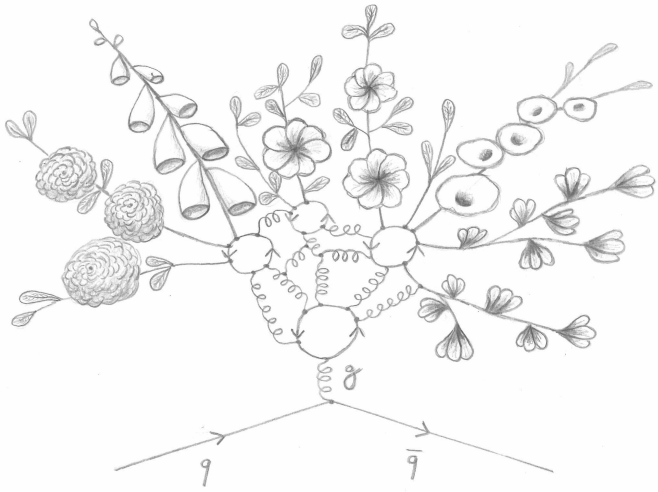
Measurements of rare processes with top quarks and electroweak bosons in proton-proton collisions at the CMS detector

Luka Lambrecht

A dissertation submitted in fulfillment of the requirements for the degree of
Doctor in the Sciences: Physics

Ghent University
Faculty of Sciences
Department of Physics and Astronomy



September 2024

Author Luka Lambrecht (Ghent University)

Supervisors: prof. dr. Didar Dobur (Ghent University)
prof. dr. Filip Moortgat (CERN)

This thesis was submitted to the board meeting of the Faculty of Sciences of Ghent University on May 22, 2024, presented to the examination board on July 10, 2024, and presented to the public on October 4, 2024.

Chairperson of the examination board: prof. dr. Nele Vandersickel (Ghent University)

Members of the examination board: dr. Pieter Everaerts (University of Wisconsin)
dr. Alexander Grohsjean (DESY)
prof. dr. Ben Page (Ghent University)
dr. Kirill Skovpen (Ghent University)

Door de wolkenweelde worden overmand
wanneer de morgenster de dag bereidt,
wanneer de stilte van het grijze land
de verbeelding over de horizon leidt;

een wereld zien in een korrel zand
en een hemel in een bloementapijt;
oneindigheid vatten in de palm van je hand
en in elk moment een eeuwigheid. [1]

Met klokken en latten zal je slechts leren
dat de maat van de muziek der sferen
zich niet in solsleutels vangen laat.

Hoe meer we proberen de wereld te beheren
en in systemen categoriseren,
hoe groter dat wat onze grip ontgaat.

[1] Naar W. Blake - *Auguries of Innocence*

Contents

Abstract

Introduction This dissertation studies several rare elementary particle processes involving top quarks, in an attempt to test and challenge our current knowledge of the fundamental workings of the Universe. The standard model of particle physics is a comprehensive and quantitative description of all known elementary particles and their interactions. Despite its tremendous success in accurately predicting probabilities and characteristics of elementary particle processes, tested and verified in many experiments, it is known to be incomplete. This dissertation fits in the context of pushing the standard model to its limits, by studying rare processes involving the heaviest known elementary particle, the top quark, with the aim of finding clear deviations from the prediction, to be interpreted either as signs of new physics effects or of specific modeling issues.

Data The analyses in this dissertation were performed on data recorded with the CMS detector at the CERN LHC. The Large Hadron Collider (LHC) is currently the world's largest and most powerful particle accelerator and collider. The Compact Muon Solenoid (CMS) is a composite detector that aims to detect the outgoing particles and reconstruct the underlying fundamental interaction of the proton-proton collisions in its center. The data sample used in this work was collected by CMS in the years 2016–2018 and consists of 138 fb^{-1} of proton-proton collisions at a center-of-mass energy of 13 TeV.

Processes In particular, this dissertation focuses on tZq (simultaneous production of a single top quark and a Z boson) [1] and $\text{t}\bar{\text{t}}\text{W}$ (production of a top quark pair together with a W boson) [2]. Both processes are rare because of the heavy final state and the electroweak interaction(s), and are expected to be sensitive to several scenarios for extensions of the standard model. The tZq process is particularly interesting because of its purely electroweak production mode and direct coupling between the top quark and Z boson. The $\text{t}\bar{\text{t}}\text{W}$ process is notoriously difficult to model because of the large electroweak and higher order QCD corrections.

tZq The theoretically predicted cross-section for the tZq process where the Z boson decays to a pair of electrons, muons or τ leptons is $\sigma^{SM}(\text{tZq}) = 94.2^{+1.9}_{-1.8}$ (scale) ± 2.5 (PDF) fb [3]. The calculation performed at next-to-leading order in the five-flavour scheme and also includes non-resonant lepton pair production with an invariant mass above 30 GeV. In this work, we measure a corresponding cross-section of $\sigma^{\text{observed}}(\text{tZq}) = 87.9^{+7.5}_{-7.3}$ (stat) $^{+7.3}_{-6.0}$ (syst) fb. The total uncertainty is about 11%, which makes this the most precise measurement of the tZq cross section so far. The measurement is in agreement with the standard model prediction within experimental and theoretical uncertainties, and hence no signs of new physics effects are found for this process so far.

$\text{t}\bar{\text{t}}\text{W}$ In the case of $\text{t}\bar{\text{t}}\text{W}$, we perform differential cross-section measurements as a function of several variables such as the kinematic properties of the leptons and jets in the events. The results of these measurements can be used as a reference to test modeling improvements against. This analysis is currently in CMS internal review, and therefore the results shown in this work are expected outcomes based on simulation.

Lepton identification In both the tZq and $\text{t}\bar{\text{t}}\text{W}$ analyses, a vital ingredient is the accurate identification of leptons (electrons and muons), and the estimation of background processes with fake or secondary leptons. The extensive research reported in this dissertation regarding the estimation of this background constituted an important contribution to several other related measurements, such as the simultaneous production of four top quarks [4] and a search for heavy neutral leptons [5].

Displaced vertexing calibration In addition to the above, this dissertation contains a study of displaced tracking and vertexing [6–8]. This study is carried out in the context of searches for long-lived heavy neutral leptons or other exotic experimental signatures involving hypothesized particles with relatively long lifetimes, decaying at macroscopic distances from the proton-proton interaction point. The reconstruction efficiency of these displaced particles is typically estimated from simulation, but is difficult to calibrate. In this dissertation, we use the decay of the neutral K meson to two pions ($K^0_S \to \pi^+\pi^-$) to assess the accuracy of the simulation of displaced tracking and vertexing and to derive calibration factors. The reconstruction of displaced K^0_S vertices is compared between data and simulation, as a function of the radial displacement distance. We obtain calibration factors generally in the order of 10–15% for large displacements, with a strong dependency on dynamic detector inefficiencies. The study is performed for several reprocessing versions of the CMS data and simulation. In particular, we observe the strong impact of an issue with the tracker readout electronics in the early half of 2016 data taking, which is reduced by a reprocessing of the simulation.

Automatic data quality monitoring Finally, this dissertation shows a few case studies of applying machine learning to DQM and DC [9,10]. Data quality monitoring (DQM) and data certification (DC) consist of ensuring the good quality of data recorded and reconstructed by CMS, and are vital for its correct interpretation and for the reliability of physics analyses using it. We apply machine learning tools, in particular (convolutional, residual) autoencoders to automatically detect potential detector issues in the CMS pixel tracker, with a finer time granularity than would be feasible with the conventional techniques. We show that an anomaly detection efficiency of essentially 100% can be achieved, with a sufficiently low false alarm rate in the order of 10^{-5}, though the test set to obtain these numbers cannot be unambiguously defined. Furthermore, important technical advances were made to provide the quality monitoring data with this finer time granularity as a standard workflow in CMS.

Samenvatting (Dutch)

Inleiding Dit proefschrift onderzoekt verschillende zeldzame elementaire deeltjesprocessen met top quarks, in een poging onze huidige kennis van de fundamentele werking van het universum te testen en uit te dagen. Het standaardmodel van de deeltjesfysica is een veelomvattende en kwantitatieve beschrijving van alle bekende elementaire deeltjes en hun wisselwerkingen. Ondanks het enorme succes van dit model in het nauwkeurig voorspellen van de waarschijnlijkheden en eigenschappen van elementaire deeltjesprocessen, getest en geverifieerd in vele experimenten, is het aantoonbaar onvolledig. Dit proefschrift past binnen de context van het testen van het standaardmodel in zo extreem mogelijke omstandigheden, via het bestuderen van zeldzame processen met het zwaarst bekende elementaire deeltje, het top quark, met als doel het vinden van duidelijke afwijkingen van de voorspellingen, en deze te interpreteren hetzij als tekenen van nieuwe fysische effecten, hetzij als specifieke modelleerproblemen.

Data De analyses in dit proefschrift werden uitgevoerd met data die gecollecteerd werden met de CMS detector op de CERN LHC. De Large Hadron Collider (LHC - grote hadronenbotser) is momenteel 's werelds grootste en krachtigste deeltjesvesneller en -botser. De Compact Muon Solenoid (CMS - compacte muonsolenoïde) is een samengestelde detector die tot doel heeft de uitgaande deeltjes te detecteren en de onderliggende fundamentele interactie te reconstrueren van de proton-protonbotsingen in het middelpunt. Het datasample gebruikt in dit werk werd gecollecteerd door CMS in de jaren 2016–2018 en bestaat uit 138 fb^{-1} aan proton-protonbotsingen met een massamiddelpuntsenergie van 13 TeV.

Processen In het bijzonder focust dit proefschrift op tZq (de gelijktijdige productie van een enkel top quark en een Z boson) [1] en op $\text{t}\bar{\text{t}}\text{W}$ (de productie van een top quark paar samen met een W boson) [2]. Beide processen zijn zeldzaam wegens de zware eindproducten en de elektrozwakke interacties(s), en worden verwacht ontvankelijk te zijn voor verscheidene scenarios voor uitbreidingen van het standaardmodel. Het tZq proces is bijzonder interessant vanwege het puur elektrozwakke productiemechanisme en de rechtstreekse koppeling tussen het top quark en het Z boson. Het $\text{t}\bar{\text{t}}\text{W}$ proces is moeilijk te modelleren vanwege de grote elektrozwakke en hogere-orde QCD correcties.

tZq De theoretisch voorspelde werkzame doorsnede voor het tZq proces waarbij het Z boson vervalt naar een paar van leptonen (hetzij elektronen, muonen of τ leptonen) is $\sigma^{SM}(\text{tZq}) = 94.2^{+1.9}_{-1.8}$ (scale) $\pm$ 2.5 (PDF) fb [3]. Die berekening werd uitgevoerd in op-een-na-hoogste orde en met de five-flavour aanpak, en bevat ook niet-resonante leptonparen met invariante massa groter dan 30 GeV. In deze thesis meten we een corresponderende werkzame doorsnede van $\sigma^{\text{observed}}(\text{tZq}) = 87.9^{+7.5}_{-7.3}$ (stat) $^{+7.3}_{-6.0}$ (syst) fb. De totale meetonzekerheid is ongeveer 11%, wat dit de meest nauwkeurige meting van de tZq werkzame doorsnede tot nu toe maakt. De gemeten waarde is in overeenstemming met de voorspelling volgens het standaardmodel binnen de experimentele en theoretische onzekerheden, en bijgevolg worden er tot nu toe geen sporen van nieuwe fysische effecten aangetroffen voor dit proces.

$\text{t}\bar{\text{t}}\text{W}$ In het geval van $\text{t}\bar{\text{t}}\text{W}$ meten we de differentiële werkzame doorsnede als een functie van verscheidene variabelen, zoals de kinematische eigenschappen van de leptonen en jets in de events. De resultaten in dit proefschrift kunnen worden gebruikt om verbeteringen in het modelleren tegen te testen. Deze analyse bevindt zich momenteel nog in het stadium van intern nazicht binnen CMS, en daarom zijn de resultaten die in dit werk worden getoond enkel de verwachte resultaten gebaseerd op simulatie.

Leptonidentificatie In zowel de tZq- als $\text{t}\bar{\text{t}}\text{W}$-analyse is het nauwkeurig identificeren van

leptonen (elektronen en muonen) een cruciaal ingrediënt, evenals het inschatten van achtergrondprocessen met verkeerd gereconstrueerde of secundaire leptonen. Het uitgebreide onderzoek gerapporteerd in dit proefschrift betreffende het schatten van deze achtergrond heeft een belangrijke bijdrage gevormd tot verschillende andere gerelateerde metingen, waaronder de observatie van de gelijktijdige productie van vier top quarks [4] en een zoektocht naar zware neutrale leptonen [5].

Kalibratie van excentrische vertexreconstructie Bovenop het reeds genoemde, bevat dit proefschrift een studie van excentrische track- en vertexreconstructie. Die studie past in de context van zoektochten naar metastabiele zware neutrale leptonen of andere exotische experimentele topologieën met hypothetische deeltjes met een relatief lange levensduur, vervallend op macroscopische afstand van het proton-proton interactiepunt. De reconstructie-efficiëntie van deze excentrische deeltjes wordt typisch ingeschat met behulp van simulatie, maar is moeilijk te kalibreren. In dit proefschrift gebruiken we het verval van het neutrale K meson naar twee pionen ($K^0_S \to \pi^+\pi^-$) om de nauwkeurigheid van het simuleren van excentrische track- en vertexreconstrucie in te schatten, en om kalibratiefactoren af te leiden. De reconstrucie van excentrische K^0_S vertices wordt vergeleken tussen data en simulatie als een functie van de radiële excentriciteitsafstand. We bekomen kalibratiefactoren die typisch rond de 10–15% liggen voor grote excentriciteitsafstanden, maar erg gevoelig zijn voor dynamische detectorinefficiënties. De studie wordt uitgevoerd voor verschillende herwerkte versies van de CMS data en simulatie. In het bijzonder zien we een grote impact van een probleem met de uitleeselektronica van de tracker in de eerste helft van 2016 data, die wordt verkleind in een latere herwerking van de simulatie.

Automatische datakwaliteitsmonitoring Ten slotte bevat deze thesis een aantal gevallenstudies van het toepassen van machinaal leren op DQM en DC. Data quality monitoring (DQM - datakwaliteitsmonitoring) en data certification (DC - datacertificering) bestaan uit het garanderen van de goede kwaliteit van de data verzameld en gereconstrueerd door CMS en zijn cruciaal voor de correcte interpretatie ervan en voor de betrouwbaarheid van de fysica-analyses die ervan gebruik maken. We gebruiken machinaal leren, in het bijzonder (convolutionele, residuele) autoencoders om automatisch potentiële detectorproblemen te detecteren in de CMS pixel tracker, met een fijnere tijdsresolutie dan mogelijk zou zijn met de conventionele technieken. We tonen aan dat anomalieën kunnen worden gedetecteerd met een efficiëntie van praktisch 100%, met een voldoende lage vals-alarmfractie in de orde van 10^{-5}, hoewel de test set om deze waarden te bekomen niet volledig eenduidig gedefinieerd kan worden. Bovendien werd er grote technische vooruitgang geboekt in het beschikbaar maken van de data voor kwaliteitsmonitoring met deze fijnere tijdsresolutie als een standaardprocedure in CMS.

Reading guide

Comprising about 300 pages, this thesis is a rather big chunk of reading homework that I present to my busy jury members. That was totally unintentional and I'm sorry for it, but it just came about this way. To lighten the burden, I hereby provide a few words on how to most efficiently read this thesis, in case time runs short:

- The prologue and epilogue are completely optional and not directly relevant for the scientific content of this work. (On the other hand, they might be experienced as a slightly more exciting read than the strictly scientific content of this work.)
- Chapters 1 and 2 constitute a customary introduction to the standard model of particle physics and the CMS detector respectively, and contain only very little of my own work. They are not necessary for understanding the remainder of this thesis, provided one has at least a vague notion of both concepts.
- Chapters 3 and 4 cover almost exclusively my own work, and hence are important in this thesis.
- Chapter 5 contains a broader overview of the research field in which the following chapters take part, but contains little or none of my own work and is not necessary for understanding the remainder.
- Chapter 6 discusses some experimental techniques (but no major results yet) used in the two following chapters, containing partly but not exclusively work of my own.
- Chapters 7 and 8 contain the main topics and results, focus exclusively on work of my own, and hence are an important part of this thesis.

Acknowledgements

First of all, an immense gratitude goes towards professor Dobur, who, already many years ago when I was still an innocent engineering student, told me on our first coincidental meeting: "Sure, I can take you to CERN! Oh and I would like you to do a master thesis in our group as well!" This is where it all started for me.

Next, I want to thank my colleagues and office mates, who provided me with guidance, useful tips and tricks, much-needed distraction and relaxation, but above all the feeling of not having to struggle alone. I am very grateful for the exceptionally cordial and friendly atmosphere in our hallway; I always enjoyed and looked forward to being in the office even if it meant cycling through rain or morning frost.

This thesis is like a colouring picture for toddlers (in painfully many ways). I really just applied the final colours to lines and shapes that an uncountable number of engineers, physicists, data scientists and policy makers laid out before me. The CERN institute, the LHC accelerator and the CMS experiment are worldwide collaborative efforts spanning multiple decades, and I was only very fortunate to be able to use the carefully planned, painstakingly collected and conveniently pre-processed data on which this thesis is based.

A perhaps uncommon but very important acknowledgement goes out to all taxpayers in the Kingdom of Belgium. Your generous contributions have kept me alive during the past years and allowed me to bring this extremely important and terribly relevant work to a good end. Though it may bring only little consolation for the high tax burden in our country, I do believe that the degree of success of a society can be measured by the fraction of its national product spent on 'profitless' activities such as fundamental science, arts, and social security. I am happy that I was able, through my very unproductiveness, to highlight the (otherwise sometimes difficult to spot) difference between humanity and a rabble of mindless chimpanzees.

I would not have managed without the indirect support of friends and family. It may have looked like this whole PhD trajectory was just smooth sailing, but trust me it was not, and at the worst moments it was your care and love and laughter that helped me out. Incidentally, this also holds the other way around. The period of my PhD research was filled with tremendous waves and developments on a personal level. Some doors opened, others closed. I would like to reverse-acknowledge my research for providing a safe haven to retreat from the real world when it became burdensome.

Last but not least, I would like to sincerely thank anyone who reads (a part of) this thesis. The sense of usefulness, of fulfillment, of having changed other people's lives for the better, which is so vital for personal happiness, has been particularly hard to find these last years as a PhD student. Any person who reads this thesis and either mildly enjoys it or, who knows, learns something from it, greatly contributes to my mental health.

Luka
May 2024

P.S.: This work was supported by a grant from FWO. I would like to express my gratitude for positively evaluating my application, and for the relatively smooth administrative procedures.

Research Foundation Flanders
Opening new horizons

List of publications and presentations

Here I provide an overview of publications and related academic work that I directly contributed to during my PhD, either as main author or as contributing author.

Academic publications

- CMS Collaboration, "Inclusive and differential cross section measurements of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **02** (2022) 107, doi: https://doi.org/10.1007/JHEP02(2022)107 [1] (primary contribution)

- CMS Collaboration, "Measurements of the differential cross-section and charge asymmetry of a top quark pair in association with a W boson in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". `https://cms.cern.ch/iCMS/analysisadmin/cadilines?id=2772&ancode=TOP-24-003&tp=an&line=TOP-24-003`. (in preparation, currently in CMS internal review) [2] (primary contribution)

- CMS Collaboration, "Observation of four top quark production in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Physics Letters B* **847** (2023) 138290, doi: https://doi.org/10.1016/j.physletb.2023.138290 [4] (secondary contribution)

- CMS Collaboration, "Search for heavy neutral leptons in final states with electrons, muons, and hadronically decaying tau leptons in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Journal of High Energy Physics* **2024** (Jun, 2024) doi: http://doi.org/10.1007/JHEP06(2024)123 [5] (secondary contribution)

- CMS Collaboration, "Search for long-lived heavy neutral leptons with displaced vertices in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **07** (2022) 081, doi: https://doi.org/10.1007/JHEP07(2022)081 [7] (secondary contribution)

- CMS Collaboration, "Search for long-lived heavy neutral leptons in proton-proton collision events with a lepton-jet pair associated with a secondary vertex at $\sqrt{s} = 13\,\text{TeV}$", 2024. (submitted to JHEP). `https://arxiv.org/abs/2407.10717` [8] (secondary contribution)

Published CMS documents

- CMS Collaboration, "Displaced tracking and vertexing calibration using neutral K mesons". CERN-CMS-DP-2024-010 [6] (sole author)

- CMS Collaboration, "Prospects for computer-assisted data quality monitoring at the CMS pixel detector". CERN-CMS-DP-2022-013 [9] (sole author)

- CMS Collaboration, "Tracker DQM machine learning studies for data certification". CERN-CMS-DP-2021-034, 2021 [11] (secondary contribution)

- CMS Collaboration, "CMS tracker data quality certification with new machine learning tools". CERN-CMS-DP-2024-070, 2024 [12] (secondary contribution)

Conference proceedings

- L. Lambrecht on behalf of the CMS Collaboration, "Prospects for computer-assisted data quality monitoring at the CMS pixel detector", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated*

Equipment **1045** (2023) 167495, doi: https://doi.org/10.1016/j.nima.2022.167495 [10] (sole author, presenting own work)

- V. Wachirapusitanand on behalf of the CMS Collaboration, "Machine Learning applications for Data Quality Monitoring and Data Certification within CMS". CMS-CR-2022-014, 2022 [13] (secondary contribution)

Selected conference presentations and posters

- *Displaced vertex reconstruction with Ks and Lambda*, talk at CMS Long-lived Exotica Workshop, 30 January 2020, link (in digital version) (restricted access)
- *Improved inclusive cross-section measurement of a top quark in association with a Z boson at the CMS experiment*, poster at EPS-HEP 2021, link (in digital version)
- *CMS results of top quark electroweak production, including associated productions*, talk at EPS-HEP 2021, 26 July, 2021, link (in digital version)
- *New $t\bar{t}\gamma$ and tZq measurements in CMS*, talk at EOS be.h Equinox Meeting, 9 September 2021, link (in digital version)
- *Inclusive and differential cross-section measurements of tZq at CMS*, talk at LHC TOP WG Meeting, 1 December 2021, link (in digital version)
- *Prospects for automatic data quality monitoring at the CMS pixel detector*, poster at 15th Pisa Meeting on Advanced Detectors, 24 May 2022, link (in digital version)
- *Prospects for automatic data quality monitoring at the CMS pixel detector*, poster at EOS be.h Summer Solstice Meeting, 21 June 2022, link (in digital version)
- *Machine learning techniques for data quality monitoring at the CMS detector*, poster at ACAT 2022, 25 October 2022, link (in digital version)

Internal CMS notes

- CMS Collaboration, "Measurement of single top quark production in association with a Z boson". CMS-AN-20-170. (CMS internal) [14] (CMS internal note, restricted access)
- CMS Collaboration, "Differential cross section measurements of $\mathrm{t\bar{t}W}$ production at $\sqrt{s} = 13\,\mathrm{TeV}$". CMS-AN-23-077. (CMS internal) [15] (CMS internal note, restricted access)
- CMS Collaboration, "Study of displaced vertex reconstruction using light neutral hadrons in Drell-Yan events". CMS-AN-20-111. (CMS internal) [16] (CMS internal note, restricted access)

Prologue: the justification of high energy physics experiments

Ever since human consciousness came into being, mankind has been searching for meaning and purpose. For not yet fully clear reasons, we seem to be unable to take existence for granted, and instead ask for its reasons and objective, both on the macro-level, for example 'Why and wherefore does human life exist on Earth, and where will it lead to, or could it lead to if we try?', and on the micro-level, for example 'What does my personal life mean and how does it fit in the whole?'. These questions have been asked and inconclusively answered by many great philosophers, theologists, scientists and other thinkers; and though modern life, with its distracting luxuries and petty delusions is very good at (or perhaps even unconsciously designed for) ignoring these questions, they sometimes still haunt a humble PhD student like myself. They were the motivation for studying physics in the first place and for contributing to its endeavour in the form of this PhD. Fortunately, being a student of physics rather than philosophy, I can divert many of the *why*-questions to *how*-questions, which, however involved their solutions may be, are much more likely to be at least in principle answerable. Instead of asking *for what purpose* the Cosmos was created, one can investigate *how* it functions and *how* it came into being. In the last couple of centuries, this has proven to be the more fruitful approach for the progress of mankind.

And yet, one (or rather at least one) major *why*-question remains when studying elementary physics: 'Why am I doing this?'. I am quite sure this is a question many physics students will recognize as having been asked by themselves, with an amount of desperation that fluctuates with about the same time scale as recurring exam periods. But one could also more generally ask 'What's the use of all these *how*-questions, what will the answers to them gain us (if discovered at all) and should we not rather dedicate our life and efforts to more urgent and important matters?'. These questions, in various forms, variations and intensities, kept floating around in my head, in particular during my PhD research time at the CMS experiment at the Large Hadron Collider at CERN. Hence arose the need to justify the field of high energy physics, and its role in society. This is not in the last place motivated by frequent questions from laymen in the form of 'But what are you actually doing?' or 'But what could possibly the practical use of that?' or more generally "I'm just trying to figure out why you would do this; why anyone would do this."[2] It is the answer to that question that forms the topic of this essay.

The meta-question

The experiments at the Large Hadron Collider (among which the CMS detector where I did my research), or more generally those in the field of high energy physics, ultimately try to

[2]The answer in short: "The reason anyone would do this, if they could, which they can't, would be because they could, which they can't." Question and answer from *Rick and Morty*.

answer the most fundamental questions of 'Life, the Universe and Everything'[3]. What is all matter around us really made of? What are the fundamental laws governing the behavior of everything we see, hear, taste, smell and feel? Why, out of all possible hostile configurations, do we live in a universe that supports complex life forms? Is our world essentially random, determined, or controllable?

These questions, inspiring and intriguing as they may be, do not in any way seem to affect the practicalities of daily life, at least not that of most people on this planet (leaving out the unfortunate batch of PhD students professionally burdened with these questions on a daily basis). The knowledge that there are exactly three kinds of ordinary neutrinos, does not tell me whether or not I should take a raincoat to the office today. The fact that the observed probability of the simultaneous production of a top-quark and a Z-boson in proton-proton collisions does nicely agree with the standard model predictions[4], barely eases the pain if I hit my toe against the table leg (again). But perhaps more importantly, this knowledge does not even seem to help us any further in major contemporary world problems, such as the transition to renewable energy sources, the optimization of sustainable agriculture, ensuring world-wide access to clean water, or developing vaccines (to only name a few that come to mind first). The ugly truth is that these are essentially 'only' engineering problems, to which the forefront of elementary physics has long since lost all claims of relevancy.

It is in this respect that the meta-question, the question about the question, then arises: why do we keep asking ourselves these esoteric and abstract questions, that seem to be so far removed from the road towards a better society? Or from a more practical point of view: why do we keep funding expensive high energy physics experiments, devouring tax money and person power, while at the same time there are so many unresolved threats and problems in the 'real' world, that equally require attention and financial resources? Scientists usually question everything, except whether their questions are somehow relevant and whether spending whatever money they can get on trying to find the answers to them is somewhat justified. In this essay I will try to collect and formulate some arguments for and against the spending of public funds on high energy physics.

Of course I do not have the final answer to this question, so I invite all readers to formulate replies to the arguments given here, or construct (counter-) arguments of their own. Fortunately, my humble mind is not the first one to have pondered on these topics. Therefore I can rely on the reflections of physicists and philosophers who asked similar questions and wrote down their results. Here are the most important ones that have inspired this writing:

Richard Feynman, a Nobel prize winning theoretical physicist known for the development of quantum electrodynamics, gave a series of three lectures in 1963 entitled 'The meaning of It All' [17], in which he discusses the role of science in society, focusing on the importance of freedom of thought, open-mindedness and accepting uncertainty.

Herwig Schopper was Director-General of CERN in the days of the Large Electron-Positron Collider (LEP), the predecessor of the Large Hadron Collider. In his memoirs on that period [18], he shortly discusses its justification, and many more interesting related arguments are implicit in the rest of the text.

Hallam Stevens in his article 'Fundamental science and its justifications, 1945–1993' [19], in-

[3] From *The hitchhiker's guide to the galaxy* by D. Adams.

[4] See chapter 7.

vestigates the arguments brought up by high energy physicists to justify ever larger and more expensive accelerators, and how these were shaped by Cold War politics and other societal phenomena.

The all-round physicist and humanist Viktor Weisskopf wrote many essays on physics and its role in society [20]. Although many of his thoughts and arguments are dominated by the contribution of fundamental science to the atomic bomb, a theme which feels a little outdated by now, many of his ideas can be readily translated to contemporary issues.

Critique of practical reason - some arguments from critics

Let us start by summarizing and bundling some arguments typically advanced against spending financial or human capital on high energy physics research (such as the kind performed at CERN). The most pressing one is always a variant of the form 'it has no practical use'. That would not be an issue in a perfect world, but unfortunately this world is plagued by many problems, threats and challenges that cause acute human suffering. To put it with some pathos: all the accumulated knowledge about the nature of reality can never justify a single child dying from hunger. If only more people had investigated resilient farming methods in harsh climates instead of W bosons, this might have been avoided. More rationally: hoping that fundamental research will lead to useful results is always uncertain, a 'leap of faith' on behalf of the scientists. At the same time, a number of practical research fields can be easily identified that will almost certainly lead to improvement of the human condition. To again only name a few: cancer treatment, water purification, thorium reactors, hydrogen-based fuels, humanoid robotics and genetic modification. From a utilitarian perspective, the choice of for fundamental high energy physics rather than 'practical science' is irrational.

Closely related to the above is the exorbitant cost of modern high energy physics experiments, both because of the huge, exotic and high-tech infrastructure that is required, and because of the large number of people operating these experiments. The funds for these activities are often, potentially with some financial detours, essentially drawn from taxes paid by everyone. This implies that the money that goes to high energy physics is in competition with public health care, infrastructure, pensions, and all other benefits of the welfare state.

A more abstract argument disputes the privileged position of fundamental science with respect to, say, religion, or arts and literature. This argument leaves aside the question of practical applications, and instead focuses on fundamental science as one potential way towards the fulfillment of human existence. It denies however, that science has any particular claim to truth; it is essentially just a hypothesis like any other, a 'language game' of professionals inventing a more or less self-consistent jargon with no special relation to reality, or even a mass delusion of collectively self-brainwashed minds. And even if one admits that the results of science are probably 'true', a different kind of true than those of religion or creative culture, then still there is no value in this truth, as it does not tell us how to live well and, at least for most people, does not satisfy the need for sense and purposefulness.

These are, shortly summarized, the typical main arguments advanced against the studying of elementary particles or high energy physics, especially when experiments are getting expensive. Many variations and nuances of these arguments exist, at least some of which I hope will be clarified in the remainder of this text, when discussing the responses to them.

How much for one LHC, please? - the price tag

In its most practical formulation, the problem of the justification of high energy physics comes down to the question of how much money (if any) can be reasonably spent on it. I hinted before that high energy physics experiments are expensive. As a first step in the discussion, we should ask ourselves whether that is really true and make this claim a little more quantitative. In this paragraph I will keep the focus on CERN specifically (while it is certainly not the only player in the field of high energy physics), in order to have clear numbers to compare against each other.

Schopper discusses the finances of LEP, and of CERN in those days in general. According to him, the CERN yearly total expenditures (including operation, investment and personnel) are comparable to those of a single large European university. Dozens of those are funded by the CERN member states, so the marginal additional cost of CERN to academics at large is negligible. Let's consider my own Ghent University as an example (though it is rather medium-sized than large). As can be retrieved in the yearly budget reports for universities and colleges collected by the Flemish government[5], Ghent University has had a yearly budget of about 700–750 million euros in the last couple of years, the major part of which is coming, via several detours, from tax income. In comparison, the annual CERN budget[6] in recent years is about 1300–1400 million Swiss francs, most of which is coming directly from member state contributions. The factor of two difference and the Swiss francs to euros conversion rate do not significantly alter the conclusion that the CERN budget is indeed in the same order of magnitude as that of (only one) large university.

Belgium's contribution to the CERN yearly buget is approximately 2.7% [7] of the total, amounting to about 40 million euros. This number can be compared to the total yearly expenditures of tax income by the Belgian government(s), amounting to about 250 billion euros in the last few years, of which about 12% (29 billion euros) was spent on education (partly on universities) and about 2% (5 billion euros) directly on fundamental research[8]. To further put the number of 40 million euros into perspective, it can be compared with, for example, the direct yearly contribution of Belgium to NATO (about 66 million[9]), the total yearly salary of professional football players in the Belgian competition (about 140 million[10]), or the direct subsidies of the Belgian federal government for fossil fuels (10 billion[11]). Compared to these numbers, Belgium's contribution to the CERN budget seems to be only a marginal additional expenditure. This fact in itself is of course not a justification for it, but it gives more weight to other arguments that will follow. On the other hand, in absolute terms, 40 million euros per year is quite an impressive number (think of the possibilities if you could spend it on projects that relieve immediate human suffering), and moreover CERN is by far not the only institution dedicated to fundamental research that would qualify for those with which this essay essentially deals.

One important note: care should be taken when discussing CERN's personnel cost: many researchers at CERN who are not part of the core accelerator staff are not paid by CERN it-

[5] Flemish Government, *Repertorium* (in Dutch).

[6] CERN, *Final Budget of the Organization (2023)*, and *CERN budget - Taking a closer look at the LHC*.

[7] W. Van Doninck et al., *Flanders, Belgium and CERN*.

[8] Department of Finances of the Flemish Government, in an addendum to the tax letter every Flemish taxpayer got in 2021, data from Eurostat.

[9] NATO, *Funding NATO*, and Business AM, *NAVO-budget blijft stijgen, maar België staat voorlaatste in rangschikking van defensie-uitgaven* (in Dutch).

[10] *FootyStats* (in Dutch).

[11] FOD Financiën en FOD Volksgezondheid, Veiligheid van de Voedselketen en Leefmilieu, *Federale inventaris van subsidies voor fossiele brandstoffen* (in Dutch).

self, but by national funding agencies or universities. The ratio of permanent staff to external researchers using the CERN infrastructure (among which myself) is currently about 1/5[12]. Hence the actual operating costs for CERN are higher than the budget mentioned above, and moreover these additional funds are in direct competition with other academic research. The expenses for permanent staff constitute about 6% of CERN's budget[13]; taking into account the 1/5 ratio, a more realistic estimate of CERN's annual costs, including external researchers, increases the former 1400 to about 1800 million francs. This does however not significantly alter the conclusions so far.

CERN is not a black hole simply absorbing any amount of money that is thrown into it; it also generates economic added value and return on investments, mainly through the development of new technologies, as will be discussed in more detail in the following paragraph. Furthermore, about 40% of the CERN budget is spent on materials and infrastructure, implying this part of the budget is returned directly to the member states, whose companies provide these[14].

In conclusion, I think it is fair to say that, though it is hard to quantify exactly the total (tax) expenditures on CERN and especially the economic return generated from it, the total price tag of CERN is relatively modest, especially given the fact that it is a world-leading laboratory in one of the most technologically complex fields of present day science. It is unlikely that the hypothetical ending of CERN funding and redirecting of the money to more concrete solutions of world problems would make any significant impact in that direction. This is only a consequence *in practice*, and does not yet justify *in principle* the absolute expenses; for that we need additional arguments.

Unexpected fruit - technology spin-offs

'Science and technology' is a common phrase in education and advertisements. The two subjects are indeed closely connected. Hence, perhaps the most obvious argument in favour of high energy physics experiments that we encounter is the following: fundamental science research leads almost inevitably to the development of new technologies, that overall improve the quality of life for humanity.

This relation is a bidirectional one. On the one hand, results obtained in fundamental research open the gates towards new technologies. Think for example of the electronics in computers, smartphones, cars, and almost every household appliance imaginable, whose existence would not have been possible without studying electrons as elementary matter particles, and quantum mechanics as a fundamental theory of nature. On the other hand, the ever increasing complexity of high energy experiments pushes technology to and often beyond its current limits. CERN is a forerunner in this respect, often engaging with private high-tech companies in joint research efforts that benefit both experimental particle physics and other fields. The most famous and relevant example in this context is probably the invention of the world wide web, that originally arose from the very practical need to share LEP data and software with researchers around the world.

A first point of critique that often arises with respect to these technologies is that technology does not always benefit society. In fact there are technologies that surely degraded or at least threaten our quality of living. The prototypical example here is the atomic bomb, particularly

[12] W. Van Doninck et al., *Flanders, Belgium and CERN*.
[13] CERN, *CERN budget - Taking a closer look at the LHC*.
[14] W. Van Doninck et al., *Flanders, Belgium and CERN*.

relevant here since fundamental physics played a very prominent and explicit role in its development. But even in an age where the threat of a nuclear war seems to have lessened (with respect to the cold war era), there are plenty of examples to be found of technologies whose total contribution to our quality of life seems at least questionable. Think for example of cloning and genetic modification of humans, social media, artificial intelligence, fully automatic drone bombers, all kinds of chemical and atmospheric pollution by the industry, implantable body augmentations for the rich, the medical stretching of the length of life beyond the point of comfort, etc.

One way to reply to this accusation is to point to the fact that science only provides means, not the ethics on how to use them. The decision to use the atomic bomb is a political one, not a scientific one. Many of the other more modern examples given above are also just side effects of innovations that could be used for the better. It could be argued that having a particular possibility is always intrinsically worth more than not having it; whether to actually make use of it is a different question that lies outside the realm of science. This does not fully redeem science however, for one can argue that scientists should have realized by now, after centuries of man-made atrocities, that any technology that can be used for worse, will probably be used so at some point by someone. Should any potential new technology that is perceived within reach at some moment in time be actually developed? No technology has ever been one-sided and isolated from any other. Science does not progress in leaps but in small steps, and it is practically very hard, if not impossible, to draw a line between intrinsically *good* and *bad* technologies. To stick to the example of the atomic bomb, this technology cannot be regarded decoupled from nuclear energy, which, despite some doubts and critique, clearly has beneficial effects as well. The same holds for the more recent example of artificial intelligence, where the huge potential benefits (e.g. safer and more efficient traffic with self-driving cars, better medial diagnosis and treatment, automation of difficult, tedious or dangerous work, etc.) cannot be separated from potential abuse.

Considering the balance over the last couple of centuries, I think it is hard to deny that technology has overall improved our quality of life rather than degraded it. Progress in agriculture has reduced famine and hunger and the basic human concern for food. Medicinal science has dramatically increased the physical quality of our life. Some more recent examples: telecommunication and aviation have brought the whole wide world within reach. The internet has radically lowered the threshold to knowledge, information and self-improvement. The fact that many of these examples have not actually been realized in large parts of the world is, again, mainly a political problem rather than a technological one. Of course it is not a priori guaranteed that the balance will remain positive in the next centuries. The question whether we are at some kind of turning point in the history of technology, where the negative aspects will start to outweigh the positive ones, is in my opinion not a ridiculous one, and should be evaluated carefully in each epoch.

Apart from focusing on the negative aspects of technology, one can also consider it not having any aspects at all: one can rightly ask if each and every endeavour in high energy physics is bound to lead to new technologies that directly affect daily life. Especially in a time where high energy experiments investigate matter in ever more exotic states as would never occur naturally on present-day Earth, doubt can arise as to how this could ever lead to something directly useful. Irrelevance is a more painful accusation to science nowadays than causing the destruction of the world, as it was in the early days of the atomic bomb. I think this point of view neglects however the second direction of the twofold relation between science and technology as mentioned above. It may be true that the discovery of the Higgs boson will not lead to a cure for cancer (although, looking at the history of science, one learns to be careful with

those claims), but the technologies developed to make the discovery possible surely can have an impact on daily life.

Moreover, the list of applications that are direct consequences of research conducted without a particular aim except a deeper understanding of the subject matter is impressive. Just to give few examples: semiconductors and transistors (basis components of virtually every piece of electronic equipment), chips and integrated circuits (without which computers would still be the size of a large room), nuclear power, induction coils in motors, telecommunication via electromagnetic waves, medical diagnostics with radioisotopes, LED lights, solar power, etc. With in the future perhaps quantum computing, thorium reactors, and nuclear fusion. Many of these applications originated from fields of research that seemed utterly without practical use a few years or decades before. For example, the established physicists Rutherford said in 1933: 'Anyone who expects a source of power from transformation of these atoms is talking moonshine.' [20]. Little more than a decade later there was the atomic bomb.

Aesthetics, enlightenment and purpose

Another category of justifications for fundamental physics is more abstract than technology. It does not focus on tangible results or measurable effects on human day-to-day life. It comes in different shades and variations, but usually comes down to something like this: 'The search for the underlying structure of our universe, for a full understanding of our physical reality at the most fundamental level is an inalienable part of our human nature and mission. Moreover, the results we have so far, though not yet complete, are particularly beautiful'. High energy physics is regarded as a path to enlightenment of the human species, a striving point where, either by reaching it or by continuously getting closer to it, humanity can reach the pinnacle of its evolution and realize its potential to the fullest. Furthermore, it is not a business but an art, the only difference being that we do not create the artwork but rather discover it. The symmetries and harmonies of nature, its (alleged) simplicity and elegance, are reasons enough for studying it *in se*. *La science pour la science*.

The obvious weak point of this argument is that it is a very subjective one. Many people seem not to be bothered at all by questions of what everything is made of, why we are here, and what is out there in the remotest regions of the cosmos. And who are physicists to judge? They tend to be quite biased on this topic anyway. Many people also seem not to be sensitive to the beauty of fundamental physics, most likely because of the ominous mathematical barrier, the opaqueness of which I personally still perceive every day. Yet I would boldly argue that the disinterestedness of most people in society for fundamental physics is rather a problem of most people in society than of physicists. As mentioned before, the peculiarities of modern life seem to be particularly suited for wooing 'annoying' questions of sense and structure to sleep, for dismissing them as obsolete and unimportant with respect to the order of the day, where this or that soccer team has beaten some other one, or where a new television series features this or that famous actor. The remarkable efficiency with which modern society fills our lives with flashing lights and turbulent noises, in order to mask the black void behind it all, gives me the impression that it was specifically designed for facilitating the runaway from it and the unanswerable questions waiting there. This seems to me one of the reasons why most people are not really fond of physics, apart from the fact that they just don't like maths.

And yet the big questions are important, though their answers may seem unsettling at first. Are we just a random collection of molecules assembled in such a way as to reproduce, on a small rock in an uncaring universe? Are concepts such as choice, free will and responsibility an

illusion? To what degree is our perception of the world a mental creation? These are daunting philosophical questions, that could bring about a total re-evaluation of the *conditio humana*, and should not be run away from[15]. No contemporary philosophical discussion about these topics is possible without a major influence of fundamental physics and its claims about the nature of reality. And for this reason, though not everyone should be a physicist, everyone should have a basic understanding of its principles, to the degree that everyone should think about the world they live in.

One could argue that a final goal of physics, something like a full understanding of the universe and its history and future at the most basic level, is not achievable. Yet I believe it cannot reasonably be denied that science progresses. Every day, the accumulated knowledge of mankind is larger than the day before. The fact that, especially in the field of high energy physics, everything we pretend to know now might be proven essentially false at some point in the future is irrelevant in this respect. That is just the normal way of scientific progression. Also the question of whether the final goal is anything close to conceivable, is only marginally relevant here, since there is only one way to find out.

The next line of defense, though admittedly a desperate one, is a *reductio ad absurdum*. Many people are not interested in literature, art, philosophy, theology, psychology, even history. Should we abolish them all? I do not think that is a good idea, as I believe that fundamental science, together with art and philosophy in the broad sense of the word, are essentially what makes us human. If we only create and use technology in our daily lives, without deeper questioning, we're just another (though slightly more fancy) kind of chimpanzees.

Another point of critique on any aesthetics-based justification of physics, is about the timing. Yes, it my be true that understanding the big bang would be a nice accomplishment for humanity, but should this really have priority, now that we are faced by so many imminent threats and problems on a global scale? Apart from wars, famine, diseases, poor housing and related problems that have been around for ages, I think climate change is currently the most pressing example here. How can one justify quietly studying abstract elementary particles while there is the urgent need for more research on sustainable energy, climate change impact mitigation, carbon capture, etc.? Without denying the severity and acuteness of climate related (and other) problems of the present day world, I think it is fair to say that each and every epoch in history has considered itself as being in a state of imminent crisis different from any other in centuries before. It is also fairly certain that postponing fundamental research to 'a more suitable time' would be equivalent to abolishing it altogether, as the hypothetical moment where all global threats and crises have been averted is probably even more unachievable than the point of answering all questions in physics.

Anti-scientific tendencies in the broad sense have always been present in society, as a counter-reaction, since the start of the historic Enlightenment. A common criticism on science from their side (despite their huge differences) is that it has a dehumanizing effect. Science reduces the immense and unfathomable human soul to a collection of numbers forming a genetic sequence, the breathtaking beauty and balance of our planet to a collection of arrows moving carbon in and out our 'ecosystem', and the unspeakably improbable and unintelligible cosmos to a computer model. I think this criticism is superficial and unjust: rather than reducing reality, the study of fundamental physics augments it. The beauty of a rainbow is not in any way diminished by the knowledge of the electromagnetic spectrum and the laws of refraction and reflection, nor that of the solar radiance by insight in the fundamental interactions causing

[15] See also the epilogue to this dissertation for an elaboration on these questions.

the fusion process between hydrogen nuclei. The real target of the dehumanizing criticism, if any, should rather be the political or economic system exploiting the results and means of fundamental science for their own agenda, while often this distinction is not made. The link between fundamental science and its applications is rather complex and has already partly been discussed in the previous paragraph. The 'dehumanizing' criticism is, in my opinion, especially cynic when it comes from the side of theology or religious people; just count and compare the war crimes, genocides and reification of the enemy committed in the name of some pope or deity to those for the sake of science.

Building a better world bottom-up - the scientific mindset

Another argument to justify the study of fundamental physics, without the direct or guaranteed objective of new technologies and without resorting to something as insubstantial as aesthetics, is similar to the one used to justify for example the study of ancient 'dead' languages, such as Greek and Latin, in secondary education. It has a positive impact on the minds of the students in the broad sense, which in turn benefits society in the long term.

The study of the underlying mechanisms of the universe or ancient languages and culture have in common that they open up one's mind, sharpen one's recognition for similarities and structure, even provoke a sense of modesty in relation to something that transcends one's own limited space, time and perspective. On a conceptual level, the benefits for society are tolerance and respect for other cultures and viewpoints, respect towards nature, having a sense of the genuine long-term (on a cosmological scale), having a notion of balance, harmony and unity. On a more practical level, one can think of enhanced problem-solving skills, technical expertise, critical thinking, solid, honest and uncompromising reasoning, a sense for measurability and accuracy, the ability and willingness to learn new things quickly, and many more things.

There are several counter-arguments to this view. One category does not deny these benefits, but rather asks whether studying such remote topics such as merging black holes or some extremely rare decay of a Higgs boson are really the only, let alone most efficient or cost-effective way to achieve these. A second category denies the positive impact of studying physics on the human mind altogether, arguing that fundamental science is just another business sector like all other, equally poisoned by politics, competition, self-centeredness and tunnel visions, and with no stronger claim to transcendence or truth than any other.

As to this latter criticism: it is certainly true that physicists remain human, and therefore, imperfect. And it is equally true that fundamental science is a business with many competitors, where money is involved, and so inevitably arguments have to be made and strife may arise. However, I do think it is possible to decouple these rather practical and technical circumstantial effects from the core mindset of physics. Some of the positive influences of this state of mind on human beings, apart from the ones enumerated above, I will discuss in some more detail below. Before that, let's consider more closely the former criticism, that fundamental science may be beneficial for the mind, but not efficient in this respect. It is hard to compare those alleged 'efficiencies'; high energy physics is quite unique it its setting and scope, and its effects quite different, when looking in more detail, from that of studying ancient languages or history for example (which is undoubtedly cheaper). Both ways are very different, and both should be cherished, upheld and not directly compared to each other in terms of efficiency. Moreover, especially because of these differences, there is an aspect of personal affinity as well, where the way of science and the way of humanities, even if leading to similar ends, may be

open or shut to different individual minds.

The comparison with some more applied sciences, such as biotechnology, that entail both the scientific mindset and the promise of a direct impact on society, is more problematic. The practical side of the argument (problem-solving skills, critical thinking and creativity) can not really distinguish between fundamental and applied sciences. And the more conceptual side (tolerance and harmony) starts to overlap with the ethereal and elusive concepts of aesthetics discussed in the previous paragraph. But in my opinion there is still a category of intermediate concepts, such as abstraction, fundamental doubt and the (de-)correlation between intuition and reason that are quite unique for fundamental science and distinguish it from both applied sciences and cultural studies.

There are a few specific aspects of the scientific method and mindset that I think have an undeniably positive impact on society. The first is the hypothetico-deductive method, which is the prototype for how scientific theories are (or should be) constructed. Very schematically, the method goes as follows. Suppose one has a given number of observations (for example the position of stars in the sky over the course of time as observed through a telescope). One tries to come up with a theory (or model, or hypothesis) that explains these observations (for example, the law of gravitation). Then, given this theory, one tries to devise an experiment or observation that is different from all the ones on the basis of which the theory was conceived, and one predicts the outcome according to the current hypothesis (for example: the acceleration with which an apple falls to the ground). Finally, one carries out the experiment. Any hypothesis that did not predict the outcome correctly is rejected (for the time being), and among all the ones that did, one chooses the simplest one. Then the process starts over. Further details of the method do not matter for this discussion, neither even does the question whether this is really how modern scientific theories have been constructed. The point is that this idealized method implements at least in principle the concepts of testability, falsificationism and critical reasoning. It is an antidote against gut feelings and selective blindness.

Another focal point of the scientific mindset is accepting ignorance and uncertainty, which is important since it is a vital ingredient for progression, freedom and tolerance. True science implies accepting the fact that every particular piece of knowledge we have today, every theory, every mental image of the universe, might someday be proven false. (As a sidenote: though many people have pointed out the similarities, or at least compatibility between science and religion, this is one of the most fundamental differences between them.) But it also provides the means with which one can keep believing in progressive research, and appreciate the value and status of truth under these at first sight pessimistic conditions. This is especially true for high energy physics, which operates at the frontier of the most existential questions. What is everything made of? How did it all come into being? What is the underlying structure, what are the hidden principles that govern the Universe? The changing answers to these questions do not only concern a handful of experts in the field, they touch every one of us (well, at least they should, in my opinion). Accepting that nothing is absolutely certain directly leads to freedom of thought and tolerance towards other ideas and worldviews. This of course does not imply that any idea should be approved of; but objective observation must be the judge, not a priori prejudice. Furthermore, fixed certainties are incompatible with the imaginative and creative powers needed for both scientific and societal progression. Only the possibility to doubt in a responsible, scientific way the current beliefs and theories about how things are or should be opens the door towards real change.

Between dream and act

The paragraphs above mostly discussed the field of high energy physics from an external and conceptual (one could ask: idealized?) point of view. The question now arises to what extent all these arguments relate to the very personal experience of doing a PhD, with all accompanying joys and frustrations, small successes and moments of desperation, momentary glimpses of a higher truth and (many) down-to-earth computer problems.

It is certainly true that, if one does not pay attention to take a step back once in a while, one tends to get lost in the details of daily work, one 'cannot see the forest for the trees', as the saying goes. In a pessimistic mood, it is not difficult to reject all the arguments given above to justify high energy physics at large, when looking at a more personal level: CERN may give rise to useful technologies in the long term, but my contribution to those is hard to discover, physics may be beautiful but my daily work rather consists of writing very dirty computer code, and it is not very obvious in what way my 'notion of balance, harmony and unity' (if at all present) will benefit society in any way. Even worse, the initially felt passion for the research topics risks to be smothered by the very ugly and practical everyday issues, missed deadlines, long working days, detachment from 'real' society and the people around you, the feeling of being at a standstill in the learning process, and many more effects that are absent from the intuitive representation of fundamental scientific research.

Just as loneliness can be paradoxically more heartfelt in cities, surrounded by crowds of people, big scientific collaborations such as CMS can give the particular feeling of not being really part of it, while at the same time making it very hard to distinguish a personal contribution. Everything is constantly moving, ever new results, recommendations and suggestions for analysis techniques are presented at high rate by bright people in many more meetings than the hypothetical upper limit that should be imposed rather sooner than later by the World Health Organization. At the same time, you are alone at your little desk (also covid is to blame here), staring at your laptop screen, for the fourth day in a row trying to solve a small technical issue that should have taken a few minutes. Under these conditions, it is difficult to view yourself as being a useful part of the scientific community. It is this sense of uselessness and obsolescence, both of my personal life and that of physics in general, that has been the most tormenting aspect of my PhD time, and one of the biggest inspirations for this text.

There are surely some factual points of improvement to be implemented in the way PhD work is organized at CMS. For example, the amount of time and effort I had to spend on purely conventional style issues and related nonsense, raised in innumerable rounds of feedback and comments on a paper draft, is really beyond all justification and can drive saner minds than my own to desperation. Next, the people who wrote the 'documentation' on the parts of the software that can in principle be shared between all CMS collaborators should be put in jail for flagrant lies and deception, or at least brought to a mental hospital for an urgent examination of their empathetic and linguistic capabilities. The obsession with organizing meetings without any real content or contact has reached unprecedented heights. Just naming a few examples here.

However, the biggest issue lies of course within one's own mind. It is well known that 'the road to success is paved with many failures', and that finding out that something doesn't work is equally important as finding something that does, and that being stuck or having to start over are natural parts of the research process. I know all this, but still. Even these de-idealizing statements have something idealistic or even heroic in them, which is very hard to relate to the confused and inconclusive struggles of everyday PhD work. The psychology of PhD students,

notably the so-called 'impostor syndrome' has been well documented [16]. Reading about and having knowledge of the impostor syndrome does however not really help, as it is often introduced as occurring to people who 'despite reaching significant intellectual milestones ranging from advanced degrees to professional awards, cannot internalize their success or convince themselves they deserve it.' But it is a rather central feature of the syndrome that those who suffer from it do not identify themselves as having reached significant intellectual milestones. Speaking with fellow PhD students, it seems that they all share very similar concerns and feelings. But even this brings little and only momentary relief, since the mind is essentially solipsistic and always finds new ways of convincing itself that its situation is special and incomparable to that of those suffering from exactly the same feelings.

A vital question in this paragraph is the following: supposing I could justify high energy physics at large, does that also justify my particular involvement in it as a PhD student? This question might seem strange, since high energy physics (or any other research field for that matter) could not exist without researchers; but it arises from a seeming cognitive dissonance between a mental image of what this kind of research should be like, and the day-to-day experiences during my PhD time. This discrepancy I have no doubt will be confirmed by many PhD students from any research field, and this is exactly the main argument to state that it is most likely a mental discrepancy rather than a factual one. So let's conclude this personal intermezzo and turn to two interesting case studies of high energy physics experiments, to see if the arguments above hold in real-life situations.

The SSC catastrophe and what to learn from it

The Superconducting Super Collider (SSC) was a planned American (who could have guessed by the name?) particle collider, the construction of which was first approved, but then, after billions of dollars had already been invested on the infrastructure, finally stopped and canceled in 1993. The reasons for this catastrophic waste and disgrace for the high energy physics community are diverse. The hubris in the original project proposal, the technical challenges, and the exploding budget estimates are surely part of it, but another one is constituted by the fact that the traditional justifications used by physicists to convince governments to provide funding were at a breaking point. The case study of SSC is particularly interesting since it shows how strongly the convincing power of the same arguments depends on the political stage of the time frame.

The main concern in this case study that is relevant to the justification of high energy physics, is the seeming relativism of all justifications. The arguments pointing to (mostly military) technology spin-offs that were successfully used during the first phase of the Cold War were replaced by those emphasizing symmetry and beauty when humanist, anti-war and anti-industry tendencies became stronger in the later 1960's and 1970's. Those arguments were in turn replaced by a focus on economic advantage and national pride in the neo-liberal and revived Cold War atmosphere of the 1980's, ultimately failing after the Cold War ended. So is there any truth in any of the listed arguments that will stand the test of time, or are they all just clever tricks making use of the particular political situation to draw money from governments?

A distinction should be made here between truth and political relevance. Or put more sharply: between what happened and what should have happened. What the SSC catastrophe shows is that the same arguments do not always suffice to convince governments, not that the arguments in themselves are invalid. I truly believe that there is at least a little universal truth

[16] See e.g. Ref. [21] and K. Weir, *Feel like a fraud?* as a starting point.

in all the arguments given in this essay, although this statement is of course open for critical evaluation. Nevertheless, an important message here is that the justifications given by scientists advocating the SSC, however true they may have been, were not as clear and self-evident as to convince politicians deciding on budgets.

So what can we learn from this story? First of all, let's keep budgets for high energy physics within limits. It is better to secure a medium-sized but long-term secured baseline financial support than embarking on an ambitious but risky multi-billion project. Experimental particle physics is simply not at that point anymore where breakthroughs and discoveries come frequently and spectacularly enough to keep the general audience engaged and motivated. Secondly, care should be taken with arguments that blend physics with politics; they may be particularly powerful today, but are just as fleeting and transient as political commitment and might turn out completely unconvincing tomorrow. Any science should be justifiable from itself, or from whatever directly follows from it, such as technology. If a particle physics experiment is only attractive because of a particular political configuration (though this may be hard to recognize in practice), one should probably consider other options. And finally, it's usually a good idea to scale down whatever Americans propose by a factor of three to end up with something realistic.

A final remark: was it in the end a *good* or a *bad* decision to cancel SSC at the time? There is no clear-cut answer to this, since there is no way to tell what would have happened if it had been finished and operated. Probably the Higgs boson would have been discovered a few years earlier than 2012, but whether that difference matters in the light of the centuries is another question. We can only hope that the budgets that became available after the canceling of SSC did not go to funding high-tech military research or highly addictive painkiller drugs. Instead of discussing the past, let's talk about a very similar decision for a future collider in the next paragraph.

The elephant in the room - FCC

To conclude, let's talk about one very particular but important example of justifying experimental high energy physics: to build or not to build the Future Circular Collider (FCC). This 100-km circumference particle collider, the natural extension of the Large Electron Positron collider (LEP) and the Large Hadron Collider (LHC), will dominate the global budget and person power in the field of experimental high energy physics for the decades to come, potentially at the cost of canceling many alternative smaller scale experiments. Moreover, with all due respect for the particular technical difficulties, it is conceptually not very different from the LEP/LHC, only larger (for example, the proton-proton collision energy will be close to an order of magnitude larger than that of the current LHC). A clear objective, such as the discovery of the electroweak gauge bosons in the case of LEP, or the discovery of supersymmetry and/or the Higgs boson in the case of LHC, seems not clearly defined. This is not to deny that there's a whole spectrum of interesting physics to be investigated at the FCC [22, 23] (ranging from precision tests of the standard model, over the exclusion/discovery of large classes of dark matter candidates, to an order of magnitude increase in the discovery reach for heavy particles), but the relevancy of the outcomes is more uncertain than for LEP or the LHC. This matters a lot, considering that FCC will not be just a sidetrack, but the dominating entity in experimental particle physics in the lives of people from the next generation.

So what about all the arguments given in the previous paragraphs? Do they apply here? When looking more closely, we see that most of the justifications for high energy physics experiments

are not absolute; they usually deal with balance and compromise. We should ask what new technologies are expected to be developed, both in the context of building FCC, and based on its potential results, how they would benefit society, and compare that to FCC's price tag. It is in this respect that my remark above matters, that FCC is conceptually not very different from the LHC but bigger. This reduces somewhat the probability that revolutionary new technologies will be developed for its construction and operation. In fact, one of the preconditions for a project of this vast scale, is that the essential technologies either exist or are in a reasonably advanced stage of development. CERN has always promoted and actively taken part in synergies between experimental particle physics and industry for developing technologies, and will continue to do so in the context of FCC. There are many technical challenges that I am gravely insulting by saying that FCC is just a bigger LHC, the solution of which will lead to new techniques in civil engineering, superconducting magnet technology, distributed computing grids, and many other fields. And yet it is difficult to point out how these are not just very particular and technical extensions of what was already developed for the LHC or its predecessors, and how they will decisively and positively impact society.

Next, we can consider the ongoing search for fundamental knowledge as part of our humanity, but not at all costs; to care for others is also a part of our humanity and we should continuously evaluate if the expenses on high energy physics (in competition with more direct societal investments) are not surpassing reasonable limits. It is in this view that we should consider the not very clearly defined goals of FCC. There is, at this moment and to my limited knowledge, no single mature theory favored by many physicists, that could be decisively experimentally tested at the FCC, such as (minimal) supersymmetry was for the case of LHC. The scientific community is somewhat confused, ideas are dispersed, proposed models are exotic and far-fetched. Of course, one of these ideas might still turn out to be true, and FCC could be the breakthrough experiment that leads to a significantly improved understanding of our universe. But strong evidence for this is lacking. On the other hand, staying within the safe confines of the standard model, it is guaranteed that its fundamental parameters and rich phenomenology can be tested with unprecedented precision at the FCC, sometimes multiple orders of magnitudes better than what could be achieved at the LHC. It is likely (though difficult to quantify), that this experimental program will at least provide some hints on the direction in which groundbreaking new theories could be sought for.

Thirdly, there might be other ways than fundamental physics to shape the people who would need to tackle societal problems tomorrow (or rather today). But this argument does not really distinguish between FCC or any other direction the high energy physics community might go. Whether physicists are trained and shaped at the FCC or at any alternative experiment does probably not matter much for the potential advantages for society; the basic ingredients of experimental high energy physics research, and the impact on the mind of those involved, would remain the same. The general case was already discussed before, and for the specific case of FCC or no FCC, this argument is inconclusive.

My personal opinion is the following. I do not think we should ask to increase the overall budget for high energy physics for the sake of building FCC, if that ever was an option. There is simply not enough conclusive evidence that FCC will be worth it, for society and humankind. The question is then reduced to one which merely deals with physics, instead of physics within society: given the limited and fixed budget available, is FCC, among all candidates, really the most promising experiment to lead to new clues in the search for a deeper understanding of the universe? At this point I am sorry to disappoint those expecting a clear yes or no, but I would leave the discussion to more senior physicists than myself.

Conclusion

The question of *why experimental high energy physics matters* is relevant in a world with limited resources and where many more concrete problems equally require funds and attention. In this essay I tried to cover some of the arguments and counter-arguments in the justification of high energy physics. As this is not a scientific article, I will not use the space of this section for neatly summarizing everything I said before, but rather for encouraging anyone reading this, whether or not involved in the field, to critically reflect on the given arguments and form an own reasonable view on the presented facts and opinions, and to come up with other ones. I'll be happy to hear them and expand my scope.

Is the spending of public tax money on high energy physics experiments justified? Despite its form, this is hardly a *yes* or *no* question. It will always be a matter of interpretation, perspective and different value-scales. What is important is not that we find a definitive answer to the question, but rather that we keep asking it. Perhaps this insight is the most profound value of high energy physics.

Chapter 1

The standard model of particle physics

This introductory chapter provides a very short overview of the so-called standard model of particle physics. This model constitutes the theoretical framework in the context of which all work in this thesis was carried out. A more detailed contextualization of this thesis within the larger theory is given at the end of this chapter, after a discussion of the shortcomings of the standard model.

1.1 Summary of everything

The standard model of particle physics (SM) is a model that bundles all of our knowledge of elementary particles and their interactions in a consistent theory. Its development can be acknowledged among the greatest achievements of physics in the 20th century, next to the theories of relativity and quantum mechanics (upon both of which the SM is built), and forms an important step towards a 'theory of everything', though many problems and open questions remain. Despite its shortcomings, to be discussed later in this chapter, the SM is mindbogglingly successful on the experimental side. For multiple decades now, the main goal of experimental high energy physicists has been to find discrepancies between the predictions of the standard model and the corresponding measurements, but no such discrepancies have yet been found conclusively. The work in this thesis fits in the endeavour of pushing the SM to its limits in the hope of finding hints for yet unknown physical effects beyond it.

For a more complete introduction to the standard model with all its mathematical intricacies, the reader is referred to textbooks such as [24–26]. Only the most important concepts and phenomena will be summarized in the paragraphs below.

1.1.1 Particles and fields

Fermions The standard model is most conveniently introduced by enumerating the elementary particles described by it. Ordinary matter is essentially made of 'down' quarks (d), 'up' quarks (u), and electrons. The d and u quarks form a pair, that is supplemented by two similar but heavier pairs, introducing the strange (s), charm (c), bottom (b) and top (t) quarks. These heavier quarks are unstable and quickly decay to their lighter counterparts after they have been produced in energetic reactions in nature or in particle accelerators. They also occur as virtual intermediate contributions to interactions, even if they are not in the initial or final state. Similarly, the electron (e) forms a pair with a neutral electron-neutrino (ν_e), with two supplementary pairs consisting of the muon (μ) and the tau (τ) leptons (both with corresponding neutrinos). Muon and tau leptons are unstable and quickly decay to lighter particles[1]. All of the above particles are categorized as fermions (with half-integer spin). For each fermion, the SM also contains an anti-fermion, a particle with the same mass but opposite charges.

Bosons The second category consists of bosons (with integer spin), particles that mediate the interactions between (anti-)fermions. This category comprises the gluon (for the strong interaction, also called quantum chromodynamics (QCD)), the W^+, W^- and Z bosons (for the weak interaction) and the photon (for the electromagnetic interaction). Finally, the H boson is the quantum of the H field, generating mass for some of the bosons and fermions. This brings the total number of elementary particles to about 30 (with some ambiguity on whether or not antineutrinos are actually distinct from neutrinos - a so far unresolved question).

The strong interaction (mediated by gluons) acts only on fermions with a property called 'colour charge'[2], i.e. only on quarks. Gluons carry colour charge as well, and hence undergo self-interaction. The electromagnetic interaction (mediated by photons) acts on all fermions with electric charge, i.e. quarks and charged leptons. The W boson couples to fermions with a property called 'weak isospin', i.e. all left-handed fermions, and the Z boson couples to all fermions except right-handed neutrinos (which are not predicted to exist in the standard model). Furthermore, the W and Z bosons and the photon can interact with each other in several combinations. Finally, the H boson interacts with all massive particles.

From this relatively small amount of particles and their mutual interactions, the visible world around us is constructed. Typically, the fermions are thought of as 'matter particles' and bosons as 'force carriers', but the division is not very clear-cut. For example, the majority of the mass of protons and neutrons is generated by gluons, not quarks. On the other hand, fermions enter as mediators in all interactions just as bosons do.

Fields The standard model is a particular implementation of a quantum field theory, in which the particles are regarded as quanta (think of elementary quantities, or more correctly elementary excitations) of a field, a continuous entity spread out in space-time. Field theory circumvents the problem of 'action at a distance', while the quantum nature preserves the discrete (rather than continuous) behaviour observed at very small scales. The standard model expressed in this way is compatible with the fundamental principles from classical physics

[1] Neutrinos, on the other hand, do not really decay, but rather display a particular property called oscillation, where the different flavour states undergo oscillations into each other in both directions. This phenomenon, experimentally observed, is however not explicitly part of the standard model; see section 1.2.1.

[2] Colour charge is a property analogous to electric charge, but with a ternary instead of binary structure, conventionally called red, green and blue. Colour charges are unrelated to actual colours, which are not even defined at these small distance scales.

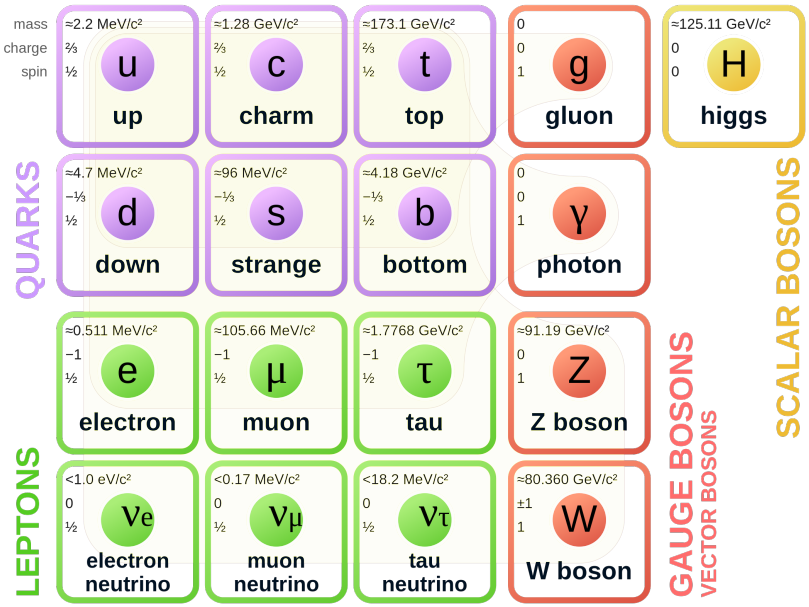


Figure 1.1: Summary of the particle content of the standard model. Both quarks (in purple) and leptons (in green) are organized in three generations (from left to right) of similar pairs with increasing mass. Within each quark pair, there is a quark with electric charge $+2/3e$ and one with $-1/3e$; within each lepton pair, we find a charged lepton ($-1e$) and a neutral lepton (called neutrino). All quarks and leptons are fermions, with spin 1/2. Ordinary matter consists mostly of u quarks, d quarks and electrons, with the other fermions being produced in high-energy reactions. Vector bosons (in red) have spin 1 and mediate the strong (g), electromagnetic (γ) and weak (W, Z) interactions, and a mass mechanism is provided by the field associated with the scalar H boson (in yellow). Figure from Ref. [27].

(notably conservation of energy and momentum), with quantum mechanics (fields being in some sense the extension of wavefunctions), and with special relativity.

1.1.2 Symmetries and the Lagrangian

Symmetries are at the core of the standard model. In the context of particle physics, the term 'symmetry' refers to an invariance of all observables with respect to a degree of freedom in the description of a system. The SM is generated from a composition of several fundamental local gauge symmetries, shortly summarized as $SU(3) \times SU(2) \times U(1)$. The $SU(3)$ symmetry refers to color charge and encodes the fact that the effects of the strong interaction are unchanged if the color charges of all quarks and gluons in a system are consistently replaced. The $SU(2) \times U(1)$ symmetry can be associated to the unified electroweak interaction but is more difficult to interpret since it is spontaneously broken. A modified $U(1)_{EM}$ symmetry remains however, implying that the electromagnetic interaction does not change if all electric charges are consistently interchanged. Other symmetries of the standard model that are not specific to it include invariance with respect to the Poincare-group (rotations, translations and boosts), implying that the physical laws are the same on this side and on the other side of the galaxy, as well as invariance with respect to CPT-reversal, i.e. inverting time, space and charges, implying that the physical laws remain the same if the entire Universe is mirror-imaged around any arbitrary point in space and time and if each particle is replaced by its antiparticle.

One of the reasons why symmetries are quintessential to the standard model is their relation to conserved quantities via Noether's theorem. Invariance with respect to space-time coordinates (expressed by the group of Poincare transformations) implies conservation of linear momentum, angular momentum and energy. Similarly, the $SU(3)$ symmetry implies the existence of conserved color charge, while $U(1)_{EM}$ provides the familiar conserved electric charge.

In classical field theory, the central entity of any physical system is its Lagrangian (or more correctly Lagrangian density), an object from which the laws governing the behaviour of that system can be derived in a formalized way by the principle of least action. Symmetries of the Lagrangian are transformations of the fields that leave the total Lagrangian, and hence the physical laws derived from it, invariant. The Lagrangian of the standard model is the most general renormalizable expression containing all known fields respecting the local $SU(3) \times SU(2) \times U(1)$ gauge symmetries and global Poincare symmetries[3]. In one of its most concise forms, it can be written down as [24][4]:

$$
\begin{aligned}
\mathcal{L} = & -\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} - \frac{1}{4} W^a_{\mu\nu} W^{a\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \\
& + i \sum_j \left(\overline{\Psi}_{(q_j,L)} \not{D} \Psi_{(q_j,L)} + \overline{u}_{j,R} \not{D} u_{j,R} + \overline{d}_{j,R} \not{D} d_{j,R} + \overline{\Psi}_{(l_j,L)} \not{D} \Psi_{(l_j,L)} + \overline{e}_{j,R} \not{D} e_{j,R} \right) \\
& + (D_\mu \phi)^\dagger D^\mu \phi - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \\
& + \sum_j \left(y_{u_j} \overline{\Psi}_{(q_j,L)} u_{j,R} \tilde{\phi} + y_{d_j} \overline{\Psi}_{(q_j,L)} d_{j,R} \phi + y_{e_j} \overline{\Psi}_{(l_j,L)} e_{j,R} \phi \right) + h.c.
\end{aligned}
$$

The first line involves the dynamics of the gauge bosons, with G being the field associated to the strong interaction (i.e. gluons) and W and B representing the fields of the electroweak interactions (which transform into the W^+, W^-, Z and γ fields after spontaneous symmetry

[3] As well as an observed, seemingly accidental CP symmetry for the strong interaction, mentioned here for completeness.

[4] Also thanking Thomas Mertens' course material for the Quantum Field Theory course at Ghent University, edition 2023

breaking). The second line describes the kinematics of the fermions, left-handed (L) SU(2) doublets and right-handed (R) SU(2) singlets of leptons and quarks, and their interactions with the gauge bosons; the third line introduces the Higgs (H) field ϕ and gives mass to some of the vector bosons and the fourth line to some of the fermions.

The further intricacies are not important for the remainder of this thesis. The main point of interest here is that the Lagrangian of the standard model fully characterizes the laws of nature at the scales of elementary particles interacting in the LHC[5].

1.1.3 Some phenomenological aspects of the standard model

Hierarchy of the interaction strengths The interaction strength of a fundamental interaction can be summarized in a single dimensionless number called the coupling constant. It differs strongly in magnitude between the strong ($\sim$1), electromagnetic ($\sim$1/137) and weak ($\sim$1/30) interactions [28]. This has implications for practical calculations in the perturbation framework (see below), where higher-order QCD corrections are often needed with respect to generally smaller lower-order electroweak corrections. Notice that this hierarchy is not fixed but depends on the energy scale. For example, the strong coupling constant is not actually constant but becomes larger at lower energies and larger distances, leading to colour charge confinement, and smaller at higher energies and smaller distances, leading to asymptotic freedom. As another example, the weak interaction is usually much weaker than the electromagnetic one despite its larger intrinsic coupling strength, due to the mass of the W and Z bosons (compared to the massless photon); they only become comparable in strength at high energies where the masses of the W and Z bosons become negligible.

Hierarchy of particle masses The lightest elementary particle in the standard model, except for neutrinos which are known to have a very small mass despite their being predicted massless, is the electron with a mass of 0.511 MeV/c^2; the heaviest is the top quark with a mass of about 172.5 GeV/c^2. Hence the particle masses span more than 5 orders of magnitude. This behaviour is a consequence of the different Yukawa coupling strengths of the H field to the different fermion fields. The top quark, being the heaviest elementary particle, has a very strong coupling to the H boson and is of special importance in this thesis. It is discussed in more detail in chapter 5

1.1.4 From the Lagrangian to measurements

The Lagrangian of the standard model determines how particles interact, but it is not a directly measurable quantity. Instead, most experiments collide particles and only measure, essentially, a number of collision events with a specified signature (e.g. where a specific particle of interest was produced). Given the total number of initiated collision events, this corresponds to measuring the probability for a specific process to happen. This probability is technically most conveniently expressed as a cross-section σ (see section 2.1 for more details from the experimental side). The road from the standard model Lagrangian to predicted cross-sections that can be directly experimentally verified is rather long and technical [26], hence we limit ourselves again to the highlighting the main points.

Perturbation The laws of physics derived from the SM Lagrangian cannot be solved exactly except in a very limited number of examples, such as free fields (no interactions, just isolated

[5]That is, the Lagrangian together with some external experimental input. For example, the values of the Yukawa couplings for quarks (y_{u_j} and y_{d_j}) and for leptons (y_{e_j}), that determine their mass, are free parameters not predicted by the theory.

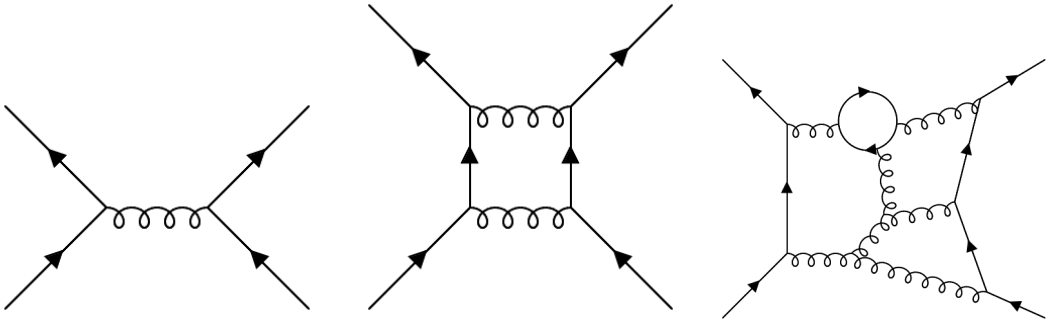


Figure 1.2: Example of Feynman diagrams and the perturbation expansion. Left: the lowest order (LO) diagram, representing the simplest possible interaction with given initial and final state. Middle: a representative next-to-leading order (NLO) diagram, representing the next-to-simplest possible interaction between the same initial and final state. Right: a higher order diagram. Every additional order in the calculation (represented by diagrams with additional vertices) has a smaller contribution to the total probability than the previous one (if the coupling strength is smaller than 1). Hence, practical calculations can often be truncated at LO or NLO, ignoring higher order terms.

particles floating through space). For all interesting cases, an approximation approach called perturbation theory has to be applied, where the Lagrangian is split in a free part and an interacting part, and the interacting part is written as a series expansion and calculated up to an arbitrary order of precision (typically truncated at the lowest order, LO, or next-to-leading order, NLO).

Feynman diagrams This truncation/approximation procedure naturally leads to Feynman diagrams, that are pictorial representations of the terms in the series expansion of the interaction Lagrangian. Each of these terms consists of a particular set of interactions leading from a given initial state to a final state, with LO terms containing the most simple interactions with the fewest interaction vertices, and higher orders containing paths between initial and final state with additional vertices. Typically, every additional order in the calculation (represented by diagrams with additional vertices) has a relatively smaller contribution than the previous one, which is the motivation behind perturbation theory and LO Feynman diagrams. In some cases, such as low-energy QCD interactions, this does not hold anymore and other numerical methods have to be applied. The perturbation expansion and Feynman diagrams are illustrated in Fig. 1.2. Apart from being schematic representations of calculations, LO Feynman diagrams are also often used (e.g. in chapters 7 and 8) to gain a qualitative understanding of the probability, experimental signature and potential peculiarities of a given process.

Matrix element and cross-section Using a fixed set of rules [24], the terms in the standard model Lagrangian can be translated into multiplicative expressions associated to the vertices and lines in a Feynman diagram. The result obtained by summing these expressions over all possible Feynman diagrams (in practice truncated, as explained above) with a given initial and final state, is called the matrix element. And finally, the infinitesimal cross-section $d\sigma$ is proportional to the square of the modulus of this matrix element: $d\sigma \propto |\mathcal{M}|^2$. To obtain a single number that can be directly verified in experiments, one needs to integrate $d\sigma$ over the phase space of interest.

Non-perturbative effects The calculation of the matrix element with perturbation theory is only applicable for a high-energy interaction between elementary particles. However, in practice, the initial state particles are not necessarily elementary (e.g. protons in the LHC), nor are the observable final state particles (e.g. hadronization of quarks). Hence, the calculation of the matrix element as coarsely outlined above is only a part of the calculation, to be complemented with non-perturbative heuristic techniques. The discussion of these effects is deferred to section 2.13, where event simulation is detailed.

1.2 Shortcomings of the standard model

The standard model has proven to provide a very accurate description of interactions between elementary particles in many tests and experiments over the last few decades, ranging over large energy and probability ranges. A small anthology of those test results, from the CMS experiment (see chapter 2), are summarized in Fig. 1.3. Despite its theoretical elegance and experimental success, the standard model is known not to be a valid theory of everything. It describes only a part of reality, and some phenomena that are experimentally observed are not predicted by it.

1.2.1 Indirect experimental problems

Gravity The most striking example of such a phenomenon is gravity. While the SM is compatible with special relativity, it does not include general relativity, nor gravity altogether. This is of no direct concern for the experiments at the LHC as gravitation has no measurable impact on the measurements, but it clearly makes the SM not suitable to describe e.g. planetary motion or cosmology. Adding gravity to the standard model unfortunately proves to be more difficult than simply postulating a 'graviton' and adding a few terms to the Lagrangian (e.g. because these terms would be non-renormalizable, meaning that practically every calculation yields a divergent result and the usual mathematical tricks to circumvent these divergences do not work) [29]. Despite extensive efforts from theoretical physicists, this is still very much an open problem.

Dark matter Dark matter poses another puzzle. The hypothesis of dark matter originates from some cosmological observations that seem to point to a source of gravitation (i.e. mass) that cannot be detected with other (typically electromagnetic) means. Explanations for these observations in the form of non-luminous stellar-like objects made from ordinary baryonic matter have been largely ruled out [30, 31]. A favourable alternative explanation for this phenomenon is the existence of WIMPs, weakly interacting massive particles, a new type of particle (e.g. supersymmetric particles or heavy neutral leptons) that interacts only very feebly with ordinary matter. This type of particle is missing from the SM however. Though many searches for dark matter candidates of this form have not found anything [32], the WIMP hypothesis remains popular and therefore a challenge for the SM.

Neutrino mass The third observation at odds with the SM is neutrino oscillation [33], i.e. the observation that the weak eigenstates of neutrinos do not correspond to their propagating eigenstates. For this to hold, neutrinos must have a nonzero mass, while the SM predicts them to be massless[6]. Moreover, from oscillation experiments, the mass of neutrinos is inferred to be much smaller than the typical mass scales of other elementary particles in the SM, hinting

[6]The most minimalistic way to account for neutrino masses in the SM is by including right-handed sterile neutrinos and a corresponding Yukawa interaction. While this would require only a minimal extension to the SM, the existence of these particles is an open question, and they are not generally considered to be part of it.

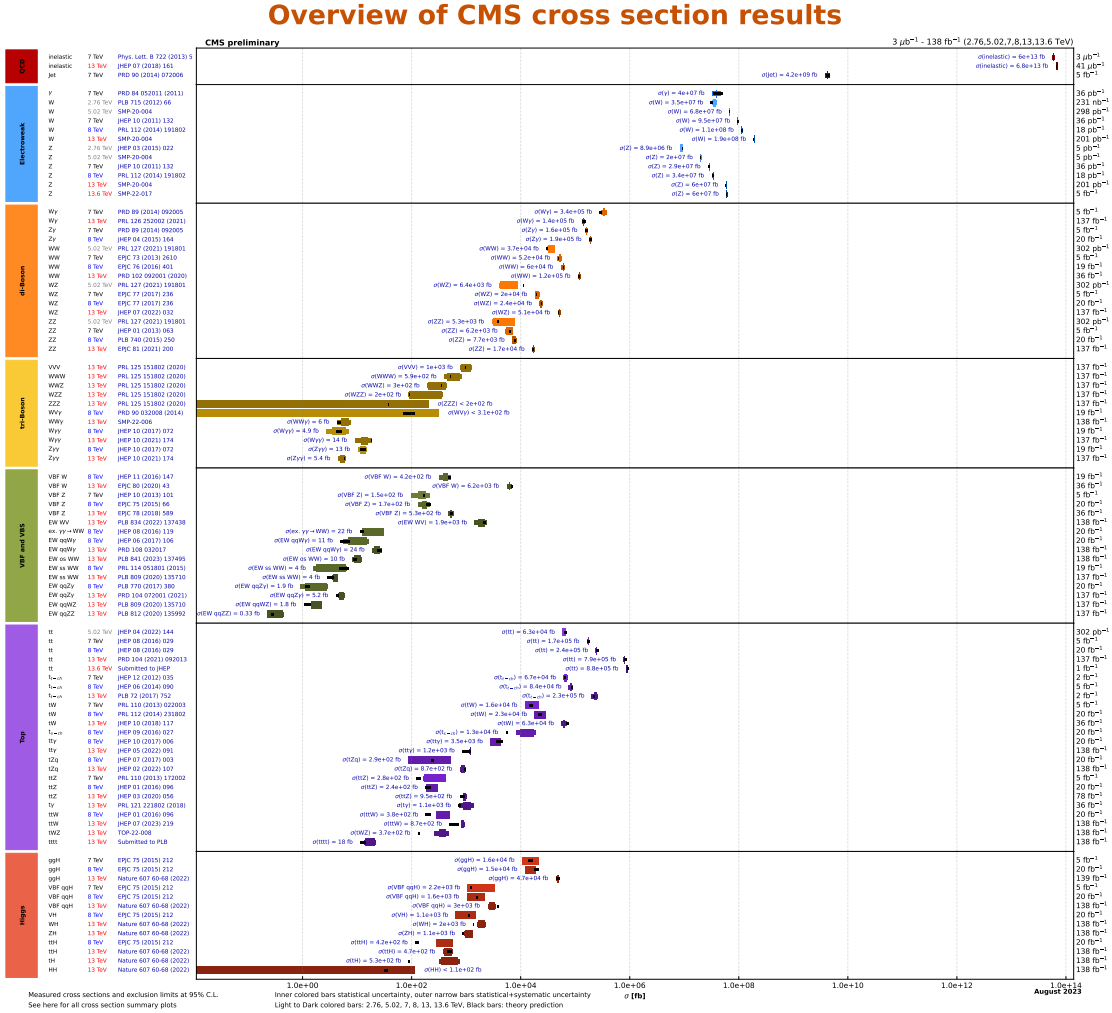


Figure 1.3: Overview of predictions and corresponding measurements of the cross-section of a collection of elementary particle processes, by the CMS experiment. Each line in the figure represents a separate prediction (black bar) and corresponding measurement (coloured bar). The predictions for these processes can differ strongly in technical implementation but are all ultimately based on the standard model. This overview highlights the accurate predictions over a rich phenomenology and many orders of magnitude in probability, which shows the strength of the standard model. Figure from the *CMS standard model summary plots*.(*)

(*) `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined`

that the mass of neutrinos is not generated by the same mechanism as for the other fermions.

Vacuum energy Next, the accelerating expansion of the Universe despite gravitational contraction tendencies poses a problem. The energy density of the vacuum is the most likely candidate for 'dark energy' or 'the cosmological constant', enhancing the expansion of space. But when its value is calculated within the SM, the obtained value is *by far* too large (between 50 and 120 orders of magnitude) to explain the observed expansion rate [34].

Matter and antimatter The approximate symmetry of the SM with respect to matter and antimatter is another point of concern, since the visible Universe seems to be devoid of antimatter and consist of regular matter only. While the SM does show small asymmetries between matter and antimatter in the form of CP symmetry violation, these are too small to explain the huge imbalance observed in our Universe [35, 36].

1.2.2 Theoretical problems

Hierarchy problem Of a more theoretical nature is the so-called hierarchy problem, concerning the mass of the H boson. In the SM, the physical mass of the H boson is composed of the difference between a bare mass term as a free parameter and a set of loop corrections. The loop correction terms can be calculated in the context of the SM, and turn out to be far larger than the observed H boson mass. This implies that the free mass term must be tuned very accurately such that the difference with the almost equally large loop corrections yields the much smaller observed mass. Though not forbidden, this fine-tuning seems unnatural and hints at effects from new physics that cancel the large corrections in a natural way. Of a similar nature, but less severe, is the Yukawa coupling hierarchy problem, i.e. the unnatural smallness of the interaction strengths between lighter fermions and the H boson.

Renormalization and regularization Most of the calculations in the SM contain divergences if carried out intuitively ab initio. In order for the SM to be informative and make finite predictions, several techniques have to be applied to remove the divergences, including renormalization (replacing bare masses with self-interactions by the experimentally observed dressed masses without self-interactions) and regularization (introducing cut-off scales in the calculations). The need for these techniques can be regarded as a hint that the current standard model is an effective approximation of an underlying theory. Also the running of the coupling constants, i.e. the varying of elementary interaction strengths with the energy scale, fits within this viewpoint.

1.2.3 Direct experimental tensions

W boson mass And finally, a number of direct measurements at collider experiments or similar setups seem to be in tension with the standard model predictions, though not yet conclusively. Most notably, the very precise measurement of the W boson mass by the CDF Collaboration at the Tevatron accelerator yields a value that is larger than the SM prediction by 7σ [37]. The mass of W bosons is a central parameter in the SM and can be computed to a high precision from other parameters such as the elementary electric charge and the masses of other particles. A higher observed value could point to additional corrections to the W boson mass from yet unknown heavy particles. Care must however be taken, as this measurement is very complex both from a theoretical and from and experimental side, and moreover the CDF measurement seems to be in tension not only with the SM, but also with earlier (less precise) measurements from the LEP experiments and D0, as well as with recent measurements by

LHCb [38], ATLAS [39], and CMS [40], all of which are consistent with the standard model prediction.

Muon magnetic dipole moment Next, the magnetic dipole moment of the muon has been measured to a value that deviates 4.2σ from the SM prediction [41]. Similar to the W boson mass, the muon magnetic dipole moment receives corrections from loop diagrams involving other particles, so extensions to the standard model including additional particles could explain this discrepancy. On the other hand, the calculation of the magnetic dipole moment is very complicated, and it is not fully clear yet whether systematic errors in the computation could generate this difference. Recently, an updated experimental result using a larger data set [42] yielded 5.0σ deviation from the SM prediction, yet this value is to be cited with caution as several predictions yield strongly differing results, so it is not very clear what 'the' SM prediction is exactly. In particular, recent approaches with ab-initio lattice QCD calculations (as opposed to phenomenological calculations with several experimental inputs) make the tension essentially disappear [43]. And even for the data-driven calculations, new measurements of properties used as an input in the magnetic moment calculation are in disagreement with earlier results and significantly reduce the tension as well [43, 44]. A better theoretical understanding of the computation of this quantity is the most likely road to resolving this tension.

Lepton flavour universality Another direct discrepancy between observation and SM predictions that until recently caused quite some excitement among physicists concerns the violation of lepton flavour universality in b-meson decays. While the SM predicts electrons, muons and tau leptons to behave similar except for effects arising from their different mass, decay rates of b-mesons were observed to violate this rule [45]. However, updated versions of these measurements revealed that the observed discrepancy was likely the result of errors in the measurement - new results are compatible with the SM predictions [46, 47].

1.3 Beyond the standard model and the context of this thesis

Due to its remaining mysteries and plain defects, a multitude of possible extensions to the SM has been developed by theorists and phenomenologists, and sought for by experimentalists, collectively called beyond-standard-model (BSM) theories. The general idea is that the currently known standard model is an approximation valid at (relatively) low energies, of a more fundamental yet-to-be-discovered underlying theory with less free parameters. On the experimental side, two distinct strategies can be discerned in this respect. Firstly, one can look for the production (or other observable effects) of particles beyond the standard model directly. This strategy is typically employed when looking for signatures of specific BSM hypotheses, such as a given configuration of supersymmetry or heavy neutral leptons, both in general purpose detectors such as CMS and ATLAS, and in many dedicated, smaller scale experiments.

However, as direct searches have remained without (clear) positive results, indirect methods become important. In this strategy, the self-consistency of the SM is pushed to its limits by measuring its known predictions as accurately as possible and looking for small discrepancies. The measurement of the W boson mass mentioned above fits in this strategy, as do the main results presented in this thesis. The tZq process had already been discovered and found to roughly agree with SM predictions, but this work improves the precision of the measurement. The same is true for the $t\bar{t}W$ process, where a mild discrepancy with respect to the SM prediction was found prior to this work. However, $t\bar{t}W$ is notably difficult to model correctly,

even assuming the correctness of the SM. This thesis reports the first differential measurements of $t\bar{t}W$ in CMS, providing useful insights in the nature of this modeling issue.

Chapter 2

The CMS detector

The experimental work reported in this thesis was carried out with the Compact Muon Solenoid (CMS) detector. For the proper understanding of the subsequent work, a short introduction to the basic layout and functioning of the CMS detector, as well as to the reconstruction and identification algorithms for detected particles, is of vital importance. This chapter will focus mainly on the points of interest for understanding the data analysis strategies described in later chapters; a more unbiased, complete and thorough description can be found in [48], with important updates in [49].

2.1 The Large Hadron Collider

The need for particle accelerators The energy scale of the physical processes studied in this thesis is not easily experimentally accessible. In the natural world around us, the typical energies of particles and particle reactions (such as the decay of radioactive isotopes or

photo-ionization) are *much* lower than those we are interested in. In fact, the vast majority of what we experience around us consists of u-quarks, d-quarks and electrons[1], making up only the lowest energy/mass part of the 'existing'[2] particles in the standard model. One way to study very high energy particle interactions that does not involve 'artificially' accelerating them is with cosmic rays, i.e. protons and other particles hitting the Earth with energies many times higher than what can be achieved in man-made accelerators; however, the flux of these particles is far too low for the precision physics envisioned by the CERN program. Particle accelerators offer a controlled experimental environment with access to relatively high energy scales and sufficiently high collision rates for statistically significant precise measurements.

LHC The Large Hadron Collider (LHC) is currently the world's largest particle accelerator, located at the CERN research facility near Geneva in Switzerland. It is ring-shaped, with a circumference of about 27 km. It collides two beams of protons at a record-breaking center-of-mass energy of 13.6 TeV[3]. The LHC is not the only accelerator at CERN; the site hosts a network of linear and circular accelerators (shown in Fig. 2.1), many of which were state-of-the-art machines at their time of construction, that now serve as pre-accelerators for the LHC. In particular, the protons for the LHC are obtained from hydrogen atoms (abundant in nature) by stripping their electrons; they are then accelerated in the Linear accelerator 4 (Linac4), enter into the circular Proton Synchrotron Booster (PSB), or shortly called Booster, next into the Proton Synchrotron (PS) and next into the Super Proton Synchrotron (SPS) before finally being injected in the LHC[4].

Circular particle accelerators employ electric fields to accelerate the particles, and magnetic fields to bend their trajectories into a circular shape. In the LHC, the electric fields for acceleration are provided by radiofrequency cavities [50]. The oscillating behaviour of the electric field inside these resonant cavities ensures that the protons are kept at the target energy and are spaced in discrete packages (called bunches) along the ring. The magnetic fields for bending are provided by a set of superconducting Nb-Ti dipole magnets, while quadrupole magnets are used for focusing the beam [51].

Proton accelerators The LHC collides protons, which are not elementary particles. This implies that, apart from fully elastic proton-proton collisions, the actual collisions occur between the constituents (called partons) of the protons, i.e. gluons, valence quarks or sea quarks. Since these constituents do not carry a fixed momentum fraction of the proton, but rather follow a probability distribution (called parton distribution function, or PDF), the actual collision energy can vary. Hence, the LHC inherently scans a range of energies, which is often useful in searches for new phenomena of which the energy or mass is not exactly known. Furthermore, since protons are relatively heavy (compared to e.g. electrons, which are sometimes also used in accelerators), they emit less synchrotron radiation and can therefore be accelerated to higher energies. On the other hand, only a part of that energy is available for the actual collision between partons, following the PDF. Furthermore, the parts of the shattered proton that do not take part in the collision of interest, as well as additional QCD interactions, pollute the

[1] This is making abstraction of gluons, which in fact are responsible for the major part of the proton and neutron mass, but which can be absorbed in the concept of *valence* quarks with their *constituent* mass.

[2] There are several ways in which the higher energy/mass particles of the standard model 'exist' even if they do not seem to occur in our everyday environment. They contribute in virtual interactions to the mass of other particles, as well as to the vacuum energy. They can also occur as real particles in high energy reactions such as the ones studied in this thesis, which may not be so abundant in our everyday environment, but are important to study the earlier stages of the Universe.

[3] This is the current value, valid since the 2022 data taking. The work in this thesis was carried out with data collected at a slightly lower 13 TeV.

[4] An animation of this process is shown *here* (digital link).

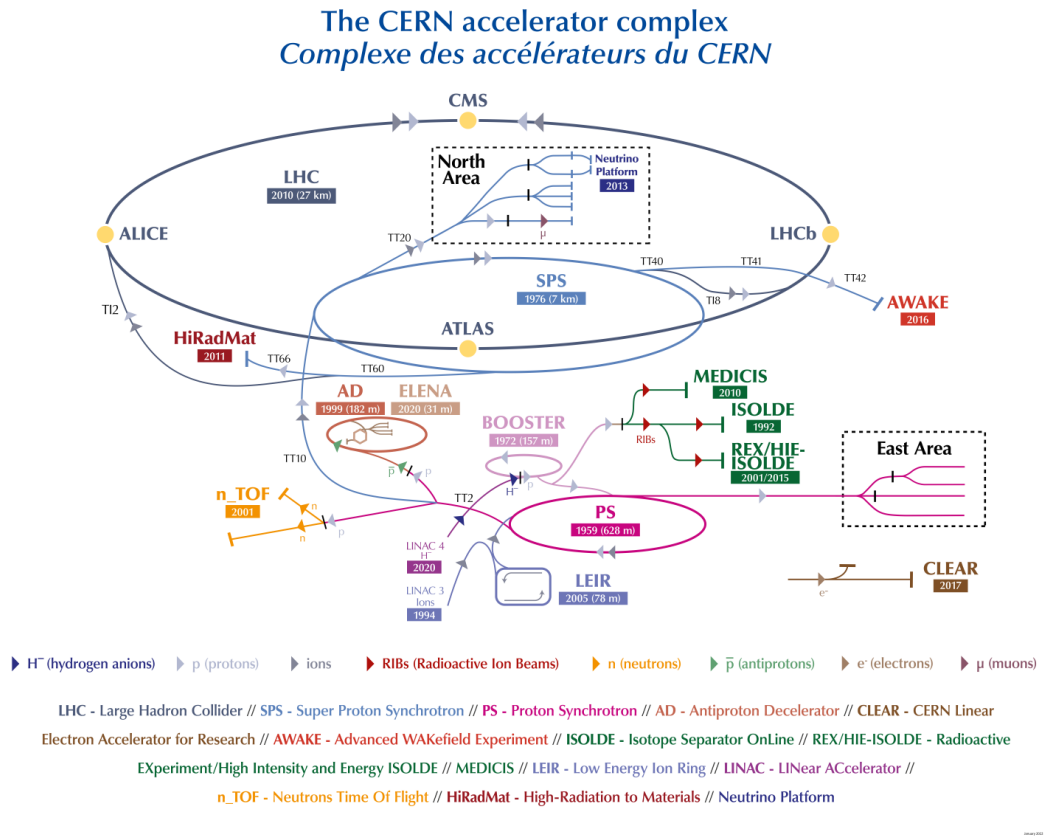


Figure 2.1: Overview of the particle accelerator complex at CERN. Figure from Ref. [52]

experimental signature with respect to 'cleaner' electron-positron collisions.

Luminosity The unique selling points of the LHC are not limited to its center-of-mass energy, but also include its very high rate of collisions, expressed in a quantity called luminosity. For colliding particle bunches with a gaussian density profile (equal for both), the luminosity can be expressed as [53]:

$$\mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y} \tag{2.1}$$

with N_1 and N_2 the number of particles in each bunch, f the revolution frequency, N_b the number of bunches and σ_x and σ_y the width of the gaussian density profile in the two transverse directions to the beam direction. The units of (instantaneous) luminosity are $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Integrating this quantity over the time of data taking, one obtains the integrated luminosity, which is a measure of the size of the collected dataset. Integrated luminosity is often expressed in inverse femtobarns (fb^{-1}) where a barn is 10^{-28} m^2. The LHC was designed to deliver an instantaneous luminosity of 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, but regularly ran at twice that value during the 2016–2018 period of data taking [54]. The delivered integrated luminosity for that period amounts to[5] 190 fb^{-1}, almost 20 times the size of all hadron colliders prior to the LHC.

Cross-section The luminosity of a particle collider is directly proportional to the rate of expected events of a given type ($\dot{N}$). The proportionality factor is the cross-section σ which expresses the probability for an event of that type to happen, given a collision:

$$\dot{N} = \mathcal{L} \cdot \sigma \tag{2.2}$$

Pileup To achieve the highest possible luminosity, the proton bunches are focused strongly such that the probability for a collision increases. In fact, the average number of proton-proton collisions per bunch crossing (called pileup) is usually larger than 1, about 30 on average during the 2016–2018 data taking, peaking up to 60 [55]. This means that, synchronously with the particles of a potentially interesting events, the detectors are traversed by many other particles from pileup (or strictly speaking in-time pileup) interactions. Although required for high luminosity, high pileup has a detrimental effect on the performance of particle identification and accuracy of their measurement, and special techniques are employed to mitigate the effect of pileup on measured quantities. Additionally, out-of-time pileup consists of detector signals being polluted by previous bunch crossings, typically because the reset time after each bunch crossing is non-negligible with respect to the time between bunch crossings (about 25 ns).

2.2 The Compact Muon Solenoid: introduction

CMS The Compact Muon Solenoid (CMS) detector is one of the general-purpose experiments installed on the collision points on the LHC. It consists of several specialized subdetectors to measure the various properties of the particles emerging from the collision point at its center. The various subdetectors are indicated in Fig. 2.2. The CMS detector has an overall approximate cylindrical symmetry and consists of so-called 'barrel' layers in the middle, supplemented with 'endcap' disks at both ends. It has a modular structure, with different kinds of subdetectors arranged together in the following order (from the center outwards): tracking detectors for reconstructing the flight trajectory of charged particles, electromagnetic and hadron calorimeters for energy determination of electrons/photons and hadrons respectively,

[5] This number includes several special runs not used for regular proton-proton physics analysis and does not take into account inefficiencies in data taking by the experiments. Hence the integrated luminosity of the dataset used in this work is somewhat lower.

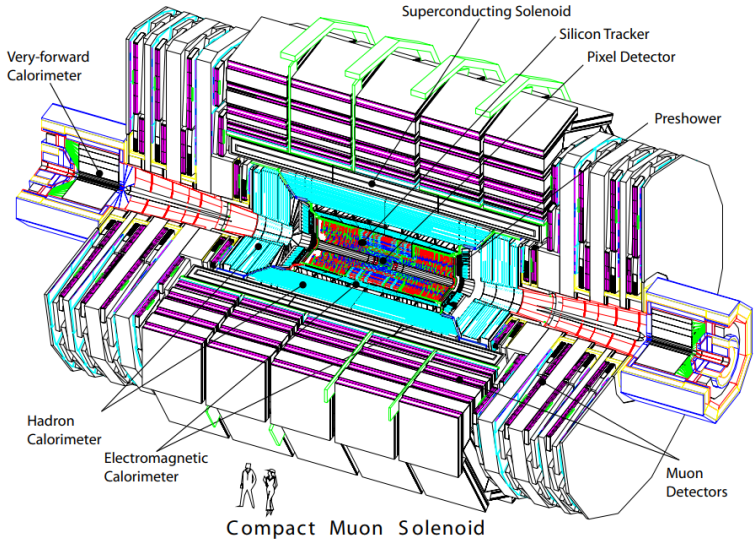


Figure 2.2: A cutaway view of the CMS detector, with an indication of the different types of subdetectors. Figure from Ref. [48].

and muon systems for identifying and characterizing muons passing all previous detection layers.

Solenoid magnet Apart from particle detection systems, CMS consists of a superconducting solenoid magnet providing a strong magnetic field of 3.8 T in its interior along the beamline. This feature is vital for the measurement of the momentum of charged particles by the degree of curvature of their trajectory inside this field, as well as the sign of their electric charge by the direction of the bend.

Some facts and figures The CMS detector is a miracle of scale. Despite being called compact, it is about 20 m long, 15 m in diameter and weighs more than 10,000 tonnes. Its interior is cooled to -25°C[6] and its magnet even to -269°C[7]. The beam pipe is operated in a vacuum of below 10 mbar[8]. Proton bunches are made to collide every 25 ns, i.e. 40 million times per second. The innermost detector (the pixel tracker) starts at a radius of only 2.9 cm away from the beamline and consists of 124 million readout channels and pixels of size 100×150 μm^2 [49].

2.3 Principles of particle detection

Particle detection Every detection is based on interaction of the to-be-detected object with a piece of sensitive detector material [56, 57]. This does not only apply to technologically advanced particle detectors installed on the LHC, but even to the most familiar kind of detection, human vision, i.e. photon detection by the rods and cones in the eye. The basic operational principle of the eye is the interaction of impending photons with detector material in the form of photosensitive pigments in the cells in the retina. While the eye, for obvious reasons, has been optimized by evolution to detect photons in the wavelength range emitted by the Sun, it is not suited to detect other particles at other energy scales directly. Therefore, other types

[6] Further reading: `https://home.cern/news/news/experiments/cool-running-cms-tracker`
[7] Further reading: `https://cms.cern/news/cms-magnet-warm`
[8] Further reading: `https://cerncourier.com/a/the-ls2-vacuum-challenge/`

of sensitive materials and detector layouts have been devised to make us 'see' these particles.

In the following sections in this chapter, the most important sub-detectors of CMS are enumerated and described. Despite their many differences in active material, shape, size, layout etc., they are based on a not-very-large number of categories in which particles can physically interact with matter to allow detection. These categories are the subject of this section.

Ionization One important category of interaction between particles and their environment is ionization, a type of electromagnetic interaction that involves a charged particle converting a part of its kinetic energy into separating positive and negative charges in the medium it passes through. Under the influence of an externally applied electric field, the positive and negative charges are pulled towards opposing electrodes, resulting in a measurable electric current. This is the basic principle behind the muon systems in CMS, where muons ionize gas molecules and the free electrons and ions drift towards opposing electrodes. The same principle also underlies the operation of the silicon tracker, where passing charged particles excite electrons from the valence to the conduction band in the depletion region of a reverse-biased pn-diode. Ionization was also at the basis of cloud chambers and bubble chambers, where the local ionization of an undercooled gas or overheated liquid by a particle passing by resulted in directly visible tracks of condensation or bubbles respectively.

Bremsstrahlung and pair production Electromagnetic interactions may further appear in the form of bremsstrahlung and pair production. The former process is the emission of a photon by a charged lepton (mainly electrons); the latter is the conversion of a photon into a pair of charged leptons (mainly electron-positron). Both processes are forbidden in vacuum but can occur in detector material via an interaction with a heavy nucleus. The combination of both interactions (at lower energies supplemented by ionization, Compton scattering and the photoelectric effect) gives rise to a cascade effect in which incoming electrons or photons lose all their energy by subsequent interactions. This cascade can be exploited for electromagnetic calorimetry. For example, at the low-energy end of the avalanche, the final electrons and photons can be absorbed by a scintillation material that in turn emits low-energy photons, e.g. visible light, which can be measured quantitatively and used to determine the original energy of the incoming particle.

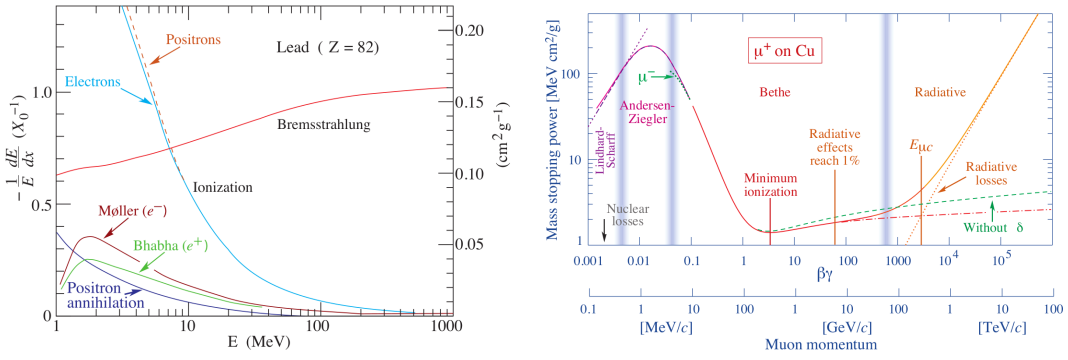


Figure 2.3: Passage of electrons (left) and muons (right) through matter. For typical electron energies in CMS of at least a few GeV, bremsstrahlung is the dominant interaction mechanism, supplemented by ionization at the low energy end of the shower. Muons in CMS are typically in the minimum-ionization regime. Figure from Ref. [58].

Strong interactions Strong interactions occur between hadrons, i.e. particles composed of quarks and/or antiquarks. In practical applications, this is typically set up as incoming hadrons interacting with densely packed atomic nuclei. Similarly to electromagnetic cascades, hadronic interactions may be chained one after the other into the development of a 'shower' or 'spray' of decay products in which the incoming hadron gradually looses all its energy. The final, low-energy particles at the end of the shower could for example be converted and detected in the form of light by the process of scintillation. In general, hadronic showers are more chaotic and less easily predictable than electromagnetic showers. This is because hadrons come in many different species, each with their own masses, decay properties and other characteristics. Moreover, a stochastic fraction of the decay products in hadronic interactions consists of photons, electrons, muons and neutrinos. The latter two types of particles typically escape hadronic calorimeters and degrade the energy resolution.

Elastic scattering Another possibility for particles to interact is by elastic scattering (the underlying mechanism of which can be both electromagnetic and strong in nature). In this interaction, the particles only recoil against each other like billiard balls, without transforming or emitting other particles. While this kind of interaction is exploited in some searches for dark matter (where one looks for the effects of recoiling nuclei caused by otherwise invisible particles), it is typically an unwanted effect in CMS as it changes the direction of particles stochastically and therefore reduces the momentum measurement resolution.

Others Several other types of interactions between particles and a sensitive material exist and are exploited for detection, such as inverse beta decay (via the weak interaction), Cerenkov radiation emission, the photo-electric effect, etc. They are not discussed here however as they are less relevant to the CMS sub-detectors described below.

2.4 Coordinate system and kinematic variables

Particles detected by CMS are often characterized in terms of their transverse momentum p_{T}, pseudorapidity η and azimuthal angle ϕ. These are kinematic quantities fully specifying the direction and momentum of a particle, specifically tailored to the cylindrical symmetry of CMS with the nominal proton-proton collision point at its origin:

- **Transverse momentum** The transverse momentum (p_{T}) is defined as the (magnitude of the) projection of the full momentum vector on the plane perpendicular to the axis of symmetry, i.e. the beamline. It is usually preferred over the (full) momentum, since the longitudinal component of the initial colliding partons is unknown; hence the kinematics of particles in terms of transverse momenta can be more straightforwardly compared to theoretical predictions. Moreover, the vector sum of transverse momenta of all particles created in a collision is expected to be zero, which can help to characterize particles not directly detectable by CMS.

- **Pseudorapidity** The pseudorapidity is defined as $\eta = -\ln(\tan(\theta/2))$ where θ is the angle of the particle momentum with respect to the beamline. It is preferred over θ since differences in pseudorapidity are (approximately) Lorentz-invariant with respect to boosts along the direction of the beamline, which is useful as the center-of-mass frame of the colliding partons does not necessarily correspond to the lab frame of CMS.

- **Azimuth** Finally, the azimuth (ϕ) is the angle in the transverse plane with respect to the axis pointing radially inward towards the center of the LHC. This choice is however arbitrary, and most physical quantities are expected to have no ϕ-dependence because of the radial symmetry of CMS.

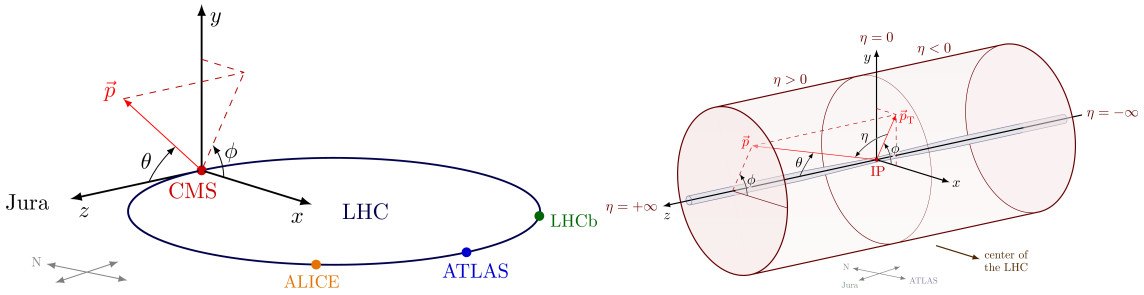


Figure 2.4: The conventional coordinate system used in CMS. Left: CMS coordinates with respect to the LHC and surroundings. Right: zoom-in on CMS, highlighting the cylindrical geometry. Figures from the *CMS coordinate system*.(*)

(*) `https://tikz.net/axis3d_cms/`

The cylindrical coordinates and their relation to Euclidean xyz-coordinates are displayed in Fig. 2.4. The spatial proximity of objects (tracks, clusters, particles, etc.) in CMS is often expressed in terms of their ΔR, which is defined as their distance in η-ϕ coordinates: $\Delta R = \sqrt{(\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2}$.

