



FACULTY OF SCIENCES
DEPARTMENT OF PHYSICS AND ASTRONOMY

Exploring top-quark and Z boson coupling in top pair production in association with a Z boson

Steven Mortier

Promotor: Prof. Dr. Didar Dobur

Supervisor: Dr. Marek Niedziela



Thesis submitted in partial fulfilment of the requirements for the degree of
MASTER OF SCIENCE IN PHYSICS AND ASTRONOMY

Academic year 2019-2020

We are just an advanced breed of monkeys on a minor planet of a very average star. But we can understand the Universe. That makes us something very special.

-

Stephen Hawking

Acknowledgements

Firstly, I would like to thank my promotor, Didar Dobur, for giving me the amazing opportunity to perform an internship at CERN, which was an incredible experience I will never forget. I would also like to thank her for defining this interesting thesis topic for me to work on.

I also wish to express my sincere gratitude to my supervisor, Marek, for helping me in whichever way he could, be it early in the morning or past midnight. It must have been quite a challenge to handle my numerous questions, but I am very thankful he did. I would also like to thank Willem, for assisting me in machine learning related topics and general issues as well.

Next, I would like to thank Bernd, Ian and Margot, for the table tennis breaks and general fun conversations. On this note, I would like to express my gratitude to everyone at the INW for creating a generally pleasant and motivating work environment. Even though the discussion topics over lunch did not always exactly suit lunch, they were certainly interesting.

I would also like to thank a group of friends who are very special to me, some of whom I've known since I was merely six years old. Thank you, Christiaan, Jelle, *Pally*, Reinier, Sam, Simon and Wolfgang, for being an amazing group of friends.

I am deeply grateful to my parents, brother and sister. Without them, none of this would have ever been possible. Thank you for enabling me to do what I like to do and supporting me no matter what.

Last but not least I would like to thank my girlfriend, Luna, for always being there for me when I needed her most. Her unconditional love and support truly is what kept me going in times when I was struggling. We might not know what the future holds, but I know we can overcome everything together.

Steven Mortier, June 2020

Summary

During the past century, humanity’s understanding of the processes happening at the smallest scales apparent in our universe has grown tremendously. This progress is summarized in the standard model of particle physics, which describes the elementary particles and three of their four fundamental interactions. Although the standard model is a remarkably successful theory, it leaves some phenomena unexplained. The purpose of this thesis is to study a tiny piece of this very large puzzle, contributing to the efforts of thousands of theoretical and experimental physicists.

The first chapters of this thesis aim to establish a general framework within which this thesis is performed. We will start off by discussing the standard model and its imperfections. Special attention is given to the heaviest particle apparent in the standard model, the top quark. Hereafter, a description of the CMS detector follows, as this thesis was performed as a part of the CMS collaboration. Simultaneously, we will discuss how physical objects are reconstructed from raw detector data. Subsequently, an introduction to machine learning is presented, as machine learning techniques play a key role in the analysis presented in this thesis. The two machine learning algorithms used in this thesis, boosted decision trees and neural networks, are covered in more detail.

This thesis presents a measurement of the inclusive production cross section of top quark pair production in association with a Z boson in the two-lepton final state. The analysis demonstrated in the penultimate chapter utilizes data from proton-proton collisions at a centre-of-mass energy of 13 TeV collected by the CMS detector in 2018. We combine event selection criteria with machine learning algorithms in order to achieve a good distinction between signal and background. A fit is performed by comparing the output distributions of the machine learning techniques of simulated Monte-Carlo data with actual detector data, taking into account a multitude of sources of systematic uncertainty. This fit results in the signal strength, which can then be used to calculate the production cross section of the process under study. All results of the analysis presented in this thesis are discussed in the final chapter.

Nederlandstalige samenvatting

Gedurende de voorbije eeuw is het menselijk inzicht in de processen die zich afspelen op de kleinste schalen in ons universum enorm toegenomen. Al deze kennis zit vervat in een elegant wiskundig model, namelijk het standaardmodel van de deeltjesfysica. Dit model beschrijft de fundamentele deeltjes en hun interacties. Het standaardmodel is een uitzonderlijk succesvolle theorie, maar faalt toch bij de beschrijving van sommige fenomenen. Het doel van deze thesis is om een klein stukje van deze enorme puzzel te onderzoeken en op deze manier bij te dragen aan de inspanningen van duizenden theoretische en experimentele fysici.

De eerste hoofdstukken van deze thesis schetsen een algemeen kader waarbinnen deze thesis tot stand kwam. Om te beginnen wordt het standaardmodel besproken, samen met zijn onvolkomenheden. Er wordt speciale aandacht besteed aan het zwaarste deeltje aanwezig in het standaardmodel, de top quark. Daarna volgt een beschrijving van de CMS detector, aangezien deze thesis werd uitgevoerd als onderdeel van de CMS samenwerking. In hetzelfde hoofdstuk bespreken we ook hoe we fysische objecten kunnen reconstrueren aan de hand van onbewerkte detectordata. Vervolgens wordt er een inleiding tot machine learning gegeven, aangezien machine learning een sleutelrol speelt in het onderzoek dat werd uitgevoerd in het kader van deze thesis. De twee algoritmen die gebruikt worden in deze thesis, *boosted decision trees* en *neural networks*, worden in meer detail besproken.

Deze thesis bespreekt een meting van de werkzame doorsnede van de productie van een top-antitop quark paar in combinatie met een Z boson, gemeten in de eindtoestand met twee leptonen. De analyse die geïllustreerd wordt in het voorlaatste hoofdstuk maakt gebruik van data van proton-proton botsingen, verzameld door de CMS detector in 2018 bij een zwaartepuntsenergie van 13 TeV. We hebben zorgvuldig gekozen selectiecriteria gecombineerd met machine learning algoritmen om een zo goed mogelijk onderscheid tussen signaal en achtergrond te kunnen maken. De output-verdelingen van de machine learning algoritmen van gesimuleerde Monte-Carlo data werden gefit aan detectordata, rekening houdend met verschillende bronnen van systematische

onzekerheid. Dit resulteerde in de signaalsterkte, die vervolgens gebruikt kon worden om de werkzame doorsnede van het onderzochte proces te berekenen. Alle resultaten van dit onderzoek worden besproken in het laatste hoofdstuk.

Contents

List of Figures	xiii
List of Tables	xvii
List of Abbreviations	xix
1 Introduction	1
1.1 Notations and conventions	2
2 The Standard Model	3
2.1 Particles and interactions	3
2.1.1 Fermions	3
2.1.2 Gauge bosons and fundamental forces	5
2.1.2.1 Electromagnetic interaction	5
2.1.2.2 Weak interaction	5
2.1.2.3 Electroweak unification	6
2.1.2.4 Strong interaction	6
2.1.3 The Higgs boson	7
2.2 Limitations of the standard model	7
2.2.1 Gravity	7
2.2.2 Dark matter and dark energy	8
2.2.3 The hierarchy problem	8
2.2.4 Matter-antimatter asymmetry	8
2.2.5 Neutrino masses	9
2.3 The top quark	9

3	The CMS Detector at the LHC	11
3.1	The LHC	11
3.2	The CMS detector	13
3.2.1	Coordinate system and basic variables	13
3.2.2	Tracker	14
3.2.3	Electromagnetic Calorimeter	15
3.2.4	Hadronic Calorimeter	17
3.2.5	Muon system	18
3.3	Triggering and DAQ	19
3.4	Object and event reconstruction: the particle flow algorithm	20
3.4.1	Iterative tracking	20
3.4.2	Calorimeter clustering	21
3.4.3	Link algorithm	21
3.4.4	Particle reconstruction and identification	22
4	Machine Learning	25
4.1	Introduction to Machine Learning	25
4.2	Boosted Decision Trees	28
4.2.1	Decision trees	28
4.2.2	Gradient boosting	29
4.3	Deep Neural Networks	32
4.4	Genetic algorithms	36
5	Search for the two lepton decay mode of the $t\bar{t}Z$ process	39
5.1	Signal signature	39
5.1.1	Production mechanisms	40
5.1.2	Decay mechanisms	40
5.1.3	Signature of the two lepton decay mode	42
5.2	Dominant backgrounds	44
5.2.1	Drell-Yan	44
5.2.2	$t\bar{t}$	45
5.2.3	Other backgrounds	45
5.3	Monte Carlo samples	46

5.4	Lepton identification	46
5.4.1	Kinematic cuts	47
5.4.2	Isolation cuts	48
5.4.3	Vertex cuts	48
5.4.4	Lepton MVA	49
5.4.5	Extra muon identification cut	50
5.5	Jet identification	50
5.6	Event selection	51
5.7	Triggers	52
5.8	Signal region	52
5.9	Control regions	54
5.9.1	Drell-Yan	54
5.9.2	$t\bar{t}$	55
5.10	Cutflow	56
5.11	Discriminating variables	58
5.12	Machine learning algorithms	60
5.12.1	Input variables	61
5.12.2	Optimization of hyperparameters	62
5.12.3	Boosted decision tree	64
5.12.4	Deep neural network	65
5.13	Cross section calculation using the <i>Higgs combine tool</i>	66
5.13.1	The combine tool	66
5.13.2	Systematic uncertainties	67
5.13.2.1	Rate uncertainties	67
5.13.2.2	Shape uncertainties	68
5.13.3	Cross section calculation	69
6	Conclusions and outlook	75
6.1	Future prospects	75
	Appendices	77
A	Used MC samples	79

B	Signal region plots	83
C	Drell-Yan control region plots	87
D	$t\bar{t}$ control region plots	89
E	Used hyperparameters for the BDT and NN	93
F	Shape uncertainty corrections	97
G	Science popularization	101
	Bibliography	107

List of Figures

2.1	The standard model particles.	4
2.2	Results of a fit of the SM parameters as a function of the W boson and top quark masses. The contours at 68% and 95% CL for the fit including (blue) and excluding (grey) the M_H measurement can be compared to direct measurements (green) of the M_t and M_W	10
3.1	Schematic overview of the CERN accelerator complex. All pre-accelerators are shown, as well as the four main LHC experiments: CMS, ATLAS, ALICE and LHCb.	12
3.2	Overview of the CMS detector depicting its different parts. A person is displayed to serve as a reference scale.	13
3.3	The CMS coordinate system.	14
3.4	A cross-section of the CMS detector. The parts of the detector in which particles typically deposit their energy are also depicted.	15
3.5	One fourth of the cross section of the tracker along the z-axis. Solid blue lines represent pixel detectors. Solid pink lines represent single-sided silicon strip modules, while void blue lines represent double-sided strip modules. The tracker's coverage of pseudorapidity values up to 2.5 is also visualized.	16
3.6	Transverse section of the ECAL showing the barrel, preshower and endcaps. . . .	17
3.7	The different parts of the HCAL. In this schematic overview, HB is the hadronic barrel, HE the hadronic endcaps, HF the hadronic forward calorimeters and HO the hadronic outer calorimeter.	17
3.8	Layout of the muon chambers.	19
4.1	The bias-variance trade-off	27

4.2	A very simple example of a decision tree.	28
4.3	Neural network with one hidden layer	33
4.4	The AUC with respect to the learning rate for six generations of models.	37
5.1	Leading order Feynman diagrams of the $t\bar{t}Z$ process. Figure (a) shows $t\bar{t}Z$ production from initial state radiation, while figures (b), (c) and (d) show final state radiation production mechanisms.	41
5.2	A pie chart showing the distribution of the different final states resulting from the $t\bar{t}Z$ process. OS indicates opposite sign leptons, while SS signifies same sign leptons.	43
5.3	The Feynman diagram for the decay channel where the Z boson decays to two charged leptons and the W bosons both decay hadronically.	43
5.4	The Feynman diagram for the Drell-Yan background process.	44
5.5	The Feynman diagram for the $t\bar{t}$ background process.	45
5.6	The ROC curve for three different lepton MVAs. This analysis made use of the Ghent lepton MVA.	49
5.7	The amount of jets and the amount of b-tagged jets, achieved by removing the amount of jets and amount of b-tagged jets cuts respectively. In both cases, all remaining event selection criteria listed in table 5.6 are applied.	52
5.8	Reconstructed invariant dilepton mass and leading lepton transverse momentum with signal region selections applied.	53
5.9	Reconstructed invariant dilepton mass and leading lepton transverse momentum with DY control region selections applied.	54
5.10	Reconstructed invariant dilepton mass and amount of jets with $t\bar{t}$ control region selections applied.	55
5.11	Plots of the discriminant variables used in this analysis: the invariant dilepton mass, missing transverse energy, amount of b-tagged jets and reconstructed W boson mass.	59
5.12	A comparison between a severely overfitted BDT and a BDT with a lot less overfitting. For both BDTs, their “overfitting parameter” is given, for which lower means better.	63
5.13	The output shape of the boosted decision tree.	63

