HAL
open science

Measurement of triboson production in the multilepton final state and electron studies with the CMS experiment at the LHC

Jonas Rembser

To cite this version:

Jonas Rembser. Measurement of triboson production in the multilepton final state and electron studies with the CMS experiment at the LHC. High Energy Physics - Experiment [hep-ex]. Institut Polytechnique de Paris, 2020. English. NNT : 2020IPPAX083 . tel-03720899

HAL Id: tel-03720899

https://theses.hal.science/tel-03720899v1

Submitted on 12 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2020IPPAX083

Thèse de doctorat

ÉCOLE POLYTECHNIQUE

IP PARIS

Measurement of triboson production in the multilepton final state and electron studies with the CMS experiment at the LHC

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à l'École polytechnique

École doctorale de l'Institut Polytechnique de Paris (ED IP Paris) n°626
Spécialité de doctorat : Physique des particules

Thèse présentée et soutenue à Palaiseau, le 18 Novembre 2020, par

JONAS REMBSER

Composition du Jury :

Jürgen Reuter Senior Staff Scientist, DESY Theory Group, Hamburg	Rapporteur
David Rousseau Directeur de Recherche, IJCLab, Orsay (UMR9012)	Rapporteur
Lucia Di Ciaccio Professeur, LAPP, Annecy (UMR 5814)	Examinateur
Wolfgang Adam Senior Staff Scientist, HEPHY, Wien	Examinateur
Claude Charlot Directeur de Recherche, LLR, Palaiseau (UMR7638)	Examinateur
Philippe Busson Directeur de Recherche, LLR, Palaiseau (UMR7638)	Directeur de thèse

Abstract

This thesis reports on the measurement of the production of three massive gauge bosons in proton-proton collisions at 13 TeV with the CMS experiment at the CERN Large Hadron Collider (LHC). The thesis work focuses on the production of two W bosons and one Z boson (WWZ), decaying to either electrons or muons. This process has relatively little reducible background because of the fully reconstructed Z boson, and a higher production cross-section than processes with more Z bosons. The measurement considers 137 fb^{-1} of collision data recorded during the the second LHC run (2015 – 2018).

The production of three massive gauge bosons is a so-far unobserved phenomenon predicted by the standard model of particle physics. The process is sensitive to the trilinear and quartic gauge couplings, which might be modified by physics beyond the standard model. Additionally, triboson production is an increasingly important background to both direct searches for new physics and standard model precision measurements.

While trilinear gauge couplings are also accessible via diboson production, probing quartic couplings requires the measurement of Vector Boson Scattering (VBS) or triboson production. The VBS process better isolates the vector boson self-couplings, whereas triboson production is mainly affected by diagrams without any vector boson self-interaction. However, this plays a minor role in searches for new physics at energy scales beyond the electroweak scale, where triboson production has the advantage of probing a more massive final state.

The analysis presented in this thesis establishes evidence for the WWZ final state with a significance of 3.35 standard deviations (1.31 standard deviations if the vector bosons are required to be on-shell). Combined with other analyses covering the remaining massive triboson final states, this measurement contributed to observing the production of three massive gauge bosons with a significance of 5.67 standard deviations (2.90 standard deviations for on-shell vector bosons). All results are in agreement with the standard model predictions. Beyond that, this thesis presents studies of the WWZ processes sensitivity to anomalous quartic gauge couplings in an effective field theory framework.

Triboson production is a rare process that requires high electron and muon identification efficiencies. Among these two types of leptons, electrons are more challenging to measure. Therefore, this thesis includes a rigorous update and optimization of the multivariate electron selection based on the Boosted Decision Tree (BDT) algorithm, applied within the CMS collaboration since the beginning of LHC data-taking. As a result, the identification of electrons remains very performant, even at a significant increase in the number of proton-proton interactions per bunch crossing. The multivariate electron identification algorithm produced for this thesis is now the default for CMS analyses ot the second LHC run.

Résumé

Cette thèse présente la mesure de la production de trois bosons de jauge massifs dans des collisions entre protons à 13 TeV avec l'expérience CMS auprès du Grand collisionneur de hadrons (LHC) du CERN. Le travail de cette thèse se concentre sur la production de deux bosons W et d'un boson Z (WWZ), où les trois bosons se désintègrent en électrons ou en muons. Ce processus a un bruit de fond réductible relativement faible car le boson Z est entièrement reconstruit, et une section efficace plus élevée que les processus avec plusieurs bosons Z. L'analyse couvre 137 fb^{-1} de données enregistrées pendant la deuxième phase d'exploitation du LHC (2015 – 2018).

La production de trois bosons de jauge massifs n'a pas encore été observé, mais elle est prédit par le modèle standard de la physique des particules. Ces mesures sont donc un test du modèle standard et un défis de sensibilité. Ce processus est sensible aux couplages de jauge trilinéaire et quartiques, qui pourraient être modifiés par la physique au-delà du modèle standard. En outre, la production de tribosons est un bruit de fond de plus en plus important pour les recherches directes de nouvelles physiques et les mesures du modèle standard.

Les couplages de jauge trilinéaires sont également accessibles dans le processus diboson, mais le sondage des couplages quartiques nécessite la mesure de la diffusion de bosons massifs (VBS) ou de la production de tribosons. Le processus VBS isole bien les auto-couplages des bosons massifs, mais la production de triboson est dominée par des diagrammes sans aucun auto-couplage. Cependant, cela joue un rôle mineur dans les recherches de nouvelles physiques aux échelles d'énergie au-delà de l'échelle électrofaible, où la production de tribosons a l'avantage de sonder un état final plus massif.

L'analyse présentée établit la production WWZ avec une signification statistique de 3.35 écarts-types, qui est réduit à 1.31 écarts-types si les bosons sont non-virtuels. Combinée avec d'autres analyses couvrant les autres états finals de tribosons, cette mesure a contribué à une observation d'une signification statistique de 5.67 écarts-types de la production de trois bosons de jauge massifs (2.90 écarts-types pour les bosons non-virtuels). Les résultats sont en accord avec les prédictions du modèle standard. De même, cette thèse présente des études sur la sensibilité du processus WWZ aux couplages de jauge quartiques anormaux dans le cadre d'une théorie effective des champs.

Comme la production de triboson est un processus rare, sa détection requiert une bonne efficacité d'identification des électrons et des muons. Parmi ces deux types de leptons, les électrons sont les plus difficiles à mesurer. Par conséquent, ce travail de thèse inclut une amélioration rigoureuse de la sélection d'électrons multivariée basée sur l'algorithme Boosted Decision Tree (BDT), utilisé dans le cadre de la collaboration CMS depuis la première période d'exploitation du LHC. Cette optimisation permet ainsi de conserver une identification des électrons très performante dans la deuxième période malgré le fait que le nombre d'interactions proton-proton par croisement de faisceaux de particules a considérablement augmenté. L'algorithme d'identification d'électrons multivariée développé dans cette thèse est désormais devenu l'algorithme standard dans les analyses de CMS pour toutes les données de la deuxième phase du LHC.

Summary for the general public

The standard model of particle physics contains matter particles (fermions) and particles mediating interactions between matter (gauge bosons). Electrons and the so-called quarks found in protons and neutrons are matter particles. The massless photon, which mediates electromagnetic interaction, is the most well-known gauge boson. The so-called weak interaction, responsible for radioactivity, is also mediated by gauge bosons. Unlike photons, these weak bosons are massive, decay rapidly, and directly interact among themselves. This thesis studies the production of three weak bosons in high-energy proton collisions at the CERN Large Hadron Collider, where sufficient energy is available to produce the heavy bosons. This process is sensitive to weak boson-self interaction. As they are unstable, they can only be detected indirectly by their decay products, e.g. electrons. Therefore, this thesis also presents improvements in electron detection with the CMS particle detector.

Résumé vulgarisé pour le grand public

La théorie de la physique fondamentale contient des particules de matière (fermions) et des porteurs des interactions élémentaires (bosons de jauge). Les électrons et les quarks constituant les protons et les neutrons présents au sein du noyau atomique sont des fermions. L'interaction faible, responsable de la radioactivité, est portée par des bosons de jauge dits bosons *faibles*. Contrairement aux photons qui portent l'interaction électromagnétique, les bosons faibles sont massifs, se désintègrent rapidement et interagissent entre eux. Cette thèse étudie la production de trois bosons faibles dans les collisions entre protons au Grand collisionneur de hadrons du CERN, où l'énergie est suffisante pour observer ce processus sensible à l'auto-couplage des bosons faibles. Ces bosons ne peuvent être détectés qu'indirectement par leurs produits de désintégration, par exemple des électrons. Cette thèse inclut des améliorations de la détection des électrons dans le détecteur CMS.

Acknowledgements

I'm very grateful to have done this doctorate in a supportive environment: the CMS group at LLR. Thanks for all your help and encouragement, Jean-Baptiste, Florian, Yves, Roberto, Alex, Amina, and Olivier! Thanks for all the discussions. I always enjoyed learning about the CMS experiment and its history from you. But there are two LLR-CMS members I want to highlight. First of all, my thesis advisor Philippe. I couldn't have succeeded without his help in many crucial moments. His wisdom and experience in handling the dynamics of large collaborations are invaluable, and I strive to replicate aspects of his attitude in my career. Second, I want to particularly thank Claude, who has followed my theses since the beginning, advising me for the electron work and the physics analysis. As an unbiased thinker with a good sense of priorities and detailed knowledge of many aspects of the CMS experiments, he always inspired me.

Many members of the lab, mostly other doctoral students and postdocs have become good friends of mine during these years. I'm thrilled to have met all of you on the journey and wish you the best in your careers! In particular, I want to thank here Duje, with whom I shared my love for music and many other interests, and Artur, the postdoc I looked up to and who inspired me often.

One of the highlights of the last three years were the teaching activities I did on the side. Tutoring students is always exciting! Thanks so much, Alex, Chiara, Cristina, Matteo, and Géraud, for the TREX experience, and thanks to the students choosing my projects in the experimental high-energy physics lecture.

I want to show gratitude also to my amazing colleagues from the CMS collaboration from all over the world. I have not met most of you personally, but you still played an essential role in this journey. I have to highlight the EGM physics object group. Thanks for warmly accepting my contributions and promoting my efforts! Working with you, Livia, Shilpi, and Swagata, was always great fun. But in particular, I want to thank Francesco and Sam for greatly supporting my career.

The software and computing group was another group within CMS where I could learn from fantastic people. In particular, I'm incredibly thankful for the support and guidance of Slava and Andrea, and also happy to have worked with Chris and Matti. And thanks a lot to Danilo, who played a key role towards the end! There are more members of the CMS software community that I looked up to, but unfortunately, I didn't have enough time to collaborate with them. I'm still grateful for the inspiration.

The final CMS group that was important for my work was the standard model physics analysis group, where I particularly acknowledge Kenneth, Josh, and Pietro. But most importantly, I need to thank my friends from the *VVV analysis team*, particularly the postdocs Philip, Sapta, and Hannsjörg, but also Michael, Sergo, and Frank. I learned a lot from you guys and wished we had met in person more often!

In 2020, the first year of the Covid-19 pandemic and last year of my doctorate, I did not see many different people. Still, I saw my flatmate and very good friend Alberto almost every day. I think without him, this year would have been incredibly dull and more challenging than it already was.

Finally, I need to thank my family for their support because it definitely would not have been possible without them. They were always there if I wanted to complain; sorry about my sometimes very negative attitude! And they were always there with great advice, and without their support, I would not have made it.

Contents

1 Theoretical foundation and motivation **1**
1.1 The standard model of particle physics . . . 1
1.1.1 Particles and interactions . . . 3
1.1.2 Free fermions . . . 4
1.1.3 Free vector bosons . . . 4
1.1.4 The internal U(1) symmetry for electromagnetism . . . 5
1.1.5 The internal $SU(2)_L$ symmetry for weak interactions . . . 6
1.1.6 Minimal scalar sector and unification of SU(2) and U(1) . . . 7
1.1.7 Fermion masses . . . 9
1.1.8 SU(3) and quantum chromodynamics . . . 10
1.2 Physics phenomena beyond the standard model . . . 10
1.3 Triboson production as a test of the standard model . . . 12
1.3.1 Triboson production at the LHC . . . 14
1.3.2 The effective field theory framework . . . 15

2 The CMS experiment at the Large Hadron Collider **21**
2.1 The Large Hadron Collider . . . 21
2.1.1 The CERN accelerator infrastructure . . . 22
2.1.2 Design and parameters of the LHC . . . 22
2.1.3 Operations history and future . . . 25
2.2 The CMS experiment . . . 26
2.2.1 The CMS coordinate system . . . 28
2.2.2 Detector subsystems . . . 29
2.2.2.1 Solenoid magnet . . . 29
2.2.2.2 Inner charged-particle tracker . . . 29
2.2.2.3 Electromagnetic calorimeter . . . 32
2.2.2.4 Hadronic calorimeter . . . 34
2.2.2.5 Muon detectors . . . 36
2.2.3 Trigger and data acquisition system . . . 37
2.2.4 Luminosity measurement . . . 39

3 Reconstruction and simulation of collision events with the CMS detector **41**
3.1 Global event reconstruction with the particle-flow algorithm . . . 42
3.2 Electron and photon reconstruction . . . 43
3.2.1 Reconstruction steps . . . 44
3.2.2 Electron charge identification . . . 45
3.2.3 Electron and photon identification . . . 45
3.2.4 Selection efficiencies in data versus simulation . . . 46
3.2.5 $Z \to e^+e^-$ invariant mass with full energy corrections . . . 48
3.3 Muon reconstruction . . . 48
3.3.1 Muon track reconstruction . . . 48
3.3.2 Muon identification and isolation . . . 49
3.4 Jet reconstruction . . . 50
3.4.1 Jet clustering . . . 51

3.4.2 Jet energy scale and resolution corrections . . . 51
3.4.3 Missing energy reconstruction . . . 52
3.4.4 Tagging of heavy-flavor jets . . . 53
3.5 Statistics of in-time pileup extrapolation . . . 54
3.6 Simulation of proton-proton collisions . . . 58
3.7 CMS reconstruction software . . . 59

4 Machine learning fundamentals 61
4.1 Introduction to machine learning . . . 61
4.1.1 Training a machine learning model . . . 61
4.1.2 Classification problems . . . 62
4.1.3 The bias-variance tradeoff . . . 63
4.2 Boosted decision trees and the XGBoost algorithm . . . 65
4.2.1 Gradient boosting for decision trees . . . 65
4.2.2 Hyperparameters of the XGBoost algorithm . . . 67
4.3 Sample reweighting for classification . . . 68

5 Multivariate electron identification 71
5.1 General training strategy and requirements . . . 71
5.2 Effect of the new pixel detector on electron observables . . . 73
5.3 First optimization of the electron MVA for 2017 data . . . 74
5.4 Reoptimization of the electron MVA with XGBoost . . . 78
5.5 Implementation in the CMS reconstruction software . . . 84
5.6 Hyperparameter tuning with Bayesian optimization . . . 85
5.6.1 Bayesian optimization with Gaussian processes . . . 86
5.6.2 Training and optimization setup . . . 87
5.6.3 Results and interpretation . . . 88
5.7 Conclusion and outlook . . . 92

6 Measurement of triboson production 93
6.1 Analysis overview . . . 94
6.2 Analysis samples . . . 95
6.2.1 Trigger selection . . . 95
6.3 Object selection . . . 95
6.3.1 Lepton isolation variable . . . 95
6.3.2 Common loose lepton selection . . . 97
6.3.3 Selections for W and Z candidate leptons . . . 98
6.3.4 Jet selection and b-tagging . . . 99
6.3.5 Missing transverse momentum . . . 100
6.4 Event preselection . . . 102
6.5 Event categorization . . . 102
6.6 Event classification with Boosted Decision Trees . . . 104
6.6.1 Input variables . . . 104
6.6.1.1 $t\bar{t}Z$ discrimination . . . 105
6.6.1.2 ZZ discrimination . . . 105
6.6.1.3 Common variables . . . 107
6.6.2 BDT training . . . 116
6.7 Signal extraction with BDT classifiers . . . 122
6.7.1 Introduction . . . 122
6.7.2 Classifier calibration diagnostics . . . 122
6.7.3 Classifier recalibration . . . 124
6.7.4 Ceiling study with mixture model and likelihood free inference . 126

6.7.5 Outlook on likelihood-free inference methods 129
6.7.6 Signal extraction with multidimensional histograms 130
6.8 Data-driven background estimation 131
6.8.1 Principles of data-driven background estimation 133
6.8.2 Estimate of ZZ and $t\bar{t}Z$ backgrounds from control regions 134
6.8.3 Treatment of other backgrounds 136
6.9 Systematic uncertainties 137
6.9.1 Theory uncertainties 138
6.9.2 Experimental uncertainties 139
6.10 Combination and statistical interpretation 141
6.10.1 Statistical uncertainties only 142
6.10.2 Systematic uncertainties in the likelihood function 143
6.10.3 Likelihood scans and nuisance parameter impact 145
6.11 Combined triboson measurement results 147

7 Conclusion **153**

Acronyms List **155**

Bibliography **159**

Chapter 1

Theoretical foundation and motivation

The theory of fundamental physics is the so-called *standard model of particle physics*, which describes the fundamental particles and their interactions. The limits of the standard model can be explored in different experimental environments, one of them being the *high-energy frontier*, where the energy refers to the center-of-mass energy of particle collision events. The state-of-the-art particle collider that explores the current energy frontier is the Large Hadron Collider (LHC), which collides proton beams at a center-of-mass energy of $\sqrt{s} = 13$ TeV. So far, the LHC experiments have confirmed all predictions by the standard model at this energy scale and have not seen anything contradicting it. This means that high energy physics at the energy frontier is entering an era of precision physics, intending to conceive high-quality measurements that are sensitive to effects of physics beyond the standard model even if there is no smoking-gun-like signal.

This thesis explores the energy frontier by studying the production of three massive vector bosons (W or Z bosons). The rate of this *triboson production*[1] process is sensitive to self-interactions among the vector bosons, their coupling to the Higgs boson, and possible new physics beyond the standard model affecting the electroweak sector. Before elaborating further on the motivation for this measurement in Section 1.3, this chapter gives an introduction to the standard model in Section 1.1 and reviews the phenomena that it does not explain in Section 1.2.

1.1 The standard model of particle physics

At the very foundation of physics stands an observation by the Michelson-Morley experiment: the speed of light is the same in all inertial reference frames. This result was rigorously interpreted by Einstein and other physicists of the early 20th century. Mathematically, the invariance of the speed of light can be enforced by describing reality with Minkovski spacetime, where translations and Lorentz transformations permit us to change from one reference frame to another. The equations of physics should be covariant, meaning that they should have the same form in all reference frames. To ensure this, the building blocks of the equations must live in some representation of the Poincaré symmetry group (Lorentz transformations plus translations), just like vectors in spacetime do. These representations can be labeled by the so-called *spin*, referring to different representations of the Lorentz group, and a *mass*, which needs to be supplemented to parametrize representations of the full Poincaré group. The elementary objects in our visible Universe can be described in precisely these representations. This model was built and extensively tested in the last decades and is now known as the standard model of particle physics [1, 2, 3].

[1] The W and Z bosons are not the only fundamental bosons in the standard model. However, in this thesis, the term *triboson production* always refers to the production of three W or Z bosons.

In the standard model, quantum fields of spin-1/2 (fermionic) make up stable and unstable matter. Spin-1 fields (bosonic) mediate interactions between the fermion fields. A spin-0 field, the Higgs field, gives rise to mass terms for almost all fields. Besides spin and mass, there are also other field labels that correspond to *internal* symmetries. Internal means that spacetime itself is not affected by these symmetry transformations, but only the fields are concerned. The additional labels (or quantum numbers) corresponding to the internal symmetries are the key to build a consistent theory that accounts for all the experimentally observed interactions.

The standard model is a *quantum field theory*. The additional abstraction with field operators makes it possible to mathematically formulate a theory of interacting particles that can turn into each other. A given field can create or annihilate its corresponding particles by acting on the quantum state of our Universe, sticking to the interaction rules encoded by the standard model. Figure 1.1 shows all the particles that are described in this way.

Before diving into the details of the theory, we will shift to a more experimental perspective and discuss the standard model particles listed in Figure 1.1. Then, we will review the standard model starting with the fields that live in the lowest-dimensional nontrivial representation of the Poincaré group: the fermionic matter fields. Afterwards, we will introduce an internal symmetry that causes the fermions to interact with photons, leading us to quantum electrodynamics. We will take a look at the full electroweak theory, including the Higgs mechanism (also known as electroweak symmetry breaking). The strong interaction part of the standard model, Quantum Chromodynamics (QCD), is reviewed only briefly, as this is not the physics that was primarily targeted in this thesis. The presented review is strongly inspired by [4].

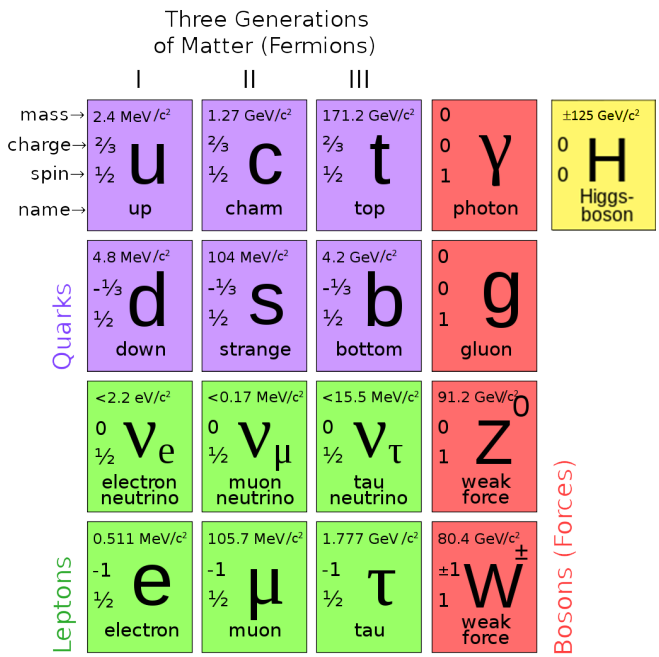


Figure 1.1: The particles of the standard model, showing the three generations of fermions that are divided into quarks (purple) and leptons (green). The four gauge bosons that mediate fundamental interactions are shown in red. The Higgs boson indicated by the yellow box on the right is the only fundamental scalar particle in the standard model.

1.1.1 Particles and interactions

The standard model particles listed in Figure 1.1 can be grouped into four types, as indicated by the colors in the figure. In the matter sector, there are quarks and leptons. Then, there are the gauge bosons that mediate interactions between matter. Finally, there is a distinct scalar particle, the Higgs boson, whose corresponding field is responsible for giving mass to all particles. It is instructive to talk about the interactions – or forces – and their associated gauge bosons first. Being aware of the possible interactions helps to understand the differences between the fermions later on. All we need to know about fermions for the moment is that they come in three generations of increasing mass, where only the first generation particles are stable and hence make up the visible matter in our Universe. The standard model knows three kinds of interactions. The well-known electromagnetic force is mediated by the massless photon, denoted with γ. Then there is the so-called weak force, mediated by the W and Z bosons, which permits different fermions to interact wich each other. For example, nuclear beta decays happen through the weak interaction. The weak force is very short-ranged because the W and Z bosons are massive. Finally, the standard model describes the strong interaction, mediated by the gluons, that binds together one type of fermions, namely the quarks.

Quarks and leptons are the two fermion-types that come in three generations, with a set of two intimately related leptons or quarks in each generation. A corresponding opposite-charged antiparticle matches all fermions. Quarks all have a fractional electric charge and are the constituents of hadrons, such as the proton or the neutron. The proton contains one down quark and two up quarks, whereas the neutron consists of one up quark and two bottom quarks. The third-generation top quark takes a special place in particle physics phenomenology because it is with 172.1 GeV the most massive particle of them all. This is important for two major reasons. First, it means that the top quark is the lepton that interacts most with the Higgs boson, dominating quantum corrections to its mass. Second, the large mass makes the top quark the only quark that decays via the weak interaction before it can hadronize, its decay products being a W boson and a bottom quark.

Among the leptons, there is one charged particle and a corresponding neutral neutrino per generation. The first-generation charged lepton is the ubiquitous electron that binds with neutrons and protons to form atoms. In the second generation, we have the heavier muon that can be observed in nature as the decay product of hadrons from cosmic-ray interactions in the upper atmosphere. This makes the muon an important actor in astroparticle physics [5], but it also has a distinct role in collider physics. Its lifetime is long enough for it to reach particle detectors, and it interacts very little before it decays as $\mu^- \to e^- + \bar{\nu}_e + \nu_\mu$. The annotated particle masses listed in Figure 1.1 indicate that the muon is more massive than the two first-generation quarks. This raises the question of possible muon decays to quarks. As we will discuss later in Section 1.1.8, quarks can never be the final decay products because of color confinement. Instead, the final decay product must be hadrons, where the lightest hadrons – the pions – are more massive than the muon. Hence, hadronic muon decays are forbidden by energy conservation.

The third generation tau-lepton is massive enough to not only decay to first- and second-generation leptons, but also to hadrons. Its lifetime is relatively short, so taus produced in collider experiments decay close to the interaction point. Tau detection is particularly challenging, as there is at least one neutrino among the decay products and the hadronic decays can be easily confused with a hadronic jet stemming from a quark or a gluon.

Now that the standard model particles have been presented, it is time to explore the theory that describes them.

1.1.2 Free fermions

The Lorentz group is covered by two copies of SU(2), plus two distinct transformations that bridge its four disconnected subgroups: time-reversal and the reversal of spatial coordinates, commonly called parity transformation. The most fundamental objects that can be Lorentz transformed in a non-trivial way live in the two-dimensional representation of one copy of SU(2) and the trivial representation of the other copy. This representation is labeled $(\frac{1}{2}, 0)$, and the objects that live in it are called *left-chiral spinors* However, the parity transformation swaps this representation with $(0, \frac{1}{2})$, making the object a *right-chiral* spinor. Therefore, to describe a physical system invariant under parity transformations, we need to use both kinds of spinors. They are usually put together into a single object called a *Dirac spinor*, which hence lives in the $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ representation and stands for fermion fields:

$$\psi = \begin{pmatrix} \chi_L \\ \zeta_R \end{pmatrix}. \tag{1.1}$$

It is assumed in the standard model that when transitioning from one state to another, nature chooses the path that minimizes the integral of some object, called the *Lagrangian* $\mathcal{L}$. This integral is the so-called *action*, and the equations obtained by minimizing it are named the *equations of motion*. The Lagrangian has to be Lorentz invariant, and its terms may include derivatives of the fields. However, terms with higher-order field derivatives can only appear if first-order-derivative terms do not affect the equations of motion. Otherwise, the resulting theory would be unstable. Terms in the third order of the fields – or higher – result in interactions between the fields. Following these restrictions, the Lagrangian to describe free fermions must be

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi} \left(i\gamma^\mu \partial_\mu - m \right) \psi. \tag{1.2}$$

where the matrices γ^μ have to satisfy the anticommutation relation $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$ with the Minkovski metric $\eta^{\mu\nu}$, and $\bar{\psi} = \psi^\dagger \gamma^0$. The hermitian conjugate of Equation 1.2 would work as well, but it is not necessary to keep it. It would lead to the same equations of motion.

It is now clear that fermions are very fundamental in theoretical physics, so it is satisfying to know that these objects can describe the building blocks of matter. In reality, there are many distinct fermions with masses spread over orders of magnitudes. The different fermions also live in different representations of the internal symmetries.

1.1.3 Free vector bosons

In the next higher-dimensional representation of the Lorentz group, both copies of SU(2) are represented by the spin-1/2 representation. The total spin is the sum of both individual spins, so the $\left(\frac{1}{2}, \frac{1}{2}\right)$ representation gives us spin-1 bosons. It transforms exactly like the four-vector representation, meaning that this *is* the four-vector representation. Therefore, the objects living in it are called *vector bosons*.

Since the spacetime derivative is also a four-vector, it is possible to write the general Lagrangian for a free vector boson without any additional matrices:

$$\mathcal{L}_{\text{Proca}} = \frac{1}{2} \left(\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu \right) + m A^\mu A_\mu. \tag{1.3}$$

The $\partial^\mu A_\mu$ term does not affect the equations of motion, so the terms with two derivatives have to be kept this time. If the vector boson is massless, this Lagrangian is called $\mathcal{L}_{\text{Maxwell}}$. The kinetic terms for vector bosons are often written in terms of the field strength tensor:

$$F^{\sigma\rho} := \partial^\sigma A^\rho - \partial^\rho A^\sigma, \tag{1.4}$$

which makes the Maxwell Largrangian look very compact:

$$\mathcal{L}_{\text{Maxwell}} = \frac{1}{4} F^{\mu\nu} F_{\mu\nu}. \tag{1.5}$$

The minus sign in front of the second term in Equation 1.3 seems at first glance unmotivated, but with this choice, the Lagrangian has an internal U(1) symmetry for massless fields. This has far-reaching implications that we will explore in the following.

1.1.4 The internal U(1) symmetry for electromagnetism

So far, we have used spacetime symmetry to arrive at a description of free fermions and vector bosons. We will now see how interacting theories arise naturally from imposing internal symmetries on the Lagrangian.

The Maxwell Lagrangian allows us to gauge the vector potential A_μ with the derivative of an arbitrary function:

$$A_\mu \to A'_\mu = A_\mu + \partial_\mu g a(x). \tag{1.6}$$

The arbitrary function $a(x)$ can be different for each point in space-time, making this gauge symmetry a *local symmetry*. The constant g is factored out because it will be interpreted as a coupling constant later.

For the fermions, we can attempt to transform the fields with a phase rotation, corresponding to a U(1) internal symmetry as well:

$$\psi \to \psi' = e^{iga(x)}\psi, \quad \bar{\psi} \to \bar{\psi}' = e^{-iga(x)}\bar{\psi}. \tag{1.7}$$

This is *not* a local symmetry of $\mathcal{L}_{\text{Dirac}}$. It is only a *global symmetry*, meaning that it is only a symmetry if the parameter $a(x)$ does not depend on the position, i.e., $a(x)$ is just a.

However, if we add an additional coupling term to $\mathcal{L}_{\text{Dirac}}$ and apply the transformations in Equations 1.6 and 1.7 *together*, the local symmetry works out. In other words, the following Lagrangian is invariant under local U(1) transformations:

$$\mathcal{L}_{\text{Dirac+Extra-Term}} = -m\bar{\psi}\psi + i\bar{\psi}\gamma_\mu\partial^\mu\psi + gA_\mu\bar{\psi}\gamma^\mu\psi. \tag{1.8}$$

To make the Lorentz-invariance of the terms more explicit, one usually defines the covariant derivative $D_\mu = \partial_\mu + igA_\mu$, which transforms like a four-vector even under the local symmetry. Using this definition and adding the Maxwell Lagrangian – for which Equation 1.6 has already been a local symmetry –, one arrives at the Lagrangian for Quantum Electrodynamics (QED):

$$\mathcal{L}_{\text{QED}} = -m\bar{\psi}\psi + i\bar{\psi}\gamma_\mu D^\mu\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{1.9}$$

Noether's theorem connects each internal symmetry with a conserved quantity. The internal U(1) symmetry results in the conservation of electric charge.

1.1.5 The internal $SU(2)_L$ symmetry for weak interactions

The success with the U(1) symmetry motivates the investigation of other possible internal symmetries to model the remaining fundamental interactions. In the standard model, there is not only one fermionic field, so we must also consider possible internal symmetries of Lagrangians with multiple fermions. For working with two fermions, it is convenient to organize them in a *doublet*:

$$\Psi := \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}, \quad \bar{\Psi} := \begin{pmatrix} \bar{\psi_1} & \bar{\psi_2} \end{pmatrix}. \tag{1.10}$$

The Dirac Lagrangian $\mathcal{L}_{\mathrm{D_1+D_2}} = i\bar{\Psi}\gamma_\mu\partial^\mu\Psi$ for such a doublet in the massless case is invariant under global SU(2) transformations:

$$\Psi \to \Psi' = e^{ia_i\frac{\sigma_i}{2}}\Psi, \quad \bar{\Psi} \to \bar{\Psi}' = \bar{\Psi}e^{-ia_i\frac{\sigma_i}{2}}, \tag{1.11}$$

where a_i are arbitrary real constants and $\frac{\sigma_i}{2}$ are the generators of SU(2), with the Pauli matrices σ_i. Again, if we want to have a *local* SU(2) symmetry, the derivative needs to be replaced with the covariant derivative:

$$D^\mu = \partial^\mu - ig\frac{\sigma^i}{2}W_i^\mu. \tag{1.12}$$

This time, three spin-1 fields are needed, one for each generator of SU(2). These fields usually appear in combination with the generators, so it is useful to define

$$\mathcal{W}_\mu := \left(W_\mu\right)_i \frac{\sigma_i}{2}. \tag{1.13}$$

To get a locally SU(2) invariant Lagrangian, we still need a kinetic term for these three vector fields. One might naively attempt to use the simplest possible derivative terms, just like for a single vector field (compare Section 1.1.3), but then the Lagrangian would not be SU(2) invariant. Instead, the kinetic term needs to be defined in terms of the covariant derivative:

$$\mathcal{W}^{\mu\nu} = \frac{i}{g}\left[D^\mu, D^\nu\right] = \frac{i}{g}\left(D^\mu D^\nu - D^\nu D^\mu\right). \tag{1.14}$$

Applying this object to a test function $\mathcal{W}_{\mu\nu}f(x)$ reveals that the actual derivatives cancel out and how the field strength tensor can be expressed in terms of the fields:

$$\mathcal{W}^{\mu\nu} = \partial^\mu\mathcal{W}^\nu - \partial^\nu\mathcal{W}^\mu - ig\left[\mathcal{W}^\mu, \mathcal{W}^\nu\right]. \tag{1.15}$$

Note that because of the multiplication with the generators, the field strength tensor is a matrix. However, to get a SU(2) invariant kinetic term for these three spin-1 fields, it is sufficient to take the trace of the analogon to the Maxwell term.

Now, we can write down the locally SU(2) invariant Lagrangian:

$$\mathcal{L}_{\text{locally SU(2) invariant}} = i\bar{\Psi}\gamma_\mu D^\mu\Psi - \frac{1}{4}\mathrm{Tr}\left(\mathcal{W}_{\mu\nu}\mathcal{W}^{\mu\nu}\right). \tag{1.16}$$

The additional commutator term in the field strength tensor (Equation 1.15) appears for all non-abelian gauge groups. It introduces new terms in the Lagrangian with the third or fourth power of the field $\mathcal{W}_\mu$. These terms correspond to the trilinear and quartic self-interactions of vector bosons, which are studied in this thesis.

The weak nuclear force is modeled by such a SU(2) interaction, more specifically a parity-violating $SU(2)_L$ interaction because only left-handed leptons interact via the weak force. This weak force is extremely short-ranged compared to the electromagnetic force, which can be modeled by making the spin-1 fields massive. However, it is not possible to add mass terms for the vector bosons without breaking the SU(2) symmetry. Another problem are the fermion masses, as the mass terms for fermion doublets are only SU(2) invariant if both fermions in the doublet have the same mass. Mixing the left-handed SU(2) fermion doublets with the right-handed SU(2) singlets to write down mass-terms that are symmetric under parity transformation does not work.

Although from a symmetry perspective, the addition of mass-terms for the fermions appears to be just as problematic as for vector bosons, the introduction of massive vector bosons entails more theoretical problems. A massless vector boson can only be transversally polarized, but a massive vector boson can also have longitudinal polarization. This longitudinally polarized component causes the self-interaction processes to violate unitarity around 1.2 TeV [6, 7]. In general, the local SU(2) symmetry is needed to get a renormalizable quantum field theory. The bottom line is that Equation 1.16 with added mass-terms cannot be the correct theory for weak interactions. Next, we will discuss how all these problems can be addressed with minimal changes to the theory.

1.1.6 Minimal scalar sector and unification of SU(2) and U(1)

It is not possible to directly add mass terms for vector bosons and fermions without explicitly breaking the internal SU(2) symmetry of the theory. A way to solve this problem without giving up the SU(2) symmetry altogether is via a mechanism that spontaneously breaks the symmetry at lower energies while preserving it at high energy. The Higgs mechanism – or Brout-Englert-Higgs (BEH) mechanism to give credit to all its inventors – achieves this by introducing a complex scalar SU(2) doublet with a self-interaction term in the Lagrangian that causes this scalar field to have a nonzero vacuum expectation value at lower energies.

Combining the findings of the previous sections, we so far have a locally U(1) and SU(2) invariant Lagrangian without mass-terms:

$$\mathcal{L}_{\text{SU(2) and U(1)}} = i\bar{\Psi}\gamma_\mu D^\mu \Psi - \frac{1}{4}\text{Tr}\left(\mathcal{W}_{\mu\nu}\mathcal{W}^{\mu\nu}\right) - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}, \tag{1.17}$$

with the covariant derivative

$$D^\mu = \partial^\mu - ig\frac{1}{2}B^\mu - ig'\mathcal{W}^\mu. \tag{1.18}$$

The gauge field of the U(1) symmetry is now called B^μ instead of A^μ, where A^μ is the common name for the photon field. In a minute, we will see that the mass eigenstates will mix the SU(2) and U(1) gauge fields, with the photon field being the mixture's massless eigenstate. The factor of $\frac{1}{2}$ is introduced to unclutter the notation later.

The BEH mechanism adds the following minimal scalar sector to the Lagrangian:

$$\mathcal{L}_{\text{BEH}} = (D^\mu\Phi)^\dagger\left(D_\mu\Phi\right) + \rho^2\Phi^\dagger\Phi - \lambda\left(\Phi^\dagger\Phi\right)^2 = (D^\mu\Phi)^\dagger\left(D_\mu\Phi\right) - V(\Phi). \tag{1.19}$$

In this context, *minimal* means not merely a non-interacting scalar field, but also the necessary changes to realize electroweak symmetry breaking. The scalar has to be a SU(2) doublet. Otherwise, it would be again not possible to get different masses for

the elements of the left-handed fermion doublets. It is also necessary to add the next highest power in Φ to get a Higgs potential $V(\Phi)$ with the required properties. We will also find out soon why the scalar field needs to be complex.

The minimum of the potential $V(\Phi)$ can be found by equating its derivative with zero, and it turns out to be at

$$\Phi^\dagger\Phi = \frac{\rho^2}{2\lambda}. \tag{1.20}$$

This minimum is symmetric in all four components of Φ, but the system has to choose one of the solutions. This is the crux of the BEH mechanism. The SU(2) symmetry is spontaneously broken for perturbations around the minimum, even if the original Lagrangian is symmetric. The spontaneous symmetry breaking is depicted in Figure 1.2.

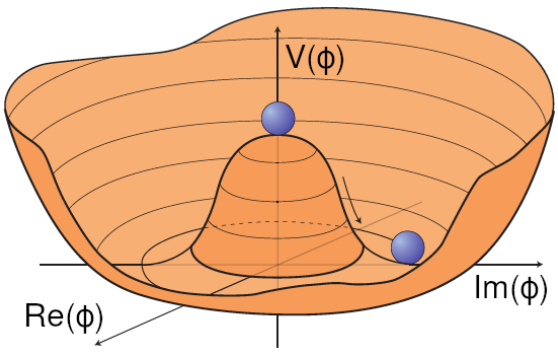


Figure 1.2: A visualization of the Higgs potential $V(\Phi)$ [8]. Just like a ball would roll into the symmetric valley of minima and eventually come to a halt at a random position, the Higgs field's vacuum expectation value assumes a random minimum, breaking the SU(2) symmetry. Note that since the actual Higgs field is not only a scalar but also a SU(2) doublet, so it has four components, unlike what is depicted in the figure.

A convenient choice for the minimum of the potential for demonstration purposes is

$$\Phi_{\text{min}} = \begin{pmatrix} 0 \\ \sqrt{\frac{\rho^2}{2\lambda}} \end{pmatrix} \equiv \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}. \tag{1.21}$$

In quantum field theory, computations are done as expansions around the minimum:

$$\Phi = \begin{pmatrix} \phi_{1r} + \phi_{1c} \\ \frac{v}{\sqrt{2}} + \phi_{2r} + \phi_{2c} \end{pmatrix} \approx e^{i\theta_i \frac{\sigma_i}{2}} \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix}. \tag{1.22}$$

The last redefinition involves some linear transformation from the $\phi_{\{1,2\},\{l,r\}}$ to the newly introduced θ_i fields and is a valid approximation because the first order of the series expansion is equivalent.

The exponential in Equation 1.22 looks exactly like a general local SU(2) transformation of a scalar field. For a global SU(2) symmetry, the massless fields θ_i, which depend on spacetime, cannot be gauged away and are called *Goldstone bosons*. However, with the added spin-1 fields W_i^μ the Lagrangian has a local SU(2) symmetry, and the Goldstone bosons can be gauged away. In more graphical language, the three SU(2) gauge bosons *eat up* the Goldstone bosons, explaining how the massive vector bosons get their longitudinal degree of freedom. This argument also explains why the scalar doublet has to be complex, as at least three additional degrees of freedom are required. The fourth remaining degree of freedom h is the *Higgs field*.

Plugging Equation 1.22 into the BEH Lagrangian reveals gauge boson interactions with the Higgs field and Higgs self-interactions. If we leave the Higgs field aside and only consider the vacuum expectation value, we can understand how the SU(2) gauge bosons become massive:

$$\left(D^\mu \frac{v}{\sqrt{2}}\right)^\dagger \left(D_\mu \frac{v}{\sqrt{2}}\right) = \frac{v^2}{8}\left((g')^2\left(\left(W_1^\mu\right)^2 + \left(W_2^\mu\right)^2\right) + \left(g' W_3^\mu - g B^\mu\right)^2\right). \tag{1.23}$$

All the matrix multiplications related to SU(2) representations have been done at this stage. In the next step, the gauge fields must be redefined to obtain terms that can be interpreted as mass terms:

$$\begin{aligned} W_\pm^\mu &= \frac{1}{2}\left(W_1^\mu \mp i W_2^\mu\right), \\ \begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix} &= \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} g W_3^\mu + g' B^\mu \\ g' W_3^\mu - g B^\mu \end{pmatrix} \end{aligned} \tag{1.24}$$

The mass terms look as follows:

$$\frac{1}{4} v^2 g'^2 (W^+)_\mu (W^-)^\mu + \frac{1}{8} v^2 (g^2 + g'^2) Z_\mu Z^\mu. \tag{1.25}$$

As experimentally observed, the photon field A^μ is massless. If we had not used the unitary gauge in which the Goldstone bosons from Equation 1.22 disappear, they would have emerged in Equation 1.24 as parts of the massive vector bosons. The masses of the W and Z bosons are related by the Weinberg angle or weak mxing angle θ_W:

$$\cos\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}} = \frac{m_W}{m_Z}. \tag{1.26}$$

We have seen how the mass terms obtained from the vacuum expectation value of the Higgs field can spontaneously break the SU(2) symmetry at lower energies and how the SU(2) and U(1) gauge bosons mix. In the standard model, the internal $\mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ symmetry is broken to $\mathrm{U(1)_{EM}}$. The subscript Y denotes the hypercharge, which is the conserved quantity of the U(1) symmetry before electroweak symmetry breaking, and EM stands for electromagnetic.

A scalar boson compatible with the Higgs boson has been discovered by the ATLAS and CMS collaborations in 2012 [9, 10], and most of its couplings to vector bosons and fermions have been measured. The measurements confirm that the discovered boson indeed exhibits the couplings to the W and Z bosons expected from the Higgs boson.

1.1.7 Fermion masses

To understand the fermion mass terms, one must first understand the $\mathrm{SU(2)_L}$ representations of the fermions in the standard model. The left-handed leptons come in doublets: a charged massive lepton together with a corresponding neutral massless neutrino. The right-handed charged lepton is a $\mathrm{SU(2)_L}$ singlet. The right-handed neutrino does not exist in the standard model because when the theory was developed, the neutrino appeared massless. For one lepton generation, we write the left-handed doublet and right-handed singlet as

$$l_L = \begin{pmatrix} \nu_{e,L} \\ e_L \end{pmatrix}, \quad e_R. \tag{1.27}$$

In the quark sector, both the up-type and down-type quarks are massive, so we require two right-handed singlets per generation:

$$q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad u_R, \quad d_R. \tag{1.28}$$

Introducing so-called Yukawa interactions with the Higgs field in the Lagrangian gives mass to the fermions:

$$\mathcal{L}_{\text{Yukawa}} = -\lambda_d^{ij} \bar{q}_L^i \Phi d_R^j - \lambda_u^{ij} \bar{q}_L^i i\sigma_2 \Phi^* u_R^j - \lambda_l^{ij} \bar{l}_L^i \Phi e_R^j. \tag{1.29}$$

The indices i and j are generational indices. For the quarks, the Yukawa couplings mix the generations, describing the experimentally observed mixing between quark mass and flavor eigenstates [11].

1.1.8 SU(3) and quantum chromodynamics

Analogous to the discussion in Section 1.1.5 about the SU(2) symmetry, one can impose a local SU(3) symmetry on the standard model. Like SU(2), the SU(3) symmetry group is non-abelian. It has eight generators commonly denoted by the Gell-Mann matrices $\lambda^{a=1,\dots,8}$. The eight generators give rise to eight massless gauge boson fields G_μ^a named gluons. Most standard model fields live in the trivial representation of this gauge group, except for the quarks. Hence, the gluon fields make quarks interact and form hadrons, corresponding to the strong interaction. The quarks are represented by SU(3) triplets, but this time, the triplet's components are not representing fundamentally different particles with different masses, different from what we saw for the SU(2)_L symmetry for the leptons. The three components of the SU(3) triplets are labeled by the *color*, which can be either red, blue, green, or a corresponding anticolor. The coupling constant in the SU(3) extension of the covariant derivative is called the strong coupling constant g_s.

It is necessary to mention that all coupling strengths in the standard model are not constant, but they vary with the energy scale of the probed interaction. This *running* of the coupling constants comes from quantum corrections to the interaction. Different from the electroweak coupling constants, the strong coupling decreases with the energy scale, which is known as *asymptotic freedom* [12, 13]. At lower energies, the large coupling makes it impossible to solve the equations of motions in the perturbation theory framework, and the quarks are subjet to *color confinement*. Therefore, quarks cannot be observed individually, but they are always bound together in color-neutral hadrons. This behavior dictates how quarks and gluons are measured at collider experiments. These particles always hadronize before interacting with the detector. The resulting hadrons decay further, causing in a collimated spray of particles that is known as a *jet*.

1.2 Physics phenomena beyond the standard model

The standard model is an incredibly successful theory that can coherently describe almost all phenomena observed in nature and does not make any predictions that contradict observations. Still, are several directly or indirectly observed phenomena that remain unexplained. These include neutrino masses, the matter-antimatter asymmetry in the Universe, dark matter, dark energy, cosmic inflation, and gravity.

That neutrinos have a mass was decisively concluded from neutrino flavor oscillation experiments at the Super-Kamiokande Observatory [14] and the Sudbury Neutrino Observatory [15]. Flavor oscillation can only be explained if at least two neutrino generations have a small nonzero mass. As discussed in Section 1.1.7, it would be necessary to have corresponding right-handed $SU(2)_L$ singlet to give masses to the left-handed neutrinos via the Yukawa mechanism. Such low-mass electrically neutral $SU(2)_L$ singlets are undetectable. Therefore, it is not easy to test if neutrinos acquire mass in the same fashion as the other standard model particles. This is also highly disputed because of the many magnitudes between the charged lepton and neutrino masses. Possible solutions include more Higgs fields that live in different representations of the standard model symmetries or a theory in which the left- and right-handed neutrinos have vastly different masses. The neutrino might also be its own antiparticle (Majorana fermion), in which case it is possible to write a mass term that does not break gauge invariance without adding right-handed neutrinos. This possibility can be studied experimentally by measuring neutrinoless double-beta decays [16]. Any of these solutions would profoundly extend the standard model. Hence, it is hard to dispute that the neutrino masses are physics beyond the standard model.

Our Universe contains much more matter than antimatter, which can only be explained if there are CP-violating processes that do not have the same probability after a simultaneous charge conjugation and parity transformation. Such CP violation was first observed indirectly via kaon-antikaon oscillations and later directly measured in the quark sector, where the mixing of flavor and mass eigenstates can cause CP violation. In 2020, CP violation in the lepton sector has been reported by the T2K collaboration with a statistical significance above 3 σ [17]. It is not clear yet whether CP violation is more significant in the quark or lepton sector, but the CP violation in the standard model is not sufficient to explain the matter-antimatter asymmetry in our Universe.

The baryonic matter described by the standard model makes up only about 5 % of the Universe's energy content. Another $\sim$27 % corresponds to dark matter, which has only been measured via the gravitational effects it causes. Two prominent dark-matter effects are the modified rotation velocity of galaxies as a function of the radius and gravitation lensing. After decades of searches, there is no indication that dark matter interacts via any force other than gravity, and the order of magnitude of the mass of potential dark matter particles is uncertain.

The ongoing expansion of the Universe can not be explained by the baryonic and dark matter alone. The observed redshift of the Cosmic Microwave Background (CMB) – as measured by the Planck satellite [18] – necessitates an additional energy term, corresponding to the remaining $\sim$68 % in the energy balance. However, this can be modeled by introducing an intrinsic energy of space, called the cosmological constant.

The two unexplained cosmological observations – dark matter and dark energy – are challenging to consolidate with the standard model's existing particle content, even though they are often mentioned as motivations for standard model precision measurements at the collider energy frontier. There may be an *energy desert* between the electroweak and the Planck scale, meaning that no unknown particles have a mass somewhere in between. However, the Higgs boson discovery opened a new avenue that connects the energy frontier with cosmology that might even benefit from the energy desert, as no additional physics needs to be considered in extrapolations to very high energy scales. Cosmological observations hint that brief exponential expansion of space took place around 10^{-36} s after the Big Bang, seeding the formation of large-scale structures in our Universe [19]. The particle-physics explanation for this *cosmic inflation* is unknown, but scalar *inflaton* fields are ubiquitous in early Universe cosmology.

Being the only scalar fundamental field that is experimentally confirmed, the Higgs field might have profound implications on primordial universe cosmology, and the other way around [20]. The possible metastability of the standard model vacuum expectation value might also have consequences on the evolution of the Universe.

Gravity is challenging to model as a quantum field theory because the internal symmetry related to gravity is spacetime itself. However, gravity is much weaker than the other forces of nature, and quantum-gravitational effects would only appear near the Planck energy scale, which cannot be directly probed experimentally. For energies below, gravity is very well modeled by the theory of general relativity. Therefore, the standard model not including gravity is not a problem for our understanding of nature at accessible energy scales, even though quantum gravity is an intriguing theoretical challenge. It should be highlighted that general relativity is a gauge theory just like the standard model, so the two theories are not as far away from each other as one might think.

This section outlined most avenues that can be taken to study new physics beyond the standard model. At the energy frontier, Higgs physics is particularly interesting, as the presence of a fundamental scalar field raises various questions. These questions concern the shape of the Higgs potential with its implications on vacuum stability, the connection between the Higgs field and cosmology, and also the problem whether a model with a fundamental scalar is even self-consistent at all energies [21]. Currently, the energy frontier is around the electroweak scale, presenting the opportunity to precisely measure interactions between the Higgs boson, the weak bosons, and the most massive fermion that dominates the quantum corrections for scalar fields, i.e., the top quark.

1.3 Triboson production as a test of the standard model

The introduction to the standard model explained how the non-abelian electroweak gauge group leads to self-interactions among the weak bosons and how the interactions between the weak bosons and the Higgs fields are fundamental in electroweak symmetry breaking. This enables a rich physics program in the electroweak sector, studying the production of vector bosons in proton collisions confirms the standard model's gauge structure and the electroweak symmetry breaking process. Furthermore, physics beyond the standard model might affect the vector boson self-coupling, causing anomalous triple gauge couplings (aTGCs) or anomalous quartic gauge couplings (aQGCs). Limits of aTGCs can be inferred by measuring diboson production and Vector Boson Fusion (VBF) processes. To probe the quartic coupling of vector bosons and constrain the aQGCs, one can measure Vector Boson Scattering (VBS) or triboson production.

Vector boson scattering probes the interaction of bosons emitted from two colliding quarks with two vector bosons in the final state. The two quarks can be measured in the detector as two jets at small scattering angles in both directions, which is the distinct signature to identify VBS processes. At leading order, the probability of VBS processes contains four powers of α_{EW} if one does not consider the decay of the weak bosons. The same VVjj signature can also be caused by processes proportional to $\alpha_{\mathrm{EW}}^2\alpha_{\mathrm{QCD}}^2$, for example, if two gluons and two vector bosons are radiated from the interacting fermions. This QCD background is the main challenge in VBS measurements. However, the large angle between the jets for the purely electroweak process allows for a good separation of the VBS process, making it possible to isolate the vector boson self-interaction that is regularized by the Higgs boson.

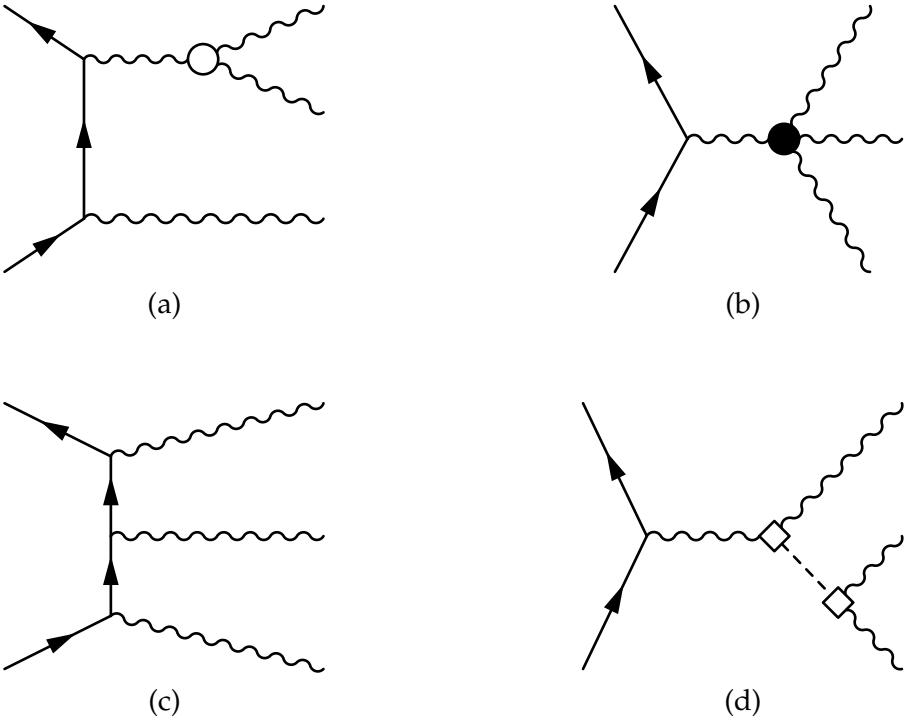


Figure 1.3: Example tree-level Feynman diagrams for the production of three massive vector bosons. The solid line represents the fermions in the initial state, i.e., quarks in the case of proton collisions. The wiggly lines indicate either a W or a Z boson, and the dashed line stands for the Higgs boson. This figure highlights the trilinear gauge couplings (○), the quartic gauge coupling (●), and the interactions between Higgs and gauge bosons (◇).

Triboson production – for which most relevant tree-level diagrams are shown in Figure 1.3 – generally has lower cross-sections than VBS because of more kinematic constraints and the more massive final state, even though the probability is only proportional to α_{EW}^3. Furthermore, the contributions from the radiation of weak bosons from the fermion line (as in Figure 1.3c) cannot be separated via the kinematic properties of additional jets. Still, probing the triboson final state that does not overlap with VBS and is kinematically very different is essential to complete the study of aQGCs at the LHC.

Measuring the triboson final state also opens up several unique opportunities. Potential aQGCs are thought to be caused by new physics at higher energies beyond the electroweak scale, so probing a final state with a higher invariant mass than the diboson system in the VBS final state might be an advantage. With three vector bosons in the final state that were directly involved in the self-interaction process, more observables are available to probe the physics that might be at the origin of aQGCs. The role of the Higgs diagram is also different in triboson production. For VBS, the diagram with a Higgs boson in the T channel strongly interferes with the other diagrams, unitarizing the cross-section. For triboson production, the Higgs diagram – shown in Figure 1.3d – has an on-shell Higgs boson in the S channel, and the interference with the other diagrams is negligible. The diagrams without the Higgs produce on-shell vector bosons, while one vector boson from the Higgs decay is highly off-shell. Because of this little overlap, one should make an disctinction between *on-shell* and *Higgs mediated* (or *Higgs associated*) boson production.

Before the latest data-taking runs of the LHC, triboson production could not be measured because the process is very rare if one considers the leptonic final states, and the (semi)hadronic final states have ample background.

1.3.1 Triboson production at the LHC

Measurements sensitive to triple and quartic gauge couplings date back to the LEP and Tevatron experiments [22, 23, 24, 25], but there were no direct searches for triboson production before the LHC era. The LHC experiments so far carried out only a few measurements. The ATLAS experiment published a search for WWW production based on the LHC Run 1 dataset with $\sqrt{s}$ = 8 TeV [26] and a seach for the more general WVV production with parts of the LHC Run 2 dataset [27], showing evidence for the WVV process. The CMS collaboration published a search for WWW production with data from the first Run 2 data-taking year [28].

The cross-section for triboson production in proton-proton collisions at $\sqrt{s}$ = 13 TeV are listed in Table 1.1, combining predictions for on-shell VVV production and the VH $\rightarrow$ VVV process. The predictions for on-shell VVV production include next-to-leading-order (NLO) QCD corrections. These are often the dominant corrections because of the larger value of the interaction strength α_{QCD} compared to α_{EW}, the coupling strength of the electroweak interaction. Even if not considered in this thesis, it is important to mention that recently, computations of the VVV cross-sections that also include NLO electroweak contributions were published [29, 30, 31, 32] The VH $\rightarrow$ VVV cross-sections listed in Table 1.1 consider both QCD and electroweak corrections at NLO. The uncertainties on all the listed cross-sections are approximately 10 %.

Quantity	WWW	WWZ	WZZ	ZZZ	Reference
$\sigma_{pp\rightarrow\mathrm{VVV\ non\text{-}VH}}$ (fb)	293.4	165.1	36.0	23.1	NLO QCD [33, 34, 35, 36]
$\sigma_{\mathrm{VH}\rightarrow\mathrm{VVV}}$ (fb)	216.0	188.9	55.7	14.0	NNLO QCD + NLO EW [37, 38, 39]
$\sigma_{\mathrm{tot.}}$ (fb)	509.4	354.0	91.6	37.1	
$\mathcal{B}_{\mathrm{VVV}\rightarrow SS}$ (%)	7.16	-	-	-	
$\mathcal{B}_{\mathrm{VVV}\rightarrow 3\ell}$ (%)	3.46	4.82	6.37	-	
$\mathcal{B}_{\mathrm{VVV}\rightarrow 4\ell}$ (%)	-	1.16	0.81	3.22	branching fractions from [40]
$\mathcal{B}_{\mathrm{VVV}\rightarrow 5\ell}$ (%)	-	-	0.39	-	
$\mathcal{B}_{\mathrm{VVV}\rightarrow 6\ell}$ (%)	-	-	-	0.13	
$\sigma_{\mathrm{tot.}} \times \mathcal{B}_{\mathrm{VVV}\rightarrow SS}$ (fb)	36.4	-	-	-	
$\sigma_{\mathrm{tot.}} \times \mathcal{B}_{\mathrm{VVV}\rightarrow 3\ell}$ (fb)	17.6	17.1	5.83	-	
$\sigma_{\mathrm{tot.}} \times \mathcal{B}_{\mathrm{VVV}\rightarrow 4\ell}$ (fb)	-	4.12	0.74	1.19	
$\sigma_{\mathrm{tot.}} \times \mathcal{B}_{\mathrm{VVV}\rightarrow 5\ell}$ (fb)	-	-	0.36	-	
$\sigma_{\mathrm{tot.}} \times \mathcal{B}_{\mathrm{VVV}\rightarrow 6\ell}$ (fb)	-	-	-	0.05	
$\sigma_{\mathrm{tot.}} \times \mathcal{B}_{\mathrm{VVV}\rightarrow SS} \times L$	4987	-	-	-	
$\sigma_{\mathrm{tot.}} \times \mathcal{B}_{\mathrm{VVV}\rightarrow 3\ell} \times L$	2411	2343	799	-	
$\sigma_{\mathrm{tot.}} \times \mathcal{B}_{\mathrm{VVV}\rightarrow 4\ell} \times L$	-	564	101	163	
$\sigma_{\mathrm{tot.}} \times \mathcal{B}_{\mathrm{VVV}\rightarrow 5\ell} \times L$	-	-	49.3	-	
$\sigma_{\mathrm{tot.}} \times \mathcal{B}_{\mathrm{VVV}\rightarrow 6\ell} \times L$	-	-	-	6.85	

Table 1.1: Summary of signal process cross-section, branching fractions, and total expected number of events produced in the LHC Run 2, with an integrated luminosity of L = 137 fb^{-1}. The uncertainties on the cross-sections are typically 10 %. For the branchings, the subscript SS stands for a pair of *same-sign* charged leptons. The full WWW to dileptons final state cannot be measured because backgrounds are too dominant.

Table 1.1 also shows the branching fractions for final states with different numbers of charged leptons. Because of the abundant QCD background at lepton hadron colliders, final states with more charged leptons have a higher signal purity. Finally, the table shows the number of predicted events in all relevant final states with 137 fb^{-1} of collision data, which corresponds to the LHC Run 2 dataset. A large fraction of the events is lost because not all leptons go in the detector's fiducial volume. In the WWZ $\rightarrow 4\ell$ channel, approximately 50 % of the events are not fiducially accepted by the CMS detector.

Having at least one produced Z boson is very helpful for background suppression, as both charged leptons from the Z boson decay are reconstructed, and invariant-mass requirements can be imposed to select Z boson candidates purely. This makes the WWZ process very appealing for precise measurements, particularly in the four-lepton channel where all the weak bosons decay leptonically. Processes with more leptonically-decaying Z bosons are much rarer and difficult to access even with the LHC Run 2 dataset. Therefore, the WWZ $\rightarrow 4\ell$ channel is particularly promising for triboson measurements, which is why it is in the focus of the measurements presented in this thesis.

1.3.2 The effective field theory framework

Many theories of Physics beyond the Standard Model (BSM) theories were developed in the last decades to explain the unexplained phenomena listed in Section 1.2, or make the standard model mathematically and philosophically more appealing by introducing new fields that make the theory have less numerical coincidences. However, it is not practical to test all of the numerous BSM models at the LHC. A more generic framework to search for BSM physics that became very popular is the *effective field theory (EFT)* framework [41, 42]. Complementing *bump hunt* analyses that look for new particle resonances at the mass scale probed by the collider, the EFT framework allows for a consistent interpretation of process rate enhancements in the high-mass tails of distributions. In other words, energy scales that can not be probed directly because they are beyond the collision energy can still be probed indirectly.

If a new particle with mass Λ would mediate an interaction between standard model particles with a dimensionless coupling constant f, the propagator that enters the scattering amplitude and generally depends on the scattering momentum p can be approximated by a constant if $\Lambda^2 \gg p^2$:

$$\frac{f^2}{p^2 - \Lambda^2} = -\frac{f^2}{\Lambda^2}\left[1 + \frac{p^2}{\Lambda^2} + \frac{p^4}{\Lambda^4} + \dots\right] \xrightarrow[\Lambda^2 \gg p^2]{} -\frac{f^2}{\Lambda^2}. \tag{1.30}$$

Hence, the BSM effect can be modeled with a coupling constant of mass dimension -2. If such dimensionful couplings are allowed in the Lagrangian of the theory – which has to be of mass dimension 4 – it is possible to introduce new dimensionful operators, albeit generally at the cost of the theory's renormalizability for a finite number of expansion terms.

Operators of odd dimension break lepton or baryon number conservation, so they are rarely considered. Therefore, as suggested by the example in Equation 1.30, the lowest-dimensional effective operators are of dimension 6. Operators that consistently extend the standard model to a *standard model effective field theory (SMEFT)* need to keep the Lagrangian invariant under the standard model gauge symmetries, so there is a limited set of operators $\mathcal{O}_i^d$ of dimension d that might extend the standard model Lagrangian:

$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \sum_{d>4}\sum_i \frac{f_i \mathcal{O}_i^d}{\Lambda^{d-4}}. \tag{1.31}$$

The new coupling strengths f in the EFT Lagrangian are usually called *Wilson coefficients*. Some of the dimension 6 operators modify the trilinear gauge couplings, so the EFT framework can be used to set interpretable limits on the aTGCs. The same connection exists between the dimension 8 operators and the aQGCs, so dimension 8 operators are vital for the general interpretation of VBS and triboson measurements. The relevant dimension 8 operators, as defined in [43], are:

$$
\begin{aligned}
\mathcal{L}_{S,0} &= \left[\left(D_\mu\Phi\right)^\dagger D_\nu\Phi\right] \times \left[\left(D^\mu\Phi\right)^\dagger D^\nu\Phi\right] \\
\mathcal{L}_{S,1} &= \left[\left(D_\mu\Phi\right)^\dagger D^\mu\Phi\right] \times \left[\left(D_\nu\Phi\right)^\dagger D^\nu\Phi\right] \\
\mathcal{L}_{M,0} &= \mathrm{Tr}\left[\hat{W}_{\mu\nu}\hat{W}^{\mu\nu}\right] \times \left[\left(D_\beta\Phi\right)^\dagger D^\beta\Phi\right] \\
\mathcal{L}_{M,1} &= \mathrm{Tr}\left[\hat{W}_{\nu\beta}\hat{W}^{\mu\nu}\right] \times \left[\left(D_\beta\Phi\right)^\dagger D^\mu\Phi\right] \\
\mathcal{L}_{M,2} &= \left[B_{\mu\nu}B^{\mu\nu}\right] \times \left[\left(D_\beta\Phi\right)^\dagger D^\beta\Phi\right] \\
\mathcal{L}_{M,3} &= \left[B_{\mu\nu}B^{\nu\beta}\right] \times \left[\left(D_\beta\Phi\right)^\dagger D^\mu\Phi\right] \\
\mathcal{L}_{M,4} &= \left[\left(D_\mu\Phi\right)^\dagger \hat{W}_{\beta\nu}D^\mu\Phi\right] \times B^{\beta\nu} \\
\mathcal{L}_{M,5} &= \left[\left(D_\mu\Phi\right)^\dagger \hat{W}_{\beta\nu}D^\nu\Phi\right] \times B^{\beta\mu} \\
\mathcal{L}_{M,6} &= \left[\left(D_\mu\Phi\right)^\dagger \hat{W}_{\beta\nu}\hat{W}^{\beta\nu}D^\mu\Phi\right] \\
\mathcal{L}_{M,7} &= \left[\left(D_\mu\Phi\right)^\dagger \hat{W}_{\beta\nu}\hat{W}^{\beta\mu}D^\nu\Phi\right] \\
\mathcal{L}_{T,0} &= \mathrm{Tr}\left[\hat{W}_{\mu\nu}\hat{W}^{\mu\nu}\right] \times \mathrm{Tr}\left[\hat{W}_{\alpha\beta}\hat{W}^{\alpha\beta}\right] \\
\mathcal{L}_{T,1} &= \mathrm{Tr}\left[\hat{W}_{\alpha\nu}\hat{W}^{\mu\beta}\right] \times \mathrm{Tr}\left[\hat{W}_{\mu\beta}\hat{W}^{\alpha\nu}\right] \\
\mathcal{L}_{T,2} &= \mathrm{Tr}\left[\hat{W}_{\alpha\mu}\hat{W}^{\mu\beta}\right] \times \mathrm{Tr}\left[\hat{W}_{\beta\nu}\hat{W}^{\nu\alpha}\right] \\
\mathcal{L}_{T,5} &= \mathrm{Tr}\left[\hat{W}_{\mu\nu}\hat{W}^{\mu\nu}\right] \times B_{\alpha\beta}B^{\alpha\beta} \\
\mathcal{L}_{T,6} &= \mathrm{Tr}\left[\hat{W}_{\alpha\nu}\hat{W}^{\mu\beta}\right] \times B_{\mu\beta}B^{\alpha\nu} \\
\mathcal{L}_{T,7} &= \mathrm{Tr}\left[\hat{W}_{\alpha\mu}\hat{W}^{\mu\beta}\right] \times B_{\beta\nu}B^{\nu\alpha} \\
\mathcal{L}_{T,8} &= B_{\mu\nu}B^{\mu\nu}B_{\alpha\beta}B^{\alpha\beta} \\
\mathcal{L}_{T,9} &= B_{\alpha\mu}B^{\mu\beta}B_{\beta\nu}B^{\nu\alpha}
\end{aligned}
\tag{1.32}
$$

These operators group into scalar (or longitudinal) operators that only involve the scalar field and the covariant derivatives, transversal operator with only the gauge fields, and mixed operators that contain both covariant derivatives of the scalar field and the gauge fields. Triboson production and VBS measurements are both most sensitive to the transverse operators. Figure 1.2 indicates which operators affect a given quartic gauge coupling. It can be seen that the WWZ process, which involves the WWZZ coupling, is sensitive to most of the aQGC operators, except for $\mathcal{L}_{T,8}$ and $\mathcal{L}_{T,9}$, which give rise to fully neutral gauge couplings that are otherwise forbidden in the standard model.

Nonzero aQGC operator values would be manifest in the tail of the distributions of observables sensitive to the scattering process's energy scale, for example, the invariant mass of all leptons in the final state. With three bosons in the final states, it is also possible to extract more information that distinguishes the EFT contribution from the standard model contribution, most notably from observables sensitive to the angular separation between the bosons. A promising observable is, for example, the sphericity of the triboson momenta, or – more appropriate for hadron colliders – the *transverse sphericity* [44]. The WWZ process that has all the three weak bosons in the final state[2], so it also opens up for the first time the possibility to compare the distributions of angles between different bosons in the same event selection. This might be very useful for the interpretation of a possible high-mass event excess. For example, the operators $\mathcal{L}_{T,0}$, $\mathcal{L}_{T,1}$, and $\mathcal{L}_{T,2}$ cause different angular distributions for the three bosons. The current best aQGC limits as measured by CMS are listed in Figures 1.4, 1.5 and 1.6, which also showcase some ATLAS results for comparison.