2.5 Solenoid magnet and return yoke

Apart from being (relatively) compact and having an extensive muon detection system (described in section 2.9), CMS derives its name from the solenoid magnet in which the tracker and calorimeters are contained. The main purpose of the solenoid magnet is to provide a strong magnetic field in the tracker volume, with the aim of deflecting the trajectory of traversing particles which allows a determination of their momenta by the measurement of the radius of curvature. Additionally, the magnetic field outside the magnet is contained spatially by a return yoke. This ensures that the magnetic field pervading the muon system (situated just outside the magnet) is still strong enough to bend the muon trajectories and assist in their momentum determination. The magnet consists of superconducting niobium-titanium (NbTi) cables, wound in a cylindrical shape providing a uniform magnetic field of 3.8 T over a length of 12.5 m and with a diameter of 6.3 m. The superconducting state is maintained by cooling the magnet down to 4 K (-269°C) using liquid helium [48].

2.6 Pixel and strip trackers

Trackers The innermost subdetectors of CMS are silicon-based trackers. Their goal is to allow track reconstruction of individual particles on the basis of a three-dimensional spatial sequence of 'hits', which are localized excesses of free charge carriers in the semiconductor cells, induced by a traversing charged particle. Tracking allows to distinguish individual particles and helps clustering them into groups likely to originate from the same parton (so-called 'jets', also described further on). Furthermore, the curvature of the tracks, immersed in a 3.8 T magnetic field, allows to measure the momentum and the sign of the charge of the particle. Note that photons and neutral hadrons do not leave a track in these subdetectors as the ionization mechanism requires charged particles.

Design criteria Since the number of charged particles per bunch crossing is on average about 1000 [59], and since a bunch crossing takes place every 25 ns, high-granularity, fast-response trackers are needed, with highly integrated on-detector electronics. However, this requires

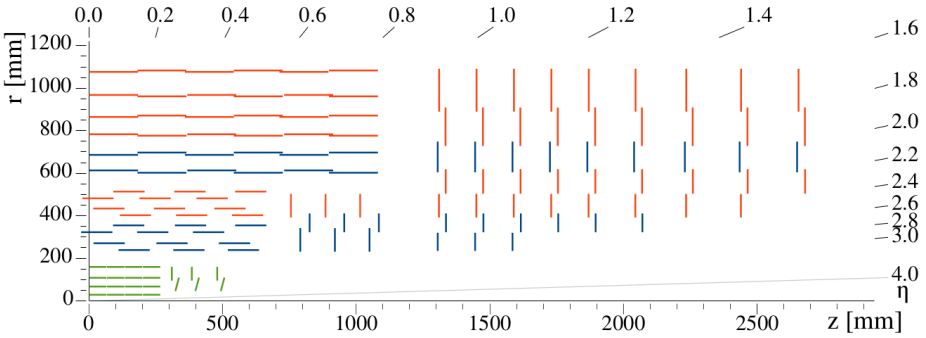


Figure 2.5: Layout of the current CMS tracker. The pixel detector is shown in green, and the strip detector in red (single-sided modules) and blue (double-sided modules). Horizontal layers form the barrel, (nearly) vertical layers form the endcap disks. An earlier version of the tracker, used before the start of 2017 data-taking, had one pixel barrel layer and one pixel endcap disk fewer. Figure from Ref. [61]

additional cooling material, which in turn conflicts with the idea of having as little intervening material as possible to avoid multiple scattering and unwanted interaction processes (e.g. bremsstrahlung and photon conversion). Moreover, the extreme charged particle fluxes may cause severe radiation damage, such as the trapping of free positive charge carriers near the insulating oxide layers, the disruption of the silicon crystal creating additional energy levels in the band gap, and unwanted ionization in the readout circuitry. The final design is a compromise between all these conflicting criteria [48].

Layout The tracker is divided into a *pixel tracker* and a *strip tracker*; the spatial layout of both subsystems is summarized in Fig. 2.5. The pixel tracker consists of 4 barrel layers and 3 endcap disks, arranged in a cylindrical geometry. It consists of 1856 modules of 160×416 silicon pixels each, where each individual pixel has a size of 100×150 μm^2 [49], highlighting the focus on the very fine spatial granularity for this subsystem.

The strip tracker has a similar structure and technology as the pixel tracker, but with more layers, and with larger, strip-shaped silicon sensors of varying size between 10 cm × 80 μm and 25 cm × 180 μm [48], chosen as a compromise between optimal resolution and a manageable number of readout channels. The determination of the coordinate of a traversing particle along the direction of the strips is very crude because of the length of the strips. To mitigate this, some of the modules in the strip tracker are double-sided, containing two back-to-back silicon strip sets with a small relative rotation (stereo angle of about 6 degrees), allowing a coarse measurement of the third spatial coordinate along the strip direction (z in the barrel, r in the endcap).

Performance In total, the tracker covers the full azimuthal range and a pseudorapidity up to $|\eta| < 2.5$, which corresponds to about 98.7% of the full solid angle. The cluster hit efficiency, i.e. the efficiency of detecting (a cluster of) sensor hits if a charged particle traverses a detector layer, is well above 99% for all barrel layers and endcap disks at an instantaneous luminosity of $1{\times}10^{34}$ cm^{2}s^{-1}, and remains above 99% at double that luminosity except in the innermost pixel layer. The position resolution of hits in the pixel tracker depends on the track angle and the radial position of the layer, but is generally in the order of 10 μm in the $r\phi$-direction and 20 μm in the z-direction [49]. For non-isolated particles of $1 < p_\mathrm{T} < 10\,\mathrm{GeV}$, the track resolutions are typically 1.5% in p_T and 20–75μm in the transverse impact parameter [60].

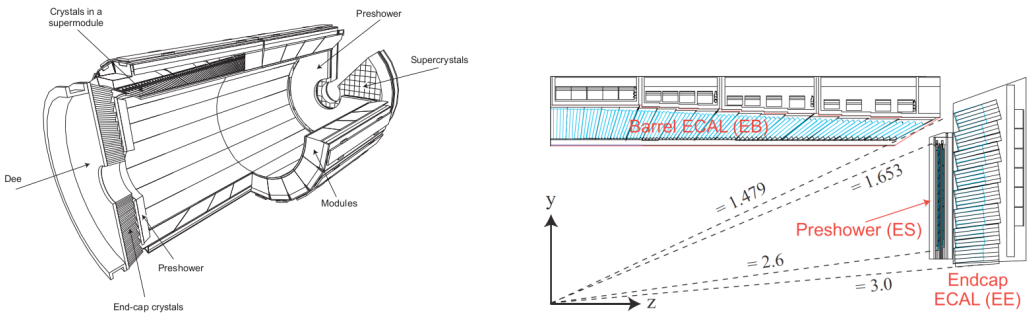


Figure 2.6: Layout of the current CMS electromagnetic calorimeter (ECAL), with indication of barrel, endcap and preshower. Figure from Ref. [62].

2.7 Electromagnetic calorimeter

ECAL The electromagnetic calorimeter (ECAL) measures the energy of photons and electrons. The interaction of photons with the ECAL material causes them to convert in electron-positron pairs, while electrons (and positrons) emit bremsstrahlung photons. This causes a chain reaction (or 'shower') in which incoming electrons and photons lose all their energy and are totally absorbed. The energy measurement is based on the amount of light that is emitted in the chain reaction by the process of scintillation. Lead tungstate ($PbWO_4$) is chosen as scintillation material because of its high density, its radiation hardness, its fast scintillation time and its ability to contain the electromagnetic showers within small longitudinal and transverse distances [48].

Layout The shower confining capacity of lead tungstate allows for a compact calorimeter with fine spatial granularity. The ECAL in CMS consists of about 60,000 crystals of size 22×22 mm^2 (corresponding to approximately $\Delta\eta \times \Delta\phi = 0.0174 \times 0.0174$) in the barrel ($|\eta| < 1.479$) and about 7,000 of size 29×29 mm^2 in each of the endcaps ($1.479 < |\eta| < 3.0$) [48]. The crystals are tightly designed to cover the full space of the ECAL, with a truncated pyramid shape that gets wider towards the back side. The blue-green scintillation light generated in those crystals is converted into an electrical signal by avalanche photodiodes (in the barrel) or vacuum phototriodes (in the endcaps) mounted at the back of the crystals.

Preshower The ECAL endcap is supplemented with a 'preshower' detector in the range $1.653 < |\eta| < 2.6$, with the principal aims of identifying neutral pions (decaying to two photons), improve electron identification and enhance the position determination of both electrons and photons. It consists of two layers, each of which contain a lead radiator to initiate the shower and silicon strip sensors to measure the shower profile and energy [48]. The overall geometry of the ECAL, including preshower, is displayed in Fig. 2.6.

Energy resolution The energy resolution of the ECAL is strongly η-dependent, varying from about 1% at low $|\eta|$ to about 5% at high $|\eta|$, measured for electrons in $Z \to e^+e^-$ decays [62,63]. For photons, similar values between 1% and 5% are reported, depending on its pseudorapidity and conversion time [63, 64].

2.8 Hadron calorimeter

HCAL The energy of charged and neutral hadrons is measured in the hadron calorimeter (HCAL), which envelopes the ECAL. The HCAL layer makes CMS an almost hermetic detector: apart from muons and particles not directly detectable by CMS (like neutrinos), all particles produced in the central collisions that make it to this stage are finally absorbed and measured by the HCAL (up to small geometric and other inefficiencies). The HCAL units consist of brass absorbing layers that initiate hadronic showers, and plastic scintillation layers with photomultipliers that convert the shower into an electrical signal. The distinctive property of the HCAL is that it detects and measures the energy of neutral hadrons, which do not leave tracks in the tracker and deposit only a little amount of energy in the ECAL. These particles' energy and momentum yield important contributions to the total transverse momentum budget of each collision event. By combining energy and momentum estimates from all subdetectors, one can determine the amount of momentum carried by particles which escape direct detection, notably neutrinos.

Layout The HCAL consists, as usual, of a barrel (up to $|\eta| < 1.39$) and endcap part (up to $|\eta| < 3.0$). The segmentation of the modules is coarser than for the ECAL: $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$ for $|\eta| > 1.6$ [48, 65].

Forward and outer HCAL The HCAL barrel and endcaps are supplemented by a very forward part and an outer part. The forward HCAL is a module that covers high pseudorapidity values up to $|\eta| < 5.2$, in order to allow identification of jets that are almost parallel to the beamline and increase the total coverage of the detector. Because of the requirement to resist the high radiation levels at the largest $|\eta|$ values, the design of the forward HCAL is different from the other parts, with steel absorber plates and quartz fibers capturing the Cherenkov light emitted by the electromagnetic component of the shower. The outer HCAL consists of additional absorbers and scintillators, situated outside of the CMS magnet. Its function is to capture the tails of highly energetic jets, for which the main barrel HCAL, confined between the ECAL and the inside of the CMS magnet, is too shallow to fully contain the shower [48].

Energy resolution The energy resolution of the HCAL depends on the subsystem and energy of the incoming particle. It can be parameterized as follows [66]: $\sigma_E/E = 0.85\,\mathrm{GeV}^{1/2}/\sqrt{E} \oplus 0.07$ (for the barrel and endcaps) and $\sigma_E/E = 1.98\,\mathrm{GeV}^{1/2}/\sqrt{E} \oplus 0.09$ (for the forward HCAL), where the energy is expressed in GeV and the two terms are added in quadrature. For typical energies of around 50 GeV, this yields a resolution of about 14% in the barrel and endcaps, and about 30% in the forward HCAL.

2.9 Muon system

The importance of muons Muons are of special importance to CMS, as implied by its middle name, the reason being that muons are relatively easy to identify and measure accurately. Many known or hypothesized processes in the standard model and beyond can result in final states involving muons, making them an important object of analysis. Muons undergo only little interaction in the calorimeters and hence are the only non-(quasi-)sterile particles that penetrate to the outer layers of CMS (apart from relatively small backgrounds, such as exceptional punch-through hadrons), making muon identification and triggering relatively easy. Furthermore, muons exhibit less radiative losses in the tracker material than electrons, allowing more accurate momentum measurements. The outer CMS subdetectors, specifically focused on muon detection, are located outside of the magnet, interwoven with the iron return

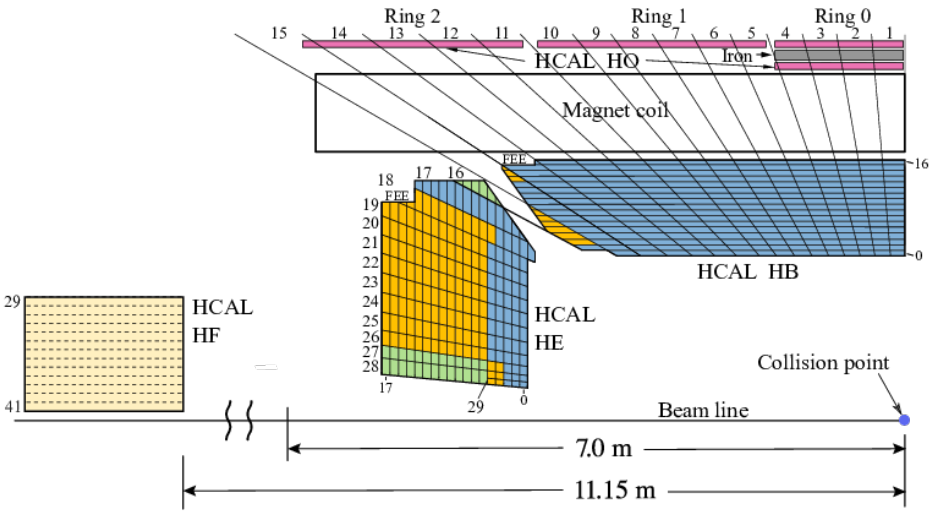


Figure 2.7: Layout of the current CMS hadron calorimeter (HCAL), with indication of the barrel, endcap, outer and very forward component. Figure from Ref. [65].

yoke that contains the magnetic field.

Gas-based detectors Contrary to the solid-state trackers and calorimeters, the muon system is made of gas-based detectors, that collect the ionization charge induced by a muon traversing the gas. Gas-filled detectors are in general lighter and less expensive than solid-state detectors and can more easily be scaled to large detection volumes. Though all muon detectors in CMS are based on ionization in gas, the implementation details differ, and four different detector types can be distinguished: DTs, CSCs, RPCs and GEMs[9], discussed separately in the paragraphs below. The layout of the muon detection systems are shown in Fig. 2.8 and a sketch of their operational principle is shown in Fig. 2.9.

DT Because of the relatively low flux of particles that make it beyond the HCAL in the barrel region, a relatively coarse and slow drift tube (DT) system meets the demands sufficiently. Drift tubes are gas-filled tubes with a thin positively charged anode wire spanned in the middle. When the gas atoms are ionized, the free electrons drift towards the anode wire and form an electrical signal. There are 4 stations of DTs, interspersed with the iron return yoke for the magnet. Each station consists of 12 layers of tubes (or 8 in the outermost station), stacked perpendicular for optimal position resolution. The DT system covers the range $|\eta| < 1.2$ [68].

CSC In the endcaps, where muon and background rates are higher, cathode strip chambers (CSC) are installed. Instead of individual cells containing a single wire in the middle, CSC chambers contain many parallel anode wires, as well as perpendicularly oriented cathode strips. Due to this geometry, CSCs provide better position resolution than DTs. Furthermore, because of the shorter drift path, CSCs have a fast response time. CSCs are installed in the region $0.9 < |\eta| < 2.4$. Each CSC chamber consists of 6 layers, each of which providing a two-dimensional muon position measurement [68].

The DT and CSC systems together cover the whole $|\eta| < 2.4$ range and provide good muon detection efficiency. Furthermore, their response time is generally fast enough to identify the

[9] Only DTs, CSCs and RPCs were used during the data taking used in this thesis. GEMs are a relatively new addition to the CMS muon system; the first ones being installed for commissioning only in 2019 [67].

correct bunch crossing where the muon originated from, and their position resolution fine enough to provide an accurate muon momentum estimate. Both properties are essential for muon triggers. Yet, to further increase the trigger efficiency and redundancy, an additional type of detector is used.

RPC Both in barrel and endcaps (up to $|\eta| < 1.6$), resistive plate chambers (RPC) are added to the DT and CSC layers to provide reliable trigger information due to their intrinsic fast response and good time resolution down to O(1 ns). Contrary to DTs and CSCs, RPCs are not tracking detectors consisting of many layers; instead they consist of only two stacked thin gaps over which a very high voltage is applied. Ionization electrons are accelerated and generate an avalanche of secondary ionizations, and the total charge is collected in readout strips. The timing resolution of RPCs is better than for DTs and CSCs, at the cost of a worse spatial resolution; hence they provide complementary information and trigger redundancy.

The entire muon system, consisting of DTs, CSCs and RPCs (before GEMs were installed), shows a very high muon identification efficiency. In terms of individual hits or segments in the detectors, the efficiency is in the range 94-98%, depending on the subsystem, and has remained stable over many years of data taking. In terms of full muon reconstruction, the efficiency is even higher, typically 95-100% depending on the kinematic properties of the muon [69].

GEM The latest addition to the CMS muon system consists of gas electron multipliers (GEM). Like RPCs, GEMs are avalanche detectors where an initial primary ionization generates a cascade of secondary ionizations because of the strong electric field. The special feature of GEMs is a stack of thin polymer foils coated with copper, perforated with 70 μm wide holes. Inside the holes, the electric field is stronger than in the bulk of the drift gap, enhancing the ionization avalanche. GEMs display a high rate capability and radiation hardness, which is why the GEM system, when completed, will extend up to $|\eta| < 2.8$, thus extending the muon coverage of CMS. In the regions where the GEMs overlap with other muon detectors, they will increase the number of hits produced by a muon, improving the reconstruction accuracy and maintaining the trigger efficiency even at higher values of instantaneous luminosity [67].

2.10 Runs, upgrades and high-lumi LHC

Continuous LHC development Though the LHC and its experiments have been operational for about 15 years, they have not yet reached a 'final' form, but rather are in continuous development. The active time of the LHC is divided in so-called 'Runs'[10], with major upgrades in between them. After an extensive commissioning and testing period, the first real physics data was taken during 2011–2012 (Run I) at a center-of-mass energy of 7 and 8 TeV. Run I resulted, among other things, in the discovery of the Higgs boson [70, 71]. After a long shutdown and gradual re-commissioning of the accelerator and its detectors, Run II spanned the years 2016–2018, at 13 TeV. It is the data of Run II that is used in the analyses discussed in this dissertation. After another long shutdown, Run III is currently ongoing at 13.6 TeV. The delivered luminosity per data-taking year is displayed in Fig. 2.10. Note that though the total integrated luminosity for 2022 data is low compared to a typical Run II data-taking year, the instantaneous luminosity (the slope of the curve) is higher, which will lead to larger datasets in further data-taking years of Run III.

[10] This term has multiple meanings. It can refer to a single period of uninterrupted data acquisition, typically lasting a few hours. But in this context, it refers to a time period of typically a few years, in between long shutdowns.

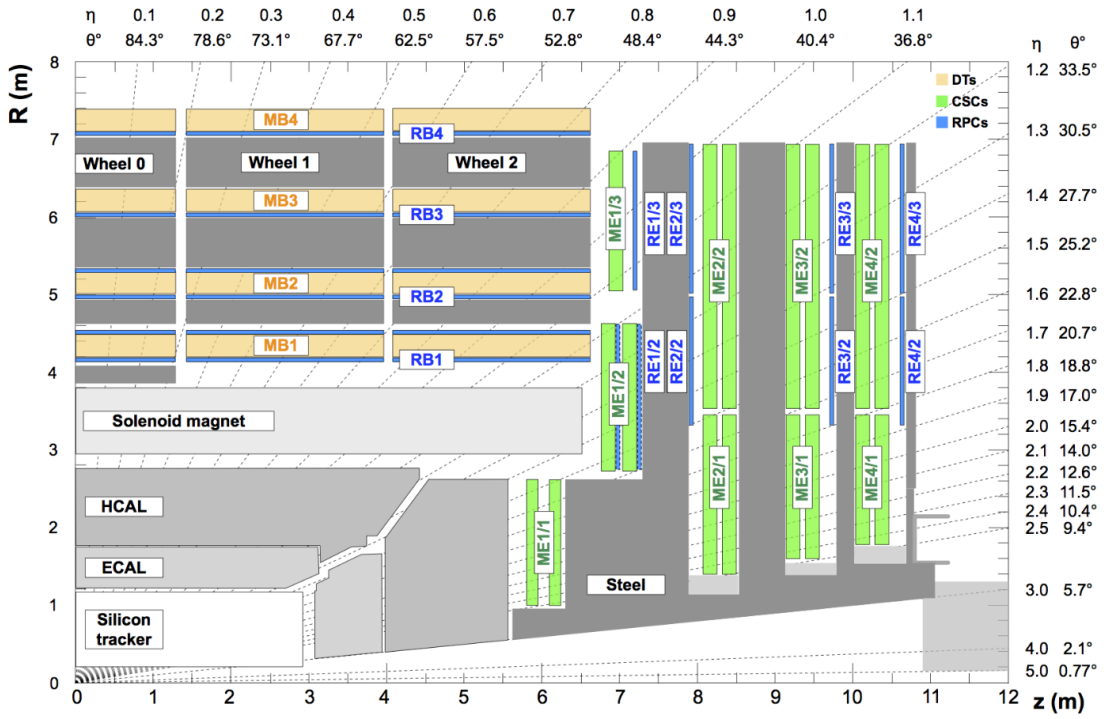


Figure 2.8: Layout of the CMS muon detectors. This is a snapshot from 2018 and does not include the GEM detectors yet. Figure from Ref. [69].

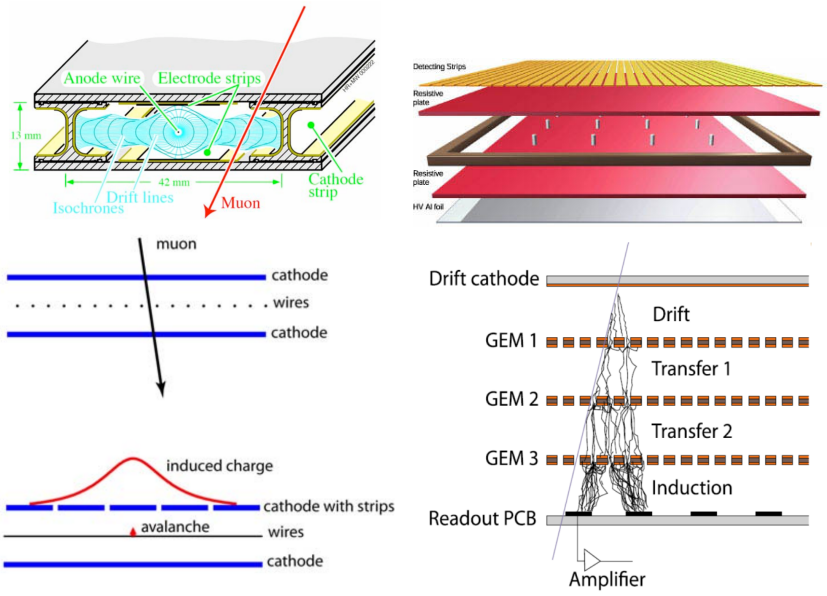


Figure 2.9: Sketch of the design and/or operational principle of the muon detectors in CMS: drift tube cells (upper left), resistive plate chambers (upper right), cathode strip chambers (lower left), and gas electron multipliers (lower right). Figures from Ref. [48, 67] and the *CERN website*.(*)

(*) `https://cms.cern/detector/detecting-muons/resistive-plate-chambers`

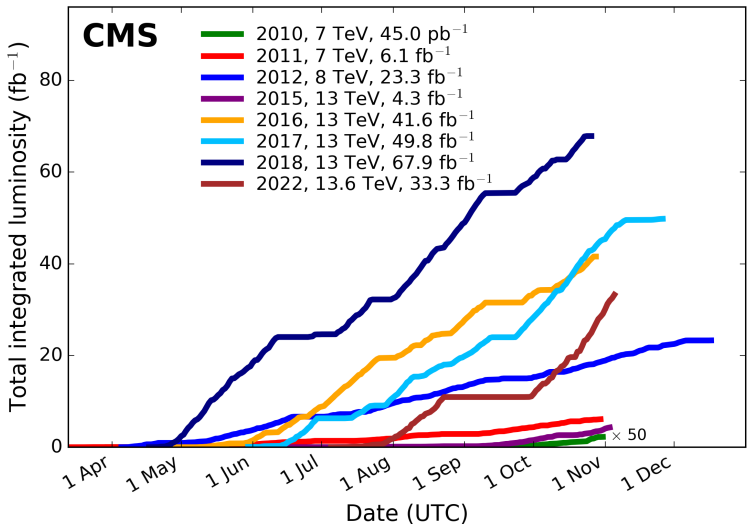


Figure 2.10: Cumulative luminosity versus time delivered to CMS during stable beams for proton-proton collisions at nominal center-of-mass energy. Figure from the *CMS luminosity public results*.(*)

(*) `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`

High-lumi LHC The next phase for the LHC is called 'high-lumi LHC'. After another long shutdown with major upgrades, data taking will resume with even higher instantaneous luminosity, leading to larger datasets for precision measurements of standard model properties and potentially for the discovery of beyond-standard-model effects that were up to now hiding in limited statistics. After that, the LHC will have reached the end of its lifetime, and the role of flagship particle accelerator will be taken over by even larger accelerators such as the FCC, for which the LHC could potentially serve as a mere injector.

CMS Upgrade The detectors installed on the LHC have to follow and adapt to the changing accelerator conditions. For example, higher luminosity goes hand in hand with higher pileup, imposing more stringent requirements on the detector granularity and reconstruction software to identify particles from individual collisions. Moreover, increasing luminosity causes more radiation damage by highly energetic particles that cross the detector, to which the components have to made resistant. A small part of the general upgrade project consists of new RPC stations with an improved design, which will be shortly revisited in section 2.14.

2.11 Triggering

The need for triggers A crossing of proton-proton bunches takes place at the center of CMS every 25 ns. The resulting data flow consists of 1 MB-events [72] at a rate of 40 MHz, which is many orders of magnitude larger than what can be stored or fully processed. Furthermore, the vast majority of collision events in CMS consist merely of soft QCD scattering, because of the very large cross-section of such interactions. These events are not particularly interesting to CMS, as instead its physics program is focused on electroweak production of heavy bosons. Therefore, a two-tiered trigger system [73, 74] filters potentially interesting events for full reconstruction and storage.

Level-one trigger In a first stage, a custom hardware level-one trigger (L1T) reduces the rate of events from 40 MHz to a maximum of 100 kHz. The L1T is designed for filtering events quickly to sustain the high input rate. It uses coarse-granularity detector data from the calorimeters and muon systems only (i.e. tracks are not yet reconstructed at this stage), resulting in a latency of maximum 4 μs per event. On a software level, the L1T consists of a collection of various filtering algorithms (called 'seeds'), each targeting a specific signature of potentially interesting events. These signatures typically consist of electrons, muons, photons and/or jets with relatively high p_{T} and low $|\eta|$ (with respect to typical 'uninteresting' soft QCD signatures), or a large missing transverse momentum. The output rate of some L1T seeds with loose selection requirements is managed by a 'prescale', meaning that only one out of several events that pass the selection is retained and sent to the next processing steps. The value of the prescale can be fine-tuned dynamically to keep the L1T output rate as close to the maximum as possible, even in varying detector conditions (e.g. decreasing instantaneous luminosity as a long run progresses).

High-level trigger Further filtering is performed by a high-level trigger (HLT), bringing the final event rate at a level of about 100 Hz in the original design, which in practice has grown to about 1 kHz during Run II. In contrast to the L1T, The HLT uses inputs from all CMS subdetectors (including the tracker) and applies a fast version of the full event reconstruction software. This allows for more complicated filter criteria but takes a longer time (about 100 ms for most events). Whereas the L1T hardware was designed specifically for CMS, the HLT consists of custom software only, running on commercially available hardware (a large collection of processors). This design choice resulted from the observation that processor electronics, data-acquisition and storage systems are evolving rapidly. Contrary to most other components of CMS, the HLT is designed in such a way as to allow hardware modifications without much ado. This flexibility also allows to use different triggers over time, in order to increase the signal efficiency for some analyses or to compensate for unforeseen backgrounds or experimental conditions.

Parking and scouting Two alternative trigger strategies, called parking and scouting [72] have been implemented by CMS, aimed at overcoming potential blind spots in the default strategy[11]. Data scouting consists of storing the output of the HLT reconstruction algorithm of selected events directly (and removing all lower-level data), rather than performing a second reconstruction using a slower but more accurate reconstruction algorithm as is usually done. This strategy drastically reduces the event size and removes the need for the second reconstruction altogether, and hence allows to process and store events at a much higher rate, i.e. with lower trigger thresholds. In data parking, the second reconstruction is not skipped, but rather delayed: the full content of events selected by the HLT are moved to tape storage and wait there until computing resources are available to reconstruct them (e.g. during long shutdowns). As the bottleneck in regular data taking is not the storage bandwidth, but rather the computing capacity for reconstruction, this strategy also allows to store a higher rate of events than is possible with the default triggering approach.

2.12 Event reconstruction

From detector response to reconstructed particles There is a gap that needs to be bridged between the raw detector response (e.g. localized electric charge excesses in the tracker

[11] These alternative trigger strategies are mentioned here for completeness, but are generally more relevant for searches with exotic signatures or for b-physics specifically, rather than for standard model measurements using top quarks, and hence they are not used in this thesis.

pixels) and the input for higher level data analyses (e.g. an electron with a given p_{T}, η and ϕ). The response of the CMS tracker, calorimeters and muon systems to a given collision event has to be translated into a collection of particles with determined kinematic properties. This step is called event reconstruction, and its algorithms are briefly described in this section. Paragraph 2.12.1 deals with the local aggregation of ionization hits or energy deposits into tracks and clusters respectively. These elements form the inputs for the particle-flow algorithm, discussed in paragraph 2.12.2. The subsequent paragraphs 2.12.3 and 2.12.4 present some additional details for muons and electrons. Finally, paragraph 2.12.5 presents jet clustering. Muons, electrons and jets constitute the basic objects that the higher-level data analyses described in this dissertation start from.

2.12.1 Tracks and clusters

Hits Charged particles traversing the tracker produce free charge carrier excesses in the semiconductor pixels and strips they pass through. The deposited charge is digitized into a so-called 'hit'. Additional complications arise as particles typically deposit charge in several adjacent pixels or strips as they pass a single detector layer. Therefore the excited pixels or strips are combined into a single hit and the position of the particle passage is estimated from a comparison with simulated charge depositions, with a spatial resolution better than the physical size of a single pixel or strip [59].

Tracks After all hits have been determined, they are used as input for the track finding and fitting procedure. A track is a sequence of hits in subsequent tracker layers likely to originate from a single particle traveling through the detector. Accurate track identification and characterization is important since tracks provide information on the direction, momentum, electric charge and point of origin of the particles in a collision event. They are also vital for jet clustering and heavy-flavor tagging.

CMS employs an iterative track reconstruction procedure. In a first step, only reconstructed tracks with high quality and confidence are retained, and the corresponding hits are removed from the hit collection. In following steps, the same procedure is applied on the remaining hits, but the quality requirements are gradually relaxed. This procedure simultaneously allows a higher reconstruction efficiency for genuine tracks and a lower number of fake tracks (from incorrectly grouped hits) than what could be achieved with a single iteration [59].

Within a given step of the iterative procedure, tracks are reconstructed as follows. Seeds are formed from a small number of nearby hits, providing initial track candidates and a rough estimation of their direction. Starting from the seeds, hits from more outward detector layers are added one by one, using a Kalman filter approach [75, 76]. In a nutshell, the Kalman filter predicts the expected position of a hit in the next tracker layer based on its current estimate of the track parameters and hit position in the current layer. If a hit is found within the predicted region, it is added to the sequence of hits and used to update the estimate of the track parameters before continuing to the next layer. When a track is found in this way, it is refitted two times using an inward and an outward going Kalman filter, with more complicated propagators than in the track finding stage, and the final track parameters are determined from a combination of both fits [59].

Muon tracks Hits in the muon systems are combined into tracks as well, in a conceptually similar way as described above for the silicon tracker, but simplified since the occupancy is generally much lower. Seeds are formed from a small number of nearby hits in the DT or CSC systems, and muon tracks are built from these seeds by combining all hits in the DT, CSC

and RPC systems [77].

Vertices The point in space where a proton-proton collision takes place is called a vertex[12]. Vertex reconstruction is important to disentangle several collisions and their products that might occur in the same bunch crossing (pileup). The procedure for vertex reconstruction consists of selecting good tracks (with sufficient hits and low transverse impact parameter with respect to the beam line), clustering their z-coordinate using a deterministic annealing algorithm, and finally fitting the exact vertex position from all tracks assigned to this vertex [59].

From all reconstructed vertices in a bunch crossing, the one with the hardest scattering is chosen to be the primary vertex as it has the highest potential for containing an interesting proton-proton interaction from the point of view of the CMS physics program [78]. Tracks and other detector signals associated with other proton-proton interaction vertices than the primary vertex are assumed to come from pileup interactions and are mainly used for pileup mitigation.

Clusters Apart from tracks in the tracker and muon systems, clusters of energy deposits in the calorimeters are an important input to the particle-flow algorithm. They are vital for detecting photons and neutral hadrons (that do not leave tracks), as well as electrons (see section 2.12.4). The clustering procedure starts from a number seeds, which are calorimeter cells with locally maximal energy deposition above a certain threshold, and then proceeds to assign neighboring cells to the growing clusters if their energy deposit is sufficiently high with respect to the expected noise level. If a cluster constructed in this way encompasses several seeds, it is assumed they come from multiple nearby particles and their individual energy deposits are reconstructed by fitting a sum of Gaussian distributions to the observed cluster. Both the ECAL and the HCAL are calibrated with samples of simulated photons and neutral hadrons respectively to account for nonlinear and nonuniform detector response [77].

2.12.2 Particle flow

The particle-flow algorithm [77, 79] is a central feature of the event reconstruction at CMS. It aims to reconstruct different types of detected particles (i.e. electrons, photons, muons, charged hadrons and neutral hadrons) by using an optimal combination of the CMS subdetectors. In a simplified form, the algorithm proceeds as follows:

- **Elements** First, a list is made of all 'elements' belonging to an event. An element is a locally reconstructed object that can either be a track in the silicon tracker or in the muon chambers, or a cluster of energy deposits in the ECAL or HCAL units.
- **Blocks** Next, a distance measure is calculated between each pair of elements. For example, the distance between an energy deposit and a track can be defined as the angular separation between the deposit and the track position extrapolated to the radial depth of the calorimeter. All elements with a sufficiently small distance in between them are grouped in 'blocks'.
- **Muons and electrons** Within these blocks, muons and electrons are first identified as a combination of specific elements, e.g. tracker and ECAL elements with compatible energy estimates.

[12] Vertex can also refer to the point in space where one of the particles produced in a collision decays, if that particle has moved a distinguishable distance away from its production vertex before decaying. These decay vertices are typically called secondary vertices and are used for several purposes, such as heavy flavor tagging (section 2.12.7) and searches for hypothesized long-lived particles.

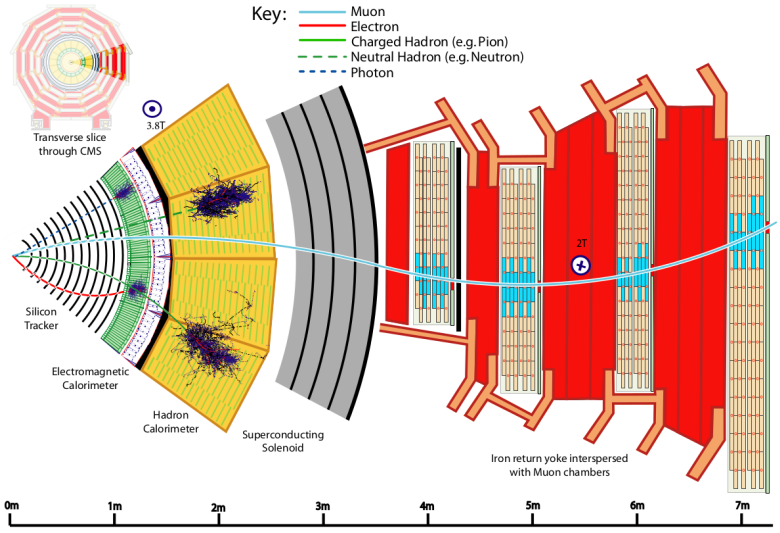


Figure 2.11: Schematic view of the CMS detector: tracker layers (black lines), ECAL (green), HCAL (yellow), superconducting magnet (grey) and muon chambers (between the red areas). Indication of several types of detected particles with their experimental signature: a muon (blue line), an electron (red line), a photon (blue dashed line), a charged hadron (green line) and a neutral hadron (green dashed line). Figure from Ref. [77].

– **Charged hadrons** The remaining tracks in the block are attributed to charged hadrons, and are combined with energy deposits in the ECAL and HCAL using a combination of angular distance and energy compatibility.

– **Photons and neutral hadrons** Finally, the remaining energy deposits not matched to any track, as well as the excess energy in clusters matched to tracks, are associated to photons or neutral hadrons.

Detector signatures The typical signatures of the different types of particles traversing CMS are sketched in Fig. 2.11. Muons are reconstructed from a track in the muon system and a matching track in the tracker, while they typically do not leave significant energy deposits in the calorimeters. Electrons are reconstructed from a track in the tracker and a matching shower in the ECAL, with additional matching deposits from bremsstrahlung photons. Photons are absorbed by the ECAL but do not leave a trace in the tracker since they are electrically neutral. Hadrons typically leave only a part of their energy in the ECAL and are fully absorbed by the HCAL instead; charged hadrons additionally leave a track while neutral hadrons do not. Exploiting these differences in experimental response, a global combination of the information from all subdetectors leads to an accurate identification and characterization of detected particles.

2.12.3 Muons

Specifically for muons, the dedicated muon chambers at the outer edge of CMS provide very accurate identification information, since the background from other types of particles is very low in this region. Therefore, the particle flow algorithm reconstructs muons by matching hits in the muon system to hits in the tracker. Two complementary approaches are used: either the reconstruction is done outside-in, where reconstructed tracks in the muon system (so-called

'standalone muons') are matched to those in the inner tracker (the resulting objects are called 'global muons'), or an inside-out approach is used, where each tracker track is extrapolated to the muon system and matching segments in the muon chambers are sought for (in which case the result is called a 'tracker muon').

Global muons have both a high purity (mainly thanks to the muon systems) and an accurate momentum determination (mainly thanks to the tracker). The granularity of the muon systems is generally too coarse to contribute to an accurate momentum estimate (compared to the tracker), except for very high-p_{T} muons (with almost straight tracks), where the larger spatial extent of the muon systems allows for a better estimation of the radius of curvature.

2.12.4 Electrons

Electrons are difficult to reconstruct since they emit bremsstrahlung photons on their way through the tracker. This deteriorates both their energy measurement and their track determination. In order to retain accurate track reconstruction, possible electron tracks are refit using a Gaussian-sum filter [80] instead of the standard Kalman filter [75, 76]. This type of filter uses a mixture of Gaussian distributions to estimate the energy loss due to bremsstrahlung rather than a single one, and can better follow the significant changes in direction of an electron track due to this emission. Moreover, in an attempt to catch the electron energy lost by bremsstrahlung, a tangent line to the electron track is extrapolated from each tracker layer towards the ECAL. If an energy deposit is found in this place, it is considered to originate from the decelerating electron. An electron may therefore be reconstructed from a track and multiple ECAL deposits.

2.12.5 Jets

Showering and hadronization Quarks and gluons that may originate from the primary interaction between the proton constituents or from secondary decays cannot propagate as free particles due to confinement and self-interaction respectively. They therefore undergo processes of showering and hadronization, resulting in more-or-less collimated collections of particles called jets. Jet reconstruction is useful to recover the properties of the original quarks or gluons emerging from the primary interaction. Secondarily, jet clustering greatly reduces the complexity of an event (as opposed to a description at the level of individual hadrons).

The anti-k_{T} algorithm CMS uses the anti-k_{T} algorithm [81] for jet clustering. Starting from a list of all particle-flow candidates, the distances between each pair of them (i and j) is calculated as follows:

$$d_{ij} = \min(p_{T,i}^{-2}, p_{T,j}^{-2}) \cdot \frac{\Delta_{ij}^2}{R^2}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and where $p_{T,i}$, y_i and ϕ_i are the transverse momentum, rapidity and azimuthal angle of particle i respectively. The 'cone size' parameter R can be tuned and is typically set to 0.4 for the analyses discussed in this dissertation. Furthermore, for each particle, the following property is calculated:

$$d_{iB} = p_{T,i}^{-2}$$

Among the list of all distances d_{ij} and d_{iB}, the minimum is chosen. If it is a d_{ij}, particle i and j are taken together (their four-vectors summed) and considered as a single particle or pseudo-jet in the next steps. If it is a d_{iB}, particle or pseudo-jet i is taken to be jet and removed from the further steps. After this, the distances are recalculated and the procedure

repeated until all particles are clustered.

Properties The anti-k_T algorithm has the property that large-p_T particles serve as clustering centers, around which neighboring smaller-p_T particles are clustered, due to the combination of the negative squared p_T and the directional difference in the distance metric. The algorithm is infrared and collinear safe (meaning that the resulting clusters are nearly invariant under the addition of low-p_T particles and that it counts jets originating from almost collinear particles correctly).

Jet energy The energy contained in a jet is ideally the sum of the energies of the particles contained in it. In practice, measuring the jet energy is not straightforward for a number of reasons. Individual particles inside jets might have overlapping calorimeter deposits and not be well distinguished; background particles from other proton-proton interaction vertices (pileup) contaminate the detector response; the calorimeter response is typically not uniform, nor linear with respect to the 'true' energy. An involved procedure of jet energy corrections (JEC) tries to mitigate these difficulties. Corrections are applied to subtract pileup, to account for the detector response (using simulation), and finally for residual differences between data and simulation. On top of these, the jet momenta in simulation are smeared out to account for the worse jet energy resolution (JER) observed in data compared to simulation [82–84]. These corrections are significant and come with sizable uncertainties, the handling of which is discussed in section 6.5.

2.12.6 Missing transverse momentum

CMS is designed to be as close as possible to a hermetic detector, meaning that it tries to catch all particles produced in a collision event, to have an as complete as possible description of the full event. Neutrinos are however not directly detectable by CMS, as their interaction with ordinary matter is too weak to be captured by any of the CMS sub-detectors. Apart from these SM particles, there are a number of hypothesized BSM particles that would not leave a trace in CMS either (for example, dark matter candidates). A handle on their presence can be obtained from the conservation of momentum: as the net transverse momentum of the incoming particles is essentially zero, this is expected to hold for the collection of outgoing particles as well[13]. The missing transverse momentum ($\vec{p}_T^{\,\text{miss}}$) is defined as the negative vector sum of all directly detected objects (leptons, photons and jets):

$$\vec{p}_T^{\,\text{miss}} = -\sum_i \vec{p}_{T,i} \tag{2.3}$$

Under the assumption of the standard model, $\vec{p}_T^{\,\text{miss}}$ represents the vector sum of the transverse momenta of all neutrinos in an event (modulo detector inefficiencies or inaccuracies in measuring the visible $\vec{p}_T$). Individual neutrinos cannot be distinguished from one another however, nor from other hypothesized invisible particles.

2.12.7 Heavy-flavor tagging

The analyses discussed in this dissertation target processes with top quarks. As top quarks decay almost exclusively to b quarks, identifying jets originating from a b quark is of major importance. This procedure is called b tagging[14].

[13] The same does not hold for the longitudinal momentum, the exact value of which in the initial state is unknown due to the stochastic sampling from the PDF.

[14] The more comprehensive term heavy-flavor tagging also includes identifying jets originating from c quarks as opposed to jets from b quarks and light partons (u, d, s, or g). In this dissertation, c tagging was not used, so b tagging and heavy-flavor tagging are used interchangeably.

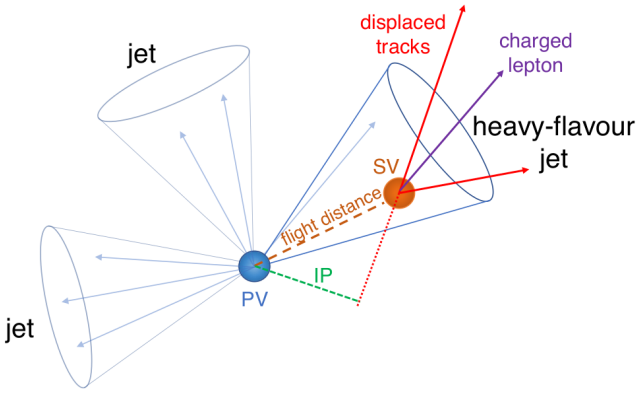


Figure 2.12: Sketch of a b jet with its characteristic features: secondary vertex, high impact parameter, not strongly collimated decay products, and the presence of a charged lepton. Figure from Ref. [85].

Several properties of b hadrons (i.e. hadrons containing a b quark) are exploited [85]:

- **Secondary vertex** The lifetime of b hadrons is in the order of 1.5 ps, which, depending on the momentum, translates into a traveled distance before decaying of a few mm to one cm. This distance is resolvable by the CMS reconstruction algorithms, so that a secondary vertex can usually be reconstructed from the b hadron decay products.
- **Impact parameter** The impact parameter of a track with respect to the primary vertex is defined as the distance of closest approach between the primary vertex and the (possibly extrapolated) track. It is expected to be small for tracks originating from the primary vertex, and larger for tracks originating from b decay at a secondary vertex, as can be seen in Fig. 2.12
- **Transverse momentum** Because of the b quarks higher mass, its decay products have on average a larger transverse momentum with respect to the jet axis than those of jet constituents in light jets.
- **Presence of leptons** In approximately 20% of the cases, the decay of a b hadron results in a charged lepton, the presence of which can also be exploited to tag b hadron decays.

The b-tagging algorithms used in this thesis are DeepCSV [85] and DeepJet [86, 87]. For the DeepCSV algorithm, several input variables are constructed based on the general considerations mentioned above, and then combined in a deep neural network. The DeepJet algorithm improves upon this approach by having a more comprehensive and low-level set of input features, and a more involved network architecture including convolutional and recurrent layers.

2.13 Event simulation

Measuring a given number of events with given kinematic properties conveys only little information if there is no prediction that specifies how many one would have expected to see on the basis of the standard model. In most cases, that standard model prediction is encapsulated in simulated events (with calculated cross-sections). These simulated events are analyzed in the same way as the observed events, so that a final comparison between both can be made,

both in terms of the total number and the kinematic distributions. Event simulation is done in three distinguishable steps: matrix element calculation, hadronization, and detector simulation, each of which are discussed in slightly more detail in the following paragraphs.

Matrix element calculation The term matrix element is derived from terminology involving a 'scattering matrix' between a set of initial and final states. In our current paradigm, a matrix element is basically the mathematical translation of a Feynman diagram (see section 1.1.4), or more accurately, the sum of all possible Feynman diagrams with the given initial and final state. Matrix element level calculation then proceeds conceptually by defining an initial and final state, drawing all possible Feynman diagrams between those states (starting from lowest order and typically truncated at a given order), translating them into mathematical expressions, and summing them up. This perturbative procedure works well for the core process of a parton-parton scattering event, since these interactions occur at high energy scales where the strong coupling constant α_S is effectively small enough for low-order approximations to be somewhat accurate.

Of vital importance for matrix element calculations is the factorization formula [88], which describes how to factor out non-perturbative effects between the partons of a proton into parton distribution functions (PDFs). For a requested final state n, the cross-section σ can be written (approximately, valid at a sufficiently high interaction energy) as follows:

$$\sigma = \sum_{a,b} \int_0^1 dx_a dx_b \int f_a(x_a, \mu_F) f_b(x_b, \mu_F) d\sigma_{ab\to n}(\mu_F, \mu_R) \tag{2.4}$$

where a and b are two partons, x_a and x_b their respective momentum fractions in the proton, f_a and f_b their respective PDFs and $d\sigma_{ab\to n}$ the differential cross-section for the process $ab \to n$, which is being integrated over the phase space of the final-state particles.

The factorization formula shows a dependency on the factorization scale (μ_F) and renormalization scale (μ_R). Both are in principal nonphysical parameters only needed for calculation technicalities, which all observable quantities are independent of. However, when calculations are performed by cutting off the perturbation expansion at some finite order (which is in practice always needed), a residual dependency on these scales remains; in that case, repeating the calculation with varied scales gives a rough estimate of the impact of higher order effects ignored in the calculation and this procedure can be used to estimate the corresponding uncertainty. From a physical point of view, the factorization scale μ_F separates the perturbative, high-energy, short-distance effects in the matrix element from the non-perturbative, lower-energy, longer-distance effects encapsulated in the PDFs [89]. Technically, it removes collinear divergences from the initial state by absorbing these parts of the calculation into finite PDFs. The renormalization scale μ_R is introduced to absorb ultraviolet divergences into the coupling constant.

The PDFs cannot be calculated in a perturbative way and need to be provided as external experimental input (see e.g. [90, 91]) to matrix element generators. Varying the PDFs, next to varying μ_F and μ_R, is an important way of estimating the systematic uncertainty on the theoretical predictions.

Parton showering and hadronization Apart from the PDF, there are other important contributions to the production of a physical final state in a proton-proton interaction that cannot directly be included in the matrix element calculation of the hardest sub-process, either because they are too many (making the matrix element calculation too long for practical

use) or because they are non-perturbative (making the whole Feynman diagram perturbative expansion break down). Therefore, these contributions are added using heuristic models in a stage after the generation of the hardest sub-process. These effects include:

- **Parton shower** Parton showering is the process of a cascade of strong interactions (i.e. gluon emission and splitting) that originates from the primary partons in the hardest sub-process, both in the initial and final state, evolving from the high energy scale of the primary interaction down to the lower energy scale where hadronization takes place [88, 92]. Though these interactions are in principle still in the perturbative regime, phenomenological models are used in practice.
- **Hadronization** Since quarks are confined to colourless objects at lower energy scales, a process of hadronization takes place that binds coloured quarks into colourless composites. These processes are non-perturbative, and phenomenological models have to be applied. The most popular one is the 'string model' [88, 93], modeling the confinement as a linear string under tension between two quarks, which may break by the formation of a q $\bar{\text{q}}$ pair recombining with the original partons into colour singlet states. As a part of hadronization, color reconnection models are employed to form the initial strings between coloured final state partons; the physical mechanism for this is however not yet fully understood.
- **Underlying event** The remnants of the two protons that provided a parton for the hard interaction can undergo secondary interactions, called the underlying event.

Perhaps better than in a well-structured list, the various mechanisms contributing to the evolution from the hardest parton-level interaction to the physically observable final state are sketched in Fig. 2.13.

Some of these modeling aspects come with non-negligible uncertainties on the total event yield and kinematics. Uncertainty in the parton shower model can be captured by varying the renormalization scale (effectively varying α_S) for initial and final state radiation. Uncertainty in color re-connection (an important ingredient in hadronization) are estimated using dedicated simulated samples using different color re-connection models. Underlying event uncertainties are estimated with dedicated samples as well, where the relevant parameters have been varied within their experimental uncertainties [94–96].

Detector simulation The final step of the simulation converts the hadrons, leptons and photons that emerge as final-state particles from the previous step into the corresponding detector response. The response of the detector to various kinds of particles passing through its active material is simulated using the GEANT4 toolkit [98, 99]. This program is used to simulate the physical effects of particles passing through various kinds of matter, essentially a technical and detailed implementation of the general principles described in section 2.3, as well as the digitization and reading-out of the signals induced by this passage. After this step, data and simulation are handled essentially equally, with the same event reconstruction algorithms (as discussed in section 2.12) being applied to both real and simulated detector signals.

The simulation of the CMS detector response is by far the most computationally intensive step of the whole simulation chain. An alternative, much faster detector simulation approach (called FastSim, as opposed to FullSim) exists [100], based on a simple phenomenological parameterization of the detector tuned to match the FullSim output. However, accuracy is currently still an issue for this approach (despite active developments using machine learning [101]) and it has not been used in this work.

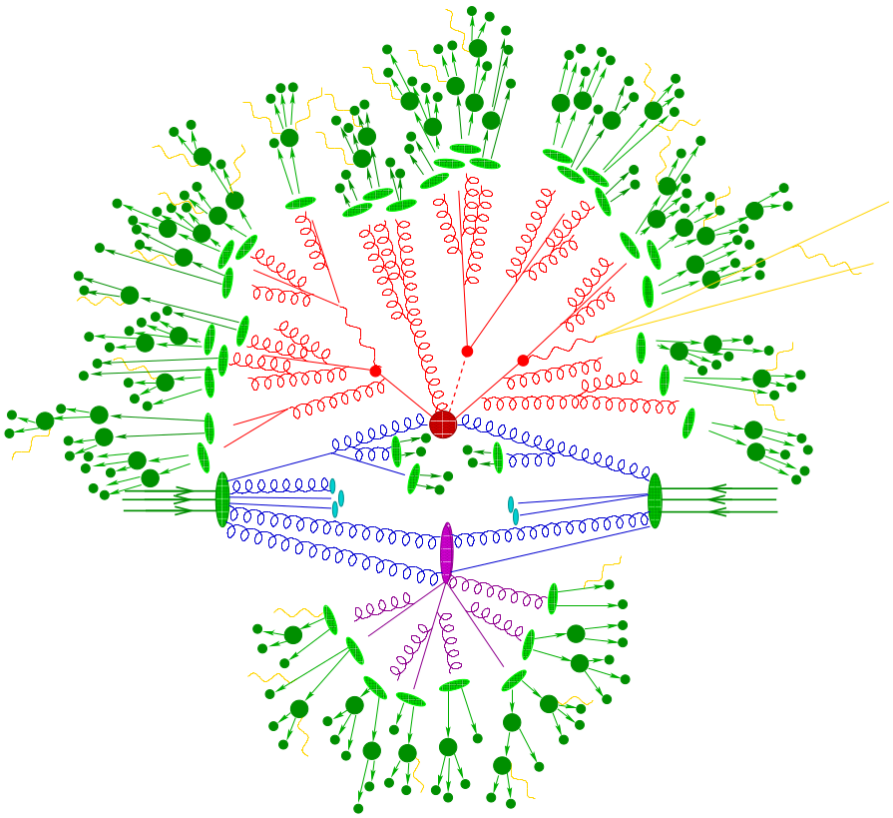


Figure 2.13: Sketch of the different steps in the simulation of a proton-proton collision. The dark red blob is the hard interaction (the contents of which are simulated with matrix element calculations). Final state parton showers and decay of heavy resonances are shown in red. Initial state radiation is also shown on the parton lines leading up to the hard interaction. Hadronization occurs in the light green blobs, followed by further decays of the produced hadrons in dark green. The lower half of the sketch also shows interactions of the remnants of the proton apart from the main interacting partons, collectively called the underlying event. Figure from Ref. [97].

Even the full GEANT4 -based simulation is known to show discrepancies with respect to observed data at a significant level. This is why simulated samples are subjected to corrections derived from calibrations to data, discussed in more detail in section 6.5.

2.14 Addendum: the RPC upgrade project

This chapter has been dealing mainly with the hardware and software of the CMS detector, which is a necessary introduction for the work that follows, but in which I have very little personal contribution. To end this chapter on a different note, I shortly mention the RPC upgrade project, and more specifically my personal involvement in it.

RPC upgrade The RPC upgrade project consists of the replacement of the readout system for the already installed RPC detectors (not further discussed here), as well as the design and installation of two new RPC stations in the endcap of CMS. As can be seen in Fig. 2.8, the current RPC wheels in the endcap do not extend all the way to the beamline, but are limited to about $|\eta| < 1.9$ at maximum. For the upgrade project, two new stations called RE3/1 and RE4/1 will be installed in the third and fourth endcap disks, extending the pseudorapidity range up to $|\eta| < 2.4$. The new RPC stations will improve the muon reconstruction by adding potential hits in the region that is currently only instrumented with CSCs. Furthermore, the trigger efficiency and background rejection will be enhanced by the good time resolution of the RPCs. The improvements of the new RPCs with respect to the old ones include [102]:

- The muon signal is extracted from both sides of the readout strips, and from the time difference the position of the passing muon in the direction of the strips can be inferred.
- The thickness of the gaps and electrodes, as well as the operating voltage, and hence the avalanche charge, are reduced. This modification increases the rate capability and detector lifetime.
- The reduction of the avalanche charge is compensated by more sensitive front-end electronics, avoiding a reduction in efficiency.

Assembly and quality control Together with other members of the Ghent University CMS group, I participated in the assembly and quality control of RE3/1 chambers to be used in the upgrade. Figure 2.14 shows some example results of a quality check on individual gas gaps (i.e. before chamber assembly). The gaps are filled with argon gas up to a specified pressure (15 mbar), the gas inlets are sealed, and the gap is left untouched for a 10 minutes stabilization + 10 minutes measuring time, during which the pressure is monitored. Large pressure drops (above 0.4 mbar / 10 mins) are indicative of gas leaks, in which case the gap is either repaired or rejected for further use. In a second part of the test, the gap is pressed on the positions of the spacers that keep the electrodes apart. Broken spacers are manifested as large pressure spikes due to the large deformation induced by the pressing. In Fig. 2.15, a similar test is shown for the two gaps in a fully assembled chamber. The gas leak test has to be repeated after assembly since leaks could be introduced in the various connecting parts between the gas inlets on the gap and the gas inlets on the outside of the chamber.

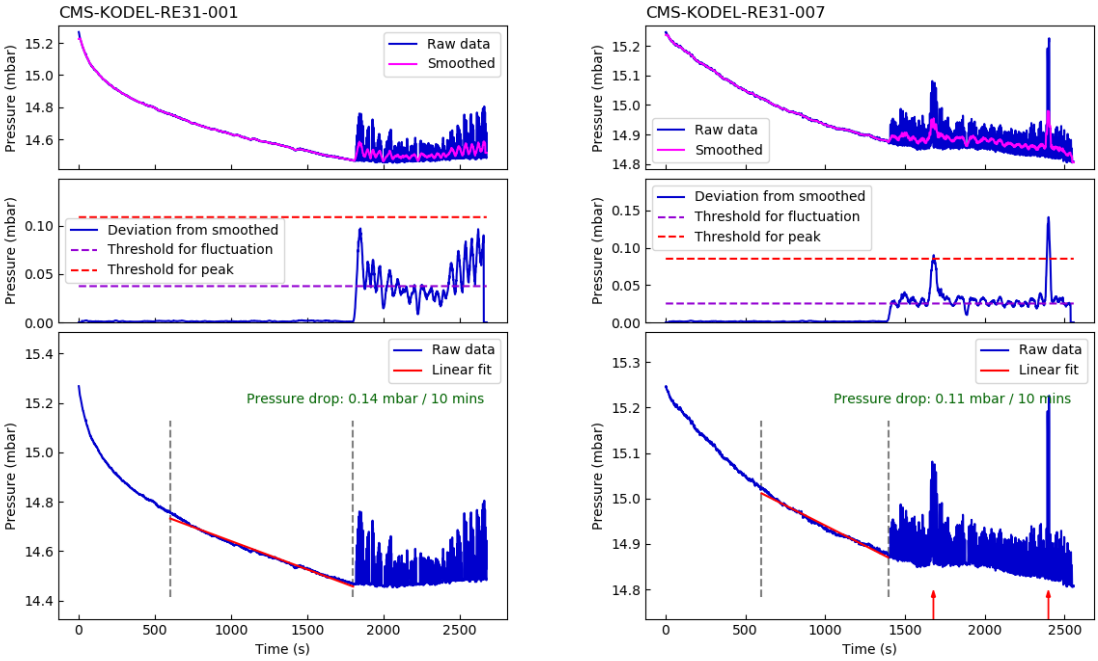


Figure 2.14: Examples of the gas leak test and spacer test for gaps to be used in RPC chambers. Left: a good gap, with acceptable pressure drop (the rejection threshold is 0.4 mbar / 10 mins) and no apparent broken spacers. Right: a rejected gap, with acceptable pressure drop but two broken spacers.

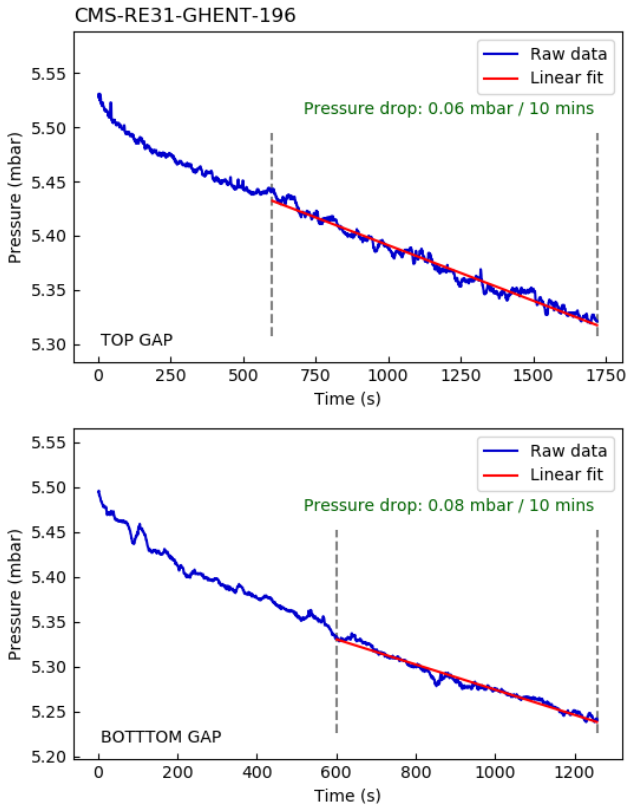


Figure 2.15: Example of the gas leak test of the two gaps in an assembled chamber. Both the top and bottom gap show an acceptable pressure drop (the rejection threshold is 0.4 mbar / 10 mins).

Chapter 3

Calibration of displaced tracking and vertexing efficiency

This chapter describes a study of displaced tracking and vertexing efficiency in the CMS detector, in the context of searches for hypothesized long-lived particles, especially heavy neutral leptons. The study has been published to the CMS Collaboration in an internal analysis note [16] and to the outside world in a detector performance note [6]. Furthermore, the results have been used in an innovative search for heavy neutral leptons in the fully leptonic channel [7] (published) and in the semi-leptonic channel [8] (currently in the final stages of CMS internal review). Moreover, the method presented here is planned to be used as a centralized tool for other analyses (e.g. Ref. [103]) and for Run III data.

3.1 Long-lived particles and displaced vertexing

Long lived particles In recent years, searches for long-lived particles (notably heavy neutral leptons) have become a main field of research of the CMS Collaboration [7, 104–118]. In the context of the CMS experiment, a particle is referred to as long-lived if its lifetime is relatively long compared to the typical strong and electroweak timescales of the primary interaction, to the extent that this particle will travel a macroscopic distance (in the order of a few centimeters up to meters) before it decays. When, in addition, this particle is itself invisible to the CMS tracker (e.g. because it is electrically neutral) but decays to visible particles, the typical experimental signature in this type of analysis consists of a secondary vertex[1] with a large transverse displacement from the center of CMS ('transverse' in this context meaning the plane orthogonal to the beamline), from which two or more tracks originate. This signature

[1] 'Secondary' with respect to the primary vertex where the initial interaction between the partons in the colliding protons takes place, defined in more detail further on.

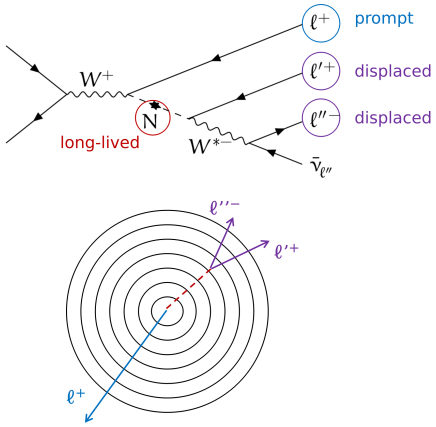


Figure 3.1: Typical Feynman diagram with a long-lived heavy neutral lepton N (above) and as sketch of the corresponding experimental signature (below). The concentric circles represent a transverse cross-section of the CMS tracker, in the center of which the protons are made to collide and the primary interactions take place. The long-lived heavy neutral lepton is not directly detectable, but decays to a pair of charged leptons (that leave a trace in the tracker) at macroscopic distances from the primary interaction vertex. Feynman diagram from Ref. [7].

is sketched in Fig. 3.1.

Tracker The innermost sub-detector of CMS is the pixel tracker, the layout of which is shown in Fig. 3.2. The barrel pixel layers are situated at 29, 68, 109 and 160 mm from the beamline (or at 44, 73 and 102 mm before the Phase-1 upgrade). This implies that, if the transverse displacement of an invisible long-lived particle is larger than a few centimeters, the tracks resulting from its decay could originate beyond the first few layers of the pixel tracker, potentially affecting the tracking and vertexing efficiency.

Displaced tracking efficiency An important ingredient in analyses searching for long-lived or displaced invisible particles (notably heavy neutral leptons) is the reconstruction of (pairs

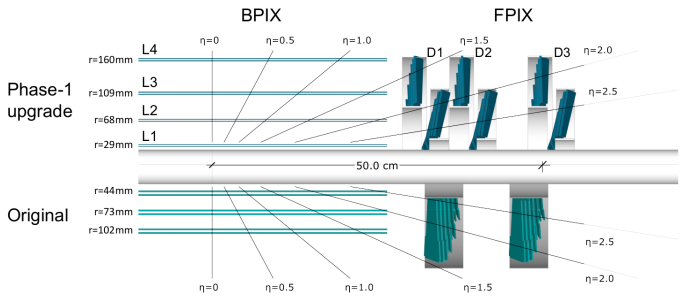


Figure 3.2: Layout of the CMS pixel tracker, the sub-detector closest to the beamline. The label "Original" (lower half of the figure) refers to the layout during data taking in 2016; "Phase-1 upgrade" (upper half) refers to the layout during 2017 and 2018 data taking, after a thorough upgrade of the pixel tracking system. Figure from Ref. [49].

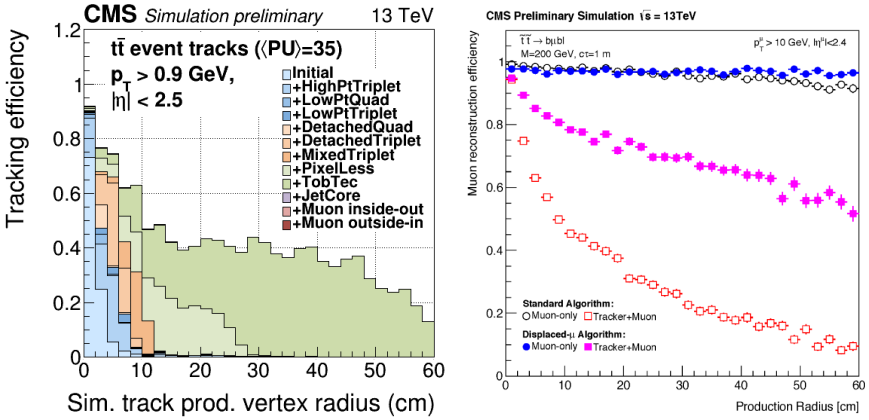


Figure 3.3: Left: reconstruction efficiency for tracks as a function of radial displacement (obtained from simulation). The various contributions, shown in different colours, represent consecutive iterations of the iterative tracking algorithm (see section 2.12.1), with progressively looser track quality requirements. The total efficiency is observed to decrease significantly with the production radius. Figure from Ref. [119]. Right: reconstruction efficiency for muons as a function of radial displacement (obtained from simulated events with supersymmetric partners of top quarks undergoing displaced decays to muons). Similar to tracks, the efficiency drops with the production radius for reconstruction algorithms relying on hits in the CMS tracker (labeled "Tracker+Muon" in the figure). Figure from Ref. [120].

of) tracks and their common vertices originating far from the beam line, primary vertex[2], and proton-proton collision point. The CMS detector and reconstruction software were not primarily designed to accurately detect these heavily displaced objects, and therefore it can be expected that the efficiency will be significantly lower than for tracks originating closer to the primary vertex. As an example, this effect can be observed in the reconstruction efficiency for both displaced tracks and muons, displayed in Fig. 3.3. In Ref. [7] (in the context of which this study was first carried out), the reconstruction efficiency for displaced muons from HNL decays drops from 85-90% at mild displacements to 40-50% for large displacements. For displaced electrons, the reconstruction efficiency is even lower, amounting to 20–40% at mild displacements and 15–20% at large displacements.

Objective of this study Developing a data-driven technique to estimate this displaced tracking and vertexing efficiency is not straightforward, as the typical tag-and-probe techniques [121] in Z boson production events are not suitable for heavily displaced objects. Therefore, one has to rely on studies using simulated signal samples to extract the track and vertex reconstruction efficiency. Yet this type of study introduces a source of uncertainty, as the agreement of the simulation to data in the case of displaced objects is not a priori guaranteed. Hence, an uncertainty factor has to be determined and taken into account, as a translation of the fact that the modeling of tracks and vertices might be less accurate at large radial distance from the primary vertex. It is the determination of these uncertainty factors that forms the major objective of this study.

[2]The primary vertex can be thought of as the place where the hard scattering between partons in the colliding protons occurs. Technically, it is defined as the vertex corresponding to the hardest scattering in the event, evaluated using tracking information alone, as described in section 9.4.1 of Ref. [78].

3.2 The probing particles

To investigate the magnitude and possible kinematic dependencies of the uncertainty factor for displaced tracking and vertexing, we look at the decays of well-known standard-model (SM) particles that give approximately the same signature as the more exotic beyond-standard-model (BSM) vertices that are typically searched for in the corresponding analyses. Similar methods have been used before in searches for long-lived heavy neutral leptons or other exotic particles at ATLAS [122–124], where the resulting uncertainty factor proved to be the dominant one of all those investigated. Hence, this source of uncertainty is non-negligible and should be estimated carefully by analyses relying on displaced tracking and vertexing.

Two processes have been used as a probe for displaced tracking and vertexing:

- The decay of the neutral K meson (K_S^0) to two pions:

$$K_S^0 \to \pi^+\pi^-$$

- The decay of the neutral lambda baryon (Λ^0) to a proton and a pion:

$$\Lambda^0 \to \mathrm{p}\pi^-$$

$$\overline{\Lambda^0} \to \overline{\mathrm{p}}\pi^+$$

The Feynman diagrams corresponding to these decay processes are displayed in Fig. 3.4. The K_S^0 meson[3] is a weak eigenstate corresponding to a superposition of the strong eigenstates $K^0(\mathrm{d}\bar{\mathrm{s}})$ and $\overline{K^0}(\bar{\mathrm{d}}\mathrm{s})$. The decays of both K^0 and $\overline{K^0}$ to a positively and a negatively charged pion are completely equivalent from the detector point of view, and in the remainder of the text we will not distinguish between both. Similarly, when referring to the Λ^0 baryon or its decay to a proton and a negatively charge pion, the conjugate $\overline{\Lambda^0}$ and its decay to an antiproton and a positively charged pion is implicitly included as well.

The K_S^0 and Λ^0 particles have suitable lifetimes to decay within the volume of the silicon tracker of CMS, at macroscopic distances from the collision point, as can be estimated as follows. From the mass shell formula:

$$E = \sqrt{m^2 + p^2} = \gamma m$$

(with E the particle's energy, m its rest mass, p its momentum and $\gamma = \frac{1}{\sqrt{1-v^2}}$, all in natural units), it follows that the speed of the particle v can be calculated as:

$$v = \frac{1}{\gamma}\sqrt{\gamma^2 - 1} = \frac{1}{\gamma}\frac{p}{m}$$

Therefore, the typical decay distance r is given by

$$r = vt = v\gamma\tau = \frac{p}{m}\tau$$

where τ is the proper lifetime of the particle. If we substitute the average transverse momentum p_T for p, we get an estimate of the average transverse decay distance (called here Δ_2D in agreement with [7]). The results are shown in Tab. 3.1.

[3] Colloquially called "K-short" meson, in distinction to the K_L^0 or "K-long" meson, which consists of the same quarks but has different symmetry properties and a much longer lifetime.

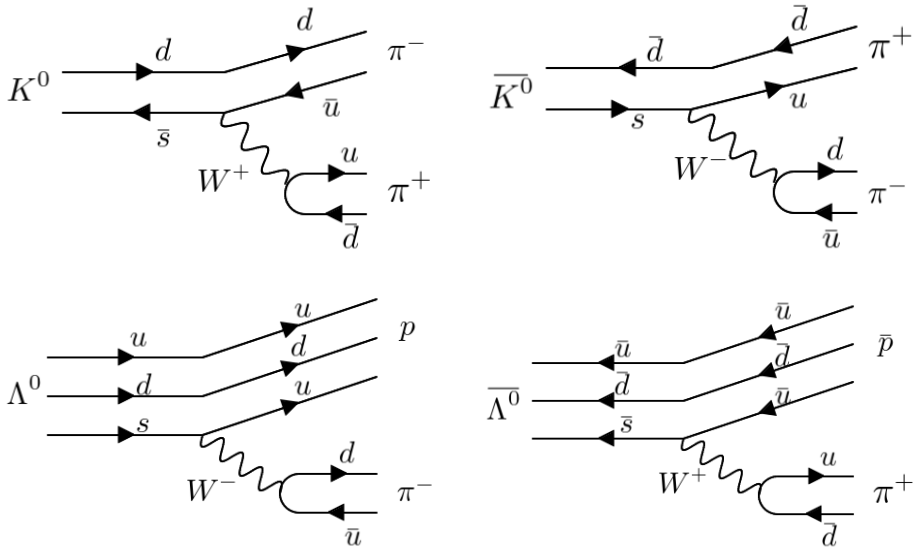


Figure 3.4: Diagrams for the decays of K^0_S (above), Λ^0 and $\overline{\Lambda^0}$ (below) that are used in this study.

Table 3.1: Estimate of the average radial decay distance for K^0_S and Λ^0 particles based on their proper lifetime, mass and kinematic properties. The lifetime and mass were taken from the PDG references [125] (for K^0_S) and [126] (for Λ^0). Average transverse momenta were estimated from the data, see e.g. Fig. 3.8.

Particle	Lifetime (s)	Mass (GeV)	p_{T} (GeV)	Δ_{2D} (cm)
K^0_S	8.95e-11	0.498	3	16
Λ^0	2.63e-10	1.116	5	35

3.3 The method in a nutshell

Essentially, this study consists of comparing the number of reconstructed K^0_S and Λ^0 hadrons between data and simulation, as a function of the radial displacement (Δ_{2D}) of the reconstructed vertices. In this way, differences between data and simulation in terms of displaced tracking and vertexing show up as systematic discrepancies at relatively large radial displacement.

Event selection The K^0_S and Λ^0 hadrons are reconstructed from pairs of tracks in additional jets and hadronic activity in the underlying event of a sample of selected Drell-Yan events with dimuon decays (i.e. $Z \to \mu^+\mu^-$ events). The main reason for looking at K^0_S and Λ^0 decays specifically within dimuon Drell-Yan events is that this process is simple, theoretically well-understood and calculable, and experimentally rather clean. Moreover, by triggering on the dimuon signature, the trigger is unbiased with respect to the presence of K^0_S and Λ^0 hadrons, removing some potential systematic biases that are not of interest for this study. Finally, focusing on this particular process, which has no significant backgrounds in its relevant phase space, allows to use a single sample of simulated events (as opposed to a collection of several samples, when the background processes are not negligible). This simplifies the analysis workflow.

3.4 Data and simulated samples

This analysis uses a set of proton-proton collision events delivered by the LHC at a center-of-mass energy of 13 TeV and collected by the CMS detector. We use the full Run II data set, consisting of data collected in 2016 (36.3 fb^{-1}), 2017 (41.5 fb^{-1}) and 2018 (59.7 fb^{-1}), adding up to a total integrated luminosity of 138 fb^{-1} . Details on the luminosity measurement and calibration for the three data taking years can be found in [127–129].

APV saturation effect Of special interest in this data sample is the APV saturation effect [130–132]. This dynamic decrease in tracking efficiency in early 2016 data was caused by a saturation effect in the APV readout chips in the strip tracker electronics[4]. Around the middle of 2016 data taking, the APV settings (in particular: the preamplifier feedback voltage bias) were reconfigured to increase the readout speed and remove the saturation. The APV saturation effect was not included in the initial simulation and will be shown to have a strong impact on the displaced tracking and vertexing calibration.

Legacy reprocessing During the long shutdown between Run II and Run III, CMS undertook a reprocessing of all Run II data and simulation (the so-called "legacy" campaign [133, 134]) with the most accurate values for detector alignments, calibrations and dynamic inefficiencies, either to be mitigated in data reconstruction, or correctly taken into account in the simulation. In particular, the APV saturation effect was incorporated in the simulation (in an approximate way) in order to reduce the discrepancy between data and simulation because of that issue [132, 134]. Data and simulation from before the legacy campaign are referred to as pre-legacy[5]. This study was originally carried out with pre-legacy data, as the legacy reprocessing was not yet available at that time. However, the study was later repeated with legacy data, and the results between both reconstruction versions are compared. The pre-legacy and legacy

[4] When the issue was first noticed, its cause was unknown, and it was hypothesized that the saturation was caused by a heavy ionizing particle (HIP). Therefore, this effect is sometimes called the HIP effect in CMS jargon, but the term is not accurate and will be avoided in this work.

[5] In CMS jargon, "legacy" is often referred to as "ultra-legacy (UL)", and "pre-legacy" as "pre-UL", but we will avoid those terms in this work

version of the method are almost identical, with only minor differences in event the selection, as detailed in section 3.5.

Pre-legacy data and simulation As this study targets the Drell-Yan process in the dimuon channel, we use a primary dataset that is collected from dimuon triggers. The data reconstruction version dates from 2018[6], and was the most recent and recommended version before the legacy reprocessing. Suitable simulated Drell-Yan samples were selected in correspondence with the data samples We used centrally provided standard samples with the MadGraph5_aMC@NLO generator at next-to-leading order [135] and FxFx merging [136] for the matrix element calculations, and pythia [93] for the parton showering. The CUETP8M1 tune [95, 96] is used for the 2016 sample, while the CP5 tune [94] is used for the 2017 and 2018 samples. The interaction of the generated particles with the detector material and the resulting detector response signal is simulated with Geant4 [98].

Legacy data and simulation The data samples used in the legacy version of this study are the same as in the pre-legacy version, but in the legacy reconstruction version[7]. The generation of simulated samples happens identically to the pre-legacy simulation, except for the updated usage of the CP5 underlying event tune instead of the CUETP8M1 tune in 2016 simulation. The main difference with the pre-legacy samples is in the updated detector simulation. The APV saturation effect splits the 2016 data and simulation in a part with old APV settings, where the APV saturation effect was present (and mitigated in legacy reprocessing), and a part with new APV settings, where it was not.

Triggers In correspondence with the targeted experimental signature, a cocktail of several dimuon triggers is used, chained in a logical OR, i.e. the event is kept for further analysis if it triggers at least one of all dimuon triggers. They are listed in Tab. B.2 (in the appendix).

3.5 Event selection

The event selection strategy is aimed at $Z \to \mu^+\mu^-$ events. The idea is to look for K_S^0 and Λ^0 particles in the jets that are produced next to the primary process of Z boson production and decay, while the latter can be used for normalization. The events must satisfy the requirements listed below:

- The number of muons in the event satisfying a standardized muon identification is required to be exactly 2. The "medium" muon ID is used, as defined by the Muon POG [69], providing a good trade-off between selection efficiency and fake rate[8].
- The leading muon must have a transverse momentum of at least 30 GeV, the trailing muon of at least 25 GeV.
- The selected muon invariant mass is required to be within 10 GeV of the Z boson mass.

[6] In CMS technical jargon: the primary dataset is called DoubleMuon, and the reconstruction versions are ReReco_17Jul2018 for 2016 (B-H) data, ReReco_31Mar2018 for 2017 (B-F) data, ReReco_17Sep2018 for 2018 (A-C) data, and PromptReco for 2018 (D) data, where A-H refer to chronological data-taking periods within a single year.

[7] In CMS technical jargon: the reconstruction versions are: HIPM_UL2016 for early 2016 data (B-F) with the APV saturation effect present, UL_2016 for the second half of 2016 data (F-H), UL_2017 for 2017 (B-F) data, and UL_2018 for 2018 (A-D) data, where A-H refer to chronological data-taking periods within a single year. The fixing of the APV saturation effect took place towards the end of 2016 F, hence most of it is included in early 2016 and the rest in late 2016.

[8] A technical description of the identification requirements can be found *here* (digital link) (CMS internal).

- Events containing one or more loosely tagged b jets are vetoed. A loose b-tagged jet is defined by the BTV POG [85] as a jet of which the DeepCSV values for single and double b tagging add up to a threshold that is tuned per data-taking year[9].

In order to check the quality of the event selection, we plot the data to simulation ratio for a couple of control variables that are independent from the K_S^0 and Λ^0 yields. In Fig. 3.5, we show the transverse momentum and the pseudorapidity of the muons that pass the event selection and the invariant mass of the dimuon system. The agreement is found to be overall satisfactory. However, we want to factor out any remaining discrepancy of this kind to focus only on the potential displaced track and vertex reconstruction mismodelling. Rather than fine-tuning the data-to-simulation agreement by refining the event selection or applying correction factors, we choose to follow a more in-situ and robust approach by applying an additional normalization (see section 3.7).

As a cross check, an alternative event selection targeting the $\mathrm{Z} \to e^+e^-$ process was implemented, with the appropriately modified data sets, triggers and event selection criteria. No systematic differences were observed between both event selections, and for ease of presentation we restrict ourselves to the $\mathrm{Z} \to \mu^+\mu^-$ case.

3.6 Vertex reconstruction

The vertex reconstruction algorithm developed for this study resembles the built-in V^0-vertex fitter in the CMS software[10] (where V^0 is used to refer to K_S^0 and Λ^0 simultaneously), with however some alterations to account for the fact that we are working on a reduced data format[11]. Furthermore, we re-optimized some of the cut values in order to retain more V^0 candidates at small radial displacements from the primary vertex, at the cost of a slightly larger overall background. This design choice was inspired by the need for a large number of candidates close to the primary vertex, as these will be used for normalization (see section 3.7).

The vertex reconstruction starts from all tracks in the event. The rationale is to apply three-tiered selection system, sketched in Fig. 3.6, first on individual tracks, next on pairs of tracks, and finally on fitted vertices. This structuring helps to increase the computing efficiency, e.g. by performing the computationally intensive vertex fitting only on pairs of high-quality tracks that originate close to each other and have opposite charge.

[9] The numerical values can be found *here* (digital link) (CMS internal). The values are different for the pre-legacy and the legacy processing, as the working points were re-optimized on the legacy simulation.

[10] A description of the algorithm can be found *here* (digital link) (CMS internal). The implementation can be retrieved from *here* (digital link).

[11] In CMS technical jargon: the original V^0-vertex fitter was designed for the RECO data tier, while the analysis presented here runs on MINIAOD. The RECO level contains more low-level and detailed information, but is impractically large for most analyses. The MINIAOD level has a strongly reduced size per event, at the cost of the loss of some information.

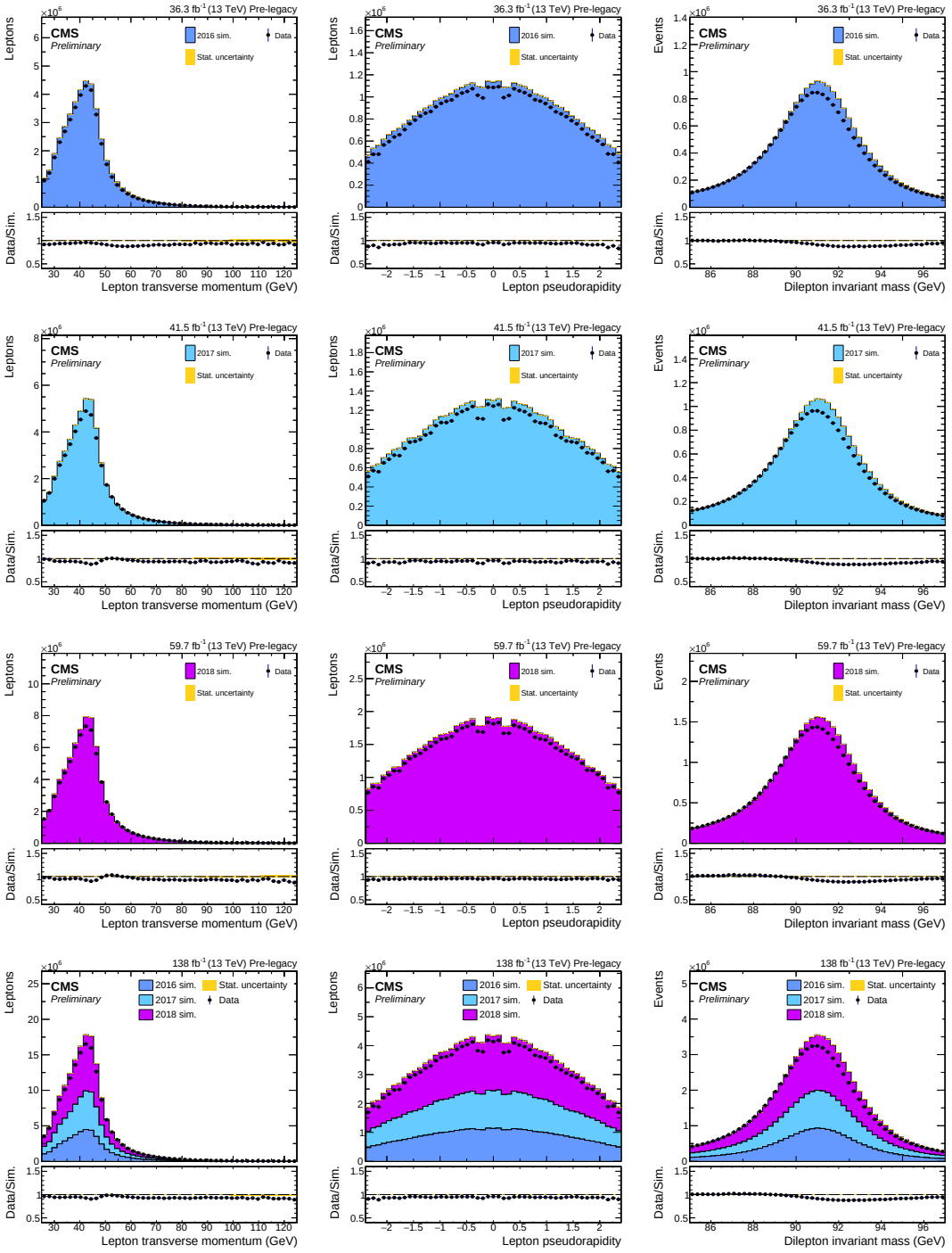


Figure 3.5: Data to simulation ratio plots for three control variables: the muon transverse momentum (left column), the muon pseudorapidity (middle column), and the dimuon invariant mass (right column). The rows correspond to different data-taking years: 2016 (first), 2017 (second), 2018 (third) and combined for Run II (fourth). The figures shown here correspond to the pre-legacy version of this study, with very similar results obtained for the legacy version (not shown here).

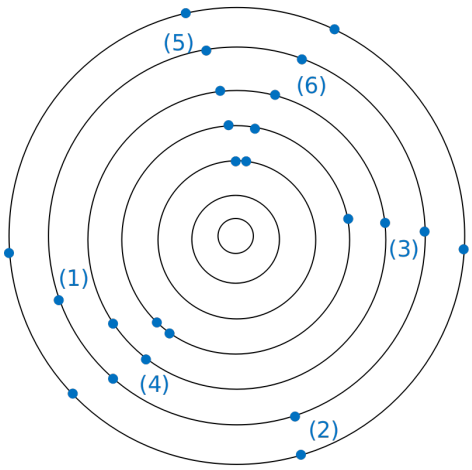


Figure 3.6: Schematic representation of the tiered vertex reconstruction and selection approach. Step 1: filter individual tracks (e.g. remove (1) because of too low p_{T} and (2) because of too few hits). Step 2: filter all pairs of remaining tracks (e.g. remove (3) and (4) since they do not form a pair). Step 3: fit and select a vertex to remaining pairs of tracks (e.g. to (5) and (6)).

Individual tracks The first step in the vertex reconstruction consists of grouping all tracks available in the event information. These consist of the tracks belonging to the collection of particle-flow candidate tracks, and additionally a collection of high-purity tracks that were not associated to any particle-flow candidate[12]. In order to be taken into account for further processing, the tracks need to fulfill a set of pre-selection conditions:

- The track must have the quality label called "loose" as defined by the Tracking POG [59]. This selection imposes relatively mild constraints on the minimum number of hits, the maximum number of missing expected hits, the normalized χ^2 of the track fit, and track the impact parameters.
- The normalized χ^2 of the track fit must be smaller than 5.
- At least 6 valid hits must be associated to the track.
- The transverse momentum of the track must be larger than 1 GeV.
- The transverse impact parameter significance of the track must be larger than 2.

The requirement involving the transverse momentum of the track is not present in the original V^0-fitter. It was implemented for this study to more closely resemble the object selection typically used in HNL or other long-lived particle searches, at the cost of reducing the amount of K_S^0 and Λ^0 candidates passing the final selection. For K_S^0 mesons, this is not a problem as a sufficient number of candidates remain; the method with Λ^0 baryons will be shown to suffer from other problems.

Pairs of tracks In the next step, all combinations of selected tracks are considered if the following conditions are met:

- The tracks must have opposite charge.

[12] In CMS technical jargon, these collections are called PACKEDPFCANDIDATES and LOSTTRACKS respectively.

- The distance of closest approach of both tracks with respect to each other must be smaller than 0.2 cm.
- The point of closest approach must be situated within a cylinder of radius 120 cm and length 300 cm around the nominal center of the detector.

Vertices A Kalman vertex fitter [137] is applied on the selected combinations of tracks. Before assigning the vertex to a specific category, some additional requirements need to be satisfied:

- The normalized χ^2 of the vertex fit must be smaller than 7.
- The cosine of the pointing angle must be larger than 0.99, where the pointing angle is defined as the angle between the transverse position vector of the vertex and the combined transverse momentum vector of the tracks (see Fig. 3.7).

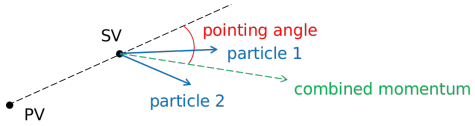


Figure 3.7: Definition of the pointing angle.

- The invariant mass of the track pair is calculated under different hypotheses ($\pi^+\pi^-$, $\mathrm{p}\pi^-$ and $\overline{\mathrm{p}}\pi^+$). The result is required to be within 70 MeV of the K^0_S mass (using the $\pi^+\pi^-$ assumption) or within 50 MeV of the Λ^0 mass (using the $\mathrm{p}\pi^-$ or $\overline{\mathrm{p}}\pi^+$ assumption).

If a vertex passes these requirements, it is considered to be a K^0_S-candidate if the $\pi^+\pi^-$-invariant mass is closer to the K^0_S mass than the $\mathrm{p}\pi^-$-invariant mass is to the Λ^0 mass and the $\overline{\mathrm{p}}\pi^+$-invariant mass to the $\overline{\Lambda^0}$ mass, and vice versa for a Λ^0 or a $\overline{\Lambda^0}$ candidate.

Optimization Two important differences with respect to the built-in V^0 fitter should be noted here. First, a requirement involving the inner hit position is replaced by a requirement on the pointing angle, since the former information is not available the data and simulation used for this study[13]. Secondly, a requirement on the transverse distance divided by its uncertainty was left out. This leads to a higher combinatorial background but is necessitated by the demand to retain a sufficient amount of candidates at small radial distances.

Validation The transverse momentum, pseudorapidity and invariant mass of the reconstructed K^0_S and Λ^0 candidates resulting from this procedure are displayed in Fig. 3.8. From this figure, the following observations can be made:

- Typical p_T values of K^0_S and Λ^0 hadrons reconstructed in this way are much lower than those of exotic resonances typically targeted by searches. This difference in p_T spectrum forms a drawback of this study as estimator of displaced tracking and vertexing efficiencies. We revisit this point in section 3.9.
- A discrepancy in η spectrum can be observed for K^0_S candidates between 2016 on the one hand, and 2017/2018 on the other. This is a result of the APV saturation effect and will be revisited in section 3.8.

[13] In CMS technical jargon: the information on the inner hit position is only available up to the RECO data tier, not in the MINIAOD data tier used in this study.

- The remaining combinatorial background is small, especially for K_S^0 candidates, but not negligible, particularly for Λ^0 candidates. Moreover, the relative background level depends on the radial distance of the reconstructed vertex, as shown in section 3.7. Therefore, a background subtraction procedure is still needed.

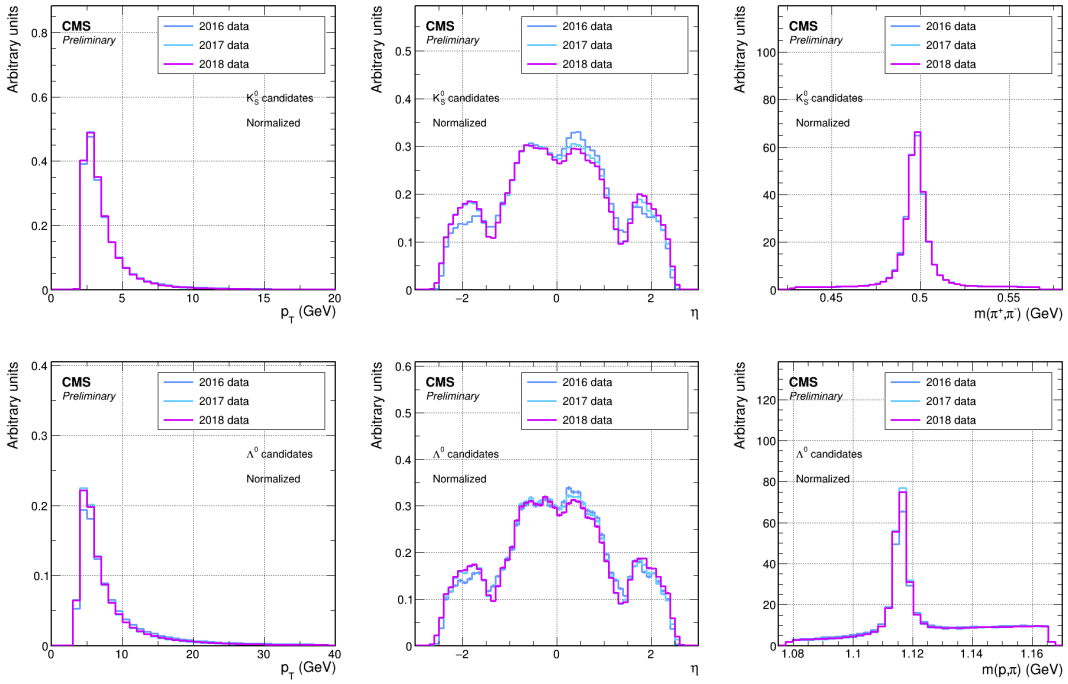


Figure 3.8: Reconstructed kinematic variables of the K_S^0 candidates (upper row) and Λ^0 candidates (lower row): transverse momentum (left column), pseudorapidity (middle column), and invariant mass (right column). All distributions were normalized to unit surface area. The figures shown here correspond to the pre-legacy version of this study, with very similar results obtained for the legacy version (not shown here).

To further validate our vertex reconstruction and selection procedure, we plot the reconstructed invariant mass of the K_S^0 and Λ^0 particles and apply a fit to the spectrum. The result is shown in Fig. 3.9. Note that the fitted function is a polynomial for the background, superimposed with a double gaussian peak with a common mean for the signal (which is often used to represent a Breit-Wigner mass peak convoluted with detector effects), i.e.:

$$f(x) = \sum_{n=0}^{N} a_n x^n + A_1 \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma_1}\right)^2\right) + A_2 \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma_2}\right)^2\right)$$

where N is the degree of the polynomial. This approach is based on [138, 139], where similar plots for K_S^0 and Λ^0 invariant masses can be found. We fix the degree of the polynomial to 1 (i.e. a straight line), as no improvement in the fit quality is observed for higher orders. The fitted K_S^0 and Λ^0 mass are in agreement with the respective PDG masses [125, 126].

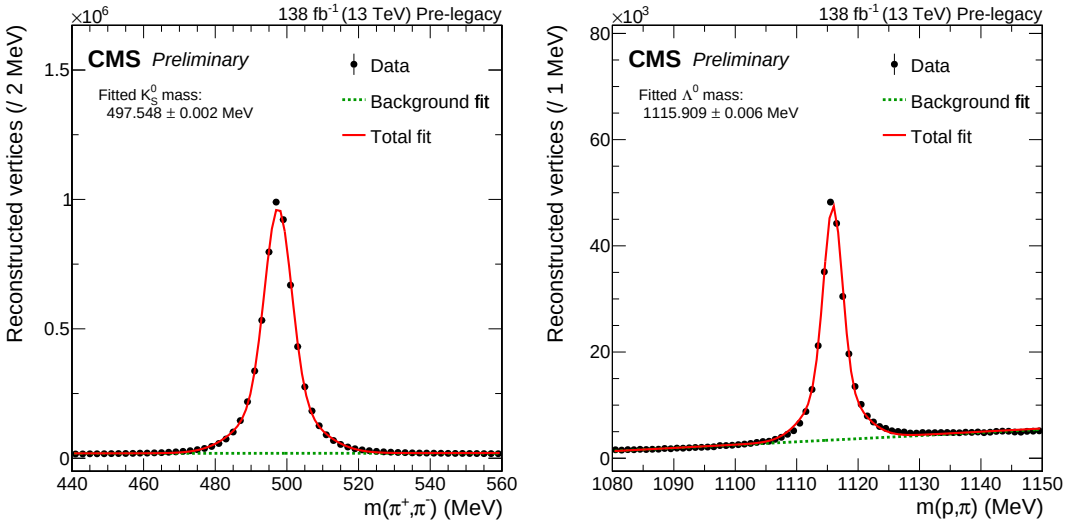


Figure 3.9: Invariant mass distribution of K_S^0 candidates (left) and Λ^0 candidates (right) with a fit consisting of a linear background contribution and a double gaussian signal contribution. The figures show the full Run II data and correspond to the pre-legacy version of this study, with very similar results obtained for the legacy version (not shown here).

3.7 Background subtraction and normalization

As can be seen in Figs. 3.8 and 3.9, the combinatorial background, consisting of random combinations of spatially close-by tracks that accidentally mimic a V^0 candidate, is small but non-negligible (especially for Λ^0 candidates). Furthermore, the relative magnitude of the background depends strongly on the radial displacement Δ_{2D}, since the density of originating tracks is much higher closer to the proton-proton interaction point. Several methods have been attempted to subtract the background instances from the collection of K_S^0 and Λ^0 candidates.

Background subtraction The most straightforward approach consists of selecting candidates that pass a tighter mass criterion than the one used to reconstruct the candidates in the first stage. K_S^0 candidates are required to have a mass within 10 MeV of the K_S^0 mass, Λ^0 candidates within 5 MeV of the corresponding mass. This approach poses a difficult compromise between keeping a sufficient number of genuine candidates and enhancing the relative purity, and is not used further. A more involved approach consists of fitting a first order polynomial to the background (using only the sidebands to the invariant mass peak) per bin in radial displacement, subtracting it from the peak and count the remaining number of events within it. This approach is used in what follows, and illustrated in Figs. 3.10 and 3.11.

Normalization Next, the simulation needs to be normalized in some sense to match the data. In the first normalization strategy, the normalization factor is chosen in such a way as to make the total number of events (not particle yields) passing the selection criteria in simulation and data equal. This method has an important drawback however: it is sensitive to overall K_S^0 or Λ^0 mismodelling, an effect which we would like to factor out of the comparison. A non-negligible discrepancy in yields of K_S^0 and Λ^0 particles has been reported before [138]. Assuming that this potential mismodelling is more or less independent of the radial decay distance, one can opt to directly normalize the number of K_S^0 in simulation to data, but taking into account only those instances at small displacements. This way, one factors out most of the overall discrepancies and what remains is the effect that we are trying to assess in this study:

a systematic discrepancy between simulation and data as a function of the radial position of the vertex.

Discussion Before turning to quoting results for all data taking eras, we will highlight the importance of these analysis design choices. In Fig. 3.10, we show several different plots for the yields of K_S^0 particles in data and simulation with different choices of normalization and background subtraction. The key points from this figure are the following:

- Performing the background subtraction is necessary, despite the overall small background level, since it is relatively much more present at low displacement.
- As anticipated, some overall mismodelling in the number of K_S^0 particles might be present, as we see a systematic disagreement between data and simulation at low displacement values, even when normalizing the total number of events in simulation to that in data.
- With appropriate handling of background and normalization, the number of reconstructed K_S^0 vertices in data matches that in simulation at low displacements, but deteriorates slightly with respect to simulation at higher displacement values, indicating that the displaced tracking and vertexing efficiency is slightly overestimated in simulation. This will be expanded upon in section 3.8.

The same plots are shown for Λ^0 particles in Fig. 3.11. The conclusions are overall similar, yet the relative background levels at low displacement are much higher for Λ^0 candidates than for K_S^0 candidates, and overall signal statistics are lower. Since the low displacement bins are needed for accurate normalization, we conclude that the Λ^0 particle is less suited as a probing particle for this method than the K_S^0 particle and will consider the latter one only. Potentially, the Λ^0 particle can be recovered as a suitable probe with a re-optimization of the vertex reconstruction and selection procedure that focuses on reducing the background at low displacement. An additional structure is visible in the upper left plot of Fig. 3.11, which can be correlated with the positioning of the layers that constitute the pixel tracker of CMS, as detailed in section 3.8.

3.8 Results and discussion

In this section, we show the obtained data-to-simulation ratios and discuss the observations, in particular regarding the APV saturation effect, the legacy reprocessing, and the impact of the pixel detector structure.

Pre-legacy In Fig. 3.12, we consider the agreement between data and simulation for displaced K_S^0 vertices, for the pre-legacy reconstruction of the combined Run II data (lower right) as well as split per data-taking year (other panels). A difference in data-to-simulation agreement can be distinguished between 2016 on the one hand, and 2017/2018 on the other hand, which gets averaged out in the combined Run II plot. The 2016 data shows a clear loss in displaced tracking and vertexing efficiency that is not present in the simulation, nor in 2017/2018 data and simulation.

Summarizing the distributions seen in Fig. 3.12, we observe an agreement between data and simulation at large displacements within about 40% for 2016 data, 15% for 2017 data, and 10% for 2018 data. This averages out in the Run II results to about 10-15%.

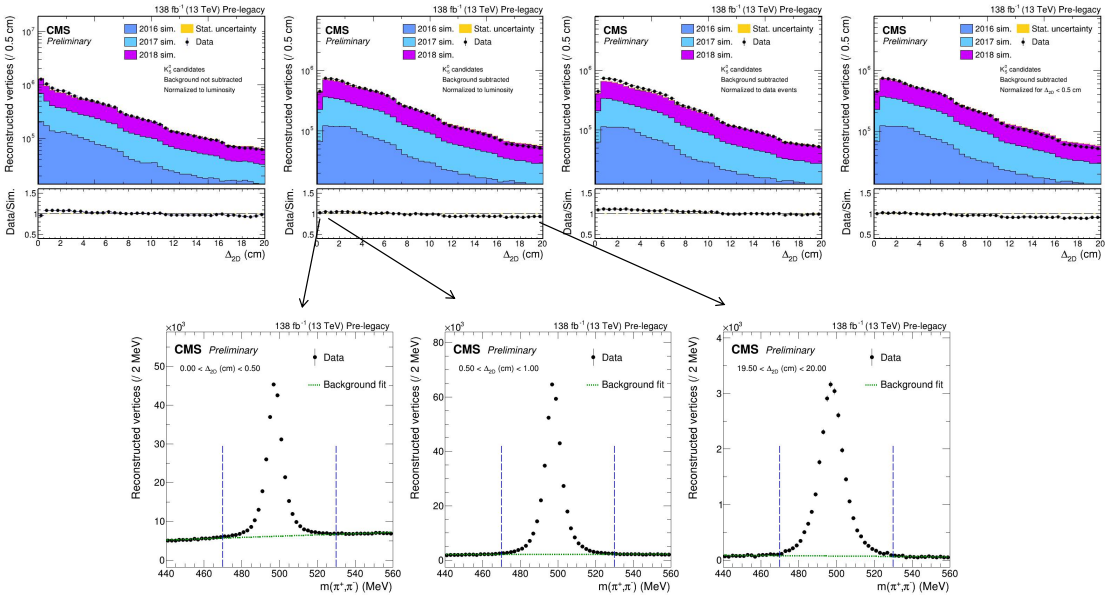


Figure 3.10: Background subtraction and normalization considerations for K^0_S candidates. The upper row shows several choices for background subtraction and normalization. From left to right: (1) no background subtraction, indirect normalization to the luminosity; (2) background subtraction using a linear fit to the invariant mass sidebands; (3) normalization to the number of events in data; (4) direct normalization in the range 0 - 0.5 cm. The lower row shows some exemplary invariant mass distributions of K^0_S candidates, per bin in radial displacement, showing the relative amount of combinatorial background. The background is subtracted by performing a linear fit to the sidebands of the invariant mass peak (shown by blue dashed lines) and subtracting this contribution from the distribution. The figures shown here correspond to the pre-legacy version of this study, with very similar results obtained for the legacy version (not shown here).

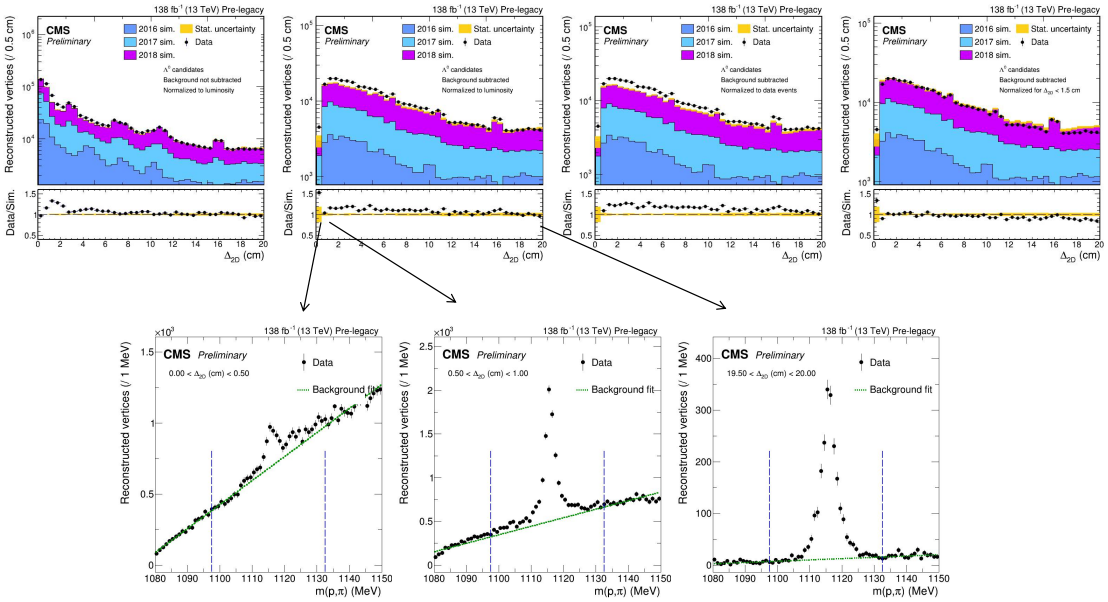


Figure 3.11: Background subtraction and normalization considerations for Λ^0 candidates. The upper row shows several choices for background subtraction and normalization. From left to right: (1) no background subtraction, indirect normalization to the luminosity; (2) background subtraction using a linear fit to the invariant mass sidebands; (3) normalization to the number of events in data; (4) direct normalization in the range 0 - 1.5 cm. The lower row shows some exemplary invariant mass distributions of Λ^0 candidates, per bin in radial displacement, showing the relative amount of combinatorial background. The background is subtracted by performing a linear fit to the sidebands of the invariant mass peak (shown by blue dashed lines) and subtracting this contribution from the distribution. The figures shown here correspond to the pre-legacy version of this study, with very similar results obtained for the legacy version (not shown here).

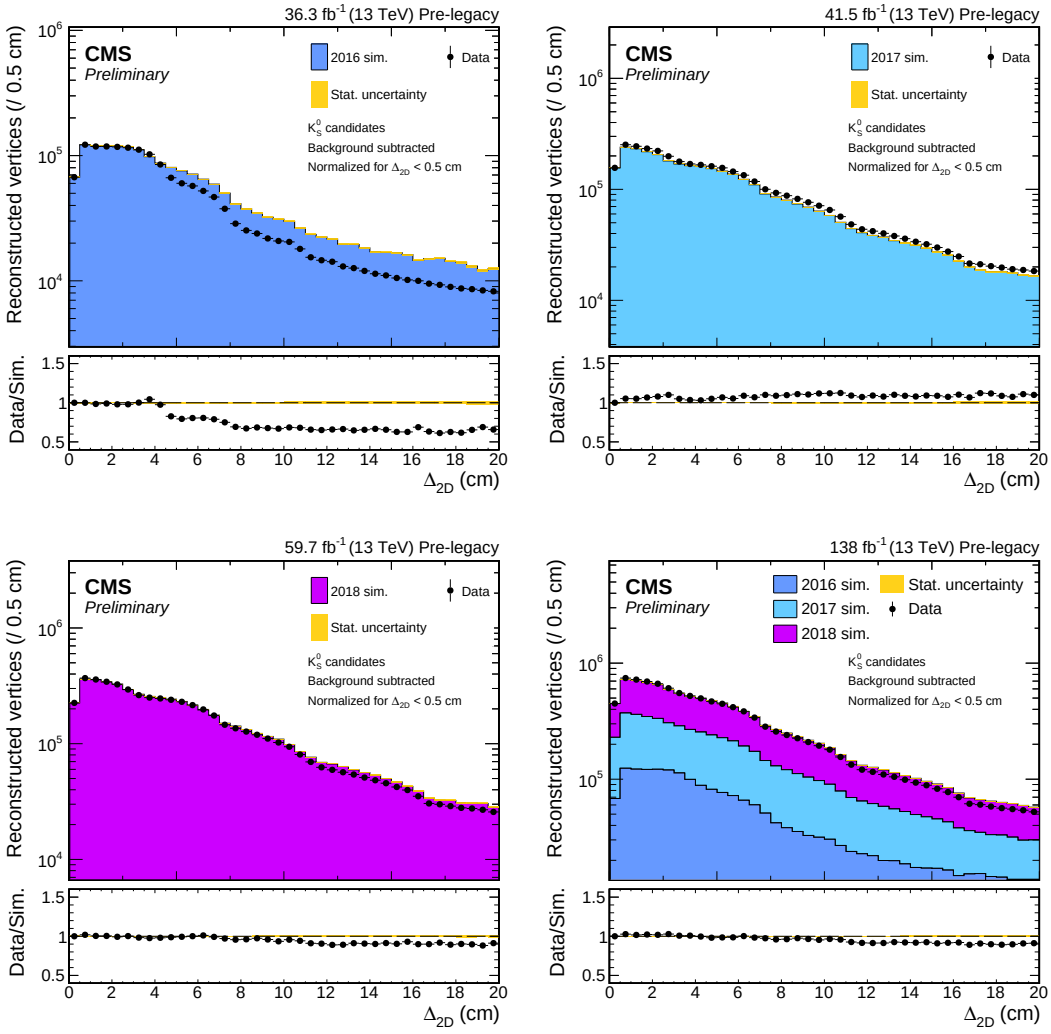


Figure 3.12: Data-to-simulation agreement for displaced tracking and vertexing of K_S^0 mesons, for the pre-legacy reconstruction of the combined Run II data (lower right) and split per data-taking year (other panels).