5.14	The variables that were used as an input to the BDT and NN, ranked by importance. This feature importance plot was obtained from training the BDT.	64
5.15	The output shape of the neural network.	65
5.16	The impact plot for the BDT.	72
5.17	The impact plot for the NN.	73
5.18	Pre- and post-fit plots of the BDT output showing the observed (points) and predicted (histograms) yields.	74
5.19	Pre- and post-fit plots of the NN output showing the observed (points) and predicted (histograms) yields.	74
E.1	A schematic overview of the structure of the NN.	96
F.1	The effect on the BDT output distribution of the different shape uncertainties varied one standard deviation up and down.	97
F.1	The effect on the BDT output distribution of the different shape uncertainties varied one standard deviation up and down.	98
F.2	The effect on the NN output distribution of the different shape uncertainties varied one standard deviation up and down.	98
F.2	The effect on the NN output distribution of the different shape uncertainties varied one standard deviation up and down.	99
G.1	De deeltjes van het standaardmodel.	102
G.2	Links: Foto van de CMS detector tijdens onderhoud. Rechts: Een typisch event waargenomen door de CMS detector. In dit event vervalt een Higgs boson naar twee τ leptonen.	104

List of Tables

5.1	Decay modes of the Z boson with their respective end products and BRs.	40
5.2	Decay modes of the top quark with their respective end products and BRs. The BRs do not add up to exactly 100%, this is due to the relatively large errors on the measurements of the BRs.	40
5.3	Event generators used to simulate events for the various processes.	46
5.4	Lepton selection criteria	47
5.5	Jet selection criteria	50
5.6	Event selection criteria	51
5.7	The cutflow table. The cutflow of each group of backgrounds is represented, along with the fractions of events which remain after applying the event selection criteria, where the reference point is the amount of events left after the basic object selections. In the rightmost column, the signal-to-background ratio is visible. . .	57
5.8	All input variables given to the BDT and NN. The lepton variables are given for both the leading and sub-leading lepton, while jet variables are provided for the five jets with highest p_T . An exception is made for $\Delta R_{l1/2}$ and $\Delta R(b)_{l1/2}$, as only the $\Delta R_{j/b}$ of the (b)jet which minimizes $\Delta R_{j/b}$ is used, both for the leading and sub-leading lepton.	62
5.9	All included rate systematic uncertainties.	68
5.10	The signal strength measurements, together with the cross section calculated using this signal strength. For both quantities, 68% CL errors are provided. . .	71
A.1	List of all used $t\bar{t}Z$ datasets along with their respective cross sections.	79
A.2	All “other” used datasets along with their respective cross sections.	80
A.3	List of all used $t\bar{t}X$ datasets along with their respective cross sections.	81
A.4	List of all used $t\bar{t}W$ datasets along with their respective cross sections.	82

A.5	List of all used $t\bar{t}$ datasets along with their respective cross sections.	82
A.6	List of all used Drell-Yan datasets along with their respective cross sections. . . .	82
E.1	The set of hyperparameters used to train the BDT.	94
E.2	The set of hyperparameters used to train the NN.	95

List of Abbreviations

AI Artificial Intelligence

AUC Area under curve

BDT Boosted Decision Tree

BEH Brout-Englert-Higgs

BSM Beyond standard model

CERN European Organization for Nuclear Research

CL Confidence level

CMS Compact Muon Solenoid

COM Center of mass

CR Control region

CSC Cathode strip chambers

DAQ Data acquisition

DNN Deep Neural Network

DOF Degrees of freedom

DT Drift tubes

DY Drell-Yan

ECAL Electromagnetic Calorimeter

FSR Final state radiation

GS Grid search

HCAL Hadronic Calorimeter

HLT High level trigger

IP Interaction point

ISR Initial state radiation

JEC Jet energy scale corrections

L1 Level 1

LEP Large Electron-Positron Collider

LHC Large Hadron Collider

MC Monte Carlo

MET Missing transverse energy

ML Machine Learning

MSE Mean squared error

MVA Multivariate Analysis

NN Neural Network

NP Nuisance parameter

OF Opposite flavour

OS Opposite sign

OSSE Opposite sign same flavour

PF Particle flow

POG Physics object group

PU Pileup

PV Primary vertex

QCD Quantum chromodynamics

QED Quantum electrodynamics

QFT Quantum field theory

RF Radio frequency

ROC Receiver operating characteristic

RPC Resistive plate chambers

RSS Residual sum of squares

SBR Signal-to-background ratio

SF Same flavour

SM Standard Model

SR Signal region

SS Same sign

Chapter 1

Introduction

Throughout the course of history, humanity has always been profoundly interested in the fundamental structure of matter. The ancient Greeks thought all matter was composed of discrete, indivisible units. However, it is important to note that these ideas were based on metaphysical reasoning rather than empirical evidence. In the 19th century, John Dalton named these particles *atoms*. The origin of this word lies in the Greek word *ατομος* (atomos), which means unsplitable: Dalton thought this was the end of the story. The first real breakthrough was the discovery of the electron by Thomson in 1897, as its size was about 1800 times smaller than the smallest atom known at that time. Ever since then, the combined effort of thousands of theoretical and experimental physicists has resulted in a better understanding of the fundamental structure of matter. All matter in the universe is made off a few basic building blocks called fundamental particles, governed by four fundamental forces [1]. Currently, the most accurate mathematical theory describing all these particles and three of the four fundamental forces is the standard model of particle physics. The standard model is an incredibly successful theory, but it does have some imperfections. We will discuss the standard model in chapter 2.

New theories trying to fill the gaps in the standard model are constantly being proposed, with an example of such a theory being supersymmetry. Most of these theories are expected to have a strong coupling to the heaviest particle in the standard model, the top quark. For this reason, among many others, the top quark is of particular interest to physicists all around the globe. In an attempt to improve our understanding of the top quark, the analysis presented in chapter 5 of this thesis aims to study the coupling of the top quark to the Z boson. In doing so, we will make use of various machine learning techniques, which will be addressed in chapter 4. The data used for this analysis was taken at the Compact Muon Solenoid (CMS) detector at the

European Organization for Nuclear Research. The CMS detector will be discussed in chapter 3.

1.1 Notations and conventions

Throughout this thesis, natural units will be used. This means that two universal constants, the speed of light c and the reduced Planck constant $\hbar$, will be set to 1:

$$\hbar = c = 1 \tag{1.1}$$

Using natural units enables us to simplify quite a lot of equations, as well as allow us to express energies, momenta and masses in the same unit, namely electron volts (eV).

In the last chapters of this thesis, many histograms will be shown. Every plotted histogram will include overflow binning, unless otherwise stated. This means that each event which falls outside of the plotted range will be placed in the last bin.

Chapter 2

The Standard Model

The standard model (SM) of particle physics describes all known elementary particles and three of the four fundamental forces in a mathematical way. The gravitational force is the only fundamental interaction which is not included in the SM. In the first part of this chapter, the elementary particles and their interactions will be discussed. In the next part, some of the shortcomings of the SM will be covered. The last part of this chapter will discuss the top quark, which is a central part of the analysis presented in this thesis.

2.1 Particles and interactions

Firstly, we will go over the particle constituents of the SM. There are two different types of particles: fermions and bosons. Fermions are the building blocks of all matter we see around us, while bosons are responsible for mediating the fundamental interactions and generating the mass of the SM particles through the Brout-Englert-Higgs (BEH) mechanism. Bosons and fermions can be distinguished from each other by their spin. The particle content of the SM is depicted in figure 2.1.

2.1.1 Fermions

Fermions are particles which have half-integer spin and follow Fermi-Dirac statistics. They also follow the Pauli exclusion principle, meaning that only one fermion can occupy a particular quantum state at any given time [3]. The fermions can be further divided into two types of particles: leptons and quarks. These two types can in turn be divided into three generations of leptons or quarks. The three charged leptons are the electron (e), muon (μ) and tau (τ). They

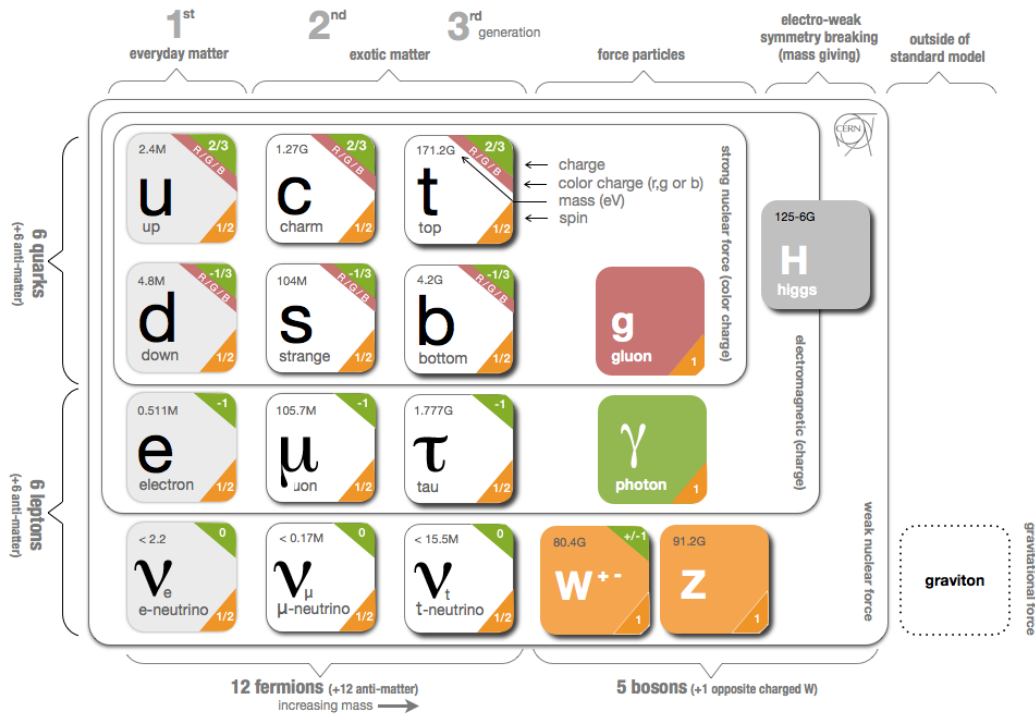


Figure 2.1: The standard model particle content. Figure taken from Ref. [2].

all have neutrino counterparts, which are very light neutral particles: the electron-neutrino (ν_e), muon-neutrino (ν_μ) and tau-neutrino (ν_τ). Leptons interact through the weak, electromagnetic and gravitational forces. However, as neutrinos are electrically neutral and basically massless¹, they only interact through the weak force. The three quark generations are made up of the up (u) and down (d) quark, charm (c) and strange (s) quark and lastly the top (t) and bottom (b) quark. Apart from their masses, these families have exactly the same properties. Quarks interact through the weak, electromagnetic, gravitational and strong forces. While leptons carry integer charge, a positively charged quark has charge $\frac{2}{3}e$ and a negatively charged quark has charge $-\frac{1}{3}e$. All of the previously mentioned fermions also exist in their antiparticle form. These antiparticles are completely identical to their particle counterparts, except for their inverted charge. Although there are 12 different fermions (24 when their antiparticle forms are included), virtually all matter in the world as we know it is composed out of only the first generation of quarks and leptons. It is only when studying particle physics at higher energies that we start encountering the particles of the second and third generation, for instance in particle colliders or cosmic ray physics.

¹Neutrinos are not actually massless, but their mass is so small that it can be neglected for gravitational effects. However, it is important that they are not massless but we will come back to this later.

2.1.2 Gauge bosons and fundamental forces

Bosons are particles which follow Bose-Einstein statistics and therefore have integer spin. Unlike leptons, an unlimited amount of bosons can occupy the same quantum state. Currently, four different groups of bosons are known to exist: the Higgs boson, the W and Z bosons, gluons (eight in total) and the photon [4]. All of these have spin 1, except for the Higgs boson which has a spin of 0. These bosons are all mediators of one of the fundamental forces, with the Higgs boson once again being an exception. There is also a fifth, hypothetical boson: the graviton, which is the candidate for being the mediator boson of the gravitational force. Thus far, no evidence of its existence has been found.

2.1.2.1 Electromagnetic interaction

The electromagnetic interaction is the fundamental force that acts between electrically charged particles. It is mediated by the photon, which in itself has no electrical charge. As a result of the photon being electrically neutral, the range of the electromagnetic interaction is infinite. However, its strength does decrease with growing distance. Quantum electrodynamics (QED) is the quantum field theory (QFT) which describes the electromagnetic interaction. QED is an abelian gauge theory with the symmetry group $U(1)$ [5].

2.1.2.2 Weak interaction

All particles in the SM interact through the weak force. Contrary to the electromagnetic force, the weak force only acts on subatomic distance scales. It is also weaker than the electromagnetic force at low energy reactions, although they are of about equal strength for high energy processes. The weak force is mediated through the Z , W^+ and W^- bosons, which are all massive. These masses are the reason for the low strength of the weak interaction at low energy scales. With W and Z boson masses of 80.379 GeV and 91.188 GeV respectively² [6], it takes a lot of energy to produce these bosons. The weak interaction is quite unique, in that it is the only interaction which can change quark flavors. Another exclusive property to the weak interaction is that it is the only interaction which violates parity and charge-parity (CP) conservation [7, 8].

²For comparison: the W and Z boson masses are five orders of magnitude greater than the electron mass and three orders of magnitude greater than the up and down quark masses.

