generated events after showering. In contrast with the reconstructed data processed through the detector simulation, these datasets are not comparable to measured data. Still, it is reasonable to verify the expected SM e/μ symmetry based on them as a first step. In this comparison, no correction (efficiency or other) is applied.

A comparison of ee vs. $\mu\mu$ and $e\mu$ vs. μe for different kinematic distributions of the truth datasets is shown in Figure 56 (more distributions are displayed in appendix D.1, Figure 113). In the ee vs. $\mu\mu$ case, we find an asymmetry that is not very large on average (less than 2%) but presents singular trends around the Z peak in the $m_{\ell\ell}$ distributions and the leptons' p_T distributions. We identified that these features originate from the lepton energy loss via bremsstrahlung and FSR, implemented in Pythia, which is much more pronounced for electrons than muons. As a result, the m_{ee} peak is shifted to the left compared to the $m_{\mu\mu}$ one, leading to the asymmetric ratio profile displayed in the plot. This also affects the $e\mu$ vs. μe comparison to some degree. Indeed, although both datasets contain an electron and a muon, this energy loss varies with the lepton's p_T. As a result, we observe slightly increasing slopes towards the higher values in the ratio of the $m_{\ell\ell}$ and p_T distributions. Since this effect is corrected in ATLAS during electron reconstruction [71], we didn't consider it any further. Instead, we verified that deactivating "lepton showering" in Pythia leads to symmetric truth datasets, as displayed in Figure 57 (more distributions are shown in appendix D.1, Figure 113).

5.4.3 Symmetry in reconstructed data

The first step towards probing the symmetry within the reconstructed data is to measure the efficiencies introduced by the detector simulation. Since we have access to the truth data, the lepton efficiencies are measured directly from the ratio of reconstructed over truth distributions.

Following our experience using ATLAS data, our first attempt was to measure separate efficiency maps for electrons and muons and for leading and subleading leptons, parameterized in the lepton's p_T and η. But this method led to significant residual asymmetries related to how the Iso constraint is applied in Delphes. We found this becomes minor when using dilepton efficiencies instead. Such an effect hasn't been observed within ATLAS MC data. Therefore, the efficiencies that we use for the efficiency correction are dilepton efficiencies, with separate maps for the ee, $\mu\mu$, $e\mu$ and μe datasets, parameterized as a function of $p_T^{\ell_0}$, η^{ℓ_0}, $p_T^{\ell_1}$, and η^{ℓ_1}. The efficiency correction is applied event-by-event following the description in section 3.3.3, using the efficiency-ratio CFs.

The ee vs. $\mu\mu$ and $e\mu$ vs. μe reconstructed distributions after efficiency correction show very similar results than with the truth distributions, as is shown in Figure 58 (more distributions are displayed in appendix D.1, Figure 115). The same asymmetry caused by the energy loss via bremsstrahlung is present. An additional asymmetry can be seen around the Z peak in the ee vs. $\mu\mu$ comparison due to the smearing applied in Delphes to simulate lepton momentum resolution. In contrast, the same comparison using the datasets with "lepton showering" deactivated leads to more symmetric distributions, as is shown in Figure 59 (more distributions are displayed in appendix D.1, Figure 115). Still, the asymmetry due to momentum resolution is unaffected. A possible method to account for this effect could be applying the electron p_T smearing to the muons and vice-versa, but this hasn't been attempted.

5.4.4 Discussion

In this study, we reproduced all the steps required to obtain efficiency-corrected datasets within our simulation framework: event generation, detection simulation, measurement of efficiencies, and efficiency correction. Such datasets can be used in a symmetry-based DDP implementation.

Both the $e\mu$ vs. μe and ee vs. $\mu\mu$ comparisons are inspected, demonstrating that the e/μ-symmetry method can be applied to generic searches in different selections.

We identified two sources of asymmetries not efficiency-related: lepton energy loss via bremsstrahlung and lepton momentum resolution. Depending on the selection, these could also affect the symmetry between measured datasets and should be corrected to permit a meaningful DDP implementation. To this end, in the following section, we present a generic method to account for any asymmetry provided it is accurately modeled within the MC samples.

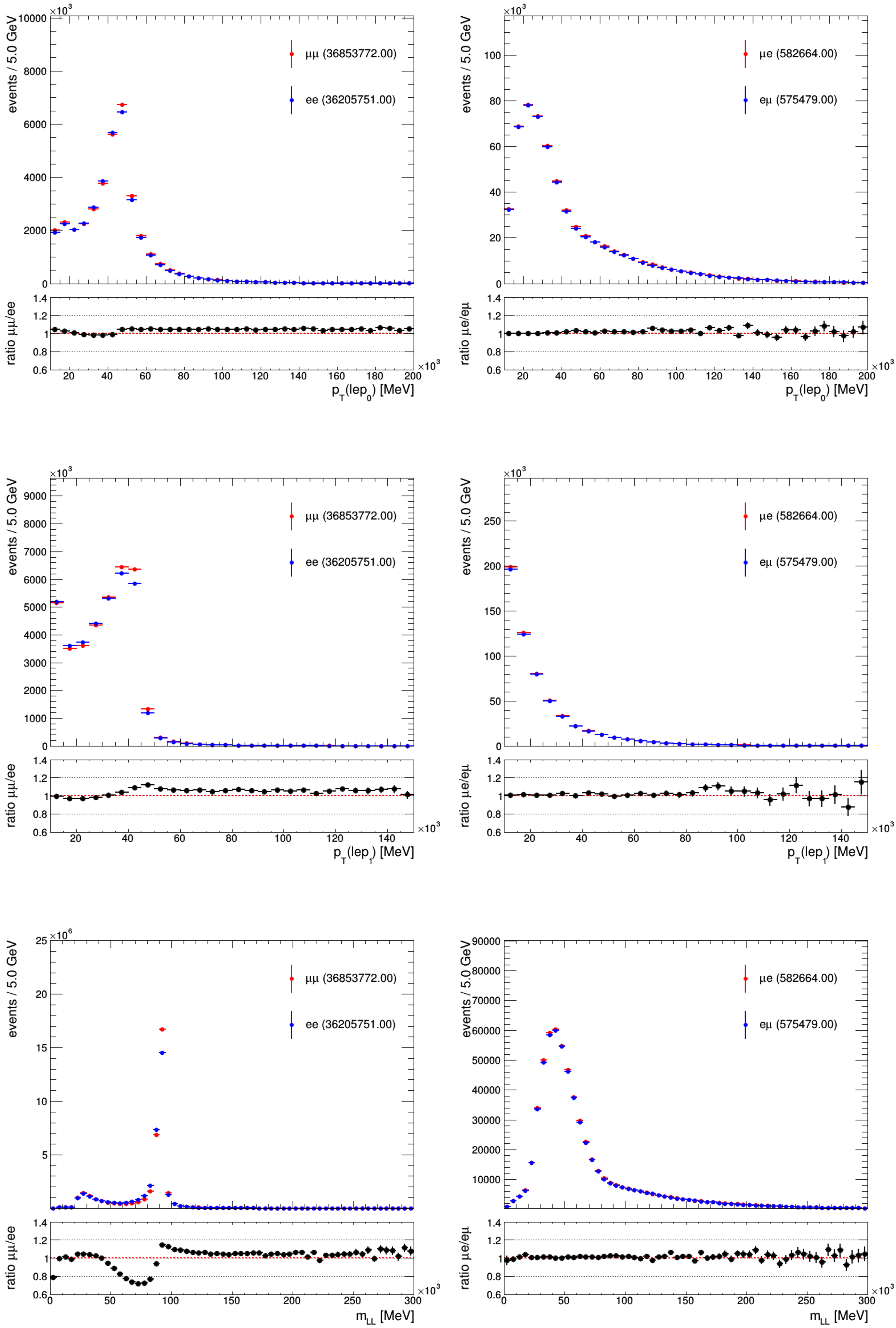


Figure 56: Comparison of ee vs. $\mu\mu$ (left) and $e\mu$ vs. μe (right) truth datasets (more distributions are displayed in Figure 113).

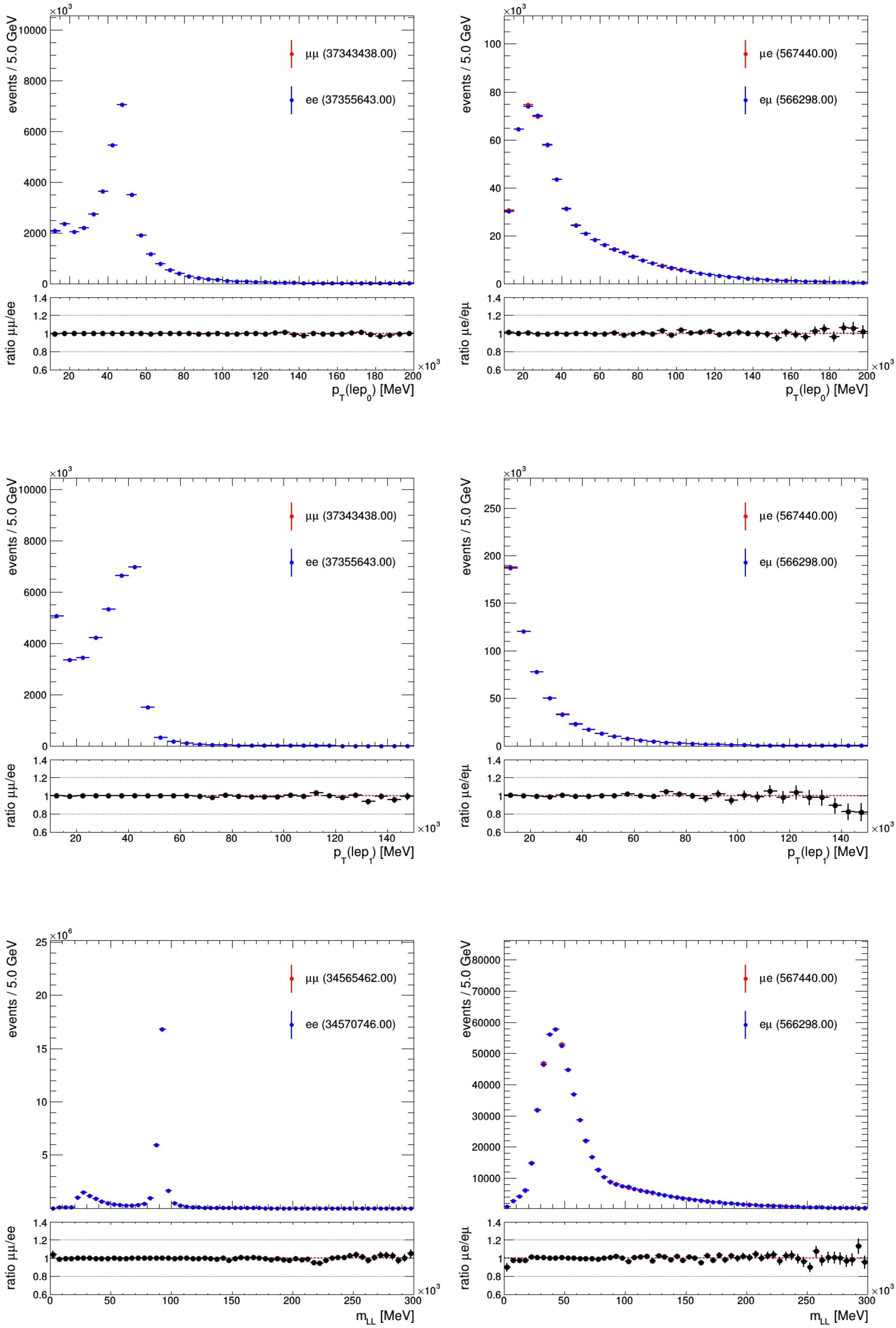


Figure 57: Comparison of ee vs. $\mu\mu$ (left) and $e\mu$ vs. μe (right) truth datasets with "lepton showering" deactivated (more distributions are displayed in Figure 114).

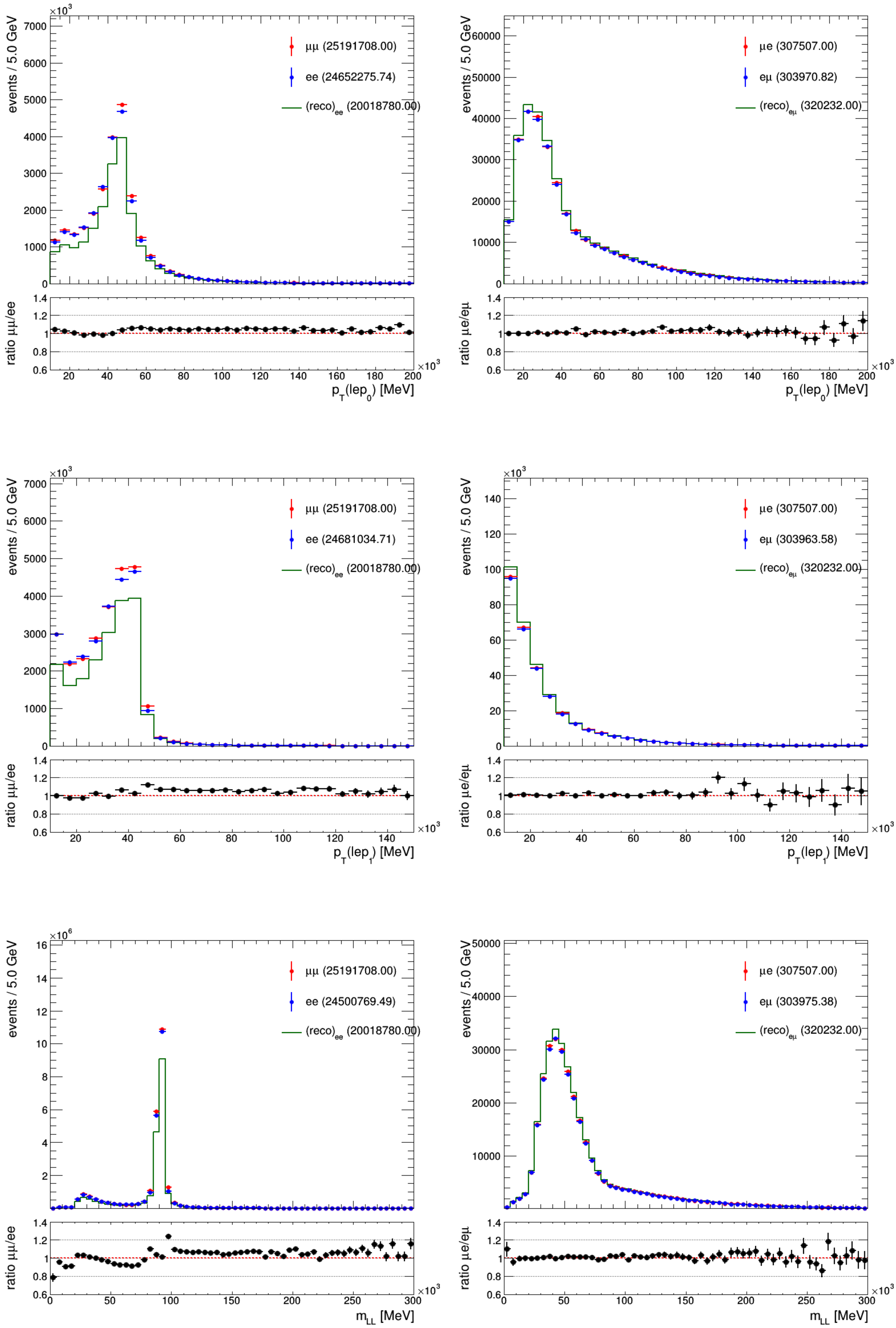


Figure 58: Comparison of ee vs. $\mu\mu$ (left) and $e\mu$ vs. μe (right) reconstructed datasets after the efficiency-based correction is applied to ee or $e\mu$ (more distributions are displayed in Figure 115).

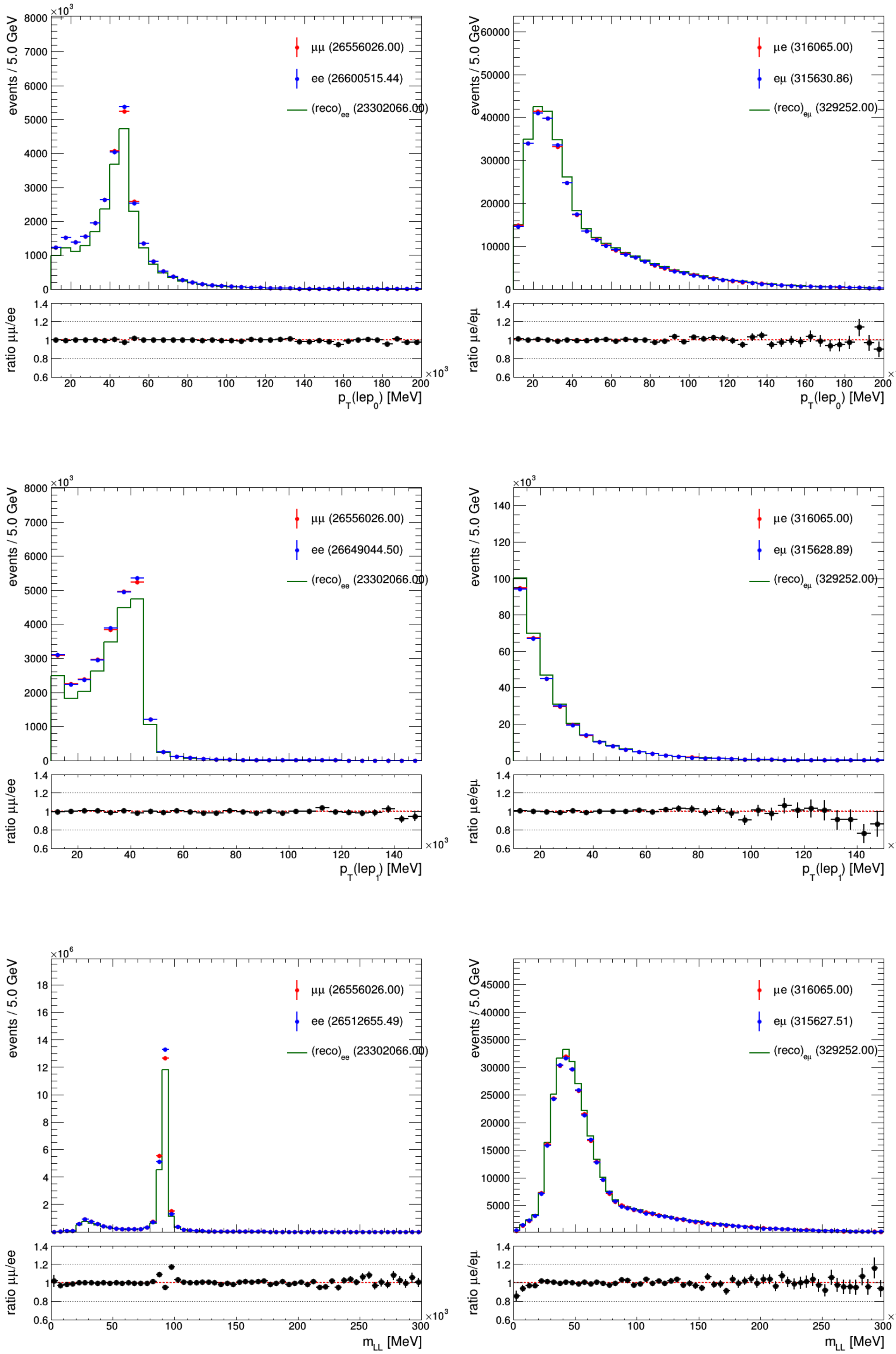


Figure 59: Comparison of ee vs. $\mu\mu$ (left) and $e\mu$ vs. μe (right) reconstructed datasets with "lepton showering" deactivated after the efficiency-based correction is applied to ee or $e\mu$ (more distributions are displayed in Figure 116).

5.5 Generic symmetry restoration

In this section, we present a different approach to restore the symmetry between two datasets, based on the method introduced in the Higgs LFV example study at the beginning of section 3.3.3. It relies on MC simulation but can also be applied efficiently to the measured data. In addition, it accounts for any asymmetry accurately modeled in MC. Being generic and efficient, it can be especially useful in a symmetry-based DDP implementation that aims to scan many different selections, searching for hints of BSM physics.

This alternative method for symmetry restoration between two datasets is described in section 5.5.1. In section 5.5.2, we implement it using the standalone MC samples, comparing datasets with various leptonic final states after correction. In section 5.5.3, based on the ATLAS Run-2 datasets from the Higgs LFV search, we compare this approach to the efficiency-based correction method implemented in the analysis. A discussion of the results is given in section 5.5.4.

5.5.1 Description

We describe this method in the context of the e/μ-symmetry method. Let us consider the e- and μ-datasets introduced in section 3.3. They are expected to be symmetric at interaction point but include some efficiency-induced asymmetry at detection level. To restore the symmetry, we scale the e-dataset to the μ-dataset by applying the event-by-event efficiency-ratio CF:

$$R^{\epsilon} = \frac{\epsilon_{\mu}}{\epsilon_{e}} \tag{71}$$

Up to now, R^{ϵ} is determined by estimating the ϵ_e and ϵ_μ efficiencies separately. We propose to estimate it from the ratio of μ- over e-dataset event distributions (N_μ/N_e) instead. Indeed, based on the symmetry assumption, if N_0 is the number of events at interaction point, then $N_e = \epsilon_e \cdot N_0$ and $N_\mu = \epsilon_\mu \cdot N_0$, leading to the equalities:

$$\frac{N_e}{\epsilon_e} = \frac{N_\mu}{\epsilon_\mu} \Leftrightarrow \frac{\epsilon_\mu}{\epsilon_e} = \frac{N_\mu}{N_e} \tag{72}$$

This approach can be challenging with measured data due to the presence of fake and potential signal events. The solution considered is to measure R^{ϵ} from MC and use the SFs provided by the distinct ATLAS performance groups to obtain CFs applicable to data distributions. Since $\epsilon^{\text{data}} = \text{SF} \cdot \epsilon^{\text{MC}}$, this is described by:

$$R^{\epsilon} = \frac{\epsilon_{\mu}^{\text{data}}}{\epsilon_{e}^{\text{data}}} = \frac{\text{SF}_{\mu} \cdot N_{\mu}^{\text{MC}}}{\text{SF}_{e} \cdot N_{e}^{\text{MC}}} \tag{73}$$

To implement this, we derive R^{ϵ} correction maps from the ratio of n-dimensional SF-scaled MC distributions, where n is the number of variables chosen to parametrize the correction. The correction is then applied event-by-event directly using the values obtained from this map as CFs.

In the following, we sometimes refer to this approach as the MC distribution-based correction, compared to the efficiency-based correction previously implemented.

The main advantage of this method is its simplicity. Indeed, estimating efficiencies – which can vary depending on the selection – is no longer necessary. In contrast, SFs are much more stable (as discussed in section 4.3.4). Without much effort, dedicated maps for different selections can be derived with varying parametrizations if needed. Another advantage is that all asymmetries accurately modeled in MC can be corrected using this method, even if they aren't efficiency-related. This is illustrated in section 5.5.2.

On the other hand, this approach relies on MC, rendering the intended data-driven analysis method simulation-dependent. But even with the efficiency-based method implemented in the Higgs LFV analysis, electron efficiencies are measured from MC (see section 4.4). And since we use a ratio of MC distributions, we can expect semi-cancellations of potential mismodelings. The main limitation comes from the amount of MC statistics available. If too low, it limits the number of parametrizations that can be included in the correction, rendering it imprecise. If confronted with this scenario, a careful choice of the binning can help to some degree, or additional MC samples could be requested.

Finally, this approach does not necessarily replace the efficiency-based method used up to now. Indeed, a hybrid implementation can be envisioned. For example, one can use the muon and trigger efficiencies provided by the ATLAS performance groups in combination with this approach to correct the residual asymmetry related to electron efficiencies. In the context of the $e\mu$ vs. μe comparison, we illustrate this when correcting the $e\mu$ dataset. We first derive the MC-based correction map with muon and trigger efficiency correction applied to the $e\mu$ events:

$$R^{\epsilon}_{e,\mathrm{map}} = \frac{\mathrm{SF}_{\mu e} \cdot N^{\mathrm{MC}}_{\mu e}}{\frac{\epsilon^{\mathrm{data}}_{\mu 0} \cdot \epsilon^{\mathrm{data}}_{trig,\mu e}}{\epsilon^{\mathrm{data}}_{\mu 1} \cdot \epsilon^{\mathrm{data}}_{trig,e\mu}} \cdot \mathrm{SF}_{e\mu} \cdot N^{\mathrm{MC}}_{e\mu}} \tag{74}$$

Similarly, this can be implemented as follows:

$$R^{\epsilon}_{e,\mathrm{map}} = \frac{\epsilon^{\mathrm{data}}_{\mu 1} \cdot \epsilon^{\mathrm{data}}_{trig,e\mu} \cdot \mathrm{SF}_{\mu e} \cdot N^{\mathrm{MC}}_{\mu e}}{\epsilon^{\mathrm{data}}_{\mu 0} \cdot \epsilon^{\mathrm{data}}_{trig,\mu e} \cdot \mathrm{SF}_{e\mu} \cdot N^{\mathrm{MC}}_{e\mu}} \tag{75}$$

In this ratio, the numerator (denominator) is the n-dimensional MC distribution of the μe ($e\mu$) dataset scaled event-by-event with various factors. The resulting $R^{\epsilon}_{e,\mathrm{map}}$ map only accounts for differences related to electron efficiencies since the muon and trigger efficiency corrections are applied. To correct the measured $e\mu$ data, the following CF is applied event-by-event:

$$R^{\epsilon}_{\mathrm{CF}} = \frac{\epsilon^{\mathrm{data}}_{\mu 0} \cdot \epsilon^{\mathrm{data}}_{trig,\mu e}}{\epsilon^{\mathrm{data}}_{\mu 1} \cdot \epsilon^{\mathrm{data}}_{trig,e\mu}} \cdot R^{\epsilon}_{e,\mathrm{CF}} \tag{76}$$

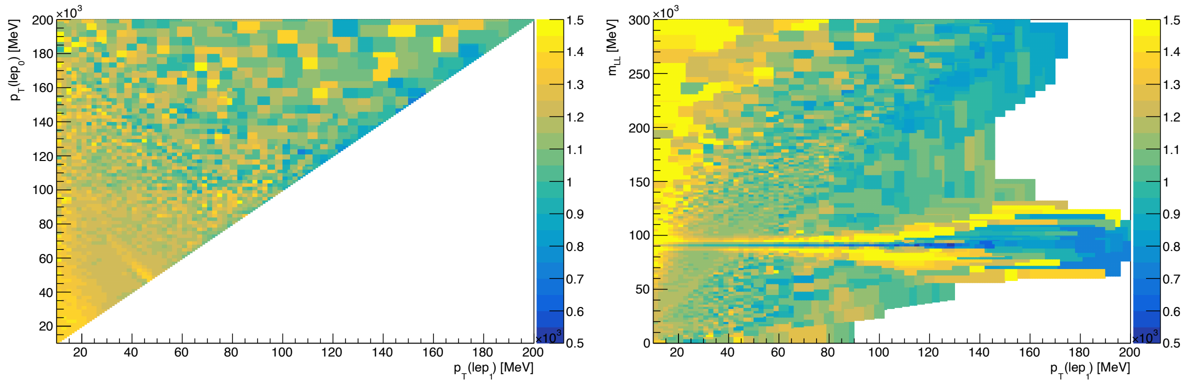
where the $R^{\epsilon}_{e,\mathrm{CF}}$ CF is obtained from the $R^{\epsilon}_{e,\mathrm{map}}$ map derived in (74) or (75).

5.5.2 Implementation in standalone MC samples

We implement this alternative symmetry-restoration approach to the standalone MC samples described in section 5.4, particularly to the reconstructed data with "lepton shower" activated. In this context, no data samples are available; therefore, no SFs are applied. Instead, we verify the level of restored symmetry within the same datasets used to generate the CF maps. This approach is compared to the efficiency-based method implemented in section 5.4.3.

ee vs. $\mu\mu$

In this example, we first derive a 2D CF map parametrized in ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) from the ratio of $\mu\mu$ over ee distributions. For simplicity, we didn't include any η dependency. As a result, we don't expect the lepton η distributions to be symmetric after correction. Still, we expect the symmetry to be restored in the other distributions, such as $|\Delta\eta_{\ell\ell}|$. A specific bin-merging algorithm (that can be improved) is applied to adapt the size of the CF bins to the available statistics: for candidate bin sizes ranging from 2×2 - 40×40 GeV, we select, in each sub-region of the map, the CF value from the smallest bin in which both the $\mu\mu$ and ee datasets have a relative statistical uncertainty smaller than $110^{-1/2}$ (meaning at least 110 events since no event weights are applied in these samples). The resulting map is shown in Figure 60(a).



(a) (b)

Figure 60: CF maps derived from the ratio of the $\mu\mu$ over ee ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) (a) and ($m_{\ell\ell}$ vs. $p_{\mathrm{T}}^{\ell_1}$) (b) distributions.

We apply the derived CFs event-by-event to the ee dataset and review the level of restored symmetry. As seen in appendix D.2, Figure 117, although the lepton p_{T} distributions agree well, the asymmetry caused by the lepton energy loss via bremsstrahlung and the momentum resolution isn't completely resolved in other distributions such as $m_{\ell\ell}$ or $|\Delta\eta_{\ell\ell}|$.

In a second attempt, we include $m_{\ell\ell}$ in the parametrization of the CFs. This choice is motivated by the fact that the $m_{\ell\ell}$ distribution is the most sensitive to effects such as p_{T} resolution or lepton energy loss and that the dilepton system is characteristic of the event's topology in Drell-Yan processes. To this end, we derive a 3D CF map parametrized in ($m_{\ell\ell}$ vs. $p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) from the ratio of $\mu\mu$ over ee distributions, using the same method as before, except that the bins are now cubic instead of square. To illustrate the $m_{\ell\ell}$ dependency of the resulting CFs, Figure 60(b) displays a similar map but obtained from the 2D ($m_{\ell\ell}$ vs. $p_{\mathrm{T}}^{\ell_1}$) distributions instead. Distinctive features are observed around the Z peak at $m_{\ell\ell} = 90$ GeV, which correspond to the fluctuations due to p_{T} resolution and lepton energy loss.

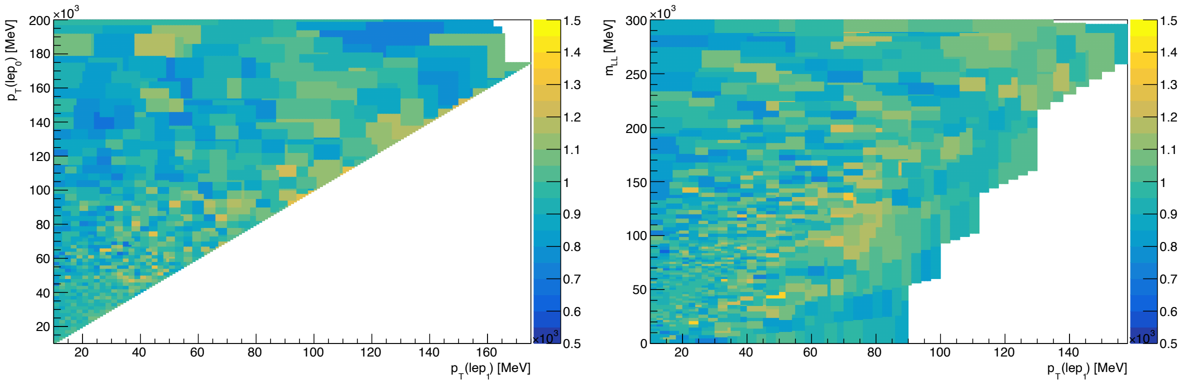
Using the CFs from the 3D map results in a significantly improved agreement, shown in appendix D.2, Figure 118, especially for the $m_{\ell\ell}$ distribution that is now symmetric. But asymmetries appeared in the higher range of the leptons' p_{T} distributions ($p_{\mathrm{T}} > 80 \sim 100$ GeV), which we attribute to an imprecise correction due to the reduced statistics in these regions.

To address this issue, we attempt a correction that uses CFs from the 3D ($m_{\ell\ell}$ vs. $p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) map for events with $p_{\mathrm{T}}^{\ell_0} < 90$ GeV and CFs from the 2D ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) map for events with $p_{\mathrm{T}}^{\ell_0} > 90$ GeV. The threshold value (90 GeV) is chosen empirically from observing the residual asymmetries previously noticed and doesn't have any special motivation otherwise. In this case, the symmetry is precisely restored in all the distributions considered (except for the lepton η distributions), as is shown in Figure 63.

In summary, in this example, the symmetry is well restored using this approach. The asymmetries from different sources (efficiencies, momentum resolution, and energy loss) are effectively and simultaneously corrected. In comparison, the efficiency-based correction applied in Figure 58 results in significant residual asymmetry. Another advantage of this method is the flexibility to choose the parametrization of the correction to address any residual asymmetry.

$e\mu$ vs. μe

Similarly, in this example, we derive CF maps with the same method and parametrizations as in the previous case but from the ratio of μe over $e\mu$ distributions. These are shown in Figure 61.



(a) (b)

Figure 61: CF maps derived from the ratio of the μe over $e\mu$ ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) (a) and ($m_{\ell\ell}$ vs. $p_{\mathrm{T}}^{\ell_1}$) (b) distributions.

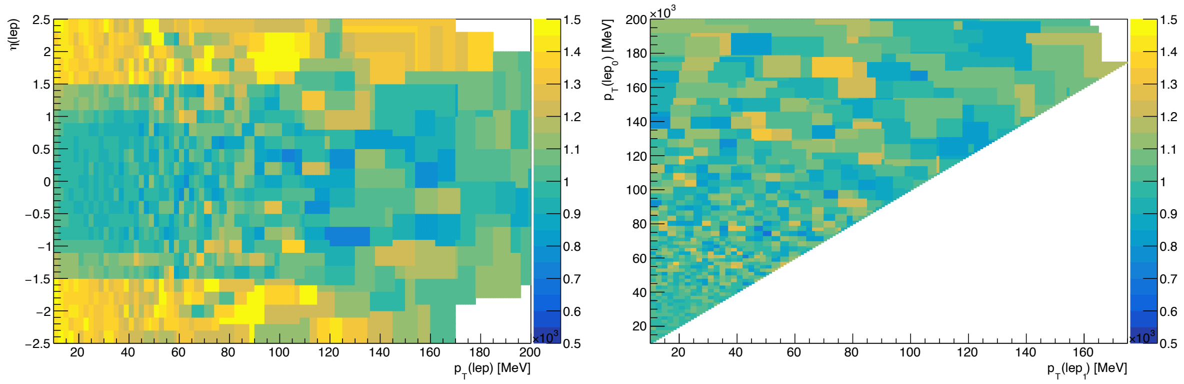
In this case, already with the 2D ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) correction, good agreement is achieved for almost all distributions (excluding the lepton η distributions), as shown in appendix D.2, Figure 119. In the $|\Delta\eta_{\ell\ell}|$ distribution, some small asymmetry ($< 10\%$) is present in a few bins, although it is also generally symmetric. Including η parametrizations in the CF map could potentially resolve this asymmetry.

With the same correction strategy implemented in the ee vs. $\mu\mu$ example (using CFs from the 3D ($m_{\ell\ell}$ vs. $p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) map for events with $p_{\mathrm{T}}^{\ell_0} < 90$ GeV and CFs from the 2D ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) map for events with $p_{\mathrm{T}}^{\ell_0} > 90$ GeV) the symmetry is improved further although the differences are small compared to the statistical fluctuations. Still, some improvement in the $m_{\ell\ell}$ and $|\Delta\eta_{\ell\ell}|$ distributions is visible, as seen in Figure 64.

$e\tau_{\mathrm{had}}$ vs. $\mu\tau_{\mathrm{had}}$

We also consider the case of $e\tau_{\mathrm{had}}$ vs. $\mu\tau_{\mathrm{had}}$. Here, only one light lepton is present per event, and we don't expect asymmetries related to the hadronic τ since it is present in both datasets. As a result, only single-lepton corrections are needed, simplifying the correction's parametrization. In the following, the subscript $_0$ ($_1$) indicates the light lepton (τ_{had}).

We derive a 2D CF map parametrized in (η^{ℓ_0} vs. $p_{\mathrm{T}}^{\ell_0}$) (of the light lepton) from the ratio of $\mu\tau_{\mathrm{had}}$ over $e\tau_{\mathrm{had}}$ distributions. The method used is the same as in the previous examples but with candidate bin sizes in the η-axis ranging from 0.02 - 1. The resulting CF map is shown in Figure 62(a). After applying the correction to the $e\tau_{\mathrm{had}}$ dataset, the symmetry is well restored, as is shown in Figure 65.



(a) (b)

Figure 62: (a) CF map derived from the ratio of the $\mu\tau_{\mathrm{had}}$ over $e\tau_{\mathrm{had}}$ (η^{ℓ_0} vs. $p_{\mathrm{T}}^{\ell_0}$) distributions of the light leptons. (b) CF map derived from the ratio of the $e^-\mu^+$ over $e^+\mu^-$ ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) distributions.

$e^+\mu^-$ vs. $e^-\mu^+$

Finally, we consider the case of $e^+\mu^-$ vs. $e^-\mu^+$. These two datasets only differ by the charge of the two leptons. Since Delphes doesn't simulate charge-dependent efficiencies or momentum smearing, and lepton energy loss is charge-independent, we expect the two datasets to be symmetric, even without correction. Still, we apply the same procedure as in the other examples. In the following, the subscript $_0$ ($_1$) indicates the leading (subleading) lepton, even though it can be an electron or a muon in both datasets.

We derive a 2D CF map parametrized in ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) from the ratio of $e^-\mu^+$ over $e^+\mu^-$ distributions, which is shown in Figure 62(b). As seen in Figure 66, the two datasets are symmetric in all distributions, even without correction. Applying the correction to the $e^+\mu^-$ dataset slightly improves the overall agreement.

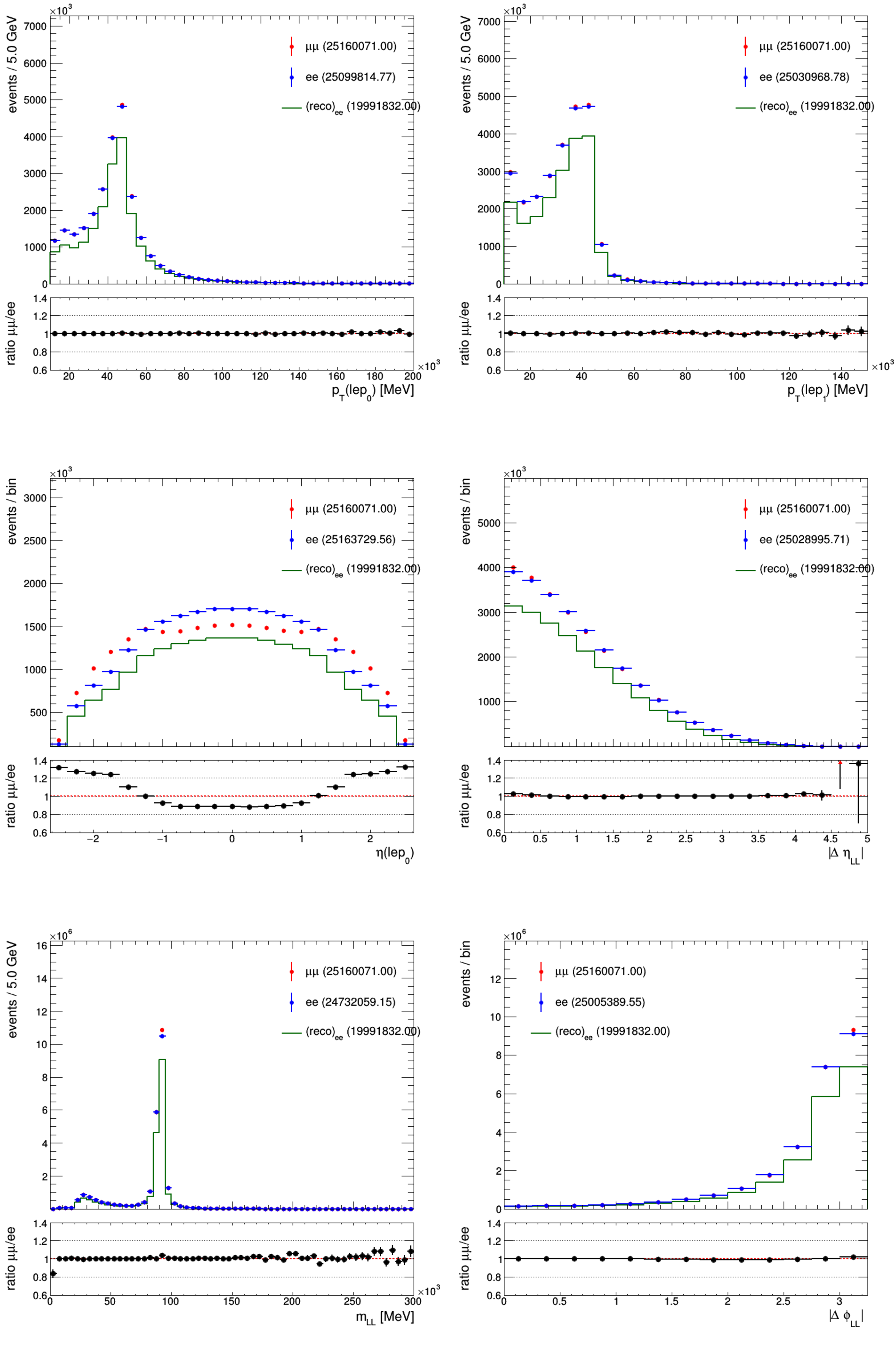


Figure 63: Comparison of the ee vs. $\mu\mu$ reconstructed datasets after the mixed ($m_{\ell\ell}$ vs. $p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) and ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) MC distribution-based correction is applied to ee.