The observed loss off tracking and vertexing efficiency in 2016 is an average effect, that gets enhanced when splitting between data-taking periods, in particular between the period with old APV settings (where the APV saturation effect was present) and the period with new APV settings. The reconfiguration of the APV chips is visible in the form of a much better data-to-simulation agreement in the corresponding data-taking period in Fig. 3.13.

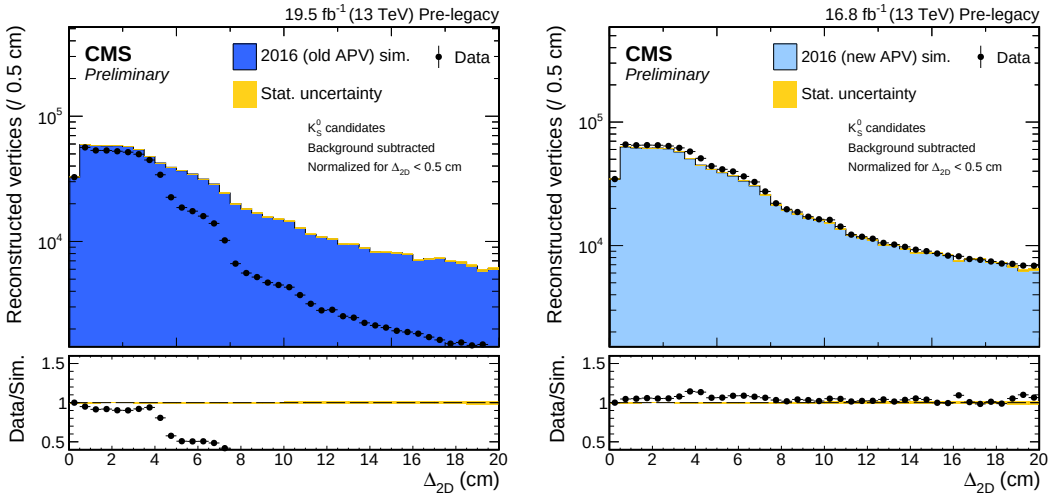


Figure 3.13: Data-to-simulation agreement for displaced tracking and vertexing of K_S^0 mesons, for the pre-legacy reconstruction of 2016 data, split between the data-taking period with old APV settings (left) and new APV settings (right).

Legacy Analogous results for the legacy reprocessing are shown in Fig. 3.14, for the combined Run II data (lower right) and split per data-taking period (other panels)[14]. Summarizing these distributions, we observe an agreement between data and simulation at large displacements within about 20-25% for 2016 data, 10% for 2017 data, and 10% for 2018 data. This averages out in the Run II results to about 10-15%.

The APV saturation effect in the early 2016 data is much less pronounced in the legacy reprocessing with respect to the pre-legacy version, as can be seen from comparing Fig. 3.14 with Fig. 3.12. However, the APV effect mitigation strategy does reduce, but not completely resolve, the data-to-simulation discrepancy and a significant effect is still observed even in legacy data and simulation.

Detector structure As a side result, we notice how the structure of the CMS pixel tracker is discernible in the distributions of K_S^0 vertices as a function of radial displacement. This is highlighted in Fig. 3.15. The difference between 2016 and 2017/2018 is caused by the upgrade of the pixel detector in between the two data-taking periods, as introduced in section 3.1.

Uncertainty In extension of the previous paragraph, we also investigate the uncertainty on the radial distance measurement, mainly coming from uncertainties in the measured track parameters propagated to a position uncertainty of the fitted secondary vertex. The magnitude of this uncertainty and its correlation with the radial displacement and the structure of the pixel detector are shown in Fig. 3.16 and 3.17 for 2016 (new APV) and 2018 data respectively.

[14]The luminosity value for 2018 data taking is slightly larger in the legacy reprocessing as compared to pre-legacy, because of the recovery of previously excluded data. The impacts on the results is however negligible.

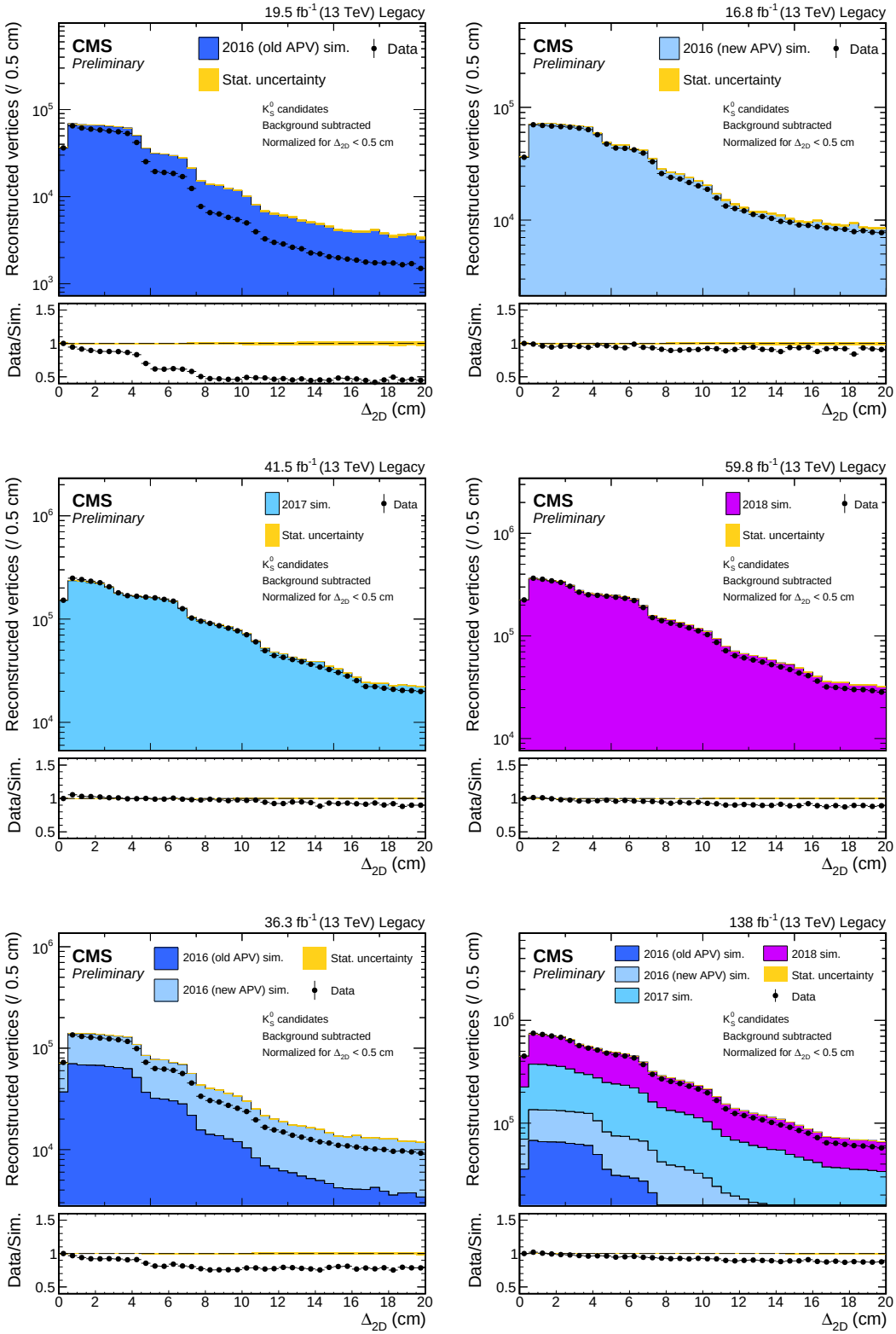


Figure 3.14: Data-to-simulation agreement for displaced tracking and vertexing of K_S^0 mesons, for the legacy reprocessing of the combined Run II data (lower right) and split per data-taking period (other panels). The lower left panel shows the combination of the old APV settings and new APV settings data-taking period in 2016, that are also shown separately in the upper row.

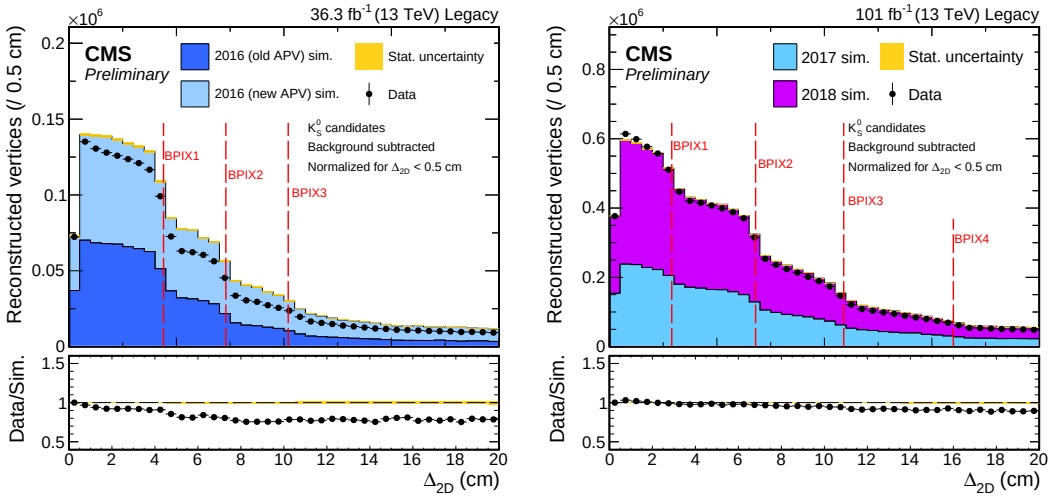


Figure 3.15: Data-to-simulation agreement for displaced tracking and vertexing of K_S^0 mesons, highlighting the physical position of the barrel layers of the CMS pixel tracker, in 2016 data taking (left) and 2017/2018 data taking (right). The figures show the legacy reprocessing, with similar results obtained for the pre-legacy reconstruction (not shown here).

In these figures, we also show a quantity called the significance of the radial displacement, defined as the radial displacement itself divided by its uncertainty.

The uncertainty on the radial displacement is strongly affected by the structure of the pixel detector: the closer a secondary vertex is positioned to the next tracker layer (in an outward direction), the smaller the uncertainty on its position. Directly after each tracker layer, the uncertainty increases because of the larger distance to the next tracker layer, yielding less precise track parameters and vertex fits.

Correction factors The data-to-simulation ratios shown in Figs. 3.12 and 3.14 can be applied as correction factors or quantitative sources of uncertainty to reconstructed displaced vertices, if the experimental signature of interest is close enough to the one studied here. In extension of this, we also calculate these correction factors in a two-dimensional binning, additionally splitting in the p_T of the reconstructed K_S^0 meson. The results are shown in Fig. 3.18 for the pre-legacy reconstruction and Fig. 3.19 for the legacy reprocessing.

3.9 Conclusions and outlook

Summary The reconstruction efficiency of tracks and vertices that originate at a relatively large radial displacement from the center of the CMS detector is a source of uncertainty for analyses working with displaced objects. While these efficiencies can be estimated from simulation, it is unclear how accurate and reliable displaced objects and their reconstruction are simulated. The study described in this chapter quantifies the resulting uncertainty by considering K_S^0 and Λ^0 hadrons, and comparing the amount of reconstructed vertices between data and simulation as a function of the radial displacement. Correction factors are derived, that can be used to scale the simulation and/or to add as an uncertainty source in the statistical analysis of searches involving displaced objects. The magnitude of the discrepancy between data and simulation is usually in the order of 10-15%, but in early 2016 data a larger difference

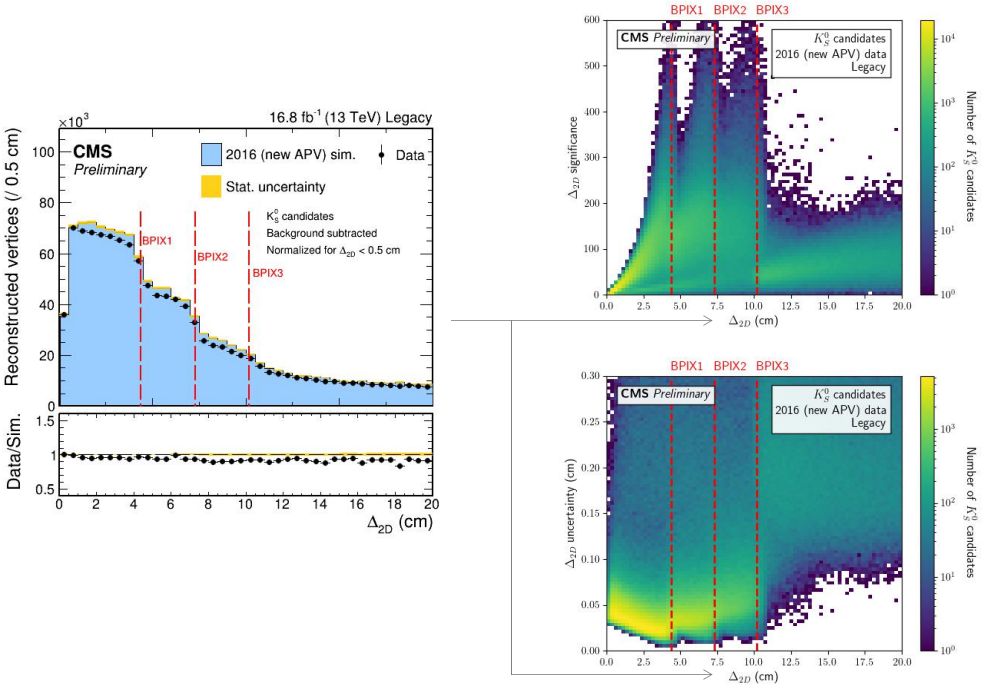


Figure 3.16: Radial displacement of K_S^0 vertices (left), its uncertainty (lower right), and its significance (upper right) in 2016 (new APV) data in the legacy reprocessing. The positions of the pixel detector layers are indicated with red dashed lines.

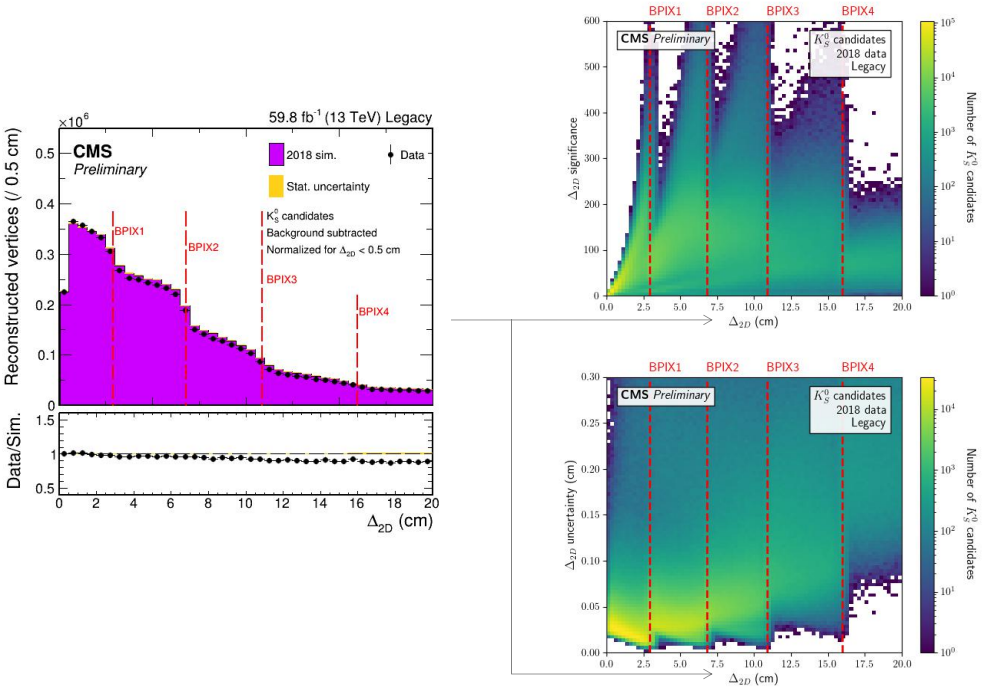


Figure 3.17: Radial displacement of K_S^0 vertices (left), its uncertainty (lower right), and its significance (upper right) in 2018 data in the legacy reprocessing. The positions of the pixel detector layers are indicated with red dashed lines.

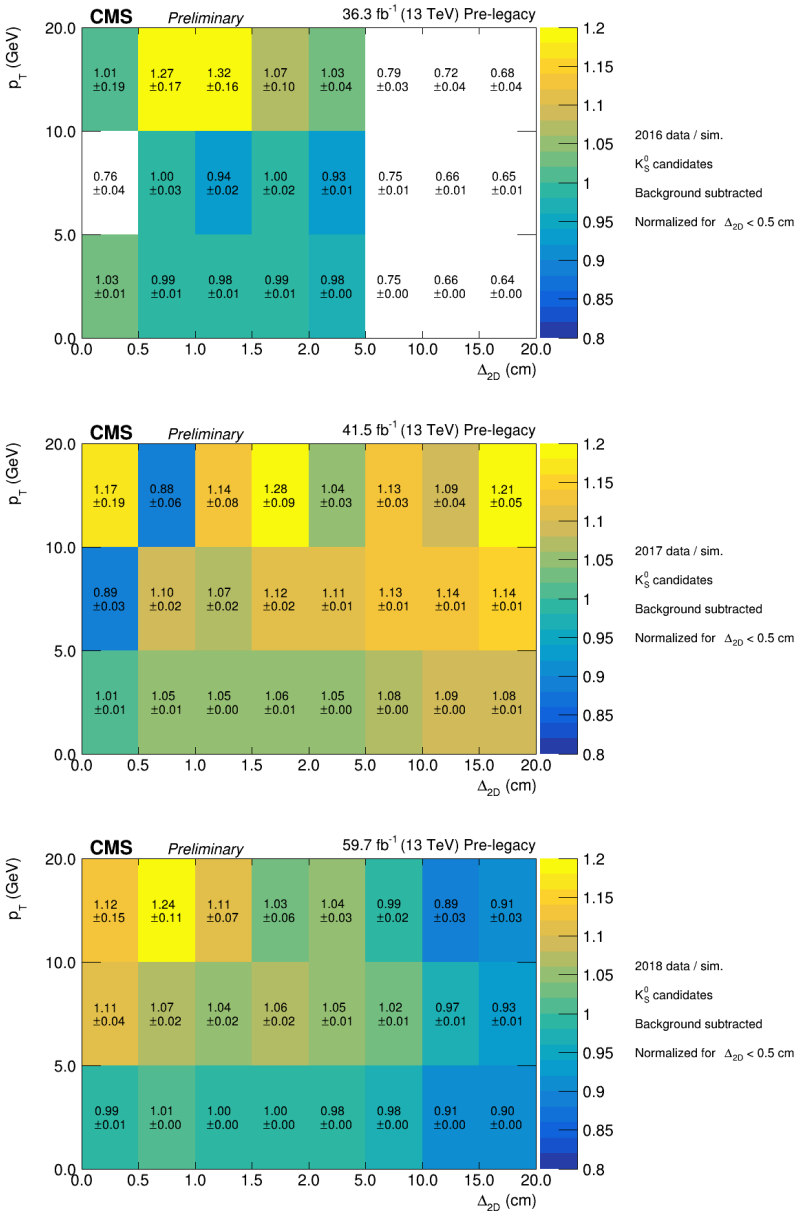


Figure 3.18: Data-to-simulation agreement for displaced tracking and vertexing of K_S^0 mesons as a function of radial displacement and transverse momentum, split per data-taking year (chronologically from top downwards), with pre-legacy reconstruction.

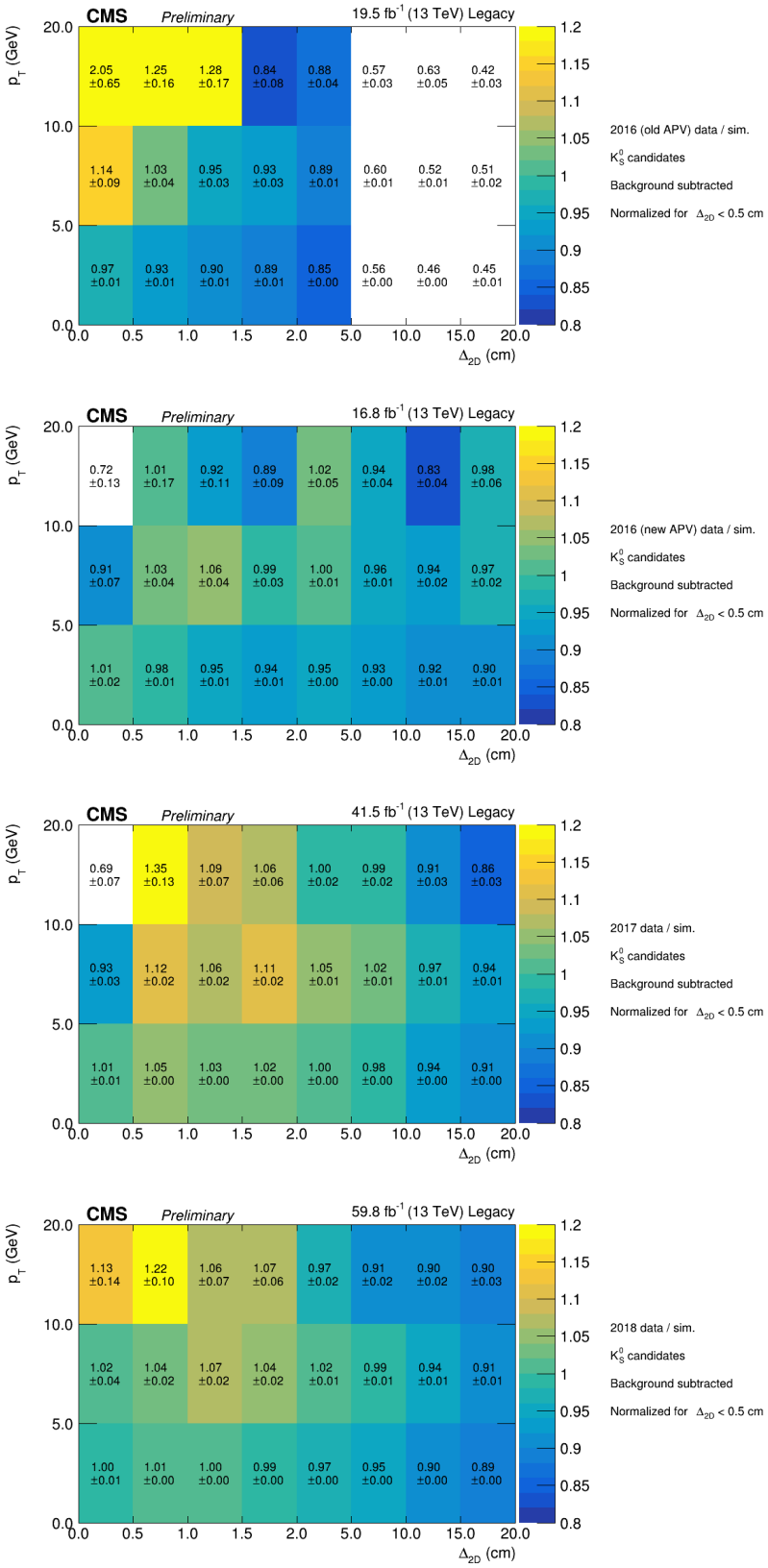


Figure 3.19: Data-to-simulation agreement for displaced tracking and vertexing of K_S^0 mesons as a function of radial displacement and transverse momentum, split per data-taking year (chronologically from top downwards), with legacy reprocessing.

is observed because of the APV saturation effect, that is reduced, but not completely solved in the legacy reprocessing. The results of this study have been documented in a CMS internal analysis note [16] and a public detector performance note [6], were incorporated into several CMS analyses [7, 8], and are planned to be used as a centralized tool for other analyses (e.g. Ref. [103]) and Run III data.

To conclude, we discuss some of the identified shortcomings of the presented study, as well as an outlook on how to overcome them.

Experimental signature The most fundamental drawback of this study consists in the difference in signature between the processes studied here (low p_{T} hadrons) versus the ones typically of interest in HNL searches (high p_{T} leptons). Hadrons and leptons are reconstructed in CMS using different sets of sub-detectors (though the silicon tracker is the most important sub-detector for hadrons as well as for electrons and muons as far as tracking goes) and tracking algorithms, which could affect the displaced vertexing calibration. For example, electron tracks are reconstructed using a Gaussian-sum filter instead of a Kalman filter because of bremsstrahlung emissions (as discussed in section 2.12.4). It is as yet not very clear how well the results in this study can be extrapolated to electron tracks, because of the different tracking algorithm. Also the difference in transverse momentum spectrum could have a non-negligible impact, since the tracking efficiency is known to be p_{T} dependent [59].

A solution to this issue consists of finding a relatively well-known and well-modeled SM process that involves displaced dilepton resonances. A suitable candidate could be the J/ψ meson decaying to a pair of muons or electrons. While this meson is too short-lived to have any direct appreciable contribution to the displacement range of interest, it can be produced at larger displacements in the decay of B hadrons [140].

Vertexing procedure A shortcoming that is intricately related to the previous one, is the custom vertexing algorithm and selection criteria used in this study. Ideally, the displaced vertexing procedure should follow as closely as possible the one employed for the BSM vertex search. This is difficult to achieve when working with different objects (e.g. hadrons versus leptons), but could be implemented when this study is repeated with a SM displaced lepton resonance. On the other hand, while the details of the vertexing algorithm can be expected to impact the displaced vertexing efficiency, they are not expected to have an important impact on the data-to-simulation agreement in this respect.

Corrections and uncertainties A further potential improvement to this study consists of applying more rigorously and exhaustively all recommended corrections, correction factors and associated systematic uncertainties to the simulation[15]. While these corrections and uncertainties are not expected to alter the results significantly because of the additional normalization performed in this study, including them would enhance the completeness and consistency, and would allow for a rough estimate of the uncertainties on the obtained correction factors. (In the current version, only statistical uncertainties were considered.[16])

[15]The correction for matching the pileup distribution in simulation to that in data, expected to be the most important systematic correction in this context, was already taken into account in this study. Other relevant corrections, not yet taken into account in this version of the study, include for example lepton identification and momentum corrections, and b-tagging corrections, which are expected to improve the simulation-to-data agreement in Fig. 3.5, to separate more unambiguously the displaced tracking and vertexing effects from any other effects.

[16]Preliminary studies were performed with two sources of systematic uncertainty that can be expected to have a significant impact. The first one consists of propagating the estimated uncertainties the background fit to

Normalization and background subtraction Next, the strategy for background subtraction and normalization could be revisited. While for the K_S^0 as probing particle, the situation is manageable (but likely not optimal), for Λ^0 a difficult compromise has to be made: because of high background levels and a low number of signal events in small displacement bins, the background subtraction method in these bins results in too large uncertainties, and hence the bins at slightly larger displacement are used for normalizing the simulation to the data; this could however normalize away a part of the effect we are trying to measure. In a next iteration, the vertex reconstruction cut values could be optimized more systematically, with special attention to the resulting signal-to-background ratio in the low displacement bins. Potentially this could already drastically improve the situation and allow to use the Λ^0 particle as a cross-check for the K_S^0 method.

Measuring efficiency Finally, we have only considered the difference in efficiency between data and simulation, and have not touched upon the efficiency itself. In the current setup for this study, calculating the efficiency for K_S^0 or Λ^0 reconstruction as a function of radial displacement is non-trivial, since not all K_S^0 and Λ^0 particles are stored in the generator-level information in the simulation. To overcome this difficulty, one might consider generating custom samples with all K_S^0 and Λ^0 particles stored at generator-level. On the other hand, assuming that other probing particles are found, decaying to leptonic final states (to overcome some of the other issues mentioned above), this would likely no longer be a problem.

the final results. The second one is the uncertainty in the pileup reweighting procedure (see section 6.5 for more details). The effect of these systematic uncertainties is to approximately double the total uncertainty (with respect to statistics only), but have no significant impact on the observed simulation-to-data discrepancies.

Chapter 4

Towards automated data quality monitoring using machine learning

This chapter describes a collection of studies performed in the context of the automation of data quality monitoring and data certification. The results have been published in the form of a detector performance note [9], a number of conference posters[1] and proceedings [10].

4.1 Introduction

DQM and DC In complex and composite detectors such as CMS, extensive monitoring of all detector components is paramount to assure the quality of the data and to spot detector problems when they arise. This is a vital prerequisite for providing reliable, high-quality data

[1] 15th Pisa Meeting on Advanced Detectors (PM 2022) and 21st International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2022)

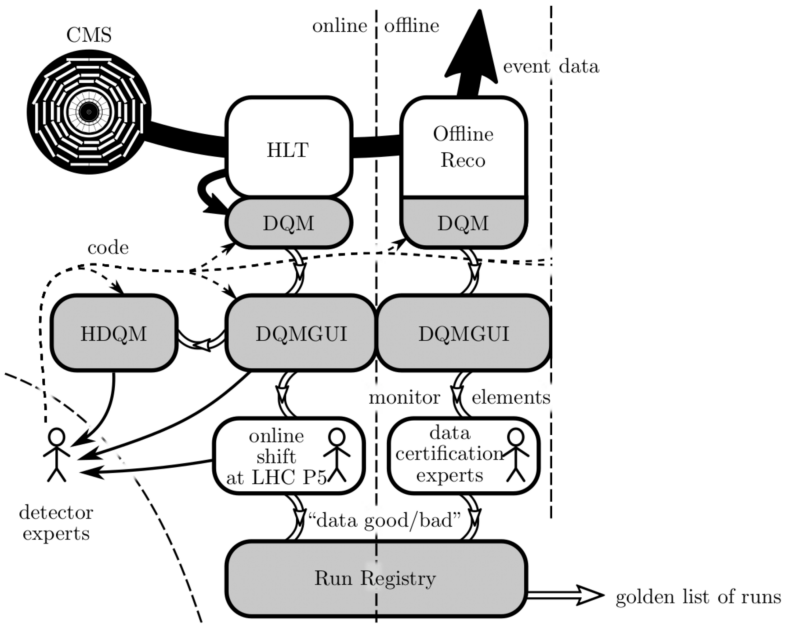


Figure 4.1: Summary of the current DQM ("online") and DC ("offline") procedures in CMS. The final output is a list of good ("golden") runs, i.e. sections of data taking with validated good data quality. Figure (slightly modified) from [141].

and for accurate modeling of dynamic detector effects in the simulation, and hence for all analysis efforts using CMS data. An introduction to the data quality monitoring (DQM) and data certification (DC) procedures in CMS can be found in [141, 142], with the main points of interest shortly summarized below and in Fig. 4.1. While CMS is running, dedicated data streams (comprising about 10% of the data rate stored for physics analyses) are processed into a set of reference histograms called monitoring elements, that summarize the status and performance of the detector. These monitoring elements are the input for both the DQM and DC workflows, with the distinction being the following:

- **Data quality monitoring (DQM)**: the monitoring elements are inspected during data-taking in real-time by shifters, with the purpose of spotting detector issues that need immediate action.
- **Data certification (DC)**: a more extensive set of monitoring elements is re-inspected after data-taking by subsystem experts, in order to spot subtle detector or data quality issues and certify the validated data as good for physics analyses.

Limitations The efficiency of the DQM and DC procedure is defined as the fraction of certified data with respect to all CMS recorded data, and it is typically in the order of 90 - 95%, as exemplified in Fig. 4.2. Despite the high efficiency, the current workflow has some disadvantages and bottlenecks related to the largely manual procedures. Shifts are carried out 24/7 in eight-hours time slots, which may lead to fatigue and loss of focus. Furthermore, the number of monitoring elements to be visually inspected by shifters and experts is necessarily large because of the complexity and number of components of CMS[2], which limits the time resolution of the data certification. Finally, there is a strong dependence on the visualization of the

[2] As an order of magnitude, a quick counting shows 302 monitoring elements divided over 21 subsystems that should be in principle monitored continuously, and this is only the summary for real-time data taking. For offline certification, there are multiple thousands of monitoring elements that need to be inspected.

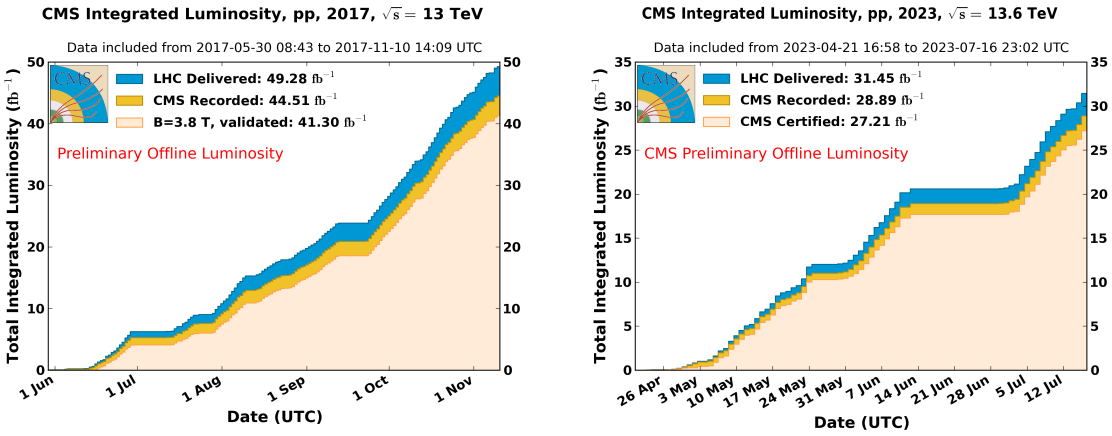


Figure 4.2: Overview of delivered, recorded, and certified integrated luminosity for data taking in 2017 (left) and 2023 (right). The difference between delivered and recorded data is caused by downtimes of the CMS detector (or specifically of its trigger or data acquisition system). The difference between recorded and certified/validated data is caused by various kinds of problems in the subdetectors of CMS that jeopardize the data quality. Figure from the *Public CMS Data Quality Information*.(*)

(*) `https://twiki.cern.ch/twiki/bin/view/CMSPublic/DataQuality`

underlying data: sub-optimal plotting may obscure the presence of an anomaly to a human eye.

Automation Automatic anomaly detection methods could mitigate these issues. For example, certification could be performed per luminosity section (LS, a data taking period of about 23 seconds used as elementary time unit in CMS data taking) as opposed to per run (a data taking period typically lasting hundreds of luminosity sections - a couple of hours). The goal of this project is not to replace the current DQM and DC procedures, but rather to support them by pointing shifters and subsystem experts towards automatically detected potential anomalies, in order to more efficiently direct their attention and focus. While the baseline procedure is the same for DQM and DC, the former is particularly challenging because the methods have to be applied during real-time data taking; they should be stable, reliable, and ideally easily interpretable by non-expert shifters (while the latter can be trained and applied on a much more flexible timescale and investigated by subsystem experts). For this reason, the studies in this chapter start from an offline DC perspective, with a few feasibility studies for an application in online DQM towards the end. This could lead on the one hand to a higher DQM efficiency, as detector issues could be spotted sooner, and on the other hand to a more reliable DC result, as anomalies can be spotted on a much finer time granularity.

The project of (partly) automating the DQM and DC workflows is being pursued in parallel by several subsystems. In the next sections, I will mainly describe my own studies, focusing on some particular monitoring elements of the CMS pixel tracker, postponing a short overview of related work in different subsystems to section 4.5.

4.2 General considerations, strategy, and autoencoders

Lack of labels An important consideration dictating the overall strategy for these studies is the lack of reliable data labels. Though the used data sets (2017 data in one study, 2023 data in the other) were of course certified (via the largely manual procedure summarized in

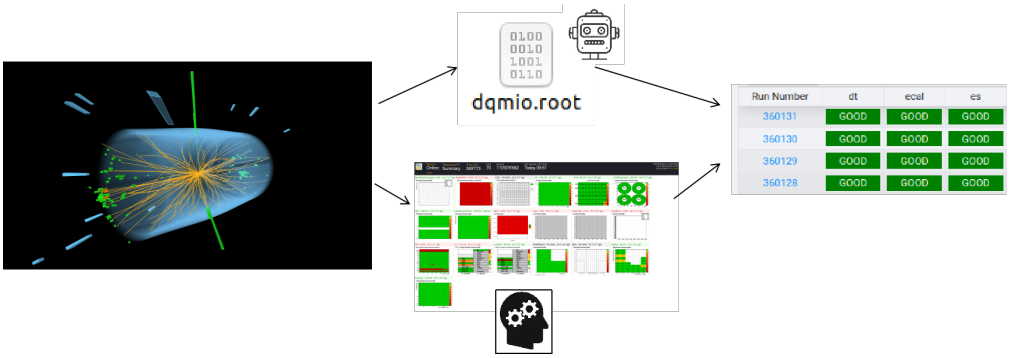


Figure 4.3: *Very* short summary of data quality monitoring (DQM) and data certification (DC) at CMS. Left: representative event display of CMS while taking data. Below: plots in the DQM GUI summarizing the detector status and performance, inspected by shifters and subsystem experts. Right: register of detector status per run and per subsystem. Above: can the manual procedure be improved, augmented or streamlined by automated methods?

section 4.1), these labels usually apply on a per-run granularity, while we intend to go to per-LS level. Furthermore, labels are assigned based on the estimated overall status of an entire sub-detector (the pixel tracker in our case study), which does not necessarily correspond to the status as inferred from the small selection of monitoring elements used here. And finally, even assuming the available labels to be fully correct, they are not guaranteed to cover each potential failure scenario. For these reasons, we approach this problem as an unsupervised one.

Class imbalance The large majority of the data is expected to be good (based on operational experience and previous certification cycles). This results in a very large class imbalance between good and bad instances, potentially biasing any classification method. Furthermore, we prefer remaining agnostic with respect to the behaviour of bad instances, as different kinds of issues might result in different phenomenologies at monitoring element level, some of which may not be present in our training set. For these reasons, we employ anomaly detection techniques; not limited to classifying predefined good and bad instances, but detecting anything that is deviating from the normal behaviour.

Anomaly detection In summary, we apply unsupervised anomaly detection techniques. We train the method on an unlabeled dataset with some minimalistic filters applied and check the performance on a set of known good and bad luminosity sections (see details further on). The method relies on the algorithm learning the generally accepted behaviour of the data (in the training set) and flagging any anomalous deviation. These considerations apply grosso modo to all studies presented in this work, as well as to those by other subsystems briefly mentioned in section 4.5.

Global vs. local Two distinct approaches have been followed: one referred to as 'global training' and one as 'local training'. The former employs a large data set (e.g. the full 2017 data) for training the models and is used to find self-inconsistencies, i.e. anomalies within the training set. This approach could assist in certification cycles at the end of a longer data-taking period. The latter one mimics the procedure during live data-taking, where a small set of recently recorded data would be used to train a model instantly applied on the current ongoing recording. The next sections mainly focus on global training, while the differences with local training are highlighted in section 4.3.6 (for one case study) and 4.4.6 (for the second one).

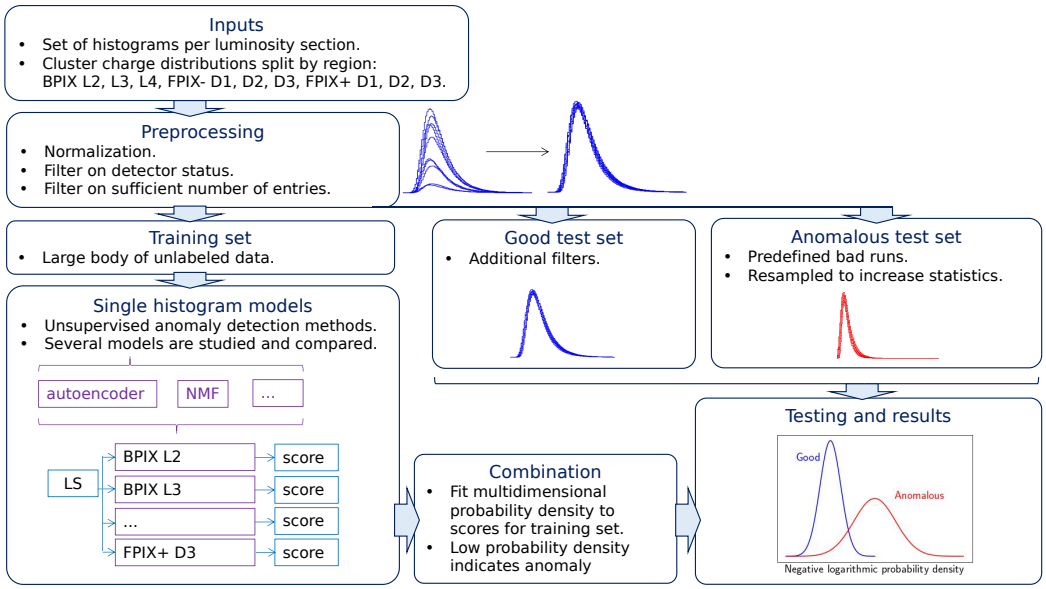


Figure 4.4: Schematic overview of the method: from input monitoring elements (essentially just histograms) to a quantitative anomaly score.

Overview The approach for global training is conceptualized in Fig. 4.4, with each of the steps detailed in the paragraphs that follow. In summary: an unsupervised anomaly detection model is trained on a large set of unlabeled data and tested on a good subset of the training data as well as on a collection of known bad runs. The models operate on single monitoring elements, so in order to obtain a final score per luminosity section (characterized by multiple monitoring elements), the individual scores are fitted with a multidimensional probability density function.

Autoencoders For the task of unsupervised anomaly detection, these studies rely on autoencoders. An autoencoder [143, 144] is a neural network that has the same output dimension as input dimension. Whereas 'regular' neural networks in their most simple form are usually trained to output one single number for signal instances and another number for background instances, autoencoders are trained to reconstruct their input as accurately as possible. They are restrained from learning the identity transform by having intermediate layers of smaller dimensions than the input and output layers. As such, autoencoders are forced to learn an efficient representation of the data, encapsulating the essential information in a smaller format than the original data (illustrated in Fig. 4.5). Common applications of autoencoders include clustering, dimensionality reduction, data compression and denoising. In the slightly modified form of variational autoencoders [145], they can be used as generative models as well.

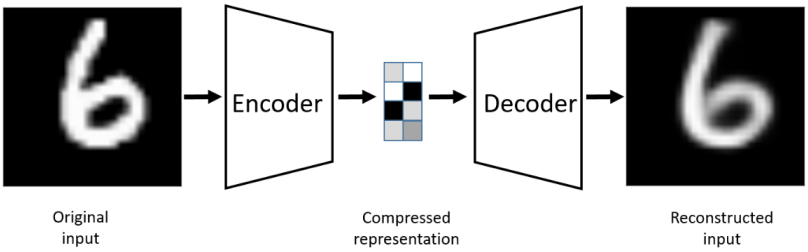


Figure 4.5: Illustration of an autoencoder trained on handwritten digits. The intermediate layer in between the encoder and decoder is trained to contain an efficient representation of the data, from which the original input can be reconstructed with minimal losses. Figure from Ref. [144].

Autoencoders for anomaly detection Autoencoders are suitable for our purpose because they are implicit anomaly detectors. Since they are trained on reconstructing the largest possible number of input instances as accurately as possible, their learned transformation will focus on the general, most prevalent behaviour of the data in the training set. For example, in the simple hypothetical case where all good monitoring elements have the same shape and differ only in their integral, a single number (namely a scaling constant) suffices to accurately reconstruct a given good monitoring element. An anomalous monitoring element will not be accurately reconstructed by the autoencoder, as only its integral will be preserved and not its shape. Hence the mean-squared-error between an input monitoring element and its autoencoder reconstruction can be used as a measure for anomality of the input monitoring element.

4.3 Anomalous cluster charge in the pixel tracker

This section presents the details of the first of the two case studies discussed in this chapter; the second one is described in section 4.4.

4.3.1 Monitoring elements and data set

The case study reported here uses a set of monitoring elements from the CMS pixel tracker [49]. In more detail, we use the distributions of the collected electric charge (in elementary charge units) per cluster for three of the pixel barrel layers (denoted as BPIX L2 to L4), as well as for the pixel forward disks on each side of the barrel (denoted as FPIX± D1 to D3)[3]. The distribution of cluster charge in the inner pixel barrel layer (BPIX L1) was not included for now, as its shape is typically unstable and varying rather strongly, while not necessarily indicating unexpected detector problems[4].

These monitoring elements are sensitive to a wide spectrum of potential detector issues or other anomalies that cause a loss of electric charge collection efficiency (such as high-voltage failures, readout issues or silicon degradation), which would be manifested by a deviation from the expected Landau shape. The monitoring elements for BPIX L2 are displayed for one good

[3]In CMS technical jargon, these monitoring elements are called PixelPhase1/Tracks/PXBarrel/charge_PXLayer_* (where * ranges from 2 to 4) and PixelPhase1/Tracks/PXForward/charge_PXDisk_* (where * ranges from -3 to +3, excluding 0).

[4]Because of its position closest to the interaction point, L1 is inherently more prone to dynamic inefficiencies at high instantaneous luminosity, such as radiation effects and single-event upsets (where a single ionizing particle strikes a sensitive part of the electronics) [49,146]. These are however known issues, for which automatic recovery procedures are already in place, and therefore fall outside the scope of this anomaly detection study.

and two problematic example runs in Fig. 4.6. The distributions for the other barrel layers and endcap disks are similar.

The monitoring elements used in this case study result from a special reprocessing of the 2017 data, where they are stored with a per-LS time granularity (while usually only the per-run elements are stored after the real-time DQM step). While this was an ad-hoc procedure, it has now been formalized for Run III data-taking, detailed in section 4.6. The dataset used in this case study is the ZeroBias dataset (applying only minimal trigger selections, typically used for tracking studies), reconstructed with legacy reprocessing[5] [134].

4.3.2 Preprocessing, training set and testing sets, and network architecture

Some preprocessing is required to obtain a good-quality dataset of representative distributions for the monitoring elements used in this study. We perform two basic selections on the available set of all luminosity sections in the 2017 data:

- **DCS** The detector control system (DCS) [147] is part of the monitoring and data acquisition system of CMS. It summarizes (among other things) the readiness of the components of CMS for data-taking (whether the high-voltage is switched on, etc.). The DC team of CMS provides a convenient overview of all runs and luminosity sections where the DCS status is ok (i.e. the detector is 'switched on'), which is used as a first and basic filter in this study. Note that the creation of this filter is fully automatic and does not require manual intervention.
- **Entries** We also filter out luminosity sections that have monitoring elements with a low number of entries and large statistical fluctuations, as could happen for example when there is a problem in the LHC or with triggering in CMS. Since the autoencoder (and other methods) are not expected to behave well on low-statistics monitoring elements and since many other and simpler methods are able to 'detect' almost empty monitoring elements, they are removed from the study. For the monitoring elements used in this study, we found the criterion (number of entries) / (number of bins) > 100 to be a satisfactory compromise between quality and quantity of passing instances.

Normalization Furthermore, all individual monitoring elements are normalized to unit surface area, as it was observed that the integral depends strongly on instantaneous data-taking conditions (e.g. luminosity and pileup) while the pure shape gives a more reliable distinction between good and bad instances. This will be shortly revisited in section 4.7.

Template filter This selection and normalization suffices to obtain a reasonably clean set of trainable instances, but does not prevent a non-negligible amount of anomalous monitoring elements from being still present in the training set. At the same time, we prefer to have a method that does not rely on a previous, manual certification cycle to define the training set as a list of good runs (which, moreover, might not be accurate enough because manual certification is largely run-based). To purify the training set and hence improve the autoencoder performance, an additional template filter is applied to remove clear outliers. This is done by calculating the mean-squared-error (MSE) between each monitoring element in the training set and those in a collection of reference template histograms, and rejecting those which differ too strongly from the templates. The templates are obtained as averaged partitions from the training set, and the rejection threshold is tuned to a rather tight value in order to obtain a clean training set. The procedure is illustrated in Fig. 4.7, and the number of remaining

[5] In CMS technical jargon, the used datasets are /ZeroBias/Run2017*-09Aug2019_UL2017-v1/DQMIO where * stands for the era indicator, B, C, D, E or F.

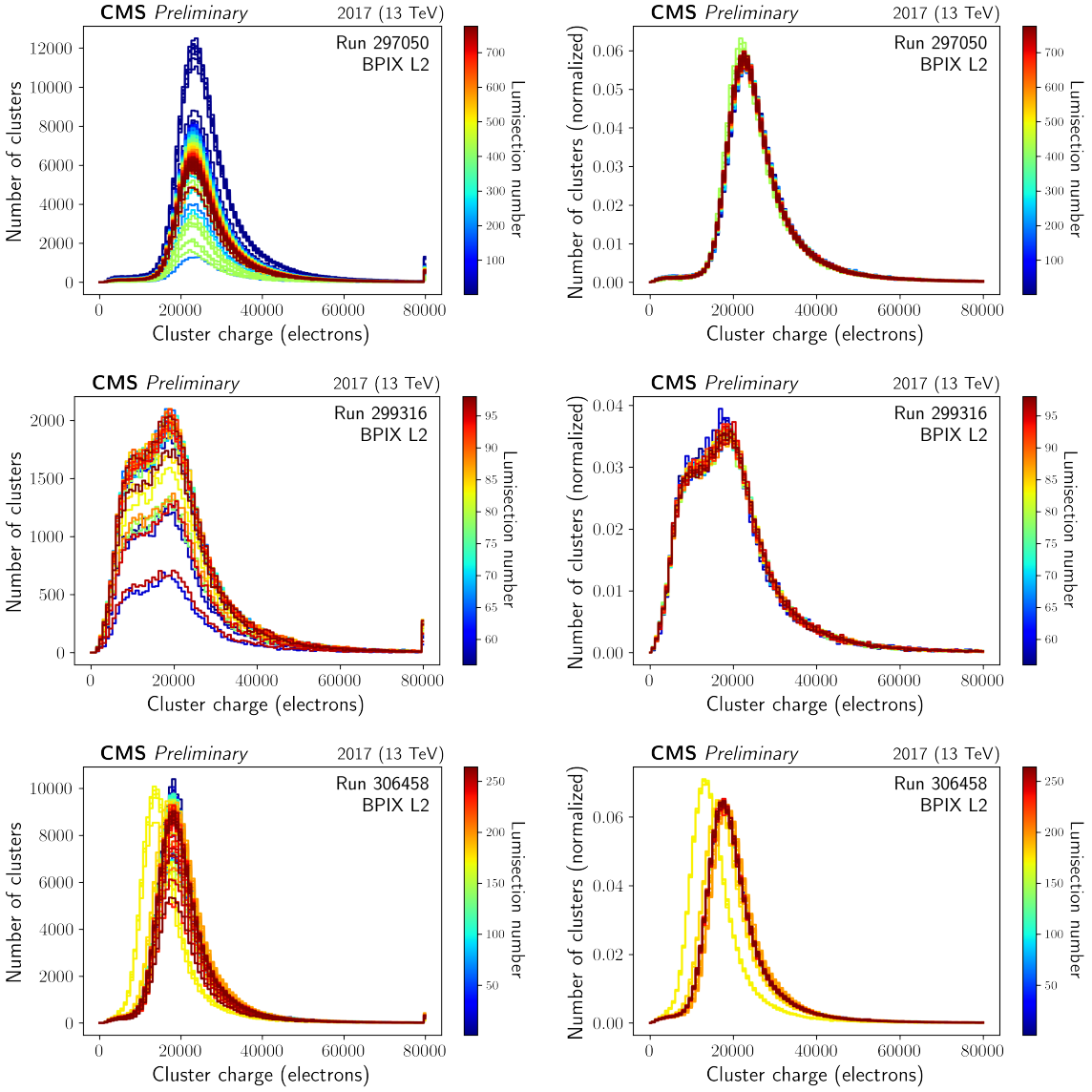


Figure 4.6: Distributions of collected electric charge per cluster for the second pixel barrel layer (BPIX L2). Left column: raw distributions. Right column: same distributions, but normalized to unit surface area for easier shape comparison. Upper row: run 297056, showing the expected behaviour for this type of monitoring element. Middle row: run 299316, showing the potential effect of a detector problem. Lower row: run 306458, showing a more subtle detector problem, that lasts only a few luminosity sections and would not be seen in run-averaged quantities. The color scale shows the luminosity section number, equivalent to the time dimension within the run. While the figures are for BPIX L2, the cluster charge distributions in other pixel barrel layers and disks have similar shapes.

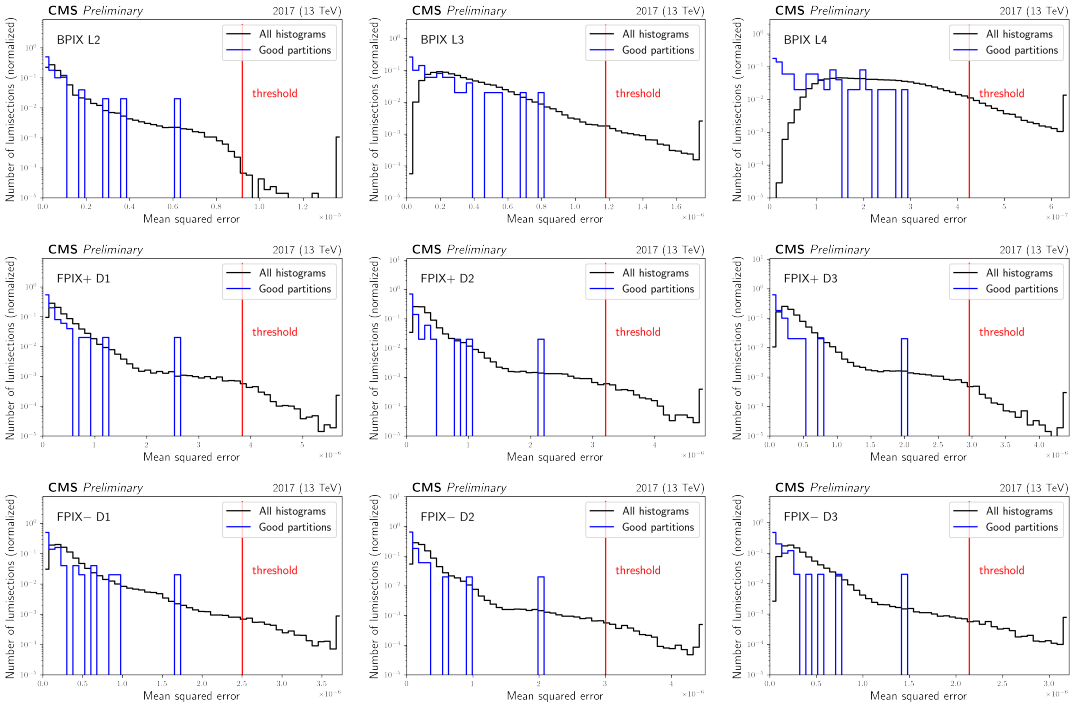


Figure 4.7: Template filtering to remove outliers from the training set for the autoencoder. The plots show the MSE between the monitoring elements in the training set and a set of reference templates, which serves as a crude anomaly score for that monitoring element. The black line represents the distribution of the MSE for all monitoring elements in the training set; the blue one is the MSE for a set of averaged partitions, that represent good monitoring elements. The threshold is chosen as 1.5 times the maximum MSE value in the blue distribution, and monitoring elements with an MSE above this threshold are not used in the autoencoder training.

luminosity sections after all selections is listed in Tab. 4.1. All luminosity sections that pass these selections and preprocessing steps are used in the training.

Test set: good instances To construct a test set consisting of good instances, several approaches have been attempted and compared:

- **Golden json**: The so-called 'golden json' selection is a list of runs and luminosity sections that have manually been certified as good for physics analyses (see Fig. 4.1). Still, it cannot be used directly as a test set of good instances because the manual certification is largely run-based instead of LS-based and some anomalies lasting only a short time might be missed. Furthermore, some shape anomalies in the monitoring elements under consideration in this study might prove non-crucial for the physics usage of the data, hence being included in the golden json despite the anomaly. Yet, we still check that our method performs well on the golden json (see section 4.3.4) in the sense that the number of anomalies found in it is small, and that they correspond to actual anomalies rather than false alarms.

- **Predefined runs**: A fixed collection of good runs and/or luminosity sections is selected based on criteria such as the pileup, trigger rate and above all visual appearance of the

Table 4.1: Number of luminosity sections passing the selections for the training set.

Selection	Number of remaining luminosity sections	Fraction of total (%)
All LS (2017 data)	225954	100
- era 2017 B	28335	
- era 2017 C	61522	
- era 2017 D	28747	
- era 2017 E	44870	
- era 2017 F	62480	
DCS-on	215144	95.2
Sufficient number of entries	210361	93.1
Template filter	191663	84.8

monitoring elements. This method does not suffer from the disadvantages of using the golden json, but has larger potential biases due to the specifics of the chosen runs.

- **Averaged partitions**: In this approach, the training set is divided into partitions and the monitoring elements within each partition are averaged. This creates a set of good histograms on the condition that the number of partitions is not too large, so that short-lived anomalies are averaged out. On the other hand, the resulting averages are often artificially good and not a complete representation of the full phase space. Moreover, the test set of good instances is inconveniently small (in the order of 50 instances).

- **Resampled averaged partitions**: To overcome the difficulties in the previous method, the averaged partitions can be resampled into a larger set with more variation (see section 4.3.5). This does improve but not completely solve the problem of missing phase space.

- **Training set**: Finally, the good part of the test set can be chosen equal to the training set. This method has the disadvantage of being blind to potential remaining anomalies in the training set, but this can be mitigated by having a large variation in anomalous luminosity sections for testing. It can be furthermore blind to good luminosity sections that do not pass the template filtering, but this can be mitigated by checking the number of false alarms raised when evaluating on the golden json. The combination of these methods is expected to give a reliable estimate of the performance and will be used when quoting the results in section 4.3.4.

Test set: anomalies The part of the test set consisting of anomalous instances is constructed from (resampled) manually observed bad runs, detailed in Tab. 4.2. Luminosity sections from these runs are selected if they pass the DCS and number-of-entries criteria (but no template filter is applied). As illustration to these considerations, the good part of the test set defined in terms of averaged partitions is displayed in Fig. 4.8, while one of the anomalous runs is shown in Fig. 4.9.

Autoencoder architecture The architecture of the autoencoder model for this study is kept deliberately simple, to ensure that the model cannot overfit to the (rare but unavoidable) anomalous instances in the training set. Its properties are summarized in Tab. 4.3. Other advantages of the shallow network architecture are the short time and modest resources required for training. On a regular single CPU, the training of this model takes about 80 seconds (per

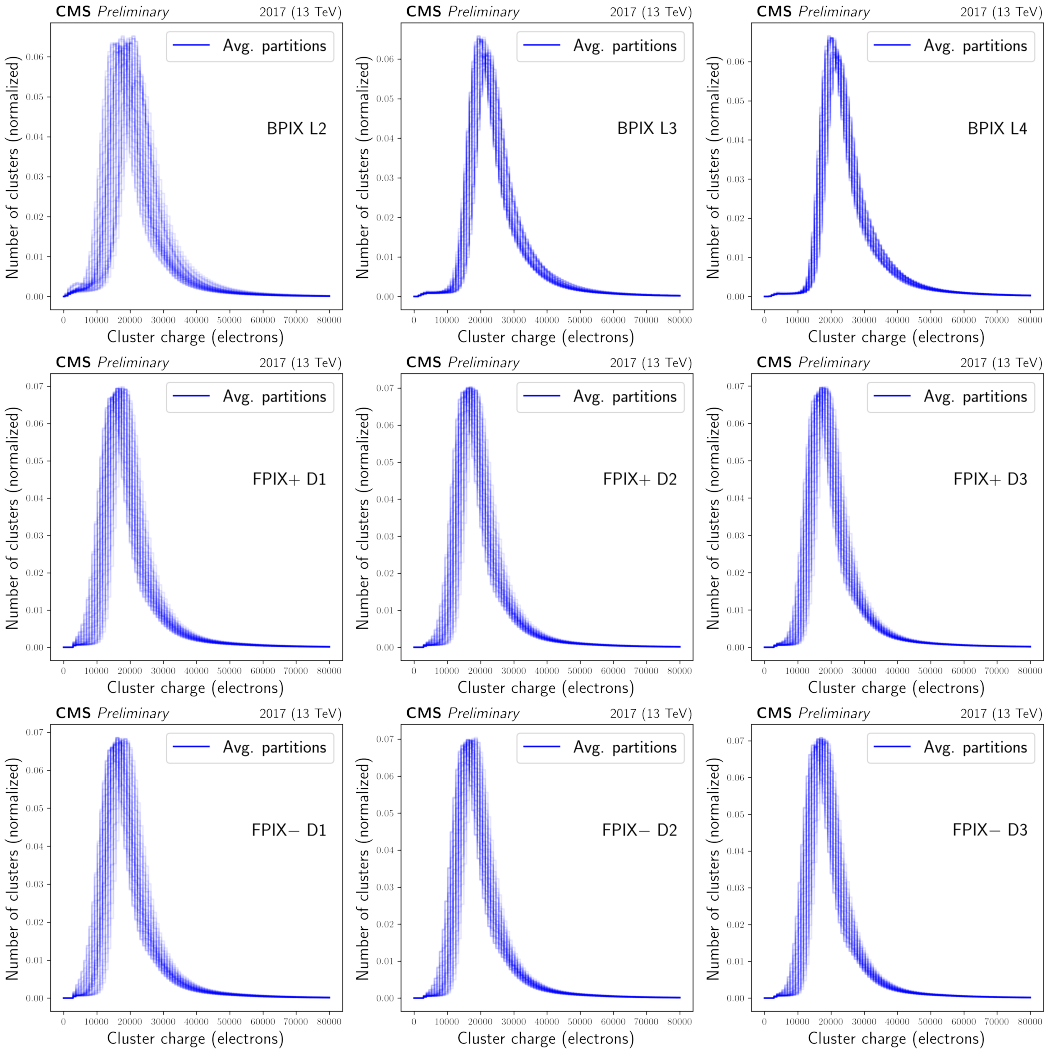


Figure 4.8: Expected shapes for the distributions of collected cluster charge in the barrel layers and forward disks of the pixel tracker. The shapes have been obtained as 50 averaged partitions from the entire set of 2017 luminosity sections and show the range and variability of acceptable distributions depending on the detector conditions.

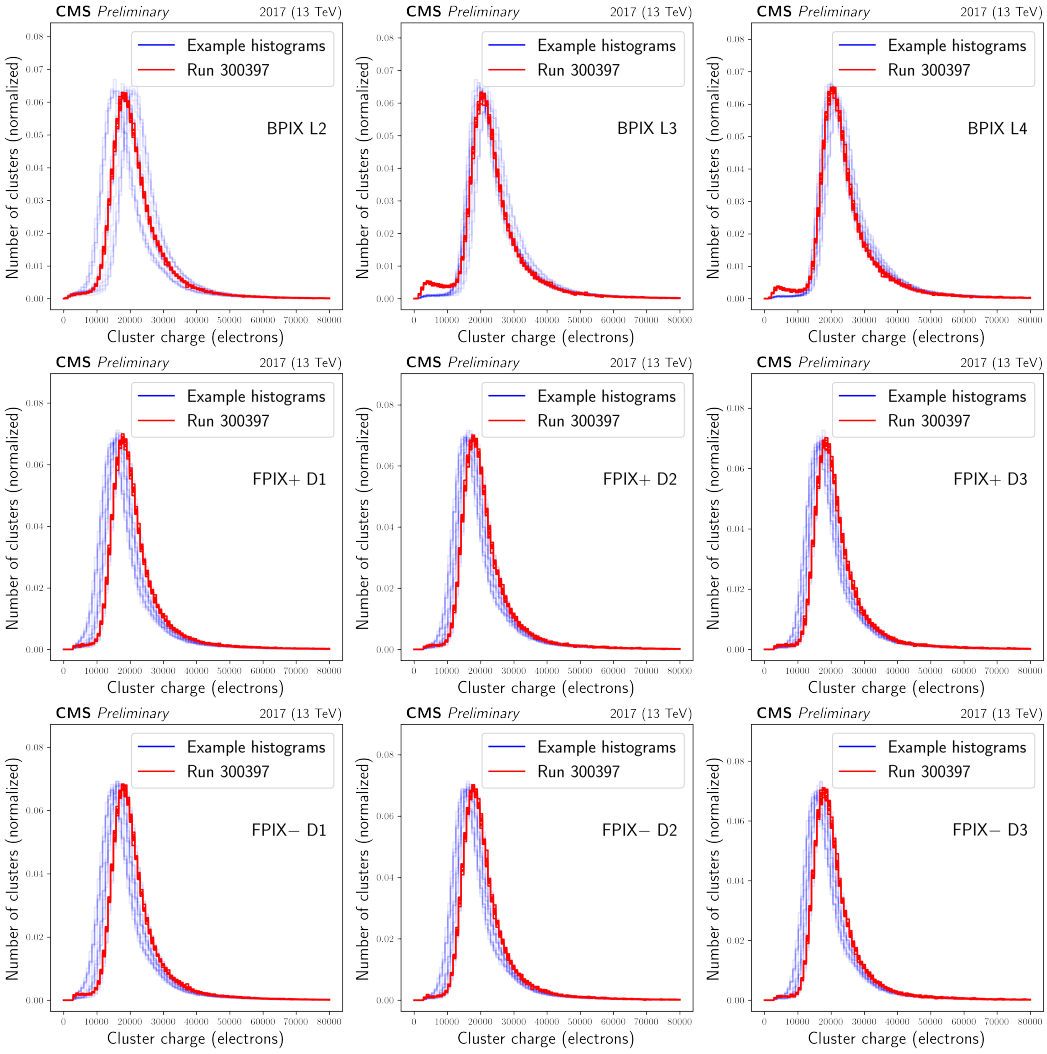


Figure 4.9: The luminosity sections from run 300397 (red) compared to examples of good monitoring elements (in blue) in the barrel layers and forward disks of the pixel tracker. During this run, a timing scan caused a loss in charge collection efficiency, visible as secondary bumps at low cluster charge values in barrel layer 3 and 4.

Table 4.2: Definition of runs constituting the anomalous part of the test set.

Run number	Total number of luminosity sections	Selected number of luminosity sections	Cause of anomaly
297287	16	15	Pixel timing scan
297288	15	15	Pixel timing scan
297289	15	15	Pixel timing scan
299316	93	43	Pixel timing scan
299317	35	35	Pixel timing scan
299318	46	46	Pixel timing scan
299324	22	9	Pixel timing scan
299325	190	190	Pixel timing scan
299326	22	22	Pixel timing scan
300373	13	13	Pixel timing scan
300374	13	12	Pixel timing scan
300379	13	13	Pixel timing scan
300398	14	14	Pixel timing scan
Total	507	442	

Table 4.3: Summary of autoencoder design and training parameters used in this study.

Property	Value
Architecture	One hidden layer between input and output, with half as many nodes as there are bins in the input monitoring elements.
Activation	Hyperbolic tangent (see the keras activations documentation).
Optimizer	Adam (see the keras optimizers documentation) [148].
Loss function	Mean squared error (MSE) between original and reconstructed monitoring element.
Training strategy	Double training, with decreasing batch size, usually 20 epochs with a batch size of 2000 followed by 20 epochs with a batch size of 200.

monitoring element). This implies that frequent re-training (e.g. when detector conditions have changed) is very feasible.

4.3.3 Combining monitoring elements

The autoencoders detailed in section 4.3.2 are trained separately for each of the monitoring elements considered in this study, i.e. for the cluster charge distributions in the pixel layers (all but the innermost one) and the endcap disks. When applying the method in practice, this results in 9 output scores per luminosity section (one for each of the monitoring elements considered). These are combined in order to get a single anomaly score per luminosity section, by fitting a multidimensional probability distribution to the cloud of 9-dimensional points in the space spanned by the output scores for each monitoring element. Several forms of the probability distribution have been attempted:

- Uniform multi-rectangular distribution:

$$f(\bar{x}) = \begin{cases} a & \text{if} \quad x_1 < t_1, x_2 < t_2, ... \\ 0 & \text{elsewhere} \end{cases}$$

  With a a normalization constant and $t_1, t_2, ...$ a set of threshold values. Note that the values of $x_1, x_2, ...$ are non-negative as they represent squared differences. This form of the probability distribution corresponds to defining a threshold for each of the monitoring elements and performing a logical OR, i.e. a luminosity section is considered anomalous if at least one of its monitoring elements is above threshold. This method is conceptually simple but has the disadvantages of being binary and requiring to fine-tune multiple thresholds.

- Gaussian kernel density:

$$f(\bar{x}) = a \sum_i K(\bar{x} - \bar{x}^i) \text{ with } K(\bar{x}) = \exp\left(-\frac{1}{2}\bar{x}^T \Sigma_K^{-1} \bar{x}\right)$$

  With a a normalization constant and Σ_K the kernel covariance matrix or in this context also called the bandwidth matrix. It is estimated according to Scott's rule of the thumb[6] by multiplying the data covariance matrix with the factor $n^{\frac{-1}{d+4}}$ where n is the number of points and d is the number of dimensions. This method can be thought of as centering a little gaussian distribution on each of the points, and summing them all up. This method has the disadvantage of being computationally rather slow due to the large number of points.

- Centered gaussian:

$$f(\bar{x}) = a \cdot \exp\left(-\frac{1}{2}\bar{x}^T \Sigma^{-1} \bar{x}\right)$$

  With a a normalization constant and Σ the data covariance matrix. This is a single multidimensional gaussian distribution centered on the origin and with a covariance matrix derived from the data. Note that only the fraction of space where all coordinates are non-negative is meaningful, hence the normalization constant is non-trivial, but its value does not matter for the remainder of the method as only relative differences in probability values are of interest.

This centered gaussian proves to give optimal performance and will be used in the remainder of this section. The method is illustrated (in two dimensions) in Fig. 4.10

4.3.4 Results

First, the method is applied to a test set consisting of good instances (training set) and bad instances (resampled known bad runs). The resulting score distributions are used to tune a threshold value. Next, the method with the chosen working point is applied on the golden json.

Performance on test set The performance of the method on the test set is summarized in Fig. 4.11, showing the distribution of output scores for good and bad instances, the ROC curve, the optimal working point and the corresponding confusion matrix. For determining the optimal working point, the criterion of maximal pseudo-AUC is used, defined as follows:

$$\text{Pseudo-AUC} = \text{signal efficiency} \times (1 - \text{background efficiency})$$

[6] See e.g. the scipy.stats.gaussian_kde documentation and references therein.

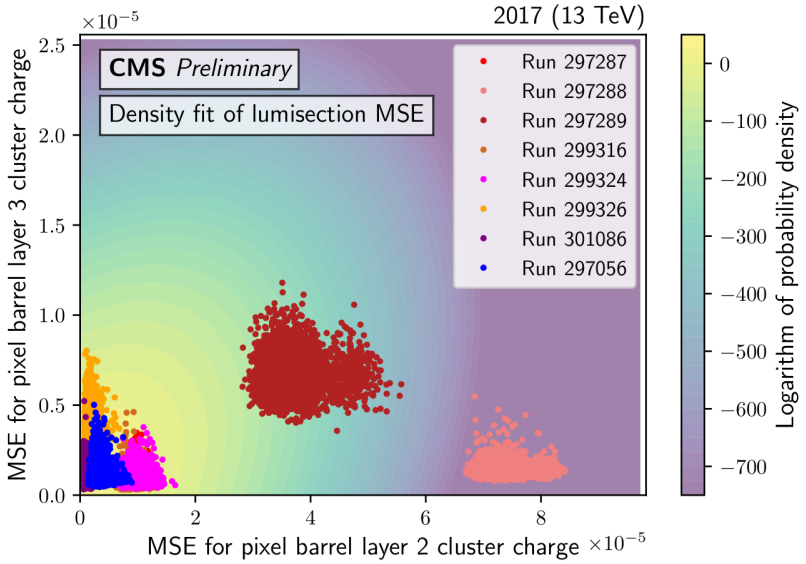


Figure 4.10: Illustration of the method used to combine the autoencoder output scores for several input monitoring elements. Each point in this figure represents one luminosity section. The points in blue are luminosity sections run 297056 (considered to be of good quality and non-anomalous according to the conventional, manual certification procedure), while the other points, in shades of red, pink, orange and brown, are luminosity sections from runs displaying anomalies in the input monitoring elements. The anomalies are caused by high voltage issues in run 301086, and timing scans for all other cases (see also Tab. 4.2). The x- and y-axes feature the mean squared error (MSE) between the original monitoring element and its autoencoder reconstruction, for the cluster charge in the pixel barrel layer 2 and 3 respectively. The background contour is a centered gaussian probability density function fitted to a similar point cloud from the autoencoder training set (not shown in the figure). A threshold can be imposed on this probability density to distinguish good from anomalous luminosity sections. The blue points are on average closer to the origin, as they have lower reconstruction errors. Some of the anomalies may not be manifest in these particular two input monitoring elements, and hence the corresponding point clouds are not expected to be separated from the blue one in this figure. However, the point clouds for runs 297288 and 297289 (with visible anomalies in at least one of these two monitoring element types) are clearly separated by this method from the point cloud for run 297056.

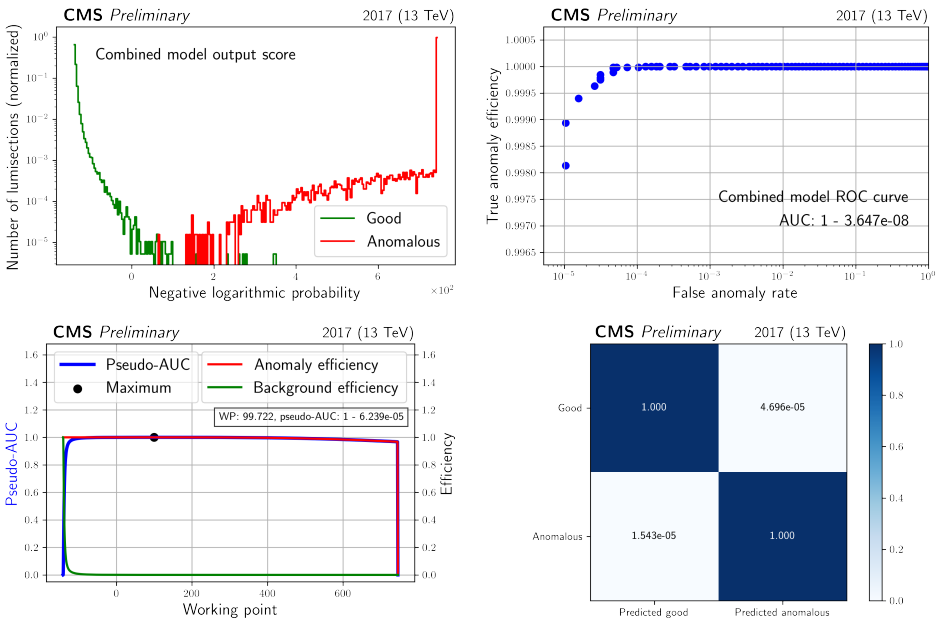


Figure 4.11: Performance metrics for the autoencoder model on the test set. Upper left: distribution of output scores on good luminosity sections (green) and known anomalous luminosity sections (red). Upper right: corresponding ROC curve. Lower left: automatic determination of the optimal working point on the ROC curve. Lower right: confusion matrix for the optimal working point.

In other words, the pseudo-AUC is the rectangular area enclosed by a point on the ROC curve. Maximizing it is one way of choosing the working point, though it should be emphasized that other working point criteria could be used as well, depending on the application. The (few) overlapping output scores between good and bad instances in the test set correspond to luminosity sections from runs 299317 and 299318. They show a particular kind of anomaly in BPIX L2, which the autoencoder is able to reconstruct relatively well. Potentially this type of anomaly is present in the training set in non-negligible quantities, and a more extensive filtering strategy could be attempted as a solution.

Performance on golden json The method with tuned working point is now applied on the luminosity sections in the golden json, which is in principle expected to contain only good luminosity sections (with caveats mentioned in section 4.3.2). We observe O(100) luminosity sections out of $O(2 \times 10^5)$ luminosity sections being tagged as anomalous, corresponding to a naive false alarm rate of about 5×10^{-4}, which is higher than the one found during optimization of the working point on the custom evaluation sets. However, upon closer inspection, the majority of flagged luminosity sections do in fact contain an anomalous distribution for at least one of the monitoring elements. These anomalies have either been missed during the manual certification procedure, or were deemed to be non-critical.

Categorization of anomalies in golden json About half of the flagged luminosity sections in the golden json have a small number of entries, causing large reconstruction errors because of statistical noise while not necessarily having an anomalous shape. This could be remedied in future studies by re-tuning the number-of-entries threshold, or summing consecutive monitoring elements together until a requested level of statistical precision is reached. For example, by increasing the minimum number-of-entries to number-of-bins ratio from 100 to

250, the fraction of flagged luminosity sections decreases from 100 in 2×10^5 to 60 in 1.95×10^5.

The remaining flagged luminosity sections can be categorized as follows:

Table 4.4: Categorization of the flagged luminosity sections (LS) in the golden json.

Visible anomaly, last LS in run
Run 297227 LS 192, run 297488 LS 424, run 299649 LS 332, run 301998 LS 1704, run 305590 LS 1020, run 306051 LS 415, run 306139 LS 1112
Visible anomaly, high voltage test
Run 306458, 47 LS spread across the entire run.
Visible anomaly, not further specified
Run 300389 LS 18, run 300389 LS 19
Other
Run 306139 LS 346, run 306459 LS 1652, run 306460 LS 23, run 306460 LS 32

Discussion Most of the flagged luminosity sections contain visible anomalies, either from being the last luminosity section in the run (e.g. the luminosity section where something goes wrong and the run is stopped) or from a specific run with a high voltage test. There are two luminosity sections with a visible but not yet further identified anomaly. And finally, a small number of luminosity sections that seem to be real false alarms, with distributions that are typically at the boundary of what can be considered normal or acceptable. This remaining false alarm rate is in the order of 4 in 1.95×10^5, consistent (within the large statistical uncertainty on this value) with what was found with the custom evaluation set.

4.3.5 Resampling

In the evaluation as described above, luminosity sections from hand-picked anomalous runs are resampled in order to increase the statistical precision by making multiple copies of a given monitoring element while introducing a reasonable degree of variation with respect to the original one. The varied histogram is defined as the original one with an added variation term, that consists of a collection of random Fourier-series-like components, i.e. sine functions with random amplitude, wave number and phase. For an original distribution $h(x)$, a varied distribution $h'(x)$ is defined as follows:

$$h'(x) = h(x) + \sum_{i=1}^{n} A_i \cdot f(x) \cdot \sin(k_i x + \phi_i)$$

with

- n (the number of sine components) set to 3.
- A drawn randomly from a uniform distribution between 0 and $2/n$.
- $f(x)$ amplitude scaling, chosen as $h(x)$ divided by a scalar.
- k drawn randomly from a uniform distribution between 0 and $\pi/4$. For integer x values (i.e. bin numbers), this ensures a sufficiently large period of at minimum 8 bins.

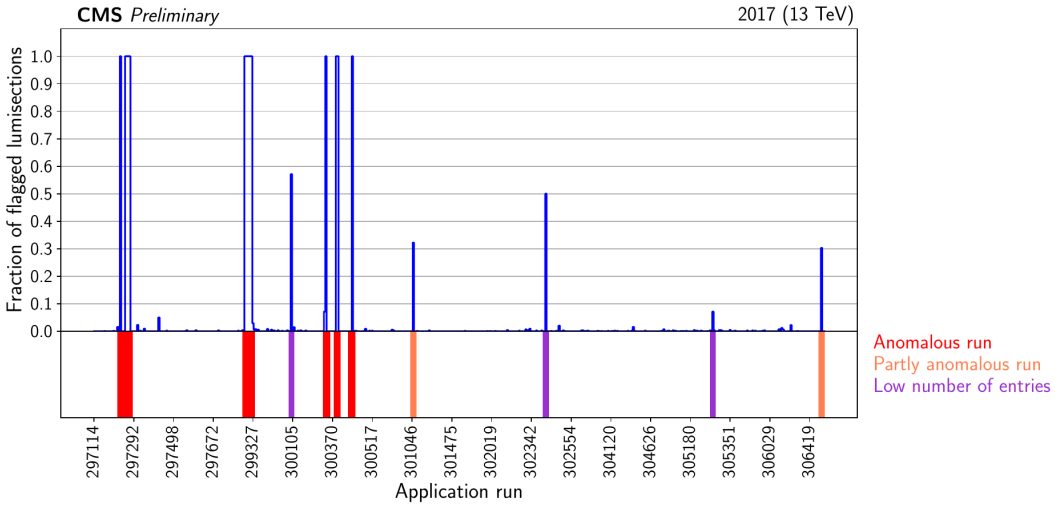


Figure 4.12: Results for local training: automatic flagging of anomalies in an operational approach.

- ϕ drawn randomly from a uniform distribution between 0 and 2π.

This procedure of resampling was found to provide adequate variation while keeping essential shape characteristics on each of the studied monitoring elements. However, it is rather ad-hoc and not motivated by physical causes, e.g. a detailed modeling of the impact of changing detector conditions on these distributions. We revisit this point in section 4.7.

4.3.6 Local and sequential training

An alternative training and testing procedure is also attempted, that represents more closely the operational situation in practical real-time DQM applications of the method. In this so-called local training, for each application run, the model is updated with a dedicated training on the runs preceding the application run. The results, displayed in Fig. 4.12 show that the fraction of flagged luminosity sections is low for good runs and higher for known anomalous or otherwise special runs (indicated with vertical coloured lines in the bottom pad), showing that the model flags anomalous luminosity sections accurately in local training as well as in global training. This opens a promising perspective on applying the method in 'online' DQM.

4.3.7 Comparison to other methods

Several other anomaly detection methods were tested in order to put the performance of the autoencoders into perspective. They are shortly introduced in order of increasing complexity, and the comparison between all methods is discussed at the end of this section.

Moments method This simple method is based on the observation that many anomalies are manifested in the cluster charge monitoring elements in the form of an additional peak at lower charge values. This results in a smaller mean value and larger spread than for non-anomalous distributions. When training this model, the moments of all training distributions are calculated up to some specified order as follows:

$$m_i = \frac{\sum_j b(j) \times j^i}{\sum_j b(j)}$$

with m_i being the i'th order moment, j the index running over the bins in the monitoring element and $b(j)$ being the corresponding bin content. The mean μ_i and standard deviation σ_i

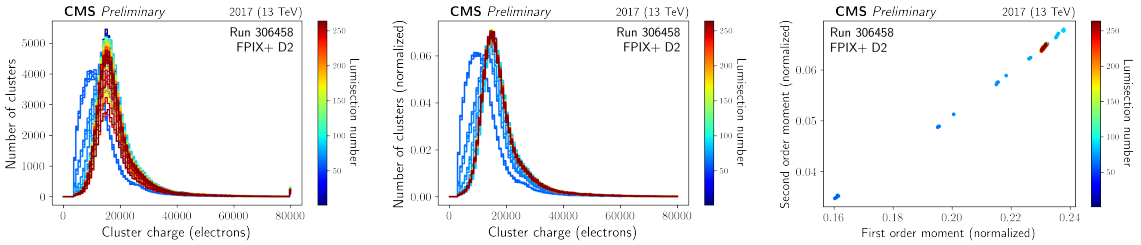


Figure 4.13: Illustration of the moments method on run 306458 in FPIX+ D2. A fraction of the luminosity sections show distributions that are shifted to the left with respect to the expected shape, which reflects in first and second order moments distinct from the bulk. Left: raw monitoring elements, middle: normalized monitoring elements, right: moments. Note that the moments are calculated after rescaling the range of the x-axis to (0,1) as implicit normalization.

for each order moment over the training set are calculated. A given distribution is evaluated by computing the amount of standard deviations between its moments and those of the training set (sum-squared over all orders):

$$\text{anomaly score} = \sum_i \left(\frac{m_i - \mu_i}{\sigma_i} \right)^2$$

The model performance is observed not to increase significantly after including the second order moment. Figure 4.13 illustrates the idea of the method.

This model has the advantage of being conceptually simple and deterministic in nature. However, it is not able to accurately distinguish subtle shifts in moments due to small anomalous peaks from the expected shifts in distributions due to changing detector conditions.

Landau fit method The Landau fit method fits an analytical approximation for the Landau distribution to the observed monitoring elements. In more detail, the fitted distribution is:

$$L(p, w, n) = \frac{n}{\sqrt{2\pi}} \exp\left(-\frac{y + \exp(-y)}{2} \right) \text{ with } y = \frac{x - p}{w}$$

In this expression, the parameter p represents the peak position, w its width and n the overall normalization. The mean-squared-error (MSE) between a monitoring element and its fitted distribution is used as a measure for anomality. In extension, the Landau shape can be convoluted with a Gaussian distribution, but this is not observed to increase the performance significantly. The method is illustrated in Fig. 4.14.

The particular advantage of this model is that it does not require training; it can be directly applied on any monitoring element presented to it. On the downside, the theoretical Landau shape is a rather coarse approximation for the observed monitoring elements, and the model has difficulty in distinguishing bad instances from good (but not perfectly Landau-like) instances.

Templates method This method is inspired by manual certification procedures, where monitoring elements to be judged are compared to reference histograms from known good runs. In our application, the choosing of reference histograms is automated by partitioning the training set into subsets of consecutive luminosity sections, and averaging the monitoring elements

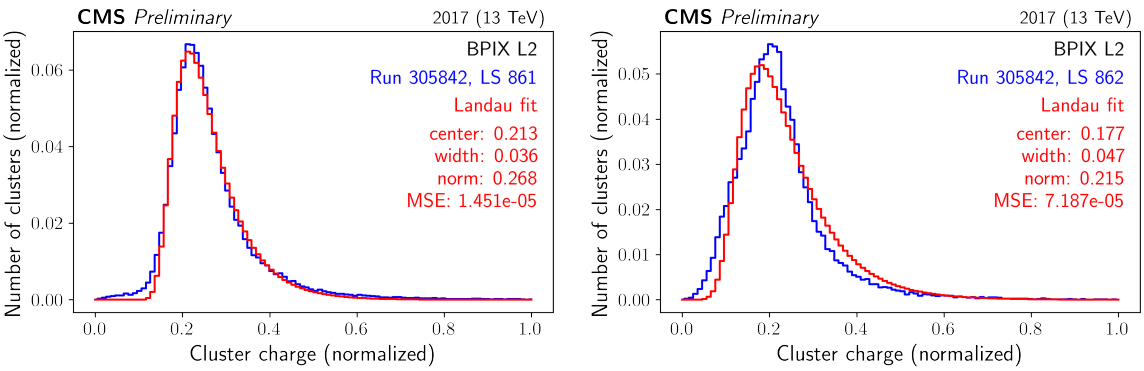


Figure 4.14: Illustration of the Landau fit method on the last two luminosity sections of run 305842 in BPIX L2. Left: good luminosity section, right: anomalous luminosity section. The best-fit values of the parameters (as explained in the text) are indicated, as well as the resulting MSE, which is seen to be significantly higher for the anomalous distribution than for the good one.

within each partition to a single reference histogram. This partitioning and averaging intends to average-away isolated anomalies while still retaining gradual shape shifts due to changing detector conditions. When evaluating this model, the MSE between the test monitoring element and all reference histograms is calculated and the minimum is chosen as anomality score.

The interpretability and similarity to manual certification of this method is a benefit. On the other hand, the details on how to select or construct the reference histograms adds some complexity to the model while it is still outperformed by the NMF and autoencoder methods. However, rather than as a final model in itself, it can be used as a filter in the preprocessing step to obtain cleaner training and/or testing sets (see section 4.3.2).

NMF method This method employs non-negative matrix factorization (NMF) [149, 150] to construct an optimized set of basis components for the monitoring elements in the training set. Upon evaluation, the presented monitoring element is reconstructed as closely as possible with a linear combination of the basis components. The MSE between the original monitoring element and its reconstruction is chosen as an anomality measure. The motivation and operation of this method is very similar to the autoencoder method discussed in greater detail before, the only difference being that the reconstruction of a monitoring element by means of an autoencoder is replaced by the NMF reconstruction. The method is illustrated in Fig. 4.15.

Discussion Figure 4.16 summarizes the comparison between the anomaly detection methods. The separation between the output scores for good and anomalous luminosity sections can be seen to increase when going from more simple methods to more involved ones, with the autoencoder performing best.

4.4 Anomalous cluster occupancy in the pixel endcaps

In this section, we discuss a second case study of automated anomaly detection using autoencoders. An important difference with respect to the previous case study (discussed in section 4.3), is that the monitoring elements under study are two-dimensional spatial occupancy maps instead of one-dimensional summary distributions. This requires a more involved

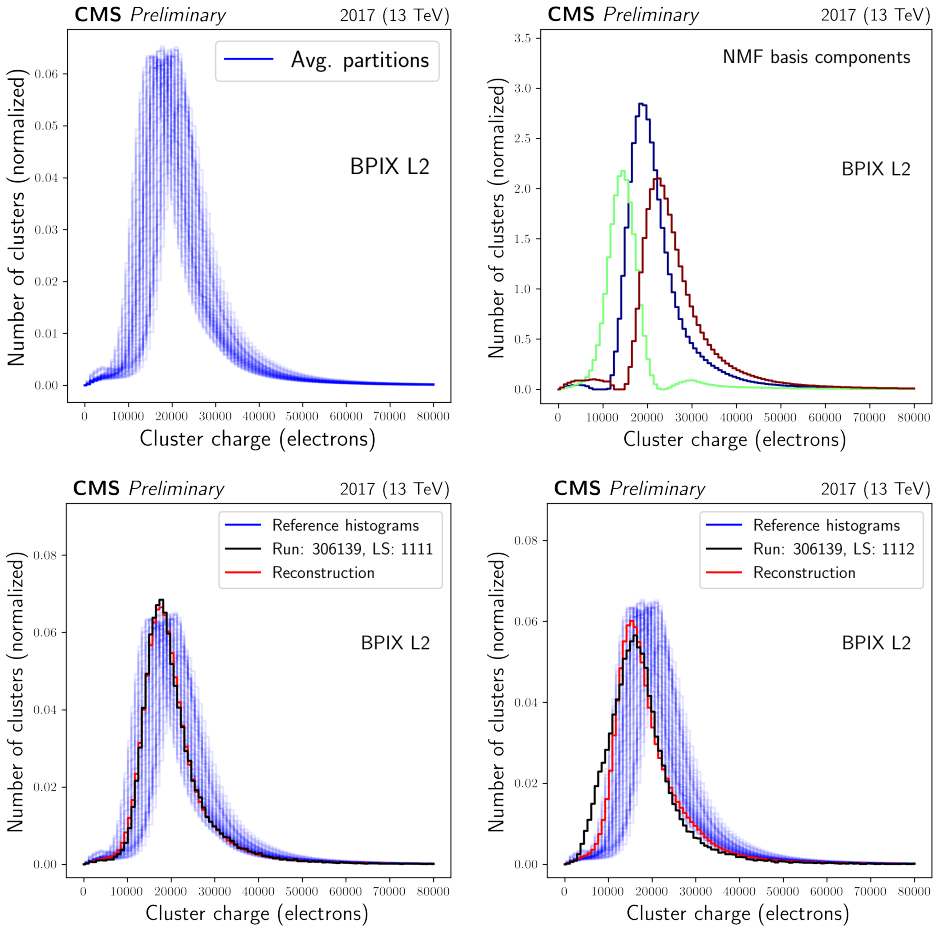


Figure 4.15: Illustration of the NMF method. Upper left: example good monitoring elements to give an idea of the expected shapes. Upper right: NMF basis components extracted from the training set. Lower left: example of a good luminosity section and its NMF reconstruction. Lower right: example of an anomalous luminosity section and its NMF reconstruction.

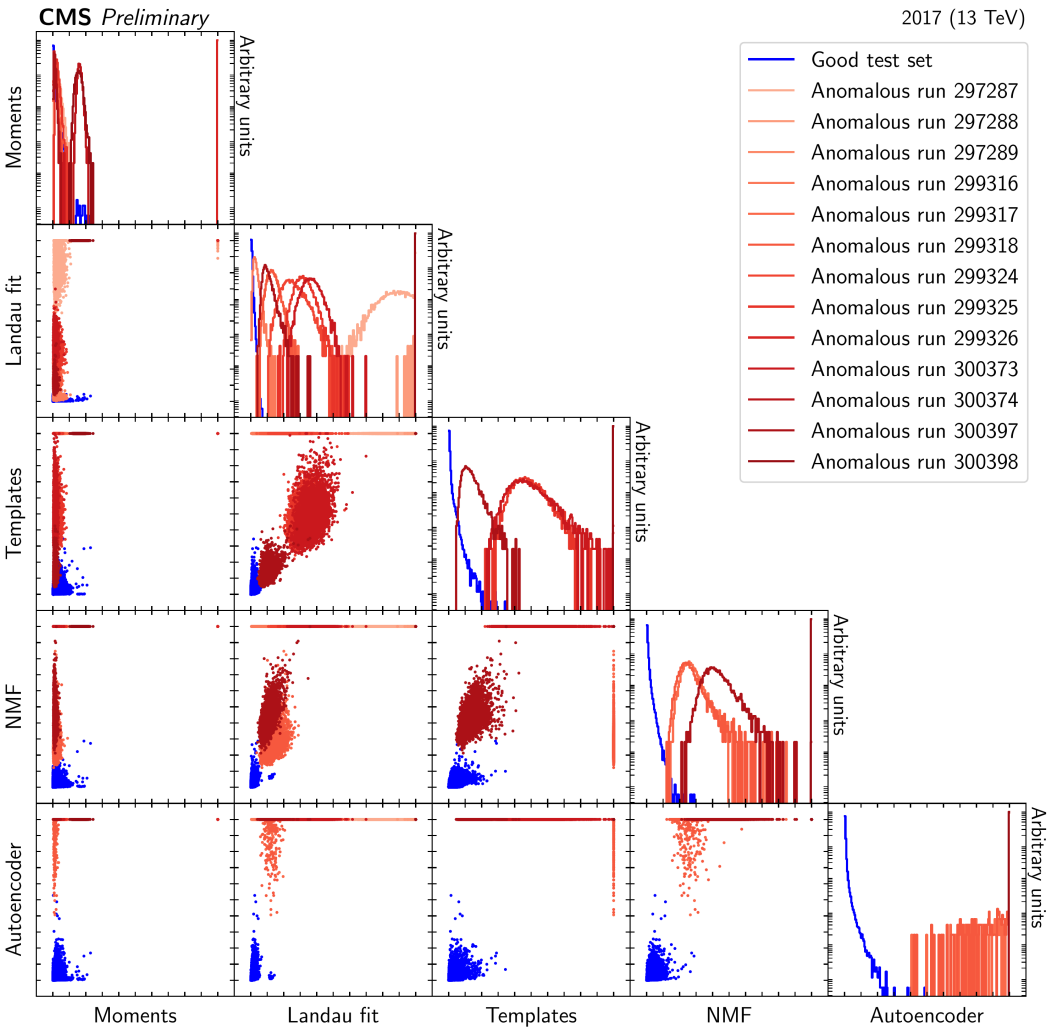


Figure 4.16: Output score distributions (diagonal) and correlations between the anomaly detection methods (below diagonal). The pads on the diagonal line display the distributions of the final output score for a given anomaly detection method, with the good test set in blue and the runs forming the anomalous part of the test set in shades of red and orange. The distributions have been normalized to unit surface area and the y-axis scale is logarithmic. The pads under the diagonal show the same distributions, but in a two-dimensional representation. Each dot represents one luminosity section with its assigned score according to one model on the y-axis, and according to another model on the x-axis. The horizontal and vertical arrays of points correspond to lower bounds on the fitted probability density where it cannot be numerically distinguished from zero.

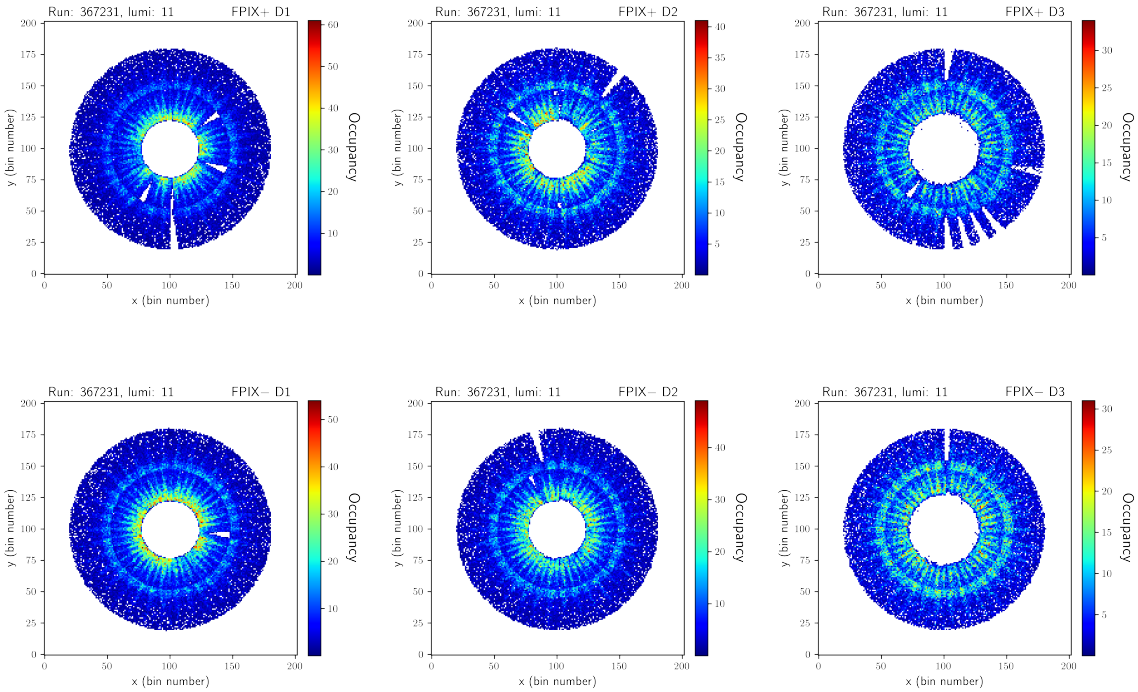


Figure 4.17: Examples of the monitoring elements used in this study. The distributions show the occupancy of reconstructed cluster positions as a function of spatial coordinates transverse to the beamline. The same luminosity section (run 367231, luminosity section 11) is shown for the 6 pixel endcap disks (+1 to +3 on the upper row, -1 to -3 on the lower row).

autoencoder architecture, with a convolutional structure. It should be noted that this study is work in progress at the time of writing, and there are a few loose ends that need more investigation.

4.4.1 Monitoring elements and data set

We study the reconstructed position of clusters of electric charge in the endcap disks of the pixel tracker, as a function of the x and y coordinates perpendicular to the beamline[7]. All 6 endcap disks were included in this study, but the barrel layers were not, as they have been covered by other studies already (see section 4.5).

The position of reconstructed clusters on tracks is sensitive to a wide range of operation issues that could affect the data quality, but especially to problems with the front-end drivers, power supply, and token bit managers (managing the readout chips), all of which typically result regions with zero occupancy in the region of space covered by the particular problematic component[8]. Some representative examples of these monitoring elements are shown in Fig. 4.17.

[7] In CMS technical jargon, the monitoring elements are called PixelPhase1/Tracks/PXForward/clusterposition_xy_ontrack_PXDisk_*, where * ranges from -3 to -1 and from +1 to +3.

[8] There are other monitoring elements that provide cluster occupancy binned according to electronic channel rather than spatial coordinates; this will be revisited in section 4.7.

The data used in this study was recorded by the CMS detector in 2023. We use the prompt reconstruction, that is available within O(48h) after the data has been recorded[9]. The majority of all 2023 data has been used[10], with the exception of a section of data-taking in the beginning of the year, when per-LS storage was not yet enabled (see section 4.6 for more details).

4.4.2 Preprocessing

We apply a rather extensive set of preprocessing operations, with the goal of making the data more robust against statistical fluctuations and of reducing the computational cost of training and evaluating an autoencoder. These steps are shown on an example in Fig. 4.18.

Normalization We divide each monitoring element by the pileup for the corresponding luminosity section. The integrated occupancy of each of the monitoring elements under study was observed to have an approximately linear behaviour as a function of pileup; hence the division by pileup results in an approximately constant integrated occupancy per luminosity section (for non-anomalous monitoring elements) and serves as normalization. Deviations from the expected linear trend can be attributed to special run conditions or trigger menus.

Averaging The occupancy in the outer rim of the monitoring elements is rather low and hence subject to relatively large statistical fluctuations. To mitigate these, a running average is taken over monitoring elements of consecutive luminosity sections. For a given monitoring element, the n preceding luminosity sections are used in the averaging, where n manages the trade-off between statistical significance and time granularity. It is set to 2 in the inner ring and 5 in the outer ring.

Rebinning The monitoring elements are rebinned (by averaging blocks of 5x5 bins) to reduce the computational cost of further processing. In this rebinning procedure, original bins with zero occupancy are exactly propagated to the merged bins rather than being averaged out with neighbouring non-zero occupancy bins. This is needed to retain sensitivity to small zero-occupancy regions[11].

4.4.3 Autoencoder architecture

The autoencoder architecture used in this study is based on similar networks used for studies with the pixel barrel and with the ECAL barrel and endcaps. It comprises the principles of residual connections [151] and convolutional layers [143]. The latter are often used for the processing of data types with some form of spatial locality and translational invariance (such as images); the former are introduced to make the training of deep networks more stable as the network can focus on learning residuals with respect to the identity transform. A sketch of the architecture used in this study is shown in Fig. 4.19. The model is trained with the mean-squared-error loss function, the 'adam' optimizer [148], a batch size of 32, 30 epochs, and a validation split of 10%. The training is stopped if the validation loss does not decrease significantly for more than 10 epochs, and the best performing snapshot of the model during

[9] Typically, the prompt reconstruction is replaced, at least for physics analysis purposes, by re-reconstruction versions at the end of the data-taking year or later. The prompt reconstruction is used here both because it was the only version available at the time of this study, and because it corresponds better to what would be used during live data-taking

[10] In CMS technical jargon, we use the /ZeroBias/Run2023*-PromptReco-v*/DQMIO where * represents C (v1 – v4), D (v1 – v2), E (v1) and F (v1).

[11] Admittedly, small regions with anomalously low but non-zero, or with anomalously high occupancy, might still be averaged away by this procedure. This is revisited in section 4.7

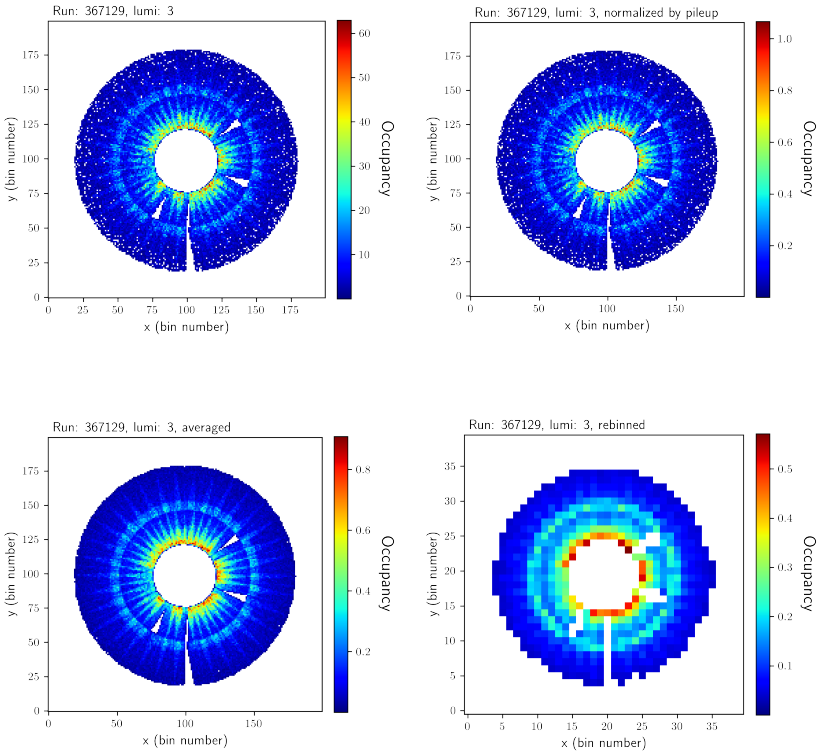


Figure 4.18: Illustration of the preprocessing steps on a single monitoring element. Upper left: raw monitoring element. Upper right: normalized by pileup. Lower left: averaged over preceding luminosity sections. Lower right: rebinned. The figure is for run 367129, luminosity section 3, pixel disk +1, but is representative for other luminosity sections and monitoring elements as well.

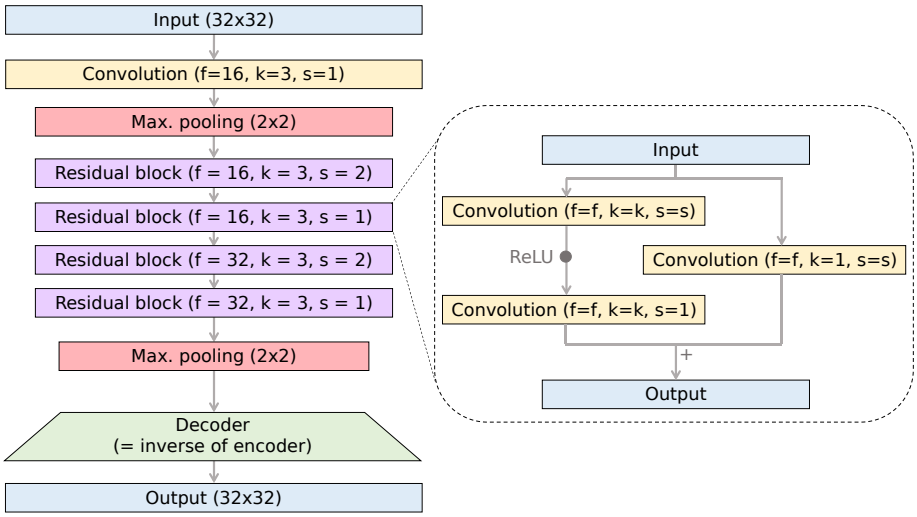


Figure 4.19: The model architecture of the autoencoder used in this case study. Left: full model layout. Right: detailed view of one residual block. The parameters of the convolutional layers and residual blocks are defined as follows: f = number of filters, k = kernel size, s = strides.

training (not necessarily the last one) is saved.

The advantage of the preprocessing, especially the rebinning, becomes clear when training the network, which typically only takes about 3 hours on a regular single-CPU machine when using the full 2023 data as training set (and much shorter when using only a part). This is in contrast with, for example, a similar study on the barrel layers instead of the endcap disks (see section 4.5), where training on only a fraction of the data typically took multiple hours on a GPU cluster. On the other hand, the potential bias introduced by the rebinning should be investigated in more detail in a continuation of this study.

4.4.4 Post-processing

When applied on a single monitoring element, the autoencoder provides as an output a reconstructed image that, ideally, is almost identical to the input image, except for any anomaly present in the input. Hence, the bin-per-bin squared difference between the input and the output of the autoencoder, typically called a loss map, is a suitable anomaly measure. However, to enhance the performance of the anomaly detection, several post-processing steps are applied on the loss maps, summarized in Fig. 4.20. These techniques are heavily inspired by similar work in ECAL.

Spatial correction From the fact that the average occupancy decreases systematically with the radial distance from the center of the disk, it follows that the same holds for the squared error. This causes a non-uniformity in the loss map, which can be corrected for by dividing it by the occupancy map averaged over a long data-taking period.

Time correction Many 'anomalies' or other fluctuations appear only for a short time (≤ 1 luminosity section) and can safely be ignored. To increase the robustness of the model against such transitory phenomena, for each loss map a running average is taken with the 2 preceding luminosity sections.

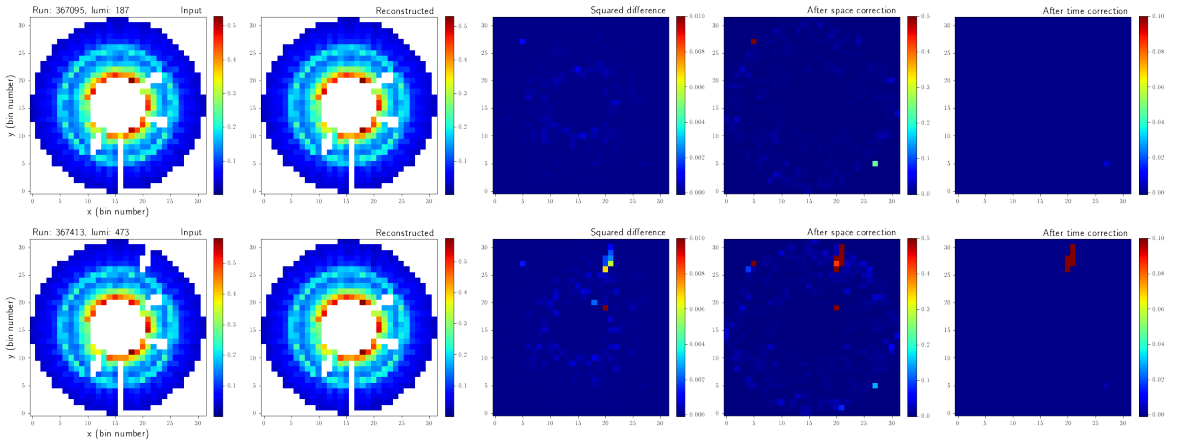


Figure 4.20: Example of a monitoring element with (upper row) and without (lower row) anomaly. Leftmost column: input data. Second column, autoencoder response. Third column: squared difference between input and response. Fourth column: after applying the spatial correction. Rightmost column: after applying the time correction. The anomaly (in the lower row) is correctly and clearly identified, while the input without anomaly (in the upper row) shows low loss.

Masking Known zero-occupancy regions in the input distributions are masked in the loss map (i.e. set to have zero loss). This is done in order not to raise constant alarms for already known issues, and rather focus on newly appearing problems. In principle, masking is not necessary if the autoenecoder is both trained and evaluated on monitoring elements where these known problems are present, as the model will learn to reconstruct the behaviour of these issues; but it is needed when the training set and evaluation set have a different set of known problems.

4.4.5 Results

Definition of the training set As was the case in section 4.3, a proper definition of a training (and more importantly, a testing set) is a major point of ambiguity. For the training set, we use the 2023C-v1 data with some selections applied that are summarized in Tab. 4.5. It is important to note that, just as in section 4.3, none of these selections rely on earlier manual certification work. The requirement on the number of entries is applied before normalizing to the pileup and used to remove monitoring elements with large statistical fluctuations. The first 5 luminosity sections in every run are not used because the averaging preprocessing cannot be properly performed there. Finally, we veto monitoring elements that have zero-occupancy regions of size (at least) 2x2, 3x1 or 1x3 (except when they overlap with known zero-occupancy regions). We note that, up to now, the discussion was applicable to each of the 6 monitoring elements considered in this study. However, an autoencoder is trained on each of them separately, and the performance tests discussed below are focused on the first positive endcap disk (FPIX+ D1) only. We return to all monitoring elements in section 4.4.6.

Testing without labels As a first check, we apply the trained model on the unlabeled data with minimal selections (see Tab. 4.5). We apply a gradually increasing threshold and check the number of cells in the loss map for each monitoring element that are above the threshold. This results in the distributions shown in Fig. 4.21. By focusing on the peak and tail of these distributions, some examples without and with anomalies can be isolated, shown in Fig. 4.21 as well. This procedure shows promising results, where anomalies are accurately flagged and good instances generally have a low loss map.