2.1.2.3 Electroweak unification

After the huge success of QED, attempts were made to find a similar self-consistent gauge theory for the weak force [9]. During the 1960s, Sheldon Lee Glashow, Abdus Salam, and Steven Weinberg independently discovered such a gauge-invariant theory for the weak and electromagnetic force [10, 11, 12]. They found out that the theory could be described using the $SU(2)_L \otimes U(1)_{Y_W}$ symmetry group. This required the existence of four massless carrier bosons: two electrically charged bosons, W^+ and W^- , and two electrically neutral bosons, Z and the photon. However, as the weak interaction is such a short range force, this interaction must be mediated by massive carrier bosons [13]. This mass problem was solved using spontaneous symmetry breaking, giving mass to the carrier bosons of the weak interaction while keeping the photon massless. The mechanism that is accountable for the mass of the carrier bosons is the Brout-Englert-Higgs (BEH) mechanism, which will briefly be described in section 2.1.3.

2.1.2.4 Strong interaction

The strong interaction is most commonly known as the force that holds nuclei together. Only quarks interact through the strong interaction. Much like the weak interaction, the strong interaction only acts on subatomic (femtometer) scales. It is however very strong, as its name would suggest: it is by far the strongest out of all the fundamental interactions. The strong interaction is mediated by gluons and is described by a non-Abelian $SU(3)$ theory called quantum chromodynamics (QCD). The name QCD comes from the fact that there are three linearly independent strong charges, which draws an analogy to the three fundamental colors. As a result of QCD being a non-Abelian theory, the strong force carriers also carry a colour charge, which allows them to interact with each other. Quarks also have a colour charge which can be red, green or blue (antiquarks carry the anticolours). However, particles are required to be colour neutral. This means that quarks can never be free, in the sense that they can never be isolated from other quarks. They will clump together to form composite particles like mesons (a quark and antiquark of the “same” colour³) or baryons (three quarks with each a different colour). This phenomenon is called colour confinement. When two quarks are getting further and further separated from each other, the energy required to separate them even more will grow enormously to the point where it is energetically more beneficial to create a new quark-antiquark pair which will accompany the “isolated” quark. This makes it effectively impossible to isolate

³For instance, if the quark has red colour charge, the antiquark must have antired colour charge.

a quark, or any particle with non-neutral color charge.

2.1.3 The Higgs boson

In solving the mass problem for the weak interaction by using spontaneous symmetry breaking, a complex scalar field is introduced [13]. This field is called the Brout-Englert-Higgs field, or Higgs-field for short. The main idea behind this is that instead of having a single ground state at zero, the scalar potential has a continuous ground state, resulting in an infinite amount of possible groundstates. After applying spontaneous symmetry breaking, the field will have settled in one of these ground states, called the vacuum expectation value [13]. Initially, the field has four degrees of freedom (DOF). However, after spontaneous symmetry breaking, three of these DOF result in unphysical, massless Goldstone bosons. These Goldstone bosons can be gauged away, resulting in a mass term in the Lagrangian for the gauge bosons of the weak interaction. The still remaining DOF is the one corresponding to the Higgs boson. This mechanism of giving mass to gauge bosons, resulting in the existence of the Higgs boson, is called the previously mentioned Brout-Englert-Higgs mechanism. Despite being theoretically predicted in 1964, the Higgs boson was first observed only in 2012 at the two largest experiments located at CERN's LHC, CMS and ATLAS. They reported a mass of 125.3 GeV and 126.0 GeV respectively [4, 14]. At the time of writing, the Higgs mass is estimated to be 125.10 ± 0.14 GeV [6]. The Higgs boson is an electrically neutral particle and the only currently known boson that has a spin of zero.

2.2 Limitations of the standard model

Despite being an incredibly successful theory, there are still a number of phenomena which can not be explained by the SM. I provide a non-exhaustive list below.

2.2.1 Gravity

The gravitational force is by far the weakest of all fundamental forces. As its effects in high energy particle physics are so small, it can safely be neglected in experiments at the LHC. So far, no microscopic theory of gravity which can be unified with the SM has been found. A lot of different attempts to construct such a theory have been made, for instance supergravity [15], but so far none of them have been successful.

2.2.2 Dark matter and dark energy

Numerous cosmological and astrophysical studies point to the existence of dark matter. The name dark matter arises from the fact that this type of matter does not seem to interact through the electromagnetic interaction. It does however interact gravitationally, making it detectable through a number of techniques including but not limited to studying the rotation curves of galaxies or using a phenomenon called gravitational lensing. The latest Planck results state that ordinary matter only accounts for 4.9% of the universes mass-energy density, while dark matter makes up about 26.5% of the mass-energy density [16]. The rest of the universe is made up of dark energy, which drives the expansion of the universe [17]. It is called dark energy only due to the fact that we are almost completely ignorant about its nature. The Planck results show that dark energy is consistent with the assumption of some cosmological constant like a constant vacuum energy density [16]. Dark energy is not compatible with the SM, on the contrary: trying to compute this cosmological constant leads to a mismatch of more than 100 orders of magnitude [18].

2.2.3 The hierarchy problem

While the discovery of the Higgs boson was a giant success for the SM, it brought some problems along with it. The observed mass of the Higgs boson of 125 GeV does not directly correspond to the bare mass appearing in its mass term in the Lagrangian. This bare mass has to be corrected by radiative corrections, also called loop corrections. These radiative corrections have to be calculated up to the Planck scale, at which the SM theory is expected to break down. Taking all possible loops up to the Planck scale into account yields a correction of the order of 10^{19} GeV. This implies that the bare mass of the Higgs boson has to be corrected by a factor of this magnitude such that the total mass including correction can be compatible with the observed mass of 125 GeV. A fine-tuning of the theory of this order is deemed unnatural and this is known as the hierarchy problem.

2.2.4 Matter-antimatter asymmetry

The Big Bang should have created equal amounts of matter and antimatter [19]. However in current times, everything we observe, from micro-scale objects to the largest stars, seems to be made up of regular matter. This matter-antimatter asymmetry leaves us with the up until today unanswered question of what happened with all this antimatter created in the Big Bang.

As shown by Sakharov [20], there are three requirements which have to be met to generate this asymmetry: Charge (C) and Charge-Parity (CP) violation, the absence of a thermal equilibrium and at least one baryon number violating process. Within the SM, there is some CP violation apparent in the weak interaction. But even if the other two criteria were met, this amount of CP violation would still not be able to account for the matter-antimatter asymmetry we observe today.

2.2.5 Neutrino masses

Neutrino's of different flavours have been observed to oscillate into one-another [21]. In order to be able to oscillate, they have to possess a nonzero mass. The actual masses are not known yet, only the squared differences between the masses are known. However, an upper limit for the masses of eV scale has been found, lying many orders of magnitude under the masses of other fermions [22]. In the SM, neutrinos are massless, which is not compatible with the observed neutrino oscillations.

2.3 The top quark

Ever since the top quark was discovered by the CDF and DØ collaborations in 1995 at the Tevatron at FermiLab [23, 24], studies of the top quark have formed a distinct field within high energy particle physics, attracting the attention of hundreds of scientists all around the world. There is a plethora of reasons as to why studying the top quark is such an interesting research domain. The fact that the top quark is the heaviest fundamental particle currently known, with a mass of 173.0 ± 0.4 GeV [6], might be the most obvious reason why studying the top quark is compelling. With a decay width of $1.42^{+0.19}_{-0.15}$ GeV [6], the top quark has a lifetime of about $5 \cdot 10^{-25}$ s, which is much shorter than the hadronization time at 10^{-23} s. In 99.8% of top quark decays, it decays to a W boson and a b quark, with the latter hadronizing and forming a jet. Due to this extraordinary short lifetime, the top quark can be considered a free particle and it is possible to be studied on its own. This is a major difference with all other quarks, which will almost instantly hadronize and thus always have to be studied in hadronic bound states.

The mass of the top quark is of great importance in global fits of all SM parameters together. For instance, an important consistency test of the SM is the simultaneous indirect determination of the W boson and top quark masses. The results of such a fit are displayed in figure 2.2 [25]. Thus far, it seems that such indirect fits are indeed compatible with direct measurements of

these masses, confirming the consistency between experiment and the SM as well as its predicted relations between the parameters.

As previously mentioned, the top quark is the heaviest fundamental particle in the SM. The coupling of a massive fermion to the Higgs field is proportional to its mass, making the top quark the fundamental particle with by far the largest coupling to the Higgs boson. This raises the question whether the top quark plays a special role in electroweak symmetry breaking [26]. The top quark production also plays a crucial role in many scenarios for new, beyond standard model (BSM) physics. A number of theories predict either the existence of new particles decaying to the top quark or with large coupling to the top quark [26]. Processes like this would manifest themselves as a resonance in $t\bar{t}$ production. Accordingly, studying top quarks may hint towards the presence of new BSM physics, or allow us to calculate exclusion limits for the production rate of these particles predicted by several BSM theories.

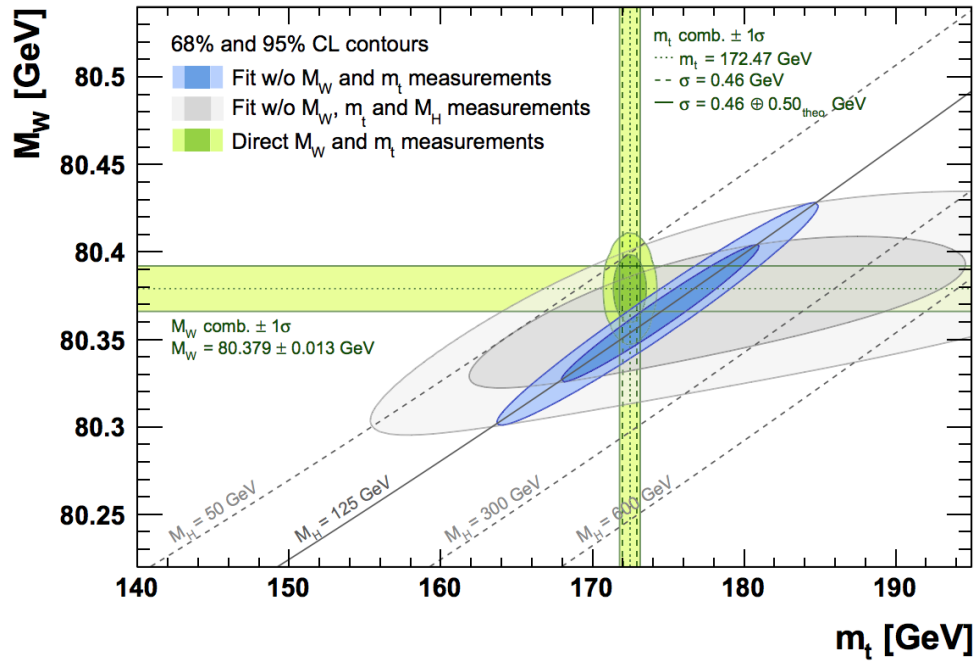


Figure 2.2: Results of a fit of the SM parameters as a function of the W boson and top quark masses. The contours at 68% and 95% CL for the fit including (blue) and excluding (grey) the M_H measurement can be compared to direct measurements (green) of the M_t and M_W . Figure taken from Ref. [25].

Chapter 3

The CMS Detector at the LHC

The analysis presented in this thesis was performed using data collected at the Compact Muon Solenoid (CMS) detector. In this chapter, we will discuss the Large Hadron Collider (LHC), the CMS detector and how we can reconstruct physical events from detector data.

3.1 The LHC

The LHC is the world's largest and most powerful particle accelerator and collider. It is located at the European Organization for Nuclear Research (CERN), near Geneva. The LHC lies in a 27 kilometres long tunnel, which previously housed the Large Electron-Positron Collider (LEP). Inside the LHC, two high energy hadron beams travel at near the speed of light, colliding with each other in the four points where the main experiments are located. Most of the time, the hadrons used are protons, but on occasion heavy ion beams are used as well. The proton beams are inserted into the ring of the LHC after being pre-accelerated by a number of smaller accelerators. A schematic overview of CERN's accelerator complex can be seen in figure 3.1. The LHC is using 1232 dipole magnets and 392 quadrupole magnets, along with 688 sextupole magnets and many additional magnets [27]. The dipole magnets are used for bending the beam, while the quadrupole and sextupole (and the other magnets) have the task of focusing the beam. All of these magnets are cooled down to -271.3 degrees Celsius, at which point these magnets become superconducting, which allows the generation of immense magnetic fields. Liquid helium is used to cool the magnets to this extraordinary low temperate, even colder than outer space [27]! Additionally to the magnets, the LHC makes use of 16 radio frequency (RF) cavities, which are used to accelerate the beams. As the beam of protons is forced to change its acceleration

as to stay on a circular track, a so-called *Bremsstrahlung* is emitted, also referred to as braking radiation. This causes the particles to lose some energy and decelerate, an effect which has to be compensated and necessitates the usage of large radius colliders, as the effects increase with the decreasing turn radius. This is done using the previously mentioned RF cavities, which will re-accelerate the beam, maintaining its speed to be near the speed of light. One of the reasons for replacing the LEP with the LHC was the previously mentioned Bremsstrahlung: the energy loss due to Bremsstrahlung increases proportionally to m^{-4} , resulting in a very high energy loss for light particles like electrons. As protons are about 2000 times heavier than electrons (or positrons), using protons reduces the energy loss by about twelve orders of magnitude, making maintaining their energy easier. In its current state, the LHC is able to run at a center of mass (COM) energy of 13 TeV. It is expected that by Run 3, scheduled to commence in 2021, the LHC will run at a COM energy of 14 TeV [28].