	WWWW	WWZZ	ZZZZ	WWγZ	WW$\gamma\gamma$	ZZZγ	ZZ$\gamma\gamma$	Z$\gamma\gamma\gamma$	$\gamma\gamma\gamma\gamma$
$\mathcal{L}_{S,0}$, $\mathcal{L}_{S,1}$	✓	✓	✓						
$\mathcal{L}_{M,0}$, $\mathcal{L}_{M,1}$ $\mathcal{L}_{M,6}$, $\mathcal{L}_{M,7}$	✓	✓	✓	✓	✓	✓	✓		
$\mathcal{L}_{M,2}$, $\mathcal{L}_{M,3}$ $\mathcal{L}_{M,4}$, $\mathcal{L}_{M,5}$		✓	✓	✓	✓	✓	✓		
$\mathcal{L}_{T,0}$, $\mathcal{L}_{T,1}$ $\mathcal{L}_{T,2}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
$\mathcal{L}_{T,5}$, $\mathcal{L}_{T,6}$ $\mathcal{L}_{T,7}$		✓	✓	✓	✓	✓	✓	✓	✓
$\mathcal{L}_{T,8}$, $\mathcal{L}_{T,9}$			✓			✓	✓	✓	✓

Table 1.2: Checklist showing which quartic gauge boson coupling is affected by a given dimention 8 operator [43].

[2]The W^+ and W^- bosons are excitations of the same quantum field, so they count as separate bosons.

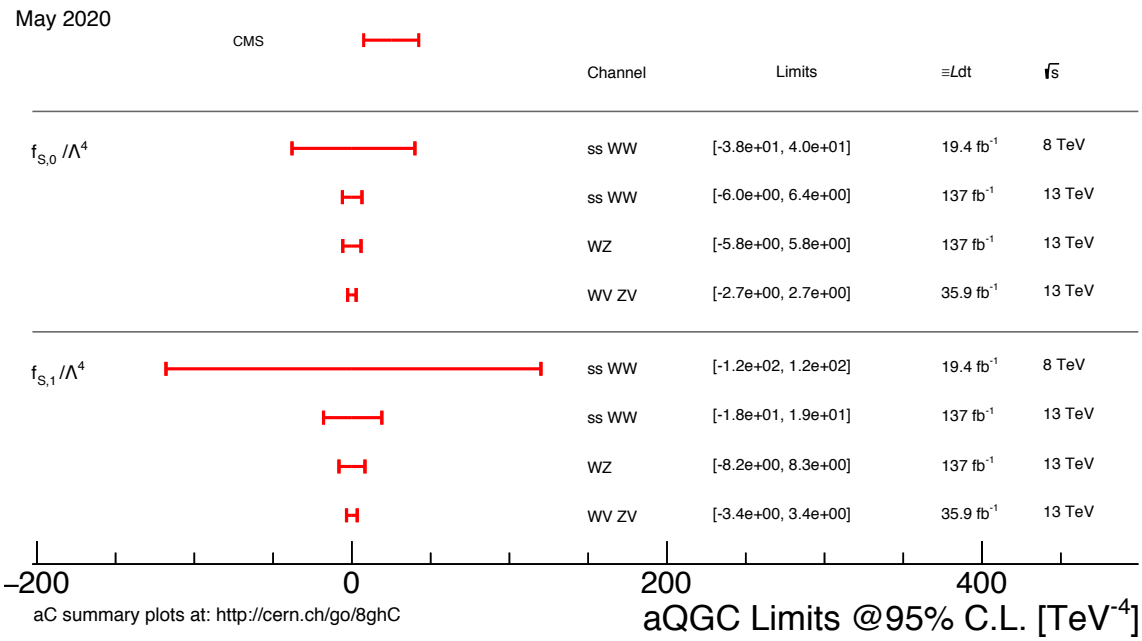


Figure 1.4: Limits on dimension 8 scalar/longitudinal parameters $f_{S,i}$ [45].

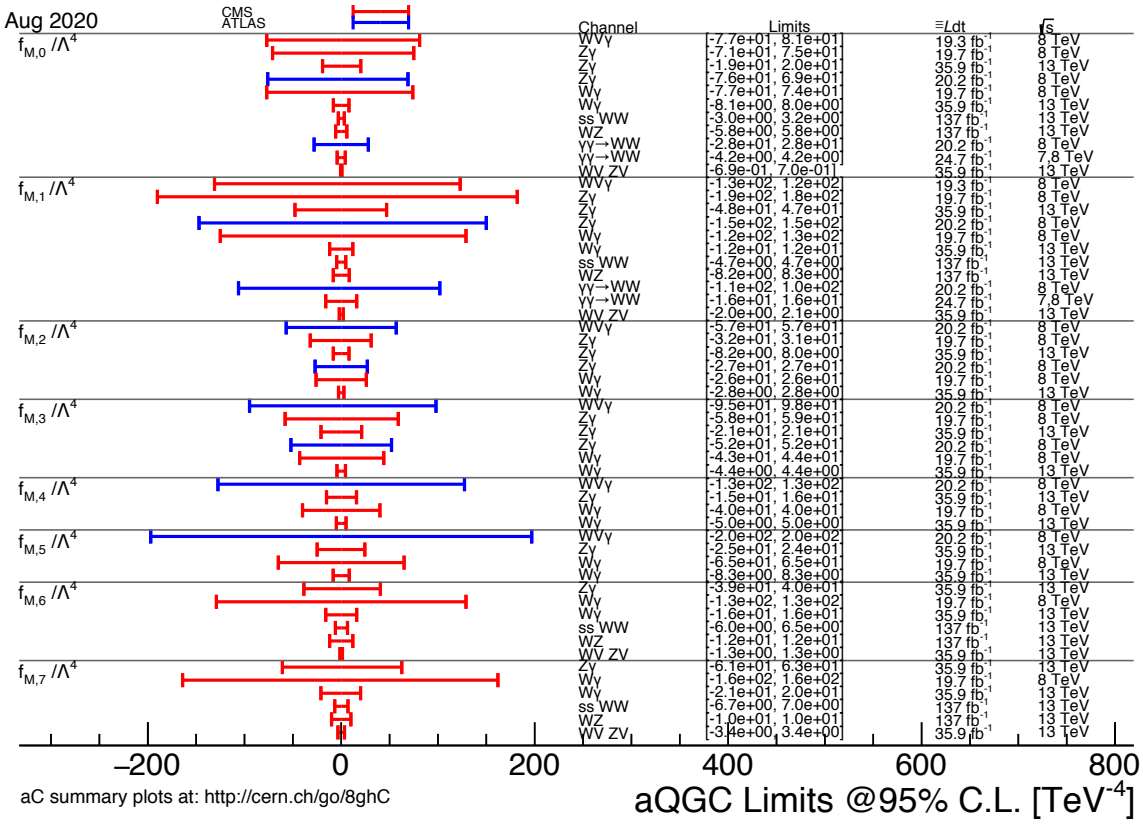


Figure 1.5: Limits on dimension 8 mixed parameters $f_{M,i}$ [45].

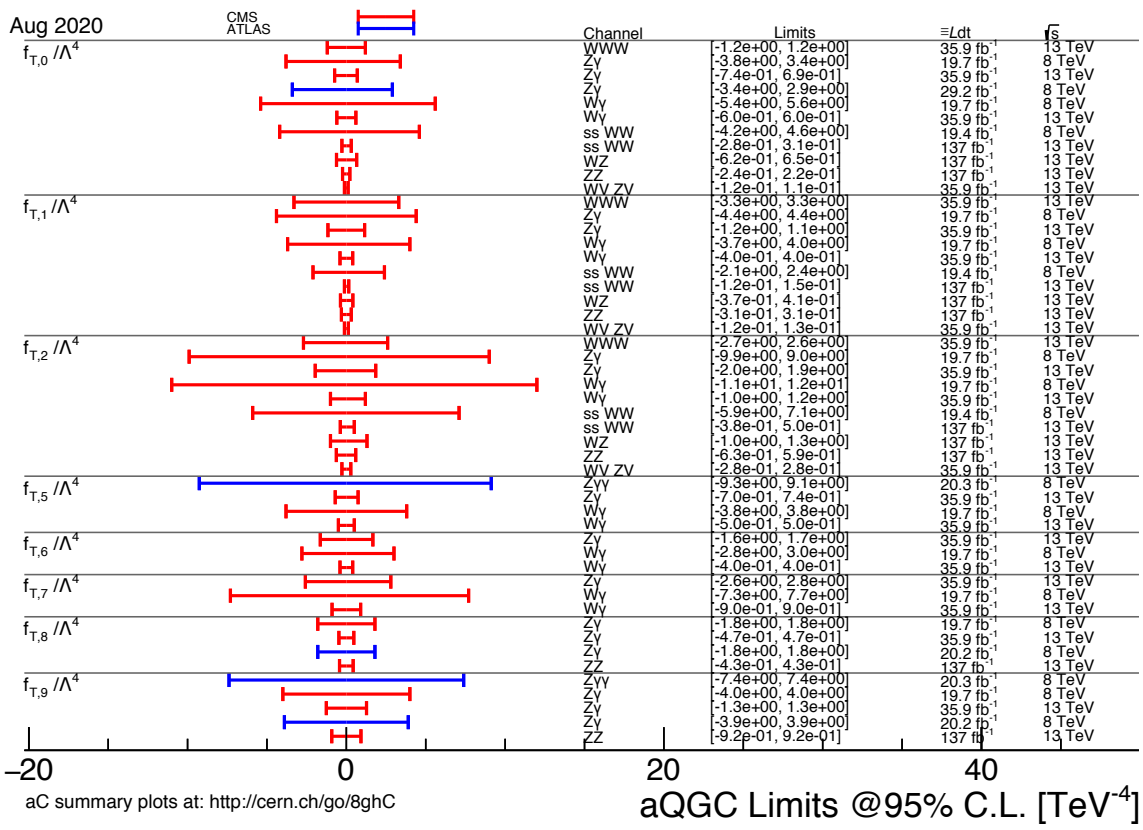


Figure 1.6: Limits on dimension 8 transversal parameters $f_{T,i}$ [45].

Chapter 2

The CMS experiment at the Large Hadron Collider

The Large Hadron Collider (LHC) is a collider for subatomic particles built and operated by the European Organization for Nuclear Research (CERN) underneath the French-Swiss border near Geneva. It primarily produces proton-proton (pp) collisions at the center of mass energy of 13 TeV, the current world record. It can also accelerate beams of heavy ions and regularly conducts lead-lead or proton-lead collisions. A large part of the existing CERN accelerator complex serves as an injector chain to feed the LHC with protons at an energy of 450 GeV. Superconducting magnets in the LHC then force the particle beams onto a circular trajectory, while superconducting radiofrequency cavities accelerate the beam on each turn.

The accelerated particle bunches collide at four different interaction points, where particle detectors measure the collision remnants. The four detectors specialize in different areas of high-energy physics. The LHCb experiment [46] is dedicated to the physics of hadrons that contain a bottom quark (B hadrons), for example, rare decays, excited states, and CP properties. The A Large Ion Collider Experiment (ALICE) detector [47] focuses on lead-lead collisions, studying strong interaction physics and the properties of quark-gluon plasma to understand the structure of hadronic matter. Finally, the two general-purpose detectors ATLAS [48] and CMS [49] reconstruct almost the complete collision final state, so they are often associated with a variety of questions within high-energy frontier physics. However, they were designed primarily to learn how the electroweak symmetry is broken at the TeV scale, be it via the Higgs mechanism or otherwise.

In this thesis, LHC proton-proton collision data recorded by the CMS detector in 2016, 2017, and 2018 have been analyzed. This period is referred to as LHC Run 2. Section 2.1 reviews the design and performance of the LHC to an extent necessary for the understanding of collision data. Section 2.2 gives a full overview of the CMS detector, which covers the various subdetectors and notions of the trigger and data acquisition system.

2.1 The Large Hadron Collider

The year 1984 is often mentioned as a milestone in the genesis of the LHC: the first LHC workshop took place in Lausanne, and the first LHC studies were presented to the International Committee for Future Accelerators (ICFA) [50]. Twenty-five years later, in 2009, the ALICE collaboration published the first scientific results based on LHC collision data [51]. In the context of the CMS and ATLAS experiments, which require a higher collision rate, the year 2011 marks the start of data-taking for physics. In the 2020s, the LHC will be upgraded to the High Luminosity LHC (HL-LHC), which goes along with significant detector upgrades. With the HL-LHC upgrade, the collider should also reach the nominal collision energy of 14 TeV.

2.1.1 The CERN accelerator infrastructure

The European Organization for Nuclear Research (CERN) was founded in 1954 and operates a very successful family of hadron accelerators – among other facilities. The original Proton Synchrotron (PS) provided the 25 GeV beams for the Gargamelle experiment that discovered weak neutral current interactions in 1973. Another notable achievement of CERN related directly to the weak interaction was the W and Z boson discovery by the UA1 and UA2 experiments in 1983. This discovery was powered by the Super Proton Synchrotron (SPS)[1], which can reach proton energies of 450 GeV.

Different effects dominate beam dynamics in different energy ranges. There is no single accelerator design that is adequate for all energies. Hence, the PS and SPS now serve as sequential pre-accelerators for the LHC. The LHC resides in a 26.7 km underground tunnel that is, on average, approximately 100 m below ground level. Initially, the tunnel housed the Large Electron-Positron Collider (LEP), decommissioned in the year 2000.

The protons are obtained from hydrogen by stripping off the electrons with electric fields. They consecutively pass through a linear accelerator (LINAC 2), the Proton Synchrotron Booster (PSB), the PS, and the SPS before finally entering the LHC. With the HL-LHC upgrades, the linear accelerator will be replaced with the new LINAC 4 [52]. Instead of accelerating protons directly, the LINAC 4 will accelerate negative hydrogen atoms, which lose all electrons by going through a stripping foil in the PSB. Using H^- ions allows for charge-exchange injection to the synchrotron, where the beam can be accumulated over many turns [53]. This makes it possible to reach the beam brightness[2] required by the HL-LHC.

Figure 2.1 shows an overview of the CERN accelerator complex. One recent highlight besides the LHC is the Advanced WAKEfield Experiment (AWAKE), which conducts research and development on plasma wakefield acceleration. Another new highlight is the Extra Low ENergy Antiproton (ELENA) storage ring, which decelerates antiproton beams to the energy of 0.1 MeV for antihydrogen production. Having such electrically neutral anti-atoms enables the measurement of gravitational effects on antimatter.

2.1.2 Design and parameters of the LHC

A circular particle accelerator like the LHC increases the energy of charged particles by sending them through radiofrequency (RF) resonator cavities. Dipole magnets force the particles onto a circular path where they repeatedly pass through the same accelerating field. Heavier particles like hadrons require stronger magnetic fields but only lose little energy via synchrotron radiation. Therefore, the LHC's most important building blocks are the cylindrical vacuum vessels with the superconducting dipole magnets, shown in Figure 2.2. The particles in the two parallel beam pipes travel in opposite directions to collide head-on at the interaction points. There are 1232 of these approximately 15 m long dipoles distributed along eight circular arcs. Liquid helium at 1.9 K cools down the niobium-titanium superconductor, which has to conduct up to 11.7 kA of electrical current. The magnetic field is ramped up from 0.535 T up to 7.8 T in the case of 6.5 TeV proton beams. The 16 RF cavities are within one of the straight sections between the dipole arcs. Their field oscillates with around 400 MHz depending slightly on the proton energy, which is increased by 0.5 MeV over each turn. For comparison, the energy loss by synchrotron radiation is 7 keV per revolution for 7 TeV beams.

[1] Or, more precisely, the Super Proton–Antiproton Synchrotron (S$\bar{p}$pS), which is how the SPS was called while it operated as a proton-antiproton collider.

[2] The beam brightness is the current per transverse emittance: $B = \frac{I}{8\pi\epsilon_x\epsilon_y}$

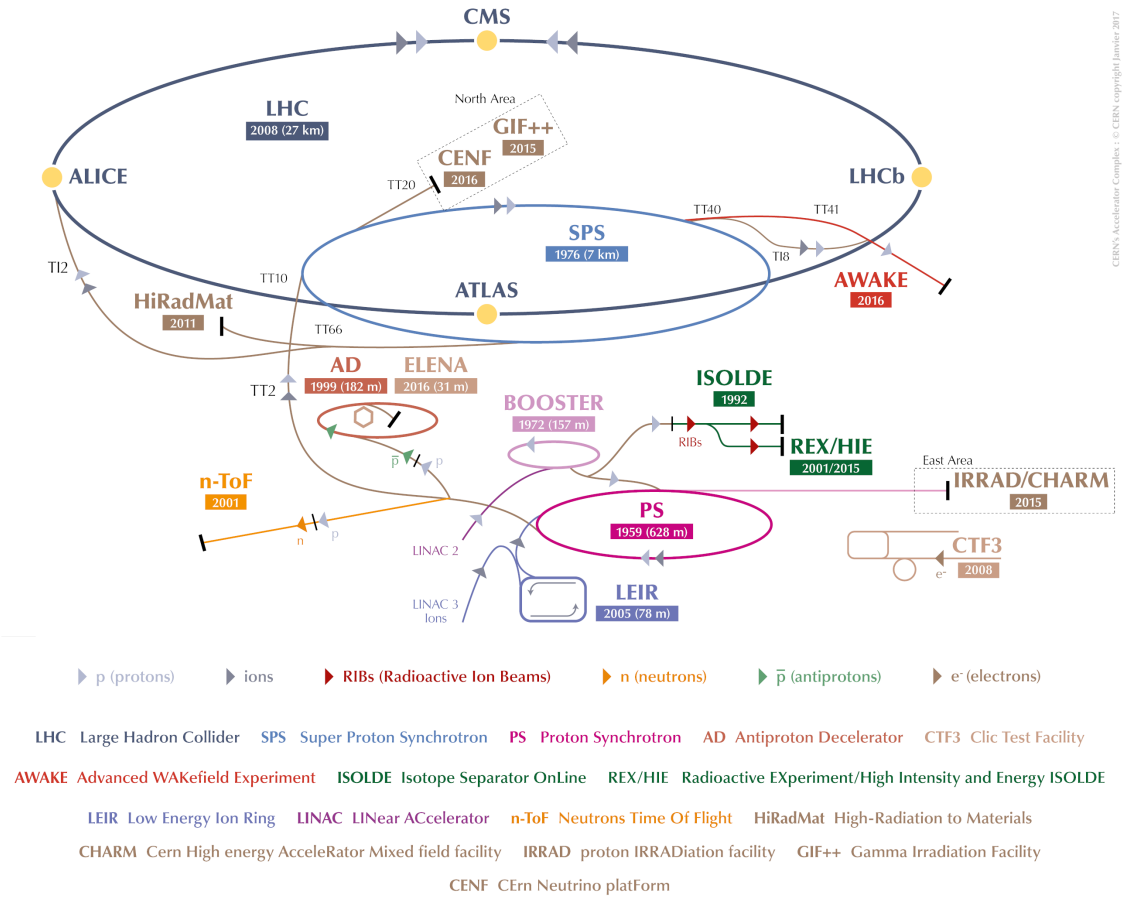


Figure 2.1: The CERN accelerator complex [54]. Note that after the second long LHC shutdown in 2019–2020, LINAC 2 will be replaced with LINAC 4.

Besides acceleration, the RF cavities have a bunching effect on the beam. The particle injection into the LHC is timed such that they are synchronized with the falling slope of the oscillating electric potential. At this phase, relativistic particles that are injected a bit ahead are accelerated slightly more. The radius of their trajectory extends because of the relativistic mass increase, and the stretched path causes the particles to fall behind. Particles that are initially behind are accelerated less, which brings them on a shorter trajectory so they can catch up. Therefore, the protons are stabilized in so-called *RF buckets*. The RF frequency is 35640 times higher than the accelerated hadrons' revolution frequency, so theoretically, there are as many RF buckets to fill. However, the PS provides protons with a bunch spacing of 25 ns, and one needs an *abort gap* in the bunch structure to have a time window for switching the kicker magnets in case of a beam abort. Hence, 2808 RF buckets are filled.

Dipole magnets and RF cavities alone are not sufficient for a stable beam. A particle-beam is smeared in both longitudinal and transverse position-momentum phase space. In general, the spread of the beam in phase space (*emittance*) is be preserved – according to the Liouville theorem –, and the *beta functions* β_x and β_y describe the evolution of the relative transverse emittance. It was discussed before how RF cavities have a focusing effect in the longitudinal direction. A focusing effect in *one* transverse direction is achieved with quadrupole magnets at the cost of defocusing in the other direction. Therefore, the magnetic lattice consists of alternating focusing and defocusing quadropole magnets, resulting in a net focusing effect on the beam. Along most of the beamline, the beta functions oscillate with the spacing of the focusing and defocusing quadrupoles.

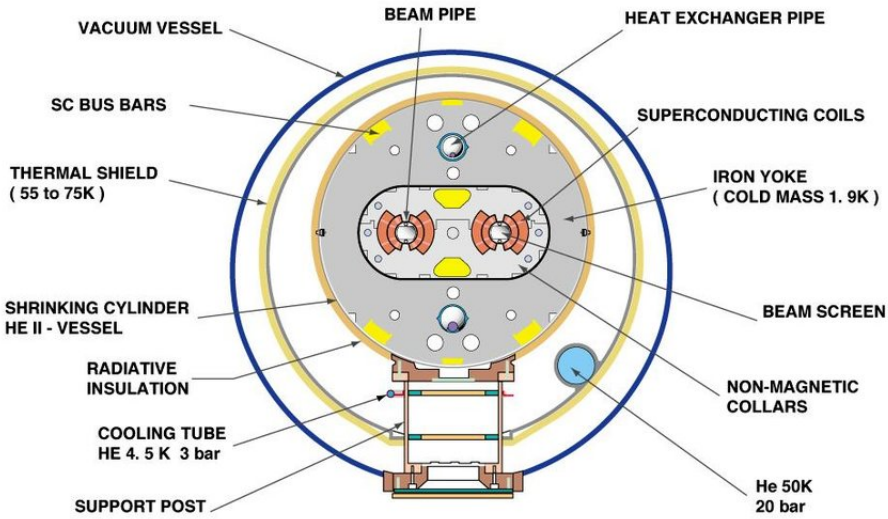


Figure 2.2: Cross-section of LHC vacuum vessel with superconducting dipole magnet and beam pipes [55].

The beta functions value at the interaction point, called *beta star*, is of particular importance for a high collision rate. Note that there also are sextupole, octupole, and decapole magnets in the LHC, which are mandatory to correct for undesired effects on the beam dynamics, such as the energy-dependent focusing of the quadrupoles (chromaticity) or dipole edge effects.

The collision rate at the LHC is the collision processes cross-section σ times the instantaneous luminosity L of the machine. For a Gaussian beam distribution, it is

$$L = \frac{N_p^2 n_b f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} F. \tag{2.1}$$

The meaning of most parameters is explained in Table 2.1, and γ_r is the relativistic Lorentz factor that is around 7460 for a beam energy of 7 TeV. The geometric factor F accounts for the crossing angle between the beams at the interaction point. If the beams profile is assumed to be circular in the transverse plane, this factor is

$$F = \left[1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right]^{-1/2}. \tag{2.2}$$

The crossing angle θ_c is necessary to constrain the interaction region in the longitudinal direction. During an LHC fill, the number of protons per bunch decreases as they collide and – to a lesser extent – interact with the gas in the beam pipe. To ensure the luminosity delivered to the experiments is still approximately constant, the LHC can counter this effect by adapting the beta function or the beam overlap. If the instantaneous luminosity is high enough, multiple inelastic pp collisions can pile up in one bunch crossing. The average pileup is

$$\langle PU \rangle = \frac{L\sigma}{n_b f_{\text{rev}}}. \tag{2.3}$$

For the high luminosity experiments CMS and ATLAS, the nominal luminosity of pp collisions is $L = 10^{34}$ cm^{-2}s^{-1}, so the average pileup for an inelastic cross-section of 79.5 mb [56] should be around 25. For some subdetectors, the response is longer than the

bunch spacing of 25 ns. Therefore, from the point of view of the experiments, pileup is often categorized into *in-time* and *out-of-time* pileup, with different correction algorithms for both types (see Section 3.5 for a discussion of in-time pileup correction).

		Beam 7 TeV
RF system frequency		400.790 MHz
Bunch separation	Δt	25 ns
Number of bunches	n_b	2808
Number of protons per bunch	N_p	1.15×10^{11}
Revolution frequency	f_{rev}	11245 Hz
Transverse bunch size RMS at the interaction point	σ^*	16.7 µm
Longitudinal bunch size RMS	σ_z	7.55 cm
Beta function at the interaction point	β^*	0.55 m
Crossing angle at the interaction point	θ_c	285 µrad
Transverse emittance	ϵ_n	3.75 µm
Synchrotron radiation loss/turn		7 keV

Table 2.1: LHC parameters for the nominal pp collision energy of 14 TeV [55].

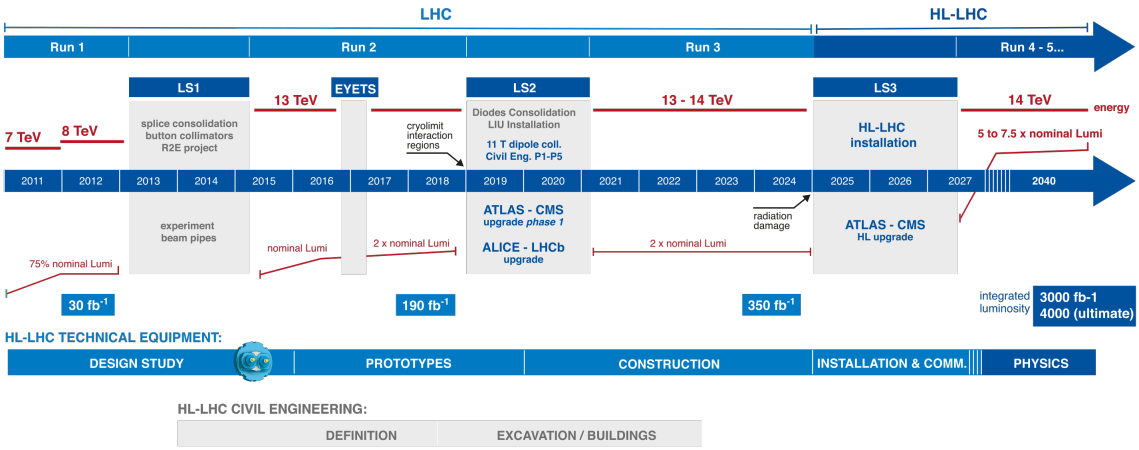


Figure 2.3: The LHC schedule between 2011 and the 2040s [57]. Between the data-taking periods, there are long shutdowns (LS) and an extended end-of-year technical stop before operation in 2017. In the third LS, the High Luminosity LHC (HL-LHC) will be installed, and the ATLAS and CMS detectors will be upgraded to deal with the increased luminosity.

2.1.3 Operations history and future

On September 19th, 2008, a few days after the first beam circulation, a discharge of a faulty electrical connection inflicted severe damage to the collider, affecting 53 dipole magnets. The electrical resistance of the electrical connections between the dipole magnets was too high to sustain the current required for nominal beam energy, causing them to overheat and melt during a ramp-up test without beam. At this point, CERN had to decide if it should either replace all of the problematic connectors – causing further delays – or start operating at lower energy after minimal repairs. The management opted for a quick restart at a collision energy of 7 TeV in 2010. In 2012, the collision energy was raised to 8 TeV before the electrical connections were finally replaced in the first long shutdown that marked the end of the LHC Run 1.

For the LHC Run 2 between 2015 and 2018, the collision energy was not raised to the nominal 14 TeV because the training of the dipole magnets would have taken too long. Figure 2.4 shows the instantaneous and integrated luminosity delivered to the CMS experiment for both Run 1 and Run 2. The design luminosity of 10^{34} cm^{-2}s^{-1} was not reached in 2015, while it has been significantly exceeded in the following years. Accordingly, only a few percent of the collisions recorded by CMS and ATLAS experiments in Run 2 was produced in 2015. Furthermore, the 2015 collisions are distinguished by their lower pileup, as shown in Figure 2.5. This asymmetric dataset size is undesirable for an experimentalist: the combination of datasets measured under different conditions introduces additional systematic uncertainties and requires dedicated simulations for the different conditions. The ATLAS collaboration nevertheless decided to combine all four years for most of their full Run 2 analyses. The CMS collaboration generally does not consider the 2015 dataset, which is also the case for this thesis's analyses.

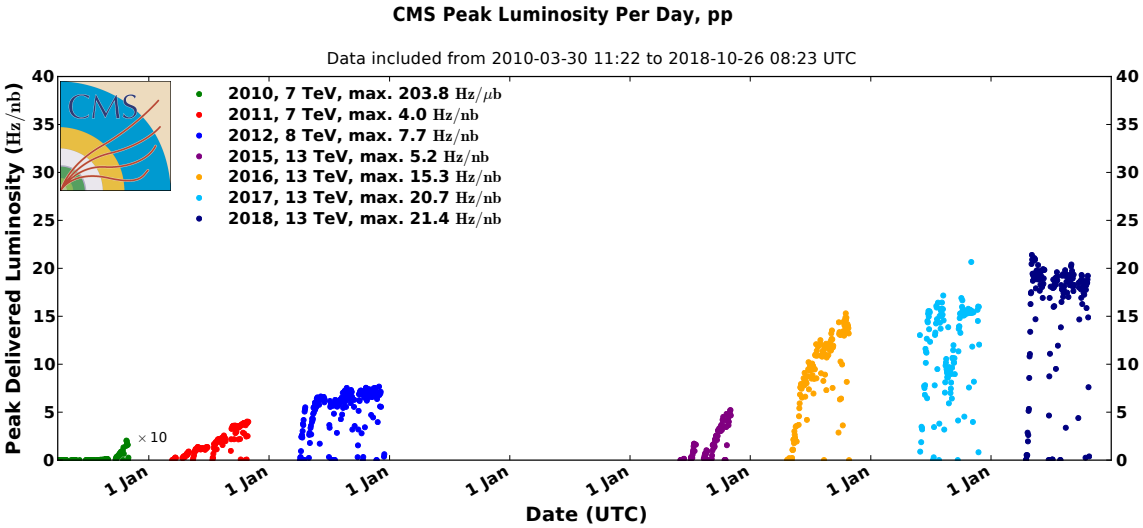
After the current long shutdown (LS2), the LHC will enter another three years of data-taking (Run 3), during which the design luminosity is expected to be exceeded by at least a factor of two. With the upgrade to the HL-LHC before the subsequent Run 4, the collider should nominally reach an instantaneous luminosity of $L = 5 \times 10^{34}$ cm^{-2}s^{-1}, which is five times higher than the design luminosity of the original LHC. The total integrated luminosity delivered to each of the high-luminosity experiments before the long shutdowns is indicated in Figure 2.3. The ultimate target for a successful HL-LHC is 4000 fb^{-1}, which would be around 21 times more than the luminosity collected so far at the end of Run 2.

2.2 The CMS experiment

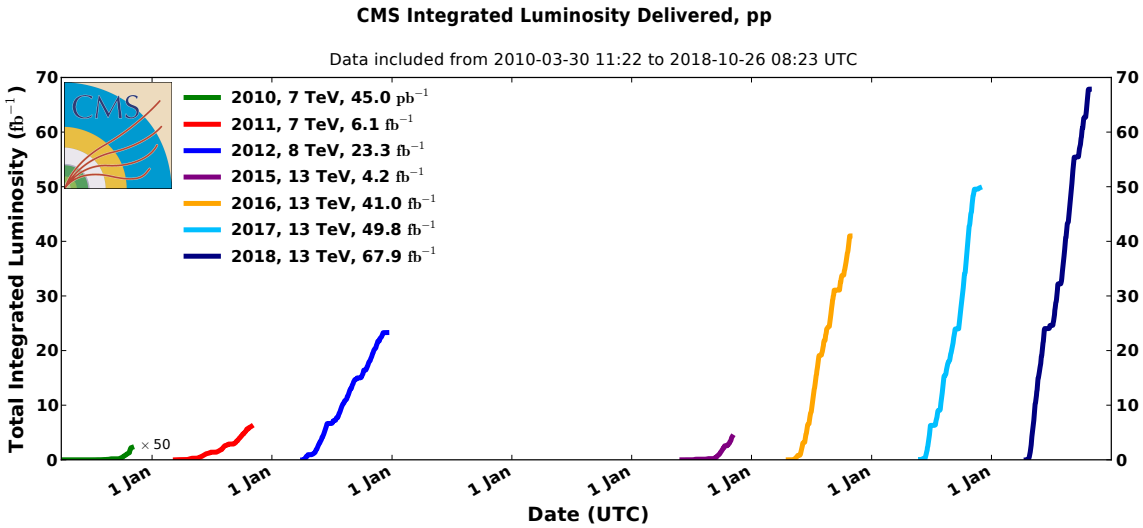
The Compact Muon Solenoid (CMS) experiment is a large general-purpose particle detector at one of the four interaction regions of the LHC.

The CMS concept was first discussed publicly in 1990 at the Large Hadron Collider Workshop in Aachen [58]. Initially, it was thought that the high-radiation environment of the LHC would prohibit the tracking of charged particles other than muons. However, bending the muon trajectories with a powerful solenoid magnet – surrounding the calorimeters that measure the energy of particles other than muons – was already part of the early proposals. This compact design and the focus on muons gave the experiment its name: the Compact Muon Solenoid. Higgs physics has always been the main benchmark that influenced the design decisions taken for CMS, even when it was not clear if the Higgs boson predicted by the minimal scalar sector would exist or if electroweak symmetry is broken by another mechanism.

The prospect of measuring Higgs boson decays into four muons dictated the field strength of the solenoid and the dimensions of the experiment. The detector is 21.6 m long, has a diameter of 14.6 m and weighs 12,500 tons. Together with the 3.8 T magnetic field, this size allows for the measurement of muons with the necessary momentum resolution. A cutaway view of the detector is shown in Figure 2.6, where one can clearly see how the muon spectrometers that surround the superconducting solenoid make up most of the detector volume. Within the solenoid are the calorimeters and the inner charged particle tracker. As is the standard for collider experiments, the calorimeter system is split between an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL). The ECAL is another subdetector that was strongly influenced by Higgs physics. It was designed to achieve the necessary diphoton mass resolution for the measurement of Higgs boson decays to two photons.



(a)



(b)

Figure 2.4: Peak luminosity (a) and cumulative luminosity (b) for proton-proton colisions delivered to the CMS experiment during LHC Run 1 (2010–2012) and Run 2 (2015–2018).

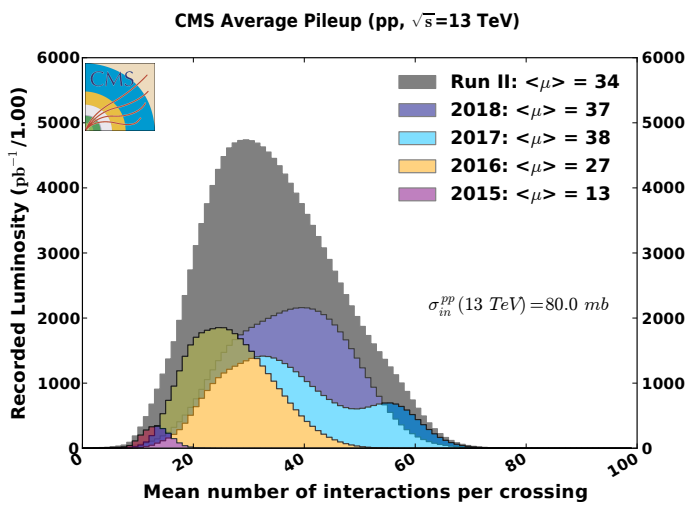


Figure 2.5: Interactions per bunch crossing (pileup) for CMS data from LHC Run 2.

With the very high collision rate of 40 MHz, it is not possible to reconstruct and store all collision events. Dedicated trigger hardware that decides which events should be kept is strictly necessary. A brief explanation of the trigger and data acquisition system will be given in this detector-hardware-oriented chapter, following the detector subsystem explanation. At the end of the chapter, an overview of CMS luminosity measurement is given. Chapter 3 will discuss how the information form the CMS subdetectors are combined to reconstruct the objects in the final states of collision events.

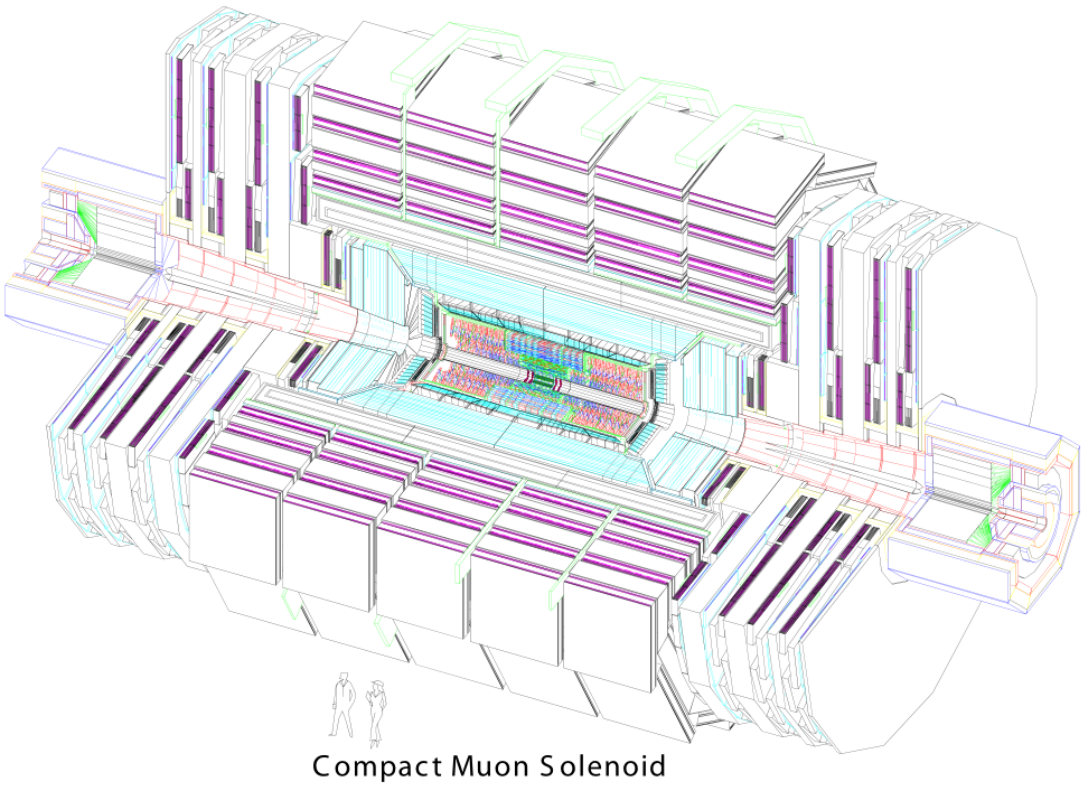


Figure 2.6: A cutaway view of the CMS detector [59].

2.2.1 The CMS coordinate system

The origin of the CMS coordinate system [49] is the nominal interaction point. The x-axis of this right-handed coordinate system points up to the surface. The y-axis points towards the center of the LHC ring and the z-axis consequently points towards the Jura mountains. Positions in the detector are often given in cylindrical coordinates. The polar angle θ is measured from the z-axis, and the azimuthal angle ϕ represents an anticlockwise rotation in the x-y plane relative to the positive x-axis. Another coordinate that is very common in hadron collider physics is the so-called *pseudorapidity* used instead of the polar angle θ:

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

Explaining the name, the pseudorapidity in the limit of a massless particle converges to the rapidity:

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right). \tag{2.5}$$

Most final state particles in hadron collisions come out at small angles relative to the beamline, which translates to an approximately uniform distribution in pseudorapidity. Accordingly, the distance between two particles in the detector, ΔR, is very often measured in the η-ϕ plane. Furthermore, differences in pseudorapidity are Lorentz invariant.

In the following description of the CMS detector subsystems, any given coordinates refer to the global coordinate system that was defined in this section.

2.2.2 Detector subsystems

2.2.2.1 Solenoid magnet

The CMS detector design was originally optimized for muon detection. An essential ingredient to get an excellent muon momentum resolution is a powerful magnetic field that bends the trajectory of charged particles. This field is generated by a solenoid since this design can best resist the magnetic forces exerted on the conductor by a very strong field. The CMS superconducting solenoid is made up of 2168 turns of a high-purity aluminum-stabilized conductor. Its length is 12.9 m at a radius of 3 m, it is cooled down to 4.6 K, and a current of about 18 kA runs through the coil to generate a 3.8 T magnetic field. The outer muon systems (see Section 2.2.2.5) are interleaved with an iron return yoke that contains the magnetic field inside the detector volume. Another advantage of the solenoid design is the homogeneous magnetic field in the barrel, which only bends the particles in the transverse plane. Two interesting design ideas related to the magnet problem were present in the first CMS concepts but were abandoned because of cost and engineering difficulties. One idea was to use extra toroid magnets in the forward regions for precise charge determination up to $|\eta| < 4$, and another idea was to use iron absorbers in the HCAL, which would have amplified the central value of the magnetic field to up to 6 T, further increasing the bending power [60].

A map of the magnetic field generated by the superconducting solenoid at the central value of 3.8 T is shown in Figure 2.7. The magnet will presumably not be exchanged or upgraded during the remaining lifetime of CMS [61]. Some consolidation and upgrade work has been done before 2020, in particular, to minimize magnet strain from unnecessary field-ramping cycles [62].

2.2.2.2 Inner charged-particle tracker

The tracking detector is the detector subsystem closest to the interaction point. It measures the helix trajectories that are described by charged particles in the magnetic field of the solenoid. These measured tracks are used to determine the particles' momenta and charges and also allow for the reconstruction of primary and secondary particle interaction vertices. This vertexing is essential for the CMS physics program, as many of the most interesting collision events at the LHC produce b quark jets from the decay of heavy particles. Tagging these jets as well as other objects, like c quarks and τ leptons, asks for track measurements as close to the interaction vertex as possible[3]. Vertexing also makes it possible to reject tracks from pileup collisions. Among the other design

[3]Both the tagging of heavy flavor (b, c) jets and the tagging of τ leptons requires measurement of secondary vertices very close, albeit for slightly different reasons. For τ leptons (lifetime corresponding to $c\tau = 87.11$ µm), one wants to measure the vertex of the τ-decay directly [64]. For heavy flavor jets, one is interested in the decay vertices of short-lived heavy flavor hadrons that emerge from the b and c quark hadronization.

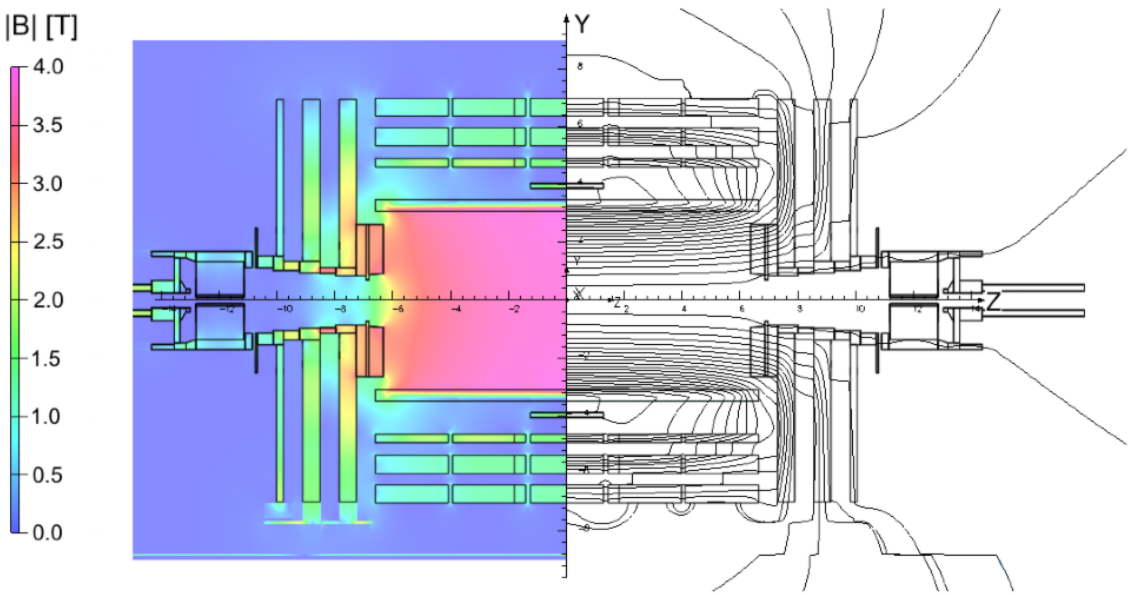


Figure 2.7: Map of the magnetic field of the CMS detector (left) and the corresponding field lines (left) at 3.8 T [63].

goals of the tracker was a high position resolution to match hits to tracks correctly and a high granularity to maintain a low enough occupancy in the high-pileup LHC environment. Furthermore, the tracker material budget should be minimized not to degrade the calorimeter measurements. Putting the calorimeters inside the magnet coil has the advantage of little material in front of them, and this advantage should not be nullified by excessive tracker material. Initially, a tracker system with three subdetectors was considered, with the outermost one based on Micro-Strip Gas Chambers (MSGCs) [65]. It was later decided that the simplest and most elegant way to meet the requirements is an all-silicon system with a pixel detector at the center and a silicon strip detector surrounding it. The design and performance of the tracking detector are detailed in Reference [66]. Figure 2.8 shows the layout of the inner tracking system.

The pixel detector covers $|\eta| < 2.5$ and was replaced at the end of 2016 during the technical stop of the LHC [67]. The previous pixel detector was also used throughout Run 1 and is often referred to as the *Phase 0 pixel detector*, while the new one – meant to serve until the end of Run 3 – is called the *Phase 1 pixel detector*. This change was made to cope with the increasing luminosity of the LHC, maintaining equal or better performance. The original design with three layers in the barrel and two discs in the endcaps has been replaced by a design with four layers and three disks (Figure 2.9). The radial coverage was increased as well. While the old pixel detector had barrel layers at 4.4, 7.3, and 10.2 cm, the new one has layers at 3.0, 6.8, 10.2, and 16.0 cm. As a result, the new pixel detector has improved momentum resolution, heavy flavor tagging performance, and pattern matching capabilities. Another notable change is the reduced material budget in the new Pixel detector, which is shown in Figure 2.10a. The material budget of the old pixel detector is also put into the context of the full inner tracker in Figure 2.10b. The silicon sensor thickness of 285 µm did not change with the new detector.

The pixel detector is surrounded by a silicon strip detector, which consists of several submodules (Figure 2.8). The Tracker Inner Barrel (TIB) has four layers with silicon strips of dimension 10 cm x 80 µm, at a thickness of approximately 320 µm. It is complemented by the Tracker Inner Disks (TID), with three disks on each side and the same

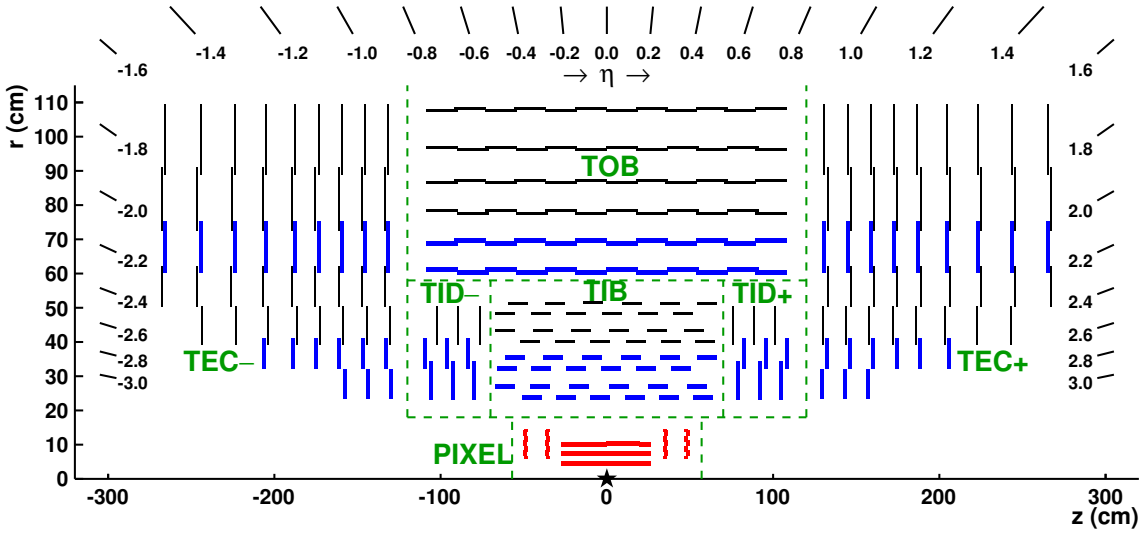


Figure 2.8: Schematic of the original CMS inner tracker [66]. Pixel modules are drawn in red, and strip tracker modules providing 2D hits are drawn in black. Blue lines indicate two back-to-back strip modules, where one module is rotated by a *stereo* angle such that 3D hit positions can be reconstructed. The tracker subsystems are labeled in green.

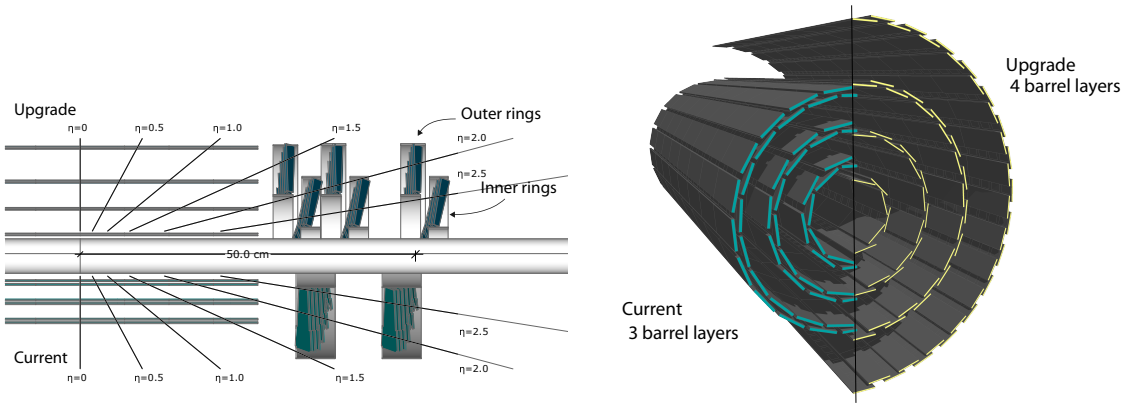
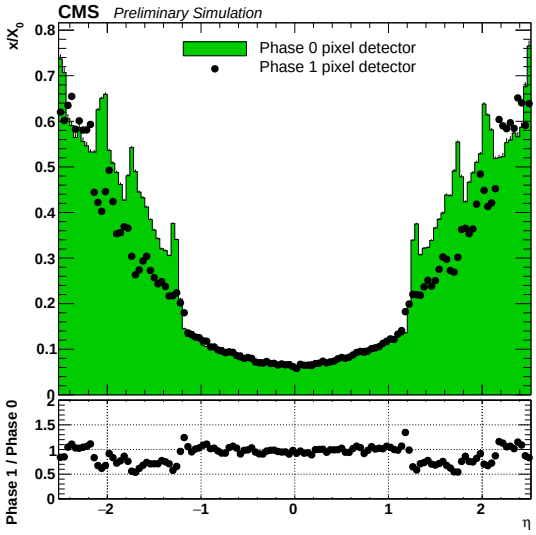
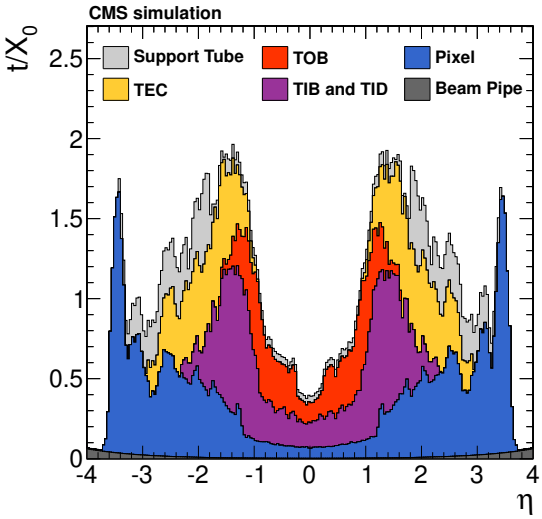


Figure 2.9: Left: the layout of layers and disks in the old and in the new pixel detector. Right: the barrel pixel layers of the old and new pixel detectors [67].

sensor thickness. The thickness is chosen as a trade-off between active volume and susceptibility to radiation damage. The particle flux that causes radiation damage reduces with the distance from the interaction point. Therefore, the strips in the six layers of the Tracker Outer Barrel (TOB) are thicker, with a thickness of approximately 500 µm. Finally, the Tracker Endcaps (TEC) provides nine more disks on each side, equipped with 500 µm sensors as well. The strips in the barrel detectors are oriented parallel to the beam direction, and the strips in the disks of the endcap detectors are oriented radially. Some layers and disks carry two detector modules back-to-back, where one is rotated with a slight *stereo* angle relative to the other. The relation between hits from both planes constrains the last coordinate (z in the barrel and r in the endcaps), albeit with a resolution that is around an order of magnitude worse than for the coordinate perpendicular to the strips.



(a) Material budget of the old (Phase 0, solid green) and the new (Phase 1, black dots) pixel detectors in units of radiation length X_0 [67].



(b) Total material crossed by a partical originating from the nominal interaction point as a function of pseudorapidity η, expressed in units of radiation lengths X_0 [66]. This figure corresponds to the original setup with the Phase 0 pixel detector.

Figure 2.10: The material budget of the Phase 0 and Phase 1 pixel detectors compared and put into confext of the full inner detector.

2.2.2.3 Electromagnetic calorimeter

The homogeneous crystal Electromagnetic Calorimeter (ECAL) has been part of the CMS concept almost from the beginning, even before it was even thought possible to measure charged particles with an inner tracking detector in the LHC environment [58]. It was designed with a firm emphasis on an excellent energy resolution to be maximally sensitive to $H \rightarrow \gamma\gamma$ resonances on top of an overwhelming exponential background.

The ECAL resides in the solenoid volume between the inner tracker and the hadronic calorimeter. It is a homogeneous scintillating crystal calorimeter, relying on the electromagnetic shower formation in a dense material that is, at the same time, an active scintillating medium. The ECAL is made up of a barrel part and two endcaps, as shown in Figure 2.11. The barrel part (EB) covers $|\eta| < 1.479$ and contains 61,200 lead tungstate ($PbWO_4$) crystals that are read out by avalanche photodiodes (APDs). The two endcaps (EE) cover $1.479 < |\eta| < 3.0$ and contain 7324 crystals each, this time read out by vacuum phototriodess (VPTs) which are more radiation hard than APDs. The endcap crystal calorimeter is complemented by a two-plane silicon strip preshower detector (ES) with two radiation lengths of lead in front of the first plane and one radiation length in front of the second. The silicon sensors measure 6.2 cm x 6.2 cm with an active area of 6.1 cm x 6.1 cm, divided into 19 strips.

The ECAL is highly granular in η and ϕ to separate well between electrons and jets. Also, if the transverse area for energy summing can be kept as small as necessary for accurate energy measurement, the impact of pileup contributions can be minimized. The front face cross-section of the EB crystals is around 2.2 cm x 2.2 cm, which approximately matches the $PbWO_4$ crystals Molière radius of 2.19 cm. The EE crystals have a slightly larger front face of 2.86 cm x 2.86 cm. The choice of larger EE crystals seems to contradict

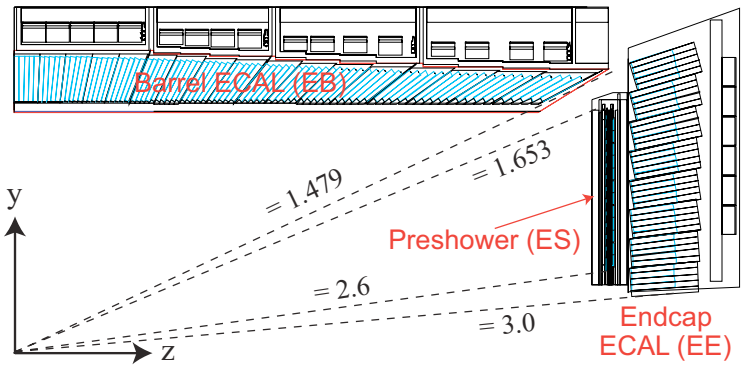


Figure 2.11: Transverse section through the CMS electromagnetic calorimeter [59].

with pileup rejection being the primary motivator for the high granularity, as particles from pileup are more abundant in the endcaps. Indeed, this increased EE crystal front face was not forseen in the original ECAL design [68], but was only decided for the CMS Technical Design Report in 2006 [59]. The purpose was to reduce cost and compensate for the weaker light collection capabilities of the VPTs. A width of 2.86 cm was just about possible for the crystal producers. The barrel's crystals are 22 cm long and the ones in the endcaps 23 cm, with one radiation length X_0 being approximately 0.89 cm.

In beam tests, the ECAL barrel was shown to have an energy resolution that fits the typical functional dependence [49]:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \tag{2.6}$$

$$S = 2.8\ \%\ \mathrm{GeV}^{1/2}, \quad N = 120\ \mathrm{MeV}, \quad C = 0.3\ \%.$$

The stochastic term S represents fluctuations in the amount of produced and collected scintillating light. The noise term N corresponds to mostly electronic noise and pileup contributions. The constant term C accounts for inter-calibration errors, nonuniformity in the longitudinal light collection, and leakage of energy from the back of the crystal. Measuring the energy resolution of the ECAL in the context of LHC Run 2 collisions is more complicated than in beam tests, but it can be done using known dielectron resonances. The ECAL resolution will be discussed more in Section 3.2.5 which covers the electron reconstruction.

The ECAL performed very reliably during the LHC Run 2 [69]. The high radiation environment and consequent transparency loss of the crystals necessitate constant monitoring and updating of various calibration parameters. A laser system monitors the transparency for light with a wavelength of 440 nm, which corresponds to the scintillating light's peak. The transparency loss, measured by the laser system, is shown in Figure 2.12.

The most notable ECAL related change between Run 1 and Run 2 was done at the energy reconstruction level. The signal amplitude is no longer estimated using a digital filter, but by a template fitting method dubbed *multifit* [71]. It accounts for pileup contributions from other bunch crossings, the so-called out-of-time (OOT) pileup. The scintillator decay time of about 40 ns is longer than the time between bunch crossings (25 ns). Hence, to measure the bulk of the pulse shape, the ECAL reads out ten samples of 25 ns, where the triggering bunch crossing corresponds to the 6th sample. The multifit fits up to 15 pulse shapes simultaneously to disentangle the signal pulse from the pulses caused

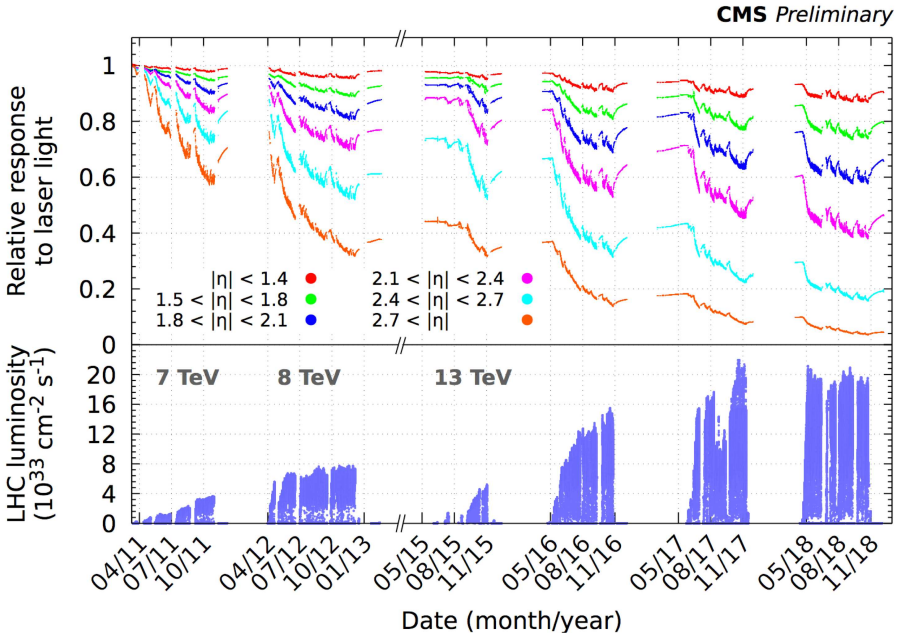


Figure 2.12: Relative response to laser light (440 nm in 2011 and 447 nm from 2012 onward) injected in the ECAL crystals, measured by the ECAL laser monitoring system, averaged over all crystals in bins of pseudorapidity [70].

by OOT pileup, solving a non-negative least squares (NNLS) problem in which all amplitudes are constrained to be positive. This algorithm is also used at the High Level Trigger (HLT) where fast execution time is mandatory, which is why an adaptation for Graphics Processing Units (GPUs) is envisioned for Run 3. Because the NNLS algorithm is iterative, the GPU accelerated implementation of the multifit procedure might also come with some changes to the algorithm. Often, the solution already converges in the first few steps while only a few channels iterate for much longer, which is not friendly to parallelization. It should be noted that not every channel is treated with the full multifit reconstruction in each event: the ECAL uses a zero-suppression readout and typically only reads out around 8×10^3 crystals per typical Run 2 event.

2.2.2.4 Hadronic calorimeter

Directly behind the ECAL and for the most part still within the magnet cryostat resides the Hadronic Calorimeter (HCAL). Its job is to measure charged and neutral hadrons of any kind to ensure a good resolution on jet energy and missing transverse energy measurements. However, it should be kept in mind that in the CMS particle-flow reconstruction, the charged hadrons energy is mainly estimated from track measurements. Accordingly, the HCAL is mostly used for hadron identification and to measure the energy of neutral hadrons. Most hadrons are able to pass through the ECAL and interact in the brass absorber layers of the HCAL, creating a hadronic shower. The brass layers are interleaved with plastic scintillator tiles to sample a fraction of the shower energy deposit by converting it to visible light. The resolution attainable on hadronic shower measurements is strongly limited by the non-Poissonian shower fluctuations caused by the many possible hadronic interactions, shower leakage behind the HCAL, and the electromagnetic shower component from potential $\pi_0 \to \gamma\gamma$ decays to which the calorimeter responds differently. It is also possible – in particular for heavy-flavor jets – that undetectable neutrinos are produced in the hadronic shower. Therefore, the

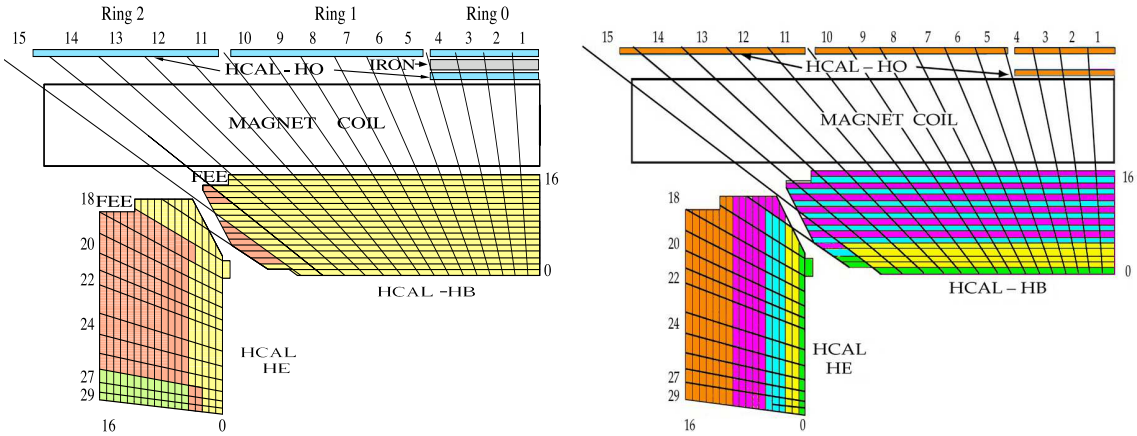


Figure 2.13: Schematic views of the HCAL showing the HB, HE and HO calorimeters [72], where (a) shows the original HPD readout segmentation and (b) shows the segmentation that is envisioned for Run 3 with the SiPM readout. The "FEE" labels indicate the locations of the front-end electronics. The HF calorimeter is not shown.

sampling calorimeter design is very typical for hadron calorimeters because of its cost-effectiveness. The penalty from the sampling measurement on the resolution does not weigh in too much compared to the effects described above.

The HCAL barrel (HB) contains 15 brass layers, covers the pseudorapidity range $|\eta| < 1.4$ and is segmented in 2304 *towers* of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. A few more scintillator layers outside the magnet make up the outer barrel (HO) system that is sometimes also referred to as a *tail catcher*. The HO extends up to $|\eta| < 1.26$. The hadronic endcaps (HE) cover $1.3 < |\eta| < 3.0$ and overlap slightly with the HB. There are 1152 modules in each endcap, spread between up to 18 brass absorber layers. Below $|\eta| < 1.74$, the tower segmentation matches the one in the barrel, while at high pseudorapidity, the azimuthal segmentation is reduced by a factor of two. Each of the HCAL towers in the barrel covers 5×5 ECAL crystals,

The plastic scintillator tiles of the HCAL are connected to the front-end electronics with wavelength-shifting fibers. The signal is collected with Hybrid Photodiodes (HPDs). Light collection efficiency and size constraints of the HPDs require the bundling of multiple fibers for a single HPD. However, the HPDs are gradually replaced with Silicon Photomultipliers (SiPMs) that have higher efficiency and require fewer fibers bundled together, allowing for a more granular readout. During the LS1, the replacement has already been done for the HO. By the start of Run 3, all HPDs in the full HCAL will be replaced, improving the longitudinal segmentation of the readout, as shown in Figure 2.13.

There are two additional calorimeters inside the detector that do not belong to the ECAL or HCAL subsystems. The pseudorapidity range $-6.6 < \eta < -5.2$ is covered by the Centauro and Strange Object Research Calorimeter (CASTOR), a electromagnetic/hadronic Cherenkov sampling calorimeter. The CASTOR not only helps to validate the modeling of the underlying event[4] in hadron collisions, but also has its own physics program of exotic searches and QCD phenomenology [73]. Located at $z = \pm 140$ m are the two stations of the Zero-Degree Calorimeter (ZDC), which is especially used for heavy-ion studies [74].

[4]The underlying event is everything else that happens in the collision besides the primary high energy scattering process.

2.2.2.5 Muon detectors

The name of CMS, the Compact Muon Solenoid, suggests that the muon detection system is central to the detector concept [75]. The muon system is the outermost subdetector system, interleaved with the iron return yoke for the magnetic field. On the one hand, the inner tracking detector described in Section 2.2.2.2 is optimized for track separation and vertexing near the interaction point, which demands the high resolution and readout channel number that can be achieved with silicon detectors. On the other hand, the muon system has to measure muon track sections outside the solenoid volume that are long enough to measure the track curvature for muon momentum determination as precisely as necessary. These measurements do not require such a high resolution for individual measurement points, but rather a large detector volume. Gaseous detectors are more appropriate than silicon detectors in this case.

Like most subdetectors covered in the previous sections, the muon system is segmented into a barrel ($|\eta| < 1.2$) and endcap ($0.9 < |\eta| < 2.4$) region. A schematic to accompany this overview of the muon system can be found in Figure 2.14. The example of the ECAL has shown already how the different radiation levels in the barrel and endcaps necessitate different readout electronics. Unlike the calorimeters, located within the solenoid volume, the muon system is exposed to different magnetic field strengths in the barrel and endcaps. The differences in the radiation level and the magnetic field strength (Figure 2.7) illustrate why the CMS muon system uses different detector types in barrel and endcap. The barrel is equipped with Drift Tubes (DTs), and the endcaps are made of Cathode Strip Chambers (CSCs) that use different readout electronics. Both regions are complemented by Resistive Plate Chambers (RPCs) operating in avalanche mode, which have an excellent timing resolution of the order of 1 ns. In total, the muon system amounts to about 25,000 m^2 of detector surface.

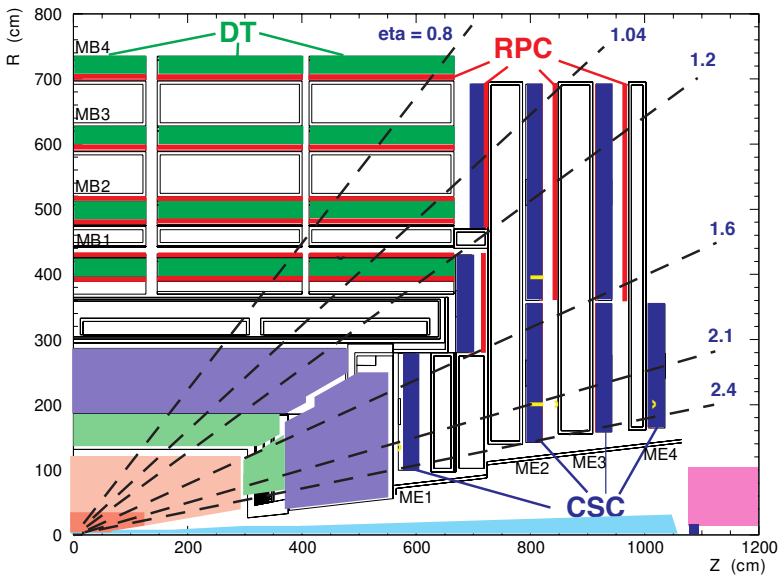


Figure 2.14: Schematic of a quadrant of the CMS detector, highlighting the muon system. Text labels indicate the DTs, Resistive Plate Chambers (RPCs), and CSCs [59].

Like most gaseous detectors, the DT detector in the barrel records muon hits by measuring the electronic avalanche that forms at the stretched anode wire when a muon ionizes the gas in the detector. As can be inferred from Figure 2.14, which shows a quarter of the CMS detector, the DT is segmented into five wheels with four *stations* each. The

three inner stations each have 12 layers of wires, where 8 of them contain wires parallel to the beam pipe, measuring the hit position in the r-ϕ plane. The other four layers have perpendicular wires to measure the z-coordinate. In the outermost station, there are only eight layers with wires parallel to the beamline. With this setup, the DT achieves a hit position resolution of about 100 µm in the r-ϕ plane.

The CSC chambers all have concentric wires that extend in the r direction, and they measure the remaining coordinate more indirectly than by wire position. The cathode is segmented into strips perpendicular to the anode wires, and the mirror charge at these cathode strips is measured. In this way, the CSC detector achieves a resolution of about 75 to 150 µm in the ϕ direction and around 200 µm in the radial direction. This resolution is similar to the one of the DT detector, hinting that the advantage of the cathode strip measurements is not necessarily the position resolution but rather a better timing resolution that makes the CSCs more suitable for triggering in the endcap region where the muon density is higher than in the barrel.