Table 4.5: Definition of the training and testing sets. The left column details the selection applied. The middle column shows the number of remaining luminosity sections (LS) after applying the selection (exclusive with respect to other selections). The right column shows to what fraction of the total that number of remaining luminosity sections corresponds.

Selection	Number of LS	Fraction (%)
Run 2023C-v1	18065	100
Training set		
Number of entries $> 1 \times 10^4$	16503	91.4
Veto first 5 LS in every run	17794	98.5
Veto dead regions	8210	45.5
All of the above	8081	44.7
Testing set (unlabeled)		
Number of entries $> 1 \times 10^4$	16503	91.4
Veto first 5 LS in every run	17794	98.5
All of the above	16359	90.6
Testing set (good instances)		
Number of entries $> 1 \times 10^4$	16503	91.4
Veto first 5 LS in every run	17794	98.5
Veto dead regions	6417	35.5
All of the above	6303	34.9
Testing set (fake anomalies)		
Random good instances with fake anomalies	6000	-

Definition of test set To make the results of the previous paragraph more quantitative, a test set of good and bad instances needs to be defined. A pure set of good instances is isolated by applying the same selections as for the training set, but with a stricter veto on zero-occupancy regions (vetoing patterns of (at least) 2x1 and 1x2). Bad instances are drawn as random instances from the good testing set, with the introduction of fake anomalies in the form of randomly positioned and sized zero-occupancy regions. This procedure is likely too optimistic and should be revisited (see section 4.7), but serves as a baseline check of the functioning of the model.

Testing with labels The results of the evaluation on the test set discussed above are shown in Fig. 4.22. The performance is essentially perfect, but unfortunately the interpretation is not very straightforward. First of all, the ad-hoc definition of fake anomalies does not necessarily represent realistic failure scenarios. Secondly, the definition of the good instances for testing, with a strict veto on dead regions, is problematic as it only represents a small part ($\pm 35\%$) of the data. Hence the model flags large parts of the data as anomalous. On the other hand, transitory zero-occupancy regions are in fact present in these flagged monitoring elements. Hence this can be interpreted as not a failure of the method, but as an invitation for rethinking how to define actual anomalies that need to be flagged. Perhaps, a simple increase of the time correction window will already improve the situation, but this remains to be seen. Despite these caveats, the method is shown to accurately catch anomalous zero-occupancy regions.

4.4.6 Local training

The results of the previous section are now generalized to all pixel endcaps and all data taking periods within 2023 data taking. As before, we use the same data-taking period for training and evaluation, but in addition, we pursue a more operational approach where the models used for evaluation on a given data-taking period are trained on the previous data-taking period. The combined results are summarized in Fig. 4.23. For most data-taking periods and monitoring elements, the performance of the method is good both when the training is performed on the same data or on previously taken data. The exceptions to this rule can be traced down to the following causes:

- **Failed trainings** Insufficiently trained models give inaccurate reconstructions for both good and anomalous instances. This happens in particular for some monitoring elements in the 2023D-v1 period, and can likely be remedied by changing the training parameters, or, in case of too few available training instances, by merging this data-taking period with other periods.
- **False alarms** Genuine false alarms can be caused by non-negligible and systematic differences between monitoring elements of different data-taking periods. This is a more tricky issue, but updating the model training with the beginning of the next data taking period could potentially help in these cases.

This is the current status of this case study. Several particular points of improvement or continuation have already been identified in the text. A more general discussion on the prospects of this approach is postponed to section 4.7.

4.5 Related work

Several other CMS subsystems are performing similar studies to the ones presented in this chapter, as well as auxiliary work to facilitate, streamline and synchronize them. A short

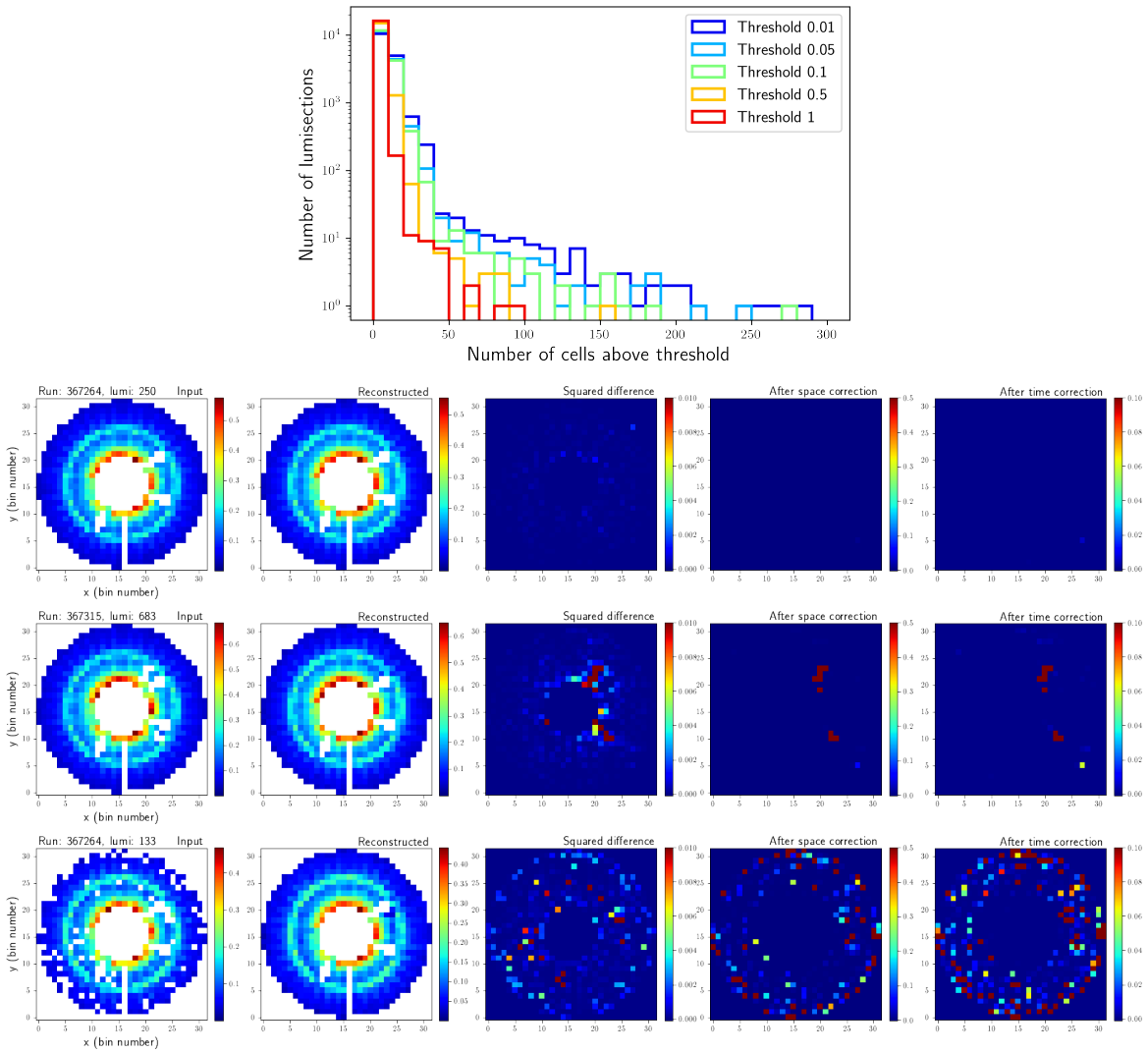


Figure 4.21: Qualitative evaluation of the anomaly detection method. The upper figure shows the distribution of number of cells per loss map (corresponding to a monitoring element) that are above a given numerical threshold (with several threshold values shown in different colors). The three lower rows show examples, obtained by isolating monitoring elements from the left, middle and right part of the upper graph. It is observed that cells above threshold in the final loss map (on the right) are correctly coming from anomalies in the input data (on the left).

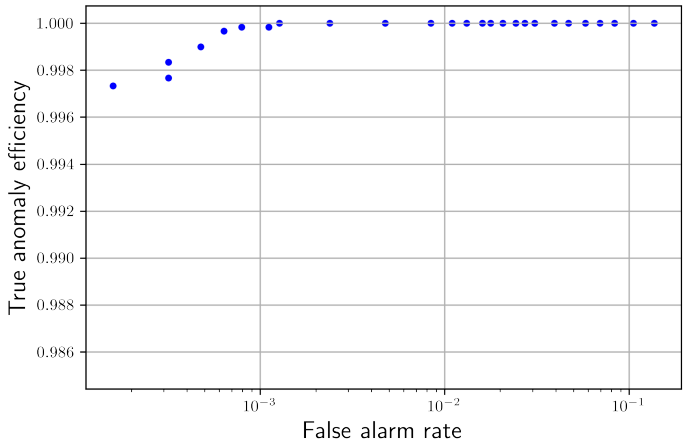


Figure 4.22: True anomaly tagging efficiency versus false alarm rate (aka the ROC curve) of the anomaly detection method using an autoencoder. The false alarm rate is calculated on a set of real monitoring elements with a filtering applied to remove problematic zero-occupancy regions, and the anomaly efficiency is calculated by introducing fake anomalies in the data, as explained in more detail in the text.

overview is given in this section.

ECAL The electromagnetic calorimeter (ECAL) in CMS conducted studies using autoencoders on occupancy monitoring elements in both the barrel and endcaps at ECAL tower level filled with digitized hit counts [152, 153]. Typical anomalies consist of dead towers (registering no hits), hot towers (registering anomalously high hit counts) and missing supermodules (collection of dead towers), all of which are accurately caught by the autoencoder. Of special interest is the extensive preprocessing (masking of persistent issues and normalization with respect to pileup) and post-processing (spatial and temporal response correction).

HCAL The hadronic calorimeter (HCAL) in CMS has conducted similar studies on digitized hit occupancy maps per channel [154]. The model consists of a variational autoencoder with builtin convolutional, recurrent and graph neural networks. In this study, special care is given to capture time-dependent anomalies such as slowly degrading channels. Furthermore, the several layers of the HCAL are stacked together in a three-dimensional data structure and processed by a single global network rather than having separate networks for each of the layers.

JetMET The JetMET group (studying the performance of jet and missing transverse energy reconstruction) developed an autoencoder-based tool for spotting anomalous distributions [155]. Application of this method in 2022 data allowed the recovery of two runs (about 350 pb^{-1} of data), that were marked as bad before, but in which the observed anomaly turned out to be limited to single luminosity sections. The JetMET group also conducts machine learning studies for certification using higher-level scalar input features with a per-run granularity and with both supervised and unsupervised methods [156].

Pixel tracker: barrel occupancy Another anomaly detection study is being performed for the pixel tracker in CMS [11,13,157]. In this case, the input monitoring elements consist of the reconstructed cluster position in the barrel layers of the detector (rather than in the endcaps

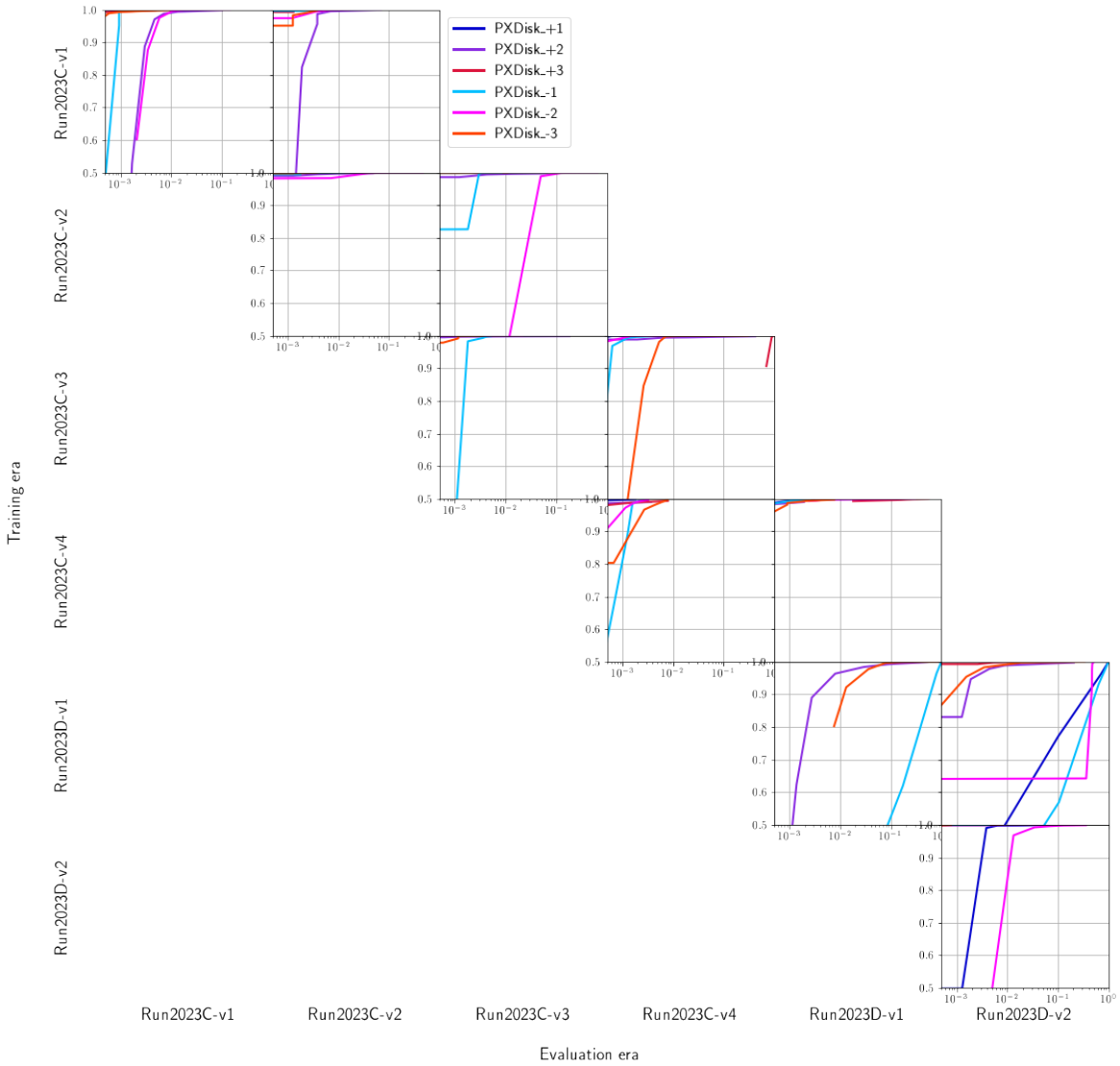


Figure 4.23: Summary of the results of the anomaly detection method with autoencoders for all monitoring elements and data-taking periods considered in this study. Each individual panel shows 6 ROC curves, corresponding to the cluster occupancy monitoring elements of the 6 layers of the pixel endcap disks (separated by different colours). Each panel corresponds to a different combination of training set (vertical direction) and evaluation set (horizontal direction). For each evaluation set, the method is tested with training on the same set (diagonal panels) and on the previous set (above diagonal panels).

as in section 4.4). Of special interest is the usage of principal component analysis to combine the output scores of several monitoring elements.

ML Playground / DIALS Infrastructural works include the development of DIALS (Data Inspector for Anomalous Luminosity Sections, formerly known as ML Playground), a web service intended for hosting the per-LS monitoring elements in an easily accessible format, for data exploration, and potentially for streamlining model training and evaluation [13]. This tool, in combination with the per-LS DQMIO data tier (see section 4.6), will greatly simplify accessing the per-LS monitoring elements, which was one of the biggest bottlenecks in these studies so far.

AutoDQM The AutoDQM tool[12] is used for running a collection of statistical tests to compare a set of monitoring elements from a query run to a predefined reference run. Recently, the tool was extended with machine-learning-based comparisons.

4.6 The per-LS DQMIO dataset

An important aspect of these studies that was not yet sufficiently emphasized is the technical availability of the input data. During the course of these studies, I contributed to the effort of figuring out the best way to make per-LS monitoring elements readily available in a standardized format, and to developing tools for reading and exploring these data. This resulted in the per-LS DQMIO dataset[13]. This is an extension of the existing DQMIO dataset (containing monitoring elements stored per run in a custom ROOT format), where a subset of monitoring elements is stored at the per-LS level. The subselection is managed by an easily modifiable configuration file, so new monitoring elements can be added on demand. These datasets are centrally available via DAS[14] for all CMS users. An earlier version of this approach was piloted in the official reprocessing of 2022 data, with the current mature implementation being fully integrated in central CMS production since early 2023 data taking.

4.7 Conclusion and outlook

The studies reported in this chapter show promising prospects for automatic anomaly detection methods assisting the data quality monitoring (DQM) and data certification (DC) procedures in the CMS pixel tracker. With a simple autoencoder model, anomalies in the cluster charge distributions for the pixel barrel layers and endcap disks can be efficiently detected. The autoencoder method is shown to perform better than several benchmark methods including direct comparisons to reference distributions. This section presents some thoughts on the next steps and on potential improvements to the studies discussed in this chapter.

DC vs. DQM The methods presented here have been mainly used from a DC point of view, and more specifically as an a posteriori approach where large data samples are available for training, with the intention to spot anomalies in that same dataset, that had been missed during live data taking. A real-world application of this approach would be a dedicated re-certification campaign at the end of a data-taking period of one or multiple years. The next step is to implement these methods in online DQM, to raise alarms in real time when anomalies occur. This chapter includes successful proofs-of-concept of this approach, by doing an

[12] `https://autodqm-official.readthedocs.io/en/latest/index.html`

[13] For a full description of the details, history, and technicalities, see this TWiki page (digital link) (CMS internal)

[14] `https://cmsweb.cern.ch/das/` (CMS internal)

artificial replay of historical data and updating the autoencoder training as the data taking progresses. However, more studies are needed to guarantee the low false alarm rate and high anomaly efficiency in practice, and if needed to re-tune the parameters of the method. Apart from that, the remaining work to achieve that goal should be mainly technical in nature, implementing the method in the CMS data-taking software.

Number of monitoring elements The number of monitoring elements studied in this chapter is very small compared to those typically looked at during shifts and certification. It is as yet not very clear by how much this could be scaled up, i.e. how much additional studies and fine-tuning would be needed to apply these methods to other monitoring elements not yet considered so far. Also the feasibility of the extra computational load when training and evaluating autoencoders for a large number of monitoring elements has not yet been assessed. In the most likely scenario, automatic anomaly detection methods will remain limited to a small subset of monitoring elements, carefully chosen as to catch as many anomalies as possible, with the remaining monitoring elements being looked at occasionally by subsystem experts for diagnosing the exact problem once an anomaly has been spotted by the automatic method.

Evaluation One of the fundamental difficulties in the studies presented here, is the ambiguous and arbitrary definition of anomalies for the evaluation of the methods. Reliable labels at the per-LS and per-monitoring-element level are lacking, hence self-consistent methods or artificial evaluation sets have to be used. This is the weak spot of the performance reporting of these methods. The easiest way forward is to carefully start applying the automatic methods in the DQM and DC procedures, in close contact with experienced subsystem experts, to estimate their performance in real applications, and from there to design a good benchmark evaluation procedure for future improvements.

Resampling A more specific area of improvement is the resampling method as described in section 4.3.5. The current approach adds some random smooth variation to existing monitoring elements, but bin-per-bin statistical variations are not taken into account. Nor is this form of adding variation guaranteed to be physically meaningful or representative for the expected shapes of the distributions in question. This is a typical case of 'it seems to work good enough for now, but I might revisit it more thoroughly later', which is then never revisited because of lack of time. Future studies could attempt to develop a more mathematically sound method of resampling, as well as more realistic variations (or artificial anomalies).

Normalization In the study using cluster charge distributions (see section 4.3), the monitoring elements were normalized to unit surface area to focus on shape comparison rather than normalization effects. With this procedure, there is a remaining non-uniformity in the bins of each monitoring element (meaning that the peak has a higher event count than the tails), which can lead to a non-uniform sensitivity to anomalies depending on where in the monitoring element they appear. Similar arguments led to the spatial response correction in the cluster occupancy study (see section 4.4). As an alternative solution (as used by the JetMET group) the bins could be normalized independently from each other, such that the event counts in each bin follow a normal distribution across the entire dataset.

Cluster occupancy monitoring elements Another specific topic of improvement is the choice of monitoring elements for the cluster occupancy in the pixel tracker. As explained in section 4.4.1, the monitoring elements used so far contain the reconstructed cluster position as a function of spatial coordinates. However, for quickly identifying the exact faulty component of the pixel tracker causing a given anomaly, it would be better to have a cluster occupancy map as a function of electronic channel rather than spatial coordinates. As a matter of fact,

these monitoring elements exist[15], but were not yet available per luminosity section at the time of this study. I updated the per-LS DQMIO configuration in order to enable per-LS storage for these monitoring elements, starting from 2024 data taking, to be used for further studies in this direction. Another advantage is that these monitoring elements have a smaller number of bins than the ones currently used, hence likely removing the need for controversial rebinning procedures. Moreover, they have a rectangular topology, which is likely to further simplify the autoencoder training with respect to the circular geometry of the monitoring elements used so far.

[15] In CMS technical jargon, they are called
PixelPhase1/Phase1_MechanicalView/PXBarrel/
clusters_per_SignedModuleCoord_per_SignedLadderCoord_PXLayer_*
(with * ranging from 1 to 4) and
PixelPhase1/Phase1_MechanicalView/PXForward/
clusters_per_SignedDiskCoord_per_SignedBladePanelCoord_PXRing_*
(with * ranging from 1 to 2).
Or equivalently, using the number of digitized hits instead of cluster occupancy:
PixelPhase1/Phase1_MechanicalView/PXBarrel/
digi_occupancy_per_SignedModuleCoord_per_SignedLadderCoord_PXLayer_*
(with * ranging from 1 to 4) and
PixelPhase1/Phase1_MechanicalView/PXForward/
digi_occupancy_per_SignedDiskCoord_per_SignedBladePanelCoord_PXRing_*
(with * ranging from 1 to 2)

Chapter 5

Theoretical and experimental status in top quark physics

This chapter aims to provide a short overview of the current status and open questions in theoretical and experimental top quark physics. It does not feature my own work, but rather a selection of relevant results from the literature, mainly to serve as a broader context in which to regard the more specific chapters that follow. More extensive and general overviews can be found for example in Refs. [158, 159], and for CMS specifically in Ref. [160].

5.1 Introduction

The top quark is the heaviest known elementary particle in the standard model[1]. Several of its properties make the top quark an interesting object of study for various reasons.

Lifetime The top quark is the only type of quark whose lifetime (5×10^{-25}s) is shorter than the typical time scale of hadronization ($10^{-24} - 10^{-23}$s), therefore it "decays before it hadronizes" [158, 161]. This implies that no hadrons containing top quarks or (meta-stable) $\mathrm{t\bar{t}}$ quarkonium bound states can be formed[2], and that the decay of the top quark is a very pure

[1] The mass of the top quark cannot be completely unambiguously defined, as explained in section 5.4.1, but it is in the order of 173 GeV.

[2] With the exception of $\mathrm{t\bar{t}}$ production close to the threshold energy of two times the top quark mass, in which case the elusive and yet to be discovered 'toponium' (an extremely short-lived bound state of a top quark and antiquark) is predicted to have a non-negligible impact [162, 163].

and relatively simple process, the kinematics and dynamics of which are not complicated by the non-perturbative QCD interactions that affect hadronic bound states.

Decay modes The standard model predicts the top quark to decay almost exclusively to a W boson and a b quark, the non-diagonal elements in the CKM matrix being close to zero for the t-row [161, 164]. Due to this simple decay structure, potential anomalous decays could be relatively easily and unambiguously detectable. Moreover, being heavier than the W boson and b quark together, the top quark is the only type of quark to decay to a real (on-shell) W boson.

Coupling to the H boson The top quark plays a special role in combination with the Higgs boson. The Yukawa coupling of the top quark to the Higgs field is in the order of unity, which is much larger than that of any other particle in the standard model (this is a translation of the large mass of the top quark) [161]. Therefore, potential beyond-standard-model (BSM) effects arising in the Higgs sector are likely to affect top quark physics as well.

We start with a discussion of various top quark production mechanisms at hadron colliders in section 5.2. In top quark physics, as in most of high energy particle research, the most extensively studied quantity is the cross-section of a process, which is a measure for its probability (see section 2.1). Cross-sections of processes involving top quarks have been predicted and measured both inclusively and differentially, and these will be the topics of section 5.3. Other properties of the top quark will be discussed in section 5.4. In section 5.5, we touch upon the subject of effective field theory as general parameterization of deviations from the standard model. Finally, section 5.6 lists some of the remaining questions and riddles involving the top quark.

5.2 Top quark production mechanisms at hadron colliders

Due to its large mass, only a limited number of high-energy particle colliders have direct experimental access to the top quark. The discovery of the top quark took place in 1995 by the CDF and D0 Collaborations at the Tevatron collider [165, 166]. After the decommissioning of the latter, the LHC is currently the only facility where top quarks can be produced, with the general-purpose detectors CMS and ATLAS performing most of the current experimental top quark research.

There are several mechanisms for top quark production at hadron colliders: $\mathrm{t\bar{t}}$ production, single top-quark production in the s-channel, t-channel and tW-channel, and associated productions involving top quark(s) produced together with other particles. The most important properties of these mechanisms are discussed in the paragraphs below.

Top quark pair production The production of a pair of a top quark and a top antiquark ($\mathrm{t\bar{t}}$) has the highest cross-section at hadron colliders, since it is a pure QCD process, i.e. it can be mediated using quarks and gluons only (see representative diagrams in Fig. 5.1). Because of the large coupling strength of the strong interaction, higher order corrections are typically needed to obtain reasonable predictions. The cross-section and kinematics of $\mathrm{t\bar{t}}$ production can be calculated to a high precision, up to NNLO in QCD [167, 168], potentially including electroweak NLO corrections [169]. The remaining theoretical uncertainties are mainly stemming from the choice of renormalization and factorization scales, parton distribution functions and the strong coupling constant [159]. From the experimental side, several complementary analysis channels are accessed, depending on the further decays of the W bosons in both of the initial $\mathrm{t} \to \mathrm{Wb}$ decays (charge conjugation for one of them is implicit): the dilepton

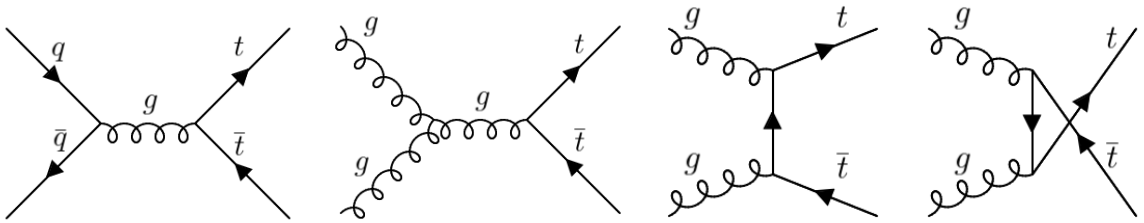
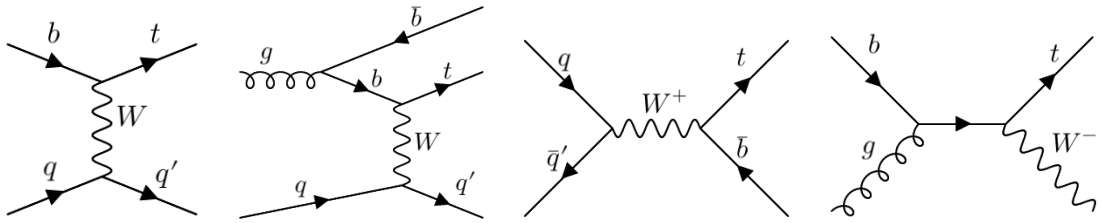


Figure 5.1: Leading order Feynman diagrams for the production of $t\bar{t}$ at the LHC in the $q\bar{q}$-channel (a), gg s-channel (b), t-channel (c), and u-channel (d). The production of $t\bar{t}$ is a pure QCD process (mediated by the strong interaction only) and hence has a relatively high cross-section in proton-proton collisions.

channel when both W bosons decay to electrons or muons (including intermediate τ leptons), the single lepton channel when one W boson decays as before and the other to hadrons, the fully hadronic channel when both W bosons decay to hadrons, and the remaining possibilities involving at least one hadronically decaying τ lepton. Each of these decay channels has characteristic advantages and challenges, and are typically studied in parallel.

Single top quark production Single top quark production proceeds mostly via W boson exchange, either in the t-channel, s-channel or tW-channel (see representative diagrams in Fig. 5.2). All of these channels contain electroweak interactions and hence their cross-section is typically lower than the $t\bar{t}$ cross-section. The t-channel is characterized by a light quark recoiling against the top quark, typically detected at high $|\eta|$ values. The presence of an additional b quark from gluon splitting could give another handle on recognizing this final state, but it is often outside the kinematic acceptance of detectors (because of too low p_T). The tW-channel is interesting from a modeling perspective because the final state is similar to $t\bar{t}$ (but with one fewer b-jet). In fact, the NLO diagrams for tW and $t\bar{t}$ interfere, and care must be taken to avoid double counting [159, 170]. From the experimental side, this makes the tW process challenging to measure because of the large $t\bar{t}$ background. The s-channel production of single top quarks is suppressed in the LHC because of the antiquark in the initial state, which must be taken from the quark-gluon sea in the proton rather than from among the valence quarks. Its characteristic signature is the presence of the (decay products of a) top quark and an additional b-jet.

Associated productions Associated productions include the production of either a $t\bar{t}$ pair or a single top quark, together with one or more additional particles. Typically, the cross-sections of these processes are lower than for pure $t\bar{t}$ or single top quark production because of the additional electromagnetic or weak interaction that is required and because of the heaviness of the final state. The most prominent examples are, in the category "$t\bar{t}$ + extra": $t\bar{t}\gamma$, $t\bar{t}Z$, $t\bar{t}W$ and $t\bar{t}H$, and in the category "single top + extra": tZq, tγ and tHq. Of special interest among the more exotic examples is $t\bar{t}t\bar{t}$, currently the heaviest final state experimentally accessible [4]. All of these processes have been studied in CMS and ATLAS, since each of them could be affected in different ways by different BSM scenarios. Among these, tZq and $t\bar{t}W$ are discussed in more detail in chapters 7 and 8 respectively.



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

Figure 5.2: Leading order Feynman diagrams for the purely electroweak single top quark production at the LHC in the t-channel with (a) and without (b) a b quark in the initial state, the s-channel (c), and the tW-channel (d). Charge conjugate states are implied.

5.3 Cross-sections of top quark processes

Overview The current state-of-the-art of top quark physics, at least as far as inclusive cross-sections are concerned, can be summarized concisely in overview graphs such as Fig. 5.3. This figure shows the agreement between measured and predicted probabilities for a large number of processes involving top quarks. Of special interest is the large range over which predictions and measurements can be made, spanning almost five orders of magnitude in cross-section. In this paragraph, we discuss some of these measurements in more detail.

Inclusive $\mathrm{t\bar{t}}$ production The inclusive $\mathrm{t\bar{t}}$ production cross-section has been studied extensively by the Tevatron and LHC experiments. Good agreement between data and prediction is observed, with no statistically significant deviation for any of the studied center-of-mass energies and decay channels[3]. The reference cross-section is calculated at NNLO in QCD with soft gluon resummation[4] up to NNLL [171].

Single top quark production Single top quark production in the t-channel, s-channel and tW-channel has been studied by the Tevatron and LHC experiments as well, with no statistically significant discrepancies between data and prediction being observed so far in these channels either[5]. Note that the s-channel single top quark production has so far escaped observation at the LHC (being suppressed by the required antiquark in the initial state); the measurements targeting this process did not yet reach a sufficient statistical significance above background processes. The process was however observed at the Tevatron experiments [172].

Associated productions Some of the most prominent inclusive cross-section measurements of top quark pairs and single top quarks in association with vector bosons are shown in Fig. 5.4. These processes are rather rare because of the required electroweak interactions and heavy final states. Hence, experimental uncertainties are large, and both larger datasets and improved analysis techniques will be needed to perform precision measurements of these processes and test the standard model more stringently. Important examples not shown in Fig. 5.4 are $\mathrm{t\bar{t}H}$ and tHq (i.e. a top quark pair or single top quark in association with a H boson). The former of these, $\mathrm{t\bar{t}H}$, is of particular importance because of its direct sensitivity to the coupling of

[3]The results are concisely summarized in the *LHC Top WG Summary Plots* (digital link). The total $\mathrm{t\bar{t}}$ production cross-section (at $\sqrt{s} = 13\,\mathrm{TeV}$) is about 800 pb. In perspective, the tZq and $\mathrm{t\bar{t}W}$ cross-sections that are the subjects of chapters 7 and 8 are less than a pb (see also Fig. 5.3).

[4]Soft gluon resummation is a technique to go beyond fixed order calculations for the emission of low-energetic gluons, that can lead to relatively large logarithmic contributions to the total cross-section.

[5]The results are concisely summarized in the *LHC Top WG Summary Plots* (digital link).

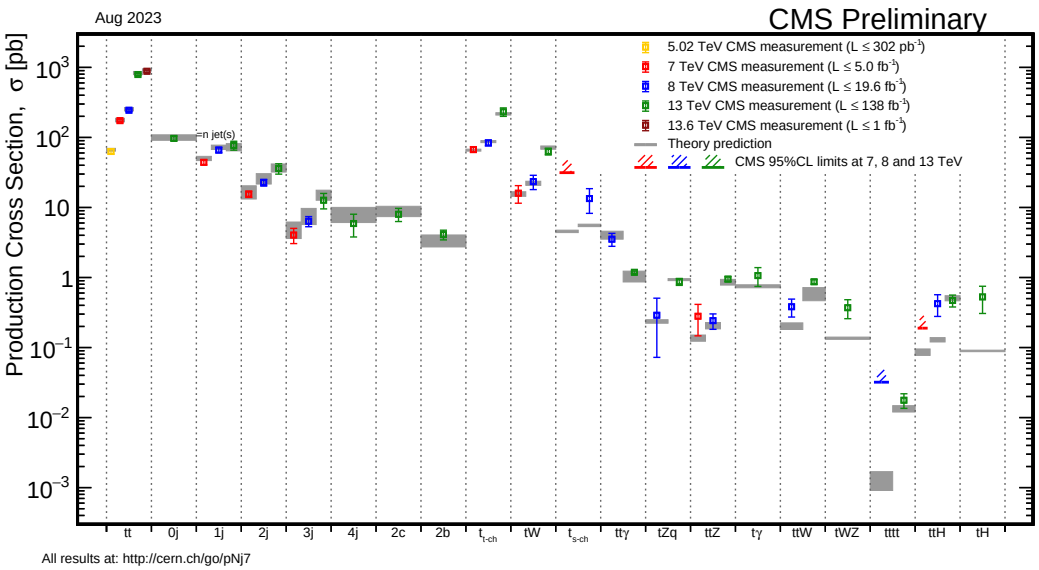


Figure 5.3: Recent overview of the experimental top quark physics program at the CMS detector (with cross-section measurements at different center-of-mass energies in red, blue and green) in comparison with theory predictions (in gray). Figure from the *CMS Standard Model Summary Plots*.(*)

(*) `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined`

the top quark and the H boson. It has been observed to agree within the uncertainties to the standard model prediction by both ATLAS and CMS [173, 174] and is currently the topic of a very active area of research interest. The latter process, tH, is even more challenging and has not yet been conclusively observed so far [175].

Differential cross-sections In differential cross-section measurements, the cross-section of a given process is measured not as a single number, but as a function of one or multiple kinematic variables. This type of measurement provides information on the dynamical aspects of the process under study that might be averaged out in the inclusive cross-section. Hence, differential cross-section measurements are more stringent tests of the standard model (or at least of the computer programs that are supposed to encapsulate the standard model for making predictions) and lead to a more thorough understanding of modeling difficulties or potential BSM effects.

Because of the large number of events that is required for differential measurements with reasonable statistical uncertainties, they were typically limited to the $t\bar{t}$ process until recently. However, the large data volume of the LHC Run II allows for differential measurements of more exotic processes such as $t\bar{t}Z$ [176, 177], $t\bar{t}\gamma$ [178, 179] tZq [1] and $t\bar{t}W$ [180]. These have so far remained without clear signal of BSM effects[6].

[6] The case of $t\bar{t}W$ is somewhat special, as there is some discrepancy between measurement and prediction, but this is most likely caused by modeling difficulties rather than BSM effects. This will be revisited in chapter 8

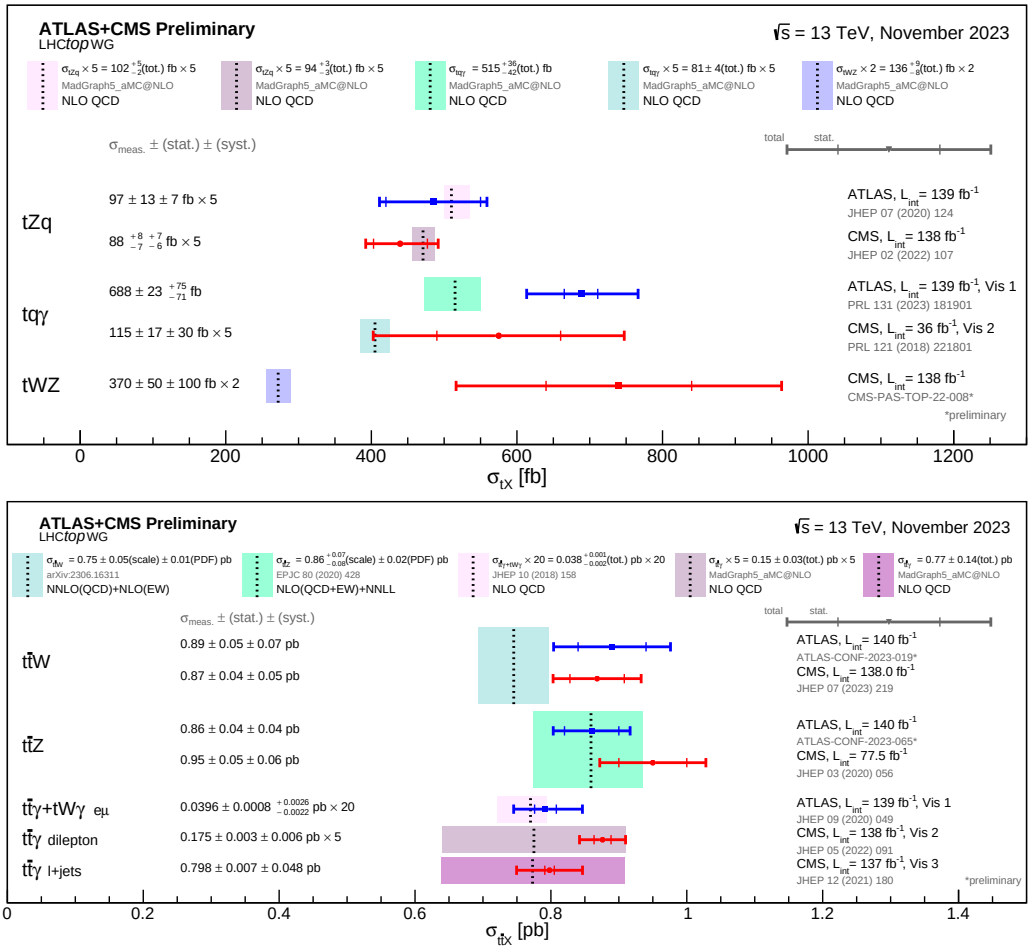


Figure 5.4: Combined ATLAS and CMS results in the field of tX (above) and ttX (below) measurements, i.e. associated production of a single top quark or a top quark pair with additional bosons. Figures from the *LHC Top WG Summary Plots*.(*)

(*) `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots`

5.4 Other properties

5.4.1 Mass

The mass of the top quark is an important parameter of the standard model. As for all other fermion masses, the top quark mass is not directly predicted by it, and hence it should be inferred from experimental measurements. Its precise determination is important, as it enters in the calculation of loop corrections to many other important and measurable quantities such as the W and H boson masses. Furthermore, it is related to the Yukawa coupling between the top quark and the Higgs field, which is larger than for any other type of quark. The relatively clean decay mode of the top quark (into a W boson and a b quark before hadronization) allows for mass measurements more precise than for any other type of quark. The current most precise measurement, consisting of a combination of LHC experimental data[7], gives a value of $m_{\mathrm{t}} = 172.52 \pm 0.14\ (\mathrm{stat.}) \pm 0.30\ (\mathrm{syst.})$ GeV [181], with a total uncertainty of less than 0.2%.

Theory From the theoretical point of view, the mass of the top quark is not uniquely defined, nor directly observable, since the top quark can neither exist as a free particle (because of QCD confinement) nor form bound hadronic states (because of its extremely short lifetime). It is a matter of some controversy to what extent the m_{t} parameter used in MC matrix element and parton shower generators corresponds to more theoretical mass definitions such as the pole mass (based on the pole of the renormalized propagator, as if the top quark was a free on-shell particle) or the $\overline{\mathrm{MS}}$ mass (energy scale dependent mass used in regularization) [182, 183].

Experiment Experimentally, the currently most accurate way of measuring the top quark mass is to reconstruct the $\mathrm{t\bar{t}}$ system from its detected decay products, and compare the measured distribution of some well-chosen variable with template distributions obtained from event generators with different input settings for the top quark mass. For example, in the latest CMS dilepton analysis [184], the observable of choice is the minimum invariant mass between any of the two selected leptons and selected b-tagged jets. The complementary lepton + jets analysis [185] employs as observable a reconstructed top quark mass obtained from a kinematic fit to the four-momenta of the selected leptons, jets and missing transverse energy under the assumption of a decaying $\mathrm{t\bar{t}}$ system. A similar procedure was employed in an analysis focusing on single top quark production rather than $\mathrm{t\bar{t}}$ [186], where the observable is the analytically reconstructed top quark mass.

5.4.2 Charge asymmetry

A counter-intuitive feature of the $\mathrm{q\bar{q}} \to \mathrm{t\bar{t}}$ process is that it displays a small charge asymmetry, i.e. it is not symmetric under the exchange of t and $\bar{\mathrm{t}}$ [187–189]. It is counter-intuitive since, at LO, the process is fully symmetric; asymmetries are only caused by interference effects at NLO, namely those between initial-state and final-state radiation and between box and tree diagrams.

Phenomenology In the LHC, $\mathrm{t\bar{t}}$ production is mainly gg initiated (rather than $\mathrm{q\bar{q}}$), because of the fact that initial-state antiquarks are only available in the parton sea, with lower probabilities. Hence the expected asymmetry is small, in the order of 1%, as the gg initiated process remains symmetric at all orders. Moreover, the expected asymmetry gets smaller as the center-of-mass energy is increased, since the gg initiated process becomes even more dominant. The asymmetry is manifested as a difference in absolute rapidities between the positively charged

[7] For an overview of the individual measurements and combinations, see the *LHC Top WG Summary Plots* (digital link).

top quarks and the negatively charged top antiquarks. It is usually quantified as follows:

$$A_C = \frac{N(\Delta|y| > 0) - N(\Delta|y| < 0)}{N(\Delta|y| > 0) + N(\Delta|y| < 0)}$$

where $\Delta|y| = \left|y_{\mathrm{t}}\right| - \left|y_{\bar{\mathrm{t}}}\right|$. In other words, the observable charge asymmetry is defined as the relative difference in number of events where the top quark has a larger absolute rapidity than the top antiquark, compared to the number of events where the opposite is the case. It is expected to be a small but positive number, as the top quark is on average more aligned with the incoming (valence) quark, which has on average a larger longitudinal momentum than the incoming (sea) anti-quark.

Measurements The charge asymmetry A_C has been measured extensively by both CMS and ATLAS [190], with no significant deviation from the standard model prediction observed. Because of smallness of A_C in $\mathrm{t\bar{t}}$ production at the LHC, a special importance is placed on the measurement of this quantity in similar processes where the $\mathrm{q\bar{q}}$ as opposed to the gg initial state is enhanced, for example $\mathrm{t\bar{t}W}$ [2, 191]. Sometimes, slightly different definitions of the quantity of interest are used, replacing the rapidity by the pseudorapidity and/or the top quarks by the leptons stemming from their decays.

5.4.3 Spin polarization observables

Top quark pair production Several observables related to the polarization of spin can be defined in $\mathrm{t\bar{t}}$ or single-top processes. In $\mathrm{t\bar{t}}$ production, the spins of the top quark and antiquark are largely unpolarized, but mutually correlated [192]. The W bosons from the top quark decay are however polarized with respect to the top quark [193].

Single top quark production In contrast to $\mathrm{t\bar{t}}$, top quarks produced in single-top processes through the Wtb vertex are fully polarized in the standard model, because of the maximally parity-violating V-A nature of the weak interaction involved [194, 195]. An observed deviation in the polarization observables could be an important hint for BSM effects entering as anomalous couplings in the top sector.

Phenomenology In the dominant single top quark production mode, the t-channel, the top quark spin is polarized in the direction of the light quark that recoils against the top quark. The polarization of the top quark affects the angular distribution of its decay products, most notably in case of leptonic decay. Hence, the kinematics of the lepton from t-channel top quark decay can be used to measure the polarization. The most common variable to study this effect is the polarization angle $\cos\left(\theta^{\star}_{\mathrm{pol}}\right)$:

$$\cos\left(\theta^{\star}_{\mathrm{pol}}\right) = \frac{\vec{p}_q^{\,\star} \cdot \vec{p}_l^{\,\star}}{\left|\vec{p}_q^{\,\star}\right| \cdot \left|\vec{p}_l^{\,\star}\right|}$$

where $\vec{p}_q$ is the momentum vector of the recoiling quark, $\vec{p}_l$ the momentum vector of the lepton from the top quark decay, and the star indicates that the vectors are evaluated in the rest frame of the top quark. Since top quarks are predicted to be polarized in the direction of the recoiling quark, and the leptons are preferentially emitted in the polarization direction of the top quarks, $\cos\left(\theta^{\star}_{\mathrm{pol}}\right)$ is predicted to be skewed towards high values. In more detail, it can be shown that the differential cross-section as a function of $\cos\left(\theta^{\star}_{\mathrm{pol}}\right)$ is a linear function:

$$\frac{d\sigma}{d\cos\left(\theta^{\star}_{\mathrm{pol}}\right)} \propto \frac{1}{2} + A_l \cos\left(\theta^{\star}_{\mathrm{pol}}\right)$$

where A_l is called the spin asymmetry and is predicted to be in the order of 0.5.

Measurements The spin asymmetry was measured in t-channel single top quark production, and verified to be in good agreement with the standard model prediction [194, 195]. While this measurement is the most sensitive one because of the relatively large statistical precision, the spin asymmetry can be measured in rarer processes such as tZq as well [1], to serve as a consistency check and to capture the potential impact of additional particles in the final state.

5.4.4 CKM matrix element

Phenomenology The CKM matrix element V_{tb} encapsulates the coupling strength between the top quark and the b quark, especially when compared to V_{ts} and V_{td}, the coupling with respect to the lower-generation s and d quarks respectively. It is different from the other elements in the CKM matrix in that it is close to unity (which also implies, because of unitarity, that V_{ts} and V_{td} must be very small), resulting in the fact that the top quark decays almost exclusively to a b quark, and only very rarely to s or d quarks directly.

Measurements The single top quark production processes provide a simple way to measure the size of the CKM matrix element $|V_{\mathrm{tb}}|$. As both the production and decay of the top quark occurs via the Wtb vertex, the cross-section for this process depends quadratically on the CKM matrix element [159, 160]: $\sigma_{\text{single-t}}/\sigma^{\mathrm{SM}}_{\text{single-t}} \propto |V_{\mathrm{tb}}|^2$. Hence, the cross-section measurements can be directly translated in $|V_{\mathrm{tb}}|$ measurements (all other parameters affecting the cross-section being held at their nominal values). Extensive results have been provided by the CMS and ATLAS Collaborations[8], all consistent with $|V_{\mathrm{tb}}| = 1$ within the uncertainties.

5.4.5 Strong coupling constant

The coupling constant of the strong interaction (α_S) is a very important parameter in the standard model. It can be derived from top quark-antiquark pair measurements since it affects the cross-section for this process quite strongly (with a simple $\sigma_{\mathrm{t\bar{t}}} \propto \alpha_S^2$ relation at lowest order) [160]. Hence, an inclusive $\mathrm{t\bar{t}}$ cross-section measurement can be directly interpreted as an α_S measurement (all other parameters affecting the cross-section being held at their nominal values). The current most precise result yields $\alpha_S = 0.117$, evaluated at the Z mass scale, with 1.8% relative uncertainty [196]; it is obtained from a combination of 7 and 8 TeV data from ATLAS and CMS, and it is compatible with the world average measured value[9].

5.4.6 Top quark Yukawa coupling

The top quark Yukawa coupling y_{t} has a strong impact on the inclusive $\mathrm{t\bar{t}t\bar{t}}$ production cross-section, which contains a term proportional to $\left|y_{\mathrm{t}}/y_{\mathrm{t}}^{\mathrm{SM}}\right|^4$, as well as an interference term proportional to $\left|y_{\mathrm{t}}/y_{\mathrm{t}}^{\mathrm{SM}}\right|^2$ [160,197]. Measuring the inclusive $\mathrm{t\bar{t}t\bar{t}}$ cross-section can therefore be translated into measuring y_{t} [198]. The top quark Yukawa coupling also affects some kinematic distributions of $\mathrm{t\bar{t}}$ production, and hence its value can be extracted from a measurement of differential $\mathrm{t\bar{t}}$ cross-sections, especially binned in the mass of the $\mathrm{t\bar{t}}$ system and the rapidity difference between the top quark and antiquark [199, 200]. The latter approach (yielding the more precise results) obtains a top quark Yukawa coupling consistent with the standard model value within uncertainties, and an upper limit $y_{\mathrm{t}}/y_{\mathrm{t}}^{\mathrm{SM}} < 1.54$ at 95% confidence level.

[8] For an overview, see the *LHC Top WG Summary Plots* (digital link).

[9] For an overview of individual measurements, see Ref. [160]

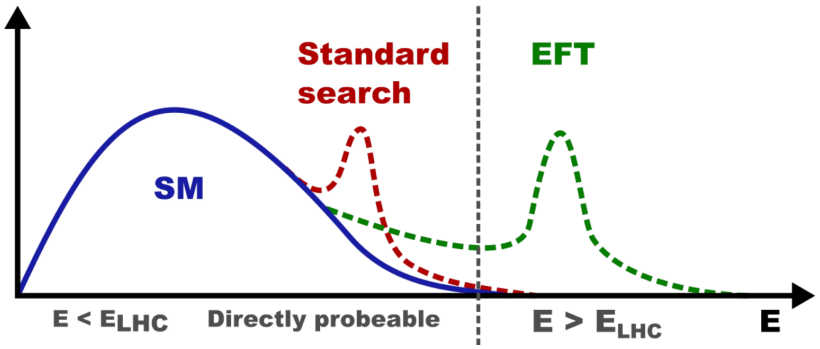


Figure 5.5: An exaggerated picture illustrating how directly searching for new physics, sometimes also called "bump hunting", differs from the EFT approach in terms of the assumed energy scale of new physics processes. Figure and caption by G. Mestdach [201].

5.5 Effective field theory

Indirect signatures In the absence of any direct experimental evidence for specific BSM scenarios in any of the dedicated searches carried out at the LHC so far, indirect signatures become the target of investigation. The central hypothesis of effective field theory (EFT) is that the new physics phenomena are occurring at a too high energy scale (i.e. the masses of the associated particles are too high) to be probed directly at the current generation of accelerators, but that their existence is manifested at the currently experimentally accessible energy scale by small changes in the observed cross-sections and kinematic distributions of standard model processes. This conceptual shift between direct and indirect signatures is illustrated in Fig. 5.5.

Parameterization Furthermore, these deviations can be comprehensively and consistently modeled using a set of effective operators added to the standard model Lagrangian density [158]:

$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{C_i^{(6)} O_i^{(6)}}{\Lambda^2} + O(\Lambda^{-4})$$

Here, $\mathcal{L}_{\text{SM}}$ is the Lagrangian density of the standard model. The added terms consist of dimension-6 operators[10], representing higher-order interactions between standard model particles. These terms are forbidden in the standard model because they are not renormalizable, but they are valid as low-energy approximations of high-energy phenomena involving interactions with new particles. The interaction strength of each operator is regulated by the corresponding Wilson coefficient $C_i^{(6)}$ divided by the energy scale Λ squared, used to bring the dimension of the term as a whole back from 6 to 4, as required. Λ is interpreted as the energy scale where the new physics phenomena become too significant for the effective field approach to be a valid approximation.

Experimental approach This theoretical framework translates into the following experimental challenge: calculate the impact of each possible EFT operator on each possible signal or background process, and interpret all data-to-simulation comparisons in terms of constraining the magnitude of the Wilson coefficients, ideally until only a small number of them are found to be compatible with values different from zero; this could indicate in which direction

[10] All standard-model operators in the Lagrangian density are of dimension 4 or less. Hence, dimension-5 operators form the first possible extension, but they violate baryon or lepton number and are typically not considered.

new physics should be sought for. The large number of operators, processes and experimental signatures make this a difficult endeavour.

Results in the top quark sector Not all operators are expected to impact all processes significantly; hence one can isolate more-or-less orthogonal subsets of operators and processes, and study them independently. The subset of operators involving couplings to top quarks is most sensitively probed in experimental signatures enriched in top quark processes. The EFT approach is still relatively young, and is expected to benefit from more comprehensive and systematic studies in the future. But so far, no coefficients incompatible with zero have been observed[11].

EFT-analyses involving tZq and $\mathbf{t\bar{t}W}$ Of special interest for this work are the tZq and $t\bar{t}W$ process. The latter is not really suitable to extract strong constraints on EFT operators, both because it is more difficult to model even when limited to the standard model and because it does not have particular sensitivities not testable in other processes, but can be included in EFT analyses for completeness nonetheless. The former however is interesting from an EFT point of view as it is affected by many EFT operators with sizable impact, especially at the high energy tail of kinematic distributions (e.g. of the top quark p_{T}) [202]. Hence, this process has been targeted in a number of EFT analyses [203, 204].

5.6 Open questions

Despite the extensive theoretical progress and experimental success in top quark physics in the last decades, the last of the quarks remains shrouded in mystery, and many open questions remain [159]. Much is known and understood about the "how" (e.g. its decay mechanisms, its virtual corrections to interaction diagrams and its phenomenology in proton-proton collisions) but much less about the "why".

Mass Why is the top quark so much heavier the other quarks? The standard model does not predict the top quark mass as the Yukawa couplings are free parameters, to be determined experimentally. Yet it is difficult to imagine that these are to be chosen randomly, that there isn't some kind of generating mechanism behind the masses of the three generations of quarks and fermions that necessitates the masses they are observed to have. But the heaviness of the top quark compared to the other quarks seems to defy any kind of quasi-regular pattern of increasing masses one would naively expect in that case. It is also suspicious that the mass of the top quark is in the same scale as that of the W, Z and H bosons, insinuating a potential special role of the top quark in electroweak symmetry breaking.

Yukawa coupling Why is the Yukawa coupling of the top quark numerically very close to unity? Is there a special connection between the top quark and the H boson? This is a compelling question, as it relates to the H boson mass puzzle as well. Calculated loop corrections to the H boson mass, especially coming from virtual top quarks, are extremely large with respect to its observed mass; it seems there is a so-far undiscovered symmetry that exactly cancels these corrections, but it has as yet eluded experimental searches.

Stability of the Universe An even more radical consequence of the interaction between the top quark and the H boson is the possible meta-stability of our Universe: these interactions reshape the H potential to such an extent that our Universe is perhaps not in the absolute

[11] The results are concisely summarized in the *LHC Top WG Summary Plots* (digital link).

ground state (i.e. stable), but rather in a meta-stable secondary potential minimum[12]. Measurements indicate the we are close to the borderline between stability and meta-stability, with however a 99.3% CL preference for the latter [205]. Why is our Universe in this critical state? Is the explanation just anthropic or is there a special mechanism that favours Universe configurations close to this boundary? Furthermore, even if we assume that the decay time of the Universe due to this particular instability is much larger than the probable lifetime of a life-sustaining Universe anyway, the question remains how our Universe ended up here, given that it used to be much denser and hotter, and a transition to the absolute minimum would have been much more likely to occur.

The context of this thesis Chapters 7 and 8 of this work aim to take a few small steps towards the elucidation of these mysteries. The questions on the special role of the top quark in electroweak symmetry breaking and its connection with the H boson, the various numerical coincidences and seemingly accidental cancellations, are potential signs of new physics effects beyond the standard model. As introduced in chapter 1, these effects can either be searched for in direct, dedicated searches, or by precision measurements of the standard model in challenging phase spaces. It is the latter approach that is followed in chapter 7 and 8.

[12]Further reading: `https://cerncourier.com/a/the-higgs-and-the-fate-of-the-universe/`

Chapter 6

Some experimental techniques in top quark physics

This chapter groups the description of methods and techniques that are common to the two top quark analyses that will be described in chapters 7 and 8. It does not aim to give a full overview of experimental methods commonly used by analyses involving top quarks, but rather has the much more humble purpose of avoiding unnecessary repetition in the chapters that follow. None of the methods described in this chapter are radically innovative, but I adapted them to the specific needs of chapters 7 and 8, and sometimes made a few improvements or extensions.

My implementation of the method used for estimating the background yield and kinematics with nonprompt leptons (see section 6.2) and its results were used not only in the tZq [1] and $t\bar{t}W$ [2] measurements discussed in later chapters, but also in the first observation of the $t\bar{t}t\bar{t}$ process [4] and in a search for heavy neutral leptons [5] performed in our research group.

6.1 Prompt lepton identification using an MVA method

Prompt vs. nonprompt leptons Most of the analyses involving top quarks target experimental signatures with one or more leptons (i.e. electrons or muons). The reason is that leptons can be identified with high accuracy and measured with high precision, using a combination of the tracker and the ECAL (for electrons) or gaseous outer detectors (for muons). For these analyses, it is of vital importance to distinguish prompt from nonprompt leptons. In this context, the term 'nonprompt lepton' refers to any lepton originating from secondary decays (e.g. b hadron decays) or any non-lepton object that is wrongly reconstructed and identified as a lepton. Contrary to this, the so-called 'prompt leptons' are genuine leptons originating from the primary process in the event (mostly from the decay of any of the top quarks or accompanying electroweak bosons).

Discriminating features Figure 6.1 shows an example diagram of a $\mathrm{t\bar{t}}$ event, illustrating the definitions of prompt and nonprompt leptons. Figure 6.2 sketches the experimental signatures of prompt and nonprompt leptons, the difference between which can be exploited to distinguish between the two categories of leptons:

- **Isolation** Nonprompt leptons are often less 'isolated', i.e. there are more particles in a spatial cone centered on the lepton, because of the hadronic activity in the decay chain that they are a part of. Typically, the scalar sum of transverse momenta of all particles in a cone around the lepton is taken as a measure for isolation.

- **Impact parameter** Secondary vertices can be macroscopically displaced from the primary vertex. This leads to the effect that the distance of closest approach of the (extrapolated) lepton trajectory with respect to the primary vertex is on average larger for nonprompt leptons. This distance is called the impact parameter.

- **Number of hits** The number of pixels or strips in the CMS tracker where a passing lepton deposited electric charge (called a 'hit'), is on average smaller for nonprompt leptons, because of the macroscopic displacement between their point of origin and the center of CMS.

- **Jet parameters** All reconstructed particles in CMS are clustered in jets. One can consider the jet geometrically closest to a lepton: for prompt leptons this jet will essentially coincide with the lepton, while nonprompt leptons are often just a small part of a jet. Typical variables are the $p_{\mathrm{T}}^{\mathrm{ratio}}$, the ratio between the p_{T} of the lepton and that of its closest jet, and $p_{\mathrm{T}}^{\mathrm{rel}}$, the relative p_{T} of a lepton with respect to the direction of its closest jet.

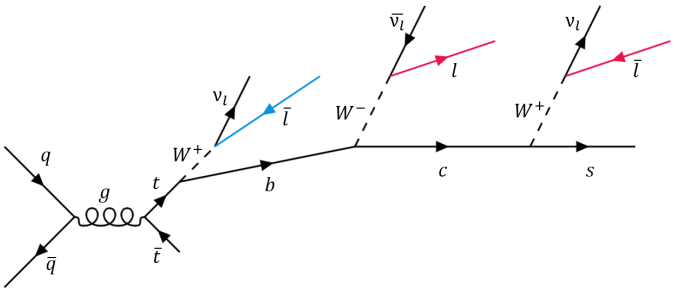


Figure 6.1: Example diagram of a $t\bar{t}$ event with a prompt lepton (in blue) and several nonprompt leptons (in red). The prompt lepton originates from the decay of the top quark, which happens instantaneously and at the primary vertex (compared to the typical experimental time and length scales of CMS), while the nonprompt leptons originate from subsequent decays in the chain. The decay of the top antiquark is not shown in the figure but could happen analogously.

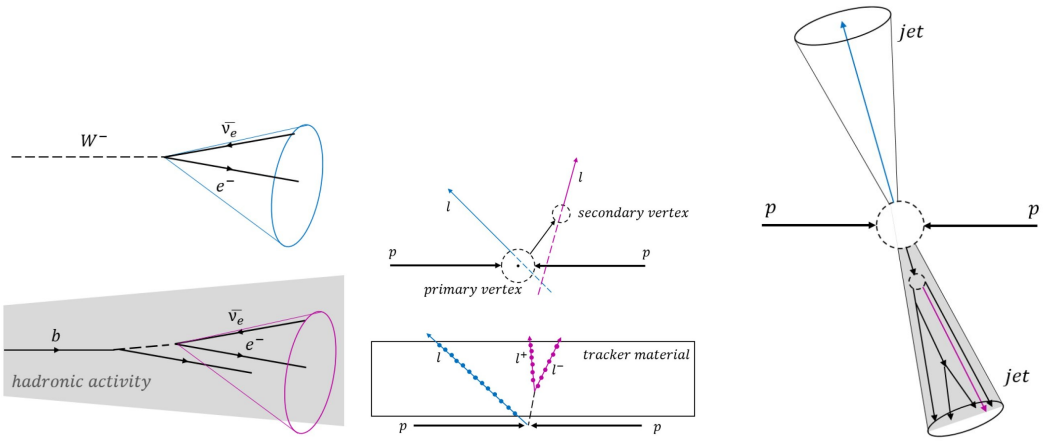


Figure 6.2: Experimental signatures of prompt leptons (in blue) versus those of nonprompt leptons (in purple). Left: isolation, upper middle: impact parameter, lower middle: number of hits in the tracker, right: closest jet parameters.

GBDT Based on these observations, a number of discriminating variables are constructed and combined in a multivariate analysis (MVA) discriminator, more specifically in a gradient boosted decision tree (GBDT). The training sample consists of prompt leptons from simulated $t\bar{t}Z$, $t\bar{t}W$, tZq, $t\bar{t}t\bar{t}$ and $t\bar{t}H$ samples, and nonprompt leptons from simulated $t\bar{t}$ events. The hyperparameters of the GBDT are optimized using cross-validation, class imbalance is accounted for via reweighting, and overtraining is avoided using early stopping if no further improvement in the validation loss is observed. More details can be found in [206][1].

Calibration The efficiency of the lepton identification based on this MVA has been calibrated in data, and a set of correction factors and their associated uncertainties have been measured[2], and have been approved by the Muon POG and EGamma POG. The correction factors are used to reweight the simulation and the associated uncertainties are propagated using weight variations.

Lepton identification in practice Using a well-tuned threshold score of an MVA discriminator such as the one described above, prompt leptons can be identified and selected with a given efficiency and nonprompt background rate, depending on the needs of a specific measurement. Apart from an MVA score threshold, several other variables are typically used to define high-fidelity leptons in chapters 7 and 8, to apply very basic selection criteria and limit the lepton phase space to the one used for training the MVA. The definition of these additional variables is summarized in Tab. 6.1. The impact parameter is the distance of closest approach between a track (linearly extrapolated from its innermost measured position towards the center of CMS) and the primary vertex in an event. It can be measured in three-dimensional space (d), in a projection on the transverse plane (d_{xy}), or in a projection on the longitudinal axis (d_z). Examples of lepton identification criteria, combining baseline cuts on these variables with a tuned MVA score threshold, can be found in Tabs. 7.1 and 8.1 for muons, and Tabs. 7.2 and 8.2 for electrons.

6.2 Nonprompt lepton background estimation using the fake rate method

Origin of nonprompt lepton backgrounds A considerable source of background to the tZq, $t\bar{t}W$ and $t\bar{t}t\bar{t}$ processes (in the multilepton channel) consists of events with at least one nonprompt electron or muon. Although the prompt lepton MVA (see section 6.1), and the lepton identification criteria using this MVA, were designed to target prompt leptons and reject nonprompt leptons and are highly efficient at that task, the number of events that pass the lepton selections and yet contain one or more nonprompt leptons is not negligible, mainly because of the relatively large cross-sections of these processes with respect to the signal and prompt backgrounds.

The main physical processes contributing to this background category are $t\bar{t}$ and Drell-Yan (DY, i.e. Z boson production), with different relative contributions in different signal selections. Since these processes have very low selection efficiencies in our typical signal regions (because of the multi-lepton and (b-tagged) jet requirements), they are difficult to estimate

[1] The lepton MVA for prompt lepton identification discussed in this section was originally developed mainly by Willem Verbeke (in the context of the first tZq observation [207]) and optimized, extended to the full Run II, and generalized to multiple signal processes involving top quarks by Kirill Skovpen [206]. Since I was less directly involved in its development, but mostly in its usage, the technical description here is kept rather limited.

[2] By Tom Cornelis and Kirill Skovpen, from our research group.

Table 6.1: Definition of variables used in lepton identification. Note: in these variables, proximity (e.g. of a jet to a lepton) is defined in η-ϕ space, i.e. in terms of their direction.

Observable	Definition
	Muons and electrons
$\lvert d_{xy}\rvert$	Transverse impact parameter with respect to the primary vertex.
$\lvert d_z\rvert$	Longitudinal impact parameter with respect to the primary vertex.
$\lvert d\rvert/\sigma_d$	Impact parameter divided by its estimated uncertainty.
Relative isolation	Scalar p_T sum of physics objects in a cone around the lepton, divided by the p_T of the lepton itself. The physics objects include charged hadrons, neutral hadrons and photons, and a correction for pileup contributions. The cone size is variable, scaling with $1/p_\mathrm{T}$ of the lepton [208, 209].
DeepJet of nearest jet	The DeepJet b-tagging score [86, 87] of the jet nearest to the lepton.
Jet p_T ratio	The scalar ratio of the p_T of the lepton to the p_T of the jet nearest to the lepton.
	Muons only
Muon POG ID	Standard muon identification criteria defined by the Muon POG [69].
	Electrons only
Missing hits	Number of missing hits in the track associated with the electron. Missing hits are positions where a hit is expected based on the estimated track parameters, but where no hit is found.
$\sigma_{i\eta i\eta}$	Measure for the width of an ECAL shower in the η direction. It is computed as the energy-weighted standard deviation of ECAL crystal η around the most energetic crystal [210].
H/E	Energy deposited in the HCAL divided by that in an ECAL shower, for a region of HCAL in a small cone around the direction of the ECAL shower [210].
$1/E - 1/p$	Difference between the energy measured in the ECAL supercluster and the momentum measured in the tracker at the point of closest approach to the vertex (both inverted) [210].
Conversion rejection	Criterion for rejecting electrons likely to originate from photon conversions, based on the presence of missing hits in the innermost layers of the tracker. In addition, if a pair of nearby electron tracks can be formed, a common vertex is fitted and the conversion probability is assessed using the χ^2 of the fit [211].
Charge consistency	Agreement between three methods for determining the charge of the electron (curvature of the gaussian-sum-filter track, curvature of the kalman-filter track, position of ECAL cluster with respect to track) [211].
EGamma POG MVA score	Output score of a set of MVAs (in p_T and η bins) developed by the EGamma POG for prompt electron identification [210].

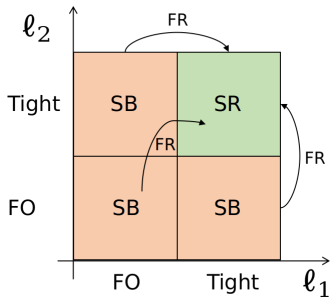


Figure 6.3: Schematic overview of the fake rate method (in this case for a dilepton signature). The signal region (SR) selects only events where all leptons pass the tight definition. The sidebands (SB) consists of events where all leptons pass the FO definition but at least one lepton does not pass the tight definition. Using the fake rate (FR), these events can be used to estimate the nonprompt contribution to the signal region.

reliably from simulation. Hence, we apply a data-driven technique to estimate the nonprompt background.

The fake rate method in a nutshell The nonprompt background is estimated using the 'fake rate method' (also called 'fake factor method' or 'tight-to-loose ratio method'). The idea is to measure the probability, called the fake rate, that a nonprompt lepton passing a loosened selection[3] (called 'fakeable object' or FO) also passes the tight identification, and then apply that probability to a data sample similar to the signal region, but selected with FO leptons instead of tight leptons. The fake rate is measured in a data sample enriched in QCD multijet events, as a function of p_{T} and η. It is then applied in a data sample passing the signal region selections, but with the difference that the leptons are required to pass the FO definition instead of the tight definition, while on the contrary events with only tight leptons are vetoed (to avoid overlap with the signal region). This very short summary is displayed in Fig. 6.3 and discussed in more detail in the following paragraphs. Note that similar methods have been used in other recent analyses involving top quarks, e.g. [207, 212–214].

6.2.1 Applying the fake rate

The fake rate f is defined as follows:

$$f = P(\text{lepton passes tight} \mid \text{lepton passes FO and lepton is nonprompt}).$$

In other words, it is the probability that a nonprompt lepton that passes the FO definition also passes the tight definition. The measurement of the fake rate is discussed in the following paragraphs; here we assume that the fake rate has been measured already and discuss how it is applied. We consider an application region (in data), which consists of the same selections as the signal region with the exception that leptons are required to pass the FO definition instead of the tight definition, and furthermore events where all leptons additionally pass the tight definition are vetoed in order to avoid overlap. The estimation of the nonprompt contribution to the signal region is obtained by weighting the events in the application region, where the weight depends on the fake rate. The full derivation of the correct weighting factor is discussed

[3] Loosened with respect to the full lepton selection used for the main part of the analysis in question. The full lepton selection is typically referred to as the 'tight' selection, and the loosened one as the 'fakeable object (FO)' selection, a convention which is also used in this text. For examples of tight and FO selections, see Tabs. 7.1, 7.2, 8.1 and 8.2.

in Ref. [215] (in a formal way), and by myself in Ref. [216] (in a more intuitive way). The result is:

$$\text{weight} = (-1) \prod_{\text{failing}} \frac{-f}{1-f}, \tag{6.1}$$

where the product runs over leptons in the event failing the tight definition and f is the corresponding fake rate. The procedure is applied on data events in the application region, but also on simulation and the resulting contribution is subtracted from the data, in order to correctly take into account prompt leptons failing the tight definition. The next paragraphs discuss in more detail how the fake rate is measured.

6.2.2 Cone correction

Mother parton The fake rate is measured as a function of the p_{T} and η of the nonprompt lepton. However, since the probability of a nonprompt lepton passing the full selection depends in part on its environment (e.g. isolation), the fake rate for two nonprompt leptons with the same measured p_{T}, but originating from partons with different p_{T}, might differ between the two. To mitigate this effect, the leptons in the measurement are cone-corrected [217] to their mother parton p_{T}.

Cone correction As a good proxy for the p_{T} of the mother parton of a lepton (which cannot be retrieved in data), we can employ the p_{T} of the jet that is geometrically closest to the lepton (see Fig. 6.4). Hence, in practice, the cone correction consists of replacing the p_{T} of the lepton by that of its geometrically closest jet. This cone-corrected p_{T} is then used instead of the uncorrected p_{T} in all further steps of the fakerate method.

Correction factor Note that this correction is applied only to leptons passing the FO definition but failing the tight definition, as the leptons used for further analysis should remain unbiased. This causes a discontinuity in the p_{T} spectrum of the leptons as a function of the lepton MVA discriminant at the cut value of the tight selection. To mitigate this difference, an additional multiplicative factor is added to the cone-corrected p_{T} and tuned such that there is no discontinuity. These factors are typically in the order of 0.6 - 0.9, depending on the value of the lepton MVA discriminant for tight lepton selection.

6.2.3 Simulated closure tests

Before measuring the fake rate in a QCD multijet enriched data sample and applying it to signal-like data, we first perform simulated closure tests to prove the principle of this method. These tests are also used to tune the FO definition for the leptons in this analysis, for it is not guaranteed that a measured fake rate in QCD multijet events yields an accurate prediction in the application region (mostly DY- and $\mathrm{t\bar{t}}$-like events) for any choice of FO definition. More details on the optimization procedure are given in section 6.2.4. Here, we report the procedure and the result for the optimized FO definition.

Event selection for simulated fake rate measurement Simulated closure tests are performed by measuring the fake rate in simulated QCD multijet samples and applying it in simulated DY and $\mathrm{t\bar{t}}$ samples. The events in the QCD multijet samples are subjected to the following selections:

- The event must have exactly one lepton passing the FO definition. This lepton is required to be muon/electron when measuring the fake rate for muons/electrons, respectively.

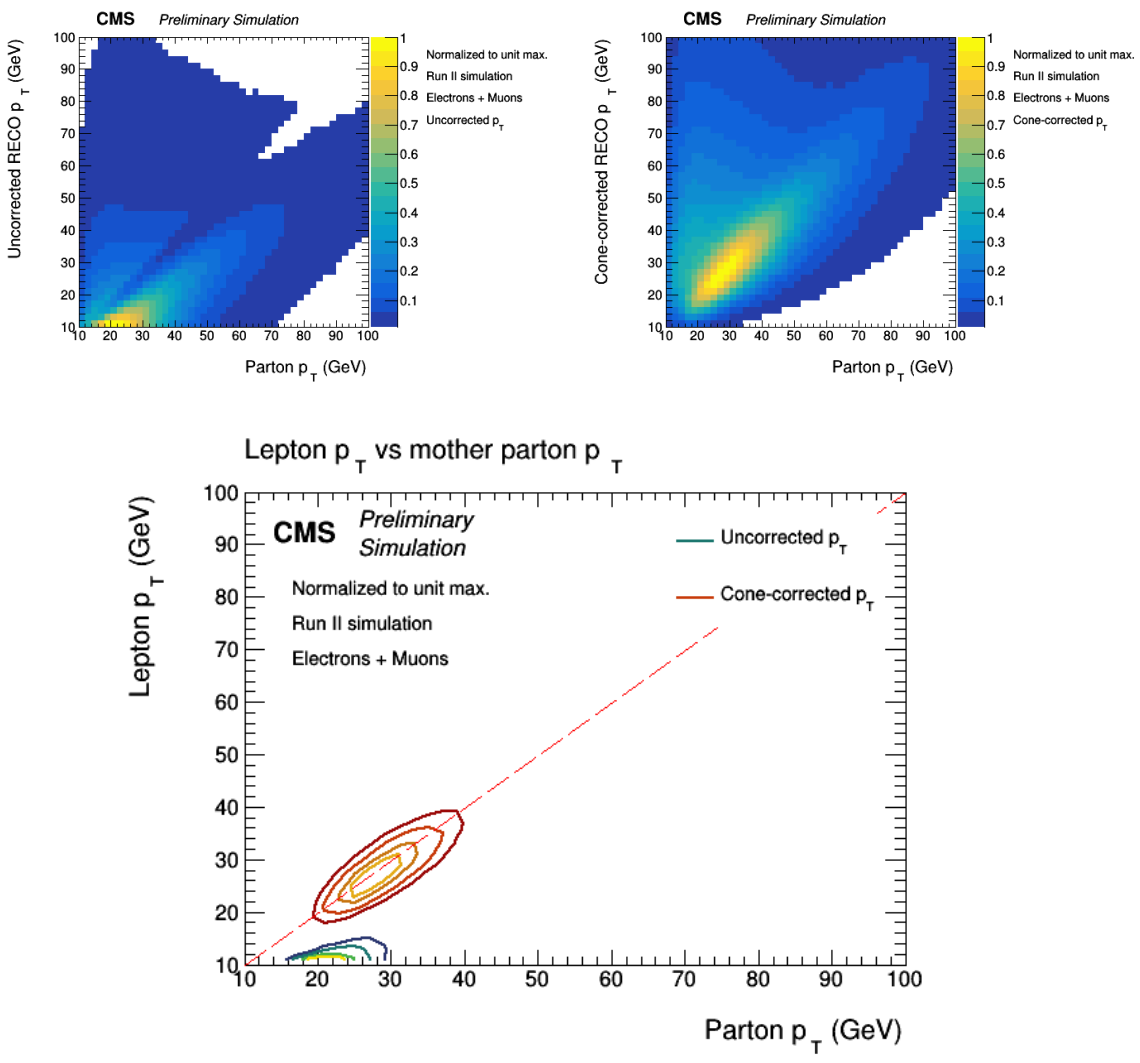


Figure 6.4: Proof-of-concept of the cone correction. Upper left: correlation between the uncorrected p_{T} of the nonprompt lepton and the p_{T} of the mother parton. Upper right: correlation between the cone-corrected p_{T} of the nonprompt lepton and the p_{T} of the mother parton. Lower: both results plotted together using contour lines. The cone-corrected p_{T} is seen to approximate the mother parton p_{T} much better than the uncorrected p_{T} does.

- The event must contain at least one jet with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$, separated from the lepton by at least $\Delta R = 0.7$.

The remaining nonprompt leptons go into the denominator, and if they additionally pass the tight definition they go into the numerator as well. The ratio of both (per bin of p_{T} and η) is the simulated fake rate.

Event selection for simulated closure test The events in the DY and $\mathrm{t\bar{t}}$ samples are required to pass the following conditions:

- The event must have at least two leptons passing the FO definition.
- In case the number of leptons is two, they must be same sign.
- At least one of these leptons must be nonprompt.

If all leptons additionally pass the tight definition, the event is 'observed', meaning that it is a genuine event with at least one nonprompt lepton still passing the signal region lepton selection. If at least one of the leptons is non-tight, the simulated fake rate is used to calculate a weight for this event (according to formula 6.1). Ideally the 'observed' and predicted distributions should align, implying that we can use simulated QCD multijet to accurately predict the fake rate in simulated DY and $\mathrm{t\bar{t}}$.

Results of the closure test The results of these closure tests are shown in Fig. 6.5 for samples corresponding to 2018 data-taking conditions (with very similar results for samples corresponding to other data-taking years, not shown here explicitly). Overall, decent closure is observed within the uncertainties (which are quite large especially for DY due to the efficient nonprompt background rejection). The systematic error bands in these plots correspond to a flat 30% a priori uncertainty (on top of the statistical uncertainties). Note that this uncertainty is chosen here arbitrarily as a guide for the eye; the actual uncertainties provided to the statistical model in two use cases will be discussed in sections 7.5.1 and 8.5.2.

6.2.4 Optimization of the FO definition

Motivation The fake rate method is not guaranteed to work for any FO definition. In more detail: it is not guaranteed that the ratio of nonprompt tight leptons to nonprompt FO leptons, as measured in QCD multijet enriched events, can be used to accurately predict the nonprompt background in the signal region (consisting mostly of $\mathrm{t\bar{t}}$ and DY) for just any FO definition, as the origin and composition of nonprompt leptons could be different between measurement and application region. In particular, the $\mathrm{t\bar{t}}$ process is expected to give mainly nonprompt leptons originating from heavy-flavor partons (e.g. leptons from b-hadron decays), whereas the DY process is expected to give mainly nonprompt leptons originating from light-flavor partons (e.g. light jets faking leptons). The fake rate measured in QCD multijet events should therefore be tuned to be independent of the mother parton flavor.

Procedure In order to find a suitable FO definition, the procedure of the simulated closure test (section 6.2.3) is repeated for a grid of FO definition candidates. The template FO definition has the form 'if leptonMVA > threshold for tight, pass; else, apply some additional cuts'. The grid of FO definition candidates is then constructed by defining some lepton variables to apply additional cuts on, as well as a range of numeric cut values. In this study, we have the following variables:

- $p_{\mathrm{T}}^{\mathrm{ratio}}$: the ratio of the transverse momentum of the lepton to that of its geometrically closest jet. The $p_{\mathrm{T}}^{\mathrm{ratio}}$ of a lepton is required to be greater than a threshold value.

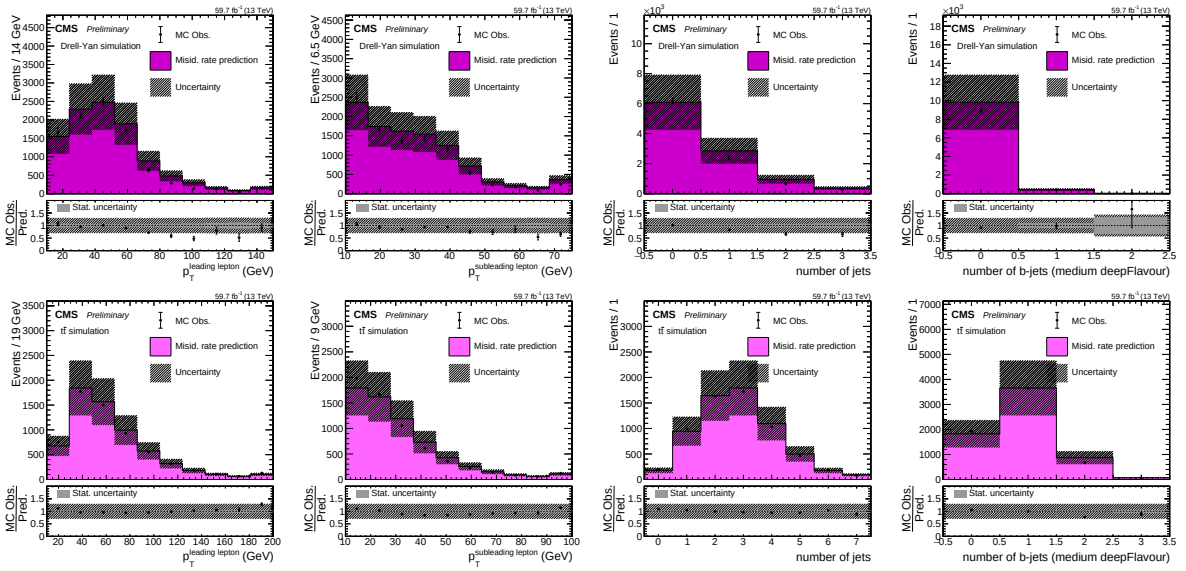


Figure 6.5: Simulated closure tests for the 2018 FO definition, in DY (upper row) and $\mathrm{t\bar{t}}$ (lower row). Closure is shown as a function of the p_{T} of the leading and subleading lepton in the event, and as a function of the number of jets and b-tagged jets.

- $\mathrm{DF_{jet}}$: the DeepFlavour b-tagging score of the jet geometrically closest to the lepton. The $\mathrm{DF_{jet}}$ of a lepton is required to be less than a threshold value.
- MVA_e: a POG-defined MVA used for prompt electron identification. The electron is required to pass the MVA selection, using four different working points (no MVA selection, WPLoose, WP90 and WP80). This cut is not applicable for muons. In a previous iteration of this study (with an older version of the lepton MVA), the WPLoose working point was found to be optimal. This variable was kept fixed to this value during the new iteration of this study. Since good fake rate maps and closure in both simulation and data are observed, there is no need to extend the search to other MVA_e working points.

The closure for each candidate FO definition is quantified as the normalized χ^2-divergence between the observed and predicted histogram in the simulated closure test. We used the histograms of the leading lepton p_{T}, but very similar optimal FO definitions are found using any other variable as a function of which the closure can be determined.

Results The results of this procedure are shown in Fig. 6.6 (for electrons) and 6.7 (for muons). For simplicity, we again only show the results for samples corresponding to 2018 data-taking conditions, with analogous results for other data-taking years. Three (sometimes conflicting) criteria are used to determine the optimal FO definition: the closure in DY, the closure in $\mathrm{t\bar{t}}$, and the magnitude of the fake rate. The latter criterion is important since too large fake rates produce large statistical uncertainties when applying the method in data.

Discussion The results of the FO optimization can be interpreted in terms of the impact that the tuneable cuts have on the selection of nonprompt leptons in both the measurement region (simulated QCD multijet) and application region (simulated $\mathrm{t\bar{t}}$ and DY). A distinction should be made between nonprompt leptons originating from a light parton (e.g. light jets faking leptons) and those originating from a heavy parton (e.g. leptons from b or c hadron decays).

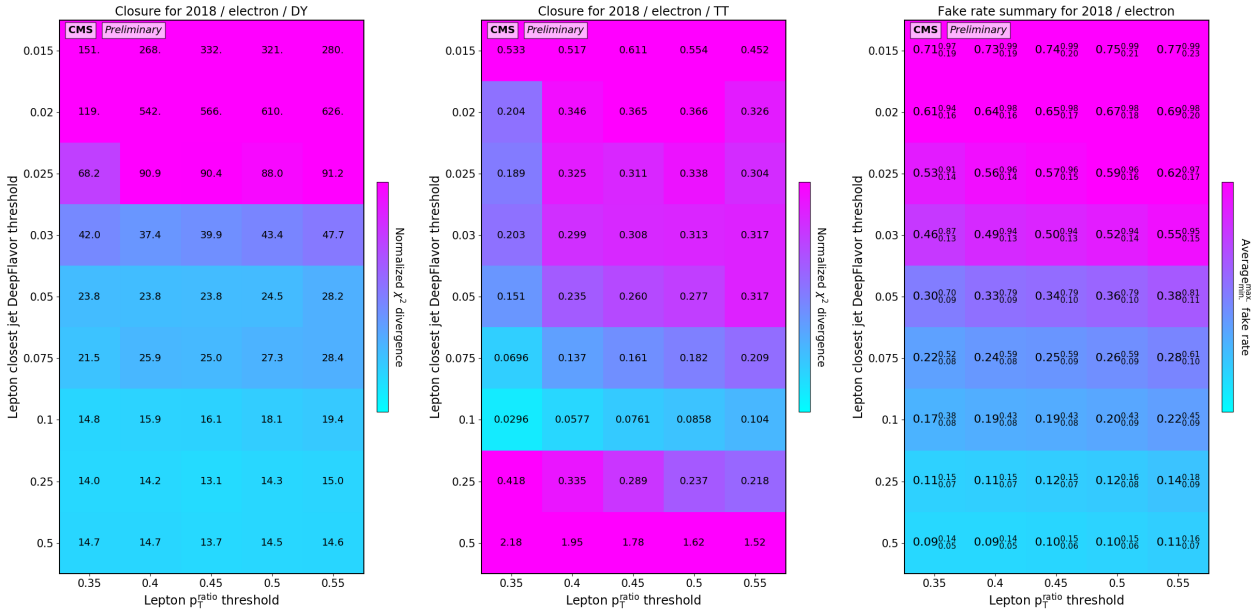


Figure 6.6: Results of simulated closure tests for 2018 electrons, for a grid of potential FO definitions, in DY (left) and $t\bar{t}$ (middle). In addition, the resulting fake rate maps are summarized for each working point (right) in the following format: $\text{average}^{\text{maximum}}_{\text{minimum}}$ fake rate.

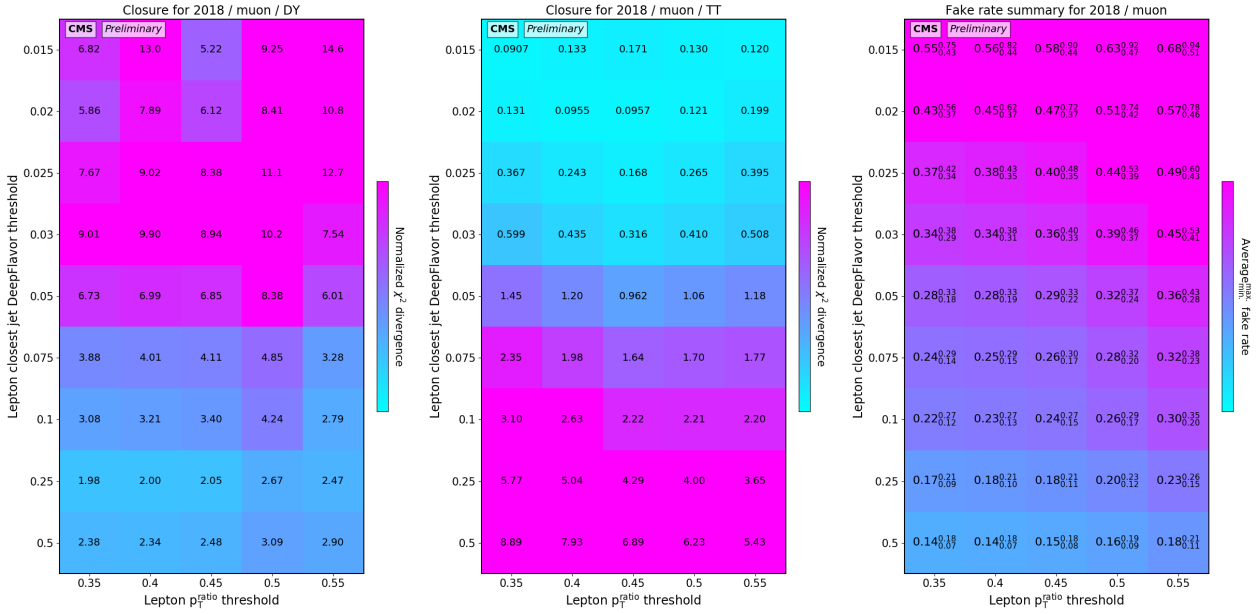


Figure 6.7: Results of simulated closure tests for 2018 muons, for a grid of potential FO definitions, in DY (left) and $t\bar{t}$ (middle). In addition, the resulting fake rate maps are summarized for each working point (right) in the following format: $\text{average}^{\text{maximum}}_{\text{minimum}}$ fake rate.

In the QCD multijet measurement region, heavy-flavored partons are the dominant source of nonprompt muons passing the FO and tight definition ($gg \to b\bar{b}$, $gg \to c\bar{c}$), with about equally large contributions from b- and c-partons at sufficiently large transverse momenta. Nonprompt electrons on the other hand are mostly coming from light-flavored partons (mostly jets faking leptons) for transverse momenta larger than 25 GeV. In the DY application region, we find that nonprompt muons are predominantly coming from heavy-flavored partons with large contributions from both b- and c-flavored partons. Nonprompt electrons, on the other hand, have about equal contributions from light and heavy-flavor partons. In the $t\bar{t}$ application region, both nonprompt electrons and muons are mainly coming from b-flavored partons.

6.2.5 Measuring the fake rate in data

Event selection Next, we measure the fake rate in a QCD multijet enriched data sample. Essentially, the fake rate is measured as the ratio between the number of tight leptons to the number of FO leptons, in a phase space meant to select mainly nonprompt leptons. To this end, we apply the following event selections:

- **Triggers** The event must pass any of the single lepton triggers listed in Tab. B.1 (in the appendix). Since these triggers are highly prescaled[4], the prescale value for each of them was measured first in order to correctly scale the simulation.

- **Lepton** The event must have exactly one lepton passing the FO definition. This lepton is required to be muon/electron when measuring the fake rate for muons/electrons, respectively. Furthermore, the lepton must pass an additional p_T selection based on the fired trigger, in order to be situated at the trigger efficiency plateau: the lepton p_T is required to be greater than the p_T value implied in the trigger name, plus a threshold of 1 GeV for muons and 2 GeV for electrons.

- **Jet** The event must contain at least one jet, with $p_T > 30$ GeV, separated from the lepton by at least $\Delta R = 0.7$. An additional jet p_T selection criterion is applied depending on the fired trigger (if applicable): the p_T of the jet is required to be greater than the p_T value implied in the trigger name, plus a threshold of 10 GeV. The presence of this extra jet is imposed to enhance the purity of the selection in nonprompt leptons from QCD multijet events (as opposed to prompt lepton contamination, for example from DY), and also serves as an implicit quality selection on the lepton, as the lepton and the jet kinematics are typically correlated in QCD multijet events.

- **Missing transverse energy** The missing transverse momentum p_T^{miss} must be smaller than 20 GeV (to suppress electroweak contamination).

The same conditions are applied to simulated samples of QCD multijet and the relevant electroweak processes (mainly W + jets, DY, $t\bar{t}$, and diboson production). Leptons are binned in p_T and η and for each bin a histogram of the modified transverse mass m_T^{mod} is filled for both denominator (as defined by the selections above) and numerator (where the lepton additionally passes the tight definition). The modified transverse mass is defined similarly to the usual transverse mass m_T:

$$m_T = \sqrt{2 p_T^{lepton} p_T^{miss} (1 - \cos(\Delta\phi))}, \tag{6.2}$$

with $\Delta\phi$ the azimuthal angle between the lepton and the missing transverse momentum. The modified transverse mass uses the same formula but where the lepton p_T is replaced by a fixed

[4]This means that not every event that fires this trigger is fully reconstructed and stored, as the resulting rate would be too high; instead only one out of every n events is selected, where n is a tuneable parameter that can vary over time.

value (35 GeV in this analysis). This variable is preferred over the usual transverse mass as it is less strongly correlated with the lepton p_{T} and hence with the fake rate.

Fake rate measurement Next, we determine the fake rate as the ratio between numerator (tight selection) and denominator (FO selection) of amounts of nonprompt leptons in the measurement region. The amount of nonprompt leptons in the numerator and denominator of each p_{T}/η bin in the measurement region is estimated in three different ways.

- **Subtraction** The amount of nonprompt leptons is estimated by subtracting the prompt leptons from simulation (mainly DY, $\mathrm{t\bar{t}}$ and W + jets) from the observed data. This method has the advantage that one does not need to rely on the simulation of nonprompt leptons from QCD multijet events. On the other hand, shape information from the distributions of simulated prompt and nonprompt leptons is not used optimally. This method is illustrated in Fig.6.8.
- **Maximum likelihood fit** Alternatively, we apply a maximum likelihood fit using the COMBINE tool [218–220], fitting the templates from prompt and nonprompt contributions to the observed data. This fitting procedure is done for each p_{T}/η bin and for numerator and denominator independently. This method is illustrated in Fig.6.9.
- **χ^2 fit** Thirdly, we employ a χ^2 fit to scale the yield of the nonprompt lepton contribution to optimally fit the data. Contrary to the maximum likelihood fit, the shapes of the templates are fixed, and only the nonprompt lepton yield is susceptible to change. This method is illustrated in Fig.6.10.

The three methods are observed to yield fake rate maps compatible within statistical uncertainties, hence the subtraction method is finally chosen for its simplicity. In both flavours of fits, the freely floating nonprompt normalization, together with the rather large statistical uncertainties (mainly of the QCD multijet simulation), gives sufficient freedom to achieve a good fit in all bins. Therefore, no additional systematic uncertainties are added as this could potentially introduce secondary local minima in the fit likelihood. A sufficiently large uncertainty on the nonprompt prediction obtained in this way will anyway be assigned based on the closure in data. The post-subtraction / post-fit amounts of nonprompt leptons in numerator and denominator are divided to obtain the fake rate. The resulting fake rate maps are shown in Fig. 6.11.

At high p_{T}, the measurement region becomes dominated by prompt contamination, which makes it hard to probe the fake rate in that region. When the prompt contribution matches the data or is very close to it, there are no events left after prompt subtraction. This is the reason for the bins with 0 fake rate or high uncertainty. In the analysis the fake rate maps are truncated at 45 GeV so these high p_{T} bins are not used.

6.2.6 Validation in data control regions

The final validation of this method is performed by verifying good data-to-prediction agreement in control regions with large contributions from nonprompt backgrounds. The definition of these control regions are however analysis-specific, as they should be as close as possible to the signal region of interest in a specific analysis, to keep the extrapolation error from control region to signal region small. For the case of the tZq and $\mathrm{t\bar{t}W}$ analyses, the relevant control regions and results will be discussed in more detail in the respective chapters.

$\mathrm{t\bar{t}t\bar{t}}$ In the case of the $\mathrm{t\bar{t}t\bar{t}}$ observation [4], several control regions with significant nonprompt contributions are defined, and the data-to-prediction agreement in these regions is shown in

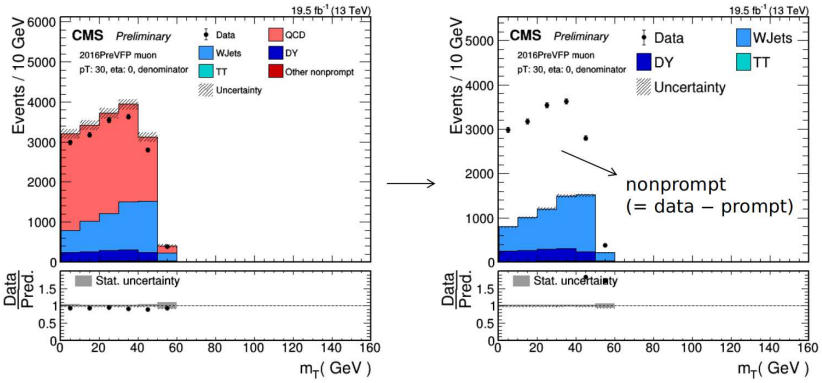


Figure 6.8: Illustration of the fake rate measurement using the subtraction method. Left: prefit distributions of data and simulation. Right: determination of the nonprompt yield as difference between data and prompt simulation. The nonprompt simulation is ignored in this method.

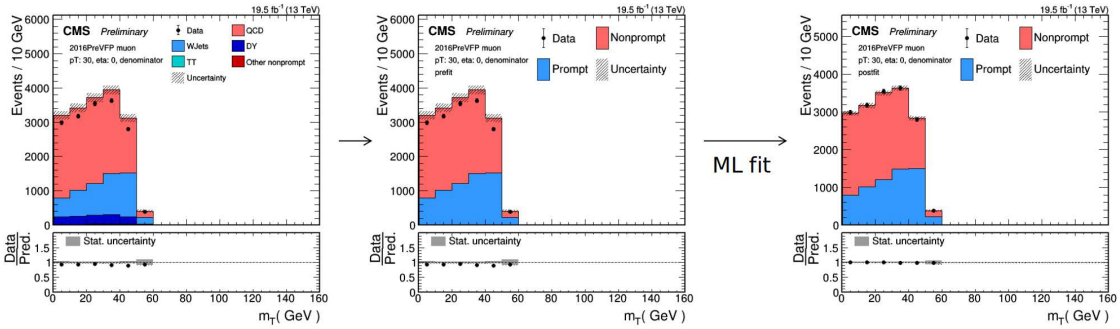


Figure 6.9: Illustration of the fake rate measurement using the maximum likelihood fit method. Left: prefit distributions of data and simulation. Middle: same distributions but grouped into one prompt and one nonprompt category. Right: determination of the nonprompt yield as integral of the postfit distribution, after performing a maximum likelihood fit of the simulation to the data.

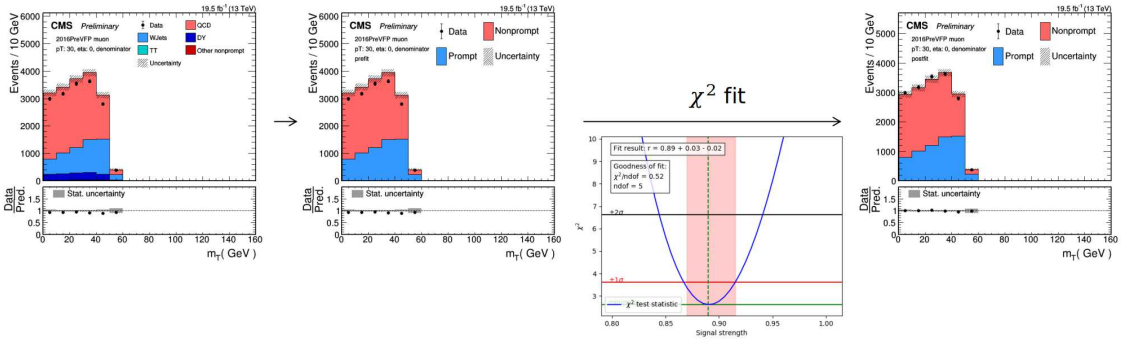


Figure 6.10: Illustration of the fake rate measurement using the χ^2 fit method. Left: prefit distributions of data and simulation. Middle: same distributions but grouped into one prompt and one nonprompt category. Right: determination of the nonprompt yield as integral of the postfit distribution, after performing a χ^2 fit of the simulation to the data. Bottom: the χ^2 value as a function of the nonprompt normalization parameter.

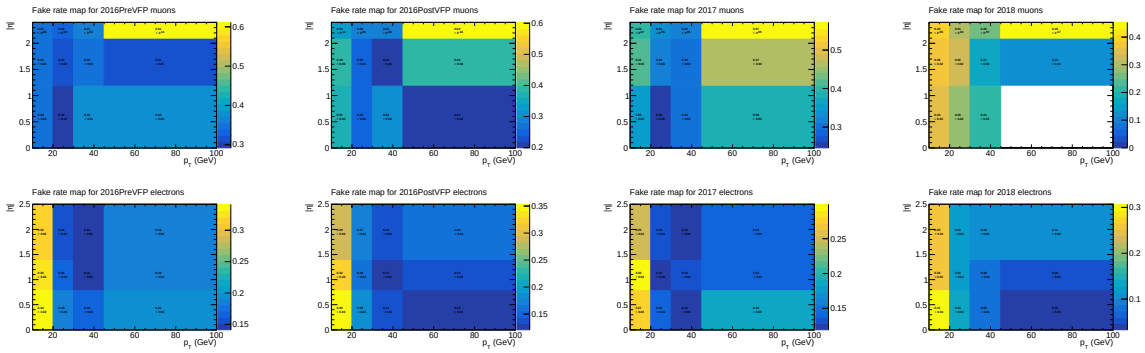


Figure 6.11: Fake rate maps obtained from data for muons (upper row) and electrons (lower row), for 2016PreVFP (first column), 2016PostVFP (second column), 2017 (third column), and 2018 (fourth column). The fake rate is determined as the ratio between the postfit nonprompt yield in the numerator to the one in the denominator.

Fig. 6.12. The definition of these control regions and their correspondence to the signal regions used in the analysis can be found in Ref. [4]. The accurate prediction of the background contribution with nonprompt leptons using the fake rate method, and its thorough validation in control regions, was an important ingredient for the first ever observation of the $t\bar{t}t\bar{t}$ process.

HNL In the case of the HNL search [5], the validation of the fake rate method is illustrated in Fig. 6.13. Two important input variables to the signal BDTs used in that analysis are shown, for the signal region optimized for low-mass HNLs and without hadronically decaying taus.

6.3 Charge misidentification background estimation using the charge flip method

Charge misidentification In CMS, the electric charge of a detected particle is generally determined from the direction of curvature of its track. For an electron specifically, the position of the associated ECAL supercluster with respect to the track can additionally be used to determine its charge. However, even when selecting only electrons for which the different charge determination methods give consistent results, there is still a small but nonzero probability of assigning the wrong charge to a reconstructed electron. This is mainly caused by bremsstrahlung emission, electrons radiating photons while passing through the detector. The emitted photons can in turn convert to electron-positron pairs, the tracks of which are difficult to distinguish from the original track.

Charge misidentification is mainly problematic in same-sign dilepton signatures, since processes such as DY and $t\bar{t}$ with large cross-sections produce opposite sign dilepton signatures that, because of charge misidentification, will have non-negligible spill-overs into the same-sign signal region.

The charge flip method in a nutshell The charge flip method is conceptually similar to the fake rate method for nonprompt leptons (see section 6.2). It consists of measuring the probability for charge misidentification (mostly in simulation, but with a correction factor from data), and then applying it in data to a sideband to the signal region, containing opposite-sign dilepton events, in order to obtain the corresponding contribution in the (same-sign dilepton) signal region. Note that similar methods have been used in e.g. [213, 214].

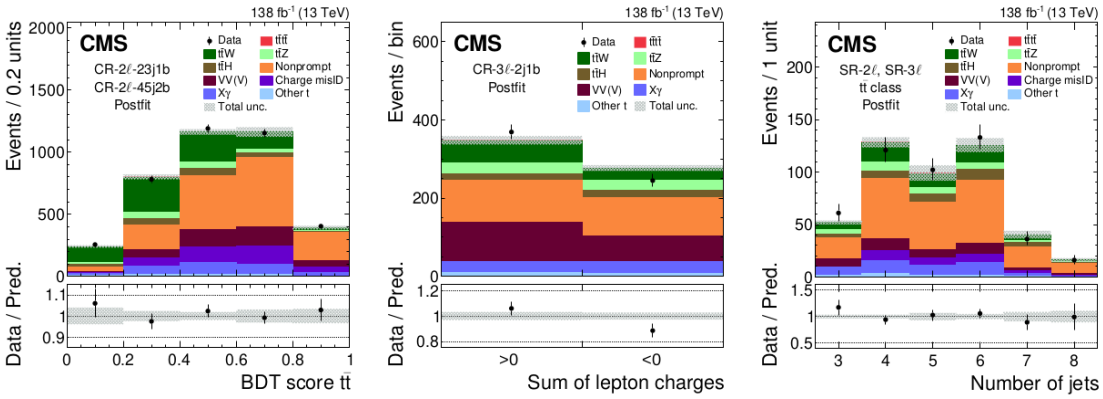


Figure 6.12: Data-to-prediction agreement in several control regions with significant nonprompt contributions in the $\mathrm{t\bar{t}t\bar{t}}$ analysis, highlighting the good prediction provided by the fake rate method. Left: score of the $\mathrm{t\bar{t}}$ output node of a multi-class BDT, in a dilepton control region. Middle: sum of lepton charges in a trilepton control region. Right: number of jets in the part of the combined dilepton and trilepton signal region where the $\mathrm{t\bar{t}}$ output score of a multi-class BDT is highest. Figure from Ref. [4].

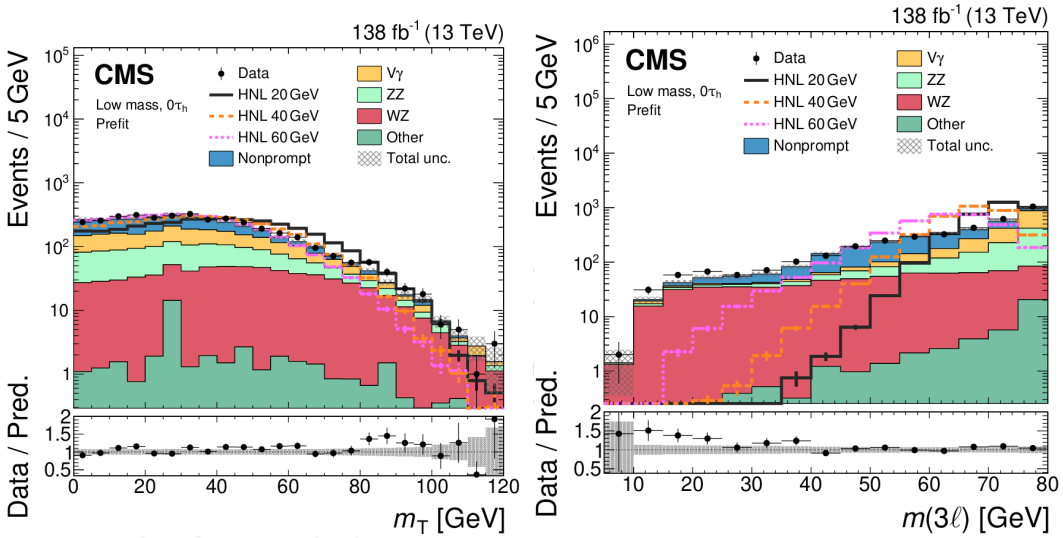


Figure 6.13: Data-to-prediction agreement in a signal region with a significant nonprompt contribution in an HNL search, highlighting the good prediction provided by the fake rate method. Left: transverse mass of the system formed by $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and a specified lepton. Right: invariant mass of the three-lepton system. Figure from Ref. [5].

6.3.1 Measuring the charge misidentification rate in simulation

Procedure We measure the charge misidentification rate in simulated DY samples using a selection requiring exactly two electrons passing the tight definition, forming an invariant mass within 15 GeV of the Z boson mass. Both electrons are furthermore required to be prompt, in order to remove potential overlap with the fake rate method: nonprompt electrons are estimated using the fake rate method regardless of whether they are assigned the correct charge. The charge misidentification rate is now defined as the number of electrons for which the reconstructed electric charge is opposite to its true electric charge, divided by the total number of electrons. The resulting charge misidentification rates are shown in Fig. 6.14 as a function of the p_{T} and $|\eta|$ of the two leptons in the event.

Choice of binning This measurement is inherently limited by the low charge misidentification rates, causing relatively low statistical precision in the measured rates. As a mitigation strategy, a binning is chosen that is coarse enough to keep the statistical precision in each bin sufficiently high, while still capturing the $|\eta|$ and p_{T} dependence.

Cross-check in $\mathrm{t\bar{t}}$ As a cross-check, the charge misidentification rates are measured in $\mathrm{t\bar{t}}$ samples instead of DY samples (omitting the event selection requirement on the invariant mass of the two leptons). As shown in Fig. 6.14, the obtained rates are consistent with those from DY except in a small number of bins, in which no real systematic effect (e.g. consistent over the years) can be discerned. These bins only have a minor impact however, and the closure in both DY and $\mathrm{t\bar{t}}$ samples is satisfactory with the charge misidentification rates as obtained from DY (see further on).

Muons Up to now, muons have been neglected in this discussion. Their charge misidentification rates can in principle be measured in the same way as for electrons, but the values are typically in the order of 10^{-6} and consistent with zero. These rates are much lower than for electrons, on the one hand because of the higher mass of muons causing them to emit less bremsstrahlung photons, and on the other because of the gaseous outer detectors allowing a track curvature determination over a much larger trajectory length than for electrons. The muon charge misidentification rate will be ignored (i.e. assumed to be zero) in the further analysis.

6.3.2 Simulated closure tests

The charge misidentification rates are used in the form of weights applied to an opposite-sign dilepton selection. Three distinctions can be made based on the flavor of the two leptons:

- ***μμ* channel**: Since the charge misidentification rate for muons is negligibly small, the weight for this type of events is zero.
- ***eμ* channel**: The weight w for this type of events is determined by the electron charge misidentification rate $P(e)$:
$$w = \frac{P(e)}{1 - P(e)} \tag{6.3}$$
- ***ee* channel**: Each of both electrons could be charge-misidentified, therefore the weight is given by:
$$w = w(e_1) + w(e_2) - w(e_1) \cdot w(e_2) \tag{6.4}$$
where $w(e_1)$ and $w(e_2)$ are calculated as in the previous case.

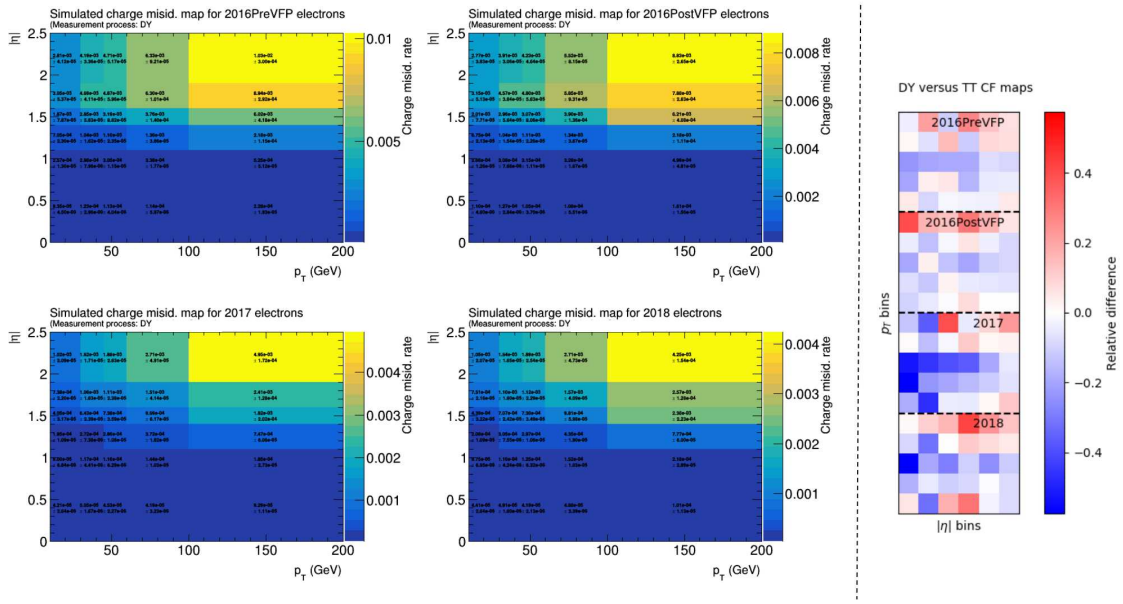


Figure 6.14: Charge misidentification rates for electrons as a function of p_{T} and $|\eta|$ for the data-taking years used in this work. The right overview figure shows the relative difference between the charge misidentification rates as obtained from DY versus $\mathrm{t\bar{t}}$ samples.