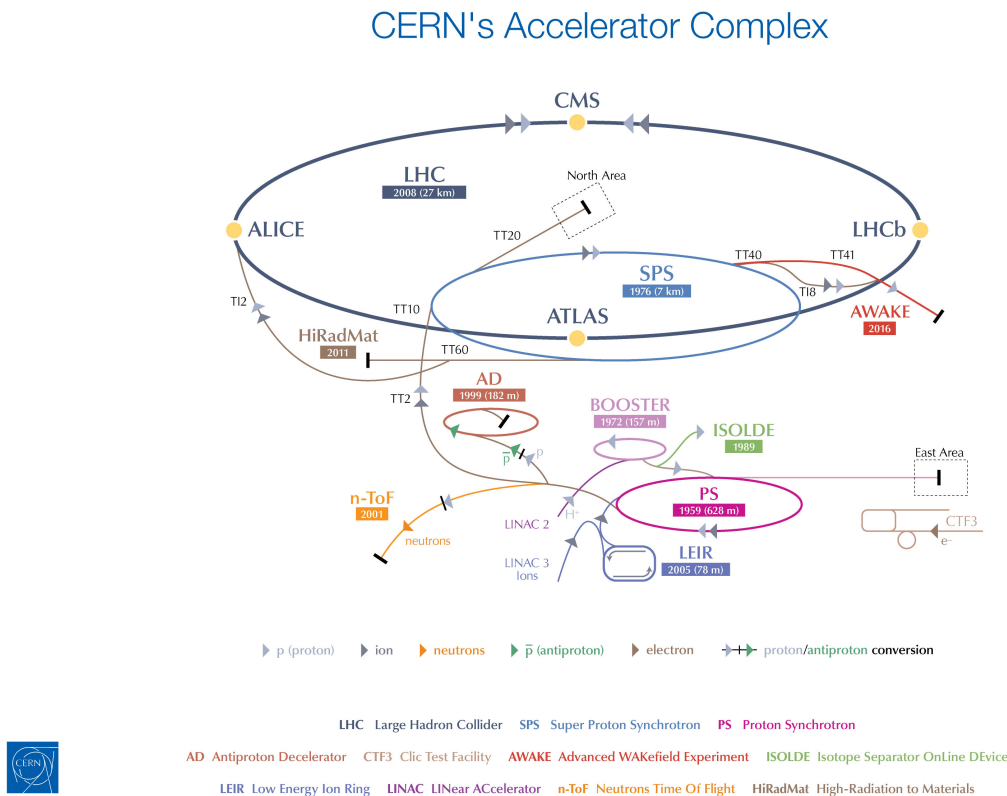


Figure 3.1: Schematic overview of the CERN accelerator complex. All pre-accelerators are shown, as well as the four main LHC experiments: CMS, ATLAS, ALICE and LHCb. This figure was taken from Ref. [29].

3.2 The CMS detector

CMS is a general purpose detector, used for a range of particle physics studies which include Higgs physics, new physics searches and many more. Its standout feature is the solenoid magnet giving rise to a magnetic field of about 4 Tesla, enabling the detector to make very precise transverse momentum (p_T) measurements based on the track curvature of charged particles [30]. CMS was designed to accurately detect muons, which together with the relative compactness of the detector and the solenoid gives rise to the name Compact Muon Solenoid. A 3D-overview of the CMS apparatus is given in figure 3.2, while a cross-section of the detector can be found in figure 3.4. The CMS detector is a cylindrical detector, built to be as hermetic as possible. In the following sections, we will discuss the coordinate system, the tracker, the calorimeters and the muon system.

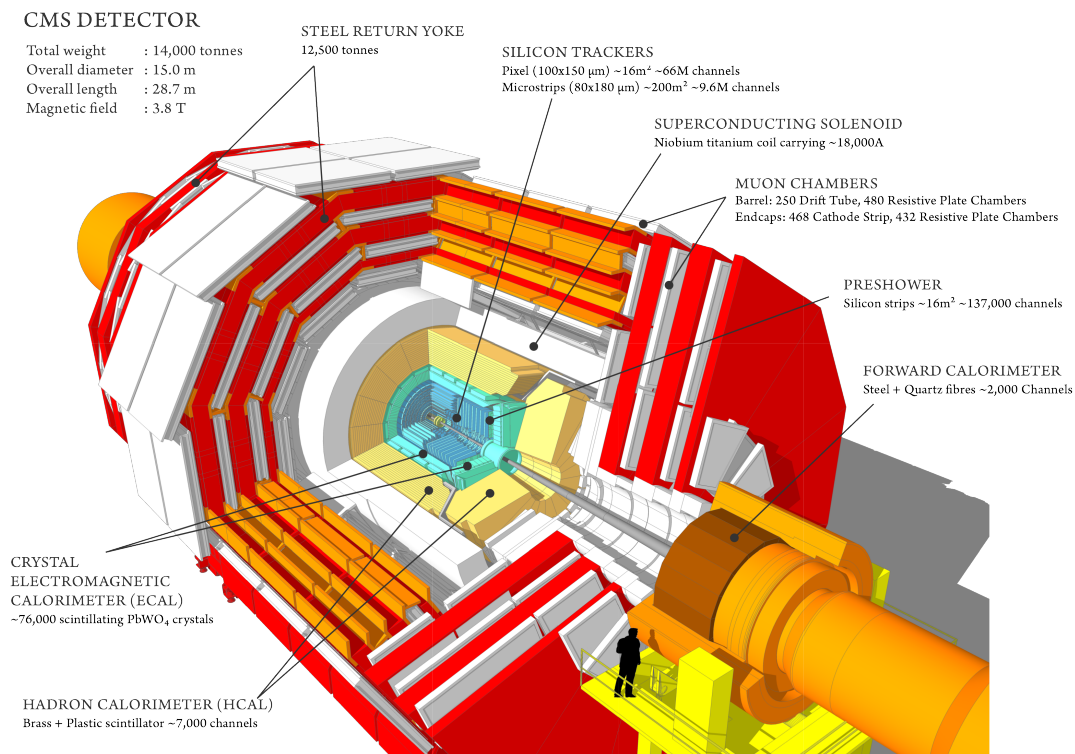


Figure 3.2: Overview of the CMS detector depicting its different parts. A person is displayed to serve as a reference scale. Figure taken from Ref. [31].

3.2.1 Coordinate system and basic variables

The origin of the CMS coordinate system lies in the collision point, with the y -axis pointing vertically upward. The x -axis points radially inward towards the center of the LHC, while finally

the z -axis points along the beam direction towards the Jura mountains [32]. The azimuthal angle ϕ is measured from the x -axis in the xy -plane, while the polar angle θ is measured from the z -axis. Using this right-handed coordinate system, depicted in figure 3.3, the pseudorapidity η is defined as follows:

$$\eta = -\ln \left[\tan \frac{\theta}{2} \right] \quad (3.1)$$

The pseudorapidity is often used instead of the polar angle, as differences in pseudorapidity are Lorentz-invariant, meaning that the value will not change when boosted along the beam line. To quantify the angular separation of two tracks in the detector, we define ΔR as follows:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad (3.2)$$

As both η and ϕ are Lorentz invariant under a boost along the z axis, ΔR is a Lorentz invariant quantity as well. The momentum and energy of a certain object are measured transverse to the beam direction because this is the only component of the momentum which can be measured as the magnetic field lies along the z axis. The transverse momentum and energy are denoted by p_T and E_T respectively.

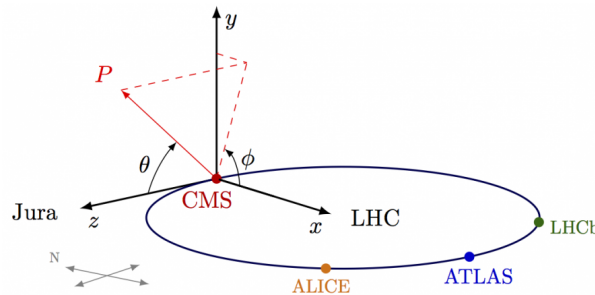


Figure 3.3: The CMS coordinate system. Figure taken from Ref. [33].

3.2.2 Tracker

The tracker is the part of the detector located the closest to the interaction point (IP), as can be seen in figure 3.4. Its purpose is to precisely measure the momenta of charged particles emerging from the LHC collisions, as well as reconstructing primary and secondary vertices [30]. By measuring the positions of charged particles passing through the detector, tracks can be reconstructed. The tracks of particles will be bent due to the homogeneous magnetic field created by the CMS solenoid surrounding the entire tracker. Tracks from particles with a higher

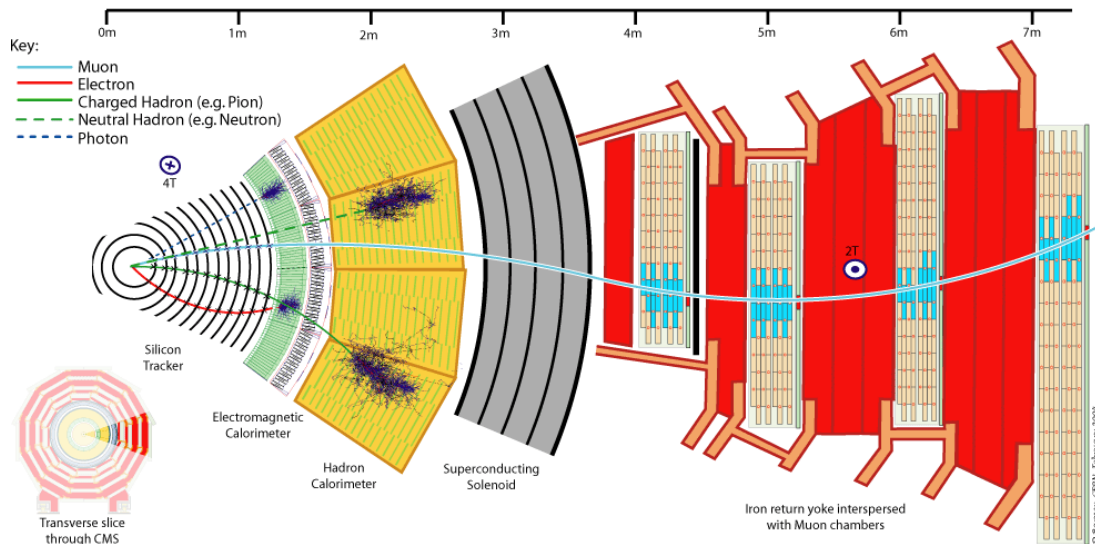


Figure 3.4: A cross-section of the CMS detector. The parts of the detector in which particles typically deposit their energy are also depicted. Figure taken from Ref. [34].

momentum will be bent less, allowing the tracker to measure the momentum of these particles. As this method to measure the momentum requires the particles to be electrically charged, neutral particles will pass through this part of the detector unnoticed. At the LHC design luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$, a lot of particles have to be detected in a short amount of time. This asks for a detector with high granularity (high spatial resolution) and very fast response times. The tracker also has to be capable to handle huge amount of radiation. All of these requirements can be satisfied by using a silicon-based detector. The tracker is made up of three sub-components. In the innermost component close to the IP, silicon pixels are being used as they ensure high granularity and can withstand the high flux of particles [32]. The intermediate component is made up of silicon microstrip detectors. The outer part of the tracker also consists of silicon microstrip detectors, but with an increased strip size due to the lower particle flux. A sketch of a cross section of the tracker detector is presented in figure 3.5.