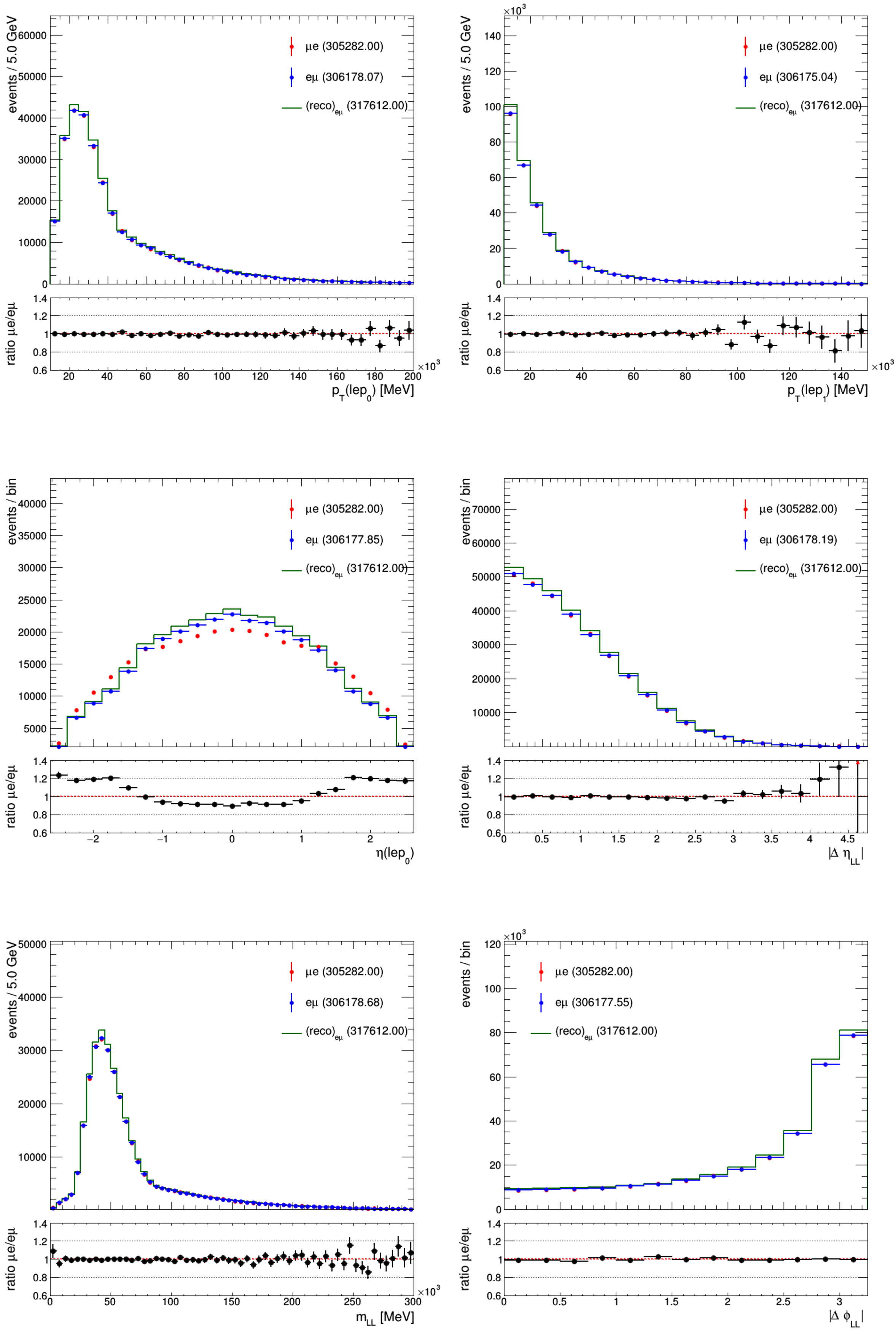


Figure 64: Comparison of the $e\mu$ vs. μe reconstructed datasets after the mixed ($m_{\ell\ell}$ vs. $p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) and ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) MC distribution-based correction is applied to $e\mu$.

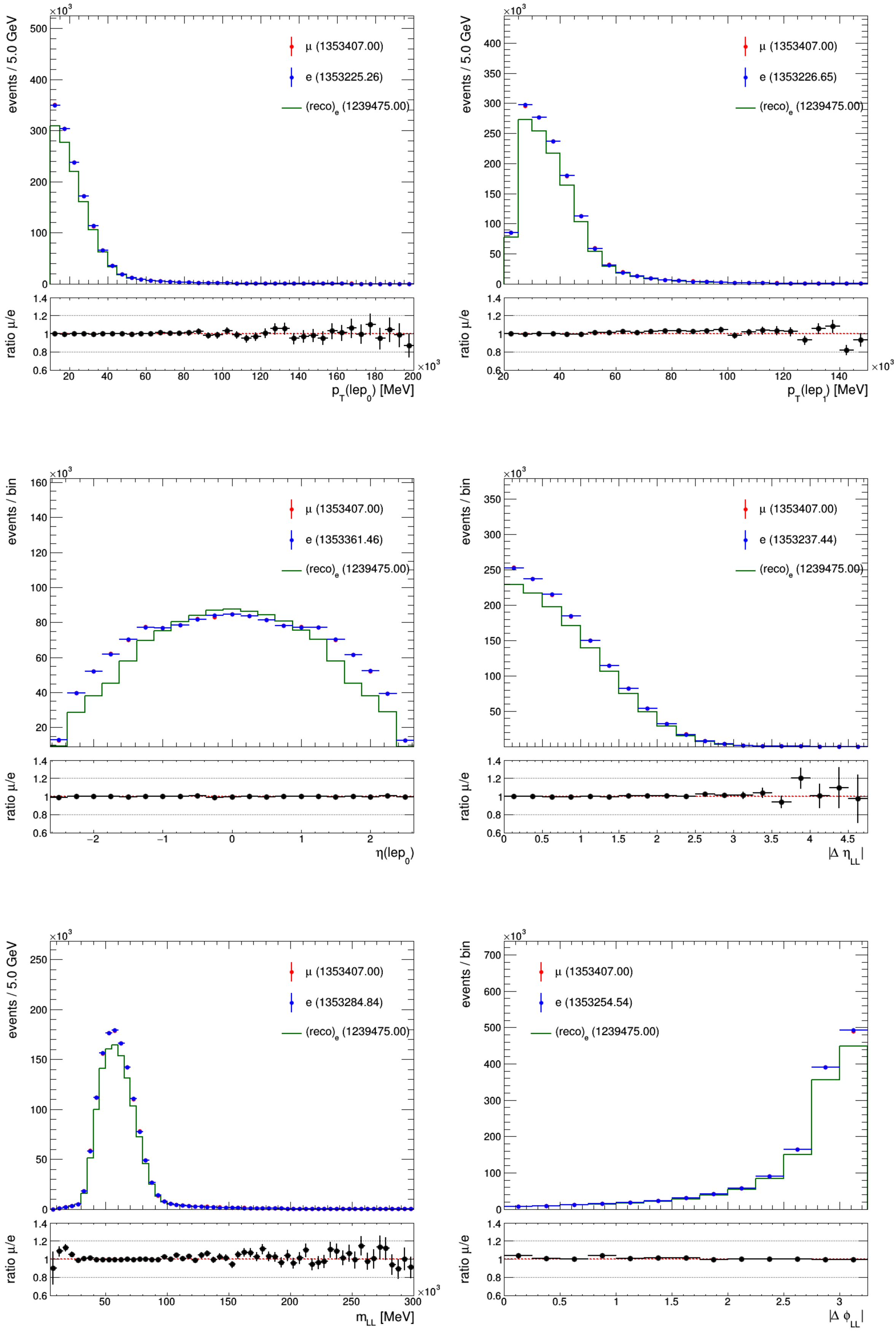


Figure 65: Comparison of the $e\tau_{\text{had}}$ vs. $\mu\tau_{\text{had}}$ reconstructed datasets after the (η^{ℓ_0} vs. $p_{\text{T}}^{\ell_0}$) MC distribution-based correction is applied to $e\tau_{\text{had}}$. The subscript $_0$ ($_1$) indicates the light lepton (τ_{had}).

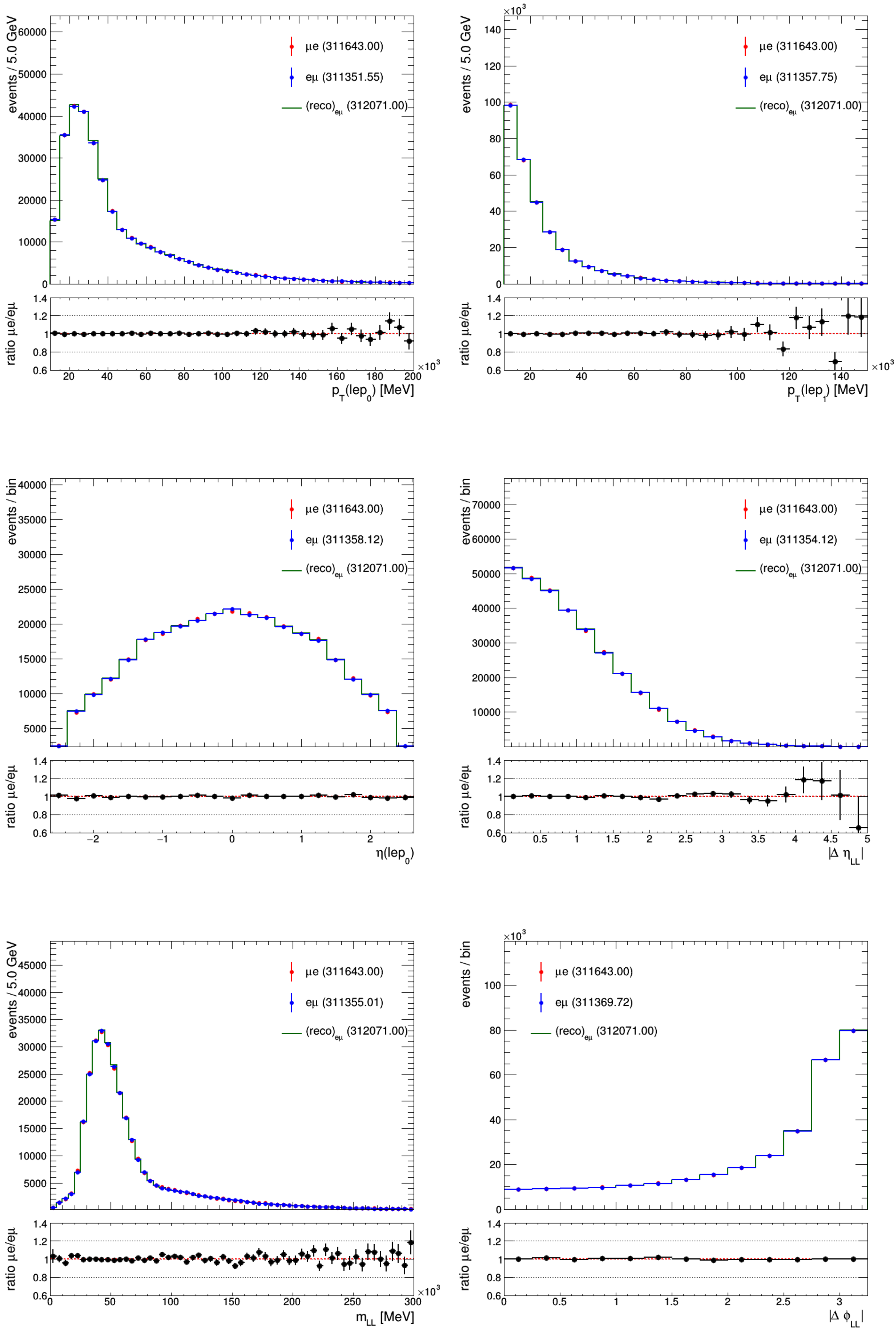


Figure 66: Comparison of the $e^+\mu^-$ vs. $e^-\mu^+$ reconstructed datasets after the ($p_T^{\ell_0}$ vs. $p_T^{\ell_1}$) MC distribution-based correction is applied to $e^+\mu^-$. In the legend, $e\mu$ stands for $e^+\mu^-$ and μe for $e^-\mu^+$.

5.5.3 Implementation in ATLAS Run-2 samples

In this section, we evaluate the alternative approach for symmetry restoration using the ATLAS Run-2 data (measured and simulated) that was used in the Higgs LFV analysis presented in section 4 (see section 4.1 for a description of the samples). In addition, we use the same object reconstruction (section 4.2), event classification (section 4.3.2), event selection (section 4.3.3), fake estimate (section 4.5), and background combination strategy (section 4.3.5). Only the obtention of the efficiency-CFs is different (section 4.3.4) even though the way they are applied remains unchanged.

In this manner, we can compare this approach to the efficiency-based method implemented in the analysis[11]. To this end, we compare the resulting symmetry after correction between the $e\tau_\mu$ and $\mu\tau_e$ datasets based on the MC samples (section 4.6.1) and the measured data samples (section 4.6.3). Following the analysis strategy, this comparison is made separately in the VBF and non-VBF SRs.

To derive the CF maps, we use truth-matched prompt $e\tau_\mu$ and $\mu\tau_e$ events from the SM MC sample ($Z \to \tau\tau$, $t\bar{t}$, single top-quark, diboson, $H \to \tau\tau$ and $H \to WW$), which pass the baseline selection requirements. Separate maps are generated for the VBF and non-VBF SRs. In addition, two distinct parametrizations are considered:

- $p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$
- $p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$

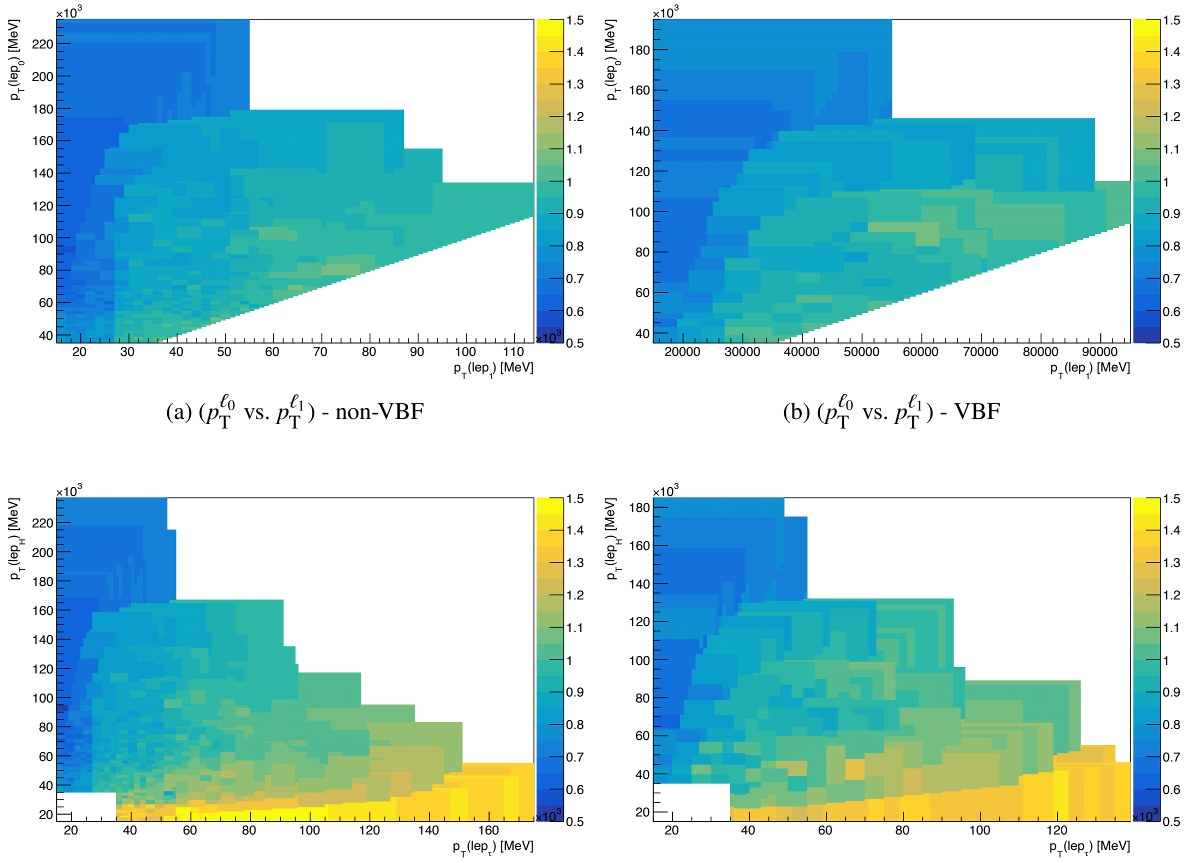
As a reminder, ℓ_H (ℓ_τ) is the leading (subleading) lepton in the Higgs estimated rest frame (see section 4.3.2). Therefore, in the ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) case, events with switched lepton p_{T} ordering in the Higgs and laboratory frame contribute to separate regions of the CF map than the other events. The goal is to better model the correction for these events if their topology is special in some way.

The CF maps are derived from the ratio of the $\mu\tau$ over $e\tau$ distributions. To produce CFs applicable to data events, all the usual SFs are applied to the MC distributions, as described in (73). The same bin-merging strategy used in section 5.5.2 is also implemented here, except that the relative statistical uncertainty threshold is lowered in the VBF SR from $110^{-1/2}$ to $40^{-1/2}$ due to lower MC statistics. The resulting CF maps are shown in Figure 67. We compare the $e\tau$ and $\mu\tau$ datasets after applying the CFs event-by-event to $e\tau$ (the same CFs are used in MC and data since the MC events are weighted by their usual SFs).

The resulting distributions of the SM MC sample (data vs. background predictions) when using the ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) maps in the non-VBF and VBF SRs are shown in appendix D.3, Figures 120 (122) and 121 (123), respectively. The resulting distributions of the SM MC sample (data vs. background predictions) when using the ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) maps in the non-VBF and VBF SRs are shown in appendix D.3, Figures 124 (126) and 125 (127), respectively.

Excluding the lepton η distributions, this alternative approach successfully restores the expected symmetry in all the samples, selections, and distributions considered. As a general trend, compared to the efficiency-based correction, the modeling achieved is improved in regions with high statistics but can be slightly worse in regions with fewer events. Similarly, the ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) correction is more precise than with ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$), provided there are enough statistics. As such:

[11] This alternative symmetry-restoration approach was developed when the analysis was nearly complete; therefore, implementing it in the analysis hasn't been considered



(a) ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) - non-VBF

(b) ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) - VBF

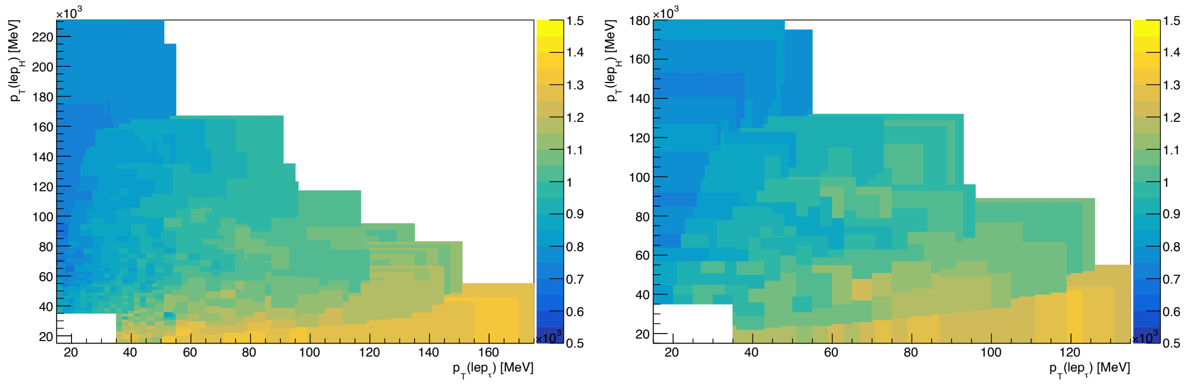
(c) ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) - non-VBF

(d) ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) - VBF

Figure 67: CF maps derived from the ratio of $\mu\tau$ over $e\tau$ SM MC distributions with the ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) or ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) parametrizations and in the non-VBF or VBF SRs.

- With MC data in the non-VBF SR, the MC distribution-based correction performs better than the efficiency-based one for all distributions. However, in some specific regions, the modeling is worse, e.g., lower range of $m_{\mathrm{T}}(\ell_0, E_T^{\mathrm{miss}})$ and $\Delta\phi(\ell\ell)$. The agreement is better with ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) than ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$), especially in the lepton p_{T} distributions
- With MC data in the VBF SR, the modeling is also generally improved with the new approach, except for the lower range of $m_{\mathrm{T}}(\ell_0, E_T^{\mathrm{miss}})$ once again. Using ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) or ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) maps leads to similar results on average, with some distributions more symmetric with the first (e.g., $p_{\mathrm{T}}^{\ell_H}$) and others with the second (e.g., $p_{\mathrm{T}}^{\ell_\tau}$)
- With measured data in the non-VBF SR, the agreement is very similar with all methods, with differences at the level of the uncertainties, although $p_{\mathrm{T}}^{\ell_\tau}$ is noticeably improved with the alternative approach
- With measured data in the VBF SR, the agreement is also similar with all methods and difficult to compare due to the large fluctuations. Excluding the lepton p_{T} distributions, the efficiency-based correction performs slightly better on average

Finally, we also implemented the hybrid approach proposed at the end of section 5.5.1, where we apply muon and trigger efficiency-based corrections and rely on the MC distribution-based correction to account for the residual electron efficiency-related asymmetry. Following the description in (75), we derive the CF maps from the MC distributions scaled by the “switched” muon and trigger efficiencies ($\epsilon_{\mu 1} \cdot \epsilon_{trig,e\mu}$ for μe events and $\epsilon_{\mu 0} \cdot \epsilon_{trig,\mu e}$ for $e\mu$ events). These are shown in Figure 68 with the ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) parametrization. As expected, the CFs are closer to unity in this case.



(a) ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) - non-VBF (b) ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) - VBF

Figure 68: CF maps derived from the $\mu\tau$ over $e\tau$ SM MC distributions scaled by their “switched” muon+trigger efficiencies in the non-VBF or VBF SRs.

To restore the symmetry between $e\tau$ and $\mu\tau$, the $e\tau$ dataset is corrected event-by-event with the product of the muon+trigger efficiency-ratio times the CFs from the maps just derived, as described in (76). The resulting distributions of the SM MC sample (data vs. background predictions) in the non-VBF and VBF SRs are shown in Figures 69 (71) and 70 (72), respectively.

Also here the symmetry is restored successfully in the different samples and selections. Compared to the standalone MC distribution-based correction, the modelling is improved in many distributions and similar in the others, excepting the lepton p_{T} ones where it performs slightly worse on average.

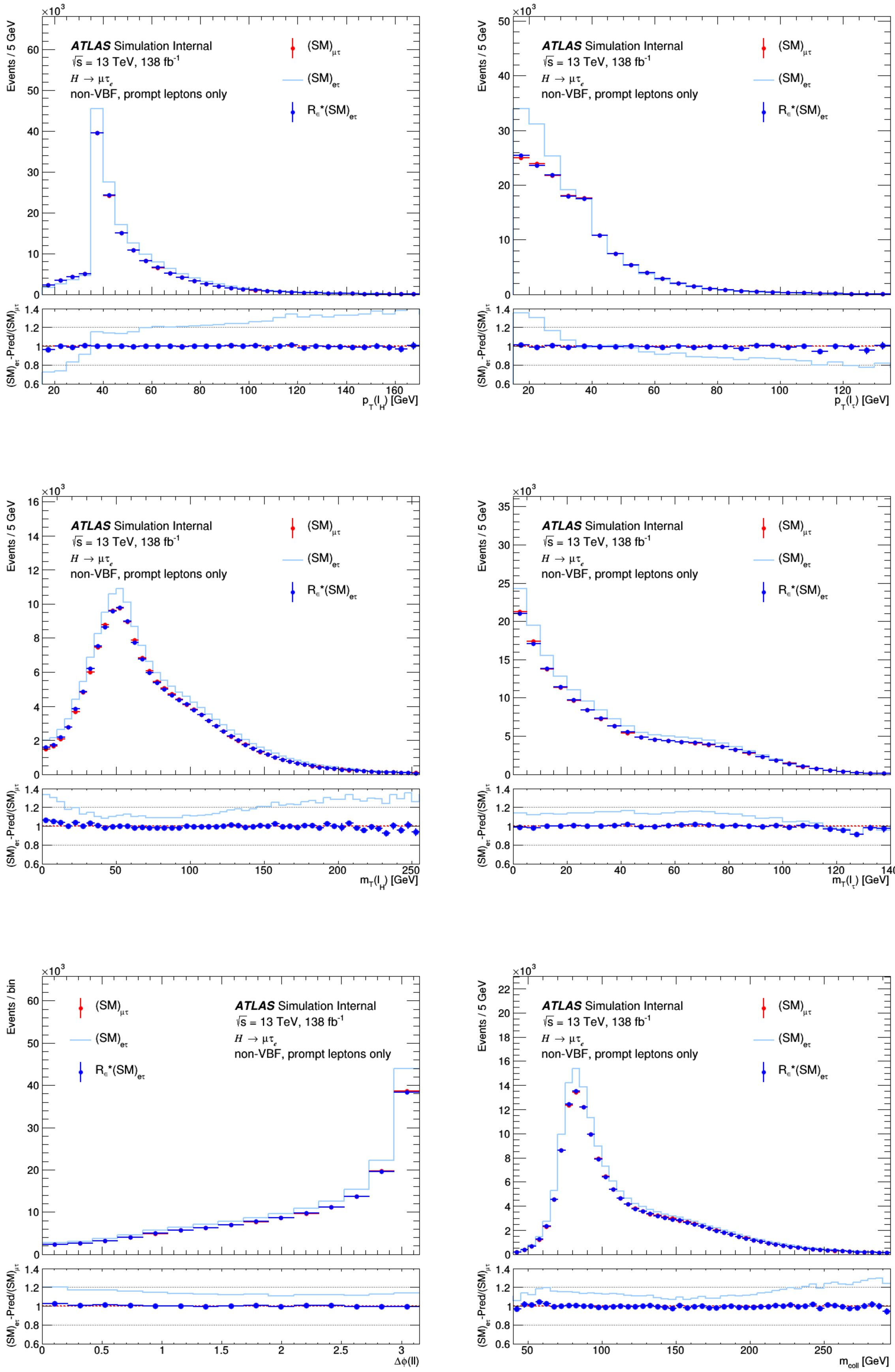


Figure 69: Comparison of $e\tau$ (dark blue) and $\mu\tau$ (red) SM MC datasets in the non-VBF selection after the ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) hybrid correction is applied to $e\tau$. The uncorrected $e\tau$ dataset is also displayed (light blue).

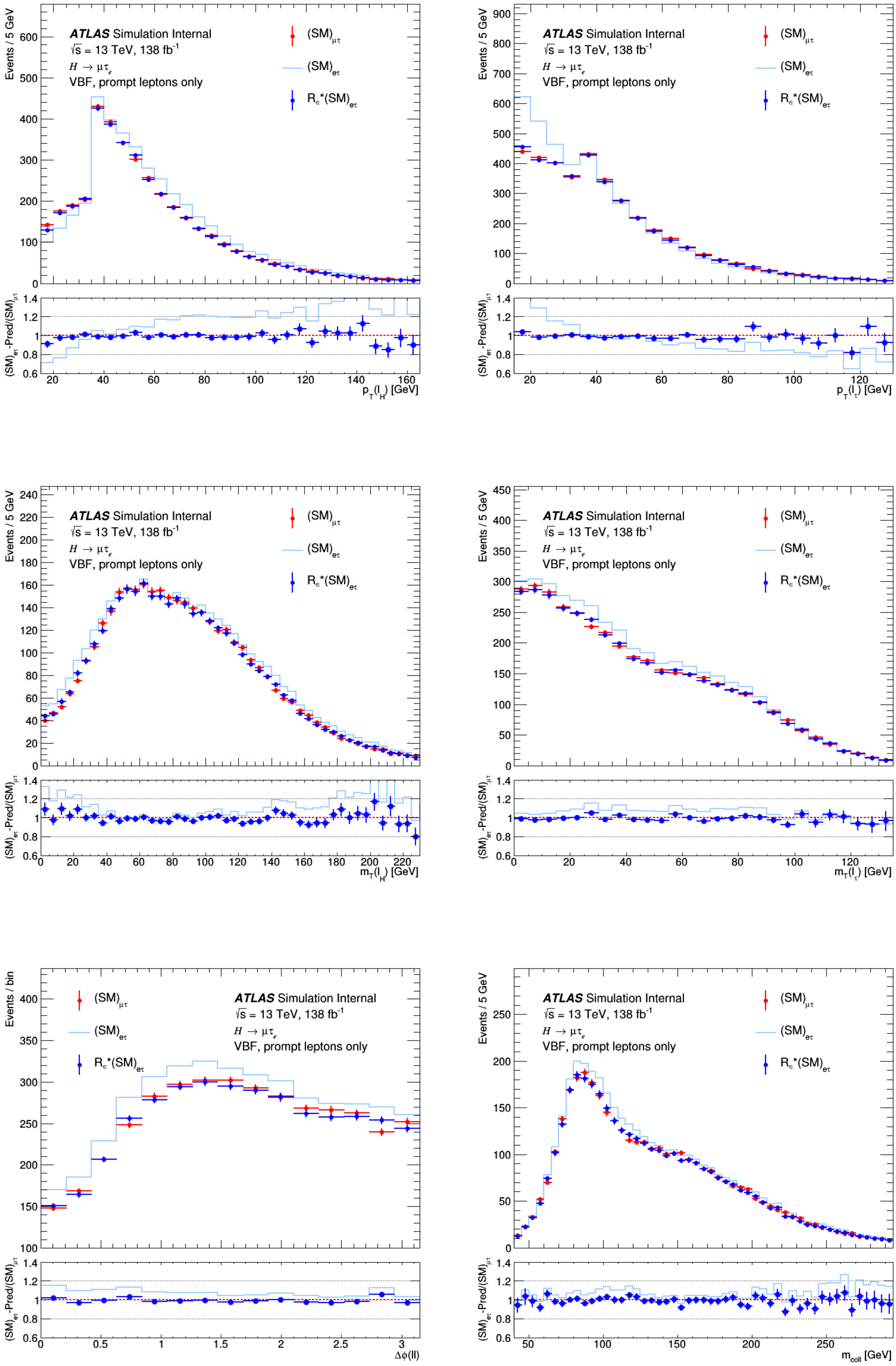


Figure 70: Comparison of $e\tau$ (dark blue) and $\mu\tau$ (red) SM MC datasets in the VBF SR after the ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) hybrid correction is applied to $e\tau$. The uncorrected $e\tau$ dataset is also displayed (light blue).

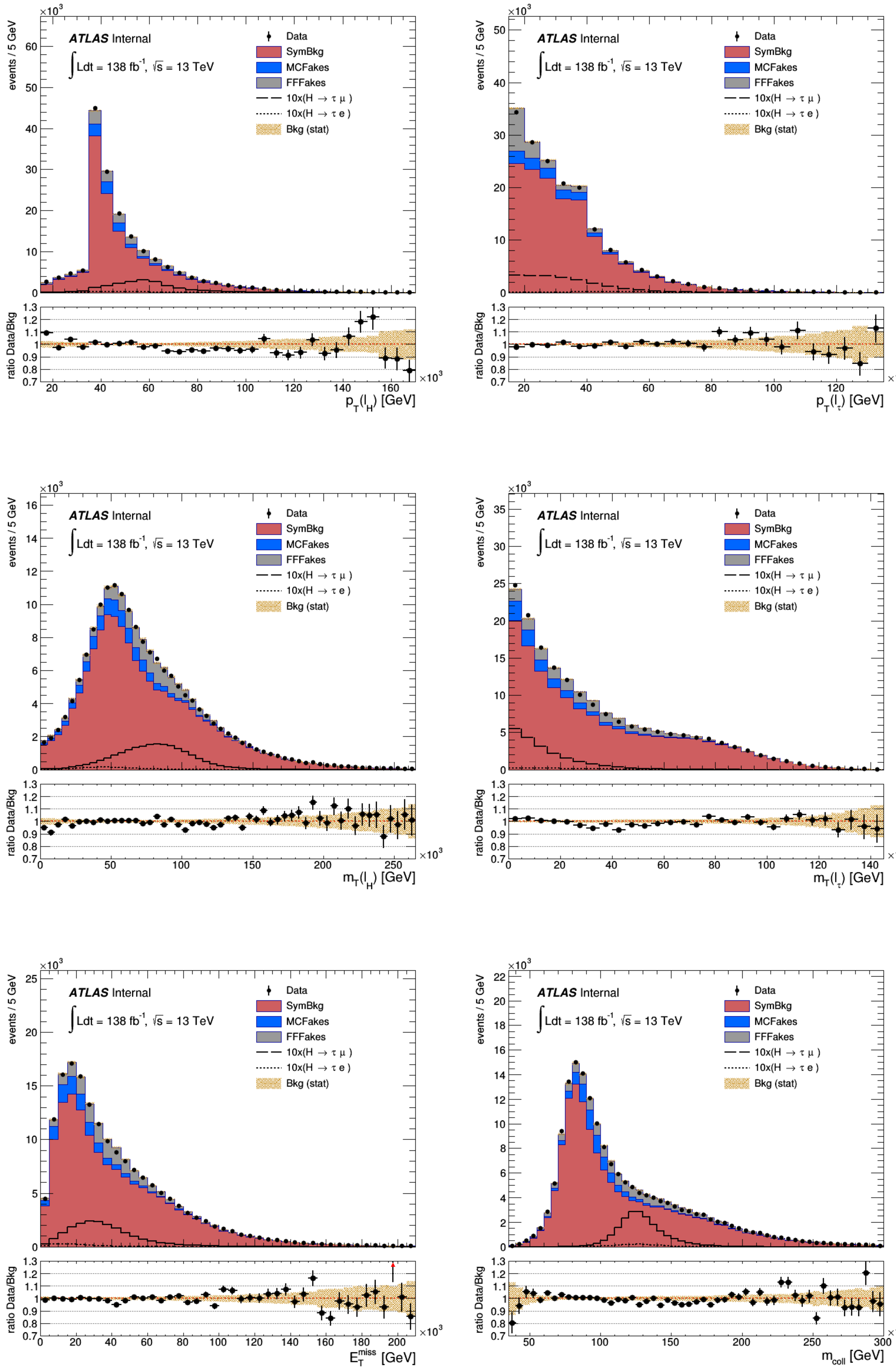


Figure 71: Comparison of the $e\tau$-based predicted background and $\mu\tau$ data in the non-VBF SR when the ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) hybrid correction is used to derive the symmetric background. Only statistical uncertainties are displayed.

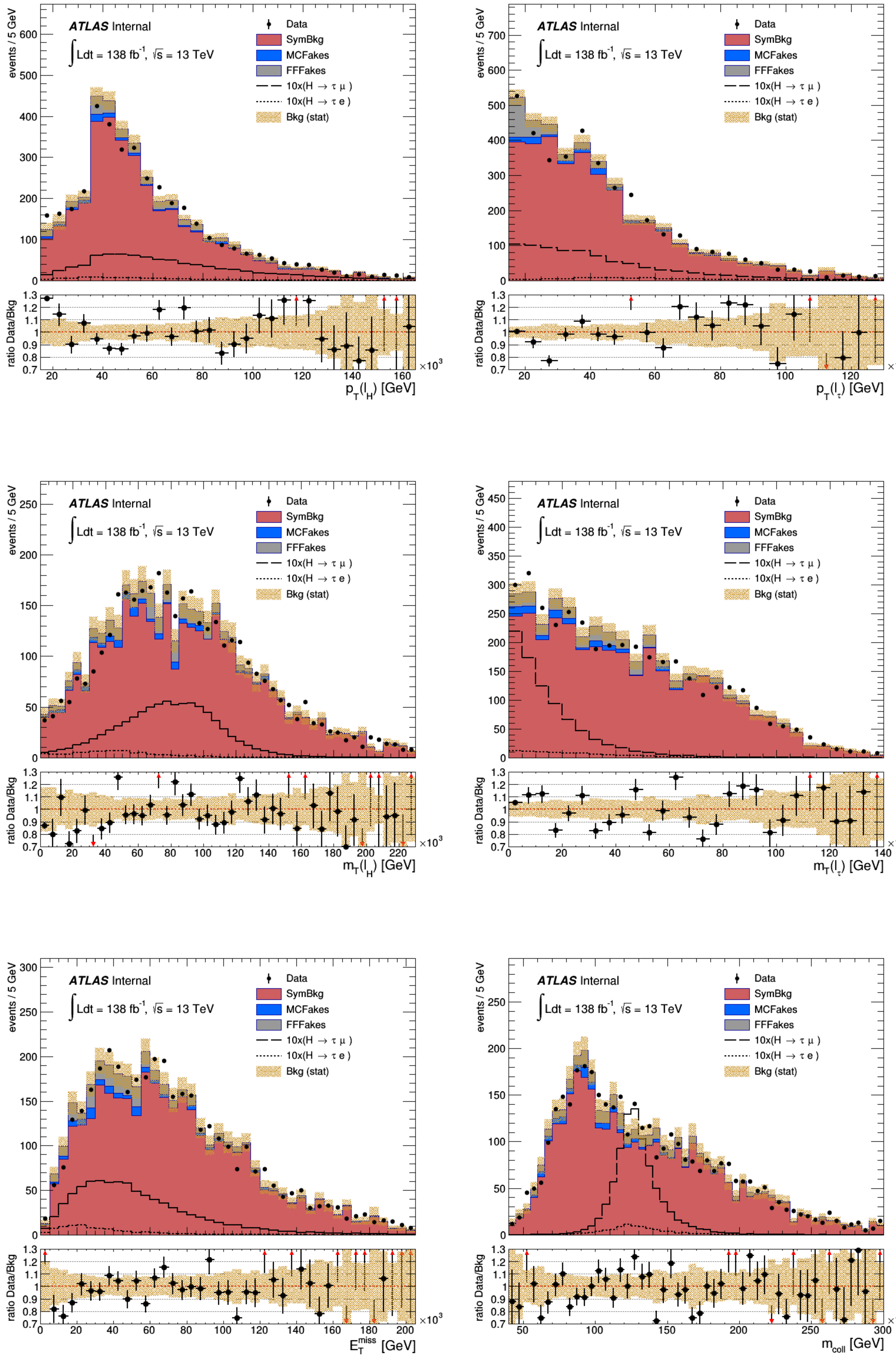


Figure 72: Comparison of the $e\tau$-based predicted background and $\mu\tau$ data in the VBF SR when the ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) hybrid correction is used to derive the symmetric background. Only statistical uncertainties are displayed.

5.5.4 Discussion

In this section, we considered an alternative approach to restore the symmetry between two datasets. Relying on MC distributions and their associated SFs, it provides a generic and simplified method to account for any source of asymmetry accurately modeled in MC, in particular efficiency differences between the compared datasets.

Implemented with the standalone MC samples, it successfully restores the expected symmetry between datasets with various dilepton final states: ee vs. $\mu\mu$, $e\mu$ vs. μe, $e\tau_{\text{had}}$ vs. $\mu\tau_{\text{had}}$ and $e^+\mu^-$ vs. $e^-\mu^+$. This illustrates to some degree the range of selections that can be probed for NP in a symmetry-based DDP implementation, although many more selections can be considered. Using this correction approach, we effectively accounted for the different sources of asymmetry present in the simulated samples: lepton efficiencies, energy loss, and momentum resolution. Depending on the datasets, different parametrizations were considered to improve the modeling.

Based on the ATLAS Run-2 samples from the Higgs LFV search, the performance of this approach is compared to the efficiency-based correction applied in the analysis. With only a 2D parametrization as a function of the leptons' p_{T}, the overall performance is improved in MC and similar in data. Adding more parameters to the correction, such as the lepton η or ϕ, could further improve the agreement since small residual asymmetries are present in some MC distributions.

A hybrid implementation which combined the muon+trigger efficiency-based correction and the MC distribution-based correction was also conducted. This improved the modeling in many distributions, leading to the best average performance among all the implementations considered.

6 Conclusions

Searching for hints of BSM physics is the main goal of the LHC experiments at CERN. Thus far, clear indications of deviations from SM predictions have yet to be uncovered. With most signatures predicted by well-motivated SM extensions (within the current LHC reach) already searched for, there is no clear guidance as to which searches should be conducted in the future. Yet, the amount of recorded data is the greatest ever accumulated and is largely unexplored. With limited resources at hand, novel and complementary approaches for discovery should be considered. This is the purpose of the proposed DDP; letting the data itself guide us towards its regions of interest, significantly enhancing our potential for discovery.

Symmetries of the SM can be exploited for this endeavor; by splitting the data into two datasets expected symmetric under the SM-only assumption, we become sensitive to any asymmetric contribution from potential BSM processes. This can be exploited to scan large portions of the observables space, efficiently uncovering regions more likely to include non-SM contributions. Such regions would be marked for further study using traditional data analysis methods. The research presented in this thesis aims at laying the groundwork for such implementations.

In this context, the e/μ-symmetry method is developed, which exploits the approximate symmetry between electrons and muons in SM processes to search for BSM physics. This data-driven analysis method was first used in the ATLAS Run-1 search for Higgs LFV decays. Since then, significant improvements have been achieved. In particular, a simplified statistical model implementation was developed. As a result, the efficiency correction – which accounts for the asymmetry induced by the different electron and muon efficiencies – is now applied event-by-event. This approach enables restoring the expected SM symmetry with much higher precision. In addition, the SR is not split into separate channels, leading to reduced statistical uncertainties on the background estimate. With these improvements, the analysis sensitivity is enhanced, and its application range is broadened.

One of the main efforts presented in this thesis is the application of the e/μ-symmetry method to the ATLAS Run-2 search for Higgs LFV decays based on a 138.42 fb^{-1} dataset of pp collisions at $\sqrt{s} = 13$ TeV. LFC is an accidental symmetry of the SM that is violated in nature via neutrino oscillation, but evidence of LFV has yet to be uncovered within LHC experiments. Two distinct searches were conducted in collaboration with other ATLAS members, the searches for $H \to e\tau$ and $H \to \mu\tau$. The main challenges, shared among the analyzers, included estimating the lepton efficiencies, developing the efficiency correction procedure, the data-driven estimation of the fake background, implementing a dedicated NN for signal enhancement, and conducting the statistical analysis itself. The final results are combined results of this analysis with two other MC-based searches for the same signals. Some tension with the SM assumption was observed at the level of 2.5σ, while no evidence for the $H \to e\tau$ signal was found. Upper limits on the BRs were set at 95% CL: 0.230% (0.192%) for $H \to e\tau$ and 0.163% (0.182%) for $H \to \mu\tau$ when the two searches were conducted independently (simultaneously).

The successful completion of the direct searches for Higgs LFV decays is an important endorsement of the data-driven e/μ-symmetry method. It shows that systematic effects induced at detection level, such as the different lepton efficiencies, can be effectively corrected and that the expected SM symmetry can be restored, in this case, between the $e\mu$ ($e\tau_\mu$) and μe ($\mu\tau_e$) datasets. In addition, the achieved sensitivity is comparable to that of more traditional analysis techniques. However, these searches still follow the blind-analysis paradigm where only a specific signal is searched for in a small theoretically motivated sub-region of the observables space. In terms of the DDP proposed, no specific signal is searched for. Instead, the full $e\mu$

and μe datasets are compared, and any significant asymmetry observed is considered a potential sign for BSM physics. Hence the $e\mu/\mu e$ comparison becomes a general test for LU. Furthermore, any two datasets with a switched number of electrons and muons in their final state can similarly be compared, each sensitive to different BSM manifestations. And other SM symmetries, exact or approximate, can also be tested using a similar approach.

Implementing such data-directed and generic searches based on symmetries of the SM is still at an initial stage. To demonstrate proof of concept, a procedure was developed in a simplified framework: identifying asymmetries between two measurements, represented by 2D histograms, using the generic N_σ test statistic. Relying on simulated data and neglecting, for now, systematic detector effects, we show that with little optimization, the sensitivity to detect asymmetries using this generic test is only slightly lower than that of optimal likelihood-based tests that have full knowledge of the signal. This approach has the advantage of being extremely fast, and the generalization to n-dimensional histograms is straightforward, enabling to efficiently scan large portions of the observables space for hints of BSM physics. Another approach that relies on weakly-supervised NNs, implemented by a collaborator, is also presented, paving the way towards NN implementations for the symmetry-based DDP. These studies, reported in [1], were developed on a practical case – the search for Higgs LFV decays, constructed from standalone MC data generated using dedicated software.

The procedure employed to generate these simulated samples is described. Making use of this simulated data, additional studies were conducted towards implementing the e/μ-symmetry method in datasets with various dilepton final states: ee vs. $\mu\mu$, $e\mu$ vs. μe, $e\tau_{\text{had}}$ vs. $\mu\tau_{\text{had}}$ and $e^+\mu^-$ vs. $e^-\mu^+$. We identified two sources of asymmetries not efficiency-related: lepton energy loss via bremsstrahlung and lepton momentum resolution. These mainly affected the ee vs. $\mu\mu$ symmetry and couldn't be resolved using the efficiency-based correction.