In order to check that the chosen binning is sufficiently fine to model the kinematics in the prediction, we perform a simulated closure test by applying the method to simulated DY and $\mathrm{t\bar{t}}$ samples in an opposite-sign selection, and compare the distributions of some observables to the same-sign selection. These results are shown in Fig. 6.15 (for samples corresponding to 2018 data-taking conditions, with similar results for other data-taking years). The good agreement as a function of the lepton kinematic variables shows that the measured charge misidentification maps can model the kinematic properties sufficiently accurately. Furthermore, the good agreement in the $\mathrm{t\bar{t}}$ samples in particular shows that the charge misidentification rate determined in DY is also applicable to other signatures.

6.3.3 Correction factors in data

Event selection Next, we perform a closure test in data. The event selection for this test is the same as before, but omitting the promptness requirement (as it uses generator-level information not available in data):

- The event must contain exactly two selected electrons.
- The invariant mass of both electrons is sufficiently close to the Z boson mass: $\left|m(\mathrm{ee}) - m_{\mathrm{Z}}\right| < 15$ GeV.
- The two electrons must have the same electric charge (or opposite electric charge for the data sample from which the same-sign prediction is obtained using the charge flip method).

Results The results are shown in Fig. 6.16. An under-prediction of the order of 40-50% is observed in both 2017 and 2018, while the agreement in 2016 is better (with a slight over-prediction in 2016PreVFP however). These differences and discrepancies have been observed before [216, 217, 221] and are caused by the strong dependence of the charge misidentification

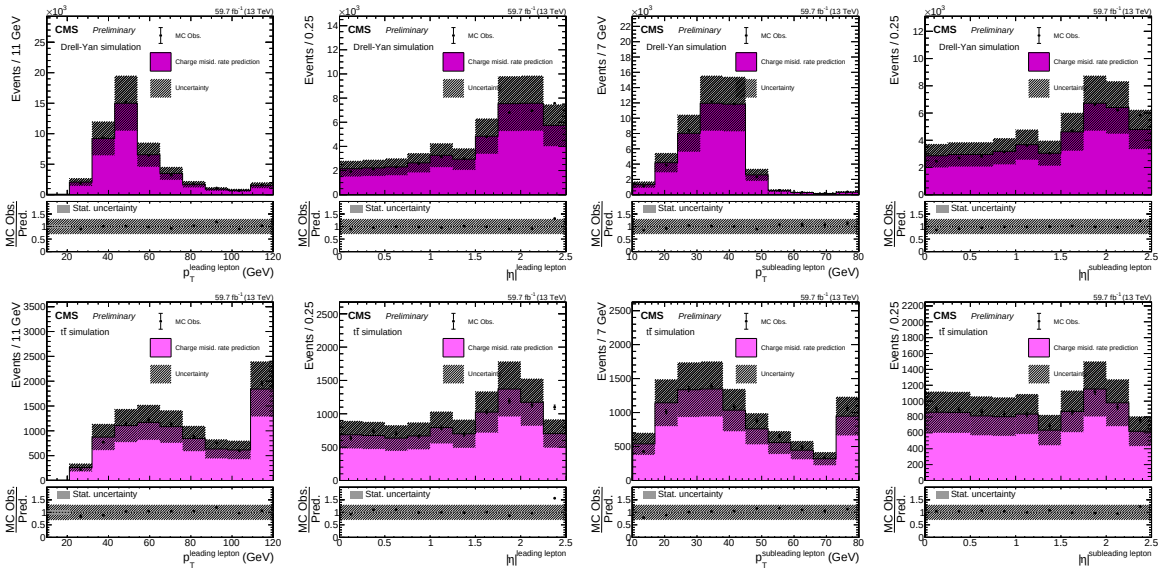


Figure 6.15: Simulated closure tests for the 2018 charge flip method, in DY (upper row) and $t\bar{t}$ (lower row). Closure is shown as a function of the leading and sub-leading lepton p_T and $|\eta|$.

rate on the amount and type of detector material that the electrons have to cross. While in principle this is included in the simulation, small effects such as misalignments, radiation damage, inhomogeneities in the magnetic field, and multiple scattering of electrons off detector material are not simulated perfectly and can have a large impact on the charge misidentification rate. The qualitative difference between 2016 and 2017/2018 is most likely caused by the upgrade of the pixel tracker in between those two data-taking periods.

Calibration factors In order to sensibly apply this method in signal regions, it has to be calibrated by deriving correction factors from the measurements in Fig. 6.16. These correction factors are flat, i.e. they are one scale parameter per data-taking year that impacts the integral but not the shape of the distributions. Good agreement between data and prediction is observed after applying this correction. The right column of Fig. 6.16 shows that it does not matter which observable is used to determine the correction factor; they all give the same result. The $m_{\ell\ell}$ observable constitutes an exception to this, because its shape is badly modeled by the charge flip method (see Fig. 6.17); this is expected since electrons whose charge was misidentified have lost more energy to bremsstrahlung photons on average, causing the mass peak to be shifted to lower central values than the expected value of the mass of the Z boson.

6.4 Trigger efficiency measurement

Phase space and triggers The baseline lepton phase space in this work consists of a dilepton selection (two same-sign leptons with transverse momenta larger than 25 and 15 GeV) and a trilepton selection (three leptons with transverse momenta larger than 25, 15 and 15 GeV)[5]. Additionally, we employ several control regions with two, three or four leptons. The trigger strategy is uniform across all these regions and consists of a combination of single lepton,

[5]There is a slight inconsistency between the trilepton signal region in the tZq analysis, with p_T thresholds of 25, 15 and 10 GeV, and the trilepton signal region in the $t\bar{t}W$ analysis, with p_T thresholds of 25, 15 and 15 GeV. The results for the trigger efficiency measurement are however found to be not significantly different between both selections, so only the latter ones are shown here

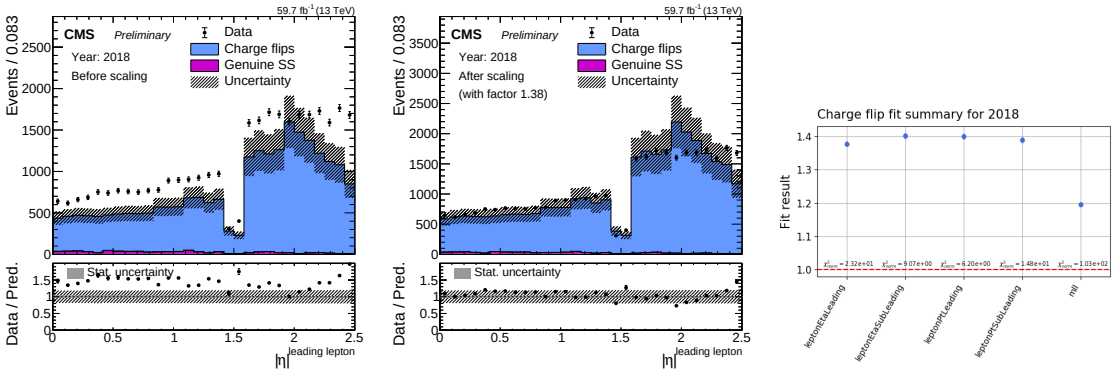


Figure 6.16: Derivation of correction factors for the charge flip method. The left column shows the distribution of $|\eta|$ of the leading lepton in data and prediction (including the uncorrected charge flip method prediction in blue). The middle column shows the same distributions after optimally scaling the charge flip probabilities. The right column shows the scaling factors obtained from several observables. The black points represent the data selected with the same-sign requirement; the blue histogram represents the same-sign prediction obtained from an opposite-sign data sample using the charge flip method; and the purple histogram contains genuine same-sign contamination estimated from simulation (DY and $t\bar{t}$).

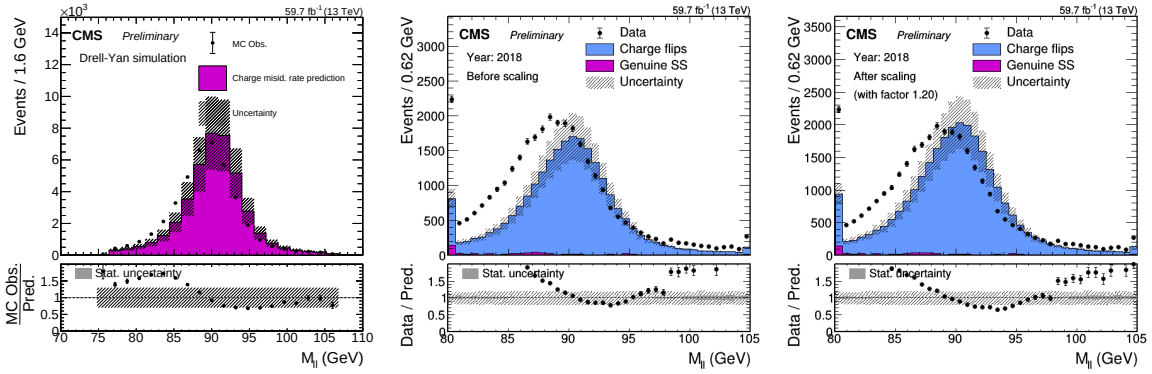


Figure 6.17: Distribution of the invariant mass of the two leptons. Left: simulated closure test in a DY sample. Middle: closure test in data before applying the correction factor. Right: closure test in data after applying the correction factor.

dilepton and trilepton triggers[6]. Events both in simulation and data are required to pass at least one of the triggers. The trigger efficiency for the leptonic phase space in this work is usually high but not perfect, implying that a non-negligible fraction of events that fall within the phase space of the event selection are not triggered on, and hence are not stored in the CMS datasets. A simulation of the trigger procedure is also applied in the simulated samples to be consistent with data, but residual differences in trigger efficiency between data and simulation still remain. These are measured and then applied as correction factors to the simulated samples.

Orthogonal trigger method The efficiency of the trigger mix in data (and simulation) is measured using the orthogonal trigger method. The idea of this method (summarized in Fig. 6.18) is to apply a lepton selection (consisting of identification criteria, p_{T}-thresholds etc.) similar to the one that will be used in the main analysis, to data sets not triggered on the presence of any lepton. By checking what fraction of the selected events additionally passes the lepton triggers, we obtain an unbiased estimate of the efficiency of these triggers for the events of interest in this analysis. This procedure is performed for dilepton, trilepton and four-lepton event selections separately.

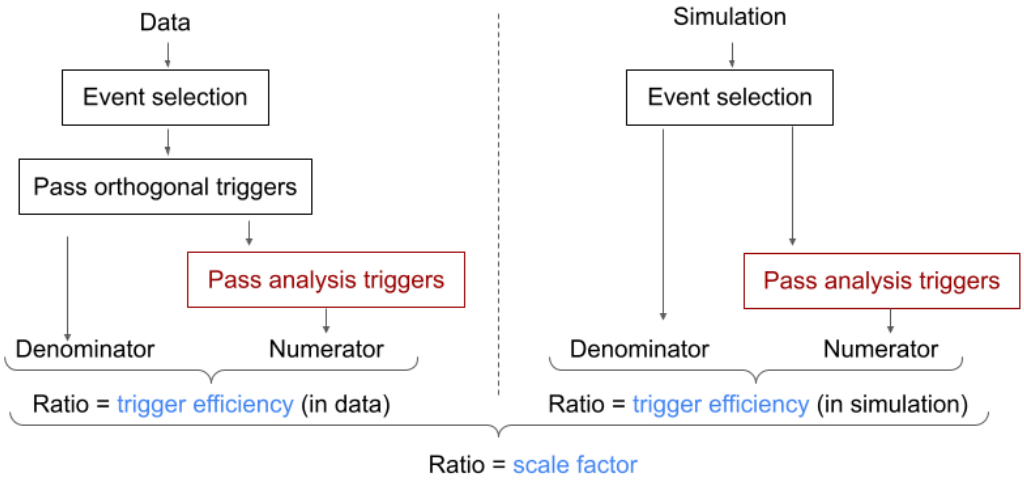


Figure 6.18: Schematic overview of the orthogonal trigger method for estimating trigger efficiencies and scale factors.

6.4.1 Dilepton selection

Datasets and event selection We use the MET, JetHT and HTMHT primary data sets[7], since they are unbiased with respect to lepton selection or triggering. We select events with two same-sign tight leptons passing the transverse momentum thresholds of 25 and 15 GeV respectively. Furthermore, events are required to pass a set of orthogonal reference triggers, consisting of $p_{\mathrm{T}}^{\mathrm{miss}}$ and (missing) HT triggers. In simulation, we apply the same lepton selection as in data, but not the orthogonal trigger requirement.

Results The resulting efficiencies are shown as a function of one-dimensional lepton p_{T} variables in Fig. 6.19. The plots in this figure show the efficiency of the lepton trigger mix, which is the ratio of events that pass the lepton selection and additionally at least one of the lepton

[6] A detailed list of all lepton triggers used in this work can be found in Tab. B.2 (in the appendix).

[7] These primary datasets collect events triggered by large missing transverse momentum or jet transverse momentum.

triggers to the events that are only required to pass the lepton selection. The plots also show normalized spectra of data and simulation passing the lepton selections (but not necessarily the lepton triggers). These spectra are mainly plotted as a sanity check and for highlighting the statistically important ranges for the different variables.

In practice, we use more fine-grained results, where the binning is two-dimensional (as function of the transverse momenta of the two leptons in the event), and additionally split in lepton flavour channels[8]. The differences in trigger efficiency between data and simulation are applied as correction factors to the simulation and the corresponding uncertainties are propagated to the statistical analysis (see section 6.5).

6.4.2 Trilepton and four-lepton selection

The procedure is repeated for alternative event selections with, instead of two same sign leptons, three leptons (p_{T} thresholds of 25, 15, 15 GeV respectively) and four leptons (p_{T} thresholds of 25, 15, 15 and 10 GeV respectively). The trigger efficiencies for these selections are found to be compatible with 100% within uncertainties, and hence no scale factors are applied. An uncertainty factor of 2% is however propagated to the statistical analysis (see section 6.5).

6.4.3 Orthogonality check

We also check that our orthogonal triggers are indeed unbiased with respect to lepton triggers, using the correlation ratio defined as follows:

$$c = \frac{N_{\mathrm{lepton}}/N_{\mathrm{all}}}{N_{\mathrm{lepton\ +\ orthogonal}}/N_{\mathrm{orthogonal}}} = \frac{N_{\mathrm{lepton}} \times N_{\mathrm{orthogonal}}}{N_{\mathrm{all}} \times N_{\mathrm{lepton\ +\ orthogonal}}} \tag{6.5}$$

where N_{lepton} is the number of events passing the lepton triggers, $N_{\mathrm{orthogonal}}$ is the number of events passing the orthogonal triggers, and N_{all} is the number of events without trigger requirement. The results are summarized in Fig. 6.20, for different data-taking years, event selections and sets of lepton triggers. The correlation ratio is everywhere close to 1, showing that our orthogonal triggers are indeed unbiased, as needed for the orthogonal trigger method to work properly.

6.5 Corrections and systematic uncertainties

6.5.1 Experimental uncertainties

In the simulation of various processes in proton-proton collisions, a rough distinction can be made between the simulation of the elementary particle interaction in vacuum (i.e. the process of interest), and the simulation of the interactions of all generated particles with the detector material (see section 2.13 for more details). The raw predictions obtained in this way are often found to describe the observed data not sufficiently accurately. Many of the discrepancies can however be attributed to imperfections in the detector simulation, rather than to the pure event generation, and are hence typically corrected for in order to obtain more realistic predictions with which the fundamental theory can be tested. The appropriate correction strategies, as well as the sources of systematic uncertainty that affect them, are detailed in this section.

Luminosity Simulated samples are normalized using the nominal luminosity of a given data-taking year. As some sources of systematic uncertainty affect the luminosity measurements

[8] These results are more fully documented in Ref. [15].

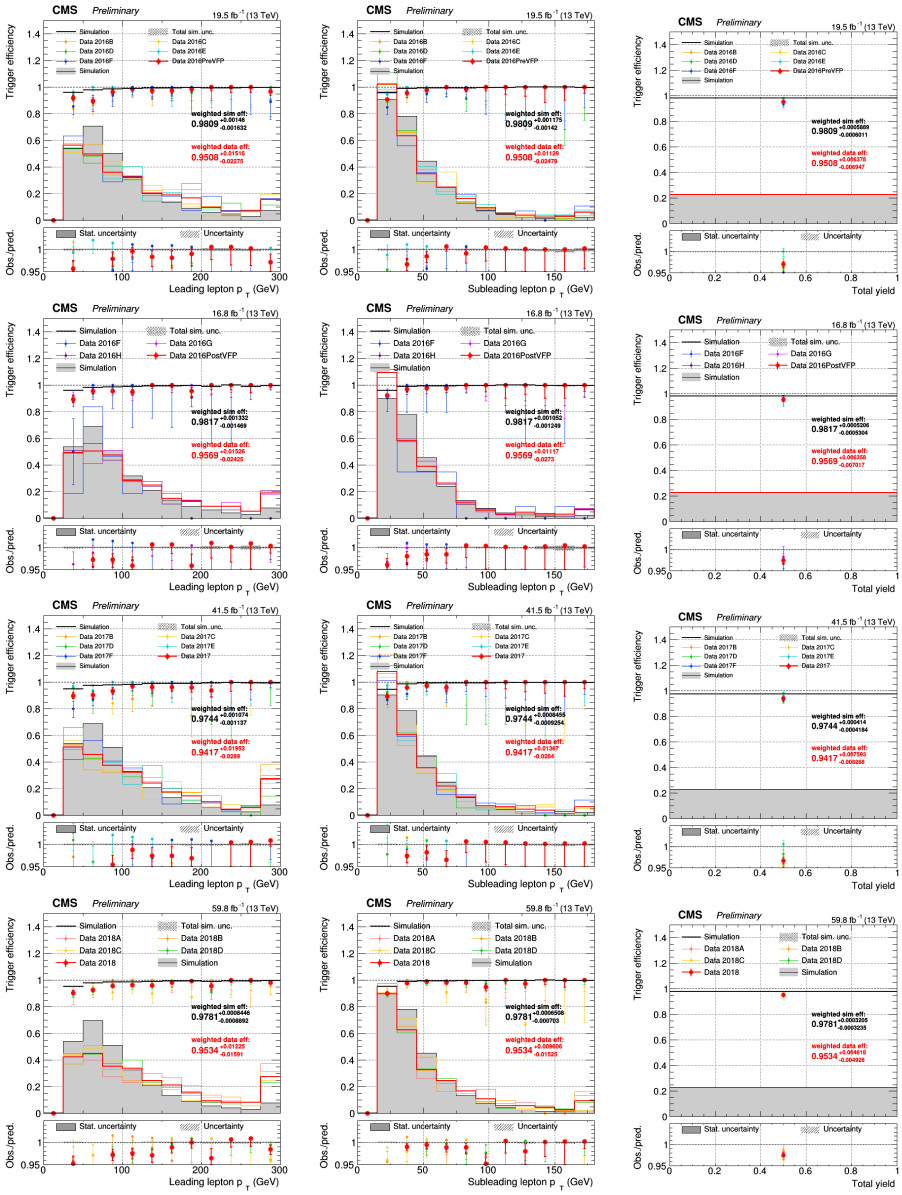


Figure 6.19: Trigger efficiency for events with two same sign tight leptons. The rows represent 2016PreVFP, 2016PostVFP, 2017 and 2018 data respectively, while the columns show the efficiency as a function of leading and sub-leading lepton p_T and total yield respectively. Normalized spectra of data and simulation passing the lepton selections are shown as well.

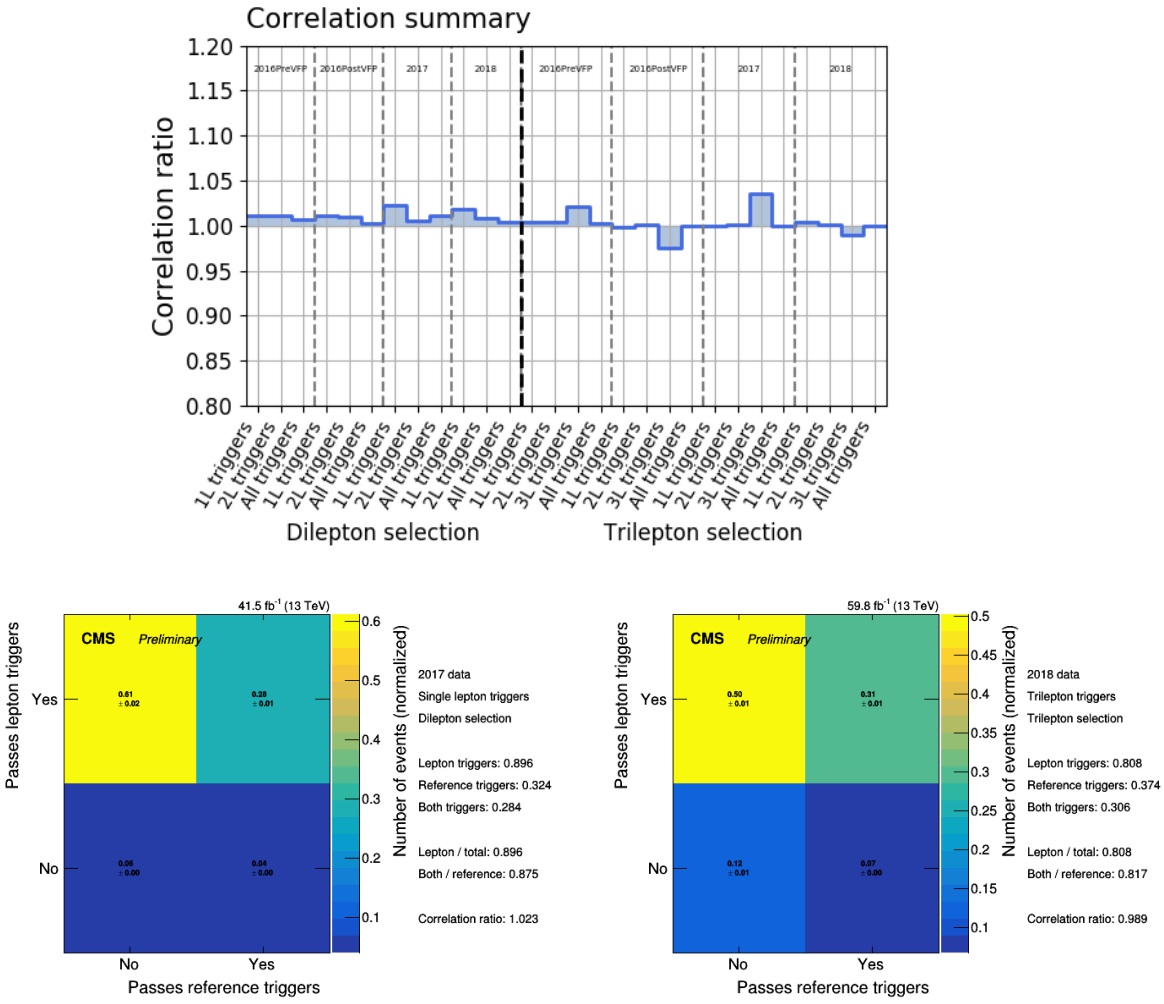


Figure 6.20: Check of the correlation between the lepton triggers and orthogonal triggers used in the orthogonal trigger method. Upper: summary of the correlation ratio across all data-taking years, lepton selections and triggers. Lower row: two examples of calculations for single bins in the summary.

(see [127–129]), we take into account uncertainty factors of 1.2%, 2.3% and 2.5% for 2016, 2017 and 2018 respectively, partly correlated and partly uncorrelated between the data-taking years depending on the source of uncertainty. The luminosity uncertainty affects all simulated samples equally and is applied as flat normalization uncertainty.

Trigger Uncertainty in the efficiency with which events passing the signal selection are triggered in data is taken into account as well. The trigger efficiency measurement is described in more detail in section 6.4. For the tZq analysis (chapter 7), the trigger efficiency is found to be compatible with 100% in both data and simulation, within the statistical limitations of the measurement. These statistical fluctuations are taken into account in the form of a systematic normalization uncertainty of 2%, uncorrelated between the data-taking years. In case of the $t\bar{t}W$ analysis (chapter 8), in particular the same-sign dilepton region, a reweighting procedure is applied to the simulation and the uncertainties are propagated by varying the reweighting factors within their uncertainties. For the trilepton and four-lepton regions, the same procedure as for the tZq analysis is applied.

Pileup The number of proton-proton collisions in a single bunch crossing is called pileup. Simulated samples are typically generated before the data-taking period they apply to is finished, and hence they have to make an initial assumption on the pileup distribution. This assumption might not match perfectly the pileup distribution in real data, which can lead to sizable differences in kinematic distributions of reconstructed events, as high-pileup events are typically harder to reconstruct accurately because of the higher noise fraction. When the data-taking period is finished, the pileup distribution in real data can be calculated from the measurement of the instantaneous luminosity and the total inelastic proton-proton collision cross-section (note that this is not equal to measuring the number of reconstructed vertices per event, where other effects and inefficiencies play a role). The simulated samples are then reweighted to make their distribution of proton-proton interactions match the calculated pileup distribution in data. The uncertainty on this correction is taken into account via a modified reweighting procedure, where the events in simulated samples are reweighted in such a way that the distribution of the number of proton-proton interactions matches a distribution that is obtained from varying the cross section of minimum bias proton-proton collisions by 4.6% up or down with respect to its nominal value. This reweighting procedure is applied to all simulated samples and can affect their shape as well as their normalization. Since the proton-proton inelastic interaction cross-section is a physical quantity, the pileup uncertainty is treated as fully correlated between all years and processes.

Prefiring An unexpected dynamic inefficiency in Run II was caused by ECAL energy deposits being assigned to the wrong bunch crossing. Because of a radiation-induced transparency loss of the ECAL crystals, the level-1 trigger (L1T) system displayed a small probability of assigning an ECAL energy deposit to the previous bunch crossing, rather than the current one containing the actual source of the energy deposit [73]. This effect can lead to "prefiring", i.e. the L1T incorrectly accepts the event in the previous bunch crossing, and, because of limited rate and bandwidth, this automatically leads to an incorrect rejection of the event in the current bunch crossing. As a secondary effect, the missing transverse energy in the events in both the previous and the current bunch crossing will be affected. The probability for prefiring can be measured for different types of physics objects (photons, jets, leptons) as a function of their kinematic properties. Using these measurements, the prefiring is accounted for as a reweighting correction in simulation, where each simulated event is given a weight that represents the probability for that event not to prefire (typically close to 1), calculated as a product of the non-prefire-probabilities of the individual physics objects in that event. The uncertainty on the prefire probability is propagated through the analysis by varying the

prefiring weights up and down within their uncertainties. Similar to the pileup correction and uncertainty, the reweighting is performed for all simulated samples and is treated fully correlated across the years and processes.

Lepton identification The identification efficiency of prompt leptons is typically observed to be slightly different in simulation and data. This is corrected for by measuring these differences in efficiency as a function of kinematic properties of the leptons, using the tag-and-probe method in DY events, and applying these as per-lepton reweighting factors to simulated events. Details on the measurement and its results for the lepton identification used in this work can be found in [206]. The resulting scale factors have associated uncertainties, both from statistical limitations (both in the data and simulated samples used for the scale factor measurement) as well as systematic effects such as a different signal peak template used in the fitting procedure and the definition of the dilepton invariant mass window. The effect of these uncertainties is evaluated via a reweighting procedure, where the scale factors are varied up and down within their uncertainties. We consider systematic uncertainties on the scale factors to be fully correlated among the data-taking years, while statistical uncertainties are treated uncorrelated. Furthermore, we treat electron and muon scale factors as being uncorrelated among each other, since although the scale factor measurements are conceptually similar in both cases, the sources of uncertainty are mostly due to technical aspects that differ between both approaches.

Jet energy scale and resolution Typically, the measurement of the jet energy shows some non-negligible differences between data and simulation. The correction of these effects was already discussed in section 2.12.5. We take into account the jet energy correction (JEC) uncertainty by varying the jet energy scale up and down by one standard deviation for all jets in the event, for a number of independent sources of uncertainty in the JEC. This modification is done at a stage before the event selection, in order to take correctly into account any impact on both the selection and numerical values associated with the event (including the computation of the BDT score). The effect of modifying the jet energies on the $p_{\mathrm{T}}^{\mathrm{miss}}$ is taken into account consistently. These uncertainties are correlated across all processes, and either correlated or uncorrelated between data-taking years, depending on the source of uncertainty. The similar jet energy resolution (JER) uncertainty is treated in exactly the same way. In a similar manner as for the JEC and JER uncertainties, we propagate the uncertainty related to unclustered energy contributions on the $p_{\mathrm{T}}^{\mathrm{miss}}$ through the event selection.

Electron and muon energy The simulated energy measurements of both electrons [63] and muons [69] are calibrated to data. For electrons, the DY process ($\mathrm{Z} \to \mathrm{e}^{+}\mathrm{e}^{-}$) is used, and the comparison between the resonant Z mass peak between simulation and data is used to derive correction factors. For muons, several resonances (Z, J/Ψ and $\Upsilon(1S)$) and calibration techniques are employed in parallel. Uncertainty in the energy scale and resolution corrections for electrons and muons are treated in a similar way as for jets, and are treated correlated between all data-taking years and processes. These effects are however very small in the analyses described in this work.

b tagging Tagging jets as originating from a hadron containing a b quark, a.k.a. b tagging, is also characterized by a given efficiency and mistagging rate, that are typically slightly different between data and simulation. The difference in b-tagging efficiency for jets in data and simulation is measured using a variety of methods in a variety of measurement regions, yielding consistent correction factors within their uncertainties [85]. The per-event correction factors are calculated as a product over the per-jet correction factors for all jets in the event, that represent the measured difference in efficiency of correctly b tagging a jet with given kinematic properties in data with respect to simulation. The associated uncertainty is taken into account

by varying the data-to-simulation scale factors for the b-tagging efficiency up and down within their uncertainty. Several independent variations of the scale factors are considered, treated as each being fully correlated across the processes and correlated/uncorrelated across the data-taking years depending on whether they have a systematic or statistical origin.

6.5.2 Theoretical uncertainties

Apart from the experimental (i.e. detector- or analysis-related) uncertainties detailed above, there are some sources of uncertainty of a more theoretical nature, i.e. in the pure event generation or cross-section calculation. This section explains how these are estimated in the analysis. For a little more background on the origin and context of these uncertainties, see section 2.13.

Background normalization The cross-section values used to scale all relevant background processes are subject to some uncertainty. Dedicated measurements of the WZ, ZZ, Zγ, t$\bar{\text{t}}\gamma$ and t$\bar{\text{t}}$Z processes yield current uncertainties in the order of 4-8% [176,178,222–224]. However, since the phase space of these dedicated measurements does not necessarily correspond to the tZq/t$\bar{\text{t}}$W measurement phase space, we apply larger flat a priori normalization uncertainties to these background processes or keep their normalization freely floating in the fit. The details are discussed in more detail in the respective chapters.

PDF Imperfect knowledge of the parton distribution functions (PDFs) for the colliding protons yields a source of systematic uncertainty. The PDF uncertainty is evaluated by reweighting all simulated samples with a set of weights corresponding to variations in the PDFs (where available) [91]. The distributions of all kinematic variables as well as the BDT score are evaluated for each PDF variation, and the root-mean-square of these variations is computed as final shape uncertainty. Note that in principle, varying the PDF influences both the (a priori) cross-section and the (a posteriori) acceptance and shape of all processes. However, for processes that already take a fixed and flat normalization uncertainty or whose normalization is the quantity of interest, we take into account only the acceptance/shape effect. Both normalization and acceptance/shape effects are treated as fully correlated between the data-taking years and between the processes to which they are applicable.

Renormalization and factorization scales Uncertainty in the modeling of simulated processes resulting from missing higher-order contributions is estimated by varying the renormalization and factorization scales. All simulated processes are reweighted using both independent and correlated variations of both matrix element calculation scales (where available), and the envelope is taken as a final shape uncertainty[9]. The same remark concerning the shape and cross-section uncertainty holds as was the case for the PDF uncertainty. These sources of uncertainty are similarly correlated between the data-taking years but are treated uncorrelated between the processes to which they are applicable.

Initial state and final state radiation Uncertainties related to the initial-state and final-state radiation (ISR and FSR respectively) are evaluated by a reweighting procedure as well. All simulated samples for which these weights are available are reweighted corresponding to variations in the matrix element renormalization scale (effectively varying the strong coupling constant) for emissions in ISR and FSR of a factor 2 (up and down). These variations are

[9]The recommended procedure is a little ambiguous. However, it was checked that switching between three major options, namely using the envelope of all correlated and uncorrelated variations, using only correlated variations, and using only uncorrelated variations, had a negligible impact on the final results.

considered to be fully correlated across the data-taking years and processes.

Underlying event The uncertainty in the tuning of the parton shower settings [94] related to the underlying event is estimated from separate samples, generated with alternative sets of these settings. These alternative samples are processed in the same way as the nominal samples, resulting in a slightly different distribution from which the cross-section is extracted. Contrary to the uncertainties above, the underlying event uncertainty was only evaluated in the tZq analysis (and only for the tZq and $t\bar{t}Z$ processes), as its effect is expected to be negligible in other backgrounds. This uncertainty is treated fully correlated across the data-taking years and processes.

Color reconnection Similarly as for the underlying event uncertainty, color reconnection uncertainty is estimated using an alternative set of samples for tZq and $t\bar{t}Z$, where the default color reconnection model is replaced by alternative models [225]. This uncertainty is treated fully correlated across the data-taking years and processes.

6.6 Cross-section measurement using the maximum likelihood fit method

6.6.1 Introduction

Cross-section measurements Despite the seeming complexity and variety of measurements performed at CMS, many of them are essentially cross-section measurements, i.e. measurements of the elementary probability for a given process to occur. From 'simple' cross-section measurements, more complex and composite information can often be extracted. For example, the top quark spin asymmetry can be determined from the cross-section of the tZq process, measured as a function of the top quark polarization angle [1]. In this section, we discuss the most common technique in CMS to measure a cross-section: the maximum likelihood fit, which will be used to obtain the main results in chapters 7 and 8.

Parameter of interest Consider a set of measured data points, and a model that is supposed to describe these data. Often, the model contains one or more parameter of which the value is not fixed a priori; these free parameters can be tuned to a value such that the model optimally describes the data. For the simple example of a linear model $y = a \cdot x + b$ (where some variable y is measured as a function of another variable x), the slope a and intercept b are examples of free parameters. Typically, we are not only interested in how well the model fits the data, but also in the specific values of the parameters for which the fit is best. For realistic models in CMS analyses, the main free parameter of interest (POI) will be the cross-section of a given process; measuring the cross-section then implies finding out for which value of the cross-section the model optimally describes the observed data.

Likelihood Likelihood is a quantitative measure of how well a model describes the data. It is defined as the sampling probability of the observed data from the model, or in statistics notation:

$$L(\text{model} \mid \text{data}) = P(\text{data} \mid \text{model}) \tag{6.6}$$

One can tune the free parameters of a model to the data by performing a maximum likelihood fit, that is to find the values of the free parameters for which the likelihood of the model with respect to the observed data is maximized.

6.6.2 The likelihood

Form of the likelihood In CMS analyses, the likelihood is typically binned, and, in its most basic form, has the following shape:

$$L = \prod_i P(y_i^{\text{obs}}|y_i^{\text{pred}}) = \prod_i \frac{\exp\left(-y_i^{\text{pred}}\right)(y_i^{\text{pred}})^{y_i^{\text{obs}}}}{y_i^{\text{obs}}!} \tag{6.7}$$

In each bin (labeled by i), one calculates the Poisson probability of observing y_i^{obs} events when the model predicts y_i^{pred}. The predicted number of events in each bin depends on the model parameters as follows:

$$y^{\text{pred}}(\sigma_s, \vec{\theta}) = s(\sigma_s, \vec{\theta}) + \sum_j b_j(\vec{\theta}) \tag{6.8}$$

The expected number of events (in each bin) is the sum of all expected contributions: the signal process contribution s (depending on its cross-section σ_s) and the various background process contributions b_j. Both signal and backgrounds are affected by a number of other model parameters $\vec{\theta}$, typically called nuisance parameters [219, 220]. They encode the effect of the various systematic uncertainties, as discussed in section 6.5. The likelihood now becomes:

$$L = \prod_i P(y_i^{\text{obs}}|y_i^{\text{pred}}) \cdot \prod_k p(\theta_k) \tag{6.9}$$

where the added factor represents the probability distributions of the nuisance parameters, typically Gaussian distributions. This additional factor is a translation of the fact that we often have *some* idea what the value of the parameter is; the higher the degree of uncertainty, the wider the probability distribution is chosen.

Nuisance parameters The effect of including systematic uncertainties in the statistical model is sketched in Fig. 6.21. By using the likelihood in the form of equation 6.9, fitting the model to the data takes into account effects both from the signal cross-section of interest and from the relevant nuisance parameters. While Fig. 6.21 shows only one distribution and one nuisance parameter, real CMS analyses typically use many distributions (in signal and control regions) with O(100) nuisance parameters (see chapters 7 and 8 for examples).

Statistical uncertainties Apart from systematic uncertainties, statistical uncertainties are included in the model as well. The statistical uncertainty on the data is already implicitly included in the model by using Poisson distributions in the likelihood. The statistical uncertainty on the prediction can be taken into account by adding a set of nuisance parameters to the model, each of which varies the yield of the prediction within a single bin with a magnitude determined by a weighted summed Poisson error from all contributions [220, 226].

6.6.3 Maximum likelihood fit and profiling

Maximum likelihood fit Once the likelihood is fully defined (most of the work in which is typically taken up by estimating how the prediction varies under the influence of the nuisance parameters), it can be maximized in order to find the best value for the POI. For computational reasons, one does not maximize L directly, but rather minimizes the negative log likelihood.

Profiling Minimizing the negative log likelihood finds the combination of POI (e.g. the signal cross-section σ_s) and nuisance parameters ($\vec{\theta}$) that optimally describe the data. These best-fit

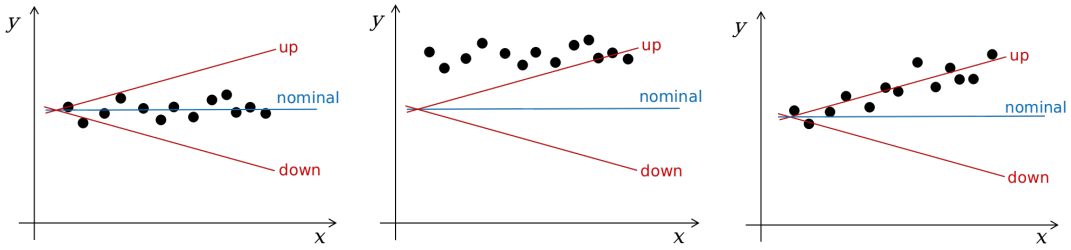


Figure 6.21: Example of systematic variation and its impact on a maximum likelihood fit. The three plots show the same nominal prediction (in blue) and the varied predictions in the case of varying a parameter θ of the model up or down with respect to its nominal value (in red). Left: nominal model matches the data. This suggests that both the cross-section and θ are predicted accurately. Middle: the model is shifted with respect to the data by a normalization constant. This suggests the cross-section is higher than predicted, but θ is modeled accurately. Right: the up-varied model matches the data. This suggests the cross-section is modeled accurately, but θ is higher than its nominally expected value.

values are denoted as $\hat{\sigma}_s$ and $\hat{\vec{\theta}}$. The naive minimization does not provide an estimate on the uncertainty. Therefore, a scan of the profile likelihood ratio is performed, defined as:

$$-2\Delta \ln L(\sigma_s) = -2 \ln \left(\frac{L(\sigma_s, \hat{\vec{\theta}}_{\sigma_s})}{L(\hat{\sigma}_s, \hat{\vec{\theta}})} \right) \tag{6.10}$$

In this expression, the denominator is the likelihood evaluated at its best-fit point. In the numerator, the nuisance parameters are re-optimized for a given and fixed value of the POI σ_s. The conventional factor 2 is added since the profile likelihood ratio defined in this way is approximated (in the large sample limit) by a χ^2 distribution [227]. This implies that the single-standard-deviation uncertainty on the POI can be estimated by evaluating the profile likelihood ratio for a range of POI values and finding the intersection with 1 [228, 229]. This procedure is illustrated in Fig. 6.22. When there are multiple POIs, the profile likelihood ratio is approximately distributed as a χ^2 with a number of degrees of freedom equal to the number of POIs.

6.6.4 Statistical significance

Before it becomes feasible to measure the cross-section of a given process using a maximum likelihood fit, one has first to observe the process, i.e. establish a discrepancy between data and the background-only prediction in a specific part of the phase space where the signal is expected to be. This discrepancy is quantified in terms of statistical significance σ[10], which is linked to a p-value via the standard normal distribution:

$$\sigma = \Phi^{-1}(1-p) \tag{6.11}$$

$$\Phi(\sigma) = \frac{1}{\sqrt{2\pi}} \int_{-\inf}^{\sigma} \exp\left(-x^2/2\right) dx \tag{6.12}$$

In practice, either toys or analytic approximations are used to construct a predicted distribution of the likelihood ratio test statistic. The p-value is calculated by comparing the observed

[10] Not to be confused with the σ representing a cross-section; the context will make this clear.

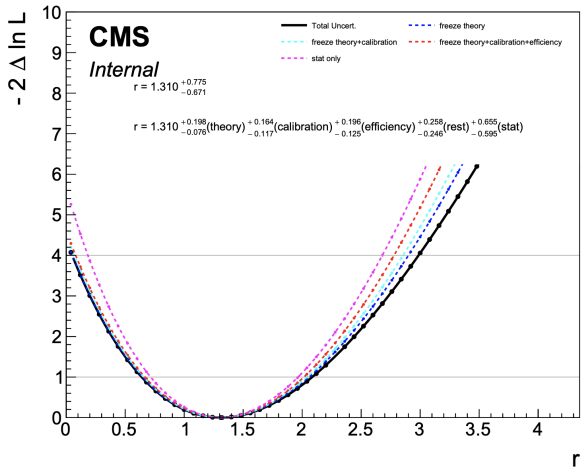


Figure 6.22: Example scan of the likelihood near the value that maximizes it. The different curves are obtained by fixing (or 'freezing') some nuisance parameters to investigate the effect of the remaining ones in more detail. The intersections of the curves with the lower and upper horizontal lines define a 1σ and 2σ uncertainty interval respectively. Figure from the COMBINE tool website.(*)

(*) `https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit`

value of the test statistic to the predicted distribution. Finally, the p-value is converted into a significance for easier interpretation. The criterion for discovery in this field of science is typically 5σ, or $p < 3 \times 10^{-7}$. Both the tZq and $t\bar{t}W$ process have been observed with $> 5\sigma$ significance in earlier work, hence these values for tZq in chapter 7 are reported for completeness and as a confirmation only.

6.7 Differential cross-section measurements

In differential measurements, rather than determining a single number, the cross-section of the process of interest is measured as a function of a given observable. This can be readily incorporated into the framework of binned maximum likelihood fits by considering not one signal (denoted as s in equation 6.8) but rather a collection of signals, one per bin in the observable, and fitting the cross-section for each of these signals simultaneously. The details of this maximum-likelihood approach to differential measurements are discussed in this section[11][12].

6.7.1 Definition of the particle level and fiducial phase space

Unfolding The data distributions as measured by the CMS detector are biased and transformed with respect to the 'true' distributions, because of unavoidable inefficiencies, imperfections and spatial constraints. In order to obtain differential cross-section distributions that represent more closely the true shapes (i.e. as they would be in vacuum, before any interaction

[11] This is just one method to perform differential cross-section measurements; other methods exist but are not discussed in this work.

[12] Both the tZq and the $t\bar{t}W$ analyses contain differential measurements. However, the discussion in this section is focused on the case of the $t\bar{t}W$ analysis, as the differential part of the tZq analysis was mainly carried out by David Walter from the DESY group. Still, the focus here is on the methodology only, while analysis-specific details and results are given in section 8.8.

with detector material), that can be compared to other experiments or theory predictions, a so-called 'unfolding' procedure is needed. Unfolding is the opposite operation of the forward 'folding', i.e. convoluting the true distributions with detector effects. Unfolding can be performed to several levels of simulated truth. In this study, we choose the so-called 'particle level', because of the relatively small extrapolation from detector level and small dependency on the choice of generator.

Particle level The particle level used in this study corresponds to stable final state particles including parton showering and hadronization but before any detector simulation is performed[13]. Basic selection criteria are applied to the particle level objects and events in order to reduce the extrapolation between detector level and particle level. This set of particle level selections forms the fiducial phase space. The particle level objects are defined as follows:

- **Leptons** are dressed by clustering prompt electrons and muons with stable photons not arising from hadron decays, using the anti-k_T algorithm with a cone size of $R = 0.1$. Only particle level leptons with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$ (electrons) or $|\eta| < 2.4$ (muons) are selected for further processing.
- **Jets** are clustered using the anti-k_T algorithm with a cone size of $R = 0.4$, starting from all stable particles excluding prompt dressed leptons and prompt neutrinos. Only particle level jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.4$ are selected for further processing.
- **b-jets** are identified in the collection of jets using the so-called 'ghost clustering' procedure, where decayed b-hadrons are not removed from the particle collection but instead scaled to an infinitely soft particle. A jet is identified as a b-jet if such a b-hadron is found among its constituents.
- **MET** is defined as the vector sum of all neutrinos from W, Z or prompt τ decays projected on the transverse plane.

We apply some loose selection requirements on the events at particle level, in order to delimit a theoretically well defined measurement region and signal definition (typically called the 'fiducial' phase space). As this event selection is very analysis-specific, it is discussed in section 8.8.

6.7.2 The principle of likelihood based unfolding

Discretization In a differential cross-section measurement, the quantity of interest is the distribution of the cross-section σ over a variable x, written as $d\sigma/dx$. Integrating over x, we retrieve the inclusive cross-section:

$$\sigma = \int \frac{d\sigma}{dx} dx \tag{6.13}$$

The problem of measuring $d\sigma/dx$ is transformed into a discrete procedure by dividing the full range of x into a number of bins and measuring partial cross-sections, e.g. in (generator-level) bin k, we measure σ_k, defined as:

$$\sigma_k = \int_{x_k^{\mathrm{low}}}^{x_k^{\mathrm{high}}} \frac{d\sigma}{dx} dx \tag{6.14}$$

where the integration is done over the bin k. This partial cross-section cannot be measured directly in the corresponding detector-level bin, because of measurement imperfections.

[13] From a technical point of view, we use the Rivet-based ParticleLevelProducer (CMS internal), as in Refs. [1] and [178].

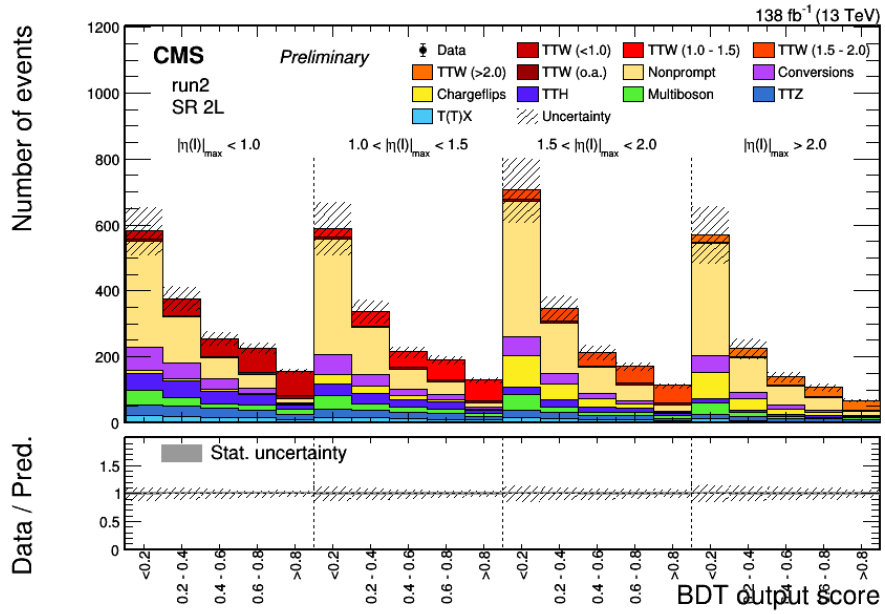


Figure 6.23: Example of how a differential cross-section measurement is performed using the maximum likelihood fit method. The events are binned at detector level in both the BDT score (to discriminate $t\bar{t}W$ from its backgrounds) and the measured variable (H_T in this case). Furthermore, the $t\bar{t}W$ signal is split at particle level in different H_T components (different shades of red), where each particle level contribution is dominant in the corresponding detector level bin.

Splitting at generator level To mitigate these effects, a maximum-likelihood based unfolding technique is applied. The signal sample is split into several non-overlapping contributions according to the generator level bins of a given variable x. Each of these contributions has its own predicted cross-section value σ_k^{SM} and to each we assign a signal strength parameter μ_k defined as $\sigma_k^{\text{meas}}/\sigma_k^{SM}$. The signal strength modifiers are directly determined in a simultaneous maximum likelihood fit of all background and signal contributions to the data.

Distributions used in the measurement The distributions used in the fit are designed to simultaneously separate the $t\bar{t}W$ signal from the backgrounds, and the different $t\bar{t}W$ contributions from each other. The first separation is obtained by using the event BDT, while the second one can be achieved by using the variable x. This variable is expected to give a good separation between the signal contributions, since while the reconstruction is not perfect, we still expect the reco-level bin k to be dominated by the corresponding generator-level component k, at least when the binning is chosen suitably. An example is shown in Fig. 6.23.

6.7.3 Response matrices and choice of binning

Response matrices Response matrices are a tool to assess the smearing of measured quantities at detector level with respect to generator level. They can be used to devise a reasonable binning for the differential measurement, by demanding that the matrix is sufficiently diagonal, i.e. that its non-diagonal entries are sufficiently close to zero. In this way, one ensures that a given particle level component of the signal sample is dominant in the corresponding detector level bin, and hence that it can be measured accurately using the fit procedure described in the previous paragraph. An example is shown in Fig. 6.24.

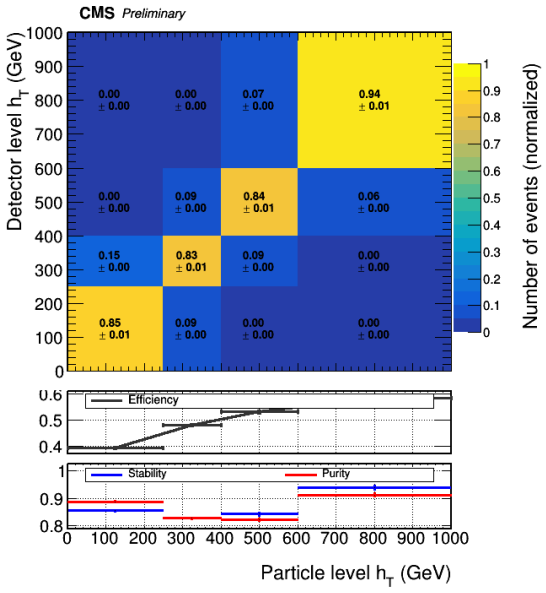


Figure 6.24: Example of a response matrix (for the H_{T} variable in this case). The x-axis shows the particle level bins, and the y-axis the corresponding detector level bins. The degree of diagonality of the response matrix gives an indication of the reconstruction accuracy of this variable.

Diagonality metrics Several quantities are defined to quantify the diagonality of the response matrix:

- The stability s_i for a given bin i is defined as the fraction of events from a generator level bin y_i that are observed in the corresponding detector level bin x_i:

$$s_i = \frac{N(\text{events in } x_i \text{ and } y_i)}{N(\text{events in } y_i)} \tag{6.15}$$

- The purity p_i for a given bin i is defined as the fraction of events from a detector level bin x_i that are coming from the corresponding generator level bin y_i:

$$p_i = \frac{N(\text{events in } x_i \text{ and } y_i)}{N(\text{events in } x_i)} \tag{6.16}$$

- The efficiency e_i for a given bin i is defined as the fraction of events from a generator level bin y_i that are reconstructed and selected at detector level (in any detector level bin):

$$e_i = \frac{N(\text{events in } y_i \text{ and selected at detector level})}{N(\text{events in } y_i)} \tag{6.17}$$

Some care needs to be taken regarding events that do not pass the selection at either particle or detector level. Events that are not selected at particle level are (by definition) not in the fiducial region. They are therefore not considered as signal and do not enter the response matrix anywhere. Some ambiguity arises with respect to events that are in the fiducial region but that are not selected at detector level: they can be either included or excluded from the response matrix and/or the definition of stability and purity. We follow the convention of not including them, so that the efficiency and smearing are more clearly factored and distinguished from one another.

6.7.4 Normalized differential cross-sections

Shape effects In the previous paragraph, we discussed (absolute) differential cross-sections. In case of $t\bar{t}W$ mismodelling (or BSM effects), there can be a mismatch between data and prediction in both the shape and the overall normalization of these distributions. The sensitivity to shape effects can be enhanced by normalizing the differential cross-sections. As an additional advantage, many uncertainties that (partly) affect the overall normalization of the prediction and measurement cancel out, resulting in smaller overall uncertainties.

Procedure The normalization of the theory prediction is trivially performed by normalizing the predicted distribution to unity. This is done both for the nominal prediction and for each of the varied templates of all systematic uncertainties, before calculating the uncertainty band. The normalization of the measured distributions is more difficult because care needs to be taken with the propagation of statistical and systematic uncertainties. For an absolute differential distribution with bins 1 to n, the cross-section value x_i in bin i is replaced by its normalized value as follows:

$$x_i \rightarrow f_i(x_1, ..., x_n) = \frac{x_i}{x_1 + ... + x_n} \tag{6.18}$$

Error propagation For any function $f(x_1, ..., x_n)$ of a number of uncertain variables $x_1, ..., x_n$, the uncertainty σ_f can be approximated by:

$$\sigma_f^2 \approx \sum_{ij} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} C(x_i, x_j) = \begin{pmatrix} \frac{\partial f}{\partial x_1} & ... & \frac{\partial f}{\partial x_n} \end{pmatrix} C \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ ... \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \tag{6.19}$$

with C the covariance matrix between the variables x_1 to x_n. For our form of f, the derivatives can be evaluated as:

$$\frac{\partial f_i}{\partial x_i} = \frac{\Sigma - x_i}{\Sigma^2} \tag{6.20}$$

$$\frac{\partial f_i}{\partial x_{j \neq i}} = \frac{-x_i}{\Sigma^2} \tag{6.21}$$

with Σ defined as $x_1 + ... + x_n$. The covariance matrix can be obtained from the correlation matrix ρ via $C_{ij} = \sigma_{x_i} \sigma_{x_j} \rho_{ij}$. The correlation matrix in turn can be obtained from the fit using the COMBINE tool [218–220].

The error propagation detailed above is performed separately for up and down uncertainties. Furthermore, it is performed separately for statistical and systematic uncertainties, the results of which are added in quadrature after propagation to retrieve the total uncertainty. Since the COMBINE fits only provide the correlation matrix of the full measurement and statistical-only measurement, the covariance matrix of the systematic uncertainties is calculated as the difference between the full and statistical-only one.

Chapter 7

Single top quark in association with a Z boson

In this chapter we study the electroweak production of a top quark or antiquark in association with a Z boson (shortly called tZq or the tZq process). The results of this research have been documented in a CMS internal note [14] and published in a paper [1]. This research project was carried out in close collaboration with David Walter from the DESY group, with his emphasis being on the differential cross-section measurements and mine on the inclusive one; therefore the discussion here will be focused on the latter aspect.

7.1 Introduction

The electroweak production of a top quark or antiquark in association with a Z boson (tZq) was observed for the first time a few years ago, in proton-proton collisions at a center-of-mass energy of 13 TeV, in both the CMS and ATLAS experiments [3, 230–232]. In both cases the measured cross-section was found to agree with the standard model prediction within the experimental uncertainties of about 15%. The study presented in this chapter confirms the observation of the tZq process and improves the accuracy of its measured cross-section.

Some representative leading order (LO) diagrams contributing to this process are shown in Fig. 7.1. The tZq production modes consist essentially of t-channel single top quark creation, modified by an additional Z boson, either radiated off from one of the quarks, or from W boson exchange. Note that, throughout the text, we use the name tZq to refer both to top quark and top antiquark production (i.e. charge conjugate states are implied), unless specified otherwise. Nonresonant lepton pairs are also included in the signal definition. The tZq process has several features that make it an interesting probe for potential BSM effects or modeling issues.

Electroweak production The tZq process is referred to as an electroweak production process, since the primary interactions (e.g. tZ, tWb, qWq', WWZ) are electroweak in nature. Therefore, the quantum chromodynamics (QCD) modeling uncertainties on the cross-section and kinematics of the tZq process are expected to be small [202], contrary to for example the $\mathrm{t\bar{t}Z}$ process, where the $\mathrm{t\bar{t}}$ pair is produced via the QCD interaction.

Couplings The tZq process is sensitive to the coupling strength between top quarks and Z bosons, as well as to the triple gauge-boson coupling (WWZ). Both of these couplings are probes of new physical phenomena not included in the standard model. For example, the measured value of the tZq cross-section could deviate from its standard model prediction in the presence of leading-order flavour-changing neutral currents [233] or more generally in the context of effective field theory [202]. The t-Z coupling can also be studied in the $\mathrm{t\bar{t}Z}$ process[1], but, as mentioned, QCD modeling uncertainties are expected to be larger. Moreover, tZq contains t-W-b vertex (and $\mathrm{t\bar{t}Z}$ does not), further enhancing the potential sensitivity to BSM effects.

Sensitivity to PDFs Furthermore, the tZq process can be split in contributions with a top quark and with a top antiquark, and the ratio of their cross-sections can be measured. This quantity is sensitive to the parameterization of the parton distribution functions (PDFs) of the proton and could help constrain PDF models. (However, as we will see, the statistical precision of the measurement in the currently available dataset is too low to contribute significantly to the PDF constraining.)

Polarization Because of the electroweak production mechanism, the top quark in tZq production is expected to be strongly polarized. While measurements of this property in t-channel single top quark production show better statistical precision [234, 235], the tZq channel provides complementary information.

This study in a nutshell This study extends on the first observations by including the full Run II data set, consisting of data taken during 2016, 2017 and 2018 at a center of mass energy

[1] Incidentally, the $\mathrm{t\bar{t}Z}$ process and the tZq process have similar cross-sections at $\sqrt{s} = 13\,\mathrm{TeV}$, as the heavier final state in $\mathrm{t\bar{t}Z}$ is compensated by the fact that it is produced via the QCD interaction rather than the electroweak one.

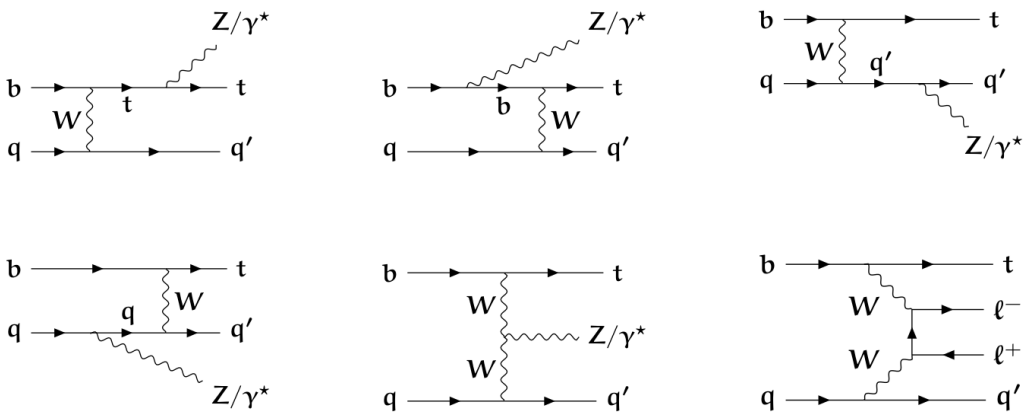


Figure 7.1: Representative leading order Feynman diagrams for the tZq process. In these diagrams, $\gamma*$ represents an off-shell (virtual) photon, ℓ is a lepton, and q and q' are generic light-flavour quarks. Charge conjugate states are implied for each diagram. The tZq production mechanism is essentially the same as t-channel single top quark production, but with an additional Z boson being radiated off from one of the initial or final state quarks, or via a triple gauge boson (WWZ) vertex. Contributions with nonresonant lepton pairs (lower right), are also included in the signal definition. Figure reproduced from [1].

$\sqrt{s}$ of 13 TeV and corresponding to an integrated luminosity of 138 fb^{-1} , in order to reduce the statistical uncertainty on the measurement. Furthermore, we apply several improvements to the analysis strategy, such as an optimized prompt lepton identification and extensive validations of background predictions in data, in order to further enhance the precision of the cross-section measurement. Finally, this work performs the first measurement of Z boson production in association with a top quark and with a top antiquark separately, as well as their ratio[2].

This measurement targets the trilepton final state, where the top quark decays leptonically and the Z boson to a pair of oppositely charged leptons of the same flavour. We exclude τ leptons from the analysis[3], and by leptons we will in this entire chapter refer to electrons and muons only, unless specified otherwise. The event selection strategy is tailored and optimized to extract events containing exactly three prompt leptons: an opposite-sign same-flavour pair from the Z boson decay[4] and a third lepton from the top quark decay. Furthermore, we split the events into three categories based on the number of jets and b-tagged jets, since the composition of the background processes is rather different in these three categories. The objects relevant for event selection and classification are depicted schematically in Fig. 7.5 and further discussed in section 7.4.

This chapter is structured as follows. In section 7.2, we list the data sets and simulation samples used in this analysis. Sections 7.3 and 7.4 are devoted to object selection and event

[2]The paper [1] also goes beyond the first observations by including differential measurements. But these are, as mentioned, not the focus of this dissertation.

[3]That is to say, hadronically decaying τ are excluded from the analysis, but electrons and muons originating from promptly decaying τ leptons are implicitly included.

[4]As specified further on, our signal definition is not limited to the on-shell Z resonance, and non-resonant lepton pairs with invariant mass larger than 30 GeV are included in the signal simulation; but the event selection targets on-shell Z resonances in order to reject background processes.

selection respectively. In both cases, a major role is played by multivariate analysis (MVA) techniques. Both the prompt lepton identification and the signal extraction rely strongly on boosted decision trees (BDTs). Then we turn to a discussion on the most important background processes in section 7.5. In section 7.6, we specify the sources of systematic uncertainty that affect this measurement and how they are taken into account. Next, the results of this measurement are quoted in section 7.7. Finally, section 7.8 presents a conclusion and an outlook for further measurements of this process.

7.2 Data and simulated samples

7.2.1 Data

This analysis uses a set of proton-proton collision events delivered by the LHC at a center-of-mass energy of 13 TeV and collected by the CMS detector. We use the full Run II data set, consisting of data collected in 2016 (36.3 fb^{-1}), 2017 (41.5 fb^{-1}) and 2018 (59.7 fb^{-1}), adding up to a total integrated luminosity of 138 fb^{-1} . Details on the luminosity measurement and calibration for the three data taking years can be found in [127–129]. We select only data samples that was certified to be of excellent quality for physics analysis[5].

Datasets In order to collect as many relevant data events as possible, we combine several primary datasets (defined as collections of triggers), where overlap between them (i.e. duplicate events) is removed based on run / lumiblock / event number. The relevant primary datasets for this analysis are those that contain trigger selections of one or more light leptons in the final state[6]. There are different reconstruction versions of each dataset, as a consequence of the fact that the original raw data can be reprocessed by CMS once important updates or improvements to the event reconstruction software (see section 2.12) have been made. The data reconstruction version used here dates from 2018[7], which was the most recent and up-to-date version at the time when this study was carried out.

Triggers On the trigger level, all events are required to pass at least one of a collection of single-lepton, dilepton and trilepton triggers[8]. Using the orthogonal trigger method, it has been checked that the trigger efficiency for the phase space targeted in this analysis is essentially 100%, with no statistically significant discrepancy between simulation and data.

7.2.2 Background simulation

Generators and parton showering In this analysis, we include Monte-Carlo (MC) simulated samples corresponding to all standard model processes relevant in the phase space of

[5] In CMS technical jargon, this certification is summarized in the so-called "golden json files". In this case, the used golden json files are the following

- `https://cms-service-dqmdc.web.cern.ch/CAF/certification/Collisions16/13TeV/ReReco/Final/Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt`
- `https://cms-service-dqmdc.web.cern.ch/CAF/certification/Collisions17/13TeV/ReReco/Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON_v1.txt`
- `https://cms-service-dqmdc.web.cern.ch/CAF/certification/Collisions18/13TeV/ReReco/Cert_314472-325175_13TeV_17SeptEarlyReReco2018ABC_PromptEraD_Collisions18_JSON.txt`

[6] In CMS technical jargon: the primary datasets are SingleElectron, SingleMuon, DoubleEG, DoubleMuon and MuonEG for 2016 and 2017 data, and EGamma, SingleMuon, DoubleMuon and MuonEG for 2018 data.

[7] In CMS technical jargon: the versions are ReReco_17Jul2018 for 2016 data, ReReco_31Mar2018 for 2017 data, ReReco_17Sep2018 for 2018 A-C data, and PromptReco for 2018 D data.

[8] The full list of used triggers per data-taking year is given in Tab. B.2 (in the appendix).

three prompt leptons[9]. Most samples used in this analysis are generated at next-to-leading-order (NLO) precision in QCD, using the MadGraph5_aMC@NLO [135] or POWHEG [236] programs. For some cases, in which NLO samples are not available or suffer from low statistics, leading order (LO) samples, generated with MadGraph [135] are used[10]. The modeling of parton showering, hadronization and the underlying event is performed using PYTHIA (v8.2) [93], with the CP5 tune [94] (or, if unavailable, the CUETP8M1 [95, 96] tune for some 2016 background samples). Two separate matching schemes are used to avoid double counting of partons generated in the different simulation steps: the FxFx [136] scheme for NLO samples, and the MLM [237] scheme for LO samples. Most samples are generated with the NNPDF3.1 PDF set [91] (with NNLO precision in perturbative QCD), while some 2016 background samples employ the NNPDF3.0 set [90].

Detector simulation All simulated events are processed by a Geant4-based [98] simulation of the CMS detector, and in a next step reconstructed using the same algorithms applied to data. For the modeling of additional proton-proton collisions present in data events (pileup), the simulated events are extended with simulated pileup events, and the distribution of the number of additional interactions in the simulated samples is reweighted to match the distribution observed in data.

Categorization The simulated samples are grouped per background category[11]. The samples listed under 'Nonprompt background' contribute to the signal region only via nonprompt leptons; these samples are only used for cross-checking intermediate results; in the final analysis, the nonprompt background is estimated directly from data (see section 7.5).

7.2.3 Simulation of the tZq process

Cross-section The predicted cross-section for the tZq process has been calculated at next-to-leading order (NLO) in perturbative QCD, yielding $\sigma_{\mathrm{tZq}}^{\mathrm{SM}} = 94.2^{+1.9}_{-1.8}$ (scale) ± 2.5 (PDF) fb for a dilepton invariant mass greater than 30 GeV [3]. The systematic uncertainty associated with the energy scale used in the calculations arises from variations of the factorization and renormalization scales, and the PDF uncertainty from variations in the PDF sets.

Simulated events Simulated tZq events for modeling the kinematics are generated at NLO in perturbative QCD as well, using the MadGraph5_aMC@NLO generator (v2.4.2) [135]. Non-resonant dilepton production and Z/γ^* interference is also included in the simulation. More technical information on the tZq sample can be found in the McM database entry[12] and the corresponding datacards[13]. The TZQ_LL_4F sample is generated as $\mathrm{pp} \to \mathrm{t\bar{b}}\ell^+\ell^- j$, where ℓ is any lepton flavour (also including τ), and j the light-flavour recoiling jet. The charge conjugate process is included as well. The minimum invariant mass of the opposite-charge same-flavour lepton pair is set to 30 GeV. The top quark is taken into account at resonance, with a mass set to 172.5 GeV, and all its allowed decay modes are included. The renormalization and factorization scales were both set to the CMS conventional choice of the sum of the

[9] A detailed list of the specific samples used to model each process is provided in Tabs. A.1–A.2 (in the appendix).

[10] A more detailed description of which generators are used for which (set of) simulated samples in this analysis is given in Ref. [1]

[11] See Tabs. A.1–A.2 (in the appendix) for details.

[12] `https://cms-pdmv-prod.web.cern.ch/mcm/requests?prepid=TOP-RunIIFall18wmLHEGS-00133&page=0&shown=262271` (CMS internal).

[13] `https://github.com/cms-sw/genproductions/tree/master/bin/MadGraph5_aMCatNLO/cards/production/2017/13TeV/tzq_ll_4f_ckm_NLO`

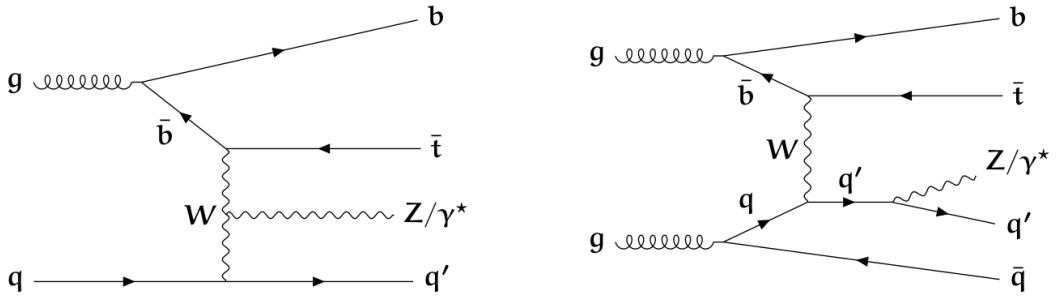


Figure 7.2: Representative LO (left) and NLO (right) Feynman diagrams for the tZq process in the 4FS. These diagrams correspond to those in the 5FS (Fig. 7.1) but the b quark in the initial state is obtained at the matrix-element level from gluon splitting rather than taken from the proton PDF. Figure reproduced from [1].

transverse masses of all final state particles and partons, divided by two.

4FS vs 5FS Two different simulation approaches exist regarding the origin of the b-quark needed in the initial state (see Fig. 7.1). In one approach, called the five-flavor-scheme (5FS), the b quark is taken directly from the corresponding parton distribution function (PDF) of the proton. In the other, called the four-flavour-scheme (4FS), the b quark is generated at the matrix-element level by the splitting of a gluon from the parton sea of the proton. The diagrams in Fig. 7.1 correspond to the 5FS; representative diagrams in the 4FS are shown in Fig. 7.2. The 5FS is expected to yield a more accurate cross-section calculation as the difficult-to-model gluon splitting is absorbed in the experimentally established PDF. On the other hand, the 4FS is expected to better model the kinematics, as gluon-gluon and quark-antiquark initial states are explicitly taken into account at the matrix element level and as the final state contains an additional particle [238]. Therefore, the cross-section prediction mentioned above was performed in the 5FS, while the events are generated using the 4FS. In the analysis, we use the 4FS sample to extract kinematic distributions, but normalize them according to the 5FS cross-section.

7.3 Object selection

This analysis relies on the particle-flow (PF) algorithm [77] to perform the reconstruction of individual particles in proton-proton collision events from the corresponding detector signals. This algorithm uses an optimized combination of information from all CMS subdetectors in order to identify muons, electrons, photons, charged hadrons or neutral hadrons. Next, the anti-k_T algorithm [81] is used to cluster reconstructed particles into jets, using a distance parameter of 0.4[14]. All these objects are used as an input for further selection and identification, as detailed in the following subsections.

7.3.1 Jets and missing transverse energy

Jets are required to pass the following selection criteria:

- $p_{\mathrm{T}} > 25\,\mathrm{GeV}$

[14]The particle-flow and anti-k_T algorithm are described in more detail in sections 2.12.2 and 2.12.5 respectively.

- $|\eta| < 5$
- The jet must be separated from any identified lepton by $\Delta R > 0.4$.
- The jet must pass the tight jet ID as defined by the JetMET POG[15].
- $p_{\mathrm{T}} > 60\,\mathrm{GeV}$ in the range $2.7 < |\eta| < 3$

The tight jet ID imposes conditions on the jet substructure (e.g. number of constituents and relative amount of charged and neutral components), aimed at removing anomalous energy deposits. In this analysis we keep jets up to $|\eta| = 5$ (which is relatively large compared to the standard jet selection of up to $|\eta| = 2.4$ or 2.5), since tZq events tend to contain a forward jet from the recoiling light quark. However, for jets with $|\eta|$ between 2.7 and 3, we impose an additional p_{T} threshold of 60 GeV to mitigate an excess of jets caused by noise in the endcap of the electronic calorimeter (ECAL)[16]. It can be checked in section 7.5 that all analysis variables show reasonable agreement between data and simulation, despite the relatively high $|\eta|$ threshold. To account for known and calibrated differences between jets in simulation and in data, jet energy corrections [82–84] are applied to all simulated samples.

Jets originating from b-hadrons, shortly called b-jets, are identified using the DeepJet/DeepFlavour algorithm [85–87]. In this analysis, we employ the medium working point as a threshold for defining b-tagged jets, which is designed to have a b jet efficiency of about 85% (for jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$). The light jet misidentification rate (i.e. the probability that a light jet is identified as a b-jet) is about 1%, and the c jet misidentification rate is about 15%. As a check, we recalculated the b-jet identification efficiency as well as the c-jet and light jet misidentification rates in simulated samples for the specific jet selection criteria used in this analysis, confirming the numbers mentioned above. Scale factors to correct for known differences in b-tagging efficiency between simulation and data are taken from central measurements by the b-tagging POG, following their recommendations[17].

To every event a missing transverse momentum vector ($\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$) is assigned, defined as the negative vector sum of all objects reconstructed by the jet finding algorithm. It represents an estimate of the combined transverse momentum (direction and magnitude) carried by particles that cannot be detected by CMS, notably neutrinos. The magnitude of this vector is denoted as $p_{\mathrm{T}}^{\mathrm{miss}}$.

7.3.2 Leptons

Accurate prompt lepton identification to reject background contributions with nonprompt leptons is paramount for this analysis. Our major tool to discriminate prompt from nonprompt leptons consists of a prompt lepton MVA (discussed earlier in section 6.1). Apart from an MVA threshold, baseline selections are applied to both muons and electrons that make sure the objects are within the limits of what can be accurately reconstructed by the CMS detector and that they are within the phase space of the lepton MVA training.

Muon selection As baseline selection (also called 'loose ID' in this chapter) for muons, we apply a p_{T} threshold (to allow accurate reconstruction) and an η threshold (to make sure the muons are within the muon system acceptance), as well as some loose cuts on impact parameter, relative isolation and a POG-defined ID. These are exactly the same as the cuts

[15] `https://twiki.cern.ch/twiki/bin/view/CMS/JetID` (CMS internal)
[16] `https://twiki.cern.ch/twiki/bin/view/CMS/JetMET` (CMS internal)
[17] `https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation` (CMS internal)

used to delimit the phase space for the lepton MVA training and testing. For the final analysis we apply an additional selection on the lepton MVA score (this selection is called 'tight ID' in this chapter). In between the loose and tight ID is a 'fakeable object (FO) ID', used in the estimation of the nonprompt background from data (see section 7.5.1). The different muon identifications for this analysis are summarized in Tab. 7.1

Table 7.1: Summary of muon identifications used in this analysis. Values between parentheses indicate that the threshold is only applied if the muon fails the lepton MVA threshold. Values separated by slashes indicate different values for 2016/2017/2018. The interpolation function $I(\mathrm{WP}_1, \mathrm{WP}_2)$ is used to mean that WP_1 is chosen for $p_\mathrm{T} < 25\,\mathrm{GeV}$, WP_2 for $p_\mathrm{T} > 40\,\mathrm{GeV}$ and a linear interpolation between WP_1 and WP_2 for intermediate p_T values. The definition of the variables is given in Tab. 6.1.

Muon identification			
Observable	Loose	Fakeable	Tight
p_T	$> 5\,\mathrm{GeV}$	$> 10\,\mathrm{GeV}$	$> 10\,\mathrm{GeV}$
$\lvert\eta\rvert$	< 2.4	< 2.4	< 2.4
$\lvert d_{xy}\rvert$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$\lvert d_z\rvert$	< 0.1 cm	< 0.1 cm	< 0.1 cm
$\lvert d\rvert/\sigma_d$	< 8	< 8	< 8
Relative isolation	< 0.4	< 0.4	< 0.4
Muon POG ID	$>$ WP-medium	$>$ WP-medium	$>$ WP-medium
DeepJet of nearest jet	-	- ($< I(0.02/0.025/0.025, 0.015)$	-
Jet p_T ratio	-	- (> 0.45)	-
Lepton MVA score	-	> 0.4 (< 0.4)	> 0.4

Electron selection The baseline selection thresholds for electrons ('loose ID') are motivated analogously as for muons. For final analysis electrons (passing the 'tight ID') we apply additional thresholds on top of the lepton MVA score, since they were observed to significantly increase the signal to background ratio. The 'fakeable object' (FO) ID is used to estimate the nonprompt background from data (see section 7.5.1). The different electron selections used in this analysis are summarized in Tab. 7.2.

Table 7.2: Summary of electron identifications used in this analysis. Values between parentheses indicate that the threshold is only applied if the muon fails the lepton MVA threshold. The $\sigma_{i\eta i\eta}$ values separated by slashes apply to barrel ($|\eta| < 1.479$) and endcap electrons respectively. The definition of the variables is given in Tab. 6.1.

Electron identification			
Observable	Loose	Fakeable	Tight
p_T	$> 7\,\text{GeV}$	$> 10\,\text{GeV}$	$> 10\,\text{GeV}$
$\|\eta\|$	< 2.5	< 2.5	< 2.5
$\|d_{xy}\|$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$\|d_z\|$	< 0.1 cm	< 0.1 cm	< 0.1 cm
$\|d\|/\sigma_d$	< 8	< 8	< 8
Relative isolation	< 0.4	< 0.4	< 0.4
Missing hits	< 2	< 2	< 2
$\sigma_{i\eta i\eta}$	-	$< 0.011/0.030$	$< 0.011/0.030$
H/E	-	< 0.1	< 0.1
$1/E - 1/p$	-	$>$ -0.04	$>$ -0.04
Conversion rejection	-	✓	✓
Charge consistency	-	✓	✓
EGamma POG MVA score	-	- ($>$ WP-loose)	-
DeepJet of nearest jet	-	- (< 0.5)	-
Jet p_T ratio	-	- (> 0.5)	-
Lepton MVA score	-	> 0.4 (< 0.4)	> 0.4

Prompt lepton MVA strategy The entire analysis presented in this chapter was carried out with two lepton identifications, differing in the working point of the prompt lepton MVA. In the first stages of the analysis, a relatively tight working point proved to be advantageous, yielding the better signal-to-background ratio after full event selection. This tight working point has a prompt lepton efficiency of 90%, while suppressing nonprompt lepton background to a level of about 1%. Apart from having the better signal-to-background ratio, there is a historical reason to suppress the nonprompt background as much as possible: since the nonprompt prediction comes with a rather large uncertainty, it proved to be the dominant source of uncertainty in the previous tZq analysis [230]. However, at a later stage, we employed a much looser working point and found the final uncertainty on the cross-section measurement to be reduced. The main reason is that both the prompt and nonprompt control regions profit from the increased number of passing events, and when included in the final maximum-likelihood fit (see section 7.7) are able to better constrain the systematic uncertainties, despite the larger nonprompt contribution to the signal region. The looser working point (with which all results in the current version of this chapter are obtained) corresponds to a signal efficiency of about 92%, and a background contamination level of about 2-3%, for both muons and electrons. These numbers are obtained from measurements in simulated tZq events for leptons with $p_\text{T} > 10\,\text{GeV}$, and are calculated with respect to the baseline selections discussed above.

In Figs. 7.3 and 7.4, we show typical ROC curves, for muons and electrons respectively, comparing the performance of the lepton MVA used in this analysis to other prompt lepton MVAs, notably the earlier tZq MVA (to which this one is a direct extension) and the MVA used in the $\text{t}\bar{\text{t}}\text{H}$ analysis. The curves in Fig. 7.3 and 7.4 were obtained from a simulated sample of

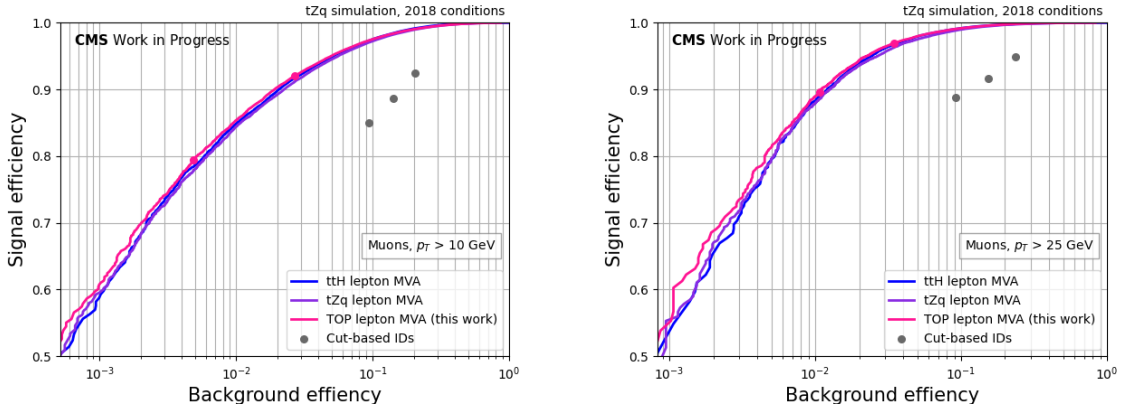


Figure 7.3: Lepton MVA ROC curves for muons and comparison with earlier lepton MVAs and cut-based selections. Left: muons with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$, right: muons with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$. In these plots, 'signal' refers to prompt muons, while 'background' refers to nonprompt muons. Efficiencies are calculated with respect to the baseline selection used in this analysis. The indicated working points of the TOP lepton MVA (pink circles) correspond to the loose and tight working point employed in this analysis, as discussed in the text. The cut-based selection criteria (grey circles) correspond to the loose, medium and tight isolation criteria, as defined by the POG [69].

tZq events corresponding to the 2018 detector conditions (with very similar results obtained for the other data-taking years). Note that the efficiencies are calculated with respect to electrons and muons that pass the baseline selection discussed in previous paragraphs, not with respect to all reconstructed electrons and muons. From these figures, we conclude that the new lepton MVA outperforms both older MVAs (though the difference for muons is negligible) and cut-based identifications.

7.4 Event selection

7.4.1 Signal region

Signal region definition The signature of interest in this analysis targets the fully leptonic decay of tZq. In a first stage, events are rejected if they do not pass any of the triggers (see section 7.2.1). Furthermore, we apply the MET filters as recommended by the JetMET POG[18] to suppress events with large amounts of missing transverse momentum coming from noise, cosmic rays and beam-halo. We also do not consider events in which no valid primary vertex could be reconstructed. Next, events are selected only if they contain exactly three leptons that pass the FO definition, vetoing any deficit or surplus. These three leptons must pass p_{T}-thresholds of 25, 15 and 10 GeV for the leading, sub-leading and trailing lepton respectively. Next, these three leptons are required to additionally pass the tight definition. As we expect two of the three leptons to originate from a single Z boson, events are required to contain an opposite-sign same-flavour (OSSF) pair of leptons, of which the invariant mass must be situated within a 30 GeV window around the nominal Z boson mass (i.e. we require $\left|m(\mathrm{OSSF}) - m_{\mathrm{Z}}\right| < 15\,\mathrm{GeV}$). Furthermore, at least two jets are required, of which at least

[18] `https://twiki.cern.ch/twiki/bin/view/CMS/MissingETOptionalFiltersRun2` (CMS internal)

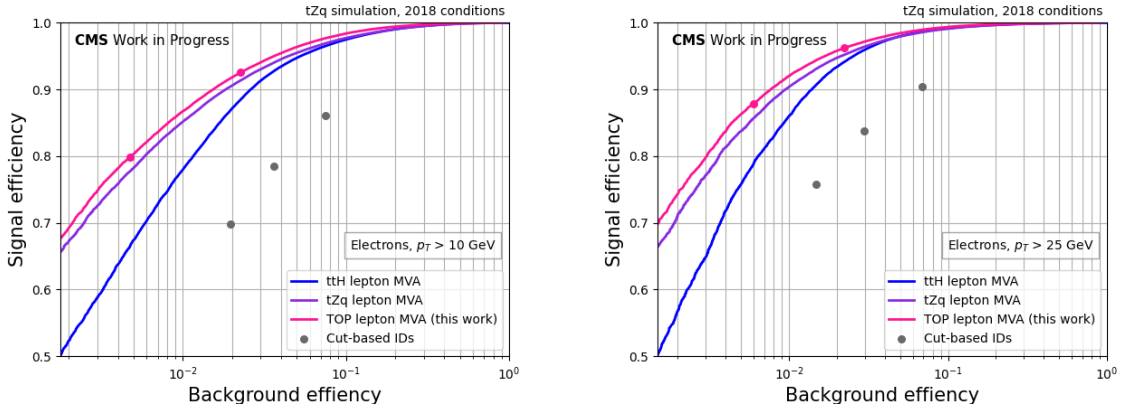


Figure 7.4: Lepton MVA ROC curves for electrons and comparison with earlier lepton MVAs and cut-based selections. Left: electrons with $p_T > 10\,\text{GeV}$, right: electrons with $p_T > 25\,\text{GeV}$. In these plots, 'signal' refers to prompt muons, while 'background' refers to nonprompt muons. Efficiencies are calculated with respect to the baseline selection used in this analysis. The indicated working points of the TOP lepton MVA (pink circles) correspond to the loose and tight working point employed in this analysis, as discussed in the text. The cut-based selection criteria (grey circles) correspond to the loose, medium and tight criteria, as defined by the POG [210].

one b-jet.

Categorization The selections detailed above make up the full signal region for this analysis. However, to increase the sensitivity, we divide the signal region into three categories based on the number of jets and b-tagged jets:

- events containing exactly 1 b-jet and 2 or 3 jets in total.
- events containing exactly 1 b-jet and 4 or more jets in total.
- events containing 2 or more b-jets.

Discussion Applying the event selection as explained above (with the objects defined in section 7.3) on the data and simulation samples listed in section 7.2, we arrive at an expected number of about 423 tZq events versus 2958 background events and an observed number of 3689 events in data, in the full 138 fb^{-1} data. Predicted distributions of signal and background events for this event selection (not split in categories) are displayed in Fig. 7.6 for some interesting kinematic variables. An overall good agreement between prediction and data is observed. However, as the background contribution to the signal region is rather overwhelming compared to the number of actual tZq events, an MVA-based signal extraction is necessary to obtain a precise cross-section determination; this MVA is discussed in section 7.4.3 after introducing the reconstruction of the top quark from its decay products in section 7.4.2.

Branching fraction The trilepton decay channel for tZq has a small branching fraction of only about 1.8%[19], with the other 98% taken up by purely hadronic, single lepton and dilepton

[19] Calculated from the measured branching fractions of the W boson and Z boson for leptonic decays. The

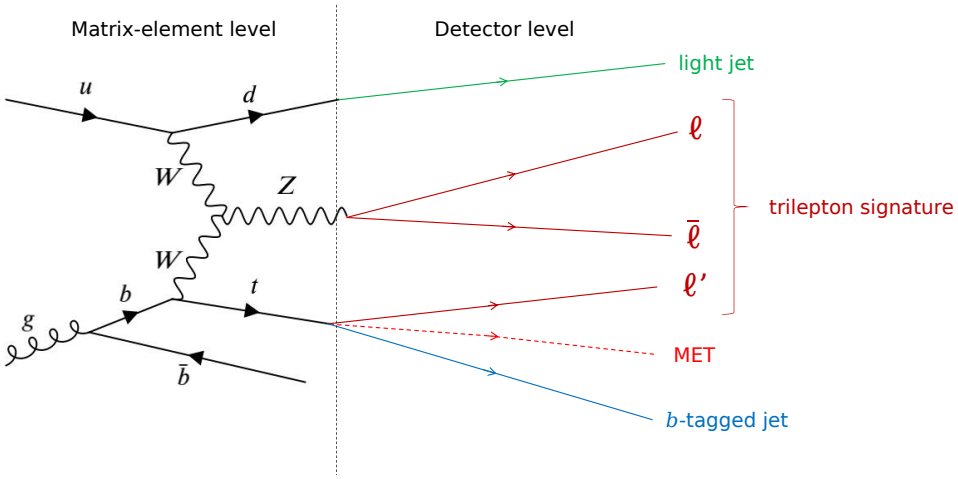


Figure 7.5: The tZq process at matrix-element level (left) and at detector level (right), showing how the analysis signature arises from the underlying event. The main characteristics of the tZq process at detector level are: an OSSF lepton pair from the Z boson decay, the presence of a third lepton, a b-tagged jet and some missing transverse energy (MET) from the top quark decay, and a light-flavoured jet at high pseudorapidity from the recoiling light quark. Potentially, a second b-tagged jet is present from gluon splitting, but it is often outside the detector acceptance. Apart from the main objects sketched here, additional jets may be present in the event via initial and final state radiation.

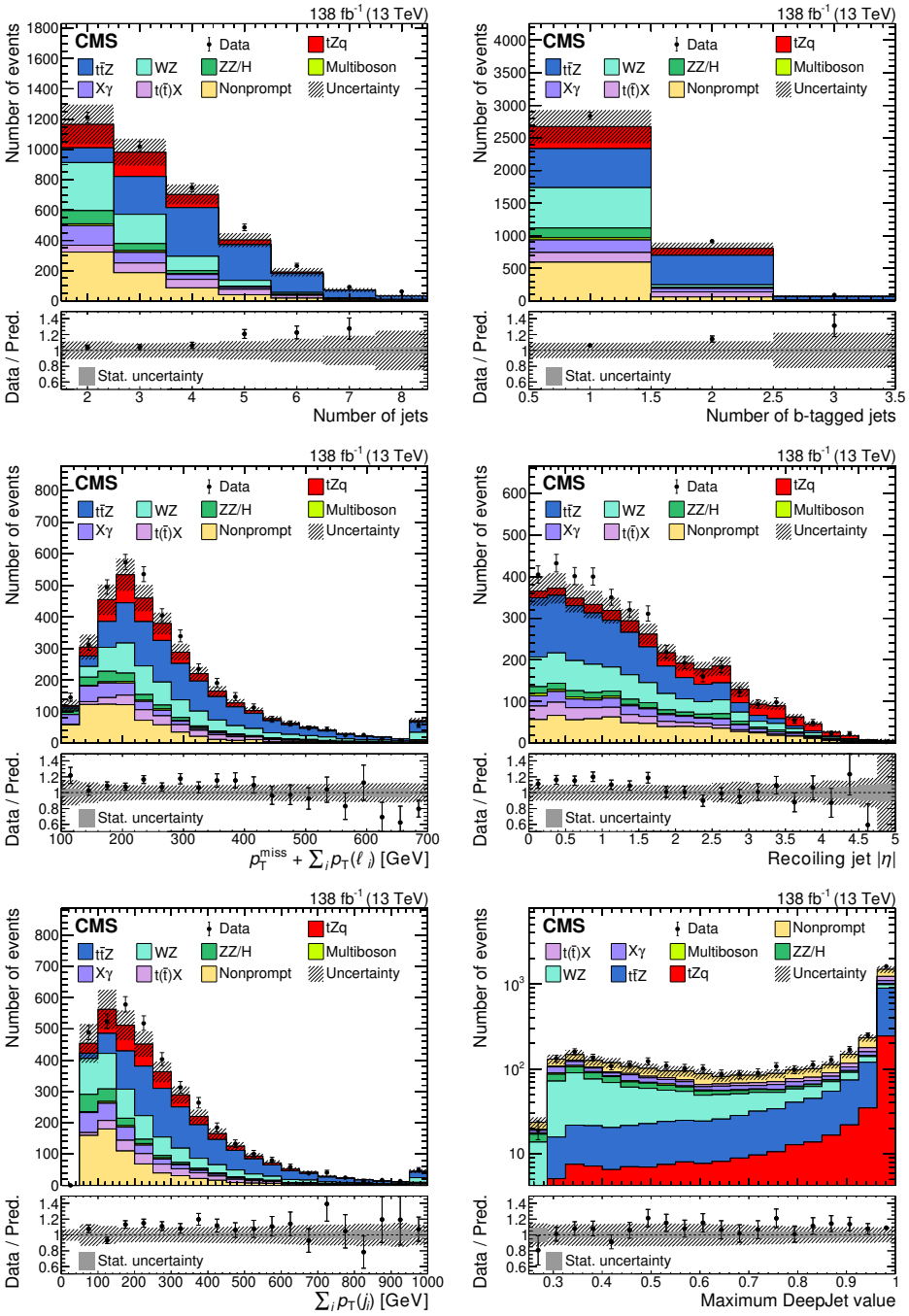


Figure 7.6: Distributions of some interesting variables in the signal region for the data (black points) and predictions (colored histograms). Upper left: number of jets. Upper right: number of b-tagged jets. Middle left: scalar sum of the p_T of the three leptons and the p_T^{miss} (also called L_T below). Middle right: $|\eta|$ of the recoiling jet (defined in section 7.4.2). Lower left: scalar sum of the p_T of all jets (also called H_T below). Lower right: maximum DeepJet score among the b-tagged jets in the event. The last bins include the overflows. The lower panels show the ratio of the data to the sum of the predictions. The vertical lines on the data points represent the statistical uncertainty in the data; the shaded area corresponds to the total uncertainty in the prediction; the gray area in the ratio indicates the uncertainty related to the limited statistical precision in the prediction.

decays. Despite its small branching fraction, the trilepton channel is the most sensitive one, because of the high lepton detection efficiency and comparatively small backgrounds. The fully hadronic channel is dominated by QCD multijet background, the single lepton channel by W + jets, and the dilepton channel by DY (in case of OSSF leptons) or a mixture of nonprompt (from DY and $t\bar{t}$), ttX and diboson production (in case of same-sign leptons).

Selection efficiency The event selection is designed to be an optimal compromise between tZq selection efficiency and background rejection. The selection efficiency as a function of sequential conditions is shown in Fig. 7.7 for tZq and in Fig. 7.8 for several important backgrounds. We observe that the main inefficiencies in tZq selection are coming from the requirements on the number of tight leptons, jets and b-tagged jets, as expected because of the limited acceptance of the detector and efficiency of the reconstruction algorithms. We observe furthermore, in accordance with our expectations, that the selection efficiency for Drell-Yan and $t\bar{t}$ events is very low because of the 3 tight lepton selection, that the efficiency for WZ events is suppressed by the (b-tagged) jets requirements, and that $t\bar{t}Z$ events are not strongly suppressed by our event selection, as they are very similar to tZq events.

The absolute numbers should be taken with a grain of salt however, as the reference point with respect to which the efficiency is calculated is not necessarily the same in all cases. In the case of tZq, the efficiency is calculated with respect to the full tZq sample used in this analysis, which includes hadronic top quark decays as well as top quark and Z boson decays to (hadronically decaying) τ leptons, which are excluded from our signal region. The selection efficiency with respect to this reference point is 3.8% (see Fig. 7.7). When restricting the reference point to tZq events where top quarks and Z bosons decay to electrons or muons only, this value should be corrected with the appropriate branching fraction, which yields about 25%. Alternatively, the reference point for the efficiency calculation can be set to the number of events that pass a fiducial event selection at generator level that mimics our signal region definition at detector level. For example, we select events with (all at generator level) exactly three light leptons within the geometric detector acceptance, that pass the p_{T} thresholds of 25, 15 and 10 GeV respectively, that contain an opposite-charge same-flavour lepton pair with an invariant mass within 15 GeV of the Z boson mass, and with at least 2 jets of with at least one b-jet with $p_{\mathrm{T}} > 25$ GeV and within the geometric detector acceptance. With respect to this reference point, the efficiency of our selection at detector level is about 45%, implying that about half of all events that can reasonably be expected to end up in our signal region effectively make it there. The remaining inefficiencies are mainly coming from the lepton and b-tagged jet identification.

7.4.2 Top quark reconstruction

The accurate reconstruction of the top quark kinematics and decay chain are not of crucial importance to this inclusive analysis. However, as some of the BDT variables (see section 7.4.3) depend on the assignment of a particular b-tagged jet as coming from the top quark decay, an algorithm that does this assignment is still needed. The method applied here is chosen for its simplicity and its being intuitively clear. The same method was also used in the previous tZq analysis [239] and earlier work [240]. We perform the top quark reconstruction in several steps:

- In each event, one of the three leptons does not belong to the OSSF pair forming the Z boson mass. (Note: if multiple OSSF pairs are present in the event, the ambiguity is

decay to τ leptons is included, but comes with an additional branching fraction of the τ leptons to electrons or muons.

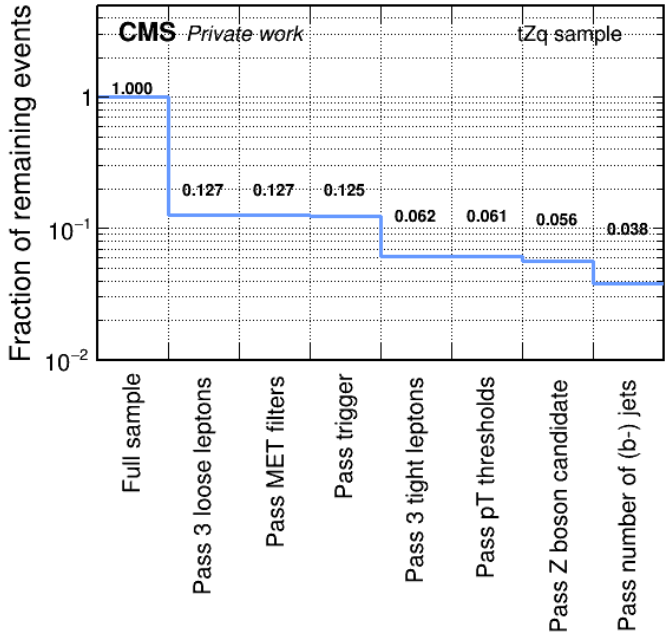


Figure 7.7: Selection efficiency for tZq events. The efficiencies are calculated with respect to the tZq sample used in this analysis, which includes hadronic top quark decays as well as top quark and Z boson decays to (hadronically decaying) τ leptons, which are excluded from our signal region.

resolved by considering the pair yielding the invariant mass closest to the mass of the Z boson as coming from the Z boson decay.) In a genuine well-reconstructed tZq event in the trilepton channel, this third lepton originates from the leptonic W boson decay, in turn from the top quark decay ($\mathrm{t} \to \mathrm{b} + \mathrm{W} \to \mathrm{b} + \ell + \nu$). This lepton we term '$\ell_\mathrm{t}$'.

- This lepton (ℓ_t) can be used together with the missing transverse momentum $\vec{p}_\mathrm{T}^{\,\mathrm{miss}}$ to reconstruct the longitudinal component of the missing momentum, assuming the neutrino from the top quark decay is the only source of missing momentum in the event ($\bar{p}_{\mathrm{miss}} = \bar{p}_\nu$). This is achieved by imposing the W boson mass constraint $(\bar{p}_{\ell_\mathrm{t}} + \bar{p}_\nu)^2 = m_\mathrm{W}^2$, yielding a quadratic equation for the longitudinal component of $\bar{p}_{miss}$ / $\bar{p}_\nu$ with two possible solutions:

$$p_{\nu,z}^{\pm} = \frac{\alpha p_{\ell,z}}{p_{\ell,T}^2} \pm \sqrt{\frac{\alpha^2 p_{\ell,z}^2}{p_{\ell,T}^4} - \frac{p_\ell^2 (p_\mathrm{T}^{\mathrm{miss}})^2 - \alpha^2}{p_{\ell,T}^2}} \tag{7.1}$$

with

$$\alpha = \frac{m_W^2}{2} + \vec{p}_{\ell,T} \cdot \vec{p}_\mathrm{T}^{\,\mathrm{miss}} \tag{7.2}$$

In case the discriminant (under the square root) is negative, the extremum of the parabola is used as solution instead, which is equivalent to putting the discriminant to zero.

- There is an additional mass constraint that can be exploited: the four-vectors of the neutrino, $\bar{p}_{\ell_\mathrm{t}}$ and the b-tagged jet from the top quark decay should together form the top

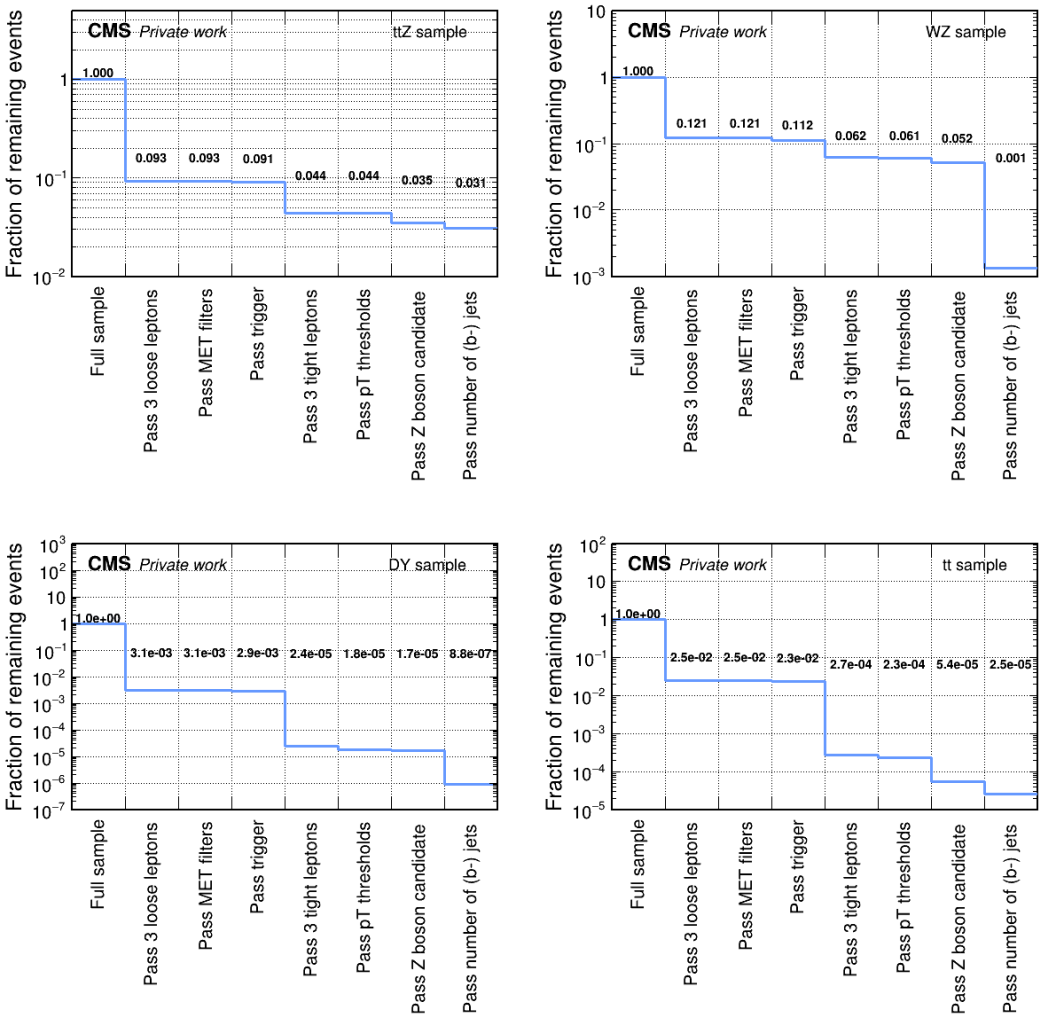


Figure 7.8: Selection efficiency for several important backgrounds. Upper left: $t\bar{t}Z$ (the full sample includes all top quark decays, and Z boson decays to τ leptons and neutrinos). Upper right: WZ (the full sample includes W and Z boson decays to τ leptons). Lower left: Drell-Yan (the full sample includes Z boson decays to τ leptons). Lower right: $t\bar{t}$ (the full sample includes top quark decays to τ leptons).

quark mass. As there are possibly two solutions for the neutrino four-vector and multiple b-tagged jets in the event, the invariant mass of each combination is calculated. The combination of which the result is closest to the nominal top quark mass[20] is considered to be the 'true' top quark signature.

- The b-tagged jet found in this way is called the '$\mathrm{b_t}$ jet' and is assumed to come from the top quark decay. The jet with the highest p_T excluding b-tagged jets is called the 'recoiling jet'. Both of them are used in the variables for the event BDT (see section 7.4.3). The reconstructed longitudinal component of missing momentum is not used further, nor is the reconstructed top quark mass.

7.4.3 BDT discriminator

Input variables The relative yield of tZq is rather small even in the signal regions described above. Therefore a boosted decision tree (BDT) discriminator was trained to extract signal events while rejecting backgrounds. The variables that serve as an input to the BDT algorithm are listed and detailed in Tab. 7.3. For the definition of the $\mathrm{b_t}$ jet and the recoiling jet, see section 7.4.2.

The variables $|\eta|_\mathrm{recoil}$, $|\eta|^{max}$, $M(j,j)^{max}$, $p_T(j,j)^{max}$ and $\Delta R(\ell_\mathrm{t}, \text{recoiling jet})$ are all ultimately motivated by the presence of a jet with large absolute pseudorapidity recoiling against the top quark. The variables deepFlavour_{max}, $\Delta R(l,b)^{min}$ and $\Delta R(\ell_\mathrm{t}, \mathrm{b_t}\text{-jet})$ are based on the presence of a b jet from a leptonically decaying top quark. The variable ℓ_t asymmetry helps to distinguish tZq from charge-symmetric backgrounds. The variables L_T, H_T, M_T and M_{3l} are generic kinematic variables. The number of jets and b-tagged jets mainly serve to make the BDT distinguish between parts of the phase space with a significantly different background composition.

Hyperparameter optimization The BDT was trained using the TMVA package [241]. The final BDT configuration results from an extensive optimization procedure. In a first step, a grid search over reasonable ranges for all relevant BDT architecture and training parameters was performed, the result of which is shown in Fig. 7.9 (a). The figure of merit is the integral under the ROC curve, shortly AUC. The optimized parameters include the number of trees, the maximum decision depth of each tree, the minimum percentage of the training set in each leaf node, and the learning rate.

Next, we cross-checked this result using a genetic algorithm approach[21]. Starting from an initial set of BDT configurations randomly chosen within a reasonable phase space, a genetic algorithm selects highly performing BDTs and combines and mutates their properties generation after generation, until no further improvement is detected. The performance metric of interest in this context is again the AUC. Using this algorithm, no improvement with respect to the grid search could be found (see e.g. Fig. 7.9 (b)).

Other training considerations In this analysis we use a single BDT to classify events belonging to all signal regions and all data-taking years (i.e. nine distinct regions). Checks were done to assess the performance of separate BDTs per region, but the increased statistics during training for a combined BDT were found to outweigh the potential loss of specificity in this

[20] As discussed in chapter 5, there are multiple definitions and corresponding values of the top quark mass. We use here 172.5 GeV, which is typically the value used in simulated samples in CMS.

[21] The technical implementation of this algorithm, as used in this work, is by the hand of by Willem Verbeke. The code is public and can be found *here* (digital link).

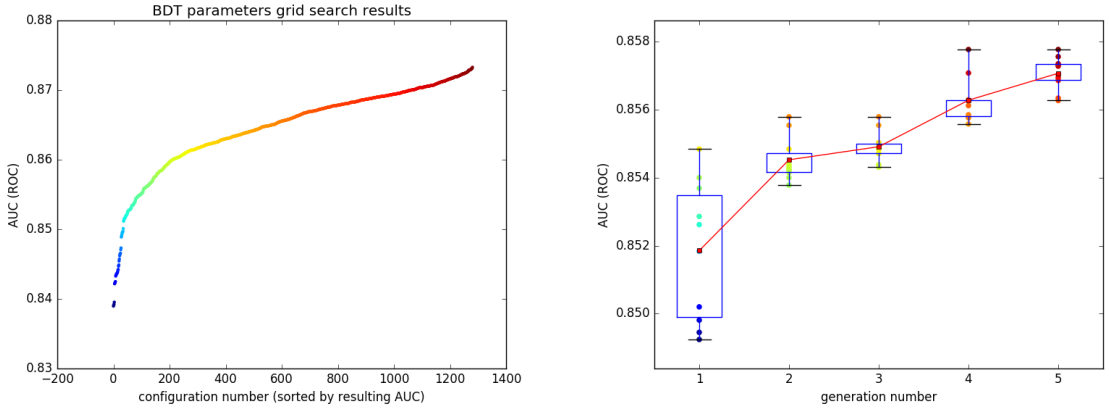
Table 7.3: Overview of event variables used in the BDT.

BDT variables	
Name	Description
$\lvert\eta\rvert_{\text{recoil}}$	Absolute value of pseudorapidity of recoiling jet
$\lvert\eta\rvert^{max}$	Absolute value of pseudorapidity of most forward jet
ℓ_{t} asymmetry	Absolute value of pseudorapidity of ℓ_{t} times electric charge of ℓ_{t}
deepFlavour$_{max}$	Maximum deepFlavour b-tagging score among all jets
L_T	Scalar sum of transverse momenta of all leptons + missing transverse momentum
H_T	Scalar sum of transverse momenta of all jets
M_T	Transverse mass of ℓ_{t} and $\vec{p}_T^{\,miss}$, defined as $\sqrt{2p_T(\ell_{\text{t}})p_T^{miss}\Big(1-\cos\Big(\Delta\Phi(\ell_{\text{t}},\vec{p}_T^{\,miss})\Big)\Big)}$
$M(j,j)^{max}$	Maximum di-jet invariant mass of all di-jet combinations
$p_T(j,j)^{max}$	Maximum di-jet transverse momentum among all di-jet combinations
$\Delta R(l,b)^{min}$	Minimum angular separation between a lepton and a b-jet
$\Delta\Phi(l,l)^{max}$	Maximum azimuthal separation between two leptons
$\Delta R(\ell_{\text{t}}, \text{recoiling jet})$	Angular separation between ℓ_{t} and the recoiling jet
$\Delta R(\ell_{\text{t}}, \text{b}_{\text{t}}\text{-jet})$	Angular separation between ℓ_{t} and the b_{t}-jet
M_{3l}	Invariant mass of the three-lepton system
nJets	Number of jets in the event
nBJets	Number of b-tagged jets in the event

case, especially since the number of jets and b-tagged jets are used as input variables.

One major issue in training the BDT is the very limited statistics of the nonprompt simulation (notably Drell-Yan) in the signal regions. This issue was addressed by allowing also events from the nonprompt background processes with only 2 instead of 3 tight leptons (but still 3 FO leptons) to be used in the training. We find that this increases the AUC slightly without causing overtraining effects.