3.2.3 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is built around the tracker detector. Its goal is to measure the energies of electrons (positrons) and photons with high accuracy, yielding a separate measurement of the momentum which can be compared with the momentum measurement from the tracker. The ECAL is the first destructive part of the CMS detector: particles which are detected in the ECAL will be destroyed and thus they will not travel to the further parts of the

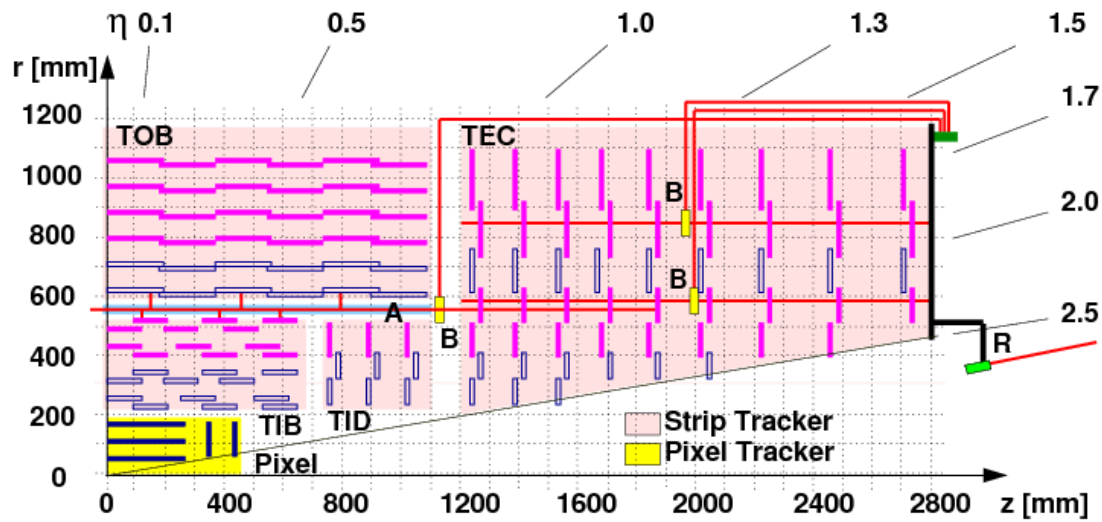


Figure 3.5: One fourth of the cross section of the tracker along the z -axis. Solid blue lines represent pixel detectors. Solid pink lines represent single-sided silicon strip modules, while void blue lines represent double-sided strip modules. The tracker's coverage of pseudorapidity values up to 2.5 is also visualized. Figure taken from Ref. [35].

detector. In the CMS apparatus, the ECAL is made up of lead tungstate (PbWO_4) crystals. These crystals have a high density and short radiation length, making them extremely fit for use in the CMS ECAL as this allows for a compact size of the detector compared to designs based on other available materials [30]. The crystals will get ionized when an electron, positron or photon passes through the ECAL, after which they will emit a scintillation light. The amount of light emitted will be proportional to the energy absorbed from the incident particle. This light is then detected by photodiodes which convert the light to an electronic signal. The ECAL consists of three main parts: the barrel, preshower and endcaps. The barrel section covers the range where $|\eta| < 1.479$ and the endcaps covers the $1.479 < |\eta| < 3$ range. The preshower covers the $1.653 < |\eta| < 2.6$ range [30]. Neutral pions are very likely to decay into two photons, and it is the job of the preshower to distinguish these photons from “actual” photons. The two photons originating from the pion decay will have an angle between them proportional to their energy. When they are directed at the endcaps, this angle will be too small for the less granular endcaps to be able to determine whether the hit originated from one photon or two closely-spaced photons. By installing the high-granularity preshower in front of the endcaps, a distinction between a single photon or two closely-spaced photons can be made. An overview of the ECAL layout is given in figure 3.6.

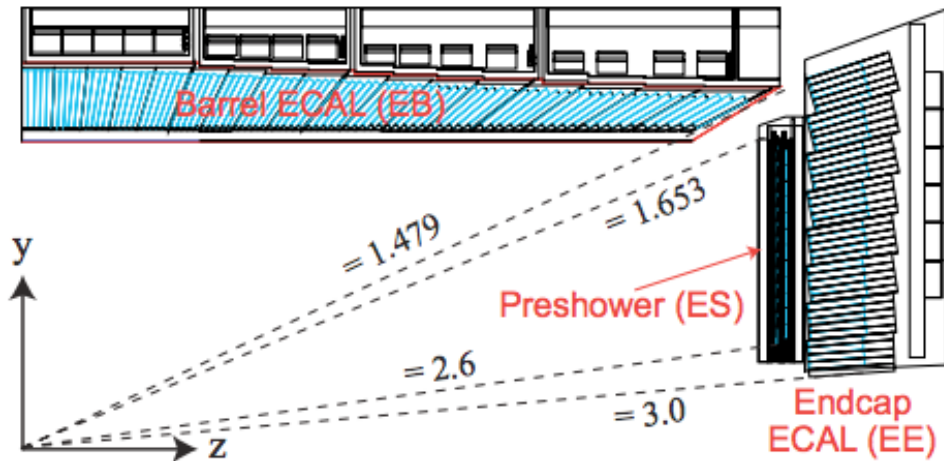


Figure 3.6: Transverse section of the ECAL showing the barrel, preshower and endcaps. Figure taken from Ref. [32].

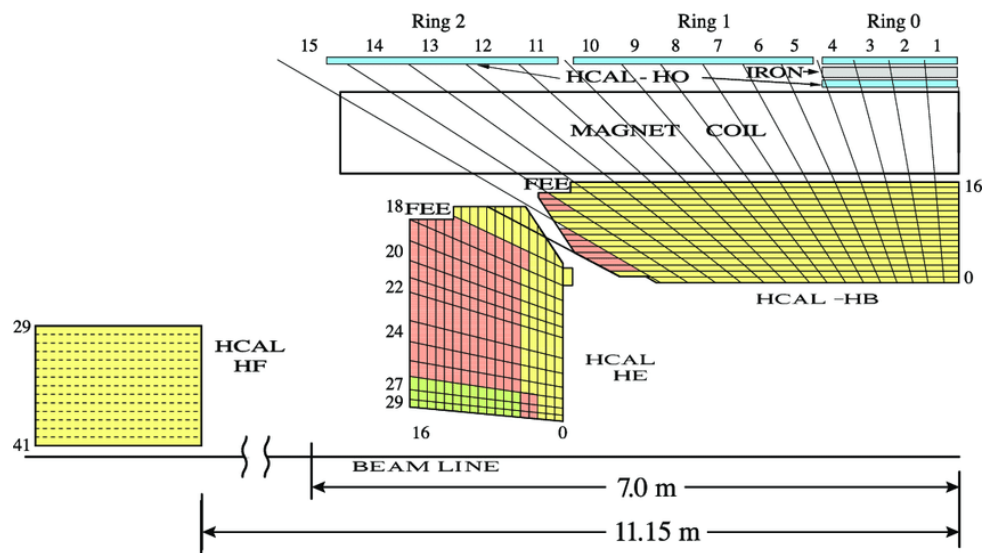


Figure 3.7: The different parts of the HCAL. In this schematic overview, HB is the hadronic barrel, HE the hadronic endcaps, HF the hadronic forward calorimeters and HO the hadronic outer calorimeter. Figure taken from Ref. [36].

3.2.4 Hadronic Calorimeter

The part of the detector surrounding the ECAL is the hadronic calorimeter (HCAL). Just like the ECAL, and any calorimeter for that matter, this is a destructive detector. As can be seen in figure 3.4, the HCAL completely surrounds the ECAL and the main part of the HCAL still lies within the solenoid. The HCAL consists of a number of different parts: the barrel covers the region up to $|\eta| < 1.4$, while the endcaps cover the overlapping range of $1.3 < |\eta| < 3.0$

[32]. The depth of the HCAL system is chosen such that all hadrons decay within it, allowing for the measurement of their full energy. In order to be able to measure energetic forward jets, the hadronic forward calorimeters are located 11.2 metres from the IP, covering a range of $2.9 < |\eta| < 5$, overlapping slightly with the endcaps. Detecting jets up to an $|\eta|$ of 5 also improves the missing transverse energy (MET) measurement, as any object left undetected gives rise to MET by definition. Lastly, the hadron outer calorimeter contains an array of scintillators covering the $|\eta| < 1.26$ range. It is positioned outside of the solenoid, capturing highly energetic hadrons which might not be contained within the HCAL barrel part, thereby improving the central shower containment in this region [32]. The layout of the HCAL is shown in figure 3.7.

3.2.5 Muon system

As is apparent by the CMS experiment's middle name, detecting muons and measuring their momenta precisely is of central importance. Muons do not leave a signal in the HCAL as they do not interact strongly, and they are too penetrating to be absorbed in ECAL, meaning that they only leave a small fraction of their energy in it. For this reason, CMS has a dedicated detector built to identify muons, measure their momentum, and to trigger on them [30]. Triggering will be discussed in section 3.3. There are three different types of gaseous detectors used to measure the muons [32]. Their layout in the CMS detector is visualized in figure 3.8. Drift tubes (DT) are used in the barrel region, measuring muons up to $|\eta| = 1.2$. The DTs consist of tubes filled with a CO_2 -Ar gas mixture [37]. When a muon passes through the detector, it ionizes the gas. The resulting electrons then travel to a positively charged wire where they are detected. From this information, the muon coordinate in the $r - \phi$ bending plane can be deduced [30], yielding a total of two coordinates. The endcaps consist of cathode strip chambers (CSC), which cover the range where $0.9 < |\eta| < 2.4$. These consist of arrays of positively-charged anode wires crossed with negatively-charged copper cathode strips [38]. The gas mixture used to fill the CSCs also contains CO_2 and Ar like we saw in the DTs, but here 10% of the gas mixture is CF_4 [39]. As muons are passing through the CSCs, they create free electrons and positive ions, which travel to the anode and cathode strips respectively. The CSCs provide us with two muon coordinates again, as the cathode wires yield the ϕ -coordinate and the anode strips the η -coordinate of the muon [39]. In both the barrel region and the endcaps, resistive plate chambers (RPC) are used as well, although they only cover the pseudorapidity range up to $|\eta| = 1.6$. The RPCs consists of two parallel plates, one of which is a positively charged anode while the other is a negatively

charged cathode. Both plates are made of a plastic material with a very high resistivity [40]. As muons pass through the RPCs, they knock electrons out of the gas modules. These electrons create an avalanche of electrons which are detected by some external wires after a small, yet very precise time delay. The spatial resolution is not as good as for the DT and CSC detectors, but they give a very fast signal and have a good time resolution of only one nanosecond [40]. This makes them ideal to use for triggering, as the system can quickly decide if the event is worth keeping or not.

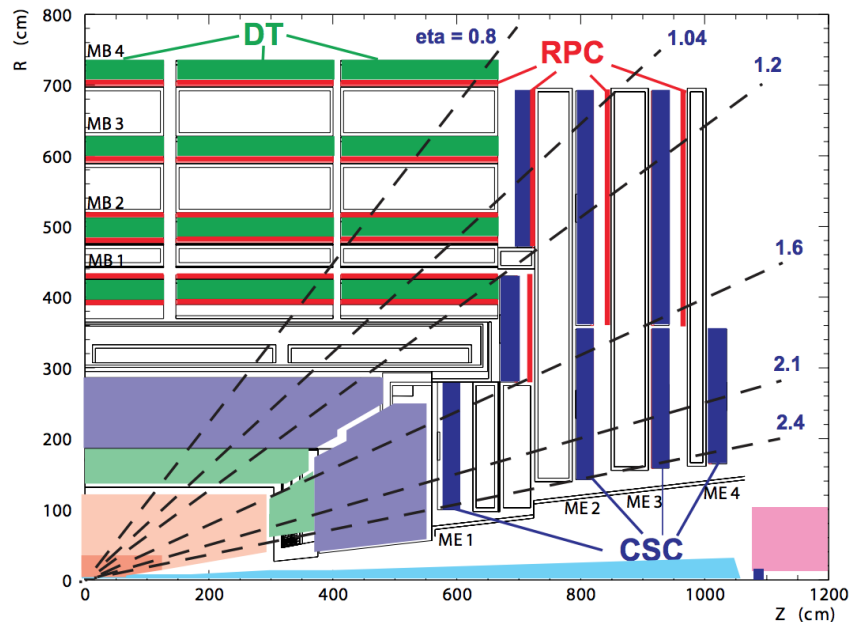


Figure 3.8: Layout of the muon chambers. Figure taken from Ref. [32].

3.3 Triggering and DAQ

When the LHC is performing at its peak, the two proton beams cross each other every 25 nanoseconds, yielding 40 million collisions per second [41]. As there are up to $1.18 \cdot 10^{11}$ particles per proton bunch, this results in about one billion proton-proton interactions taking place in the CMS detector every second [42]. It is impossible to read out all this data, and even if it were possible, most of the events would not contain interesting physics processes anyway. The high collision rate makes it necessary to employ a triggering system for data acquisition (DAQ), which will select potentially interesting events and discard all others. There are two different levels of triggers: the level 1 (L1) trigger and the high level trigger (HLT). The L1 is a very fast, hardware-based trigger. It is only allowed a time of $3.2 \mu\text{s}$ to make a decision on whether

or not the event should be kept for further processing [32]. Due to this time constraint, it is only possible to use data from the calorimeters and muon chambers to make the decision. Its decision is based only on objects such as electrons, muons, photons and jets which exceed some specific E_T and p_T cuts. Variables like MET and total E_T are used in the decision making as well. The L1 trigger reduces the event rate from 40 MHz to about 100 kHz. The HLT is a software-based trigger, which reduces the event rate from 100 kHz all the way to 100 Hz. As this system operates at much lower event rate than the L1 trigger, it has time to also use simplified particle and event reconstruction. In this way, the silicon tracker information can be used, which is by far the most precise detector present in the CMS apparatus. More than 1000 computers employing processors purposely manufactured for this task are used to apply the HLT.