When the information from the muon system is combined with the measurements of the inner tracker, a momentum resolution for muons with $20 < p_{\mathrm{T}} < 100$ GeV of about 1-2 % is reached in the barrel [63].

2.2.3 Trigger and data acquisition system

Handling the vast amount of data produced in hadron-collision experiments is a challenge. The proton bunches interact with each other in the detector every 25 ns, corresponding to a frequency of 40 MHz. The total data from the millions of the detector readout channels amounts to approximately 1 MByte per event, a number that is growing with the increasing pileup and, therefore, increasing detector occupancy. The trigger [76] and data acquisition system [77] must be able to cope with this data flux.

The bottleneck of the data acquisition is the speed at which the collision data can be written to permanent storage. This limit amounts to around 800 events per second. Thus, the collision events have to be filtered on the fly, reducing the event rate by multiple orders of magnitude. CMS uses a two-stage filtering system, where a hardware-based Level-1 (L1) trigger first reduces the 40 MHz rate to about 100 kHz. The second stage, the software-based High Level Trigger (HLT), further reduces the rate to the storable 800 Hz.

The L1 trigger is implemented in field-programmable gate arrays (FPGAs) and application-specific integrated circuits (ASICs). For more sophisticated algorithms, programmable memory look-up tables (LUTs) are employed. It reads out the muon system and calorimeters with reduced granularity and then decides in 4 µs if an event is worth passing to the HLT. Figure 2.15 gives an overview of the L1 trigger architecture, showing how calorimeter and muon information is initially handled separately. The information is then combined in the *global* trigger. To pass – or *fire* – the L1 trigger, the event must fullfill the requirements implemented in one of the so-called trigger *seeds* in the trigger *menu*. The L1 trigger system underwent a significant upgrade between 2015 and 2016, where some electronics were replaced, and more powerful FPGAs installed. The new architecture allows for a sampling of the energy with the full granularity of the calorimeter trigger readout, significantly improving the energy and position resolution for reconstructed objects in the calorimeter. For the HL-LHC, it is foreseen to also do track reconstruction with the inner tracker at L1 [78].

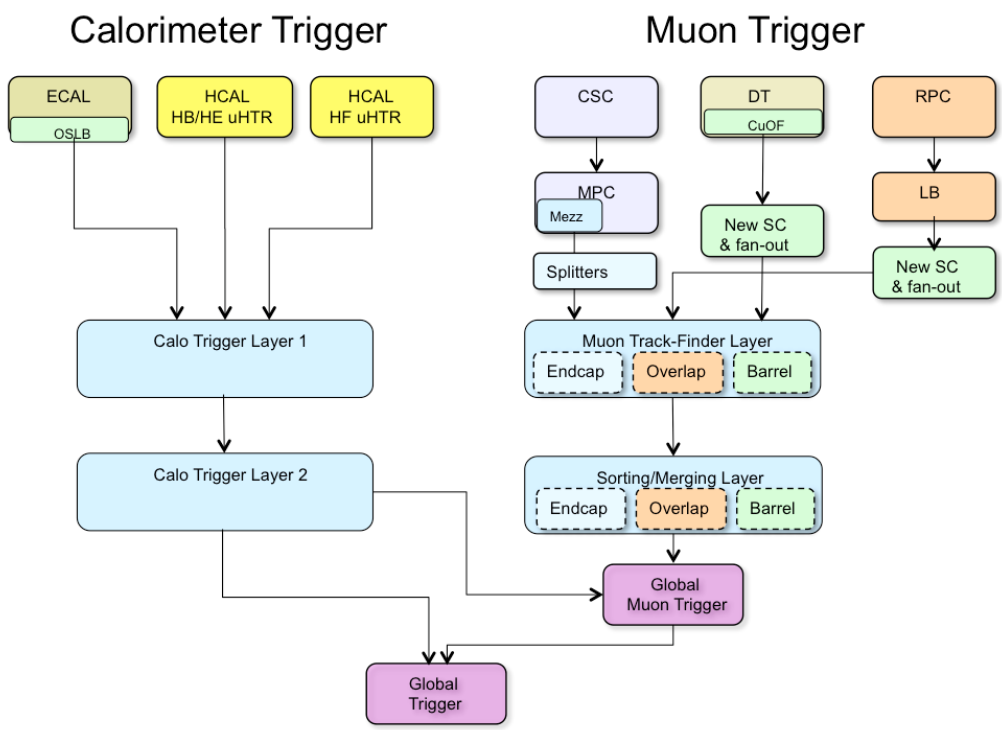


Figure 2.15: Structure of the L1 trigger system [79].

The HLT is implemented in software running on commercial computers, just like the full event reconstruction. The HLT reconstruction is often called *online* reconstruction, while the *offline* reconstruction processes the events once the raw data has been stored. Unlike the L1 trigger that uses a separate readout with reduced granularity, the full readout is available for the HLT. This means that much of the reconstruction code can be shared between online and offline reconstruction, most notably in the particle-flow reconstruction (explained in Section 3.1). While the offline reconstruction of one event takes about one second, the online reconstruction runs for approximately 150 ms per event. Similar to the L1 seeds, there are several HLT *paths* that an event can match. Examples of trigger paths are the double-electron, double-muon, or lepton cross-trigger path, which require two isolated prompt leptons passing certain energy thresholds. These lepton trigger paths were used in the triboson analysis presented in this thesis (more detail given in Section 6.2.1). The trigger paths are defined by a sequence of software modules[5], which either compute event observables or use event information for filtering decisions. For the most efficient computing resource usage, the modules with the most stringent filters are preferred to be at the beginning of the sequence, while computationally expensive modules are placed towards the end.

The timing demands of the HLT often drive innovation in the reconstruction software. A recent example is the seeding of electron candidates (see Section 3.2.1 for more details on the electron seeding step). The new HLT algorithm for Run 2 considers more hits in the pixel hit pattern when matching tracks with electromagnetic showers, exploiting the additional layer of the pixel detector. However, more significant changes to the HLT software will be introduced before Run 3. Some of the most performance-critical tasks such as track pattern matching and ECAL pulse shape fitting are envisioned to be accelerated with GPUs, which would be a first for CMS reconstruction software in general.

[5]see Section 3.7 for more information on the CMS reconstruction software framework

Trigger seeds or paths can be *prescaled*, meaning that only an adjustable fraction of the events passing a specific trigger selection is kept. Prescaling is often done for validation purposes or trigger paths that are very unrestrictive for specific physics goals. Some events are also kept independently of any trigger decisions, resulting in a dataset that is unbiased by the trigger. Such events are referred to as *minimum bias* events and are used, among other things, for validation and QCD measurements.

If an event passes the final filter of one of the HLT paths, it is piped by the DAQ to one of the data streams that fill the *primary datasets* that are ultimately stored on the CERN Tier 0, which is the computing center at the top of the hierarchy of the Worldwide LHC Computing Grid. Examples of primary datasets are double-muon or double-electron samples.

A particular primary dataset selects a small fraction of the events for Data Quality Monitoring (DQM). In the first *online* validation, the general state of the detector is monitored, including channel quality, synchronization, and data integrity. The online validation is done from the CMS control center on the experimental site. The *online* validation follows the delayed full event reconstruciton. Human validators scrutinize high-level quantities such as energy spectra, particle multiplicities, and invariant masses. The output of the DQM effort is a list of certified events that are ready to be analyzed.

2.2.4 Luminosity measurement

Measuring the instantaneous and integrated luminosity recorded with CMS is essential for operating the experiment and analyzing its data. The online luminosity measurement provides feedback to the LHC for tuning beam parameters and enables the measurement of trigger rates on the detector side. This feedback allows for the DAQ and trigger systems to adapt to the beam intensity, for example, by setting the prescaling factors of individual trigger paths. A total of seven luminometers are used in CMS [80]. Three of them are primarily used for online monitoring and are read out by a dedicated data acquisition system independent of the primary CMS readout: the Pixel Luminosity Telescope (installed at the beginning of 2015), the Fast Beam Conditions Monitor, and the already mentioned hadronic forward calorimeter (HF), but this time connected to a separate readout. For the offline reconstruction, three more estimates are provided via vertex counting and measuring the activity in the pixel detector and the DT muon system. The method based on the pixel occupancy is the most accurate but cannot be used online because of CMS trigger bandwidth limitations. Finally, the CERN radiation monitoring system can also be used for a complementary luminosity measurement.

The calibration, linearity, and stability of the luminosity observables have to be studied regularly to provide a systematic uncertainty estimate on the integrated luminosity for each data-taking year. In Run 2, from which data were analyzed in this thesis, this systematic uncertainty ranges between 2.3 % (2017) and 2.5 % (2016 and 2018).

Chapter 3

Reconstruction and simulation of collision events with the CMS detector

For the analysis of proton collision events, it is essential to reconstruct the final-state particles with high precision and produce simulations for which the reconstruction performance accurately represents the performance achieved in real collision data. The reconstruction workflows for particle detector experiments are traditionally very subdetector oriented, meaning that the different particles – or more generally *physics objects* – are reconstructed in different subdetectors. Nevertheless, the best possible performance can only be achieved when the information from all subdetectors is combined for individual object reconstruction. The most famous example for the success of this *particle-flow* approach is the combination of tracker and calorimeter information for jet measurements, which dramatically improves the energy resolution at low momentum.

It is usually not possible to unambiguously tell to what kind of particle a given detector signature corresponds. However, an unambiguous reconstruction greatly simplifies the bookkeeping of the energy distribution in the final state, allowing for consistent definitions of jets and missing transverse energy. The CMS event reconstruction aims for such a *Global Event Description (GED)*. Section 3.1 explains this global event reconstruction in the particle-flow paradigm.

The reconstruction of physics objects used in the triboson analysis is explained in dedicated sections. Section 3.2 focuses on electrons, Section 3.3 discusses the muon reconstruction, and Section 3.4 covers the reconstruction of jets and missing transverse energy. The global event reconstruction must make compromises in particle identification. For example, it is easy to confuse a prompt lepton with a leptonically-decaying tau or a nonprompt lepton within a jet, so the prompt lepton has to be tightened for the GED. Hence, the vast majority of CMS analyses consider a priori all reconstructed leptons, which is a superset of the reconstructed leptons that are also accepted as *particle-flow candidates* by the GED reconstruction. This also applies to the triboson analysis.

When selecting prompt leptons for analysis, one usually considers the lepton's isolation in the event to suppress nonprompt background. For hadron collisions with a high pileup rate, pileup contributions must be subtracted for the computation of isolation variables. In CMS, two different correction schemes are used for leptons and muons. It is difficult to conclude which correction algorithm is generally more performant, although it would be desirable to have consistent isolation definitions for both types of leptons. To contribute to this dicussion, this thesis includes a comparison of pileup correction methods from a statistical viewpoint in Section 3.5.

Finally, this chapter explains how the simulation and reconstruction of collision events is technically realized in Section 3.6 and Section 3.7 respectively.

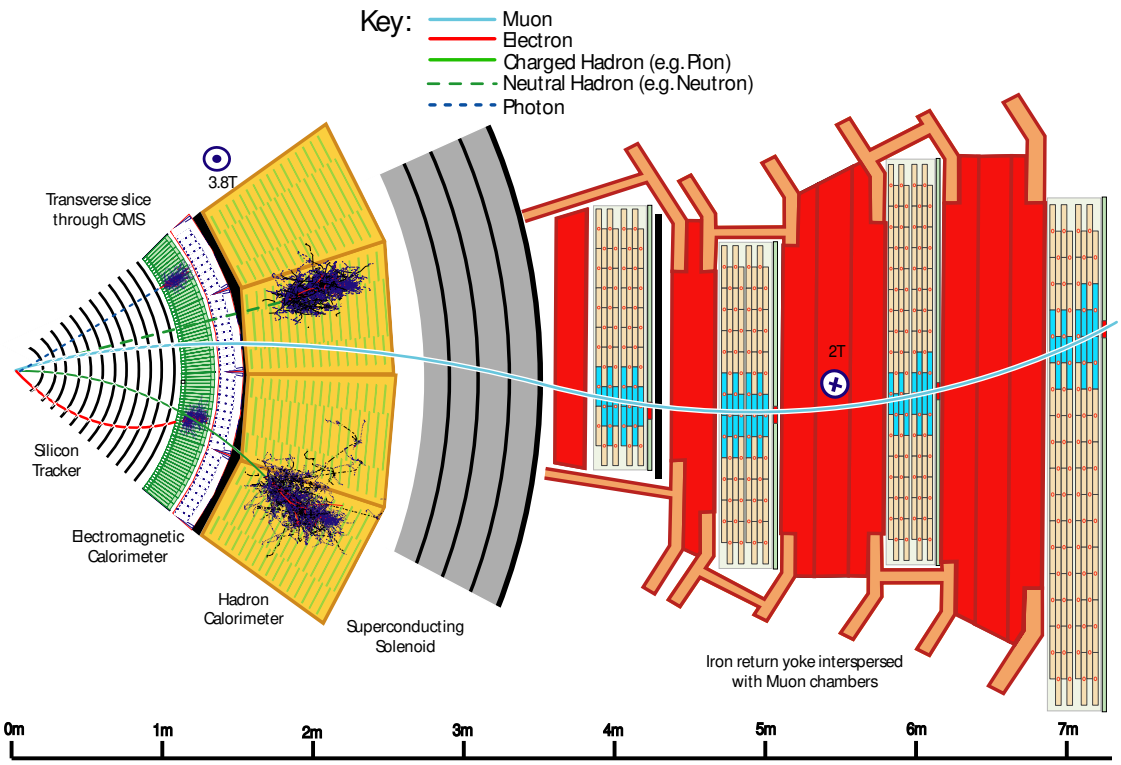


Figure 3.1: Transverse view of the CMS detecor, showing the signatures of the five physics objects reconstructed by the particle-flow algorithm [81]: muons, electrons, charged hadrons, neutral hadrons, and photons.

3.1 Global event reconstruction with the particle-flow algorithm

The particle-flow (PF) reconstruction is responsible for identifying the five basic physics objects in the event: muons, electrons, charged hadrons, neutral hadrons, and photons. Their signatures in the detector subsystems are indicated in Figure 3.1. These five types of *PF candidates* are then used for the reconstruction of the three higher level physics objects: jets, missing transverse energy, and tau leptons. The PF candidates are also used to compute isolation variables for leptons and jets.

To make use of the information from all subdetectors, the algorithm starts with the so-called PF *elements*, which are some basic reconstructed signatures that will be combined to PF candidates. The PF element types include tracks of different kinds, such as muon tracks, displaced tracks, tracks from photon conversions, and tracks reconstructed with the Gaussian Sum Filter (GSF) algorithm for electrons (see section Section 3.2). Furthermore, there are all the single-particle-like clusters from the different calorimeter systems and two more elements dedicated to electron and photon reconstruction: the so-called *superclusters* that are wider ECAL clusters corresponding to the signature of an electron or photon that starts to interact in the tracker material, and identified occurrences of bremsstrahlung interactions.

Having access to all PF elements, the PF *block* algorithm then loosely decides which elements might go together to make up a PF candidate, grouping the elements in *PF blocks*. Different *link* algorithms are plugged into the block algorithm. These linkers focus on one pair of PF element types at a time, defining the criteria that decide whether the two elements are merged into a block. The purpose of the block algorithm is to reduce

the number of element combinations that the actual particle-flow algorithm needs to consider.

Most details about the particle-flow algorithm and the reconstruction of PF elements [81] are not essential to understand the physics objects analyzed in this thesis. However, it should be noted that electrons and muons play a special role in the algorithm. These objects are reconstructed by separate algorithms, but the objects these algorithms produce are compatible with the PF elements. For example, an electron is a supercluster matched to a GSF track. These complete objects are also handed to the particle-flow algorithm, which decides where it preserves them as prompt electron or muon PF candidates or splits them up to PF elements that go into other PF candidates. Hence, the full list of reconstructed electrons and muons are a superset of the electron and muon PF candidates.

3.2 Electron and photon reconstruction

This section gives an overview of electron and photon reconstruction and identification in CMS and the performance with data taken in 2016 and 2017, mirroring a paper written by the thesis' author. [82] The performance in 2018 is similar to the 2017 performance, as the detector conditions and pileup level did not change significantly. The figures for 2017 performance are the first public electron performance results for this year. The 2016 data has been reprocessed with improved ECAL calibrations [69], hence the potential of Run 2 data after scrutinizing it is highlighted. Most plots in this section are taken from a detector performance summary note published in May 2018 [83].

The treatment of electrons and photons with the CMS detector relies primarily on the ECAL. Since the footprint of an electromagnetic object can also comprise traces in the inner tracker, information from the silicon tracker must also be considered. A pixel detector replacement became necessary between the 2016 and 2017 data taking periods [67]. The original one could not support the expectations-exceeding number of pileup interactions, both from a bandwidth and radiation perspective. The new pixel detector features an additional fourth layer of active modules in the barrel and an additional third disk per endcap, resulting in a four-hit coverage in the whole tracking region. The radius of the innermost layer has been reduced from 44 mm to 29 mm for improved vertex resolution. The material budget in the endcaps has been reduced by up to 50 %. This lowers the number of converting photons and bremsstrahlung-emitting electrons, which are challenging for reconstruction algorithms. The tracking system extends up to $|\eta| = 2.5$, defining the fiducial region for electrons and photons.

After the successful first operational run between 2009 and 2013 (Run 1), the Large Hadron Collider is now back in data-taking and closing up the second operational period from 2015 to 2018 (Run 2) at $\sqrt{s} = 13$ TeV and an ever-increasing pileup. The last published papers on electron and photon performance with the CMS detector are dated back to 2015, summarizing the performance at $\sqrt{s} = 8$ TeV during Run 1 [84, 85]. Since then, the reconstruction algorithm was not changed as far as the underlying concepts are concerned. Some improvements were made to better keep track of the energy flow in the event, i.e.,, to avoid double counting of energy. This is essential for an accurate determination of the missing transverse energy in the event, especially at elevated pileup levels. Equally important is the steady refinement of the identification algorithms to reduce the number of jets misidentified as prompt electrons or photons, as well as alignments and calibration efforts to keep the residual differences between data and Monte Carlo (MC) simulated events at a minimum.

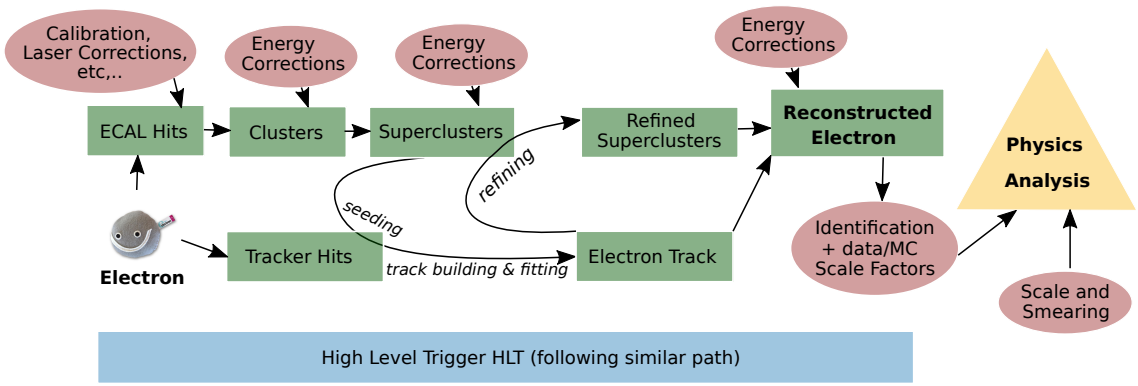
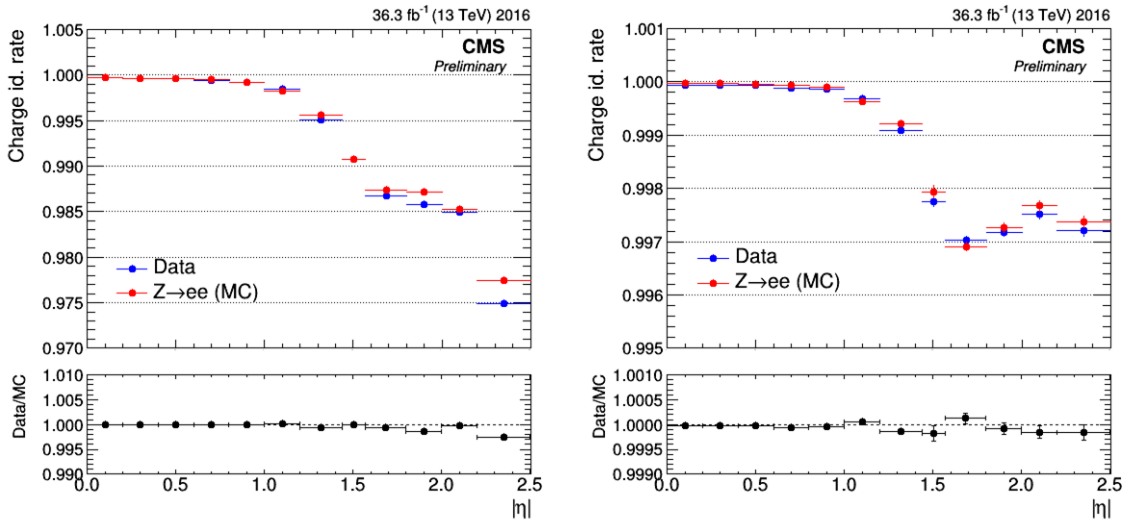


Figure 3.2: Simplified schematic of the electron reconstruction workflow [82].

3.2.1 Reconstruction steps

Figure 3.2 gives a conceptual overview of the electron and photon reconstruction algorithm, taking electrons as an example. On the ECAL side, the energy deposits in the crystals get merged to clusters that topologically match the expectation from an individual particle impacting the calorimeter (single-particle-like clusters). The energy deposit in one crystal may be split among more than one single-particle-like cluster. These clusters get further combined into so-called *superclusters*, matching the pattern of an array of clusters expected from an electromagnetic object that might have radiated or converted before reaching the ECAL. In particular, that means the individual clusters are spread in the azimuthal direction because of the magnetic field. These superclusters seed the track building and fitting algorithm. Electrons suffer large radiative energy losses in material, which are not Gaussian distributed but follow the Bethe-Heitler model [86]. Hence, electron tracks are reconstructed using a Gaussian Sum Filter (GSF) instead of the widespread Kalman Filter (KF) [87]. The tracks are, in turn, used to refine the superclusters. Clusters matching extrapolated track tangents at tracker layers are merged, while conversions are recognized and fitted. The refined supercluster and electron track make up what is called a reconstructed electron. The GSF-based track reconstruction can also be seeded from a general track that is matched to a single-particle-like cluster, complementing the reconstruction at low energies. This tracker-driven seeding is only available for offline reconstruction and not at the trigger level. As indicated in Figure 3.2, energy corrections via multivariate regression are inserted after most clustering steps, instead of just inserting one energy regression at the end of the reconstruction. This is because the intermediate reconstructed objects are used in other algorithms as well, like the CMS particle-flow algorithm [81].

The reconstruction efficiency for electrons at the Z peak is about 96%, with a slight increase in the endcaps in 2017 because of the new pixel detector with a reduced material budget. The improved ECAL calibrations in the recent 2016 data reconstruction were beneficial for the efficiency as well, on the full range of pseudorapidity. MC indicates that the new pixel detector yields not only a higher electron reconstruction efficiency but also an about 30% lower fake rate at the Z peak than in the previous year. This is attributed to the more robust quadruplet-based track seeding algorithm, made possible by the four-hit coverage, which reduces the number of fake tracks in general. So far, the algorithm targeted a maximum efficiency at a fake rate that is still computationally tolerable. It is only in the so-called identification (ID) step where the background gets effectively rejected while keeping a signal efficiency that is tolerable for the analysis at hand. This step will be explained later in more detail after discussing the charge identification for electrons.



(a) Charge identification by majority vote

(b) Tight charge ID selection

Figure 3.3: Rate of correct charge identification for Z electrons as a function of the pseudorapidity for all electrons where the charge is whichever two estimates agree on (a) and for electrons where all three estimates agree (b) [83]. In both cases, electrons pass the loose cut-based identification.

The reconstruction path of the High-Level Trigger follows the lines of the offline reconstruction as close as possible to not complicate analysis work with differently-behaving online and offline reconstruction efficiencies. The picture for photons is similar, except that tracks can come from converted photons only.

3.2.2 Electron charge identification

The rates of correct electron charge identification were measured in 2016 data. Three different charge estimates are inferred from the GSF track curvature, from the curvature of the closest KF track, and from $\Delta\phi$ between the cluster and the GSF track extrapolated to the vertex. The default charge estimation for electron candidates is taken as the majority vote of these three estimates. For Z electrons that pass a loose cut-based selection, this gives misidentification rates at the 10^{-3} level in the barrel and around 2% in the endcaps (Figure 3.3a), increasing with electron energy because of the tracks being less curved. A very high rate of correct charge assignment can be obtained by requiring all three charge measurements to agree (Figure 3.3b). Around the Z peak, this selection comes with an efficiency loss of 3 to 10 %, depending on the pseudorapidity. The charge identification rates show very good data/MC agreement. It should be pointed out that the charge misidentification rates are correlated with the ID selection. Tighter IDs, specifically those requiring good track measurements and conversion rejection, will have lower charge ID rates and better agreement of the three estimates.

3.2.3 Electron and photon identification

Everything described so far was common to almost all CMS analyses. However, each analysis has its own efficiency and fake rate requirements. This is why two different ID algorithms are implemented for both electrons and photons. The first is a sequential cut-based selection with several working points for general use. The second algorithm

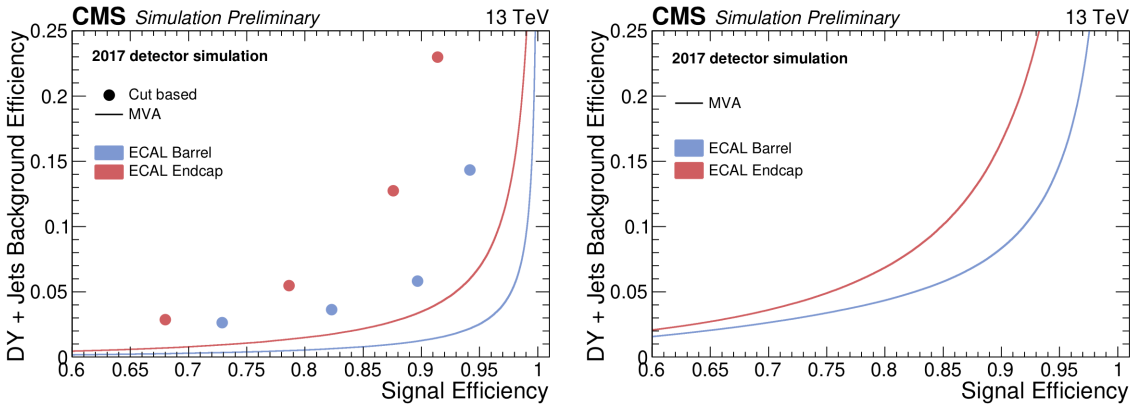
relies on Boosted Decision Trees (BDTs) [88], particularly meant for maximum separation down to low transverse momentum (p_T). The BDTs are trained in several kinematic bins. An overview of the performance of the identification step is given in Figure 3.4. The ID variables can be grouped in shower-shape, track, track-cluster matching, conversion identification, and isolation variables.

While the shower-shape variables rely on the energy deposits in the ECAL and HCAL, the track variables rely on the GSF track and the nearest KF track. An interesting example for a track variables would be $f_{\text{brem}} = 1 - p_{\text{in}}/p_{\text{out}}$, where p_{in} and p_{out} denote the momentum estimate at the beginning and at the end of the reconstructed track. Hence, the f_{brem} variable measures the momentum fraction lost by bremsstrahlung in the tracker. Apart from being a powerful ID variable, it is an excellent tool to access the material budget in data and compare it to MC. Data/MC discrepancies in this variable during Run 1 hinted at a mismodeling of the material budget, which has been corrected for in Run 2. The precise extrapolation of the electron trajectory to the calorimeter with the GSF tracking algorithm and the little material between the tracker and the ECAL result in powerful track-cluster matching variables, especially cuts on $\Delta\eta$ can be tight as no magnetic field spreads out the energy clusters in this direction. The employed conversion ID variables are the goodness of a potential conversion vertex fits and the number of missing hits at the beginning of the track, separating prompt electrons from nonprompt conversion electrons.

On the isolation side, the sequential cut IDs use the energy flow in a $\Delta R = 0.3$ cone around the object, corrected by an area–median pileup subtraction scheme. The BDT algorithm for electrons, which uses isolation variables since 2017, instead takes the underlying neutral hadron, charged hadron, and electromagnetic energy flow components separately, plus a pileup estimate. This results in better performance compared to taking only the pileup-corrected isolation sum as isolation input. Complex classifiers like BDTs benefit from lower-level input variables, which they combine in a way that is adequate to meet their target. Including isolation in the BDTs significantly improves the ID efficiency at highly efficient working points ($\gtrsim 85\%$) compared to the traditional approach where an isolation cut is applied on top of the BDT identification. Next to the new pixel detector, which reduces the fake rate in the endcaps at the reconstruction level, the new isolation-inclusive BDT selection is ensuring that the fake rate for electrons is at the same level in 2017 as it was in the year before, despite the increased pileup. It should be noted that isolation variables depend on the physics of the event. For example, they might be less appropriate for studying electrons in boosted systems. Therefore, a flavor of the BDT algorithm without isolation variables is available for these analyses as well.

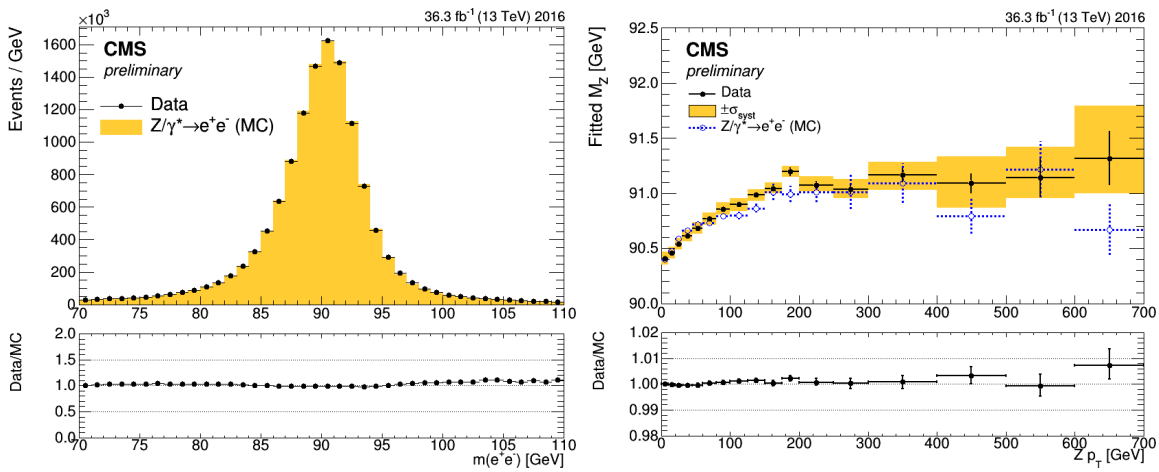
3.2.4 Selection efficiencies in data versus simulation

Selection efficiencies are measured with the so-called tag and probe method using $Z \to e^+e^-$ events both in data and simulation. In this method, probes are taken from events that are tagged by a well-identified electron selected by an invariant mass constraint ($65 < m_{ee} < 115$ GeV). Only a little background enters in this mass window around the Z peak. Consequently, it results in a sample of mostly real electrons on which selection efficiencies can be studied. The systematic effects considered are the tag selection, MC generator differences, the shape of the background fitting function as well as the signal shape (analytic fit or template from Monte Carlo). For photons, the $Z \to e^+e^-$ sample is used analogously, the only difference being that the electrons are reconstructed as photons without applying a track-based electron veto. The ratio of the identification efficiencies in data over MC (often called data/MC scale factors) are typ-



(a) Electron candidates with $p_T > 20$ GeV (b) Electron candidates with $5 < p_T < 20$ GeV

Figure 3.4: ROC curves for the electron multivariate identification (Boosted Decision Trees) and the cut-based selection working points for $p_T > 20$ (a) and low p_T (b) [83]. The signal is from Drell-Yan+Jets Monte Carlo. The background consists of reconstructed electrons from Drell-Yan+Jets Monte Carlo that do not match a generated electron within a cone of size $\Delta R = 0.1$. The signal efficiencies are not corrected for data/MC scale factors, which affect more the BDT selection.



(a) (b)

Figure 3.5: $Z \to e^+e^-$ mass distribution (a) and fitted Z mass versus Z p_T (b) where the (sub-) leading electron satisfies $p_T > 25(20)$ GeV, pass the medium cut-based identification and the pair is of opposite sign. The trigger requires a 23 and a 12 GeV online electron, and the data has been re-reconstructed with improved ECAL detector calibrations. The electron ECAL energy has been corrected post reconstruction to ensure good data/MC agreement at the Z peak. The peak is fitted with the convolution of a Breit-Wigner and a Crystal Ball function in the range 81 to 101 GeV to obtain the mass scale shift (ΔM) and resolution (σ_{CB}) [83].

ically around 0.95 in the barrel and 0.90 in the endcaps and increasingly deviate from one with tightened selection. For photons, the scale factors typically deviate less from one than in the electron case as the selection does not rely as much on the new pixel detector, which is not modeled optimally at the time of the initial 2017 calibration. The data/MC scale factors are another case where the revised ECAL calibrations for 2016 data improved data/MC agreement. This is reflected in almost all calorimetric ID input variables, mostly in complex ones like the electromagnetic energy flow isolation.

The DC-DC converters providing low voltage for the new pixel detector modules state failing in October 2017. This affected up to 10 % of the channels during data-taking and raised concerns about efficiency losses in the reconstruction of electromagnetic objects. As the effect is not present in the simulation, the efficiency loss would also show up in the data/MC scale factors for the identification step. The scale factor deviation from unity indicates the maximum possible effect. The difference between the electron and photon scale factors gives the scale of the issue since the photon identification relies almost completely on calorimetric information. From comparing periods with and without the DC-DC converter issue, the effect is estimated to be a few percent efficiency loss at the identification step. This is thought to be due to degraded tracker resolution affecting the ECAL-Tracker matching cuts.

3.2.5 $\mathbf{Z \to e^+e^-}$ invariant mass with full energy corrections

To put the reconstruction and identification to the test, one can extract the Z peak from data and Monte Calo simulation, as was done in Figure 3.5a for the recalibrated 2016 reconstruction. The resolution, obtained by fitting a Breit-Wigner convoluted with a Crystal Ball function, is approximately 1.8 GeV, which is very comparable with Run 1. Data and MC agree in Figure 3.5a by construction, as the mass scale correction in data and the resolution smearing in MC is calibrated exactly on the Z peak. Therefore, the good data/MC ratio in Figure 3.5a serves more as a validation of the calibration procedure. The scale and smearing are not calculated differential in Z boson p_{T}, so comparing the fitted Z mass as a function of the Z boson p_{T} is a benchmark where data and MC do not agree by construction. This is done in Figure 3.5b, which shows excellent stability for the mass scale agreement versus the p_{T} of the Z boson.

3.3 Muon reconstruction

Besides the undetectable neutrinos, muons are the only particles that can escape the CMS detector, leaving a trace in the muon detectors. Even if highly-energetic hadrons can punch through the HCAL, the iron return yoke ensures they rarely reach the outermost muon stations. Therefore, the CMS detector measures muons very well. Section 3.3.1 explains the basics of muon track reconstruction and momentum measurements. Afterwards, Section 3.3.2 presents widely-used muon identification and isolation criteria, motivated by their use in the WWZ analysis.

3.3.1 Muon track reconstruction

The first step in the muon reconstruction [89] is the reconstruction local to the muon system. *Standalone muons tracks* are reconstructed only with the muon spectrometer. Track segments are found among the DT and CSC hits. Together with the relevant RPC hits, a fit of these track segments gives the kinematic parameters. Not all the track segments – or *track stubs* – have enough hits to fit the kinematic parameters and their uncertainties. These track stubs are kept independently of the standalone muon tracks to combine them with inner tracker information later.

Muons are reconstructed from the standalone muon tracks, the track stubs, and the inner tracker tracks in two ways. In the first *outside-in* approach, a tracker track is matched to each standalone muon track if possible. The hits from both tracks are then fitted

together, taking into account multiple scattering and the average energy loss along the detector radius. Compared to tracker-only track fits, the resulting *global muons* can have an improved momentum resolution for high-p_T muons.

Then, there are *tracker muons*, which are reconstructed *inside-out*. All tracker tracks above $p_T > 0.5$ GeV are extrapolated to the muon system. If they are matched with a track stub, they are promoted to muons without a refit that includes the muon stations' hits. The tracker muon reconstruction improves the reconstruction efficiency below $p_T \lesssim 5$ GeV. For muons with $p_T \gtrsim 4$ GeV, reconstruction efficiency is greater than 99 % for global and tracker muons combined. The standalone muon tracks that are not matched to any tracker track are promoted to *standalone muon* candidates. However, they are usually not used because the momentum resolution is worse, and the fake rate is higher.

For global muons, the momentum of the global fit is only assigned if both the tracker fit and the global fit result in $p_T > 200$ GeV and agree within $2\sigma_{q/p}$ with the tracker-only fit. Otherwise, the result of the tracker-only fit is assigned. One might wonder why the muon system does not contribute significantly to the momentum measurement and showcases reduced resolution for muons of regular energy. This might appear counter-intuitive, given the muon system's large lever arm and the high tracker occupancy. The answer has to do with the high material budget of the muon system, more specifically the iron return yoke. Multiple scattering and muon bremsstrahlung interactions in this dense material deflect the muon from its trajectory. Furthermore, the bremsstrahlung photons can cause further hits that confuse the track pattern matching.

3.3.2 Muon identification and isolation

The muon selection used in the WWZ analysis will be explained later when the analysis is described, but it uses several standard selections from the CMS reconstruction.

After the loose muon selection, the next selection that is usually applied is the *particle-flow* muon selection [90], which is optimized to identify both prompt muons and muons in jets while maintaining a low fake rate from charged hadrons. The consideration of muons in jets is important because the particle-flow candidates are used for jet clustering, and mistaking muons for charged hadrons would underestimate the jet energy. The particle-flow selection requirements are met by applying different criteria to isolated and nonisolated muons and by considering calorimeter information. Explaining the details of this selection would go beyond the scope of this overview.

The selection of global and track muons that pass the particle-flow selection is referred to as the *loose* muon identification. When additional track-quality and muon-quality requirements are included – such as fit χ^2 and track segment matching variables –, one obtains the *medium* identification that is also used in the WWZ analysis. The selection efficiency of the medium ID is shown in Figure 3.6, differential in p_T and η. In data, the measured selection efficiency is generally around 98 %.

One more ingredient for muon selection is a relative isolation variable. Just as for electrons, the isolation is the p_T sum of charged hadrons, neutral hadrons, and photons around the lepton. The charged hadrons can be corrected for pileup contributions via vertex assignment, but the neutral components need to be corrected by pileup extrapolation. While the standard isolation for electrons extrapolates from the global pileup estimate ρ, a local pileup extrapolation from the charged hadrons in the isolation cone is done for the muons. This local charged-to-neutral extrapolation is called the $\Delta\beta$ correction. However, this $\Delta\beta$ correction is not used for the triboson analysis presented in this thesis. Instead, a custom isolation variable explained later in Section 6.3.1 is used. Since

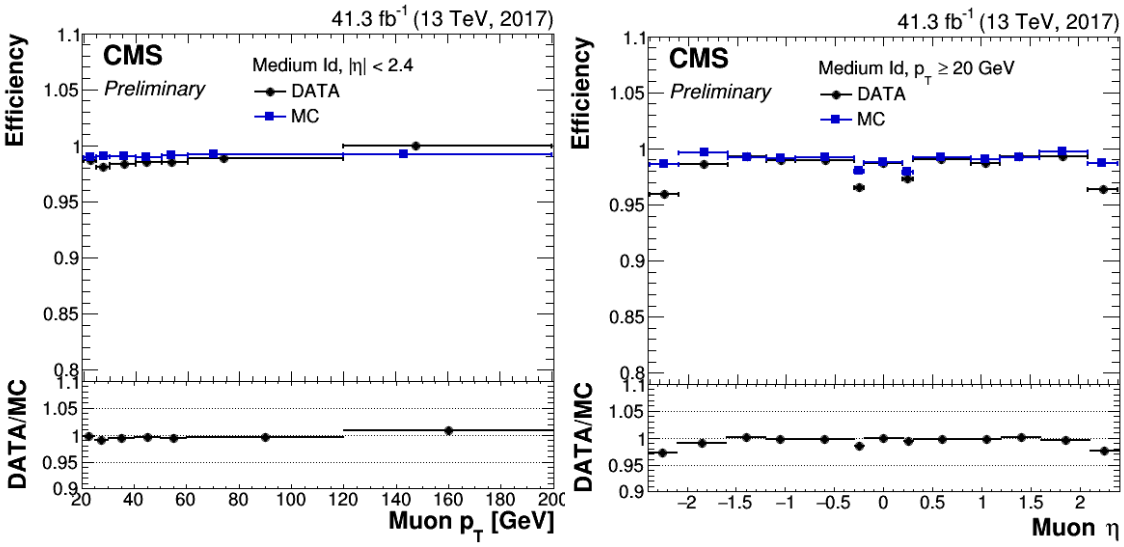


Figure 3.6: Medium muon ID efficiency as a function of p_T (left) and η for 2017 data and MC simulation [91]. In the denominator are all tracker tracks with $p_T > 20$ %. The drops at high η are due to inactive muon chambers, which are not modeled in the simulation. The dips around $|\eta| = 0.2$ are caused by the gaps between the wheels of the muon detector. The ID efficiency is also studied as a function of the number of reconstructed vertices, which shows that the efficiencies are not significantly affected by the pileup level.

it is often debated whether a local pileup extrapolation is better than a global extrapolation or not, this chapter later compares both approaches from a statistical perspective in Section 3.5, arguing that the level of correlation between charged and neutral energy in hadron collisions is such that both extrapolation methods indeed perform similarly.

3.4 Jet reconstruction

As mentioned in Section 1.1.8, quarks and gluons produced in proton collision events cannot be observed directly. They immediately hadronize, forming color-neutral states. As the hadrons move through the detector and decay further, they produce a directed group of particles called a *jet*. A jet can contain both neutral and charged hadrons and – to a lesser degree – leptons. To infer the properties of the original parton (quark or gluon), the measured particles need to be clustered to reconstruct the jets. CMS uses the so-called *anti-*k_T algorithm for jet clustering, described in Section 3.4.1. The momentum of the original parton is, in principle, the sum of the jet constituent momenta. In practice, the momentum sum does not precisely represent the partons momentum because neutrinos in the jet are not measured, not all low-p_T particles are properly reconstructed, and the nonlinearities in the detector response can cause a bias. Section 3.4.2 explains how the jet energy is calibrated to counter these effects and how the jet energy resolution is smeared in simulation to match the resolution in data. Something closely related to the jet reconstruction is the measurement of the missing transverse momentum, explained in Section 3.4.3. Finally, it is important to discuss the identification (or *tagging*) of heavy-flavor b or c jets (Section 3.4.4). All these explanations are motivated by the needs of the WWZ analysis presented in this thesis: jet reconstruction and especially b tagging are essential to reject one of the main backgrounds, namely the $t\bar{t}Z$ process. Missing energy measurements are essential to separate the ZZ background, which has no intrinsic missing energy, unlike the WWZ signal.

3.4.1 Jet clustering

In early hadron collider experiments, jets were clustered from energy deposits in the calorimeters. The last-generation experiments, such as CMS, usually combine the calorimeter measurement with track information to get a better momentum resolution on charged jet constituents. Specific to CMS, this means that the jet reconstruction uses the particle-flow candidates that were explained in Section 3.1. As a result, the overall jet energy resolution is improved, in particular at low p_{T} [81].

The CMS collaboration uses the anti-k_{T} clustering algorithm for jet reconstruction [92]. It is an iterative combination algorithm, where constituents are merged to jets based on two distances:

$$\begin{aligned} d_{ij} &= \min\left(k_{\mathrm{T},i}^{2p}, k_{\mathrm{T},j}^{2p}\right)\frac{\Delta_{ij}^2}{R^2}\,, \qquad \text{with} \quad \Delta_{ij} = \sqrt{\Delta y_{ij}^2 + \Delta\phi_{ij}^2}, \\ d_{iB} &= k_{\mathrm{T},i}^{2p}. \end{aligned} \tag{3.1}$$

Here, $k_{\mathrm{T},i}$ denotes the transverse momentum of particle i. The parameter p tweaks how momenta and geometrical separation are compared and is set to $p = -1$ in the CMS reconstruction. The radius parameter R steers the average angular size of the clustered jets. For the CMS jet reconstruction, it is set to $R = 0.4$ by default. In parallel, jets are reconstructed with to $R = 0.8$ for the clustering of wider so-called *fat jets* that aim to capture boosted decay systems, for example, the hadronic decay of boosted W bosons.

For each pair of particles (i, j), the algorithms merges them if $d_{ij} < \max\left(d_{iB}, d_{jB}\right)$. A pseudo-particle then replaces the two merged particles with the two added momenta as its momentum, and the algorithm proceeds with the next merging iteration. With the choice of $p = -1$, the distance parameter d_{ij} is small if a high-momentum particle is involved. This value ensures that the clustering is infrared safe[1] and focused on hard particles, merging the soft particles around them. The result is a set of cone-shaped jets that are not too much affected by pileup energy deposits.

3.4.2 Jet energy scale and resolution corrections

The momentum of the jets clustered with the anti-k_{T} algorithm is only a proxy for the original quark or gluon momentum, as the jet energy is biased by missing neutrinos, low-p_{T} particles that are not reconstructed, and nonlinearities in the detector response. To get the real parton momentum – or more precisely, the jet energy that would have been obtained by clustering truth-level particles, a Jet Energy Scale (JES) correction has to be applied [94]. The modeling of hadronization processes and hadronic showers that form in the calorimeter is challenging, so the jet energy resolution after applying the JES is lower in data than in simulation. This inconsistency is corrected for with Jet Energy Resolution (JER) corrections. Consecutively, most steps of the JES are applied to both data and simulation, while the JER corrections are only made for simulation.

The JES correction follows a factorized approach that is outlined in Figure 3.7. The first step is the pileup subtraction step. Out-of-time pileup corrections are calibrated by varying the bunch spacing and integration time of the calorimeter measurements. In-time pileup corrections are done separately for charged and neutral particles. Charged hadron particle-flow candidates that are not associated with the primary vertex are removed before the clustering, often referred to as the Charged Hadron Subtraction

[1] If $p = 1$ (k_{T} algorithm), the clustering would not be infrared safe [93].

(CHS). Contributions from neutral particles are estimated by extrapolation from the global pileup estimate ρ. This is the same area-median method also used for electrons, but this time the target area is not a fixed isolation cone but the actual jet area. The jet area is computed by artificially adding many infinitesimal soft particles around the jet before redoing the clustering and checking the area from which the soft particles have been included [95].

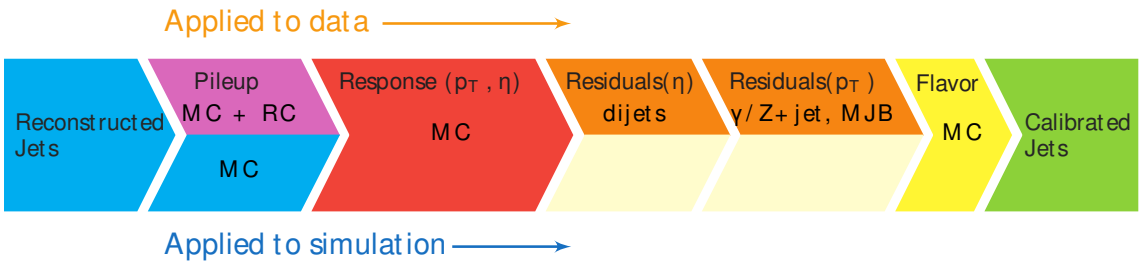


Figure 3.7: Steps of jet energy correction (JEC) for data and simulation [94]. The corrections marked with MC are derived from simulation only. RC stands for random cone, and MJB denotes an analysis with multijet events. The flavor corrections are optional and not used in the analysis presented in this thesis. Hence, they are not explained further.

Next follows the detector response calibration. The energy of *gen-jets* that are clustered from all generator-level particles is compared to the reconstructed jet energy in simulated multijet events. Response calibration factors are then computed in bins of p_T and η. The last two steps of residual corrections are only applied to data, correcting for differences between data and simulation. The first set of η-dependent residual corrections is calibrated by analyzing dijet events, making use of the momentum balance in the event. The second set of residual corrections is calibrated with Drell-Yan events where the Z boson or photon is recoiling against a jet. Once again, momentum conservation is exploited, but this time to get p_T-dependent corrections. This p_T-dependent calibration is made possible by the superior momentum resolution for the leptons from the Z boson or photon decays.

After the jet energy's central value is calibrated, there are still differences between data and simulation in the jet energy resolution. This mismatch is corrected by smearing the simulated jet four-momenta by a factor c_{JER}. This factor is inferred from a calibrated data-to-simulation scaling factor s_{JER} and the p_T resolution σ_{JER} observed in simulation. If the reconstructed jet can be matched to a gen-jet, the smearing factor is computed such that it worsens the agreement of p_T^{reco} with p_T^{gen}:

$$c_{\mathrm{JER,matched}} = 1 + (s_{\mathrm{JER}} - 1) \times \frac{p_T^{\mathrm{reco}} - p_T^{\mathrm{gen}}}{p_T^{\mathrm{reco}}}. \tag{3.2}$$

In case there is no matching gen-jet, the smearing factor is obtained from a random Gaussian variable $\mathcal{N}(0, \sigma_{\mathrm{JER}})$ around zero with variance σ_{JER}^2:

$$c_{\mathrm{JER,unmatched}} = 1 + \mathcal{N}(0, \sigma_{\mathrm{JER}}) \times \sqrt{\max(0, s_{\mathrm{JER}}^2 - 1)}. \tag{3.3}$$

The reason for the two different prescriptions is that a deterministic algorithm is preferred if possible.

3.4.3 Missing energy reconstruction

The fundamental particles that interact in proton collisions are quarks or gluons that only carry a random fraction of the proton's momentum. Therefore, the longitudinal

momentum sum of the final state particles in proton collisions is unknown on an event-by-event basis. Nevertheless, the transverse momentum sum of the colliding partons is zero because the protons collide head-on in the detector. Subsequently, if one sums all the four-momenta of the detected particles in the event, a nonzero transverse component indicates undetected particles. The only particles of the standard model that are not detected by the CMS detector are neutrinos. Accordingly, the missing transverse momentum $\vec{p_{\mathrm{T}}}^{\mathrm{miss}}$ is the only signature of neutrinos:

$$\vec{p_{\mathrm{T}}}^{\mathrm{miss}} = -\sum_{i=1}^{N_{\mathrm{PF}}} \vec{p}_{\mathrm{T}}^{(i)}. \tag{3.4}$$

The absolute value of $\vec{p_{\mathrm{T}}}^{\mathrm{miss}}$ is often referred to as MET for missing transverse energy, or simply as $p_{\mathrm{T}}^{\mathrm{miss}}$. In Equation 3.4, the summing is, in principle, done over the reconstructed particle-flow candidates in the event. For the actual computation of the particle-flow based MET, the particle-flow candidates that ended up in jets are not considered directly, but instead, the jet momentum is taken with all the corrections discussed in Section 3.4.2 applied. This special treatment of jets ensures that the MET value is not affected by neutrinos in jets from hadron decays but only by prompt neutrinos. In hadron-collision experiments, it is very challenging to achieve a fair resolution on $\vec{p_{\mathrm{T}}}^{\mathrm{miss}}$ because low-energy particles might not be reconstructed, and the precision on the measured particles is limited. The missing transverse momentum algorithms and their performance is explained in detail in Reference [96].

The particle-flow (PF) MET above is not the only MET definition used in CMS. A simpler alternative is a calorimeter-based MET. The most sophisticated MET definition available relies on the so-called pileup per particle identification (PUPPI) algorithm, which predicts a probability for each PF candidate to originate from a pileup interaction. For the analysis in this thesis, the particle-flow MET was consistently used.

3.4.4 Tagging of heavy-flavor jets

The identification or *tagging* of heavy-flavor jets aims to distinguish b or c quark jets from jets caused by gluons or light-flavor quarks. To study such heavy-flavor jets, one first needs a reliable definition of heavy-flavor jets that can be applied to the simulation, where this definition should preferably not depend on any additional parameters such as a maximum distance between the jet and the generator-level parton. In CMS, the flavor for a jet is defined at the hadron-level. The generator-level b and c hadrons that are not themselves daughters of b or c hadron decays are included in the jet-clusterin step, with their momenta scaled down to infinitesimal values. A heavy-flavor jet is then defined as a jet that picked up such a generator-level heavy-flavor hadron. If a b hadron is included, the jet is defined as a b jet, even if c hadrons are clustered too. If c hadrons but no b hadrons are included, the jet is deemed a c jet. A light-flavor jet is a jet with no such ghost b or c hadrons included, but still matching to a gen-jet. All other jets are categorized as pileup jets.

From an experimental perspective, several features distinguish a heavy-flavor from a light-flavor jet. The b or c hadrons decay a few mm to cm away from the primary vertices, forming secondary vertices that can be reconstructed from the decay products' displaced tracks. The mass of heavy-flavor hadrons is roughly one order of magnitude above the mass of light-flavor hadrons[2]. Therefore, the invariant mass of all decay products from the secondary vertex provides valuable information. The leptonic branching

[2]The average mass of b hadrons lies around 5 GeV and that of c hadrons around 2 GeV. Light-flavor hadrons such as pions and kaons have a mass of several 100 MeV

fraction is also higher for heavy-flavor hadrons than for light-flavor hadrons. About 20 % of b jets contain at least one soft electron or muon. This number reduces to 10 % for c jets.

The state-of-the-art heavy flavor tagging algorithms used in CMS are combining secondary vertex and displaced track information in a machine-learning model to obtain probabilities that a jet belongs to a given flavor category. The recommended model for Run 2 analysis used in this thesis's analysis is the *DeepCSV* tagger. This tagger uses a multiclassifier deep neural network to predict normalized probabilities $p(\mathrm{bb/b/c/udsg})$. Here, the bb output class corresponds to jets with at least two clustered b hadrons. For typical b tagging use such as in the WWZ analysis, the b-tagging score is defined as $p(\mathrm{bb}) + p(\mathrm{b})$.

The DeepCSV tagger is not the latest and greatest b-tagging model that was developed in CMS. The next-generation *DeepFlavor* or *DeepJet* tagger uses high-level jet and vertex observables and information on all PF candidates in the jets, instead of limiting the track information to six tracks like in the case of DeepCSV. It also has two additional output categories: one for b jets with leptons and one for gluon discrimination. The ROC curves that indicate the performance of the DeepCSV and DeepFlavor taggers are shown in Figure 3.8, which also shows how the DeepCSV performance improved with the new pixel detector.

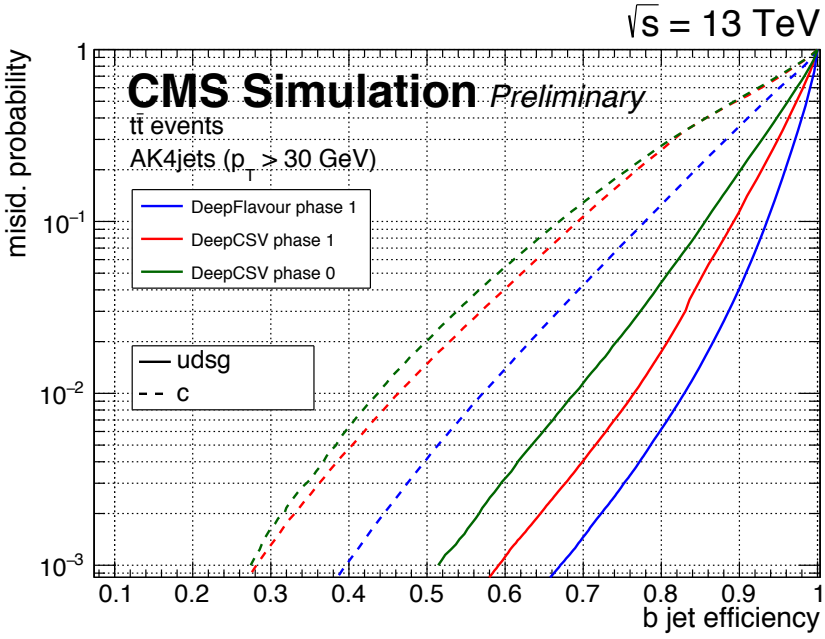


Figure 3.8: ROC curves showing the performance of the DeepCSV and DeepFlavor b tagging algorithms [97]. The jets for these curves are from simulated $\mathrm{t\bar{t}}$ events with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.5$. Here, b jets are defined as b jets from gluon splitting to two b quarks. The absolute performance for other definitions is not necessarily the same, but the figure clearly shows the relative improvement related to the Phase 1 pixel detector (green to red) and the improvement with the DeepFlavor algorithm (red to blue).

3.5 Statistics of in-time pileup extrapolation

At the LHC experiments, the reconstruction algorithms have to be made robust to pileup contamination of the event. One can use vertex or timing information to reject any ob-

jects that originate from pileup collisions. However, it is not possible to reject pileup contributions to all observables. Energy deposits from neutral particles in the calorimeters are particularly troublesome[3], as no tracks are available to assign the energy deposits to collision vertices and ignore them if the corresponding track originates from a pileup vertex. This missing information on neutral pileup deposits is problematic for the computation of MET, jet energies, or isolation variables. In this situation, one usually extrapolates to the neutral energy deposit from a correlated observable, which can be the charged energy deposits in the same region or an observable related to the global level of pileup in the event. In this section, we will discuss the general statistical challenges that underly this extrapolation problem and review the two pileup correction techniques used for lepton isolation variables in CMS. At the end of this review, it should become evident why it has been challenging to choose a one-fits-all technique within the CMS collaboration and how the pileup correction for the lepton isolation variables used in this thesis analysis was chosen.

It is handy to have a framework to estimate the performance of different estimators and compare them. Let us label the random variable that corresponds to the local pileup contribution that we want to estimate *in a given event* X_1. Another random variable X_2 shall be the observable that we use to extrapolate X_1. The number of particles from pileup in a given area follows a Poisson distribution, which will approach a Gaussian in the high pileup limit. Therefore, it is reasonable to study the general case of extrapolation between normal-distributed variables. The conditional distribution of X_1 given X_2 in the case of a bivariate normal distribution is a well-known result [98]:

$$(X_1|X_2 = x_2) \sim \mathcal{N}\left(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho\left(x_2 - \mu_2\right), \left(1-\rho^2\right)\sigma_1^2\right), \tag{3.5}$$

where $\mu_{\{1,2\}}$, $\sigma_{\{1,2\}}$ and ρ are the expectation values, standard deviations and correlation coefficients of X_1 and X_2. If all parameters would be available, we could compute the expected value of the conditional probability distribution with Equation 3.5 and interpret it as a pileup contribution measurement. The variance of the conditional probability would correspond to the uncertainty of this unbiased measurement. However, the expectation values and standard deviations depend on the number of pileup collisions in the event. As they are unknown in this statistical treatment of a single event, we have to think of another estimator that does not require this information.

Fortunately, the ratio $r = \frac{\mu_1}{\mu_2}$ is independent of the pileup level in the event, so the best estimate we can come up with is $\hat{x_1} = rx_2$. As this is not the mean value of the conditional distribution, there must be an additional bias or variance term. The additional bias is zero, which can be noticed from the linearity of the expectation value:

$$\mathrm{E}\left[rX_2 - X_1\right] = r\mathrm{E}\left[X_2\right] - \mathrm{E}\left[X_1\right] = \frac{\mu_1}{\mu_2}\mu_2 - \mu_1 = 0. \tag{3.6}$$

For the variance, we use Equation 3.6 and the definitions of variance and covariance:

$$\begin{aligned}\mathrm{Var}\left(rX_2 - X_1\right) &= \mathrm{E}\left[\left(rX_2 - X_1\right)^2\right] - \mathrm{E}\left[rX_2 - X_1\right] \\ &= r^2\mathrm{E}\left[X_2^2\right] + \mathrm{E}\left[X_1^2\right] - 2r\mathrm{E}\left[X_1X_2\right] \\ &= r^2\left(\sigma_2^2 + \mu_2^2\right) + \left(\sigma_1^2 + \mu_1^2\right) - 2r\left(\rho\sigma_1\sigma_2 + \mu_1\mu_2\right).\end{aligned} \tag{3.7}$$

[3]This will change in the future with the HGCAL, which has a timing resolution at the order of tens of picoseconds for energy clusters.

All terms with mean values cancel out. One can further define $r_\sigma = \frac{\sigma_1}{\sigma_2}$ to write down a simple equation for the variance of the extrapolation estimator:

$$\mathrm{Var}\left(rX_2 - X_1\right) = \left(1 + \frac{r^2}{r_\sigma^2} - 2\frac{r}{r_\sigma}\rho\right)\sigma_1^2 = \left(1-\rho^2\right)\sigma_1^2 + \left(\frac{r}{r_\sigma} - \rho\right)^2 \sigma_1^2. \tag{3.8}$$

This final result was written in two different forms. The first one clarifies that the variance is linear in the correlation, and the second one highlights that the variance is strictly greater than the variance of the conditional probability distribution, as expected.

Many possible conclusions can be drawn from Equation 3.8. It was constructed to be the variance of a local pileup extrapolation estimate, but nothing speaks against checking the limit of a global pileup extrapolation. This limit corresponds to $\mu_2 \gg \mu_1$, which means $r \to 0$. Using further the absence of correlation to local fluctuations, the variance of global extrapolation estimates is simply σ^2. This result can also be obtained from another perspective: a global pileup estimation has minimal variance because it sums over a large area, so it can be interpreted as a very accurate measurement of μ_1, where μ_1 is at the same time used as the pileup estimate. Therefore, the variance of the global pileup extrapolation estimate is equal to the variance of X_1, which is again σ_1^2. With these insights, we can write down the variance improvement for a local extrapolation relative to a global extrapolation, further simplified with the Poissonian relation ($r_\sigma = \sqrt{r}$). The variable X_1 will be renamed to X for a cleaner notation:

$$\frac{\mathrm{Var}\left(\hat{X}_{\mathrm{local}} - X\right)}{\mathrm{Var}\left(\hat{X}_{\mathrm{global}} - X\right)} = 1 - \rho^2 + \left(\sqrt{r} - \rho\right)^2. \tag{3.9}$$

We can also estimate the benefit of combining a local and global estimate by again interpreting the global estimate as a measurement of μ_1. In this case, one knows all parameters of the bivariate distribution and can use the actual conditional expectation value as the estimator. The variance is then given by Equation 3.5 and the improvement relative to the global estimate is:

$$\frac{\mathrm{Var}\left(\hat{X}_{\mathrm{combined}} - X\right)}{\mathrm{Var}\left(\hat{X}_{\mathrm{global}} - X\right)} = 1 - \rho^2. \tag{3.10}$$

There are two things to note before interpreting Equations 3.9 and 3.10. First, they do not depend on the global pileup level, but only on ratios and the correlations of local fluctuations. Hence, these results not only apply to the statistical description of a single event but also multiple events with an arbitrary distribution of pileup. Second, it does not matter what statistical processes are involved in the measured energy per particle, as the particles are independent. Therefore, the additional variance would scale with the number of particles, which is how the variance from the Poisson distribution scales already. In other words, r_σ does not depend on the processes related to the individual particles, as long as the Poisson assumption is valid.

Equations 3.9 and 3.10 are shown in Figure 3.9 as a function of ρ for $r = 1$ and $r = 0.5$ to visually aid the interpretation. The case $r = 0.5$ is of interest because it corresponds to the neutral-proportional-to-charge (NpC) pileup subtraction (there are two times more charged than neutral particles on average). Plugging $r = 0.5$ in the equations leads to two important predictions about the theoretical limitations of the NpC approach:

1. The correlation between the energy measured from neutral and charged particles must be greater than $\rho > 0.35$ for the NpC estimate to have a smaller variance than the global extrapolation.

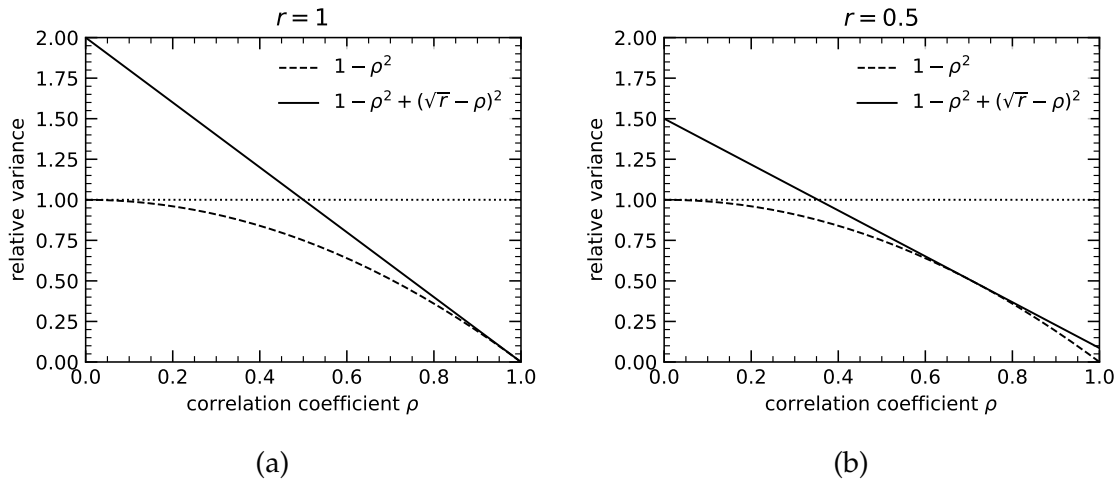


Figure 3.9: Expected variance improvements of the local (Equation 3.9, solid line) and the combined (Equation 3.10, dashed line) pileup extrapolation compared to the global estimate (dotted line) for $r = 1$ (a) and $r = 0.5$ (b). The variable ρ should be interpreted as the correlation coefficient of the local fluctuations that are exploited for the local extrapolation.

2. By combining the local extrapolation with the global extrapolation, one can maximally reduce the statistical measurement uncertainty to 93.5 % of the univariate extrapolation[4].

The first prediction explains why the local pileup extrapolation usually performs worse than the global extrapolation. In minimum bias events recorded by CMS, the correlation between neutral and charged hadrons for typical cone sizes used for jets and isolation variables is less than 35 %. The second prediction is in agreement with recent simulation studies on the use of charged-track information to subtract neutral pileup [99], which obtained $\sigma_{\text{combined}} \simeq 0.96\ \sigma_{\text{global}}$ for the combination with optimal weights.

Finally, it is essential to understand the main limitations of the model, most importantly, to know in which direction any predicted values probably change in reality. From the measurement of charged particle spectra in minimum-bias events [100], one can infer that at a mean pileup of 20, there are around 15 charged particles in typical isolation cones of $\Delta R = 0.3$, so the assumption of a Poisson process in the Gaussian limit is reasonable. Assuming that the global observable is an exact measurement of the number of pileup interactions is acceptable too, as the statistical uncertainty is very low compared to any local observable. The most questionable assumption is that the ratio of charged to neutral particles is uniformly 2 to 1, mostly because the track reconstruction efficiency is not uniform in pseudorapidity. There are also subtleties related to hadron masses: the charged-over-neutral ratio is different for transverse and longitudinal components [99]. Therefore, in reality, the global extrapolation method will probably have more of an advantage than this model predicts.

After this discussion of pileup extrapolation from a statistical perspective, comparing global pileup extrapolation methods like the area-median method to local methods like the NpC subtraction, it should be clear why local pileup extrapolation is challenging.

[4]The difference between the combined extrapolation and either the global or local extrapolation is largest around $\rho \approx 0.35$ (see Figure 3.9b). The variance is reduced by a factor 0.874, which reduces the standard deviation by a factor $\sqrt{0.874} = 93.5$ %.

3.6 Simulation of proton-proton collisions

The large dimensionality of the final state of proton-proton collisions makes Monte Carlo techniques very appropriate for simulating such collision events. This is why the terms *simulation* and *Monte Carlo (MC)* are often used interchangeably in HEP jargon. The event simulation chain comprises several steps. The first step is the simulation of the *hard scattering* event, predicting the distribution of final-state particles directly after the interaction with the high momentum exchange. The resulting particles can further split via QCD and electromagnetic interactions, which is simulated in the *parton showering* step. A thorough introduction to parton-shower event generators is given in Reference [101]. Next, the *hadronization* of the quarks and gluons at the end of the parton shower to color-neutral hadrons is simulated. The parton remnants that did not participate in the hard scattering can take part in soft interactions. These multi-parton interactions make up the *underlying event*, which is modeled as well. At the end, the interaction of the final stable particles with the detector and its electronic response is simulated in a detailed *detector simulation*. All these steps will be elaborated on in the following.

The simulation of any hard scattering process requires a precise description of the Parton Density Functions (PDFs) $f_i(x, Q^2)$ of the proton, which encode the probability to find a given parton i with momentum fraction x at the probed momentum scale Q^2. These PDFs are extracted from various measurements, such as deep inelastic scattering, Drell-Yan, and multijet measurements by collaborations like NNPDF [102], whose results are used for the CMS event simulation. The PDFs are convoluted with the partonic cross-sections, which are evaluated up to some order in perturbation theory. The maximum order is usually not the same for different simulated processes, as higher-order corrections are very computationally expensive, and not all measurements demand the same level of precision. The uncertainty from truncating the perturbation series is estimated by varying the *factorization scale* μ_F that regulates ultraviolet divergences, as well as the *factorization scale* μ_F that regulates infrared and colinear divergences and is also affecting the PDFs. Usually, the higher the maximum order in the perturbation expansion, the smaller the impact of these scale variations. The nominal value of these parameters is usually the energy scale of the simulated process.