Discussion of the final BDT The final BDT described here achieves a AUC of about 0.87, and its output distribution on the train and test set are shown in Fig. 7.10. Furthermore, we investigate the ranking of the variables according to their importance. This importance measure is calculated in TMVA by counting the number of times a variable is used to split a decision tree node, weighted by the number of instances in that node and the separation gain achieved by the splitting [241]. From this ranking we conclude that the most discriminating variables between tZq and its backgrounds are the pseudorapidity of the recoiling jet, the total transverse momentum carried by the jets and leptons in the event, and the invariant mass of the three-leptons system. However, all of the variables are important to the BDT in a non-negligible way; there is no clear cut-off value that should allow to remove some of the low-ranked variables without deteriorating the discriminating performance of the BDT. The number of jets and b-tagged jets seem to have significantly lower importance scores; however, we still include these variables for the BDT to be in principle able to distinguish between the separate background categories based on these variables.



(a) (b)

Figure 7.9: BDT hyperparameter optimization. (a): Results of a grid search, showing the (sorted) AUC for all hyperparameter configurations included in the scan. For each of the hyperparameters, a range of reasonable values is defined, combined in a grid of potential BDT configurations (1280 in total). For each configuration, a BDT is trained and its AUC evaluated. The resulting values are sorted from low to high in the figure. The AUC is relatively high for each of the configurations (with a minimum of about 84%), yet significant improvement is obtained by choosing the optimal configuration (up to 87%). (b): Results of a genetic algorithm, showing the top-10 AUCs per generation. Starting from a collection of BDTs (initially with random hyperparameters within predefined reasonable ranges), a next generation is built by combining properties of well-performing BDTs and some random variation. Both approaches, the grid search and the genetic algorithm, lead to similar optimal AUCs, but the latter is not able to reach the optimal AUC found in the grid search with the current number of generations.

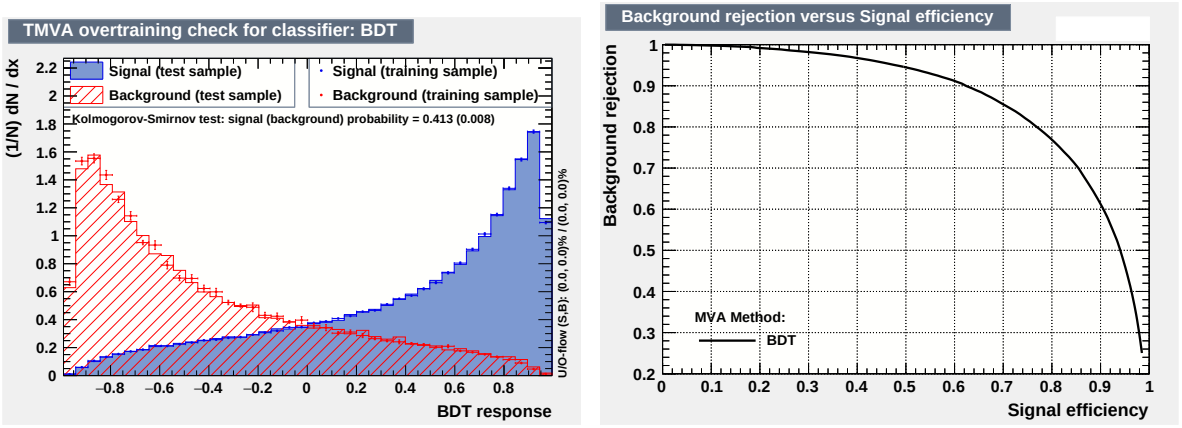


Figure 7.10: Left: BDT score distributions for signal (blue) and background (red). Both the training set (dots) and testing sample (filled/shaded histograms) are shown; the fact that they overlap indicates that the BDT was not overtrained on the specific training set and generalizes well to unseen data. Right: resulting ROC curve. The AUC was the figure of merit during the optimization procedure, and its final value is about 87%.

7.5 Background estimation and control regions

In this section we discuss the standard model background processes that are relevant in our signal region and how they are estimated. A first major category of backgrounds consists of events with three or more genuine prompt leptons. When targeting final states with relatively few jets and b-tagged jets (notably in the signal category with 2 or 3 jets of which 1 b-tagged jet), diboson processes constitute the dominant sources of background. In this category we find for example ZZ, Zγ and most prominently WZ. In final states with more jets and b-tagged jets (i.e. the signal regions with 1 b-tagged jet and 4 or more jets, or with 2 or more b-tagged jets), the main backgrounds consist of top quark pair production in association with an additional boson, notably $t\bar{t}Z$. Apart from these, some other less dominant background sources are taken into account[22]. The dominant background categories (WZ, ZZ, Zγ and $t\bar{t}Z$), and their corresponding control regions in data, are discussed more detail in section 7.5.2.

A second major category of background events consists of processes where at least one of the leptons is nonprompt, but still passes the tight selection. In our search region the main processes in this category are Drell-Yan and $t\bar{t}$. This source of background is particularly important to study, as its accurate prediction is difficult. In order to address statistical as well as potential modeling issues, we apply a data-driven technique to estimate the nonprompt background in our signal regions, as detailed in section 7.5.1.

7.5.1 Nonprompt background

Closure in simulation The method for estimating the background with nonprompt leptons is called the 'fake rate method'. It was discussed already in section 6.2; here we will only show the results of the application of the method on this specific analysis case. Figures 7.11 and 7.12 show the simulated closure on Drell-Yan and $t\bar{t}$ samples. In this simulated closure test, the prediction for the nonprompt background contribution obtained with the fake rate method (treating the simulated sample as real data) is compared to the actual nonprompt contribution obtained using the generator-level information in the simulated sample. The test is performed for two event selections: a generic three-lepton selection (Fig. 7.11), and a selection that is quasi identical to that of the signal region in this analysis (Fig. 7.12). The former selection provides better statistical precision, the latter has a better focus on the particular phase space of this analysis. Closure tests like these were used to tune the FO lepton definition for this analysis. The resulting FO definition, with which Figs. 7.11 and 7.12 were made, gives sufficiently good closure in all cases within the statistical uncertainties in addition to a (flat) systematic uncertainty of 30%, as derived from these tests and the closure in data (see next paragraph).

Closure in data A final check of the fake rate method is performed by checking the closure in data, using two separate nonprompt control regions. Events are selected into this region if they pass conditions that are very similar to the signal region, with the exception that events with an opposite-sign same-flavour pair (OSSF) are vetoed (shortly called 'no-OSSF' and shown in Fig. 7.13), or alternatively that the invariant mass of the OSSF pair is not compatible with the Z boson mass (shortly called 'no-Z' and shown in Fig. 7.14). In the latter region, we additionally reject events in the *mme* and *eee* channels that have a combined three-lepton invariant mass compatible to the Z boson mass, because of an observed inaccuracy in the modeling of conversion events in this particular region.

[22] A detailed list of the specific samples used to model each process is provided in Tabs. A.1–A.2 (in the appendix).

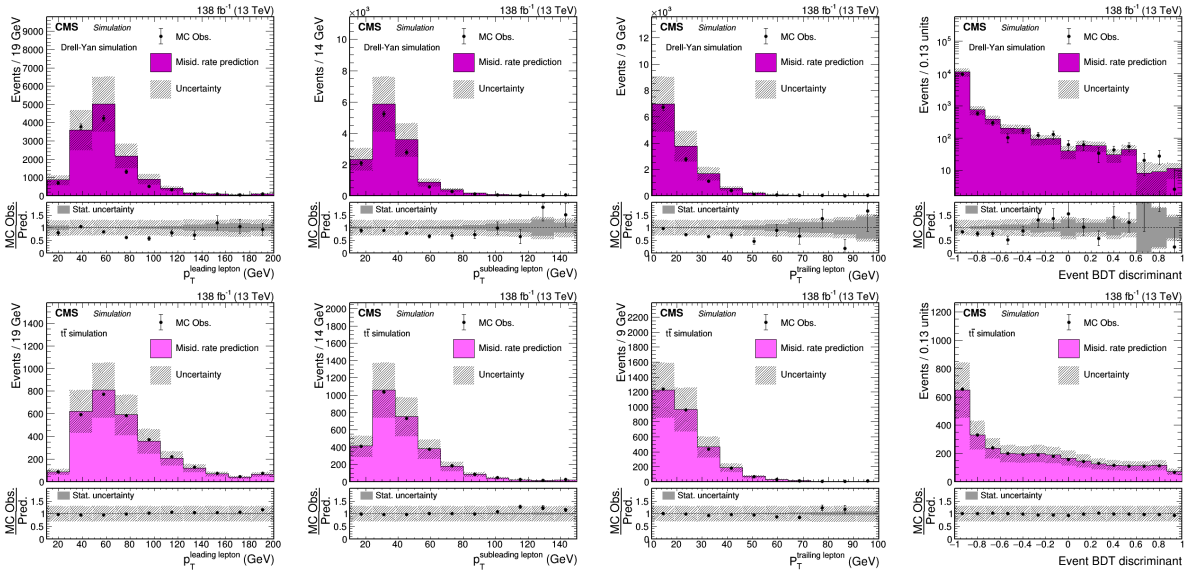


Figure 7.11: Simulated closure tests for the fake rate method and FO definitions in a Drell-Yan sample (upper row) and a $\mathrm{t\bar{t}}$ sample (lower row). Closure is shown as a function of the p_{T} of each of the three leptons in the event and additionally as a function of the BDT output aimed at extracting the tZq signal. The event selection is a generic three-lepton selection without further requirements on number of (b-tagged) jets.

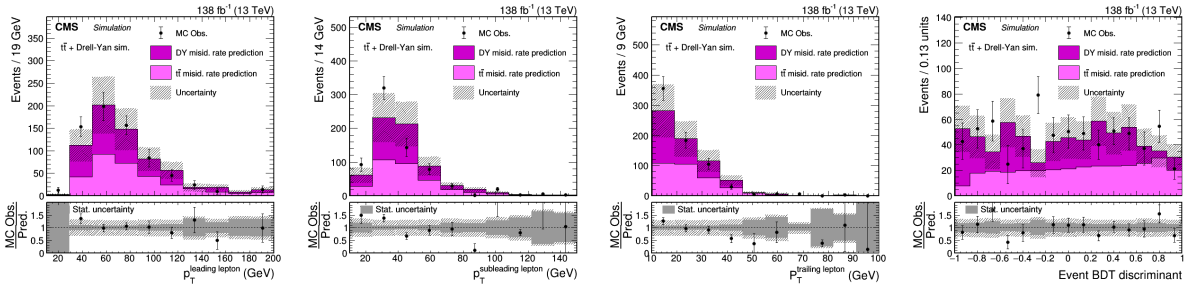


Figure 7.12: Simulated closure tests for the fake rate method and the FO definitions in a Drell-Yan and a $\mathrm{t\bar{t}}$ sample (stacked for better statistical precision). Closure is shown as a function of the p_{T} of each of the three leptons in the event and additionally as a function of the BDT output aimed at extracting the tZq signal. The event selection is quasi identical to that of the signal region. Although statistical precision is low, qualitatively good closure is still observed.

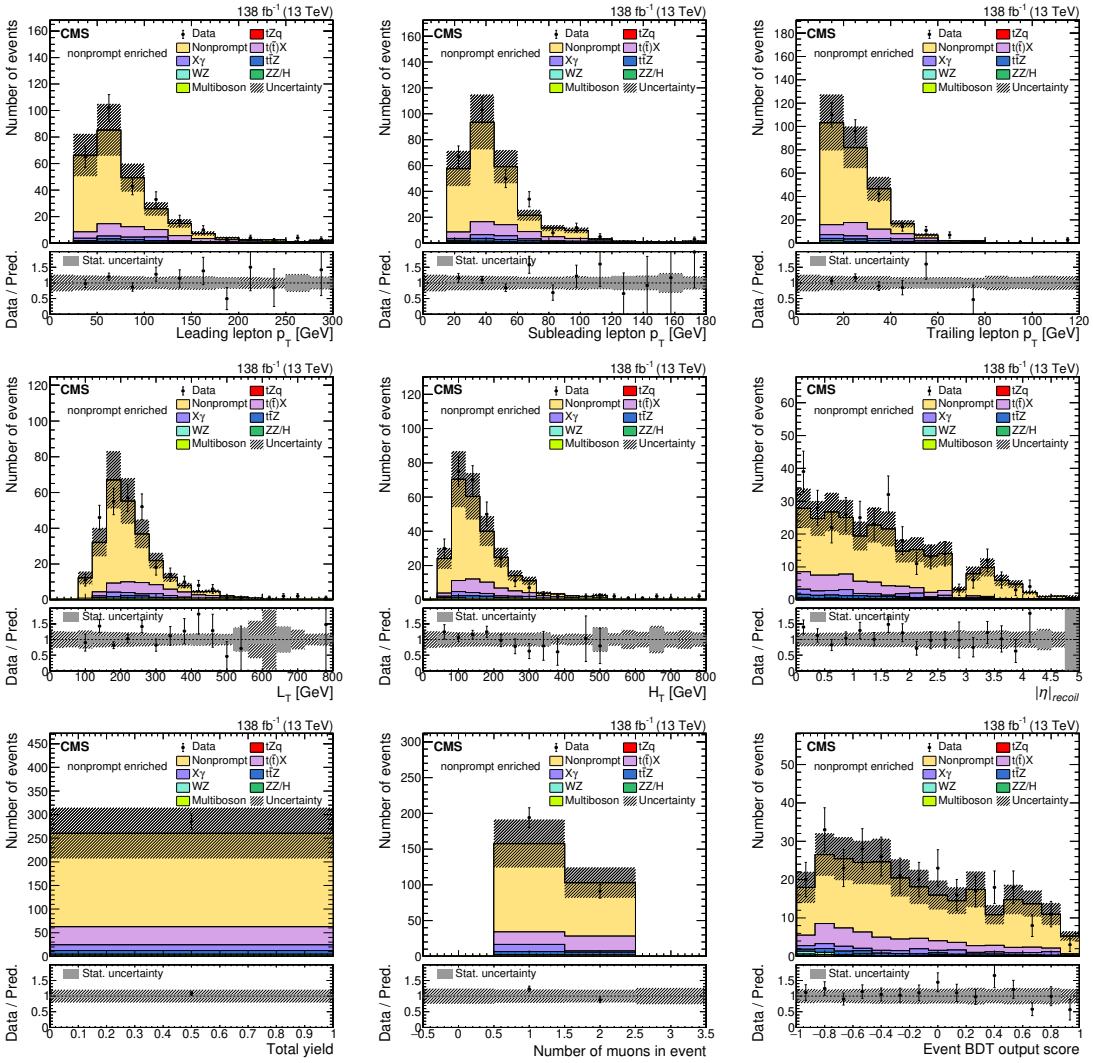


Figure 7.13: Distributions of the most relevant kinematic variables in the no-OSSF nonprompt control region for data (black points) and predictions (colored histogram). The lower panels show the ratio of the data to the sum of the predictions. The vertical lines on the data points represent the statistical uncertainty in the data; the shaded area corresponds to the total uncertainty in the prediction; the gray area in the ratio indicates the uncertainty related to the limited statistical precision in the prediction.

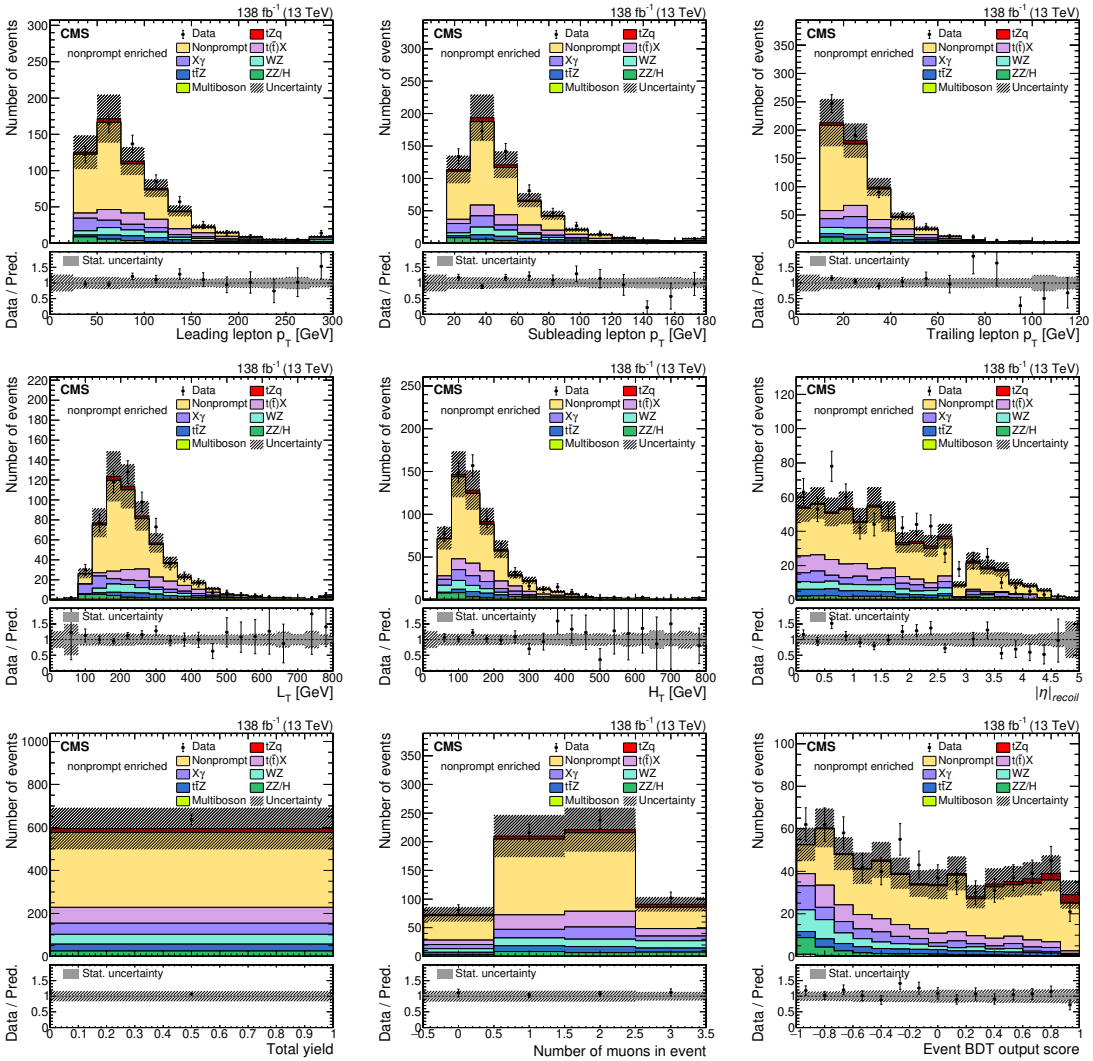


Figure 7.14: Distributions of the most relevant kinematic variables in the no-Z nonprompt control region for data (black points) and predictions (colored histogram). The lower panels show the ratio of the data to the sum of the predictions. The vertical lines on the data points represent the statistical uncertainty in the data; the shaded area corresponds to the total uncertainty in the prediction; the gray area in the ratio indicates the uncertainty related to the limited statistical precision in the prediction.

Discussion We observe good agreement in the simulated closure tests, notably in the event BDT shape (see Figs. 7.11 and 7.12). Furthermore, reasonable closure is observed in data control regions that are similar (but not overlapping) to the signal region (see Figs. 7.13 and 7.14). We therefore use the fake-rate prediction as nonprompt background in the further analysis, with an additional flat systematic uncertainty of 30% to account for observed discrepancies in the control regions.

7.5.2 Control regions for prompt backgrounds

WZ background WZ events are a dominant source of background in signal regions with relatively few jets and/or b-jets. A dedicated WZ control region can be selected with similar lepton requirements as for the signal regions but vetoing the presence of a b-jet to ensure there is no overlap with the signal regions. Furthermore we use the knowledge that some missing transverse energy must be present because of the leptonic decay of the W boson. In detail the event selections are the following:

- The event must have exactly 3 tight leptons, vetoing the presence of a fourth FO lepton (as for the signal regions).
- At least one opposite-sign same-flavour (OSSF) pair must be present in the lepton collection and its invariant mass must be compatible with the nominal Z boson mass within 15 GeV, while the invariant mass of the three-lepton system must be outside this window.
- Events that contain a b-tagged jet are vetoed.
- The missing transverse energy associated to the event must be above 50 GeV.

The distributions of the most relevant variables for this control region are shown in Fig. 7.15. Overall good agreement between data and simulation is observed, both in terms of overall normalization and kinematic distributions. Though the inclusive production cross-section of the WZ process has been predicted and measured with high precision [222], we assign a higher flat systematic uncertainty of 10% to the WZ normalization a priori (on top of all other systematic uncertainties discussed in section 7.6), to account for the higher jet multiplicity in our measurement.

All WZ events in the signal region have at least one b-tagged jet (as this is part of the signal region requirements). For about 30% of the events, this b-tagged jet actually originates from a b-quark; for the other 70% of events, it originates from a light or c-quark jet and enters the signal region only because of misidentification of this jet by the DeepJet algorithm. The modeling of WZ with additional genuine b-jets is subject to an uncertainty not covered by the control region (which is dominated by WZ events without b-jets). A dedicated study of this uncertainty was performed in DY events [176], resulting in an additional normalization uncertainty of 20% being applied to WZ events with b quarks at generator level.

ZZ background The background contribution of ZZ to a three-lepton region consists of events where both Z bosons decay leptonically but one of the leptons fails to be reconstructed or does not pass the tight selections. The ZZ control region targets events with four tight leptons in the final state. The event selections are the following:

- The event must have exactly 4 FO leptons (vetoing events with more FO leptons) that are also tight.
- The 4 leptons must make up two opposite-sign same-flavour (OSSF) pairs, both of which have an invariant mass differing not more than 15 GeV from the nominal Z boson mass.

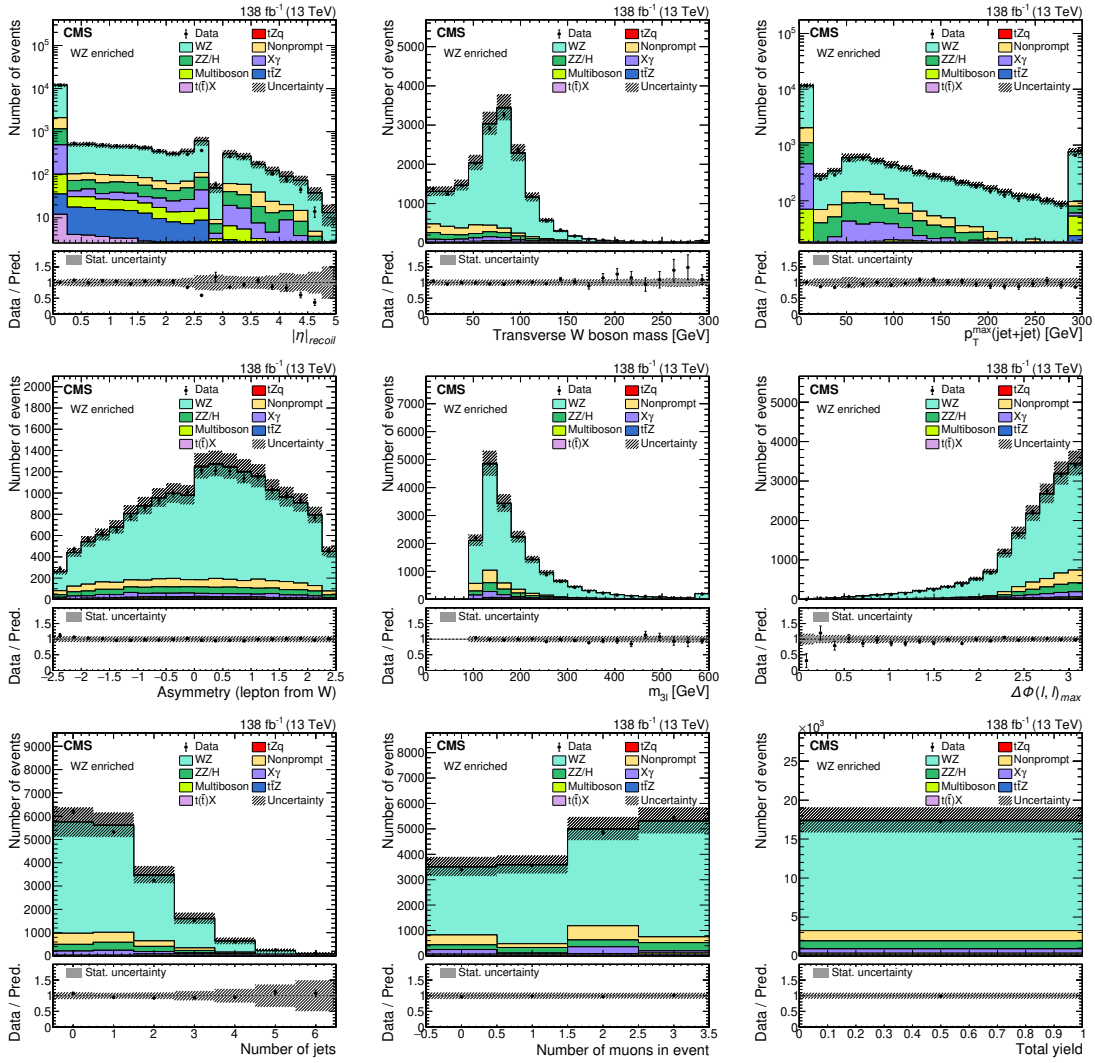


Figure 7.15: Distributions of the most relevant kinematic variables in the WZ control for data (black points) and predictions (colored histogram). The lower panels show the ratio of the data to the sum of the predictions. The vertical lines on the data points represent the statistical uncertainty in the data; the shaded area corresponds to the total uncertainty in the prediction; the gray area in the ratio indicates the uncertainty related to the limited statistical precision in the prediction.

The distributions of the most relevant variables for this control region are shown in Fig. 7.16. Overall good agreement between data and simulation is observed, both in terms of overall normalization and kinematic distributions. Though the inclusive production cross-section of the ZZ process has been predicted and measured with high precision [223], we assign a higher flat systematic uncertainty of 10% to the ZZ normalization a priori (on top of all other systematic uncertainties discussed in section 7.6), to account for the higher jet multiplicity in our measurement.

Zγ background Zγ processes can form a background in a three-lepton region via the conversion of the photon to an electron-positron pair ($\gamma \rightarrow e^+e^-$) in the detector material. These conversions can occur asymmetrically, with one of both leptons carrying most of the p_T of the original photon and passing the tight selection, and the other low-p_T one failing them. The

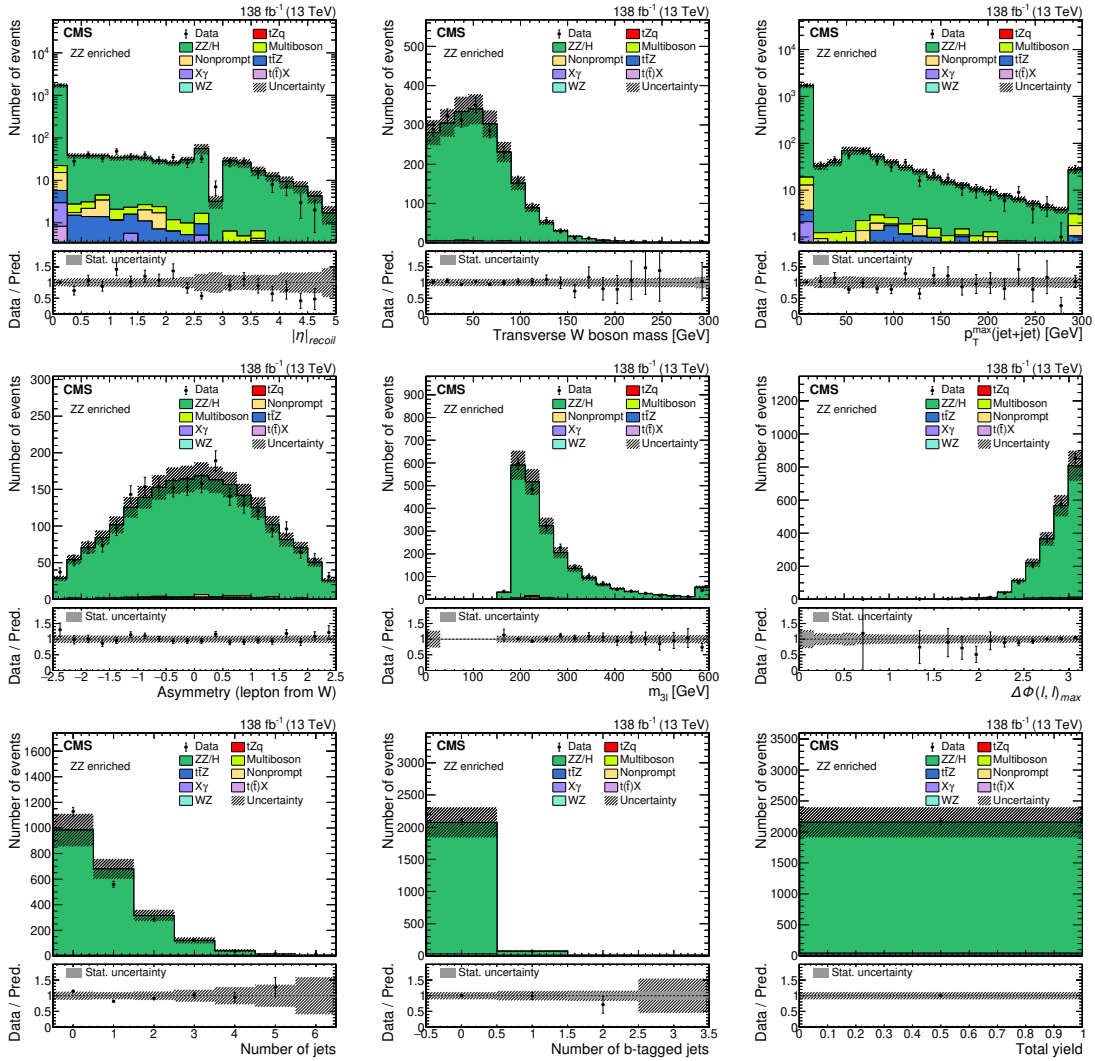


Figure 7.16: Distributions of the most relevant kinematic variables in the ZZ control for data (black points) and predictions (colored histogram). The lower panels show the ratio of the data to the sum of the predictions. The vertical lines on the data points represent the statistical uncertainty in the data; the shaded area corresponds to the total uncertainty in the prediction; the gray area in the ratio indicates the uncertainty related to the limited statistical precision in the prediction.

leptonic decay of the Z boson provides the additional two leptons to satisfy the three-lepton selection.

The control region for this type of process targets events where a Z boson decays leptonically and a photon is radiated from one of its decay products, which then converts strongly asymmetrically. In this case, the invariant mass of the three-lepton system will be close to that of the Z boson. The event selections are therefore the following:

- The event must have 3 tight leptons with at least one OSSF pair (as in the signal region).
- The difference between the invariant mass of the three-lepton system and the nominal Z mass cannot exceed 15 GeV.

- No pair of two leptons can have an invariant mass differing less than 15 GeV from the nominal Z boson mass.
- The OSSF pair of leptons must have an invariant mass larger than 35 GeV (to exclude a part of the phase space that is not covered by the simulation).

The distributions of the most relevant variables for this control region are shown in Fig. 7.17. Note that Zγ is the dominant contribution to the category labeled 'Xγ', with only minor contributions from other processes. Overall good agreement between data and simulation is observed, both in terms of overall normalization and kinematic distributions. Though the inclusive production cross-section of the Zγ process has been predicted and measured with high precision [224], we assign a higher flat systematic uncertainty of 10% to the ZZ normalization a priori (on top of all other systematic uncertainties discussed in section 7.6), to account for the higher jet multiplicity in our measurement.

In Fig. 7.17, the Zγ contribution is explicitly split in two contributions, from internal and external conversion. Internal conversion refers the prompt decay (in vacuum) of an off-shell (virtual) photon. Both $\gamma* \to e^+e^-$ and $\gamma* \to \mu^+\mu^-$ are possible. External conversion refers to the displaced decay (inside the tracker material) of a massless (real) photon. The decay to a pair of muons is strongly suppressed with respect to a pair of electrons because of the difference in mass. This behaviour can be observed in Fig. 7.17 in the plot showing the distribution of the number of muons in the events. The good data-to-prediction agreement in all variables, also the ones with strong kinematic differences between internal and external conversions, proves that both components are modeled well.

In this control region, most of the external conversion photons are radiated off from one of the leptons from the Z boson decay, i.e. it is final-state radiation (FSR). This is a consequence of the requirement on the three-lepton invariant mass being compatible with the Z boson mass, which is needed to obtain a control region sufficiently pure in Zγ (rather than mostly Drell-Yan + fakes if this requirement were to be dropped). In the signal region however, the contribution of photons from FSR is smaller, and more are coming from initial-state radiation (ISR), as a consequence of the requirement on the OSSF lepton pair mass being compatible with the Z boson mass. However, the electromagnetic interaction governing the creation of the photon is expected to be modeled well (with no difference between ISR and FSR), and the conversion of the photon is independent from whether it was radiated as ISR or as FSR. Hence this control region remains relevant for the signal region.

$t\bar{t}Z$ background The dominant background contribution to the tZq process in the signal regions with a high amount of jets and/or b-jets is $t\bar{t}Z$. We distinguish two separate control regions for this process. On the one hand, this process is so dominant in the signal regions with 1 b-jet and 4 or more jets and with 2 or more b-jets, that these signal regions serve as implicit control region for $t\bar{t}Z$. Both processes (tZq and $t\bar{t}Z$) are separated by the BDT-discriminant, and while the high-response region serves as a measurement region for tZq, the low-response region serves as an implicit control region for $t\bar{t}Z$.

On the other hand, we employ a dedicated control region targeting the fully leptonic decay of $t\bar{t}Z$. This selection requires the event to contain 4 FO leptons that also pass the tight definition, with an OSSF pair compatible with the Z boson mass. If a second OSSF pair is present, it is required not to be compatible with the Z boson mass in order to reduce the contribution from ZZ. The resulting dedicated control region suffers from low statistical precision, and is not very pure in $t\bar{t}Z$, but the distribution of the number of b-tagged jets separates $t\bar{t}Z$

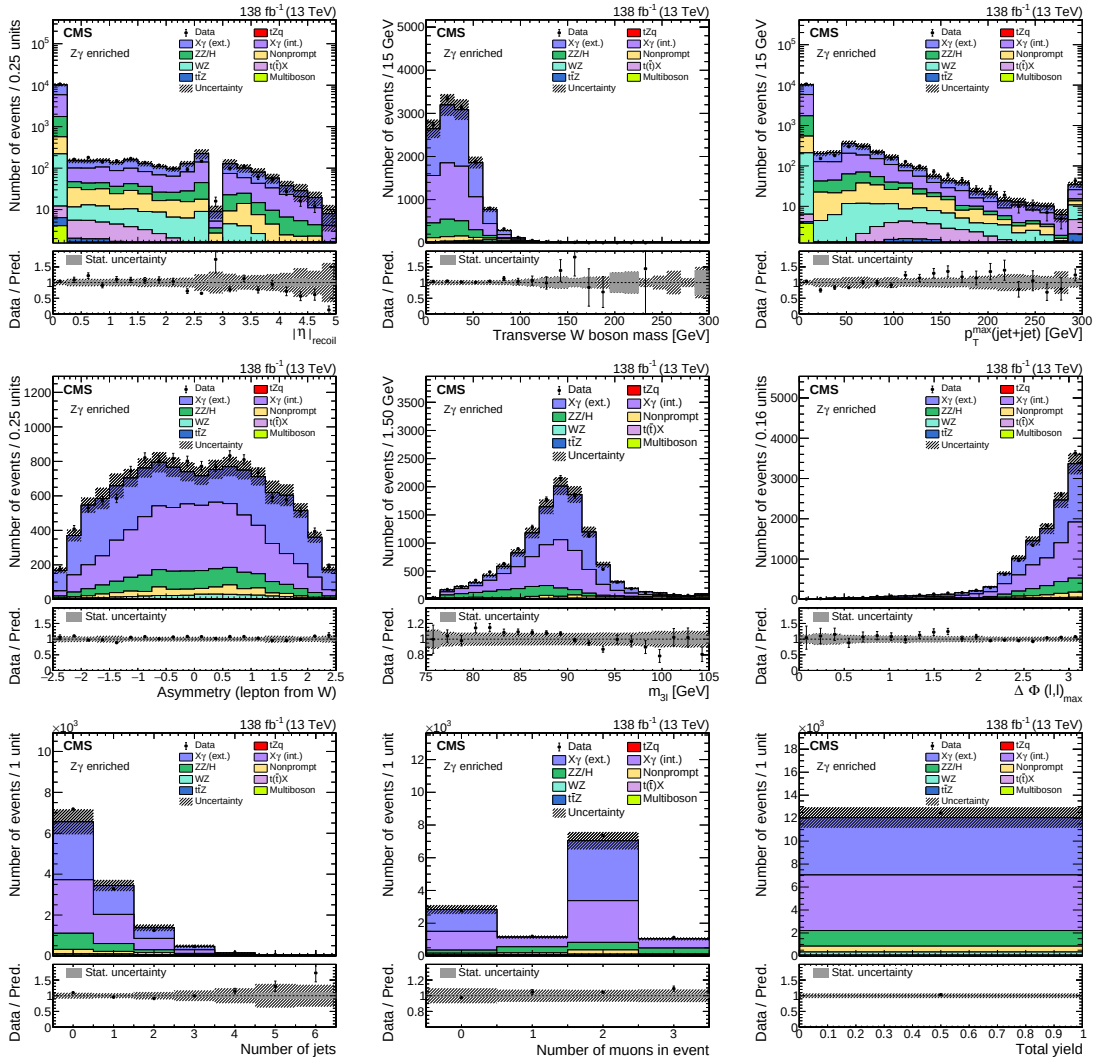


Figure 7.17: Distributions of the most relevant kinematic variables in the Zγ control for data (black points) and predictions (colored histogram). The lower panels show the ratio of the data to the sum of the predictions. The vertical lines on the data points represent the statistical uncertainty in the data; the shaded area corresponds to the total uncertainty in the prediction; the gray area in the ratio indicates the uncertainty related to the limited statistical precision in the prediction.

from ZZ (see Fig. 7.18)[23].

The distributions of the most relevant variables for this control region are shown in Fig. 7.18. Overall good agreement between data and simulation is observed, yet in some bins that are pure in $t\bar{t}Z$ (e.g. the 2 b-tagged jet bin in the number of b-tagged jets distribution) a slight excess in data is observed, consistent with earlier dedicated $t\bar{t}Z$ measurements [176]. We assign a flat systematic uncertainty of 15% to the $t\bar{t}Z$ normalization[24] a priori (based on Ref. [176]),

[23] Removing the 1 b-tagged jet category makes this control region more pure in $t\bar{t}Z$, but reduces further the statistical precision in data. Hence we keep this region inclusive in number of b-tagged jets and use that distribution in the fit (see section 7.7)

[24] In this work, the cross-section of the $t\bar{t}Z$ process is set to its theoretical prediction of $839\,\mathrm{fb}^{-1}$ [242], which is also used as a reference in the dedicated measurement [176]. A more recent prediction at full NLO precision

on top of all other systematic uncertainties discussed in section 7.6.

7.6 Systematic uncertainties

The common sources of systematic uncertainty affecting this measurement have already been discussed in section 6.5. Hence, this discussion is limited to filling in some details and specifications.

Background normalization. We apply a-priori background normalization uncertainties of 10% for the WZ, ZZ and Zγ processes, and 15% for $t\bar{t}Z$. These values are based on the general data-to-simulation agreement observed in the dedicated control regions for each of these processes (see section 7.5). They are in general larger than the uncertainties obtained in the current most precise dedicated cross-section measurements of these processes [176, 222–224], to account for differences in phase space. For the background with nonprompt leptons, we apply a flat systematic uncertainty of 30%, as discussed in section 7.5.1. All background normalization uncertainties are applied to a single background category (and hence uncorrelated between backgrounds) and treated correlated over the data-taking years.

Visualization In Figs. 7.19 to 7.21, we visualize the effect of the systematic uncertainties on the expected tZq yield as a function of the event BDT output score.

7.7 Results

In this section, we discuss the measurement of the tZq significance and cross-section. Both measurements are based on a profiled binned maximum likelihood fit (using the Combine tool [218–220]) of the simulation to the data in the various signal regions and control regions in all data-taking years. The methodology of the maximum likelihood fit has been described in section 6.6, and this section focuses on the results only.

Input distributions The input to the fit consists of the data and prediction in several distributions in signal and control regions. In detail, we use the transverse W boson mass for the WZ control region, the number of jets for the ZZ and Zγ control region, the number of b-tagged jets for the $t\bar{t}Z$ control region, and finally the total event yield for the nonprompt control regions. All of these distributions are shown in section 7.5. For the signal region, we use the distribution of the BDT output score, split into the three categories based on the number of jets and b-tagged jets (see section 7.4). These distributions are shown in Fig. 7.22.

7.7.1 Signal significance

The first set of results consist of the expected and observed significance of the tZq process. Although tZq has been observed before with $> 5\sigma$ [230, 231], we still calculate these numbers as they are a useful check of the analysis strategy and signal extraction performance. Note that the signal significance was not the metric by which the analysis strategy was optimized, as instead the uncertainty on the signal strength (see next section) is more important at this stage.

The main conclusion is that we confirm the observation of the tZq process in excess of 10σ. The observed significance is much lower than expected in the 2017 $\geq$ 2 b-jets category and in the 2018 1 b-jet, $\geq$ 4 jets category. The opposite is true in the 2017 1 b-jet, 2-3 jets

and including soft gluon resummation up to NNLL arrives at $859\,\text{fb}^{-1}$ [243], but this does not significantly alter the slight excess observed in the measurement.

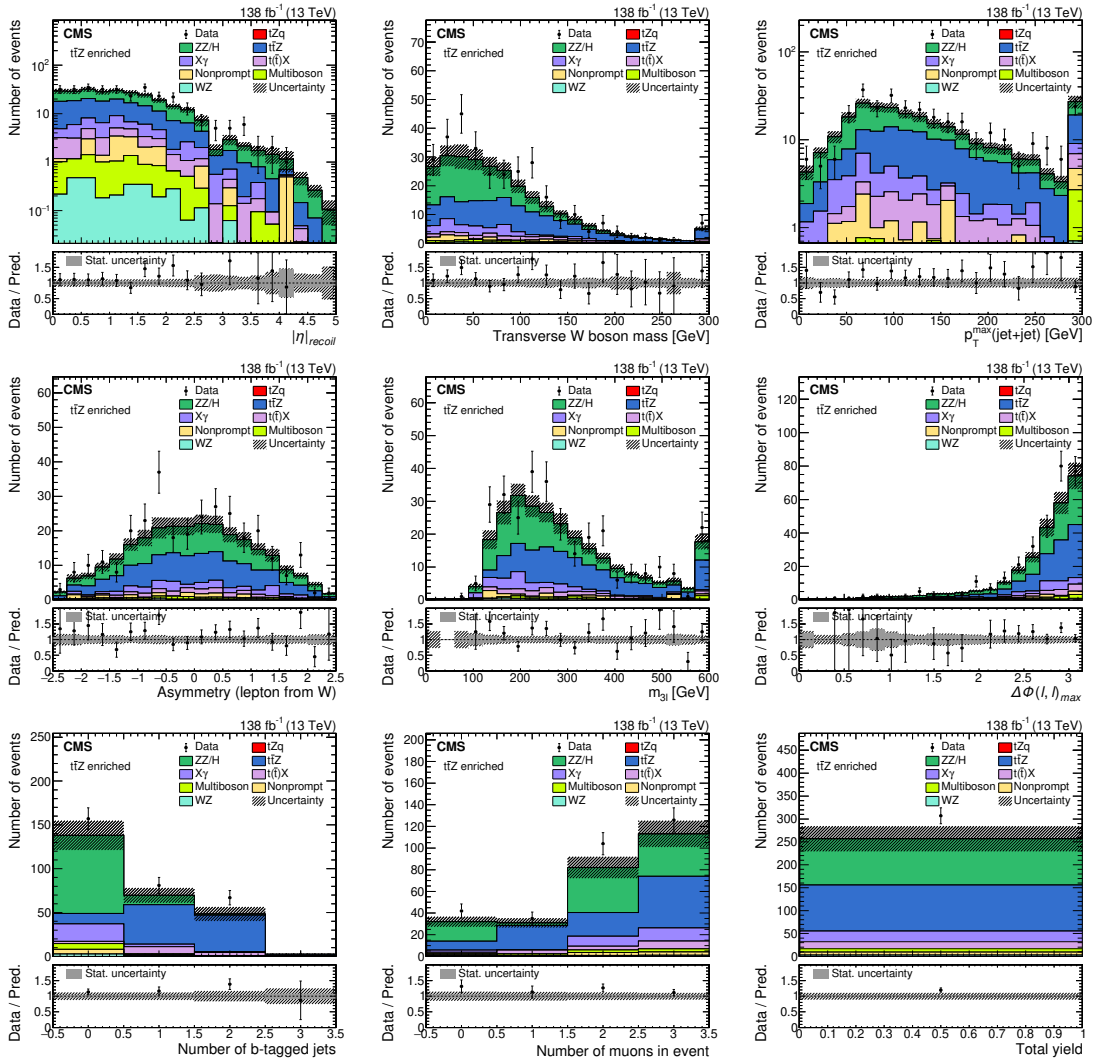


Figure 7.18: Distributions of the most relevant kinematic variables in the $t\bar{t}Z$ control for data (black points) and predictions (colored histogram). The lower panels show the ratio of the data to the sum of the predictions. The vertical lines on the data points represent the statistical uncertainty in the data; the shaded area corresponds to the total uncertainty in the prediction; the gray area in the ratio indicates the uncertainty related to the limited statistical precision in the prediction.

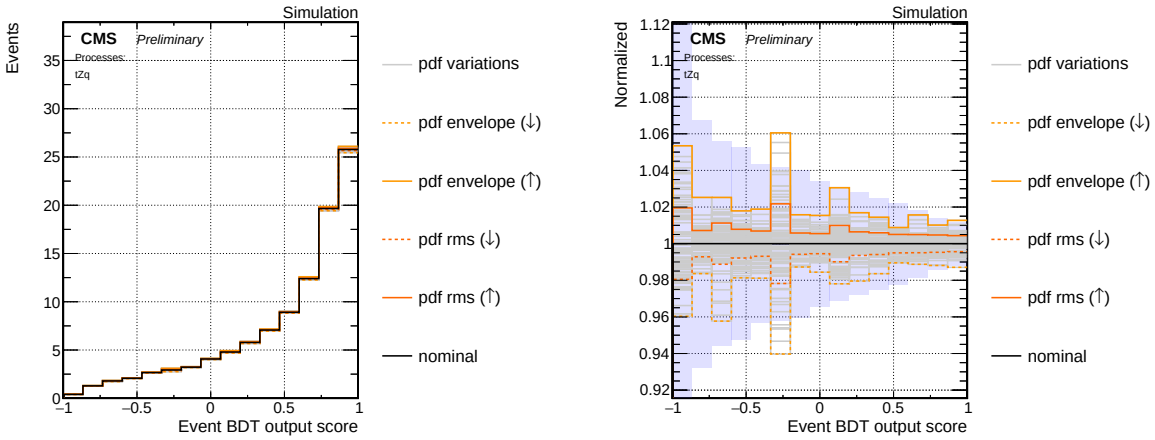


Figure 7.19: Effect of the PDF shape uncertainties on the BDT output score distribution of a tZq sample. The black curve is the nominal prediction. The gray curves represent individual PDF variations. Their envelope and root-mean-square (rms) are also shown, in shades of orange, with for both of them an up-variation (↑) and a down-variation (↓). The purple shaded area represents the statistical uncertainty. The left plot shows the absolute yield and its systematic variations, the right plot shows the yield variations relative to the nominal yield. The plots shown here correspond to a 2016 tZq sample to which the signal region selection has been applied (inclusive in the three (b-tagged) jet multiplicity categories).

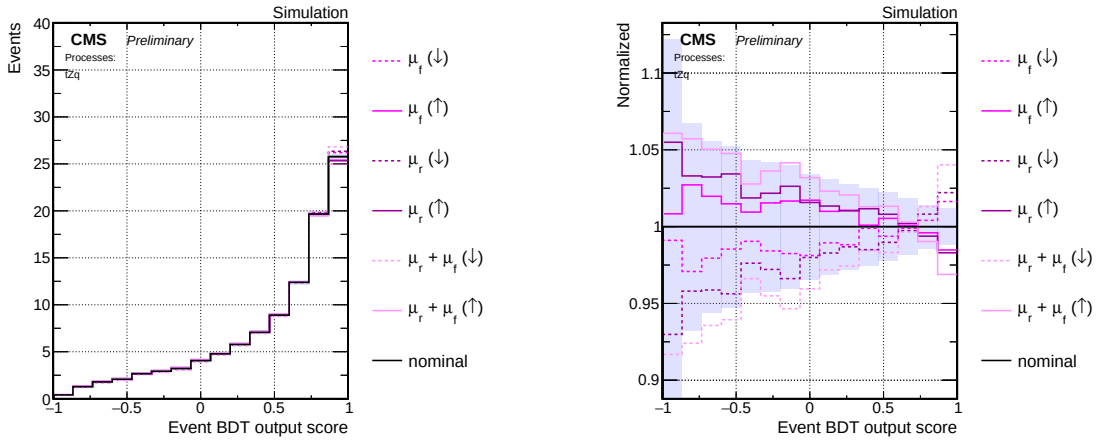


Figure 7.20: Effect of the renormalization scale (μ_r) and factorization scale (μ_f) shape uncertainties on the BDT output score distribution of a tZq sample. The black curve is the nominal prediction. The variations are shown in shades of purple, with for each of them an up-variation (↑) and a down-variation (↓). The variation labeled $\mu_r + \mu_f$ represents the simultaneous variation of both scales. The purple shaded area represents the statistical uncertainty. The left plot shows the absolute yield and its systematic variations, the right plot shows the yield variations relative to the nominal yield. The plots shown here correspond to a 2016 tZq sample to which the signal region selection has been applied (inclusive in the three (b-tagged) jet multiplicity categories).

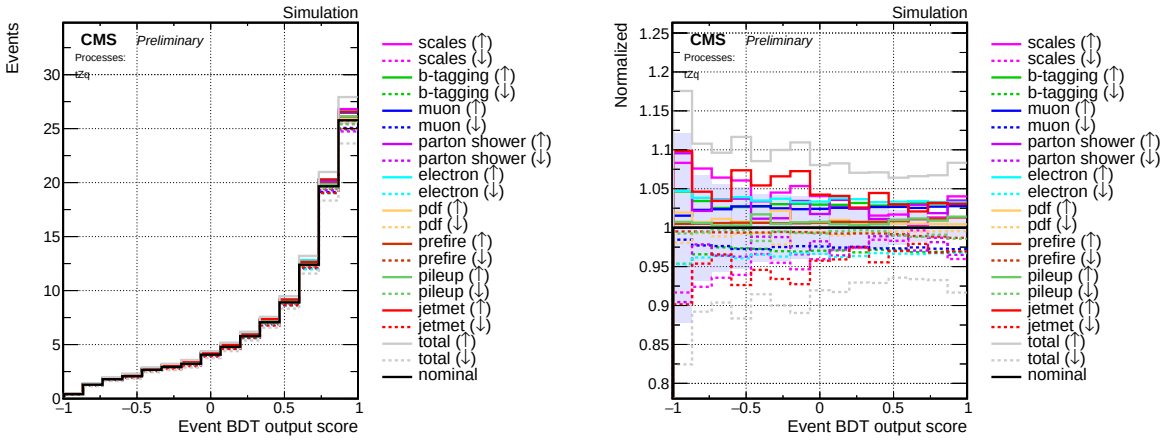


Figure 7.21: Overview of the effect of various shape uncertainties (grouped into simplified categories) on the BDT output score distribution of a tZq sample. The left plot shows the absolute yield and its systematic variations, the right plot shows the yield variations relative to the nominal yield. The plots shown here correspond to a 2016 tZq sample to which the signal region selection has been applied (inclusive in the three (b-tagged) jet multiplicity categories).

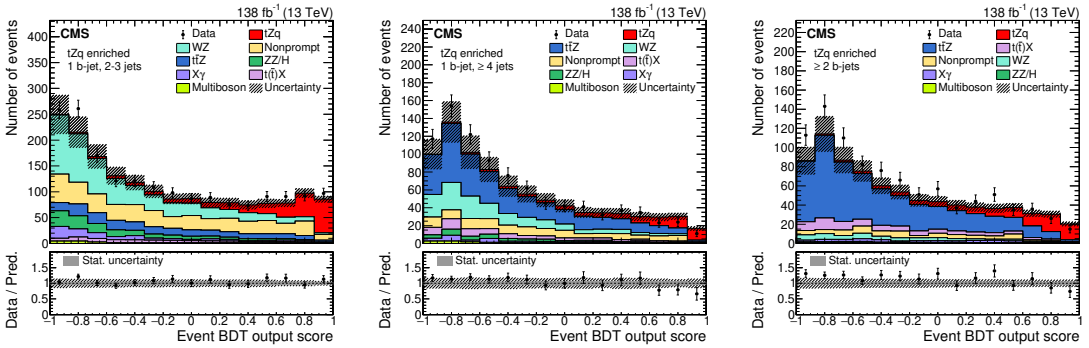


Figure 7.22: BDT output distributions for the three signal region categories considered in this analysis. Left to right: various signal categories (1 b-jet and 2-3 jets, 1 b-jet and 4 or more jets, and 2 or more b-jets respectively). The lower panels show the ratio of the data to the sum of the predictions. The vertical lines on the data points represent the statistical uncertainty in the data; the shaded area corresponds to the total uncertainty in the prediction; the gray area in the ratio indicates the uncertainty related to the limited statistical precision in the prediction.

category. In the other years/categories, the expected and observed significance agree well. These observations are all in agreement with the statistical fluctuations in data observed in the BDT output score distributions in the respective regions.

7.7.2 Inclusive cross-section

The theoretically predicted cross-section for the tZq process where the Z boson decays to a pair of electrons, muons or τ leptons is $\sigma^{SM}(\text{tZq}) = 94.2^{+1.9}_{-1.8}$ (scale) ± 2.5 (PDF) fb [3]. The calculation performed at next-to-leading order in the five-flavour scheme and also includes non-resonant lepton pair production with an invariant mass $> 30\,\text{GeV}$. The two sources of theoretical uncertainty mentioned in the result above are due to variations in the renormalization and factorization scales, and the parton distribution functions respectively. With respect to this predicted cross-section, we observe a signal strength of

$$\mu = \frac{\sigma^{\text{observed}}(\text{tZq})}{\sigma^{SM}(\text{tZq})} = 0.93^{+0.080}_{-0.077}\ (\text{stat})\ ^{+0.078}_{-0.064}\ (\text{syst})\ \text{fb}$$

corresponding to an observed cross-section of

$$\sigma^{\text{observed}}(\text{tZq}) = 87.9^{+7.5}_{-7.3}\ (\text{stat})\ ^{+7.3}_{-6.0}\ (\text{syst})\ \text{fb}$$

This result is obtained by a simultaneous maximum-likelihood fit of all signal regions and control regions for this analysis.

Sources of improvement The inclusive cross-section measurement presented here, with a total uncertainty of about 11%, is the most precise so far, comparing to the 2016+2017 result from CMS (15% [230]) and the full Run II result from ATLAS (15% [231]). The main sources of this improvement are listed below:

- **More data** With respect to the earlier CMS analysis, we use more data (full Run II as opposed to 2016-2017 only). As a check on the magnitude of this effect, we conducted our analysis on the 2016+2017 data and found an expected precision of 13-14%, which is close to the earlier CMS result on the same data set. The remaining difference comes mainly from stronger constraints on the normalization of the nonprompt lepton background. Notice that the 30% improvement on the total uncertainty follows approximately the inverse ratio between the square-root of the luminosity with respect to the previous CMS analysis. This is obvious for the statistical uncertainties, but the systematic uncertainties are affected as well, for example by having higher statistical precision in the control regions used to constrain them.

- **Lepton identification** We re-optimized both the lepton MVA and the lepton identification strategy (to a looser working point). The former optimization allows better working points, i.e. same signal efficiency at stronger background rejection, or higher signal efficiency at similar background rejection. The latter design choice allows for more signal events at the cost of a larger nonprompt background, which in turn can be stronger constrained in a dedicated control region (see also the item above). Although the magnitude of this effect is hard to quantify exactly, an important hint is obtained from conducting our analysis on the 2016+2017 dataset (the same as the previous tZq analysis at CMS). In this case, as also detailed in the item above, we obtain an expected precision of 13-14%. This can be compared to the 15% precision of the previous CMS analysis. Furthermore, from earlier experimental stages in this analysis, we noticed an improvement in expected precision when switching from a tighter working point (with a similar signal efficiency as used in the previous CMS analysis) to a looser one. Based on

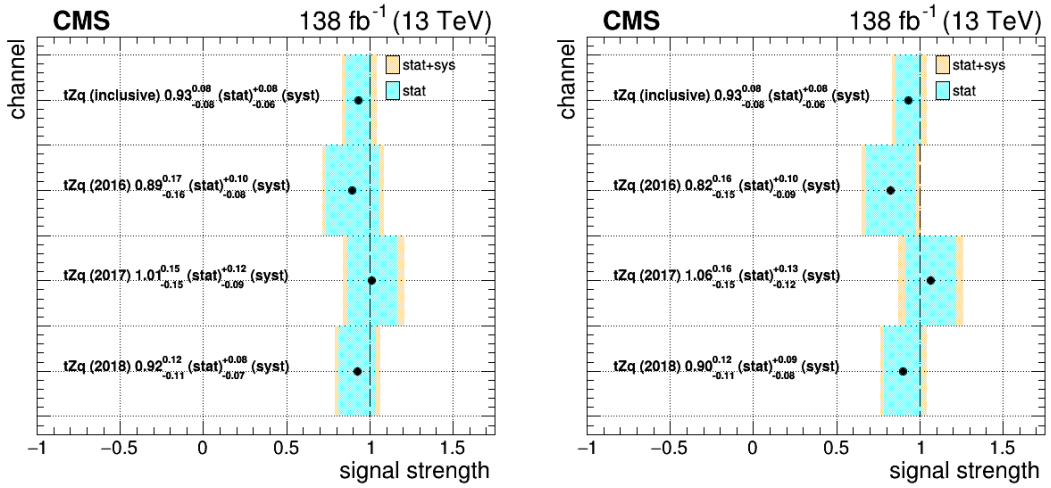


Figure 7.23: Observed signal strength for different data-taking years (labeled '2016', '2017' and '2018') in comparison with their combined measurement (labeled 'inclusive'). Left: simultaneous fit of different signal strength multipliers per data-taking year. Right: fully independent fits per data-taking year.

this, we expect the improvement from 15% to 13-14% to come largely from the different working point and to a lesser extent from the intrinsically better performing lepton MVA.

- **Larger signal region** With respect to the ATLAS analysis, we use a more extensive signal region. The ATLAS analysis is limited to the 1 b-jet, 2-3 jets region, whereas we take into account events with $\geq$ 1 b-jet and $\geq$ 2 jets. In fact we find that when taking into account only this limited signal region, we achieve an expected precision of 13-14%, which is close to ATLAS's result.

Cross-section per year and signal category A compatibility check is performed, comparing the measured signal strength and associated uncertainties in each data-taking year to each other and to the total measurement. The result is shown in Fig. 7.23 (left), where we observe signal strengths compatible with one within one standard deviation for all data-taking years.

The different signal strength multipliers per channel are fitted simultaneously. Hence the fits for the different data-taking years in Fig. 7.23 (left) are not fully independent, as this method takes into account the common scaling of background processes and the correlations between systematic uncertainties and their constraints obtained from e.g. the control regions in all three data-taking years. An alternative compatibility check is performed, where we fit the three data-taking years fully independently from each other. The result of this procedure is shown in Fig. 7.23 (right). We observe qualitatively the same behaviour for both fit results. The systematic uncertainties are however larger in the fully independent fits, as there is less statistical precision in the control regions to constrain them.

Goodness of fit We perform goodness-of-fit tests in order to check how well the observed data agrees with the simulation for the distributions included in the fit. The tests are performed per data-taking year, and the results are shown in Fig. 7.24. The results of the tests (notably the p-values) are in line with the observed signal strengths as measured per data-taking year: the data test statistic in 2017 is close to the median determined from simulated toys (corre-

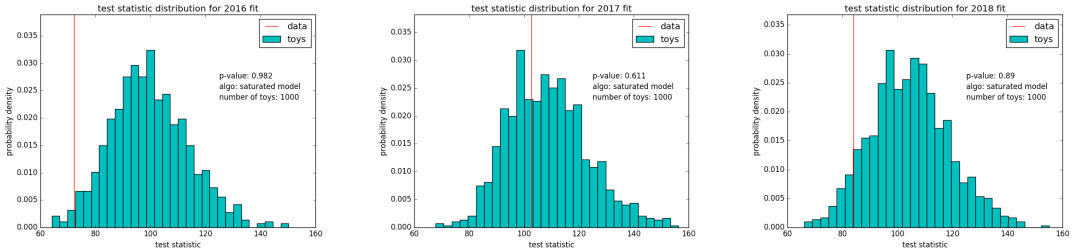


Figure 7.24: Result of goodness-of-fit tests for all data-taking years. The filled histogram shows the distribution of saturated model test statistic for 1000 toys, while the red line indicates the test statistic obtained with the observed data.

sponding to a p-value of 0.5), which corresponds to a good fit and a measured signal strength of around 1; the data test statistics for 2016 and 2018 are further removed from the center of the simulated distribution, corresponding to measured signal strengths deviating from 1.

Impact of uncertainties In Fig. 7.25, we show an overview of the impacts of the dominant sources of systematic uncertainty in this analysis. These uncertainties are: the effects on acceptance from varying the renormalization and factorization scales associated with the tZq and $t\bar{t}Z$ processes, the normalization uncertainties on the backgrounds (as discussed in sections 7.5 and 7.6), and uncertainties in the b-tagging efficiency correction, color reconnection, and parton showering. The impact of the color-reconnection model choice on the tZq signal strength is one-sided since it results from using an alternative model as opposed to a model parameter varied up and down. The best fit values of the nuisance parameters are in good agreement (within one standard deviation) of their prefit values. Also the observed impact of each nuisance parameter on the final measurement is in agreement with the expected value. The measured normalization of $t\bar{t}Z$ events is shifted by about one standard deviation with respect to the theoretical prediction, and indicates an underprediction of this background. This is expected (see section 7.5) and is consistent with earlier measurements [176]. The increased contribution from $t\bar{t}Z$ events after the fit results in a good agreement of prediction to data in both the $t\bar{t}Z$-dominated signal sub-region and the $t\bar{t}Z$ control region with four leptons.

Regarding the dominant systematic uncertainties, we observe substantial differences with respect to the previous tZq analysis [239]:

- **Normalization of nonprompt background** The impact of the normalization of the nonprompt lepton background has been reduced. This is due to a re-optimized lepton identification (see section 7.3). A slightly looser working point allows for larger statistical precision in the control regions, and so for stronger constraints on both prompt and nonprompt uncertainties when including the control regions in the fit.

- **Lepton identification** The uncertainty on the lepton identification scale factors also has a smaller impact than before, since the systematic effects for muons and electrons can be decoupled, following the latest recommendations. Furthermore, new lepton identification (mainly the new MVA) required measuring a new set of scale factors and their associated uncertainties. Furthermore, the uncertainty associated to lepton scale factors proves to be strongly correlated with the WZ normalization uncertainty, so a direct comparison is difficult.

- **b-tagging** Some of the uncertainties associated to b-tagging are larger. This is caused by moving to a different reweighting procedure, as detailed in section 6.5. When employing

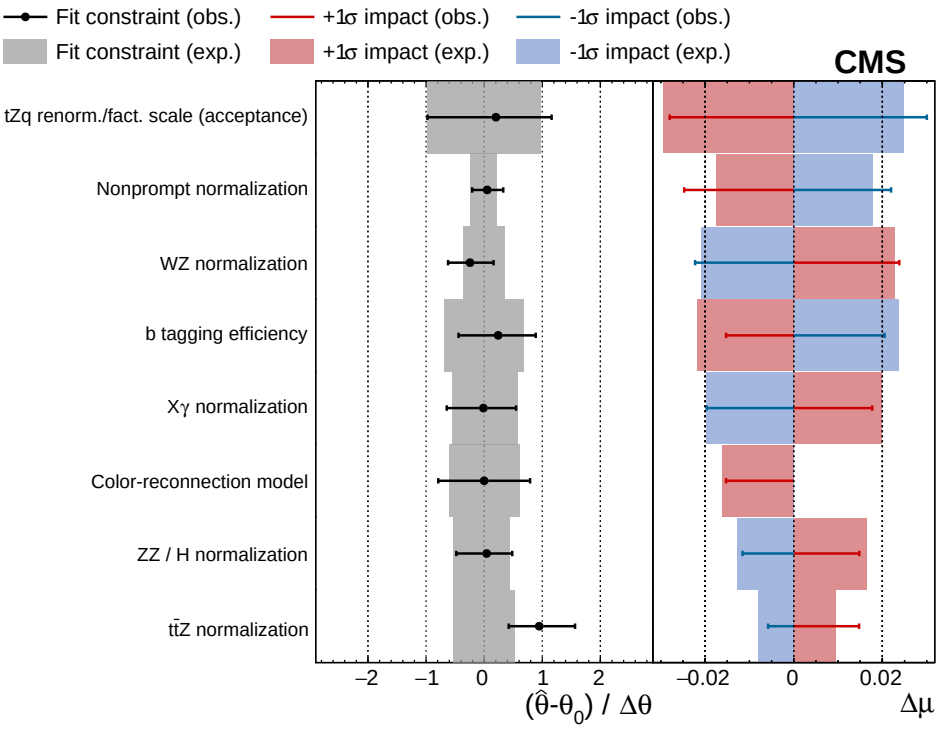


Figure 7.25: Pulls, constraints and impacts of most the important sources of systematic uncertainty for the inclusive tZq cross-section measurement. The left column lists the sources of systematic uncertainty, treated as nuisance parameters in the fit, in order of importance. In the middle column, the black points with the horizontal bars show for each source the difference between the observed best fit value ($\hat{\theta}$) and the nominal value (θ_0), divided by the expected standard deviation ($\Delta\theta$). The right column plots the change in the tZq signal strength μ if a nuisance parameter is varied one standard deviation up (red), or down (blue). The gray, red, and blue bands display the same quantity as their corresponding markers, but using a simulated data set where all nuisance parameters are set to their expected values.

a working-point based reweighting with associated uncertainties instead of a shape-based reweighting, the impacts associated to b-tagging are in the same order of magnitude as the previous tZq analysis.

- **Renormalization and factorization scales** The impacts of the renormalization and factorization scales are larger than anticipated. This is due to changing recommendations, where the envelope over several variations has to be considered (as explained in section 6.5) instead of the simultaneous variation of the renormalization and factorization scale only. When taking into account only the simultaneous variation, the impact of the QCD renormalization and factorization scales reduces to the level of the previous tZq analysis.

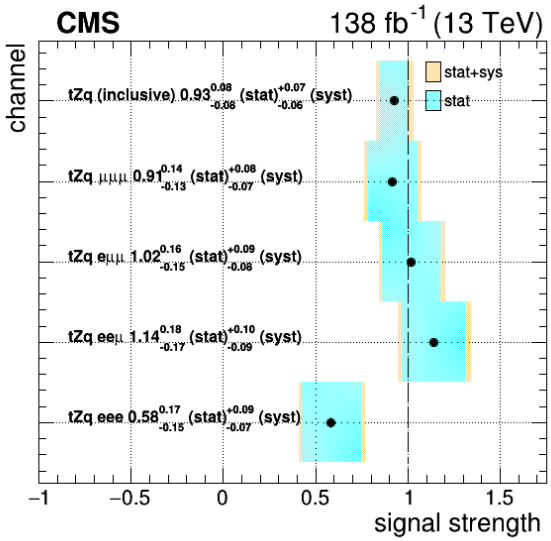


Figure 7.26: Observed signal strength for different measurement channels, split by lepton flavour composition in the final state. The labels in the plot reflect this composition, with e for electrons, m for muons, and 'inclusive' for the fully inclusive measurement.

7.7.3 Cross-sections in different channels

Lepton channels As an additional check, we measure the tZq cross-section in different lepton channels. The signal regions are split into four non-overlapping categories, based on the number of muons in the final state (from 0 to 3), and the fit is performed with each of these alternative signal regions. The result of this procedure is shown in Fig. 7.26. Note that the control regions, that help to constrain the systematic uncertainties, are not split per lepton channel, as this is not feasible for some of them (e.g. control regions containing 4 leptons); instead they are shared between the split signal regions. We observe signal strengths compatible with one within one standard deviation except in the *eee*-channel. This observation is in line with the distribution of the BDT output scores split per lepton channel, as shown in Fig. 7.27.

Top quark and antiquark Next, we measure the tZq cross-section separately for top quarks and antiquarks (hitherto commonly referred to as tZq). The procedure is similar to the measurement per lepton channel: the signal regions are split based on the charge of the lepton associated to the top quark decay and fitted with a different signal strength multiplier, while the control regions are shared. The result is shown in Fig. 7.28. The uncertainty on the top antiquark measurement is larger, since on the one hand it has a smaller cross-section than the top quark process, and on the other hand the BDT is observed to have smaller discriminating power for the top antiquark process.

Of particular interest is the ratio between both sub-processes. The ratio measurement is performed in the following way. We introduce a signal strength parameter that multiplies the normalization of both the top quark tZq and top antiquark tZq. Next, we define another signal strength parameter that applies to the top quark tZq process only. Then we perform a fit (in a similar way as for the fully inclusive measurement) for this second signal strength parameter, which now corresponds to the top quark/antiquark ratio because of the presence of the common multiplier. The result of this procedure is shown in Fig. 7.28 (bottom row). While Fig. 7.28 features signal strengths (i.e. the ratio of observed to expected quantities),

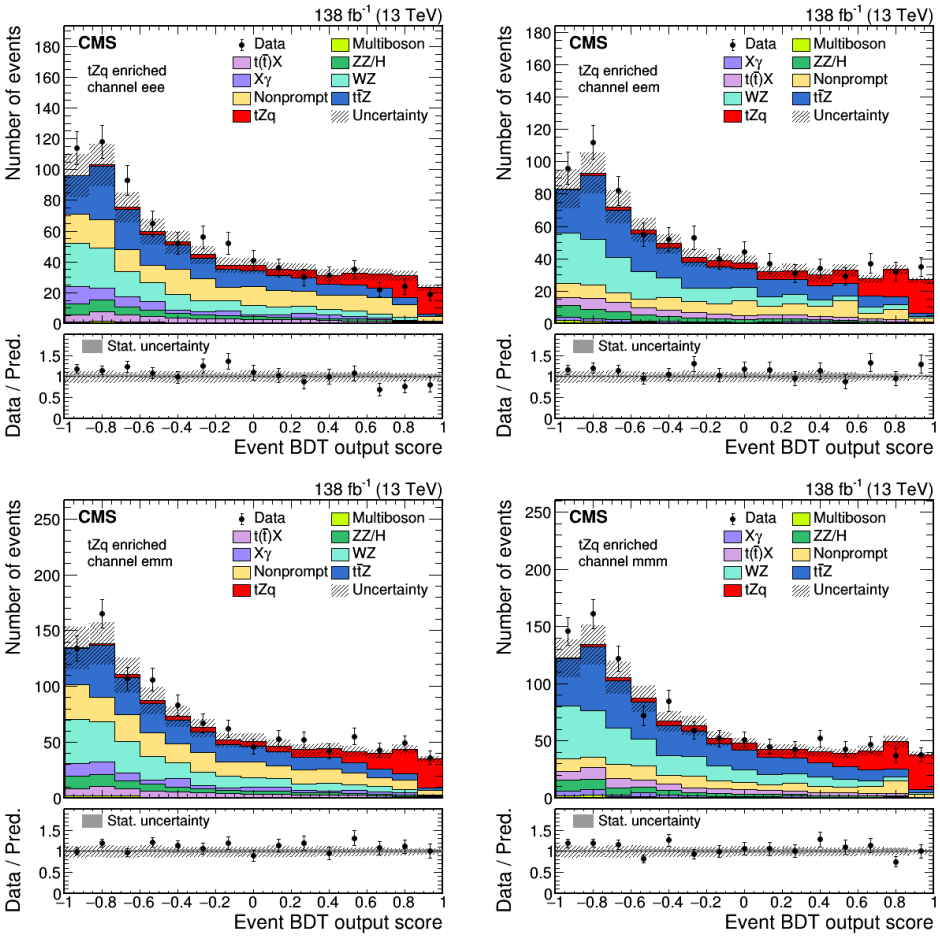


Figure 7.27: BDT output distributions for the four lepton channels considered in this analysis. Each plot contains the three signal categories, but only one specific lepton flavour final state. Upper left: *eee*-channel, upper right: *eem*-channel, lower left: *emm*-channel, lower right: *mmm*-channel.The lower panels show the ratio of the data to the sum of the predictions. The vertical lines on the data points represent the statistical uncertainty in the data; the shaded area corresponds to the total uncertainty in the prediction; the gray area in the ratio indicates the uncertainty related to the limited statistical precision in the prediction.

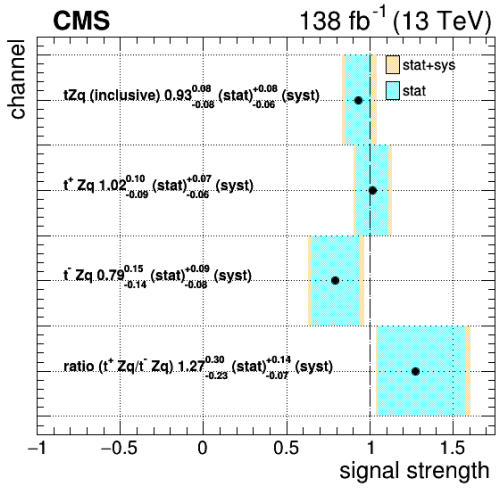


Figure 7.28: Left: observed signal strength for different measurement channels, split by the type of the top quark/antiquark. Note that all rows feature a signal strength, i.e. the observed-to-predicted ratio, not absolute cross-sections or their ratios. Hence the expected values are 1 for each row.

the absolute values for the cross-sections of top quark tZq, top antiquark tZq and their ratio are given below:

$$\sigma_{\mathrm{t}^{+}\mathrm{Zq}} = 62.2^{+5.9}_{-5.7}\ (\mathrm{stat})\ ^{+4.4}_{-3.7}\ (\mathrm{syst})\ \mathrm{fb}$$

$$\sigma_{\mathrm{t}^{-}\mathrm{Zq}} = 26.1^{+4.8}_{-4.6}\ (\mathrm{stat})\ ^{+3.0}_{-2.8}\ (\mathrm{syst})\ \mathrm{fb}$$

$$R_{\sigma_{\mathrm{t}^{+}\mathrm{Zq}}/\sigma_{\mathrm{t}^{-}\mathrm{Zq}}} = 2.37^{+0.56}_{-0.42}\ (\mathrm{stat})\ ^{+0.27}_{-0.13}\ (\mathrm{syst})$$

The relative systematic uncertainty is reduced in the ratio measurement due to partial correlations. These results, while statistically limited, show promising opportunities for the measurement of the top quark to antiquark production cross-section ratio in rare processes such as tZq, yielding additional information to similar results obtained in t-channel single top quark production [244].

7.7.4 Towards a differential measurement

The relatively large data sample used in this analysis, and the good performance of the BDT in distinguishing tZq from its backgrounds, open the window to the first differential measurement of this process. As a first step, the feasibility of differential measurements is highlighted by showing a relatively pure tZq signal region, as obtained from the default signal selection used in this analysis by extending it with the requirement that the BDT score is > 0.5. Several interesting observables in this region are shown in Fig. 7.29. The number of tZq signal events passing this selection is about 252, and the number of background events about 264, implying that the tZq purity of this region is about 49%. Furthermore, these plots show a good sensitivity to the kinematic properties of the tZq signal and a good agreement between the data and simulation (with the exception of the m(3ℓ) variable).

The imperfect simulation of reconstruction inefficiencies and detector acceptance effects (combined and labeled as 'detector level' effects) could result in discrepancies between the shape

and normalization of the measured distributions and the simulated ones. The good agreement between the measured and simulated distributions in the Fig. 7.29 shows that these possible effects are relatively small in this phase space. Furthermore, the number of events in the data associated with the tZq process means that a differential cross-section measurement is possible, once the detector-level effects are corrected for. The differential measurement itself is, as already explained, not detailed further in this dissertation, but can be found in Ref. [1].

7.8 Conclusions and outlook

Summary This chapter discussed the inclusive cross-section measurement of the tZq process (i.e. single top quark or antiquark production in association with a Z boson). It is important to measure and constrain this process in the context of self-consistency checks of the standard model via precision measurements of rare processes. The tZq process is a suitable probe for some beyond-standard-model theories, however in this measurement we do not observe a deviation from the cross-section predicted by the standard model, within the theoretical, experimental and statistical uncertainties.

Statistical limitations The measurement is currently still statistically limited, implying that collecting more data (e.g. in Run III) will 'automatically' reduce the uncertainty further. The margin for improvement is small however, as the systematic uncertainties are not much smaller than the statistical ones. Therefore, in parallel with recording more data, additional work on estimating and reducing the systematic uncertainties is required.

Modeling uncertainty The largest systematic uncertainty in the current measurement is the impact of the renormalization and factorization scales on the expected shape of the tZq signal. Those scales are technical parameters in the simulation of tZq events and do not have an objective physical meaning; improvements in the modeling techniques or using alternatively generated samples (e.g. the 5FS sample instead of the 4FS sample) could reduce this major uncertainty.

Background normalization On the experimental side, two main points for improvement can be discerned. The fake rate measurement used for estimating the nonprompt background comes with a single coarse normalization uncertainty. While this is acceptable for the initial, 'quick-and-dirty' observations of rare processes, the time has come for a better understanding and splitting of the physical sources of uncertainty associated with this method, especially since the current uncertainty gets strongly constrained in the fit. The same holds for the normalization uncertainties of important background processes. We need more reliable estimates of the data-to-simulation agreement of these backgrounds in the signal regions relevant to tZq (or similar top quark processes), by dedicated studies of these backgrounds in a phase space with a high jet and/or b-jet multiplicity.

Global fits In more general terms, in Run III and beyond, the focus of analysts is expected to shift from measuring individual processes to performing global fits of many related processes simultaneously, with a proper treatment of the correlation of systematic uncertainties and a general parameterization of potential beyond-standard-model effects (e.g. in the SMEFT framework).

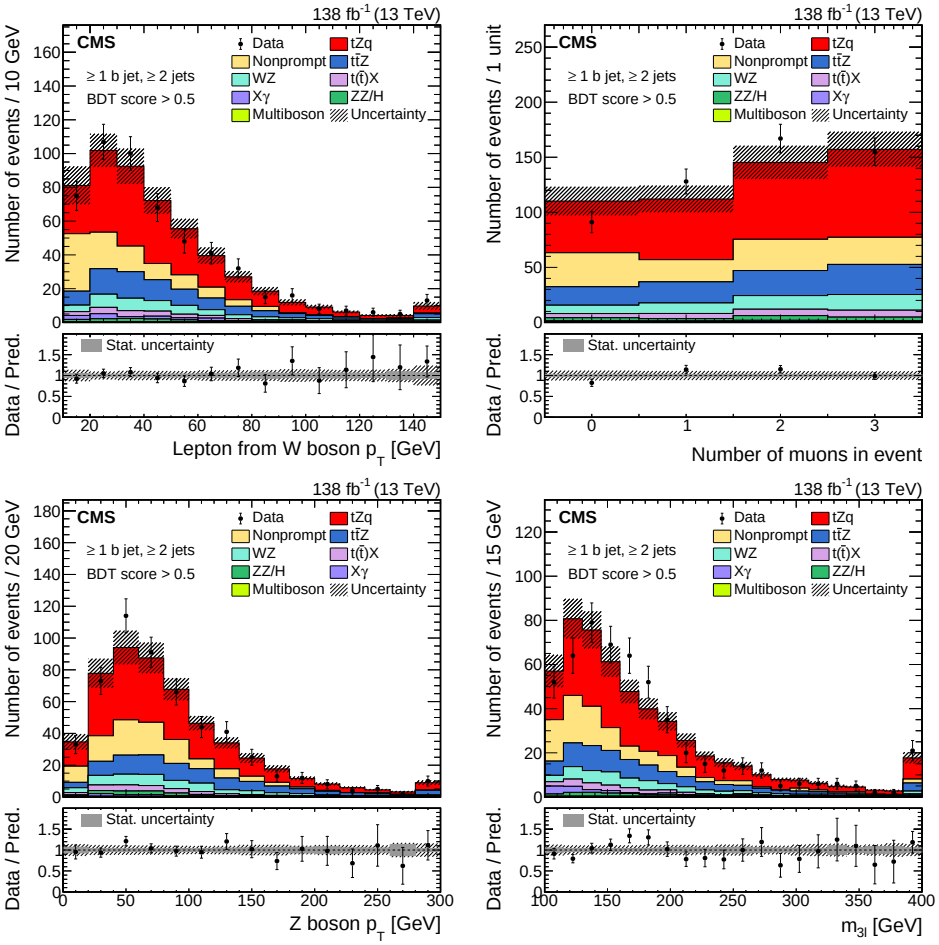


Figure 7.29: Distributions of several interesting kinematic variables in the signal-enriched region for data (black points) and predictions (colored histogram). The event selection consists of the default signal selection employed in this analysis, extended with a cut on the BDT to maximally increase the tZq purity in the selection. The lower panels show the ratio of the data to the sum of the predictions. The vertical lines on the data points represent the statistical uncertainty in the data; the shaded area corresponds to the total uncertainty in the prediction; the gray area in the ratio indicates the uncertainty related to the limited statistical precision in the prediction.

Chapter 8

Top quark pair in association with a W boson

In this chapter we study the production of a top quark-antiquark pair in association with a W boson (shortly called $t\bar{t}W$ or the $t\bar{t}W$ process). The results of this research have been documented in a CMS internal note [15] and published in a paper [2] (currently in CMS internal review)[1]. This research project was carried out in close collaboration with the CMS group at Oviedo. We pursued two conceptually different approaches as a mutual cross-check (unfolding with and without a BDT discriminator), and the Oviedo group also performed a measurement of the charge asymmetry in $t\bar{t}W$. The focus in this dissertation will be only on my direct contribution.

[1] In more detail and CMS jargon: at the time of writing, the analysis has been pre-approved and unblinding is currently ongoing.

8.1 Introduction

The leading order diagrams contributing to the production of a top quark-antiquark pair in association with a W boson ($t\bar{t}W$) are shown in Fig. 8.1. At leading order, $t\bar{t}W$ production proceeds via annihilation of an up-type quark with a down-type antiquark or vice versa, where the W boson is radiated from one of the initial state quarks. The W boson can only be radiated from the initial state, as radiation from the final state top quarks corresponds to ordinary $t\bar{t}$ production. At next-to-leading order, additional production channels are available for $t\bar{t}W$ production, initiated by a quark-gluon interaction and containing an additional jet in the final state. The $t\bar{t}W$ process is a particular process of interest to the CMS physics program for a number of reasons.

Production asymmetries The electric charge and flavour changing nature of the W boson make $t\bar{t}W$ a unique process among all $t\bar{t}V$ processes (with V representing any vector boson, e.g. Z, H, etc.), with an interesting phenomenology. Because of the dominating contribution of valence up-quarks with respect to valence down-quarks in the proton parton distribution functions, a production asymmetry occurs, where $t\bar{t}W^+$ production (with the W^+ boson emitted from an up-type quark or down-type antiquark) is more probable than $t\bar{t}W^-$ (with the W^- boson emitted from a down-type quark or up-type antiquark). For the same reason, there is a forward-central asymmetry between the top-quark and top-antiquark (see section 5.4.2 for more details).

Modeling difficulties The leading-order and next-to-leading-order cross-section of the $t\bar{t}W$ process is suppressed with respect to other $t\bar{t}V$ processes because of the fact that there are no gluon-gluon initial states, which dominate production processes at the energies of the LHC. Moreover, the relative corrections of next-to-leading order with respect to leading-order contributions are large (see section 8.2.3 for more details). This makes the $t\bar{t}W$ process a difficult and interesting process from the perspective of theory predictions.

Earlier measurements The inclusive cross-section of the $t\bar{t}W$ process has been the subject of dedicated measurements by CMS [214] and ATLAS [180], confirming a tension between the prediction and the measurement that was earlier hinted at by analyses where $t\bar{t}W$ was an important background process. This fact further increases the desirability of a differential $t\bar{t}W$ measurement, in order to gain a more detailed understanding of the potential origins of this tension. These differential measurements have already been reported by ATLAS [180, 191], but not yet by CMS. This chapter therefore presents the first differential measurements of the $t\bar{t}W$ process in CMS, extending on the inclusive results reported earlier and cross-checking the differential measurements by ATLAS.

This study in a nutshell In this study, we use the full Run II data set consisting of data taken during 2016, 2017 and 2018 at a center of mass energy $\sqrt{s}$ of 13 TeV and corresponding to an integrated luminosity of 138 fb^{-1} . We benefit from the substantial work performed since the end of Run II data taking to retro-mitigate observed detector issues and to study, understand and where possible reduce systematic experimental biases and uncertainties (all of which is summarized in the so-called "legacy" simulation campaign).

This measurement targets mainly the same-sign dilepton final state, where the W boson and one of the top quarks decay leptonically (to electrons or muons) and the other top quark hadronically. We also study the trilepton final state, but this channel suffers from a lower

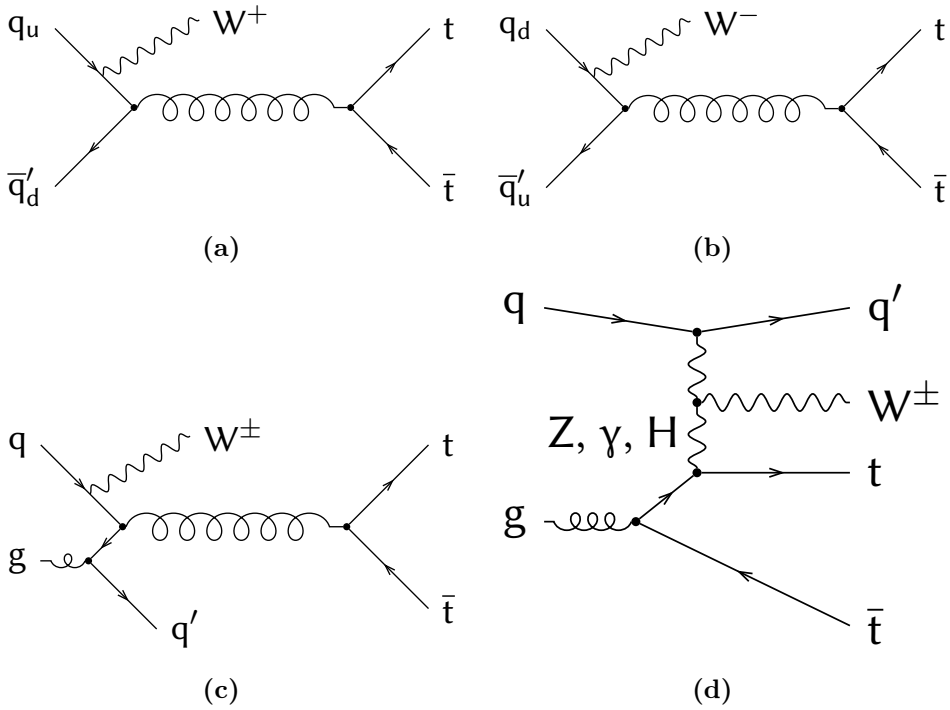


Figure 8.1: Representative feynman diagrams for the $t\bar{t}W$ process at leading order (upper row) and next-to-leading order (lower row). Figures copied from Ref. [221].

number of events and a worse signal to background ratio[2]. We exclude τ leptons from the analysis, and by leptons we will in this entire note refer to electrons and muons only[3]. We use MVA discriminants at two stages in the analysis: first, a lepton MVA is used to discriminate prompt from nonprompt leptons; secondly, an event-level MVA is used to discriminate signal events from background events. It is the interplay between both levels of MVAs that forms the main dynamic of this analysis.

After selecting events enriched in $t\bar{t}W$ signal, we perform a maximum-likelihood based unfolding procedure for the differential measurements. This procedure fits several contributions to the $t\bar{t}W$ signal (corresponding to several particle level bins) separately but simultaneously in a multidimensional maximum-likelihood fit. The unfolding is performed as a function of several variables of interest, gathered from input from both theory-oriented and experiment-oriented sources.

We compare the differential measurements to the simulated $t\bar{t}W$ signal sample currently in use by CMS, as well as to a custom sample produced with state-of-the-art improvements in NLO jet merging, expected to account for at least a part of the previously observed tension.

This chapter is structured as follows. In section 8.2, we list the data sets and simulation samples used in this analysis, and discuss the dedicated signal sample. Sections 8.3 and 8.4 are devoted to object selection and event selection respectively. In both cases, a major part is devoted to the respective MVA techniques mentioned above. Then we turn to a discussion on

[2]On the other hand, the trilepton final state is the channel of choice for the measurement of the charge asymmetry in $t\bar{t}W$, but that falls, as mentioned, outside the focus of this dissertation.

[3]As was the case for the tZq analysis, hadronically decaying τ leptons are excluded from the analysis, but promptly decaying τ leptons to electrons or muons are implicitly included.

the most important background processes in section 8.5. In section 8.6, we list the systematic uncertainties that affect this measurement and how they are taken into account. Section 8.7 deals with some inclusive measurements as cross-checks, before transitioning into differential measurements in section 8.8. The results are summarized and an outlook is given in section 8.9.

8.2 Data and simulated samples

8.2.1 Data

This analysis uses a set of proton-proton collisions delivered by the LHC at a center-of-mass energy of 13 TeV and collected by the CMS detector. We use the full Run II data set, consisting of data collected in 2016 (36.3 fb^{-1}), 2017 (41.5 fb^{-1}) and 2018 (59.8 fb^{-1})[4], adding up to a total integrated luminosity of 138 fb^{-1} . Details on the luminosity measurement and calibration for the three data taking years can be found in [127–129]. We select only data samples that was certified to be of excellent quality for physics analysis[5].

Datasets In order to collect as many relevant data events as possible, we combine several primary datasets, where overlap between them (i.e. duplicate events) is removed based on run / lumiblock / event number. The datasets of choice are those that contain trigger paths selecting one or more reconstructed light leptons in the event[6]. We use the legacy reprocessing of all data sets[7].

Triggers On the trigger level, all events are required to pass at least one of a collection of single-lepton, dilepton and trilepton triggers[8]. The efficiency of this trigger mix has been assessed using the orthogonal trigger method. An efficiency of essentially 100% with no significant deviations between data and simulation was found for selections with three or more leptons; for the dilepton selection used in this study, significant differences between the trigger efficiency in data and simulation were found and taken into account via correction factors. More details on the trigger efficiency measurement can be found in section 6.4.

[4]The attentive reader might have noticed that the 2018 luminosity here is slightly higher than in the tZq chapter. This is because the data used for the $t\bar{t}W$ analysis corresponds to a more recent reprocessing, in which some previously excluded data sections were recovered.

[5]In CMS technical jargon, this certification is summarized in the so-called "golden json files". In this case, the used golden json files are the following

- `https://cms-service-dqmdc.web.cern.ch/CAF/certification/Collisions16/13TeV/Legacy_2016/Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON.txt`
- `https://cms-service-dqmdc.web.cern.ch/CAF/certification/Collisions17/13TeV/Legacy_2017/Cert_294927-306462_13TeV_UL2017_Collisions17_GoldenJSON.txt`
- `https://cms-service-dqmdc.web.cern.ch/CAF/certification/Collisions18/13TeV/Legacy_2018/Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON.txt`

[6]In CMS technical jargon, the datasets are named SingleElectron, SingleMuon, DoubleEG, DoubleMuon and MuonEG for 2016 and 2017 data, and EGamma, SingleMuon, DoubleMuon and MuonEG for 2018 data.

[7]In CMS technical jargon, the specific versions of each dataset are: 21Feb2020_UL2016_HIPM (2016 PreVFP data), 21Feb2020_UL2016 (2016 PostVFP data), 09Aug2019_UL2017 (2017 data), 12Nov2019_UL2018 (2018 data), where the UL stands for "ultra-legacy".

[8]The full list of used triggers per data-taking year is given in Tab. B.2 (in the appendix).

8.2.2 Simulation of background processes

Generators and parton showering We employ Monte-Carlo (MC) simulated samples corresponding to all standard model processes relevant in the phase space of this analysis[9]. Most samples are generated at next-to-leading-order (NLO) precision in QCD, using the MadGraph5_aMC@NLO [135] or POWHEG [236] programs. For some cases, in which NLO samples are not available or suffer from low statistics, leading order (LO) samples, generated with MadGraph 5 [135] are used. The modeling of parton showering, hadronization and the underlying event is performed using PYTHIA [93], with the CP5 tune [94]. Two separate matching schemes are used to avoid double counting of partons generated in the different simulation steps: the FXFX [136] scheme for NLO samples, and the MLM [237] scheme for LO samples.

Detector simulation All simulated events are processed by a GEANT4-based [98] simulation of the CMS detector, and in a next step reconstructed using the same algorithms applied to data. For the modeling of additional proton-proton collisions present in data events (pileup), the simulated events are extended with simulated pileup events, and the distribution of the number of additional interactions in the simulated samples is reweighted to match the distribution observed in data. We use the legacy production of all simulated samples[10].

Categorization Similar to the approach for the tZq analysis, the samples in Tabs. A.3–A.4 are grouped per background category. The samples listed under 'Nonprompt background' contribute to the signal region only via nonprompt leptons; these samples are only used for cross-checking intermediate results; in the final analysis, the nonprompt background is estimated directly from data (see section 8.5).

8.2.3 Simulation of the $t\bar{t}W$ process

Modeling $t\bar{t}W$ The $t\bar{t}W$ process is notoriously difficult to model correctly, with large and counter-intuitive corrections at higher perturbative orders. The main reason for this is that new production channels (i.e. new initial states) become available only at higher orders. At LO, only the q $\overline{q}$' channel is available, since the fact that the W boson has to be radiated from the initial rather than from the final state (else the process is $t\bar{t}$ rather than $t\bar{t}W$) excludes the q g and gg channels. The former of these becomes available at NLO (see Fig. 8.1), while the latter enters only at NNLO [245]. Moreover, even within NLO, electroweak corrections that are supposed to be small turn out to be larger than anticipated [246]. Furthermore, at NLO, the final state contains an additional light jet (see Fig. 8.1). This increases the dependency of this process on the procedure for adding extra jets in the parton shower and the correct overlap removal between jets from the matrix element and from the parton shower [247].

Signal sample The default $t\bar{t}W$ sample (TTWJetsToLNu) in Tab. A.3 contains the LO $t\bar{t}W$ production diagrams of order $\alpha_S^2\alpha_{EW}$ (see Fig. 8.1, upper row), as well as NLO contributions in QCD of order $\alpha_S^3\alpha_{EW}$ (see Fig. 8.1, lower left), where α_S and α_{EW} are the strong and electroweak coupling constant respectively. It does however not include electroweak corrections, i.e. terms of the order α_{EW}^3 at 'LO' in EW or $\alpha_S\alpha_{EW}^3$ at NLO (see Fig. 8.1, lower right). These electroweak correction term are produced and included as an additional sample (ttWJetsToLNu_5f_EWK). Note that in the previous $t\bar{t}W$ measurement [214, 221], the same procedure was applied, but with a privately produced electroweak corrections sample

[9] A list of specific samples used to model each process is provided in Tabs. A.3–A.4 (in the appendix).

[10] In CMS technical jargon, the specific versions of each sample are: RunIISummer20UL16MiniAODAPV (2016 PreVFP conditions), RunIISummer20UL16MiniAOD (2016 PostVFP conditions), RunIISummer20UL17MiniAOD (2017 conditions), RunIISummer20UL18MiniAOD (2018 conditions), where the UL stands for "ultra-legacy"

instead of a centrally produced one (as it was not available at the time).

In general, the LO (QCD + EW) terms for $t\bar{t}W$ are of the order $\alpha_S^i\alpha_{EW}^j$ with $i+j=3$ and $j \geq 1$. No diagrams with $\alpha_S\alpha_{EW}^2$ exist, hence the only remaining contributions are given by $\alpha_S^2\alpha_{EW}$ and α_{EW}^3, both of which are included in our samples. In the case of NLO, we have $\alpha_S^i\alpha_{EW}^j$ with $i+j=4$ and $j \geq 1$. Our samples include the $\alpha_S^3\alpha_{EW}$ and $\alpha_S\alpha_{EW}^3$ contributions, but not the $\alpha_S^2\alpha_{EW}^2$ and α_{EW}^4 contributions (at matrix element level). These latter ones contain a photon in the initial or final state however, which is partly recovered in the parton showering. Moreover, their overall contribution is calculated to be small with respect to the other NLO contributions [246].

Technical information More technical information on the current, default $t\bar{t}W$ sample can be found in the McM database entry[11] and the corresponding datacards[12]. The TTWJetsToLNu sample is generated as $pp \to t\bar{t}\ell\nu(+j)$, where ℓ is any lepton flavour (also including τ), ν the appropriate neutrino, and j up to one additional jet at matrix-element level. The top quark mass is set to 172.5 GeV and all its allowed decay modes are included. The renormalization and factorization scales were set to the CMS conventional choice of the sum of the transverse masses of all final state particles and partons, divided by two. For more technical information on the electroweak corrections $t\bar{t}W$ sample, we refer to its McM database entry[13] and its datacards[14] as well.

Improved merging In addition to these standard samples, we plan to use a privately produced, state-of-the art $t\bar{t}W$ sample with an improved matching procedure tailored specifically to this process [247]. The improvement in the merging procedure between jets from the matrix element and those from the parton shower consists in explicitly distinguishing between jets originating from a QCD vertex (see Fig. 8.1, lower left) and those originating from an electroweak vertex (see Fig. 8.1, lower right). Jets from electroweak vertices, which were previously modeled inconsistently below and above the merging scale[15], are now taken from the matrix element (as opposed to from the parton shower) over the full phase space, giving a more consistent description. This sample is currently undergoing some checks before production. We refer to the Oviedo group [248] for more details on the private sample production settings and status.

Cross-section The prefit cross-section value to normalize the $t\bar{t}W$ sample(s) has been determined as follows. The agreed-upon reference $t\bar{t}W$ cross-section value for ATLAS and CMS measurements is 0.7453 pb, following a recent full NLO (QCD + electroweak) + approximate NNLO QCD calculation in Ref. [245]. This cross-section is inclusive in the decays of the top quarks and W boson, so to normalize the TTWJetsToLNu sample the branching ratio of the W boson to leptons has to be factored in. Additionally, we have to 'divide' the cross-section between the default and electroweak correction samples. Our strategy here is to again follow Ref. [245], which provides the QCD-only $t\bar{t}W$ cross-section separately. In detail, we arrive at the following values:

[11] `https://cms-pdmv.cern.ch/mcm/requests?prepid=TOP-RunIISummer20UL18wmLHEGEN-00071&page=0&shown=3932287` (CMS internal).

[12] `https://github.com/cms-sw/genproductions/tree/master/bin/MadGraph5_aMCatNLO/cards/production/2017/13TeV/TTWJets/TTWJetsToLNu_5f_NLO_FXFX`

[13] `https://cms-pdmv.cern.ch/mcm/requests?prepid=HIG-RunIISummer20UL18wmLHEGEN-01672&page=0&shown=262271` (CMS internal).

[14] `https://github.com/ataliercio/genproductions/tree/new_cards/bin/MadGraph5_aMCatNLO/cards/production/2017/13TeV/UL/TTWJets/TTWJetsToLNu_EWK_5f_NLO`

[15] In more detail, jets from electroweak vertices are not modeled by the parton shower; hence they were completely absent below the merging scale in LO calculations, and present only at LO accuracy below the merging scale in NLO calculations.

- TTWJetsToLNu sample: $\sigma(\mathrm{t\bar{t}W}, \mathrm{QCD}) \times \mathrm{Br}(\mathrm{W} \to \ell\nu)$
  $= 0.7107 \times (1 - 0.6741) = 0.2316$ pb.
- ttWJetsToLNu_5f_EWK sample: $(\sigma(\mathrm{t\bar{t}W}, \mathrm{total}) - \sigma(\mathrm{t\bar{t}W}, \mathrm{QCD})) \times \mathrm{Br}(\mathrm{W} \to \ell\nu)$
  $= (0.7453 - 0.7107 \times (1 - 0.6741) = 0.01127$ pb.

8.3 Object selection

This analysis relies on the particle flow (PF) algorithm [77] to perform the reconstruction of individual particles in proton-proton collision events from the corresponding detector signals. This algorithm uses an optimized combination of information from all CMS subdetectors in order to identify muons, electrons, photons, charged hadrons or neutral hadrons. Next, the anti-k_T algorithm [81] is used to cluster reconstructed particles into jets, using a distance parameter of 0.4[16]. All these objects are used as an input for further selection and identification, as detailed in the following subsections.

8.3.1 Jets and missing transverse energy

Jets are required to pass the following selection criteria:

- $p_\mathrm{T} > 25\,\mathrm{GeV}$
- $|\eta| < 2.4$
- The jet must be separated from any identified lepton by $\Delta R > 0.4$
- Jet must pass the tight jet ID as defined by the JetMET POG[17]

The tight jet ID imposes conditions on the jet substructure (e.g. number of constituents and relative amount of charged and neutral components), aimed at removing anomalous energy deposits. To account for known and calibrated differences between jets in simulation and in data, jet energy corrections [82–84], specifically for the legacy reprocessing [249], are applied to all simulated samples.

Jets originating from b-hadrons, shortly called b-jets, are identified using the DeepJet/DeepFlavour algorithm [85–87]. In this analysis, we use the loose working point as a threshold for defining b-tagged jets, which is designed to have a b jet efficiency of about 95% (for jets with $p_\mathrm{T} > 30\,\mathrm{GeV}$). The light jet misidentification rate (i.e. the probability that a light jet is identified as a b-jet) is about 10%, and the c jet misidentification rate is about 50%. Scale factors to correct for known differences in b-tagging efficiency between simulation and data (see section 8.6) are taken from central measurements by the b-tagging POG, following their recommendations[18].

To every event a missing transverse momentum vector ($\vec{p}_\mathrm{T}^{\,\mathrm{miss}}$) is assigned, defined as the negative vector sum of all objects reconstructed by the jet finding algorithm. It represents an estimate of the combined transverse momentum (direction and magnitude) carried by particles that cannot be detected by CMS, notably neutrinos. The magnitude of this vector is denoted as $p_\mathrm{T}^{\mathrm{miss}}$.

[16] The particle-flow and anti-k_T algorithm are described in more detail in sections 2.12.2 and 2.12.5 respectively.

[17] `https://twiki.cern.ch/twiki/bin/view/CMS/JetID` (CMS internal)

[18] `https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation` (CMS internal)

8.3.2 Leptons

As in the case of the tZq analysis, accurate lepton identification is of major importance. Our main tool to discriminate prompt from nonprompt leptons consists of a prompt lepton MVA (discussed earlier in section 6.1). Apart from an MVA threshold, baseline selections are applied to both muons and electrons that make sure the objects are within the limits of what can be accurately reconstructed by the CMS detector and that they are within the phase space of the lepton MVA training. The lepton identification in this analysis has been re-optimized for legacy data and simulation (hence the small differences with respect to the lepton identification in the tZq analysis). This was originally done in the context of the four top quark discovery [4] and the same identification is used here.

Muon selection As baseline selection (also called 'loose ID' in this note) for muons, we apply a p_{T} threshold (to allow accurate reconstruction) and an η threshold (to make sure the muons are within the muon spectrometer acceptance), as well as some loose cuts on impact parameter, relative isolation and a POG-defined ID. These are exactly the same as the cuts used to delimit the phase space for the lepton MVA training and testing. For the final analysis we apply an additional selection on the lepton MVA score (this selection is called 'tight ID' in this note). In between the loose and tight ID is a 'fakeable object' (FO) ID, used in the estimation of the nonprompt background from data (see section 8.5.2). The different muon identifications are summarized in Tab. 8.1.

Table 8.1: Summary of muon identifications used in this analysis. Values between parentheses indicate that the threshold is only applied if the muon fails the lepton MVA threshold. The definition of the variables is given in Tab. 6.1.

Muon selection			
Observable	Loose	Fakeable	Tight
p_T	> 10 GeV	> 10 GeV	> 10 GeV
$\|\eta\|$	< 2.4	< 2.4	< 2.4
$\|d_{xy}\|$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$\|d_z\|$	< 0.1 cm	< 0.1 cm	< 0.1 cm
$\|d\|/\sigma_d$	< 8	< 8	< 8
Relative isolation	< 0.4	< 0.4	< 0.4
Muon POG ID	$>$ WP-medium	$>$ WP-medium	$>$ WP-medium
DeepJet of nearest jet	-	- (< 0.025)	-
Jet p_T ratio	-	- (> 0.45)	-
Lepton MVA score	-	> 0.64 (< 0.64)	> 0.64

Electron selection The baseline selection thresholds for electrons ('loose ID') are motivated analogously as for muons. We apply additional thresholds on conversion veto and charge misidentification on top of the baseline cuts, since they were observed to significantly increase the signal to background ratio. The 'fakeable object' (FO) ID is used to estimate the nonprompt background from data (see section 8.5.2). The different electron selections used in this analysis are summarized in Tab. 8.2.

Table 8.2: Summary of electron identifications used in this analysis. Values between parentheses indicate that the threshold is only applied if the muon fails the lepton MVA threshold. Values separated by slashes indicate different values for 2016PreVFP/2016PostVFP/2017/2018. The definition of the variables is given in Tab. 6.1.

Electron selection			
Observable	Loose	Fakeable	Tight
p_T	> 10 GeV	> 10 GeV	> 10 GeV
$\|\eta\|$	< 2.5	< 2.5	< 2.5
$\|d_{xy}\|$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$\|d_z\|$	< 0.1 cm	< 0.1 cm	< 0.1 cm
$\|d\|/\sigma_d$	< 8	< 8	< 8
Relative isolation	< 0.4	< 0.4	< 0.4
Missing hits	< 2	< 2	< 2
Conversion rejection	-	✓	✓
Charge consistency	-	✓	✓
EGamma POG MVA score	-	- ($>$ WP-loose)	-
DeepJet of nearest jet	-	- (< 0.1)	-
Jet p_T ratio	-	- ($> 0.5/0.5/0/4/0.4$)	-
Lepton MVA score	-	> 0.81 (< 0.81)	> 0.81

Prompt lepton MVA strategy We employ a dedicated MVA to discriminate prompt from nonprompt leptons. It consists of a gradient boosted BDT trained with the XGBOOST package. The training set consists of simulated samples from the full Run-II simulation with legacy reconstruction, and is general enough to be applicable to multiple analyses involving top quarks. Prompt leptons are taken from a variety of processes with one or multiple top (anti-)quarks, and nonprompt leptons from $t\bar{t}$. More details are described in section 6.1. It is essentially the same lepton MVA as used in the tZq analysis, but re-optimized for legacy data and simulation.

The chosen working points correspond to a prompt lepton efficiency of 93% for muons (with a nonprompt mistagging rate of 3%) and 89% for electrons (with a nonprompt mistagging rate of 2%) [206]. These working points are comparable to the ones used in the tZq analysis (see section 7.3). They are relatively loose, with a large signal efficiency and correspondingly sizable nonprompt contribution passing the lepton MVA. The larger event yield in the nonprompt contribution is however taken care of by the event-level MVA (described in section 8.4), which benefits from a larger training set in this strategy. Moreover, the nonprompt prediction can more accurately be checked and constrained, both in the signal region (at low event-level MVA scores) and in dedicated control regions.

The scale factors for the lepton identification used in this analysis (and their associated uncertainties) have been measured by the lepton MVA developer using the official tag-and-probe recipes and conventions, and have been approved by the Muon POG and EGamma POG. The scale factors are used to reweight the simulation in the remainder of the analysis and the associated uncertainties are propagated to the final cross-section measurement.

8.4 Event selection

Signal region We consider two signal regions in this analysis, the first and major one targeting the same-sign dilepton final state, the second and minor one targeting the trilepton final state. In a first stage, events are rejected if they do not pass any of the triggers (see section 8.2.1). Furthermore, we apply the MET filters as recommended by the JetMET POG[19] to suppress events with large amounts of missing transverse momentum coming from noise, cosmic rays and beam-halo. We also do not consider events in which no valid primary vertex could be reconstructed. These selections are common to both signal regions (and control regions); the next selections are region-specific and are discussed in the subsections below.

Dilepton signal region The region-specific event selections for the dilepton signal region are the following:

- Exactly two same-sign leptons (vetoing any deficit or surplus).
- Transverse momentum thresholds of 25 and 15 GeV for leading and sub-leading lepton.
- Invariant mass safety requiring $m(\ell\ell) > 30\,\text{GeV}$.
- Veto events with $|m(\ell\ell) - m(\text{Z})| < 10\,\text{GeV}$ if both leptons are electrons (in order to reduce the charge misidentified contribution).
- $E_{\text{T}}^{\text{miss}} > 30\,\text{GeV}$ (in order to reduce the nonprompt contribution and to ensure orthogonality with one of the control regions enriched in nonprompt background).
- Number of jets ≥ 3.
- Number of b-tagged jets (using the loose working point) ≥ 2.

Distributions of signal and background processes in the dilepton signal region are shown in Fig. 8.2.

Trilepton signal region The region-specific event selections for the trilepton signal region are the following:

- Exactly three leptons (vetoing any deficit or surplus).
- Transverse momentum thresholds of 25, 15 and 15 GeV for leading, sub-leading and trailing lepton respectively.
- Invariant mass safety requiring $m(\ell\ell) > 30\,\text{GeV}$ for all pairs of leptons.
- Veto events with Z candidate, defined as an OSSF lepton pair with $|m(\ell\ell) - m(\text{Z})| < 10\,\text{GeV}$.
- Events must have an OS lepton pair (in other words: the sum of charges of the three leptons must be +1 or -1).
- Number of jets ≥ 2.
- Number of b-tagged jets (using the loose working point) ≥ 2.

Distributions of signal and background processes in the trilepton signal region are shown in Fig. 8.3.