3.4 Object and event reconstruction: the particle flow algorithm

Event reconstruction for the CMS experiment is done using the particle flow (PF) algorithm [43]. This algorithm reconstructs particles by using information from the inner tracker, the calorimeters and the muon chambers. Neutrinos can not be detected by the CMS detector, as they only interact weakly. However, in certain types of events, the MET can be used as a proxy to reconstruct neutrinos. Any undetected particles like neutrinos, but possible also dark matter, contribute to this missing energy. Yet this is not the only contribution: mismeasurements and mistakenly undetected particles are also a part of the MET, making it extremely important that the detector is as hermetic as possible. The tracker yields a very good momentum resolution for charged hadrons, which is accomplished by a precise measurement of the direction of charged particles [44]. It outperforms the calorimeters up to p_T values of hundreds of GeV. For these reasons, the tracker really is the foundation of the PF algorithm.

3.4.1 Iterative tracking

In order to reconstruct tracks with a small fake rate, PF uses iterative tracking. At first, very tight tracking constraints are used, yielding a mediocre efficiency but negligible fake rate. These tracks are then removed, and the track finding algorithm runs again, this time with slightly looser constraints, giving higher efficiency [45]. This procedure is repeated a number of times, yielding a higher track reconstruction efficiency, but the fake rate is kept low as the number

of tracks decreases every iteration. In the end, particles with a p_T as low as 150 MeV can be reconstructed, while keeping the fake rate below one percent.

3.4.2 Calorimeter clustering

As neutral particles will not be reconstructed using the previously mentioned technique, they have to be reconstructed in another way. This is done through calorimeter clustering. Not only does this reconstruct neutral particles, it also provides additional information on charged particles in case the tracker could not reconstruct their properties with high precision. Calorimeter clustering starts with a seed, which are calorimeter cells with an energy deposit higher than a set threshold and than their neighbouring cells [43]. Starting from the seeds, topological clusters are grown by aggregating adjacent cells passing an energy deposition threshold related to the level of noise. From the topological clusters, PF clusters are formed. Each seed gives rise to one PF cluster. If multiple seeds are a part of the same topological cluster, the energy of each of the calorimeter cells is shared among all clusters, relative to the cluster-cell distance.

3.4.3 Link algorithm

Particles travelling through the detector are likely to leave multiple hits in different subdetectors. Such a particle will be reconstructed as multiple PF objects, for instance a track in the tracker together with a hit in the ECAL. This phenomenon can clearly be seen in figure 3.4. For this reason, it is necessary to link multiple objects originating from the same particle together, and this is done by the link algorithm.

A link between the tracker and calorimeters is established by extrapolating the reconstructed track to the preshower, then to the ECAL at a depth corresponding to the expected maximum of a typical longitudinal electron shower and finally to the HCAL at a depth corresponding to one interaction length [43]. If this extrapolated position falls within a PF cluster, the track is linked to that cluster. The link distance is defined as the distance between the extrapolated track position and the cluster position in the (η, ϕ) plane [43]. If an extrapolated track is linked with multiple clusters or vice-versa, the link with the shortest link distance is kept.

In order to be able to attribute the energy of photons emitted from electrons through bremsstrahlung, tangents to the track of the electron at each layer in the tracker are extrapolated to the ECAL [43]. If the extrapolated track is within the boundaries of a cluster, the cluster is linked to the track as a potential bremsstrahlung photon.

Another possible linking is between calorimeter clusters. This can happen between the preshower and the ECAL as well as between the ECAL and HCAL. When the cluster position within the more granular calorimeter (the preshower or the ECAL) falls within the boundaries of a cluster in the less granular calorimeter (respectively the ECAL and HCAL), a link is established [43]. If a cluster were to link with multiple clusters in the less granular calorimeter, the link with the smallest link distance is kept.

Lastly, charged particle tracks in the tracker can be linked with a track in the muon system. The tracks are linked when a global fit between the two tracks yields an acceptable χ^2 value [43]. When multiple tracker tracks are linked with a muon track, the combination of tracks with the smallest χ^2 is chosen.

In the end, making all these links yields blocks of linked elements. These blocks can be converted into a set of identified particles, as will be explained in the next section.

3.4.4 Particle reconstruction and identification

When all blocks of linked elements are made, the final particle reconstruction can commence. The algorithm will execute the same procedure as described below on every single block. Firstly, each global muon gives rise to a particle flow muon if its combined momentum is in agreement with the momentum determined from the tracker measurement within three standard deviations [46]. The corresponding track will then be removed from the block. After the removal of the track, an estimate of the energy deposited in the calorimeters is also made. This energy is then also removed from the block.

The next particle to be reconstructed is the electron. Taking into account the possible loss of energy due to bremsstrahlung, the candidate electron track is refit using a Gaussian-Sum Filter in an attempt to follow its trajectory to the ECAL [46]. This gives rise to a fully identified electron when it passes tracking and calorimeter requirements and if its track matches with an ECAL cluster. If these requirements are met, a new PF electron can be added, while its track and ECAL clusters are removed from the block.

Now that electrons and muons are reconstructed, the remaining tracks have to satisfy tighter quality criteria: the relative uncertainty on the measured p_T has to be smaller than the relative calorimetric energy resolution for charged hadrons. Then, the remaining tracks are connected to the calorimeter clusters [46]. Neutral particles can be detected by comparing the momentum of the tracks to the calibrated cluster energy, as the excess of energy in the calorimeter compared

to the tracks originates from neutral particles like photons and neutral hadrons. If a track traverses more than one HCAL cluster, it is assigned to the closest one. Then, this charged hadron candidate track is matched with the ECAL clusters. It will be linked to the closest ECAL cluster it traverses, but it is possible that the track is linked to more than one ECAL cluster as well. In this case, the ECAL clusters are first sorted by distance, after which a loop over the ECAL clusters commences. These are added to the charged hadron candidate for as long as the total momenta of the deposits remains smaller than the track momentum [46].

If it were to happen that, after completing the loop over the calorimeter clusters, the total calibrated calorimetric energy is still smaller than the track momentum by at least three standard deviations, a relaxed search for muons and fake tracks is performed [46]. As a first step, all global muons (not already identified by the algorithm) for which the momentum is measured with a precision better than 25% are taken as PF muons.

All still remaining tracks become PF charged hadrons. Their momentum and energy are based directly on the track momentum under the charged-hadron hypothesis. In case that the track momentum is compatible with the energy deposited in the calorimeters, the charged hadron momentum and energy will be refit using the measurements in the calorimeters and tracker. However, when there is a substantial excess of calibrated calorimeter energy compared to the charged particle track momentum, photons and neutral hadrons are reconstructed [46]. Firstly, the ECAL clusters are used to reconstruct PF photons. If this still does not account for the whole calorimeter excess, PF neutral hadrons are reconstructed with the remainder of the excess.

Finally, all remaining ECAL and HCAL clusters, which have no tracks linked to them, are reconstructed as PF photons and PF neutral hadrons respectively.

Now that the end of the PF algorithm has been reached, we are left with a list of identified particles. These particles can be used to reconstruct jets using a jet clustering algorithm. We will not go deeper into this algorithm as it is out of the scope of this thesis, but more information can be found in Ref. [47].

Chapter 4

Machine Learning

This chapter will provide a brief introduction into machine learning (ML). Secondly, we will go more in-depth into the specific ML techniques that will be used in this thesis, namely boosted decision trees (BDT) and deep neural networks (DNN). Finally, we will also shortly discuss genetic algorithms.

4.1 Introduction to Machine Learning

Machine learning is a subset of artificial intelligence (AI) where computer algorithms improve automatically through experience [48]. The mathematical models built using such algorithms trained on data are then used to predict an output for a given data point, or in case of experimental particle physics, an event. This output can be some quantity we want to predict, for instance the life expectancy of a person. If this is the case, we call this **regression** [49]. However, it is also possible that we want to give a certain label to an event, based on the output of our model. For example, we might try to predict the gender of a person based on some typical properties. This is called **classification**.

A ML algorithm learns by looking at the features of a data point, or in the case of experimental particle physics, the variables of an event. It will then try to predict an output for this data point. Afterwards, it will calculate the **loss function**, which is a measure which quantifies the difference between the expected output and the predicted output. Based on the output of this loss function, the parameters of a model, often called weights, will be updated. The goal of the training process is to minimize the loss function. During model training, we will loop over all data points and calculate small weight updates for each data point. Most of the times we

will execute this loop over all data point multiple times, as the prediction of the model for a certain data point will probably change after it has calculated weight updates caused by all the other data points.

When training a ML algorithm, we will generally divide the dataset in to three parts: the training set, validation set and test set. As one could expect, the **training set** will be used to train the algorithm. The **validation set** is used to optimize the hyperparameters of the algorithm [50]. As our model improves, the outputs of the model are influenced more and more by the validation set, as this decides which hyperparameters we will use. As we iteratively change the hyperparameters based on the performance of the model on the validation set, we might end up with a model that does not generalize well and only performs good on the validation set. This is why we should always use a **test set** which will provide a fully unbiased evaluation of the final generalization performance of our model. This means that one should never use their test set to train an algorithm or optimise the hyperparameters, even when the full architecture of the model, including the choice of all hyperparameters, has been designed. After the optimal set of hyperparameters has been determined, one trains the algorithm on the training set and validation set together, and judges its performance on the test set.

Earlier in this section, we discussed how a ML model learns through experience. Given this, one might think that in order to get a great model, we could just keep training on the same data over and over again, “increasing the experience of our model”. However, this would increase the performance of the algorithm on the training set only. The model will memorize the training samples but it will fail to generalize to previously unseen samples. This is called **overfitting**. It is for this reason that we split the data in separate sets, as to get an unbiased score of the performance. Another form of overfitting is the over-tuning of hyperparameters based on results on the validation set. One possible solution to overfitting could be to stop training the model earlier [51]. However, we can not stop training the model too early, because it will not have captured the underlying process which we are trying to understand. Stopping the training process too early is called **underfitting**.

Both in over- and underfitting, the model will not perform well on samples it has never seen before. Therefore, we have to find some balance between these two. This is called the **bias-variance trade-off** [53]. When a model underfits, one says it has high bias as it does not have a lot of flexibility to capture the essence of the underlying process. On the other hand, when a model overfits, it has high variance as the model will capture not only the dependencies we want

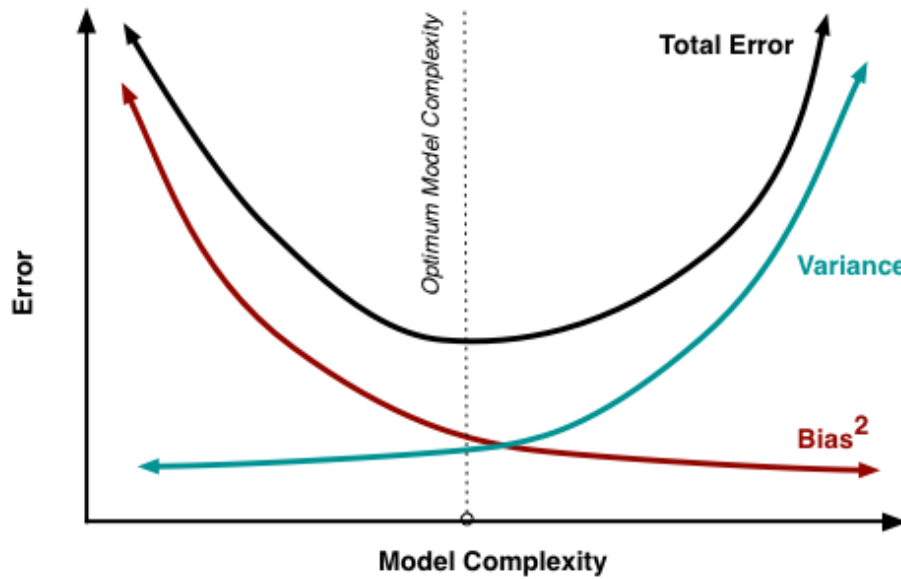


Figure 4.1: The bias-variance trade-off [52]

it to learn, but will also memorize the fluctuations of the training dataset, making it sensitive to small changes of the input features. This is summarized in figure 4.1, in which we would like to find the minimum of the error curve representing the loss function, as this configuration will have the optimal set of hyperparameters. In this figure, the underfitting region lies on the left-hand side, while the right-hand side depicts the region in which the model overfits.

There are multiple ways in which ML algorithms can learn. In **supervised learning**, an algorithm is trained on labeled examples, with the goal of predicting a label for new events in the future [54]. Contrary to supervised learning, **unsupervised learning** trains a model on a bunch of unlabeled data points, with the goal of, for example, finding some underlying structure in the data. Unsupervised learning techniques can also be used for outlier detection. Somewhere in between these two lies **semi-supervised learning**. In this case, some of the data points are labeled, while others are not. Such a scenario is for instance possible when the features of an event are relatively cheap to measure, while the correct label is much more expensive to determine [54]. Lastly, **reinforcement learning** algorithms interact with their environment, producing actions which will cause an error or yield a reward [55]. In this case, no labeled input-output pairs are needed, rather the algorithm will use a trial-and-error approach to maximize the gained reward.

In this analysis, we will be using supervised learning techniques to predict how likely a given

event is to be either a background or a signal event, based on its physical properties. The next few sections will go more in-depth on the ML algorithms which will be used.