The partons from the hard scattering can undergo a chain of softer interactions that are still in the perturbative energy regime. This showering can happen both in the initial state or the final state and is not considered in the hard scattering simulation to limit the complexity of the computation. Instead, the simulation of the parton shower recursively evaluates the probabilities for gluon radiation and colinear splittings of gluons into quark pairs until non-perturbative energies are reached. A parton shower simulator that is widely used for CMS simulations is Pythia [103], which can also evaluate hard scattering probabilities at leading order but is usually interfaced with more specialized event generators such as `MadGraph5_aMC@NLO` [104]. For the simulation of processes with multiple partons in the final state, the parton showering has to be accurately matched to the matrix element computation to avoid double-counting. This affects, for example, the simulated DY+Jets samples that are used for the optimization of the electron identification described in Chapter 5. In the CMS collaboration, the MLM matching scheme [105, 106] is usually used for leading-order simulations, while the FxFx method [107] is employed for NLO simulations.

The hadronization of the parton shower to color-neutral hadrons cannot be predicted with perturbation theory. Instead, it is simulated with phenomenological models that are tuned to measured data. A common model that is also implemented in the Pythia

library is the Lund string model [108, 109]. This model predicts an energy potential that increases linearly with the distance between quarks as they move apart. The field lines of this potential form a narrow flux tube connecting the quark pair, hence the name *string* model. If the potential energy is high enough, the string can be broken by a newly created quark-antiquark pair if the initial invariant mass is large enough. This process is repeated until only hadrons of on-shell mass remain. The full Lund sting model includes additional mechanisms to account for correlations among the flux tubes and to describe the baryon generations. It also produces no bottom and charm quarks, so all bottom and charm quarks originate from the hard scatter or gluon splitting in the parton-shower step. For the simulation of the underlying event, one must consider the proton remnants' hadronization, additional parton scattering processes that are most probable at low momentum exchange, and the color connections between the partons from the hard scattering and the underlying event. Just like the hadronization process, the underlying event simulation cannot be predicted without tuned parameters. For the simulation of proton-proton collisions in 2017 and 2018, the CMS collaboration extracted the new TuneCP5 parameters [110] for the simulation of the underlying event with Pythia.

A complete description of the CMS detector is implemented in the GEANT4 [111] framework that rigorously tracks the interactions and energy loss of particles up to a minimum energy threshold. The model includes both passive and active detector material and an accurate description of the magnetic field. The electronic response of the detector modules to the energy deposits is simulated in detail. At this point, the energy deposits from pileup simulations are mixed in, as the signals from pileup interactions and the hard scattering can interfere. From the electronic response onward, recorded and simulated events are reconstructed in the same way. This complete approach is referred to as a full simulation (*fullsim*), as opposed to a fast simulation (*fastsim*), where the simulation is simplified by high-level parametrizations to save computing time. Only fullsim samples are used in this thesis, so the fastsim approach is not explained further.

3.7 CMS reconstruction software

The CMS Software (CMSSW) framework [112] is designed following the Event Data Model (EDM) and allows for configurable implementations of event processing loops. In this EDM paradigm – sketched in Figure 3.10 –, initial event data is taken from an arbitrary source, for example, a data file or directly from the detector backend. A sequence of software modules implemented as C++ plugins then produces more event data and filters the events before an arbitrary output module writes selected event data to usually a file, in a way specified by the output module. The producer and filter modules declare what event data they require before the event loop is started, and the producer modules additionally report which event data they produce. The core framework then infers in which order the modules need to be run. The event products are usually instances of custom C++ classes that are serialized by the ROOT I/O system [113]. All modules are configured in the Python programming language.

The framework supports multithreading, meaning that it can run several producers in parallel on multiple *streams* that may get executed on different CPU threads. For this reason, the event data modules may inherit from different base classes, depending on their compatibility with multi-threaded execution. A *one* module is only instantiated once and processes all events strictly one after the other. This is useful for analysis plugins that need to accumulate information from complete datasets, such as histogramming

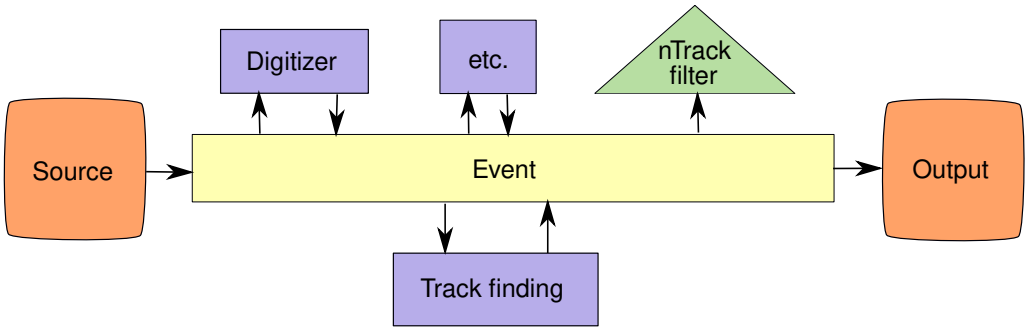


Figure 3.10: Visualization of the event data model. The source and output modules are shown in orange. The yellow rectangle in the center stands for the pool of event data produced and consumed by the producer modules in blue. Filter modules, indicated by a green triangle, only consume event data but do not produce any event data.

modules. Then, there are *stream* modules that get instantiated per stream. Most of the modules implemented in CMSSW are stream modules. However, their disadvantage is that the module cannot accumulate information about the full dataset and that module configurations are instantiated redundantly. The latter can be a problem in particular for modules that take much space in memory, for example, modules that make an inference with machine-learning models. One can also implement a *global* module, which is only instantiated once and then used by all streams concurrently. This implies that such modules cannot change their state during the event loop without causing race conditions. In other words, the modules must be *immutable*. For many algorithms used in the reconstruction, this is not given. Many algorithms are implemented in a FORTRAN-inspired style, where most helper variables and objects are instantiated at the beginning of the event loop and then mutated during the event processing.

Besides being appropriate for multithreading environments, immutable modules are often easier to understand and maintain. They are also less error-prone because information cannot leak from one event to another. Hence, global modules were always preferred when a large fraction of the electron reconstruction and identification modules were revisited in this thesis to implement the multivariate electron identification, as well as for optimization, refactoring, and cleaning of obsolete modules. This work set precedents for implementing complex algorithms in immutable plugins, besides making the electron reconstruction chain much more transparent. This is one example of the software-related work done in this thesis for the benefit of the CMS collaboration.

The CMSSW framework manages not only event data products but also so-called *event setup* products. These include detector geometry, calibration, and magnetic field information. The event setup products are valid not for one event, but for an *interval of validity*. Most of them are not produced on the fly but stored in the *conditions database*. The *global tag* set of a dataset specifies which version of the conditions data was used for the event processing, allowing for the selection of different detector and data-taking scenarios. Before Run 3, the producers were not required to declare which event setup data they need, so the framework had to run all event setup producers first. Now, event setup requirements have to be declared before the event loop, enabling the concurrent production of event data and event setup products. This evolution confirms CMSSW's position as a cutting-edge, multithreading HEP software framework.

Chapter 4

Machine learning fundamentals

Many research fields employ statistical learning – more frequently called *machine learning*. In Section 4.1.1, this chapter first explains the basics of machine learning problems, while connecting the vocabulary of the experimental High Energy Physics (HEP) community with common machine learning terminology. This introductory review is inspired by Reference [114]. Both the optimization of the multivariate electron identification and the triboson analysis make use of the Boosted Decision Tree (BDT) classification algorithm, where the BDTs are trained with the XGBoost open-source library [88]. Section 4.2 reviews BDTs and the XGBoost algorithm, in particular, its hyperparameters that are often referred to hereafter. A particular machine-learning related problem that occurs several times in this thesis is the issue of sample reweighting for the training of classification models. This subject is discussed in Section 4.3 at the end of this chapter, which presents some insights that are seldomly discussed in the literature.

4.1 Introduction to machine learning

When solving problems with computing machines, one usually starts with some input data and an algorithm to run on this data. The result of the computation is some output data. For *supervised* machine learning, the relation between the algorithm and the output data is inverted. One starts off with example inputs and outputs, and the machine runs a supervised-learning algorithm to produce a new algorithm – or *model* – that predicts the output for unseen inputs. Supervised machine learning models are usually categorized into *regression* and *classification* models. For regression, the output data are continuous variables, while for classification, the target is a discrete *class*. This section explains how to train classification models, as done in this thesis on several occasions.

4.1.1 Training a machine learning model

It is instructive to introduce some general machine learning methodology at the example of regression problems before discussing classification. For a regression problem, one aims to predict a continuous variable with one or more predictor variables. The predictors are usually called *features* or *input variables*, and the predicted variable is called *target* or *regression value*. The prediction is done with a *model* or *regressor* of the analysts choice, for example linear regression. One uses a training dataset to *fit* or *train* the model to example data. This fit involves the optimization of a *loss function*, for example, the mean squared error of the prediction.

If the model has too many parameters or the training dataset is too small, the model might be *overtrained* (more on this in Section 4.1.3). Therefore, one should not use the full available dataset for training, but hold out a fraction to *test* the performance of the

model on unseen data. Most machine learning algorithms profit from large training datasets, so the testing dataset is typically smaller than the training dataset.

Usually, a model has not only parameters that are fit to the training data but also parameters that steer the fitting process or control the shape of the model. Examples are the weights of regularization terms in the loss function or the maximum number of leaves in a decision tree. There are several ways in which these *hyperparameters* can be optimized, for example, via grid search, random search, genetic algorithms, or Bayesian optimization.

The hyperparameter performance needs to be validated in a dataset that was not used for the training. However, this hyperparameter tuning is itself subject to overtraining effects. Usually, a dataset of similar size as the test dataset is used as the *validation dataset* for hyperparameter tuning, while the actual test set is only allowed to serve for a final performance evaluation after all model-parameters and hyperparameters are fixed. To make most of the dataset used for training and hyperparemter validation, one often uses a scheme called *k-fold cross-validation*, visualized in Figure 4.1. In k-fold cross-validation, the model is trained k times, each time with a different slice of the data held out. This slice, or *fold*, serves as the validation dataset. When averaging over the different splits, one obtains a performance metric that uses the full statistical power of the dataset while the size of the training dataset for each split is maximized. The number k is usually limited by computational time. The k-fold cross-validation was used in this thesis to tune the hyperparameters of the multivariate electron identification with Bayesian optimization, which is explained later in Section 5.6.2.

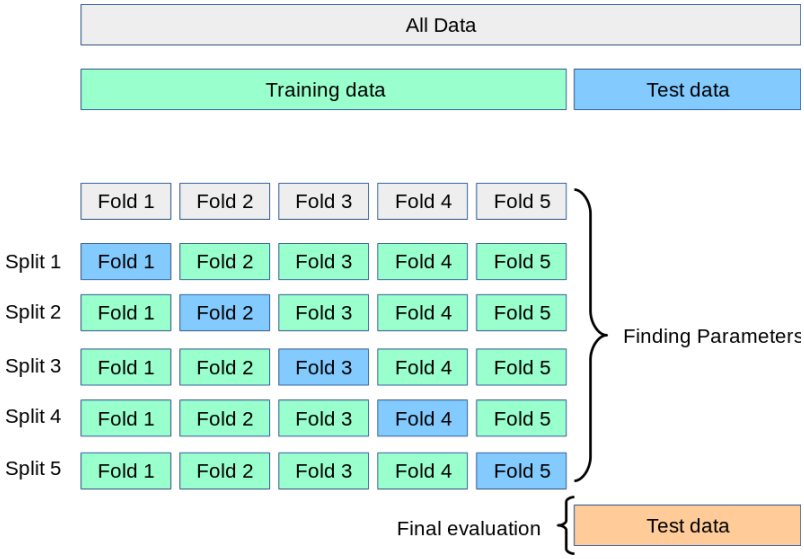


Figure 4.1: Visualization of k-fold cross-validation from [115].

4.1.2 Classification problems

The previous section introduced some machine learning terminology for regression problems. However, in this thesis, machine learning was mostly applied to *classification* problems, where the model has to predict the correct assignment to different groups or *classes* in the dataset. Examples are the separation of true reconstructed particles from fakes in a detector and the separation of recorded collision events according to the underlying interaction process.

The simplest form of classification problems is *binary classification* with only two classes. A classification model usually predicts a probability for each sample that may or may not reflect the actual probabilities in the application. The important bit is that this predicted probability – or *score* – encodes information that can be used with an appropriate cut or *working point* for a discrete class assignment. If one wants to assess the performance of a classifier without fixing a working point, one might study the *receiver operating characteristic (ROC) curve*. For the ROC curve, the true positive rate is plotted against the false-positive rate. A diagonal corresponds to no separation power, while a curve far from the diagonal indicates high separation power. These ROC curves will be shown often in Chapter 5, which discusses the optimization of the multivariate electron identification algorithm for CMS. To compress the ROC curve in a single number that can be easily compared for performance evaluation, it is common to calculate the *Area Under Curve (AUC)*. Being a performance metric independent of the working point, the AUC is very often the target of hyperparameter optimization. The area ranges between 0.5 and 1.0, where an area closer to one is better. Note that using the AUC as a performance metric also has disadvantages. Sometimes, the shape of the ROC curve will depend on the algorithm and its hyperparameters, so a larger AUC does not necessarily mean that the false-positive rate for a fixed true positive rate is higher. Therefore, it still makes sense to use working-point-dependent performance metrics such as the error rate if it is clear where the working point will be.

The training of classification models is often equivalent to a regression of the log-odds, which correspond to $\log\left(\frac{p}{1-p}\right)$ for binary classification. Therefore, one needs to do a conversion from the log-odds to probabilities at the end of the model, which is usually called the *logistic transformation* or the *sigmoid function*:

$$\hat{p}(x) = \frac{1}{1 + e^{-\hat{z}(x)}}. \tag{4.1}$$

This is a common concept for classification models, no matter if they are based on linear models (*logistic regression*), decision trees, or neural networks. The interpretation of classifier scores as log-odds is important for Section 6.7 that discusses how to get the most information out of the classification score for signal extraction.

4.1.3 The bias-variance tradeoff

The *bias-variance tradeoff* is a concept closely related to the already mentioned overtraining. A machine learning classifier never makes exact predictions, but predictions that spread away from the true value with a mean squared error (MSE) that decomposes into three terms. The first term, the *bias* term, is nonzero is the model does not have enough degrees of freedom to accurately learn the target for every point in the feature space. This can be easily understood with the extreme case of a univariate regression model with no degrees of freedom that always predicts zero. If model should predict some noiseless functional dependence $y(x)$, the model predictions will have a variance $\mathrm{Var}(x) = y(x)^2$.

The second term in the MSE is the *variance*. It appears when the model has too many degrees of freedom and learns to reproduce the noise fluctuations in the training dataset. For machine-learning models, this term rarely vanishes completely and is even present also for models with too few degrees of freedom, as these degrees of freedom might be spent to fit the noise in the training dataset if the true functional dependence is not well represented by the model. Accordingly, there is usually a sweet-spot in model complexity where the bias and variance are minimal, while for too complex models,

the variance is high (overtraining or *overfitting*), and for too simple models, the bias[1] is high (*underfitting*). This is the bias-variance tradeoff that needs to be considered in the optimization of any hyperparameter that controls the model complexity. The last term in the MSE is the *noise* in the data, which can not be predicted and is always part of the MSE for any problem.

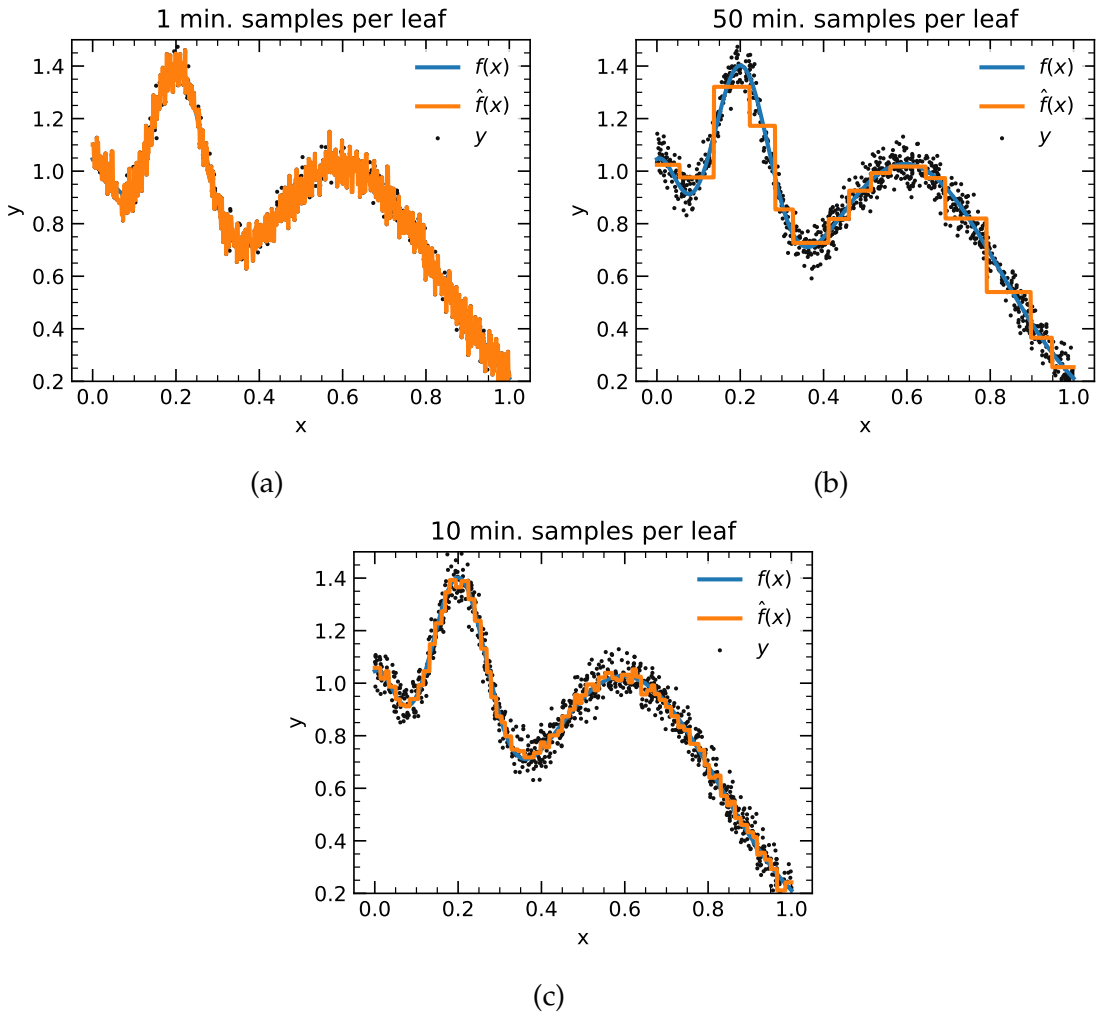


Figure 4.2: Fitting noisy data that follows a polynomial dependence on a single feature x with a regression tree for different values of the minimum samples per leaf, which is inversely proportional to the complexity of the model. Here, the fit with one sample per leaf clearly overfits the data (a), the model with 50 samples per leaf is not complex enough (b), and 10 samples per leaf are adequate to approximate the polynomial (c).

Figure 4.2 visualizes the overfitting and underfitting effects when modeling noisy data with a regression tree. The exact value of the bias and variance depends on the point in feature space. Figure 4.3 shows the values of the different terms of the MSE for different values of the model-complexity-controlling hyperparameter that was varied for this demonstration – the minimum number of samples from the training dataset in any leaf of the regression tree. If the minimum number of samples is one, the model perfectly fits the noise in the training data, so the MSE is twice the noise. As the number of minimum samples per leaf is increased to values beyond the sweet-spot, the bias term rises.

[1] and depending on the situation, also the variance

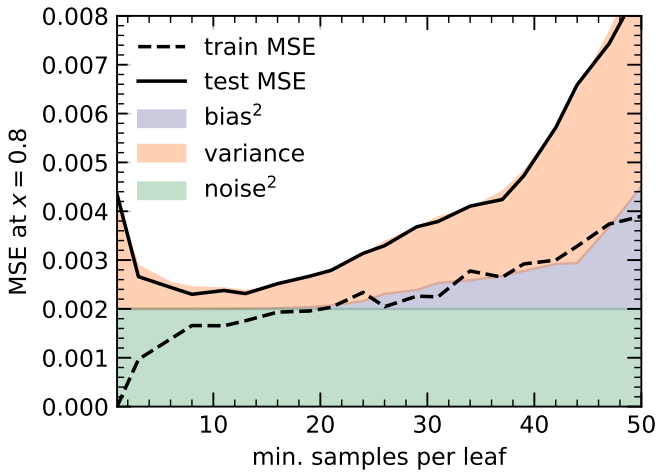


Figure 4.3: The decomposition of the MSE in the training and testing samples of the model predictions at $x = 0.8$ for the same example as shown in Figure 4.2. Note that in order to evaluate the bias (blue) and variance from the noise in the training dataset (orange), one must repeat the generation of toy data and the model fitting multiple times. The noise in the data (green) is independent of the fitted model. Here, MSE stands for *mean squared error*.

As Figure 4.3 shows both the MSE on the training and testing dataset, it conveys another important message: even for the hyperparameter value that makes the model fit best, the overtraining – which manifests itself as the difference between the train and test MSE – is not necessarily vanishing. In general, the sweet-spot of best test performance is not necessarily the point at which overtraining is minimal.

4.2 Boosted decision trees and the XGBoost algorithm

The XGBoost algorithm [88] to train decision tree ensembles is used several times in this thesis to tackle classification problems. Hence, it is worth to take a look at how this algorithm works also to understand the hyperparameters that were optimized for the Boosted Decision Tree (BDT) training in this thesis. Section 4.2.1 introduces the concept of decision trees and how gradient boosting can be used to train decision tree ensembles. Section 4.2.2 lists and explains the XGBoost hyperparameters that were considered in the hyperparameter optimizations in this thesis.

4.2.1 Gradient boosting for decision trees

A single decision tree should now be easy to understand, as we have already seen one in the previous section about the bias-variance tradeoff. As usual, one needs a loss function that will be optimized, such as the MSE for regression or the error rate for classification. Decision trees are grown iteratively, starting with a constant prediction that minimizes the loss function. Next, the training algorithm has to find the split in feature space that minimizes the loss function if separate constant predictions are made in the two *leaves* after the split. The splitting is repeated until some stopping criterion is met, for example, if the number of samples in a leaf would fall under a threshold, or if the improvement of the total loss function by splitting is not greater than the increase of an added penalty term that regularizes the model complexity.

In many cases, the variance of a model can be reduced by combining multiple models with few degrees of freedom (*weak learners*) to an ensemble model. The intuition is that because it is easier to learn the actual relation between the target and the feature than learning the more erratic noise, a model first fits the actual relation before fitting the noise. If the weak learners are given the opportunity to learn different things about the data, for example, by randomizing the training dataset, they learn – in the best case – different aspects of the true relation without wasting any degrees of freedom to fit noise. Hence, combining them might result in a better model than training a single learner with many degrees of freedom that tend to be used to learn about the noise.

Boosting means that the weak learners are trained in an iterative fashion, where each added learner learns to correct for the mistakes of the existing ensemble. A boosting method for decision trees that used to be very popular in the HEP community is the *AdaBoost (Adaptive Boosting)* algorithm [116]. In AdaBoost, wrongly predicted training examples get a higher weight when fitting the next weak learner. This very intuitive idea behind AdaBoost certainty helped to boost its popularity. *Gradient boosting* [117] is a more general[2] technique where every added weak learner is chosen such that it improves the loss function in the direction of its steepest gradient.

The loss function for the XGBoost algorithm when a new tree is added to the ensemble in the iteration t is:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t). \tag{4.2}$$

Here, l is an arbitrary per-instance loss function, for example, the squared error for regression or the logistic loss for classification. The new tree should make predictions $f_t(\mathbf{x}_i)$ that correct for the residue between the predictions of the existing ensembl, $\hat{y}_i^{(t-1)}$, and the ture target values y_i. The summing is done over the n samples in the training dataset. The term $\Omega(f_t)$ is a regularization term that will be discussed in Section 4.2.2, which covers the hyperparameters. To optimize this objective, the second-order approximation of Equation 4.2 is used:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(\mathbf{x}_i)) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i) \right] + \Omega(f_t), \tag{4.3}$$

where g_i and h_i are the first and second order gradients of the loss function at $\hat{y}_i^{(t-1)}$. The algorithm successively adds branchings of *splits* to the tree structure, as long as the loss function $\mathcal{L}^{(t)}$ reduces. The optimal predictions in each leaf can be computed from the gradients, as explained in detail in the XGBoost reference [88].

One of the caveats of boosting and the successive finding of splits is that the best tree structure might not necessarily be found by only looking one step ahead. However, it is impossible to enumerate all possible tree structures, making the *greedy* optimization a necessary compromise. The specialty of the XGBoost algorithms is speed, which is achieved with several optimizations. Some of these optimizations are enabled by approximations. The most distinct optimization is the *approximate* greedy algorithm. An exact greedy algorithm would enumerate all possible splits for all possible features to find the best next split. In the approximate algorithm implemented in XGBoost, the feature values are first histogrammed with quantile-based bins, such that the enumeration only needs to be done over the much fewer histogram bins.

[2] AdaBoost is, in fact, a special case of gradient boosting with a particular loss function (exponential loss).

4.2.2 Hyperparameters of the XGBoost algorithm

The pernalty term in the loss function (Equation 4.2) contains three regularization hyperparameters:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2 + \alpha \|w\|^1 . \tag{4.4}$$

In this penalty term, $\|w\|^2$ is the sum of squared predictions of *weights* in each leaf (L2 norm), $\|w\|^1$ is the sum of absolute values of the weights (L1 norm), and T is the number of leaves in the tree. The parameters γ, λ, and α are the hyperparameters, where λ is also called the L2 penalty and α the L1 penalty. The parameter γ corresponds to a fixed cost for adding a new split, keeping the model complexity under control. The parameters λ and α regularize the actual weights in the leaves. For classification, this means that the predictions are more conservative. The effect of the square in the L2 norm is that more extreme predictions are penalized more severely. Since many other hyperparameters are not abbreviated with greek letters, the regularization parameters will be referred to as `gamma`, `lambda`, and `alpha` hereafter.

Three more parameters that control the model complexity directly are `n_estimators`, `max_depth`, and `min_child_weight`. The effect of the `min_child_weight` parameter, which is a lower bound for the weight sum of training instances in each leaf, was already shown in Section 4.1.3 where is served as an example to illustrate the bias-variance tradeoff. The maximum depth of the tree, `max_depth`, limits the number of possible splits before arriving at a leaf. Finally, `n_estimators` is the number of boosting rounds, so it is equivalent to the number of trees in the final decision tree ensemble.

In Section 4.2.1, it was argued that even without boosting, separately-trained weak learners could be combined to a more powerful ensemble with less variance if they are allowed to learn different aspects of the training dataset. This can be enforced by the random selection of features and training instances for the training of each weak learner. This is also implemented in XGBoost, complementing the gradient boosting approach. The parameter `colsample_bytree` controls the size of the random selection of features used in a given boosting round. The default value of 1.0 means no random selection, so all features are used for each tree by default. Analogously, the `subsample` parameter controls the random sampling of training examples, also with a default value of 1.0

Finally, the `scale_pos_weight` parameter scales the weights of all instances of the positive class in the loss function. The use and implications of this parameter and the implications of per-instance reweighting – also possible with XGBoost – are presented in the next section.

4.3 Sample reweighting for classification

There are various applications of machine learning in HEP in which assigning non-equal weights to the different classes or even to the individual samples is appropriate. One of the actively studied applications of reweighting is the decorrelation of the features with model parameters. A recent example is the tagging of $H \to bb$ events with classifiers trained with reweighted samples such that the $b\bar{b}$ invariant mass distribution is uniform [118]. Therefore, the classifier is not biased towards the standard model Higgs boson mass.

Another scenario in which sample reweighting is often considered is imbalanced training data or training data in which the classes are weighted differently than in the envisioned application. For use cases in HEP, this is more the norm than the exception: training data is often simulated, and the number of events per sample is determined by computing resource constraints, which are different for each sample. The applications of statistical learning in this thesis are also affected by this problem, which was first encountered during the multivariate electron identification optimization. Therefore, the subject of reweighting for applications other than decorrelation will be discussed in the following.

Often, many more training examples are available from the positive class than from the negative class. This asymmetry is prevalent when dealing with the separation of physics processes: the signal is often generated in abundance, while preselection cuts often limit the background sample size (often just called *statistics*).

A few options come to mind:

1. apply no weights at all
2. discard samples from class with excess to even them out, do not reweight
3. normalize such that the summed weights within each class are equal
4. normalize each class to actual priors in the application, for example, cross-sections
5. apply weights from Monte Carlo event generators to individual events

To make an educated choice, one has to keep in mind several key points:

1. Every reweighting comes with a reduction of effective statistics. Hence, the performance of complex classifiers which require extensive training datasets can be impaired.
2. It is possible to calibrate the predicted probabilities a posteriori, so reweighting during the classifier's training is not strictly necessary to obtain meaningful probabilities at the end of the inference pipeline[3]. However, for many applications – like the identification of particles with a particle detector – the actual numeric value of the predicted probability is not of interest, so uncalibrated probabilities are not necessarily a problem.
3. The application of per-class weights can positively affect the performance at a given working point, a benefit that might outweigh the penalty imposed by the reduced statistics. This effect is particularly interesting for simple classifiers with few model parameters because of the absence of the statistics penalty.

[3]Beware that the calibration of probabilities is not trivial by itself, and might deteriorate final classification performance. This impairment is most apparent for histogram-based methods, where the information within a given bin is lost.

Essentially, this means that if interpretable probabilities are not a requirement, per-class weights are a mechanism to improve classification performance. They should be considered even for a training dataset that is a priori balanced, which was a lesson learned in this thesis. Traditionally, per-class weights are often only considered in the context of imbalanced datasets to balance the prior probabilities, while the effects on differential classification performance are less understood. Therefore, this section elaborates on the effects of per-class reweighting on the performance at a given working point, possibly explaining some observations made during the optimization of the multivariate electron identification.

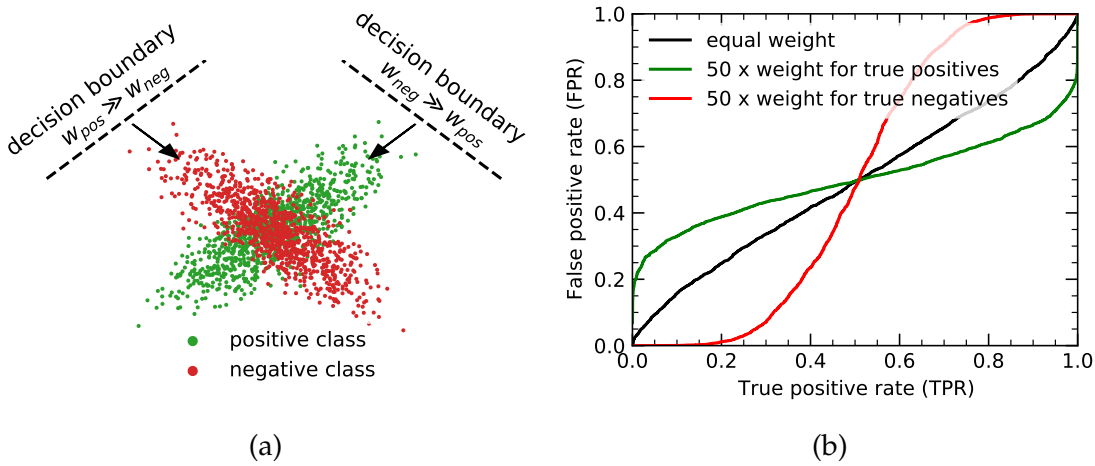


Figure 4.4: Illustration of the effect of class reweighting on the decision boundary obtained by logistic regression on a toy dataset designed to show a maximal effect size.

The effect of scaling up either the positive or negative class can be illustrated with a simple logistic regression example, but it generally applies to any classifier using a logistic loss function, for example, XGBoost or neural networks. Consider the two-dimensional example data shown in Figure 4.4a, where both classes are balanced. They cannot be separated without taking the correlation of the two features into account, which a logistic regression cannot do. The logistic regressions decision boundary is a straight line through the center of mass, with an arbitrary slope determined by statistical fluctuations. The ROC curve of this classification is indicated as the black line in Figure 4.4b. If one class is weighted much higher, the convexity of the logistic loss forces the decision boundary to be such that the distances to each sample of this class are comparable. In general, one might suspect that the decision boundary might turn out to be parallel to the first principal component of the class with the much higher weight[4]. In other words: the asymmetry of class weights breaks the symmetry of the decision boundary, which now has to choose a location where misclassification of the dominant class is avoided. In practice, this means that if false positives are penalized more, the performance improves for tight working points and vice versa, without impacting the area under the curve (see Figure 4.4b).

This differential effect is also manifest for more complicated toy datasets. The ROC curve in Figure 4.5 shows the result of the same exercise for a randomly generated classification dataset.

This line of thought should explain why the `scale_pos_weight` parameter value assumed a specific optimal value from the set $\left\{1, \frac{n_{\text{neg}}}{n_{\text{pos}}}\right\}$ considered for the electron identi-

[4]A proof of this heuristic conjecture might be the subject of future work.

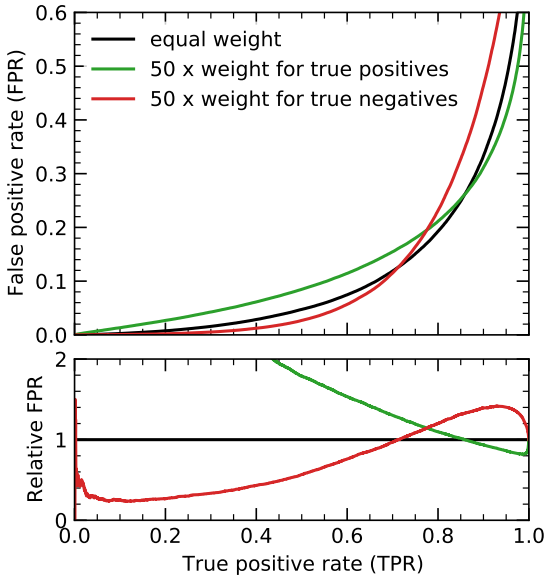


Figure 4.5: Effect of class reweighting on the ROC curve obtained from logistic regression using a randomly generated toy dataset.

fication. The electron identification aims for a high background rejection at loose working points, for which a high signal weight should be favorable. Indeed, the maximum of the two considered values, which is different for the low- and high-p_{T} training bins, maximized the background rejection at 90 % efficiency.

On a more general note, we can conclude that class scaling is an interesting hyperparameter for classification models with a logistic loss function if a specific working point is targeted. In particular, it has the potential to improve significantly the capabilities of the widely used logistic regression. For more complex models, the effect might be less dramatic, as it is counterweighted by reduced effective training sample size, and the model might fit the data so well that it is already optimal for all working points. Therefore the efficacy of artificially scaling the class weights should be studied on a case-by-case basis.

Chapter 5

Multivariate electron identification

Electrons and muons have a long enough lifetime to interact with the detector directly, making them easier to measure than tau leptons. Hence, the term *lepton* in an experimental context often only refers to electrons and muons, including their antiparticles. Most electrons start interacting in the inner tracker and might leave more than one single-particle cluster in the ECAL, or they might even produce more than one track if the emitted bremsstrahlung converts to an electron-antielectron pair.

For multilepton analyses such as the fully leptonic WWZ analysis with four leptons in the final state, the lepton selection efficiency enters the event selection efficiency at the fourth power. Therefore, a highly optimized electron identification is important to maintain a good signal efficiency at a low-enough fake rate. For this reason, the optimization of the multivariate electron identification based on Boosted Decision Trees (BDTs) – which will be referred to as *electron Multivariate Analysis (MVA)* – was an essential part of this thesis.

The LHC Run 2 had already completed one data-taking year in 2016 before this thesis started. Hence, this work focuses on the adaptation for 2017 data and simulation. Section 5.1 explains the general strategy and requirements for electron MVA training. Section 5.2 shows the dominant effects of the new pixel detector on electron track observables, motivating the retraining for 2017 data-taking. This reoptimization is described in Section 5.3, which elaborates on the newly added input variables and other changes. After this initial electron MVA update, new detector calibrations required a retuning, which has been taken as a window of opportunity to move the training from the previous framework to XGBoost [88], a highly optimized gradient boosting algorithm whose speed allowed for optimization with a more extensive training sample. Section 5.4 discusses the XGBoost training, giving a detailed performance overview of the resulting MVA that became the standard within CMS for the full Run 2 dataset. Section 5.5 comments on the software implementation of the BDT score evaluation, which is of general interest beyond the electron MVA. Finally, Section 5.6 presents a study on hyperparameter tuning with Bayesian optimization that pushes the XGBoost training to its limit. Section 5.7 concludes on the electron MVA studies and discusses how the multivariate electron identification could evolve for Run 3.

5.1 General training strategy and requirements

The CMS experiment identifies electrons with BDTs since the LHC Run 1 [85], and this strategy was steadily improved. True prompt electrons are separated from the background, which can be either nonprompt electrons (i.e., electrons in jets or from converted photon) or reconstructed electrons entirely unrelated to any real electron. Electrons from tau leptons decays are not considered.

In Run 1, the BDTs were trained with electron candidates from real Z+jets data. The electrons from the Z boson candidate were taken as the signal. The additional electrons from jets or pileup in the same Z+jets events represent the background. Kinematic cuts were applied on the Z candidate to suppress contributions from diboson processes, where the additional electron would have been a real one. Furthermore, a dedicated MVA was trained for electrons that pass the trigger thresholds to ensure optimal selection in the remaining phase space. For the 2016 Run 2 training, it was concluded that using real measurements for the training has no exact benefit, and electron candidates from Z+jets events simulated with `MadGraph5` [104] at leading order were used instead. The impact of having a dedicated training for triggering electrons was also found not significant, so it was discontinued.

The set of input variables (or *features* as they are referred to in the machine learning community) group into cluster shape, track, track-cluster matching, and isolation plus pileup variables. The BDTs are trained in six training bins, indicated in Table 5.1.

Label	Description		Selection
EB1 10	regular p_T	inner barrel	$p_T \geq 10$ GeV and $\lvert\eta\rvert < 0.8$
EB2 10		outer barrel	$p_T \geq 10$ GeV and $0.8 \leq \lvert\eta\rvert < 1.479$
EE 10		endcaps	$p_T \geq 10$ GeV and $1.479 \leq \lvert\eta\rvert < 2.5$
EB1 5	low p_T	inner barrel	$5 \leq p_T < 10$ GeV and $\lvert\eta\rvert < 0.8$
EB2 5		outer barrel	$5 \leq p_T < 10$ GeV and $0.8 \leq \lvert\eta\rvert < 1.479$
EE 5		endcaps	$5 \leq p_T < 10$ GeV and $1.479 \leq \lvert\eta\rvert < 2.5$

Table 5.1: The BDT training bins. The separate barrel and endcap models are motivated by the considerably different feature distributions and background populations. The further split inside the barrel is motivated by a jump in the material budget around $\eta = 0.8$ (see Figure 2.10b). The split in p_T is more for historical reasons: most analyses do not require electrons below $p_T = 10$ GeV, and the low p_T bins were added for analyses that require to go to such low energy, for example, Higgs to four leptons ($H \to 4\ell$). The limit of $\lvert\eta\rvert < 2.5$ is imposed by the tracker acceptance.

The BDTs return a continuous score so that each analysis could choose an appropriate working point for its selection in principle. However, each working point requires dedicated scaling factors that correct data/MC disagreements, so the working point determination is done right after the training. The final electron MVA selections are provided together with the corresponding scaling factors to the collaboration. Table 5.2 lists the required working points, which are all used in the triboson analysis covered by the later chapters of this thesis[1].

Label	Description
`wp80`	80 % signal efficiency in DY+Jets MC for electrons with $p_T > 20$ GeV
`wp90`	90 % signal efficiency in DY+Jets MC for electrons with $p_T > 20$ GeV
`wpLoose`	for multilepton analyses, with the signal efficiency matched to the efficiencies of the $H \to 4\ell$ analysis in Run 1 in each of the six training bins

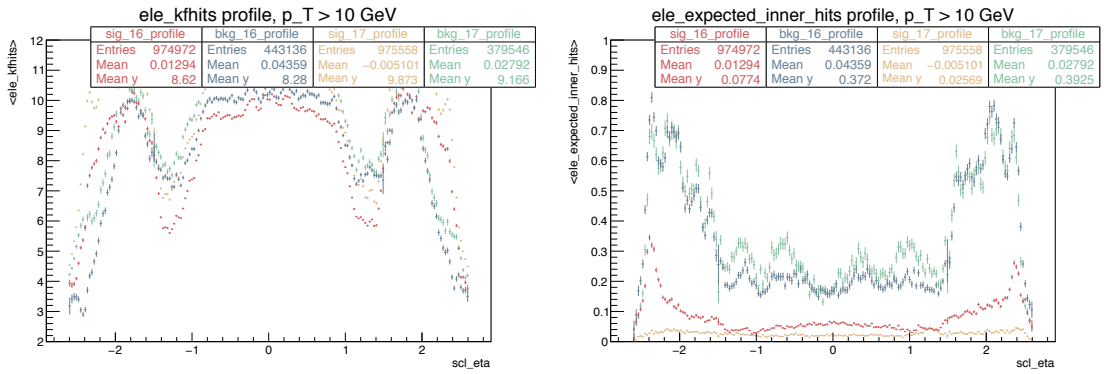
Table 5.2: Working points required for the electron MVA.

[1] Although the tightest 80 % efficient working point was not employed in the WWZ analysis that is explained in this thesis, it was used in the WWW analysis

5.2 Effect of the new pixel detector on electron observables

During the extended LHC technical stop at the end of 2016, the CMS pixel detector has been replaced (see Section 2.2.2.2). The new pixel detector has one more sensor layer in both barrel and endcaps, affecting the number of measured hits in both KF and GSF tracks. Figure 5.1a shows that the average number of hits in the closest KF track as a function of η has increased by one for genuine electrons, as expected. Unlike the number of GSF track hits, the number of KF track hits for fake electrons has not increased, making N_{KF} a more discriminating variable in 2017.

The new pixel detector's other main feature is a reduced material budget in the endcaps (see Figure 2.10a), resulting in reduced bremsstrahlung and photon conversion activity in this region. An observable very sensitive to conversion patterns in the pixel detector is $N_{\text{miss. hits}}$, the number of hits missing at the beginning of the track even though its extrapolation predicts it would have passed through active detector material. This number is low for prompt electrons but higher for nonprompt background electrons from photon conversions. However, it can also be nonzero for prompt electrons if they lost most of their energy by bremsstrahlung early, and it is a conversion leg of the bremsstrahlung photon that gets reconstructed as the electron. The pseudorapidity profile of $N_{\text{miss. hits}}$ shown in Figure 5.1b suggests that such conversions are suppressed in 2017 with the reduced material budget. Hence, $N_{\text{miss. hits}}$ gained a lot of separation power. Naturally, the reduced material budget also affects the conversion vertex fit quality and $\mathrm{f}_{\mathrm{brem}}$, the fraction of momentum lost between the inner and outer track measurements. Shower shape and cluster-track matching variables underwent only minor changes, but the large shifts in the track variables demanded a retuning of the electron MVA for 2017 data.



(a) N_{KF} (b) $N_{\text{miss. hits}}$

Figure 5.1: The electron variables for which the differences between 2016 and 2017 due to the new pixel detector are most visible: the average number of KF track hits (left) and the average number of missing expected inner hits (right) as a function of the supercluster η.

With the additional pixel layer, the seeding of track reconstruction can require up to four matching hits in the pixel detector. Asking for this additional hit decreases the combinatorial background, and therefore, the track reconstruction is much purer. Accordingly, the new detector reduces the number of fake electrons reconstructed per event, illustrated in Figure 5.2 by the number of background electrons per event as a function of the number of pileup interactions. Since the combinatorial background scales superlinearly with pileup activity, the difference is most prominent for high pileup at low p_{T}.

In conclusion, the new pixel detector improves the discrimination power of track observables for electron identification and reduces the number of electron fakes in general.

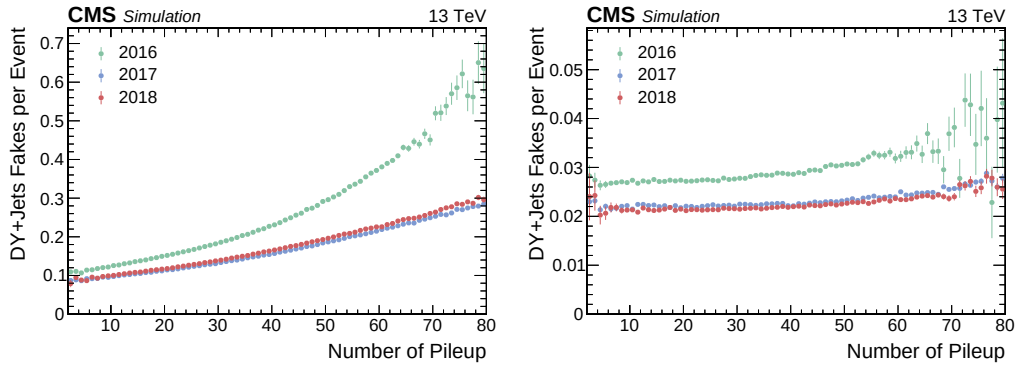


Figure 5.2: Number of reconstructed fake electrons per event as a function of the pileup for the different detector conditions of the Run 2 data-taking years in simulated DY+jets MC events. The fakes with $5 \leq p_{\mathrm{T}} < 20$ GeV are shown on the left, the ones with $p_{\mathrm{T}} > 20$ GeV on the right. No identification is applied.

5.3 First optimization of the electron MVA for 2017 data

The first optimization of the electron MVA for 2017 data focused on the adaptation to the updated detector conditions. The BDT training setup was not changed relative to the 2016 optimization, done with the gradient boosting BDT method implemented in the Toolkit for Multivariate Data Analysis with ROOT (TMVA) [119]. Since the training hyperparameters were also not changed, k-fold cross-validations for hyperparameter optimization was not employed. Instead, the DY+Jet sample was split into one training set and one testing set, where the testing set was used for the performance evaluation that included the figures presented in this section. In each training bin, the testing set nominally includes 50 % of all electrons, but as the training set was limited to 200,000 electrons from each class per bin, the testing set is larger than half of the full sample.

The features used in the electron MVA are presented in Table 5.3. They have been extended twice since Run 1. At the beginning of Run 2, the number of hits in the GSF track and the conversion rejection variables $N_{\text{miss. hits}}$ and P_{conv} have been added, resulting in a roughly 50 % reduction in background rate [120]. The pileup estimate ρ and the three particle-flow isolation components were added within the scope of this thesis. Both models with and without the isolation variables were provided since several analyses require separate identification and isolation selections to allow for background extrapolations from sidebands.

With the high average pileup in 2017 and the prospect of a further increase in 2018, selection efficiencies that are robust to changing pileup levels were required. Adding the pileup estimate ρ does not substantially affect the integrated signal and background rates of the electron MVA, but it makes the selection efficiency more uniform as a function of the pileup. This effect is visible in the endcaps and shown in Figure 5.3.

The three particle-flow isolation components are the charged hadron, neutral hadron, and photon transverse momentum sums. Their total sum is usually used as the isolation variable for electron selections (see Section 6.3.1), with the two neutral components corrected for pileup with the area-median method discussed in Section 3.5. Since a global pileup estimate has been added to the input features, the three isolation components can be added individually. Like this, the BDT itself has the opportunity to learn an appropriate pileup correction and make use of the correlations between the identification and isolation variables. Indeed, adding all the isolation components resulted in a slightly better performance than adding only the combined isolation.

Cluster shape	
$\sigma_{i\eta i\eta}$	cluster shape variance in the η direction
$\sigma_{i\varphi i\varphi}$	cluster shape variance in the ϕ direction
$\Delta\eta_{\mathrm{SC}}$	supercluster width in η
$\Delta\varphi_{\mathrm{SC}}$	supercluster width in φ
H/E	ratio of HCAL and ECAL energy in 5×5 footprint around the seed
$1-(E_{5\times1}/E_{5\times5})$	circularity variable with a 5×1 window in $i\eta i\varphi$ around seed
$R_9 = E_{3\times3}/E_{\mathrm{SC}}$	$E_{3\times3}$ is the energy in the 3×3 matrix around the seed
$E_{\mathrm{PS}}/E_{\mathrm{raw}}$	contribution of preshower to uncorrected SC energy (endcaps only)
Track information	
$\mathrm{f}_{\mathrm{brem}}$	fraction of momentum lost between inner and outer track
N_{KF}	numer of hits in nearest Kalman Filter track within $\Delta R = 0.3$
N_{GSF}	number of hits of the GSF track
χ^2_{KF}	reduced fit χ^2 of nearest Kalman Filter track within $\Delta R = 0.3$
χ^2_{GSF}	reduced χ^2 of the GSF fit
$N_{\mathrm{miss.\ hits}}$	number of expected but missing inner hits
$P_{\mathrm{conv.}}$	conversion probability obtained from the χ^2 of the conversion fit
Track-cluster matching	
$E_{\mathrm{SC}}/p_{\mathrm{in}}$	ratio of SC energy and inner track momentum
$E_{\mathrm{ele}}/p_{\mathrm{out}}$	comparing outer track momentum and energy of closest PF cluster
$1/E_{\mathrm{tot}} - 1/p_{\mathrm{in}}$	energy-momentum agreement
$\Delta\eta_{\mathrm{in}} = \lvert\eta_{\mathrm{SC}} - \eta_{\mathrm{in}}\rvert$	distance in η between energy-weighted SC center and inner track extrapolation
$\Delta\varphi_{\mathrm{in}} = \lvert\varphi_{\mathrm{SC}} - \varphi_{\mathrm{in}}\rvert$	same distance in φ
$\Delta\eta_{\mathrm{seed}} = \lvert\eta_{\mathrm{seed}} - \eta_{\mathrm{out}}\rvert$	distance between seed cluster and outer track extrapolation to calorimeter
Isolation (optional)	
$\sum_{\mathrm{chr.},\Delta R=0.3} p_{\mathrm{T}}$	p_{T} sum of charged PF hadrons from primary vertex within cone
$\sum_{\mathrm{neu.},\Delta R=0.3} p_{\mathrm{T}}$	p_{T} sum of neutral PF hadrons within cone
$\sum_{\mathrm{pho.},\Delta R=0.3} p_{\mathrm{T}}$	p_{T} sum of PF photons within cone
Pileup	
ρ	global pileup estimation variable computed with the fastjet package [121]

Table 5.3: Electron observables that are used as inputs to the multivariate identification.

The full picture of how the combination with the isolation affects the selection performance is presented in the ROC curves in Figure 5.4 for all six training bins. These figures also indicate the fake rates for the MVA trained for 2016 data in 2016 DY+jets simulations as a performance reference. As the number of fakes in 2017 is already reduced before the identification step because of the improved reconstruction with the new pixel detector, Figure 5.4 shows the number of fakes per simulated event instead of the false positive rate of the classification for meaningful comparison.

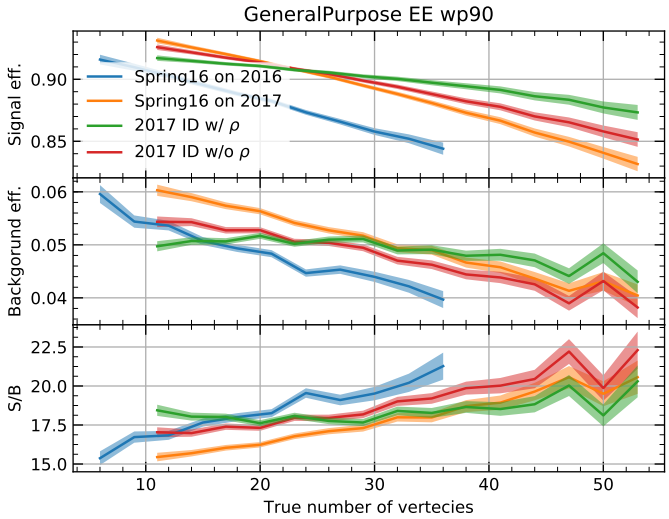


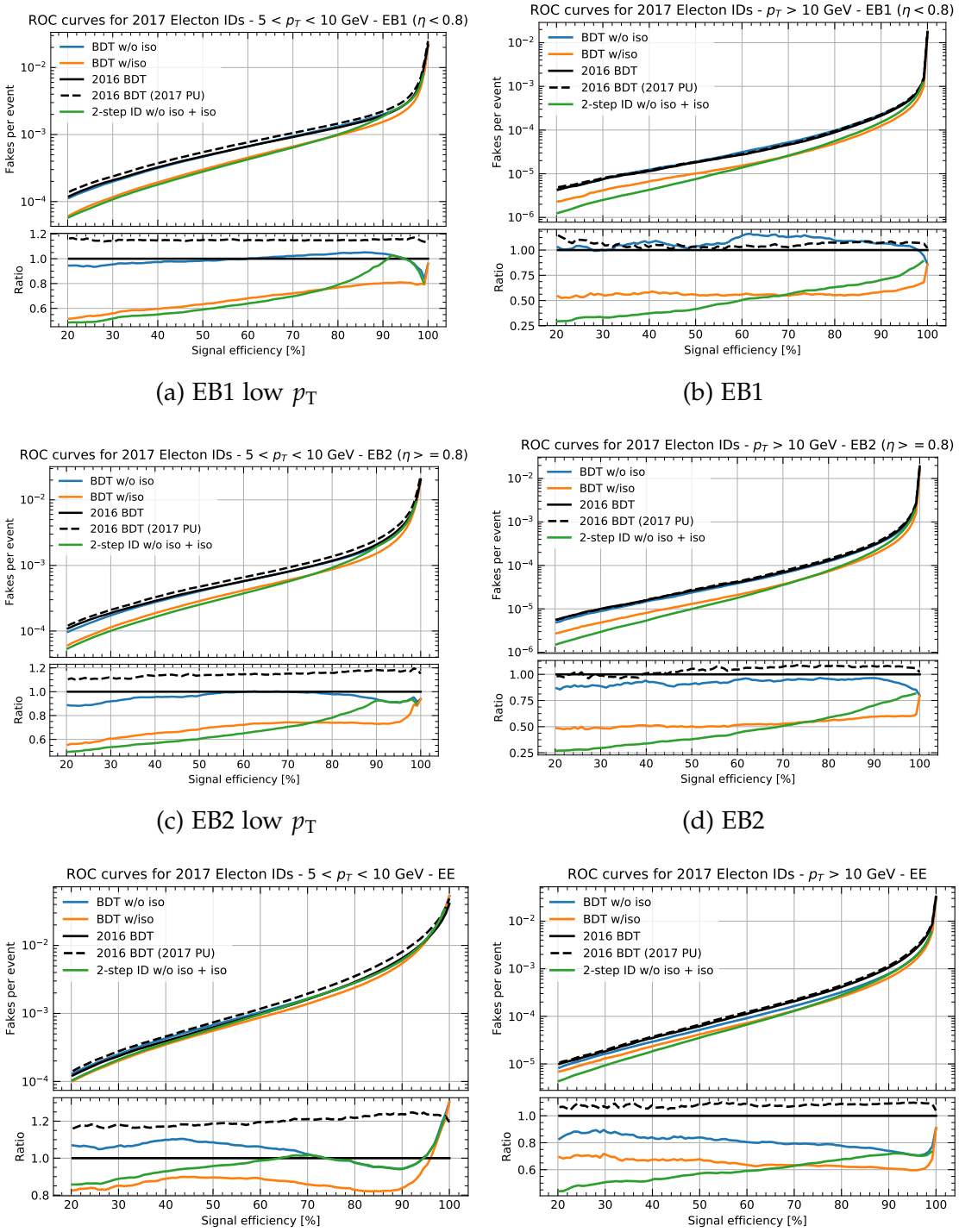
Figure 5.3: Effect of adding the pileup estimate ρ to the input variables of the electron MVA for endcap electron candidates with $p_T > 10$ GeV and within the tracker acceptance. The performance of the existing training for 2016 data is validated on DY+Jets samples simulated with 2016 (blue) and 2017 (orange) detector conditions. The efficiencies of the retrained 2017 MVA on 2017 simulation are shown for the original feature set (red) and with the added pileup estimate ρ (green).

The ROC curves for the 2016 selection are also shown with the pileup distribution reweighted to the 2017 distribution, indicating how the electron selection would likely perform without the upgraded detector. This metric is also useful to understand why the selected fakes per event have increased for high efficiencies in the low p_T endcap bin: the benefits of the new pixel detector do not entirely compensate for the effect of pileup increase.

For the new 2017 MVAs, Figure 5.4 shows the ROC curves for both the models with and without the isolation variables. To understand how the MVA with isolations compares to the traditional 2-step sequential approach – with the isolation cut applied after the MVA cut –, ROC curves for the sequential selection are shown. For a fair comparison, the sequential selection must be optimized for each point on the ROC curve. A 2D scan over both cuts results in a two dimensional manifold of points in the ROC space. The ROC curve of the optimal sequential combination is the lower bound of this population.

The combination of identification and isolation and observables in the MVA (inclusive approach) consistently improves over the sequential approach for selection efficiencies greater than 80 %, which is the relevant interval for electron selection. For high selection efficiencies in the $p_T > 10$ GeV bins, the inclusive approach results in a 15 % fake rate reduction for very high efficiency. The improvement is up to 20 % for the low p_T training bins, with the most considerable effect in the endcaps. Therefore, it was decided to use the combined training for multilepton analyses like Higgs to four leptons and to provide both models as the recommended electron MVA selections for 2017 data.

The last change made for this first retuning relates to the kinematic reweighting of the training electron candidates. In 2016, the loose working point (see Table 5.2) was realized with a model in which no per-electron weights were applied. However, the classifier for the two other working points was trained with the signal distribution reweighted to the background distribution in p_T and η, causing a smoother p_T turn-on for a fixed working point and allowing for the inclusion of p_T and η among the input variables, without introducing a kinematic bias. Since any sustainable improvement should go along with



(a) EB1 low p_T

(b) EB1

(c) EB2 low p_T

(d) EB2

(e) EE low p_T

(f) EE

Figure 5.4: ROC curves with fakes per event instead of false positive rate for the initial retuning of the electron MVA for 2017 data. The black line corresponds to the 2016 MVA applied on simulated 2016 DY+jets samples and serves as a reference for the bottom panel comparisons. The dashed black line shows almost the same, except for the sample being reweighted to match the pileup distribution of the 2017 runs. The performance of the retuned electron MVA on the 2017 sample is shown for models with (orange) and without (blue) particle-flow isolation components. The green line indicates the best possible background rejection that can be achieved by cutting first on the MVA and then on the combined isolation.

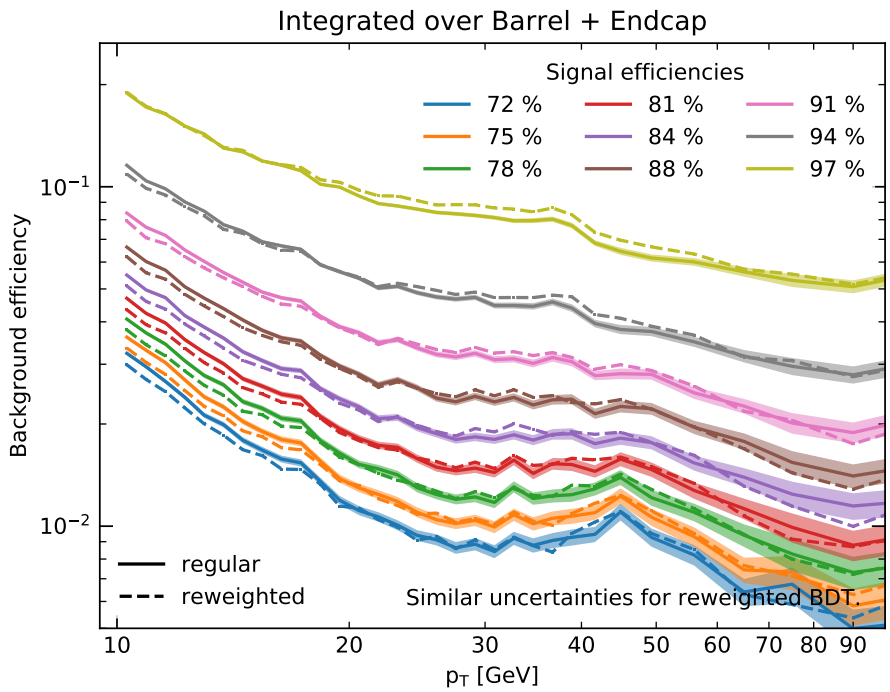


Figure 5.5: Effect of kinematic reweighting in p_T and η and including said observables in the training on the p_T-differential fake rate for uniform signal efficiency.

simplification, only a model without kinematic reweighting was trained for 2017, and the desired turn-on behavior for the 80 % and 90 % efficient working points has been achieved by working points that have an exponential rise in p_T.

The choice to leave out the kinematic reweighting was made after a careful study of the effect on differential performance, presented in Figure 5.5. It shows the background efficiency at several signal efficiency contours both for a model trained without kinematic reweighting and a model trained with kinematic reweighting and p_T and η among the input variables. As expected, the absence of reweighting results in a slightly better fake rejection around the Z-peak, while reweighting the signal to the exponentially falling p_T distribution of the background specializes more in the low p_T region. These differences are minor. Therefore, it was concluded that one single training is sufficient. Similar effects differential in η were found to be negligible, as the η distributions of signal and background are much more similar than the p_T distributions.

More performance plots and details on the exponential working point fits are not shown for this first optimization of the electron MVA for 2017 because it was not the final retuning. Updated detector calibrations required another round of optimization, which is explained in the next section that concludes with a thorough final performance overview.

5.4 Reoptimization of the electron MVA with XGBoost

After the 2017 data-taking and the initial event reconstruction, various detector calibrations have been improved. This included updated ECAL calibration constants and retuned pixel matching cuts[2]. These changes affected the electron identification observables. Hence an update of the electron MVA was required. This opportunity was taken

[2]see Section 3.2.1 for an explanation of electron seeding.

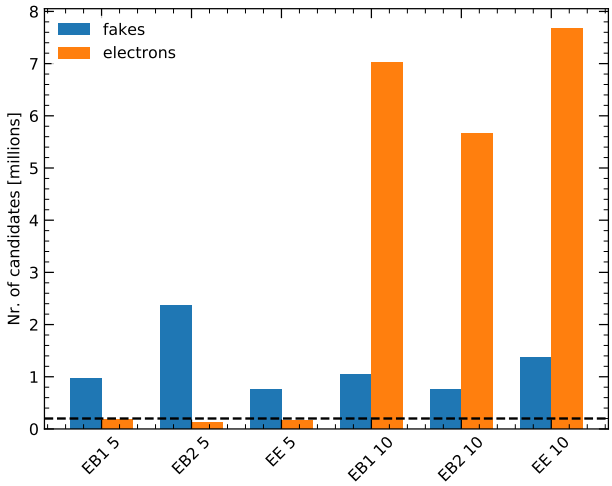


Figure 5.6: The total number of electron candidates from simulated DY+jets events available for the electron MVA training in each of the six training bins. The numbers correspond to only half of the full sample since the other half was designated for the performance tests.

to explore the XGBoost [88] algorithm as the BDT training engine to replace the previously used gradient boosting algorithm implemented in TMVA. This section explains the motivation for this decision, elaborates on the choice of hyperparameters, and concludes with several performance metrics of this final 2017 electron MVA.

In the TMVA based training framework used for all previous electron MVA optimizations, the number of electrons per signal or background class was limited to 200,000 in each training bin to keep the training time under control. However, the DY+jets simulated samples provide much more electron candidates even when only using half of it for the training and the rest for testing. The numbers of available electron candidates are visualized in Figure 5.6. Except for true electrons in the low p_{T} bins, the number of electron candidates greatly exceeds 200,000. This raises questions on how the electron MVA would improve with more training data. This question is difficult to answer with the TMVA training, as the much larger sample size asks for a reoptimization of model hyperparameters, which was not feasible with the very long TMVA training times.

The XGBoost algorithm implements several performance optimizations, for example, an approximation of the best splitting values with quantile-based histograms. Training in the high p_{T} bins takes only a few hours on 16 CPU threads for the full training data, making a grid search over a few hyperparameters possible. The optimization done separately for each training bin started with a coarse grid over `max_depth` and `min_child_weight`, before manually choosing finer grids around the most promising regions. This is illustrated in Figure 5.7 for the EB1 training bin. The number of boosting rounds was determined with a 10-rounds early stopping criterion on the test AUC with an upper bound of 1000 rounds. The final parameters were chosen such that the training performs better than a reference TMVA training for all signal efficiencies while causing only very little overtraining. After this first 2D grid search, `max_depth` has been fixed and a second 2D search was performed over `gamma` and the previously identified promising interval for `min_child_weight`. The whole procedure was done for both unweighted training samples and for a per-class reweighting with the `scale_pos_weight` parameter to balance the a priori very imbalanced dataset.

For this reoptimization, the new DY+Jets sample – which was twice as large as the sample with the original calibrations used in Section 5.3 – was split into training, val-

idation, and testing subsamples. The training and hyperparameter validation samples each amount to 25 %, with the remaining 50 % making up the testing sample. The large testing sample enabled detailed differential performance studies differential in p_{T}, η, or the number of reconstructed vertices shown at the end of this section.

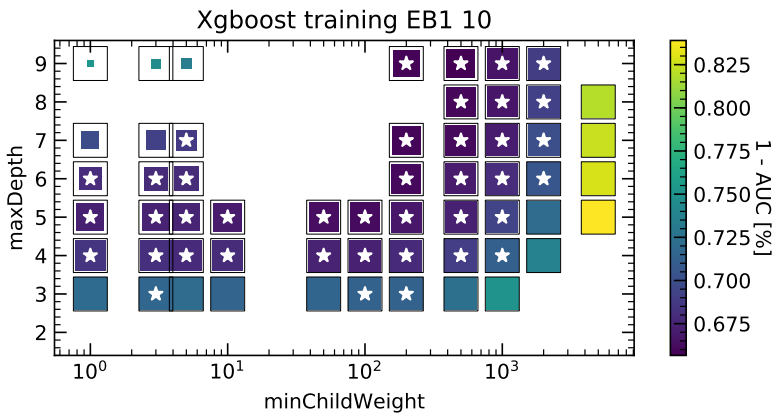


Figure 5.7: A visualization of the manually guided grid search hyperparameter optimization over the `min_child_weight` and `max_depth` parameter in the EB1 training bin. The color encodes the test AUC, the fill area the ratio between test and train AUC, and the star distinguishes the models which performed better than the reference TMVA training for all signal efficiencies.

As shown in the ROC curves in Figure 5.8 and expected from the number of electron candidates, it was not difficult to find parameters that perform better than the TMVA training for the high p_{T} training bins, where the number of both signal and background electrons exceeded 200,000. The impact is a roughly 10 % reduction in fake rate, with no increase in relative overtraining.

For the low p_{T} training bins, the XGBoost models are comparable with the TMVA models. This suggests that the performance is limited by the number of samples in the less populated class since it was below the limit for the TMVA trainings in these bins.

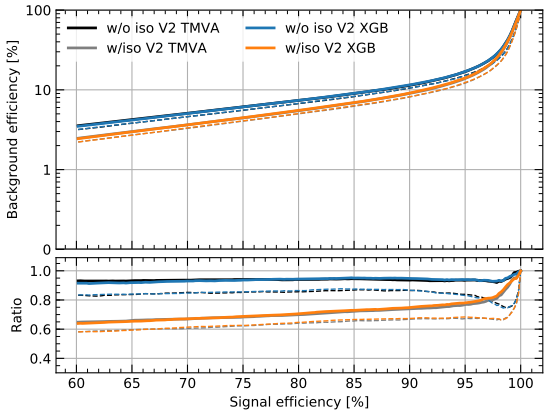
The best found hyperparameters are listed in Table 5.4. One unexpected finding was the role of `scale_pos_weight`. The usual recommendation for unbalanced datasets is to balance their weights by scaling the positive class's weight with $n_{\mathrm{neg}}/n_{\mathrm{pos}}$, so it was unexpected that the unweighted training dataset results in a better background rejection in the signal efficiency interval of interest for the electron MVA. Section 4.3 discusses sample reweighting for the training of classification models and proposes a possible explanation on why not reweighting might indeed be the better choice for models that are intended to be used at a relatively loose working point.

Section 5.3 mentioned that the 80 % and 90 % efficient working points were scaling with an exponential turn-on in electron p_{T} to achieve a more uniform selection efficiency:

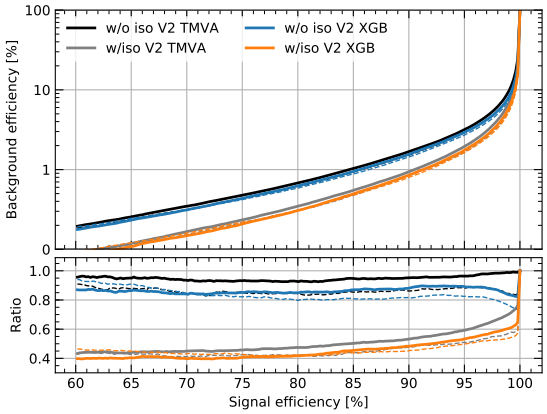
$$\mathrm{wp}(p_{\mathrm{T}}) = C - A \times \exp\left(p_{\mathrm{T}}/B\right). \tag{5.1}$$

Figure 5.10 shows these fits, demonstrating that the exponential function is well suited to yield differential selection efficiencies that match the uniform working points of the 2016 electron MVA, trained with kinematic reweighting.

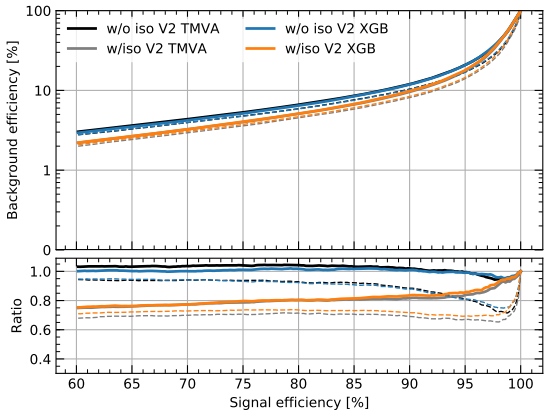
Finally, the Figures 5.11, 5.11, and 5.12 show the signal and background efficiencies for all working points.