[19] `https://twiki.cern.ch/twiki/bin/view/CMS/MissingETOptionalFiltersRun2` (CMS internal)

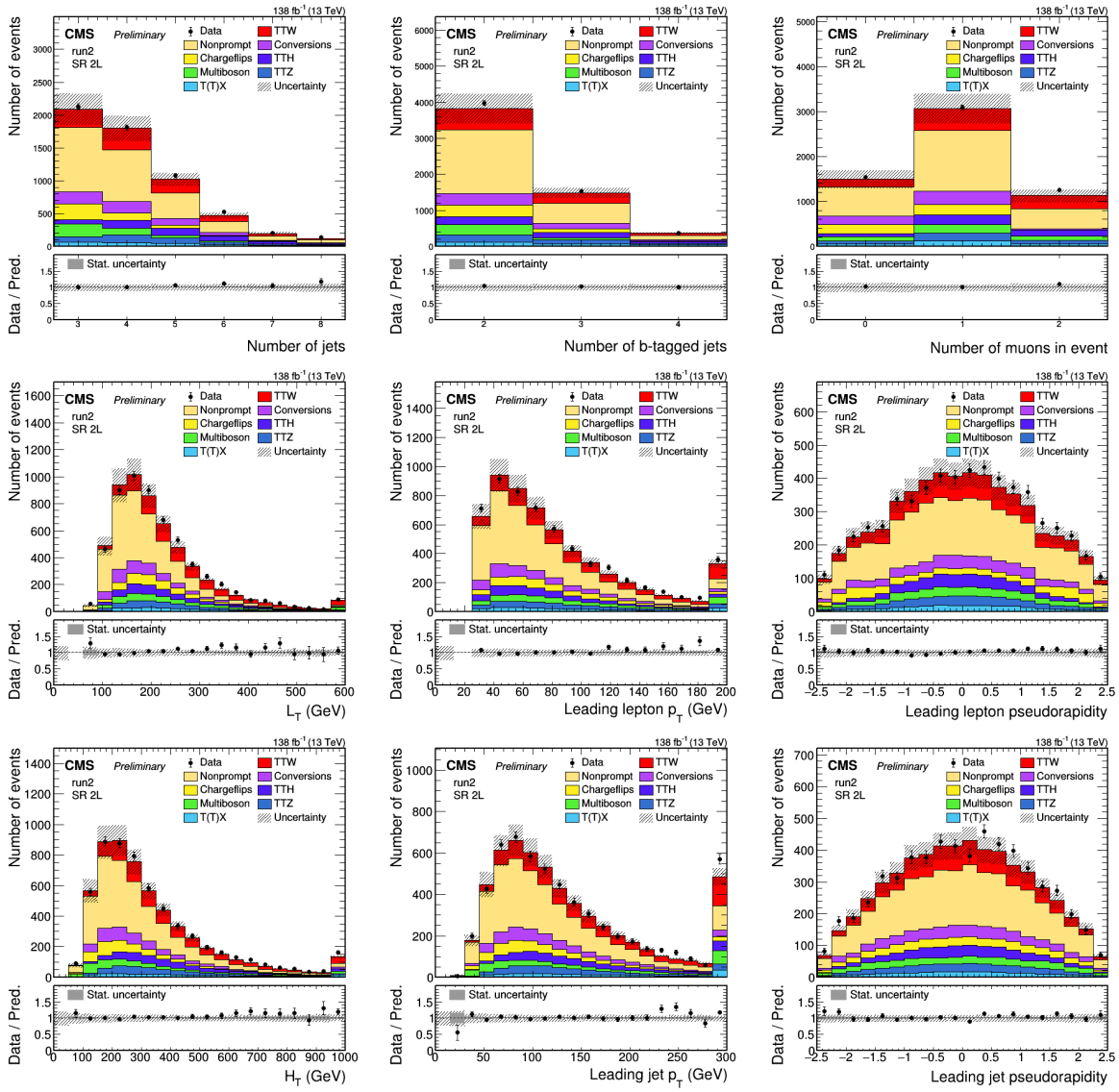


Figure 8.2: Distribution of signal ($t\bar{t}W$, in red) and background processes (other colors) in the same-sign dilepton signal region for several variables of interest.

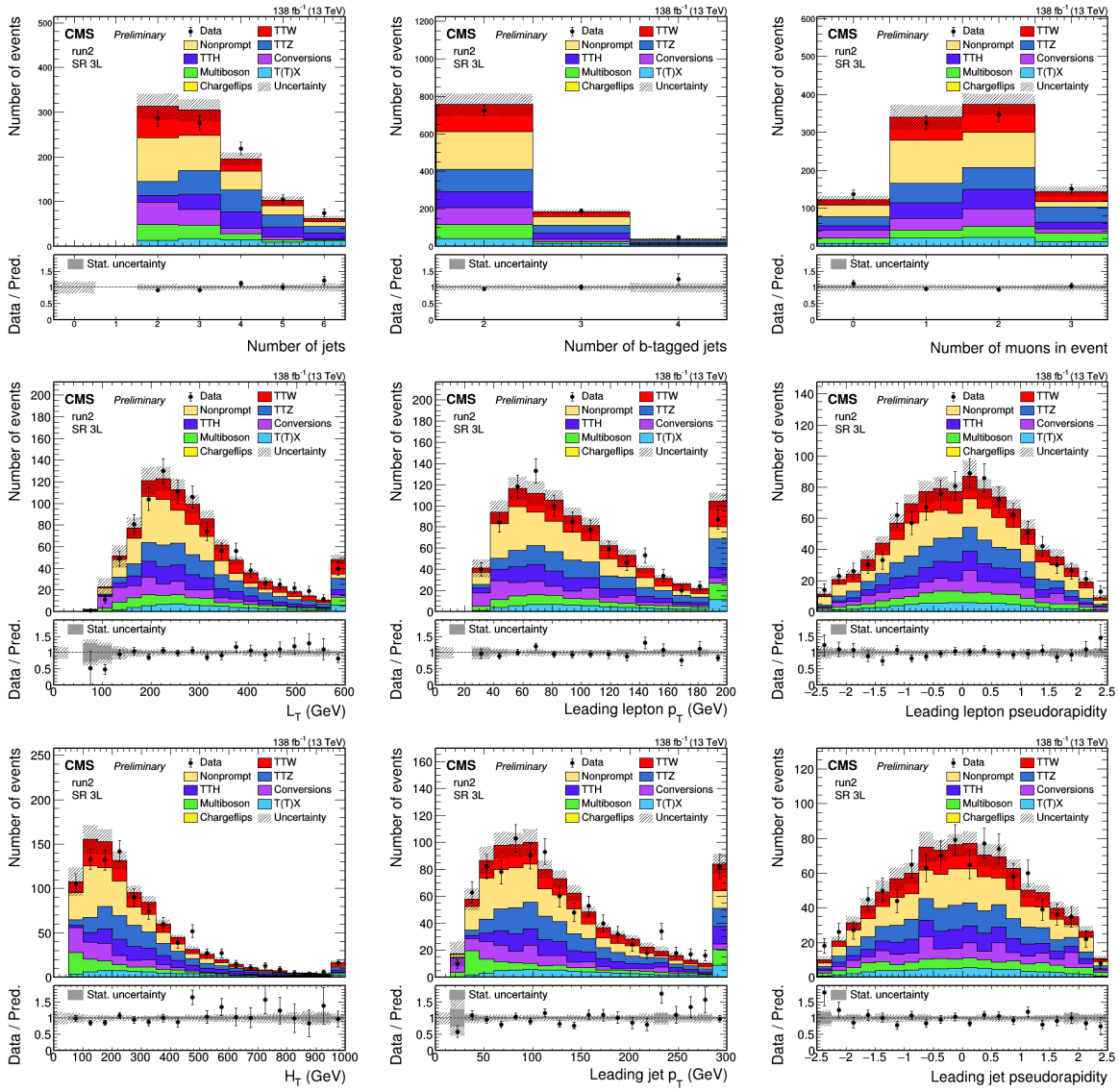


Figure 8.3: Distribution of signal ($t\bar{t}W$, in red) and background processes (other colors) in the trilepton signal region for several variables of interest.

Branching fraction With respect to all possible decay modes of top quarks and W bosons, the branching fraction of $t\bar{t}W$ decaying to two same-charge leptons is about 4.9% and to three leptons about 1.7%[20], with the other 93% being taken up by hadronic, single-lepton or opposite-charge two-lepton decay modes. Similar to the case for tZq, the multilepton channels (in this case specifically the trilepton and same-charge two-lepton channels) are the most sensitive ones, despite their relatively small branching fractions, because of the overwhelming background processes in the channels with two opposite-charge or less than two leptons.

Selection efficiency The selection efficiency for the two-lepton signal region, evaluated on the simulated samples for the $t\bar{t}W$ signal and a number of important backgrounds, is shown in Fig. 8.4. As expected, and similarly to what was observed in the case for tZq, the main selection inefficiencies for $t\bar{t}W$ events are coming from the lepton and (b-tagged) jet requirements, with other steps in the selection all having an efficiency of $> 90\%$. The backgrounds with nonprompt leptons from DY and $t\bar{t}$ are strongly suppressed by the same-charge requirement on the selected leptons. The DY background is further reduced by the required presence of (b-tagged) jets. The $t\bar{t}\gamma$ background, contributing mainly through photon conversions, is effectively suppressed by the tight lepton identification (which contains an explicit conversion veto). The $t\bar{t}H$ and $t\bar{t}Z$ processes are more difficult to separate from the $t\bar{t}W$ process with this collection of simple selections.

The same remark holds as in the case of the tZq analysis: care should be taken when interpreting the efficiency in an absolute sense, as the correct reference point should be taken into account. For the case of $t\bar{t}W$, the reference point in Fig. 8.4 is the full simulated sample, which contains W decays to all lepton flavours (including τ leptons) and fully inclusive top quark decays. With respect to this reference point, the selection efficiency is about 3.3%. When we only take into account same-charge two-lepton events in the reference point by applying the appropriate branching fractions, the efficiency is increased to about 22%. Alternatively, the reference point for the efficiency calculation can be set to the number of events that pass a fiducial event selection at generator level that mimics our signal region definition at detector level, the details of which are provided in section 8.8. With respect to this reference point, the efficiency of our selection at detector level is about 40%, implying that almost half of all events that can reasonably be expected to end up in our two-lepton signal region effectively make it there. The remaining inefficiencies are mainly coming from the lepton and b-tagged jet identification.

BDT discriminator As can be seen in Figs. 8.2 and 8.3, the relative purity of the signal regions is rather poor, and the yield of $t\bar{t}W$ events compared to its backgrounds is small. Hence, similar to the case of the tZq analysis (see chapter 7), we rely on multivariate analysis (MVA) techniques, or more specifically on a boosted decision tree (BDT), to extract the $t\bar{t}W$ signal. The BDT for this analysis was developed by David Marckx (Ghent University), and hence will not be discussed further here.

8.5 Background estimation and control regions

In this section, we discuss the standard model background processes that are relevant in our signal regions and how they are estimated. A first major category of backgrounds consists of events in which all leptons (2 in the dilepton signal region, 3 in the trilepton signal region) are genuine prompt leptons with correctly identified charge. By far the major contributions to this

[20] Calculated from the measured branching fractions of the W boson for leptonic decays, including τ leptons modified by the branching fraction for the decay of τ leptons to electrons or muons.

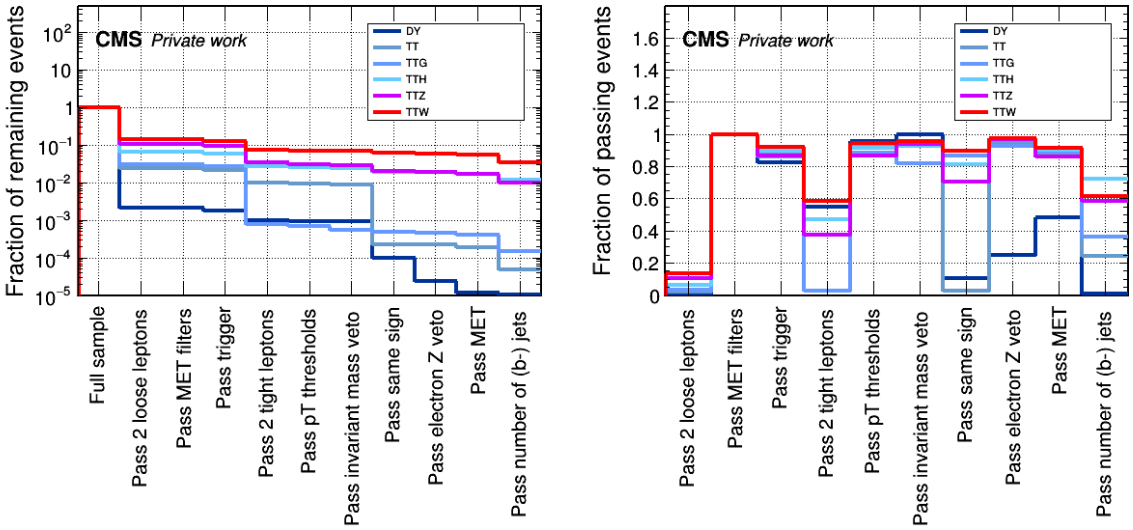


Figure 8.4: Selection efficiency in the two-lepton signal region for $t\bar{t}W$ (red) and a number of important backgrounds (other colors). Left: the fraction of events (with respect to the full simulated sample) remaining after each consecutive step in the event selection. Right: the fraction of events passing each step in the event selection (with respect to the number of events remaining after the previous step).

background category are various ttX processes, notably $t\bar{t}H$, $t\bar{t}Z$ and $t\bar{t}\gamma$. Minor additional contributions are coming from di- or multiboson processes (notably WZ) and tX processes[21]. These backgrounds are estimated directly using the simulated samples. Their modeling is checked in a number of dedicated control regions in data, discussed in section 8.5.1.

A second major category of background events consists of processes where at least one of the leptons is nonprompt, but still passes the tight selection. In our signal region the main contributing process in this category is $t\bar{t}$. This source of background is particularly important to study, as its accurate prediction is difficult. In order to address statistical as well as potential modeling issues, we apply a data-driven technique to estimate the nonprompt background in our signal regions, as detailed in section 8.5.2.

As a third category of backgrounds, we consider events with genuine prompt leptons, of which the electric charge is however wrongly identified. This background category is also estimated (partly) from data, as it is found to be not accurately modeled in simulation. The charge-misidentification estimation method is discussed in section 8.5.3.

8.5.1 Control regions for prompt backgrounds with correctly identified charges

We define two major control region for backgrounds that are taken from simulation, i.e. where all leptons are prompt and have a correctly identified charge. The first one employs a three-lepton selection, and the other one a four-lepton selection.

[21] An overview of all simulated background samples, grouped per category, can be found in Tabs. A.3 and A.4 (in the appendix).

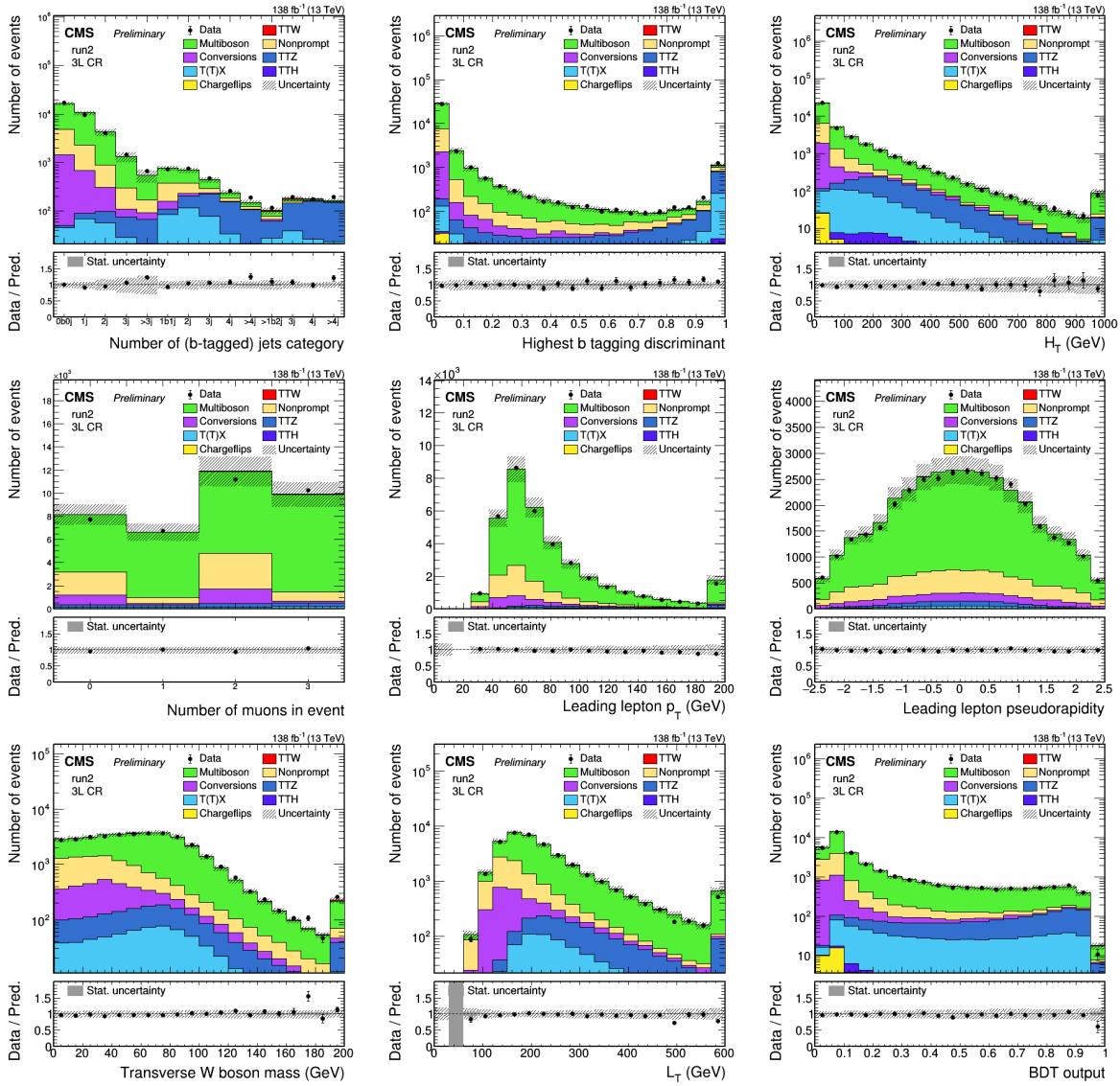


Figure 8.5: Distribution of signal ($t\bar{t}W$, in red) and background processes (other colors) in the 3-lepton control region for several variables of interest.

Three-lepton control region This control region targets mainly the WZ and $t\bar{t}Z$ backgrounds. The event selection criteria are as follows:

- Exactly 3 tight leptons with p_T thresholds of 25, 15 and 15 GeV for the leading, subleading and trailing lepton respectively.
- There has to be an OSSF pair with $|m(\ell\ell) - m(Z)| < 10\,\text{GeV}$.

Depending on the number of jets and b-tagged jets, this region is dominated by either WZ (at low multiplicities) or $t\bar{t}Z$ (at high multiplicities). Distributions in this control region compared to data are shown in Fig. 8.5.

Four-lepton control region This control region targets mainly the ZZ and $t\bar{t}Z$ backgrounds. The event selection criteria are as follows:

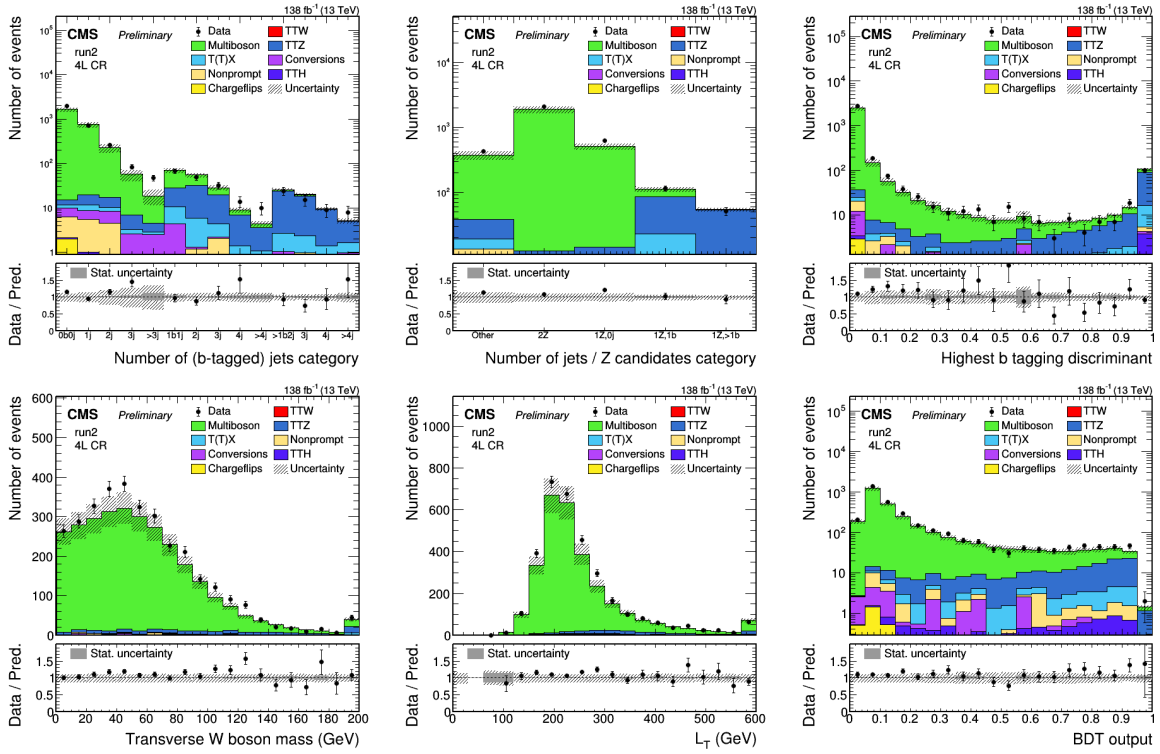


Figure 8.6: Distribution of signal ($t\bar{t}W$, in red) and background processes (other colors) in the 4-lepton control region for several variables of interest.

- Exactly 4 tight leptons with p_T thresholds of 25, 15, 15 and 10 GeV for the p_T-ordered lepton collection.
- There has to be an OSSF pair with $|m(\ell\ell) - m(Z)| < 10\,\text{GeV}$.

Depending on the number of jets and reconstructed Z boson candidates, this region is dominated by either ZZ or $t\bar{t}Z$. Distributions in this control region compared to data are shown in Fig. 8.6.

8.5.2 Nonprompt background

A considerable source of background consists of events with two or three tight leptons of which at least one is nonprompt. Although the lepton MVA and corresponding identification (see section 8.3) were designed to target prompt leptons and reject nonprompt leptons, the amount of events containing one or more misidentified leptons cannot be ignored, mainly because of the relatively large cross-sections of these processes (particularly $t\bar{t}$ and DY) with respect to genuine $t\bar{t}W$ and other prompt processes. The nonprompt background is estimated using the 'fake rate' method, as detailed in section 6.2 Here, we only show the results of the method in a dedicated control region specific to this analysis.

Control region for nonprompt leptons Events are selected into the nonprompt control region if they pass conditions that are equal to the dilepton signal region, with the exception that the E_T^{miss} cut is reversed, i.e. events with $E_T^{miss} < 30\,\text{GeV}$ are selected rather than rejected. The closure of the nonprompt estimation method in this control region is shown in Fig. 8.7, with good agreement between data and prediction observed as a function of relevant variables such as lepton flavour composition and number of (b-tagged) jets. As an additional

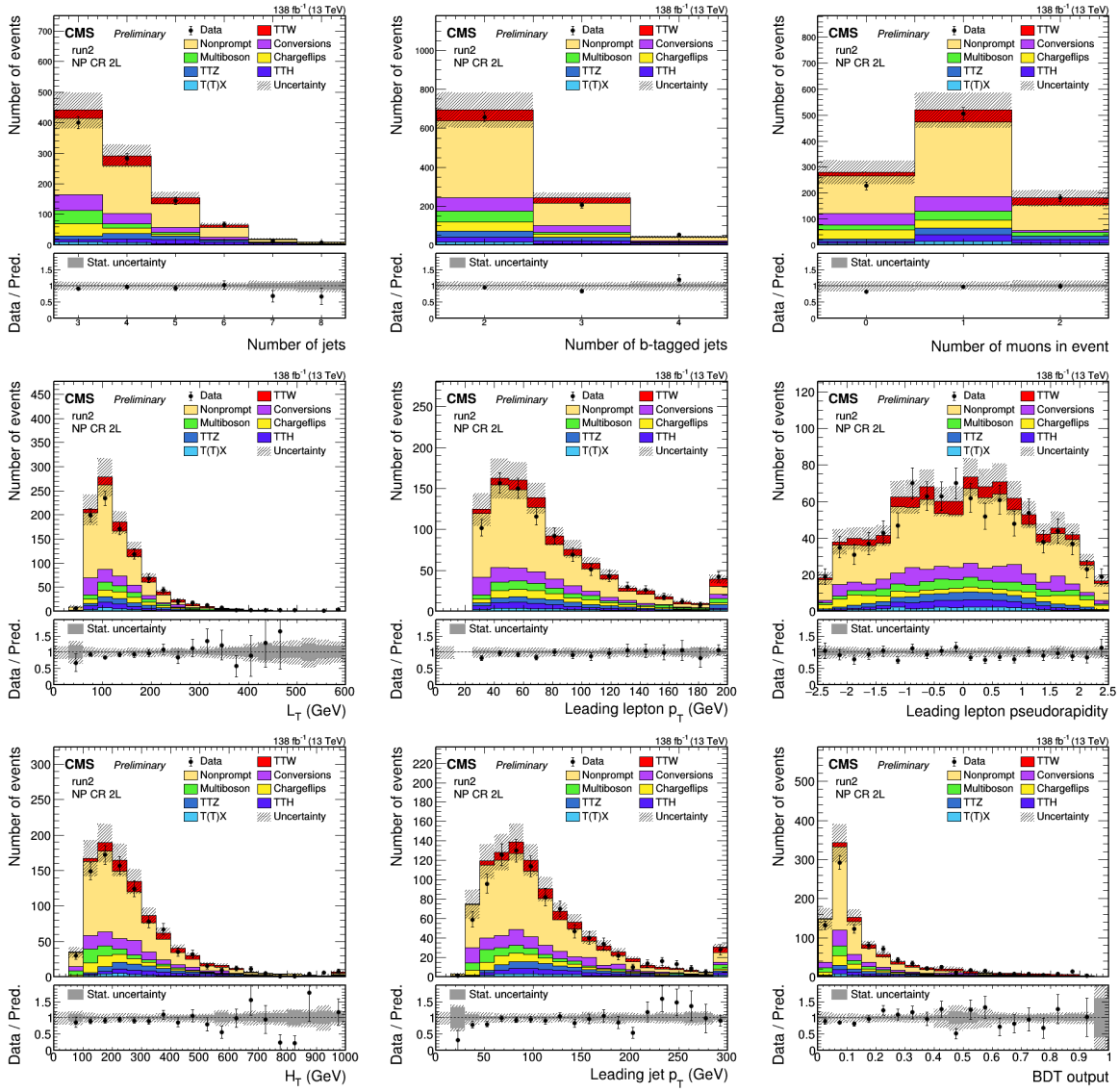


Figure 8.7: Distribution of signal ($t\bar{t}W$, in red) and background processes (other colors) in the nonprompt control region for several variables of interest.

check, we show the agreement between data and prediction in the nonprompt control region split between lepton flavour channels in Fig. 8.8.

Control region for nonprompt leptons with low number of jets The previous control region for nonprompt leptons is orthogonal to the dilepton signal region because of the inversion of the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut. In addition to this control region, we define another one where instead of the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut, the number of jets and b-tagged jets are used to make the region orthogonal to the signal region. Events are selected into this nonprompt control region if they pass conditions that are equal to the dilepton signal region, except that the number of jets must be either 1 or 2 (instead of at least 3), and the number of b-tagged jets must be at least 1 (instead of 2). The closure of the nonprompt estimation method in this control region is shown in Fig. 8.9.

Discussion We observe good agreement in the simulated closure tests, as well as in data control regions that are similar (but not overlapping) to the signal region (see Figs. 8.7–8.9).

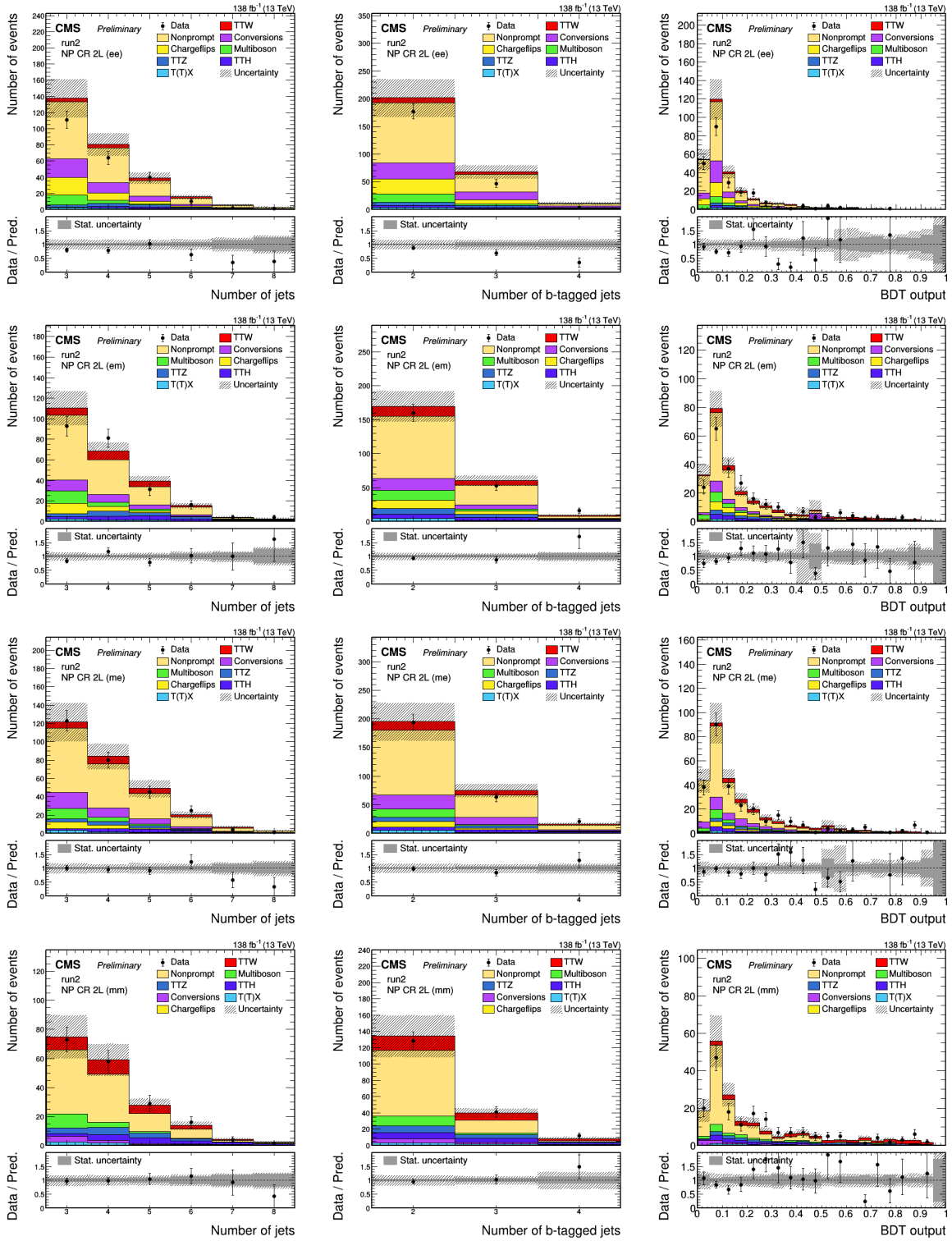


Figure 8.8: Distribution of signal ($t\bar{t}W$, in red) and background processes (other colors) in the nonprompt control region for several variables of interest and split between lepton flavour channels. From top to bottom: ee, em, me and mm channel.

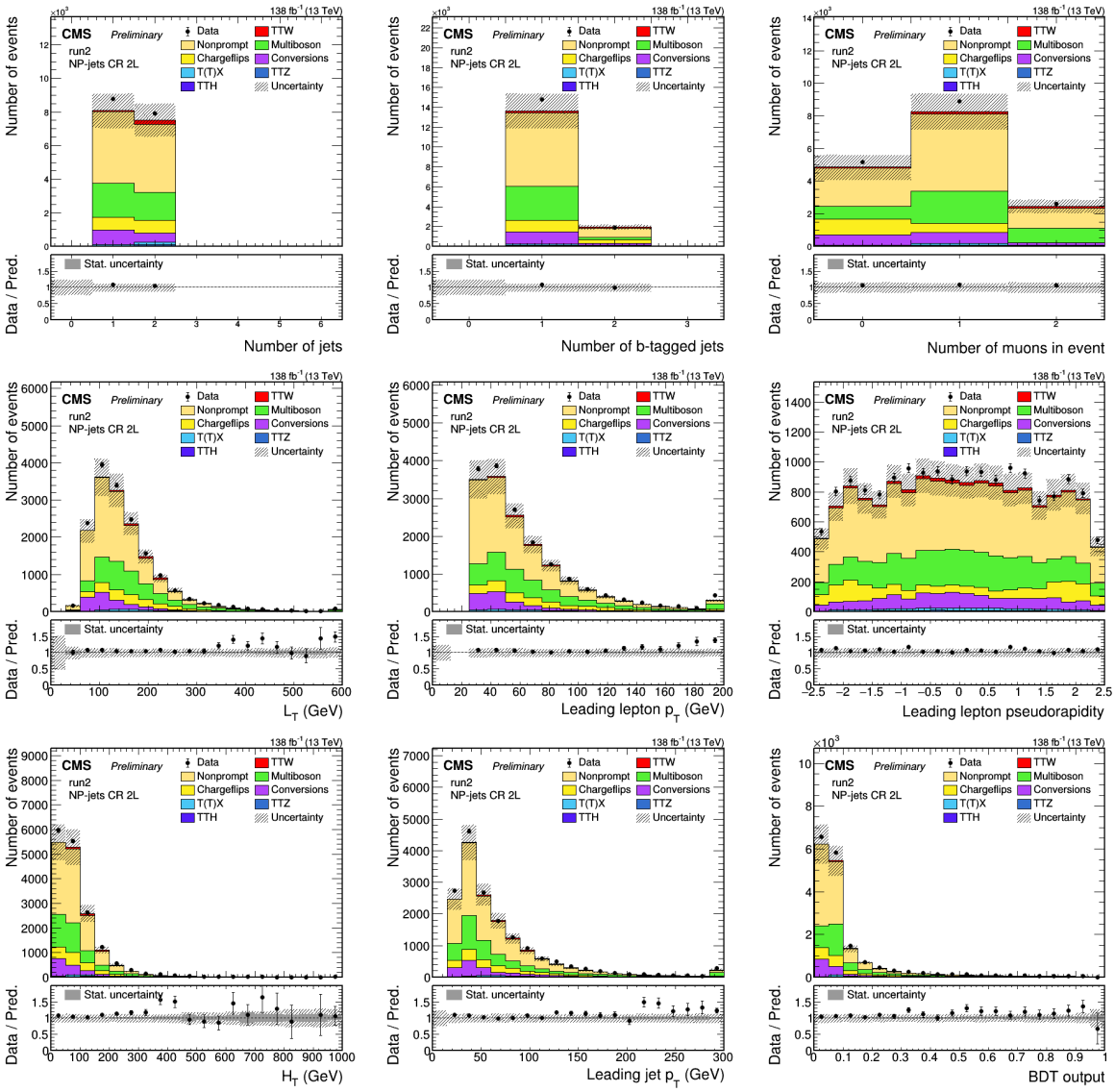


Figure 8.9: Distribution of signal ($t\bar{t}W$, in red) and background processes (other colors) in the nonprompt control region (with lower number of jets) for several variables of interest.

For the uncertainty on the fakerate prediction, we use a more involved procedure in comparison to the tZq analysis (see section 7.5.1), as the nonprompt lepton background is dominant in the $t\bar{t}W$ analysis. We apply a 20% flat normalization uncertainty correlated between nonprompt electrons and muons, and another 20% on nonprompt electrons and muons separately (and uncorrelated between both). In total, this approach corresponds approximately to the single 30% used in the tZq analysis, but it provides more flexibility to account for differences in nonprompt electrons and muons. Furthermore, we provide six varied shapes to the statistical model, obtained by varying the fake rates within their respective uncertainties, either all up or down (correlated for all p_{T} and $|\eta|$ bins), or linearly morphed from up to down (and the other way round) as a function of p_{T}, and as a function of $|\eta|$.

8.5.3 Charge-misidentification background

Another considerable source of background in this analysis consists of leptons with wrongly assigned electric charge. Although the probability for this to happen is small, the very large cross-section of (mainly) the Drell-Yan process provides a large source of opposite-sign dilepton events, a non-negligible fraction of which is wrongly identified as same-sign dilepton events and make it into the signal region. The probability for wrongly identifying the electric charge of a lepton depends strongly on the detector conditions, and is found to be not well modeled in simulation. Hence, a (partly) data-driven technique is used to estimate this background, as described in section 6.3. Here, we only show the results of the procedure in control regions in data.

Control region for charge misidentification Finally, the full procedure (including the application of the measured scale factors) can be applied in the analysis framework in a dedicated control region for charge misidentification. This control region requires exactly 2 same-sign electrons with $|m(\ell\ell) - m(\mathrm{Z})| < 10\,\mathrm{GeV}$ (where the charge-misidentified contribution is estimated from the opposite-sign counterpart). The results are shown in Fig. 8.10.

Control region for charge misidentification with jets In addition to the control region detailed and displayed in the previous paragraph, we define another region for checking the prediction of charge misidentification. Where the previous control region was inclusive in number of jets and b-tagged jets, this additional region has the requirements of at least 3 jets and at least 1 b-tagged jet (with all other selections remaining the same), bringing it closer to the phase space of the signal region in this analysis. The results are shown in Fig. 8.11.

8.6 Systematic uncertainties

The common sources of systematic uncertainty affecting this measurement have already been discussed in section 6.5. Hence, this discussion is limited to filling in some details and specifications.

Background normalization. We apply an a-priori uncertainty of 30% for the $\mathrm{Z}\gamma$ and $t\bar{t}\gamma$ processes and freely floating normalizations for the WZ, ZZ and $t\bar{t}Z$ processes (that can be constrained by including the control regions in the fit). The $t\bar{t}H$ background takes a normalization uncertainties of 10%, corresponding to the theoretical uncertainty coming from the PDF and renormalization/factorization scale variations. Other processes in the 'TX', 'TTX' (mainly $t\bar{t}VV$) and 'Multiboson' categories receive a 50% normalization uncertainty. They are however of minor importance in the signal regions. The uncertainties on the nonprompt lepton background have been discussed already in section 8.5.2.

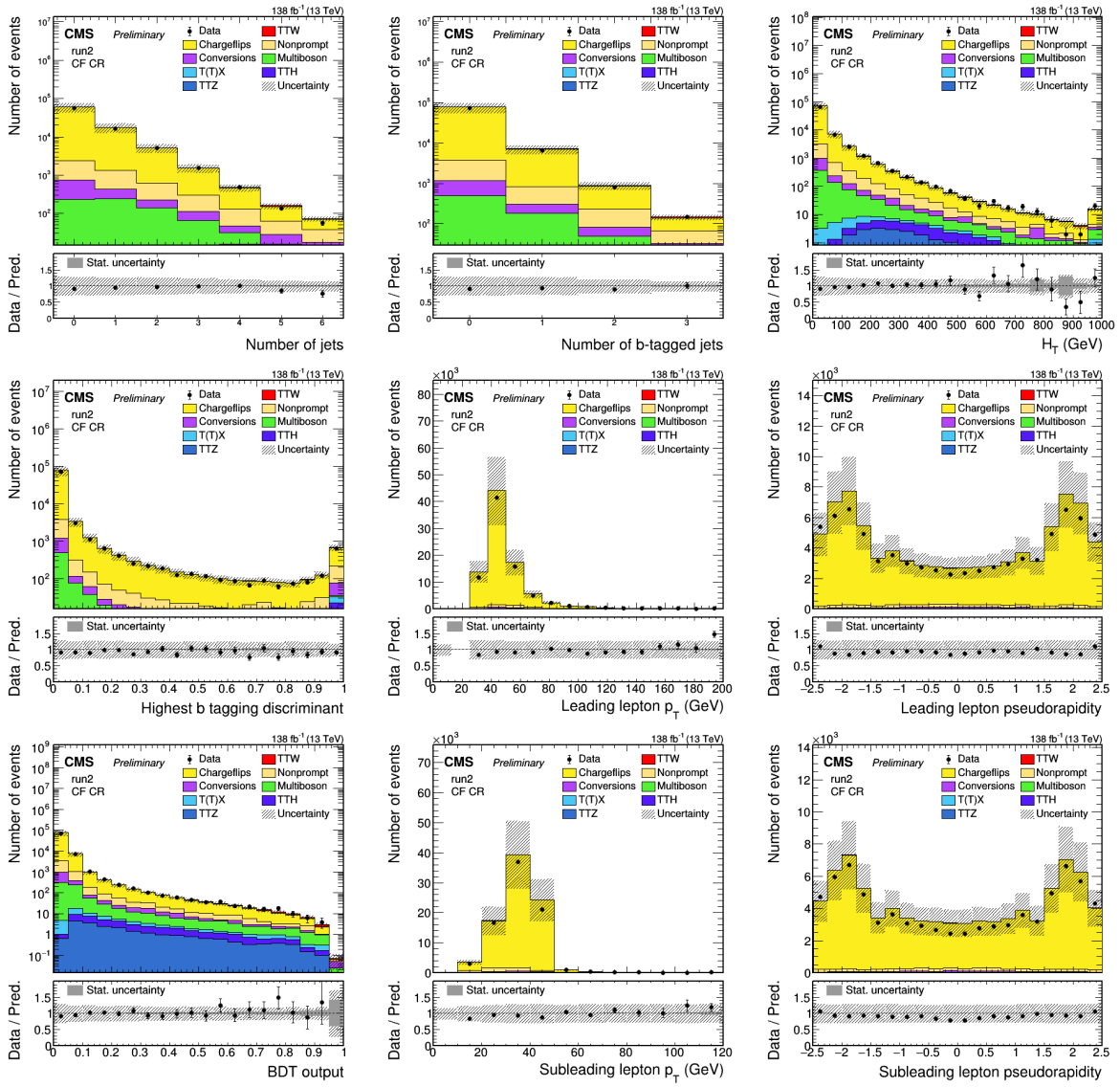


Figure 8.10: Distribution of signal ($t\bar{t}W$, in red) and background processes (other colors) in the charge-misidentification control region for several variables of interest.

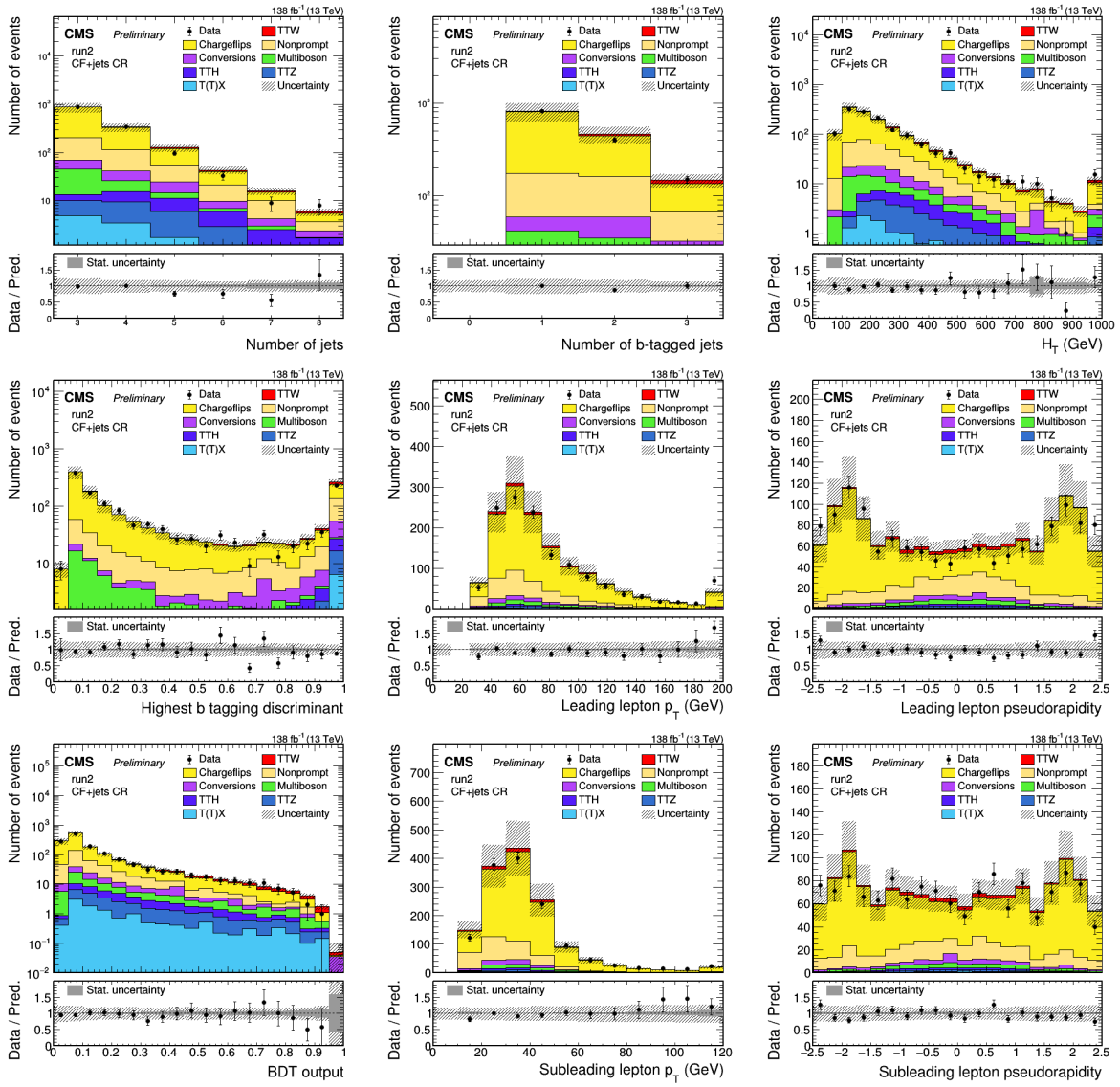


Figure 8.11: Distribution of signal ($t\bar{t}W$, in red) and background processes (other colors) in the charge-misidentification + jets control region for several variables of interest.

8.7 Inclusive measurements

In this section, we perform some inclusive cross-section measurements of the $t\bar{t}W$ process. These are mainly intended as cross-check with respect to [221] and [214] rather than original or improved results. All inclusive measurements were performed with a binned maximum likelihood fit of the prediction to the data, with sources of systematic uncertainty taken into account as profiled nuisance parameters. For all fits, the COMBINE tool [218–220] was used.

Note: all results quoted in this section are still blinded, i.e. they are obtained using an Asimov dataset that corresponds to the prediction rather than with actual observed data.

Inclusive cross-section As a first result, we discuss the fully inclusive $t\bar{t}W$ cross-section measurement. The measurement region consists of a dilepton and a trilepton event selection, as discussed in more detail in section 8.4. The distributions included in the fit are displayed in Fig. 8.15. The $t\bar{t}W$ sample is normalized (before the fit) according to the reference cross-section value discussed in section 8.2. With respect to this reference, we measure a signal strength of

$$\mu = \frac{\sigma_{t\bar{t}W}^{\text{observed}}}{\sigma_{t\bar{t}W}^{\text{expected}}} = 1 \pm 0.06 \text{ (stat.)} \pm 0.07 \text{ (syst.)} \tag{8.1}$$

The central value of the measurement is 1 by construction because of the usage of an Asimov dataset; the uncertainties are however in agreement with Ref. [214], which serves as a first cross-check for this approach.

Impact of systematic uncertainties We investigate the pulls and impacts of the most important sources of systematic uncertainty in the measurement above using the COMBINE tool. The fully inclusive results are shown in Fig. 8.12. The largest systematic uncertainty is coming from the background normalization effect of PDF variations. Other important but subdominant theoretical uncertainties include the shape effect of FSR variations and the normalization of some of the simulated backgrounds that are less well separated from the signal by the BDT, such as multiboson, $t\bar{t}\gamma$ and $t\bar{t}H$ and $t\bar{t}Z$ processes. Experimental uncertainties are mainly coming from luminosity and b-tagging uncertainties.

Split per data-taking year We investigate the stability of the fit and the self-consistency of the measurement by fitting each data-taking year separately. The input distributions per year are the same as for the fully inclusive fit, displayed in Fig. 8.13. Expected and observed results are summarized in Fig. 8.14.

Split per lepton flavour channel As an additional check, we also measure the $t\bar{t}W$ cross-section separately per lepton flavour channel, i.e. separately for the ee, em, me and mm final state. The fitted distributions are displayed in Fig. 8.15. To keep the total number of bins in the fit limited, only aggregated Run-II distributions are included, rather than the per-year distributions as before. The expected and observed results are summarized in Fig. 8.16.

Split per charge channel As a final inclusive measurement, we measure the cross-section for $t\bar{t}W^+$ and $t\bar{t}W^-$ separately (but simultaneously). To capture the correlation between both cross-sections in the measurement, we perform a likelihood scan over a two-dimensional grid in the signal strengths for $t\bar{t}W^+$ and $t\bar{t}W^-$, as shown in Fig. 8.17. The 68% and 95% confidence level (CL) contours are drawn where $-2\Delta\log(L)$ reaches 2.30 and 5.99 respectively [22].

[22] These values are obtained using Wilks' theorem, stating that the negative log likelihood ratio test statistic

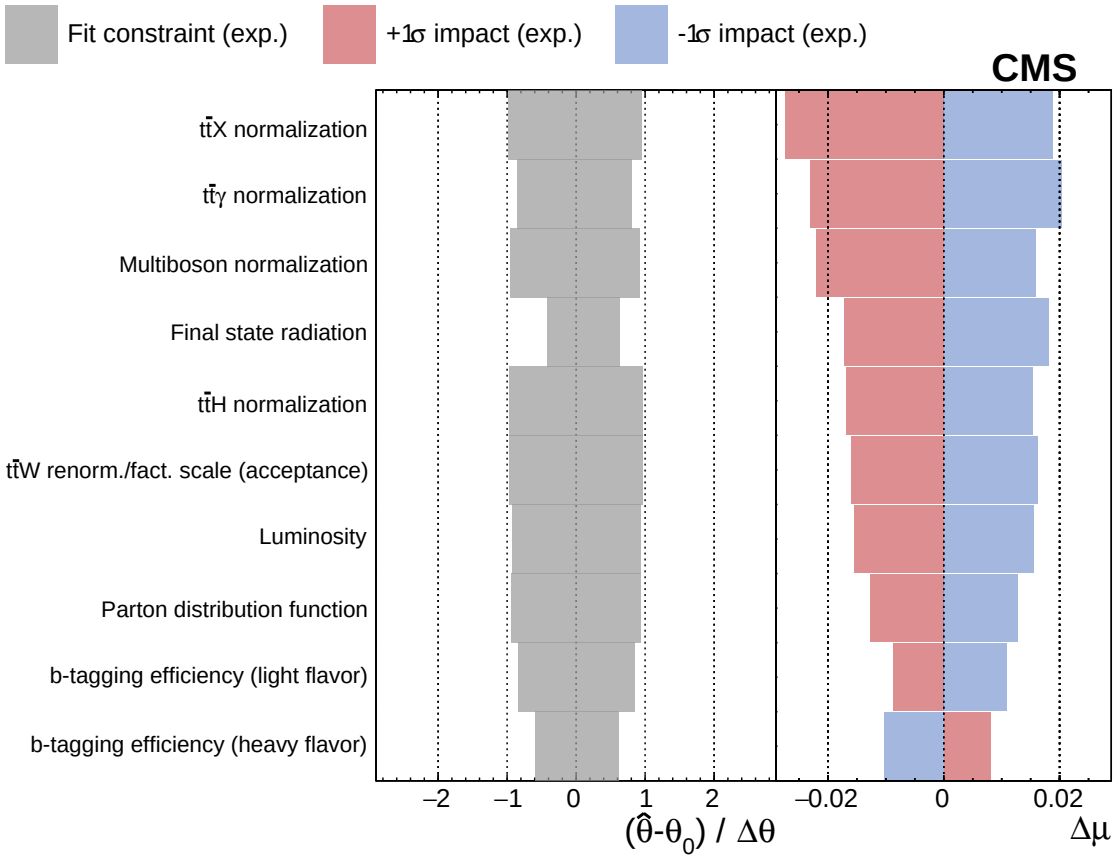


Figure 8.12: Pulls, constraints and impacts of most the important sources of systematic uncertainty for the inclusive $t\bar{t}W$ cross-section measurement. The left column lists the sources of systematic uncertainty, treated as nuisance parameters in the fit, in order of importance. In the middle column, the gray bands show for each source the difference between the expected best fit value ($\hat{\theta}$) and the nominal value (θ_0), divided by the pre-fit standard deviation ($\Delta\theta$). The right column plots the expected change in the $t\bar{t}W$ signal strength μ if a nuisance parameter is varied one standard deviation up (red), or down (blue).

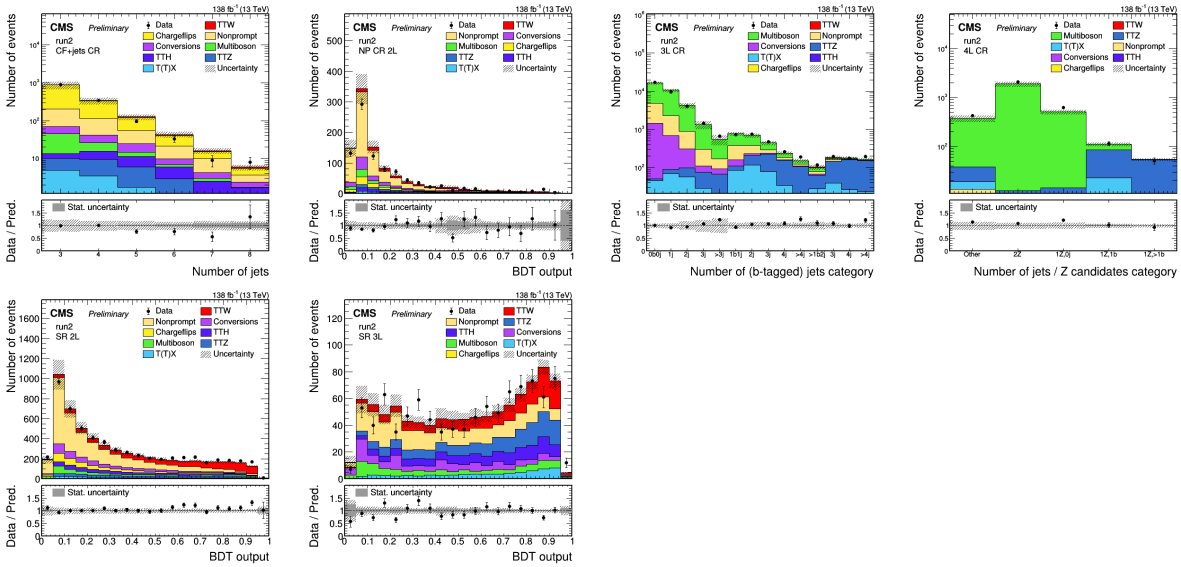


Figure 8.13: Distributions included in the inclusive cross-section measurement. For clarity of display, only Run-II aggregated plots are shown, but the actual fit is performed using the same distributions split per data-taking year. Upper row: control regions. Lower row: signal regions.

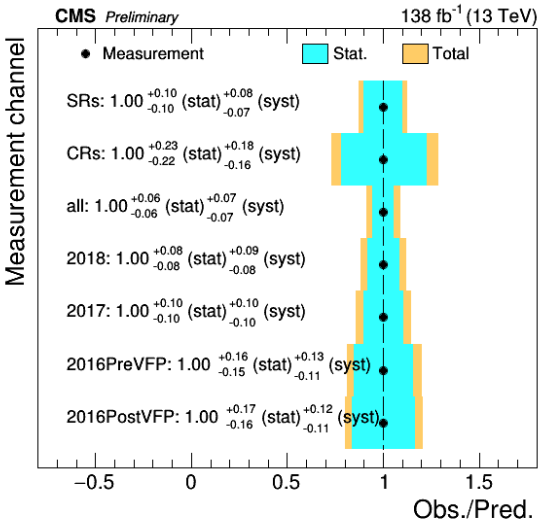


Figure 8.14: Expected signal strengths with statistical and systematic uncertainties for the measurement of the inclusive $\mathrm{t\bar{t}W}$ cross-section measurement split per data-taking year. The first row (SRs) shows the measurement in the signal regions, excluding the control regions from the fit. The second row (CRs) shows the measurement using control regions only.

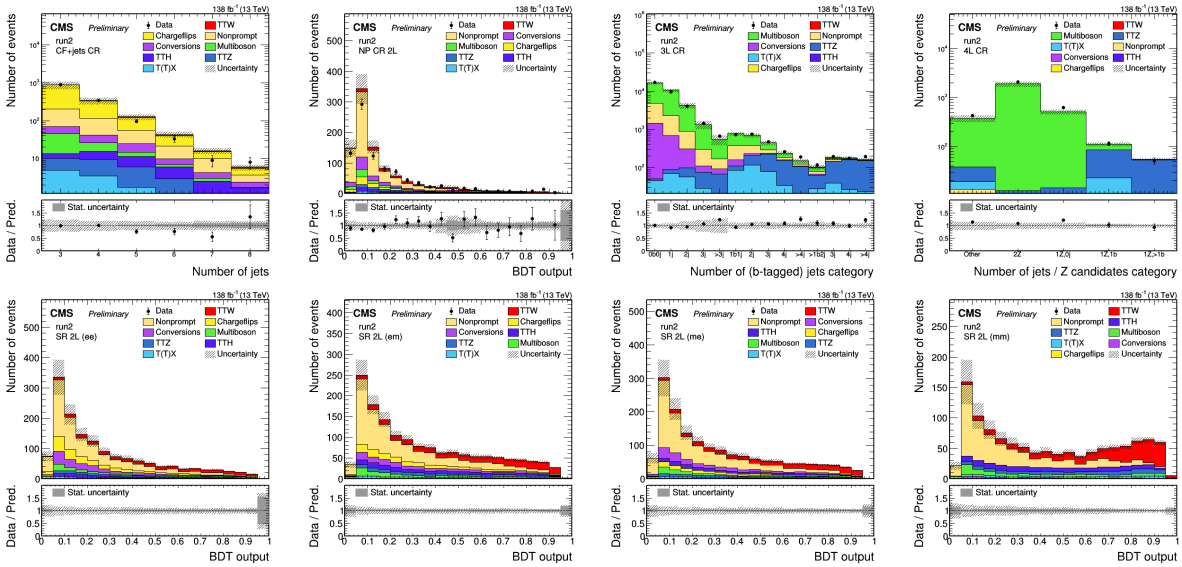


Figure 8.15: Distributions included in the inclusive cross-section measurement per lepton flavour channel. Upper row: control regions. Lower row: signal regions, split per lepton flavour channel.

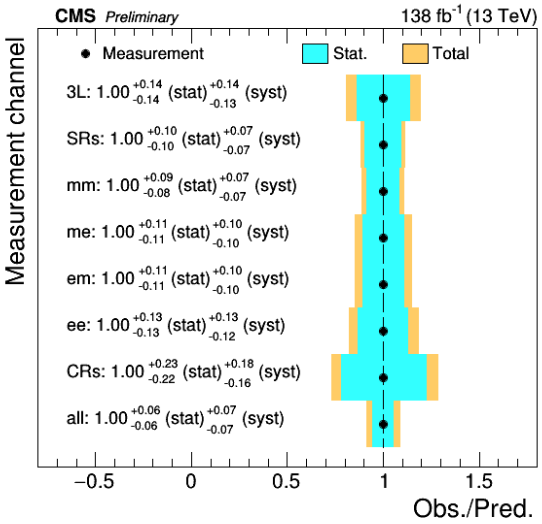


Figure 8.16: Expected signal strengths with statistical and systematic uncertainties for the measurement of the inclusive $\mathrm{t\bar{t}W}$ cross-section measurement split per lepton channel. The first row (3L) shows the measurement in the trilepton signal region (not split per lepton channel). The second row (SRs) shows the measurement in the signal regions, excluding the control regions from the fit. The seventh row (CRs) shows the measurement using control regions only.

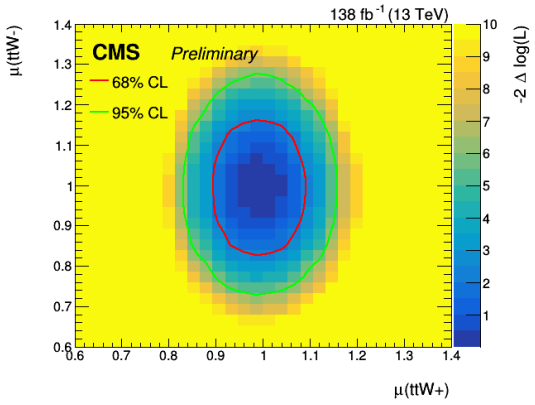


Figure 8.17: Expected likelihood scan for $t\bar{t}W^+$ and $t\bar{t}W^-$, where the signal strengths for $t\bar{t}W^+$ and $t\bar{t}W^-$ are measured separately but simultaneously. The color scale shows the quantity $-2\Delta\log(L)$ with respect to the best-fit value, and the 68% and 95% confidence level (CL) contours are superimposed as red and green lines respectively.

Consistency Next, we perform a consistency check by measuring the inclusive $t\bar{t}W$ cross-section using the setup and input distributions for the differential measurement (as described in section 8.8). Figure 8.18 shows consistent results, highlighting the accurate modeling of the BDT and of the variables of interest.

Goodness of fit To further assess the robustness of the fit and the ability of the model probability density function to accommodate the observed data, we perform goodness-of-fit tests following the CMS statistics committee recommendations[23]. The goodness-of-fit tests are performed separately for the individual fitted distributions (i.e. per signal/control region and/or per data-taking year) and the results are summarized in Fig. 8.19.

The poor modeling in the trilepton control region has been traced down to the zero-jet and b-tagged-jet multiplicity part of the phase space, which is not relevant for this measurement. Excluding this part of the phase space from this control region results in a good fit quality and is adopted as a solution.

8.8 Unfolding and differential measurements

In this section, we discuss results of the differential $t\bar{t}W$ cross-section measurements, that are the main result of this chapter. The terminology and methodology were already introduced in section 6.7, so the discussion here is limited to filling in some details and describing the results.

Fiducial selection The events must have two same-charge (light) leptons, which pass p_T thresholds of 25 and 15 GeV respectively. Furthermore, events must have at least 3 jets, of which at least 1 b jet[24]. The efficiency of this particle level selection is displayed in Fig. 8.20.

is asymptotically distributed as a χ^2 function. The number of degrees of freedom is equal to the dimension of the scan.

[23] See the COMBINE tool website.

[24] This section (as well as the whole $t\bar{t}W$ analysis) focuses on the same-sign dilepton selection; however, we repeated the method as well employing a trilepton fiducial selection rather than a same-sign dilepton one. The statistical precision in this region is however much lower than in the dilepton region. Therefore, the remainder of this section will only deal with the same-sign dilepton selection.

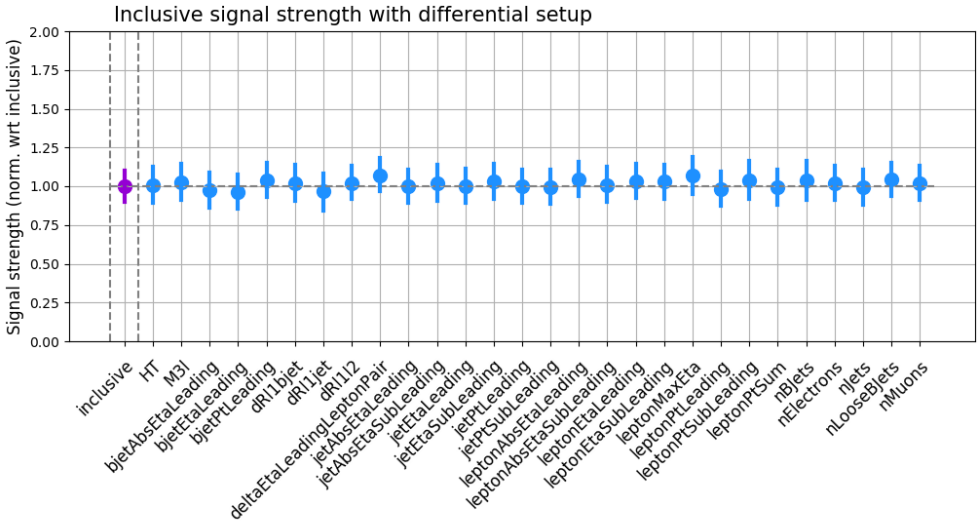


Figure 8.18: Inclusive $t\bar{t}W$ cross-section as measured with the setup for the inclusive measurement (leftmost marker, in purple) and as measured with the setup for the differential measurements, for each of the variables of interest (other markers, in blue). These measurements are unblinded, but normalized to a signal strength of 1 for the inclusive setup as to remain unbiased with respect to the actual cross-section. All measurements yield a consistent inclusive $t\bar{t}W$ cross-section.

Note that this efficiency is calculated with respect to the full TTWJetsToLNu sample, i.e. including also τ decays of the W boson and fully inclusive decays of the top quarks.

Variables The choice of variables that we unfold to is inspired by [247] (being sensitive to specific $t\bar{t}W$ modeling effects) and [250] (being useful for understanding $t\bar{t}W$ as a background to other processes). They are listed in Tab. 8.3.

Response matrices and choice of binning The response matrices are calculated for all variables as a function of which the differential cross-section is measured. Because of the large number of variables, manually fine-tuning the binning for all of them is not trivial. An attempt is made to tune the binning automatically, based on the following metric:

$$\text{metric} = \text{average}\left(\text{min. stability}, \text{min. purity}, \frac{\text{min. yield in diagonal bins}}{\text{max. yield in diagonal bins}}\right) \tag{8.2}$$

Starting from an initial guess (by manually defining a 'reasonable' binning), the bin edges are iteratively moved in the direction of smaller metric, until convergence is reached. The result of this procedure is shown in Fig. 8.21. While the effect of this optimization procedure on the stability and purity is not very large, the yield ratio in diagonal bins is significantly improved. This results in the available events being more evenly spread over the bins in the differential measurement, reducing overall statistical uncertainties. A final binning, used in the remainder of this measurement, is defined by rounding the optimal values to the nearest reasonable values.

While this whole procedure is somewhat ad-hoc, the resulting binning is satisfactory, being a good compromise between stability, purity, statistical uncertainties and human interpretability. Note that variables with a pre-defined binning, such as integer object counters (e.g. number

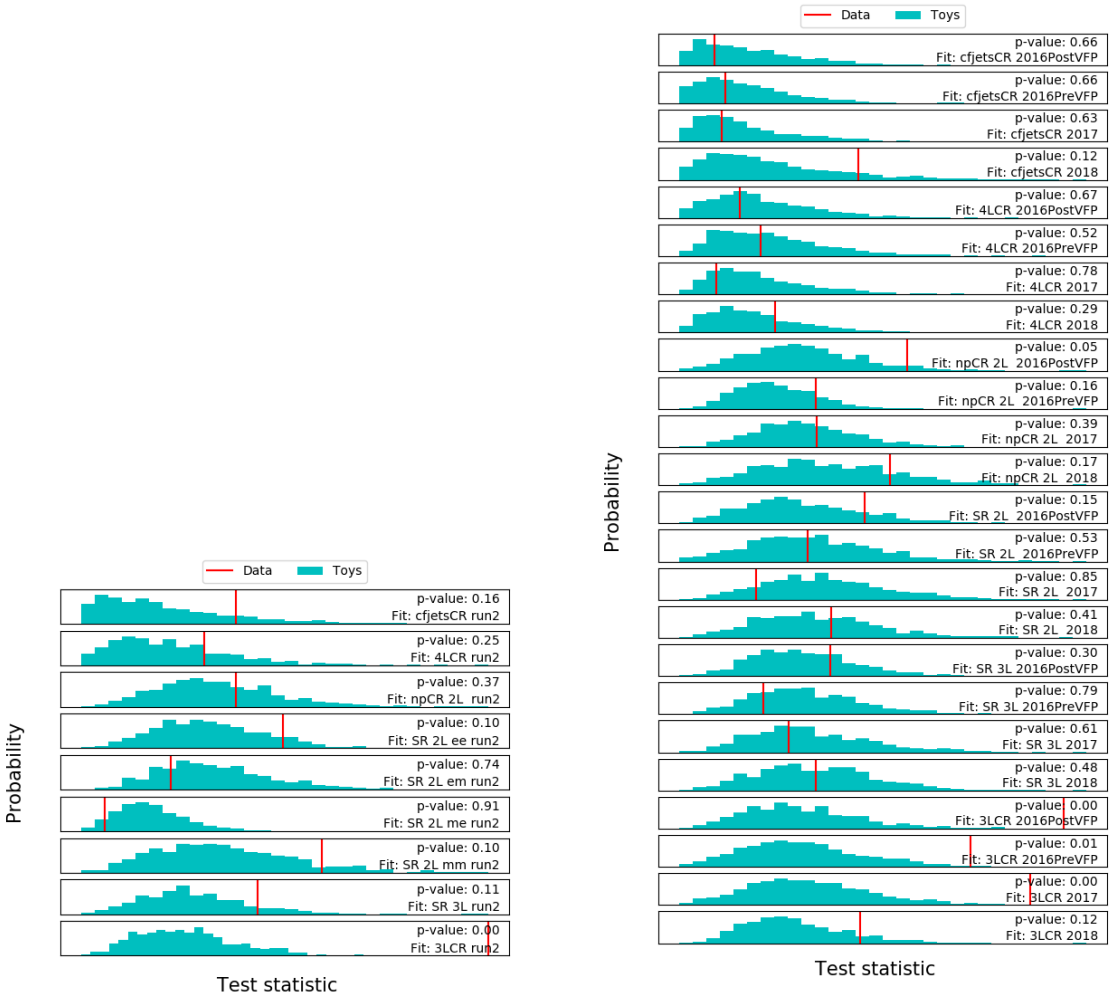


Figure 8.19: Goodness-of fit tests for the regions included in the per-channel fit (left) and the per-year fit (right).

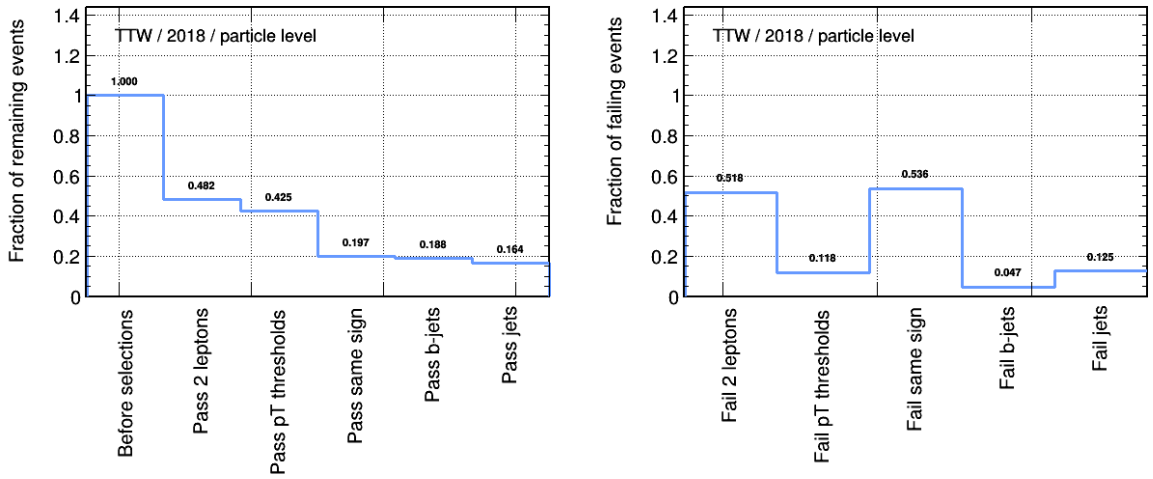


Figure 8.20: Fraction of remaining events (left) and fraction of failing events (right) for each step of the fiducial selection at particle level.

Table 8.3: Variables used in the differential measurement

Variable	Description
b-jet kinematics	
leading b-jet η	pseudorapidity of b-jet with highest transverse momentum
leading b-jet $\|\eta\|$	absolute pseudorapidity of b-jet with highest transverse momentum
leading b-jet p_{T}	transverse momentum of b-jet with highest transverse momentum
Angles between objects	
$\Delta\eta(l_1, l_2)$	absolute difference in pseudorapidity between two leading leptons
$\Delta R(l_1, l_2)$	angular separation between two leading leptons
$\Delta R(l_1, \mathrm{b-jet})$	minimum angular separation between leading lepton and any b-jet
$\Delta R(l_1, \mathrm{jet})$	minimum angular separation between leading lepton and any jet
Jet kinematics	
h_T	scalar sum of jet transverse momenta
leading jet η	pseudorapidity of jet with highest transverse momentum
leading jet $\|\eta\|$	absolute pseudorapidity of jet with highest transverse momentum
leading jet p_{T}	transverse momentum of jet with highest transverse momentum
subleading jet η	pseudorapidity of jet with next-to-highest transverse momentum
subleading jet $\|\eta\|$	absolute pseudorapidity of jet with next-to-highest transverse momentum
subleading jet p_{T}	transverse momentum of jet with next-to-highest transverse momentum
Lepton kinematics	
leading lepton η	pseudorapidity of lepton with highest transverse momentum
leading lepton $\|\eta\|$	absolute pseudorapidity of lepton with highest transverse momentum
leading lepton p_{T}	transverse momentum of lepton with highest transverse momentum
subleading lepton η	pseudorapidity of lepton with next-to-highest transverse momentum
subleading lepton $\|\eta\|$	absolute pseudorapidity of lepton with next-to-highest transverse momentum
subleading lepton p_{T}	transverse momentum of lepton with next-to-highest transverse momentum
lepton maximum $\|\eta\|$	maximum absolute pseudorapidity among leptons
$\sum$ lepton p_{T}	sum of lepton transverse momenta
m(leptons)	invariant mass of leptons
Object counters	
number of b-jets	number of b-jets
number of jets	number of jets
number of electrons	number of electrons
number of muons	number of muons

of jets) are excluded from this optimization.

Results and discussion The results are shown in Figs. 8.22–8.27. As the analysis is still blinded for now, the measurements correspond to the prediction by construction. However, the theoretical and experimental uncertainties show that these differential cross-sections will be measured with about 15-30% uncertainty for most bins, depending on the variable.

Normalized differential cross-sections The normalized differential cross-section measurements are shown in Figs. 8.28–8.33. The advantage of using normalized cross-sections is twofold. On the one hand, shape differences between measurement and prediction are more clearly highlighted as any potential overall normalization effect is canceled. On the other hand, systematic uncertainties that affect mostly the overall normalization rather than the shape are largely canceled in this procedure, resulting in overall smaller uncertainties. This effect is limited here however, since the measurements are dominated by statistical uncertainties which do not cancel out in this way.

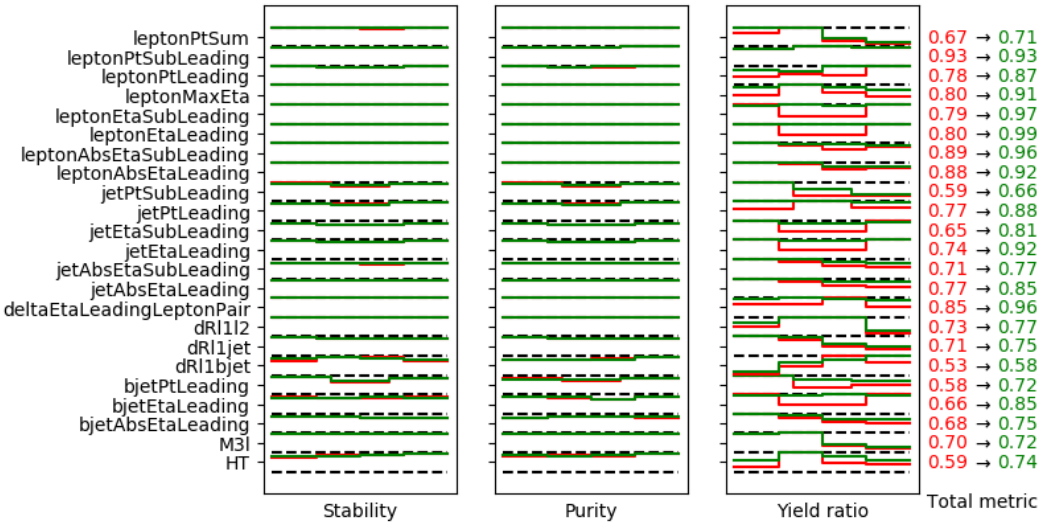


Figure 8.21: Result of automatic binning optimization

8.9 Conclusions and outlook

This chapter presents a snapshot of a project that is not yet finished. In particular, while the methodology is fully ready and tested on an Asimov dataset, the measurements are still blinded and in CMS internal review. Once the appropriate checks have been passed (in the near future), the measurements described in this chapter will be performed using observed data rather than an Asimov dataset that corresponds to the prediction by construction.

Based on earlier measurements of the $t\bar{t}W$ process, we expect to measure an overall inclusive cross-section that is higher than the predicted one (see e.g. Ref. [214]). In addition to the normalization effect, the differential cross-section measurements could highlight specific regions of the phase space where the discrepancies between prediction and measurement are largest. This provides feedback to theorists on the origins of these discrepancies and on where improvements in the modeling are likely to have the most effect. However, similar measurements by ATLAS [180] were somewhat inconclusive on this aspect, and displayed a rather good agreement to the prediction apart from a normalization effect. It remains to be seen if this behaviour will be confirmed by this CMS measurement.

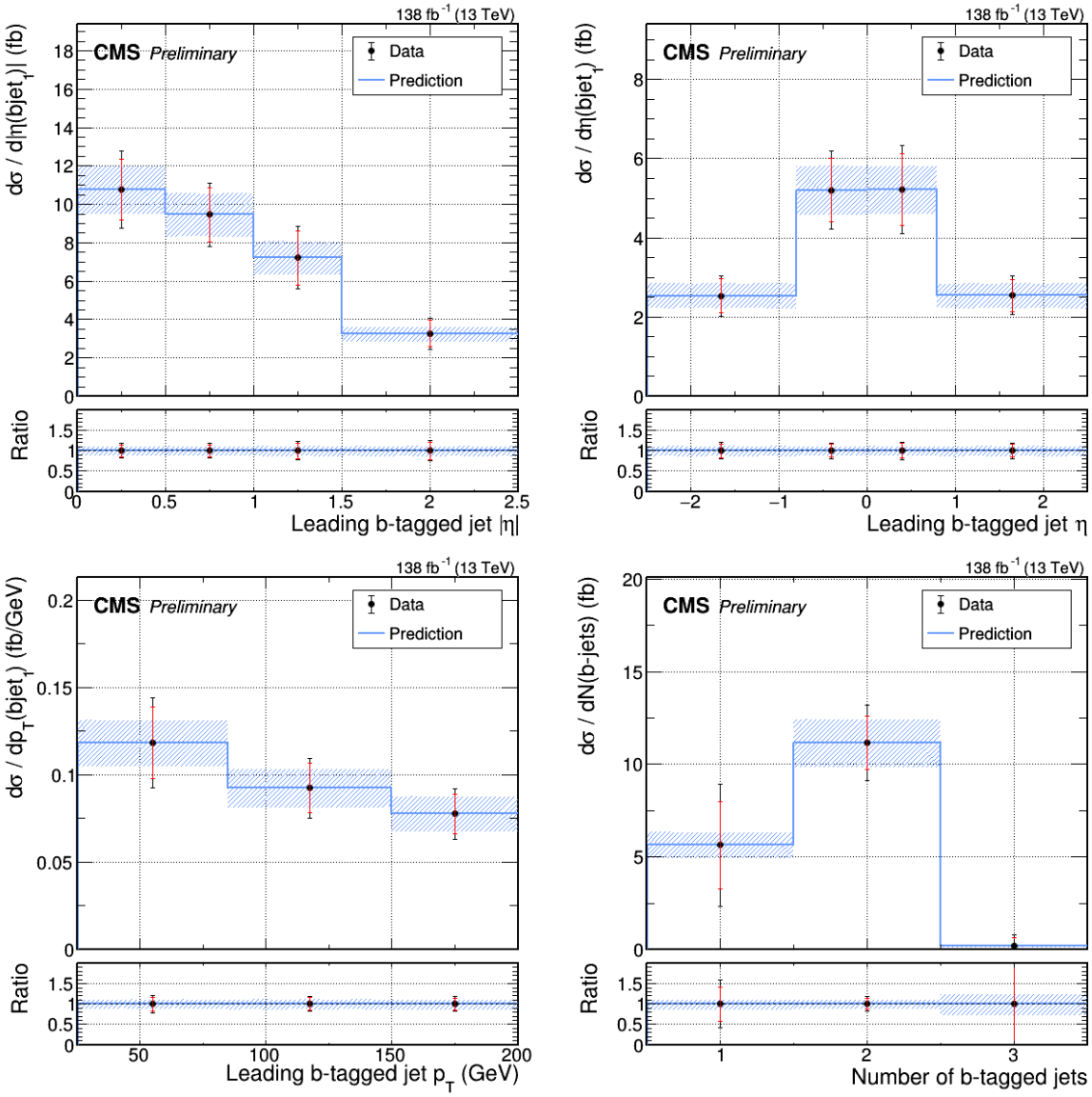


Figure 8.22: Expected differential cross-sections for b-jet variables.

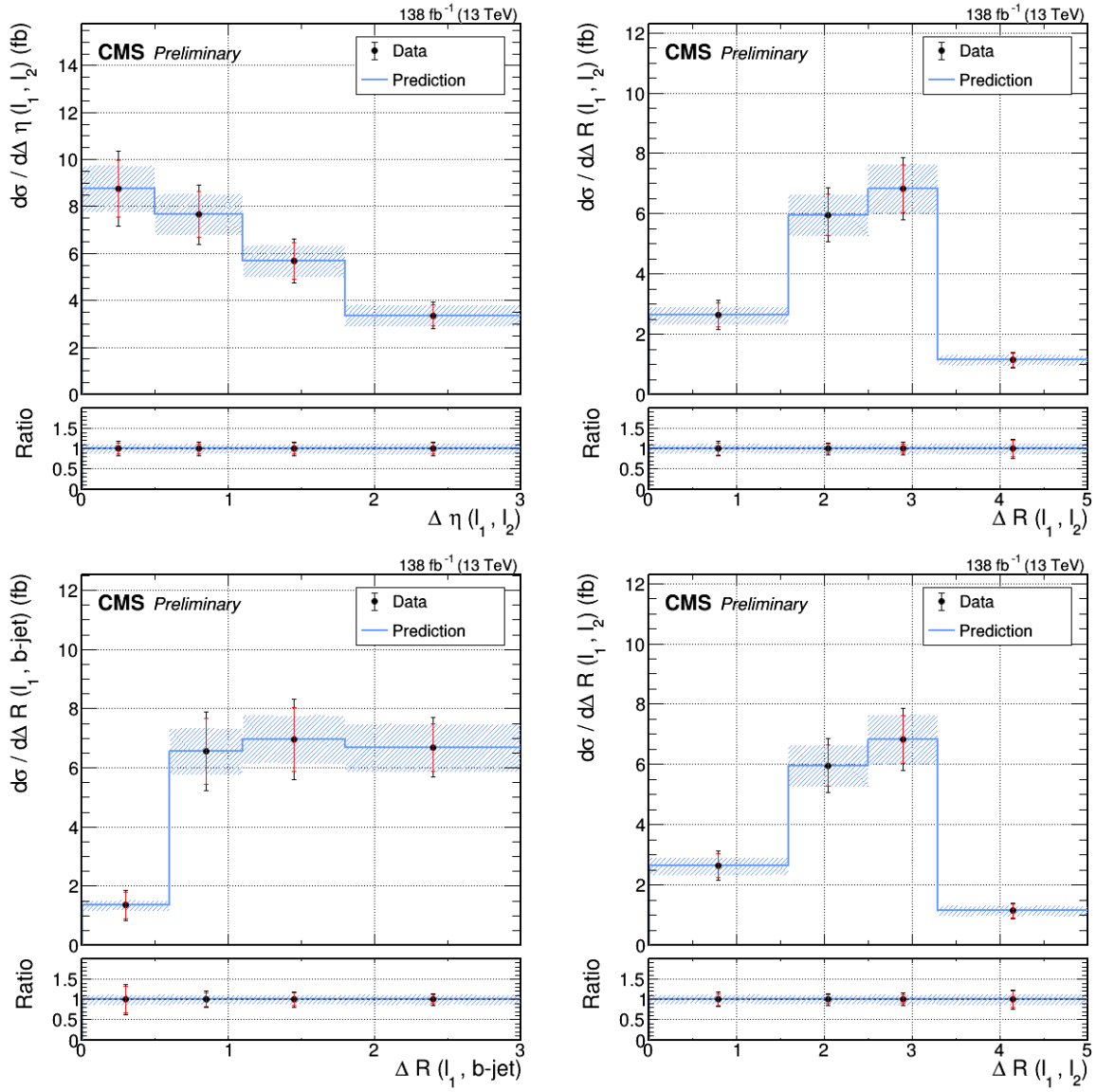


Figure 8.23: Expected differential cross-sections for angular variables.

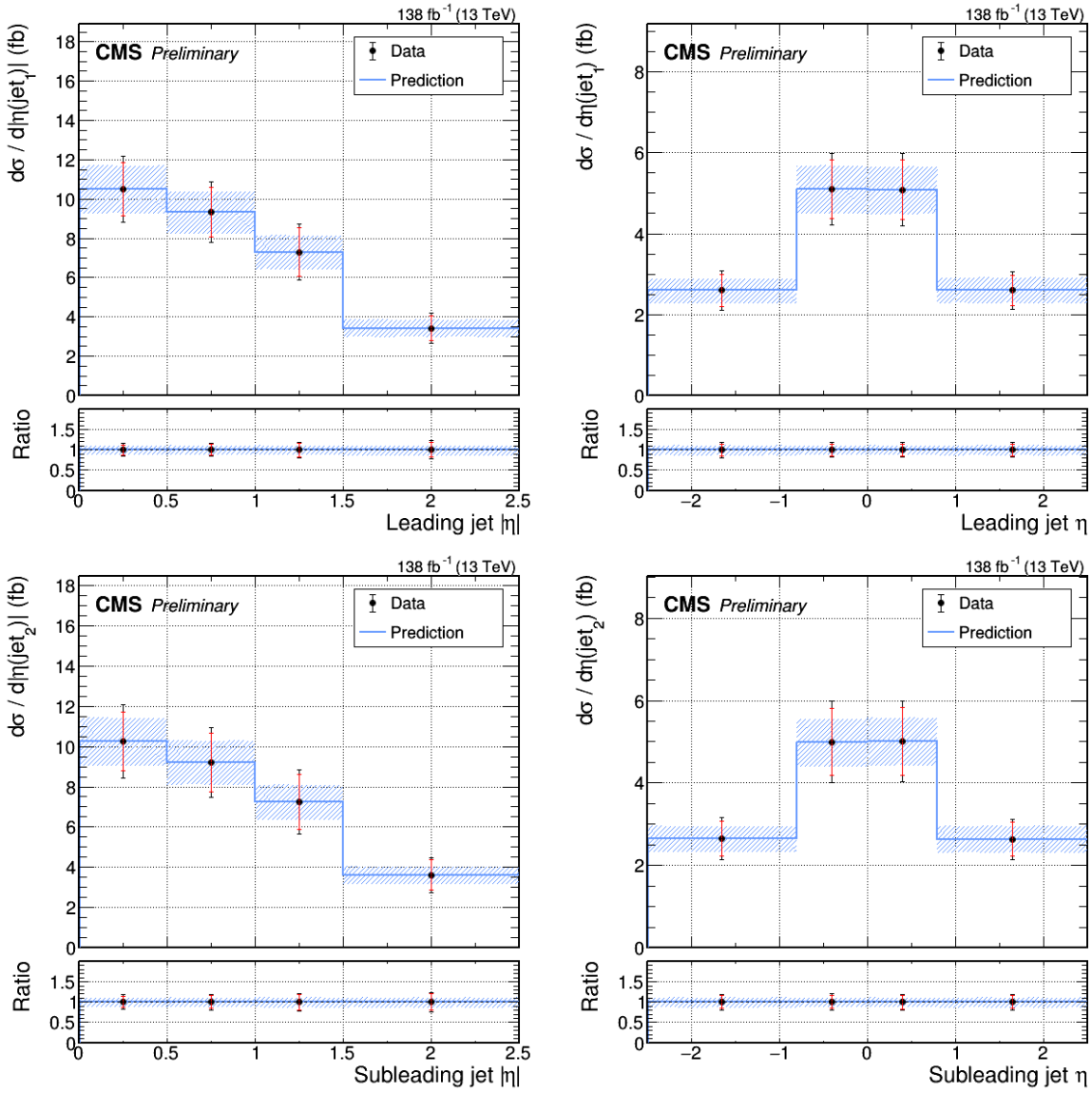


Figure 8.24: Expected differential cross-sections for jet variables (1).

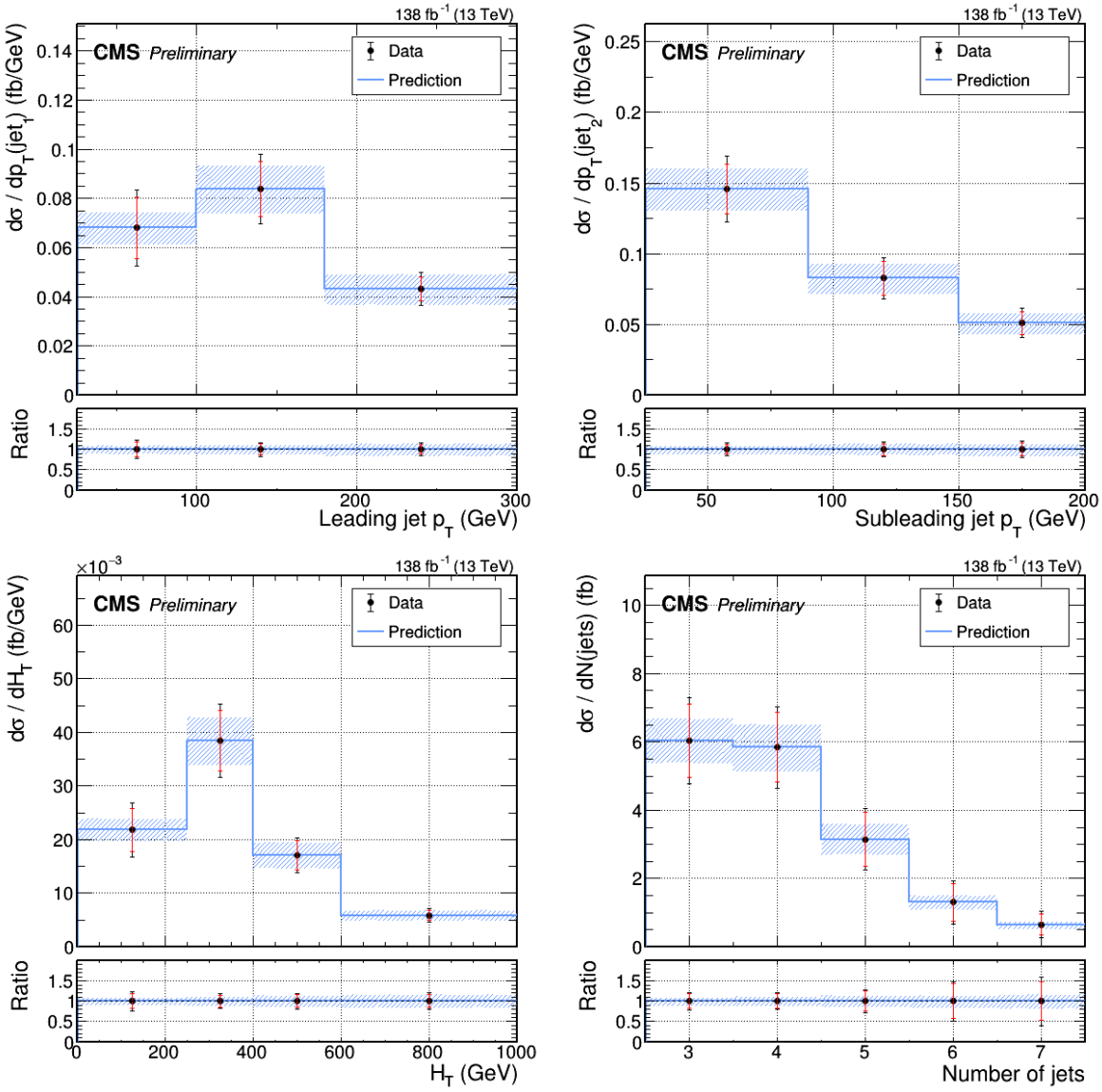


Figure 8.25: Expected differential cross-sections for jet variables (2).

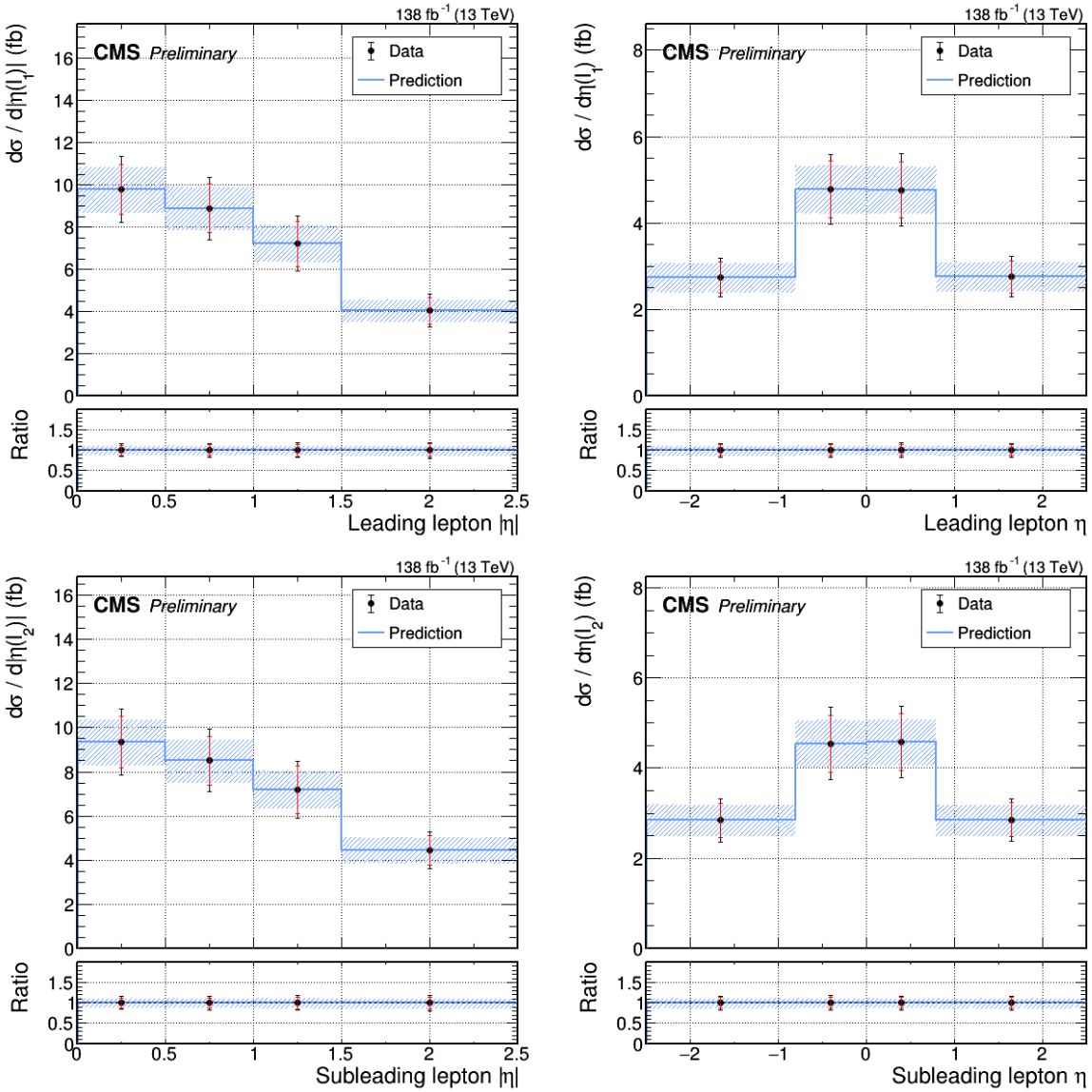


Figure 8.26: Expected differential cross-sections for lepton variables (1).

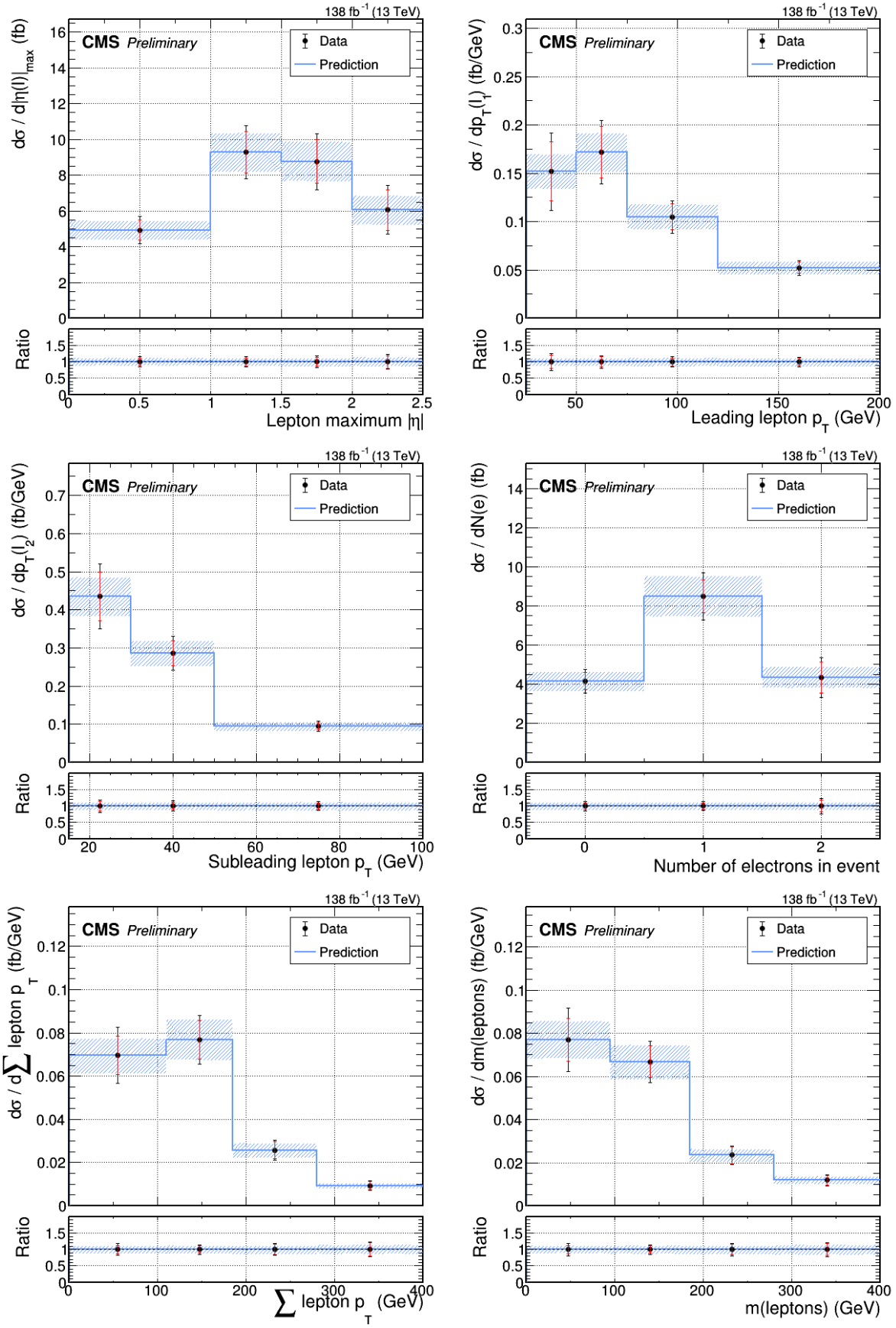


Figure 8.27: Expected differential cross-sections for lepton variables (2).

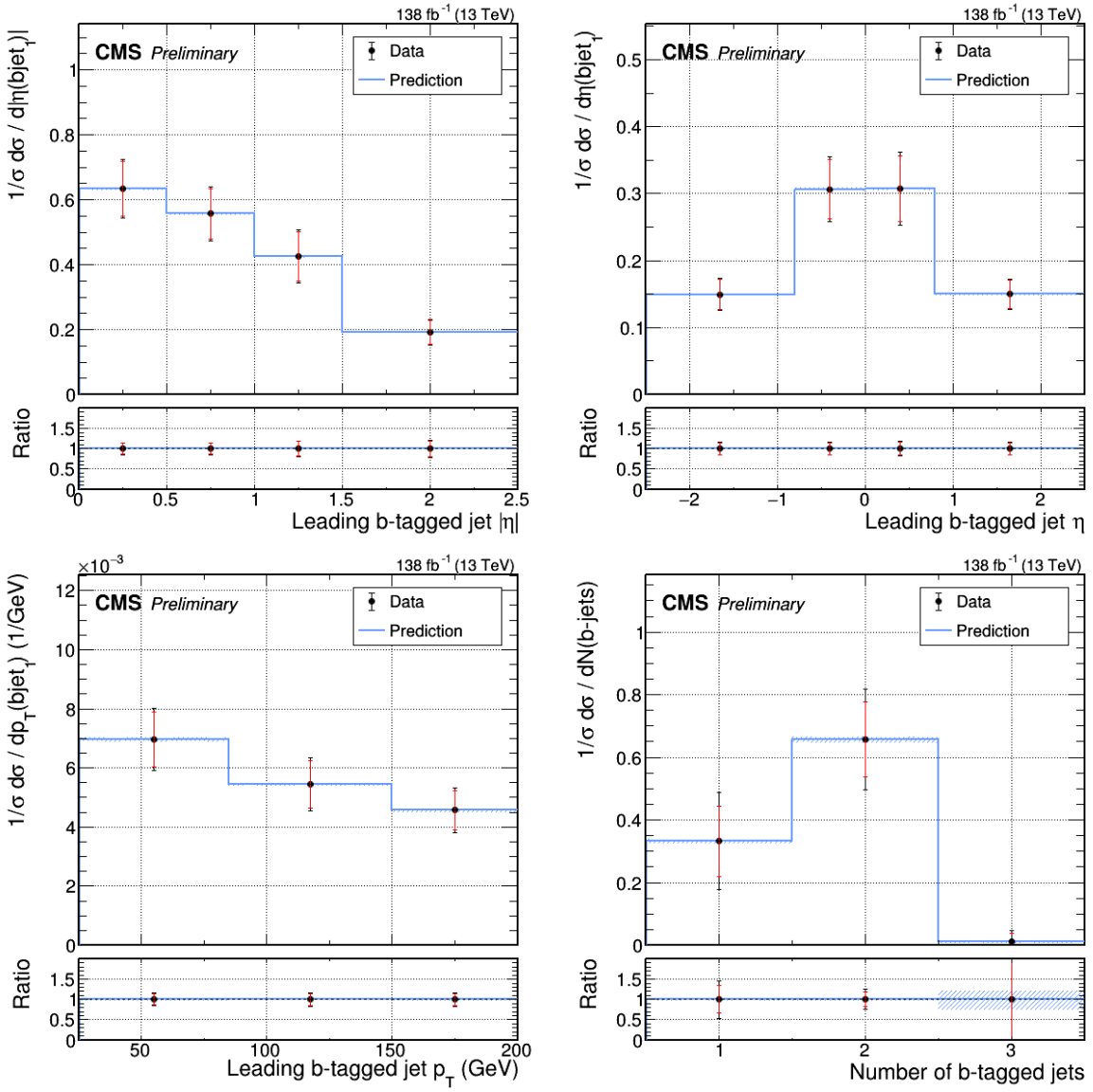


Figure 8.28: Expected normalized differential cross-sections for b-jet variables.

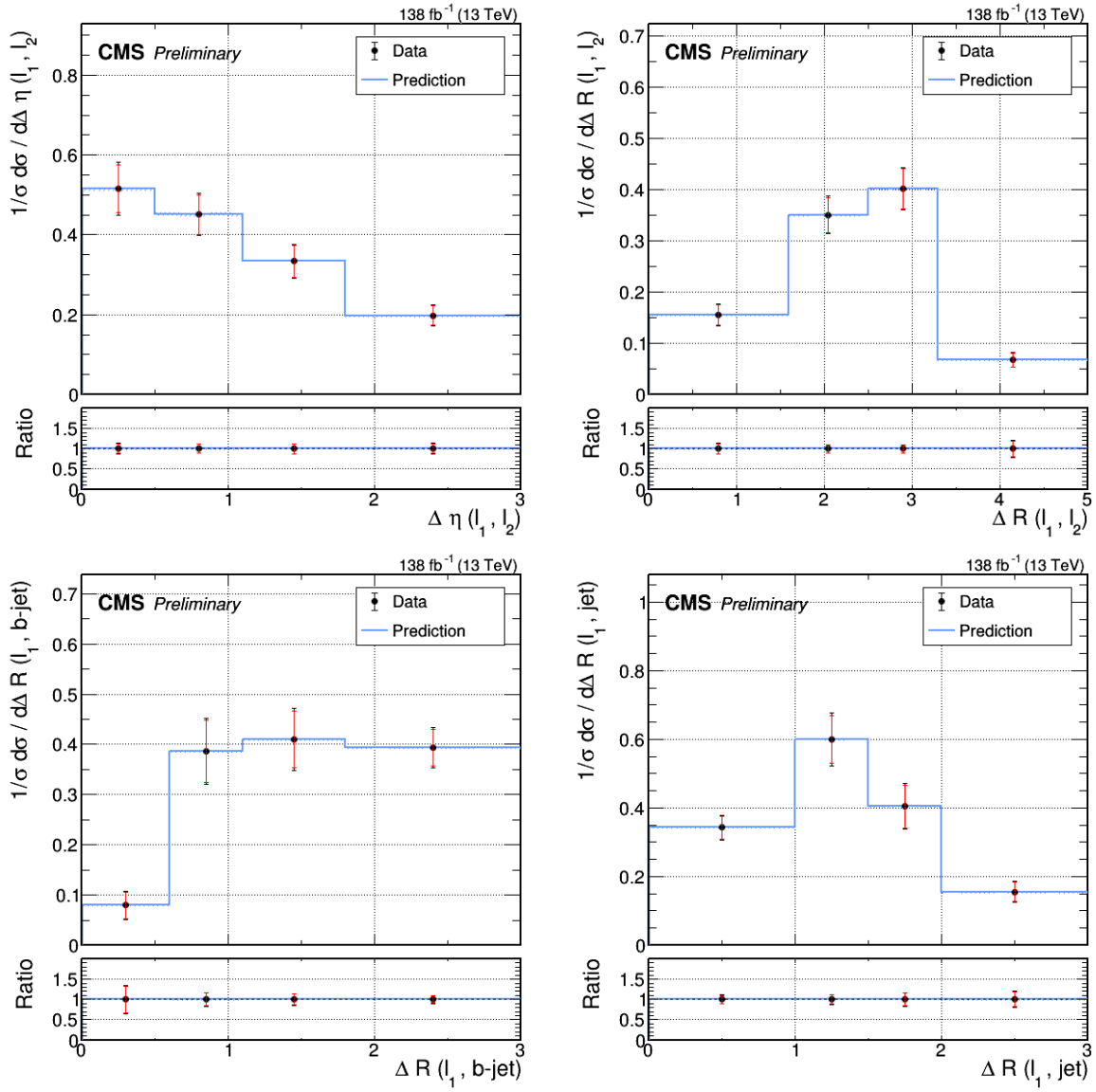


Figure 8.29: Expected normalized differential cross-sections for angular variables.

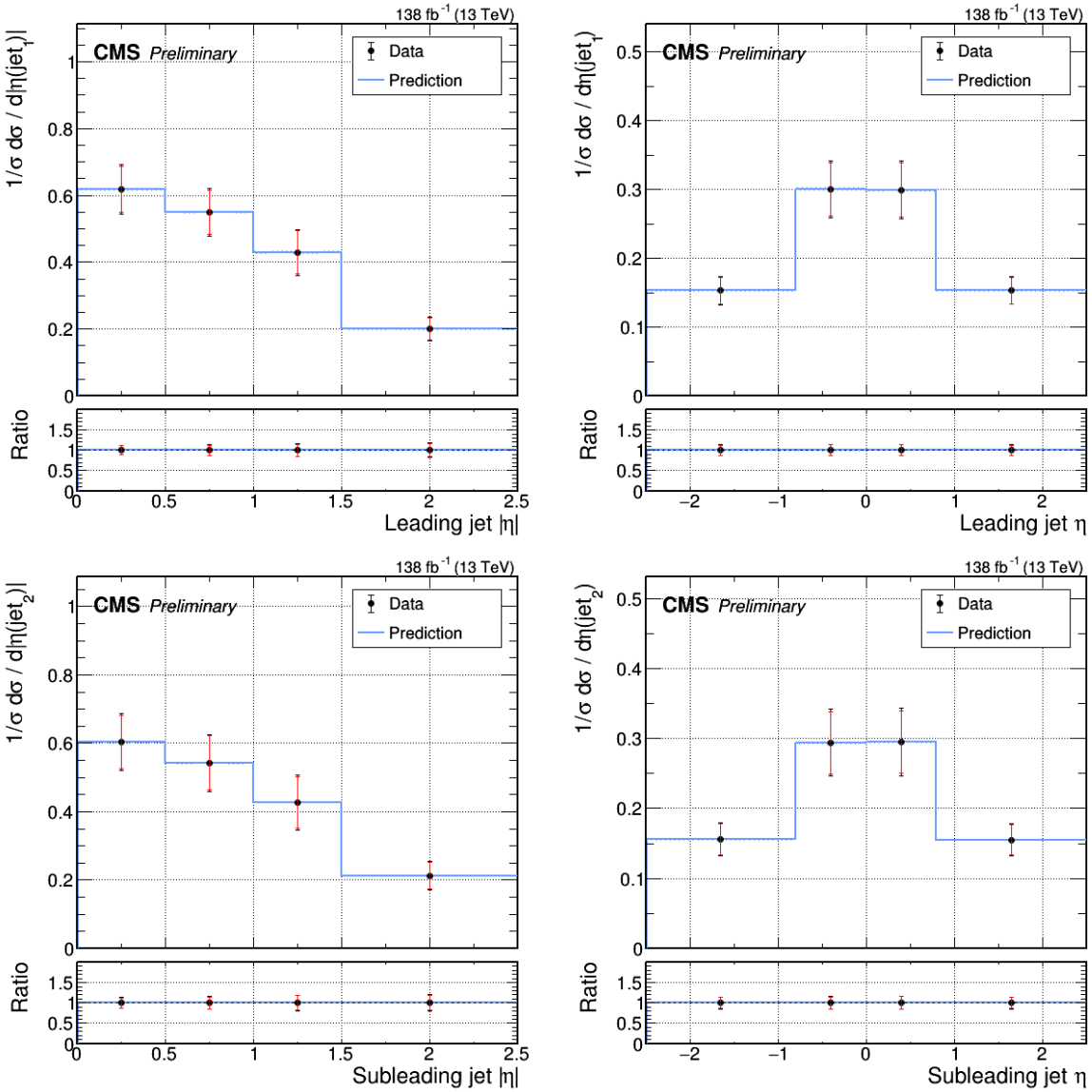


Figure 8.30: Expected normalized differential cross-sections for jet variables (1).