An alternative approach for symmetry restoration was then developed. Relying on MC simulation and its associated SFs, it provides a generic and simple method to account for any source of asymmetry accurately modeled in MC. Using this approach, the symmetry was successfully restored between the various datasets considered, accounting for all the different sources of asymmetry simultaneously. Implemented in the ATLAS Run-2 samples from the Higgs LFV search, the level of restored symmetry was improved in MC and similar in data, although in specific regions with low statistics, the efficiency-based correction was more precise. Being generic and efficient, this method can be especially useful in a symmetry-based DDP implementation. In addition, a hybrid implementation which led to the best average performance was conducted, with each approach a distinct component of the correction.

In the right hand, we hold a generic test to identify asymmetries, and in the left hand, a doublet of efficient methods to restore the expected SM symmetry. Determining how they couple, by taking into account, with the right hand, uncertainties associated with the corrections from the left hand, is the next necessary step towards setting up **data-directed searches for BSM physics based on symmetries of the SM**, which may lead to unpredicted discoveries.

References

[1] M. Birman et al., *Data-directed search for new physics based on symmetries of the SM*, Eur. Phys. J. C **82** (2022) 508, arXiv: `2203.07529 [hep-ph]` (cit. on pp. i, ii, 2, 33, 91, 134).

[2] M. K. Gaillard, P. D. Grannis and F. J. Sciulli, *The Standard model of particle physics*, Rev. Mod. Phys. **71** (1999) S96, arXiv: `hep-ph/9812285 [hep-ph]` (cit. on p. 3).

[3] M. C. Gonzalez-Garcia and Y. Nir, *Neutrino masses and mixing: Evidence and implications*, Rev. Mod. Phys. **75** (2003) 345, arXiv: `hep-ph/0202058 [hep-ph]` (cit. on p. 5).

[4] J. Schechter and J. W. F. Valle, *Neutrino Masses in SU(2) x U(1) Theories*, Phys. Rev. **D22** (1980) 2227 (cit. on p. 5).

[5] R. Barbier et al., *R-parity violating supersymmetry*, Phys. Rept. **420** (2005) 1, arXiv: `hep-ph/0406039 [hep-ph]` (cit. on p. 5).

[6] C. Arina and N. Fornengo, *Sneutrino cold dark matter, a new analysis: Relic abundance and detection rates*, JHEP **11** (2007) 029, arXiv: `0709.4477 [hep-ph]` (cit. on p. 5).

[7] J. Marsano, N. Saulina and S. Schafer-Nameki, *Compact F-theory GUTs with U(1) (PQ)*, JHEP **04** (2010) 095, arXiv: `0912.0272 [hep-th]` (cit. on p. 5).

[8] S. T. Petcov, *Neutrino mixing, leptonic CP violation, the seesaw mechanism and beyond*, Int. J. Mod. Phys. **A25** (2010) 4325, [,211(2010)] (cit. on pp. 5, 8).

[9] R. Davis Jr., D. S. Harmer and K. C. Hoffman, *Search for neutrinos from the sun*, Phys. Rev. Lett. **20** (1968) 1205 (cit. on p. 6).

[10] KamLAND Collaboration, *Measurement of neutrino oscillation with KamLAND: Evidence of spectral distortion*, Phys. Rev. Lett. **94** (2005) 081801, arXiv: `hep-ex/0406035 [hep-ex]` (cit. on p. 6).

[11] Super-Kamiokande Collaboration, *Evidence for an oscillatory signature in atmospheric neutrino oscillation*, Phys. Rev. Lett. **93** (2004) 101801, arXiv: `hep-ex/0404034 [hep-ex]` (cit. on p. 6).

[12] LSND Collaboration, *Evidence for neutrino oscillations from the observation of anti-neutrino(electron) appearance in a anti-neutrino(muon) beam*, Phys. Rev. **D64** (2001) 112007, arXiv: `hep-ex/0104049 [hep-ex]` (cit. on p. 6).

[13] K2K Collaboration, *Indications of neutrino oscillation in a 250 km long baseline experiment*, Phys. Rev. Lett. **90** (2003) 041801, arXiv: `hep-ex/0212007 [hep-ex]` (cit. on p. 6).

[14] S. Bifani, S. Descotes-Genon, A. Romero Vidal and M.-H. Schune, *Review of Lepton Universality tests in B decays*, J. Phys. G **46** (2019) 023001, arXiv: `1809.06229 [hep-ex]` (cit. on p. 6).

[15] S. Schael et al., *Precision electroweak measurements on the Z resonance*, Phys. Rept. **427** (2006) 257, arXiv: `hep-ex/0509008` (cit. on p. 6).

[16] ATLAS Collaboration, *Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 367, arXiv: `1612.03016 [hep-ex]` (cit. on p. 6).

[17] A. Pich, *Precision Tau Physics*, Prog. Part. Nucl. Phys. **75** (2014) 41, arXiv: `1310.7922 [hep-ph]` (cit. on p. 6).

[18] S. Schael et al., *Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP*, Phys. Rept. **532** (2013) 119, arXiv: `1302.3415 [hep-ex]` (cit. on p. 6).

[19] V. Cirigliano and I. Rosell, *Two-loop effective theory analysis of pi (K) —> e anti-nu/e [gamma] branching ratios*, Phys. Rev. Lett. **99** (2007) 231801, arXiv: `0707.3439 [hep-ph]` (cit. on p. 6).

[20] NA62 Collaboration, *Precision Measurement of the Ratio of the Charged Kaon Leptonic Decay Rates*, Phys. Lett. B **719** (2013) 326, arXiv: `1212.4012 [hep-ex]` (cit. on p. 6).

[21] A. Aguilar-Arevalo et al., *Improved Measurement of the $\pi \to e\nu$ Branching Ratio*, Phys. Rev. Lett. **115** (2015) 071801, arXiv: `1506.05845 [hep-ex]` (cit. on p. 6).

[22] HFLAV Collaboration, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018*, Eur. Phys. J. C **81** (2021) 226, arXiv: `1909.12524 [hep-ex]` (cit. on p. 7).

[23] LHCb Collaboration, *Test of lepton universality in beauty-quark decays*, Nature Phys. **18** (2022) 277, arXiv: `2103.11769 [hep-ex]` (cit. on pp. 7, 91).

[24] LHCb Collaboration, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, JHEP **08** (2017) 055, arXiv: `1705.05802 [hep-ex]` (cit. on p. 7).

[25] P. Ko, Y. Omura, Y. Shigekami and C. Yu, *LHCb anomaly and B physics in flavored Z' models with flavored Higgs doublets*, Phys. Rev. D **95** (2017) 115040, arXiv: `1702.08666 [hep-ph]` (cit. on p. 7).

[26] A. Carmona and F. Goertz, *Lepton Flavor and Nonuniversality from Minimal Composite Higgs Setups*, Phys. Rev. Lett. **116** (2016) 251801, arXiv: `1510.07658 [hep-ph]` (cit. on p. 7).

[27] F. J. Botella, G. C. Branco, M. Nebot and M. N. Rebelo, *Flavour Changing Higgs Couplings in a Class of Two Higgs Doublet Models*, Eur. Phys. J. C **76** (2016) 161, arXiv: `1508.05101 [hep-ph]` (cit. on p. 7).

[28] G. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher et al., *Theory and phenomenology of two-Higgs-doublet models*, Phys.Rept. **516** (2012) 1, arXiv: `1106.0034 [hep-ph]` (cit. on p. 8).

[29] P. Paradisi, *Higgs-mediated $\tau \to \mu$ and $\tau \to e$ transitions in II Higgs doublet model and supersymmetry*, JHEP **0602** (2006) 050, arXiv: `hep-ph/0508054 [hep-ph]` (cit. on p. 8).

[30] K. Agashe and R. Contino, *Composite Higgs-Mediated FCNC*, Phys. Rev. D **80** (2009) 075016, arXiv: `0906.1542 [hep-ph]` (cit. on pp. 8, 9).

[31] A. Azatov, M. Toharia and L. Zhu, *Higgs Mediated FCNC's in Warped Extra Dimensions*, Phys. Rev. D **80** (2009) 035016, arXiv: `0906.1990 [hep-ph]` (cit. on pp. 8, 9).

[32] G. Perez and L. Randall, *Natural Neutrino Masses and Mixings from Warped Geometry*, JHEP **01** (2009) 077, arXiv: `0805.4652 [hep-ph]` (cit. on pp. 8, 9).

[33] M. E. Albrecht, M. Blanke, A. J. Buras, B. Duling and K. Gemmler, *Electroweak and Flavour Structure of a Warped Extra Dimension with Custodial Protection*, JHEP **09** (2009) 064, arXiv: `0903.2415 [hep-ph]` (cit. on pp. 8, 9).

[34] A. Falkowski, D. M. Straub and A. Vicente, *Vector-like leptons: Higgs decays and collider phenomenology*, JHEP **05** (2014) 092, arXiv: `1312.5329 [hep-ph]` (cit. on pp. 8–10).

[35] G. D'Ambrosio, G. F. Giudice, G. Isidori and A. Strumia, *Minimal flavour violation: an effective field theory approach*, Nuclear Physics B **645** (2002) 155, eprint: `hep-ph/0207036` (cit. on p. 8).

[36] A. Dery, A. Efrati, Y. Nir, Y. Soreq and V. Susic, *Model building for flavor changing Higgs couplings*, Phys. Rev. D **90** (11 2014) 115022, URL: `https://link.aps.org/doi/10.1103/PhysRevD.90.115022` (cit. on p. 8).

[37] D. Aloni, Y. Nir and E. Stamou, *Large BR($h \to \tau\mu$) in the MSSM?*, JHEP **04** (2016) 162, arXiv: 1511.00979 [hep-ph] (cit. on p. 8).

[38] H. Bélusca-Maïto, A. Falkowski, D. Fontes, J. C. Romão and J. P. Silva, *Higgs EFT for 2HDM and beyond*, European Physical Journal C **77**, 176 (2017) 176, arXiv: 1611.01112 [hep-ph] (cit. on p. 9).

[39] T. P. Cheng and M. Sher, *Mass Matrix Ansatz and Flavor Nonconservation in Models with Multiple Higgs Doublets*, Phys. Rev. **D35** (1987) 3484 (cit. on p. 9).

[40] O. Eberhardt et al., *Impact of a Higgs boson at a mass of 126 GeV on the standard model with three and four fermion generations*, Phys. Rev. Lett. **109** (2012) 241802, arXiv: 1209.1101 [hep-ph] (cit. on p. 9).

[41] MEG Collaboration, *Search for the lepton flavour violating decay $\mu^+ \to e^+\gamma$ with the full dataset of the MEG experiment*, Eur. Phys. J. **C76** (2016) 434, arXiv: 1605.05081 [hep-ex] (cit. on pp. 11, 12, 20).

[42] 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, arXiv: 2105.03007 [hep-ex] (cit. on p. 12).

[43] ATLAS Collaboration, *Searches for lepton-flavour-violating decays of the Higgs boson in $\sqrt{s} =$ 13 TeV pp collisions with the ATLAS detector*, (2019), arXiv: 1907.06131 [hep-ex] (cit. on p. 12).

[44] ATLAS Collaboration, *Search for lepton-flavor-violation in Z-boson decays with τ-leptons with the ATLAS detector*, Phys. Rev. Lett. **127** (2022) 271801, arXiv: 2105.12491 [hep-ex] (cit. on p. 12).

[45] ATLAS Collaboration, *Search for the charged-lepton-flavor-violating decay $Z \to e\mu$ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, (2022), arXiv: 2204.10783 [hep-ex] (cit. on p. 12).

[46] CMS Collaboration, *Search for Lepton Flavor Violation in Z decays in pp collisions at sqrt(s)=8 TeV*, (2015) (cit. on p. 12).

[47] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001 (cit. on p. 12).

[48] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on p. 13).

[49] ATLAS Collaboration, *ATLAS inner detector: Technical design report. Vol. 1*, (1997) (cit. on p. 13).

[50] ATLAS Collaboration, *ATLAS inner detector: Technical design report. Vol. 2*, (1997) (cit. on p. 13).

[51] ATLAS Collaboration, *ATLAS liquid argon calorimeter: Technical design report*, (1996) (cit. on p. 13).

[52] ATLAS Collaboration, *ATLAS tile calorimeter: Technical design report*, (1996) (cit. on p. 13).

[53] ATLAS Collaboration, *ATLAS muon spectrometer: Technical design report*, (1997) (cit. on p. 13).

[54] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, [Erratum: Eur.Phys.J.C 73, 2501 (2013)], arXiv: 1007.1727 [physics.data-an] (cit. on pp. 14–16, 77, 93, 94).

[55] A. L. Read, *Presentation of search results: the CL_S technique*, J. Phys. G **28** (2002) 2693 (cit. on p. 16).

[56] S. Bressler, A. Dery and A. Efrati, *Asymmetric lepton-flavor violating Higgs boson decays*, Phys. Rev. D **90** (2014) 015025, arXiv: 1405.4545 [hep-ph] (cit. on pp. 18, 90).

[57] ATLAS Collaboration, *Search for lepton-flavour-violating decays of the Higgs and Z bosons with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 70, arXiv: 1604.07730 [hep-ex] (cit. on pp. 18, 19, 23, 30, 39, 90, 91).

[58] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, European Physical Journal C **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det] (cit. on pp. 21, 34).

[59] D. J.Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. (2001) (cit. on p. 21).

[60] C. Oleari, *The Powheg Box*, Nuclear Physics B Proceedings Supplements **205** (2010) 36, arXiv: 1007.3893 [hep-ph] (cit. on pp. 21, 35).

[61] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph] (cit. on pp. 21, 35, 94, 107).

[62] S. Volkovich, F. De Vito Halevy and S. Bressler, *A data-directed paradigm for BSM searches: the bump-hunting example*, Eur. Phys. J. C **82** (2022) 265, arXiv: 2107.11573 [hep-ex] (cit. on pp. 33, 90).

[63] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, (2019), arXiv: 1911.04632 [hep-ex] (cit. on p. 34).

[64] ATLAS Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, (2019), arXiv: 1909.00761 [hep-ex] (cit. on pp. 34, 42).

[65] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, JINST **15** (2020) P09015, arXiv: 2004.13447 [physics.ins-det] (cit. on pp. 34, 42).

[66] ATLAS Collaboration, *The ATLAS inner detector trigger performance in pp collisions at 13 TeV during LHC Run 2*, Eur. Phys. J. C **82** (2022) 206, arXiv: 2107.02485 [hep-ex] (cit. on pp. 34, 42).

[67] ATLAS Collaboration, *Performance of the ATLAS Level-1 topological trigger in Run 2*, Eur. Phys. J. C **82** (2022) 7, arXiv: 2105.01416 [hep-ex] (cit. on pp. 34, 42).

[68] S. A. et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250, ISSN: 0168-9002, URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688 (cit. on p. 34).

[69] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, arXiv: 1503.07589 [hep-ex] (cit. on p. 35).

[70] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, **2/2017** (2016), arXiv: 1610.07922 [hep-ph] (cit. on pp. 35, 80).

[71] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, JINST **14** (2019) P12006, arXiv: 1908.00005 [hep-ex] (cit. on pp. 35, 36, 41, 42, 48, 49, 110, 144).

[72] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex] (cit. on p. 36).

[73] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 36).

[74] 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, 2017, URL: https://cds.cern.ch/record/2261772 (cit. on p. 36).

[75] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, ATL-PHYS-PUB-2019-033, 2019, URL: https://cds.cern.ch/record/2688062 (cit. on p. 36).

[76] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 36).

[77] ATLAS Collaboration, *Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, submitted to EPJC (2020), arXiv: 2007.02645 [hep-ex] (cit. on pp. 36, 78).

[78] ATLAS Collaboration, *Forward jet vertex tagging using the particle flow algorithm*, ATL-PHYS-PUB-2019-026, 2019, URL: https://cds.cern.ch/record/2683100 (cit. on p. 36).

[79] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **79** (2019) 970, arXiv: 1907.05120 [hep-ex] (cit. on p. 37).

[80] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **78** (2018) 903, arXiv: 1802.08168 [hep-ex] (cit. on pp. 35, 37, 78).

[81] ATLAS Collaboration, *E_T^{miss} performance in the ATLAS detector using 2015–2016 LHC pp collisions*, ATLAS-CONF-2018-023, 2018, URL: https://cds.cern.ch/record/2625233 (cit. on p. 37).

[82] 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, arXiv: 2012.00578 [hep-ex] (cit. on pp. 41, 42).

[83] ATLAS Collaboration, *Measurement of the production cross section for a Higgs boson in association with a vector boson in the H -> WW* -> lvlv channel in pp collisions at s=13 TeV with the ATLAS detector*, Physics Letters B **798** (2019) 134949, ISSN: 0370-2693, URL: http://dx.doi.org/10.1016/j.physletb.2019.134949 (cit. on p. 56).

[84] F. Chollet et al., *Keras*, https://keras.io, 2015 (cit. on pp. 77, 99).

[85] M. Abadi et al., 'TensorFlow: A System for Large-Scale Machine Learning.', *OSDI*, vol. 16, 2016 265 (cit. on pp. 77, 99).

[86] 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 p. 77).

[87] A. Elagin, P. Murat, A. Pranko and A. Safonov, *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, Nucl. Instrum. Meth. **A654** (2011) 481, arXiv: 1012.4686 [hep-ex] (cit. on p. 77).

[88] S. Davidson, S. Lacroix and P. Verdier, *LHC sensitivity to lepton flavour violating Z boson decays*, JHEP **09** (2012) 092, arXiv: 1207.4894 [hep-ph] (cit. on p. 77).

[89] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV*, Eur. Phys. J. **C76** (2016) 292, arXiv: 1603.05598 [hep-ex] (cit. on p. 78).

[90] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data*, JINST **14** (2019) P03017, arXiv: 1812.03848 [hep-ex] (cit. on p. 78).

[91] 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*, Eur. Phys. J. C **79** (2019) 639, arXiv: 1902.04655 [hep-ex] (cit. on p. 78).

[92] ATLAS Collaboration, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 580, arXiv: 1705.02211 [hep-ex] (cit. on p. 78).

[93] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002, arXiv: 1703.09665 [hep-ex] (cit. on p. 78).

[94] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017 (cit. on p. 78).

[95] S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, (2011), arXiv: 1101.0593 [hep-ph] (cit. on p. 80).

[96] I. W. Stewart, F. J. Tackmann, J. R. Walsh and S. Zuberi, *Jet p_T resummation in Higgs production at $NNLL' + NNLO$*, Phys. Rev. D **89** (2014) 054001, arXiv: 1307.1808 [hep-ph] (cit. on p. 80).

[97] X. Liu and F. Petriello, *Reducing theoretical uncertainties for exclusive Higgs-boson plus one-jet production at the LHC*, Phys. Rev. D **87** (2013) 094027, arXiv: 1303.4405 [hep-ph] (cit. on p. 80).

[98] R. Boughezal, X. Liu, F. Petriello, F. J. Tackmann and J. R. Walsh, *Combining Resummed Higgs Predictions Across Jet Bins*, Phys. Rev. D **89** (2014) 074044, arXiv: 1312.4535 [hep-ph] (cit. on p. 80).

[99] I. W. Stewart and F. J. Tackmann, *Theory Uncertainties for Higgs and Other Searches Using Jet Bins*, Phys. Rev. D **85** (2012) 034011, arXiv: 1107.2117 [hep-ph] (cit. on p. 80).

[100] S. Gangal and F. J. Tackmann, *Next-to-leading-order uncertainties in Higgs+2 jets from gluon fusion*, Phys. Rev. D **87** (2013) 093008, arXiv: 1302.5437 [hep-ph] (cit. on p. 80).

[101] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph] (cit. on p. 80).

[102] ATLAS Collaboration, *A strategy for a general search for new phenomena using data-derived signal regions and its application within the ATLAS experiment*, Eur. Phys. J. C **79** (2019) 120, arXiv: 1807.07447 [hep-ex] (cit. on pp. 90, 99).

[103] CMS Collaboration, *MUSiC: a model-unspecific search for new physics in proton–proton collisions at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **81** (2021) 629, arXiv: 2010.02984 [hep-ex] (cit. on pp. 90, 99).

[104] CDF Collaboration, *Global Search for New Physics with 2.0 fb^{-1} at CDF*, Phys. Rev. D **79** (2009) 011101, arXiv: 0809.3781 [hep-ex] (cit. on pp. 90, 99).

[105] C. G. Lester and B. H. Brunt, *Difference between two species of emu hides a test for lepton flavour violation*, JHEP **03** (2017) 149, [Erratum: JHEP 08, 069 (2017), Erratum: JHEP 06, 014 (2019)], arXiv: 1612.02697 [hep-ph] (cit. on pp. 90, 91).

[106] ATLAS Collaboration, *A search for an unexpected asymmetry in the production of $e^+\mu^-$ and $e^-\mu^+$ pairs in proton-proton collisions recorded by the ATLAS detector at $\sqrt{s} = 13$ TeV*, (2021), arXiv: 2112.08090 [hep-ex] (cit. on pp. 90, 91).

[107] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*, Eur. Phys. J. C **70** (2010) 525, arXiv: 1005.1891 [physics.data-an] (cit. on p. 91).

[108] N. V. Smirnov, *On the estimation of the discrepancy between empirical curves of distribution for two independent samples*, Bull. Math. Univ. Moscou **2** (1939) 3 (cit. on p. 91).

[109] Student, *The probable error of a mean*, Biometrika (1908) 1 (cit. on p. 91).

[110] J. H. Collins, K. Howe and B. Nachman, *Anomaly Detection for Resonant New Physics with Machine Learning*, Phys. Rev. Lett. **121** (2018) 241803, arXiv: 1805.02664 [hep-ph] (cit. on pp. 91, 98).

[111] J. H. Collins, K. Howe and B. Nachman, *Extending the search for new resonances with machine learning*, Phys. Rev. D **99** (2019) 014038, arXiv: 1902.02634 [hep-ph] (cit. on pp. 91, 98).

[112] E. M. Metodiev, B. Nachman and J. Thaler, *Classification without labels: Learning from mixed samples in high energy physics*, JHEP **10** (2017) 174, arXiv: 1708.02949 [hep-ph] (cit. on p. 91).

[113] R. T. D'Agnolo and A. Wulzer, *Learning New Physics from a Machine*, Phys. Rev. D **99** (2019) 015014, arXiv: 1806.02350 [hep-ph] (cit. on pp. 91, 98, 99).

[114] R. T. D'Agnolo, G. Grosso, M. Pierini, A. Wulzer and M. Zanetti, *Learning multivariate new physics*, Eur. Phys. J. C **81** (2021) 89, arXiv: 1912.12155 [hep-ph] (cit. on pp. 91, 98, 99).

[115] O. Amram and C. M. Suarez, *Tag N' Train: a technique to train improved classifiers on unlabeled data*, JHEP **01** (2021) 153, arXiv: 2002.12376 [hep-ph] (cit. on pp. 91, 98).

[116] ATLAS Collaboration, *Dijet resonance search with weak supervision using $\sqrt{s} = 13$ TeV pp collisions in the ATLAS detector*, Phys. Rev. Lett. **125** (2020) 131801, arXiv: 2005.02983 [hep-ex] (cit. on pp. 91, 98).

[117] B. Nachman and D. Shih, *Anomaly Detection with Density Estimation*, Phys. Rev. D **101** (2020) 075042, arXiv: 2001.04990 [hep-ph] (cit. on pp. 91, 98).

[118] A. Andreassen, B. Nachman and D. Shih, *Simulation Assisted Likelihood-free Anomaly Detection*, Phys. Rev. D **101** (2020) 095004, arXiv: 2001.05001 [hep-ph] (cit. on pp. 91, 98).

[119] K. Benkendorfer, L. L. Pottier and B. Nachman, *Simulation-assisted decorrelation for resonant anomaly detection*, Phys. Rev. D **104** (2021) 035003, arXiv: 2009.02205 [hep-ph] (cit. on pp. 91, 98).

[120] A. Hallin et al., *Classifying Anomalies THrough Outer Density Estimation (CATHODE)*, (2021), arXiv: 2109.00546 [hep-ph] (cit. on pp. 91, 98).

[121] R. T. d'Agnolo, G. Grosso, M. Pierini, A. Wulzer and M. Zanetti, *Learning New Physics from an Imperfect Machine*, (2021), arXiv: 2111.13633 [hep-ph] (cit. on pp. 91, 98, 99).

[122] B. Nachman, *Anomaly Detection for Physics Analysis and Less than Supervised Learning*, (2020), arXiv: 2010.14554 [hep-ph] (cit. on p. 91).

[123] G. Karagiorgi, G. Kasieczka, S. Kravitz, B. Nachman, D. Shih, *Machine Learning in the Search for New Fundamental Physics*, (2021), arXiv: 2112.03769 [hep-ph] (cit. on p. 91).

[124] M. Feickert and B. Nachman, *A Living Review of Machine Learning for Particle Physics*, (2021), arXiv: 2102.02770 [hep-ph] (cit. on p. 91).

[125] G. Kasieczka et al., *The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics*, Rept. Prog. Phys. **84** (2021) 124201, arXiv: 2101.08320 [hep-ph] (cit. on p. 91).

[126] T. Aarrestad et al., *The Dark Machines Anomaly Score Challenge: Benchmark Data and Model Independent Event Classification for the Large Hadron Collider*, SciPost Phys. **12** (2022) 043, arXiv: 2105.14027 [hep-ph] (cit. on p. 91).

[127] J. Neyman and E. S. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, Phil. Trans. R. Soc. Lond. A **231** (1933) 289 (cit. on pp. 92, 99).

[128] 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*, JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph] (cit. on pp. 94, 107).

[129] DELPHES 3 Collaboration, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, JHEP **02** (2014) 057, arXiv: 1307.6346 [hep-ex] (cit. on pp. 94, 107).

[130] P. Chakravarti, M. Kuusela, J. Lei and L. Wasserman, *Model-Independent Detection of New Physics Signals Using Interpretable Semi-Supervised Classifier Tests*, (2021), arXiv: 2102.07679 [stat.AP] (cit. on p. 98).

[131] A. Bogatskiy et al., *Lorentz Group Equivariant Neural Network for Particle Physics*, (2020), arXiv: 2006.04780 [hep-ph] (cit. on p. 98).

[132] S. Gong et al., *An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging*, (2022), arXiv: 2201.08187 [hep-ph] (cit. on p. 98).

[133] S. Qiu, S. Han, X. Ju, B. Nachman and H. Wang, *A Holistic Approach to Predicting Top Quark Kinematic Properties with the Covariant Particle Transformer*, (2022), arXiv: 2203.05687 [hep-ph] (cit. on p. 98).

[134] B. M. Dillon et al., *Symmetries, Safety, and Self-Supervision*, (2021), arXiv: 2108.04253 [hep-ph] (cit. on p. 98).

[135] G. Barenboim, J. Hirn and V. Sanz, *Symmetry meets AI*, SciPost Phys. **11** (2021) 014, arXiv: 2103.06115 [cs.LG] (cit. on p. 98).

[136] K. Desai, B. Nachman and J. Thaler, *SymmetryGAN: Symmetry Discovery with Deep Learning*, (2021), arXiv: 2112.05722 [hep-ph] (cit. on p. 98).

[137] S. Krippendorf and M. Syvaeri, *Detecting Symmetries with Neural Networks*, (2020), arXiv: 2003.13679 [physics.comp-ph] (cit. on p. 98).

[138] R. Tombs and C. G. Lester, *A method to challenge symmetries in data with self-supervised learning*, (2021), arXiv: 2111.05442 [hep-ph] (cit. on p. 98).

[139] C. G. Lester and R. Tombs, *Stressed GANs snag desserts, a.k.a Spotting Symmetry Violation with Symmetric Functions*, (2021), arXiv: 2111.00616 [hep-ph] (cit. on p. 98).

[140] D. Kingma and J. Ba, *Adam: A method for stochastic optimization*, (2014), arXiv: 1412.6980 [cs] (cit. on p. 99).

[141] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, Ann. Statist. **7** (1979) 1 (cit. on p. 99).

[142] B. Nachman and J. Thaler, *Learning from many collider events at once*, Phys. Rev. D **103** (2021) 116013, arXiv: 2101.07263 [physics.data-an] (cit. on p. 99).

[143] D0 Collaboration, *Search for new physics in eμX data at DØ using SLEUTH: A quasi-model-independent search strategy for new physics*, Phys. Rev. D **62** (2000) 092004, arXiv: hep-ex/0006011 (cit. on p. 99).

[144] H1 Collaboration, *A General Search for New Phenomena at HERA*, Phys. Lett. **B674** (2009) 257, arXiv: 0901.0507 [hep-ex] (cit. on p. 99).

[145] Y. Afik, J. Cohen, E. Gozani, E. Kajomovitz and Y. Rozen, *Establishing a Search for $b \to s\ell^+\ell^-$ Anomalies at the LHC*, JHEP **08** (2018) 056, arXiv: 1805.11402 [hep-ph] (cit. on p. 102).

[146] ATLAS Collaboration, *Search for New Phenomena in Final States with Two Leptons and One or No b-Tagged Jets at $\sqrt{s} = 13$ TeV Using the ATLAS Detector*, Phys. Rev. Lett. **127** (2021) 141801, arXiv: 2105.13847 [hep-ex] (cit. on p. 102).

[147] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph] (cit. on p. 108).

[148] M. L. Mangano, M. Moretti, F. Piccinini and M. Treccani, *Matching matrix elements and shower evolution for top-quark production in hadronic collisions*, JHEP **01** (2007) 013, arXiv: hep-ph/0611129 (cit. on p. 108).

[149] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP **03** (2013) 015, arXiv: 1212.3460 [hep-ph] (cit. on p. 108).

Appendices

A ATLAS Higgs LFV search - supplementary electron efficiency measurements

The studies presented here complement the discussion in section 4.4.

This section presents the following supplementary electron MC efficiencies measurements:

- Comparison of efficiencies from DAOD_EGAM1 $Z/\gamma^* \to ee$ samples to reference values obtained from the EGamma performance group (appendix A.1)
- Comparison of leading and subleading electron efficiencies from DAOD_EGAM1 $Z/\gamma^* \to ee$ samples (appendix A.2)
- Comparison of efficiencies from DAOD_EGAM1 $Z/\gamma^* \to ee$ samples to those from DAOD_TOPQ1 used in our analysis (appendix A.3)
- Electron MC efficiency measurements – and validation of the efficiency correction in MC – in alternative selections (appendix A.4)

In the following, we refer to efficiencies we measure from DAOD_EGAM1 $Z/\gamma^* \to ee$ samples as *Zee* efficiencies and to efficiencies used for our analysis as *LFV* efficiencies.

A.1 Zee efficiencies compared to reference values

We compare our Zee efficiency measurements to reference values from the EGamma performance group. The measurements performed by EGamma are described in [71]. The method used to perform our measurements is described in section 4.4.1.

We list here some comments relevant to all the plots shown in this section:

- We don't provide efficiency values in the crack region ($1.37 < |\eta| < 1.52$)
- Only p_T bins up to 150 GeV are shown since the binning can vary from map to map in higher ranges
- Only statistical uncertainties are included in our Zee efficiency measurements
- The uncertainties on the reference values correspond to the TOTAL systematic uncertainty model
- We use newer datasets (p4323) for our measurements than those used for the reference values (p3918), and the AnalysisRelease is also probably different
- We use Medium Id electrons as tags instead of Tight Id for the reference nominal values
- Not all mc campaigns are shown per working point depending on whether we obtained the relevant reference values or not
- For mc16d Reco efficiencies and Id efficiencies measured with the Zmass method, the reference values are, in fact, measured from mc16c samples with the 2017 pileup profile

These can explain the residual disagreements observed between our Zee measurements and the reference values.

Reconstruction efficiencies

As detailed in section 4.4.1, the Reco efficiencies include in the denominator all reconstructed clusters in the EMCal, which are truth-matched to electrons, while only those that are reconstructed as electrons and pass the TrackQuality criteria are included in the denominator.

Figure 73 compares our Zee Reco efficiencies and the reference values. Although our measurements tend to be slightly (up to 1%) higher, the disagreement is within uncertainties.

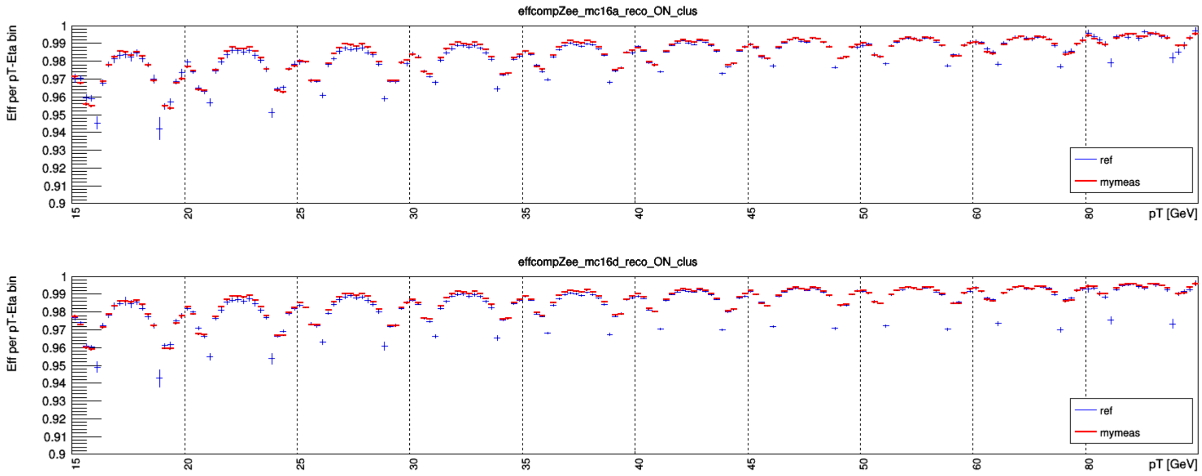


Figure 73: Comparison of our Zee Reco efficiencies with reference values from EGamma for mc campaigns mc16a and mc16d.

Identification efficiencies

For Id efficiencies, the denominator is filled with reconstructed electron probes, while only those which pass the Medium Id requirement are included in the denominator. For mc16d, reference values from both the Zmass and the Ziso measurement methods are shown.

Figure 74 compares our Zee Id efficiencies and the reference values. Although our measurements tend to be slightly (up to 1.5%) lower in the lower p_T range, the disagreement is almost always within uncertainties.

Isolation efficiencies

For Iso efficiencies, the denominator is filled with Medium Id electron probes, while only those which pass the Gradient Iso requirement are included in the denominator.

Figure 75 compares our Zee Id efficiencies and the reference values. We see that the agreement is always within uncertainties.

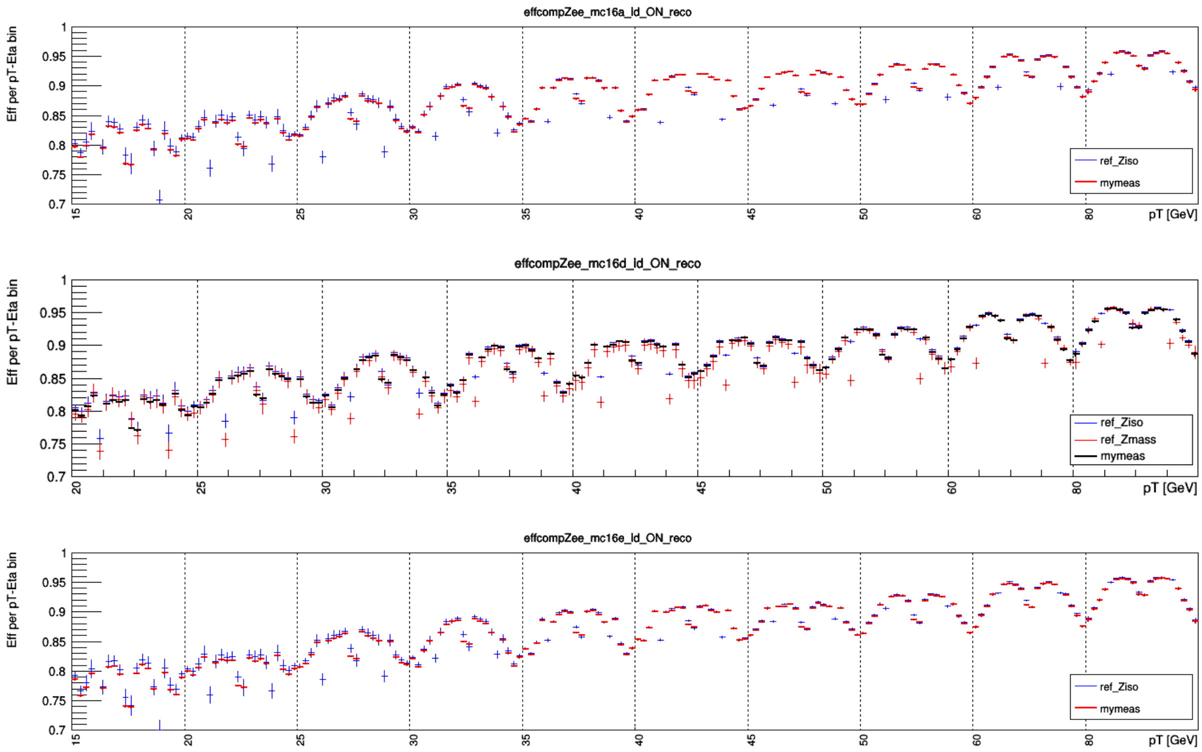


Figure 74: Comparison of our Zee Id efficiencies with reference values from EGamma for mc campaigns mc16a, mc16d, and mc16e.

A.2 Zee leading vs. subleading electron efficiencies

In section 4.4.3, Figure 17, we showed that the electron efficiencies measured for our analysis could differ whether the electron probes used in the measurement are the leading or subleading leptons in the event. Here we repeat this comparison using this time the Zee efficiencies.

Figure 76 compares the Zee efficiency measurements for leading or subleading electron probes per working point. The plots shown use combined data-taking years and $|\eta|$ bins. We can see that here as well, significant differences are observed, especially in the lower p_T range, with leading electron efficiencies values up to 5% higher for the combined (Reco*Id*Iso) electron efficiencies. The difference originates mainly from the Iso constraint but is also observed when requiring Id. On the other hand, the Reco efficiencies agree.

A.3 LFV vs. Zee electron efficiencies

As detailed in section 4.3.4, the electron efficiencies are not provided by EGamma since they were found to depend strongly on the kinematic selection and not only on the properties of the electrons. Here we compare our LFV to Zee efficiencies in order to show the differences observed in the measured values.

Figures 77-80 compare the LFV and Zee efficiencies for the different working points separately for leading and subleading electron probes. As expected, the values generally differ beyond uncertainties, with a value

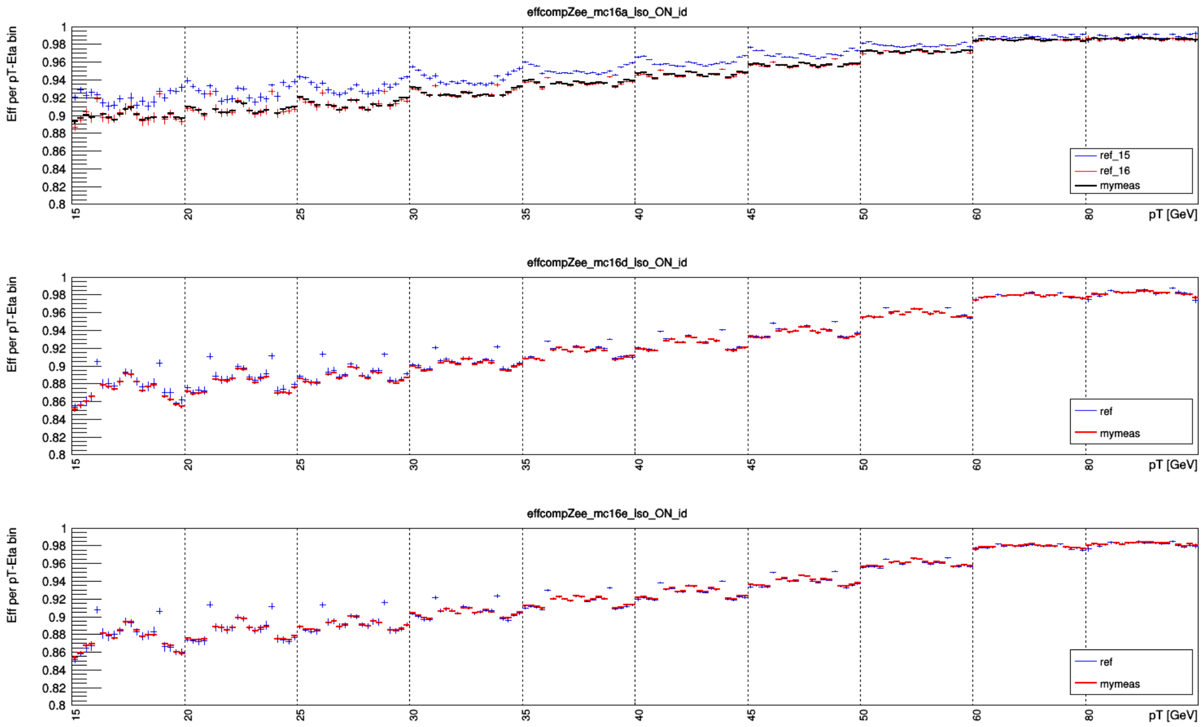


Figure 75: Comparison of our Zee Iso efficiencies with reference values from EGamma for mc campaigns mc16a, mc16d, and mc16e.

difference of up to 4-5% in specific bins for the combined electron efficiencies, mainly in the $p_{\mathrm{T}} < 25$ GeV and $p_{\mathrm{T}} > 45$ GeV ranges.