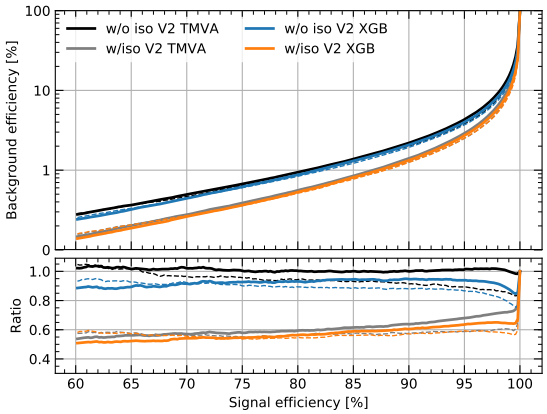
(a) EB1 low p_{T}



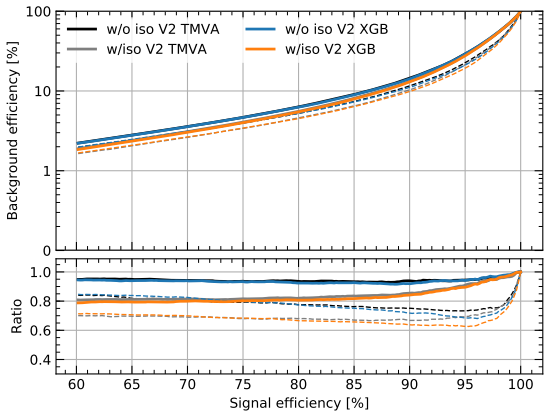
(b) EB1



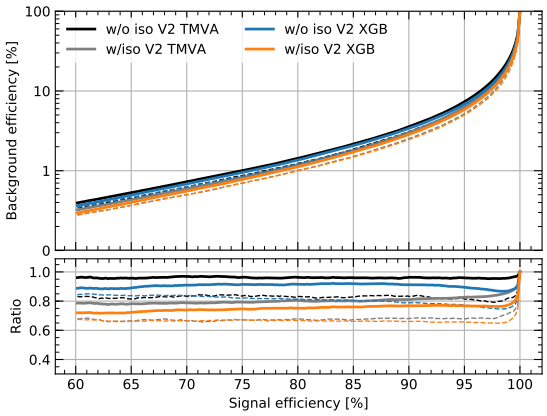
(c) EB2 low p_{T}



(d) EB2



(e) EE low p_{T}

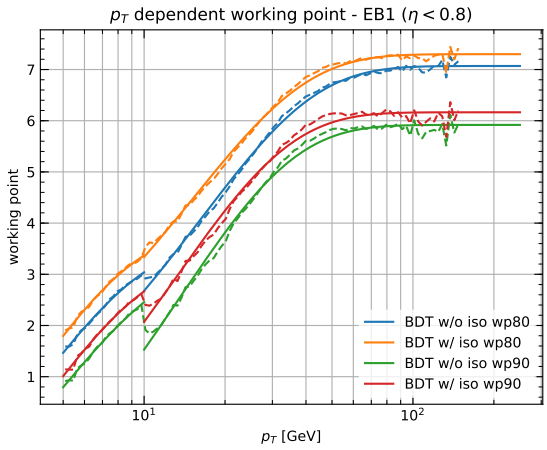


(f) EE

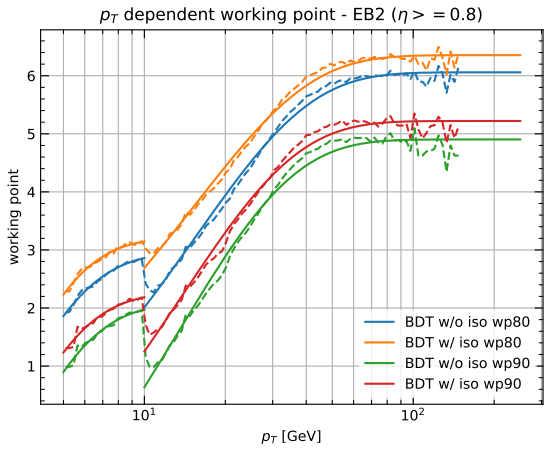
Figure 5.8: The ROC curves for the electron identification.

	max_depth	min_child_weight	gamma	scale_pos_weight	n_estimators (NoIso)	n_estimators (Iso)	Size (NoIso)	Size (Iso)	TMVA size
EB1 5	4	500	10	$n_{\mathrm{bkg}}/n_{\mathrm{sig}}$	259	269	47 KB	48 KB	1.1 MB
EB2 5	4	2000	15	$n_{\mathrm{bkg}}/n_{\mathrm{sig}}$	310	288	44 KB	42 KB	1.1 MB
EE 5	4	2000	15	$n_{\mathrm{bkg}}/n_{\mathrm{sig}}$	446	486	76 KB	84 KB	1.1 MB
EB1 10	4	500	0	1	1000	1000	165 KB	162 KB	1.1 MB
EB2 10	4	500	0	1	1000	1000	164 KB	163 KB	1.1 MB
EE 10	4	500	0	1	1000	1000	212 KB	210 KB	1.1 MB
Sum							708 KB	709 KB	6.6 MB

Table 5.4: The hyperparameters found with the manually guided grid search.

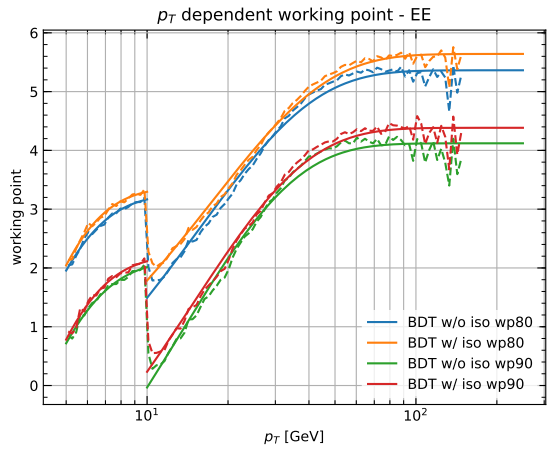


(a) EB1

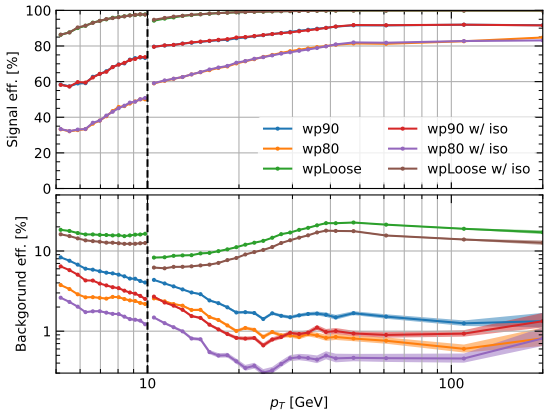


(b) EB2

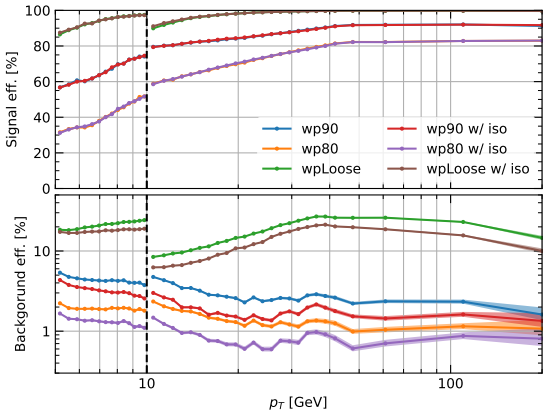
Figure 5.10: Illustration of how the 80 % and 90 % efficient working points were derived. The dashed line indicates the required working points to match the signal efficiency of the 2016 MVA for every p_{T} value. This is then fitted with Equation 5.1 for the low p_{T} and high p_{T} bins separately to obtain the working point function. Note that the isolation-inclusive training did not exist in 2016, so the efficiency target for both 2017 selections was the same.



(c) EE

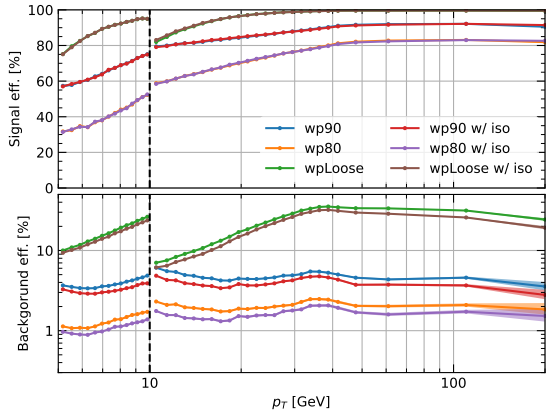


(a) EB1



(b) EB2

Figure 5.11: The signal and background selection efficiencies of the retuned electron MVA for both models and all working points in simulated DY+jets data as a function of p_T.



(c) EE

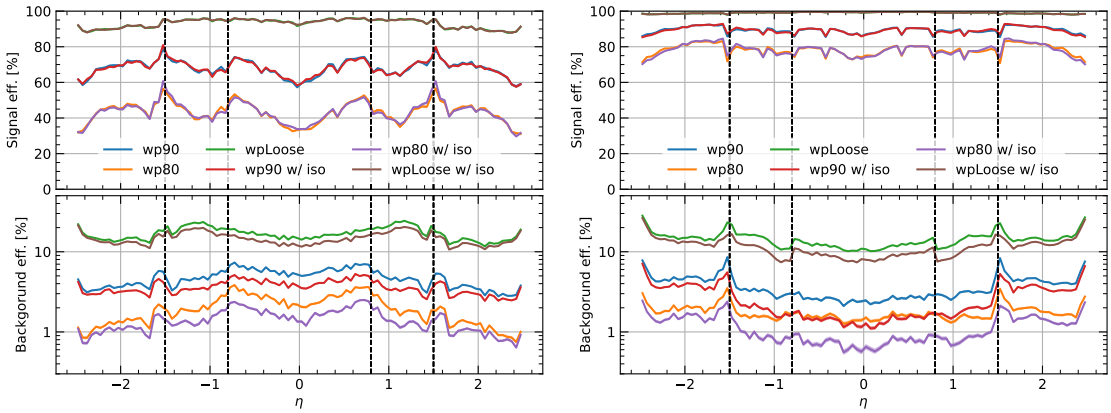


Figure 5.11: The signal and background selection efficiencies of the retuned electron MVA for both models and all working points in simulated DY+jets data as a function of η.

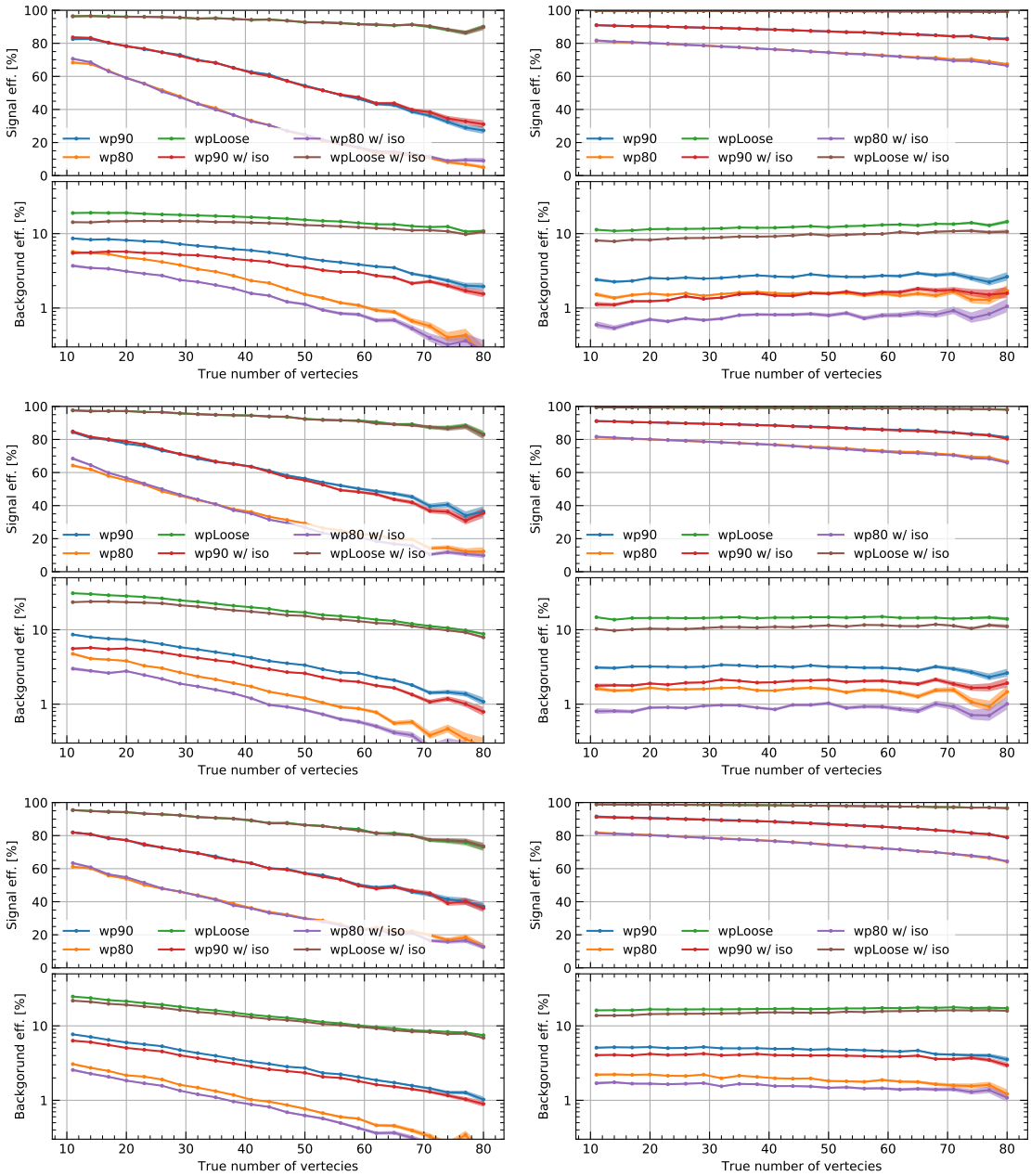


Figure 5.12: The signal and background selection efficiencies of the retuned electron MVA for both models and all working points in simulated DY+jets data as a function of the number of true collision vertices.

5.5 Implementation in the CMS reconstruction software

The implementation of the electron MVA for Run 2 in the reconstruction framework was rewritten in the scope of this work. Previously, each MVA training had been hardcoded in a separate C++ class, making it difficult to understand their differences and to ensure a consistent definition of inputs. The new implementation uses the reflection capabilities that the ROOT framework adds to C++ classes. This makes it possible to define how input variables are obtained from the electron object in a single place for all MVAs, used by a single implementation to look up how to get all variables required by a given MVA.

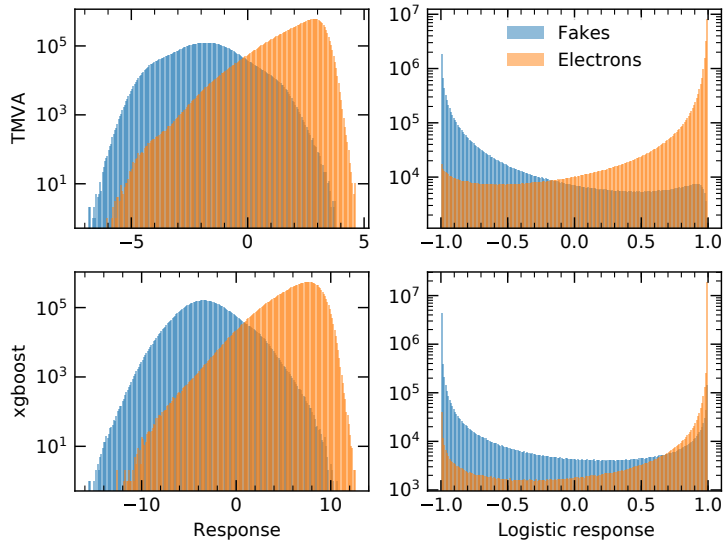


Figure 5.13: The response of the BDTs trained for the electron identification with TMVA (top) and XGBoost (bottom) to signal and background. The left plots show the direct BDT response, i.e., the sum of responses from each decision tree, and the right plots show the response after the logistic transformation $1/\left(1+\exp(-x)\right)$, linearly mapped to the interval between -1 and 1. Note that the values are usually extremely close to ± 1, which is undesirable for fitting and for encoding in a floating-point format.

The MVA implementation stores the final BDT score in a single-precision floating-point number, while the XGBoost library uses double-precision. Since this score was subjected to a logistic transformation and the electron classification is generally very discriminating, most values are very close to ± 1. This causes precision issues that sometimes lead to the wrong decision for extreme working points. Since the numeric value of the final score – which can be interpreted as a probability – is not essential for the electron MVA, the logistic transformation is now skipped, which makes the score distribution much smoother, as shown in Figure 5.13. Accordingly, all working points were determined for the score before the logistic transformation. A move to double-precision would have been prohibited by the restrictive space constraints of CMS event data files.

The CMS software implementation of decision tree inference is highly optimized. The full intermediate and terminal node information about individual decision trees is flattened into a structure of arrays, ensuring a high data locality that maximizes caching efficiency. Since the number of trees is usually much higher than the number of nodes in one tree, it has been expected that the cache efficiency can be further improved by storing the whole forest in contiguous memory. This resulted in another 20 % improvement in single-thread inference speed. This further optimized BDT inference engine was measured to be four times faster than the official XGBoost implementation – written in C – and twice as fast as other popular open-source inference engines. The CMS software implementation, including the contiguous memory optimization, has been made available in an easy-to-use open-source package that got attention from academic and non-academic data scientists and analysts [122].

5.6 Hyperparameter tuning with Bayesian optimization

Section 5.4 explained how an improved electron MVA was trained with the XGBoost algorithm. This included a manual optimization of hyperparameters to improve rela-

tive to the TMVA training in each of the six training bins. A manual hyperparameter optimization was seen as more appropriate than a fully automatic grid search or random search, as one training takes several hours, and it would be wasteful to train at a point in hyperparameter space that can be predicted to be inferior based on the results from other trials. This section presents a more sustainable automatic solution based on Bayesian optimization.

5.6.1 Bayesian optimization with Gaussian processes

Most optimization algorithms use the gradient of the function they aim to optimize, either by computing it directly or estimating it with small variations of the function's arguments. When the function is very costly to evaluate, which applies to an expensive-to-train classifier's performance metric, such an estimate of the gradient is not feasible. Instead, one needs an optimization algorithm that can work without the gradient and that generally minimizes the number of function evaluations.

Bayesian optimization fulfills these requirements by approximating the function with a probabilistic model [123]. This *surrogate model* is updated with each function evaluation and an *acquisition function* is determines which point is evaluated next. When choosing the surrogate model and the acquisition function, one must consider the trade-off between exploration and exploitation. In other words, the algorithm must balance between evaluating points in unexplored regions of the search space to avoid getting stuck in a local optimum and points near the best among the points explored so far to find the true optimum more precisely.

A common choice for the surrogate model is a Gaussian process [123, 124]. It has been chosen for this work because it has a more formal mathematical basis compared to other established approaches like the tree-structured Parzen estimator [124]. The general familiarity of researchers with Gaussian distribution makes it also very intuitive.

A Gaussian Process (GP) generalizes multivariate normal distributions to infinite dimensions. In practical terms, this means that it is a field of stochastic processes Φ that describes a bivariate normal distribution for any pair of points:

$$\begin{pmatrix} \Phi(x) \\ \Phi(y) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu(x) \\ \mu(y) \end{pmatrix}, \Sigma(x,y) \right). \tag{5.2}$$

The mean μ and covariance Σ are neither scalar values nor a vector and a matrix, but functions. For this reason, Gaussian processes are sometimes referred to as *distributions over functions*. If we sample the process at the point y, the conditional processes at every other point x make up another Gaussian process since the conditional probability of a multivariate normal distribution is also a normal distribution[3]:

$$\Phi(x) \mid \Phi(y) \sim \mathcal{N}\left(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(\phi(y) - \mu_y), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy} \right). \tag{5.3}$$

In the context of Bayesian optimization, drawing the value $\phi(y)$ means to evaluate the function we want to minimize. Computing the parameters of the conditional probability distributions corresponds to the surrogate model updating step. We also need to define an initial Gaussian process as a prior, where a simple choice is a uniform mean at zero and an exponential kernel with parameters θ_1 and θ_2 as the covariance function:

$$\begin{aligned} \mu(x) &= 0, \\ \Sigma(x,y) &= \theta_1 \exp\left(-\frac{\theta_2}{2}(x-y)^2 \right). \end{aligned} \tag{5.4}$$

[3]Function arguments were moved to subscripts for a more compact notation.

In the following study, the covariance function is the Matérn kernel [125]:

$$\Sigma_{\text{Matern}}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} r}{l} \right)^{\nu} K_{\nu} \left(\frac{\sqrt{2\nu} r}{l} \right), \tag{5.5}$$

where r is the distance between the two points, and K_ν is a modified Bessel function. The parameters were set to $l = 1$ and $\nu = 2.5$. The Matérn kernel is a standard choice for machine learning applications and is employed in the code used for the Bayesian optimization in this work, aptly named `BayesianOptimization` [126]. Note that there are also other similar open-source packages like `hyperopt` [127], which uses a tree-based surrogate model and supports optimization on distributed systems.

Some possible acquisition functions that can be used together with Gaussian processes are the *upper confidence bound*, the *expected improvement*, and sometimes the *probability of improvement*. The upper confidence bound is a profile of the Gaussian process at a given quantile, and the next point to evaluate will be at the maximum of this confidence bound. A common choice is the 2σ confidence bound, but in general, it should be adjusted for a good trade-off between exploration versus exploitation, with quantiles further in the tails favoring exploration. The expected improvement is the expected value of the improvement at a given point x under the condition that the value $\phi(x)$ assumed by the Gaussian process will be above the current best value ϕ^*:

$$\text{EI}_{\phi^*}(x) = \int_{\phi^*}^{\infty} (\phi(x) - \phi^*)\ p(\phi(x)|\phi(y_1), ..., \phi(y_n))\ d\phi(x). \tag{5.6}$$

Finally, the probability of improvement is similar to Equation 5.6, but it is only the integral over the probability. Here, the expected improvement acquisition function has been chosen because it requires no additional parameter and is intuitive to understand.

5.6.2 Training and optimization setup

This study's goal was to demonstrate how the XGBoost performance for the electron MVA would improve if the Bayesian optimization optimizes several hyperparameters. The hyperparameters in the search space are listed in Table 5.5. To break the overdetermination of the first point, a few initial points have to be chosen randomly. Here, we draw four points from a uniform distribution in the allowed parameter space. The same four points were used for each training bin to suppress variations in the optimization path due to random choice. Additionally, the default parameter values, as given by XGBoost, were evaluated, as these are tuned to perform generally well. Afterwards, the Bayesian optimization is trying to maximize the test AUC for 50 iterations.

During the training, the performance is evaluated with three-fold cross-validation to use the sample's full statistics for the test metric (see Section 4.1.1). Since the best hyperparameter values adapt to a specific training sample size, the final model trained with the best-found hyperparameters also leaves out one-third of the training data.

The number of boosting rounds (`n_estimators` in XGBoost) takes a unique role among the hyperparameters as it can be determined dynamically during the training. If the test metric fails to increase for 100 rounds, the optimal number of boosting iterations is determined to be the round before the improvement halts. If this does not happen, it is limited to 3000. Nevertheless, even with the early stopping, the optimization may choose points for which the training takes a disproportionally long time, even if it becomes apparent that the test AUC does not converge well. Therefore, some mechanisms need to be introduced to abort the training based on the previous round's results.

Another training halt criterion considers that overtraining increases monotonically and that the AUC has an upper bound of one:

$$\mathrm{AUC}_{\mathrm{train}} - \mathrm{AUC}_{\mathrm{test}} > 1 - \mathrm{AUC}^{*}_{\mathrm{test}}. \tag{5.7}$$

$\mathrm{AUC}^{*}_{\mathrm{test}}$ is the best test AUC observed so far. Put in words: if the overtraining – expressed as the difference between the train and test AUC – is larger than the residue of $\mathrm{AUC}^{*}_{\mathrm{test}}$, $\mathrm{AUC}_{\mathrm{test}}$ can not exceed $\mathrm{AUC}^{*}_{\mathrm{test}}$, even if variations in the training sample would eventually be fit entirely. Note that the comparison with past information means that the order of evaluation is making a difference, but since it only affects points that are far from the maximum, this is presumably not harmful.

The overtraining criterion can successfully abort the training of models prone to overtrain, but another criterion needs to be implemented to stop the training of overly regularized models. These points have the problem that they converge very slowly, and the regular early stopping is triggered very late as the absence of overtraining causes the test AUC to keep increasing by tiny amounts instead of reaching a sweet-spot before decreasing again. These alarmingly slow convergence cases can be caught early, considering that the evolution of the AUC is generally concave. A straight line is fit to the test AUC of the last ten rounds and is extrapolated to provide an upper bound to the AUC in the future. If this fit predicts that $\mathrm{AUC}^{*}_{\mathrm{test}}$ will only be reached in an overly long time of at least five times the training time corresponding to the default parameters, the training is stopped.

One final change with respect to the manually optimized training described in Section 5.4 was the clipping of the training data. Since training time scales approximately linearly with the training data size, signal and background have been balanced by throwing away electron candidates such that both classes are balanced (see Figure 5.6 for the number of electron candidates in each training bin). This clipping was considered to be the least harmful way to bring the training time to a level that does not prohibit Bayesian hyperparameter optimization since classifier performance is usually limited by the number of samples in the less abundant class.

This hyperparameter optimization took one week on 16 CPU cores for all six training bins, and the custom early stopping criteria were triggered several times. The study was done only with the isolation variables included because of the long run time and the minor impact of removing a small fraction of features on the optimal hyperparameters.

5.6.3 Results and interpretation

The results of the Bayesian optimization study are presented in three figures. Table 5.5 lists the best hyperparameters found for each training bin. Figure 5.14 visualizes the evolution of the test AUC during the optimization. Finally, Figure 5.15 shows the ROC curves in each of the six training bins, comparing them to various references.

The hyperparameters in Table 5.5 can be sorted in three groups. The randomization parameters `colsample_bytree` and `subsample` strongly lean towards the absence of randomization, particularly for the random sampling of training instances per tree. On the other side, the group of regularization parameter `gamma`, `alpha` and `lambda` strongly favors high values far from the default value and sometimes close to the upper bound.

The third group of parameters – `max_depth`, `min_child_weight` and `n_estimators` – is most directly related to model complexity. They allow for a comparison of the relative model complexity with expectations, raising confidence in the educated choices of the

		max_depth	min_child_weight	gamma	alpha	lambda	colsample_bytree	subsample	n_estimators	Nodes (in thousands)
default		6	1	0	0	1	1	1	-	-
parameter bounds	min	2	1	0	0	0	0.1	0.5	-	-
	max	15	20	10	10	10	1	1	-	-
optimized	EB1 5	4	15.38	3.46	8.82	4.17	0.63	0.97	760	29
	EB2 5	5	10.46	4.56	6.45	7.33	0.99	1.00	561	35
	EE 5	13	11.38	6.86	10.00	8.69	0.95	0.98	378	62
	EB1 10	14	7.18	5.27	9.64	9.59	0.78	0.98	249	200
	EB2 10	14	6.36	9.93	9.15	3.12	0.87	0.97	418	186
	EE 10	11	1.56	2.14	9.79	9.55	0.69	0.94	282	394

Table 5.5: XGBoost hyperparameters subjected to Bayesian optimization with their default values, the bounds for the optimization, and the optimized values for each training bin. The integer parameter `max_depth` is treated as a continuous parameter in the surrogate model and rounded to the nearest integer value for model training. Note that `n_estimators` is not optimized directly by the Bayesian optimization but determined by early stopping to be the boosting round after which the test AUC did not improve for 100 rounds. The final column presents the number of intermediate and terminal nodes in all decision trees to indicate the model's complexity. Results were obtained for the training with particle-flow isolation variables.

Bayesian optimization. Model complexity should scale with the number of training examples. Concretely, this means that the low p_T barrel bins presumingly require the least complex models, followed by the low p_T endcap bin, the high p_T barrel bins, and finally the high p_T endcap bin. Even though the table shows that the model complexity parameters follow this trend, it is most obviously reflected in the number of tree nodes in the models, as indicated in the table.

The evolution of the test AUC in Figure 5.14 shows the exploration versus exploitation behavior of the optimization. There are several plateaus, hinting that the expected improvement leaned towards exploitation in this study.

The ROC curves in Figure 5.15 show how the optimization translates to an improved background rejection, measured in an independent test sample that took no part in the hyperparameter optimization. The optimized models reject up to 10 % more background than the other models in the low p_T bins. The same is true for the high p_T bins, where the XGBoost models generally outperform the TMVA models because of the more extensive training datasets. Hence, the Bayesian optimization approach was a success, also since it can run fully automatically and thus has no high development cost after the initial investment made in this work. From a different perspective, this study also shows that the default XGBoost parameters generally work well for the electron MVA and that other limiting factors should be addressed, like the relatively small sample size for the low p_T bins.

With these ROC curves that integrate over a relatively large kinematic phase space, additional checks are vital for the correct interpretation. Because signal and background

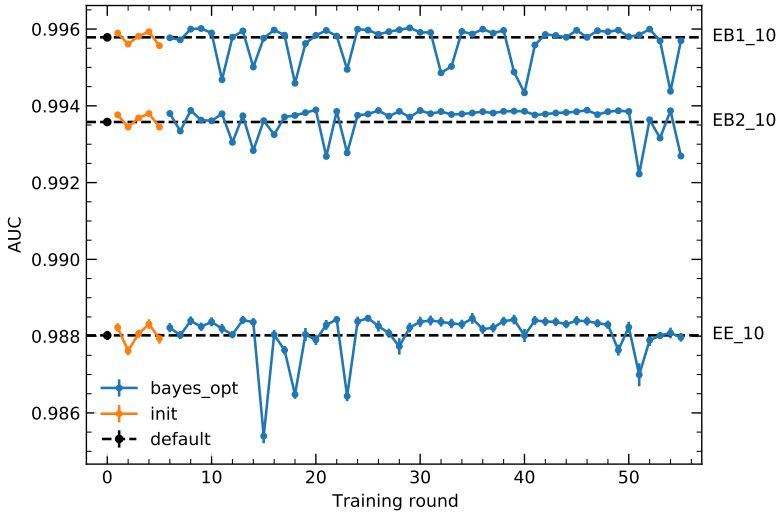
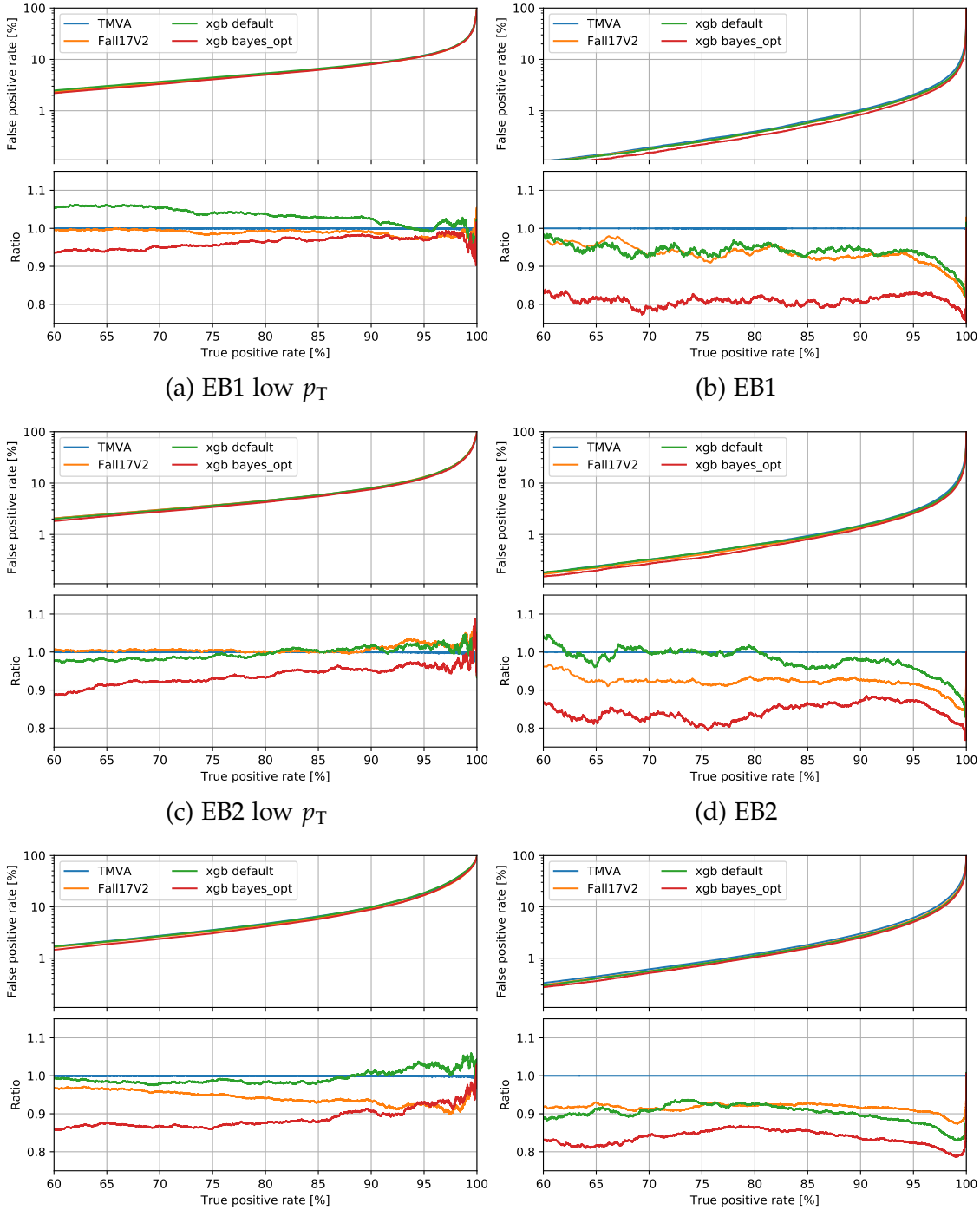


Figure 5.14: Evolution of the test AUC with the training iteration of the optimization for the high p_T training bins. The first round corresponds to the default XGBoost parameters and serves as a reference, followed by four random points and the Bayesian optimization.

are distributed differently in p_T and η, it is mandatory to check the ROC curves also for thin slices of the phase space to make sure the presumed improvement does not stem from a better separation in kinematics, but really from a better separation at a given point in the p_T-η plane. These checks were done, and they confirmed that improved fake rejection is also manifest in differential phase space, supporting the validity of the integrated comparisons.



(a) EB1 low p_T

(b) EB1

(c) EB2 low p_T

(d) EB2

(e) EE low p_T

(f) EE

Figure 5.15: ROC curves for the Bayesian optimized electron MVA (red), the default XGBoost hyperparameters (green), the manually optimized XGBoost training generally used in CMS for Run 2 analysis (yellow) and a model trained with TMVA like in the past as the reference for the ratio plot (blue). Note that the manually optimized training uses all available electron candidates instead of clipping them to balance signal and background classes (for the other XGBoost models) or limiting them to 200,000 per class (TMVA training). All trainings include the isolation variables.

5.7 Conclusion and outlook

The reconstruction of electrons with the CMS detector aims for a maximal signal efficiency, and only computational resource and storage limitations impose constraints on the fake rate. Therefore, a performant identification of real electrons after the offline reconstruction is essential. For multilepton analyses like the triboson measurements, the electron selection is one of the limiting factors, making the multivariate electron identification a crucial part of this thesis.

The multivariate electron identification has been reoptimized for 2017 data-taking, primarily to make the best use of the new pixel detector. The conditions in 2018 were very similar, and the new optimization strategies described in this chapter made the reoptimized MVA very competitive with the 2016 MVA. Therefore, analyses of the complete CMS Run 2 dataset that employ a multivariate electron identification generally use the manually optimized XGBoost training presented in this chapter.

Besides improvements in the training algorithm, the new training added a global pileup estimate to the input variables, making the response more independent of the pileup level. It further provides two different classifiers, where one additionally includes the particle-flow isolation sums. The two different algorithms allow for more flexibility in the analysis. One can choose between the more performant combined MVA or place sequential cuts on the identification and isolation to enable more possible sideband studies.

Finally, this work explored the Bayesian optimization of the XGBoost hyperparameters. The optimization showed the potential to reduce the background rate in simulated samples by another 10 %. Even though the CMS collaboration has not adopted this final optimization of the electron MVA, the training framework served other purposes. Collaborators who optimized the full electron reconstruction and identification pipeline for low-energy electrons – lepton flavor universality measurements with B-meson decays – used it successfully with only minor modifications[4].

Studies with DY+jets MC events simulated with Run 3 conditions revealed that retraining for future data-taking would be necessary, with innovation opportunities in different areas. The input variables should be revisited, possibly using the finer longitudinal readout segmentation of the HCAL (see Section 2.2.2.4). The increased noise in the ECAL and particularly the ECAL endcaps asks for a redefinition of some shower shape variables, namely those with noise cuts in their definition, like $\sigma_{i\eta i\eta}$ and $\sigma_{i\varphi i\varphi}$.

The multivariate electron identification uses many high-level variables from tracker and ECAL information, which is a perfect use case for classification with BDTs. Nevertheless, the image-like nature of the ECAL measurements makes it an appealing use case for convolutional deep neural networks. In general, major innovation might be possible with a more holistic approach to electron reconstruction and identification, keeping in mind that BDT classifiers with similar input variables are also used upstream in the reconstruction. Recently, CMS collaborators started an effort to perform the ECAL clustering with neural networks instead of clustering in η-ϕ-windows. These efforts could be a good starting point for future work on multivariate electron identification because they present the opportunity to connect cluster identification variables with the clustering step itself.

The training with XGBoost plus Bayesian optimization defines an excellent reference for performance evaluation for any future development.

[4]One further improvement was the selection of the initial points with Latin hypercube sampling.

Chapter 6

Measurement of triboson production

The central subject of this thesis was to measure the production of three massive vector bosons (W or Z bosons) in the multilepton final state. Such processes are starting to get accessible by the LHC experiments with the integrated luminosity recorded during the LHC Run 2. The CMS experiment has published a search for the production of WWW events at $\sqrt{s} = 13$ TeV in the past [28], based of the data collected in 2016 (35.9 fb^{-1}). The observed significance for WWW production was 0.60 standard deviations (1.78 standard deviations expected), and limits on anomalous quartic gauge couplings were set. The latest heavy triboson result published by the ATLAS collaboration analyzed data recorded between 2015 and 2017 (79.8 fb^{-1}) and combined a WWW analysis with an analysis that targeted the WVZ final state, where V can be either a W or Z boson [27]. This ATLAS analysis established evidence for the heavy triboson final state with a significance of 4.1 standard deviations, where 3.1 standard deviations were expected.

In this chapter, the WWZ analysis that is part of a complete VVV analysis in the –for the most part – fully leptonic final state with the CMS detector will be presented. The data are the total 137.2 fb^{-1} of collisions recorded with the CMS detector in Run 2. For this thesis, the focus was chosen to be on the WWZ final state because it is the most promising and most versatile among the massive triboson final states, even though its cross-section is lower than the one of WWW production. The fully reconstructed Z boson allows for a restrictive event selection based on invariant lepton-pair masses, strongly suppressing the reducible backgrounds[1]. The WWZ cross-section is larger than the one of the processes with more Z bosons. Unlike the WWW process, it also allows probing anomalous quartic couplings that involve Z bosons. The results presented in this chapter will confirm this sensitivity ranking.

The triboson analyses published by CMS and ATLAS mentioned earlier consider as the signal the general VVV final state, even if two bosons are off-shell from Higgs decays, where the Higgs was produced in association with a third boson (VH $\rightarrow$ VVV). Hence, to be consistent and comparable with previous results, this analysis considers the general WWZ final state as the signal. However, the final analysis flow will also be used to obtain results for on-shell production only.

This chapter first gives an overview of the WWZ analysis steps and briefly outlines the other triboson final state analyses to give the full picture. Afterwards, the basic building blocks, such as object selections, simulated samples, and trigger paths, are explained. Then, the event preselection and the BDT-based event classification are elaborated. This will be followed by an in-depth discussion on signal extraction strategies in general and the histogram-based signal extraction method used in the analysis in particular. An explanation of the background estimation and systematic uncertainties follows. Lastly, the combined fit with the analyses targeting the remaining VVV final states is explained, and the final results are presented.

[1] As the WWZ events are mainly identified by the number of leptons, true four-lepton backgrounds are referred to as *irreducible* and background with non-reconstructed or fake leptons as *reducible*.

6.1 Analysis overview

In the fully leptonic decay mode, the WWZ state decays to four charged leptons: two same-flavor opposite-sign leptons from the Z boson and two additional leptons from the W boson decays. Hereafter, the charged leptons are referred to a simply leptons. In addition, there is missing transverse momentum in the event, denoted $p_{\mathrm{T}}^{\mathrm{miss}}$. It comes from the two neutrinos among the W boson decay products. The most abundant background process is the ZZ diboson production that has a much higher cross-section. Therefore, requiring the W candidate leptons to be of different flavors is vital to reduce the background. The other main background is the $\mathrm{t\bar{t}Z}$ process, which also contains two intermediate W bosons from the top quark decays, but it can be separated with means of b-tagging. Hence, the most sensitive phase space region is the four-lepton final state with one Z candidate pair, two different flavor leptons, and no b jet candidates. This is the primary signal region in this analysis. Two Boosted Decision Trees (BDTs) are used in this region to discriminate against the remaining background, and the signal is extracted in a two-dimensional histogram of these BDTs.

The region with a second Z candidate is overwhelmingly populated by ZZ events. Nonetheless, the WWZ signal can be isolated using the missing transverse momentum primarily. Accordingly, a third BDT separates the signal from the ZZ background in this region, and some signal can be extracted from the tail of the BDT distribution.

After the ZZ and $\mathrm{t\bar{t}Z}$ backgrounds, the analysis is affected by reducible backgrounds that are three-lepton processes with one additional fake lepton in the event, mostly from WZ production. For this reason, the analysis employs relatively tight lepton selection criteria for the two leptons from the W decays, strongly suppressing the fake lepton background, which is – being an instrumental effect – particularly challenging to model.

The four-lepton final state covers in principle also WZZ with a hadronic W boson decay, but the event selection geared towards ZZ suppression leaves almost no trace of this process. The four-lepton analysis is, for this reason, mainly a WWZ analysis.

The other triboson processes are accessed via final states with a different number of leptons in the event, where the number of leptons is counted after requiring them to pass loose identification and isolation criteria. A five-lepton analysis targets WZZ, and a six-lepton analysis targets the ZZZ process. These final states have a high purity but a low signal rate. Hence no discrimination with BDTs is done. The WWW process is measured in the two- and three-lepton final states. In the three-lepton final state, the main background are four-lepton processes involving a Z boson with a non-reconstructed lepton. Hence, the events are categorized by the number of same-flavor opposite-sign (SFOS) lepton pairs, where events with zero SFOS pairs are the most sensitive. The two-lepton analysis accepts events with one or more jets from one hadronic W decay. The one jet category is important because one of the jets from the W boson decay might be too soft to pass the p_{T} thresholds, in particular for $\mathrm{WH} \to \mathrm{WWW}$ events. If there are at least two jets, the events are further categorized by the invariant mass of the jet pair that matches best a W boson candidate: one category for m_{jj} within 15 GeV inside the W mass window and one for the opposite case. Both the two-lepton and three-lepton analyses require high lepton energy thresholds and tight lepton identification and isolation criteria to suppress background from nonprompt leptons in particular.

6.2 Analysis samples

6.2.1 Trigger selection

For the signal regions in the WWZ analysis, double-electron triggers, double-muon triggers, and a cross trigger that fires on one electron and one muon are used. For the double-muon trigger, one muon must have a transverse momentum higher than 17 GeV, and another muon must be above 8 GeV. In the case of the double-electron trigger, the thresholds are 23 GeV and 12 GeV. For the cross trigger, there can either be a muon with $p_{\mathrm{T}} > 23$ GeV and an electron with $p_{\mathrm{T}} > 12$ GeV, or an electron with $p_{\mathrm{T}} > 23$ GeV and a muon with $p_{\mathrm{T}} > 8$ GeV.

The trigger information is available in the simulated samples, so the same trigger requirements as for data-taking are taken for simulation. To be not too sensitive to trigger turn-on effects, possibly not accurately modeled in the simulation, a selection that is slightly tighter than the trigger thresholds is applied at analysis level, where all the thresholds mentioned before are raised as follows:

$$\begin{aligned} 8\ \mathrm{GeV} &\rightarrow 10\ \mathrm{GeV}, \\ 12\ \mathrm{GeV} &\rightarrow 15\ \mathrm{GeV}, \\ 17\ \mathrm{GeV} &\rightarrow 20\ \mathrm{GeV}, \\ 23\ \mathrm{GeV} &\rightarrow 25\ \mathrm{GeV}. \end{aligned}$$

The remaining difference between data and simulation is taken into account as a systematic uncertainty.

6.3 Object selection

6.3.1 Lepton isolation variable

The default lepton isolation variables used in CMS analyses are based on the particle-flow (PF) measurements. The added transverse momenta of the measured photons, neutral hadrons and charged hadrons in a cone of $\Delta R < 0.3$ are devided by the leptons transverse momentum[2]. Since there is no vertexing information available to correct the two neutral components for pileup contributions, pileup correction is done on a statistical basis. Within the CMS collaboration, the standard corrections for electrons use the area-median method, often called the effective-area method. The standard muon isolation is corrected by extrapolation from the charged-hadrons that are not associated with the primary vertex, commonly called $\Delta\beta$ correction. In this analysis, it was concluded that the effective-area method is superior even for muons, as shown in Figure 6.1. As anticipated in Section 3.5 that discussed pileup extrapolation in detail, the effective-area method generally works better for smaller cone sizes with a limited correlation between the charged and neutral pileup components. Therefore, a smaller cone size of $\Delta R < 0.3$ can be also used for muons[3], like for electrons, which improves the fake discrimination power of the relative isolation variable.

To get the relative lepton isolation, only the photon and hadron PF candidates are summed, skipping the PF electron and muon candidates. However, additional leptons

[2] $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$

[3] the usual cone size for muons in CMS Run 2 analyses is $\Delta R < 0.4$

in a cone around the lepton to isolate often mean that the lepton is probably part of the background from jet fragmentation or low-mass resonaces – or in some cases not well reconstructed. In this analysis, this is dealt with by considering the PF lepton candidates in the isolation sum at the object-level and later with a threshold on the invariant mass of opposite-charge lepton pairs (as explained in Section 6.4). Both for electrons and muons, the corresponding PF objects are a subset of the full set of reconstructed electrons and muons available for analysis. If the PF lepton candidates are added to the relative isolation, the fake lepton rejection power increases. In particular, this is due to improved rejection of lepton pairs from B meson decays such as $B \to \ell D X \to \ell \ell' X$. Figure 6.1 also indicates the ROC curve for this final isolation variable dubbed $I_{\text{rel,R=0.3,EA,Lep}}$ for muons. A similar improvement was observed for electrons.

Accordingly, the final definition of the lepton isolation variable is

$$I_{\text{rel,R=0.3,EA,Lep}} = \frac{1}{p_{\text{T}}} \left[\sum_{\text{ch. had.}} p_{\text{T}} + \sum_{\text{leptons}} p_{\text{T}} + \max \left(\sum_{\text{neu. had.}} p_{\text{T}} + \sum_{\text{photons}} p_{\text{T}} - \text{EA}(\eta, \phi)\rho \right) \right]. \tag{6.1}$$

Depending of the context, p_{T} denotes the transverse momentum of the lepton to isolate or the PF candidates. The abbreviation EA stands for the *effective-area*, the extrapolation factor for the estimation of the local neutral pileup energy deposit from the global pileup estimate ρ. Since the pileup contribution density depends on the location in the detector, the effective area is calibrated in bins of lepton η. To avoid the self-contribution of the target lepton to the isolation sum, the PF lepton candidates are required to be $\Delta R > 0.0005$ away from the lepton to isolate. This isolation definition has already been used in the WWW search [28], which is now part of the more inclusive triboson analysis.

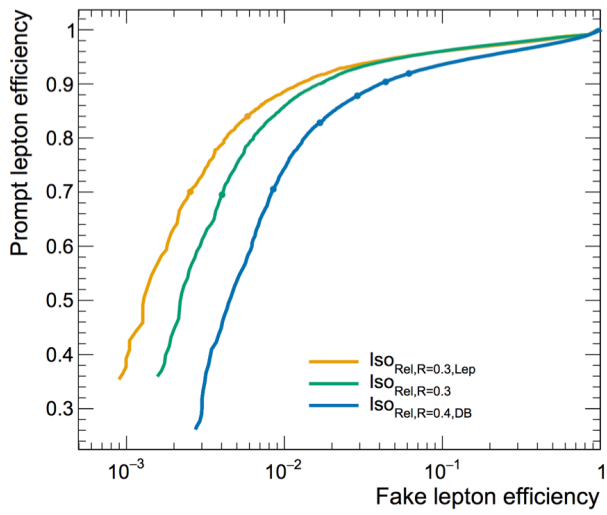


Figure 6.1: Prompt lepton versus fake muon efficiencies for different definitions of the muon isolation variable: the $\Delta\beta$ corrected relative isolation which is standard for muons ($\text{Iso}_{\text{Rel,R=0.4,DB}}$), the relative isolation corrected with the effective-area method ($\text{Iso}_{\text{Rel,R=0.3}}$), and the effective-area-corrected isolation with the added PF lepton component ($\text{Iso}_{\text{Rel,R=0.3,Lep}}$).

6.3.2 Common loose lepton selection

To ensure no double-counting of events in different channels, a common lepton selection is applied at the beginning of the analysis chain. The cuts of the common loose selections for electrons and muons are shown in Table 6.1, including a lower energy threshold of $p_{\mathrm{T}} > 10$ GeV. Note that higher energy thresholds will be applied later on in the four-lepton analysis (see Section 6.4). However, it is important to keep in mind the $p_{\mathrm{T}} > 10$ GeV threshold in the loose selection because it has an effect on the number of counted leptons, which decides to which analysis the event gets assigned.

The selection efficiency of the common loose ID, as well as the data-to-simulation agreement, is shown in Figure 6.2 for the 2018 dataset, confirming the high signal efficiency of the selection. For lepton selections, the measured data-to-simulation efficiency ratio is usually translated to a scale factor plus an associated systematic uncertainty. However, since the common loost ID is not the final lepton selection, the corresponding scale factors are not used but only monitored for validation purposes.

Common loose selection	Electron	Muon
ID	MVA no iso., loose WP	loose
$\lvert\eta\rvert$	< 2.5	< 2.4
p_{T}	> 10 GeV	> 10 GeV
$\lvert d_z\rvert$	< 0.1 cm	< 0.1 cm
$\lvert d_{xy}\rvert$	< 0.05 cm	< 0.05 cm
$I_{\mathrm{rel,R=0.3,EA,Lep}}$	< 0.4	< 0.4

Table 6.1: Common electron and muon loose selection. The electron ID is the MVA from Chapter 5. The loose muon ID selects global muons or tracker muons also accepted by the particle-flow reconstruction (i.e., the muon is also a PF candidate). The limitations in η come from the tracker acceptance for electrons and the acceptance of the muon system. The variables d_z and d_{xy} are the longitudinal and transverse impact parameters relative to the reconstructed primary interaction vertex. The isolation variable is the relative EA-corrected particle-flow isolation with added PF lepton candidates (see Section 6.3.1).

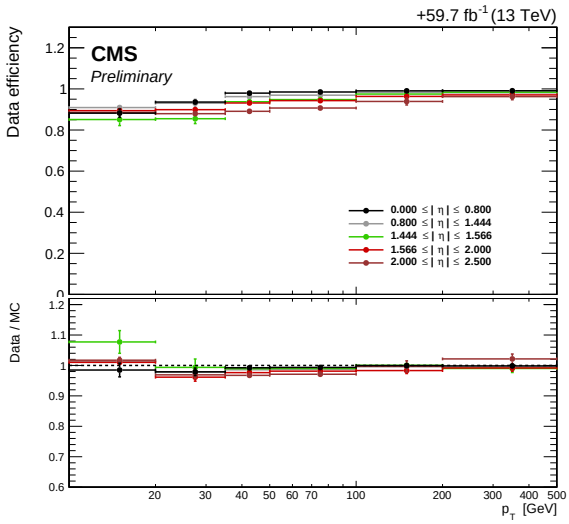


Figure 6.2: The selection efficiency of the common loose ID for electrons in 2018 data, measured in simulation and data with the tag and probe technique as a function of p_{T} and η. The bottom panel shows the data-to-simulation agreement. There are no significant qualitative differences between 2018 and the two other years.

6.3.3 Selections for W and Z candidate leptons

After selecting four-lepton events with the common loose ID, further selection criteria are needed to select the WWZ signal events purely. There are two distinct identification criteria to classify the leptons as *Z candidates* or *W candidates*. The Z candidate selection is the least restrictive because the invariant mass requirement can further reduce fakes. The W candidate selection operates at a tighter working point, resulting in almost complete suppression of fake lepton background (processes with ≤ 3 true leptons plus one reconstructed fake lepton). Since reconstructed fake leptons are a detector effect, the fake lepton background is more challenging to model correctly in simulation, which motivates these selection criteria.

The specific cuts for both the Z and W candidate selections are listed in Table 6.2 for electrons and in Table 6.3 for muons. Besides tightening the ID working points and isolation cuts, these selections introduce a cut on the so-called *impact parameter significance* $|\mathrm{IP}_{3D}/\sigma_{\mathrm{IP}_{3D}}|$, which is the closest distance between the lepton track and the reconstructed primary vertex divided by its measurement uncertainty. Cutting on impact parameters and also on $|\mathrm{IP}_{3D}/\sigma_{\mathrm{IP}_{3D}}|$ is standard for analyses that target prompt leptons. In particular, the $|\mathrm{IP}_{3D}/\sigma_{\mathrm{IP}_{3D}}|$ cut in this analysis is motivated by an improved rejection of leptons from tau decays.

Electron	Z candidate (ZID)	W candidate (WID)
Veto ID	Common loose ID	Common loose ID
ID	MVA without iso., loose WP	MVA with iso. 90% WP
$\lvert\mathrm{IP}_{3D}/\sigma_{\mathrm{IP}_{3D}}\rvert$	< 4	< 4
$I_{\mathrm{rel,R=0.3,EA,Lep}}$	< 0.2	< 0.2

Table 6.2: Final electron selections in the WWZ analysis. The MVA selections are the ones developed in this thesis, explained in Chapter 5.

Muon	Z candidate (ZID)	W candidate (WID)
Veto ID	Common loose ID	Common loose ID
ID	medium	medium
$\lvert\mathrm{IP}_{3D}/\sigma_{\mathrm{IP}_{3D}}\rvert$	< 4	< 4
$I_{\mathrm{rel,R=0.3,EA,Lep}}$	< 0.25	< 0.15

Table 6.3: Final muon selections in the WWZ analysis. Compared to the loose muon ID in the common loose selection, the *medium* ID adds track- and muon-quality requirements. Note that even though the numeric values of the $|\mathrm{IP}_{3D}/\sigma_{\mathrm{IP}_{3D}}|$ thresholds are the same as for electrons (see Table 6.2), the effective impact parameter thresholds becuase the uncertainty in the denominator is generally smaller for muons.

Impact parameters and also the impact parameter significance are among the variables that are the most challenging to model correctly in simulation. Figure 6.3 shows the $|\mathrm{IP}_{3D}/\sigma_{\mathrm{IP}_{3D}}|$ distributions for all leptons in four-lepton events in the ZZ control region (defined later in Section 6.5). These distributions suggest a satisfactory agreement between data and simulation. However, it was found that the impact parameter significance cut dominates the data-to-simulation disagreement of the selection efficiencies, especially in the endcaps of the detector. This can be seen in Figure 6.4, which shows not only the electron Z and W candidate selection efficiencies measured in data and

simulation together with their ratios but also the equivalent figures for similar selections where only the $|\mathrm{IP}_{3D}/\sigma_{\mathrm{IP}_{3D}}|$ cut was omitted.

The low simulation agreement of $|\mathrm{IP}_{3D}/\sigma_{\mathrm{IP}_{3D}}|$ is likely due to primary vertex misidentification. If the primary vertex is assigned wrongly, all lepton impact parameters increase, which reduces the selection efficiency. Because of the higher particle density in the endcaps, primary vertex misidentification happens much more frequently in this region. This explains why adding the impact parameter significance cut has a larger effect on the selection efficiency agreement in the endcaps than in the barrel region.

The data-to-simulation agreement of the W and Z candidate selections – as shown in Figure 6.4 for electrons – is used as a p_{T}- and η-dependent scaling factor that corrects the weight of simulated events. The uncertainty on the scaling factor measurement is used as a systematic uncertainty in the analysis. For the previous WWW search [28], different scale factors were applied for the individual selection steps, namely one set for the common loose ID and one set for the Z or W candidate selection. For this WWZ analysis, the electron scale factors were derived directly with unfiltered reconstructed electrons in the denominator, avoiding the double-counting of correlated systematic uncertainties. Here, this thesis work had a direct impact on the analyses of the other massive triboson final states, as the scale factors for the dedicated two-lepton and three-lepton IDs in the WWW analysis have been derived as well.

6.3.4 Jet selection and b-tagging

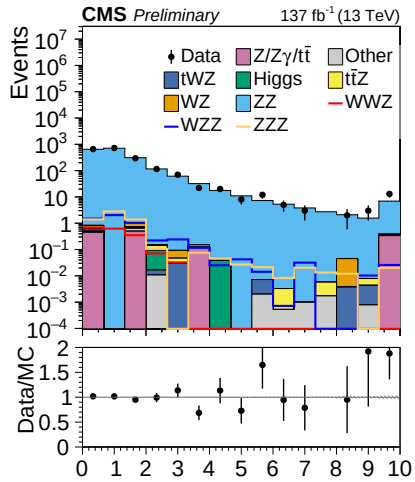
This analysis uses the standard jets reconstructed from particle-flow candidates clustered with the anti-k_{T} algorithm [92] with a cone size of $\Delta R < 0.4$. The lower jet energy threshold is 20 GeV, and only jets within the tracker acceptance of $|\eta| < 2.4$ are considered. Several identification criteria that are standard for CMS analyses are used to clean jets from noise or mismeasured jets. For 2016 data and simulation, loose jet selection criteria are taken. With the increased pileup in 2017 and 2018, the jet selections are tightened as recommended for analysis. Both selections are summarized in Table 6.4.

Jet selection	Loose (2016)	Tight (2017, 2018)
Neutral hadronic energy fraction	< 0.99	< 0.9
Neutral electromagnetic energy fraction	< 0.99	< 0.9
Number of constituents	> 1	> 1
Charged hadron fraction	> 0.0	> 0.0
Charged multiplicity	> 0	> 0
Charged electromagnetic energy fraction	> 0.99	-

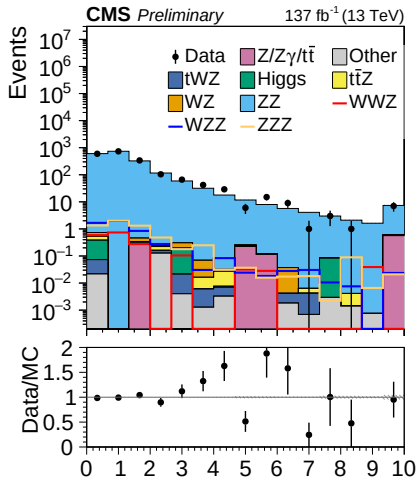
Table 6.4: Jet selection criteria used in the analysis.

The prompt leptons in the event signature are often reconstructed as jets as well, so if a jet is matched to a lepton that passes the common loose ID withing a cone of $\Delta R < 0.4$, it is not considered. On top of the selection, jet energy corrections are applied.

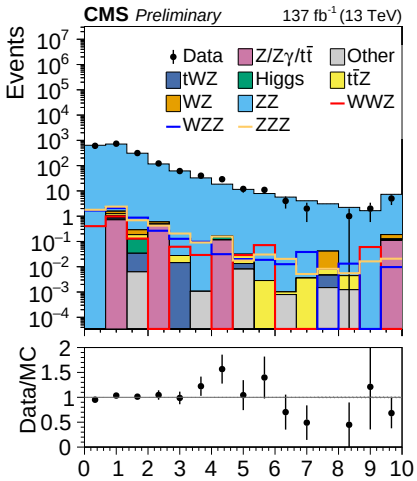
A b tagger based on a deep neural network called *DeepCSV* selects the b jet candidates in this analysis [128]. The DeepCSV tagger outputs probabilities for each jet to contain a certain number of b or c hadrons. To obtain the b jet discriminant, the probabilities for the jet to contain one or two b hadrons are summed, and the loose working point for this discriminant is used for all three data-taking years.



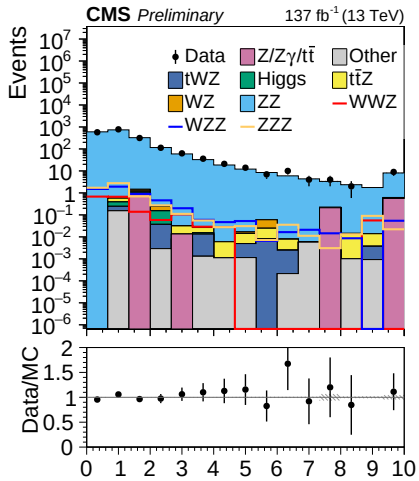
(a) Leading Z lepton candidate



(b) Subleading Z lepton candidate



(c) Leading W lepton candidate



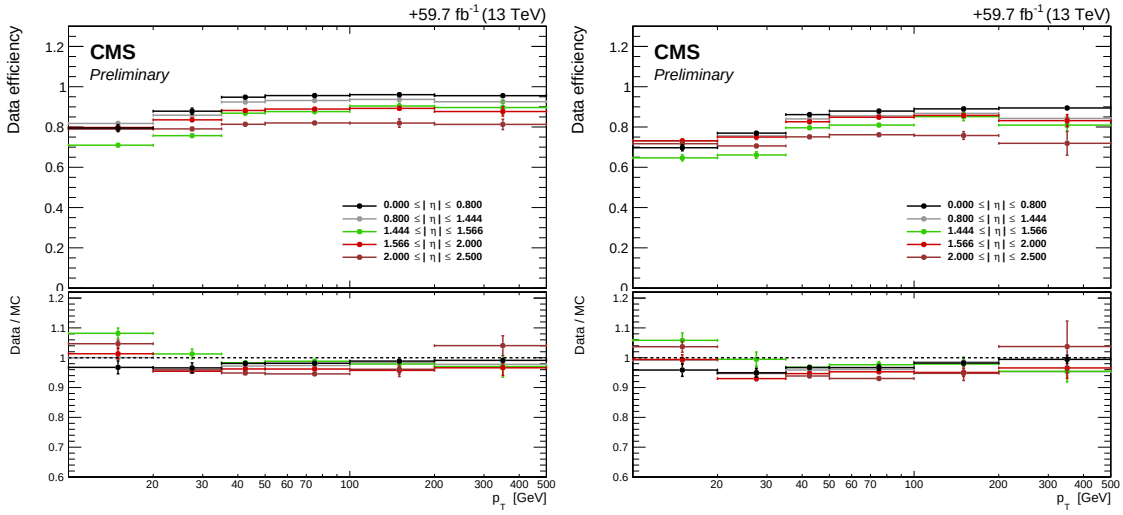
(d) Subleading W lepton candidate

Figure 6.3: $|\mathrm{IP}_{3D}/\sigma_{\mathrm{IP}_{3D}}|$ distributions of the leptons in the ZZ control region.

Similar to the treatment of lepton selection efficiencies, the b-tagging efficiencies are measured in data and simulation, and the ratio is applied as a scaling factor to the simulated event weights, with the uncertainty in the ratio as a systematic uncertainty. Different from electron and muon identification, there are two different sets of scale factors and uncertainties: one for Heavy Flavor (HF) jets – in this case b jets –, and one for Light Flavor (LF) jets. The HF and LF jet selection uncertainties translate to two independent nuisance parameters, which is important to keep in mind for the discussion of systematic uncertainties later in Section 6.9.

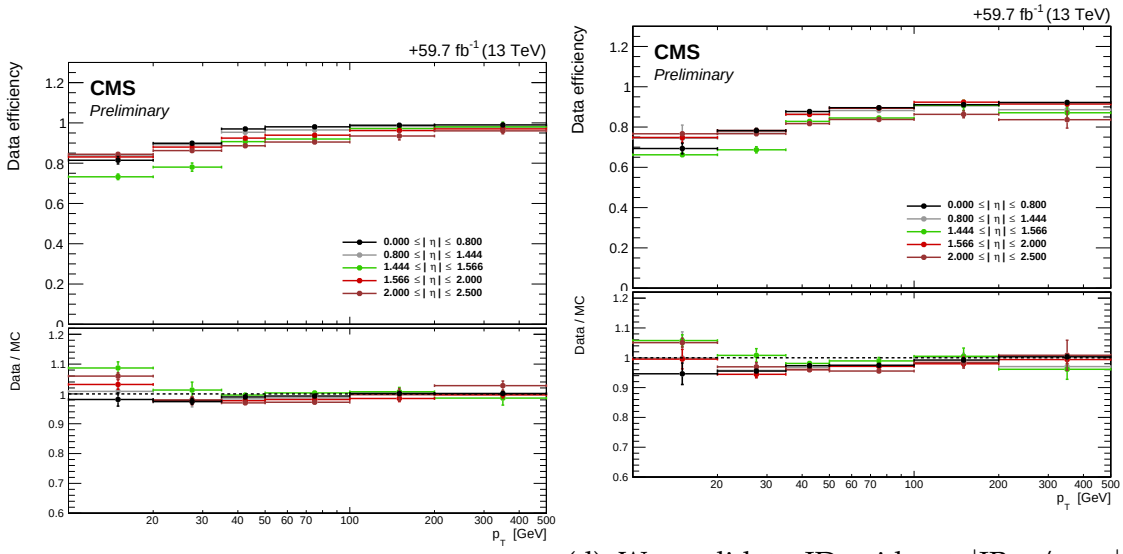
6.3.5 Missing transverse momentum

The missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$ is the negative sum of the transverse momentum of all particle-flow candidates in the event, where the jets are subject to energy corrections [96]. The $p_{\mathrm{T}}^{\mathrm{miss}}$ variable is alternatively referred to as MET for *missing* E_{T}.



(a) Z candidate ID

(b) W candidate ID

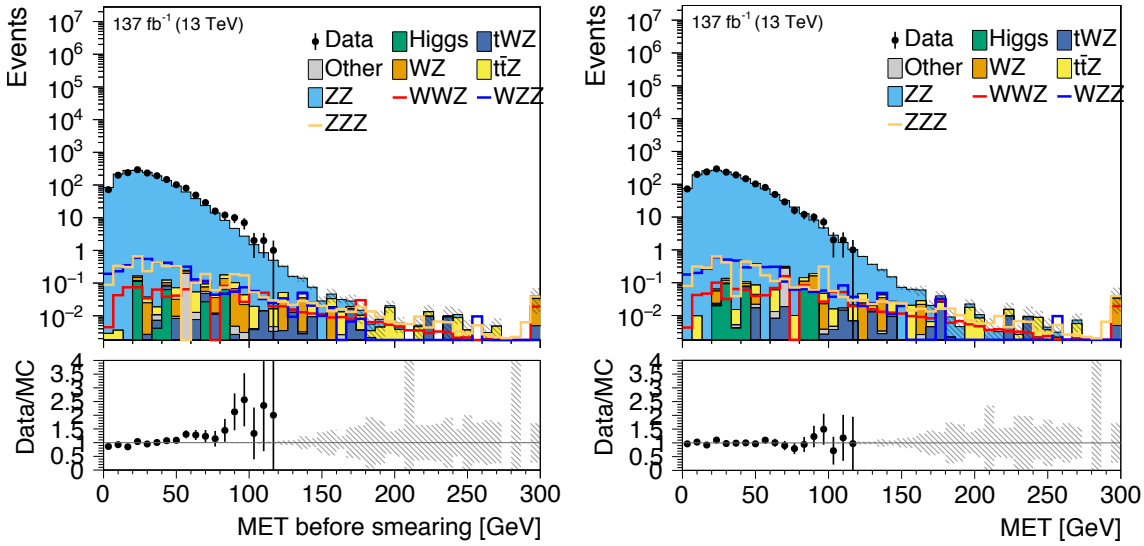


(c) Z candidate ID without $|\mathrm{IP}_{3D}/\sigma_{\mathrm{IP}_{3D}}|$ cut cut

(d) W candidate ID without $|\mathrm{IP}_{3D}/\sigma_{\mathrm{IP}_{3D}}|$

Figure 6.4: Top row: the selection efficiencies of the W and Z candidate electron IDs measured in data and simultion. Bottom row: selection efficiencies for the same IDs but with the $|\mathrm{IP}_{3D}/\sigma_{\mathrm{IP}_{3D}}|$ cut left out. This cut has the largest effect on the data-to-simulation agreement, shown in the bottom panel of each plot. The remaining disagreements were found to be mainly because of the isolation cut.

In the ZZ control region (four leptons giving two Z boson candidates) a mismodeling of $p_{\mathrm{T}}^{\mathrm{miss}}$ was observed, as shown in Figure 6.5a. The ZZ process has no intrinsic missing energy. Consequently, the MET distribution for such events is representative of the detector resolution on missing transverse momentum, which might be inaccurate in simulation. This mismodeling is corrected by smearing $p_{\mathrm{T}}^{\mathrm{miss}}$ in simulation, where the smearing is calibrated on a completely independent photon dataset. This procedure has already been applied in other CMS analyses before [129]. After the smearing, the missing transverse momentum agreement in the ZZ control region has improved (see Figure 6.5b).



(a) Before $p_{\mathrm{T}}^{\mathrm{miss}}$ smearing.

(b) After $p_{\mathrm{T}}^{\mathrm{miss}}$ smearing.

Figure 6.5: The $p_{\mathrm{T}}^{\mathrm{miss}}$ distributions in the ZZ control region before and after $p_{\mathrm{T}}^{\mathrm{miss}}$ smearing in simulated events.

6.4 Event preselection

After asking for four leptons that pass the common loose ID, the events have to be preselected to suppress reducible background before categorizing them to separate the irreducible backgrounds. First, the events are required to pass the dilepton trigger by applying an energy threshold to the leptons that are slightly above the trigger thresholds. This trigger-emulating selection has already been explained in Section 6.2.1.

Next, two leptons have to be identified as the decay products of the Z boson. These two leptons must be of same flavor, opposite charge, pass the Z candidate ID explained in Section 6.3.3 and pass the transverse momentum thresholds $p_{\mathrm{T}}^{\ell 1} > 25$ GeV and $p_{\mathrm{T}}^{\ell 2} > 15$ GeV for the leading and subleading lepton respectively. The invariant mass of the Z candidates must be within 10 GeV around the Z boson mass, both for electron and muon pair Z candidates. The two other W candidate leptons also must be of opposite charge but can be of any flavor. These leptons also need to surpass the same energy thresholds as the Z candidate leptons, namely $p_{\mathrm{T}}^{\ell 3} > 25$ GeV and $p_{\mathrm{T}}^{\ell 4} > 15$ GeV.