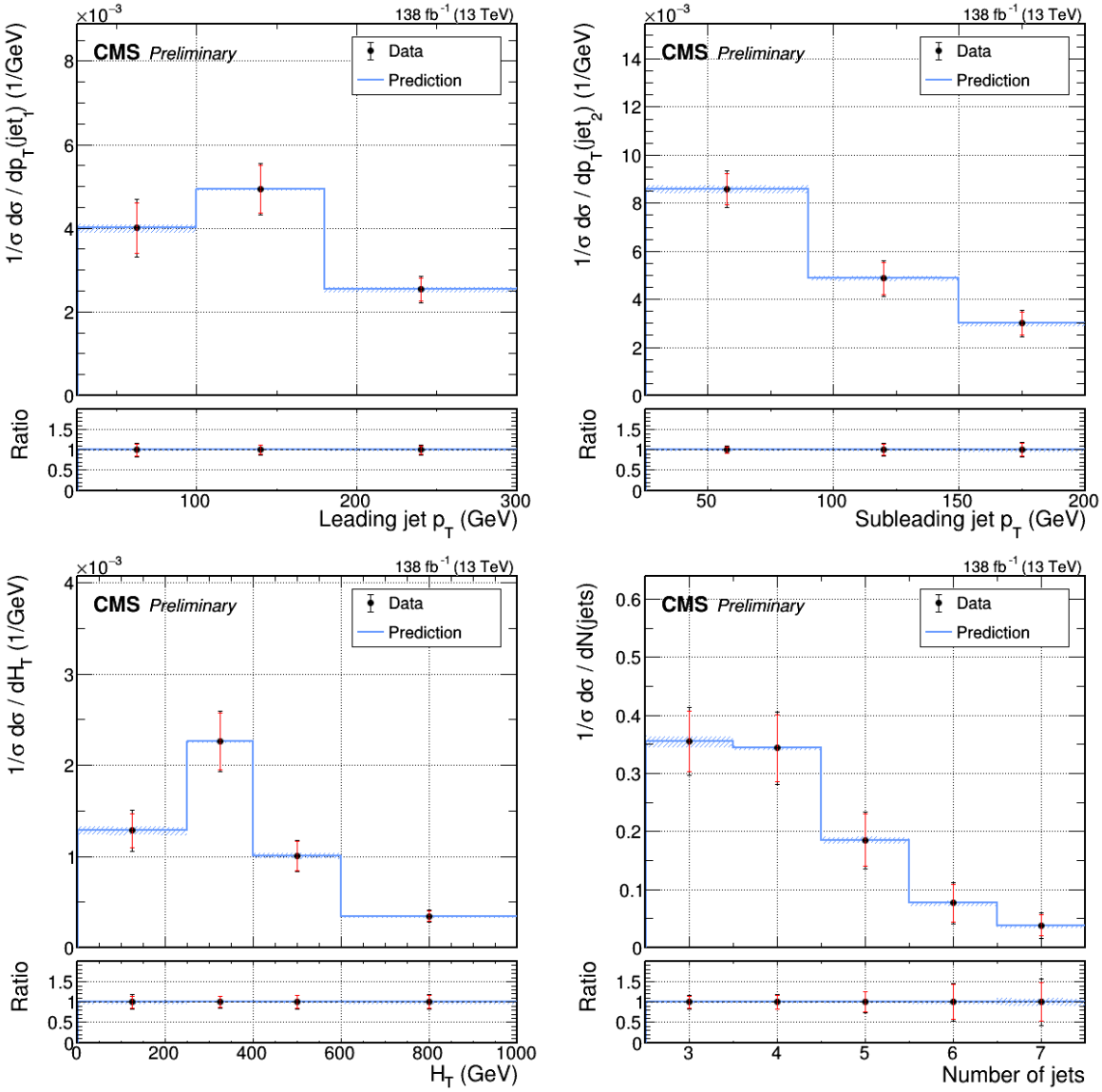


Figure 8.31: Expected normalized differential cross-sections for jet variables (2).

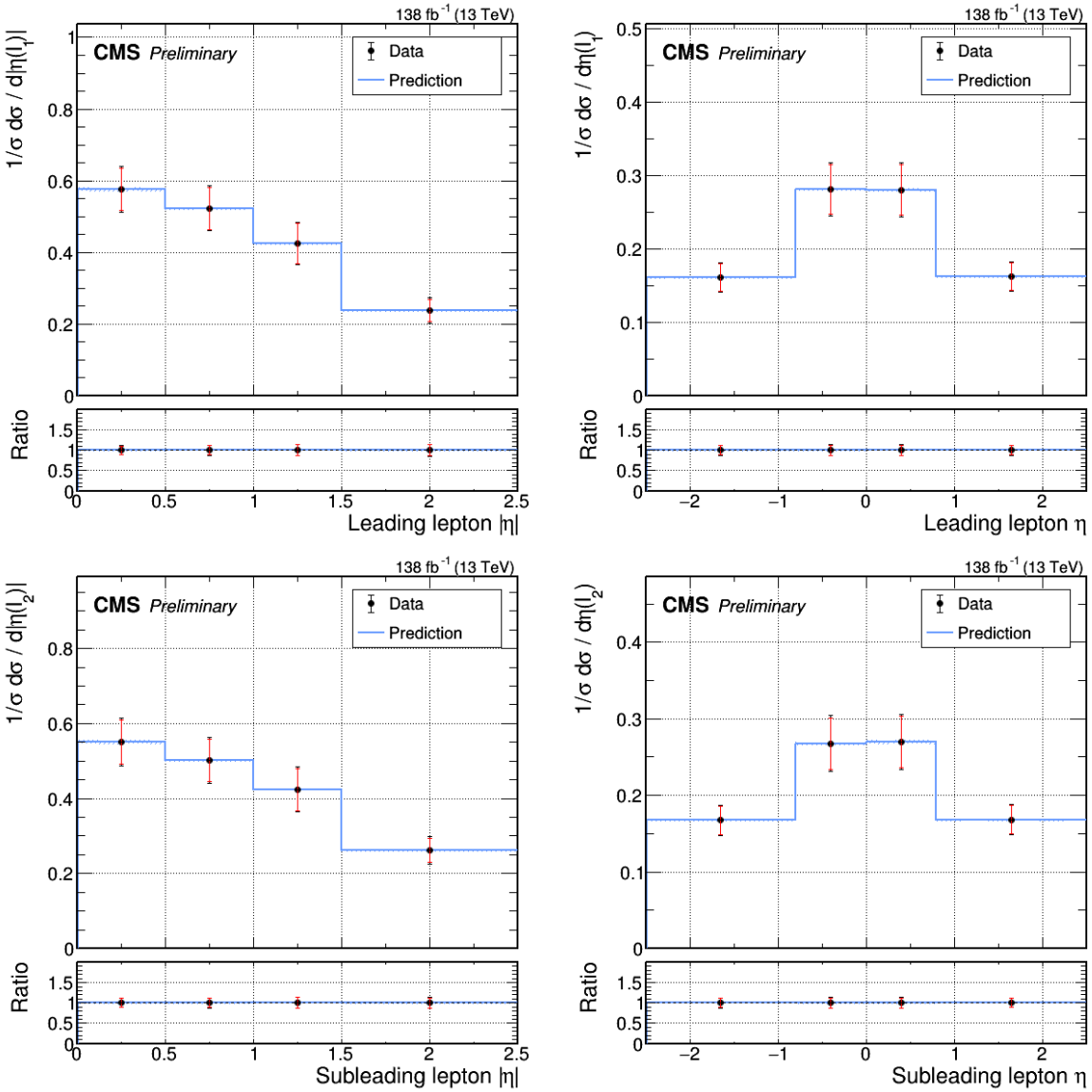


Figure 8.32: Expected normalized differential cross-sections for lepton variables (1).

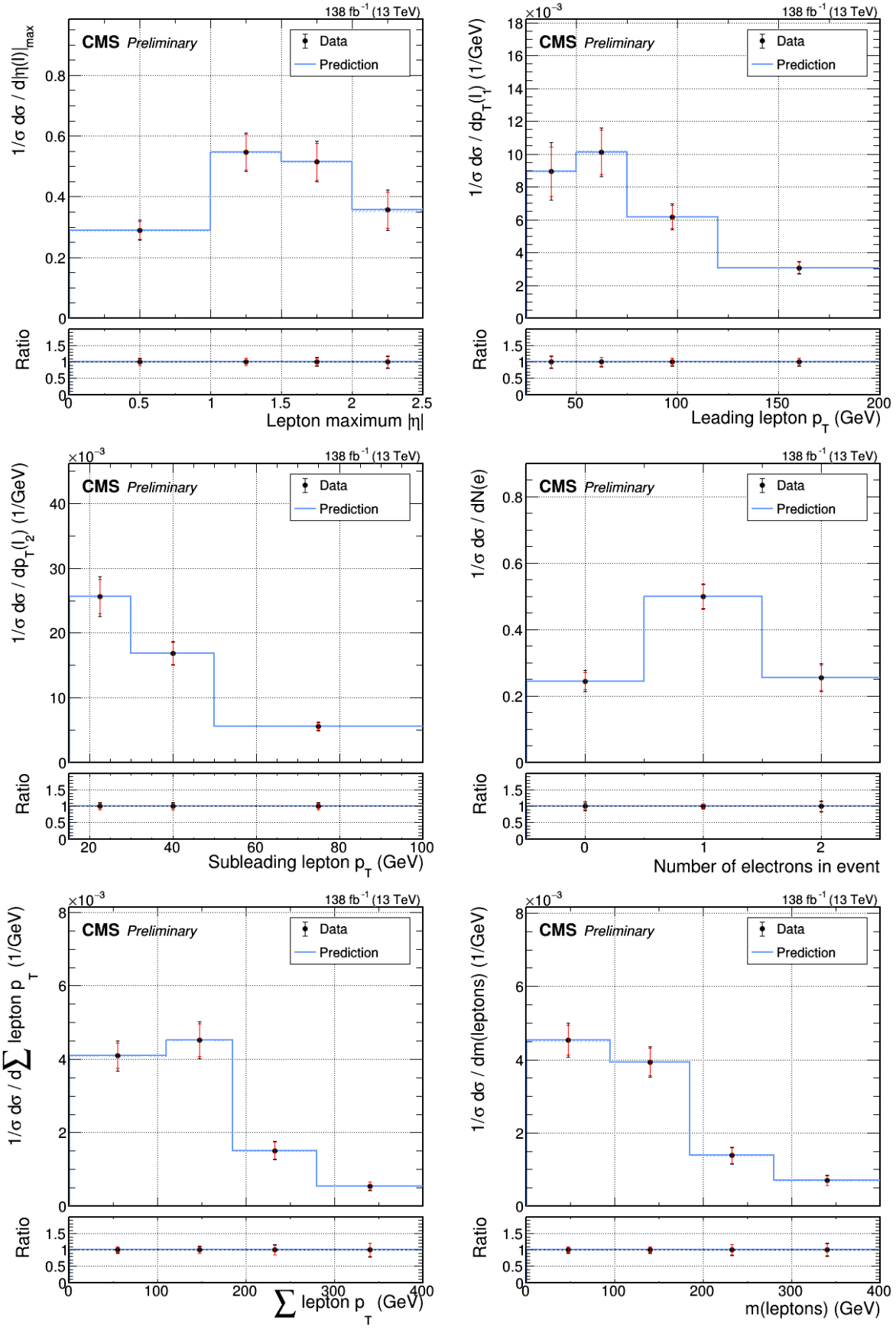


Figure 8.33: Expected normalized differential cross-sections for lepton variables (2).

Conclusions and outlook

To finish this thesis, we draw some overall conclusions, and consider what the future may bring. As the topics discussed in this thesis are rather variegated, we only present a short and broad overview here, and refer to the conclusions and outlook at the end of each specific chapter for more details. Furthermore, an abstract with the most important results of this thesis was already provided at its beginning, and only the points that are crucial here will be repeated.

Challenging the standard model All work in this thesis fits in the larger context of challenging the standard model of particle physics. At the CMS detector, signs of new physical phenomena beyond the standard model (BSM) are sought for either in a direct way, by hunting for the traces of new hypothetical particles, or in an indirect way, by measuring standard model parameters with the highest possible precision and searching for signs of inconsistencies. Chapter 3 (displaced tracking calibration) fits in the former approach, chapters 7 (tZq) and 8 ($\mathrm{t\bar{t}W}$) into the latter; chapter 4 (automatic DQM) fits in both, as ensuring good data quality is important for all CMS analyses.

Results for tZq and $\mathbf{t\bar{t}W}$ In this thesis, no clear signs of BSM physics have been observed. We confirm the validity of the standard model within the energy range and phase space studied in this thesis, and within the currently achievable experimental and theoretical uncertainties. The inclusive tZq cross-section has been measured with a total uncertainty of about 11%, which is the most precise measurement so far. The measured value corresponds well to the standard model prediction within the uncertainty. The differential $\mathrm{t\bar{t}W}$ cross-section has been measured for the first time in CMS. Though this analysis is currently still in CMS internal review and only expected results are shown here, its observed outcomes in the near future are expected to provide useful feedback to theorists for the modeling of this challenging process.

The future of tZq and $\mathbf{t\bar{t}W}$ The Run III period of data taking (currently ongoing) and High-Lumi upgrade (planned to start in 2029) are expected to yield larger datasets with respect to the ones used in this work[25], and therefore to reduce the statistical uncertainty on all measurements. For a differential cross-section measurement of $\mathrm{t\bar{t}W}$, this is relevant as the analysis is currently limited by statistical uncertainties; for an inclusive cross-section measurement of tZq, the direct effect of dataset increase is expected to be weaker, as systematic uncertainties will soon start to dominate. Both for tZq and $\mathrm{t\bar{t}W}$, since the relative statistical uncertainty only scales with the inverse square root of the number of events, early measurements with only a part of the Run III or High-Lumi data are not expected to be much more accurate than the current ones[26], and I would advise to only re-perform the tZq and $\mathrm{t\bar{t}W}$ measurements

[25] This work uses Run II data (collected in 2016–2018), corresponding to 138 fb^{-1} . Run III started in 2022, is expected to run until the end of 2025, and deliver about 250 fb^{-1}. The High-Lumi data taking is currently planned to run from 2029 to 2040 and deliver about 3000 fb^{-1}.

[26] That is under the assumption of approximately the same analysis strategy; the picture might change if more powerful data analysis methods become available.

once the full Run III data, or a significant part of the High-Lumi data, is available, in order to spend research resources efficiently. Apart from an increased luminosity, Run III and the High-Lumi LHC have a slightly increased center-of-mass energy with respect to Run II[27], but the tZq and $t\bar{t}W$ measurements are not expected to be drastically affected by this.

With more data (and improved analysis techniques), the focus in these measurements will shift from 'simple' inclusive cross-sections to more involved properties, such as the ratios between different initial states or decay channels (e.g. tZq with a top quark vs. a top antiquark), asymmetries (e.g. the enhanced charge asymmetry in $t\bar{t}W$), and kinematic distributions (that can at any stage be binned as finely as the data allows to achieve the maximum sensitivity).

Displaced tracking calibration We presented a new method for calibrating displaced tracking and vertexing that will benefit future searches involving long-lived, displaced experimental signatures by providing a quantitative estimate of simulation inaccuracies for displaced objects. The calibration factors are generally in the order of 10-15% for large displacements, but can depend strongly on dynamic detector inefficiencies and should therefore be studied carefully for each data-taking period and for the specific data reconstruction configuration of interest. The method developed in this thesis, while it has some options for potential improvements, can be (and in fact, is planned to be) readily employed in further data taking, in particular in the currently ongoing Run III.

Automatic data quality monitoring Finally, we showed some promising case studies for automated anomaly detection methods for data quality monitoring. Using an autoencoder, a detection efficiency for anomalies in the pixel tracker of essentially 100% was achieved, with a sufficiently low false alarm rate in the order of 10^{-5}, with a fine time granularity that is infeasible for the conventional purely manual procedure. Furthermore, important technical advances were made to provide the quality monitoring data with this finer time granularity as a standard procedure in CMS.

The baseline procedure and workflow for data quality monitoring in CMS is not expected to change dramatically in Run III or the High-Lumi upgrade. However, the trend of augmenting the baseline manual procedure with automatic auxiliary tools is likely to continue. In this scenario, tools such as the ones developed in this thesis would run during live data taking in parallel with the (human) shifter, and would direct the attention and focus of the shifter more efficiently to where it is needed by flagging potential anomalies. These methods are also expected to become vital tools in offline data certification, where time is less critical and dedicated studies can be performed to identify subtle anomalies with the highest possible accuracy and time granularity.

[27] 13.6 TeV and 14 TeV respectively, compared to the 13 TeV of Run II.

Epilogue: the consolation of science - a worldview based on physics

In the popular opinion, fundamental (particle) physics is sometimes depicted as an abstract and even irrelevant branch of science, dealing with extremely specialized and technical questions, the answers to which will never have any practical implications on society and human life. Although there are many arguments to vindicate the field of fundamental physics despite this criticism (see the prologue to this dissertation), its essential point remains true: answering the questions that are posed by contemporary fundamental physics will have little or no effect on the daily life of most individuals. That is, unless it is viewed in a broader context, moving away from the technical jargon and towards an interpretation of its spectacular, world-shaking and beautiful message that is relevant and understandable, and hopefully appreciable, for a wider audience.

The study of physics has come a long way since its original inception as speculative philosophy about the nature of all things. And in this process, it was gradually split into many different sub-disciplines, that in turn branched into even more specialized sub-sub-disciplines, and so on, pruning away many of the qualitative and speculative aspects and focusing on the quantitative and strictly observable ones. The results of this process are twofold: on the one hand, it has led to the tremendous and overwhelming success of science, both as explanatory and as quality-of-life improving force, but on the other hand, it has driven physics out of the general scope of society and daily life. Where the more applied fields of science are sometimes still able to spark discussions and interest in society because of the directly visible impact they have on it (e.g. artificial intelligence, nuclear power, genetic manipulation, lab-grown meat, etc.), particle physics has grown into a very specialized business that has no strong connection with the everyday thoughtscape of people anymore.

In this essay, I will try to show that modern fundamental physics can be still relevant in the same way as it was thousands of years ago: as philosophical lens through which reality can be viewed and interpreted, a way of grasping our place and purpose in the Universe, and even as a guidebook for an ethically responsible life. I will attempt this by coarsely sketching a worldview as it arises from the results of modern physics, and showing how this can be used as a compass in life. As many of the involved concepts, such as meaning, purpose, freedom and comfort are subjective and normative by definition, this will necessarily be an interpretative discussion. It is certainly not the only worldview that can be distilled from interpreting modern physics, and I'm sure there are many physicists who would disagree with me on a number of points. As a further disclaimer, I am neither a theoretical physicist, nor a philosopher (nor did I, admittedly, obtain enlightenment, harmony and perfect contentment employing the to-be-described worldview yet), but I hope this essay can still be an inspirational force driving further critical thinking.

What everything is made of

Let's start with the basics. And with this I mean the *very* basics: what is everything that we see and feel basically, fundamentally, essentially made of? It is in fact difficult to imagine that the world around us, in its full complexity and variety, consists of a relatively small number of elementary constituents. And yet it is this idea, of underlying order in apparent chaos, that science has discovered, or imposed – depending on the interpretation – in various forms and stages since its very beginnings. The early speculative attempts, that all matter consists actually of only one elementary constituent in different arrangements or degrees of compression (water according to Thales, air according to Anaximenes, etc.) might have been a little too optimistic. But even in our present knowledge, a relatively small number of elementary constituents (about 17, depending on the counting), when interacting in different structures and configurations, can generate the multitude of forms, textures and phenomena around us.

So what are these fundamental constituents? According to particle physics, they are – unsurprisingly – particles. The definition and idea of a particle is not so easy as it might seem. One can start by imagining them as small spheres, but then mentally reducing their size, until one ends up with point-like entities that have no magnitude or shape whatsoever. They do not 'consist' of anything, they are simply a collection of certain properties, such as mass and electric charge, that modify something in the space around them, such as the gravitational and electromagnetic field respectively. Put differently, a particle is not 'some thing with a so-and-so mass and electric charge', there is no deeper essence to it from which its properties follow; it is rather (and only) a collection of abstract numbers that tell how the space around it is affected by its presence. Or still differently, in some sense a single particle *isn't* anything, since it only exists in relation to other particles nearby (this thought is revisited in the next section as well).

We call a particle elementary or fundamental if it cannot be broken up into smaller parts. The designation of elementary particles has of course changed over time. As science progressed, partly driven by the technology to smash alleged elementary particles against each other with higher and higher energies, particles that were thought to be elementary at first turned out to be composite, consisting of lower level objects. A major milestone was the discovery of the substructure of the atom. After it had been discovered that all ordinary matter we know of consists of (only) about a hundred different kinds of atoms, it was now found out that actually those hundred atoms consisted of only three new elementary particles: protons, neutrons and electrons. Only the number of those particles in one atom determines its kind. Modern science has gone one step further down this ladder. Though electrons, as far as we know, are still elementary particles, protons and neutrons are not: they consist of other particles called 'quarks'. Just as different atoms can be arranged in a periodic table according to their number of protons, protons and neutrons themselves are just different representations of the same underlying particles, arranged in a different way, from which their composite properties can be derived.

Are quarks elementary? If we simply had the technology to achieve sufficiently high energies, could we not break them apart and discover yet smaller building blocks? Potentially. It is suggestive that there appear to be six types of quarks, ordered in three groups of two, that have similar properties except for larger masses; this seems reminiscent of the structure observed in atomic nuclei (the periodic table) that pointed towards the existence of the proton and the neutron. But on the other hand, there are reasons to believe quarks are the lowest step on the ladder. First, quarks cannot be isolated from each other. We can only 'see' quarks together in different arrangements (such as protons and neutrons), but we will never see a single quark on

its own[28]. Secondly, simply put: if you look deep enough inside a quark, all you see is more quarks. There is a sort of self-similarity in nature at this point: looking harder and harder (i.e. with higher and higher energies), you will see more and more quarks but nothing else. To understand this a little better, we must consider the equivalence between matter and energy. In the modern view, matter and energy are not two completely distinct concepts; they are in fact very closely related, in the sense that matter is just a form of 'condensed' energy. If you have sufficient energy, you can create particles 'out of nothing' and the other way round (though the total energy, including mass, is always conserved). Now the point is, when looking deep inside a proton for example, that necessarily means interacting with it at very high energy, this only causes the creation of more quarks. It is a little like when you zoom in on a fractal figure to find a straight edge, only to find that the same pattern is repeated on a smaller scale, no matter how strongly you zoom in[29]. This is an interesting way of providing an end to the at first sight endless sequence of breaking things down further and further; if you go far enough down this sequence, there is essentially no difference anymore between creation and destruction.

Hence, we are left with quarks and electrons. They come in different variations, and we need to add the particles that mediate the interactions between the matter particles, but overall still a relatively small number to make up all of the cosmos. The current theory describing all these particles and their interactions, is called the standard model of particle physics (see chapter 1 in this dissertation for a few more details). Neither its mathematical intricacies, nor its known shortcomings are of any concern here, the point being merely that we have, at least in principle, a fairly good understanding of how everything in our common-sense world is built up from a small number of what seems to be elementary constituents, and various interactions between them.

There is still a large and daunting gap between the mathematical description of the fundamental particles, and understanding complex phenomena in our everyday world. For example, quantum field theory is practically useless to predict the weather for tomorrow, nor will smashing protons into each other teach us anything about how thoughts and consciousness work. But the important point is that, up to now, there is no indication that this gap would be *in principle* insurmountable (even though clearly it will be *practically* very difficult). We do have a rather coherent picture of the structure of the world, starting at the fundamental basics of elementary particles, and leading all the way up to the most complex levels of biological and even mental life. Though the details of how all levels in this chain link together are in many cases not yet completely understood, we have at least a conceptual understanding of how the whole dynamics of our world is essentially reducible to a relatively small system of particles and processes, and more importantly, up to now no indication whatsoever that something 'outside' this framework of elementary particles and mathematical laws is needed to explain anything *in principle*[30]. Filling in the details is really just a problem of scale and complexity, not of concept.

[28] This causes people to question if quarks 'actually exist' or that they are simply a model, but the difference between both is anyhow vague at these scales.

[29] See for example *this animation* (digital link).

[30] This is no (attempt at a) proof that nothing outside this framework exists. Only that there is, with the current experimental data, both from specialized particle detectors and from everyday observation, no *need* for such as thing to exist to explain anything

Particles or processes?

In the previous paragraph, we touched upon the equivalence between matter and energy. This is one manifestation of a sort of complementary duality that pops up everywhere in this branch of physics. It started with the particle-wave duality in 'ordinary' quantum mechanics, stating that particles do in some circumstances not behave as point-like entities but rather as waves, known from e.g. electromagnetism or sound propagation. In one interpretation of this duality, the wave associated to a particle is a spatial and temporal probability distribution, indicating where and when the particle is likely to interact.

Now imagine a tempestuous sea, with large waves roaring randomly to and fro, constantly rising, sinking and crashing violently into each other. Behold, for such is the nature of reality. What we call particles are but the foam on the crests. When looking only at the foam, we see a very diagrammatic-like behaviour typical for particle physics: two patches of foam move towards each other, an interaction happens, and two or more deformed pieces of foam move away from each other. Yet the foam is only the tip of the wave crest, indicative of the wave but not the thing-in-itself. In the modern interpretation, more specifically in quantum field theory, this is taken one step further. All particles are considered to be quanta (elementary units of energy or vibration) of abstract fields permeating all space.

Up to now, we have described elementary particles as being a sort of 'building blocks', static constituents that one has to pile in the correct arrangement to make up reality. This image is useful in a sense, but it neglects the strongly dynamic aspect of nature. Isolated particles have no influence on their surroundings whatsoever; it is only by mutual interaction that overall change and complexity can arise from a small set of simple elementary constituents. Particles are in continuous interaction; they are never simply a static 'piece of matter' like a brick is a piece of a building. They are more like starlings in a huge swarm, continuously moving about in dynamic interaction with each other.[31] This is especially true for the strong interaction between quarks inside nucleons (or other hadrons): while the proton and neutron are typically represented as consisting of three quarks, in reality they consist of an indefinite number of quarks, antiquarks and gluons continuously engaging in mutual emission, absorption, creation and annihilation[32].

Hence one could argue that the fundamental constituents of matter are not particles, but a number of conserved quantities (such as energy and electric charge) and a set of rules describing the possible transitions between different forms of these quantities. In this picture, the particles merely enter as transient, indistinguishable and in a sense not even 'real' messengers, carrying that which is conserved from one temporary form to another. It is not true, as is sometimes believed, that fundamental particles are indestructible and eternal entities of existence, and that only their arrangement changes over time; no, both the constellation and the fundamental particles themselves are subject to change, transformation and destruction. Temporary permanence is provided by several conservation laws that limit the nature and probability of the allowed transitions, but the timescale between creation and annihilation is rather accidental; the impression is enhanced by a certain kind of observer bias, in the sense that both our physical and mental characteristics have evolved in such a way as to perceive the world and ourselves as largely permanent, since the alternative is a little less 'encouraging' in the struggle for life.

[31] See for example *this video* (digital link), and imagine moreover that the starlings are regularly crashing into each other and transform into other starlings.

[32] Another comparison, though a little more esoteric, is to the cosmic dance of Shiva. For this and many other interesting ideas, see for example F. Capra, *The Tao of Physics* [251] and *this summary* (digital link).

The image of nature sketched above leaves, in my view, little room for entities like eternal forms or souls, or even for common-sense categories like personal identity, and instead bears more resemblance to the idea of continuous metamorphosis. Though nothing comes from nothing (or goes to nothing)[33], everything is transformed far beyond any remaining identity or recognition, if one only waits long enough. Information (such as personality and memories) is at a fundamental level 'stored' or 'encoded' by particles and by the correlations between them, which can get irretrievably lost by random thermal noise if not actively maintained and backed up by various mechanisms in the brain while it is active. In this view, the idea of a remaining identity, both in people and in objects, is only a contingent illusion that arises because of the fact that we usually happen to die too soon to notice significant changes, that are progressing on a longer timescale.

Simplicity and complexity

Physical nature reveals a nontrivial and bidirectional relation between simplicity and complexity. On the one hand, the aim of physics, towards which it might be said it is rather successfully progressing, is to discover an underlying structure in apparent chaos, unity in multiplicity, a set of simple paradigms to describe the entire complex Universe. On the other hand, even the most simple, directly experienced phenomena can be broken down, at various levels of (mathematical) abstraction, into very complicated causal chains of which we often do not understand all the details yet. In other words, the deeper one dives into physics, the more clearly one starts to see both the forest for the trees and the trees for the forest[34]. The same duality occurs in the process of understanding itself, which arises from an interplay between analysis (i.e. breaking things down in more elementary constituents) and synthesis (i.e. building things up by putting elementary constituents together).

Is nature essentially simple or fundamentally complex? The answer to this question is currently yet not very clear, partly because the terms simple and complex are ambiguous. They can refer either to the multiplicity of fundamental entities, or to the possibility of understanding, and both meanings are not necessarily correlated. Though the standard model of particle physics seems to point to a simple reality, in the sense of having a fairly limited number of elementary entities and interactions between them, it takes at least a few years of dedicated study to start to grasp the basics of how it really works. The same is true in an even more extreme sense for candidate theories-of-everything such as string or brane theory, that often are (or at least claim to be) conceptually simple and all-encompassing, but prove to be technically-mathematically *very* complicated. This may be just a temporary phenomenon, and children of future generations might have no particular issues with learning string theory in high school, just as nowadays we learn derivatives or probability theory that used to be state-of-the-art mathematics several hundreds of years ago, but this is impossible to predict with any certainty.

The simplicity of nature is an important question, since the answer *might* turn out to be that nature is, at its most fundamental level, practically unintelligible to the human mind by the sheer complexity of its concepts and/or mathematics. While probably not very important in practice, since we already know how to model and exploit nature phenomenologically at various non-fundamental levels, it is a terrifying thought in principle. However, until further

[33]The several conservation laws that set the boundaries for all transformations in our Universe are, in my opinion, not specific enough to provide a pathway for the conservation of anything we would call an identity.

[34]However, in practice, we often develop a tunnel vision in the latter aspect of ultra-specialization only; therefore it's important to take an explicit step back once in a while and enjoy the scenery as a whole.

evidence, this does not seem to be the case[35], and a continuation of the quest for unification and 'simplification' (in whatever sense) remains reasonable and fruitful.

At least in one sense, physics appears to be 'simple', in that it claims to describe the Universe and everything in it with one paradigm[36]. The specific contents might be rather complicated, and moreover they might change rather drastically over time (e.g. the replacement of Newtonian physics by relativity theory), but the gist remains unaltered: the single conceptual framework of mathematical laws, that are simple enough be understandable by the human mind (or at least by some human minds), is sufficient to describe the world around us. This point is taken up in more detail in the next section.

Matter and mind

Physics describes matter and the way it engages with other pieces of matter. And it does this extremely accurately, to the extent that we can describe most of the Universe with relatively simple physical-mathematical laws. 'Physics' should be interpreted here in the wide sense of all physical sciences including e.g. chemistry, biology, geology, etc. Furthermore, over the course of the past centuries, physicists have shown that many parts of the Universe that were at first thought to fall out of the scope of physics, are actually more or less simple material processes that can be described with much of the same language as typical 'physical processes' such as balls of different sizes and weights falling off towers. Think of classical examples such as thunderstorms, but also biological procreation, the variety of animal and plant species, diseases, etc. Many parts of the Universe that are not currently understood in accurate physical terms (e.g. dark matter) are not generally believed to be intrinsically nonphysical; it is only a matter of time and sufficient research funds. The impression arises that the whole Universe and everything in it can be – at least conceptually – described by physical-mathematical laws, though in practice many systems would be so complicated that the actual description would require unfeasible computing resources.

One notable exception to this view is the human, conscious mind. We don't like to think of our own mind as being a purely material, albeit somewhat involved, calculator. Without denying the reality of our physical brain, we tend to believe that the mind must be something more than e.g. a computer program running on a physical computer; that what we think and especially what we feel and value is not purely and completely determined by electrons moving about in neurons. But in that case, what is the mind exactly? How does it relate to the brain, or to physical reality in general? This mind-matter problem is a very old and infamous one, and I do not claim to have the conclusive answer. Yet I would argue that, until further evidence, the human mind should be classified as one of those parts of the Universe that we do not yet fully understand in physical terms, but that are conceptually and fundamentally not different from more simple parts such as billiard balls hitting each other or electrons interacting with quarks.

Though at first glance it may seem too bold to 'reduce' the mind, and therefore our thoughts, memories and deepest feelings, to a purely material system characterized by input-process-output relations, we should not forget the sheer complexity of the material brain, which allows

[35] If it *were* the case, we're basically back at mysticism and religion after a long (but very useful) detour.

[36] As I have been and will be emphasizing many times in this text, this is of course just in principle. Higher-level sciences such as chemistry, biology, engineering, and psychology will always remain important for practical purposes. But the idea that (or rather the question if) there is an underlying theory from which all these 'applied' sciences could at least in principle be derived, is important for non-practical purposes

composite features to emerge that seem very far above the capabilities of the original building blocks. Consider, in analogy, a glass of water[37]: none of the phenomena that we associate with an everyday glass of water (such as its 'liquidity', its 'flowing', its 'transparency', its being 'refreshing') have a clear counterpart at the molecular level; and yet when adding sufficient molecules together, their interactions produce these emergent features. Furthermore, the alternative seems to me far more unsettling. What exactly would the mind be, if not matter governed by physical laws (and apart from dogmatically postulated non-explanatory religious entities)?

We can consider the genealogy of mental processes (e.g. emotions and consciousness), as being merely tools in the struggle for survival in the ancient world, either in their own right or simply as a corollary, a side-effect of gradually developing abstract thinking, planning and reflection skills. That all of these faculties are indeed somewhat correlated and gradual rather than binary can be clearly observed in nature, following the spectrum from the side of inorganic matter, over plants, insects, animals, all the way to primates and us. If we find it easy to believe, without resorting to anything 'supernatural', that a plant will react to external stimuli, e.g. growing stronger towards the light, is it so much more difficult to imagine that we do exactly the same, only using a more complicated chain of action, reaction and reflection? This argument will be expanded on in the next paragraph, taking the mind/matter paradox to its full extremity.

What will be, will be

The question of whether or not our Universe is deterministic is a very profound one, with severe consequences in the fields of ethics, morality, and teleology. In a deterministic Universe, the laws of physics fully determine the behaviour of all particles, and in more or less direct extension, of all we see, feel, think and do. This situation was famously formulated by Laplace as follows: if a hypothetical entity had a perfect knowledge of all elementary physical laws, and if it could determine the position and velocity of all particles in the Universe at a given moment in time, then it could in principle predict the entire future of the Universe and everything in it. Though the practical problems of creating such an entity are insurmountable, and though quantum mechanics and relativity theory have taught us that measuring the position and velocity of particles and the concept of 'a given moment in time' are more problematic then they seemed in Laplace's days, this still remains the essence of determinism. Just to already mention one of the most striking consequences, this interpretation of the Universe leaves no room for the classical, 'naive' notion of free will, the idea that 'I chose to do this but I could have chosen otherwise'. Let's see how this idea of determinism, that is so in tension with our naive everyday conception of our lives, can be built up starting from the simplest case of moving bodies in classical physics.

The deterministic view is easily accepted when one only considers the purely mechanical movement of bodies according to Newton's laws. Every first-year physics student will be able, given a physical body at a certain point in space and time, and given a set of forces acting on that body, to determine the position and velocity of that body at any given future point and time. Also, he or she will have no particular mental trouble in accepting that this is indeed possible, that the body will not 'out of its own' start to move in another direction than that predicted by Newton's laws. Note that there are a couple of practical concerns, that however do not change the relevant principle here (at least at this stage in the discussion). For example: how accurate can one know the initial position and velocity of the body, or how accurate can one

[37] Derived from J. R. Searle, *Minds, brains and science* [252]

determine the forces acting on that body? Obviously, small measurement errors in the initial conditions or forces will completely invalidate the long-term prediction. But this is only an experimental, and hence technical and temporary issue. Other example: how valid are Newton's laws? What matters here is not the exact content of the laws, but their nature, which is mathematically, exactly and unambiguously equating certain state parameters to (rates of changes of) other state parameters. Hence it does not matter if we replace classical physics by general relativity, or by any other theory yet to be conceived, as long as it satisfies this principle.

Quantum mechanics at first seems to overthrow this view of the Universe, but, on further reflection, will be seen to not significantly alter determinism's claims about the existence of free will and ethical responsibility. At the core of quantum mechanics is the principle of nondeterminism, the idea that many processes and properties of physical systems can only be defined in terms of probabilities. For example, an unstable particle will decay within a given time frame with a given probability. The properties of a very large collection of these particles can be determined and characterized using statistics, but it is impossible to predict, for any single of these particles, whether or not it will decay in any given time frame. This is an impossibility in principle, not only in practice: it is not a consequence of lacking knowledge about some property of the particle, or of inaccurate measurements. It should be noted that there are other interpretations of quantum mechanics that recover full determinism, but they have either been proven false (such as the hidden-variable theory) or do not change significantly the further discussion (such as the many-worlds interpretation) and will not be discussed further. Hence, the laws of quantum mechanics do not exactly and unambiguously determine the future of a physical system from its present state, and a Laplacian entity predicting the future of the entire Universe is not only in practice, but also in principle, inconceivable.

However, we should realize at this point that determinism as (perfect) predictability is not the complement of the existence of what we would call 'free will'. In fact, quantum mechanics puts our Universe somewhere on the spectrum between those two extremes, where our Universe is stochastic and hence not exactly determined and predictable, but yet governed by strictly mathematical laws, extended with calculus of probability and statistics. At this point, the only possibility for free will, as an escape from determinism, is to regard it as the random quantum noise that might occur in the neurons in our brain while they are exchanging the electrical signals that are our thoughts and feelings. This however, though it might make us unpredictable, would not make us free[38].

Another difficulty lies in finding the boundary between the world of physics and that of the free will. This boundary must be bidirectional interface: on the one hand, sensory inputs from the physical world must be accessible to (but not determine) the decision-making will, and on the other hand, a decision made by the will must be able to have an impact in the physical world. But this implies, in essence, that there must be a 'place' (e.g. inside our brain?) where particles from the physical world enter, are given a new position or direction by the force of the will (which can by construction not be described in the language of physics), and exit again to propagate this change to the material world. This may seem very unlikely, but it is very hard deny the existence of such an interface in one form or another if one wants to have both free will an a physical, material world. There is presently, at least as far as I know, no evidence whatsoever for this kind of interface, other than our stubborn and intuitive conception of free will. Other solutions are either rejecting the existence of a physical world outside our mind,

[38]But here we might stop and wonder what exactly we mean by 'free will', and whether this is even an a priori conceivable concept. If an action is chosen for some reason or other, it is not 'free': we can trace the sensory inputs (which are beyond our control), all the way to the choice made as a response to those inputs. However, if it is chosen for no reason at all, it is not 'willed', but rather just random.

or rejecting the existence of free will.

But in the latter case, where does the idea of free will come from, and why is it so hard to let go of? I think that a hint on the origin of the illusion of free will might be found in its naive formulation mentioned earlier: 'I chose to do this but I could have chosen otherwise'. The idea of free will can be regarded as a by-product of the human faculty for planning and advanced decision making based on expectations for the future. In comparison with animals and plants, which are in varying degree determined by instincts, i.e. decision-making processes of the form 'given the current situation, I should take this action', human beings have more advanced mechanisms of the form 'given the current situation, and given the set of conceivable actions and their expected effects, this particular one seems the most appropriate one'. Though this process can be fully deterministic, where one particular action is bound to be chosen in a given prediction-value system[39], it seems plausible that it gives the impression 'but I could have chosen otherwise' in hindsight. This advanced planning and decision-making circuitry is by itself nothing more than a very useful survival tool, materially encoded in the genes, developed in a fully stochastic evolutionary process, and explained in terms of survival of the fittest. Because of the vast complexity of the brain and the processes within it, it is difficult to see how we are driven by a combination of genetic base structure, past experiences and knowledge, and sensory inputs; much more difficult than, for example, seeing that a billiard ball is moving because it was hit by another ball. Yet as far as we currently know, this is only a difference in scale and complexity, not in principle.

So what does it mean if our Universe is indeed stochastic-deterministic? If our thoughts and actions are determined by a combination of chance and strict physical laws? Much can be, and should be, and has been said about this, but to summarize my own views very shortly: it means that there is no good or evil in the world, only good or bad luck; it means that there is a sense of peace with literally everything that happens[40], regularly interrupted by a sense of bottomless fear that there is no guarantee that anything will ever be 'set right'; there is a deep compassion and understanding of other people and their views and actions, as one sees how none of us can choose their genes, upbringing and environment, and at the same time the paradoxical intolerance to the lack of that same compassion in others; there is a devaluation of the concepts of self and sense, but fragments of a deeper understanding in recompense.

The origin and end of the Universe

How did the Universe as we experience it come into existence? While it is true that many questions about the Big Bang remain as yet unanswered, I would like to put the emphasis elsewhere: we have a fairly good understanding of the evolution of the Universe from a couple of tiny fractions of a second after the Big Bang until now, about 14 billion years later. We understand the formation of moons, planets, stars, galaxies and even larger structures from the earliest super-dense chaos right after the Big Bang. While some important pieces are missing from the theoretical understanding, notably so-called dark matter and dark energy, we can already successfully use them in calculations and simulations using simple proxy models, the substitution of which by a more realistic description seems to be just a matter of time.

In many respects, the Universe seems to be especially 'tuned' for supporting complex life such as our own. That is to say, if the value of any of the fundamental physical constants that define the behaviour of the (theories describing the) Universe were but a little different from what

[39] Derived from G. L. Drescher, *Good and Real: Demystifying Paradoxes from Physics to Ethics* [253].

[40] And even a secondary-level kind of peace with experiencing very unpeaceful emotions of joy and despair.

we observe them to be, anything like human life appears to be impossible[41]. For example, individual matter particles could not form atoms and molecules, the Universe could collapse on itself a few seconds after the Big Bang, stars could never start shining, or burn all their fuel and die within minutes, all molecular structures could be ripped apart immediately after their formation by radiation or gravitation, etc. There's always more ways in which things can go wrong than in which they can go right, and this holds especially for the Universe itself.

So why does the Universe exhibit this amazing numerical coincidence, making our wondering about it possible? There's several potential answers to this question (apart from intelligent design), of which I will list a few examples. Firstly, it is possible that many of what now seem to be free-but-fine-tuned constants (i.e. they could have taken any value but happen to have this specific one) could not in fact have had a different value, but this would only become apparent once we discover a more fundamental theory behind the models we have nowadays. Consider, in analogy, the mass of atoms: being unaware that atoms consist of discrete collections of more fundamental particles, their masses seem to be oddly fine-tuned to specific values rather than having a continuous spectrum. Another argument leverages the idea of observer bias: we could never have observed fundamental physical constants different from the ones we observe, because if they were different, we could not have existed and observed them. Potentially, this argument includes the idea of an ensemble of many parallel Universes where the fundamental constants take random combinations of values, but only in a tiny fraction of those conscious life could develop and wonder why their Universe is exactly the way it is[42]. A variant of this argument claims that the Universe in its early stages must have been a quantum-mechanical superposition of many configurations at once and that a single history or origin is impossible to define.

Fast-forward to the future. The Universe is almost certainly not going to stick around forever (in a way that supports any kind of what we would call life). There are various scenarios for how the Universe as we know it might come to an end, depending on which way the balance falls between contraction (driven by all matter in the Universe undergoing mutual gravitational interaction) and expansion (potentially driven by something called 'dark energy'). If there is a relatively large amount of matter in the Universe, attracting forces will start to prevail at some point in time, and the current expansion of the Universe will slow down, come to a standstill, and reverse into a contraction - first gradual, but ever more violently - up to a point where the extreme pressure, temperature, and radiation make life impossible. In this scenario, the sky will literally come falling down, as stars and galaxies crash into each other and are ripped apart, in this reversed Big Bang, sometimes called the Big Crunch. On the other hand, if there is too little matter, the expansion of the Universe will continue indefinitely up to the so-called heat death, a point where the last stars have been finally extinguished, and the Universe devolves into a cold, lightless, motionless, featureless nothingness. There are some other, more exotic scenarios such as the Big Rip, where all matter, even space-time itself, is finally ripped apart. But no matter which scenario our Universe will follow, it is clear that there will be a final end to everything that we could possibly care about.

[41] "So remember, when you're feeling very small and insecure, how amazingly unlikely is your birth; and pray that there's intelligent life somewhere out in space, cause there's bugger all down here on Earth!", *Monty Python*

[42] When you want to give your imagination a little challenge, try to picture the vastness of the expanse, the uncountable number of stars, galaxies, pulsars, quasars, neutron stars, black holes and other exotic objects in it, the near-endlessness of empty space in between, and the tiny fraction of that space taken by our little rock. When you succeed, then try to imagine infinitely many alternative Universes, in which literally everything works entirely different than in ours. Good luck!

This is, in a sense, a liberating realization: the heavy burden of eternity is replaced by the light feather of just a (very, very, very) long time, after which nothing we did or did not do will matter anymore. It is the bliss of total forgetfulness from the perspective of the Universe. It is obviously likely that humankind will be gone and erased from all memories long before that time, but it's still a comforting back-up option.

Our place in the Cosmos

We end up with a worldview that is both cold and comforting, familiar and terrifying, humiliating and uplifting. It all started from a bunch of abstract, shapeless, meaningless particles interacting in a stochastic-deterministic way; accidental random combinations of these particles turned out to be self-replicating; the evolutionary principle allowed these structures to grow more complicated and composite; some of these structures embedded external stimuli in intricate decision-making stratagems in the context of their struggle for life, as a side product of which the illusion of choice and consciousness arose. And so, here we are, after 14 billion years of evolving from an exploding fireball into these contingent and temporary but pretentious input-process-output systems, and there's absolutely no point to it all.

If this can teach us one thing, it's a sense of perspective. For a precious moment to get ripped out of our daily involvements of sticky life and see the bigger picture; to see all our first-world problems and successes, doubts and fears, amities and enmities, reduced to sheer trivialities not worth bothering about; to realize that there are so many factors beyond our control interfering with our lives that we might as well sit back relaxed and see where it all leads to; it's one of the greatest comforts I have ever felt (only momentarily unfortunately) and I hope I can share it with others.

We seem to be alone and we should face the consequences. Things can go wrong, and there is no indication whatsoever that at the end of days, some deity will come and make things all right again. Human beings are capable of causing tremendous suffering with each other, which remains totally unrecompensed; some bad people live long and happy lives, and many good people short and miserable ones. Our world, looked at from outside, is a very thin and fragile layer of mold on a tiny rock in an inconspicuous middle-of-nowhere. We are able, if not to utterly destroy it, than at least to make large parts of it very uncomfortable to live in, whether it is by nuclear arms, anthropogenic climate change, pollution, structural poverty, or other threats we cannot even conceive of yet.

How to live

The distinction is made at this point between the purposelessness of the random and material Universe, and the meaning people can create for their own life, for example in caring for those around them and contributing to happiness for all. While we have to accept the Universe we live in as an indifferent, random, absurd canvas that has randomly popped up from the void and will someday be gone again, so the argument goes, we can still choose how to paint it (our at least the tiny fraction of it that we inhabit). While we even understand that moral feelings of care, compassion and love are essentially electro-chemical processes in the brain that our species developed as tools in the struggle for survival (or at least as harmless byproducts of that development), that doesn't in itself mean those feelings should be disregarded as a solid basis for morality. By lack of greater cosmic purpose, we have to assert that our human happiness during the short span of our existence is a proper goal in itself, and commit to obtaining it for our fellow creatures (and usually, as a consequence, for ourselves as well).

While this is definitely the most reasonable, socially acceptable and practical view, there are a few problems with it. Firstly, it tends towards plain hedonism[43], i.e. the lazy disregard for long-term projects and the sanctification of the search for immediate and selfish sensory satisfaction as the greatest good. If transitory human happiness is indeed the one and only real basis for ethics, why not pursue one's own at the cost of others? Secondly, it is without proper theoretical or philosophical foundation. Taking stochastic-determinism seriously (by lack of proof for anything else), everything we ever think and feel and say and do is outside 'our' control, and rather predetermined by nature, nurture and chance. In its utter limit, the explicit question of 'how to live?' becomes pointless and ill-defined.

Addressing first the risk of selfish hedonism mentioned above: given that people are, at least to some degree, rational agents[44], that evaluate potential courses of actions by predicting their future effects, it is possible to develop a system of ethics that includes altruism and care for others based only on personal well-being as the deepest ground. In such a system, the fundamental argument is egocentric: what actions, or what behaviour in general, will lead, on average and on the long term, to the most beneficial reciprocal behaviour (or on an even more basic level, will give me and my offspring the highest probability for survival). It is also possible to see how these principles, which we can now explicitly reason about, were gradually built into our system of low-level instincts and intuitions (e.g. empathy) by the evolutionary mechanism. Hence, we often feel something to be right or wrong, and wrongly attribute it to some esoteric source of divine knowledge, while it is essentially just an internalized version of the categorical imperative[45]. The important point here is that purely material and essentially mechanical view of the Universe does not undermine ethical behaviour. On the contrary, it provides explanations and justifications for altruistic tendencies that are, after some thought, much more solid and comprehensible than those derived from some transcendent hypothesis, or even worse, from the fear of punishment in some afterlife.

Science, and particle physics in general, does unfortunately not provide a fully worked out, ready-made ethical framework. Yet it does provide a solid, rational and anthropocentric (even self-centric) foundation on which to build it. By lack of other criteria, personal happiness, and by generalized reciprocity, the subjective happiness of all current and future inhabitants of this planet should be the ultimate goal. By the realization that consciousness and subjectivity are not binary but gradual entities, this extends from people to animals (and further?) as well. This simple kernel can be the seed for altruism, the welfare state, care for the environment, and other characteristics typically thought of as ethical behaviour.

An important aspect of ethics derived from a scientific mindset is understanding. Understanding how each of us is to a certain degree pre-programmed to pursue their own desires. How the ancient reptile part of our brain will do anything to secure food, dominance and a reproductive partner, and how higher mental capacities such as reason and compassion, which are only secondary effects of the reptile brain growing more complex, are only partly able to subdue or mask these drives. This very understanding can help us have patience, care and compassion with those around even when they seem hostile to us at first sight. Moreover, the insight in the nature of our instincts can help us redirect their energies towards goals that they were not

[43]"Nobody exists on purpose. Nobody belongs anywhere. Everybody is gonna die. Come watch TV.", *Rick and Morty*

[44]'Rational' is meant very broadly here, as 'acting based on predicted outcomes'; the thinking process itself can of course vary from person to person, and might seem very irrational to other people.

[45]This argument, as well as some other ones in this essay, have been implicitly inspired by G. L. Drescher, *Good and Real: Demystifying Paradoxes from Physics to Ethics* [253].

originally meant for, such as explicit altruism[46].

The second criticism - that the question of how to live is pointless in itself since it is not our choice to make - is unfortunately more difficult to answer. It's not that you cannot change your life, moral behaviour, our worldview (for example, I might hope somewhat optimistically, by reading this text), only that you could not have done otherwise given who you are and the sensory inputs you receive; and those inputs, being part of the random-causal chain of the physical world, you received stochastically-deterministically. Any deviation from this pattern, however complicated, involves at some point or other a recourse to a source of the kind of freedom of choice that we have seemingly excluded. At this point I haven't gotten any further than the lazy solution of accepting some level of choice as a working hypothesis; a kind of ethical freedom that, despite at least seeming incompatible our metaphysics, we have to uphold in order to make sense of everyday life[47]. The details on how to reconcile this with the material world, I leave a an exercise to the reader.

The consolation of science

The conclusions of this (too long and yet far too short) essay are somewhat anticlimactic - not very surprising given its ambitious scope. The advances in modern high energy physics do not provide a new and revolutionary recipe for living a good and happy life, or for understanding the fortunes and misfortunes that befall us human beings. In fact, the remarks regarding our place in the cosmos and how to live, have been around for a long time[48].

What modern physics can do however, is put these intuitive truths in a new justification. The lack of ultimate purpose, the enslavement to cause and effect, and the metamorphosis of all things, are reformulated in the mathematical language of symmetries, quanta and fields, and (so far at least) verified experimentally. Many scattered insights about who we are and what we feel come together in a consistent framework, rooted in reality, with a large explanatory power, as we start to see the bigger picture between the birth and death of the Universe.

The consolation of science does not consist in discovering the meaning of it all, but rather in discovering that there is none. And hence, we recover a sense of freedom, that seemed to have gotten lost in the course of this essay; freedom in the sense of a free fall, after the floor of comfortable conventions and hard-wired illusions has been shattered under our feet.

[46] Loosely based on Schopenhauer's view of compassion as the basis for morality, as it shows us that other people are not really different from us: they have the same desires, plans and frustrations. And on his notion that the mind, while originally just a tool in the struggle for life, can rebel against its progenitor and (partly) silence the Will.

[47] With the consolation that rather illustrious philosophers such a Kant have, with more or less detours, done the same. Cf. 'Optimism is a moral duty'.

[48] Neatly summarized as: "Here's the meaning of life. Well, it's nothing very special. Try to be nice to people, avoid eating fat, read a good book every now and then, get some walking in, and try and live together in peace and harmony with people of all creeds and nations.", *Monty Python*

Appendix A

Sample lists

A.1 Sample lists for tZq

Table A.1: Simulated signal and background samples and their effective cross sections, as used in the tZq analysis (part 1). While the detector simulation changes per data-taking year to reflect the most up-to-date detector status, similar generator settings are used for all data-taking years for each sample in this analysis. For most samples with 2016 detector conditions, the older CUETP8M1 or CUETP8M2T4 underlying event tunes are used instead of the CP5 tune.

Sample name	$\sigma \times k^{NLO/LO}$ [pb]
tZq signal	
tZq_ll_4f_ckm_NLO_TuneCP5_13TeV-madgraph-pythia8	0.0942
WZ background	
WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8	5.063
ZZ/H background	
ZZTo4L_13TeV_powheg_pythia8	1.256
GluGluHToZZTo4L_M125_13TeV_powheg2_JHUGenV7011_pythia8	0.01212
VBF_HToZZTo4L_M125_13TeV_powheg2_JHUGenV7011_pythia8	0.001034
VHToNonbb_M125_13TeV_amcatnloFXFX_madspin_pythia8	0.952
GluGluToContinToZZTo2e2mu_13TeV_MCFM701_pythia8	0.0067
GluGluToContinToZZTo2e2tau_13TeV_MCFM701_pythia8	0.0067
GluGluToContinToZZTo2mu2tau_13TeV_MCFM701_pythia8	0.0067
GluGluToContinToZZTo4e_13TeV_MCFM701_pythia8	0.00334
GluGluToContinToZZTo4mu_13TeV_MCFM701_pythia8	0.00334
GluGluToContinToZZTo4tau_13TeV_MCFM701_pythia8	0.00334
$t\bar{t}$Z background	
TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8	0.2814
$t/t\bar{t}$+X background	
TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia	0.2043
ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8	0.2151
TTTT_TuneCP5_13TeV-amcatnlo-pythia8	0.009103
ST_tWll_5f_LO_TuneCP5_PSweights_13TeV-madgraph-pythia8	0.01103

Table A.2: Simulated signal and background samples and their effective cross sections, as used in the tZq analysis (part 2). While the detector simulation changes per data-taking year to reflect the most up-to-date detector status, similar generator settings are used for all data-taking years for each sample in this analysis. For most samples with 2016 detector conditions, the older CUETP8M1 or CUETP8M2T4 underlying event tunes are used instead of the CP5 tune.

Sample name	$\sigma \times k^{NLO/LO}$ [pb]
Multiboson background	
ZZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.01398
WZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.05565
WWZ_TuneCP5_13TeV-amcatnlo-pythia8	0.1651
WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8	0.2086
X+γ background	
TGJets_TuneCP5_13TeV_amcatnlo_madspin_pythia8	2.967
TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	3.697
WGToLNuG_TuneCP5_13TeV-madgraphMLM-pythia8	489
ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8	59.21
Nonprompt background	
DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8	6077.22
DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	18610
WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8	61526.7
TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	87.31
TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	182.75
ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8	3.36
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	136.0
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	80.96
ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
WWTo2L2Nu_DoubleScattering_13TeV-pythia8	0.1729
WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8	12.178

A.2 Sample lists for ttW

Table A.3: Simulated signal and background samples and their effective cross sections, as used in the $t\bar{t}W$ analysis (part 1). While the detector simulation changes per data-taking year to reflect the most up-to-date detector status, the same generator settings are used for all data-taking years for each sample in this analysis.

Sample name	Cross section [pb]
$t\bar{t}W$ signal	
TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	0.2269
ttWJetsToLNu_5f_EWK_TuneCP5_13TeV_amcatnlo-pythia8	0.016
TTH	
ttHJetTobb_M125_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8	0.4040
ttHJetToNonbb_M125_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8	0.2914974
TTZ	
TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8	0.28872
TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8	0.662
TTZToLL_M-1to10_TuneCP5_13TeV-amcatnlo-pythia8	0.08416
TTG	
TTGamma_Dilept_TuneCP5_13TeV-madgraph-pythia8	2.22
TTGamma_Hadronic_TuneCP5_13TeV-madgraph-pythia8	6.162
TTGamma_SingleLept_TuneCP5_13TeV-madgraph-pythia8	7.509
TGJets_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8	1.02
TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	88.40
TTToHadronic_TuneCP5_13TeV-powheg-pythia8	377.96
TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	365.34
TTX	
TTTT_TuneCP5_13TeV-amcatnlo-pythia8	0.01197
TTHH_TuneCP5_13TeV-madgraph-pythia8	0.0006655
TTWH_TuneCP5_13TeV-madgraph-pythia8	0.001141
TTWW_TuneCP5_13TeV-madgraph-pythia8	0.007003
TTWZ_TuneCP5_13TeV-madgraph-pythia8	0.002453
TTZH_TuneCP5_13TeV-madgraph-pythia8	0.00113
TTZZ_TuneCP5_13TeV-madgraph-pythia8	0.001386
TTTJ_TuneCP5_13TeV-madgraph-pythia8	0.0003974
TTTW_TuneCP5_13TeV-madgraph-pythia8	0.0007314
TX	
ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8	3.74
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	136.02
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	80.95
ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	21.585204
ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	21.585204
THQ_ctcvcp_4f_Hincl_TuneCP5_13TeV_madgraph_pythia8	0.07096
tZq_ll_4f_ckm_NLO_TuneCP5_13TeV-amcatnlo-pythia8	0.07358
TWZToLL_thad_Wlept_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8	0.003
TWZToLL_tlept_Whad_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8	0.003
TWZToLL_tlept_Wlept_5f_DR_TuneCP5_13TeV-amcatnlo-pythia8	0.0015
THW_ctcvcp_5f_Hincl_TuneCP5_13TeV_madgraph_pythia8	0.01561

Table A.4: Simulated signal and background samples and their effective cross sections as used in the $t\bar{t}W$ analysis (part 2). While the detector simulation changes per data-taking year to reflect the most up-to-date detector status, the same generator settings are used for all data-taking years for each sample in this analysis.

Sample name	Cross section [pb]
WZ	
WZTo1L1Nu2Q_TuneCP5_13TeV-amcatnloFXFX-pythia8	11.66
WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8	4.9173
WZJJ_EWK_InclusivePolarization_TuneCP5_13TeV_madgraph-madspin-pythia8	0.01629
ZZ	
ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8	0.6204
ZZTo4L_TuneCP5_13TeV_powheg_pythia8	1.256
GluGluHToZZTo4L_M125_TuneCP5_13TeV_powheg2_JHUGenV7011_pythia8	0.01212
GluGluToContinToZZTo2e2mu_TuneCP5_13TeV-mcfm701-pythia8	0.005423
GluGluToContinToZZTo2e2nu_TuneCP5_13TeV-mcfm701-pythia8	0.005423
GluGluToContinToZZTo2e2tau_TuneCP5_13TeV-mcfm701-pythia8	0.005423
GluGluToContinToZZTo2mu2nu_TuneCP5_13TeV-mcfm701-pythia8	0.005423
GluGluToContinToZZTo2mu2tau_TuneCP5_13TeV-mcfm701-pythia8	0.005423
GluGluToContinToZZTo4e_TuneCP5_13TeV-mcfm701-pythia8	0.0027
GluGluToContinToZZTo4mu_TuneCP5_13TeV-mcfm701-pythia8	0.0027
GluGluToContinToZZTo4tau_TuneCP5_13TeV-mcfm701-pythia8	0.0027
VBF_HToZZTo4L_M125_TuneCP5_13TeV_powheg2_JHUGenV7011_pythia8	0.001034
VHToNonbb_M125_TuneCP5_13TeV-amcatnloFXFX_madspin_pythia8	0.9561
ZG	
ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8	55.48
DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	15810.0
DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8	6077.22
Multiboson	
WWTo1L1Nu2Q_TuneCP5_13TeV-amcatnloFXFX-pythia8	45.68
WWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	11.09
WWTo4Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8	45.2
WWTo2L2Nu_TuneCP5_DoubleScattering_13TeV-pythia8	0.2232
WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8	0.2086
WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8	0.1651
WZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.05565
ZZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.01476

Appendix B

Triggers

B.1 Triggers used in the fake-rate measurement

Table B.1: Overview of the triggers used in the fake rate measurement. These triggers select events with either a muon ('Mu') or an electron ('Ele') with a transverse momentum above a given numerical threshold. For some triggers, the presence of an additional jet is also required.

Trigger	2016 data	2017 data	2018 data
HLT_Mu3_PFJet40	✓	✓	✓
HLT_Mu8	✓	✓	✓
HLT_Mu17	✓	✓	✓
HLT_Mu20	✓	✓	✓
HLT_Mu27	✓	✓	✓
HLT_Ele8_CaloIdM_TrackIdM_PFJet30	✓	✓	✓
HLT_Ele12_CaloIdM_TrackIdM_PFJet30	✓		
HLT_Ele17_CaloIdM_TrackIdM_PFJet30	✓	✓	✓
HLT_Ele23_CaloIdM_TrackIdM_PFJet30	✓	✓	✓

B.2 Lepton triggers used in the displaced vertex, tZq, and ttW analyses

Table B.2: Overview of single, double and triple lepton triggers used in the analyses presented in this dissertation. For the tZq and $t\bar{t}W$ analyses, all lepton triggers in the table below were used; for the displaced tracking and vertexing calibration study, only the dimuon triggers ($\mu\ \mu$) were used.

Trigger	2016 data	2017 data	2018 data
μ			
HLT_IsoMu24	✓	✓	✓
HLT_IsoMu24_eta2p1		✓	
HLT_IsoTkMu24	✓		
HLT_IsoMu27		✓	✓
HLT_Mu45_eta2p1	✓		
HLT_Mu50	✓	✓	✓
HLT_TkMu50	✓		
HLT_OldMu100		✓	✓
HLT_TkMu100		✓	✓
e			
HLT_Ele27_WPTight_Gsf	✓		
HLT_Ele32_WPTight_Gsf		✓	✓
HLT_Ele35_WPTight_Gsf		✓	
HLT_Ele105_CaloIdVT_GsfTrkIdT	✓		
HLT_Ele115_CaloIdVT_GsfTrkIdT	✓	✓	✓
HLT_Photon175	✓		
HLT_Photon200		✓	✓
$\mu\ \mu$			
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL	✓	✓	
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ	✓	✓	
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL	✓		
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ	✓		
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8		✓	
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8		✓	✓
HLT_TkMu17_TrkIsoVVL_TkMu8_TrkIsoVVL	✓		
HLT_TkMu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ	✓		
HLT_Mu30_TkMu11	✓		
HLT_Mu37_TkMu27		✓	✓
e μ			
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL	✓		
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	✓	✓	✓
HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ		✓	✓
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL	✓		
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ	✓		
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL		✓	✓
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ		✓	✓
HLT_Mu30_Ele30_CaloIdL_GsfTrkIdVL	✓		
HLT_Mu33_Ele33_CaloIdL_GsfTrkIdVL	✓		
HLT_Mu27_Ele37_CaloIdL_MW		✓	✓
HLT_Mu37_Ele27_CaloIdL_MW		✓	✓
e e			
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL		✓	✓
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	✓		
HLT_DoubleEle33_CaloIdL_GsfTrkIdVL	✓		
HLT_DoubleEle33_CaloIdL_GsfTrkIdVL_MW	✓		
HLT_DoubleEle25_CaloIdL_MW			✓
HLT_DoubleEle33_CaloIdL_MW		✓	
$\mu\ \mu\ \mu$			
HLT_TripleMu_5_3_3_Mass3p8to60_DZ		✓	✓
HLT_TripleMu_10_5_5_DZ		✓	✓
HLT_TripleMu_12_10_5	✓	✓	✓
e $\mu\ \mu$			
HLT_DiMu9_Ele9_CaloIdL_TrackIdL	✓		
HLT_DiMu9_Ele9_CaloIdL_TrackIdL_DZ		✓	✓
e e μ			
HLT_Mu8_DiEle12_CaloIdL_TrackIdL	✓	✓	✓
e e e			
HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL	✓	✓	✓

Bibliography

[1] CMS Collaboration, "Inclusive and differential cross section measurements of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **02** (2022) 107, doi: https://doi.org/10.1007/JHEP02(2022)107.

[2] CMS Collaboration, "Measurements of the differential cross-section and charge asymmetry of a top quark pair in association with a W boson in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". `https://cms.cern.ch/iCMS/analysisadmin/cadilines?id=2772&ancode=TOP-24-003&tp=an&line=TOP-24-003`. (in preparation, currently in CMS internal review).

[3] CMS Collaboration, "Measurement of the associated production of a single top quark and a Z boson in pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Lett. B* **779** (2018) 358, doi: 10.1016/j.physletb.2018.02.025, `arXiv:1712.02825`.

[4] CMS Collaboration, "Observation of four top quark production in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Physics Letters B* **847** (2023) 138290, doi: https://doi.org/10.1016/j.physletb.2023.138290.

[5] CMS Collaboration, "Search for heavy neutral leptons in final states with electrons, muons, and hadronically decaying tau leptons in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Journal of High Energy Physics* **2024** (Jun, 2024) doi: http://doi.org/10.1007/JHEP06(2024)123.

[6] CMS Collaboration, "Displaced tracking and vertexing calibration using neutral K mesons". CERN-CMS-DP-2024-010.

[7] CMS Collaboration, "Search for long-lived heavy neutral leptons with displaced vertices in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **07** (2022) 081, doi: https://doi.org/10.1007/JHEP07(2022)081.

[8] CMS Collaboration, "Search for long-lived heavy neutral leptons in proton-proton collision events with a lepton-jet pair associated with a secondary vertex at $\sqrt{s} = 13\,\text{TeV}$", 2024. (submitted to JHEP). `https://arxiv.org/abs/2407.10717`.

[9] CMS Collaboration, "Prospects for computer-assisted data quality monitoring at the CMS pixel detector". CERN-CMS-DP-2022-013.

[10] L. Lambrecht on behalf of the CMS Collaboration, "Prospects for computer-assisted data quality monitoring at the CMS pixel detector", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **1045** (2023) 167495, doi: https://doi.org/10.1016/j.nima.2022.167495.

[11] CMS Collaboration, "Tracker DQM machine learning studies for data certification". CERN-CMS-DP-2021-034, 2021.

[12] CMS Collaboration, "CMS tracker data quality certification with new machine learning tools". CERN-CMS-DP-2024-070, 2024.

[13] V. Wachirapusitanand on behalf of the CMS Collaboration, "Machine Learning applications for Data Quality Monitoring and Data Certification within CMS". CMS-CR-2022-014, 2022.

[14] CMS Collaboration, "Measurement of single top quark production in association with a Z boson". CMS-AN-20-170. (CMS internal).

[15] CMS Collaboration, "Differential cross section measurements of $t\bar{t}W$ production at $\sqrt{s} = 13\,\text{TeV}$". CMS-AN-23-077. (CMS internal).

[16] CMS Collaboration, "Study of displaced vertex reconstruction using light neutral hadrons in Drell-Yan events". CMS-AN-20-111. (CMS internal).

[17] R. Feynman, "The Meaning of it All - Thoughts of a Citizen-Scientist". Perseus Books, Reading, Massachusetts, 1998. ISBN 0-7382-0166-9.

[18] H. Schopper, "LEP: The lord of the collider rings at CERN 1980-2000: The making, operation and legacy of the world's largest scientific instrument". Springer, 2009. doi: 10.1007/978-3-540-89301-1, ISBN 978-3-540-89300-4.

[19] H. Stevens, "Fundamental physics and its justifications, 1945-1993", *Historical Studies in the Physical and Biological Sciences* **34** (2003), no. 1, 151–197, doi: https://doi.org/10.1525/hsps.2003.34.1.151.

[20] V. Weisskopf, "Physics in the twentieth century: selected essays". MIT Press, 1972. ISBN 0-262-23056-9.

[21] K. Kaplan, "Unmasking the impostor", *Nature* **459** (2009) 468–469, doi: https://doi.org/10.1038/nj7245-468a.

[22] M. Mangano et al., "FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1. Future Circular Collider", Technical Report 6, CERN, Geneva, 2019. doi: 10.1140/epjc/s10052-019-6904-3.

[23] CERN (edited by M. Mangano), "CERN Yellow Reports: Monographs, Vol 3 (2017): Physics at the FCC-hh, a 100 TeV pp collider", 2017. doi: 10.23731/CYRM-2017-003.

[24] F. Mandl and G. Shaw, "Quantum field theory, 2nd edition". John Wiley and Sons Ltd, 2010. ISBN 978-0-471-49683-0.

[25] A. Bettini, "Introduction to Elementary Particle Physics (2nd ed.)". Cambridge University Press, 2014. doi: doi:10.1017/CBO9781107279483.

[26] M. E. Peskin and D. V. Schroeder, "An Introduction to Quantum Field Theory". Westview Press, 1995. Reading, USA: Addison-Wesley (1995) 842 p.

[27] A. Arbuzov, "Quantum Field Theory and the Electroweak Standard Model". link, 2018.

[28] B. Povh et al., "Particles and Nuclei - An Introduction to the Physical Concepts". Springer Berlin, Heidelberg, 2015. doi: https://doi.org/10.1007/978-3-662-46321-5, ISBN 978-3-662-46320-8.

[29] J. M. Maldacena, "Gravity, particle physics and their unification", *International Journal of Modern Physics A* **15** (Jul, 2000) 840–852, doi: 10.1142/s0217751x00005449.

[30] The EROS-2 Collaboration, "Limits on the Macho content of the Galactic Halo from the EROS-2 Survey of the Magellanic Clouds", *Astronomy and Astrophysics* **469** (Jul, 2007) 387–404, doi: https://doi.org/10.1051/0004-6361:20066017.

[31] T. D. Brandt, "Constraints on macho dark matter from compact stellar systems in ultra-faint dwarf galaxies", *The Astrophysical Journal Letters* **824** (Jun, 2016) L31, doi: 10.3847/2041-8205/824/2/L31.

[32] G. Arcadi et al., "The waning of the WIMP? A review of models, searches, and constraints", *The European Physical Journal C* **78** (2018) 203, doi: 10.1140/epjc/s10052-018-5662-y.

[33] M. Gonzalez-Garcia, M. Maltoni, and T. Schwetz, "Global analyses of neutrino oscillation experiments", *Nuclear Physics B* **908** (2016) 199–217, doi: https://doi.org/10.1016/j.nuclphysb.2016.02.033.

[34] G. R. Bengochea, G. León, E. Okon, and D. Sudarsky, "Can the quantum vacuum fluctuations really solve the cosmological constant problem?", *The European Physical Journal C* **80** (Jan, 2020) doi: 10.1140/epjc/s10052-019-7554-1.

[35] M. B. Gavela, P. Hernandez, J. Orloff, and O. Pene, "Standard model CP-violation and baryon asymmetry", *Modern Physics Letters A* **09** (Mar, 1994) 795–809, doi: 10.1142/s0217732394000629.

[36] P. Huet and E. Sather, "Electroweak baryogenesis and standard model CP violation", *Phys. Rev. D* **51** (Jan, 1995) 379–394, doi: 10.1103/PhysRevD.51.379.

[37] CDF Collaboration, "High-precision measurement of the W boson mass with the CDF II detector", *Science* **376** (2022), no. 6589, 170–176, doi: 10.1126/science.abk1781, `arXiv:https://www.science.org/doi/pdf/10.1126/science.abk1781`.

[38] LHCb Collaboration, "Measurement of the W boson mass", *Journal of High Energy Physics* **2022** (Jan, 2022) doi: 10.1007/jhep01(2022)036.

[39] ATLAS Collaboration, "Measurement of the W-boson mass and width with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 7$ TeV", 2024. doi: https://doi.org/10.48550/arXiv.2403.15085.

[40] CMS Collaboration, "Measurement of the W boson mass in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". CMS-PAS-SMP-23-002, 2024.

[41] Muon g-2 Collaboration, "Measurement of the positive muon anomalous magnetic moment to 0.46 ppm", *Phys. Rev. Lett.* **126** (Apr, 2021) 141801, doi: 10.1103/PhysRevLett.126.141801.

[42] Muon g-2 Collaboration, "Measurement of the positive muon anomalous magnetic moment to 0.20 ppm", *Phys. Rev. Lett.* **131** (Oct, 2023) 161802, doi: 10.1103/PhysRevLett.131.161802.

[43] D. Castelvecchi, "Dreams of new physics fade with latest muon magnetism result", *Nature* **620** (2023), no. 7974, 473–474.

[44] F. Ignatov et al., "Measurement of the $Pe^+e^- \to \pi^+\pi^-$ cross section from threshold to 1.2 GeV with the CMD-3 detector", *Physical Review D* **109** (Jun, 2024) doi: 10.1103/physrevd.109.112002.

[45] LHCb Collaboration, "Test of lepton universality in beauty-quark decays", *Nature Physics* **18** (Mar, 2022) 277–282, doi: 10.1038/s41567-021-01478-8.

[46] LHCb Collaboration, "Test of Lepton Universality in $b \to s\ell^+\ell^-$ Decays", *Phys. Rev. Lett.* **131** (Aug, 2023) 051803, doi: 10.1103/PhysRevLett.131.051803.

[47] LHCb collaboration, "Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays", Aug, 2023. doi: 10.1103/PhysRevD.108.032002.

[48] CMS Collatoration, "The CMS experiment at the CERN LHC", *Journal of Instrumentation* **3** (Aug, 2008) S08004, doi: 10.1088/1748-0221/3/08/S08004.

[49] The Tracker Group of the CMS Collaboration, "The CMS Phase-1 Pixel Detector Upgrade". CERN-CMS-NOTE-2020-005, 2020.

[50] CMS Collaboration, "Radiofrequency cavities". link, 2012.

[51] L. Rossi, "The LHC main dipoles and quadrupoles toward series production", *IEEE Transactions on Applied Superconductivity* **13** (2003), no. 2, 1221–1228, doi: 10.1109/TASC.2003.812639.

[52] E. Lopienska, "The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022". CERN-GRAPHICS-2022-001, 2022.

[53] W. Herr and B. Muratori, "Concept of luminosity", doi: 10.5170/CERN-2006-002.361.

[54] O. Brüning et al., "The scientific potential and technological challenges of the high-luminosity large hadron collider program", *Reports on Progress in Physics* **85** (Mar, 2022) 046201, doi: 10.1088/1361-6633/ac5106.

[55] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data", *Journal of Instrumentation* **15** (Sep, 2020) P09018, doi: 10.1088/1748-0221/15/09/P09018.

[56] C. Joram et al., "Particle detectors: principles and techniques. Detector techniques". link, 2005. CERN, Geneva, 11 - 15 Apr 2005.

[57] M. Silvia, "Electromagnetic and hadronic calorimeters". 39th Heidelberg Physics Graduate Days, HGSFP Heidelberg, October 12, 2017.

[58] Particle Data Group Collaboration, "Review of Particle Physics", *PTEP* **2022** (2022) 083C01, doi: 10.1093/ptep/ptac097.

[59] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *Journal of Instrumentation* **9** (Oct, 2014) P10009, doi: 10.1088/1748-0221/9/10/P10009.

[60] CMS Collaboration, "Track impact parameter resolution for the full pseudo rapidity coverage in the 2017 dataset with the CMS Phase-1 Pixel detector". CERN-CMS-DP-2020-049, 2020.

[61] CMS Collaboration, "The Phase-2 Upgrade of the CMS Tracker", technical report, CERN, Geneva, 2017. doi: 10.17181/CERN.QZ28.FLHW.

[62] A. Benaglia, "The CMS ECAL performance with examples", *Journal of Instrumentation* **9** (Feb, 2014) C02008, doi: 10.1088/1748-0221/9/02/C02008.

[63] CMS Collaboration, "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7\,\text{TeV}$", *Journal of Instrumentation* **8** (Sep, 2013) P09009–P09009, doi: 10.1088/1748-0221/8/09/p09009.

[64] F. Ferri, "The CMS ECAL Phase-2 upgrade for high precision energy and timing measurements", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **958** (2020) 162159, doi: https://doi.org/10.1016/j.nima.2019.04.113. Proceedings of the Vienna Conference on Instrumentation 2019.

[65] CMS Collaboration, "Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13\,\text{TeV}$", *Journal of Instrumentation* **15** (May, 2020) P05002–P05002, doi: 10.1088/1748-0221/15/05/p05002.

[66] CMS Collaboration, "Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data", *Journal of Instrumentation* **5** (Mar, 2010) T03012–T03012, doi: 10.1088/1748-0221/5/03/t03012.

[67] S. Calzaferri, "Production and installation of the first GEM station in CMS", *Journal of Instrumentation* **15** (Sep, 2020) C09040, doi: 10.1088/1748-0221/15/09/C09040.

[68] T. C. collaboration, "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7\,\text{TeV}$ at the LHC", *Journal of Instrumentation* **8** (Nov, 2013) P11002–P11002, doi: 10.1088/1748-0221/8/11/p11002.

[69] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Journal of Instrumentation* **13** (Jun, 2018) P06015–P06015, doi: 10.1088/1748-0221/13/06/p06015.

[70] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Physics Letters B* **716** (Sep, 2012) 30–61, doi: 10.1016/j.physletb.2012.08.021.

[71] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Physics Letters B* **716** (Sep, 2012) 1–29, doi: 10.1016/j.physletb.2012.08.020.

[72] CMS Collaboration, "Data Scouting and Data Parking with the CMS High level Trigger", *PoS* **EPS-HEP2019** (2020) 139, doi: 10.22323/1.364.0139.

[73] CMS Collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Journal of Instrumentation* **15** (Oct, 2020) P10017, doi: 10.1088/1748-0221/15/10/P10017.

[74] CMS Collaboration, "The CMS trigger system", *Journal of Instrumentation* **12** (Jan, 2017) P01020, doi: 10.1088/1748-0221/12/01/P01020.

[75] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Journal of Basic Engineering* **82** (Mar, 1960) 35–45, doi: 10.1115/1.3662552.

[76] R. Frühwirth, "Application of kalman filtering to track and vertex fitting", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **262** (1987), no. 2, 444–450, doi: https://doi.org/10.1016/0168-9002(87)90887-4.

[77] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *Journal of Instrumentation* **12** (Oct, 2017) P10003–P10003, doi: 10.1088/1748-0221/12/10/p10003.

[78] D. Contardo et al., "Technical Proposal for the Phase-II Upgrade of the CMS Detector", technical report, CERN, Geneva, 2015. doi: 10.17181/CERN.VU8I.D59J.

[79] A. Perotta on behalf of the CMS Collaboration, "CMS event reconstruction status in Run 2", *EPJ Web of Conferences* **214** (Sep, 2019) doi: https://doi.org/10.1051/epjconf/201921402015.

[80] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC", *Journal of Physics G: Nuclear and Particle Physics* **31** (Jul, 2005) N9, doi: 10.1088/0954-3899/31/9/N01.

[81] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-kt jet clustering algorithm", *Journal of High Energy Physics* **2008** (Apr, 2008) 063–063, doi: 10.1088/1126-6708/2008/04/063.

[82] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *Journal of Instrumentation* **12** (Feb, 2017) P02014–P02014, doi: 10.1088/1748-0221/12/02/p02014.

[83] CMS Collaboration Collaboration, "Measurements of the CMS jet energy scale and resolution at 13 TeV", Technical Report CMS-CR-2018-216, CERN, Geneva, Sep, 2018. doi: 10.22323/1.340.0688.

[84] CMS Collaboration, CMS Collaboration, "Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018". CERN-CMS-DP-2020-019, 2020.

[85] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *Journal of Instrumentation* **13** (May, 2018) P05011–P05011, doi: 10.1088/1748-0221/13/05/p05011.

[86] E. Bols et al., "Jet flavour classification using DeepJet", *Journal of Instrumentation* **15** (Dec, 2020) P12012–P12012, doi: 10.1088/1748-0221/15/12/p12012.

[87] CMS Collaboration Collaboration, CMS Collaboration, "Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13 TeV with Phase 1 CMS detector". CMS-DP-2018-058, Nov, 2018.

[88] A. Buckley et al., "General-purpose event generators for lhc physics", *Physics Reports* **504** (Jul, 2011) 145–233, doi: 10.1016/j.physrep.2011.03.005.

[89] F. Maltoni, T. McElmurry, R. Putman, and S. Willenbrock, "Choosing the factorization scale in perturbative QCD", 2007. doi: https://doi.org/10.48550/arXiv.hep-ph/0703156.

[90] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi: 10.1007/jhep04(2015)040.

[91] NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J. C* **77** (2017) 663, doi: 10.1140/epjc/s10052-017-5199-5.

[92] M. Dasgupta et al., "Parton showers beyond leading logarithmic accuracy", *Physical Review Letters* **125** (Jul, 2020) doi: 10.1103/physrevlett.125.052002.

[93] T. Sjöstrand et al., "An introduction to PYTHIA 8.2", *Computer Physics Communications* **191** (2015) 159 – 177, doi: https://doi.org/10.1016/j.cpc.2015.01.024.

[94] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA 8 tunes from underlying-event measurements", *Eur. Phys. J. C* **80** (2020) doi: 10.1140/epjc/s10052-019-7499-4.

[95] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 tune", *Eur. Phys. J. C* **74** (2014) doi: 10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.

[96] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) doi: 10.1140/epjc/s10052-016-3988-x.

[97] S. Höche, "Introduction to parton-shower event generators", 2015. doi: https://doi.org/10.48550/arXiv.1411.4085.

[98] GEANT4 Collaboration, "GEANT4 – a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi: 10.1016/S0168-9002(03)01368-8.

[99] CMS Collaboration, "CMS Simulation Software", technical report, CERN, Geneva, 2012. doi: 10.1088/1742-6596/396/2/022003.

[100] S. Sekmen, "Recent developments in cms fast simulation", 2017. doi: https://doi.org/10.48550/arXiv.1701.03850.

[101] S. Bein et al., "Refining fast simulation using machine learning", 2023. doi: https://doi.org/10.48550/arXiv.2309.12919.

[102] A. Samalan et al., "Upgrade of the CMS resistive plate chambers for the high luminosity LHC", *Journal of Instrumentation* **17** (Jan, 2022) C01011, doi: 10.1088/1748-0221/17/01/C01011.

[103] CMS Collaboration, "Search for new physics with compressed mass spectra in final states with soft prompt and displaced leptons and missing transverse energy at 13 TeV with the full Run II luminosity". https://cms.cern.ch/iCMS/analysisadmin/cadilines?line=EXO-23-017&tp=an&id=2755&ancode=EXO-23-017, 2023. (in preparation, currently in CMS internal review).

[104] CMS Collaboration, "Search for long-lived particles decaying to leptons with large impact parameter in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *The European Physics Journal C* **82** (2022), no. 153, doi: https://doi.org/10.1140/epjc/s10052-022-10027-3.

[105] CMS Collaboration, "Search for long-lived particles decaying into displaced jets in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Physical Review D* **99** (Feb, 2019) doi: 10.1103/physrevd.99.032011.

[106] CMS Collaboration, "Search for long-lived particles using nonprompt jets and missing transverse momentum with proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Physics Letters B* **797** (Oct, 2019) 134876, doi: 10.1016/j.physletb.2019.134876.

[107] CMS Collaboration, "Search for long-lived particles using delayed photons in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Physical Review D* **100** (Dec, 2019) doi: 10.1103/physrevd.100.112003.

[108] CMS Collaboration, "A deep neural network to search for new long-lived particles decaying to jets", *Machine Learning: Science and Technology* **1** (Aug, 2020) 035012, doi: 10.1088/2632-2153/ab9023.

[109] CMS Collaboration, "Search for long-lived particles decaying to jets with displaced vertices in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Physical Review D* **104** (Sep, 2021) doi: 10.1103/physrevd.104.052011.

[110] CMS Collaboration, "Search for long-lived particles using displaced jets in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Physical Review D* **104** (Jul, 2021) doi: 10.1103/physrevd.104.012015.

[111] CMS Collaboration, "Search for long-lived particles produced in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Journal of High Energy Physics* **2022** (Mar, 2022) doi: 10.1007/jhep03(2022)160.

[112] CMS Collaboration, "Search for long-lived particles decaying into muon pairs in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ collected with a dedicated high-rate data stream", *Journal of High Energy Physics* (2022) doi: https://doi.org/10.1007/JHEP04(2022)062.

[113] CMS Collaboration, "Search for Long-Lived Particles Decaying in the CMS End Cap Muon Detectors in Proton-Proton Collisions at $\sqrt{s} = 13\,\text{TeV}$", *Physical Review Letters* **127** (Dec, 2021) doi: 10.1103/physrevlett.127.261804.

[114] CMS Collaboration, "Search for long-lived particles decaying to a pair of muons in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Journal of High Energy Physics* **2023** (May, 2023) doi: 10.1007/jhep05(2023)228.

[115] CMS Collaboration, "Search for long-lived particles using out-of-time trackless jets in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Journal of High Energy Physics* **2023** (Jul, 2023) doi: 10.1007/jhep07(2023)210.

[116] CMS Collaboration, "Search for long-lived heavy neutral leptons with lepton flavour conserving or violating decays to a jet and a charged lepton", 2024. doi: https://doi.org/10.48550/arXiv.2312.07484.

[117] CMS Collaboration, "Search for long-lived particles decaying in the CMS muon detectors in proton-proton collisions at $\sqrt{s} = 13$ TeV", 2024. doi: https://doi.org/10.48550/arXiv.2402.01898.

[118] CMS Collaboration, CMS Collaboration, "Search for long-lived particles decaying to a pair of muons in pp collisions at $\sqrt{s} = 13.6\,\text{TeV}$ with 2022 data". CMS-PAS-EXO-23-014, 2023.

[119] CMS Collaboration, "CMS Tracking POG Performance Plots For 2017 with PhaseI pixel detector". CERN-CMS-DP-2017-015, 2017.

[120] CMS Collaboration, "Muon Reconstruction and Identification Improvements for Run-2 and First Results with 2015 Run Data". CERN-CMS-DP-2015-15, 2015.

[121] CMS collaboration, "Measurement of the inclusive W and Z production cross sections in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ with the CMS experiment", *Journal of High Energy Physics* (2011) doi: https://doi.org/10.1007/JHEP10(2011)132.

[122] ATLAS Collaboration, "Search for displaced vertices of oppositely charged leptons from decays of long-lived particles in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector", *Physics Letters B* **801** (Feb, 2020) 135114, doi: 10.1016/j.physletb.2019.135114.

[123] ATLAS Collaboration, "Search for heavy neutral leptons in decays of W bosons produced in 13 TeV pp collisions using prompt and displaced signatures with the ATLAS detector", *Journal of High Energy Physics* **2019** (Oct, 2019) doi: 10.1007/jhep10(2019)265.

[124] ATLAS Collaboration, "Search for long-lived, massive particles in events with a displaced vertex and a displaced muon in *pp* collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector", Mar, 2019.

[125] Olive et al. (Particle Data Group). Chin. Phys. C,38, 090001 (2014) and 2015 update, `http://pdg.lbl.gov/2015/tables/rpp2015-tab-mesons-strange.pdf`.

[126] J. Beringer et al. (Particle Data Group). PRD86, 010001 (2012), `http://pdg.lbl.gov/2012/tables/rpp2012-tab-baryons-Lambda.pdf`.

[127] CMS Collaboration, CMS Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period". CMS-PAS-LUM-17-001, 2017.

[128] CMS Collaboration, CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV". CMS-PAS-LUM-17-004, 2018.

[129] CMS Collaboration, CMS Collaboration, "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV". CMS-PAS-LUM-18-002, 2019.

[130] M. Jansova, "Search for the supersymmetric partner of the top quark and measurements of cluster properties in the silicon strip tracker of the CMS experiment at Run 2". link, 2018. Presented 27 Sep 2018.

[131] E. Butz, "Mitigation of the Strip Tracker Dynamic Inefficiency (previously known as HIP)". link. presented at the CMS WGM 278, 18 Aug. 2016 (CMS internal).

[132] CMS Collaboration, CMS, "Simulation of the Silicon Strip Tracker pre-amplifier in early 2016 data". CERN-CMS-DP-2020-045, 2020.

[133] CMS Collaboration, "CMS Tracker Performance results for full Run 2 Legacy reprocessing". CERN-CMS-DP-2020-012, 2020.

[134] CMS Collaboration, "Strategies and performance of the CMS silicon tracker alignment during LHC Run 2", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **1037** (Aug, 2022) 166795, doi: 10.1016/j.nima.2022.166795.

[135] 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, doi: 10.1007/JHEP07(2014)079, `arXiv:1405.0301`.

[136] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi: 10.1007/JHEP12(2012)061, `arXiv:1209.6215`.

[137] W. Erdmann, "Vertex reconstruction at the CMS experiment", *Journal of Physics: Conference Series* **110** (May, 2008) 092009, doi: 10.1088/1742-6596/110/9/092009.

[138] CMS Collaboration, "Strange particle production in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV", *Journal of High Energy Physics* **2011** (May, 2011) doi: 10.1007/jhep05(2011)064.

[139] A. Venturi, "CMS Alignment Tracking and Physics Performance Results", *PoS* **VERTEX 2010** (2011) 016, doi: 10.22323/1.113.0016.

[140] CMS Collaboration, "J/Psi prompt and non-prompt cross sections in pp collisions at $\sqrt{s} = 7$ TeV". CMS-PAS-BPH-10-002, 2010.

[141] V. Azzolini et al., "The data quality monitoring software for the CMS experiment at the LHC: past, present and future", *EPJ Web of Conferences* **214** (Sep, 2019) doi: https://doi.org/10.1051/epjconf/201921402003.

[142] M. Rovere on behalf of the CMS Collaboration, "The data quality monitoring software for the CMS experiment at the LHC", *Journal of Physics: Conference Series* **664** (Dec, 2015) 072039, doi: 10.1088/1742-6596/664/7/072039.

[143] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning". MIT Press, 2016.

[144] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders", 2020. doi: 10.48550/ARXIV.2003.05991.

[145] D. P. Kingma and M. Welling, "An introduction to variational autoencoders", *Foundations and Trends in Machine Learning* **12** (2019), no. 4, 307–392, doi: 10.1561/2200000056.

[146] B. Vormwald, "Operational experience of the phase-1 cms pixel detector", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **958** (2020) 162018, doi: https://doi.org/10.1016/j.nima.2019.03.073. Proceedings of the Vienna Conference on Instrumentation 2019.

[147] G. Bauer et al., "Status of the CMS detector control system", *Journal of Physics: Conference Series* **396** (Dec, 2012) 012023, doi: 10.1088/1742-6596/396/1/012023.

[148] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", 2014. doi: 10.48550/ARXIV.1412.6980.

[149] D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization", *Nature* **401** (Oct, 1999) 788–791, doi: 10.1038/44565.

[150] D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization", in *Advances in Neural Information Processing Systems*, T. Leen, T. Dietterich, and V. Tresp, eds., volume 13. MIT Press, 2000.

[151] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition". `https://doi.org/10.48550/arXiv.1512.03385`, 2015.

[152] The CMS ECAL Collaboration, "Autoencoder-based anomaly detection system for online data quality monitoring of the cms electromagnetic calorimeter", 2023. doi: https://doi.org/10.48550/arXiv.2309.10157.

[153] CMS Collaboration, "An Autoencoder Based Online Data Quality Monitoring for CMS ECAL". CMS-DP-2022-043, 2022.

[154] M. W. Asres et al., "Spatio-temporal anomaly detection with graph networks for data quality monitoring of the hadron calorimeter", *Sensors* **23** (2023), no. 24, doi: 10.3390/s23249679.

[155] CMS Collaboration, CMS, "An AutoEncoder-based Anomaly Detection tool with a per-LS granularity". CERN-CMS-DP-2023-010, 2023.

[156] CMS Collaboration, CMS, "Machine Learning Techniques for JetMET Data Certification of the CMS Detector". CERN-CMS-DP-2023-032, 2023.

[157] Vichayanun Wachirapusitanand on behalf of the CMS Collaboration, "Machine Learning applications for Data Quality Monitoring and Data Certification within CMS". ACAT 2021.

[158] U. Husemann, "Top-quark physics: Status and prospects", *Progress in Particle and Nuclear Physics* **95** (Jul, 2017) 48–97, doi: 10.1016/j.ppnp.2017.03.002.

[159] M. Cristinziani and M. Mulders, "Top-quark physics at the large hadron collider", *Journal of Physics G: Nuclear and Particle Physics* **44** (May, 2017) 063001, doi: 10.1088/1361-6471/44/6/063001.

[160] C. Collaboration, "Stairway to discovery: a report on the cms programme of cross section measurements from millibarns to femtobarns", 2024. `https://arxiv.org/abs/2405.18661`.

[161] The Particle Data Group, "Review of Particle Physics, 2018-2019, 67. Top Quark", *Phys. Rev. D* **98** (2018), no. 3, 030001, doi: 10.1103/PhysRevD.98.030001.

[162] B. Fuks, K. Hagiwara, K. Ma, and Y.-J. Zheng, "Signatures of toponium formation in lhc run 2 data", *Phys. Rev. D* **104** (Aug, 2021) 034023, doi: 10.1103/PhysRevD.104.034023.

[163] CMS Collaboration, "Probing entanglement in top quark production with the CMS detector". CMS-PAS-TOP-23-001, 2024.

[164] The Particle Data Group, "Review of Particle Physics, 2022-2023, 12. CKM Quark-Mixing Matrix", *PTEP* **2022** (2022) 083C01, doi: 10.1093/ptep/ptac097.

[165] CDF Collaboration, "Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab", *Phys. Rev. Lett.* **74** (Apr, 1995) 2626–2631, doi: 10.1103/PhysRevLett.74.2626.

[166] D0 Collaboration, "Observation of the top quark", *Phys. Rev. Lett.* **74** (Apr, 1995) 2632–2637, doi: 10.1103/PhysRevLett.74.2632.

[167] S. Catani et al., "Top-quark pair production at the LHC: fully differential QCD predictions at NNLO", *Journal of High Energy Physics* **2019** (Jul, 2019) doi: 10.1007/jhep07(2019)100.

[168] S. Catani et al., "Top-quark pair hadroproduction at NNLO: differential predictions with the MSbar mass", *Journal of High Energy Physics* **2020** (Aug, 2020) doi: 10.1007/jhep08(2020)027.

[169] M. Czakon et al., "Top quark pair production at complete NLO accuracy with NNLO+NNLL corrections in QCD", *Chinese Physics C* **44** (Aug, 2020) 083104, doi: 10.1088/1674-1137/44/8/083104.

[170] S. Frixione et al., "Single-top hadroproduction in association with a W boson", *Journal of High Energy Physics* **2008** (Jul, 2008) 029, doi: 10.1088/1126-6708/2008/07/029.

[171] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$", *Physical Review Letters* **110** (Jun, 2013) doi: 10.1103/physrevlett.110.252004.

[172] CDF Collaboration, D0 Collaboration, "Observation of s-channel production of single top quarks at the tevatron", *Physical Review Letters* **112** (Jun, 2014) doi: 10.1103/physrevlett.112.231803.

[173] CMS Collaboration, "Observation of $t\bar{t}H$ production", *Phys. Rev. Lett.* **120** (Jun, 2018) 231801, doi: 10.1103/PhysRevLett.120.231801.

[174] ATLAS Collaboration, "Observation of higgs boson production in association with a top quark pair at the lhc with the atlas detector", *Physics Letters B* **784** (2018) 173–191, doi: https://doi.org/10.1016/j.physletb.2018.07.035.

[175] CMS Collaboration, "A portrait of the higgs boson by the cms experiment ten years after the discovery", *Nature* **607** (2022) 60–68, doi: https://doi.org/10.1038/s41586-022-04892-x.

[176] CMS Collaboration, "Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **03** (2020) 056, doi: 10.1007/JHEP03(2020)056, arXiv:1907.11270.

[177] ATLAS Collaboration, "Measurements of the inclusive and differential production cross sections of a top-quark-antiquark pair in association with a Z boson at $sqrts = 13\,\text{TeV}$ with the ATLAS detector", *The European Physical Journal C* **81** (2021), no. 737, doi: 10.1140/epjc/s10052-021-09439-4.

[178] CMS Collaboration, "Measurement of the inclusive and differential $\mathrm{t\bar{t}\gamma}$ cross sections in the dilepton channel and effective field theory interpretation in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Journal of High Energy Physics* **2022** (May, 2022) doi: 10.1007/jhep05(2022)091.

[179] ATLAS Collaboration, "Measurements of inclusive and differential fiducial cross-sections of $\mathrm{t\bar{t}\gamma}$ production in leptonic final states at $\sqrt{s} = 13\,\text{TeV}$ in ATLAS", *The European Physical Journal C* **79** (May, 2019) doi: 10.1140/epjc/s10052-019-6849-6.

[180] ATLAS Collaboration, "Measurement of the total and differential cross-sections of $t\bar{t}W$ production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", 2024. doi: https://doi.org/10.48550/arXiv.2401.05299.

[181] CMS and ATLAS Collaborations, "Combination of measurements of the top quark mass from data collected by the ATLAS and CMS experiments at $\sqrt{s} = 7$ and 8 TeV", 2024. doi: https://doi.org/10.48550/arXiv.2402.08713.

[182] A. H. Hoang, "What is the top quark mass?", *Annual Review of Nuclear and Particle Science* **70** (2020), no. 1, 225–255, doi: 10.1146/annurev-nucl-101918-023530.

[183] F. Jegerlehner, M. Y. Kalmykov, and B. A. Kniehl, "On the difference between the pole and the $\overline{\text{MS}}$ masses of the top quark at the electroweak scale", *Physics Letters B* **722** (May, 2013) 123–129, doi: 10.1016/j.physletb.2013.04.012.

[184] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *The European Physical Journal C* **79** (Apr, 2019) doi: 10.1140/epjc/s10052-019-6863-8.

[185] CMS Collaboration, "Measurement of the top quark mass with lepton+jets final states using pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *The European Physical Journal C* **78** (Nov, 2018) doi: 10.1140/epjc/s10052-018-6332-9.

[186] CMS Collaboration, "Measurement of the top quark mass using events with a single reconstructed top quark in pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *Journal of High Energy Physics* **2021** (Dec, 2021) doi: 10.1007/jhep12(2021)161.

[187] J. H. Kühn and G. Rodrigo, "Charge asymmetry in hadroproduction of heavy quarks", *Physical Review Letters* **81** (Jul, 1998) 49–52, doi: 10.1103/physrevlett.81.49.

[188] F. Maltoni, M. Mangano, I. Tsinikos, and M. Zaro, "Top-quark charge asymmetry and polarization in $t\bar{t}W^{\pm}$ production at the LHC", *Physics Letters B* **736** (Sep, 2014) 252–260, doi: 10.1016/j.physletb.2014.07.033.

[189] R. Lysak, "Charge asymmetry in top quark pair production", *Symmetry* **12** (2020), no. 8, doi: 10.3390/sym12081278.

[190] ATLAS and CMS Collaborations, "Combination of inclusive and differential $t\bar{t}$ charge asymmetry measurements using ATLAS and CMS data at $\sqrt{s} = 7$ and $8\,\text{TeV}$", 2017. doi: 10.48550/ARXIV.1709.05327.

[191] ATLAS Collaboration, "Search for leptonic charge asymmetry in $t\bar{t}W$ production in final states with three leptons at $\sqrt{s} = 13\,\text{TeV}$", *Journal of High Energy Physics* **2023** (Jul, 2023) doi: 10.1007/jhep07(2023)033.

[192] CMS Collaboration, "Measurement of the top quark polarization and $t\bar{t}$ spin correlations using dilepton final states in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Physical Review D* **100** (Oct, 2019) doi: 10.1103/physrevd.100.072002.

[193] CMS Collaboration, "Measurement of the W boson helicity fractions in the decays of top quark pairs to lepton + jets final states produced in pp collisions at $\sqrt{s} = 8\,\text{TeV}$", *Physics Letters B* **762** (Nov, 2016) 512–534, doi: 10.1016/j.physletb.2016.10.007.

[194] CMS Collaboration, "Measurement of top quark polarisation in t-channel single top quark production", *Journal of High Energy Physics* **2016** (Apr, 2016) doi: 10.1007/jhep04(2016)073.

[195] CMS Collaboration, "Measurement of differential cross sections and charge ratios for t-channel single top quark production in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *The European Physical Journal C* **80** (May, 2020) doi: 10.1140/epjc/s10052-020-7858-1.

[196] ATLAS and CMS Collaborations, "Combination of inclusive top-quark pair production cross-section measurements using ATLAS and CMS data at $\sqrt{s} = 7$ and 8 TeV", *Journal of High Energy Physics* **2023** (Jul, 2023) doi: 10.1007/jhep07(2023)213.

[197] Q.-H. Cao, S.-L. Chen, and Y. Liu, "Probing Higgs width and top quark Yukawa coupling from $t\bar{t}H$ and $t\bar{t}t\bar{t}$ productions", *Physical Review D* **95** (Mar, 2017) doi: 10.1103/physrevd.95.053004.

[198] CMS Collaboration, "Search for production of four top quarks in final states with same-sign or multiple leptons in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *The European Physical Journal C* **80** (Jan, 2020) doi: 10.1140/epjc/s10052-019-7593-7.

[199] CMS Collaboration, "Measurement of the top quark Yukawa coupling from $t\bar{t}$ kinematic distributions in the dilepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. D* **102** (Nov, 2020) 092013, doi: 10.1103/PhysRevD.102.092013.

[200] CMS Collaboration, "Measurement of the top quark Yukawa coupling from $t\bar{t}$ kinematic distributions in the lepton + jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. D* **100** (Oct, 2019) 072007, doi: 10.1103/PhysRevD.100.072007.

[201] G. Mestdach, "Measurement of top quark pair-production in association with a photon using the CMS detector and Effective Field Theory interpretation". link, 2022. Presented 21 Nov 2022.

[202] C. Degrande et al., "Single-top associated production with a Z or H boson at the LHC: the SMEFT interpretation", *JHEP* **10** (2018) 005, doi: 10.1007/JHEP10(2018)005, `arXiv:1804.07773`.

[203] CMS Collaboration, "Search for new physics in top quark production with additional leptons in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ using effective field theory", *Journal of High Energy Physics* **2021** (Mar, 2021) doi: 10.1007/jhep03(2021)095.

[204] CMS Collaboration, "Probing effective field theory operators in the associated production of top quarks with a Z boson in multilepton final states at $\sqrt{s} = 13\,\text{TeV}$", *Journal of High Energy Physics* **2021** (Dec, 2021) doi: 10.1007/jhep12(2021)083.

[205] D. Buttazzo et al., "Investigating the near-criticality of the higgs boson", *Journal of High Energy Physics* **2013** (Dec, 2013) doi: 10.1007/jhep12(2013)089.

[206] CMS Collaboration, "Lepton identification in the processes with top quarks". CMS-AN-2022-016. (CMS internal).

[207] CMS Collaboration, "Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at $\sqrt{s} = 13\,\text{TeV}$", *Physical Review Letters* **122** (Apr, 2019) doi: 10.1103/physrevlett.122.132003.

[208] CMS Collaboration, "Search for supersymmetry in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with 137 fb^{-1} in final states with a single lepton using the sum of masses of large-radius jets", *Physical Review D* **101** (Mar, 2020) doi: 10.1103/physrevd.101.052010.

[209] K. Rehermann and B. Tweedie, "Efficient identification of boosted semileptonic top quarks at the LHC", *Journal of High Energy Physics* **2011** (Mar, 2011) doi: 10.1007/jhep03(2011)059.

[210] CMS Collaboration, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC", *Journal of Instrumentation* **16** (May, 2021) P05014, doi: 10.1088/1748-0221/16/05/p05014.

[211] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at sqrt(s) = 8 TeV", *Journal of Instrumentation* **10** (Jun, 2015) P06005–P06005, doi: 10.1088/1748-0221/10/06/p06005.

[212] CMS Collaboration, "Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Journal of High Energy Physics* **2020** (Mar, 2020) doi: 10.1007/jhep03(2020)056.

[213] CMS Collaboration, "Search for electroweak production of charginos and neutralinos in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Journal of High Energy Physics* **2022** (Apr, 2022) doi: 10.1007/JHEP04(2022)147.

[214] CMS Collaboration, "Measurement of the cross section of top quark-antiquark pair production in association with a W boson in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Journal of High Energy Physics* **2023** (Jul, 2023) doi: 10.1007/jhep07(2023)219.

[215] A. Brinkerhoff et al., "Search for the standard model Higgs boson produced in association with top quarks and decaying to leptons". CMS-AN-2013-159. (CMS internal).

[216] CMS Collaboration, "Measurement of four top quark production in the same-sign dilepton and multilepton channels". CMS-AN-2021-182. (CMS internal).

[217] CMS Collaboration, "Search for electroweak production of charginos and neutralinos in multilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV". CMS-AN-19-114. (CMS internal).

[218] CMS Collaboration, "The CMS statistical analysis and combination tool: COMBINE", 2024. doi: https://doi.org/10.48550/arXiv.2404.06614.

[219] ATLAS Collaboration, CMS Collaboration, LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011". CMS-NOTE-2011-005, 2011.

[220] J. S. Conway, "Incorporating nuisance parameters in likelihoods for multisource spectra", 2011. doi: https://doi.org/10.48550/arXiv.1103.0354.

[221] CMS Collaboration, "Measurements of the pp to ttW process at sqrt(s)=13 TeV using the full Run-2 data sample". CMS-AN-2019-127. (CMS internal).

[222] CMS Collaboration, "Measurements of the pp $\to$ WZ inclusive and differential production cross section and constraints on charged anomalous triple gauge couplings at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **04** (2019) 122, doi: 10.1007/JHEP04(2019)122, `arXiv:1901.03428`.

[223] CMS Collaboration, "Measurements of the pp $\to$ ZZ production cross section and the $\text{Z} \to 4\ell$ branching fraction, and constraints on anomalous triple gauge couplings at $\sqrt{s} = 13\,\text{TeV}$", *The European Physical Journal C* **78** (Feb, 2018) doi: 10.1140/epjc/s10052-018-5567-9.

[224] CMS Collaboration, "Measurement of the $Z\gamma$ production cross section in pp collisions at 8 TeV and search for anomalous triple gauge boson couplings", *Journal of High Energy Physics* **2015** (Apr, 2015) doi: 10.1007/jhep04(2015)164.

[225] CMS Collaboration, "CMS PYTHIA 8 colour reconnection tunes based on underlying-event data", *Eur. Phys. J. C* **83** (2023) doi: https://doi.org/10.1140/epjc/s10052-023-11630-8, `arXiv:2205.02905`.

[226] R. Barlow and C. Beeston, "Fitting using finite monte carlo samples", *Computer Physics Communications* **77** (1993), no. 2, 219–228, doi: https://doi.org/10.1016/0010-4655(93)90005-W.

[227] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses", *The Annals of Mathematical Statistics* **9** (1938), no. 1, 60 – 62, doi: 10.1214/aoms/1177732360.

[228] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV", *The European Physical Journal C* **75** (2015), no. 212, doi: 10.1140/epjc/s10052-015-3351-7.

[229] ATLAS and CMS Collaborations, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV", *Journal of High Energy Physics* **2016** (Aug, 2016) doi: 10.1007/jhep08(2016)045.

[230] CMS Collaboration Collaboration, "Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV ", *Phys. Rev. Lett.* **122** (Apr, 2019) 132003, doi: 10.1103/PhysRevLett.122.132003.

[231] ATLAS Collaboration Collaboration, "Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector ", *Journal of High Energy Physics* (Jul, 2020) doi: 10.1007/JHEP07(2020)124.

[232] ATLAS Collaboration, "Measurement of the production cross section of a single top quark in association with a Z boson in proton-proton collisions at 13 TeV with the ATLAS detector", *Phys. Lett. B* **780** (2018) 557, doi: 10.1016/j.physletb.2018.03.023, arXiv:1710.03659.

[233] J. A. Aguilar-Saavedra, "Top flavour-changing neutral interactions: theoretical expectations and experimental detection", *Acta Phys. Polon. B35:2695-2710* (2004) doi: https://doi.org/10.48550/arXiv.hep-ph/0409342.

[234] ATLAS Collaboration, "Analysis of the Wtb vertex from the measurement of triple-differential angular decay rates of single top quarks produced in the t-channel at $\sqrt{s} = 8$ TeV with the ATLAS detector", *Journal of High Energy Physics* **2017** (Dec, 2017) doi: 10.1007/jhep12(2017)017.

[235] CMS Collaboration, "Measurement of differential cross sections and charge ratios for t-channel single top quark production in proton–proton collisions at $\sqrt{s} = 13$ TeV", *The European Physical Journal C* **80** (May, 2020) doi: 10.1140/epjc/s10052-020-7858-1.

[236] C. Oleari, "The POWHEG BOX", *Nuclear Physics B - Proceedings Supplements* **205-206** (2010) 36 – 41, doi: https://doi.org/10.1016/j.nuclphysbps.2010.08.016.

[237] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *The European Physical Journal C* (2008) doi: 10.1140/epjc/s10052-007-0490-5.

[238] D. Pagani, I. Tsinikos, and E. Vryonidou, "NLO QCD+EW predictions for tHj and tZj production at the LHC", *JHEP* **08** (2020) 082, doi: 10.1007/JHEP08(2020)082, arXiv:2006.10086.

[239] CMS Collaboration, "Search for tZq in trilepton events". CMS-AN-2018-100. (CMS internal).

[240] CMS Collaboration, "Measurement of the t-channel single top quark production cross section in pp collisions at $\sqrt{s} = 7\,\text{TeV}$", *Physical Review Letters* **107** (Aug, 2011) doi: 10.1103/physrevlett.107.091802.

[241] A. Hoecker et al., "TMVA - Toolkit for Multivariate Data Analysis", 2009. doi: https://doi.org/10.48550/arXiv.physics/0703039.

[242] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector". CERN Yellow Reports: Monographs. CERN, Geneva, 2017. doi: 10.23731/CYRM-2017-002.

[243] A. Kulesza et al., "Associated top quark pair production with a heavy boson: differential cross sections at NLO + NNLL accuracy", *The European Physical Journal C* **80** (May, 2020) doi: 10.1140/epjc/s10052-020-7987-6.

[244] CMS Collaboration, "Measurement of the single top quark and antiquark production cross sections in the t channel and their ratio in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Physics Letters B* **800** (Jan, 2020) 135042, doi: 10.1016/j.physletb.2019.135042.

[245] L. Buonocore et al., "Precise predictions for the associated production of a W boson with a top-antitop quark pair at the LHC", *Physical Review Letters* **131** (Dec, 2023) doi: 10.1103/physrevlett.131.231901.

[246] R. Frederix, D. Pagani, and M. Zaro, "Large NLO corrections in ttW and tttt hadroproduction from supposedly subleading EW contributions", *Journal of High Energy Physics* **2018** (Feb, 2018) doi: 10.1007/jhep02(2018)031.

[247] R. Frederix and I. Tsinikos, "On improving NLO merging for ttW production", *Journal of High Energy Physics* **2021** (Nov, 2021) doi: 10.1007/jhep11(2021)029.

[248] CMS Collaboration, "Differential measurements of the $t\bar{t}W$ cross-section in 13TeV proton-proton collisions using the full run2 data sample and tt asymmetry measurement using $t\bar{t}W$ production". CMS-AN-2023-079. (CMS internal).

[249] CMS Collaboration, CMS Collaboration, "Jet energy scale and resolution measurement with Run 2 Legacy Data Collected by CMS at 13 TeV". CERN-CMS-DP-2021-033, 2021.

[250] L. Ferencz et al., "Study of ttbb and ttw background modelling for tth analyses", 2023. doi: https://doi.org/10.48550/arXiv.2301.11670.

[251] F. Capra, "The Tao of Physics". Shambhala Publications, 1975. ISBN 0-87773-078-4.

[252] J. Searle, "Minds, Brains and Science". Reith Lectures. Harvard University Press, 1984. ISBN 0-674-57633—0.

[253] G. Drescher, "Good and Real: Demystifying Paradoxes from Physics to Ethics". A Bradford book. MIT Press, 2006. ISBN 9780262042338.