4.2 Boosted Decision Trees

4.2.1 Decision trees

Decision trees segment the predictor space into a number of simple regions [54]. They consist of a series of nodes which get split according to some simple rules. These rules are not pre-defined and will be learned by the algorithm. One decision tree yields a form of rule-based classification or regression. An example of a very simple decision tree can be found in figure 4.2. In this particular example, we are using a regression tree to predict an output value between zero and one, where zero means background and one means signal.

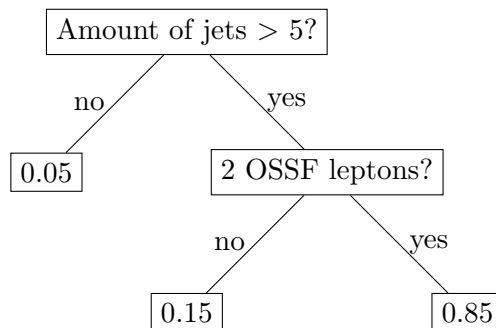


Figure 4.2: A very simple example of a decision tree.

These decision trees, of which the regression tree in figure 4.2 is an example, are trained using a top-down approach: we start from the first **node**, containing all training samples, and split this node into two¹ nodes. At each node, a certain condition is checked (e.g. the amount of jets), based on which the event will fall into one of the available **branches**. The conditions at the nodes are the parameters of the decision tree that will be trained. Finally, the outer nodes are called **leaf nodes** or **terminal nodes** and these will hold the prediction for events that follow all rules leading to this leaf node. The prediction at a leaf is determined by the mean of all training outputs reaching that leaf node.

The parameters of a node are chosen based on the impurity measure of each of the outgoing branches. The algorithm will try a number of different parameters, selecting the parameters

¹One can also split these into more than two nodes, but this can always be represented by a series of binary splits. This is why, for simplicity, it is assumed only binary splits are used here.

which decreases the impurity the most. An impurity measure frequently used is the residual sum of squares (RSS), defined in equation 4.1, where $\hat{y}_R$ is the mean output of the training samples in the considered node and x are the events in that node [54].

$$RSS = \sum_{i:x_i \in R} (y_i - \hat{y}_R)^2 \quad (4.1)$$

As we want the impurity of the (leaf) nodes to be as low as possible, the attribute A and value v at which the node will be split are decided by the combination which reduces the sum of the impurity of the two newly generated nodes most. In other words, the algorithm tries to minimize equation 4.2. This approach to training the decision tree is a so-called *greedy approach*, as we create the tree by making locally optimal decisions and it might not end up in the global minimum [56]. The training of the decision tree will continue until a stopping criterion is reached [54]. An example of a stopping criterion could be a limit on the amount of nodes between the starting point and output leaves.

$$\sum_{i:x_i \in R_1(A,v)} (y_i - \hat{y}_{R_1})^2 + \sum_{i:x_i \in R_2(A,v)} (y_i - \hat{y}_{R_2})^2 \quad (4.2)$$

Decision trees have a lot of advantages: they are relatively fast to train, are flexible with regard to the type of the input features and produce a very intuitive set of rules to decide the output value one should assign to the considered sample [54]. However, they are very prone to overfitting. There are numerous methods to prevent or reduce overtraining, such as pre- and post-pruning, bagging (which is used in random forests) or boosting [57, 58]. The latter, which we will use later in this thesis, will be explained in next section.

4.2.2 Gradient boosting

Boosted decision trees which make use of gradient boosting were first developed by J. H. Friedman [59]. In gradient boosting, the goal is to combine a set of single decision trees, which are so-called weak learners as their predictive power is limited, into an ensemble with much better predictive power, a so-called strong learner [60]. The gradient boosting algorithm will try to minimize a loss function, with a commonly used loss function being the mean squared error (MSE). Initially, the algorithm will predict the same value for all training samples, namely the average desired output value. The next step consists of calculating the residual, which is defined as the difference between the predicted output and the desired output. The following step of this

algorithm might seem a bit counter-intuitive: instead of training a decision tree with the goal of predicting the training outputs, we will train a decision tree to *predict the residuals*. When we have a prediction for the residual of all training samples, we will use the initial prediction for the output value together with the predicted residual to calculate a new output value. The predicted residual will be scaled by the learning rate before calculating a new output value. The learning rate is used to prevent overfitting and has a value between zero and one. Empirically it has been found that using small learning rates can improve the model's ability to generalize to new samples dramatically [61]. Now that we have a newly predicted output value, the algorithm can start all over again by calculating the new value of the residual, which will this time be the difference between the model output, which takes the previous residual into account, and the desired output. This process of calculating residuals, training decision trees to predict the residuals and generating a new predicted output value goes on until the desired amount of trees is reached, or until adding new trees does not further reduce the loss function.

After this basic description, I will also give a more rigorous mathematical explanation. The following section will be largely based on References [59, 62, 61, 63]. Let us assume we have a dataset $\{(x_i, y_i)\}_{i=1}^n$ with features x_i and desired outputs y_i , where n is the amount of samples. We use a differentiable loss function $L(y_i, F(x))$ with $F(x)$ being the predicted output. An example of a loss function is the previously mentioned MSE:

$$MSE = \frac{1}{2}(F(x) - y_i)^2 \quad (4.3)$$

The factor of $\frac{1}{2}$ is used to cancel out the factor of 2 achieved by taking the derivative of the loss function with respect to the predicted value. As the residual is just the predicted value minus the observed value, it is clear that the derivative of the loss function is just the negative residual, which makes math a lot easier for the algorithm as a lot of derivatives have to be calculated (hence *gradient* boosting). Now that we have a loss function and a dataset, we will initialize the model by getting a first prediction $F_0(x)$. This can be seen in the following equation, where γ refers to the predicted values:

$$F_0(x) = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, \gamma) \quad (4.4)$$

The initial predicted value we get by calculating this minimum through derivation, is a “decision tree” consisting of only one node: a leaf. After the initialisation, we get to the iterative part of the algorithm. We will start by calculating the so-called *pseudo-residuals* r_{im} ,

with $m \in \{1 \dots M\}$ and M the predefined number of trees:

$$r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)} \quad \text{for } i = 1, \dots, n \quad (4.5)$$

Due to the use of a derivative in equation 4.5, this method is referred to as *gradient* boosting. The r_{im} are called pseudo-residuals, because if we were to use another loss function (for instance the MSE without the factor $\frac{1}{2}$), we might not end up with the exact residual but something similar to it. The next part of the algorithm consists of fitting a tree h_m to the pseudo-residuals using a training set $\{(x_i, r_{im})\}_{i=1}^n$. This tree will have some terminal nodes R_{jm} , where $j \in 1 \dots J_m$ and J_m is the amount of terminal nodes in tree number m . Now we will calculate an “output value” for each leaf using the following equation:

$$\gamma_{jm} = \arg \min_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma) \quad (4.6)$$

It is also possible to calculate a single γ_m for all leaves combined, but using a separate value for each leave can further increase the quality of the fit according to Friedman [59]. This minimization is similar to the one done in the initialisation step, but here it also accounts for the previous predictions, whereas in the initialisation there was no previous prediction. Now it is time to make a new prediction for each sample. We will use the output values from the tree to update our predictions:

$$F_m(x_i) = F_{m-1}(x_i) + \nu \cdot \gamma_{jm} \quad (4.7)$$

In this equation, γ_{jm} is the output value of the leaf that the sample ends up in, while ν is the learning rate. This procedure can now be repeated until the desired number of trees M is reached. The final prediction of this algorithm for a sample x_i will then be given by $F_M(x_i)$. As each F_m requires the previous F_{m-1} to be calculated, we can see that the final prediction will use all trained trees to predict one output value for each sample. We can then use the output value to assign a class label to the considered sample, based on some threshold value (also called the working point). For instance, if an output of zero represents background and an output of one represents signal, a (naive) threshold value could be 0.5, where every sample with an output smaller than 0.5 would be classified as background, and all others would be classified as signal. It is however also possible to directly use the output values of the algorithm in your analysis, without deciding on a working point.

Using pseudocode, it is possible to summarize the algorithm as follows:

Algorithm 1: Gradient boosting

```

1  $F_0(x) = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, \gamma);$ 
2 for  $m = 1$  to  $M$  do
3    $r_{im} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$  for  $i = 1, \dots, n;$ 
4   fit learner to pseudo-residuals;
5    $\gamma_{jm} = \arg \min_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma);$ 
6    $F_m(x_i) = F_{m-1}(x_i) + \nu \cdot \gamma_{jm};$ 
7 end
8 Output predictions  $F_M(x);$ 

```

In the analysis presented in chapter 5 of this thesis, the tree boosting package XGBoost [64] developed by Tianqi Chen will be used. XGBoost is a library which includes highly flexible and optimized gradient boosting algorithms, creating a very user-friendly environment for training boosted decision trees.

4.3 Deep Neural Networks

Neural networks (NN) are machine learning algorithms which have an architecture which is based on a simplified model of the networks of neurons we find in the brain of all living creatures [65]. A NN consists of several layers of neurons. The first layer is the input layer, which has the same amount of neurons as your data has features [66]. The last layer of the NN is the output layer, which can have as many neurons as your particular case needs. In between the input and output layers, there are so-called hidden layers. When more than one hidden layer is used, we call the NN a *deep* neural network (DNN). In the analysis presented in this thesis, we will be using densely connected layers, meaning that each neuron located in a certain layer will be connected to all neurons of the layers directly next to this layer. A schematic overview of a NN with one hidden layer is given in figure 4.3, in which the layers are densely connected. The neurons of the different layers are connected using weights w_{ij} , which are the parameters of the NN which will be optimized through training. As previously mentioned, the neurons used in this analysis are only connected with neurons from the previous and next layer. More complex NN structures do exist, e.g. recurrent neural networks or convolutional neural networks, but we will not go more

in-depth into those architectures.

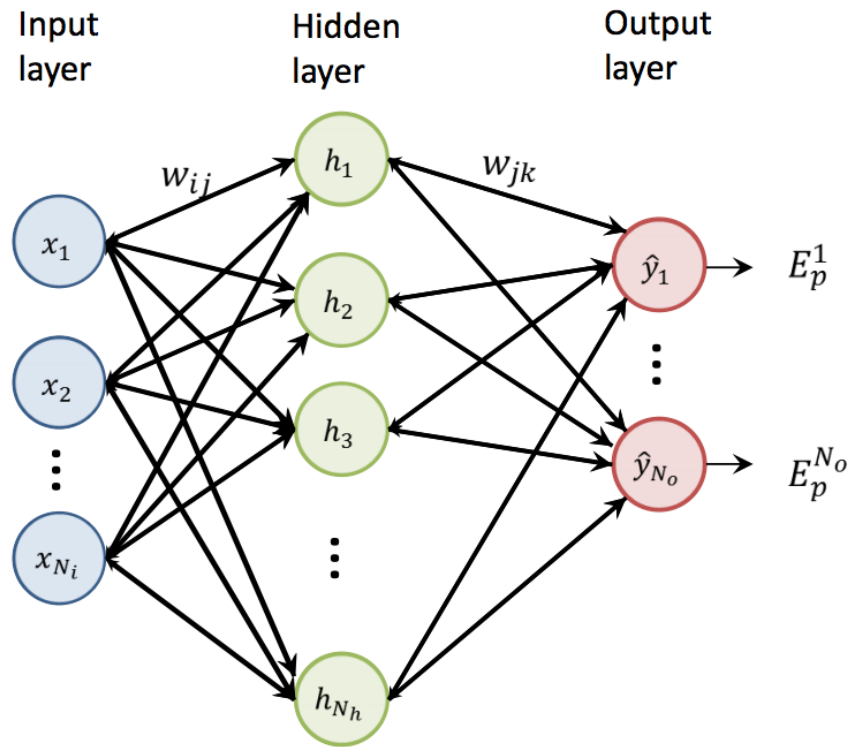


Figure 4.3: Neural network with one hidden layer [66]

Before covering the mathematics of the learning procedure of NNs, we will have a brief look at the flow of a training sample through the NN. Firstly, the features or variables of the training sample are given as an input to the input layer. In figure 4.3, the x_i values represent the features of the training sample. The next step consists of calculating the activations h_i in the hidden layer, by combining the values of the features with the weights connecting the input layer and hidden layer. The procedure for any possible extra hidden layers is the same, with the exception that the activations from the preceding hidden layer are then used to calculate the activations of the considered hidden layer. When this is done for all hidden layers, the output values in the output layer are calculated in completely the same manner, using the activations from the last hidden layer as inputs. From these outputs, the error with respect to the desired outputs can be calculated. Using these errors, updates to the weights connecting the last hidden layer to the output layer can be calculated. By using a technique called backpropagation, which we will come back to later, updates to the weights connecting the previous layers can also be calculated. This procedure is then repeated multiple times for all training samples.