Some QCD processes with large cross-sections produce B hadrons or low-mass resonances that decay to leptons, such as J/ψ and Υ mesons. To remove such events, the invariant mass of any opposite-charge lepton pair is required to be greater than 12 GeV.

The preselection criteria are summarized in Table 6.5. Four-lepton events discarded by the preselection are not used anymore in the analysis, except for estimating the fake-lepton background, which will be explained further down in Section 6.8.3. After the preselection, the ZZ background still makes up the bulk of events, followed by the tt̄Z background and then the WWZ signal in the same range as other minor backgrounds.

6.5 Event categorization

After the preselection, the events are sorted into four different categories to separate true four-lepton backgrounds. The purest category is the *eμ region*, where the two W candidate leptons are of a different flavor. This heavily reduces the ZZ background. Still, the

Features	Selections
Number of leptons	Select events with 4 leptons passing common loose ID
Triggers	Select events passing dilepton triggers
Z lepton	Find opposite charge lepton pairs, passing ZID, closest to m_Z Require Z leptons to have $p_T > 25, 15$ GeV
W lepton	Require that leftover leptons are opposite charge and pass WID Require W leptons to have $p_T > 25, 15$ GeV
Low mass resonances	Require any opposite charge pair invariant mass to be greater than 12 GeV
Z mass window	Require invariant mass of the Z leptons to be within 10 GeV of Z boson mass

Table 6.5: Summary of the event preselections in the WWZ analysis.

$e\mu$ region contains ZZ events in which one Z boson decays to two tau leptons and the taus decay to an electron and a muon. Only events with zero b-tagged jets are accepted in the $e\mu$ region to reject the $t\bar{t}Z$ background. Different-flavor events with b jets are assigned to another category, referred to as *$e\mu$ with b jets*, which is used as a control region for the $t\bar{t}Z$ background, which intrinsically has two b jets in the final state.

The second event category in which some signal can be extracted is the region where the two W candidates are of the same flavor, but their invariant mass is 10 GeV away from the Z boson mass such that ZZ contributions are reduced. This phase-space is called the *$ee/\mu\mu$ region*. The ZZ background is still dominant in this category, so it is less sensitive to the WWZ signal than the $e\mu$ category. With the invariant-mass requirement inverted, the $ee/\mu\mu$ region becomes the *$ee/\mu\mu$ on-Z region*, where the invariant mass of the W lepton candidates is on the Z peak. This category is an almost entirely pure control region for the ZZ background. Both the on-Z and off-Z $ee/\mu\mu$ regions ask for zero b-tagged jets to suppress the $t\bar{t}Z$ background.

The event categorization as outlined in the text is also presented in Table 6.6.

Category	Selection	Description
$e\mu$	different-flavor W lepton candidates, no b jet candidates	main signal category
$ee/\mu\mu$	same-flavor W lepton candidates, invariant mass 10 GeV away from Z mass	secondary signal region,no b jet candidates
$e\mu$ with b jets	different-flavor W lepton candidates, ≥ 1 b jet candidates	control region for $t\bar{t}Z$ background
$ee/\mu\mu$ on-Z	same-flavor W lepton candidates, invariant mass 10 GeV within Z mass, no b jet candidates	control region for ZZ background

Table 6.6: Summary of the event categorization with the two signal regions and the two control regions for the dominant backgrounds.

6.6 Event classification with Boosted Decision Trees

In the two signal categories, the WWZ events are separated from the backgrounds with Boosted Decision Trees (BDTs). Later in Section 6.7, it will be discussed how the discriminating power of the BDTs is used to extract the signal. This section discusses the BDT input variables (or features), the training scheme, and the validation of the BDT output by means of the control regions.

Three BDTs were trained in total. Two BDTs target the $e\mu$ region, where one separates the WWZ signal from the $t\bar{t}Z$ background, and the other separates WWZ from ZZ events. A third BDT discriminates against the ZZ background in the $ee/\mu\mu$ region.

6.6.1 Input variables

When machine learning is used for event separation in collider experiments, one usually prefers features that are well modeled in the simulation. Every added feature needs to be carefully validated, so the set of features is often reduced as much as possible without significant information loss, which results in more trustworthy BDT discriminants. For this reason, the final BDTs have been trained with a limited set of features that proved to be important in studies with more inclusive feature sets.

Since the $t\bar{t}Z$ and ZZ backgrounds are qualitatively not the same, different feature sets have been used for their discrimination. Table 6.7 gives an overview on the final list of features used for each background. What follows now is a discussion of the features, grouped by whether they are used for $t\bar{t}Z$, ZZ, or both backgrounds. Afterwards, the feature value distributions are shown as histograms.

variables	$t\bar{t}Z$ BDT	ZZ BDT
$m_{\ell\ell}$	✓	✓
$p_{\mathrm{T}}^{\mathrm{Z}}$	✓	✓
m_{T2}	✓	✓
$p_{\mathrm{T}}^{\ell 3}$	✓	✓
$p_{\mathrm{T}}^{\ell 4}$	✓	✓
$p_{\mathrm{T},4\ell}$	✓	✓
$\sum_i^4 p_{\mathrm{T}i}$	✓	✓
P_ζ		✓
P_ζ^{vis}		✓
$p_{\mathrm{T}}^{\mathrm{miss}}$		✓
$m_{\mathrm{T}}^{\ell 3}$		✓
$m_{\mathrm{T}}^{\ell 4}$		✓
$m_{4\ell}$		✓
$\min \Delta R(j, \ell 3)$	✓	
$\min \Delta R(j, \ell 4)$	✓	
leading jet p_{T}	✓	

Table 6.7: List of input variables for the BDT trainings in the WWZ analysis.

6.6.1.1 $t\bar{t}Z$ discrimination

The most discriminant variable for the $t\bar{t}Z$ background is the number of b tagged jets. However, there are relatively large systematic uncertainties on the b-tagging efficiencies, and the data-to-simulation agreement is not perfect even after b-tagging scale factors are applied. Using the number of b jets in the BDT would demand a careful validation of the BDT score distribution, and with the number of b jets used as a feature, there would be no orthogonal variable left to create a control region for $t\bar{t}Z$ background estimates.

Instead of b jet specific information, general jet information is exploited to separate the $t\bar{t}Z$ events from the signal. Since both the jets and the W candidate leptons originate from top decays, the distance ΔR between the W candidate leptons and the respective closest jet is smaller for $t\bar{t}Z$ than for WWZ events (see Figures 6.20 and 6.21). In addition to these two distance variables, the p_{T} of the leading jet is used. This variable is set to zero if no jet is present in the event, so it is clear to the BDT if there is a jet in the event or not. The distribution of the leading jet p_{T} is shown in Figure 6.22.

Other than these three jet-related variables, no other features dedicated to $t\bar{t}Z$ separation are used that are not shared with the ZZ classification.

6.6.1.2 ZZ discrimination

The $e\mu$ category contains ZZ events where one Z boson decays to two τ leptons, resulting in two leptons and large $p_{\mathrm{T}}^{\mathrm{miss}}$, which resembles the WWZ signature. In such events, the angle between the neutrinos and the leptons from the tau decays is typically small. It has been shown in other analyses that the vectorial sums of the lepton momenta (and optionally also MET), projected on the bisector of the lepton directions, contains information that can be combined with MET to obtain observables for separation of such $Z \to \tau\tau$ events [130]. This was the motivation to include the variables P_ζ and P_ζ^{vis} among the ZZ specific features:

$$P_\zeta = (\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} + \vec{p}_{\mathrm{T}}^{\,\ell 1} + \vec{p}_{\mathrm{T}}^{\,\ell 2}) \cdot \hat{\zeta} \qquad \text{and} \qquad P_\zeta^{\mathrm{vis}} = (\vec{p}_{\mathrm{T}}^{\,\ell 1} + \vec{p}_{\mathrm{T}}^{\,\ell 2}) \cdot \hat{\zeta}. \tag{6.2}$$

Here, $\hat{\zeta}$ is the normalized direction vector of the bisector of the two lepton directions. Figure 6.6 depicts a geometrical representation of these variables. Their distributions in the four categories are shown in Figures 6.14 and 6.15. Unlike in [130], no attempt was made to combine these variables with $p_{\mathrm{T}}^{\mathrm{miss}}$ to a performant discriminant. Instead, $p_{\mathrm{T}}^{\mathrm{miss}}$ is included as a feature as well, although the main reason for including $p_{\mathrm{T}}^{\mathrm{miss}}$ is that there is no missing energy in ZZ events if no tau decays are involved.

Another variable that is explicitly used for ZZ separation is the invariant mass of the four-lepton system $m_{4\ell}$. With three weak bosons in the final state instead of two, the WWZ signal has a higher invariant mass in total. Even though some energy is carried away by the neutrinos, the large angular separation between the vector bosons results in a large $m_{4\ell}$. The $m_{4\ell}$ distributions for all four categories is shown in Figure 6.19. It is especially powerful in the $e\mu$ region because for the ZZ background, most of the second Z boson momentum is carried away by the neutrinos from the τ decays. The $m_{4\ell}$ variable is not as useful for the $t\bar{t}Z$ background because the leptons also come from the decay of three vector bosons, albeit with intermediate top quark decays.

Two more features dedicated to ZZ are the transverse masses $m_{\mathrm{T}}^{\ell 3}$ and $m_{\mathrm{T}}^{\ell 4}$. Beware that the definition of transverse mass here is not $m_{\mathrm{T}}^2 = m^2 + p_x^2 + p_y^2 = E^2 - p_z^2$ which is a useful quantity that is invariant under Lorentz boosts along the z direction. In hadron collider physics, the transverse mass is an observable that is substituted for the invariant

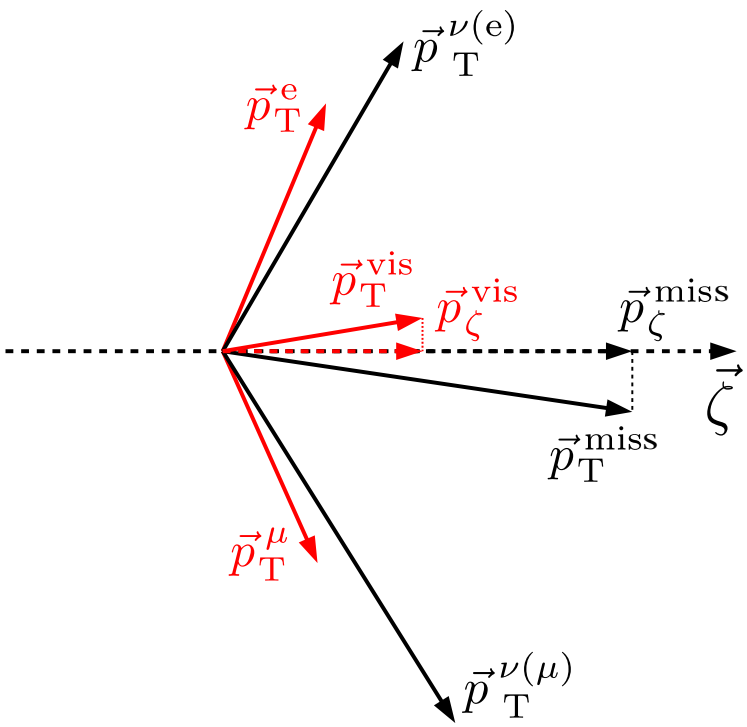


Figure 6.6: Projections of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and leptons $\vec{p}_{\mathrm{T}}$ on the $\hat{\zeta}$ direction [130].

mass of a two-particle system when one particle cannot be measured in the longitudinal direction, as is the case for missing transverse energy. It is defined as

$$M_{\mathrm{T}}^2 = (E_{\mathrm{T},1} + E_{\mathrm{T},2})^2 - (\vec{p}_{\mathrm{T},1} + \vec{p}_{\mathrm{T},2})^2, \tag{6.3}$$

where E_{T} is the transverse energy of each particle:

$$E_{\mathrm{T}} = m^2 + (\vec{p}_{\mathrm{T}})^2. \tag{6.4}$$

Since the actual masses of the daughter particles are tiny compared to their energies, the definition of the transverse mass can be simplified to

$$M_{\mathrm{T}}^2 = 2E_{\mathrm{T},1}E_{\mathrm{T},2}(1 - \cos\phi), \tag{6.5}$$

with ϕ being the angle between the daughter particles in the transverse plane.

When there are multiple two-particle systems in the event where one of the daughter particles in each system is invisible, it is not clear how the missing transverse momentum is split up between the two invisible particles. In this analysis, this applies to the two W boson decays with a neutrino produced in each decay. Still, one can make an assumption to partition the missing transverse momentum to the neutrinos, for example, the assumption that the two resulting transverse masses must be close to the value expected from W decays. However, if one is only interested in the discrimination power of such transverse mass variables and less in their interpretation, the total $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ vector might be used for both decay systems without any attempt to partition it correctly. This was done for the $m_{\mathrm{T}}^{\ell 3}$ and $m_{\mathrm{T}}^{\ell 4}$ variables in this analysis, which are the transverse masses of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ plus the leading or subleading W candidate lepton, respectively. These transverse mass variables can separate the signal from the ZZ background because the two W candidate leptons come from a more massive intermediate state for the signal, namely two W bosons instead of one Z boson.

Another transverse mass observable for tau decay discrimination is m_{T2}. In its definition, the missing energy vector is split up between the two decay systems [131]. The m_{T2} variable is the maximum of the transverse masses of both decay systems for the neutrino momentum partitioning for which the maximum of the transverse masses is minimal:

$$m_{\mathrm{T2}} = \min_{\vec{p_{\mathrm{T}}}^{\nu(1)}+\vec{p_{\mathrm{T}}}^{\nu(2)}=\vec{p_{\mathrm{T}}}^{\mathrm{miss}}} \left[\max \left(m_{\mathrm{T}}^{(1)}(\vec{p_{\mathrm{T}}}^{\nu(1)}, \vec{p_{\mathrm{T}}}^{e}), m_{\mathrm{T}}^{(2)}(\vec{p_{\mathrm{T}}}^{\nu(2)}, \vec{p_{\mathrm{T}}}^{\mu}) \right) \right] \tag{6.6}$$

The idea behind m_{T2} is that for correct transverse masses $m_{\mathrm{T}}^{(1,2)}$, even the larger of the two is less than or equal to the parent mass. The minimization over all possible splittings of $\vec{p_{\mathrm{T}}}^{\mathrm{miss}}$ to the transverse momenta of the neutrinos $\vec{p}_{\mathrm{T}}^{\nu(1,2)}$ ensures that each transverse mass does not exceed the parent mass. The parents are W bosons for the signal and τ leptons for the ZZ background. Consequently, the average m_{T2} is much lower for the background than for the signal. This difference makes m_{T2} one of the most discriminating variables, as can be seen in the histograms in Figure 6.9.

Most of the arguments related to invariant masses were oriented to the on-shell part of the WWZ signal, which is considered as the more important component. However, invariant masses are helpful for ZZ vs. ZH $\rightarrow$ WWZ separation as well, with the argument turned around. In this case, one of the two W bosons is highly off-shell, and the invariant lepton masses are *lower* than for the ZZ background.

6.6.1.3 Common variables

The list of common features in Table 6.7 includes first and foremost $m_{\ell\ell}$, the invariant mass of the W lepton candidates. This is a very powerful observable to separate the on-shell triboson signal from the Higgs-mediated production, with the Higgs process at low mass and on-shell WWZ at high values. As for the other mass-related features, the backgrounds are somewhere in between, allowing the BDT algorithm to separate the total WWZ signal from the backgrounds. Distributions of the $m_{\ell\ell}$ variable are shown in Figure 6.7.

The reconstructed transverse momentum of the first Z boson candidate $p_{\mathrm{T}}^{\mathrm{Z}}$ behaves differently for every signal or background process because the Z boson recoils against different systems. Hence, $p_{\mathrm{T}}^{\mathrm{Z}}$ is a natural choice to include among the common features. Its distributions are shown in Figure 6.8. Further lepton energy variables are also included, namely the transverse momenta of the W lepton candidates $p_{\mathrm{T}}^{\ell 3}$ and $p_{\mathrm{T}}^{\ell 4}$ (Figures 6.10 and 6.11), as well as the sum of all lepton transverse momenta ($\sum_i^4 p_{\mathrm{T}i}$, Figure 6.13) and the transverse momentum of the four-lepton system ($p_{\mathrm{T},4\ell}$, Figure 6.12).

The m_{T2} variable motivated by the ZZ background also showed some discrimination power for the $\mathrm{t\bar{t}Z}$ background. Hence it is listed as a common feature.

One might wonder why lepton impact parameters were not mentioned at all in this discussion of BDT features. After all, they are particularly useful to discriminate prompt leptons from tau-decay leptons, so in principle, they are valuable for ZZ separation in the $e\mu$ region. However, we have already seen in Section 6.3.3 when discussing lepton selection scale factors that the modeling of impact parameters in simulation is not very accurate. Therefore, it was decided not to include impact parameters among the BDT input variables. The lepton selections already use relatively tight impact parameter cuts, so omitting them at this point does not have a significant impact on the analysis performance.

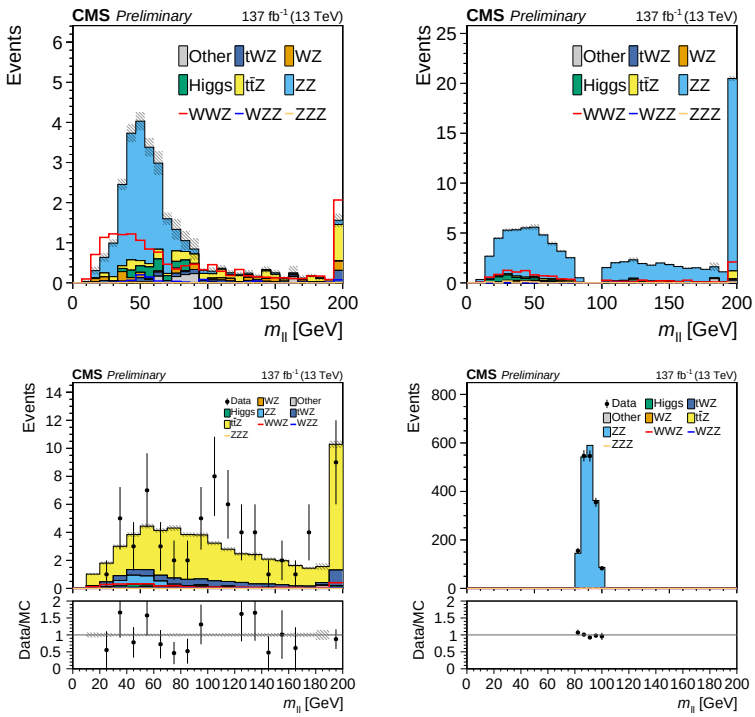


Figure 6.7: The $m_{\ell\ell}$ variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in both the $t\bar{t}Z$ and ZZ trainings.

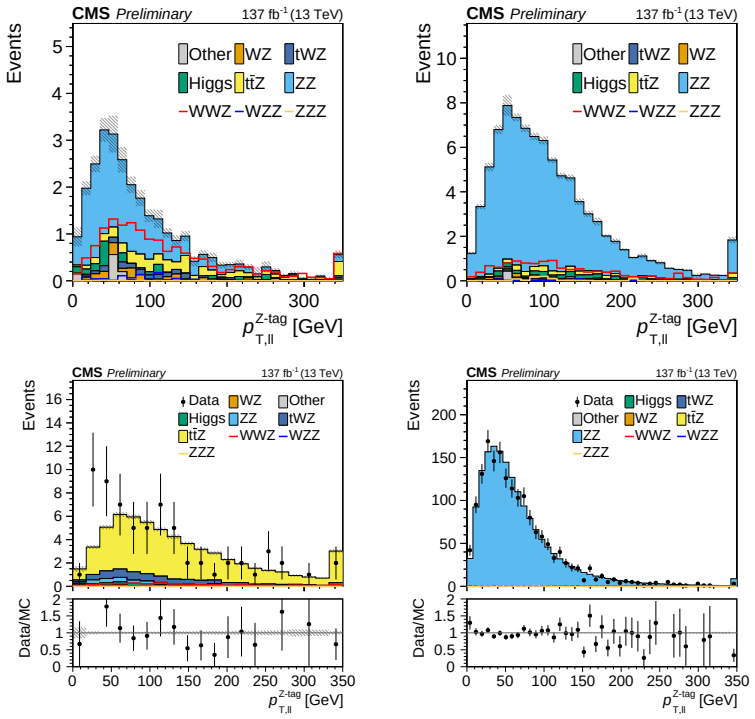


Figure 6.8: The p_T^Z variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in both the $t\bar{t}Z$ and ZZ trainings.

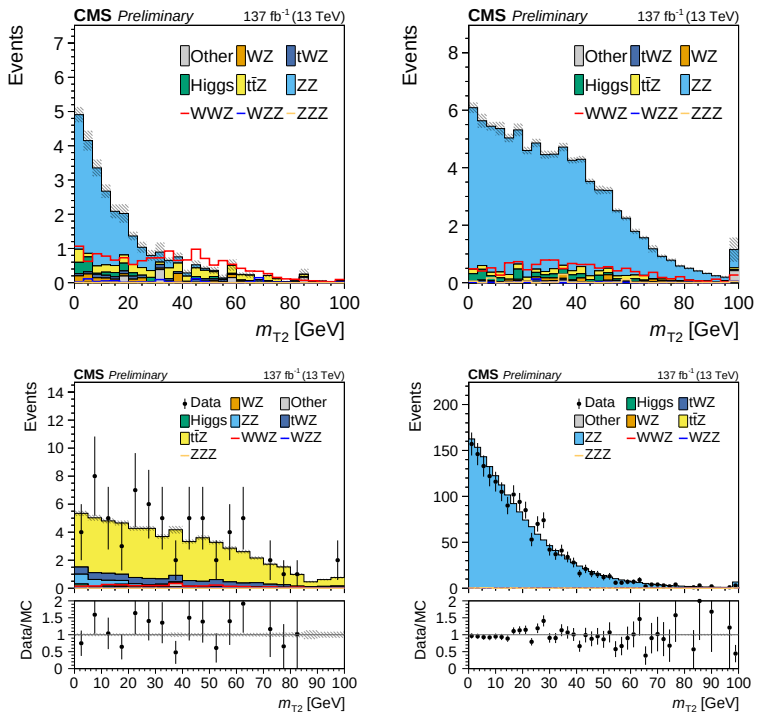


Figure 6.9: The m_{T2} variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in both the $t\bar{t}Z$ and ZZ trainings.

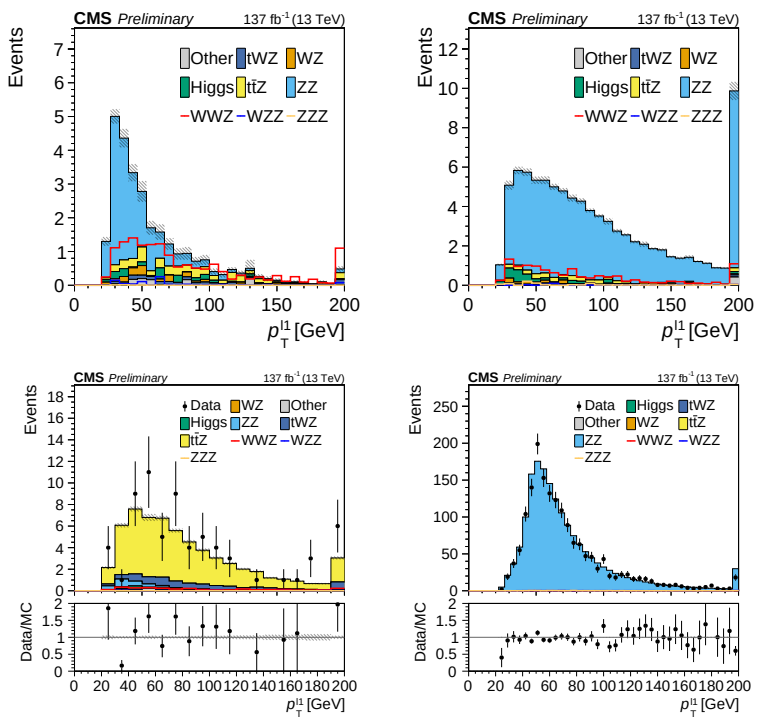


Figure 6.10: The $p_T^{\ell 3}$ variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in both the $t\bar{t}Z$ and ZZ trainings.

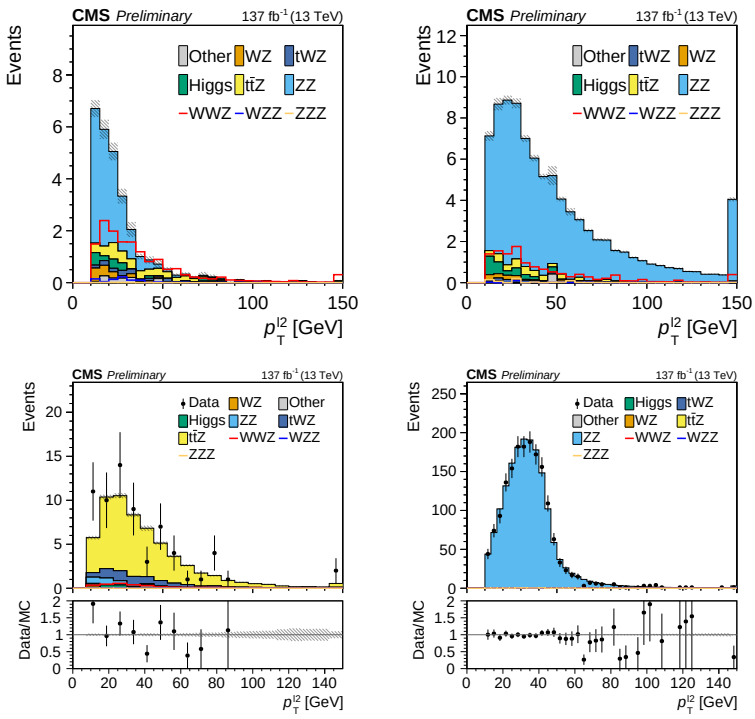


Figure 6.11: The $p_T^{\ell 4}$ variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in both the $t\bar{t}Z$ and ZZ trainings.

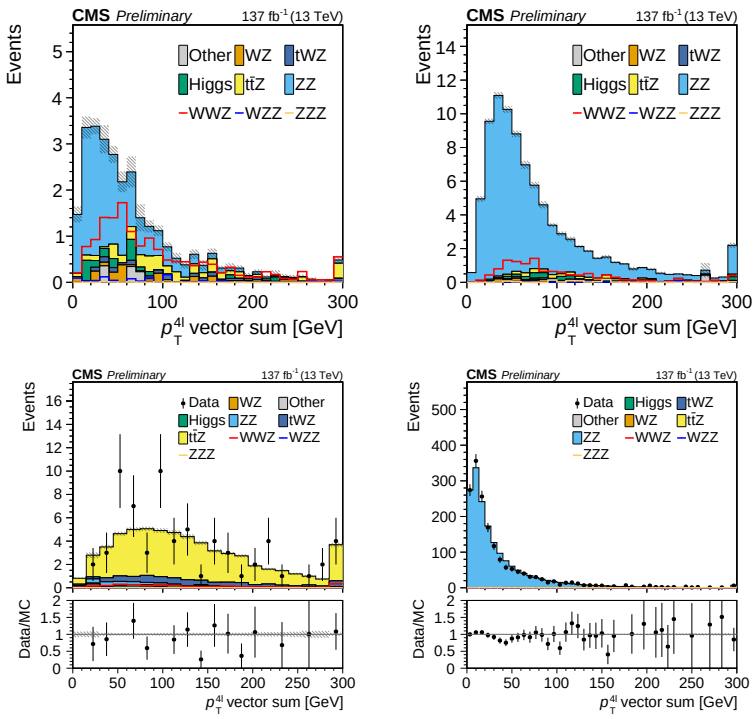


Figure 6.12: The $p_{T,4\ell}$ variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in both the $t\bar{t}Z$ and ZZ trainings.

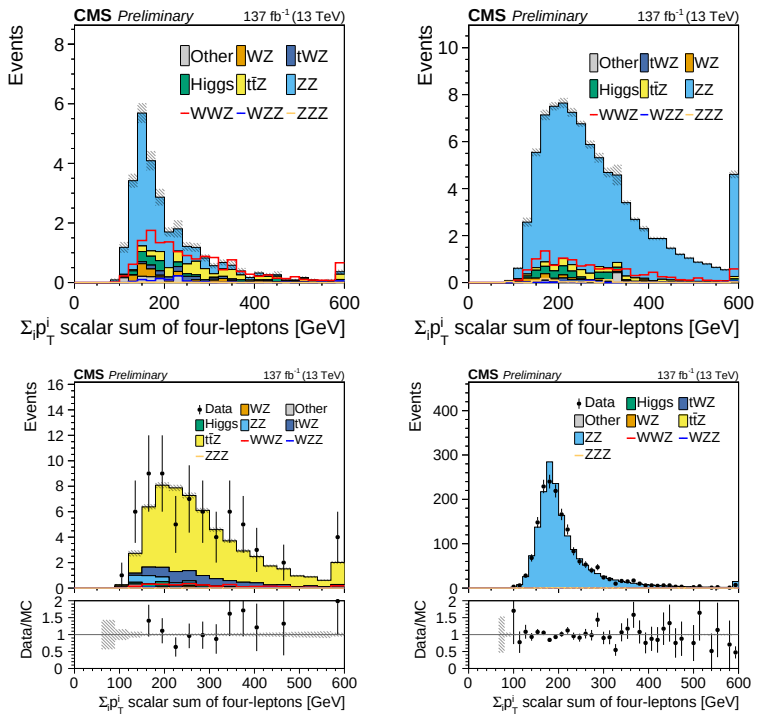


Figure 6.13: The $\sum_i^4 p_{\mathrm{T}_i}$ variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in both the $\mathrm{t\bar{t}Z}$ and ZZ trainings.

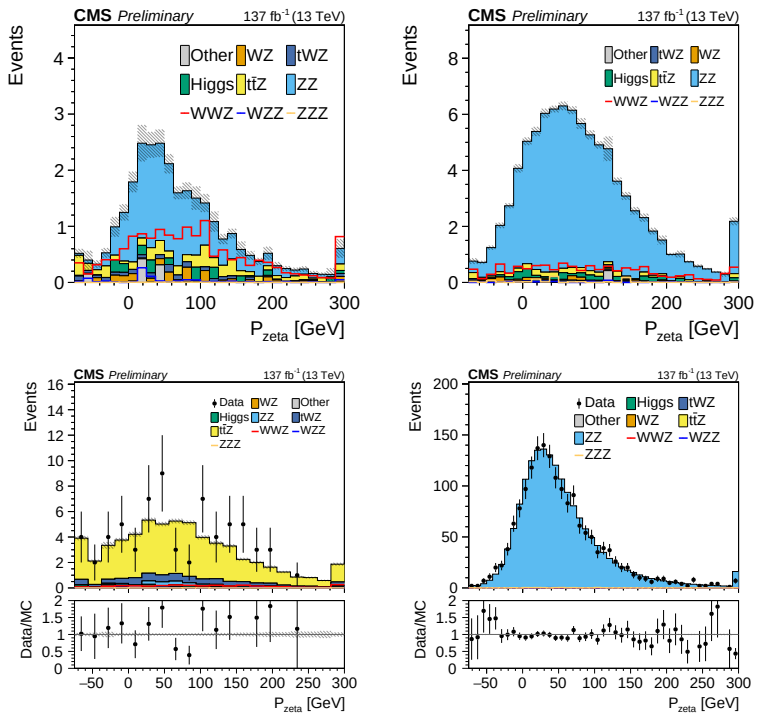


Figure 6.14: The P_ζ variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in only the ZZ trainings.

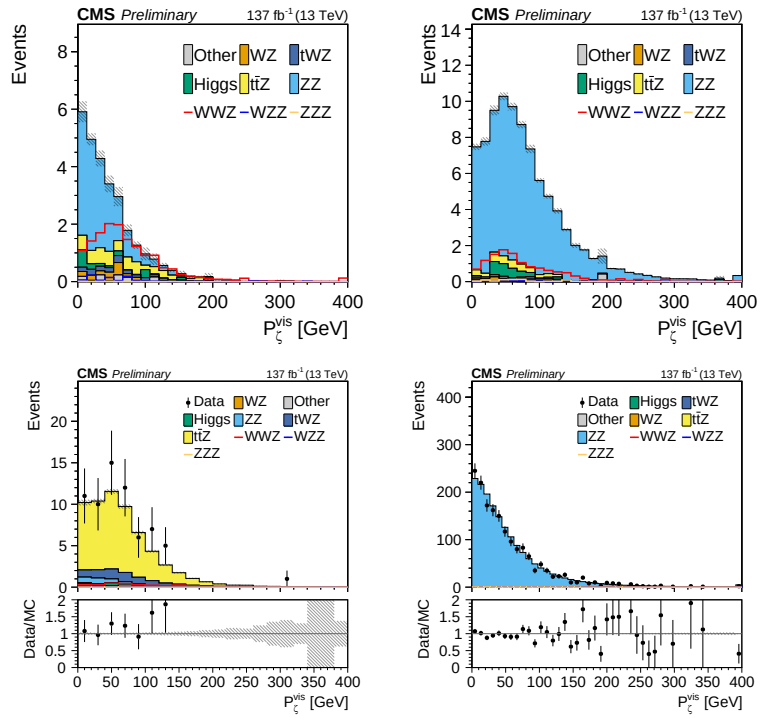


Figure 6.15: The P_{ζ}^{vis} variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in only the ZZ trainings.

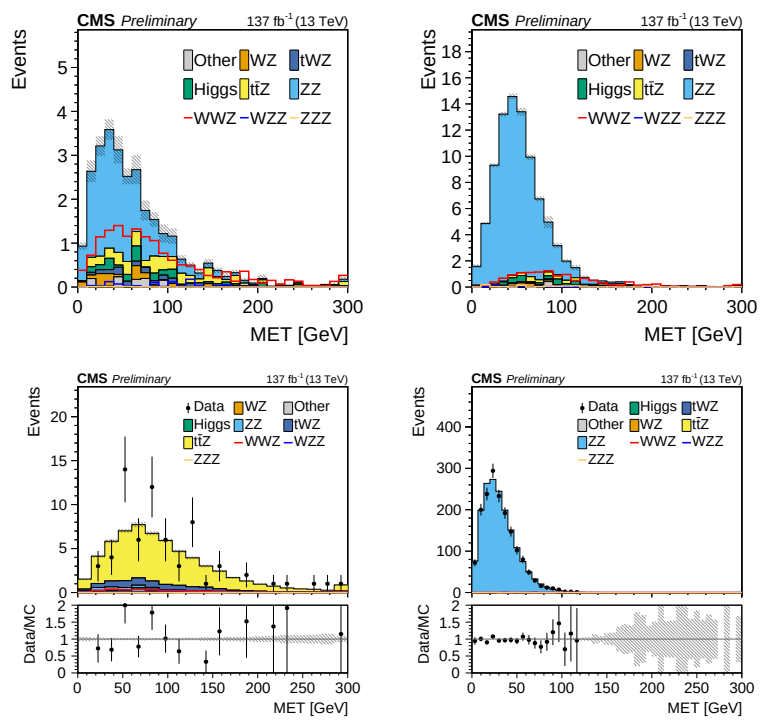


Figure 6.16: The $p_{\text{T}}^{\text{miss}}$ variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in only the ZZ trainings.

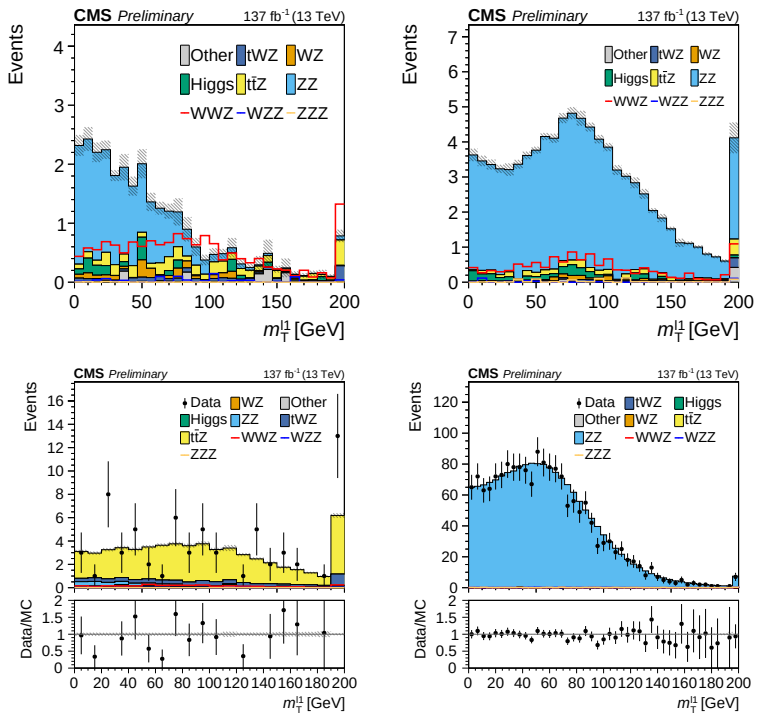


Figure 6.17: The $m_T^{\ell 3}$ variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in only the ZZ trainings.

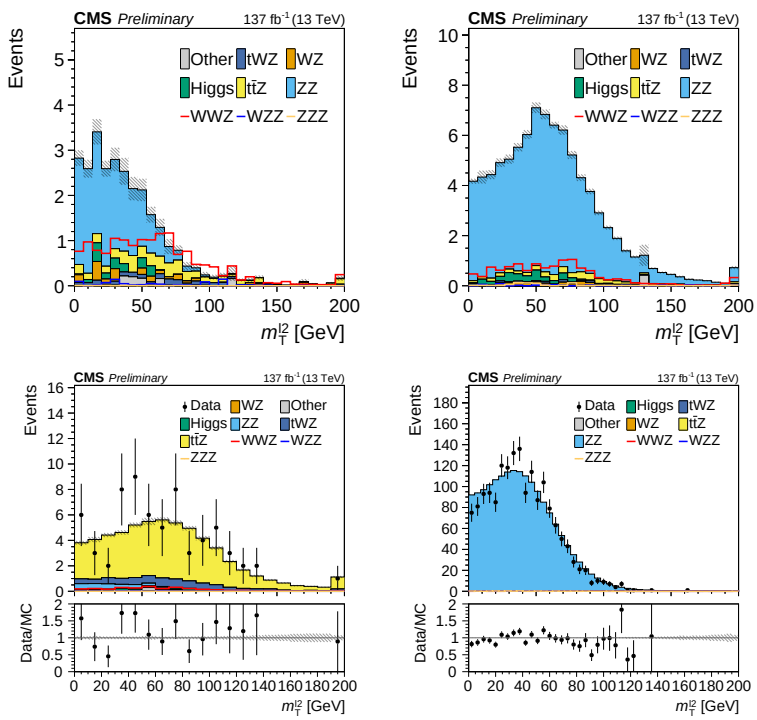


Figure 6.18: The $m_T^{\ell 4}$ variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in only the ZZ trainings.

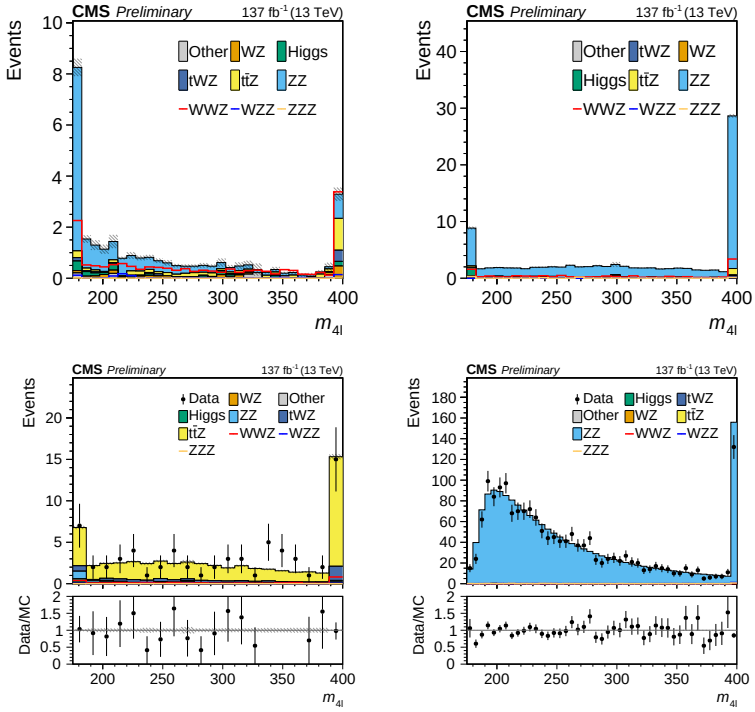


Figure 6.19: The $m_{4\ell}$ variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in only the ZZ trainings.

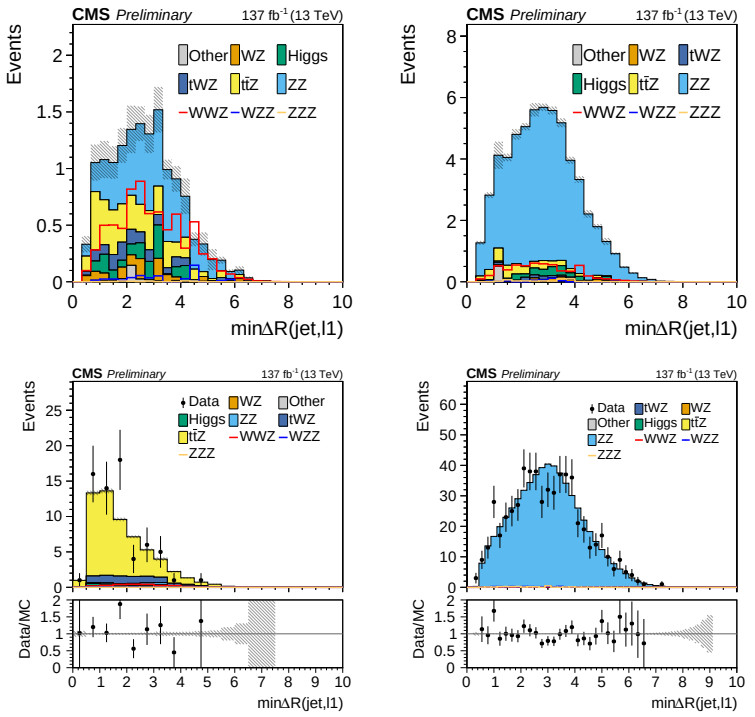


Figure 6.20: The $\min \Delta R(j, \ell 3)$ variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in only the $t\bar{t}Z$ training.

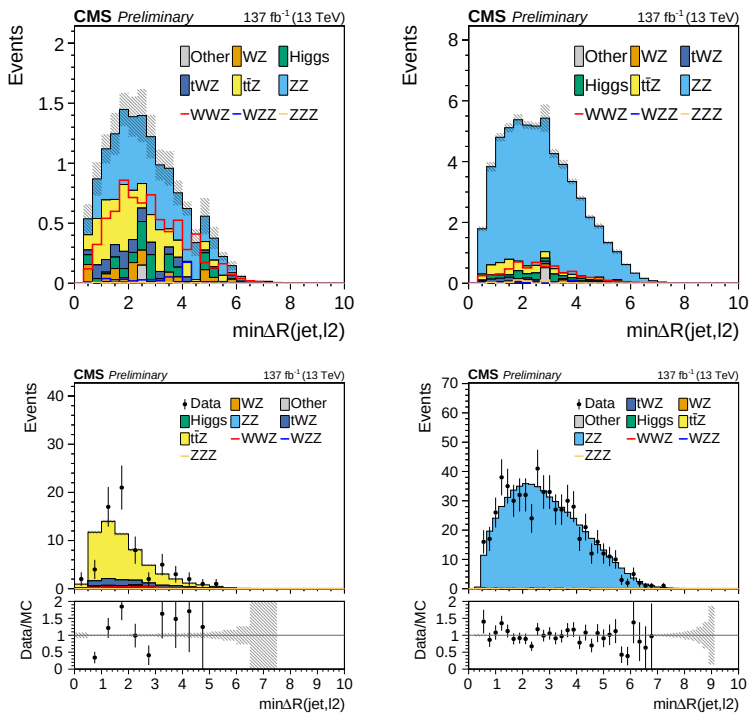


Figure 6.21: The $\min \Delta R(j, \ell 4)$ variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in only the $t\bar{t}Z$ training.

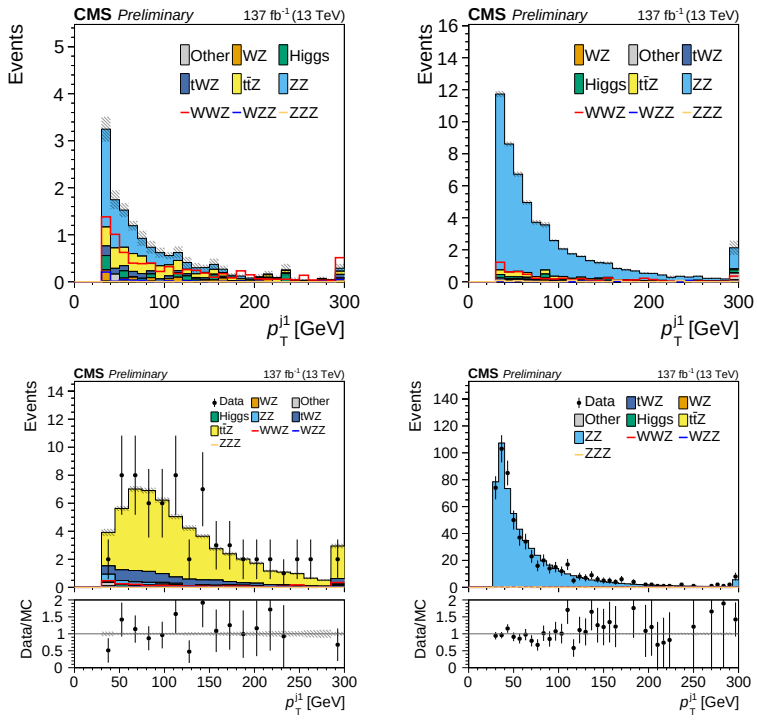


Figure 6.22: The leading jet p_T variable in the $e\mu$ region (top left), the off-Z region (top right), the $e\mu$ region with b jets (bottom left) and the on-Z region (bottom right). This variable is used in only the $t\bar{t}Z$ training.

6.6.2 BDT training

This section first explains some generalities about the BDT training in this analysis. Then, we will detail the three distinct BDTs for ZZ and $t\bar{t}Z$ separation in the $e\mu$ category and ZZ discrimination in the $ee/\mu\mu$ category.

The BDTs were trained with the simulated samples of all three years combined. As detailed in Section 4.3, it might make sense from a physics point of view to weight the events with the proper MC event weights, but this is not always the best choice to reach maximum performance. The reason is that every reweighting comes with a reduction of effective sample size, which enhances overtraining effects. The experience with this analysis was motivating this conclusion. It was observed that if the events are correctly weighted for the training, the performance on weighted MC events was worse than when all events were weighted equally for the training. This can be explained by the substantial reduction of effective sample size when the three years are combined.

For example, the 2016 on-shell WWZ signal sample has much less four-lepton events than the samples for the two other years. There is no dedicated four-lepton filtered sample with a large number of events for 2016. Consequently, the 2016 simulated signal events weights are scaled up by a large factor, heavily reducing the effective sample size. The differences between the years are not significant enough to justify this reduction.

Even though there were no event-by-event weights applied for the training, the signal and background were balanced with per-class weights. This allows for the interpretation of the BDT scores as probabilities under the assumption of equal prior probabilities, which will be useful later on when discussing likelihood-free inference techniques in Section 6.7. The number of MC events in each sample that was part of the training is shown in Table 6.8. Two-thirds of each sample was used for the training and one-third for the testing. The splitting in training and testing datasets was not done randomly but based on the event index of the simulated event. This event index is, in this context, a pseudorandom number that can be very easily reproduced. The testing uses events with an index divisible by three. The analysis steps that use the BDTs scores – such as histograms and yield tables –, never consider the training events. Accordingly, the weights of the remaining events are scaled up by a factor of three.

For the signal class, it is important to consider that the samples are on-shell WWZ and ZH $\rightarrow$ WWZ samples stitched together. The populations within each class were weighted according to the theoretical cross-sections (as in Table 1.1) to not cause a bias towards any signal population.

Very few events remain of the $t\bar{t}Z$ simulated sample in the $e\mu$ category. Therefore, the training also uses $t\bar{t}Z$ events that fell into the category $e\mu$ with b jets. As no b-tagging information is available to the BDT, the feature distributions are not significantly different in this added phase space. It was initially considered to take the b-tagging score from the DeepCSV algorithm directly as an input variable. However, the differential DeepCSV score distributions do not agree well in data and simulation. Although it is possible to recalibrate the b-tagging score[4], it would still be challenging to monitor the agreement in a separate control region convincingly. Hence the decision to keep the b jet selection as a separate cut, preserving the $e\mu$ control region with b jets.

Even when $t\bar{t}Z$ events with b jets are included in the training, a quick glance at Table 6.8 makes it clear that the training sample size for the background classes in the

[4]The calibration of the differential b-tagging score was introduced in the search for the associated production of the Higgs boson with a top-quark pair [132].

Sample	$e\mu$ region		$ee/\mu\mu$ region	
	train	test	train	test
WWZ	300,000	150,000	280,000	140,000
ZZ	18,000	7,000	810,000	405,000
$t\bar{t}Z$	25,000 (inkl. b tag $e\mu$)	≈1000	-	-

Table 6.8: Number of MC events used for the BDT training and testing in the $e\mu$ and $ee/\mu\mu$ categories. For the $t\bar{t}Z$ BDT in the $e\mu$ region, the difference between the training and testing sample sizes is not the usual factor of two because, for the training, events with b jets are included as well. The performance evaluation is done on the $e\mu$ events in the signal region with no b jets only.

more important $e\mu$ region is very limited. Just like for the electron MVA optimization, the XGBoost algorithm explained in Section 4.2 was used to train the decision tree ensemble. This makes a systematic hyperparameter optimization very challenging because the dependence of classification performance on the hyperparameters is hard to see when it is dominated by statistical fluctuations[5]. Therefore, only the maximum tree depth, learning rate, and L2 regularization parameter λ were optimized with a coarse grid search. While the much larger sample size in the $ee/\mu\mu$ region would, in principle, allow for more rigorous hyperparameter optimization, doing so was not seen as a good investment because the $ee/\mu\mu$ category plays a minor role in the analysis. The final training hyperparameters are listed in Table 6.9. Figure 6.23 shows the training- and testing-sample ROC curves for all three BDTs.

Hyperparameter	$e\mu$ region		$ee/\mu\mu$ region
	$t\bar{t}Z$ BDT	ZZ BDT	ZZ BDT
`max_depth`	3	3	5
`learning_rate`	0.1	0.1	0.1
`n_estimators`	400	400	400
`reg_lambda`	1.0	1.0	1.0
`scale_pos_weight`	$n_{\text{neg}}/n_{\text{pos}}$	$n_{\text{neg}}/n_{\text{pos}}$	$n_{\text{neg}}/n_{\text{pos}}$

Table 6.9: The hyperparameters for the three XGBoost models trained for the WWZ analysis. These parameters were found with a coarse grid search. For the regularization parameter `reg_lambda`, the best found value was 1.0, the default value for XGBoost. An explanation of the hyperparameters is given in Section 4.2.2.

Figure 6.24 shows the BDT score distributions for signal and background in training and testing samples, confirming the relatively small overtraining also visible in the ROC curves. One can see that both signal and background have two different populations, which are due to the leading jet p_{T} among the input variables that also encodes whether there was a selected jet in the event or not. The more signal-like population is the one without jets, especially for the $e\mu$ $t\bar{t}Z$ BDT. Finally, Figure 6.25 shows the BDT score distributions with the correct MC event weights for all signal and background samples used in the analysis, confirming the discrimination power of the BDT classifiers.

After the BDTs are trained, it is interesting to diagnose the importance of the models'

[5]The statistical fluctuations of classifier performance were measured as the uncertainty of the mean from k-fold cross-validation (see Section 4.1.1 for an explanation of k-fold cross-validation).

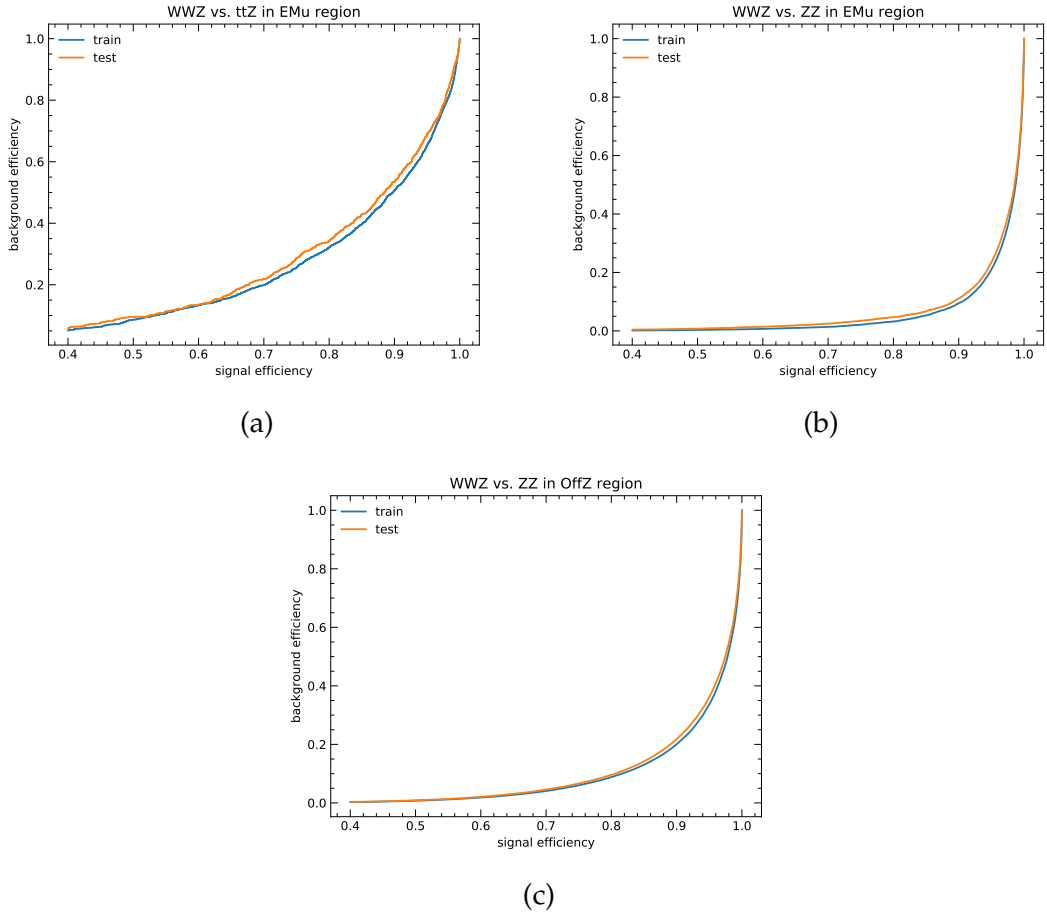


Figure 6.23: ROC curve corresponding to the four-lepton BDT cuts for the $t\bar{t}Z$ BDT (a) and ZZ BDT (b) in the $e\mu$ region, as well as for the ZZ BDT in the $ee/\mu\mu$ region (c).

different features. These studies are useful for identifying the input variables that play the most significant role in background discrimination, contain redundant information, or have little discrimination power. The XGBoost library can compute some feature importance indicators based on the tree-structure of the finished model. One frequently used simple indicator is the feature *weight*, which corresponds to how often a given feature was used for a split decision in the decision trees. This feature weight is shown in Figure 6.26a for all thre BDTs.

However, feature importance indicators based solely on the final decision tree – like the feature weight explained before – are always a compromise because it is challenging to quantify how much exploiting the correlations between features benefits the model. For example, a given feature might not have much discrimination power on its own, but it might help to partition the phase space so that some other feature can be exploited more efficiently. Hence, one of the best ways to study the impact that a feature has on a model are *n-minus-one* studies, training as many models as features, but each of them with a different feature left out[6]. The decrease in a given test performance metric caused by leaving a feature out is indicating the feature importance. The result of such a study with the AUC as the performance metric is shown in Figure 6.26b.

One caveat for such n-minus-one performance metrics is that powerful features might be ranked very low if they can be deduced from the remaining less powerful features.

[6] Another method that can be applied to the fitted estimator a posteriori is the *permutation feature importance*, where the values of a given feature are shuffled to decorrelate them from the target.

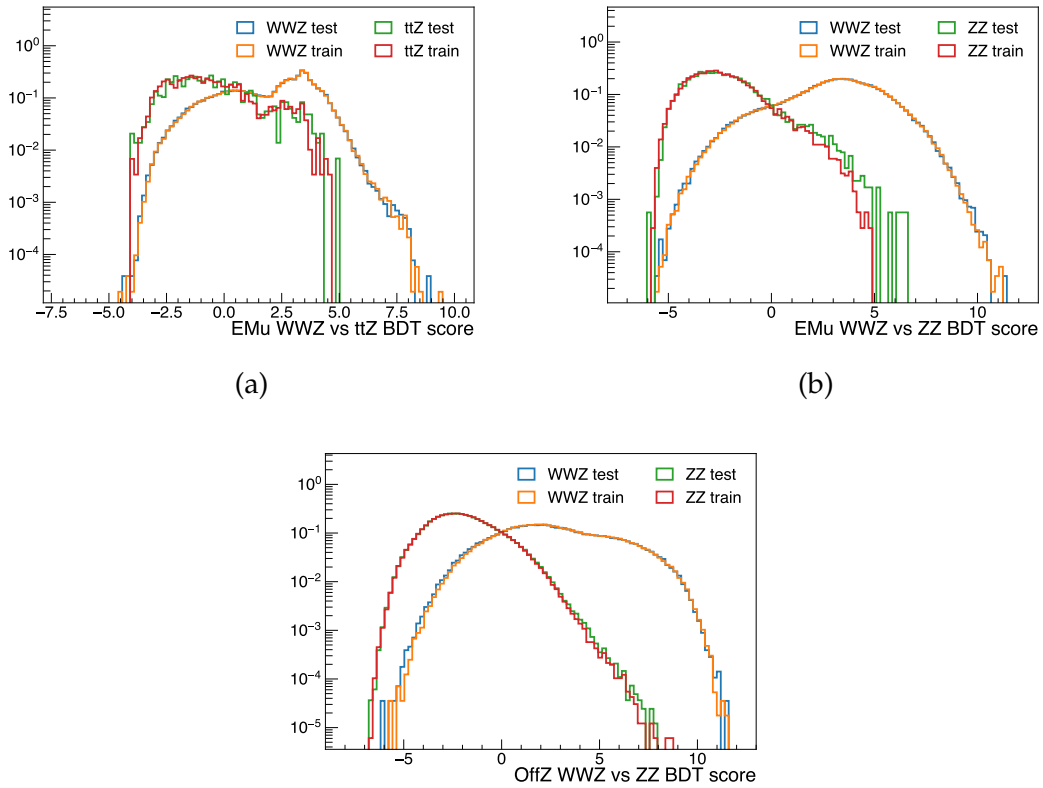


Figure 6.24: Output of the $t\bar{t}Z$ $e\mu$ BDT (a), ZZ $e\mu$ BDT (b) and ZZ $ee/\mu\mu$ (c) BDT for both the training and testing samples. All events are weighted equally.

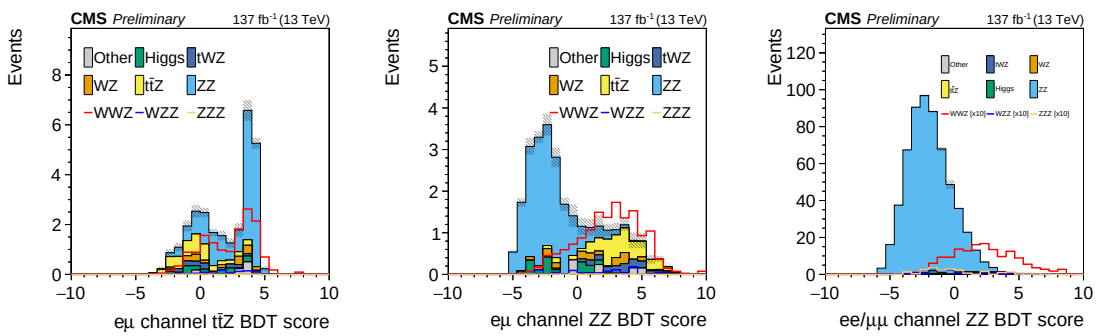
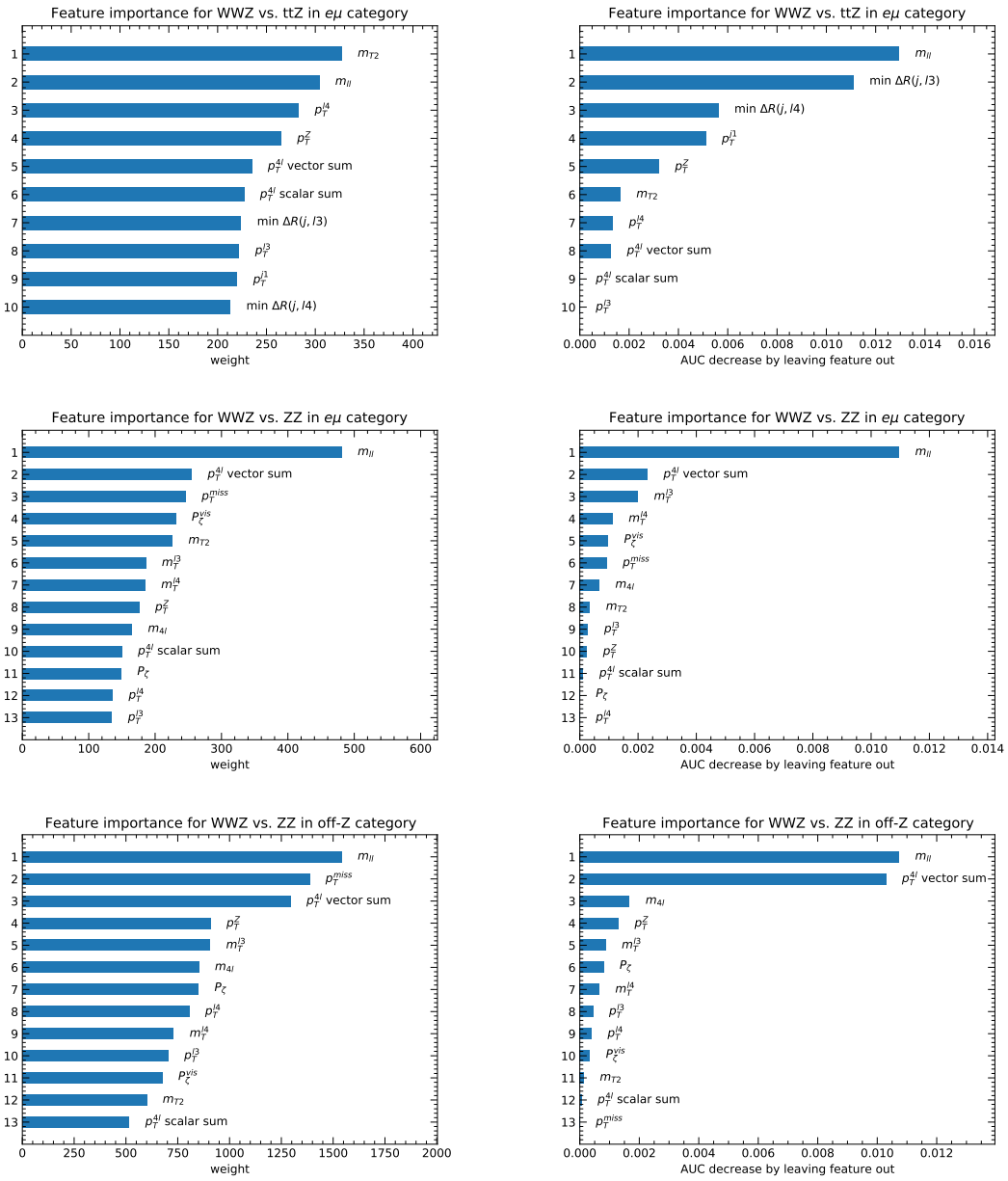


Figure 6.25: Four-lepton BDT response for signal and backgrounds in testing samples. Left: $t\bar{t}Z$ BDT; middle: ZZ $e\mu$ BDT; right: ZZ $ee/\mu\mu$ BDT.

This is nicely demonstrated by the $p_{\mathrm{T}}^{\mathrm{miss}}$ and $p_{\mathrm{T},4\ell}$ features used by the ZZ BDT in the $ee/\mu\mu$ off-Z category. Besides the $m_{\ell\ell}$ variable, one would expect $p_{\mathrm{T}}^{\mathrm{miss}}$ to appear at the top of the ranking, as the ZZ background has no intrinsic $p_{\mathrm{T}}^{\mathrm{miss}}$. However, $p_{\mathrm{T}}^{\mathrm{miss}}$ takes the last place in the n-minus-one study, presumably because the information can be easily replaced by $p_{\mathrm{T},4\ell}$, which can be seen as a MET estimate under the assumption that the four leptons are the only particles in the final state. Still, the $p_{\mathrm{T},4\ell}$ variable also seems to contain other important information because different from $p_{\mathrm{T}}^{\mathrm{miss}}$, it is ranked very high by the n-minus-one study. On the other side, the feature importance weight ranks both variables very high, confirming that missing energy is the primary discriminator against the ZZ background.

For the two $e\mu$ BDTs, the n-minus-one study confirms what was expected. For the ZZ background, the transverse masses and P_{ζ}^{vis} variables, sensitive to tau decays, play a large role. For the $t\bar{t}Z$ background, the distances between the W candidate leptons and their closest jets and the leading jet p_{T} are of substantial importance.



(a) Feature importance measured by the number of cuts on a given feature in the desicion tree ensemble, also called the feature *weight*.

(b) Feature importance measured by the decrease of the area under the ROC curve (AUC) if a given feature is left out in the training.

Figure 6.26: Feature importance ranking for the three BDTs of the four-lepton analysis: The $t\bar{t}Z$ BDT in the $e\mu$ category (top), the ZZ BDT in the $e\mu$ category (middle), and the ZZ BDT in the off-Z category (bottom). The left column (a) shows the feature importance measured with the number of cuts on a given feature, while the right column (b) shows how much the AUC would increase if a given feature is left out.

6.7 Signal extraction with BDT classifiers

6.7.1 Introduction

We now have multiple classifiers at our disposal that we can use to extract the signal from the $e\mu$ category and the $ee/\mu\mu$ category. There are several signal extraction approaches that can be grouped in *binned* and *unbinned* likelihood fits. For binned fits, the BDT score is histogrammed, resulting in a multi-bin counting experiment. This has the advantage that the likelihood for a certain number of events in a given bin follows the well-known Poisson distribution and opens up plenty of opportunities for statistically sound background estimation. One can build the full likelihood function as the product of the individual bins' Poisson likelihoods. Alternatively, one can attempt a more sophisticated *shape analysis*, where the likelihood function instead contains penalty terms that allows the differential distributions to be varied away from the shapes predicted via histograms from the simulation. For both counting experiments and shape analyses, the likelihood function has additional terms for the overall normalization.

In unbinned fits, an estimate for the probability density is directly obtained from simulation instead of predicting the outcome of a counting experiment in multiple bins. For example, one could fit the predicted histograms with analytic functions, giving an approximate closed form for the likelihood. However, handling the systematic uncertainties becomes more difficult, not to mention the additional systematic uncertainties in the extraction of such an analytic likelihood model. Therefore, unbinned fits are used less frequently, usually in situations where one cannot extract histograms from the simulation. The typical example is a parametrized search, where simulated data is only available for a few points in the parameter space. Here, it does make sense to fit the histograms for the available parameter values and then interpolate the fit parameters to obtain a probability density function for each point in the search parameter space.

Unbinned fitting methods recently got more attention from the HEP community because of their connection to machine learning. As machine learning classifiers are trained to output probabilities, wanting to use these predicted probabilities to build the likelihood function directly is a natural line of thought. In the literature, such methods are often mentioned as techniques of *likelihood-free* inference [133]. However, such methods would bring the complications of machine learning classifiers – such as overtraining and challenges with the handling of systematic uncertainties – up to the final likelihood-fitting step in the physics analysis. For that reason, likelihood-free inference directly with classification scores is generally not practiced.

This section discusses how likelihood-free inference techniques can be used to estimate the maximal signal significance obtainable with the BDTs. On the way, we will introduce new methods to diagnose the calibration of classifier scores and recalibrate them if necessary. As there is no standard way of treating systematic uncertainties in this likelihood-free inference framework, the final signal extraction in the analysis relies on a binned likelihood fit. At the end of this section, it will be explained how the binnings for the fit were in both the $e\mu$ and $ee/\mu\mu$ region before listing the expected event yields in these bins.

6.7.2 Classifier calibration diagnostics

To use likelihood-free inference methods for the estimation of the maximum possible significance, the BDT scores must be well-calibrated. This means the classifier score

should actually *be* the log of the likelihood ratio, given the classifier score as the only source of information. However, there are several reasons why this might not be the case. First of all, overtraining effects bias the score towards more extreme likelihood ratios. More specific to collider physics analysis, we might have a bias from nonuniform event weights because the weights were not considered in the training process not to lose statistical power. Finally, for the WWZ versus $t\bar{t}Z$ BDT in the $e\mu$ category, there is no reason to expect an accurate calibration to begin with because the training also considered simulated $t\bar{t}Z$ events with b jets. Applied to the $e\mu$ category, which required zero b jets, the likelihood ratio might not be well represented.