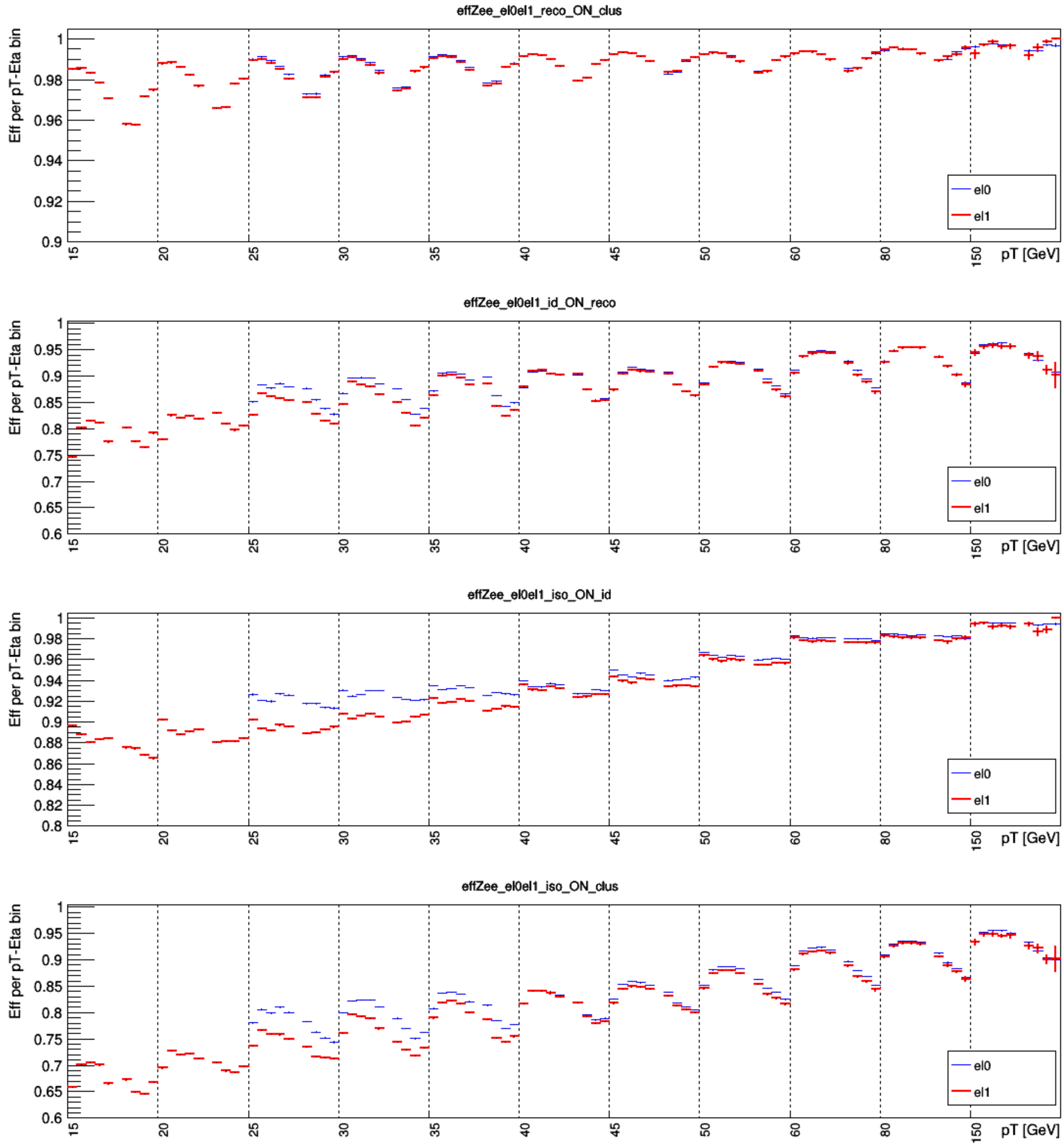


Figure 76: Comparison of Zee efficiencies between leading and subleading electron probes for the Reco, Id, Iso, and combined efficiency working points.

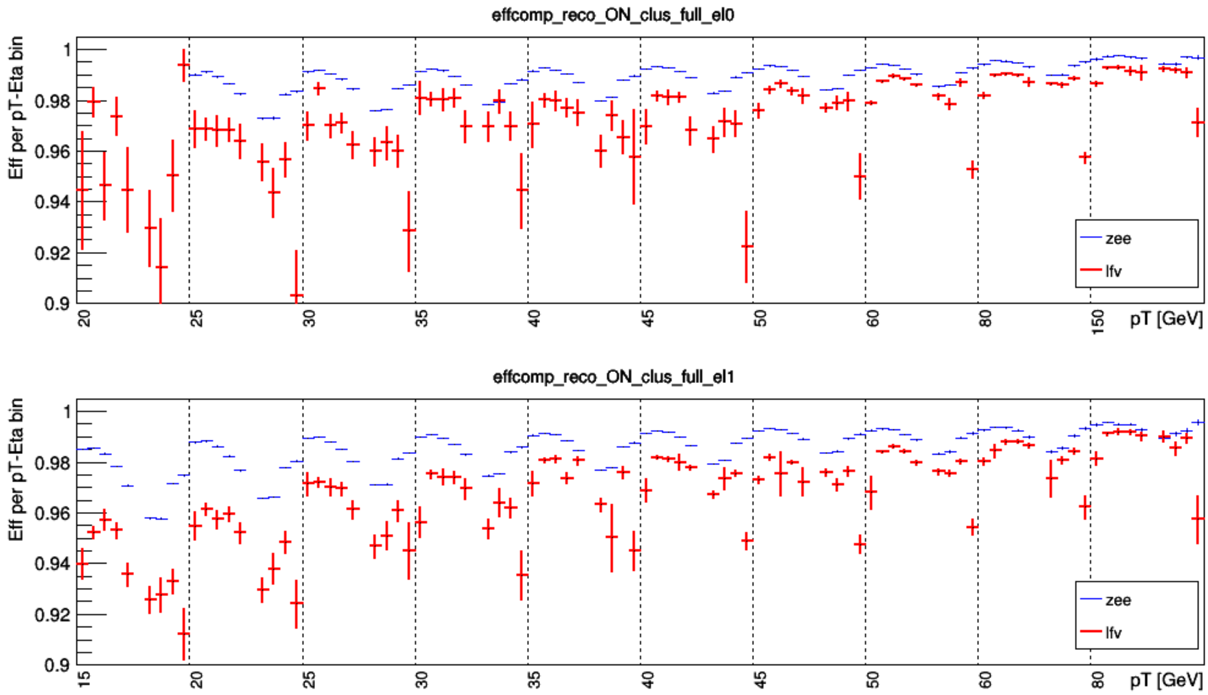


Figure 77: Comparison of LFV vs. Zee electron Reco efficiencies for leading and subleading electron probes.

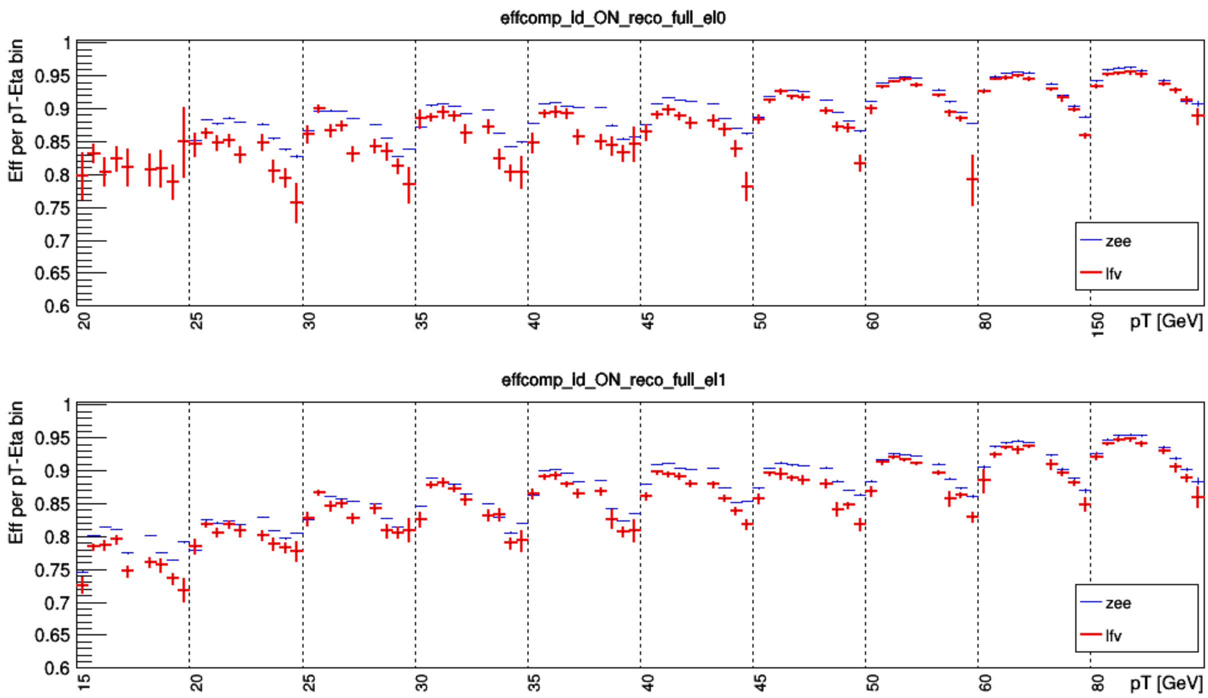


Figure 78: Comparison of LFV vs. Zee electron Id efficiencies for leading and subleading electron probes.

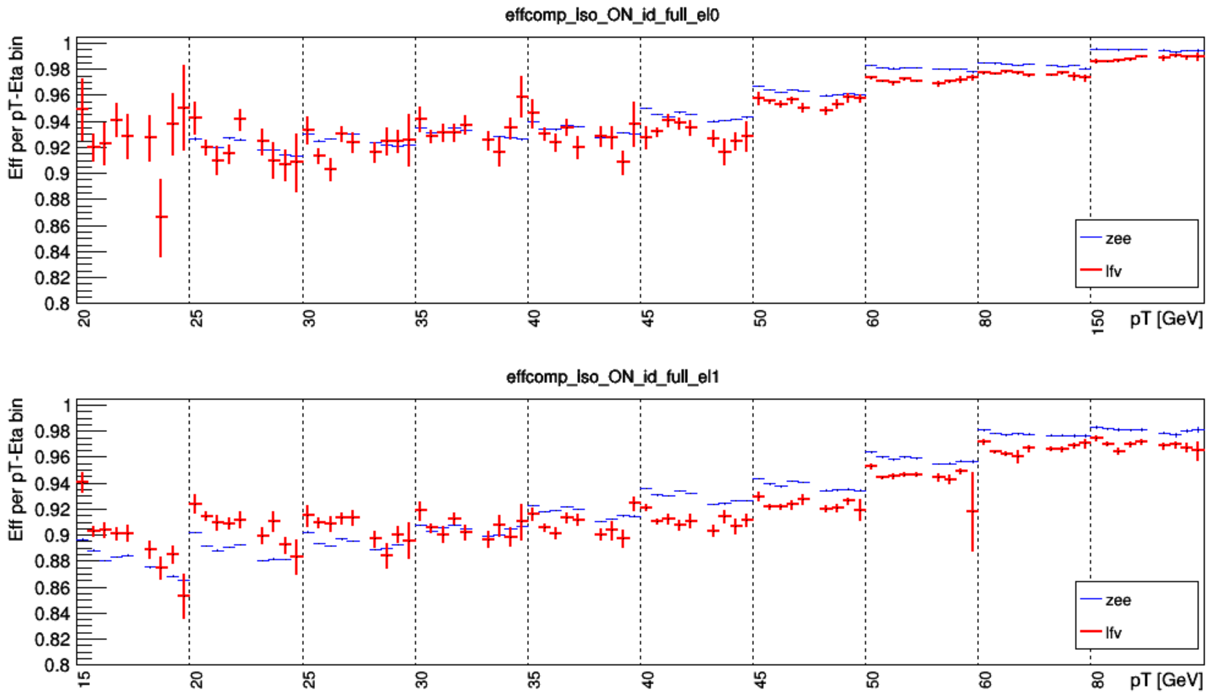


Figure 79: Comparison of LFV vs. Zee electron Iso efficiencies for leading and subleading electron probes.

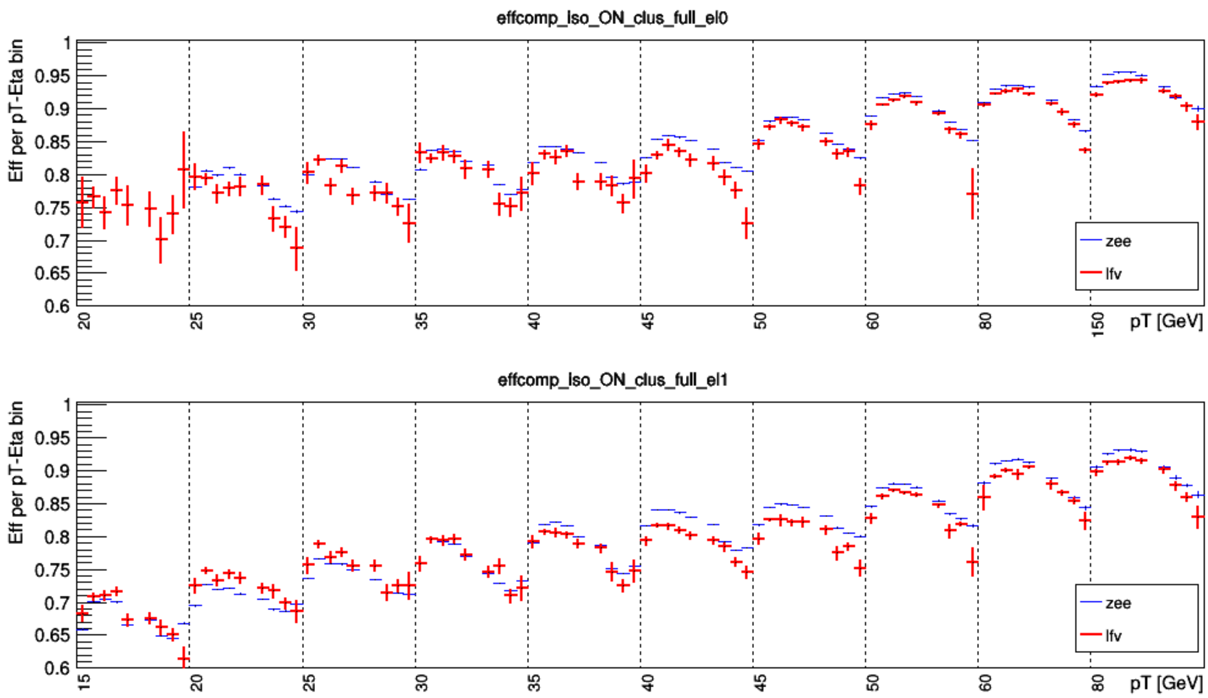


Figure 80: Comparison of LFV vs. Zee electron combined efficiencies for leading and subleading electron probes.

A.4 Alternative selections

Here we describe the measurements of electron MC efficiencies – and efficiency correction validation in MC – when applying two additional selection cuts $d0\text{sig.}(e) < 5$ and $0.2 < p_{\mathrm{T}}^{\mathrm{track}}/p_{\mathrm{T}}^{\mathrm{cluster}}(e) < 1.25$. In the following, we refer to these two cuts as $d05$ and ptr cuts, respectively.

As described in section 4.4.1, the electron MC efficiencies are measured based on NTuples produced from TOPQ1 derivations. But due to missing information in our TOPQ1-based NTuples, we can't apply there the ptr cut. Rather than reproducing the NTuples, we decided to measure the ptr efficiencies from the HIGG4D1-based NTuples, which we combine with the TOPQ1-based efficiencies. In addition, we validate this step by comparing efficiencies from TOPQ1 and HIGG4D1 at the electron working points that can be determined in both. We summarize the strategy used here based on the working points defined in section 4.4.1:

- Reco/Clus efficiency - Biased in HIGG4D1; use TOPQ1
- Base/Reco efficiency - Biased in HIGG4D1; use TOPQ1
- Id/Base efficiency - Compare HIGG4D1 vs. TOPQ1; use TOPQ1
- Iso/Id efficiency - Compare HIGG4D1 vs. TOPQ1; use TOPQ1
- d05/Iso efficiency - Compare HIGG4D1 vs. TOPQ1; use TOPQ1
- ptr/Iso or ptr/d05 efficiency - Missing in TOPQ1; use HIGG4D1

We stress that there is a difference in the efficiency measurements from the HIGG4D1 vs. TOPQ1-based NTuples. As detailed in sections 4.4.3 and 4.4.4, our TOPQ1-based efficiencies use both electron and muon tagged events, and we use the difference observed when using electron and muon tags to derive the *tag-flavor* systematic uncertainty on the measurement. But the HIGG4D1-based NTuples only include events with 1 electron and 1 muon, so only muon tags can be considered. As a result, the *tag-flavor* systematic uncertainty is based on the d05/Clus efficiencies instead of the ptr/Clus efficiencies (but the *selection* systematic uncertainty is correctly based on the ptr/Clus efficiencies). We decided to neglect this as we expect it to be a small effect; indeed, the electron MC efficiency systematics aren't ranked high in the fits.

Figure 81 shows the impact of the $d05$ and ptr cuts on the efficiencies (using signal electrons that pass the baseline selection in the denominator). We see that the $d05$ cut's impact is always under 10% and largest in the lower p_{T} range and the barrel region. The ptr cut affects mainly electrons in the endcap region, with an impact rising with p_{T} to almost 30%.

Figures 82 and 83 compare the measured efficiencies from TOPQ1 vs. HIGG4D1 for the working-points Id/Base, Iso/Id, and d05/Iso. Although the samples are different, which is made evident by the differences in statistical uncertainties, we still find good agreement between the two. There is a larger disagreement in the last p_{T} bin of the leading Id/Base efficiencies, but statistics are low, which isn't observed in the other working points. Therefore this validates that we can use the d05/Clus efficiencies from TOPQ1 combined with the ptr/d05 efficiencies from HIGG4D1.

In Figure 84, we show the overall efficiencies when applying both the $d05$ and ptr cuts, including the envelopes corresponding to the combined tag-flavor and selection systematic uncertainties.

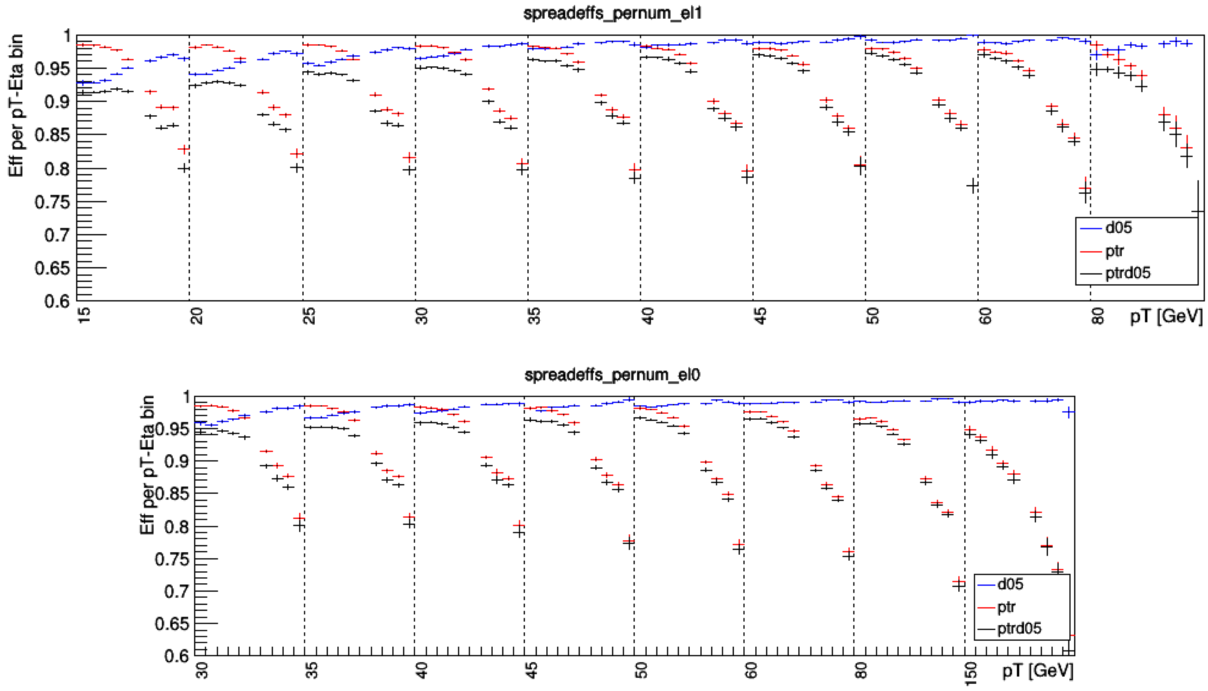


Figure 81: Electron MC efficiencies when applying the $d05$, ptr, and $d05 + ptr$ cuts to signal (Iso) electrons. The top is for subleading, and the bottom is for leading electrons. The $|\eta|$ bin efficiencies are displayed within each p_{T} bin.

In practice, depending on the alternative selection considered and the event, we need to apply either the ptr cut, the $d05$ cut, or both or none. So we also derived efficiency maps when only one of the cuts is applied and need to choose, per event, the corresponding map.

Similarly to section 4.6.1, we validate the efficiencies for the alternative selections by applying the efficiency correction to the SM MC sample with the corresponding selection and investigating the resulting symmetry between the $e\tau$ and $\mu\tau$ channels. Figure 85 shows this comparison for the case where the $d05$ cut is applied to the Higgs leptons, and the ptr cut is applied to the subleading electrons. The symmetry is successfully restored, agreeing within statistical uncertainties in most bins.

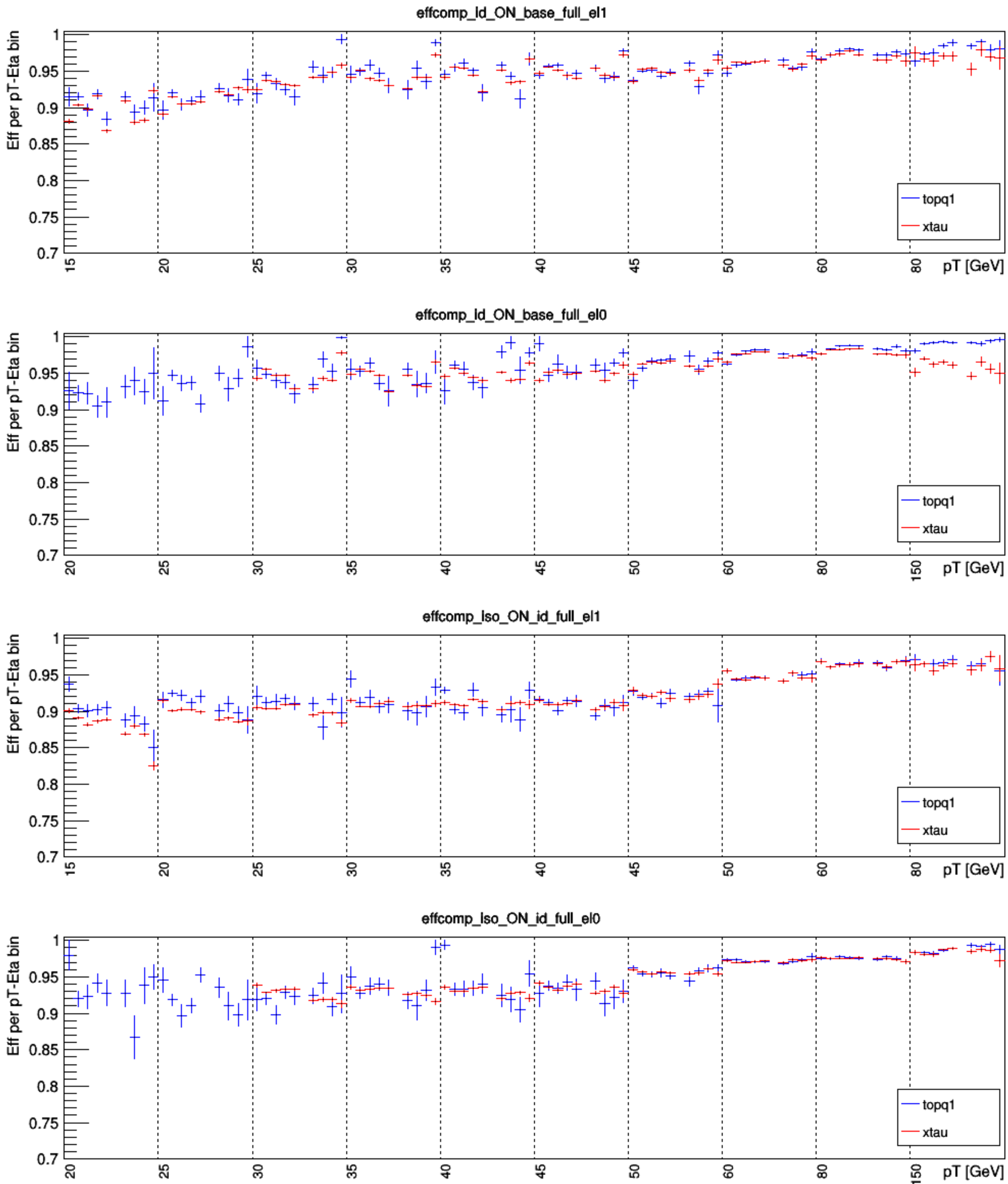


Figure 82: Comparing electron MC efficiencies measured from TOPQ1 and HIGG4D1 (labeled as *xtau*). We show in order subleading and leading electron Id/Base efficiencies, then subleading and leading electron Iso/Id efficiencies. The $|\eta|$ bin efficiencies are displayed within each p_{T} bin.

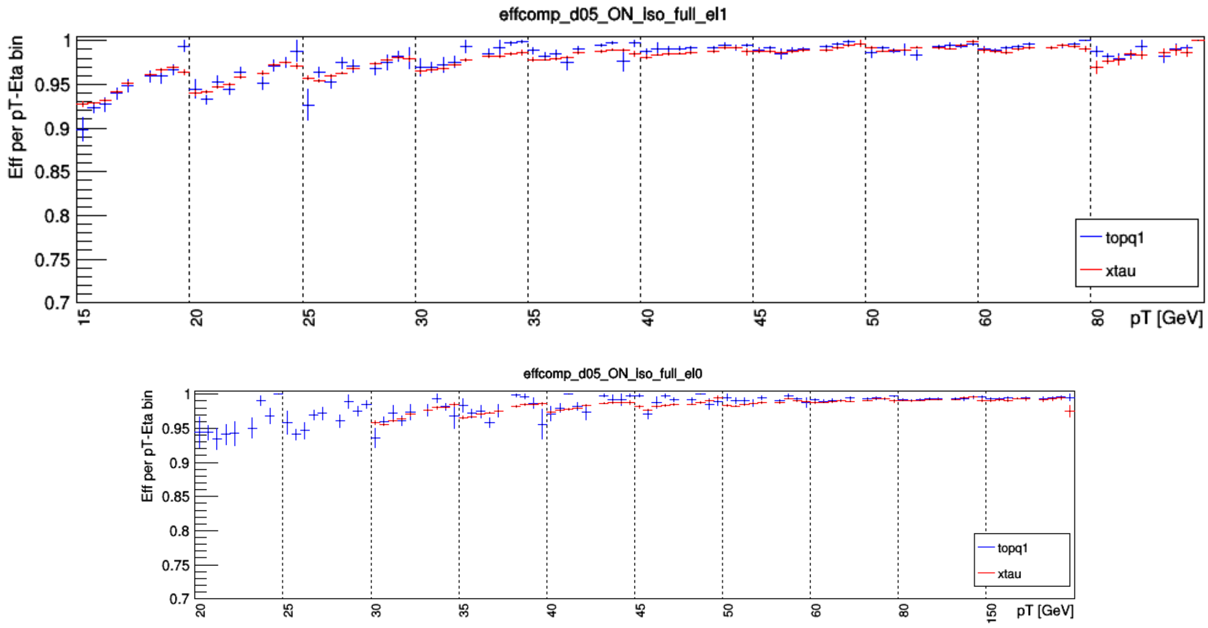


Figure 83: Same as above but for subleading and leading electron d05/Iso efficiencies.

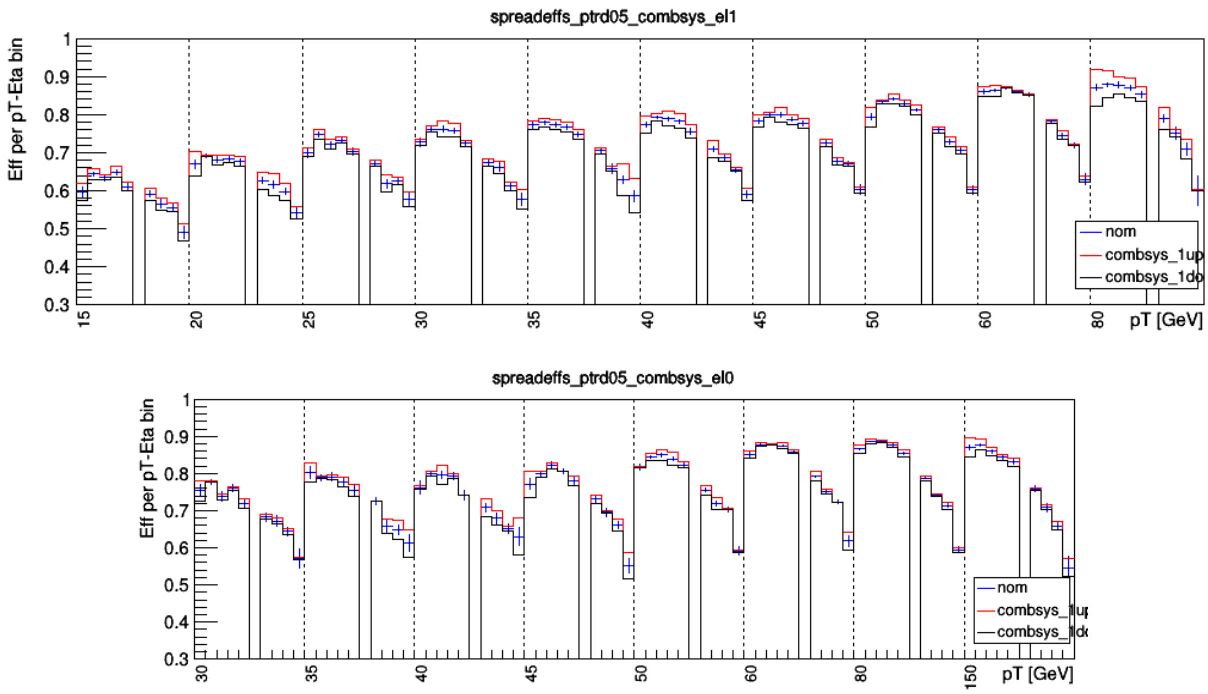


Figure 84: Electron MC efficiency measurements used for the alternative selection. Nominal values and statistical errors are in red, and upward and downward systematic variations are in blue and black. The top is for subleading, and the bottom is for leading electrons. The $|\eta|$ bin efficiencies are displayed within each p_{T} bin.

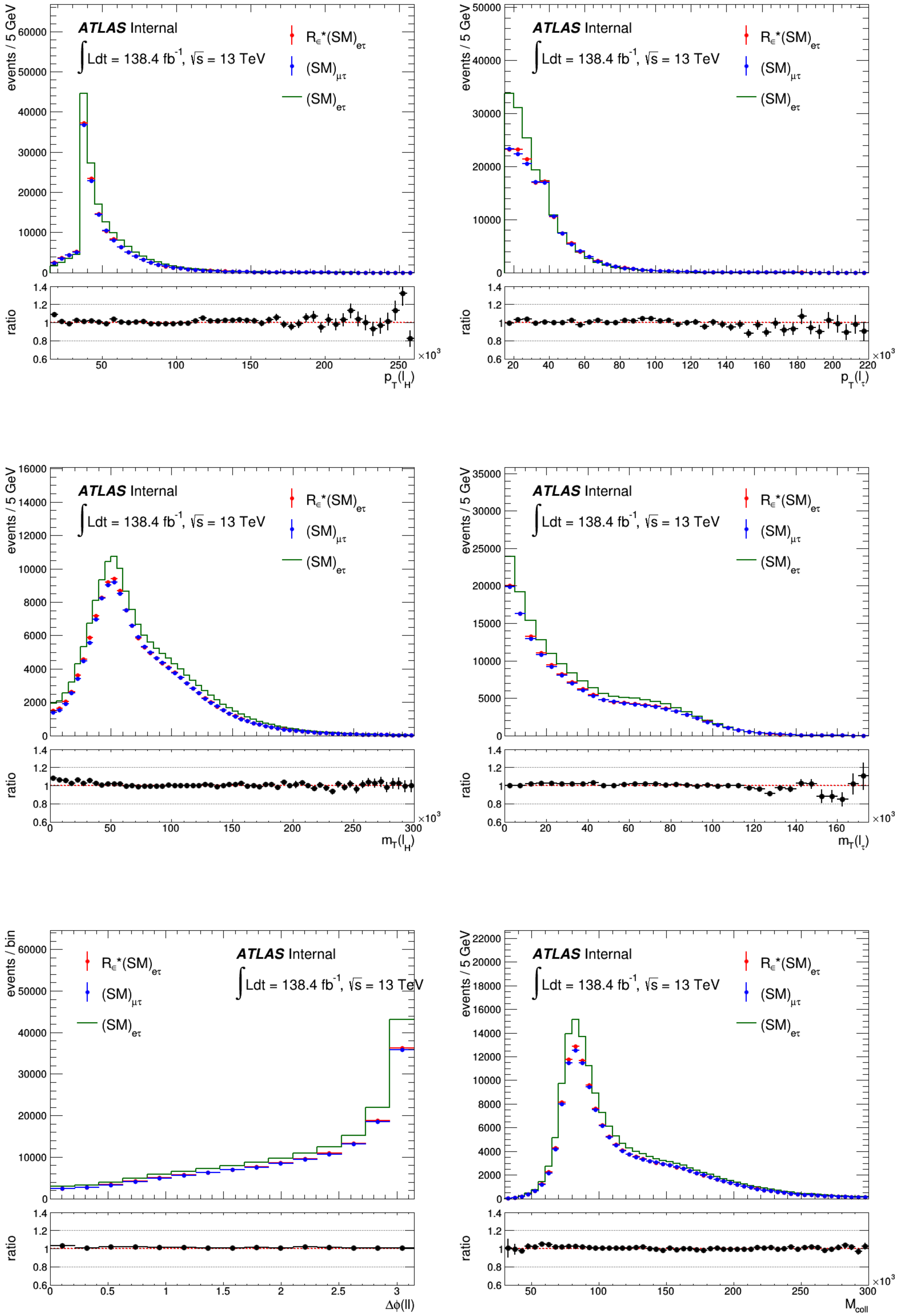


Figure 85: Comparison of efficiency-corrected $e\tau$ (red) over $\mu\tau$ (blue) events in the SM MC sample for the $p_{\mathrm{T}}(\ell_H)$, $p_{\mathrm{T}}(\ell_\tau)$, $m_{\mathrm{T}}(\ell_H, E_{\mathrm{T}}^{\mathrm{miss}})$, $m_{\mathrm{T}}(\ell_\tau, E_{\mathrm{T}}^{\mathrm{miss}})$, $\Delta\phi(\ell\ell)$ and M_{coll} distributions. The uncorrected $e\tau$ channel is also displayed (green).

B ATLAS Higgs LFV search - pre-fit symmetrization and smoothing

The studies presented here complement the discussion in section 4.9.1.

We compare fit results using different symmetrization and smoothing algorithms introduced in section 4.9.1. A description of the obtention of the results, the different inputs used, and the methods implemented to evaluate the fit performance is given in section 2.6.

In section B.1, we compare the following setups:

- MaxCorrect symmetrization and Parabolic smoothing (label *maxsym*)
- MeanCorrect symmetrization and Parabolic smoothing (label *meansym*)

The first setup was used in an older version of the analysis fits (labeled here as v0.3), while the second setup (labeled here as v0.4) was chosen for the obtention of final results. The difference lies in the symmetrization method, applied only to two-sided kinematic systematics, where either the mean or max value of |up| and |down| is used in both directions (see section 4.9.1 for more details). The main motivation for this change is that MaxCorrect seems to slightly overestimate the uncertainties in some cases. In contrast, the TwoSided symmetrization (considered in section B.2) seems to slightly underestimate them in some cases.

In section B.2, we compare our previous (v0.3) setup to the one used by our partner MC-based analysis (true at the time this study was completed). The MC-based symmetrization, applied to all systematics (kinematics and weights), uses TwoSided ((up-down)/2) for Jet, Muon, and Tau systematics; and MaxCorrect for the others but is applied only on bins that have up and down variations in the same direction. The MC-based smoothing method uses the MaxVariation algorithm. Explicitly, the setups compared are:

- MaxCorrect symmetrization and Parabolic smoothing (label *symSym_smoSym* - our v0.3 default)
- MaxCorrect symmetrization without smoothing (label *symSym_smoNo*)
- MaxCorrect symmetrization and MaxVariation smoothing (label *symSym_smoMC*)
- MC-based symmetrization and Parabolic smoothing (label *symMC_smoSym*)
- MC-based symmetrization and MaxVariation smoothing (label *symMC_smoMC*)

The comparison is made based on standalone fits of the Symmetry-based analysis. In B.1, only results with Asimov fits are shown; in B.2, mainly results with the mixed fits are shown. For each comparison, a few nuisance parameter envelope plots are displayed, selected based on two criteria:

- Choose nuisance parameters that have the largest difference of impact on POI between setups
- For each selected nuisance parameter, choose SR and sample envelopes that show the largest differences, based on a reduced chi-square test

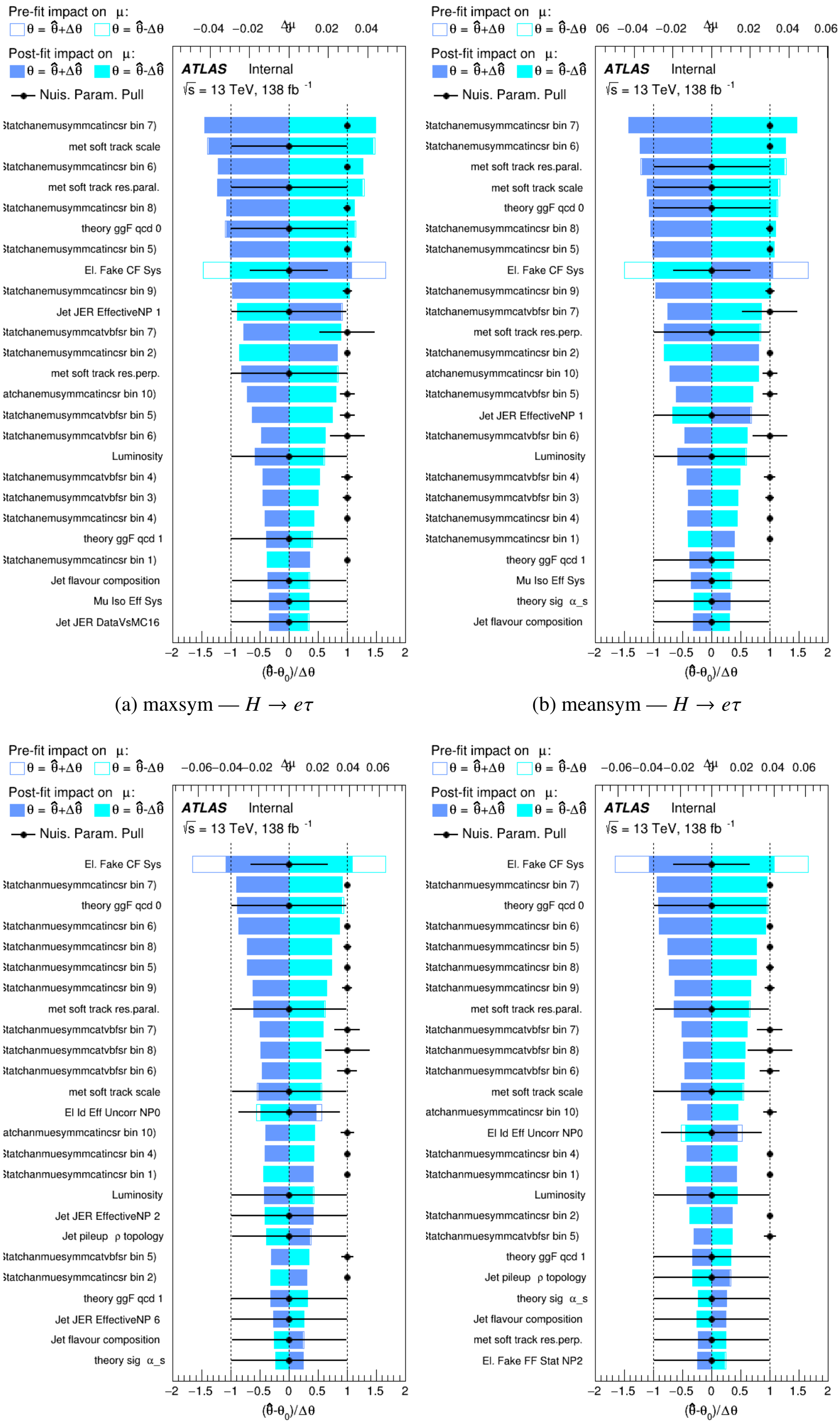
B.1 MaxCorrect vs. MeanCorrect symmetrization

Towards comparing the Asimov fits conducted with the two symmetrization options considered (MaxCorrect and MeanCorrect), Figure 86 shows the ranking of nuisance parameters in both cases, Figure 87 shows the pulls and constraints on the nuisance parameters, and Figures 88 and 89 show selected nuisance parameter envelopes. Table 22 displays the expected sensitivities.

Differences in the ranking and the constraints are minimal. Small differences are visible in the envelopes where the MeanCorrect symmetrization gives slightly less conservative uncertainties than the MaxCorrect symmetrization – as expected by construction. The effect on the expected sensitivities is small but in the direction of a slight improvement for MeanCorrect – again, as expected by the definition of this symmetrization option. The MeanCorrection symmetrization option for the nuisance parameters thus performs as intended and is therefore chosen for the final fit setup.

	maxsym	meansym
$e\tau$ limit/%	$0.193^{+0.077}_{-0.054}$	$0.189^{+0.075}_{-0.053}$
$e\tau$ signif.	8.471	8.686
$\mu\tau$ limit/%	$0.190^{+0.075}_{-0.053}$	$0.188^{+0.074}_{-0.053}$
$\mu\tau$ signif.	9.016	9.107

Table 22: Expected sensitivities comparing maxsym and meansym for the Asimov fit. The significance is given when assuming $BR = 1\%$.



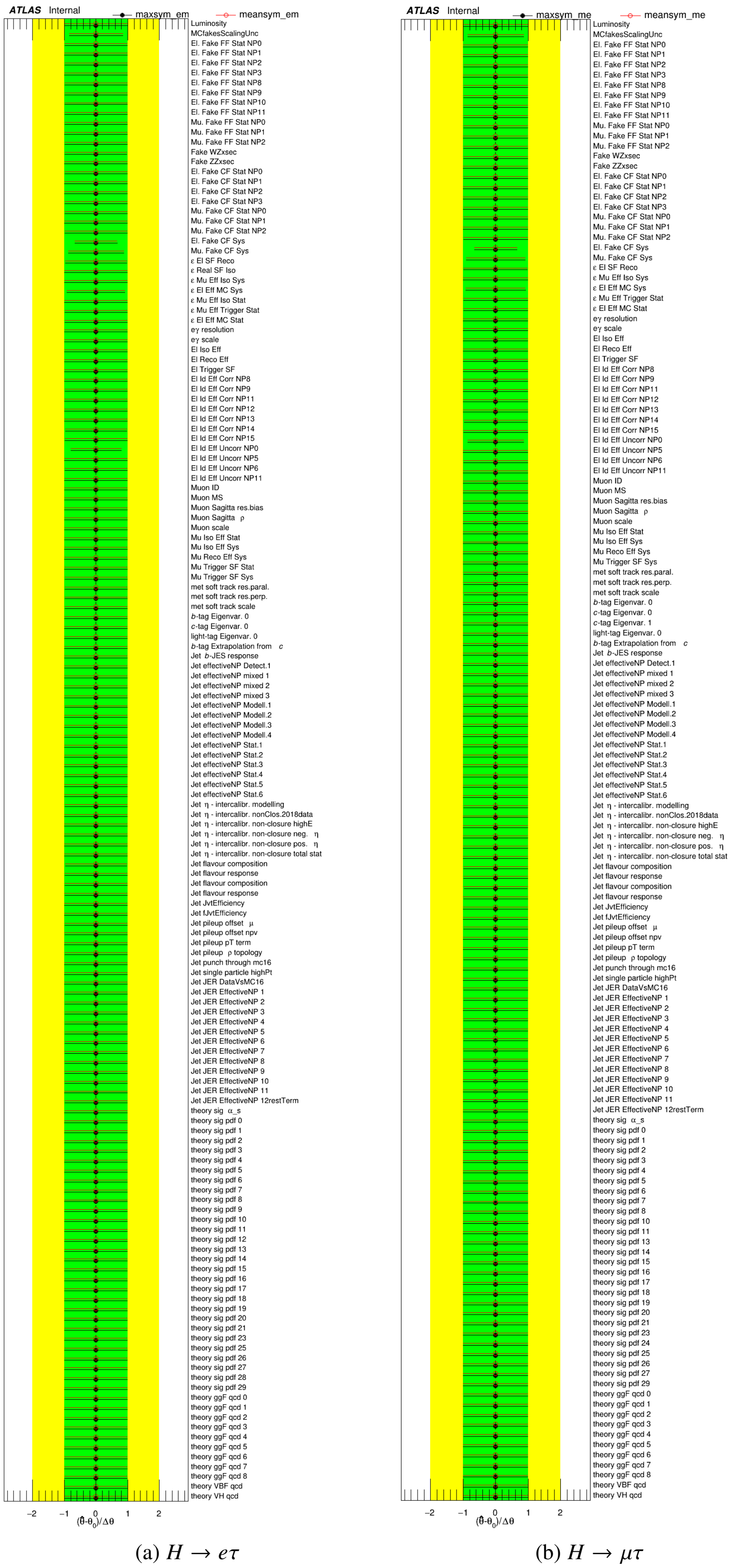
(a) maxsym — $H \to e\tau$

(b) meansym — $H \to e\tau$

(c) maxsym — $H \to \mu\tau$

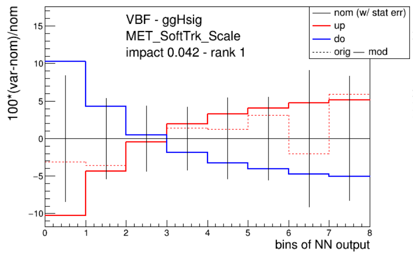
(d) meansym — $H \to \mu\tau$

Figure 86: Ranking of the nuisance parameters comparing the maxsym and meansym setups.