To diagnose classifier calibration, one would optimally use the true likelihood ratio. However, since the actual likelihood ratio is untractable, one has to resort to empirical distributions. Getting empirical probability density functions is not trivial. One could, for example, use histograms or a Kernel Density Estimate (KDE). A comparison of the BDT scores in the $e\mu$ region with the empirical likelihood ratio obtained by dividing histograms is shown in Figure 6.27. Histograms and KDEs require additional parameters, such as binning parametrization or kernel widths. The statistical uncertainties of histograms make it challenging to conclude on the quality of the calibration by eye, and further statistical tests are necessary. Another problem with histograms is the averaging within bins, which causes biases commonly known as *binning effects*. This is why empirical checks preferably make use of the empirical *cumulative* distribution function (CDF), which is simply a step function with a step at every observed value. The most famous example of a test that uses the empirical CDF is the two-sample Kolmogorov-Smirnov (KS) test. Another example is the ROC curve, which are the empirical CDFs of the positive and negative class plotted against each other.

In the same spirit as the KS test or the ROC curve, we can diagnose the classifier calibration based on the empirical CDF. Let x be the BDT score and $f(x)$ and $g(x)$ the true probability densities as a function of the BDT score for the positive and negative class. We would like to validate

$$x = \frac{f(x)}{f(x) + g(x)}. \tag{6.7}$$

Naively replacing the probability density functions by the corresponding CDFs – which we shall name $F(x)$ and $G(x)$ – is not meaningful. Instead, we should try to rewrite the equation and then integrate both sides from 0 to x:

$$f(x) = x\left(f(x) + g(x)\right). \tag{6.8}$$

Indefinite and definite integrals of $xf(x)$ can be rewritten in terms of $F(x)$ using integration by parts:

$$\int xf(x)dx = x\int f(x)dx - \int\int f(x)dx, \tag{6.9}$$

$$\int_0^x x'f(x')dx' = x\int_0^x f(x')dx - \int_0^x\int_0^x f(x')dx' = xF(x) - \int_0^x F(x')dx'. \tag{6.10}$$

With Equation 6.10, we can integrate both sides of Equation 6.8:

$$\begin{aligned} F(x) &= xF(x) - \int_0^x F(x')dx' + xG(x) - \int_0^x G(x')dx', \\ F(x) &= x\left(F(x) + G(x)\right) - \int_0^x \left(F(x') + G(x')\right)dx'. \end{aligned} \tag{6.11}$$

Finally, we can rewrite Equation 6.11 to obtain something reminiscent of Equation 6.7, but only with the CDFs, which are easy to obtain empirically:

$$x = \frac{F(x) + \int_0^x \left(F(x') + G(x')\right) dx'}{F(x) + G(x)}. \tag{6.12}$$

Alternatively, one can use the survival functions $\tilde{F}(x) = 1 - F(x)$ and the corresponding $\tilde{G}$ instead of the CDFs. This results in a slightly different relation:

$$x = \frac{\tilde{F}(x) - \int_x^1 \left(\tilde{F}(x') + \tilde{G}(x')\right) dx'}{\tilde{F}(x) + \tilde{G}(x)}. \tag{6.13}$$

If the BDT score is not well calibrated, the right-hand sides of Equations 6.12 and 6.13 are not equivalent. For the calibration check, we can plot the right-hand side of any of the two equations against the BDT score. In any case, the BDT is well calibrated if the curve is a smooth diagonal. Equation 6.12 is generally smoother for x closer to 1, while Equation 6.13 behaves better closer to $x = 0$. Therefore, for the empirical test, both equations are stitched together to obtain a smooth curve for the full range of BDT scores. These curves are shown in Figure 6.28 for the $t\bar{t}Z$ and ZZ BDTs in the $e\mu$ category. For the empirical CDFs, the correct event weights were used.

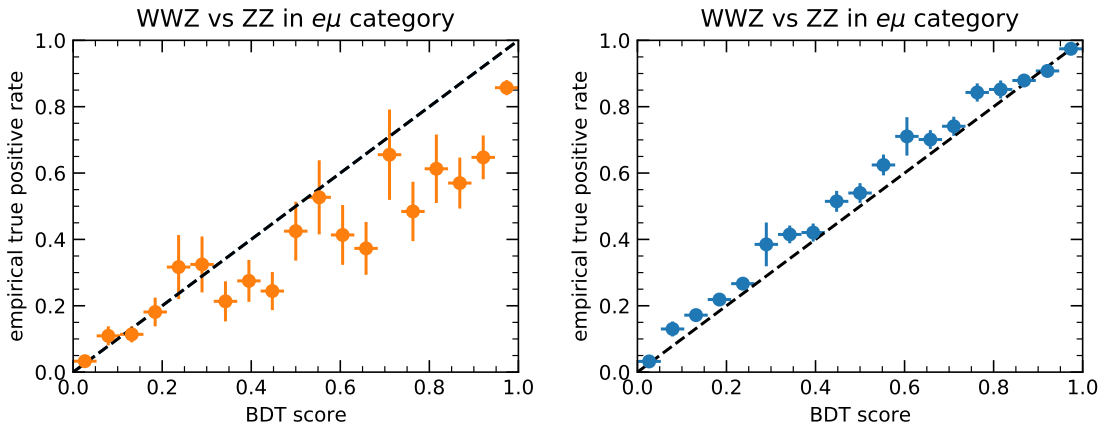
Figure 6.28 confirms what could be already anticipated from the histogram ratios in Figure 6.27. The $t\bar{t}Z$ BDT is not well calibrated because of the large overtraining due to the relatively small $t\bar{t}Z$ sample size, and because $t\bar{t}Z$ events with b jets were used for the training as well. The CDF-based check also reveals a slight miscalibration of the ZZ BDT. It was checked that this is due to the BDT training with uniform weights instead of using the proper event weights. If the calibration curve is produced with uniform weights, it aligns perfectly with the diagonal.

The calibration check makes it clear that we cannot proceed with any likelihood-free inference techniques before recalibrating the BDT scores. It was also observed that the $t\bar{t}Z$ calibration does not improve by using $t\bar{t}Z$ events with no b jets only. The approximately thousand remaining $t\bar{t}Z$ events are not sufficient to suppress overtraining, and the calibration gets even worse.

6.7.3 Classifier recalibration

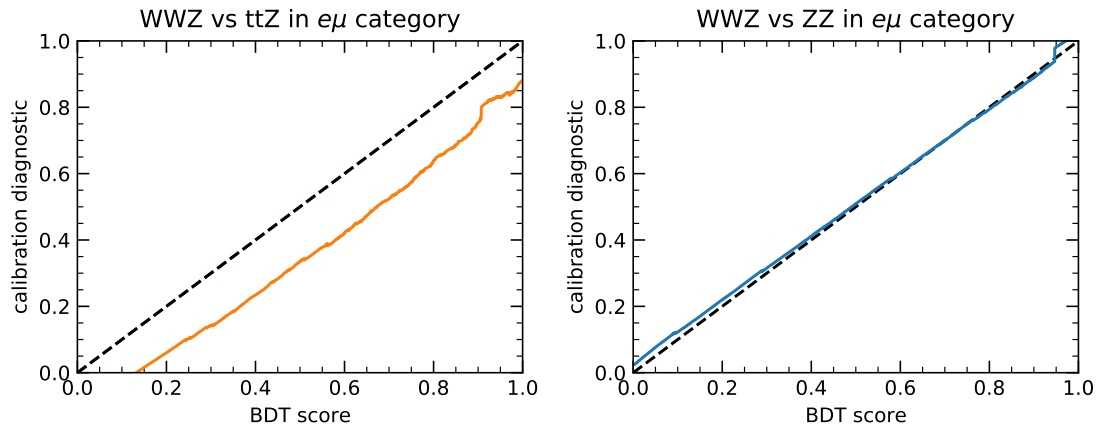
In the previous section, we have seen that it is not possible to estimate the maximum possible signal significance in the $e\mu$ region with likelihood-free inference methods without recalibrating the BDT scores. In the HEP-ML literature, classifier recalibration for likelihood-free inference is currently studied actively [134], but there is no standard procedure. Isotonic regression is an often-mentioned technique because it fits arbitrary functions under the constraint that they are monotonically increasing, as is the case for the signal probabilities. Recalibration is essentially a univariate regression problem, so even simple neural networks were suggested for this purpose.

However, for this study, it was preferred not to use another machine learning method. One would once again have to split the sample to have separate training data for the recalibration, further reducing the number of precious simulated $t\bar{t}Z$ events. Instead, we can start from the assumption that classifiers trained with logistic loss functions are generally well-calibrated, and we can think of how this calibration might be distorted. Overtraining was already mentioned as a significant effect. If a classifier is overtrained,



(a) $t\bar{t}Z$ BDT (b) ZZ BDT

Figure 6.27: Histogram-based calibration curves for the two BTDs in the $e\mu$ region ($t\bar{t}Z$ (a) and ZZ (b) BDT). The true positive rate is obtained by dividing the signal histogram by the background histogram, with the uncertainty estimated with the normal approximation implemented in the `TEfficiency` class in ROOT, as this option supports weighted histograms.



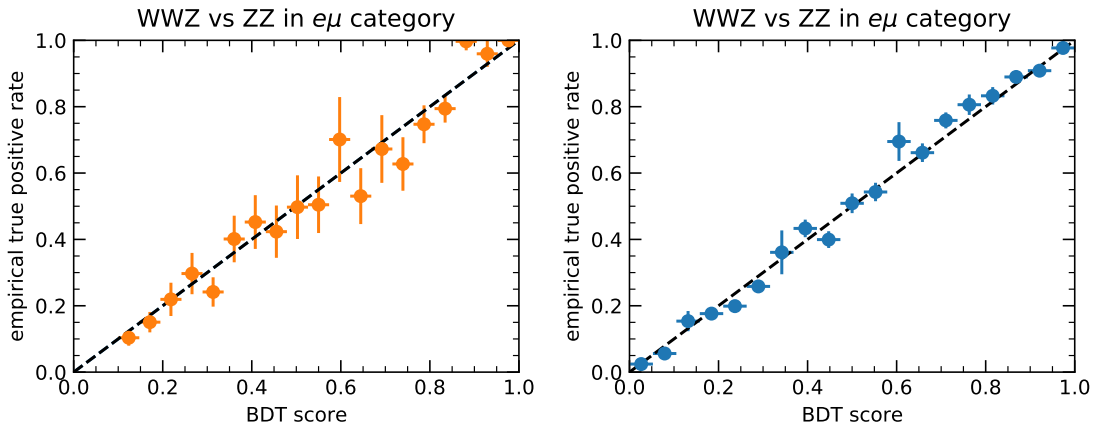
(a) $t\bar{t}Z$ BDT (b) ZZ BDT

Figure 6.28: Calibration curves obtained with the CDF-based test described in Section 6.7.2 for the two BTDs in the $e\mu$ region ($t\bar{t}Z$ (a) and ZZ (b) BDT).

the predicted log-odds are further away from zero because the classifier is too confident. Furthermore, there might be a bias in the predicted log-odds because the classifier is trained on a set of samples that qualitatively differs from the sample we want to apply it to, which is the case for the $t\bar{t}Z$ BDT in particular. Hence, we can imagine that at first order, the miscalibration has a linear effect on the predicted log-odds *before* the logistic transformation that brings the BDT score in the probability interval from 0 to 1.

To counter the linear miscalibration of the log-odds, we can attempt to find a linear transformation $t(z) = A \cdot (z - m)$ of the predicted log-odds z to get well-calibrated probabilities $x_{\text{calib.}}$ after the logistic transformation:

$$x_{\text{calib.}}(z) = \frac{1}{1 + \exp\Big(-\big(A \cdot (z - m)\big)\Big)}. \tag{6.14}$$



(a) $t\bar{t}Z$ BDT (b) ZZ BDT

Figure 6.29: Histogram-based calibration curves for the two BTDs in the $e\mu$ region ($t\bar{t}Z$ (a) and ZZ (b) BDT) after recalibration. The true positive rate is obtained by dividing the signal histogram by the background histogram, with the uncertainty estimated with the normal approximation implemented in the `TEfficiency` class in ROOT, as this option supports weighted histograms.

To find the coefficients A and m, we can use the empiric calibration diagnostic from Section 6.7.2 and find the coefficients that minimize the residue between the calibration curve and the diagonal. The obtained calibration functions for both BDTs are

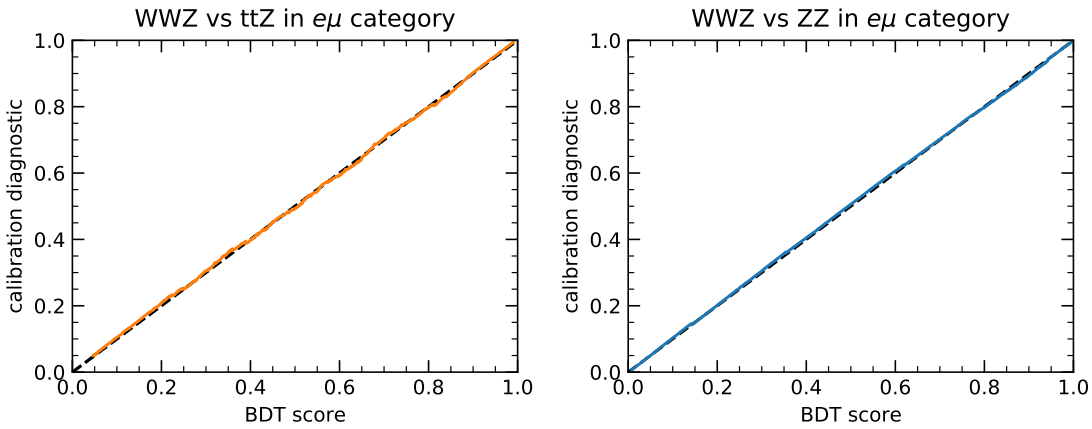
$$\begin{aligned} t_{t\bar{t}Z}(z) &= 0.5939 \cdot (z - 1.0058)\,, \\ t_{ZZ}(z) &= 0.8890 \cdot (z + 0.2238)\,. \end{aligned} \tag{6.15}$$

As expected, the scaling factor for the log-odds turns out to be smaller than one because of overtraining. Figure 6.30 shows the calibration curves after recalibration with the linear transformation of the log-odds. The calibration curves align almost perfectly with the diagonal. Figure 6.29 shows the traditional histogram-based calibration curves for comparison.

With the calibrated BDT scores, we can proceed to use likelihood-free inference methods to study the maximally obtainable signal significance. However, this calibration study represents a satisfying result on its own. The finding that the recalibration problem for logistic-loss based classifiers applied to collider physics event separation can be solved with simple linear methods might help to promote the direct interpretation of classifier scores as probabilities.

6.7.4 Ceiling study with mixture model and likelihood free inference

The previous sections discussed how the BDT scores are recalibrated to represent likelihood ratios for binary sample classification. We will now use the recalibrated scores and likelihood-free inference methods to compute an upper limit for the maximum possible WWZ signal significance obtainable with these BDTs in the $e\mu$ category. Since we have not trained BDTs for signals other than WWZ (on-shell and ZH) and backgrounds other than $t\bar{t}Z$ and ZZ, the remaining signals and backgrounds will initially be neglected in this study. The missing backgrounds are contributing more to the yields in the $e\mu$ category than the missing signal, so the upper limit we obtain will be overly optimistic.



(a) $t\bar{t}Z$ BDT (b) ZZ BDT

Figure 6.30: Calibration curves obtained with the CDF-based test described in Section 6.7.2 for the two BTDs in the $e\mu$ region ($t\bar{t}Z$ (a) and ZZ (b) BDT) after recalibration.

However, such a limit is still instrumental in assessing the efficiency of any traditional signal extraction method that we consider.

If we had only one signal and one background, it would be straight-forward to proceed to build the log-likelihood ratio surrogate:

$$l = n\left(\log(b) - \log(s+b)\right) - s + \sum_{i=0}^{n} z_{s,b}^{(i)} \ , \tag{6.16}$$

with s and b being the number of predicted signal and background events, n the number of observed events and $z_{s,b}^{(i)}$ the calibrated BDT score of event i before the logistic transformation. The terms outside the sum originate from the ratio of Poisson terms corresponding to the total number of events. Note that Equation 6.16 contains a constant term, s, that does not affect the minimum of the likelihood ratio. However, in likelihood-free inference computations, it is wise to keep all terms such that the log-likelihood ratio at the minimum should be zero. If it is not, one did not calibrate the classifiers well or made a mistake in building the likelihood surrogate. Dropping constant terms would prohibit this cross-check.

For multiple background components, we can use the formula for the likelihood ratios in mixture models [134, 135]:

$$\begin{aligned} \frac{p\left(\mathbf{x}|\theta_0\right)}{p\left(\mathbf{x}|\theta_1\right)} &= \frac{\sum_c w_c\left(\theta_0\right) p_c\left(\mathbf{x}|\theta_0\right)}{\sum_{c'} w_{c'}\left(\theta_1\right) p_{c'}\left(\mathbf{x}|\theta_1\right)} \\ &= \sum_c \left[\sum_{c'} \frac{w_{c'}\left(\theta_1\right)}{w_c\left(\theta_0\right)} \frac{p_{c'}\left(\mathbf{x}|\theta_1\right)}{p_c\left(\mathbf{x}|\theta_0\right)}\right]^{-1} \end{aligned} \tag{6.17}$$

Equation 6.17 considers the most general case. The probabilities p_c for a given component might depend on the parameters θ_0 and θ_1 of the null hypothesis and the signal hypothesis. For our application, this is not the case. Only the normalized component weights w_c depend on the hypothesis, as the weight of the signal component is zero in the null-hypothesis. Therefore, Equation 6.17 simplifies and we can replace the binary

likelihood ratios with the BDT log-odds to obtain the correct log-likelihood ratio z_{θ_0,θ_1}:

$$\exp\left(z_{\theta_0,\theta_1}\right) = \sum_c \left[\sum_{c'} \frac{w_{c'}\left(\theta_1\right)}{w_c\left(\theta_0\right)} \exp\left(z_{c,c'}(\mathbf{x})\right)\right]^{-1}. \tag{6.18}$$

By using Equation 6.18 to replace the binary log-odds $z_{s,b}$ in Equation 6.16, we obtain the correct formula for the likelihood surrogate.

We have no classifier to approximate the binary log-odds $z_{\mathrm{t\bar{t}Z,ZZ}}$ for the background components. Separation of the two background components should not be necessary since their relative weights are the same for the signal- and null-hypothesis. Naively, one could imagine setting $z_{\mathrm{t\bar{t}Z,ZZ}}$ to zero, corresponding to no discrimination between the background components. However, this results in inconsistent likelihood functions whose minimum is not at the right place. This inaccuracy was thought to be due to inconsistency in the matrix of binary log-odds. If $z_{\mathrm{WWZ,t\bar{t}Z}}$ and $z_{\mathrm{WWZ,ZZ}}$ are different, it is not consistent to assume zero separation power between the backgrounds. It *is* consistent to set $z_{\mathrm{t\bar{t}Z,ZZ}} = z_{\mathrm{WWZ,ZZ}} - z_{\mathrm{WWZ,t\bar{t}Z}}$, even though it is by no means an optimal observable to separate the backgrounds. The final consistent matrix of binary log-odds is then

$$Z = \begin{pmatrix} 0 & z_{\mathrm{WWZ,t\bar{t}Z}} & z_{\mathrm{WWZ,ZZ}} \\ -z_{\mathrm{WWZ,t\bar{t}Z}} & 0 & z_{\mathrm{WWZ,ZZ}} - z_{\mathrm{WWZ,t\bar{t}Z}} \\ -z_{\mathrm{WWZ,ZZ}} & z_{\mathrm{WWZ,t\bar{t}Z}} - z_{\mathrm{WWZ,ZZ}} & 0 \end{pmatrix}. \tag{6.19}$$

This matrix completion procedure for the backgrounds would work for an arbitrary number of background components. As far as the author is aware, the problem of consistent mixture-model likelihood ratio surrogates is not discussed in the literature so far. Note that these inconsistencies might also arise if we would have trained a classifier for the background component separation, albeit on a much smaller scale.

The likelihood-ratio surrogate was used for a scan over the signal strength. Equation 6.18 and Equation 6.16 have to be evaluated for multiple signal hypotheses, all with a different weight for the signal component. The result of the scan is shown in Figure 6.31, which also includes a signal strength scan where only the normalization term is considered, giving an idea of the signal significance in the $e\mu$ region without any further separation.

Figure 6.31 leads to several interesting conclusions, even if it only considers the WWZ signal and the two major backgrounds. The expected signal strength is $\mu_{\mathrm{WWZ}} = 1.006^{+0.356}_{-0.303}$, where the accurate representation of the central value at the permille level confirms that the binary classifiers are well-calibrated. Without the calibration described in Section 6.7.2, the central value would have been 1.036. Reading the value at $\mu_{\mathrm{WWZ}} = 0$, the expected significance is 5.04 σ. If we consider the remaining backgrounds only in the normalization term, assuming that we have no separation power at all for these events, the significance gets reduced to 4.90 σ.

Even if this likelihood-free inference cannot be carried further in the analysis because of the challenges related to systematic uncertainties and the unmodeled signals and backgrounds, it provides a very insightful ceiling to the maximum signal that could be extracted with the BDTs in the $e\mu$ category.

Understanding what the limiting factors on a full analysis are is extremely important to plan in which areas it should be improved. This section showed that classifier calibration and likelihood-free inference techniques are very useful for ceiling studies at the signal extraction level. Such ceiling studies might also be valuable for sensitivity studies of future experiments.

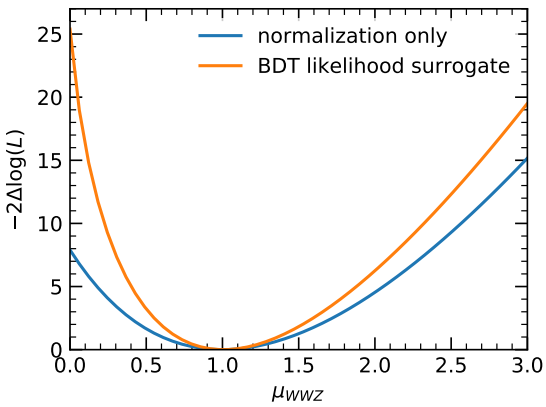


Figure 6.31: Signal strength scan with simulated events in the $e\mu$ region with the WWZ signal and the two major backgrounds $t\bar{t}Z$ and ZZ. For the orange curve, a likelihood surrogate function was built in a likelihood-free inference framework, using a mixture model to combine the binary BDT scores into an event-by-event likelihood ratio. The likelihood ratio that corresponds to only the total number of $e\mu$ events is shown in blue.

Note that the mixture model, as in Equation 6.18, can also be used to combine several calibrated binary classifiers into an optimally combined score, even if no likelihood-free inference is intended. The combined score can then used in more traditional signal extraction methods. For this application, the weights w_c for Equation 6.18 should be adapted: the signal hypothesis should be the signal only, and the null hypothesis the backgrounds only, resulting in a clean separation between signal and background instead of likelihood ratio approximation.

6.7.5 Outlook on likelihood-free inference methods

We might also imagine how systematic variations could be incorporated in such a likelihood-free inference. Usually, systematic variations are small, and the classifiers trained for the nominal values are also performant for the variations. They would still have to be recalibrated. The previous recalibration study, especially the recalibration of the ZZ BDT, demonstrated how classifiers could be recalibrated to represent the correct likelihood ratio if the correct event weights are applied, as opposed to no event weights applied during the training process. Therefore, the same method is expected to work well for recalibration after systematic variations. Since the recalibration is not computationally expensive, this would also be feasible for a large number of systematic uncertainties. The final likelihood surrogate with systematic variations would then be the envelope of the likelihood surrogates corresponding to each variation, which is equivalent to the usual profiling of systematic uncertainties.

The other challenge of such an inference method is the treatment of the remaining minor backgrounds, which need to be considered to obtain a realistic likelihood surrogate. One might train additional, simpler binary classifiers to separate them from the signal, but it should also be possible to write down a consistent mixture model that considers no discrimination between the signal and the other backgrounds. To conclude, there is a clear research path to extend, test, and validate such likelihood-free inference, which might be followed in future work.

6.7.6 Signal extraction with multidimensional histograms

The previous section showed that no matter what signal extraction method based on the BDTs in the $e\mu$ region will be used, the WWZ signal significance can not exceed 4.90 standard deviations. For the final analysis that has to consider systematic variations, the likelihood-free inference method was not used because it is unconventional. Still, it was considered to use the mixture model to combine the two binary classifiers to an optimal discriminant that is subject to a binned likelihood fit. However, the expected significance obtainable from such a combined score with a fixed number of bins was only marginally larger than the significance reached with a 2D histogram of the two BDT scores. Therefore, the more traditional and easier approach of histogramming the individual BDT scores and then performing a binned likelihood fit was followed.

In the case of two BDTs, such as in the $e\mu$ region, a common approach is to cut on the more discriminating BDT and then perform a Poisson likelihood fit – or shape analysis – based on histograms of the remaining BDT score. However, the $e\mu$ category is already relatively pure, and it is preferable to use the full category for signal extraction. Therefore, a 2D histogram approach was employed. A series of 2D boundaries have been defined to maximize sensitivity. The binning was automatically chosen by creating a simple decision tree that stops splitting when any of the following conditions based on the bins after a candidate split is met:

- relative uncertainty $1/\sqrt{s}$ for signal raises above 2 %
- relative uncertainty $1/\sqrt{b}$ for background raises above 8 %
- the s/b in the bins differs by less than 50 %
- less than one expected event in a bin

These criteria are not met anymore after a few splits, resulting in a coarse binning where the number of events in each bin is large enough for data-driven background estimation techniques. The distribution of simulated events in the 2D BDT score plane is indicated in Figure 6.32 together with the bin boundaries found by the decision tree algorithm.

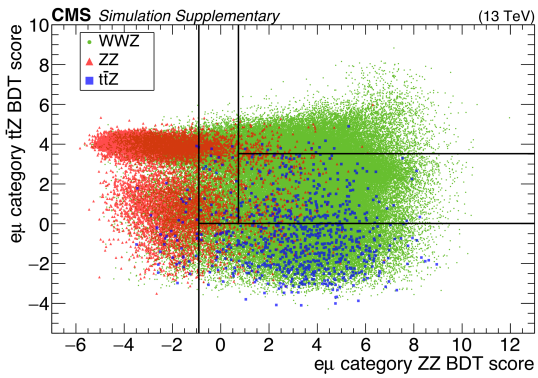


Figure 6.32: Distribution of the scores of the $t\bar{t}Z$ and ZZ BDTs in the $e\mu$ category for simulated WWZ, ZZ and $t\bar{t}Z$ events. The boundaries of the five bins used for the signal extraction are indicated as black lines.

Signal extraction in the less sensitive $ee/\mu\mu$ is less challenging given only one dominant background and one corresponding BDT discriminant. In this one-dimensional case, it is possible to make a brute-force scan for the split that results in the best combined

approximate median significance (AMS) [136] of the two resulting bins:

$$\mathrm{AMS}_2 = \sqrt{2\left((s+b)\ln\left(1+\frac{s}{b}\right)-s\right)}. \tag{6.20}$$

For the binning optimization in the $ee/\mu\mu$ category, the splits for the binning were found with a greedy iterative algorithm. The cut at which the combined significance peaks will be a first bin boundary. In the next step, the significance scan is repeated with the previous split retained for the evaluation of the combined significance. This is repeated until the improvement of the combined significance becomes negligible in the picture of the full analysis. The result of the scan is shown in Figure 6.33. Based on this scan, it was decided only to split the $ee/\mu\mu$ category at the ZZ BDT score of 3.0 and not do any further split based on the significance scan. Instead, the first bin was decided to have a lower bound at 0.0, with the events below not used for the signal extraction. In this way, it was possible to use these ZZ dominated events to constrain the normalization of the ZZ background, as described in Section 6.8.2.

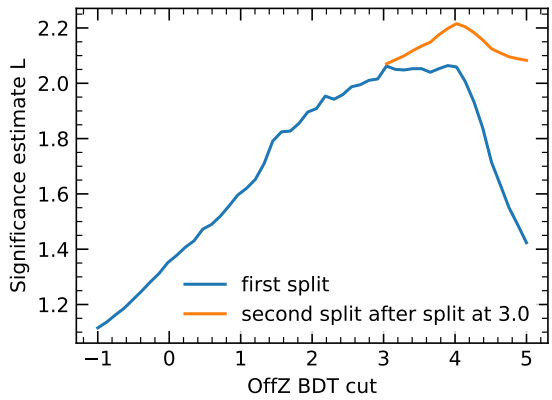


Figure 6.33: Scanning the $ee/\mu\mu$ category BDT score for the cut with maximum significance after the split.

The final binning for the signal extraction in the WWZ analysis is given in Table 6.10. The expected event yields for all signals and backgrounds in all bins are listed in Table 6.11 and Table 6.12. Finally, Table 6.13 shows the significance estimates based exclusively on simulated events in the individual bins for the signal extraction and their combination. It is particularly interesting to compare the combined 4.33 σ expected in the $e\mu$ region with the upper bound of 4.90 σ that was found with the ceiling study in the previous section. The difference corresponds to a reduction of sensitivity by the information loss in the histogram-based signal extraction method by about 10 %. In conclusion, it would be possible to increase the precision of future WWZ analyses by using a more sophisticated signal extraction method, but not by as much as needed to make this a priority.

6.8 Data-driven background estimation

In the previous section, a signal extraction scheme was optimized based on the WWZ signal and the two major backgrounds, ZZ and $t\bar{t}Z$. The cross-section of the ZZ background is roughly three orders of magnitude above the signal, while the $t\bar{t}Z$ cross-section

	ZZ BDT range	$t\bar{t}Z$ BDT range
$e\mu$ BDT bin 1	(-∞, -0.908)	(-∞ , ∞)
$e\mu$ BDT bin 2	(-0.908 , ∞)	(-∞ , 0.015)
$e\mu$ BDT bin 3	(-0.908 , 0.733)	(0.015 , ∞)
$e\mu$ BDT bin 4	(0.733 , ∞)	(0.015 , 3.523)
$e\mu$ BDT bin 5	(0.733 , ∞)	(3.523 , ∞)
$ee/\mu\mu$ BDT bin A	(0 , 3)	-
$ee/\mu\mu$ BDT bin B	(3 , ∞)	-

Table 6.10: Summary of BDT binning boundaries in the WWZ analysis.

Category	Summary			
	N_{bkg}	N_{signal}		
		N_{WWZ}	N_{WZZ}	N_{ZZZ}
$e\mu$ total	27.55 ± 0.78	14.63 ± 0.53	0.80 ± 0.20	0.0577 ± 0.0085
Bin 1	16.32 ± 0.47	1.576 ± 0.092	0.164 ± 0.076	0.0151 ± 0.0053
Bin 2	2.46 ± 0.32	1.70 ± 0.12	0.112 ± 0.087	0.0042 ± 0.0024
Bin 3	4.72 ± 0.36	2.19 ± 0.23	0.093 ± 0.032	0.0164 ± 0.0038
Bin 4	3.53 ± 0.38	5.76 ± 0.40	0.40 ± 0.16	0.015 ± 0.0034
Bin 5	0.51 ± 0.12	3.40 ± 0.22	0.028 ± 0.020	0.0069 ± 0.0035
$ee/\mu\mu$ total	86.98 ± 0.77	12.00 ± 0.56	0.28 ± 0.16	0.58 ± 0.28
Bin A	80.67 ± 0.72	6.53 ± 0.37	0.21 ± 0.15	0.35 ± 0.24
Bin B	6.30 ± 0.27	5.46 ± 0.41	0.077 ± 0.040	0.23 ± 0.15

Table 6.11: Expected number of signal events in the $e\mu$ category, estimated by MC for full Run 2 data set corresponding to 137 fb^{-1}.

Category	Composition of N_{bkg}					
	N_{ZZ}	$N_{t\bar{t}Z}$	N_{tWZ}	N_{WZ}	N_{Higgs}	N_{Other}
$e\mu$ total	17.57 ± 0.21	4.21 ± 0.28	1.59 ± 0.09	1.59 ± 0.35	1.56 ± 0.46	1.03 ± 0.38
Bin 1	14.51 ± 0.19	0.222 ± 0.059	0.080 ± 0.020	0.46 ± 0.14	1.02 ± 0.40	0.025 ± 0.011
Bin 2	1.607 ± 0.061	0.13 ± 0.05	0.073 ± 0.019	0.19 ± 0.12	0.09 ± 0.16	0.37 ± 0.29
Bin 3	0.608 ± 0.039	2.49 ± 0.23	0.63 ± 0.057	0.51 ± 0.18	0.29 ± 0.14	0.20 ± 0.15
Bin 4	0.593 ± 0.036	1.30 ± 0.14	0.754 ± 0.062	0.31 ± 0.21	0.14 ± 0.10	0.43 ± 0.25
Bin 5	0.25 ± 0.024	0.076 ± 0.031	0.057 ± 0.016	0.11 ± 0.11	0.0057 ± 0.0029	0.0094 ± 0.0046
$ee/\mu\mu$ total	77.84 ± 0.40	3.45 ± 0.15	1.202 ± 0.077	1.188 ± 0.23	2.73 ± 0.43	0.57 ± 0.41
Bin A	75.06 ± 0.40	1.377 ± 0.092	0.5141 ± 0.051	1.00 ± 0.21	2.25 ± 0.39	0.475 ± 0.41
Bin B	2.786 ± 0.075	2.07 ± 0.12	0.688 ± 0.058	0.19 ± 0.11	0.48 ± 0.19	0.093 ± 0.075

Table 6.12: Expected number of background events in the $e\mu$ category, estimated by MC for full Run 2 data set corresponding to 137 fb^{-1}.

is approximately 50 times larger. The $t\bar{t}Z$ background is strongly reduced in the $e\mu$ category, which asks for no b jets, and it can be monitored in a relatively pure control region that is equivalent to the $e\mu$ category but with one or more b jets. The $e\mu$ category almost rejects all of the ZZ backgrounds, but a sizable number of events remains – predominantly events with different-flavor leptonic tau decays. The ZZ background is almost eliminated by design in the $e\mu$ region, although a sizable number of events still contaminates said region, mostly different-flavor leptonic tau decays. Furthermore, ZZ is the

Category	N_{signal}	N_{bkg}	AMS_2	$N_{\text{signal}}/N_{\text{bkg}}$
$e\mu$ total	15.48 ± 0.57	27.59 ± 0.78	$2.72\ \sigma$	0.56
Bin 1	1.76 ± 0.12	16.32 ± 0.47	$0.43\ \sigma$	0.11
Bin 2	1.82 ± 0.15	2.46 ± 0.32	$1.05\ \sigma$	0.74
Bin 3	2.30 ± 0.23	4.72 ± 0.36	$0.99\ \sigma$	0.49
Bin 4	6.18 ± 0.43	3.53 ± 0.37	$2.70\ \sigma$	1.75
Bin 5	3.43 ± 0.22	0.51 ± 0.12	$3.04\ \sigma$	6.70
Combined	-	-	$4.33\ \sigma$	-
$ee/\mu\mu$ total	12.86 ± 0.64	86.98 ± 0.77	$1.35\ \sigma$	0.15
Bin A	7.09 ± 0.47	80.67 ± 0.72	$0.78\ \sigma$	0.09
Bin B	5.78 ± 0.44	6.30 ± 0.27	$2.04\ \sigma$	0.92
Combined	-	-	$2.18\ \sigma$	-
$e\mu$ and $ee/\mu\mu$ Combined	-	-	$4.85\ \sigma$	-

Table 6.13: Sensitivity estimate from yields predicted purely by simulated events for BDT analysis with full Run 2 data set corresponding to 137 fb^{-1}.

dominant background in the $ee/\mu\mu$ signal region, even though the second same-flavor lepton pair is required to be off the Z peak. If the second lepton pair is on the Z peak, we can get an almost entirely pure ZZ control region.

After a short introduction to data-driven background estimation techniques, this section explains how the two control regions – $e\mu$ with b jets and $ee/\mu\mu$ on-Z – are used for estimation of the $\text{t}\bar{\text{t}}\text{Z}$ and ZZ backgrounds respectively. This will be followed by a small discussion on the minor backgrounds, namely tWZ, WZ, Higgs processes, and other rare standard model processes.

6.8.1 Principles of data-driven background estimation

If possible, one should avoid trusting the simulated events blindly. One should check in control regions that are not overlapping with the signal region if the data agrees well with the simulation, and if possible, use the control regions to estimate the backgrounds in the signal region.

Take, for example, the situation where the differential distribution of some signal-region observbale – like a BDT score – is used for the measurement. There are several levels of simulation-dependence for data-driven background estimation methods. The first level is to use the data in the control region to get a background normalization scale factor:

$$B^{\text{est}}_{\text{SR},i} = B_{\text{SR},i} \times \frac{N_{\text{CR}}}{B_{\text{CR}}} = B_{\text{SR},i} \times \text{NSF}_B. \tag{6.21}$$

Here, B_{SR} and B_{CR} stand for the total MC predictions in the signal and control regions, N_{CR} is the observed number of events in the control region, and index i denotes the bin index of the differential distribution. In this scheme, one still relies on the simulation to accurately depict the differential distribution of the observable and yield ratios between the control and signal regions. Going a step further, one can perform the extrapolation for each bin of the observable, which would correspond to adding an index i to each variable in Equation 6.21 that does not have one yet. Now, one only needs the simulation to represent the *differences* in the differential distribution between the signal and control region accurately.

The final level of simulation-independence is to use additional control regions – obtained by inverting another selection criterion – to get the ratios of the differential distributions directly from data. Usually, it is not possible to find an additional cut to invert that is entirely uncorrelated. Therefore, the simulation must still be relied on to associate a certain *closure uncertainty*. Since this method involves one signal region and three control regions, it is often called the *ABCD method*.

In this analysis, the differential distributions of the BDT scores are well modeled in the control regions, so we rely on Monte Carlo to also model the distributions accurately in the signal region. Depending on the sample size of observed events in the control regions, the analysis uses bin-by-bin extrapolations or overall normalization scale factors for the estimation of the dominant backgrounds. In practice, one evaluates Equation 6.21 or its bin-by-bin equivalent as

$$B^{\text{est}}_{\text{SR},i} = B_{\text{SR},i} \times \frac{N_{\text{CR},i}}{B_{\text{CR},i}} = N_{\text{CR},i} \times \frac{B_{\text{SR},i}}{B_{\text{CR},i}} = N_{\text{CR},i} \times \text{TF}_{\text{CR}\to\text{SR},i} \tag{6.22}$$

because the effect of systematic variations often cancels out in the transfer factor $\text{TF}_{\text{CR}\to\text{SR},i}$. If the statistical uncertainties in the control region are not too large, one might even get a systematic uncertainty on the background that is smaller than the one obtained from the simulation in the signal region (including systematic variations). Note that in order to obtain the correct transfer factors, the other backgrounds and signals in both the signal and control regions must be subtracted beforehand.

6.8.2 Estimate of ZZ and $t\bar{t}Z$ backgrounds from control regions

The ZZ background can be validated with an extensive sample in the $ee/\mu\mu$ on-Z region. If one applies the two BDTs that are used for the signal extraction in the $e\mu$ region to this control sample, including the 2D splitting to the five bins for the signal extraction explained in Section 6.7.6, one obtains the histogram in Figure 6.34. With this histogram and the corresponding histogram in the $e\mu$ signal region, the transfer factors to estimate the ZZ yield in the $e\mu$ region were obtained for all three data-taking years. The estimated ZZ yields with uncertainties from both the transfer factor and the statistical uncertainty in the control region for the three data-taking years combined are shown in Table 6.14. The uncertainties in the transfer factor stem mainly from the MC statistics in the $e\mu$ region, followed by pileup reweighting and lepton selection efficiency variations.

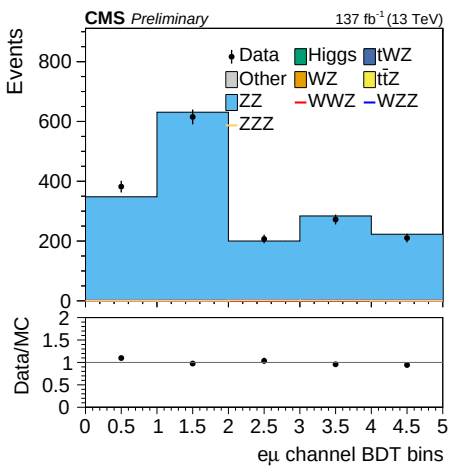
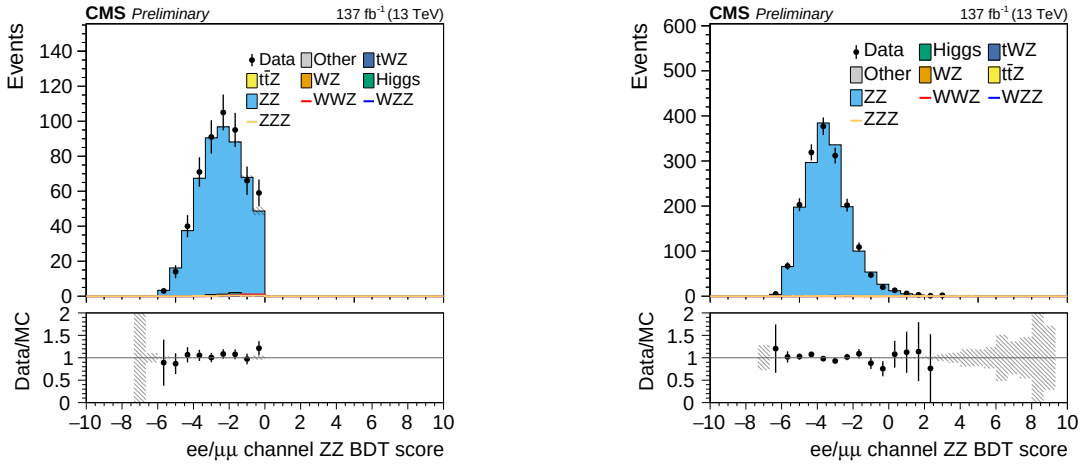


Figure 6.34: Events from the $ee/\mu\mu$ on-Z control region in the different $e\mu$ BDT bins.

Signal bin	Estimated ZZ yield	Syst. unc. [%]	Stat. unc. [%]	normalization scale factor	Simulation ZZ yield
$e\mu$ SR BDT 0	15.92	± 2.2	± 5.1	1.10 ± 0.06	14.5098
$e\mu$ SR BDT 1	1.56	± 5.2	± 4.0	0.97 ± 0.04	1.607
$e\mu$ SR BDT 2	0.62	± 7.6	± 7.0	1.02 ± 0.07	0.6077
$e\mu$ SR BDT 3	0.56	± 6.3	± 6.1	0.95 ± 0.06	0.5929
$e\mu$ SR BDT 4	0.23	± 10.0	± 6.9	0.94 ± 0.07	0.25
$ee/\mu\mu$ SR BDT A	78.32	± 3.0	± 4.3	1.04 ± 0.05	75.0584
$ee/\mu\mu$ SR BDT B	2.91	± 10.6			2.7864

Table 6.14: The estimated ZZ yields in the seven signal extraction bins (defined in Table 6.10). The $e\mu$ yields are extrapolated bin-by-bin from the on-Z control region, and the $ee/\mu\mu$ yields are extrapolated from off-Z events with a negative BDT score. The systematic uncertainty stems from systematic variations of the transfer factor computation, while the statistical uncertainty comes from the finite number of events in the control region.



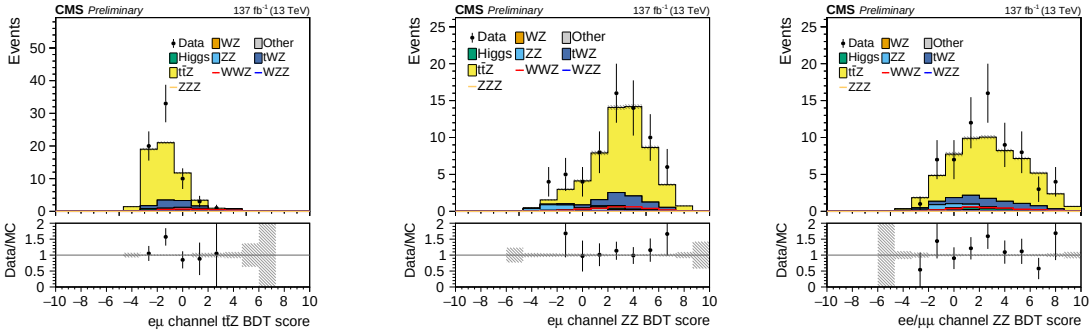
(a) Off-Z region for BDT score < 0.0.

(b) On-Z region.

Figure 6.35: The ZZ BDT score distributions for $ee/\mu\mu$ events.

For the ZZ background in the $ee/\mu\mu$ off-Z signal region, a bin-by-bin extrapolation is not possible because, for high values of the corresponding ZZ BDT score, there are very few events in the on-Z control region (see Figure 6.35b). Therefore, the extrapolation is done from the events in the off-Z region with negative BDT scores, which is not part of the signal region. This extrapolation relies on good modeling of the BDT score in the off-Z region. The good agreement in the on-Z region (Figure 6.35b) and the negative part in the off-Z region (Figure 6.35a) raises confidence that the BDT score is indeed well modeled. The estimated yields and uncertainties in the $ee/\mu\mu$ signal region are shown in Table 6.14. This time, the dominant systematic uncertainty is the pileup reweighting, while MC sample statistical uncertainties play a much more minor role.

The $t\bar{t}Z$ background can't be extrapolated bin-by-bin because there are much fewer events in its control region, the $e\mu$ category with b jets. However, the BDT scores (Figure 6.36) are well-modeled in the control region, and it is acceptable to estimate the $t\bar{t}Z$ background based on the shapes from simulation and an overall normalization scale factor measured in the control region. These estimated $t\bar{t}Z$ yields are listed in Table 6.15. The systematic uncertainties on the transfer factors are dominated by MC systematic uncertainties for the $e\mu$ bins and by b-tagging efficiency in the case of the $ee/\mu\mu$ bins.



(a) $e\mu$ $t\bar{t}Z$ BDT score (b) $e\mu$ ZZ BDT score (c) $ee/\mu\mu$ ZZ BDT score

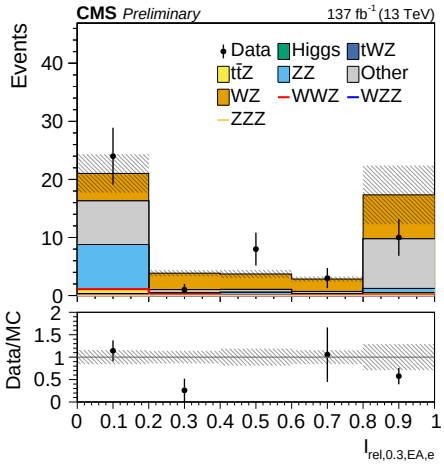
Figure 6.36: BDT distributions in the $e\mu$ with b jets control region.

Signal bin	Estimated $t\bar{t}Z$ yield	Syst. unc. [%]	Stat. unc. [%]	normalization scale factor	Simulation $t\bar{t}Z$ yield
$e\mu$ SR BDT 0	0.25	± 29.2	± 12.2	1.12 ± 0.17	0.222
$e\mu$ SR BDT 1	0.14	± 36.6	± 12.2	1.12 ± 0.17	0.1291
$e\mu$ SR BDT 2	2.79	± 13.0	± 12.2	1.12 ± 0.17	2.4912
$e\mu$ SR BDT 3	1.45	± 13.1	± 12.2	1.12 ± 0.17	1.2951
$e\mu$ SR BDT 4	0.09	± 44.6	± 12.2	1.12 ± 0.17	0.0764
$ee/\mu\mu$ SR BDT A	1.54	± 11.0	± 12.2	1.12 ± 0.17	1.3771
$ee/\mu\mu$ SR BDT B	2.32	± 10.8	± 12.2	1.12 ± 0.17	2.0728

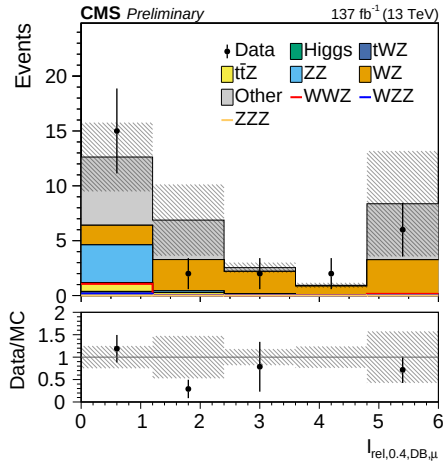
Table 6.15: The estimated $t\bar{t}Z$ background yields in the seven bins used for the signal extraction (defined in Table 6.10). All yields are extrapolated from the $e\mu$ region with b jets, relying on the simulation to represent the transfer factors accurately. The systematic uncertainty stems from systematic variations of the transfer factor computation, while the statistical uncertainty comes from the finite number of events in the control region.

6.8.3 Treatment of other backgrounds

The remaining rare backgrounds are more challenging to estimate from control regions. However, it is essential to validate WZ background modeling because it involves fake reconstructed lepton that are not necessarily represented in the simulation correctly. Lepton fake rate checks are particularly important for the W lepton candidates, as there is no invariant mass cut that drastically reduces the number of fake leptons. The number of fake leptons failing the W lepton ID is validated with a sideband of leptons not passing the nominal W lepton ID but passes the common loose ID (see Section 6.3.2) minus the isolation cut. This selection is also called the *fakeable* ID. The relative isolation variable of such leptons is shown in Figure 6.37 separately for electrons and muons. As expected, WZ events make up most of the events in the simulation, while the other backgrounds with two true electrons (DY and $t\bar{t}$) become relevant too. Even if these distributions are not well populated, the data generally agrees with the simulation. The number of fake leptons failing the W lepton ID but passing the fakeable ID appears to be modeled correctly, so the number of fake leptons that *pass* the W lepton ID likely well modeled as well. To obtain an uncertainty for the WZ normalization, we might argue that the WZ background ultimately cannot be validated more accurately than the statistical uncertainty of the data in the lepton selection sideband, which is $\pm\sqrt{73} = \pm 8.54$. There are 30.5 WZ events predicted by the simulation in this region, so $8.54/30.5 = 28\%$ is taken as an additional normalization uncertainty for the WZ background.



(a) The $I_{\mathrm{rel,R=0.3,EA}}$ of one of the W candidate electrons that failed the nominal ID but passed the fakeable ID.



(b) The $I_{\mathrm{rel,R=0.4,\Delta\beta}}$ distribution of one of the W candidate muons that failed the nominal ID but passed the fakeable ID.

Figure 6.37: Relative isolation variables for electrons (a) and muons (b) that pass the fakable ID but fail the W lepton candidate ID.

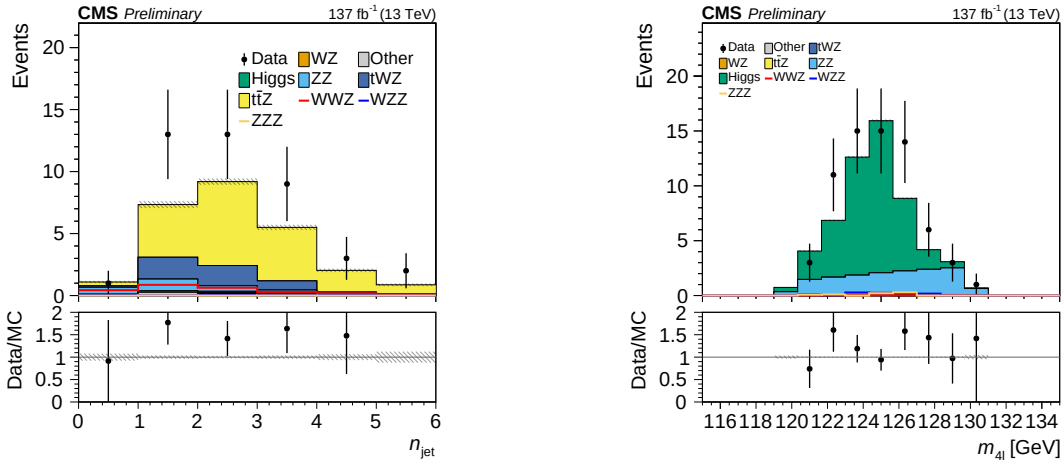
The tWZ background process has a lower cross-section than $\mathrm{t\bar{t}Z}$ and only has one b jet. One can try to isolate this background in an $e\mu$ control region with exactly one b-tagged jet, but because of b-tagging inefficiencies, the $\mathrm{t\bar{t}Z}$ background is still dominant after this selection. The n_{jet} distribution for $e\mu$ events with one b jet is shown in Figure 6.38a, and it contains information to separate between the tWZ and $\mathrm{t\bar{t}Z}$ backgrounds. If one fits this n_{jet} distribution in data with the predictions from the simulation with floating normalization parameters for the $\mathrm{t\bar{t}Z}$ and tWZ components, one obtains a normalization of 1.58 ± 0.63 for $\mathrm{t\bar{t}Z}$ and of 1.47 ± 2.13 for tWZ. The uncertainty of the tWZ normalization factor is very large because of the high $\mathrm{t\bar{t}Z}$ fraction. Nevertheless, the deviation from unity of 47 % is taken as an additional normalization uncertainty for the tWZ background.

There are several Higgs-related backgrounds, mostly from Higgs to $\tau\tau$ decays, where the Higgs is produced together with a leptonically-decaying Z boson. This process is very challenging to isolate in a control region. Some more background originates from Higgs bosons decaying into four leptons via two Z bosons, which is easier to select because the second Z boson is off-shell. Figure 6.38b compared the measured with the predicted number of events on the Higgs invariant mass peak after a cut on the invariant mass of the second Z boson candidate. The simulation of $\mathrm{H} \to 4\ell$ events describes the data well, and it is assumed that this also holds for the other Higgs-related backgrounds.

The remaining rare standard model backgrounds that have little impact on the analysis are estimated via simulation, including experimental and theoretical systematic uncertainties. The considered processes are $\mathrm{t\bar{t}}$, DY plus jets, $\mathrm{t\bar{t}W}$, WW, tZ plus jets and $\mathrm{t\bar{t}t\bar{t}}$.

6.9 Systematic uncertainties

The pre-fit signal is estimated exclusively from the simulation. For the signal bins in the $e\mu$ and $ee/\mu\mu$ categories, the average effect of the theoretical and experimental systematic uncertainties associated with the signal simulation is given in Table 6.16. For the



(a) The N_j distribution in $e\mu$ events with one b jet.

(b) The $m_{4\ell}$ distribution in $ee/\mu\mu$ category with $m_{\ell\ell} < 45$ GeV and $120 < m_{4\ell} < 130$ GeV.

Figure 6.38: Example distributions to study the minor backgrounds.

backgrounds, the systematic uncertainties are shown separately for the $e\mu$ and $ee/\mu\mu$ regions in Table 6.17 and Table 6.18 respectively. Compared with the systematic uncertainties on the signal estimate, the ZZ and $t\bar{t}Z$ backgrounds have no uncertainties from the b jet and lepton selections. The selection-related uncertainties mostly cancel out in the data-driven estimation. The remaining effects are considered in the uncertainties on the transfer factors, which incorporate both the systematic uncertainties for simulation and the statistical uncertainty on the control regions measurement.

In the following, the systematic uncertainties will be explained in more detail. They are grouped in theory uncertainties that are elaborated on in Section 6.9.1, and experimental uncertainties that are covered in Section 6.9.2.

6.9.1 Theory uncertainties

The three theory systematic uncertainties considered in the analysis relate to the renormalization and factorization scales (Q^2), the parton distribution function (PDF), and the strong coupling constant value α_{QCD}. The renormalization and factorization scales are varied simultaneously up or down by a factor two away from their nominal value. The more NLO effects are considered in the simulation, the less the impact of the Q^2 variations on the simulation. Since the triboson processes are simulated at NLO, low scale variation uncertainties are expected. Indeed, the scale uncertainty affects mostly the Higgs related backgrounds, while for the WWZ signal, the uncertainty is below 1 %.

The evaluation of PDF and α_{QCD} uncertainties follows the `PDF4LHC` prescription [137, 138]. The `NNPDF` distributions were used for all simulated samples [102]. For the complete triboson analysis, the α_{QCD} variations are fully correlated among all processes and final states. However, the PDF and Q^2 uncertainties are taken to be entirely uncorrelated for different processes, while they are still correlated for different final states from the same process. This is not relevant for the WWZ analysis, as it investigates the four-lepton final state only, but must be considered in combinations with the results from the WWW two-lepton (same sign) and three-lepton analyses. Excluding the rare backgrounds in the $e\mu$ region, the effect of the PDF and α_{QCD} uncertainties is below the percent level.

Systematic uncertainties	Uncertainty in each SR (%)	
	$e\mu$ SR	$ee/\mu\mu$ SR
b tagging scale factor (HF)	0.2	0.1
b tagging scale factor (LF)	0.8	0.8
JES	0.6	0.8
Pileup reweighting	-	2.7
α_{QCD}	0.2	0.3
PDF	0.1	0.2
Q^2	0.6	0.0
MET smearing	-	-
Luminosity	2.5	2.5
Trigger efficiency	3.2	3.2
Electron selection	3.3	3.0
Muon selection	3.2	3.3
Simulation statistics	3.7	5.0

Table 6.16: Summary of the typical systematic uncertaintes on signal events in the signal regions. The dash (-) indicates a systematic uncertainty that is either not applicable or is smaller than 0.1%.

6.9.2 Experimental uncertainties

Multiple experimental uncertainties need to be taken into account in the WWZ analysis. As for all CMS measurements, there is an overall uncertainty on the integrated luminosity of around 2.5 % (see Section 2.2.4 for more information on the luminosity measurement). The distribution of the number of pileup collisions in each data-taking year is also not known precisely. Usually, the MC simulations are produced with the expected pileup distribution and then reweighted according to the pileup distribution in the collision data that passes the quality selections, inferred from the cross-section of minimum bias events. The weights are varied according to the uncertainty on the minimum bias cross-section, propagating the pileup uncertainty to an uncertainty on the event yields predicted by the simulation. Another systematic uncertainty stems from the limited sample size of the simulation. This uncertainty is usually significant for rare backgrounds, where only a few events remain after the analysis selection.

Concerning the object selections, there are first and foremost the uncertainties on electron and muon selection efficiencies, as measured differentially in p_{T} and η with the tag-and-probe method. The uncertainty in the trigger selection efficiency was estimated by checking how many DY+jets events identified in a jet-triggered primary dataset pass the double-lepton triggers and comparing this rate with the rate in simulated DY+jets events. However, this is a very conservative estimate, as the trigger threshold effects play a lesser role in four-lepton events.

Some more experimental uncertainties come from the jet selection. As anticipated in Section 6.3.4, the uncertainty of the b-tagging selection must be considered separately for heavy-flavor and light-flavor jets. Depending on the physics process, one of the two uncertainties is more important. For the WWZ signal that has no b jets in the final state, the light-flavor selection uncertainty predominates. The opposite is the case for the tWZ background. For Higgs-related backgrounds, it depends on the signal region. In the $e\mu$ region, the Higgs background mostly consists of $\mathrm{t\bar{t}H}$ events with $\mathrm{H} \rightarrow \mathrm{WW}$ decays, so the heavy-flavor uncertainty is more relevant. In the $ee/\mu\mu$ region, there are more

Systematic uncertainties	Typical size of the uncertainty (%)					
	$t\bar{t}Z$	ZZ	tWZ	Higgs	WZ	Other
b-tag CR data statistics	12.2	-	-	-	-	-
ZZ CR data statistics	-	2.4	-	-	-	-
b tagging scale factor (HF)	-	-	4.9	1.5	-	0.5
b tagging scale factor (LF)	-	-	0.8	0.7	-	0.2
JES	-	-	2.5	0.7	-	0.3
Pileup reweighting	-	-	1.6	4.5	-	1.1
α_{QCD}	-	-	-	0.2	-	1.7
PDF	-	-	0.2	0.9	-	0.9
Q^2	-	-	-	14.5	-	1.9
Luminosity	2.5	2.5	2.5	2.5	2.5	2.5
Uncertainty on $\mathrm{TF}_{\text{b-tag CR}\to e\mu}$	29.2	-	-	-	-	-
Uncertainty on $\mathrm{TF}_{\text{ZZ CR}\to e\mu}$	-	2.2	-	-	-	-
Uncertainty on WZ validation	-	-	-	-	30.0	-
Trigger efficiency	3.2	3.2	3.2	3.2	3.2	3.2
Electron selection	-	-	3.2	2.9	-	2.8
Muon selection	-	-	3.2	3.9	-	3.3
Simulation statistics	-	-	5.7	29.6	22.1	36.3

Table 6.17: Summary of the typical systematic uncertaintes on background estimations in the $e\mu$ signal region. The dash (-) indicates a systematic uncertainty that is either not applicable or is smaller than 0.1%.

Systematic uncertainties	Typical size of the uncertainty (%)					
	$t\bar{t}Z$	ZZ	tWZ	Higgs	WZ	Other
b-tag CR data statistics	12.2	-	-	-	-	-
ZZ CR data statistics	-	2.4	-	-	-	-
b tagging scale factor (HF)	-	-	4.2	0.9	-	0.4
b tagging scale factor (LF)	-	-	0.8	1.1	-	1.8
JES	-	-	2.4	1.3	-	0.6
Pileup reweighting	-	-	1.5	3.6	-	40.8
α_{QCD}	-	-	0.5	-	0.4	
PDF	-	-	0.3	0.3	-	0.1
Q^2	-	-	0.0	0.1	-	8.3
Luminosity	2.5	2.5	2.5	2.5	2.5	2.5
Uncertainty on $\mathrm{TF}_{\text{b-tag CR}\to e\mu}$	11.0	-	-	-	-	-
Uncertainty on $\mathrm{TF}_{\text{ZZ CR}\to e\mu}$	-	3.0	-	-	-	-
Uncertainty on WZ validation	-	-	-	-	30.0	-
Trigger efficiency	3.2	3.2	3.2	3.2	3.2	3.2
Electron selection	-	-	3.0	4.6	-	3.7
Muon selection	-	-	3.4	3.3	-	0.8
Simulation statistics	-	-	6.4	15.7	19.4	72.9

Table 6.18: Summary of the typical systematic uncertaintes on background estimations in the $ee/\mu\mu$ signal region. The dash (-) indicates a systematic uncertainty that is either not applicable or is smaller than 0.1%.

$H \to 4\ell$ events with no heavy flavor jets. More jet-related uncertainties come from the jet energy scale (JES) and jet energy resolution (JER) corrections, which are merged to a jet energy correction uncertainty for Tables 6.16, 6.17, and 6.18. As the number of b-tagged jets above 20 GeV is required to be zero in both the $e\mu$ and $ee/\mu\mu$ signal regions, variations of the jet energy can cause some b jets to migrate over the energy threshold, changing the number of accepted events. Furthermore, the leading jet p_T is used in the $t\bar{t}Z$ BDT in the $e\mu$ region, but this has only a slight effect on the predicted yields. The $t\bar{t}Z$ background is extrapolated from control regions, and the signal has mostly no jets in the final state.

Finally, the uncertainties related to the background estimation, were already discussed in Section 6.8. These uncertainties cover the transfer factors, statistical uncertainties in the control regions, and additional uncertainties for backgrounds that are challenging to control. All uncertainties are fully correlated among the three data-taking years, except for luminosity and JER uncertainties, which are uncorrelated. Except for the statistical uncertainties from the simulation and the uncertainties from the background extrapolations, all uncertainties are fully correlated among the different final states measured in the full VVV analysis.

6.10 Combination and statistical interpretation

So far, this chapter mainly focused on the four-lepton analysis within the full triboson analysis. In the full analysis, all covered VVV final states are modeled simultaneously in a combined fit, primarily maximize the sensitivity to general triboson production. Furthermore, there is some mixing between the final states by lepton number and the triboson processes, so the sensitivity to specific processes like WWZ increases in a combined fit as well. For example, the four-lepton analysis can also select WZZ or ZZZ, even if the event selection actively suppresses this because it looks like ZZ background plus jets or missing energy. On the other side, the three-lepton analysis also picks up some WWZ signal. Therefore, it is essential to perform a combined likelihood fit – including all final-state oriented studies – if one wants to measure signal strengths for WWZ and the other triboson processes.

The signal strength is defined as the ratio between observed and expected event yield. Signal strength values and uncertainties are extracted by a likelihood fit of the predicted yields to the observed ones. However, the signal strengths are not the only free model parameters, even though they are the parameters of interest. Other parameters correspond to values that are only known approximately because of systematic uncertainties. These so-called nuisance parameters get constrained by auxiliary measurements. Including the likelihood functions for the auxiliary measurements is usually not feasible. Instead, one adds constraint terms based on the central values and uncertainties of the auxiliary measurements. This statistical procedure is standard for LHC analyses [139].

The signal strength fits in this analysis were performed with the collaboration-internal tools based on the `RooFit` [140] and `RooStat` [141] components of the `ROOT` library [113]. The results were cross-checked with fitting routines directly implemented with the `MINUIT` library for numerical optimization [142] (included in `ROOT`) and using its Python wrapper `iminuit` [143]. Software packages for likelihood fit must be able to service the increasingly complicated fits performed by the LHC experiments. For this reason, `RooFit` development recently became more active again [144]. There are also R&D efforts towards solutions that use modern optimization libraries such as `TensorFlow` [145] internally. A notable example is the `zfit` package [146].

In the following, we will first discuss the likelihood fit with statistical uncertainties only. Afterwards, the constraint terms corresponding to the systematic uncertainties will be motivated before showcasing the fit results for different signal strength definitions.

6.10.1 Statistical uncertainties only

A likelihood fit finds the model parameter values that maximize the likelihood $\mathcal{L}$ to observe the data given the parameterized model. For the statistically independent bins in a counting experiment, the likelihood is a product of Poisson probabilities:

$$\mathcal{L} = \prod_{j=1}^{M} p\left(N_j; P_j\right), \qquad p(N_j; P_j) = \frac{P_j^N e^{-P_j}}{N_j!}, \tag{6.23}$$

where M is the number of bins, N_j is the observed number of events in bin j, and P_j is the predicted number of events in bin j. The fit minimizes the negative log-likelihood l:

$$l = -\ln \mathcal{L} = -\ln \prod_{j=1}^{M} \left(\frac{P_j^{N_j}\, e^{-P_j}}{N_j!} \right) = \sum_{j=1}^{M} \ln N_j! - \sum_{j=1}^{M} N_j \ln P_j + \sum_{j=1}^{M} P_j. \tag{6.24}$$

The term $\sum_j \ln N_j!$ does not depend on any model parameter. Hence it can be ignored in the fit, and it is sufficient to minimize

$$l = \sum_{j=1}^{M} \Big(P_j - N_j \ln P_j \Big). \tag{6.25}$$

If the P_j were independent and free parameters, then minimization of l leads to the result $P_j = N_j$ with variance $\sigma^2_{P_j} \approx 1/N_j$.

The get the uncertainty of parameter estimators in a likelihood fit in general, one can use the Cramér–Rao bound, which is a lower bound on the variance of unbiased estimators. It relates the second derivative of the log-likelihood – the Fisher information $I(\theta)$ – to the estimator variance:

$$V(\hat{\theta}) \geq \frac{\left(1 + \frac{\partial}{\partial\theta} b(\hat{\theta})\right)^2}{I(\theta)}, \quad \text{where} \quad I(\theta) = \mathrm{E}\left[\left(\frac{\partial l(x;\theta)}{\partial\theta} \right)^2 \right]. \tag{6.26}$$

Here, θ are the model parameters, $\hat{\theta}$ the maximum likelihood, x the observations, $b(\hat{\theta})$ the bias of the estimator, and E denotes the expectation value.

The maximum likelihood estimator is unbiased and reaches the Cramér–Rao bound in the high sample limit. If one expands the negative log-likelihood around the minimum, the first derivative term vanishes by definition, while third derivative and higher-order terms are usually negligible. Hence, the log-likelihood is usually of parabolic shape around the minimum. Then, according to Equation 6.26, the estimates' uncertainty corresponds to the crossing-point of the log-likelihood with the minimum value plus 0.5.

In some cases, usually in the small sample limit, the log-likelihood is not parabolic. However, one could always imagine a transformation of the parameter θ that makes the log-likelihood parabolic in the transformed parameter. Equivalently, one can leave the parameters as they are and find the potentially asymmetric σ_θ with a parameter scan. This likelihood scan is a common practice in collider experiments. Confidence intervals

other than the 68 % interval can be found by finding the crossing-point of the likelihood function with other levels. More details about the statistical methods generally used in particle physics can be found in [147, 148].

In a typical collider physics analysis, one assumes that the predicted number of events is nearly correct. A *signal strength* parameter μ is introduced to allow for (small) modifications to the predicted number of signal events. If the prediction is based on a theoretical signal cross-section $\sigma^{\text{th.}}$, the measured cross-section is $\sigma^{\text{meas.}} = \mu\sigma^{\text{th.}}$.

The predicted number of events is the sum of the numbers of signal and background events: $P_j = S_j + B_j$. Since the purpose of the fit is to bring the prediction into line with observation, one can incorporate a signal strength into the prediction: $P_j = \mu S_j + B_j$. In this simple case, the signal strength μ is the one free parameter in the fit.

For the search for heavy triboson production in general, the signal in channel j is a sum over all four production processes:

$$S_j = S_j^{\text{WWW}} + S_j^{\text{WWZ}} + S_j^{\text{WZZ}} + S_j^{\text{ZZZ}} = \sum_{s=1}^{4} S_j^{(s)}, \tag{6.27}$$

where the index s refers to the signal sources. In one version of the fit, independent signal strengths are associated with each of the four heavy triboson final states:

$$S_j \to \sum_{s=1}^{4} \mu_s S_j^{(s)}. \tag{6.28}$$

The signal strengths μ_s do not depend on the channel j but only on the triboson process s. The negative log-likelihood with four independent signal strengths is

$$l_{[4]} = \sum_{j=1}^{M} \left\{ \left(\sum_{s=1}^{4} \mu_s S_j^{(s)} \right) + B_j - N_j \ln \left[\left(\sum_{s=1}^{4} \mu_s S_j^{(s)} \right) + B_j \right] \right\}. \tag{6.29}$$

In another version of the fit, we are interested in the overall strength of the signal from all four sources, in which case there is only one, global signal strength:

$$S_j \to \sum_{s=1}^{4} \mu S_j^{(s)} = \mu S_j, \tag{6.30}$$

which is equivalent to imposing the constraint $\mu_1 = \mu_2 = \mu_3 = \mu_4$. The one-parameter negative log-likelihood is

$$l_{[1]} = \sum_{j=1}^{M} \left\{ \mu \left(\sum_{s=1}^{4} S_j^{(s)} \right) + B_j - N_j \ln \left[\mu \left(\sum_{s=1}^{4} S_j^{(s)} \right) + B_j \right] \right\} \tag{6.31}$$

$$= \sum_{j=1}^{M} \left\{ \mu\, S_j + B_j - N_j \ln \left[\mu\, S_j + B_j \right] \right\}. \tag{6.32}$$

6.10.2 Systematic uncertainties in the likelihood function

The auxiliary measurements for nuisance parameter constraints are themselves likelihood fits in the frequentist picture. However, it is not practical to combine all detailed likelihood functions of the auxiliary measurements with the final likelihood function

that includes the signal strength. Instead, one takes the nominal result of the auxiliary measurements and their uncertainties and interpolates (or extrapolates) them to get Bayesian prior distributions for all nuisance parameters, which are multiplied with the likelihood function Poisson likelihood function from the previous section. For this interpolation of the prior, an assumption on the kind of probability distribution must be made.

For most analyses – including the triboson analysis –, all nuisance parameter constraints can be modeled with one of two distributions: the log-normal distribution or the gamma disribution [149]. Most systematic uncertainties are associated with multiplicative factors that can not be negative. Hence, a Gaussian prior would not be adequate, as it always has a nonvanishing probability for negative values. A better choice is the *log-normal distribution*, which is only supported for positive values:

$$p(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{\theta}\exp\left(-\frac{(\ln\theta - \mu)^2}{2\sigma^2}\right). \tag{6.33}$$

The parameters μ and σ^2 do not correspond to the mean and variance of the log-normal distribution of θ, but rather to the Gaussian distribution of $\ln(\theta)$. For the log-normal distribution, the mode, mean and variance are:

$$\text{mean:} \quad \exp\left(\mu + \tfrac{\sigma^2}{2}\right), \tag{6.34}$$

$$\text{mode:} \quad \exp\left(\mu - \sigma^2\right), \tag{6.35}$$

$$\text{variance:} \quad \mu^2\left(e^{\sigma^2} - 1\right). \tag{6.36}$$

It is not always clear how the nominal value μ_θ and uncertainty σ_θ of the nuisance parameter should be mapped to the parameters of the log-normal distribution. The convention followed in this analysis is to map the nominal value to the mode, and the squared uncertainty to the variance, resulting in a system of equations that can be solved numerically.

The *gamma distribution* is used for background estimation nuisance parameters. This can be understood when looking at the case of the extrapolation from a control region (CR) to a signal region (SR) with a transfer factor $1/\tau$. We want to know the posterior distribution of the expected value μ_{SR} in the signal region, given the number of events in the control region n_{CR}. A Poisson distribution connects the observed value in the control region to the expected value:

$$p(n_{\text{CR}}|\mu_{\text{CR}}) = \frac{{\mu_{\text{CR}}}^{n_{\text{CR}}} e^{-\mu_{\text{CR}}}}{n_{\text{CR}}!}. \tag{6.37}$$

With the transfer relation, one can get a probability expression that involves μ_{SR}:

$$p(n_{\text{CR}}|\mu_{\text{SR}}) = \frac{\tau{\mu_{\text{SR}}}^{n_{\text{CR}}} e^{-\tau\mu_{\text{SR}}}}{n_{\text{CR}}!}. \tag{6.38}$$

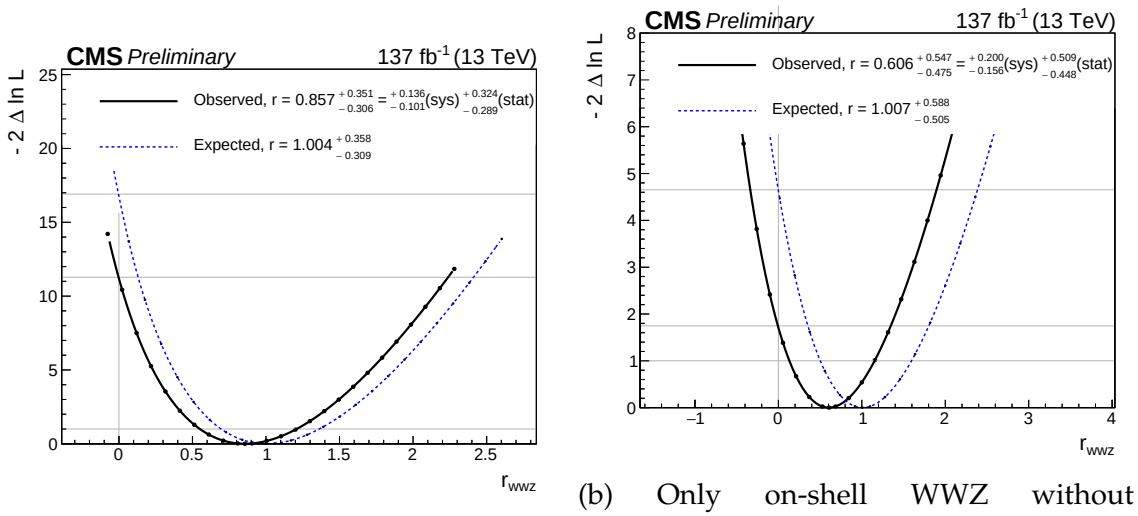
Finally, one uses Bayes theorem to get the posterior distribution $p(\mu_{\text{SR}}|n_{\text{CR}})$:

$$p(\mu_{\text{SR}}|n_{\text{CR}}) = \frac{p(n_{\text{CR}}|\mu_{\text{SR}})p(\mu_{\text{SR}})}{p(n_{\text{CR}})} = C \cdot p(n_{\text{CR}}|\mu_{\text{SR}}) = C \cdot \frac{\tau{\mu_{\text{SR}}}^{n_{\text{CR}}} e^{-\tau\mu_{\text{SR}}}}{n_{\text{CR}}!}. \tag{6.39}$$

The important step in the last equation is the second one. Since we assume to have no prior information on any of the variables, the ratio $p(\mu_{\text{SR}})/p(n_{\text{CR}})$ is simply a constant ratio C of flat priors. With this normalization factor, the posterior $p(\mu_{\text{SR}}|n_{\text{CR}})$ fits the definition of a gamma distribution.

6.10.3 Likelihood scans and nuisance parameter impact

Figure 6.39 shows the WWZ likelihood scans for different values of μ^{WWZ}. These scans consider all final states of the full triboson analysis, but for the WWZ the sensitivity is dominated by the four-lepton analysis presented in this chapter. Likelihood values are shown for two different signal definitions. In one version, the VH $\rightarrow$ VVV process is included, which is the signal that the analysis targets primarily. In the other version, only on-shell VVV production is taken as the signal. When comparing the expected and observed likelihoods that are both indicated in the figure, it becomes clear that the observed on-shell WWZ signal is weaker than expected. However, the observation is still within the 1 σ confidence interval of the prediction, so this is likely a statistical underfluctuation.



(a) All WWZ.

(b) Only on-shell WWZ without ZH$\rightarrow$WWZ.

Figure 6.39: $-2\Delta \ln L$ as a function of the signal strength for the WWZ process in general (a) and only for on-shell WWZ production without the ZH process (b).

It is also interesting to compare the inclusive fit's statistical uncertainty with the expectation from the likelihood-free inference ceiling study. For the full analysis, the statistical uncertainty on the signal strength is $^{+0.324}_{-0.289}$, while the ceiling study indicated a minimal uncertainty of $^{+0.356}_{-0.303}$ for the $e\mu$ region alone. It can be concluded that the benefit of including the $ee/\mu\mu$ region is greater than using only the $e\mu$ region and improving the signal extraction method.

To validate the statistical model and assess the effect of the nuisance parameters on the WWZ signal strength uncertainty, the *pulls* and *impacts* of the nuisance parameters are shown in Figure 6.40 for the inclusive WWZ fit, and in Figure 6.41 for the on-shell WWZ signal only. When comparing fit parameter values, one has to make the distinction between *pre-fit* and *post-fit* values. The figures only show the 30 nuisance parameters with the largest impact. The signal strengths' pre-fit values are unconstrained, while the nuisance parameters are constrained by auxiliary measurements. Comparing the pre-fit and post-fit nuisance parameter constraints, it is possible to understand how much the measurement could tighten the constraints on the nuisance parameter or pull the nominal value away from its pre-fit value. The pulls are shown on the left side of Figures 6.40 and 6.41, with the errorbars showing the post-fit uncertainty relative to the pre-fit uncertainty. The measurement is consistent with the auxiliary nuisance measurements if the pulls are small, which is the case for this analysis. Furthermore, the nuisance pa-

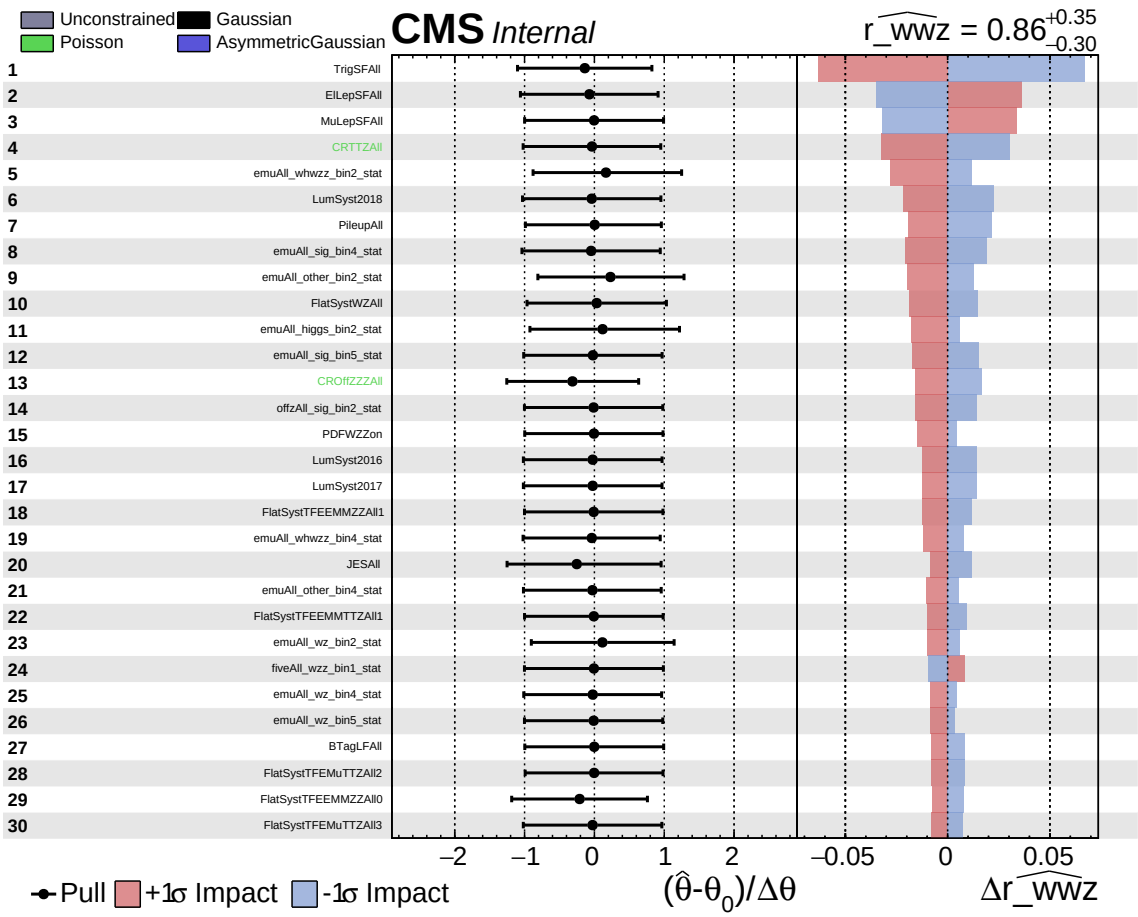


Figure 6.40: Impact plots of the dominant nuisance parameters for the total WWZ signal for the BDT analysis.

rameter constraints are not tightened by the measurement, which is generally expected. The four-lepton analysis was not designed to measure any of the nuisance parameters, although this can happen accidentally. However, if the measurement would constrain parameters that clearly can not be measured better than in their respective auxiliary measurement – such as the integrated luminosity – this would hint towards an inconsistent statistical model.

The right side of Figures 6.40 and 6.41 show the nuisance parameter impact on the signal strength measurement. The $\pm 1\ \sigma$ impacts are evaluated by varying the nuisance parameter by its $\pm 1\ \sigma$ post-fit uncertainties. The dominant systematic uncertainty is the trigger efficiency uncertainty. As argued in Section 6.9.2, the auxiliary measurement that constrains the trigger efficiency is not optimized for the four-lepton final state, so there is room for improving the full systematic uncertainty of the measurement by revisiting the trigger efficiency measurement. Just after the trigger uncertainty, the lepton selection uncertainties and uncertainty from the $t\bar{t}Z$ background extrapolation follow in place 2, 3, and 4. Since the lepton selection efficiencies were studied very carefully – following the complete auxiliary measurement procedure recommended within the CMS collaboration –, it is more challenging to reduce these uncertainties in future measurements. The $t\bar{t}Z$ background estimate uncertainty is strongly affected by the statistical uncertainty in the $e\mu$ control region with b jets, so it can be assumed that this uncertainty will reduce for future measurements based on larger datasets. Note that for the inclusive WWZ signal, the lepton selection uncertainties come before the $t\bar{t}Z$ extrapolation uncertainty,

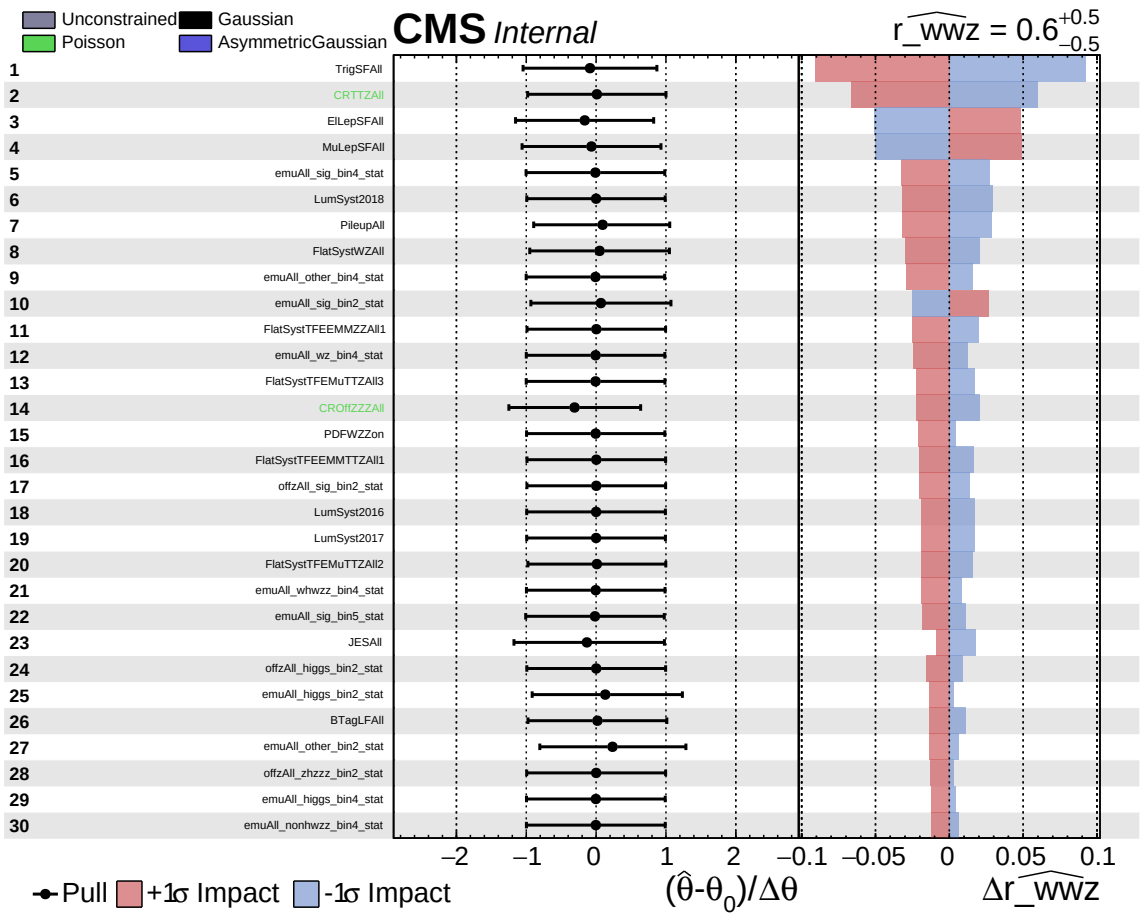


Figure 6.41: Impact plots of the dominant nuisance parameters for the WWZ signal without the Higgs process for the BDT analysis.

while for on-shell production only, this ordering is inverted. On places 6 and 7 are the pileup and luminosity uncertainties, with other systematic uncertainties of similar impact being the statistical uncertainties from the limited simulation sample size. Other uncertainties, such as jet related ones, are ranked further down at the level of 1 – 2 %.

6.11 Combined triboson measurement results

After discussing the four-lepton BDT analysis in detail, this section summarizes the final results of the combined triboson analysis that also included the WWW, WZZ, and ZZZ processes. As the four-lepton analysis dominates the sensitivity to the WWZ process by a large margin, the WWZ part of the combined results represents the outcome of the analysis presented in this chapter. Figure 6.42 shows the event yields in all bins of the complete published triboson analysis [150], including the seven four-lepton bins that were optimized in this thesis. The figure also shows the yields expected from post-fit predictions and the approximate median significance of each bin. The approximate significance confirms that the $e\mu$ region in the four-lepton final state is the most sensitive channel to study the production of three massive gauge bosons.

Table 6.19 lists the observed and expected statistical significances of the individual processes' signal strengths, as well as the significance of the on-parameter signal where all triboson processes were scaled with one single signal strength. These significances are

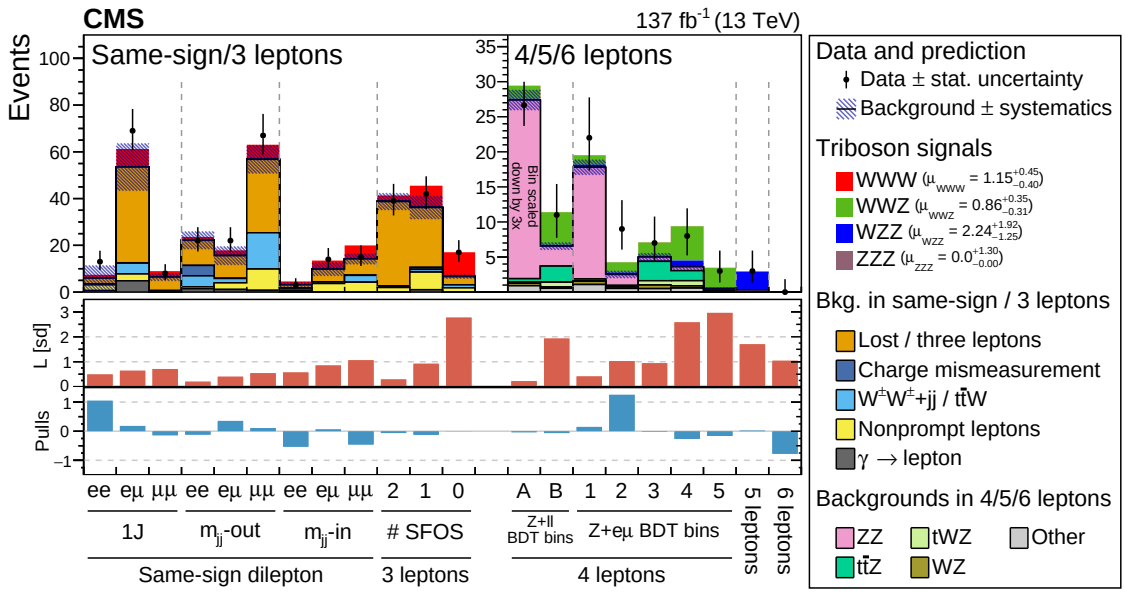


Figure 6.42: Overview of event yields in all bins of the final CMS massive triboson analysis [150], including the seven four-lepton bins that were optimized in this thesis. The observed number of events is compared to the predicted number after the fit (post-fit yields). The VVV signal is stacked on top of the total background. The observations are indicated with black markers and statistical uncertainties, while the hatched area represents the uncertainty on the background predictions. The middle panel shows the expected approximate median significance computed with Equation 6.20. The lower panel shows how much the fit result is pulled away from the nominal expectation.

listed for both the inclusive triboson signal and the on-shell VVV process only. Several milestones were reached with this measurement. It is the first measurement with a correlated triboson signal above 5 standard deviations significance, meaning that the production of three massive gauge boson has now been observed. Furthermore, the inclusive production of the WWW and WWZ final states are both measured with a significance above 3 σ, so there is evidence for these processes. If the Higgs associated production is not considered as part of the signal, none of the measurements reaches the 3 σ threshold, so more data or a more sensitive analysis is needed to measure the production of three on-shell vector bosons in the standard model in detail.

Since no ZZZ candidate events were observed in the six-lepton final state, the observed significance for ZZZ is zero. However, it is possible to put an upper limit on the signal strenth of this process, which is listed in Table 6.20. If the measured signal strenths are multiplied with the predicted cross section listed in Table 1.1, the following measured cross section measurements for inclusive triboson production are obtained:

$$
\begin{aligned}
\sigma_{\mathrm{VVV}} &= 1009.0^{+208.3}_{-195.4}(\mathrm{stat})^{+145.8}_{-121.0}(\mathrm{syst}) = 1009.0^{+255.0}_{-230.2}(\mathrm{tot})\ \mathrm{fb} \\
\sigma_{\mathrm{WWW}} &= 587.8^{+164.0}_{-153.8}(\mathrm{stat})^{+162.5}_{-129.9}(\mathrm{syst}) = 587.8^{+231.3}_{-201.2}(\mathrm{tot})\ \mathrm{fb} \\
\sigma_{\mathrm{WWZ}} &= 303.4^{+114.7}_{-102.3}(\mathrm{stat})^{+48.1}_{-35.8}(\mathrm{syst}) = 303.4^{+124.3}_{-108.7}(\mathrm{tot})\ \mathrm{fb} \\
\sigma_{\mathrm{WZZ}} &= 204.8^{+162.9}_{-113.2}(\mathrm{stat})^{+67.1}_{-16.9}(\mathrm{syst}) = 204.8^{+176.1}_{-114.5}(\mathrm{tot})\ \mathrm{fb} \\
\sigma_{\mathrm{ZZZ}} &< 198.5\ \mathrm{fb}
\end{aligned}
\tag{6.40}
$$

Process	VH as signal	VH as background
WWW	3.33 (3.09)	1.63 (1.87)
WWZ	3.35 (4.09)	1.31 (2.15)
WZZ	1.71 (0.69)	1.71 (0.76)
ZZZ	0.00 (0.89)	0.00 (0.90)
combined	5.67 (5.88)	2.90 (3.48)

Table 6.19: Observed (expected) significance (σ) for different analysis options and different signal processes.

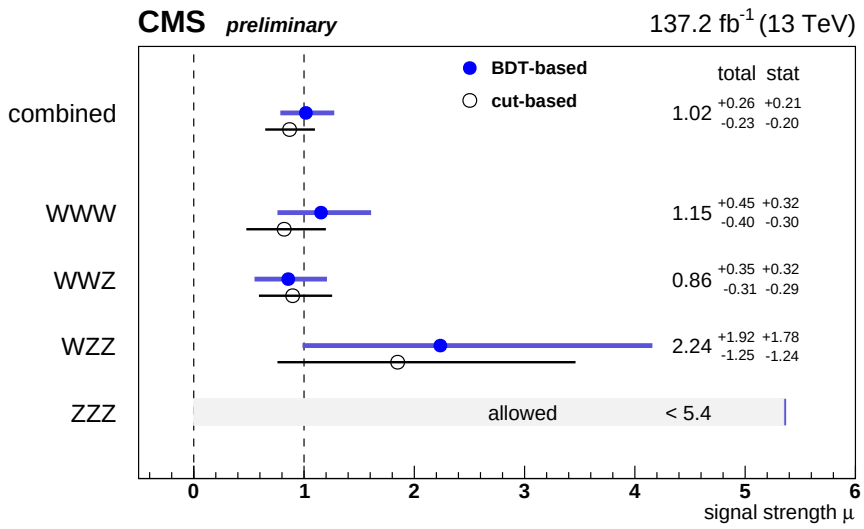
Process	VH as signal	VH as background
WZZ	6.10 (6.06, 3.83, 2.52)	5.78 (5.95, 3.69, 2.39)
ZZZ	5.35 (11.06, 6.16, 3.49)	5.69 (11.60, 6.31, 3.53)

Table 6.20: Observed (-1σ, median, $+1\sigma$ expected) 95% CL upper limit on WZZ and ZZZ signal strength for different analysis options and different signal processes.

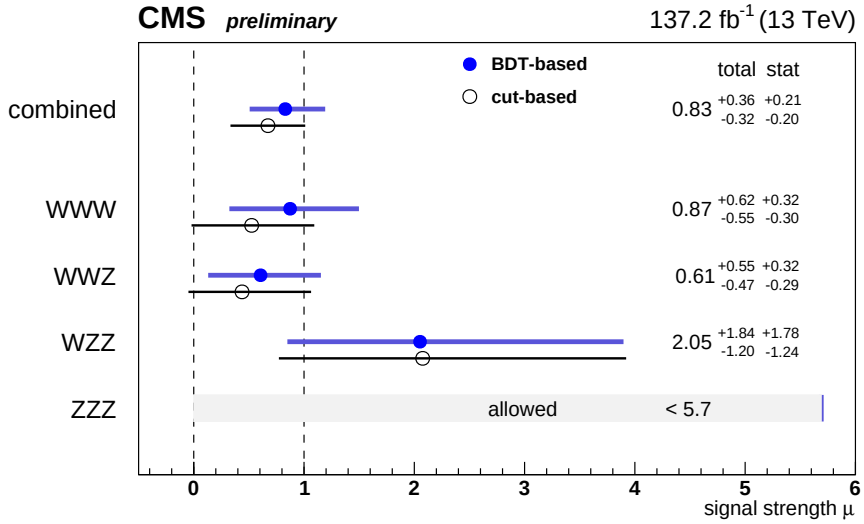
A summary of the fitted signal strengths is given in Figure 6.43, which also includes the upper limit for the ZZZ process. Once again, the results are given for both the inclusive VVV final states and the on-shell VVV signal only. The signal-strength figures also include the results of an alternative cut-based analysis that covers the same final states for a cross-check. This cut-based analysis has not been discussed so far, as it was not part of this thesis work. The cut-based analysis uses the same event preselections and categorizations in the four-lepton channel, but unlike the BDT, it only uses two discriminating variables in the $e\mu$ and $ee/\mu\mu$ categories respectively to get the bins for the signal extraction.

That the sensitivity of the BDT analysis is not dramatically higher than the sensitivity of the cut-based analysis teaches important lessons. To boost the sensitivity with machine-learning methods, it might be necessary to provide more information to the classifiers – such as object identification and isolation variables – for a more optimized event selection. This prospect has been investigated in the scope of this thesis, but such approaches are very challenging to defend because it is not clear how lepton selection uncertainties can be treated.

To conclude this presentation of the triboson analysis, two event displays are shown in Figure 6.44. Figure 6.44a shows a WWZ $\rightarrow 4\ell$ candidate event. Even though the five-lepton final state was not discussed in detail, Figure 6.44b shows one of the WZZ $\rightarrow 5\ell$ event candidates because the five leptons in this event are all electrons. As this thesis has contributed to the treatment of electrons in CMS on many fronts, such as reconstruction, identification, and selection efficiency measurements, this event display is adequate to showcase the benefits of these contributions.

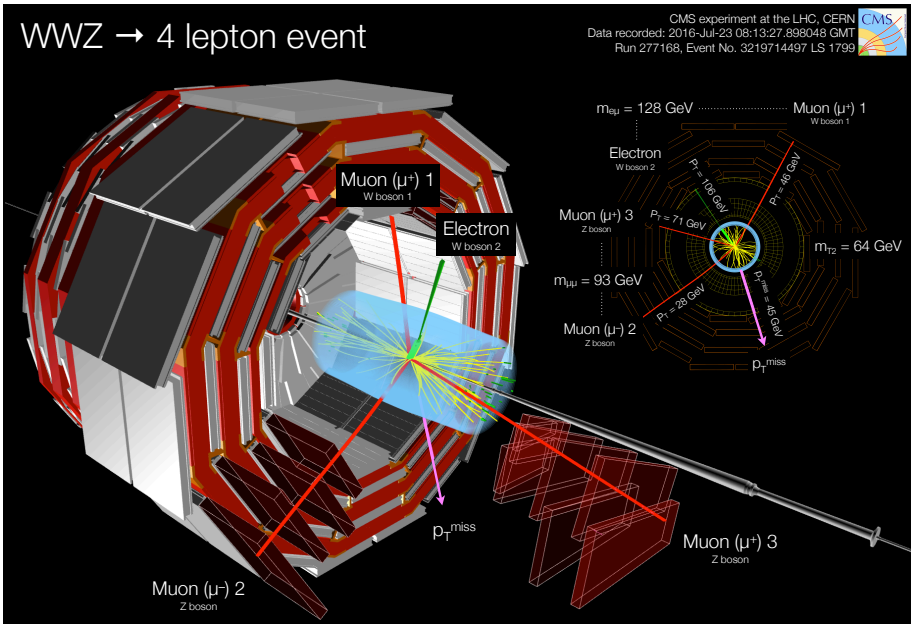


(a)

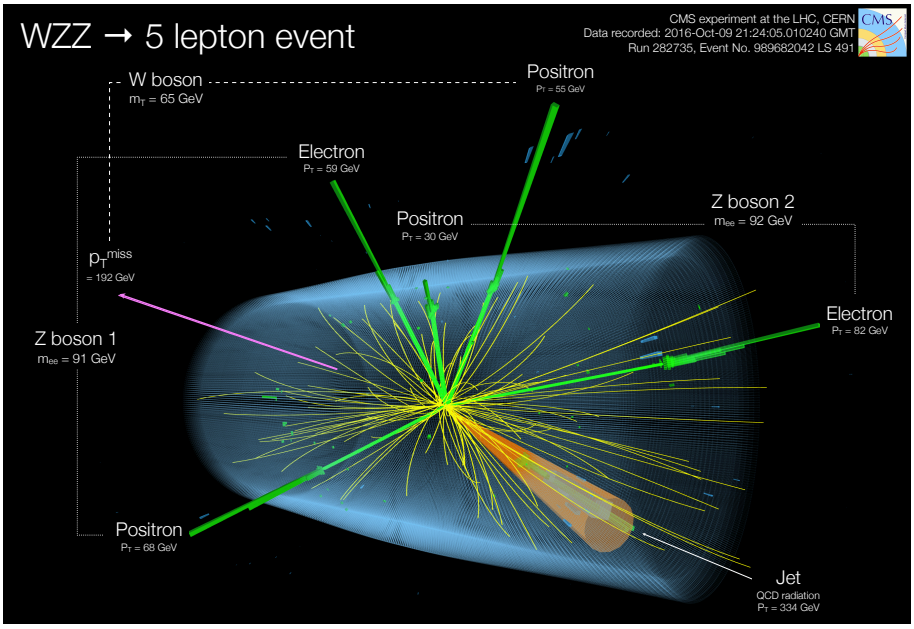


(b)

Figure 6.43: Summary plot of the fitted signal strengths for different signal process, including VH as signal (a) [150], and not including VH as signal (b). Blue dots and lines correspond to BDT-based result, black circles and lines correspond to cut-based result. The stated values come from the BDT-based analysis and include the total uncertainty and the statistical component of the uncertainty.



(a) Event display of WWZ event candidate in the four lepton final state. Two opposite-charged muons compose the Z boson candidate. The additional $e\mu$ pair associated with the two W bosons has an invariant mass of 128 GeV. The pink arrow represents the missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$ associated with the neutrinos. Only half of the muon detector is illustrated.



(b) Event display of WZZ event candidate in the five electron final state. The electrons and positron candidates are represented by the green tracks and towers that represent the energy deposits in the ECAL. The missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$ from the neutrino is represented by the pink arrow. The transverse mass m_{T} of the W boson electron and the $p_{\mathrm{T}}^{\mathrm{miss}}$ is 65 GeV.

Figure 6.44: Event displays of WWZ (a) and WZZ (b) event candidates.

Chapter 7

Conclusion

The proton-proton collision data taken during the LHC Run 2 at $\sqrt{s}$ = 13 TeV allows for the exploration of the electroweak scale at the high-energy frontier, testing the limits of the standard model of particle physics in the electroweak and scalar sector. This thesis focused on the measurement of the production of three massive gauge bosons in the four-lepton final state, which covers mostly the WWZ process. This rare process sensitive to anomalous quartic gauge couplings has not been observed to far, and this work raises the first evidence for the inclusive production of the WWZ final state at the CMS experiment. The WWZ cross-section measurement is now part of the results of the large physics program of the CMS collaboration, as shown in Figure 7.1.

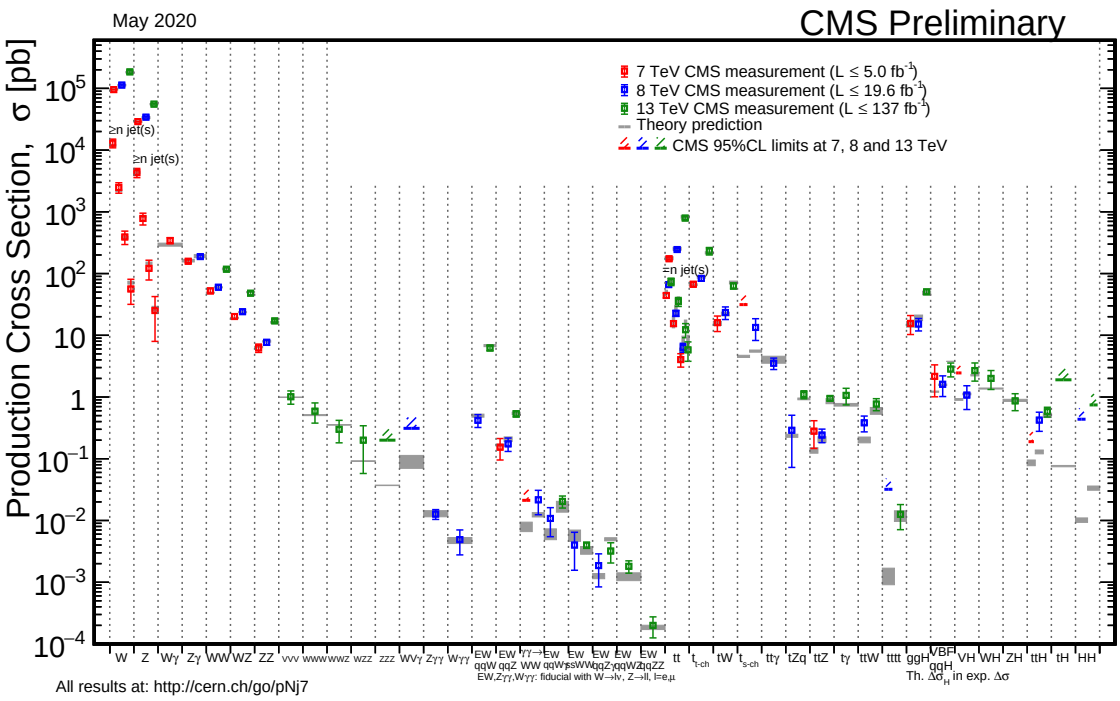


Figure 7.1: Summary of the CMS cross section measurements of standard model processes [151].

Combining results from all VVV final states, the production of three massive vector bosons has now been observed at the 5 σ level for the first time. This sensitivity milestone opens up possibilities for new standard model precision measurements, such as constraints on anomalous quartic gauge couplings in the framework of effective field theory. As no beyond-the-standard model effects have directly been observed by the LHC so far, new physics might possibly appear at energy scales beyond the LHC energy. Therefore, the future focus of triboson measurements will be on the high-mass regime, where backgrounds are reduced, and it is possible to use more inclusive selec-

tions compared to the selections used in this thesis's analysis. Semi-hadronic final states should also be considered, in particular, those signatures with unresolved jets from the decay of a boosted vector boson.

Concerning the fully leptonic WWZ $\rightarrow 4\ell$ channel as measured in this thesis, LHC measurements will still be statistically limited for many years to come. Here, the main path towards a more sensitive analysis would be at the level of the event selection, in particular the lepton selections. For the analysis presented in this thesis, the selections were chosen such that the fake-lepton induced background is suppressed. The lepton selection algorithms used in CMS are already highly optimized, which this thesis also contributed to by reoptimizing the multivariate identification of electrons for 2017 data-taking conditions. Hence, developing a reliable lepton fake rate estimation method for the four-lepton analysis would be an important next step, allowing for looser lepton selections without relying on the fake rate being well modeled in the simulation. The optimization of the multivariate electron identification presented in this thesis cumulated in the algorithm used for the analysis of the whole Run 2 dataset within the CMS collaboration.

Acronyms List

ALICE A Large Ion Collider Experiment.

AMS approximate median significance.

APD avalanche photodiode.

ASIC application-specific integrated circuit.

ATLAS A Toroidal LHC ApparatuS.

AUC Area Under Curve.

AWAKE Advanced WAKEfield Experiment.

BDT Boosted Decision Tree.

BEH Brout-Englert-Higgs.

BSM Physics beyond the Standard Model.

CASTOR Centauro and Strange Object Research Calorimeter.

CB crystal ball.

CERN European Organization for Nuclear Research (French: Organisation européenne pour la recherche nucléaire).

CHS Charged Hadron Subtraction.

CMS Compact Muon Solenoid.

CMSSW CMS Software.

CPU Central Processing Unit.

CSC Cathode Strip Chamber.

DAQ Data Aquisition.

DQM Data Quality Monitoring.

DT Drift Tube.

DY Drell-Yan.

ECAL Electromagnetic Calorimeter.

EDM Event Data Model.

ELENA Extra Low ENergy Antiproton.

FORTRAN FORmula TRANslator.

GED Global Event Description.

GP Gaussian Process.

GPU Graphics Processing Unit.

GSF Gaussian Sum Filter.

HCAL Hadronic Calorimeter.

HEP High Energy Physics.

HF Heavy Flavor.

HL-LHC High Luminosity LHC.

HLT High Level Trigger.

HPD Hybrid Photodiode.

ICFA International Committee for Future Accelerators.

ID identification.

JER Jet Energy Resolution.

JES Jet Energy Scale.

KDE Kernel Density Estimate.

KF Kalman Filter.

KS Kolmogorov-Smirnov.

L1 Level-1.

LEP Large Electron-Positron Collider.

LF Light Flavor.

LHC Large Hadron Collider.

LINAC Linear Accelerator.

LUT look-up table.

MC Monte Carlo.

MET missing transverse energy.

ML Machine Learning.

MSE mean squared error.

MSGC Micro-Strip Gas Chamber.

MVA Multivariate Analysis.

NNLS non-negative least squares.

NpC neutral-proportional-to-charge.

OOT out-of-time.

PDF Parton Density Function.

PF particle-flow.

PS Proton Synchrotron.

PSB Proton Synchrotron Booster.

PUPPI pileup per particle identification.

QCD Quantum Chromodynamics.

QED Quantum Electrodynamics.

RF radiofrequency.

ROC receiver operating characteristic.

RPC Resistive Plate Chamber.

SiPM Silicon Photomultiplier.

SPS Super Proton Synchrotron.

TEC Tracker Endcaps.

TIB Tracker Inner Barrel.

TID Tracker Inner Disks.

TMVA Toolkit for Multivariate Data Analysis with ROOT.

TOB Tracker Outer Barrel.

VBF Vector Boson Fusion.

VBS Vector Boson Scattering.

VPT vacuum phototriodes.

ZDC Zero-Degree Calorimeter.

Bibliography

[1] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[2] S. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.

[3] A. Salam, "Elementary particle physics: Relativistic groups and analyticity", in *Eighth Nobel Symposium. Stockholm*, N. Svartholm, ed., p. 367. Almquvist and Wiksell, 1968.

[4] J. Schwichtenberg, "Physics from Symmetry". Undergraduate Lecture Notes in Physics. Springer International Publishing, Cham, 2018. doi:10.1007/978-3-319-66631-0, ISBN 978-3-319-66630-3, 978-3-319-66631-0.

[5] F. Zefi, "Gamma-ray flux variation studies from the blazar B2 1215+30 with the Fermi-LAT and the Crab Nebula with the H.E.S.S. experiment". PhD thesis, Université Paris-Saclay, 2017.

[6] B. W. Lee, C. Quigg, and H. Thacker, "The Strength of Weak Interactions at Very High-Energies and the Higgs Boson Mass", *Phys. Rev. Lett.* **38** (1977) 883–885, doi:10.1103/PhysRevLett.38.883.

[7] B. W. Lee, C. Quigg, and H. Thacker, "Weak Interactions at Very High-Energies: The Role of the Higgs Boson Mass", *Phys. Rev. D* **16** (1977) 1519, doi:10.1103/PhysRevD.16.1519.

[8] J. Ellis, "Higgs Physics", in *2013 European School of High-Energy Physics*, pp. 117–168. 2015. arXiv:1312.5672. doi:10.5170/CERN-2015-004.117.

[9] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[10] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[11] H. Fritzsch and Z.-z. Xing, "Mass and flavor mixing schemes of quarks and leptons", *Prog. Part. Nucl. Phys.* **45** (2000) 1–81, doi:10.1016/S0146-6410(00)00102-2, arXiv:hep-ph/9912358.

[12] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Nonabelian Gauge Theories", *Phys. Rev. Lett.* **30** (1973) 1343–1346, doi:10.1103/PhysRevLett.30.1343.

[13] H. Politzer, "Reliable Perturbative Results for Strong Interactions?", *Phys. Rev. Lett.* **30** (1973) 1346–1349, doi:10.1103/PhysRevLett.30.1346.

[14] Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos", *Phys. Rev. Lett.* **81** (1998) 1562–1567, doi:10.1103/PhysRevLett.81.1562, arXiv:hep-ex/9807003.

[15] SNO Collaboration, "Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory", *Phys. Rev. Lett.* **87** (2001) 071301, doi:10.1103/PhysRevLett.87.071301, arXiv:nucl-ex/0106015.

[16] W. Rodejohann, "Neutrino-less Double Beta Decay and Particle Physics", *Int. J. Mod. Phys. E* **20** (2011) 1833–1930, doi:10.1142/S0218301311020186, arXiv:1106.1334.

[17] T2K Collaboration, "Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations", *Nature* **580** (2020), no. 7803, 339–344, doi:10.1038/s41586-020-2177-0, arXiv:1910.03887. [Erratum: Nature 583, E16 (2020)].

[18] Planck Collaboration, "Planck 2018 results. VI. Cosmological parameters", *Astron. Astrophys.* **641** (2020) A6, doi:10.1051/0004-6361/201833910, arXiv:1807.06209.

[19] S. Tsujikawa, "Introductory review of cosmic inflation", in *2nd Tah Poe School on Cosmology: Modern Cosmology*. 4, 2003. arXiv:hep-ph/0304257.

[20] B. Horn, "The Higgs field and early universe cosmology: a (brief) review", *MDPI Physics* **2** (2020), no. 3, 503–520, doi:10.3390/physics2030028, arXiv:2007.10377.

[21] D. J. Callaway, "Triviality Pursuit: Can Elementary Scalar Particles Exist?", *Phys. Rept.* **167** (1988) 241, doi:10.1016/0370-1573(88)90008-7.

[22] OPAL Collaboration, "W^+W^- production and triple gauge boson couplings at LEP energies up to 183-GeV", *Eur. Phys. J. C* **8** (1999) 191–215, doi:10.1007/s100529901106, arXiv:hep-ex/9811028.

[23] CDF Collaboration, "Evidence for W^+W^- production in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV", *Phys. Rev. Lett.* **78** (1997) 4536–4540, doi:10.1103/PhysRevLett.78.4536.

[24] CDF Collaboration, "Observation of WZ Production", *Phys. Rev. Lett.* **98** (2007) 161801, doi:10.1103/PhysRevLett.98.161801, arXiv:hep-ex/0702027.

[25] D0 Collaboration, "Observation of ZZ production in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96-TeV", *Phys. Rev. Lett.* **101** (2008) 171803, doi:10.1103/PhysRevLett.101.171803, arXiv:0808.0703.

[26] ATLAS Collaboration, "Search for triboson $W^\pm W^\pm W^\mp$ production in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *Eur. Phys. J. C* **77** (2017), no. 3, 141, doi:10.1140/epjc/s10052-017-4692-1, arXiv:1610.05088.

[27] ATLAS Collaboration, "Evidence for the production of three massive vector bosons with the ATLAS detector", *Phys. Lett.* **B798** (2019) 134913, doi:10.1016/j.physletb.2019.134913, arXiv:1903.10415.

[28] CMS Collaboration, "Search for the production of $\mathrm{W}^\pm\mathrm{W}^\pm\mathrm{W}^\mp$ events at $\sqrt{s} = 13$ TeV", *Phys. Rev. D* **100** (2019), no. 1, 012004, doi:10.1103/PhysRevD.100.012004, arXiv:1905.04246.

[29] Y.-B. Shen et al., "NLO QCD and electroweak corrections to WWW production at the LHC", *Phys. Rev.* **D95** (2017), no. 7, 073005, doi:10.1103/PhysRevD.95.073005, arXiv:1605.00554.

[30] D. T. Nhung, L. D. Ninh, and M. M. Weber, "NLO corrections to WWZ production at the LHC", *JHEP* **12** (2013) 096, doi:10.1007/JHEP12(2013)096, arXiv:1307.7403.

[31] Y.-B. Shen et al., "NLO QCD + NLO EW corrections to WZZ productions with leptonic decays at the LHC", *JHEP* **10** (2015) 186, doi:10.1007/JHEP10(2016)156,10.1007/JHEP10(2015)186, arXiv:1507.03693. [Erratum: JHEP10,156(2016)].

[32] H. Wang et al., "NLO QCD + EW corrections to ZZZ production with subsequent leptonic decays at the LHC", *J. Phys.* **G43** (2016), no. 11, 115001, doi:10.1088/0954-3899/43/11/115001, arXiv:1610.05876.

[33] A. Lazopoulos, K. Melnikov, and F. Petriello, "QCD corrections to tri-boson production", *Phys. Rev. D* **76** (2007) 014001, doi:10.1103/PhysRevD.76.014001, arXiv:hep-ph/0703273.

[34] T. Binoth, G. Ossola, C. Papadopoulos, and R. Pittau, "NLO QCD corrections to tri-boson production", *JHEP* **06** (2008) 082, doi:10.1088/1126-6708/2008/06/082, arXiv:0804.0350.

[35] V. Hankele and D. Zeppenfeld, "QCD corrections to hadronic WWZ production with leptonic decays", *Phys. Lett. B* **661** (2008) 103–108, doi:10.1016/j.physletb.2008.02.014, arXiv:0712.3544.

[36] F. Campanario et al., "QCD corrections to charged triple vector boson production with leptonic decay", *Phys. Rev. D* **78** (2008) 094012, doi:10.1103/PhysRevD.78.094012, arXiv:0809.0790.

[37] T. Han and S. Willenbrock, "QCD correction to the pp $\to$ WH and ZH total cross-sections", *Phys. Lett.* **B273** (1991) 167–172, doi:10.1016/0370-2693(91)90572-8.

[38] O. Brein, A. Djouadi, and R. Harlander, "NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders", *Phys. Lett. B* **579** (2004) 149–156, doi:10.1016/j.physletb.2003.10.112, arXiv:hep-ph/0307206.

[39] M. Ciccolini, S. Dittmaier, and M. Kramer, "Electroweak radiative corrections to associated WH and ZH production at hadron colliders", *Phys. Rev. D* **68** (2003) 073003, doi:10.1103/PhysRevD.68.073003, arXiv:hep-ph/0306234.

[40] Particle Data Group Collaboration, "Review of Particle Physics", *Phys. Rev. D* **98** (2018), no. 3, 030001, doi:10.1103/PhysRevD.98.030001.

[41] S. Weinberg, "Phenomenological Lagrangians", *Physica A Statistical Mechanics and its Applications* **96** (1979), no. 1, 327–340, doi:10.1016/0378-4371(79)90223-1.

[42] C. Degrande et al., "Effective Field Theory: A Modern Approach to Anomalous Couplings", *Annals Phys.* **335** (2013) 21–32, doi:10.1016/j.aop.2013.04.016, arXiv:1205.4231.

[43] O.J.P. Eboli, "Quartic anomalous gauge couplings", 2012. feynrules.irmp.ucl.ac.be/attachment/wiki/AnomalousGaugeCoupling/quartic.pdf, Last accessed on 2020-09-21.

[44] ALICE Collaboration, "Transverse sphericity of primary charged particles in minimum bias proton-proton collisions at $\sqrt{s} = 0.9$, 2.76 and 7 TeV", *Eur. Phys. J. C* **72** (2012) 2124, doi:10.1140/epjc/s10052-012-2124-9, arXiv:1205.3963.

[45] CMS Collaboration, "Limits on anomalous triple and quartic gauge couplings", 2020. https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSMPaTGC, Last accessed on 2020-09-20.

[46] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[47] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.

[48] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.

[49] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[50] C. L. Smith, "Genesis of the Large Hadron Collider", *Philosophilcal Transactions of the Royal Society A* **373** (2015) doi:10.1098/rsta.2014.0037.

[51] ALICE Collaboration, "First proton-proton collisions at the LHC as observed with the ALICE detector: Measurement of the charged particle pseudorapidity density at $\sqrt{s} = 900\,\text{GeV}$", *Eur. Phys. J. C* **65** (2010) 111–125, doi:10.1140/epjc/s10052-009-1227-4, arXiv:0911.5430.

[52] L. Arnaudon et al., "Linac4 technical design report", Technical Report CERN-AB-2006-084, CERN, Geneva, 2006.

[53] W. Weterings et al., "160 MeV H- injection into the CERN PSB", in *2007 IEEE Particle Accelerator Conference (PAC)*, pp. 1628–1630. 2007. doi:10.1109/PAC.2007.4440845.

[54] E. Mobs, "The CERN accelerator complex. Complexe des accélérateurs du CERN", Jul, 2016. General Photo. https://cds.cern.ch/record/2197559.

[55] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.

[56] TOTEM Collaboration, "First measurement of elastic, inelastic and total cross-section at $\sqrt{s} = 13$ TeV by TOTEM and overview of cross-section data at LHC energies", *Eur. Phys. J. C* **79** (2019), no. 2, 103, doi:10.1140/epjc/s10052-019-6567-0, arXiv:1712.06153.

[57] "High Luminosity LHC Project". https://hilumilhc.web.cern.ch/content/hl-lhc-project, Last accessed on 2020-06-10.

[58] M. Pimiaet al., "Compact muon solenoid", in *ECFA Large Hadron Collider (LHC) Workshop: Physics and Instrumentation*, pp. 547–563. 12, 1990.

[59] CMS Collaboration, "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software", Technical Report CERN-LHCC-2006-001, 2006.

[60] A. Sharma, M. Della Negra and Alain Hervé, "CMS Times, December 18", 2006. http://cmsinfo.web.cern.ch/cmsinfo/Media/Publications/CMStimes/2006/12_18, Last accessed on 2020-05-04.

[61] CMS Collaboration, "CMS Phase II Upgrade Scope Document", Technical Report CERN-LHCC-2015-019, 2015.

[62] CMS Collaboration, "Technical proposal for the upgrade of the CMS detector through 2020", Technical Report CERN-LHCC-2011-006, 2011.

[63] CMS Collaboration, "Performance of muon identification in pp collisions at $\sqrt{s}$ = 7 TeV", Physics Analysis Summary CMS-PAS-MUO-10-002, 2010.

[64] S. Gennai et al., "Tau jet reconstruction and tagging at high level trigger and off-line", *Eur. Phys. J. C* **46S1** (2006) 1–21, doi:10.1140/epjcd/s2006-02-001-y.

[65] CMS Collaboration, "The CMS tracker system project: Technical Design Report", Technical Report CERN-LHCC-98-006, 1997.

[66] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[67] CMS Collaboration, "CMS Technical Design Report for the Pixel Detector Upgrade", Technical Report CERN-LHCC-2012-016, 2012. doi:10.2172/1151650.

[68] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report", Technical Report CERN-LHCC-97-033, 1997.

[69] A. Zghiche, "Performance of the CMS electromagnetic calorimeter during the LHC run II", *J. Phys. Conf. Ser.* **1162** (2019) 012001, doi:10.1088/1742-6596/1162/1/012001.

[70] CMS Collaboration, "CMS ECAL Response to Laser Light", Detector Performance Note CMS-DP-2019-005, 2019.

[71] CMS Collaboration, "Reconstruction of signal amplitudes in the CMS electromagnetic calorimeter in the presence of overlapping proton-proton interactions", *JINST* **15** (2020), no. 10, P10002, doi:10.1088/1748-0221/15/10/P10002, arXiv:2006.14359.

[72] CMS Collaboration, "CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter", Technical Report CERN-LHCC-2012-015, 2012. doi:10.2172/1151651.

[73] CMS Collaboration, P. Gunnellini, "The CASTOR calorimeter at the CMS experiment", in *16th International Moscow School of Physics and 41st ITEP Winter School of Physics*. 4, 2013. arXiv:1304.2943.

[74] CMS Collaboration, "Status of zero degree calorimeter for CMS experiment", *AIP Conf. Proc.* **867** (2006), no. 1, 258–265, doi:10.1063/1.2396962, arXiv:nucl-ex/0608052.

[75] CMS Collaboration, "The CMS muon project: Technical Design Report", Technical Report CERN-LHCC-97-032, 1997.

[76] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017), no. 01, P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

[77] CMS Collaboration, "CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger", Technical Report CERN-LHCC-2002-026, 2002.

[78] C. Amstutz et al., "An FPGA-based track finder for the L1 trigger of the CMS experiment at the high luminosity LHC", in *20th IEEE-NPSS Real Time Conference*. 2016. doi:10.1109/RTC.2016.7543102.

[79] CMS Collaboration, "CMS Technical Design Report for the Level-1 Trigger Upgrade", Technical Report CERN-LHCC-2013-011, 2013.

[80] CMS Collaboration, "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV", Physics Analysis Summary CMS-PAS-LUM-18-002, 5, 2019.

[81] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017), no. 10, P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[82] J. Rembser, "CMS electron and photon performance at 13 TeV", *J. Phys. Conf. Ser.* **1162** (2019) 012008, doi:10.1088/1742-6596/1162/1/012008.

[83] CMS Collaboration, "Electron and Photon performance in CMS with the full 2017 data sample and additional 2016 highlights for the CALOR 2018 Conference", Detector Performance Note CMS-DP-2018-017, 2018.

[84] CMS Collaboration, "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s}$ = 8 TeV", *JINST* **10** (2015), no. 08, P08010, doi:10.1088/1748-0221/10/08/P08010, arXiv:1502.02702.

[85] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s}$ = 8 TeV", *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[86] W. Adam, R. Frühwirth, A. Strandlie, and T. Todor, "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC", Technical Report CERN-CMS-NOTE-2005-001, 2005.

[87] S. Baffioni et al., "Electron reconstruction in CMS", *Eur. Phys. J. C* **49** (2007) 1099–1116, doi:10.1140/epjc/s10052-006-0175-5.

[88] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system", in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pp. 785–794. ACM, New York, NY, USA, 2016. arXiv:1603.02754. doi:10.1145/2939672.2939785.

[89] CMS Collaboration, "Performance of CMS Muon Reconstruction in pp Collision Events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[90] CMS Collaboration, "Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV", Physics Analysis Summary CMS-PAS-PFT-10-003, 2010.

[91] CMS Collaboration, "Boosted jet identification using particle candidates and deep neural networks", Detector Performance Note CMS-DP-2017-049, 2017.

[92] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[93] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions", *Phys. Rev. D* **48** (1993) 3160–3166, doi:10.1103/PhysRevD.48.3160, arXiv:hep-ph/9305266.

[94] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017), no. 02, P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[95] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[96] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector", *JINST* **14** (2019), no. 07, P07004, doi:10.1088/1748-0221/14/07/P07004, arXiv:1903.06078.

[97] CMS Collaboration, "Performance of b tagging algorithms in proton-proton collisions at 13 TeV with Phase 1 CMS detector", Detector Performance Note CMS-DP-2018-033, 2018.

[98] B. Efron and T. Hastie, "Computer Age Statistical Inference". Institute of Mathematical Statistics Monographs. Cambridge University Press, 2016. ISBN 9781107149892.

[99] M. Cacciari, G. P. Salam, and G. Soyez, "Use of charged-track information to subtract neutral pileup", *Phys. Rev. D* **92** (2015), no. 1, 014003, doi:10.1103/PhysRevD.92.014003, arXiv:1404.7353.

[100] CMS Collaboration, "Measurement of charged particle spectra in minimum-bias events from proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Eur. Phys. J. C* **78** (2018), no. 9, 697, doi:10.1140/epjc/s10052-018-6144-y, arXiv:1806.11245.

[101] S. Höche, "Introduction to parton-shower event generators", in *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, pp. 235–295. 2015. arXiv:1411.4085. doi:10.1142/9789814678766_0005.

[102] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[103] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[104] 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.

[105] M. L. Mangano et al., "ALPGEN, a generator for hard multiparton processes in hadronic collisions", *JHEP* **07** (2003) 001, doi:10.1088/1126-6708/2003/07/001, arXiv:hep-ph/0206293.

[106] S. Hoeche et al., "Matching parton showers and matrix elements", in *HERA and the LHC: A Workshop on the Implications of HERA for LHC Physics: CERN - DESY Workshop 2004/2005 (Midterm Meeting, CERN, 11-13 October 2004; Final Meeting, DESY, 17-21 January 2005)*, pp. 288–289. 2005. arXiv:hep-ph/0602031. doi:10.5170/CERN-2005-014.288.

[107] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[108] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics", *Phys. Rept.* **97** (1983) 31–145, doi:10.1016/0370-1573(83)90080-7.

[109] B. Andersson, "The Lund model", volume 7. Cambridge University Press, 7, 2005. ISBN 978-0-521-01734-3, 978-0-521-42094-5, 978-0-511-88149-7.

[110] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", *Eur. Phys. J. C* **80** (2020), no. 1, 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.

[111] GEANT4 Collaboration, "GEANT4 – a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.

[112] CMS Collaboration, "CMSSW". https://github.com/cms-sw/cmssw, 2020.

[113] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework", *Nucl. Instrum. Meth. A* **389** (1997) 81–86, doi:10.1016/S0168-9002(97)00048-X.

[114] C. Adam-Bourdarioset al., "The Higgs boson machine learning challenge", volume 42 of *Proceedings of Machine Learning Research*, pp. 19–55. PMLR, Montreal, Canada, 2015.

[115] F. Pedregosa et al., "Scikit-learn: Machine learning in Python", *Journal of Machine Learning Research* **12** (2011) 2825–2830.

[116] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting", *J. Comput. Syst. Sci.* **55** (1997), no. 1, 119–139, doi:10.1006/jcss.1997.1504.

[117] J. H. Friedman, "Stochastic gradient boosting", *Comput. Stat. Data Anal.* **38** (2002) 367–378, doi:10.1016/S0167-9473(01)00065-2.

[118] E. A. Moreno et al., "Interaction networks for the identification of boosted $H \to b\bar{b}$ decays", *Phys. Rev. D* **102** (2020), no. 1, 012010, doi:10.1103/PhysRevD.102.012010, arXiv:1909.12285.

[119] H. Voss, A. Höcker, J. Stelzer, and F. Tegenfeldt, "TMVA, the toolkit for multivariate data analysis with ROOT", in *XIth International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT)*, p. 40. 2007. arXiv:physics/0703039. [PoS(ACAT)040]. doi:10.22323/1.050.0040.

[120] P. Pigard, "Electron studies and search for vector boson scattering in events with four leptons and two jets with the CMS detector at the LHC". PhD thesis, Ecole Polytechnique, 2017.

[121] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[122] J. Rembser, "XGBoost-FastForest", October, 2020. doi:10.5281/zenodo.4109412, https://doi.org/10.5281/zenodo.4109412.

[123] E. Brochu, V. M. Cora, and N. de Freitas, "A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning", 2010.

[124] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization", in *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor et al., eds., pp. 2546–2554. Curran Associates, Inc., 2011.

[125] C. K. I. W. Carl Edward Rasmussen, "Gaussian Processes for Machine Learning". Institute of Mathematical Statistics Monographs. The MIT Press, 2006. ISBN 0-262-18253-X.

[126] F. Nogueira, "Bayesian Optimization: Open source constrained global optimization tool for Python", 2014–2020. https://github.com/fmfn/BayesianOptimization.

[127] J. Bergstra, D. Yamins, and D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures", in *Proceedings of the 30th International Conference on Machine Learning*, S. Dasgupta and D. McAllester, eds., volume 28 of *Proceedings of Machine Learning Research*, pp. 115–123. PMLR, Atlanta, Georgia, USA, 17–19 Jun, 2013.

[128] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018), no. 05, P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

[129] CMS Collaboration, "Search for top squark pair production in the single lepton final state in pp collisions at $\sqrt{s} = 13$ TeV", Physics Analysis Summary CMS-PAS-SUS-16-051, 2017.

[130] CMS Collaboration, "Measurement of the $Z\gamma^* \to \tau\tau$ cross section in pp collisions at $\sqrt{s} = 13$ TeV and validation of τ lepton analysis techniques", *Eur. Phys. J. C* **78** (2018), no. 9, 708, doi:10.1140/epjc/s10052-018-6146-9, arXiv:1801.03535.

[131] C. Lester and D. Summers, "Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders", *Phys. Lett. B* **463** (1999) 99–103, doi:10.1016/S0370-2693(99)00945-4, arXiv:hep-ph/9906349.

[132] CMS Collaboration, "Search for the associated production of the Higgs boson with a top-quark pair", *JHEP* **09** (2014) 087, doi:10.1007/JHEP09(2014)087, arXiv:1408.1682. [Erratum: JHEP 10, 106 (2014)].

[133] J. Brehmer and K. Cranmer, "Simulation-based inference methods for particle physics", (2020). arXiv:2010.06439.

[134] K. Cranmer, J. Pavez, and G. Louppe, "Approximating Likelihood Ratios with Calibrated Discriminative Classifiers", (2015). arXiv:1506.02169.

[135] K. Cranmer, J. Pavez, G. Louppe, and W. Brooks, "Experiments using machine learning to approximate likelihood ratios for mixture models", *J. Phys. Conf. Ser.* **762** (2016), no. 1, 012034, doi:10.1088/1742-6596/762/1/012034.

[136] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[137] M. Botje et al., "The PDF4LHC Working Group Interim Recommendations", (2011). arXiv:1101.0538.

[138] S. Alekhin et al., "The PDF4LHC Working Group Interim Report", (2011). arXiv:1101.0536.

[139] K. Cranmer, "Practical Statistics for the LHC", in *2011 European School of High-Energy Physics*, pp. 267–308. 2014. arXiv:1503.07622. doi:10.5170/CERN-2014-003.267.

[140] W. Verkerke and D. P. Kirkby, "The RooFit toolkit for data modeling", *eConf* **C0303241** (2003) MOLT007, arXiv:physics/0306116.

[141] L. Moneta et al., "The RooStats Project", *PoS* **ACAT2010** (2010) 057, doi:10.22323/1.093.0057, arXiv:1009.1003.

[142] F. James, "MINUIT: Function Minimization and Error Analysis Reference Manual", (1998). CERN Program Library Long Writeups.

[143] H. Dembinski et al., "scikit-hep/iminuit: v1.4.9", July, 2020. doi:10.5281/zenodo.3951328, https://doi.org/10.5281/zenodo.3951328.

[144] S. Hageboeck, "A Faster, More Intuitive RooFit", in *24th International Conference on Computing in High Energy and Nuclear Physics*. 3, 2020. arXiv:2003.12875.

[145] M. Abadiet al., "TensorFlow: Large-scale machine learning on heterogeneous systems", 2015. Software available from tensorflow.org. https://www.tensorflow.org/.

[146] J. Eschle, A. Puig Navarro, R. Silva Coutinho, and N. Serra, "zfit: scalable pythonic fitting", *SoftwareX* **11** (2019) 100508, doi:10.1016/j.softx.2020.100508, arXiv:1910.13429.

[147] R. J. Barlow, "Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences (Manchester Physics Series)". WileyBlackwell, reprint edition, 1989. ISBN 0471922951.

[148] G. Cowan, "Statistical data analysis". Oxford University Press, USA, 1998.

[149] B. Cousins, "Probability density functions for positive nuisance parameters", 2010. http://www.physics.ucla.edu/~cousins/stats/cousins_lognormal_prior.pdf, Last accessed on 2020-08-12.

[150] CMS Collaboration, "Observation of the Production of Three Massive Gauge Bosons at $\sqrt{s}$ =13 TeV", *Phys. Rev. Lett.* **125** (2020), no. 15, 151802, doi:10.1103/PhysRevLett.125.151802, arXiv:2006.11191.

[151] CMS Collaboration, "Summaries of CMS cross section measurements", 2020. https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined, Last accessed on 2020-09-18.

ECOLE DOCTORALE

Titre : Mesure de la production de trois bosons massifs et identification des électrons avec le détecteur CMS auprès du LHC

Mots clés : production de tribosons, identification des électrons, éxperience CMS, LHC, modèle standard, apprentissage profond

Résumé : Cette thèse présente la mesure de la production de trois bosons de jauge massifs dans des collisions entre protons à 13 TeV avec l'expérience CMS auprès du Grand collisionneur de hadrons (LHC) du CERN. Le travail de cette thèse se concentre sur la production de deux bosons W et d'un boson Z (WWZ), où les trois bosons se désintègrent en électrons ou en muons. Ce processus a un bruit de fond réductible relativement faible car le boson Z est entièrement reconstruit, et une section efficace plus élevée que les processus avec plusieurs bosons Z. L'analyse couvre 137 fb^{-1}de données enregistrées pendant la deuxième phase d'exploitation du LHC (2015 – 2018).
La production de trois bosons de jauge massifs n'a pas encore été observé, mais elle est prédit par le modèle standard de la physique des particules. Ces mesures sont donc un test du modèle standard et un défis de sensibilité. Ce processus est sensible aux couplages de jauge trilinéaire et quartiques, qui pourraient être modifiés par la physique au-delà du modèle standard. En outre, la production de tribosons est un bruit de fond de plus en plus important pour les recherches directes de nouvelles physiques et les mesures du modèle standard.
Les couplages de jauge trilinéaires sont également accessibles dans le processus diboson, mais le sondage des couplages quartiques nécessite la mesure de la diffusion de bosons massifs (VBS) ou de la production de tribosons. Le processus VBS isole bien les auto-couplages des bosons massifs, mais la production de triboson est dominée par des diagrammes sans aucun auto-couplage. Cependant, cela joue un rôle mineur dans les recherches de nouvelles physiques aux échelles d'énergie au-delà de l'échelle électrofaible, où la production de tribosons a l'avantage de sonder un état final plus massif.
L'analyse présentée établit la production WWZ avec une signification statistique de 3.35 écarts-types, qui est réduit à 1.31 écarts-types si les bosons sont non-virtuels. Combinée avec d'autres analyses couvrant les autres états finals de tribosons, cette mesure a contribué à une observation d'une signification statistique de 5.67 écarts-types de la production de trois bosons de jauge massifs (2.90 écarts-types pour les bosons non-virtuels). Les résultats sont en accord avec les prédictions du modèle standard. De même, cette thèse présente des études sur la sensibilité du processus WWZ aux couplages de jauge quartiques anormaux dans le cadre d'une théorie effective des champs.
Comme la production de triboson est un processus rare, sa détection requiert une bonne efficacité d'identification des électrons et des muons. Parmi ces deux types de leptons, les électrons sont les plus difficiles à mesurer. Par conséquent, ce travail de thèse inclut une amélioration rigoureuse de la sélection d'électrons multivariée basée sur l'algorithme Boosted Decision Tree (BDT), utilisé dans le cadre de la collaboration CMS depuis la première période d'exploitation du LHC. Cette optimisation permet ainsi de conserver une identification des électrons très performante dans la deuxième période malgré le fait que le nombre d'interactions proton-proton par croisement de faisceaux de particules a considérablement augmenté. L'algorithme d'identification d'électrons multivariée développé dans cette thèse est désormais devenu l'algorithme standard dans les analyses de CMS pour toutes les données de la deuxième phase du LHC.

Title : Measurement of triboson production in the multilepton final state and electron studies with the CMS experiment at the LHC

Keywords : triboson production, electron identification, CMS experiment, LHC, standard model, deep learning

Abstract : This thesis reports on the measurement of the production of three massive gauge bosons in proton-proton collisions at 13 TeV with the CMS experiment at the CERN Large Hadron Collider (LHC). The thesis work focuses on the production of two W bosons and one Z boson (WWZ), decaying to either electrons or muons. This process has relatively little reducible background because of the fully reconstructed Z boson, and a higher production cross-section than processes with more Z bosons. The measurement considers 137 $fb^{-1}$of collision data recorded during the the second LHC run (2015 – 2018).
The production of three massive gauge bosons is a so-far unobserved phenomenon predicted by the standard model of particle physics. The process is sensitive to the trilinear and quartic gauge couplings, which might be modified by physics beyond the standard model. Additionally, triboson production is an increasingly important background to both direct searches for new physics and standard model precision measurements.
While trilinear gauge couplings are also accessible via diboson production, probing quartic couplings requires the measurement of Vector Boson Scattering (VBS) or triboson production. The VBS process better isolates the vector boson self-couplings, whereas triboson production is mainly affected by diagrams without any vector boson self-interaction. However, this plays a minor role in searches for new physics at energy scales beyond the electroweak scale, where triboson production has the advantage of probing a more massive final state.
The analysis presented in this thesis establishes evidence for the WWZ final state with a significance of 3.35 standard deviations (1.31 standard deviations if the vector bosons are required to be on-shell). Combined with other analyses covering the remaining massive triboson final states, this measurement contributed to observing the production of three massive gauge bosons with a significance of 5.67 standard deviations (2.90 standard deviations for on-shell vector bosons). All results are in agreement with the standard model predictions. Beyond that, this thesis presents studies of the WWZ processes sensitivity to anomalous quartic gauge couplings in an effective field theory framework.
Triboson production is a rare process that requires high electron and muon identification efficiencies. Among these two types of leptons, electrons are more challenging to measure. Therefore, this thesis includes a rigorous update and optimization of the multivariate electron selection based on the Boosted Decision Tree (BDT) algorithm, applied within the CMS collaboration since the beginning of LHC data-taking. As a result, the identification of electrons remains very performant, even at a significant increase in the number of proton-proton interactions per bunch crossing. The multivariate electron identification algorithm produced for this thesis is now the default for CMS analyses ot the second LHC run.

Institut Polytechnique de Paris
91120 Palaiseau, France