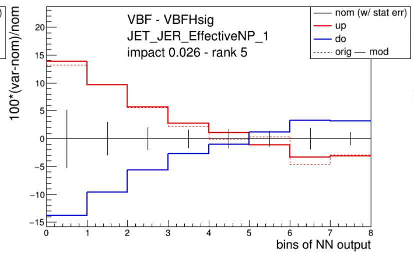


(a) $H \to e\tau$ (b) $H \to \mu\tau$

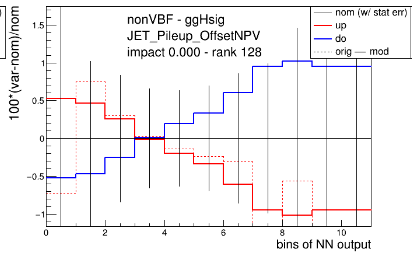
Figure 87: Pulls of the nuisance parameters comparing the maxsym and meansym setups.



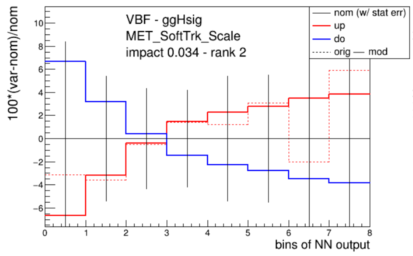
(a) maxsym - NP1



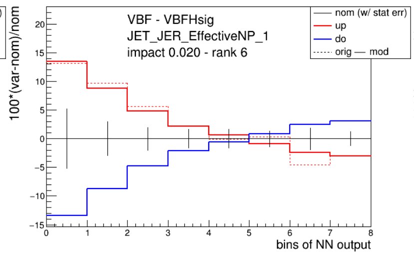
(b) maxsym - NP2



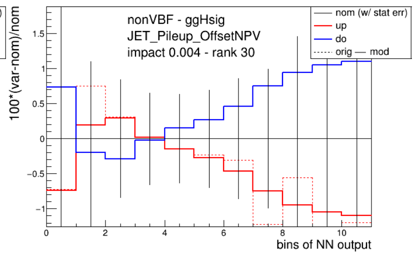
(c) maxsym - NP3



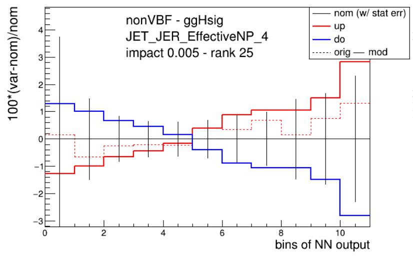
(d) meansym - NP1



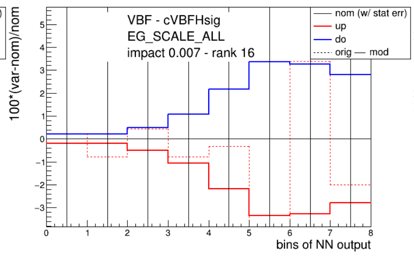
(e) meansym - NP2



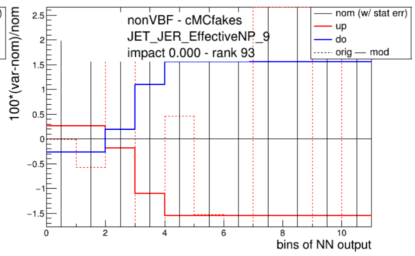
(f) meansym - NP3



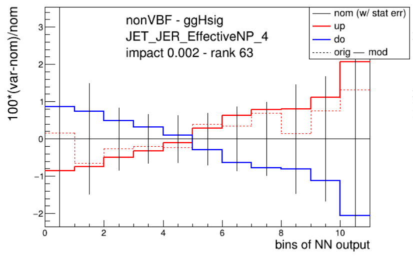
(g) maxsym - NP4



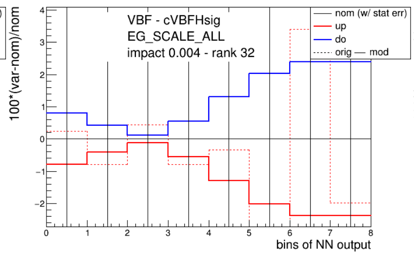
(h) maxsym - NP5



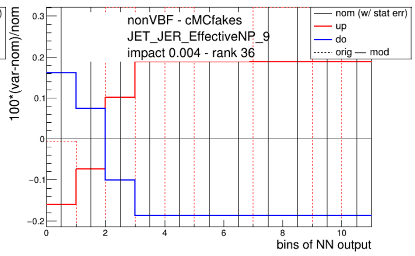
(i) maxsym - NP6



(j) meansym - NP4

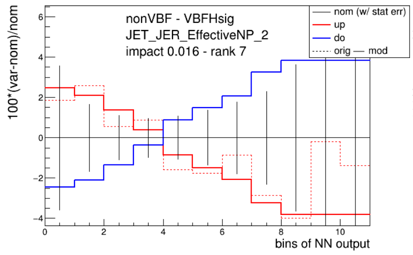


(k) meansym - NP5

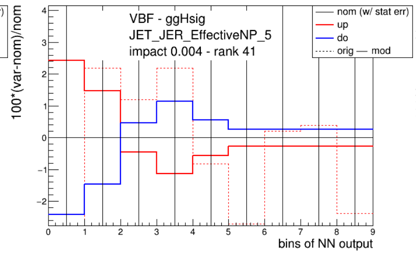


(l) meansym - NP6

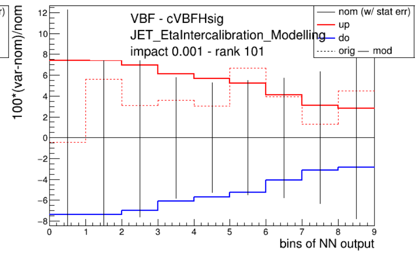
Figure 88: Selected envelope plots comparing maxsym and meansym $H \to e\tau$ Asimov fits, ordered by decreasing differences of the impact on the POI.



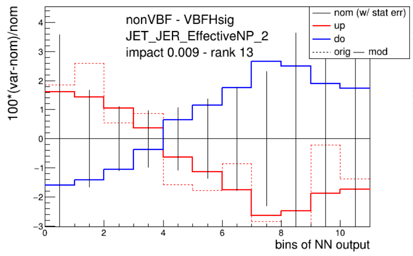
(a) maxsym - NP1



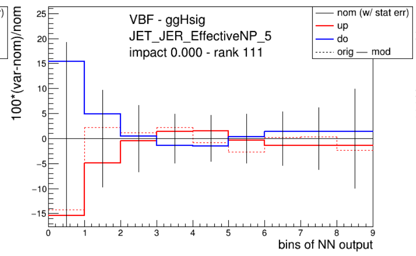
(b) maxsym - NP2



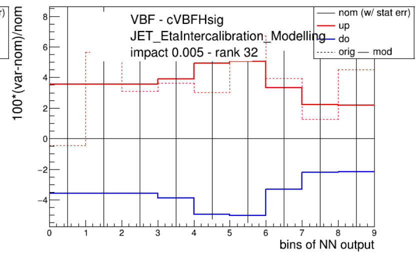
(c) maxsym - NP3



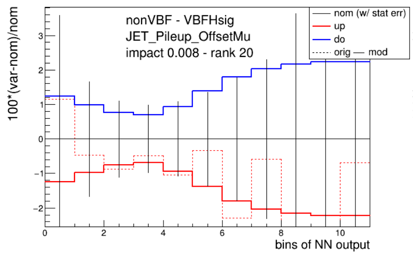
(d) meansym - NP1



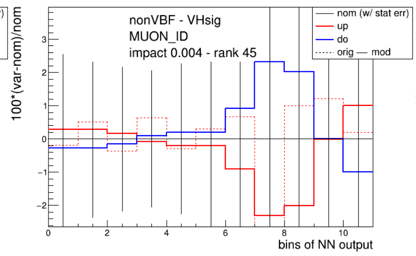
(e) meansym - NP2



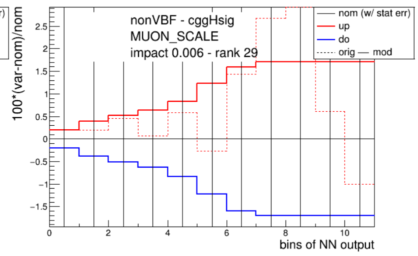
(f) meansym - NP3



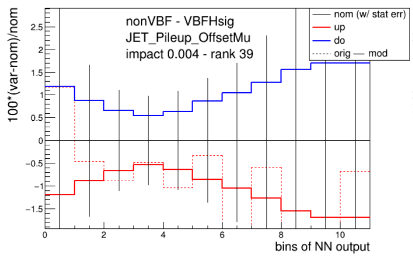
(g) maxsym - NP4



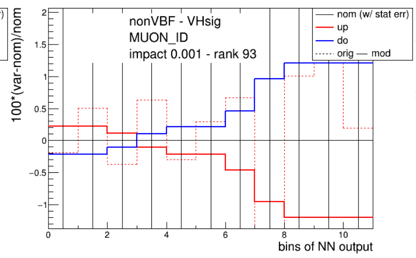
(h) maxsym - NP5



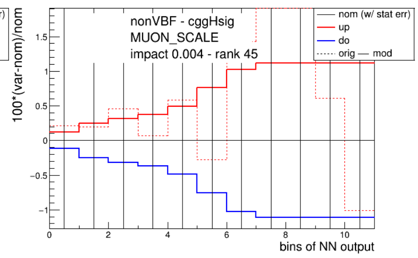
(i) maxsym - NP6



(j) meansym - NP4



(k) meansym - NP5



(l) meansym - NP6

Figure 89: Selected envelope plots comparing maxsym and meansym $H \to \mu\tau$ Asimov fits, ordered by decreasing differences of the impact on the POI.

B.2 MC-based vs. Sym-based setups

This study is based on the previous (v0.3) setup, so it cannot be directly compared to the study above. In particular, the VBF selection still uses $m_{jj} > 300\text{GeV}$ as a requirement, and the electron SFs between MC and efficiency correction are not yet fully correlated. Still, the MaxCorrect setup above, and the symSym_smoSym here, use the same symmetrization and smoothing options.

The nuisance parameter ranking, the pulls, the constraints, and the example nuisance parameter envelopes are shown in sections B.2.1 without any smoothing, B.2.2 with the MC-based smoothing, B.2.3 with the MC-based symmetrization, and B.2.4 with both MC-based smoothing and symmetrization. In each case, the fit is performed with the mixed dataset. The achieved sensitivities with the different symmetrization and smoothing options are summarized in Table 23 for the Asimov fit and Table 24 for the mixed fit.

	symSym_smoSym	symSym_smoNo	symSym_smoMC	symMC_smoSym	symMC_smoMC
$e\tau$ limit/%	$0.206^{+0.082}_{-0.057}$	$0.201^{+0.080}_{-0.056}$	$0.210^{+0.084}_{-0.059}$	$0.201^{+0.081}_{-0.056}$	$0.207^{+0.083}_{-0.058}$
$e\tau$ signif.	8.372	8.636	8.227	8.647	8.394
$\mu\tau$ limit/%	$0.187^{+0.073}_{-0.052}$	$0.187^{+0.073}_{-0.052}$	$0.192^{+0.075}_{-0.054}$	$0.184^{+0.072}_{-0.051}$	$0.190^{+0.074}_{-0.053}$
$\mu\tau$ signif.	8.975	8.955	8.609	9.148	8.698

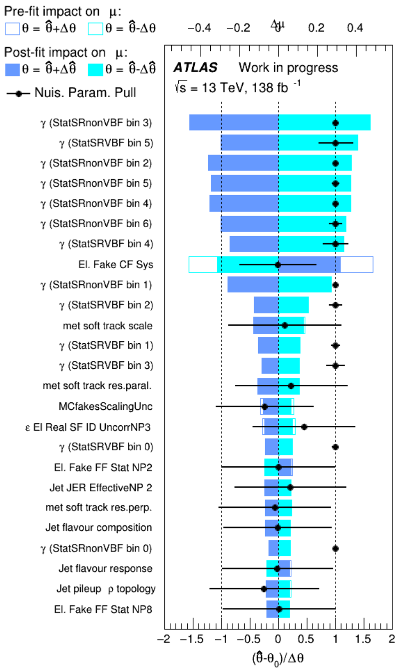
Table 23: Expected sensitivities comparing the different setups for the Asimov fit. The significance is given when assuming a 1% BR on the signal.

	symSym_smoSym	symSym_smoNo	symSym_smoMC	symMC_smoSym	symMC_smoMC
$e\tau$ limit/%	$0.200^{+0.080}_{-0.056}$	$0.198^{+0.079}_{-0.055}$	$0.207^{+0.082}_{-0.058}$	$0.198^{+0.079}_{-0.055}$	$0.205^{+0.081}_{-0.057}$
$e\tau$ signif.	0.968	0.983	0.934	0.977	0.946
$\mu\tau$ limit/%	$0.187^{+0.074}_{-0.052}$	$0.184^{+0.072}_{-0.051}$	$0.192^{+0.075}_{-0.054}$	$0.183^{+0.072}_{-0.051}$	$0.190^{+0.074}_{-0.053}$
$\mu\tau$ signif.	1.018	1.030	0.991	1.038	1.000

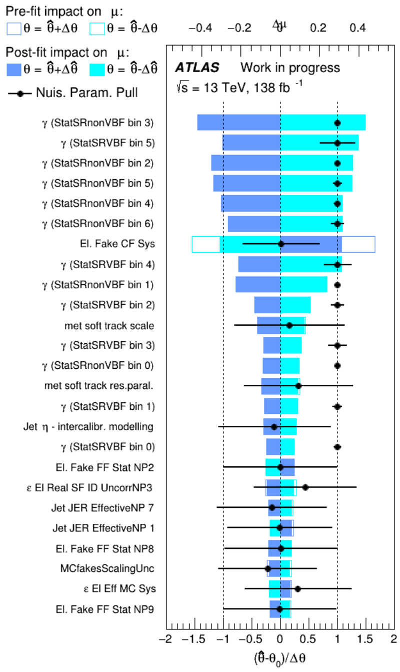
Table 24: Expected sensitivities comparing the different setups for the mixed fit. The significance is given when assuming a 1% BR on the signal.

The different setups for symmetrization and smoothing show consistent and well-behaving results. The pulls and the sensitivities are generally similar, and the shifts in the highest-ranked nuisance parameters are minor. In terms of the symmetrization option, the MaxCorrect is the most conservative, while the TwoSided sometimes slightly underestimates the uncertainties. The MeanCorrect, considered in section B.1, is meant to be in the middle. Among the smoothing options tested, MaxVariation is slightly more conservative than Parabolic.

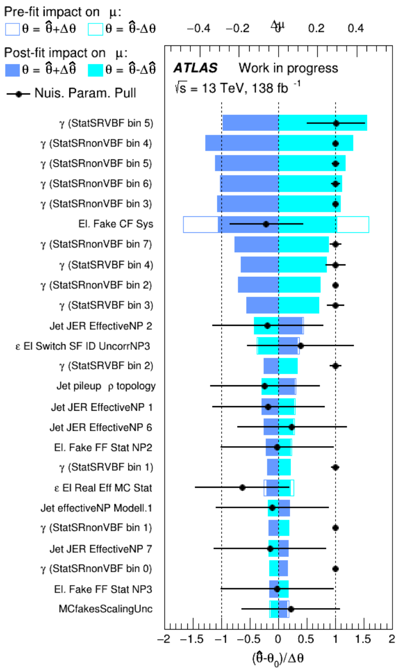
B.2.1 No Smoothing



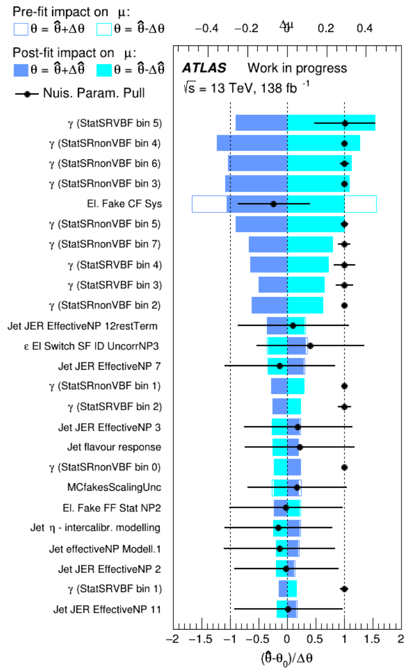
(a) symSym_smoSym — $H \to e\tau$



(b) symSym_smoNo — $H \to e\tau$

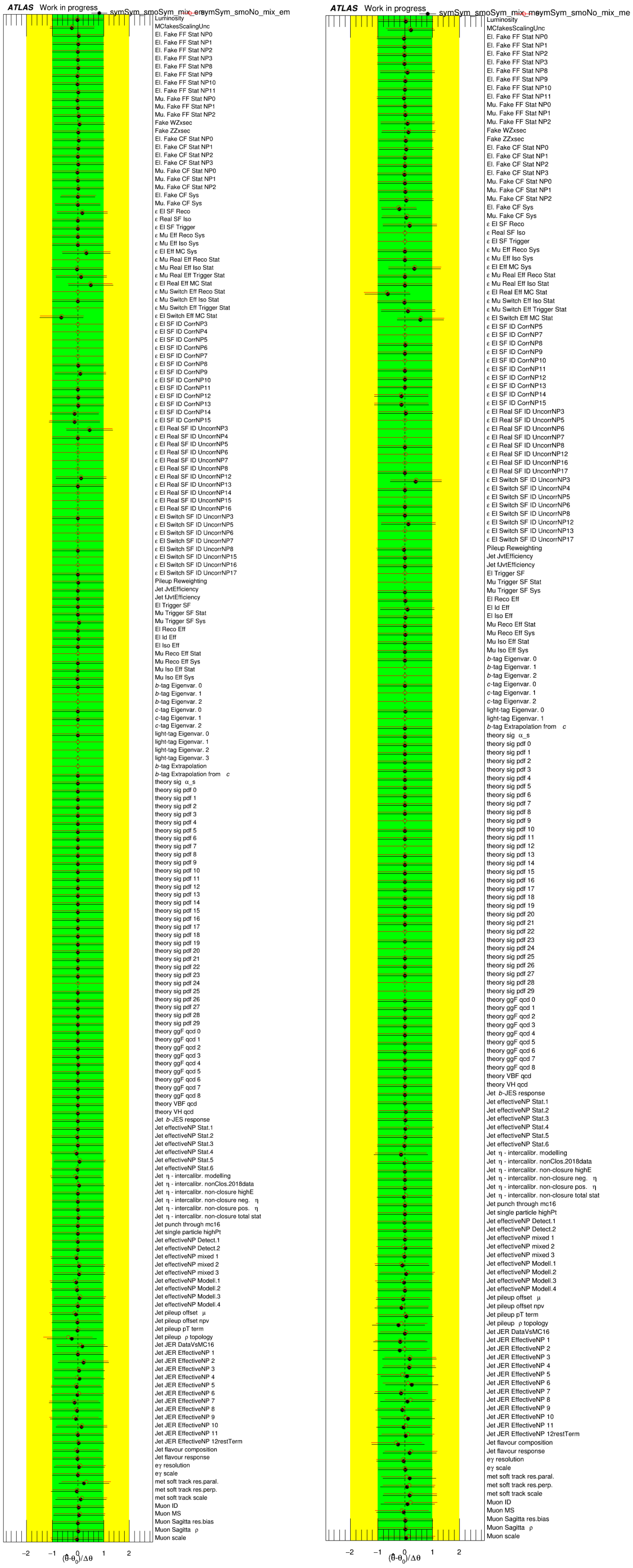


(c) symSym_smoSym — $H \to \mu\tau$



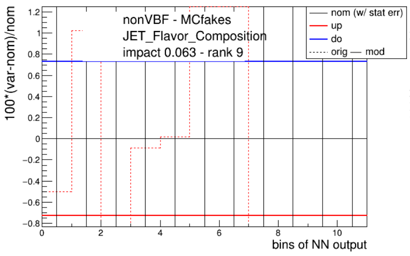
(d) symSym_smoNo — $H \to \mu\tau$

Figure 90: Ranking of the nuisance parameters comparing the symSym_smoSym and symSym_smoNo setups.

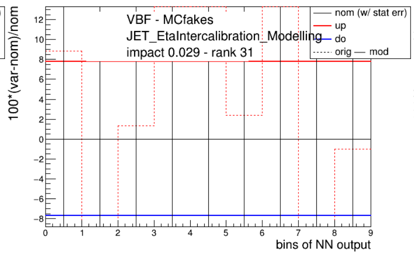


(a) $H \to e\tau$ (b) $H \to \mu\tau$

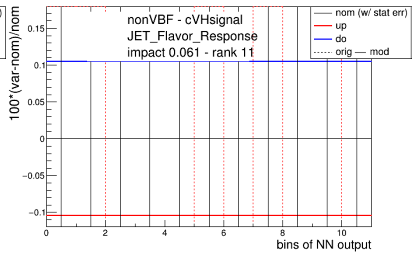
Figure 91: Pulls of the nuisance parameters comparing the symSym_smoSym and symSym_smoNo setups.



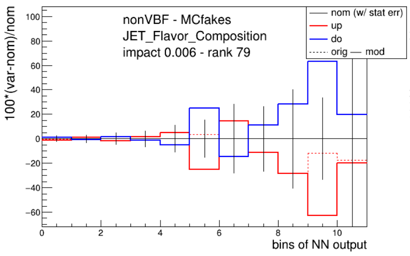
(a) symSym_smoSym - NP1



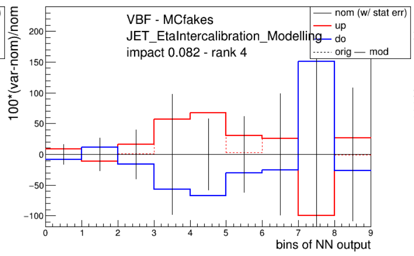
(b) symSym_smoSym - NP2



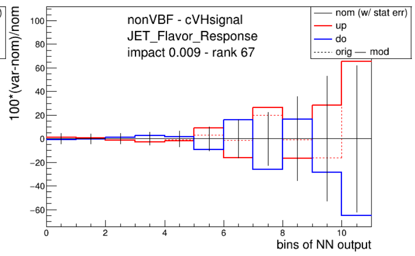
(c) symSym_smoSym - NP3



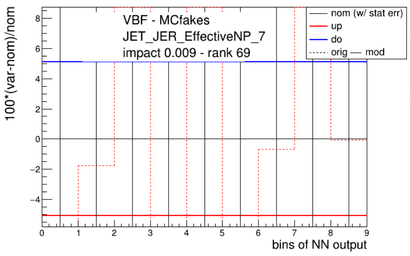
(d) symSym_smoNo - NP1



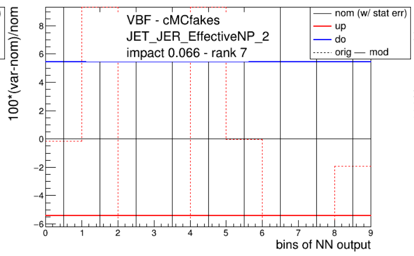
(e) symSym_smoNo - NP2



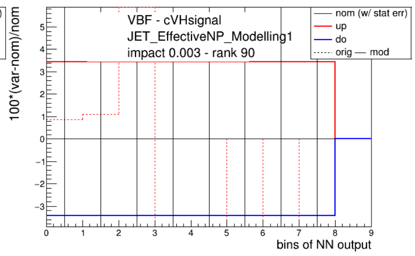
(f) symSym_smoNo - NP3



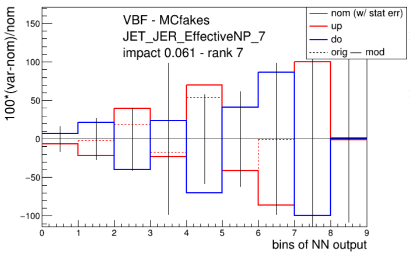
(g) symSym_smoSym - NP4



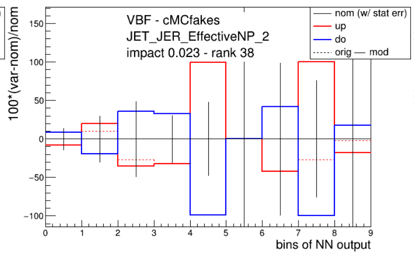
(h) symSym_smoSym - NP5



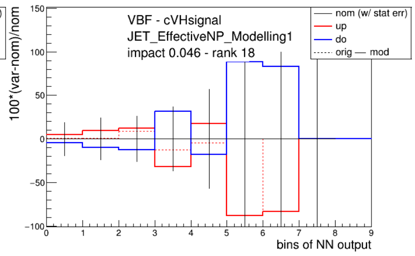
(i) symSym_smoSym - NP6



(j) symSym_smoNo - NP4

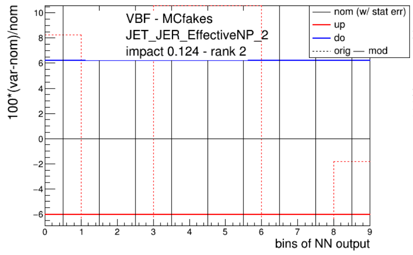


(k) symSym_smoNo - NP5

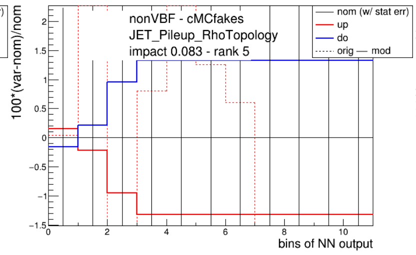


(l) symSym_smoNo - NP6

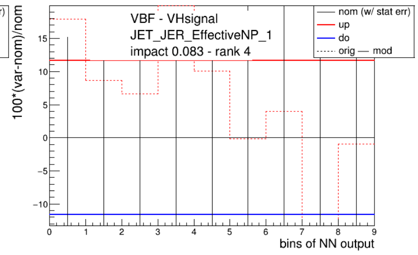
Figure 92: Selected envelope plots comparing symSym_smoSym and symSym_smoNo $H \to e\tau$ mixed fits, ordered by decreasing differences of the impact on the POI.



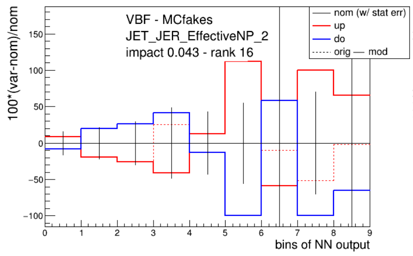
(a) symSym_smoSym - NP1



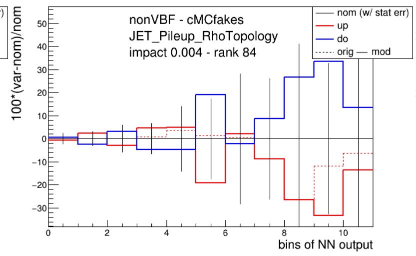
(b) symSym_smoSym - NP2



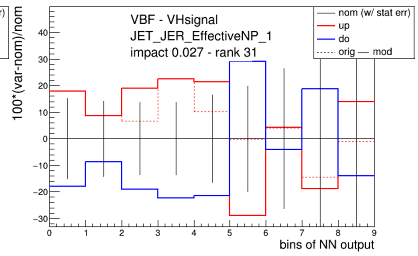
(c) symSym_smoSym - NP3



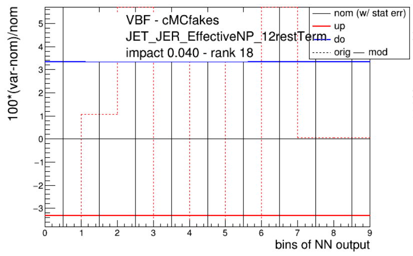
(d) symSym_smoNo - NP1



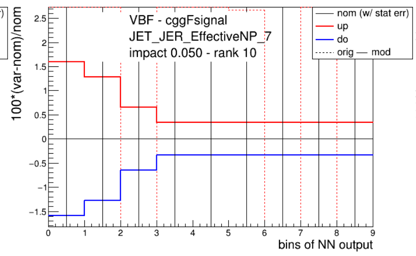
(e) symSym_smoNo - NP2



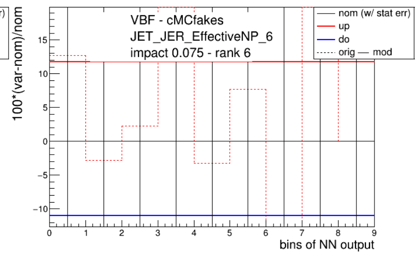
(f) symSym_smoNo - NP3



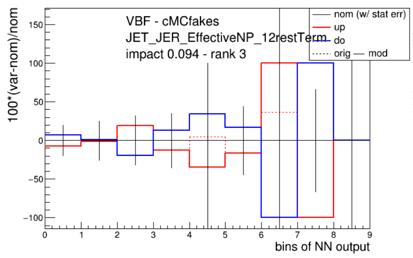
(g) symSym_smoSym - NP4



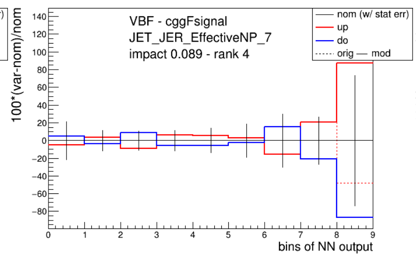
(h) symSym_smoSym - NP5



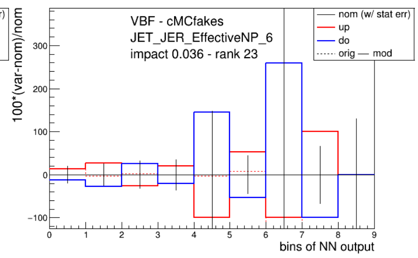
(i) symSym_smoSym - NP6



(j) symSym_smoNo - NP4



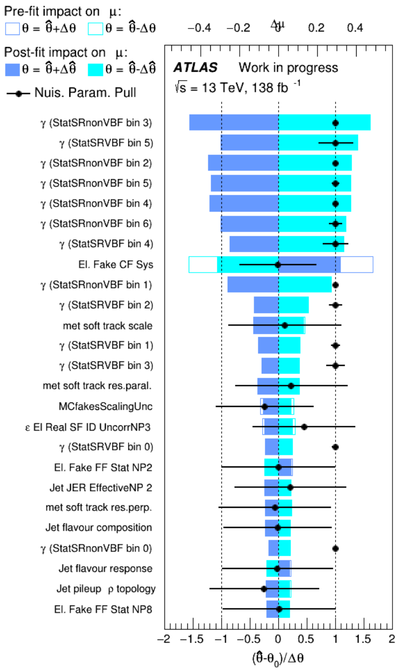
(k) symSym_smoNo - NP5



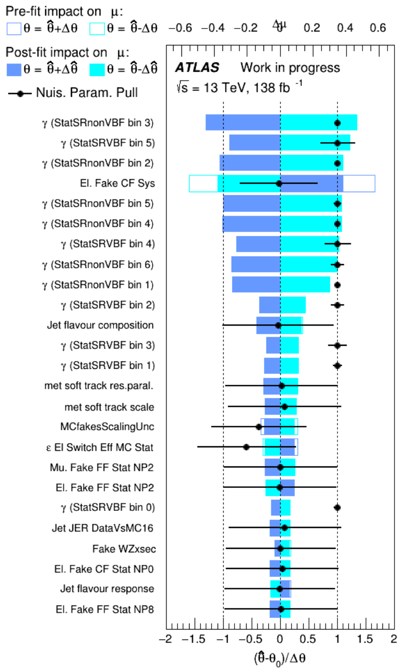
(l) symSym_smoNo - NP6

Figure 93: Selected envelope plots comparing symSym_smoSym and symSym_smoNo $H \to \mu\tau$ mixed fits, ordered by decreasing differences of the impact on the POI.

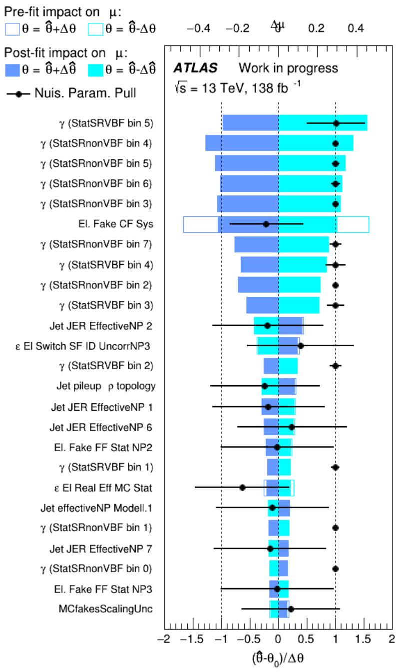
B.2.2 MCBased Smoothing



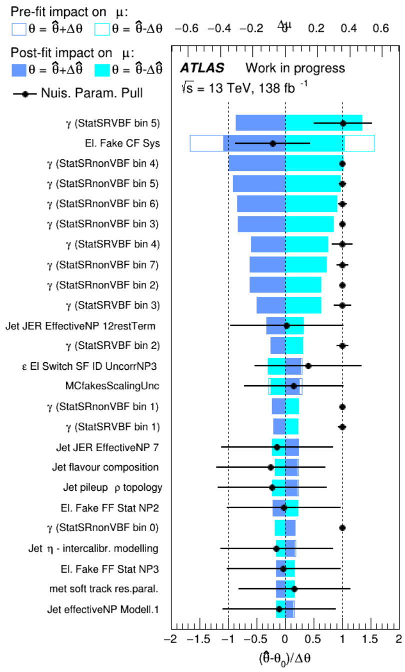
(a) symSym_smoSym — $H \to e\tau$



(b) symSym_smoMC — $H \to e\tau$

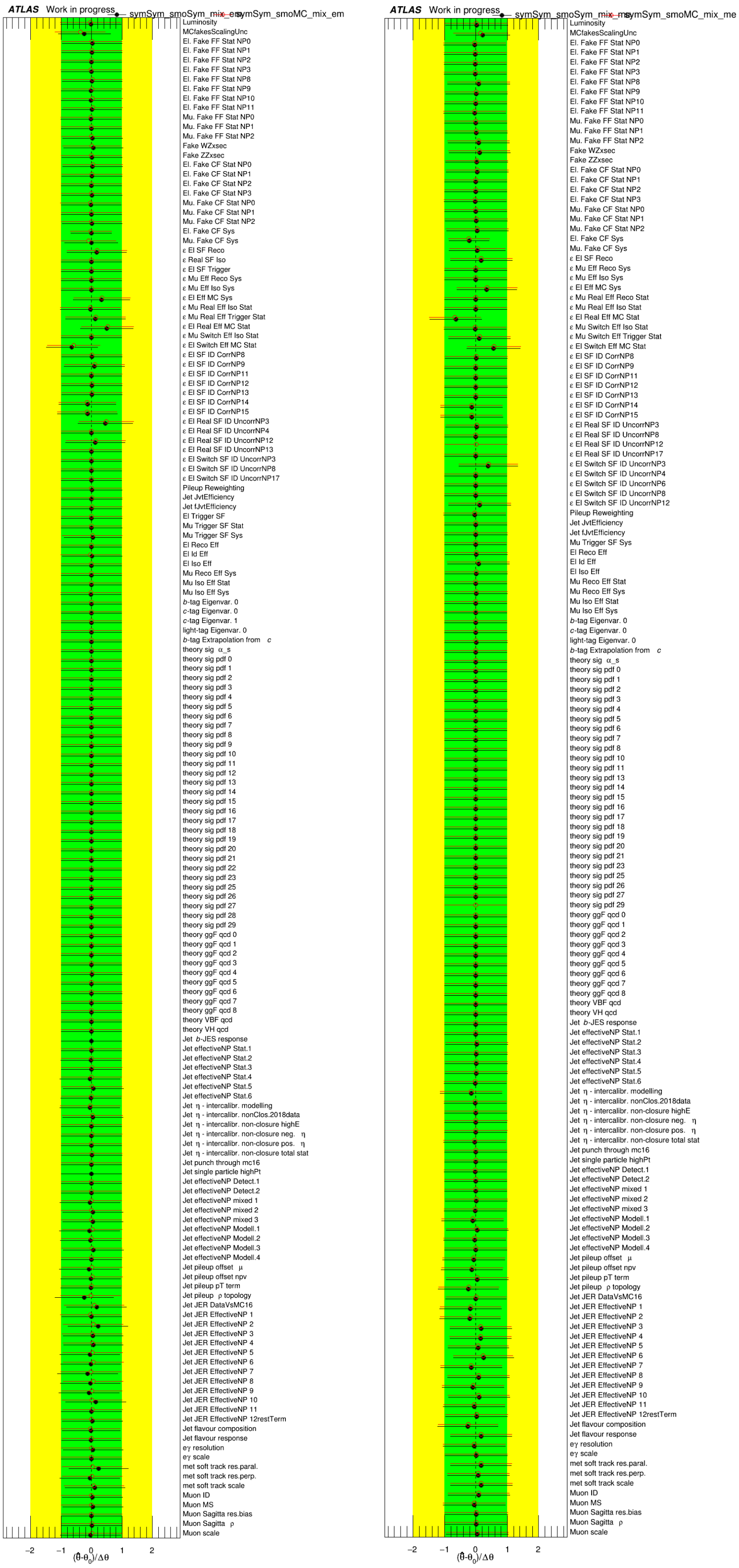


(c) symSym_smoSym — $H \to \mu\tau$



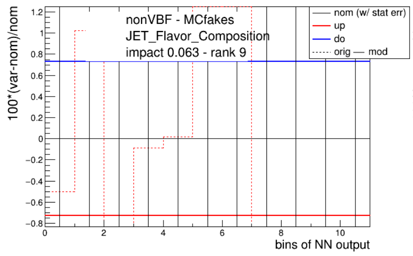
(d) symSym_smoMC — $H \to \mu\tau$

Figure 94: Ranking of the nuisance parameters comparing the symSym_smoSym and symSym_smoMC setups.

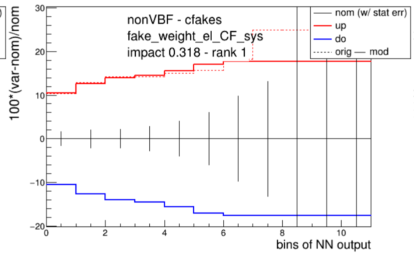


(a) $H \to e\tau$ (b) $H \to \mu\tau$

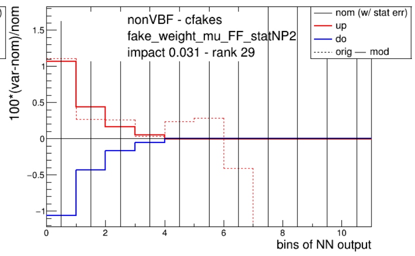
Figure 95: Pulls of the nuisance parameters comparing the symSym_smoSym and symSym_smoMC setups.



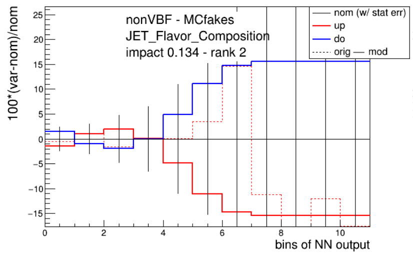
(a) symSym_smoSym - NP1



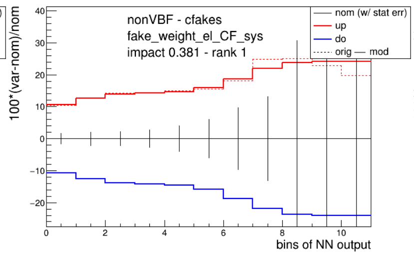
(b) symSym_smoSym - NP2



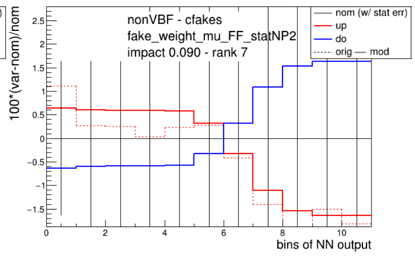
(c) symSym_smoSym - NP3



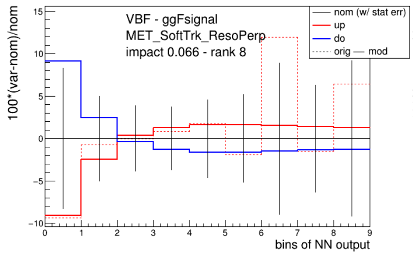
(d) symSym_smoMC - NP1



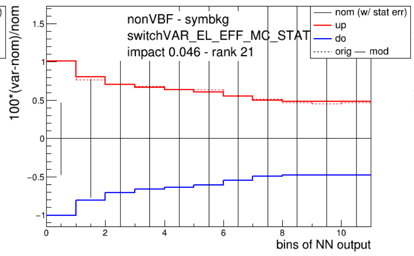
(e) symSym_smoMC - NP2



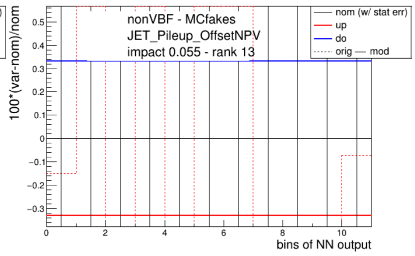
(f) symSym_smoMC - NP3



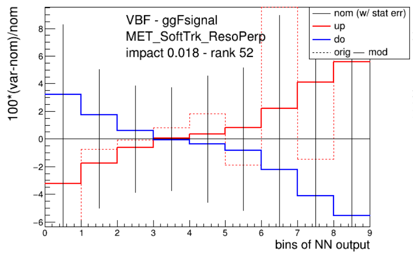
(g) symSym_smoSym - NP4



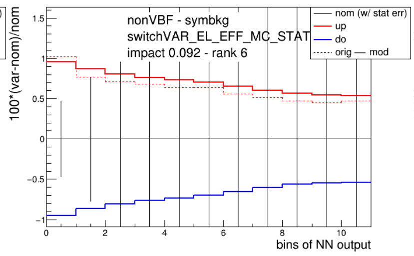
(h) symSym_smoSym - NP5



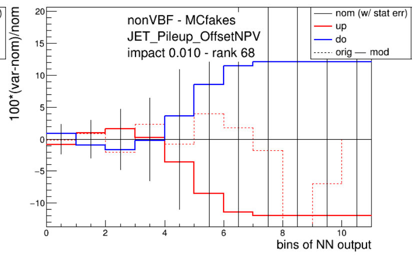
(i) symSym_smoSym - NP6



(j) symSym_smoMC - NP4

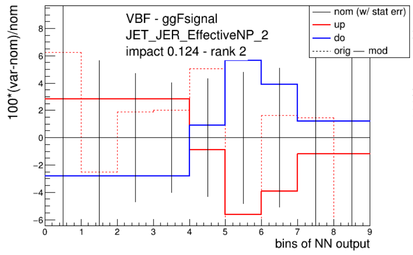


(k) symSym_smoMC - NP5

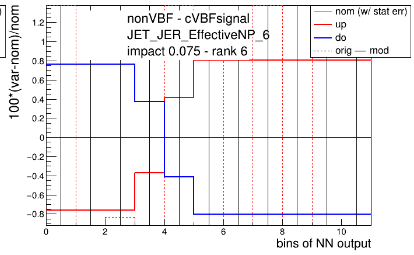


(l) symSym_smoMC - NP6

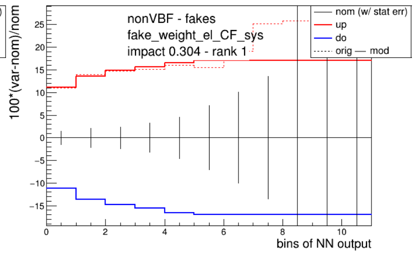
Figure 96: Selected envelope plots comparing symSym_smoSym and symSym_smoMC $H \to e\tau$ mixed fits, ordered by decreasing differences of the impact on the POI.



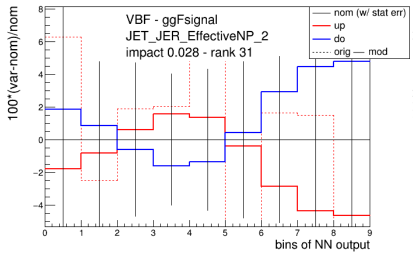
(a) symSym_smoSym - NP1



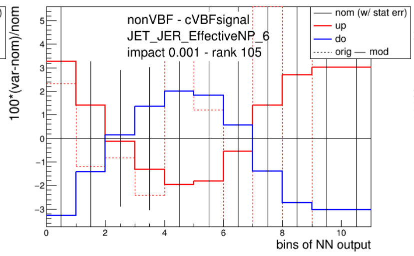
(b) symSym_smoSym - NP2



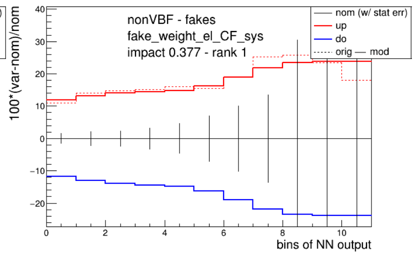
(c) symSym_smoSym - NP3



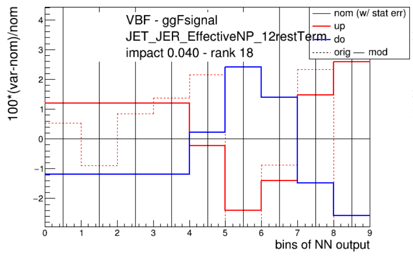
(d) symSym_smoMC - NP1



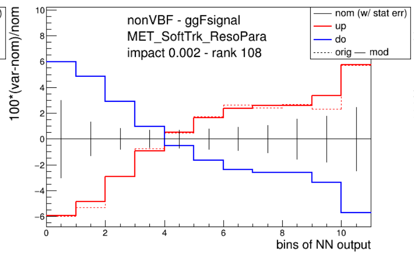
(e) symSym_smoMC - NP2



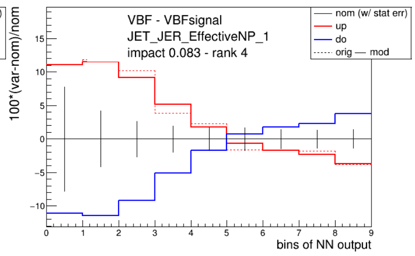
(f) symSym_smoMC - NP3



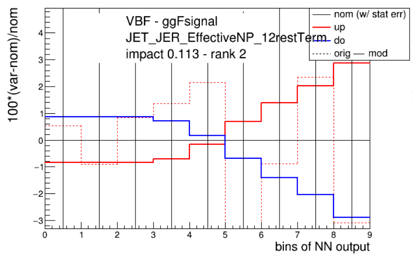
(g) symSym_smoSym - NP4



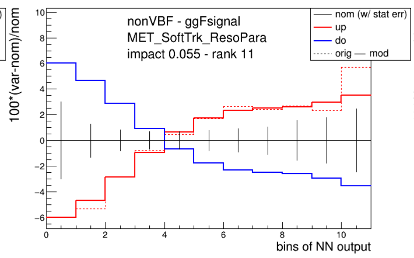
(h) symSym_smoSym - NP5



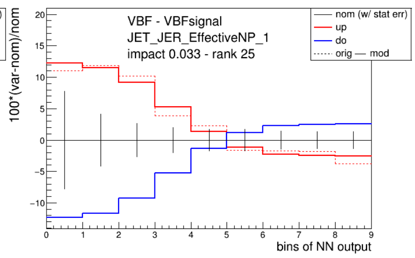
(i) symSym_smoSym - NP6



(j) symSym_smoMC - NP4



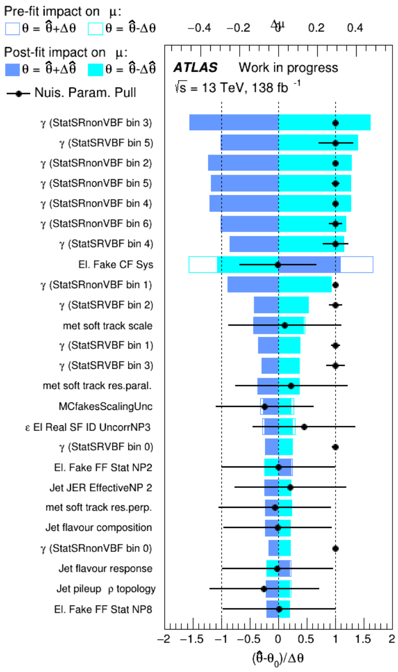
(k) symSym_smoMC - NP5



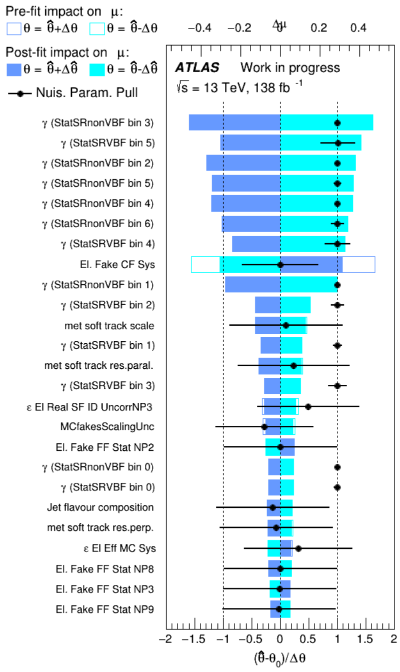
(l) symSym_smoMC - NP6

Figure 97: Selected envelope plots comparing symSym_smoSym and symSym_smoMC $H \to \mu\tau$ mixed fits, ordered by decreasing differences of the impact on the POI.

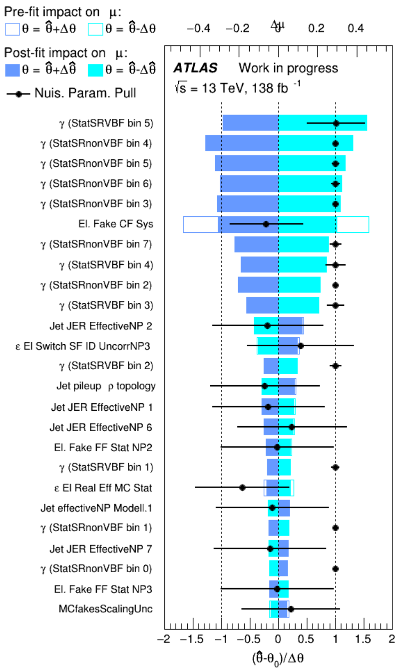
B.2.3 MCBased Symmetrization



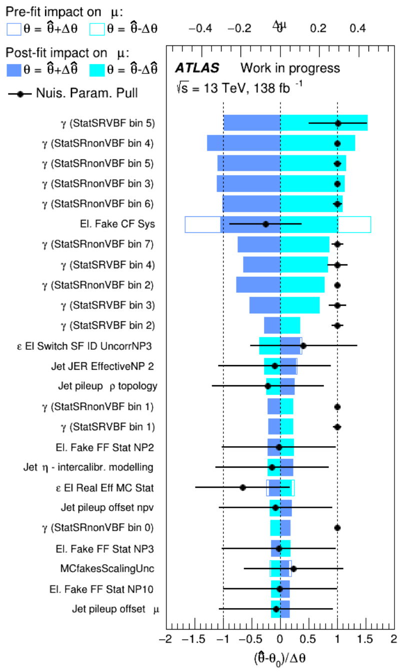
(a) symSym_smoSym — $H \to e\tau$



(b) symMC_smoSym — $H \to e\tau$

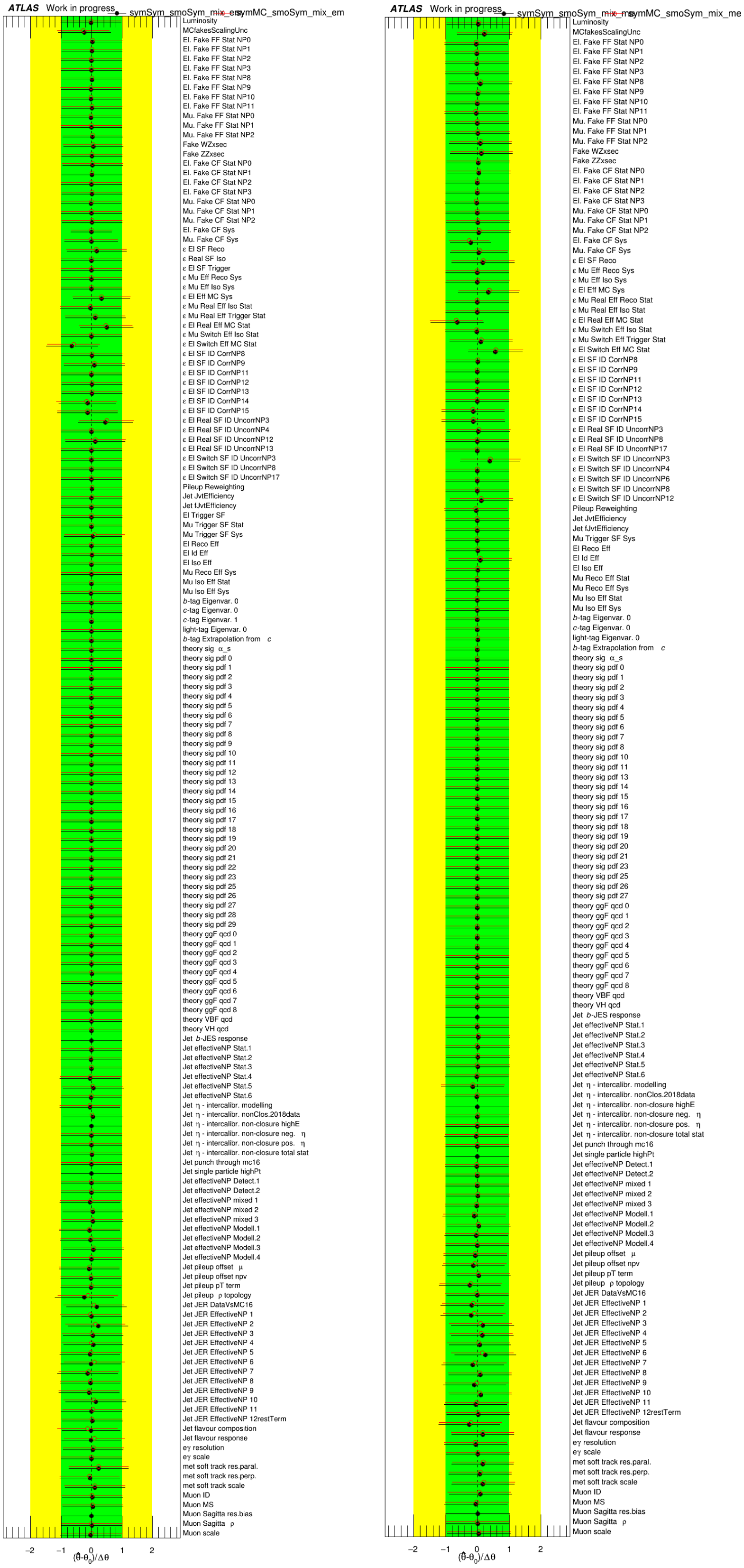


(c) symSym_smoSym — $H \to \mu\tau$



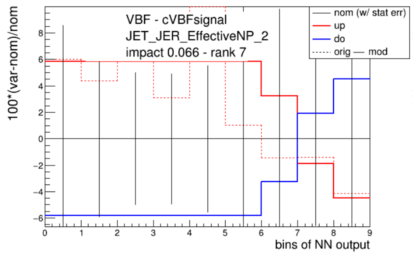
(d) symMC_smoSym — $H \to \mu\tau$

Figure 98: Ranking of the nuisance parameters comparing the symSym_smoSym and symMC_smoSym setups.

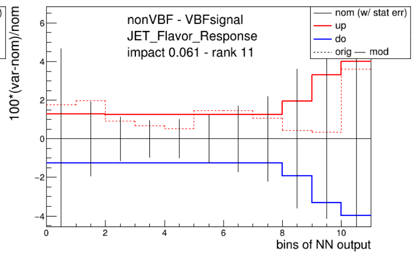


(a) $H \to e\tau$ (b) $H \to \mu\tau$

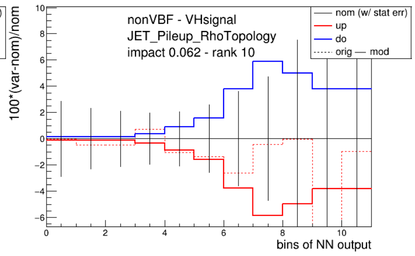
Figure 99: Pulls of the nuisance parameters comparing the symSym_smoSym and symMC_smoSym setups.



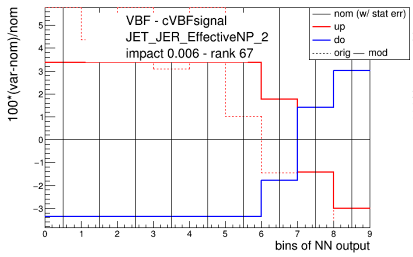
(a) symSym_smoSym - NP1



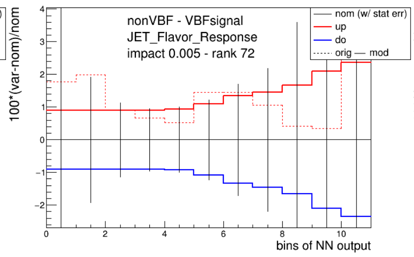
(b) symSym_smoSym - NP2



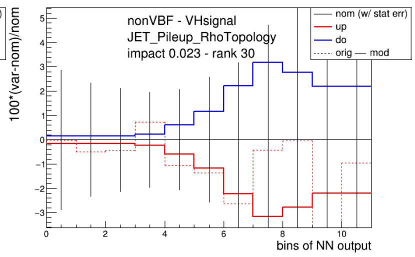
(c) symSym_smoSym - NP3



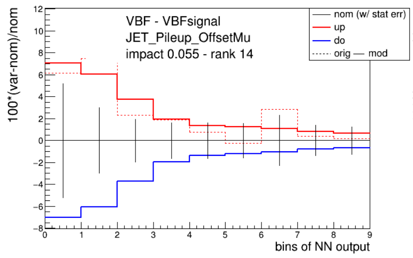
(d) symMC_smoSym - NP1



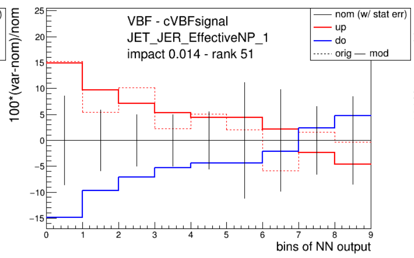
(e) symMC_smoSym - NP2



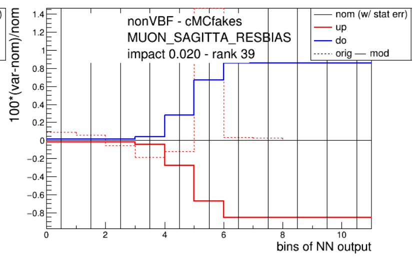
(f) symMC_smoSym - NP3



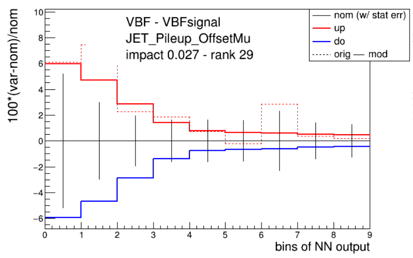
(g) symSym_smoSym - NP4



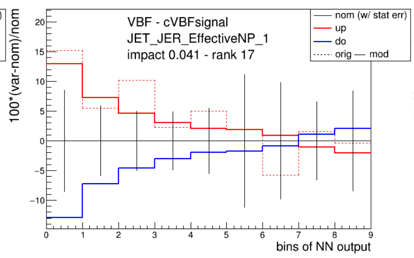
(h) symSym_smoSym - NP5



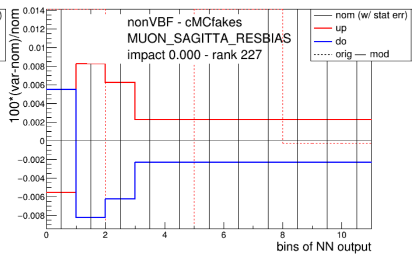
(i) symSym_smoSym - NP6



(j) symMC_smoSym - NP4

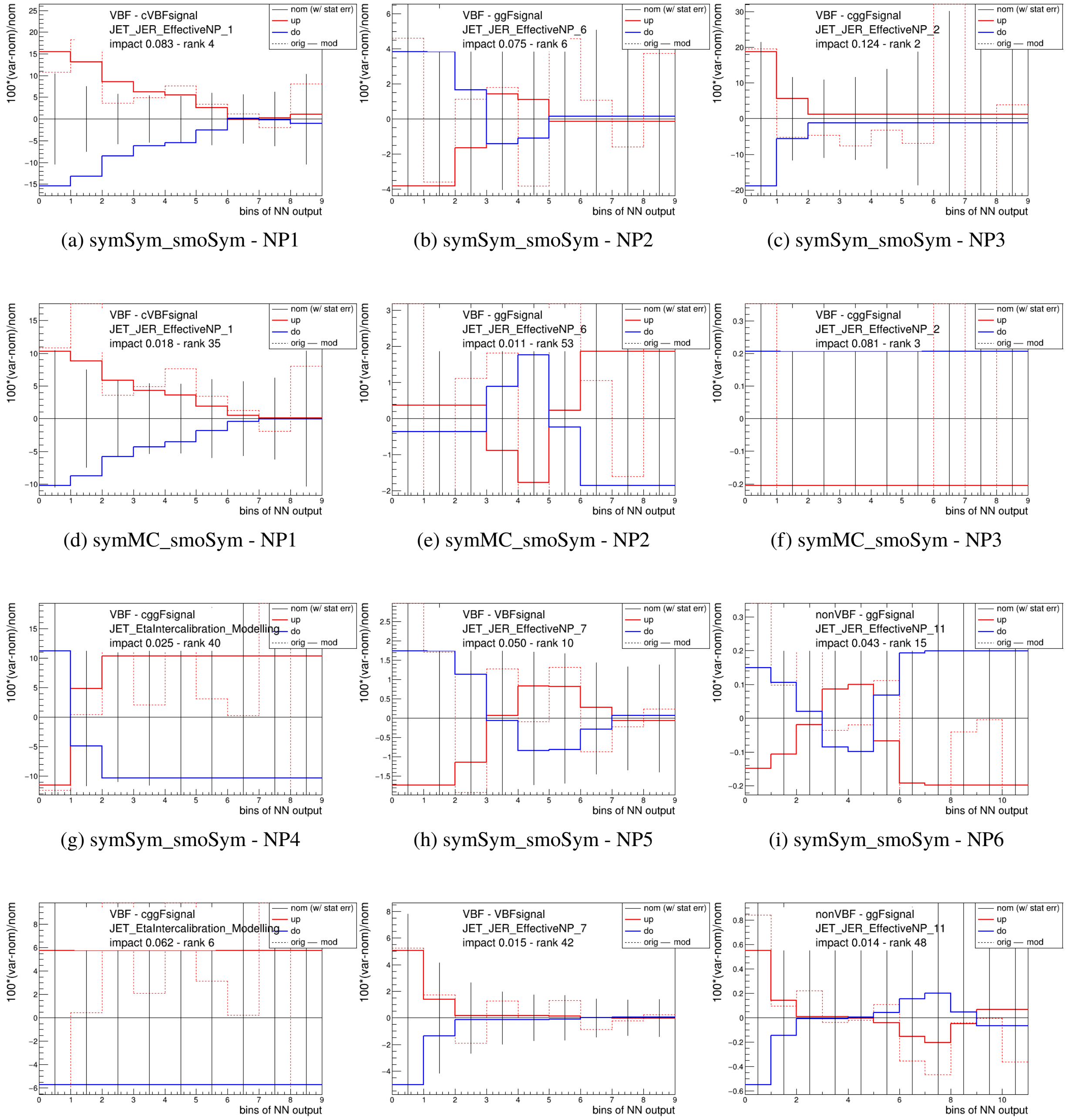


(k) symMC_smoSym - NP5



(l) symMC_smoSym - NP6

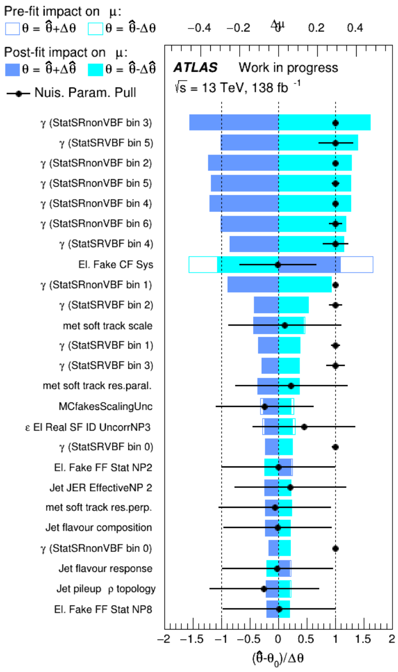
Figure 100: Selected envelope plots comparing symSym_smoSym and symMC_smoSym $H \to e\tau$ mixed fits, ordered by decreasing differences of the impact on the POI.



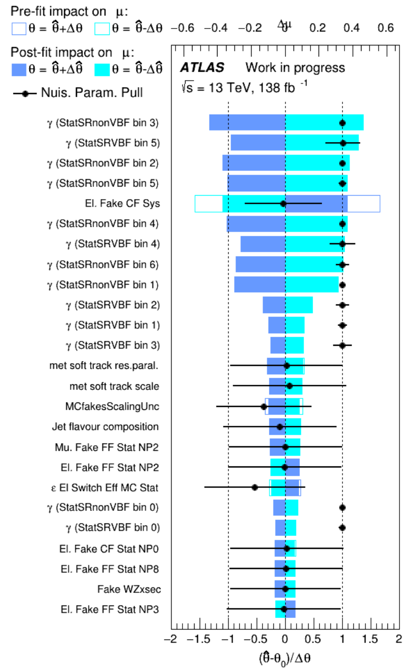
(a) symSym_smoSym - NP1
(b) symSym_smoSym - NP2
(c) symSym_smoSym - NP3
(d) symMC_smoSym - NP1
(e) symMC_smoSym - NP2
(f) symMC_smoSym - NP3
(g) symSym_smoSym - NP4
(h) symSym_smoSym - NP5
(i) symSym_smoSym - NP6
(j) symMC_smoSym - NP4
(k) symMC_smoSym - NP5
(l) symMC_smoSym - NP6

Figure 101: Selected envelope plots comparing symSym_smoSym and symMC_smoSym $H \to \mu\tau$ mixed fits, ordered by decreasing differences of the impact on the POI.

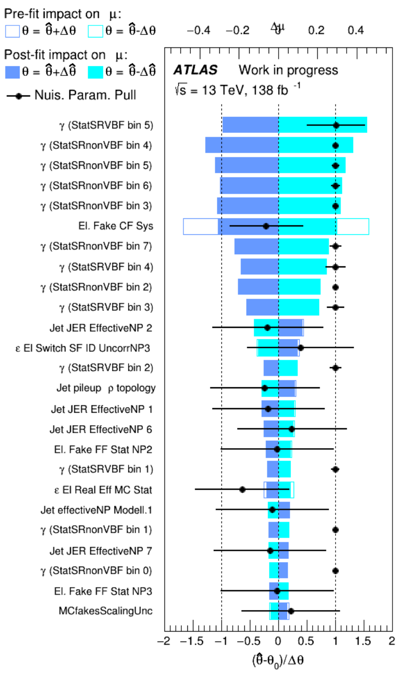
B.2.4 MCBased Symmetrization and Smoothing



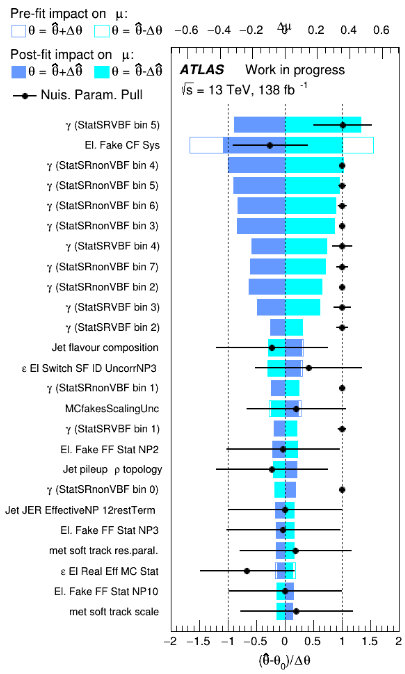
(a) symSym_smoSym — $H \rightarrow e\tau$



(b) symMC_smoMC — $H \rightarrow e\tau$

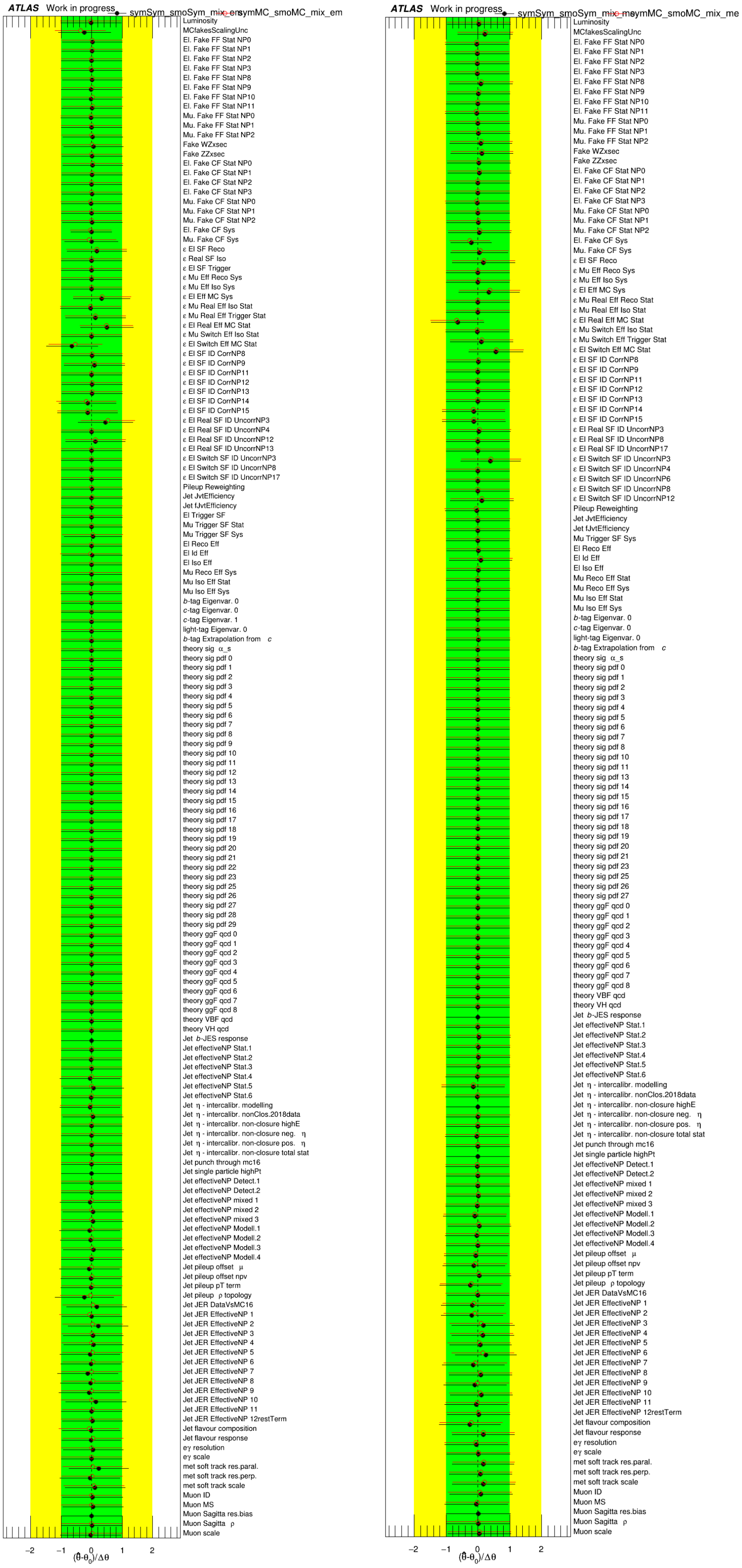


(c) symSym_smoSym — $H \rightarrow \mu\tau$



(d) symMC_smoMC — $H \rightarrow \mu\tau$

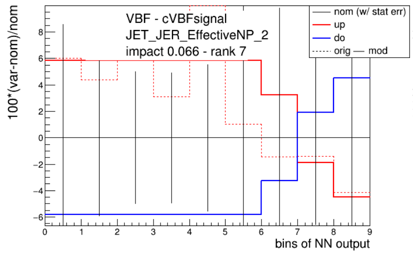
Figure 102: Ranking of the nuisance parameters comparing the symSym_smoSym and symMC_smoMC setups.



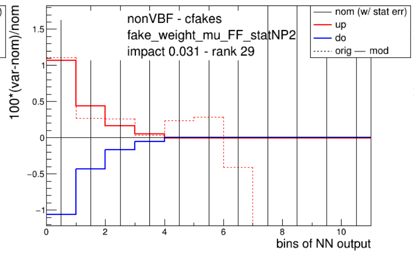
(a) $H \to e\tau$

(b) $H \to \mu\tau$

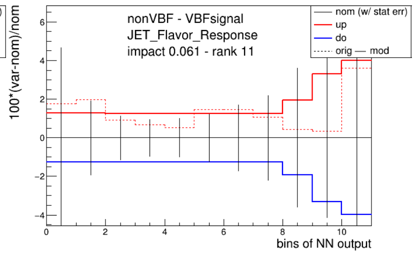
Figure 103: Pulls of the nuisance parameters comparing the symSym_smoSym and symMC_smoMC setups.



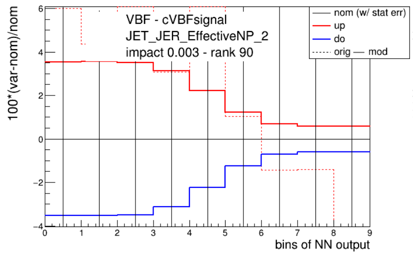
(a) symSym_smoSym - NP1



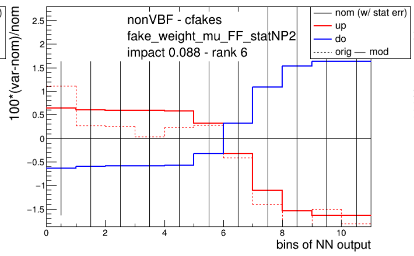
(b) symSym_smoSym - NP2



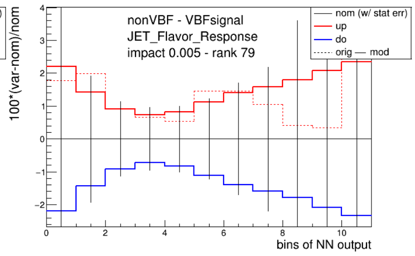
(c) symSym_smoSym - NP3



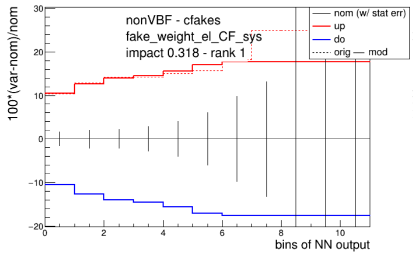
(d) symMC_smoMC - NP1



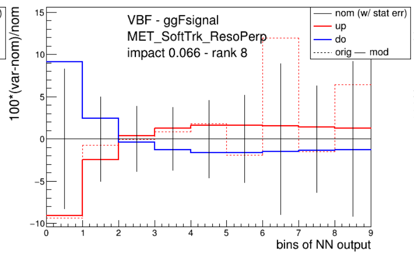
(e) symMC_smoMC - NP2



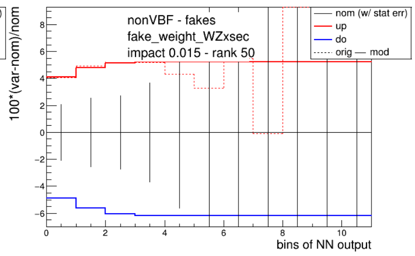
(f) symMC_smoMC - NP3



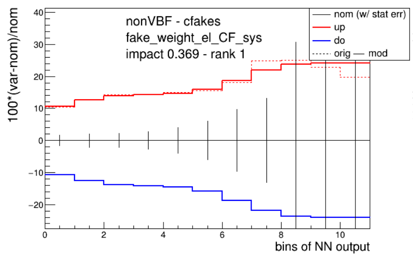
(g) symSym_smoSym - NP4



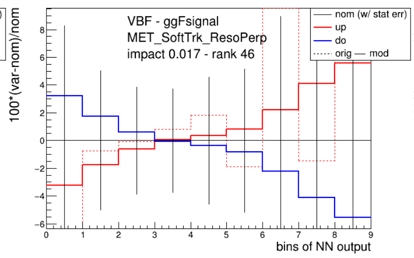
(h) symSym_smoSym - NP5



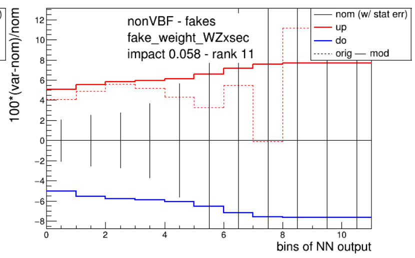
(i) symSym_smoSym - NP6



(j) symMC_smoMC - NP4

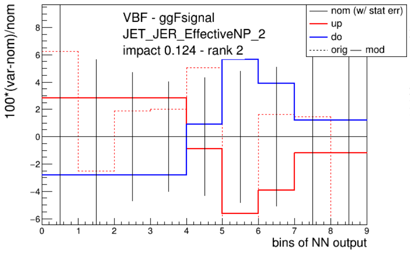


(k) symMC_smoMC - NP5

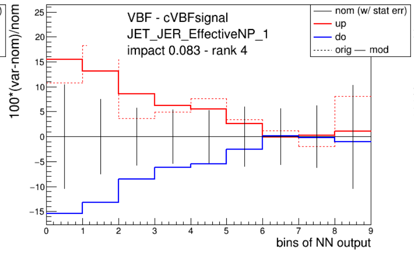


(l) symMC_smoMC - NP6

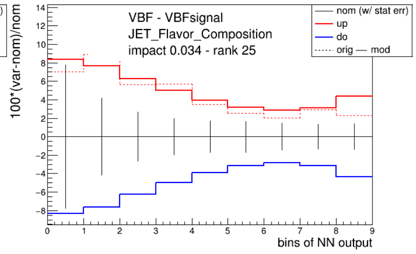
Figure 104: Selected envelope plots comparing symSym_smoSym and symMC_smoMC $H \to e\tau$ mixed fits, ordered by decreasing differences of the impact on the POI.



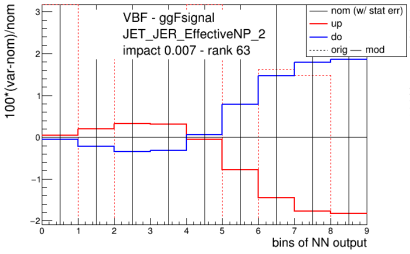
(a) symSym_smoSym - NP1



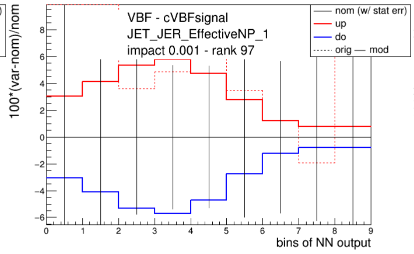
(b) symSym_smoSym - NP2



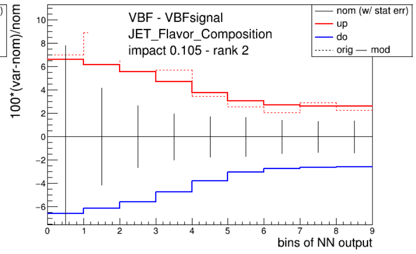
(c) symSym_smoSym - NP3



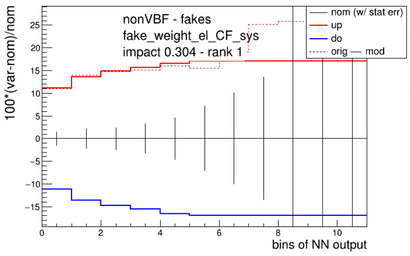
(d) symMC_smoMC - NP1



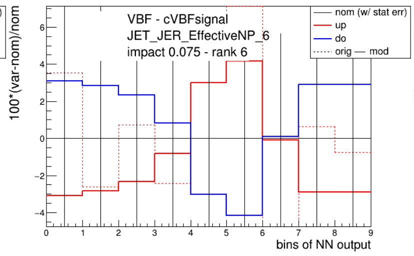
(e) symMC_smoMC - NP2



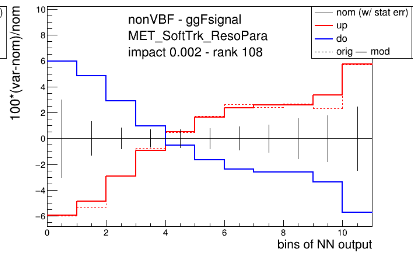
(f) symMC_smoMC - NP3



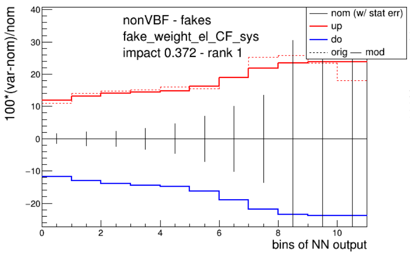
(g) symSym_smoSym - NP4



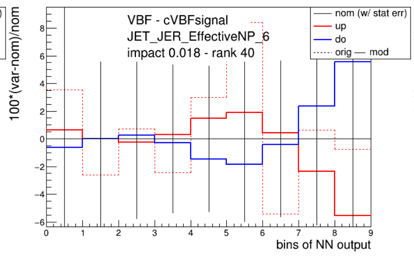
(h) symSym_smoSym - NP5



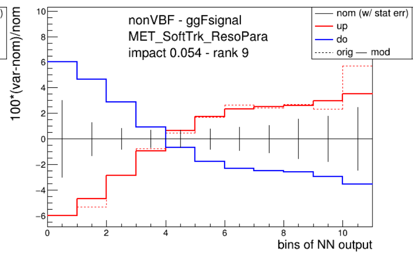
(i) symSym_smoSym - NP6



(j) symMC_smoMC - NP4



(k) symMC_smoMC - NP5



(l) symMC_smoMC - NP6

Figure 105: Selected envelope plots comparing symSym_smoSym and symMC_smoMC $H \to \mu\tau$ mixed fits, ordered by decreasing differences of the impact on the POI.

C ATLAS Higgs LFV search - blind fit results

The studies presented here complement the discussion in section 4.9.2.

This section presents the Run2 Higgs LFV analysis fit results using the blinded datasets – Asimov and mixed. As a reminder, the blinded datasets are used to evaluate the statistical model and fit's performance prior to unblinding the data in the sensitive regions for the obtention of the final analysis results. Expected results are also derived from the Asimov datasets, although those displayed in section 4.9.2 use a modified Asimov dataset built with the post-fit (unblinded) predictions, while the Asimov dataset used here is pre-fit. A description of the obtention of the results, the different inputs used, and the methods implemented to evaluate the fit performance is given in section 2.6.

Results for the $H \to e\tau_\mu$ and $H \to \mu\tau_e$ searches with the Asimov dataset are described in section C.1 and with the mixed dataset in section C.2.

C.1 Fit results with the Asimov dataset

As a reminder, the Asimov dataset is constructed as the sum of the background prediction and the signal times a fixed signal strength μ^s_{inj}. We set $\mu^s_{\text{inj}} = 1$ for the obtention of expected significance and $\mu^s_{\text{inj}} = 0$ for the expected limit. The signals are normalized to 1% BR.

The Asimov fit results were already presented in appendix B.1 (under the label MeanCorrect or "meansym") in the context of comparing different symmetrization algorithms. We refer to these results when needed; only complementary results are displayed here.

The post-fit NN output distributions are shown in Figure 106, where the blinded bins are marked with a gray hashed area. In the plots, the black dots represent the actual data shown for illustrative purposes, although it is not used in the fit.

Pulls and constraints on the nuisance parameters are shown in Figure 87 for the systematic uncertainties and Figure 107 for the γ-parameters. Their correlations are shown in Figure 108. Since the prediction is used instead of the data, there is no pull of the nuisance parameters' nominal values, but constraints and correlations appear, reducing their uncertainties. As in the unblinded fit, the only noticeable constraint is for "El. Fake CF sys", and the largest correlation is between "El. Fake CF sys" and "Mu. Fake CF sys".

The nuisance parameter impact ranking is shown in Figure 86, and the ranking per uncertainty group (the same groups defined in section 4.9.2) in Tables 25 and 26. The impact from systematics related to MC simulation – in particular jets, $E_{\text{T}}^{\text{miss}}$, and signal theory – is much higher than in the unblinded fit since here, the signal contribution is much larger ($\mu^s = 1$).

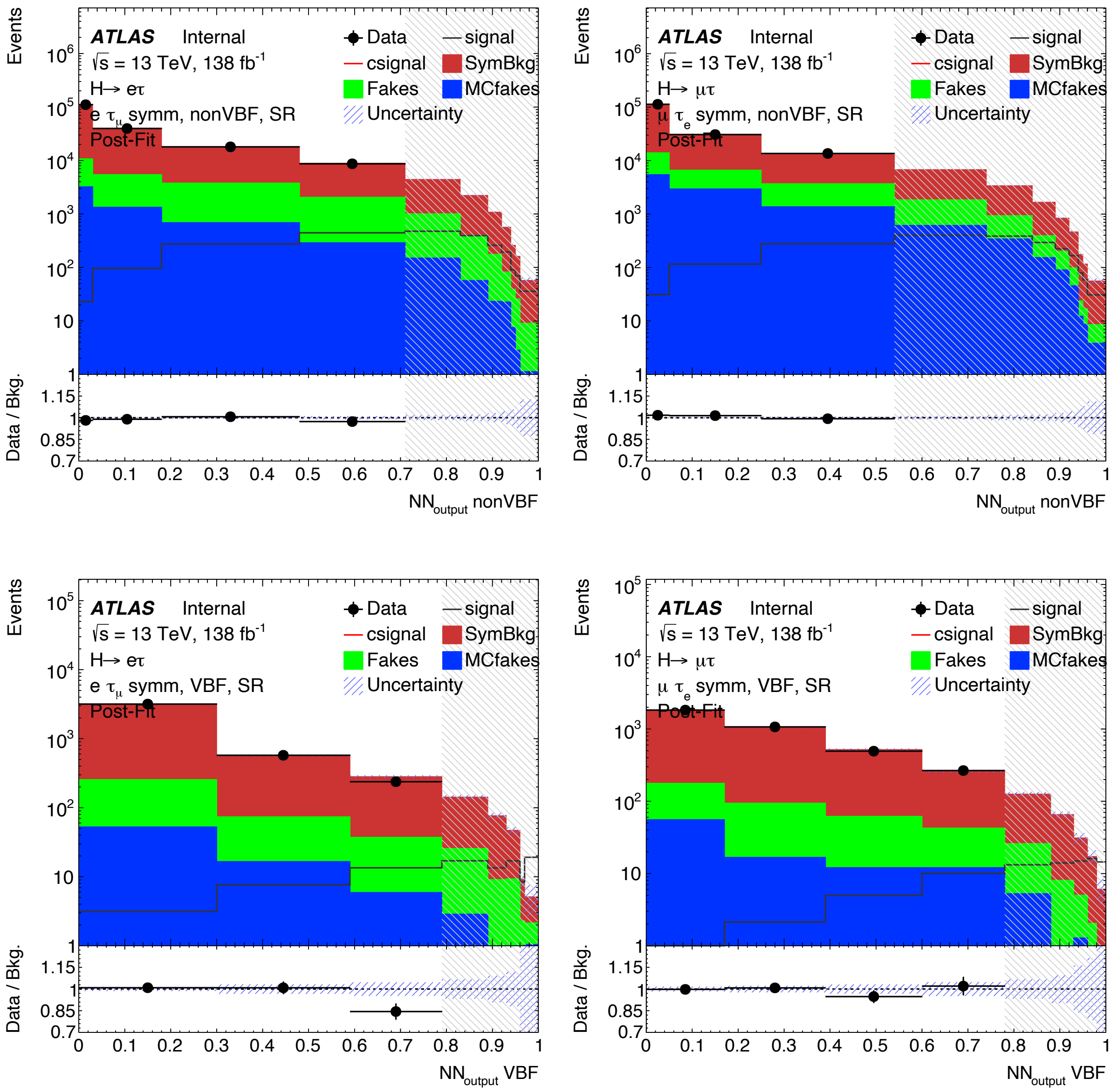


Figure 106: Post-fit NN output distributions of the $H \to e\tau_\mu$ (left) and $H \to \mu\tau_e$ signals (right) in the non-VBF (top) and VBF (bottom) categories. Statistic and systematic uncertainties are included. The BR of the signal is normalized to 1%.

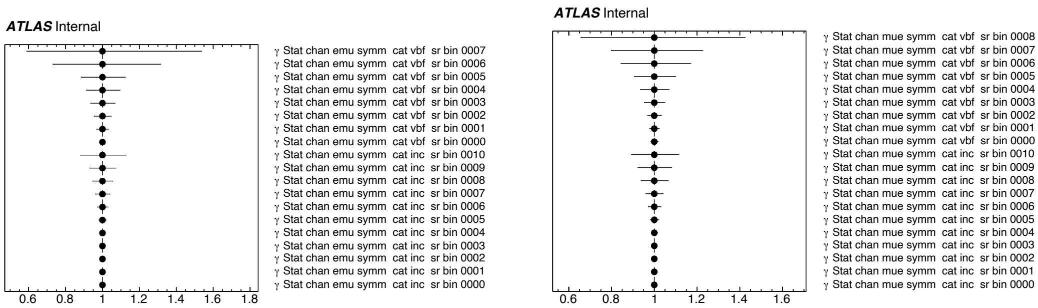


Figure 107: Pulls on the γ-parameters in the $H \to e\tau_\mu$ (left) and $H \to \mu\tau_e$ (right) Asimov fits.

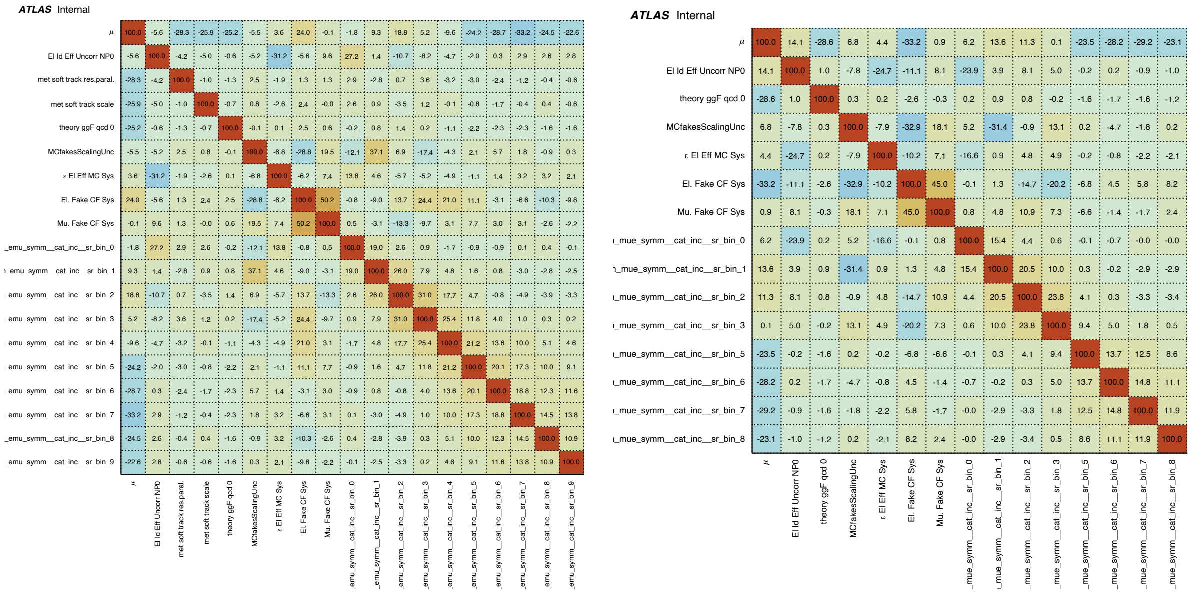


Figure 108: Correlations of the nuisance parameters in $H \to e\tau_\mu$ (left) and $H \to \mu\tau_e$ (right) Asimov fits. Only nuisance parameters with correlations above/below $+/-20\%$ are shown.

Group	Impact on unc. of μ	
Full unc.	+0.136	−0.128
Data unc.	+0.058	−0.058
Prediction unc.	+0.123	−0.113
Gammas	+0.076	−0.073
BTag	+0.003	−0.003
JETMET	+0.068	−0.057
Lepton	+0.017	−0.014
Lumi	+0.020	−0.016
SigTheory	+0.044	−0.035
Fakes + Eff. Corr.	+0.057	−0.053

Table 25: Impact of the different uncertainty groups on the uncertainty of the signal strength μ in the $H \to e\tau_\mu$ Asimov fit.

Group	Impact on unc. of μ	
Full unc.	+0.101	−0.099
Data unc.	+0.056	−0.055
Prediction unc.	+0.084	−0.083
Gammas	+0.067	−0.068
BTag	+0.001	−0.000
JETMET	+0.020	−0.018
Lepton	+0.018	−0.016
Lumi	+0.005	−0.002
SigTheory	+0.013	−0.004
Fakes + Eff. Corr.	+0.053	−0.051

Table 26: Impact of the different uncertainty groups on the uncertainty of the signal strength μ in the $H \to \mu\tau_e$ Asimov fit.

The expected significances and limits are shown in Table 27 separately for the non-VBF, VBF, and combined non-VBF + VBF categories.

	non-VBF+VBF	non-VBF	VBF
$e\tau$ limit/%	$0.189^{+0.075}_{-0.053}$	$0.230^{+0.092}_{-0.064}$	$0.338^{+0.139}_{-0.095}$
$e\tau$ signif.	8.685	7.787	3.977
$\mu\tau$ limit/%	$0.188^{+0.074}_{-0.053}$	$0.214^{+0.084}_{-0.060}$	$0.404^{+0.164}_{-0.113}$
$\mu\tau$ signif.	9.107	8.258	3.812

Table 27: Expected sensitivities. The significance is given for assuming $BR = 1\%$.

C.2 Fit results with the mixed dataset

The mixed dataset includes the data in the non-blinded bins (see Figure 106). The blinded bins include Asimov dataset yields but modified with post-fit background yields from a background-only fit in the non-blinded bins.

The nuisance parameter correlations are shown in Figure 109, and their pulls and constraints in Figure 110 (Figure 111) for the $H \to e\tau_\mu$ ($H \to \mu\tau_e$) search. Except for the pulls, these are similar to the Asimov fit results. The pulls are similar to those observed in the unblinded fits – all under or very close to 1σ – except for the γ-parameters in the blinded bins, which are strongly pulled after unblinding but not here. The nuisance parameter impact ranking is shown in Figure 112, similar to the ranking observed with the Asimov fits. Overall, the fits with the mixed datasets show very reasonable results. The measured best-fit signal strengths are $\mu^s_{e\tau} = 0.87 \pm 0.98$ and $\mu^s_{\mu\tau} = 0.93 \pm 0.96$ for the $H \to e\tau_\mu$ and $H \to \mu\tau_e$ searches, respectively, in line with the signal injected in the input datasets ($\mu^s_{\text{inj}} = 1$).

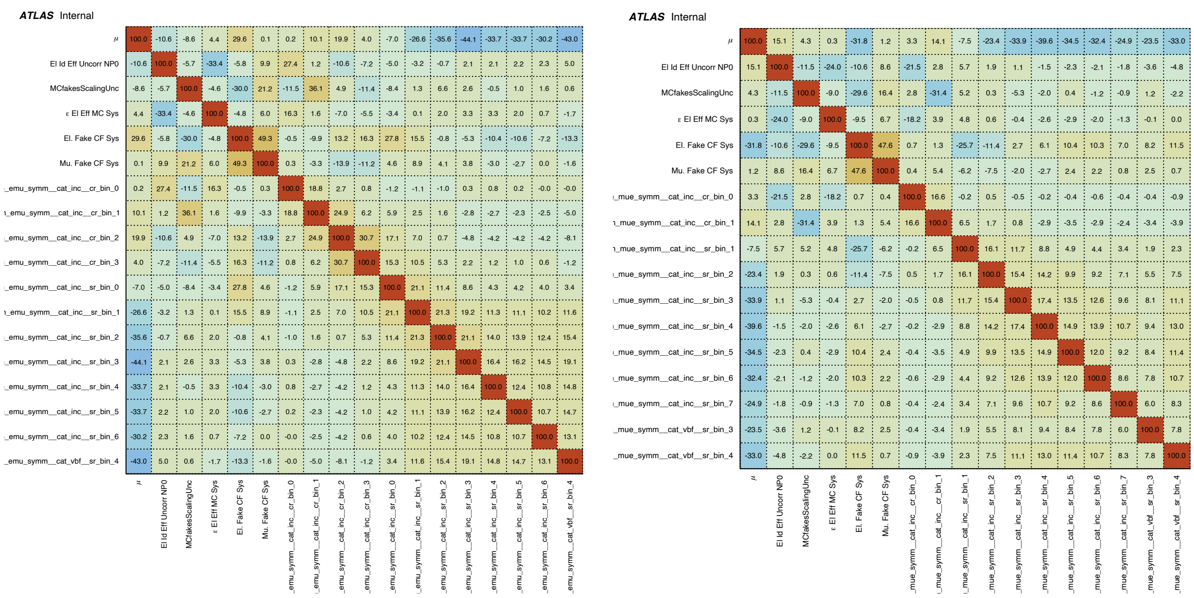


Figure 109: Correlations of the nuisance parameters in the $H \to e\tau_\mu$ (left) and $H \to \mu\tau_e$ (right) mixed fits. Only nuisance parameters with correlations above/below $+/-20\%$ are shown.

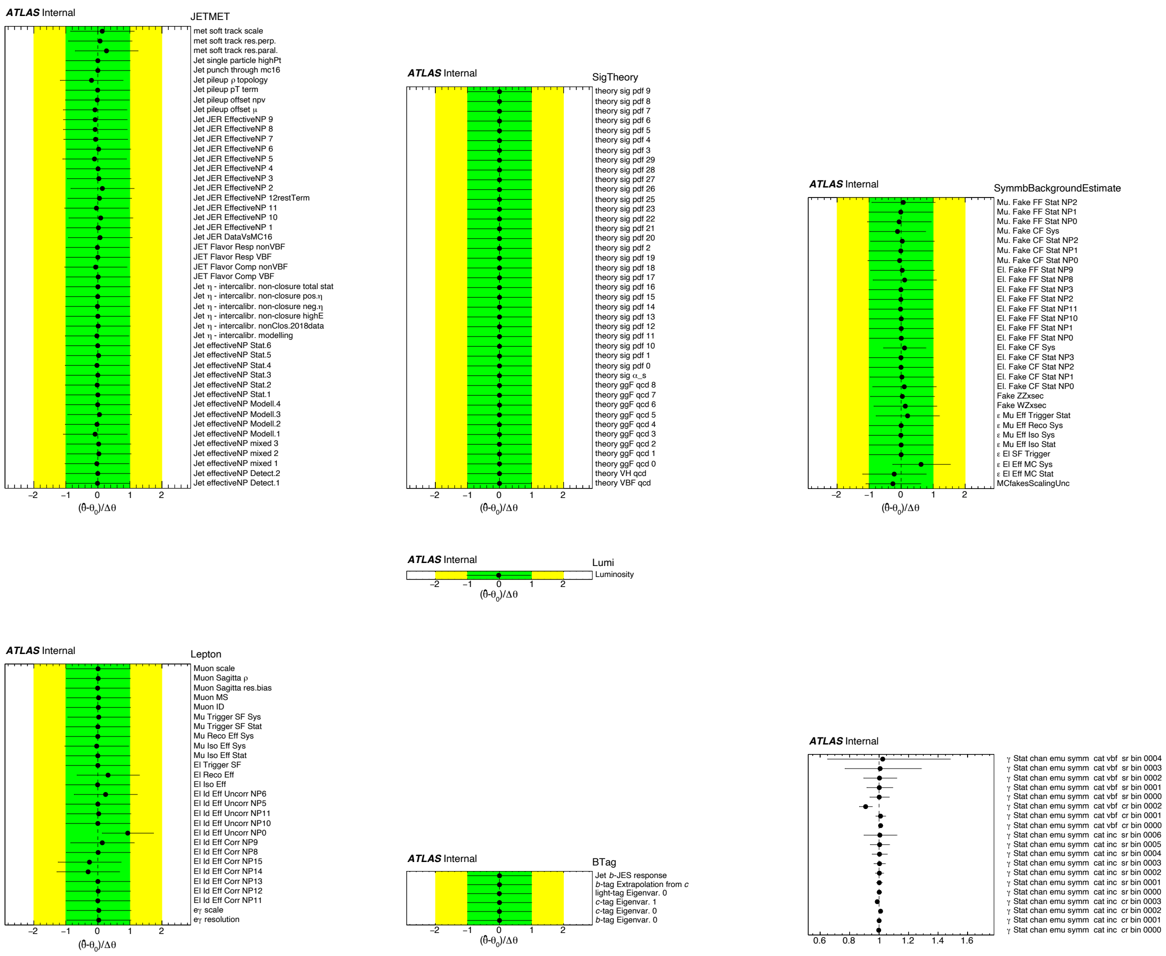


Figure 110: Pulls and constraints of the nuisance parameters in the $H \to e\tau_\mu$ mixed fit. The dots indicate , and the lines represent the constraints. The green band indicates the 1σ band relative to the pre-fit values.

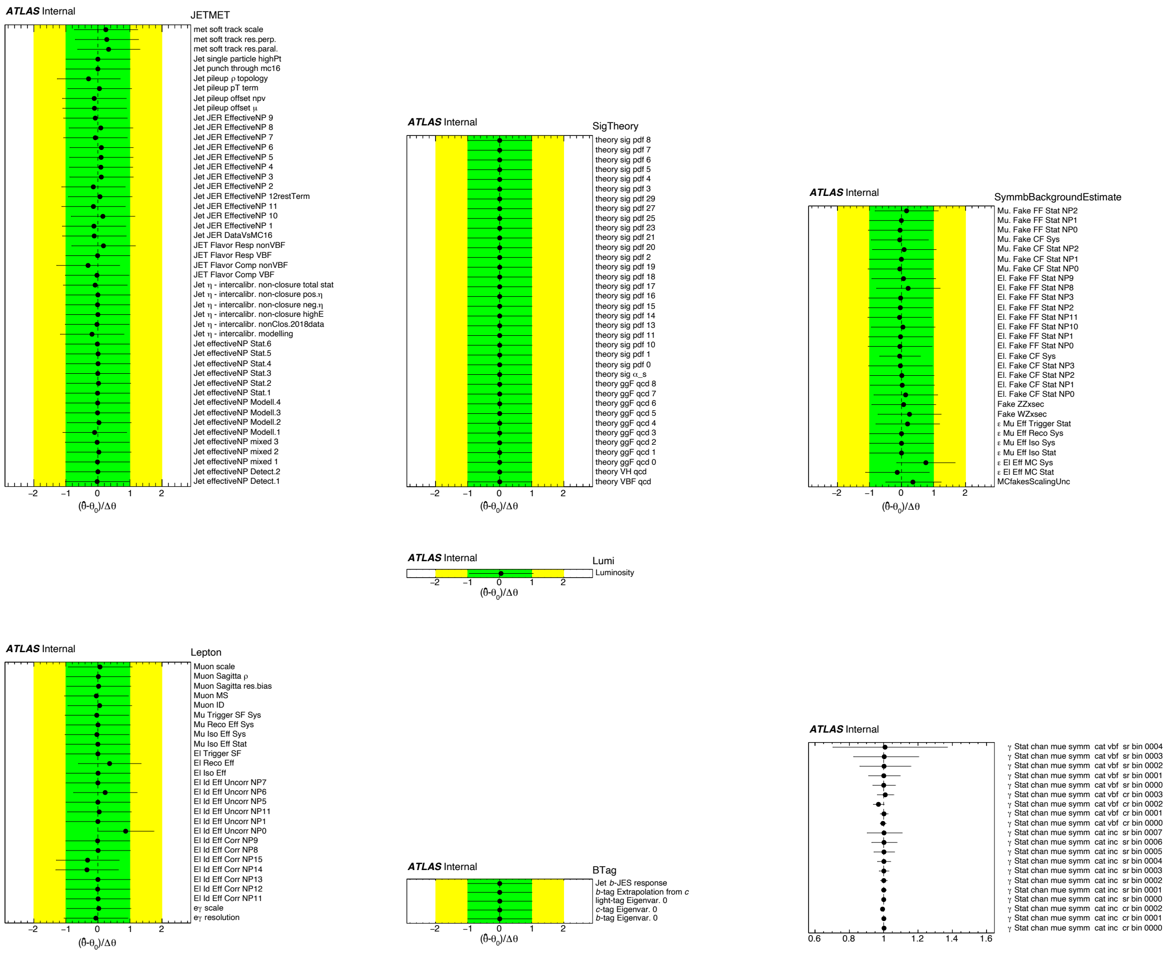
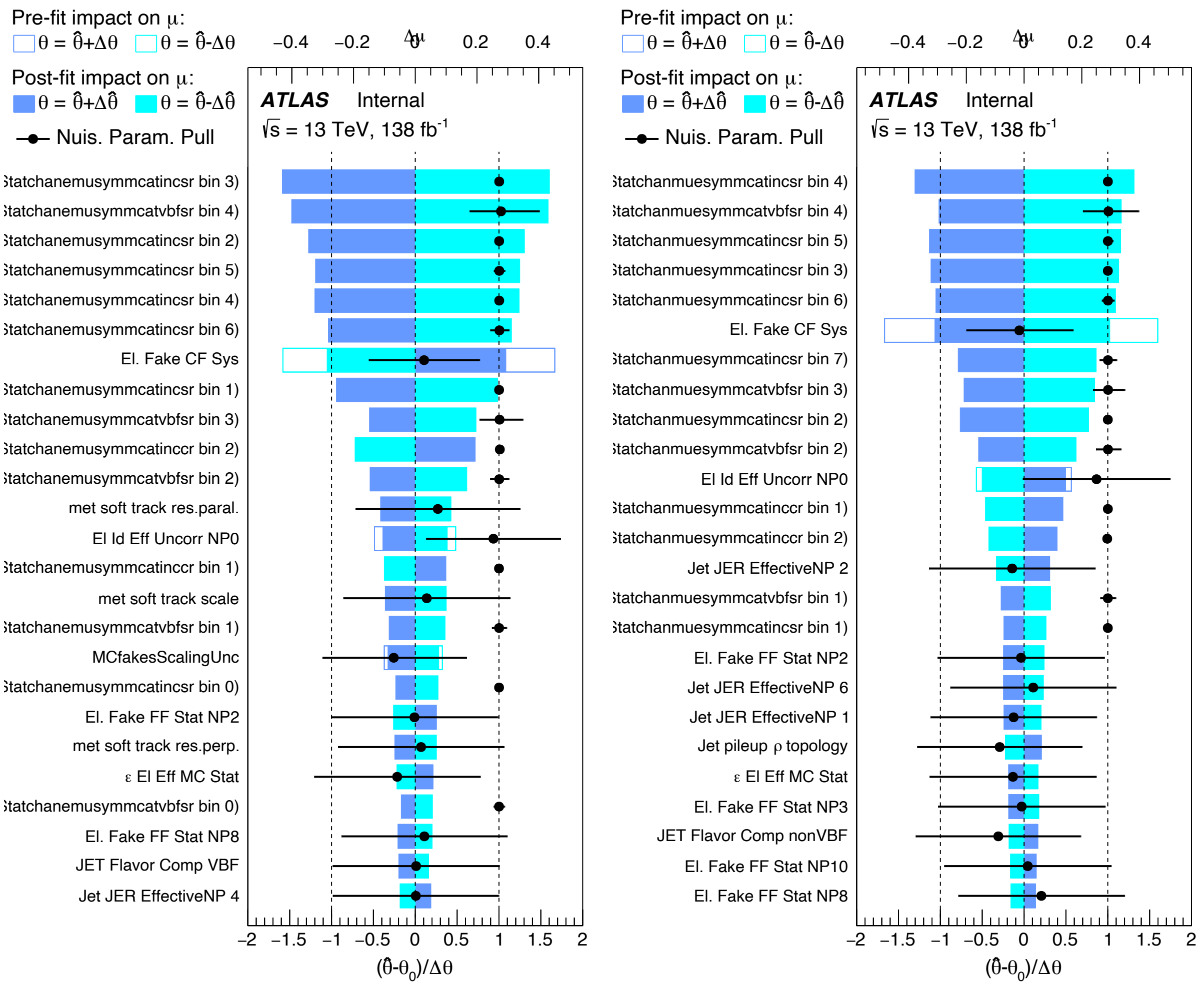


Figure 111: Pulls and constraints of the nuisance parameters in the $H \to \mu\tau_e$ mixed fit. The dots indicate the pulls, and the lines represent the constraints. The green band indicates the 1σ band relative to the pre-fit values.



(a) NP Ranking $e\tau$

(b) NP Ranking $\mu\tau$

Figure 112: The ranking of the nuisance parameters in the search in the $H \to e\tau_\mu$ (left) and $H \to \mu\tau_e$ (right) mixed fits. The postfit impact is indicated by the filled bars, while the pre-fit impact is indicated by the empty bars. In addition, the pulls are indicated by the black dots, and the black lines show the constraints.

D Symmetry DDP - supplementary kinematic distributions

In this section, we provide additional kinematic distribution comparisons, probing the restored symmetry after efficiency correction in various selections and with the different methods presented in sections 5.4 and 5.5.

D.1 Standalone MC simulated samples - efficiency-based correction

The following figures are referred to in sections 5.4.2 and 5.4.3.

- Figure 113 complements the comparison of the truth datasets (see Figure 56)
- Figure 114 complements the comparison of the truth datasets with “lepton showering” deactivated in Pythia (see Figure 57)
- Figure 115 complements the comparison of the reconstructed datasets after the efficiency-based correction (see Figure 58)
- Figure 116 complements the comparison of the reconstructed datasets with “lepton showering” deactivated in Pythia after the efficiency-based correction (see Figure 59)

D.2 Standalone MC simulated samples - MC distribution-based correction

The following figures are referred to in section 5.5.3.

- Figure 117 compares the ee vs. $\mu\mu$ reconstructed datasets after the ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) MC distribution-based correction (see section 5.5.2, Figure 63)
- Figure 118 compares the ee vs. $\mu\mu$ reconstructed datasets after the ($m_{\ell\ell}$ vs. $p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) MC distribution-based correction (see section 5.5.2, Figure 63)
- Figure 119 compares the $e\mu$ vs. μe reconstructed datasets after the ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) MC distribution-based correction (see section 5.5.2, Figure 64)

D.3 ATLAS Run-2 Higgs LFV samples - MC distribution-based correction

The following figures are referred to in section 5.5.3.

- Figures 120 and 121 compare the $e\tau$ and $\mu\tau$ SM MC datasets after the ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) MC distribution-based correction in the non-VBF and VBF SRs, respectively.
- Figures 122 and 123 compare the $e\tau$-based predicted background and $\mu\tau$ data when the ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) MC distribution-based correction is used to derive the symmetric background in the non-VBF and VBF SRs, respectively.
- Figures 124 and 125 compare the $e\tau$ and $\mu\tau$ SM MC datasets after the ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) MC distribution-based correction in the non-VBF and VBF SRs, respectively.
- Figures 126 and 127 compare the $e\tau$-based predicted background and $\mu\tau$ data when the ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) MC distribution-based correction is used to derive the symmetric background in the non-VBF and VBF SRs, respectively.

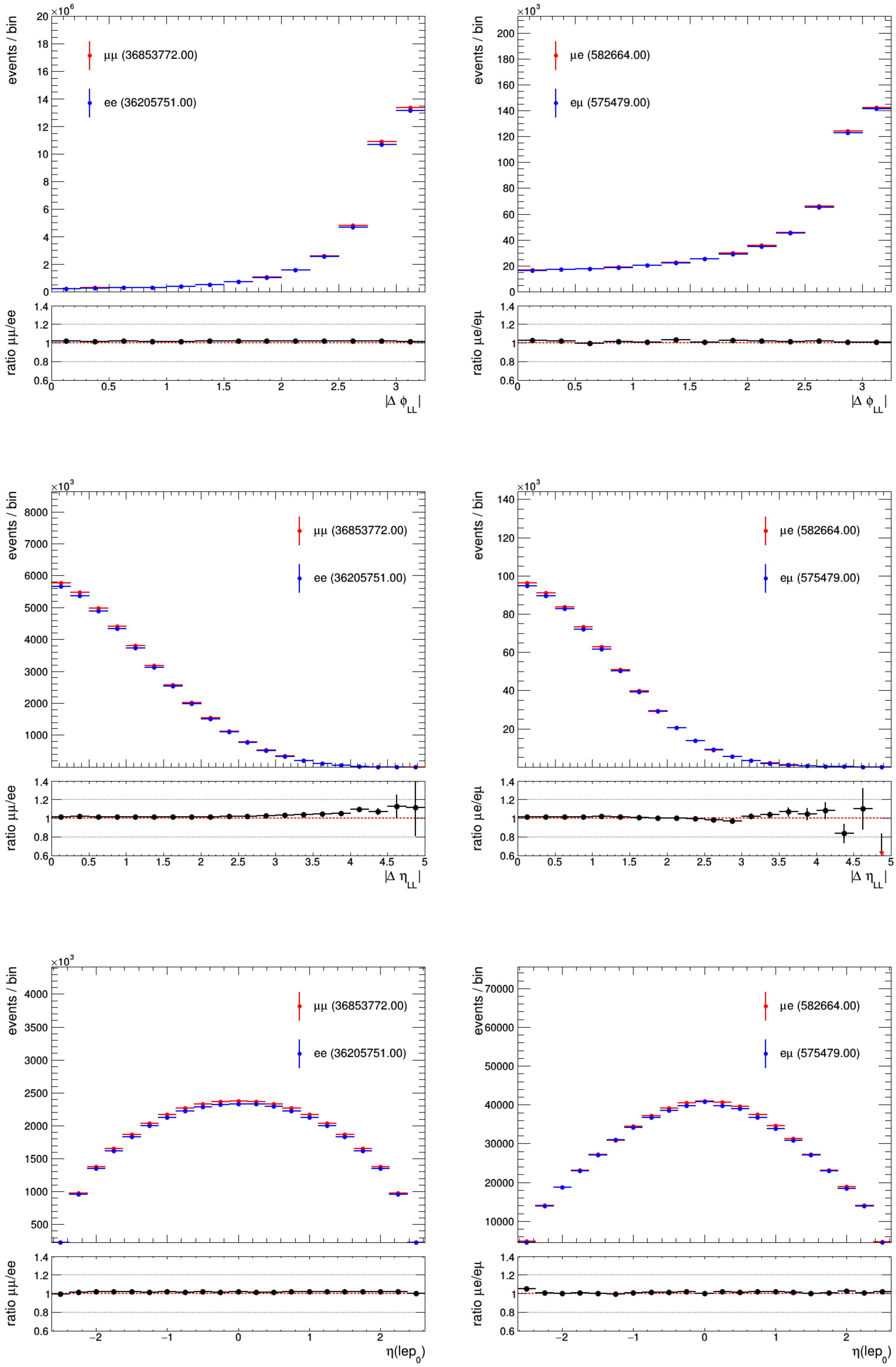


Figure 113: Comparison of ee vs. $\mu\mu$ (left) and $e\mu$ vs. μe (right) truth datasets (more distributions are displayed in Figure 56).

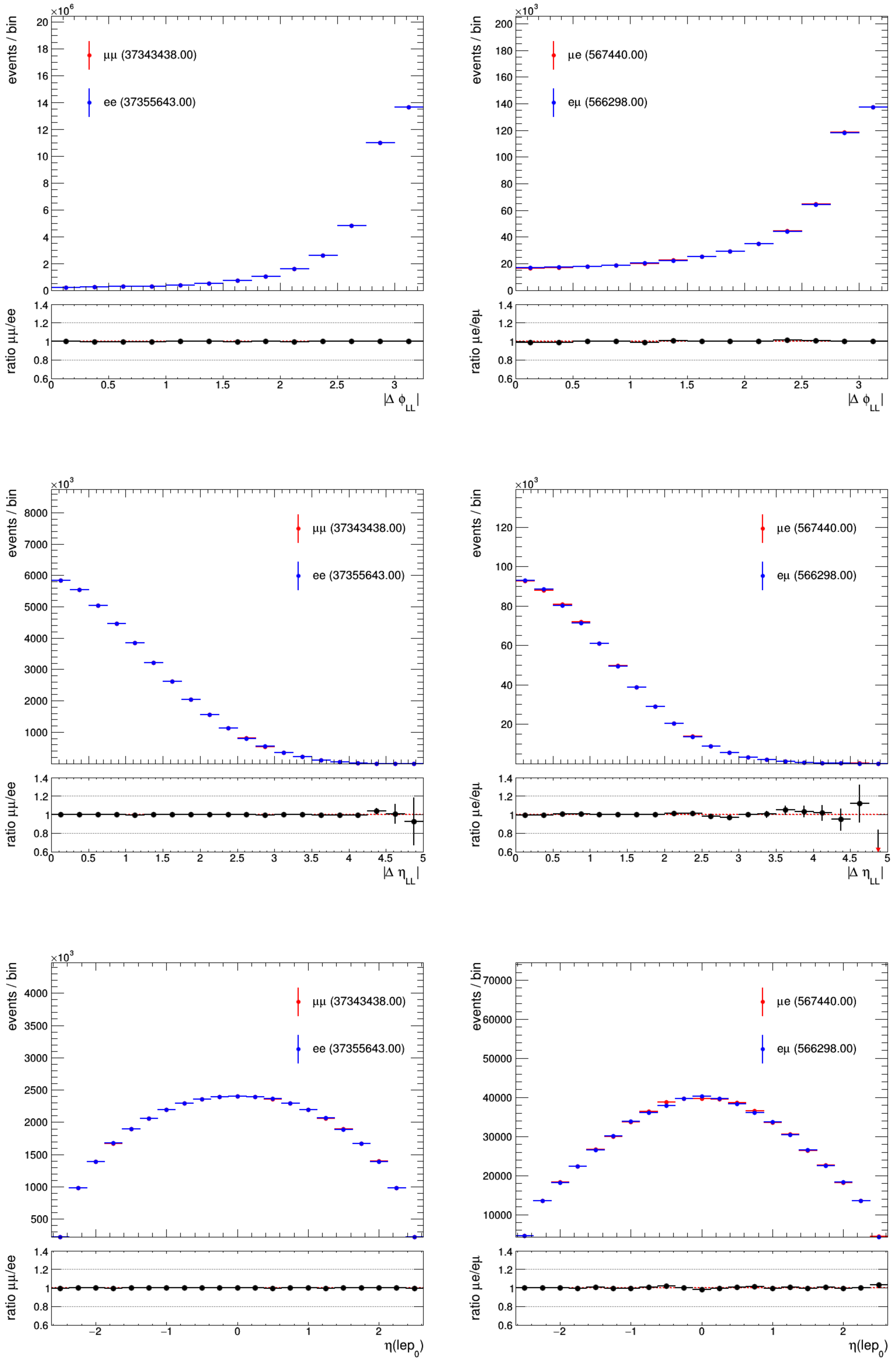


Figure 114: Comparison of ee vs. $\mu\mu$ (left) and $e\mu$ vs. μe (right) truth datasets with “lepton showering” deactivated (more distributions are displayed in Figure 56).

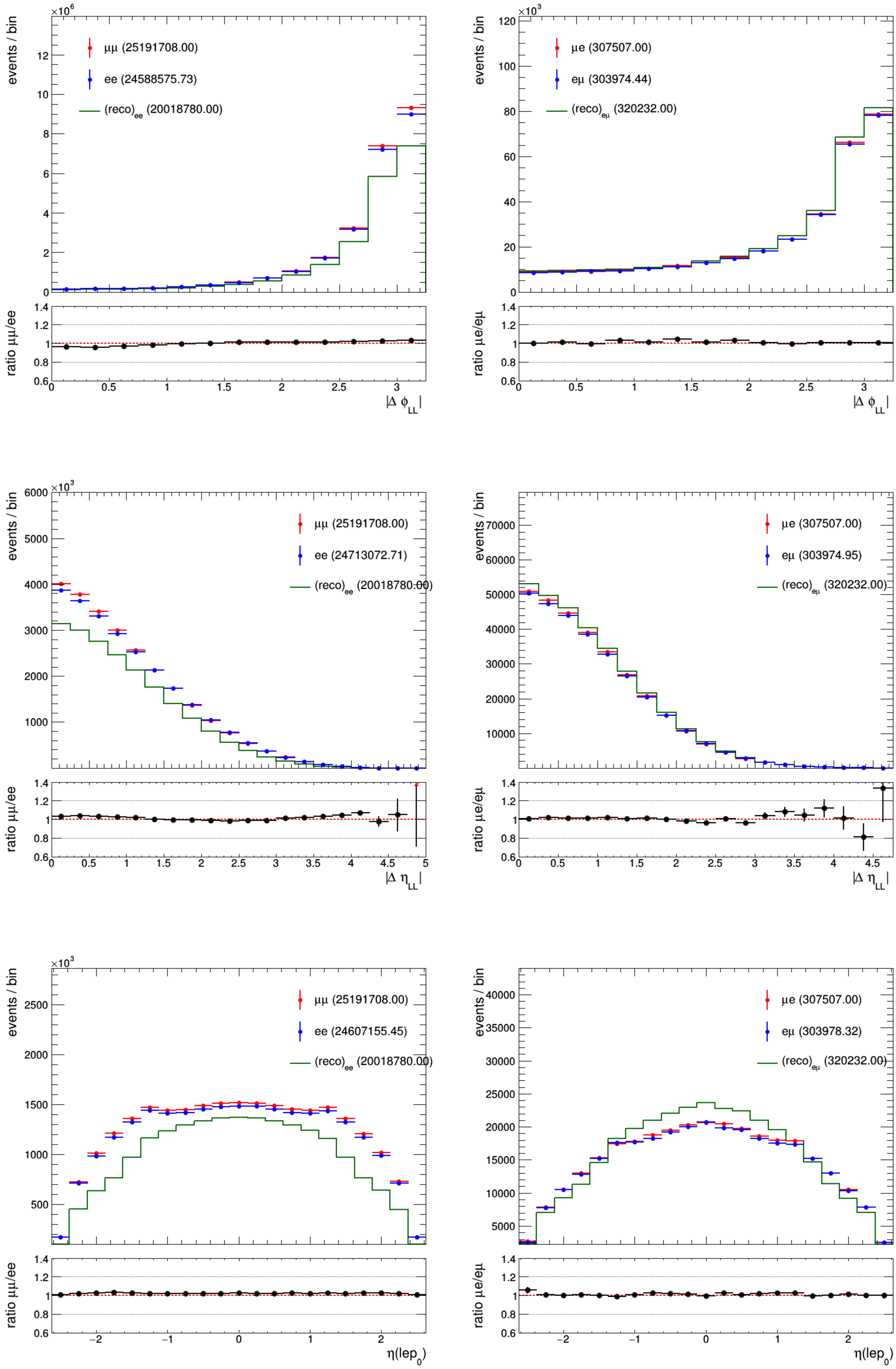


Figure 115: Comparison of ee vs. $\mu\mu$ (left) and $e\mu$ vs. μe (right) reconstructed datasets after the efficiency-based correction is applied to ee or $e\mu$ (more distributions are displayed in Figure 58).

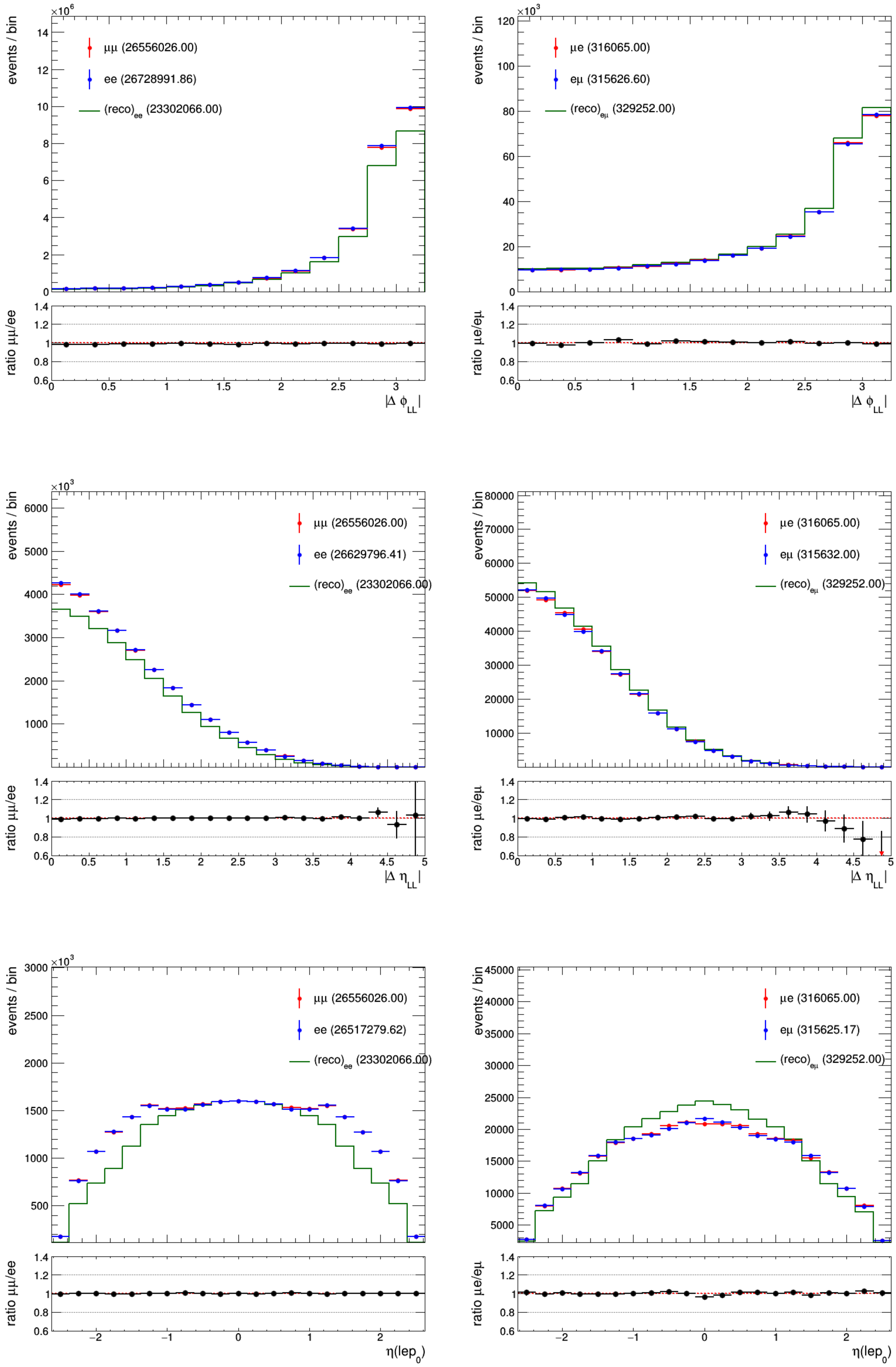


Figure 116: Comparison of ee vs. $\mu\mu$ (left) and $e\mu$ vs. μe (right) reconstructed datasets with "lepton showering" deactivated after the efficiency-based correction is applied to ee or $e\mu$ (more distributions are displayed in Figure 58).

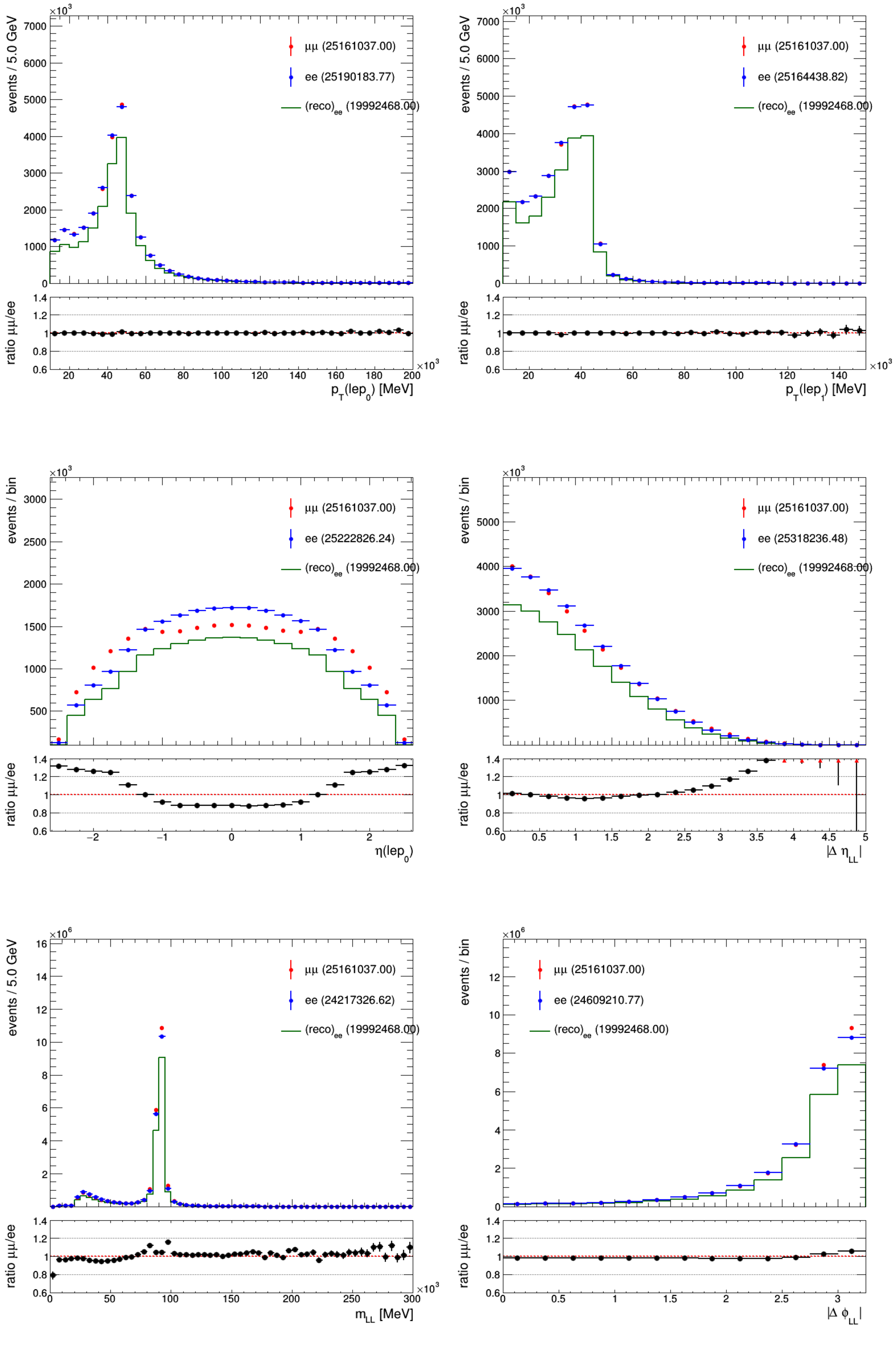


Figure 117: Comparison of the ee vs. $\mu\mu$ reconstructed datasets after the ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) MC distribution-based correction is applied to ee.

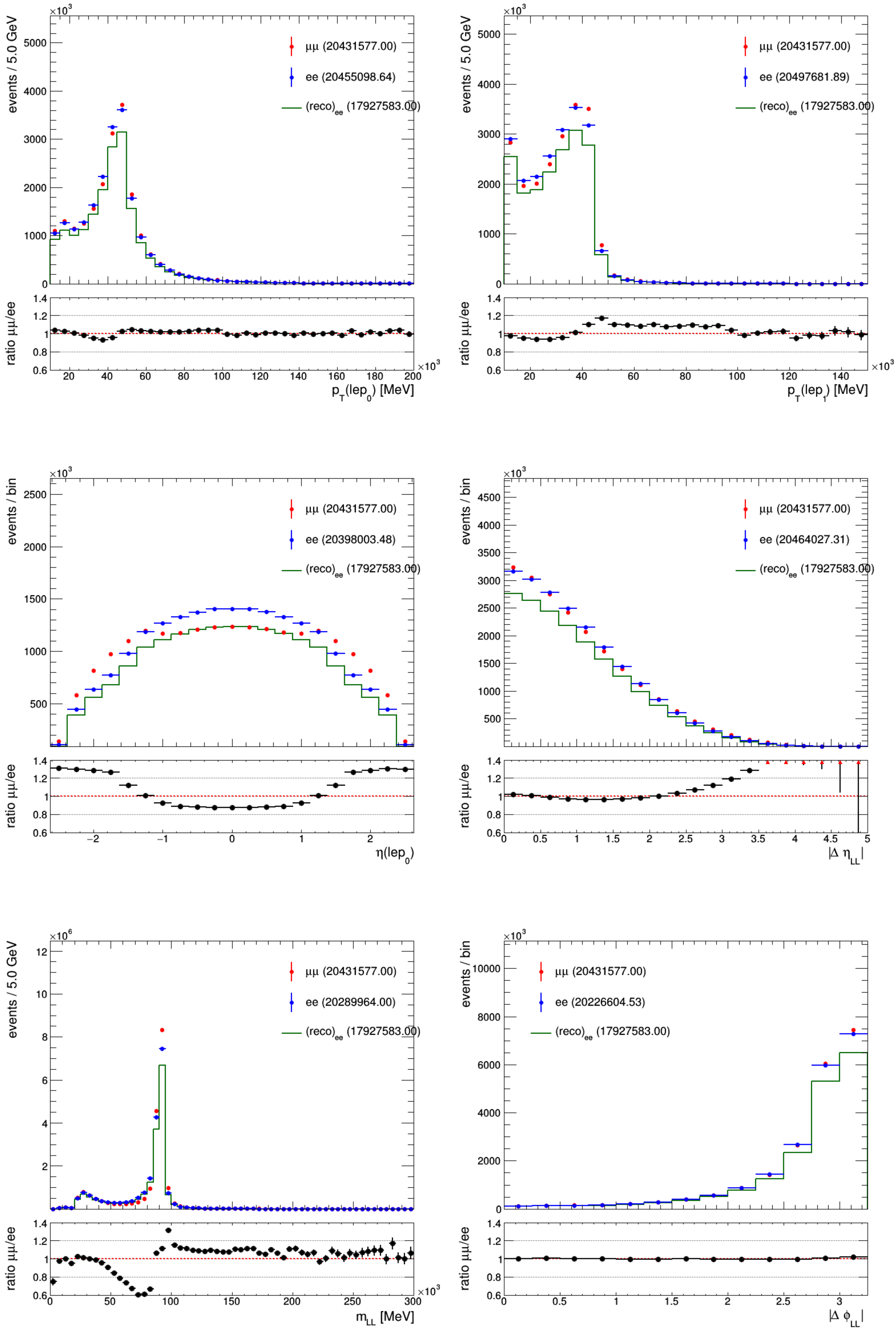


Figure 118: Comparison of the ee vs. $\mu\mu$ reconstructed datasets after the ($m_{\ell\ell}$ vs. $p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) MC distribution-based correction is applied to ee.

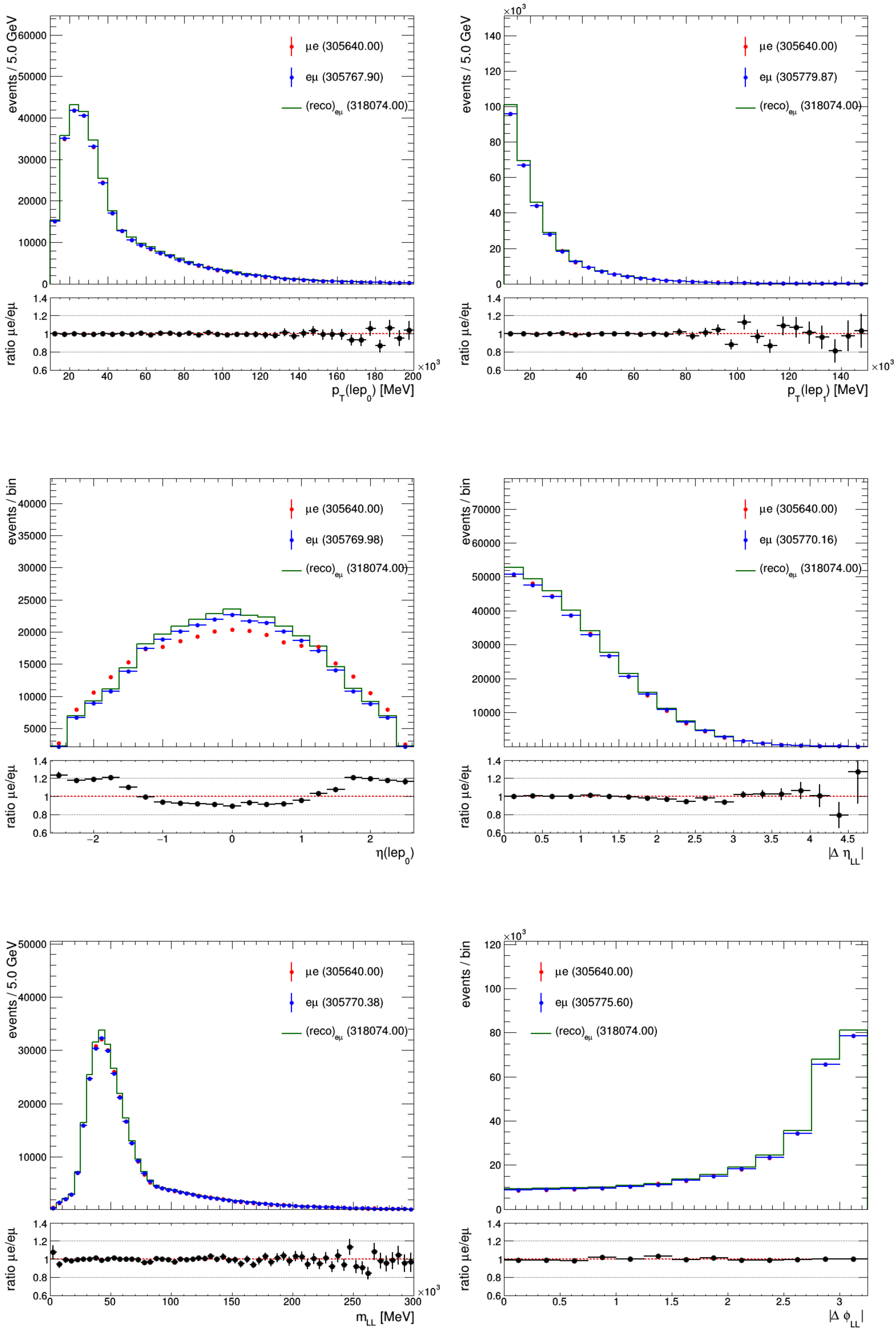


Figure 119: Comparison of the $e\mu$ vs. μe reconstructed datasets after the ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) MC distribution-based correction is applied to $e\mu$.

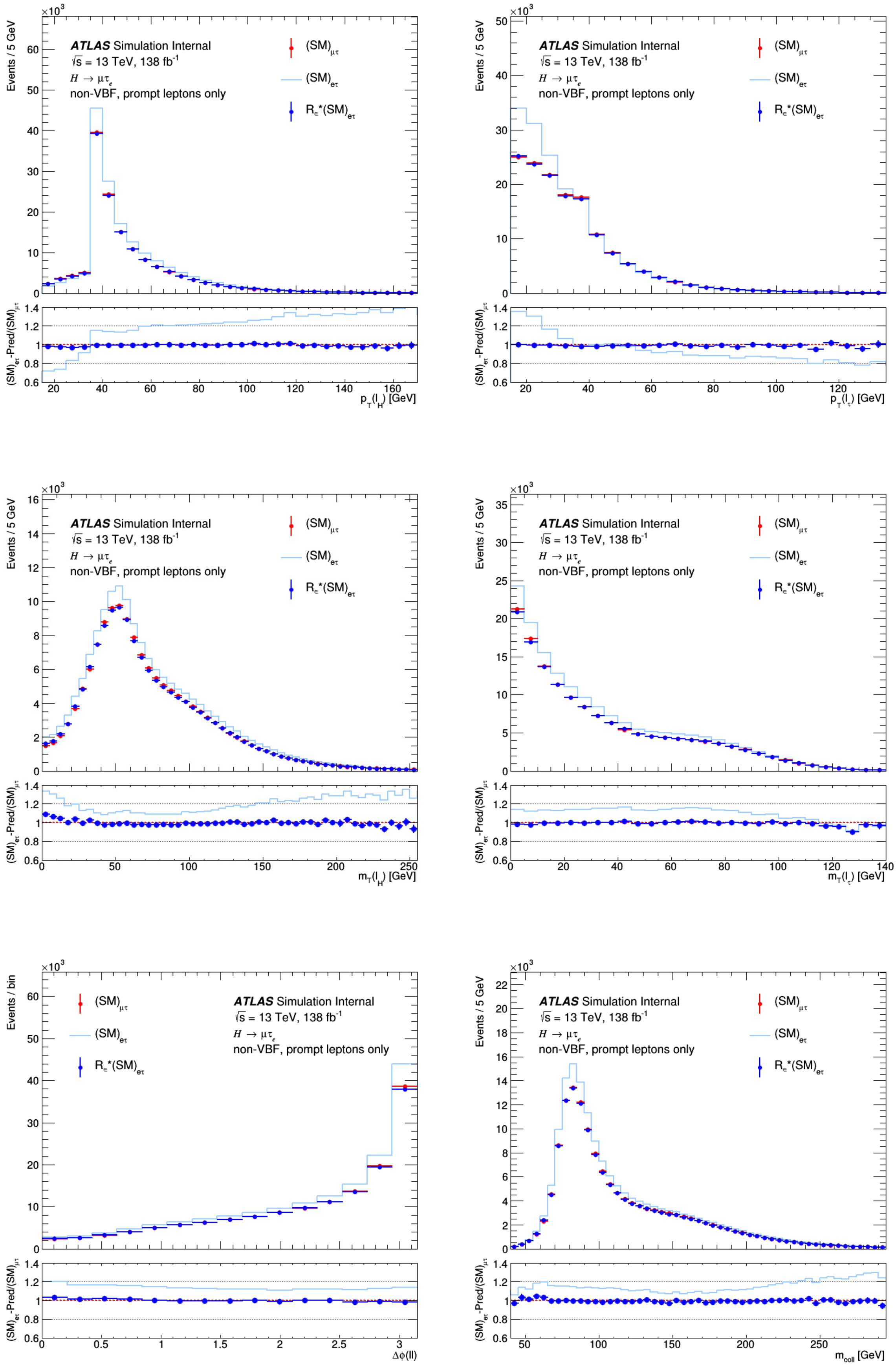


Figure 120: Comparison of $e\tau$ (dark blue) and $\mu\tau$ (red) SM MC datasets in the non-VBF SR after the ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) MC distribution-based correction is applied to $e\tau$. The uncorrected $e\tau$ dataset is also displayed (light blue).

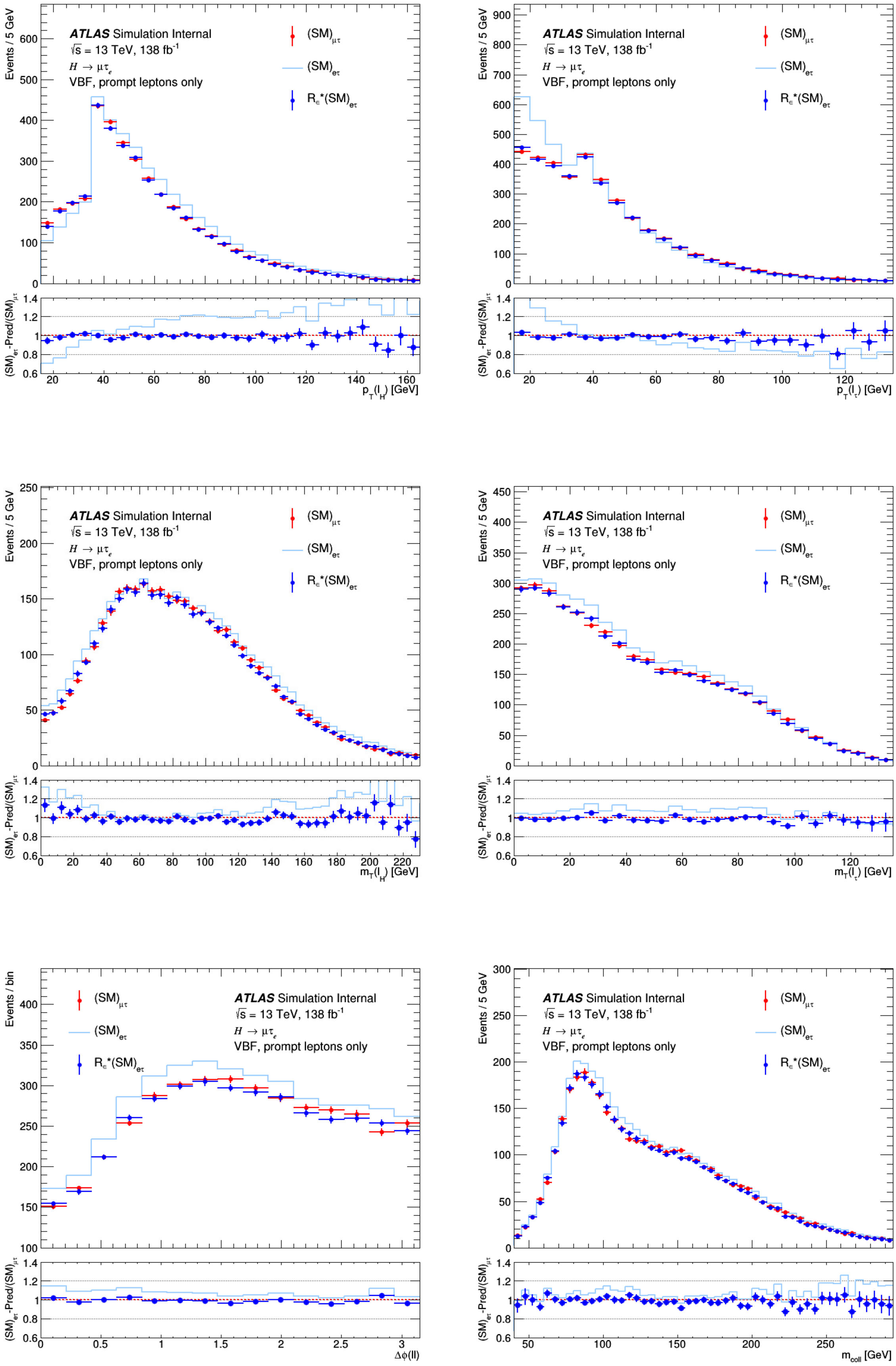


Figure 121: Comparison of $e\tau$ (dark blue) and $\mu\tau$ (red) SM MC datasets in the VBF SR after the ($p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1}$) MC distribution-based correction is applied to $e\tau$. The uncorrected $e\tau$ dataset is also displayed (light blue).

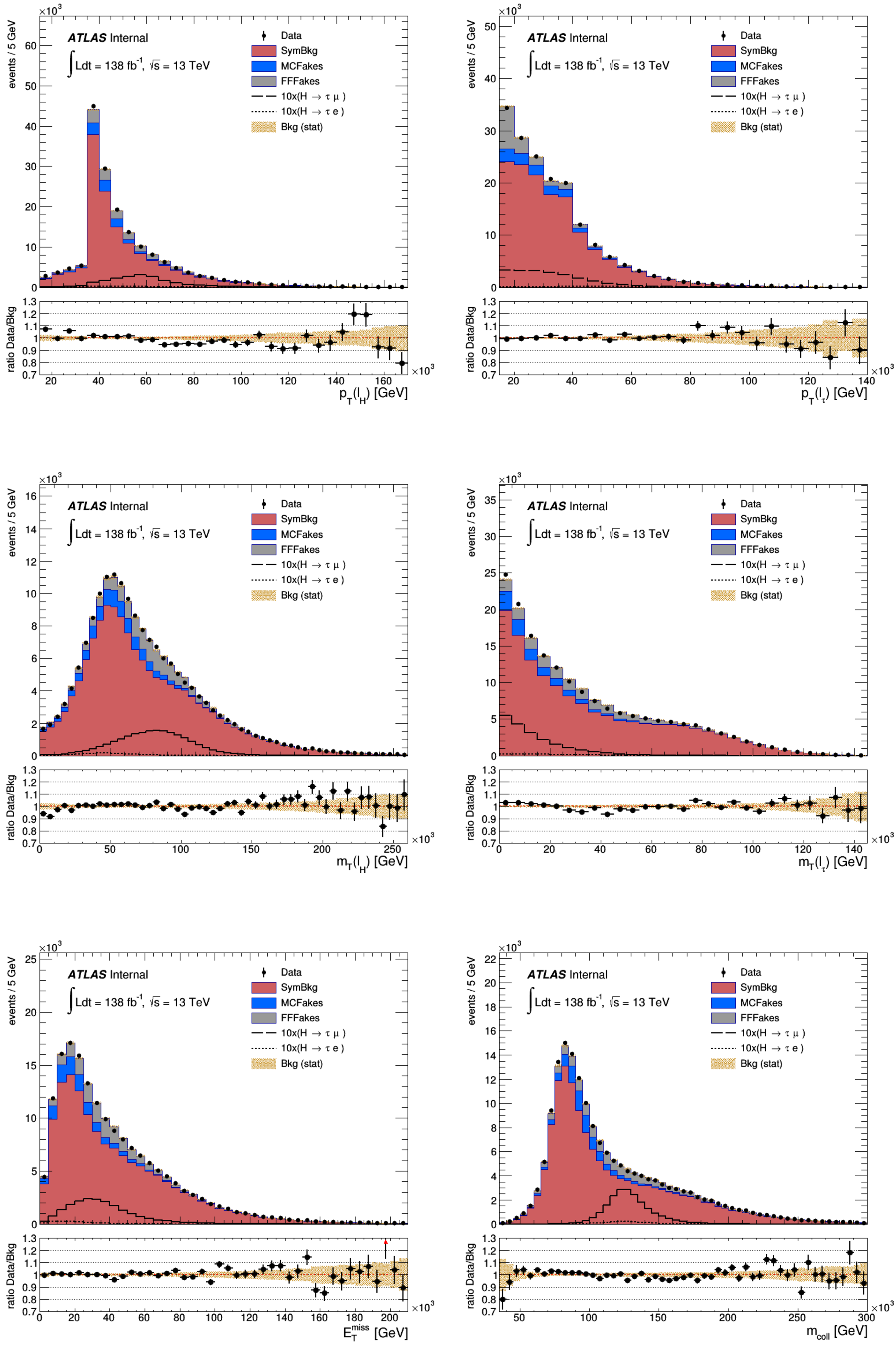


Figure 122: Comparison of the $e\tau$-based predicted background and $\mu\tau$ data in the non-VBF SR when the $(p_{\mathrm{T}}^{\ell_0}$ vs. $p_{\mathrm{T}}^{\ell_1})$ MC distribution-based correction is used to derive the symmetric background. Only statistical uncertainties are displayed.

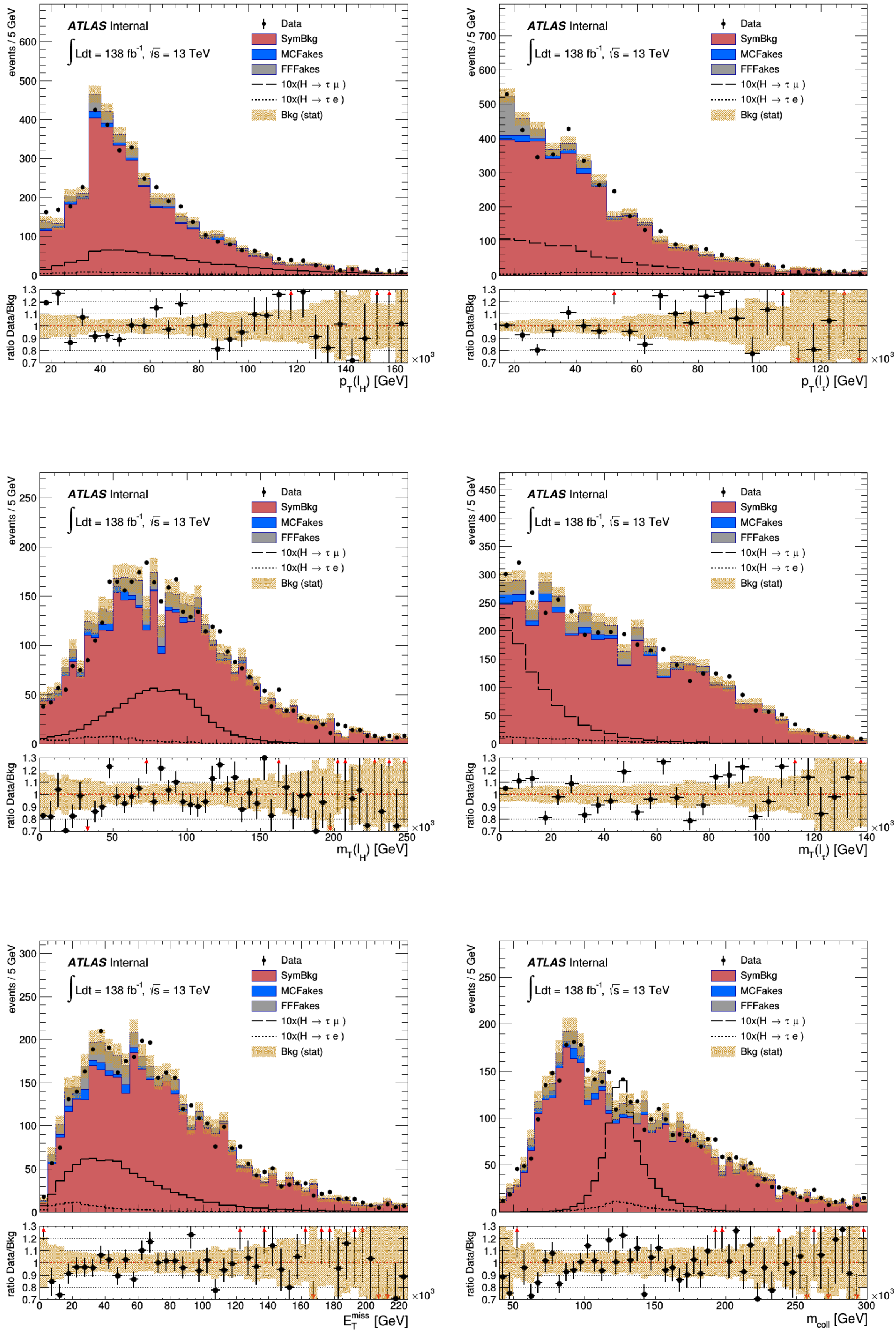


Figure 123: Comparison of the $e\tau$-based predicted background and $\mu\tau$ data in the VBF SR when the ($p_T^{\ell_0}$ vs. $p_T^{\ell_1}$) MC distribution-based correction is used to derive the symmetric background. Only statistical uncertainties are displayed.

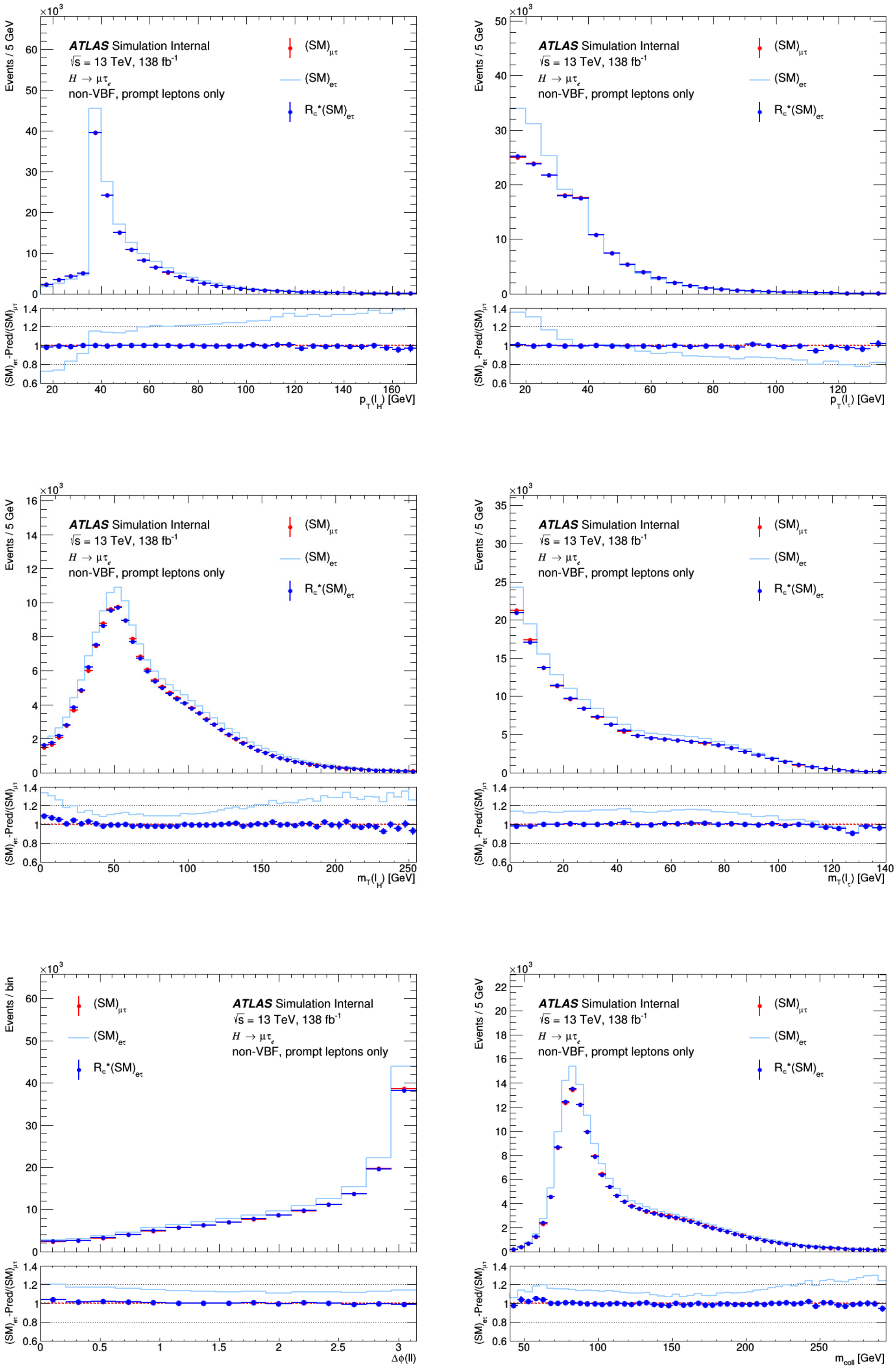


Figure 124: Comparison of $e\tau$ (dark blue) and $\mu\tau$ (red) SM MC datasets in the non-VBF selection after the ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) MC distribution-based correction is applied to $e\tau$. The uncorrected $e\tau$ dataset is also displayed (light blue).

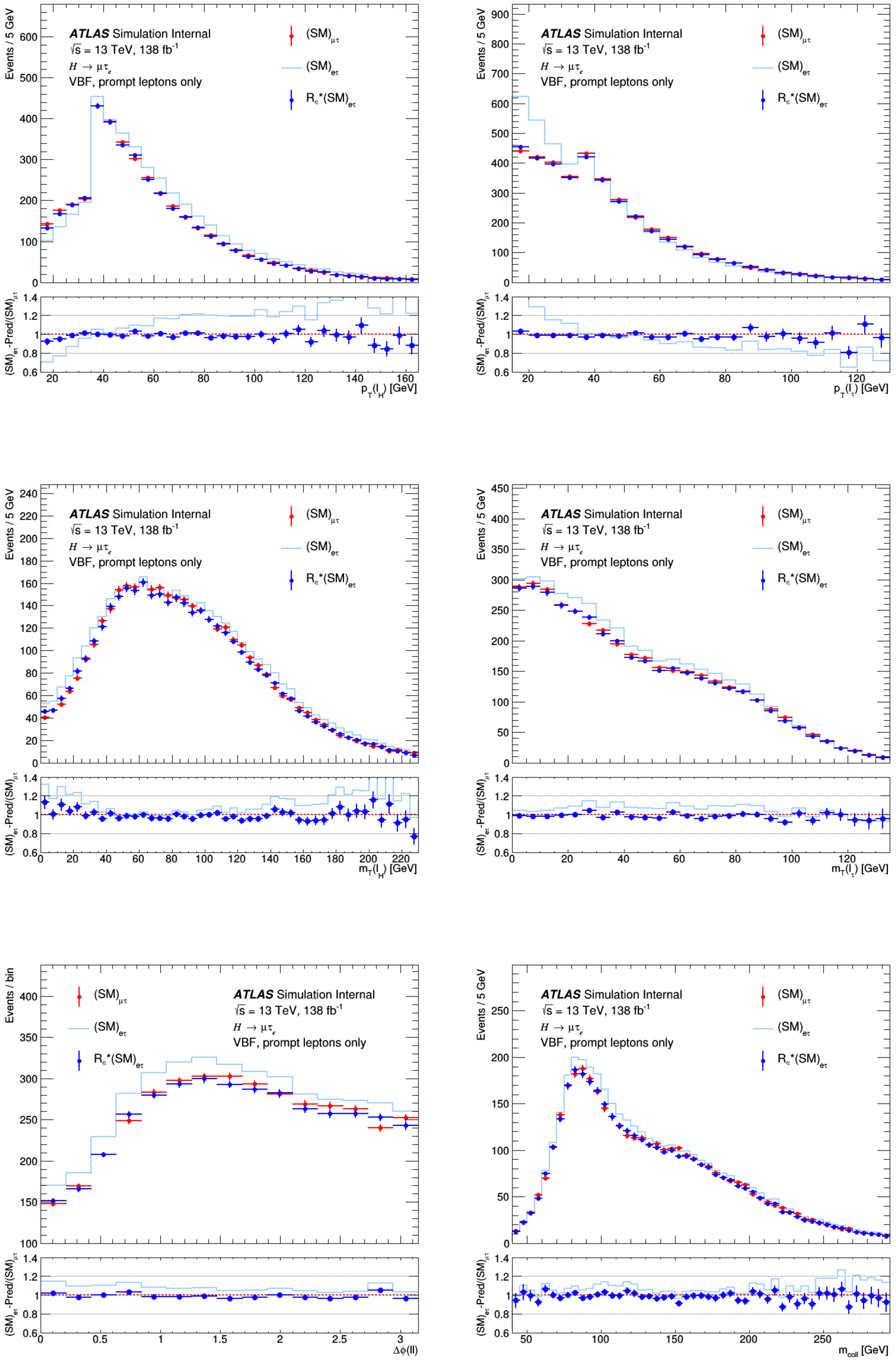


Figure 125: Comparison of $e\tau$ (dark blue) and $\mu\tau$ (red) SM MC datasets in the VBF SR after the ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) MC distribution-based correction is applied to $e\tau$. The uncorrected $e\tau$ dataset is also displayed (light blue).

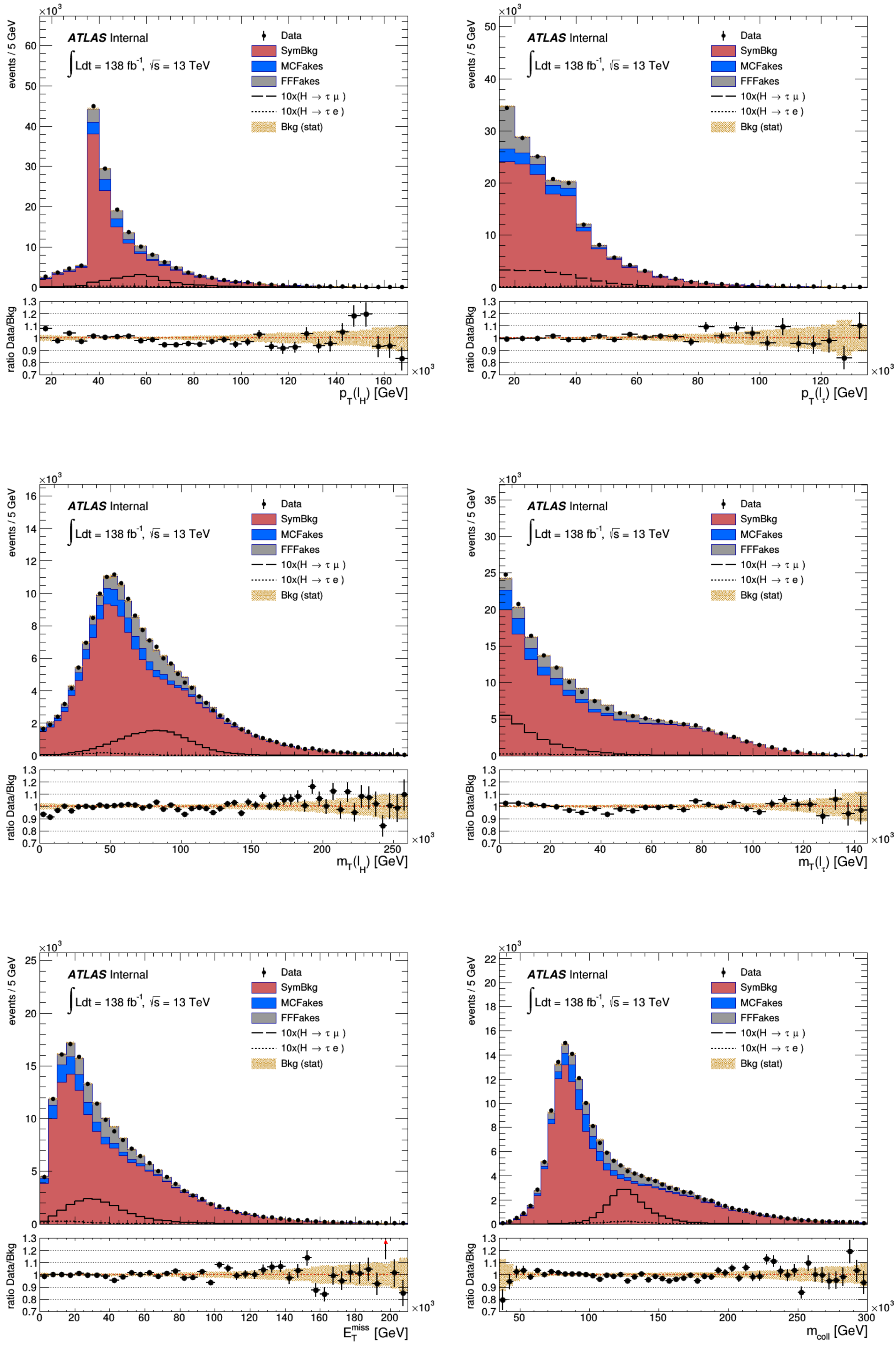


Figure 126: Comparison of the $e\tau$-based predicted background and $\mu\tau$ data in the non-VBF SR when the ($p_{\mathrm{T}}^{\ell_H}$ vs. $p_{\mathrm{T}}^{\ell_\tau}$) MC distribution-based correction is used to derive the symmetric background. Only statistical uncertainties are displayed.

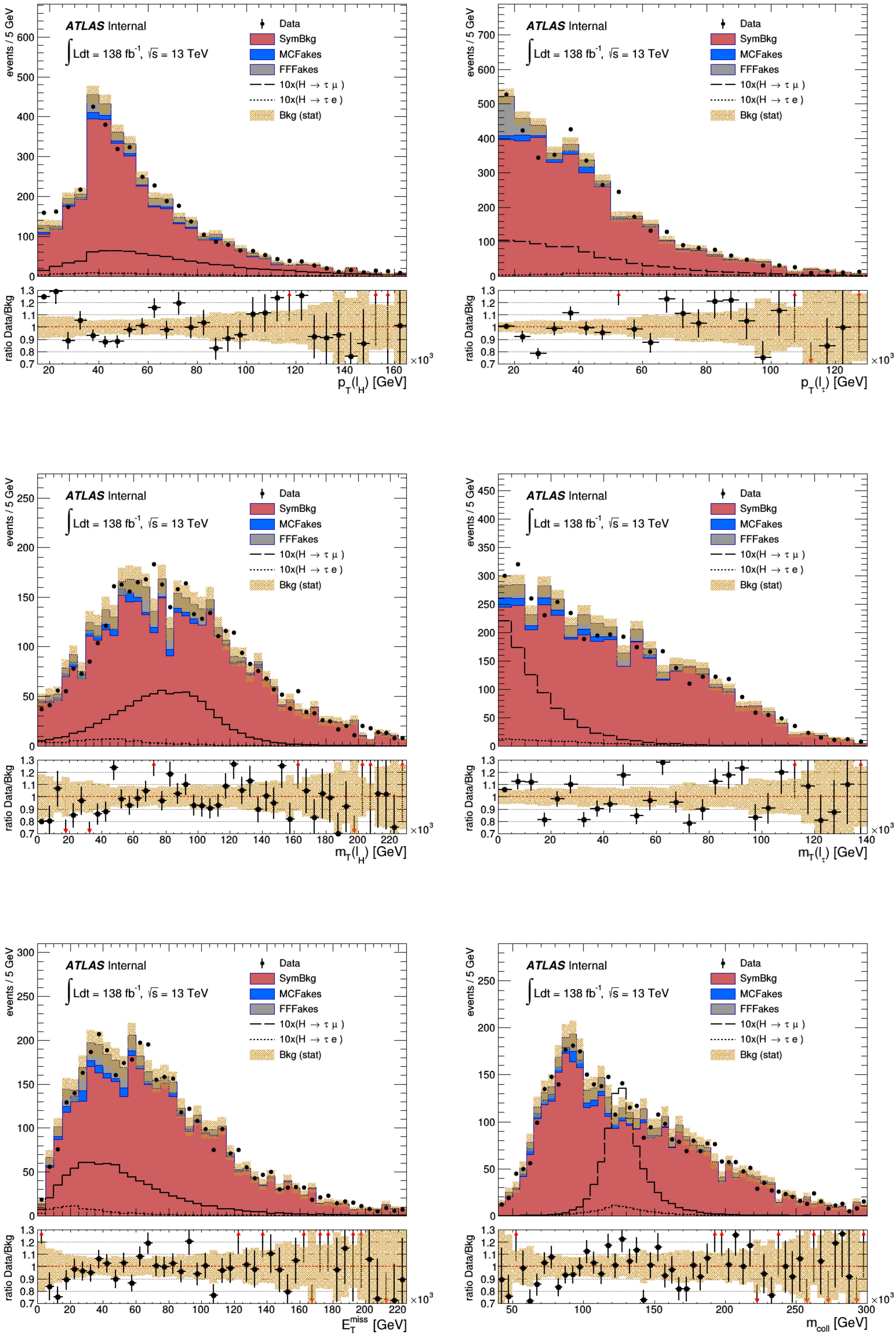


Figure 127: Comparison of the $e\tau$-based predicted background and $\mu\tau$ data in the VBF SR when the $(p_T^{\ell_H}$ vs. $p_T^{\ell_\tau})$ MC distribution-based correction is used to derive the symmetric background. Only statistical uncertainties are displayed.