We will now look at a more mathematical description of the NN learning process. Training

a NN can be formulated as an optimization problem: much as in gradient boosting, we define a loss function which we want to minimize with respect to the parameters of the model. A commonly used loss function is the MSE:

$$E = \sum_p E_p = \frac{1}{2} \sum_p (\hat{y}_p - y_p)^2 \quad (4.8)$$

In equation 4.8, $\hat{y}_p$ is the prediction of the model for learning sample p , while y_p is the desired output. Each neuron in a NN returns a value, called the activation h of this neuron. The calculation of this activation goes as follows. Firstly, we will take the sum over a linear combination of the activations of the neurons from the preceding layer and the weights w_{ij} connecting these to the considered neuron. Next, we add the so-called bias θ to this sum, giving rise to the quantity s_k^p in equation 4.9, in which N_h is the amount of neurons on the previous layer. Lastly, the activation of the considered neuron can be calculated by feeding s_k^p to the activation function ϕ as in equation 4.10. This will transform the weighted sum of the input into the activation h of the neuron in question [67].

$$s_k^p = \sum_{j=1}^{N_h} h_j^p w_{jk} + \theta_k \quad (4.9)$$

$$\hat{y}_k^p = \phi(s_k^p) \quad (4.10)$$

Now that we can calculate the activations for all neurons, we can start training the algorithm. A NN learns by updating its weights, giving rise to different activations for the neurons. The goal is to minimize a loss function such as the one in equation 4.8. However, E_p is a complex non-linear combination of the input parameters and cannot be solved analytically [66]. Instead, the gradient of E_p can be calculated numerically, which allows to iteratively approach a local minimum. This procedure is known as gradient descent [61], in which weights are adapted by taking a step in the negative direction of the gradient every iteration, with the goal of ending up in a (local) minimum of the loss function. The change of a certain weight $\Delta_p w_{jk}$ due to a training sample p would then be given by:

$$\Delta_p w_{jk} = -\gamma \frac{\partial E_p}{\partial w_{jk}} \quad (4.11)$$

The γ in equation 4.11 is the learning rate. Computing all gradients can be a very expensive task, but for the last layer, the gradient of the loss function can be calculated quite easily.

When this gradient is calculated, we can use the chain rule of calculus to give the gradient in the previous layer, and so on [66]. This is called backpropagation. Now we will calculate the change of a weight connecting the output layer to the last hidden layer. Using 4.8 we get:

$$E_p = \frac{1}{2}(\hat{y}_p^2 - 2\hat{y}_p y_p + y_p^2) \quad (4.12)$$

$$\Rightarrow \frac{\partial E_p}{\partial \hat{y}_p} = \hat{y}_p - y_p \quad (4.13)$$

Using the chain rule twice, as well as equations 4.9, 4.10 and 4.13, we get the following expression for the change to the weight:

$$\Delta_p w_{jk} = -\gamma \frac{\partial E_p}{\partial w_{jk}} \quad (4.14)$$

$$= -\gamma \frac{\partial E_p}{\partial \hat{y}_k^p} \frac{\partial \hat{y}_k^p}{\partial s_k^p} \frac{\partial s_k^p}{\partial w_{jk}} \quad (4.15)$$

$$= -\gamma(\hat{y}_k^p - y_k^p) \phi'(s_k^p) h_j^p \quad (4.16)$$

$$= \gamma(y_k^p - \hat{y}_k^p) \phi'(s_k^p) h_j^p \quad (4.17)$$

We can interpret $\Delta_p w_{jk}$ as follows: γ is the step size or the learning rate of the gradient descent. The difference between the expected and observed output is given by $(y_k^p - \hat{y}_k^p)$. $\phi'(s_k^p)$ is the derivative of the activation function evaluated at output node k and lastly, h_j^p is the output of hidden node h_j for training sample p . Now that we calculated the weight updates for the transition between the last hidden layer and the output layer, we can use this result to calculate the weight updates connecting to the previous layer, and so on, until we have reached the input layer. Training the NN now consists of updating all weights for each training sample n times, with n called the number of *epochs*².

Neural networks come in a plethora of configurations. We can change the loss and activation functions, the amount of neurons in the layers, the amount of (hidden) layers and much, much more. These types of parameters of the NN are called hyperparameters. There is no golden standard for the optimal configuration of hyperparameters: they are highly dependent on the case at hand. We might be able to find a configuration of hyperparameters which suit our problem very well, but there is no guarantee that it will give good results in another context.

²Actually, the training data is usually divided into batches and the weights will only be updated when a batch is finished.

One way of optimizing the choice of hyperparameters is through the use of a genetic algorithm, which we will discuss in more into detail in the next section.

In this analysis, we will be using a standard DNN with densely connected layers. The amount of layers used will not be predefined, we will use algorithms like a grid-search and genetic algorithms to find the ideal amount. However, the output layer will have a predetermined number of neurons: one. The interpretation of the output of the DNN will be analogous to that of the BDT, although the output shape (histogram of all model outputs from one ML algorithm) might be completely different. To construct the DNNs, we will be using Keras, an open-source deep learning library for Python [68].

4.4 Genetic algorithms

A genetic algorithm is a search heuristic inspired by the theory of natural evolution by Charles Darwin [69]. In this application, a population of ML algorithms (DNNs or BDTs) will be allowed to create a population with a better average performance. This requires some kind of metric to judge the goodness of a given model, the fitness function. The metric of choice in this analysis will be the AUC, which is short for the area under the curve (AUC) of the receiver operating characteristic (ROC). The ROC grades the performance in a two-class classification problem (signal-background) by plotting the true negative rate (= background rejection) with respect to the true positive rate (= signal efficiency). There are different definitions of the ROC, but this is the one we will be using here.

The genetic algorithm starts by taking a number of randomly initialized sets of hyperparameters and training models with these sets. After this, we grade the model's performances using the AUC. Then, the genetic algorithm sorts these models by performance and gets rid of the worst 50%. The remaining 50% will be allowed to mutate with each-other, giving rise to the same number of algorithms we started with, but the hyperparameters should fit our training data better. It is clear how Darwin's "survival of the fittest" principle comes into play here. The effect of a genetic algorithm on the learning rate is visible in figure 4.4. Here, six generations of models were trained, each one evolving from the previous one. It is evident that, after a few generations, the learning rate seems to converge to the lower end of the spectrum.

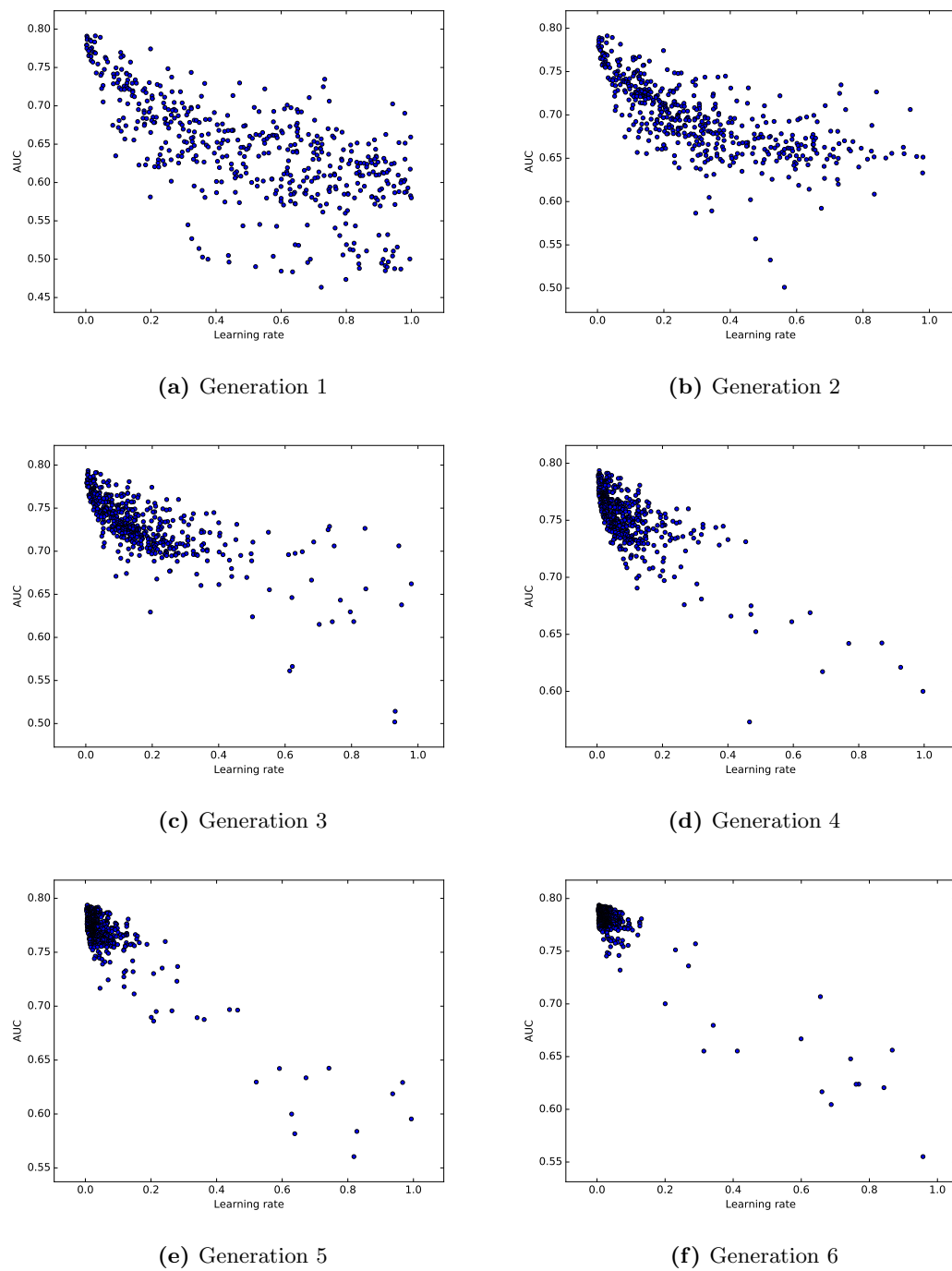


Figure 4.4: The AUC with respect to the learning rate for six generations of models.

Chapter 5

Search for the two lepton decay mode of the $t\bar{t}Z$ process

From the discussion of the top quark in chapter 2, it is clear that measurements of the top quark interactions are a very interesting research field. The analysis presented in this thesis will focus on the process where a top-antitop quark pair is produced with a Z boson radiating off one of the two top quarks. In short, this process is called the $t\bar{t}Z$ process. The $t\bar{t}Z$ process has already been studied at CMS, with the most recent cross-section measurement at a COM of 13 TeV being $\sigma(t\bar{t}Z) = 0.95 \pm 0.05$ (stat) ± 0.06 (syst) pb [70]. The purpose of this thesis is to present a similar analysis, but using a different topology. The data used to measure the cross section was taken in 2018 at a COM of 13 TeV at the CMS detector. This chapter will present an overview of the signal signature and describe its main backgrounds. It will also cover the object and event selections which were used in this analysis. Subsequently, the implementation and results of ML algorithms will be discussed. Finally, an estimate for the cross section of the $t\bar{t}Z$ process is measured.

5.1 Signal signature

We will start the analysis part of this thesis with a discussion of the signal, $t\bar{t}Z$. Firstly, we will discuss the leading order production mechanisms for the $t\bar{t}Z$ process. Secondly, the decay modes of the signal will be described. Lastly, we will go more in-depth into the specific signal topology which is being studied in this analysis.

5.1.1 Production mechanisms

As with most processes, $t\bar{t}Z$ production can be achieved through multiple mechanisms. In figure 5.1, the leading order Feynman diagrams for $t\bar{t}Z$ production can be seen. These consist of one initial state radiation (ISR) and three final state radiation (FSR) processes. In ISR, the incoming particles radiate a particle prior to the interaction. In the topology studied in this thesis, the incoming particles are quarks and the radiated particle is the Z boson. Contrary, in FSR, the Z boson is radiated by the scattered particles, which are the top quarks in the $t\bar{t}Z$ process. Interestingly, for the associated production of a W boson, the only leading order contribution would be similar to 5.1a, with the Z boson interchanged with a W boson. We will not discuss $t\bar{t}W$ production here, but more information can be found in Refs. [71, 72].

5.1.2 Decay mechanisms

The CMS detector will not detect Z bosons and top quarks directly, rather it will detect their decay products. There are three main decays modes for the Z boson: leptonically, hadronically and invisible. These decay modes, along with their respective branching ratios (BR), are given in table 5.1. It should be noted that the $q\bar{q}$ pair from the hadronic decay will, due to the colour confinement phenomenon discussed in chapter 2, hadronize and form jets, which are the objects that can be reconstructed from data.

Decay mode	Resulting particles	BR
Leptonic	$Z \rightarrow l^+ l^-$	10.09%
Hadronic	$Z \rightarrow q\bar{q}$	69.91%
Invisible	$Z \rightarrow \nu\bar{\nu}$	20.00%

Table 5.1: Decay modes of the Z boson with their respective end products and BRs. The BRs were found in Ref. [6].

Decay mode	Resulting particles	BR
Leptonic	$t \rightarrow bW \rightarrow bl\nu$	33.8%
Hadronic	$t \rightarrow bW \rightarrow bq\bar{q}$	66.5%

Table 5.2: Decay modes of the top quark with their respective end products and BRs. The BRs do not add up to exactly 100%, this is due to the relatively large errors on the measurements of the BRs. The BRs were found in Ref. [6].

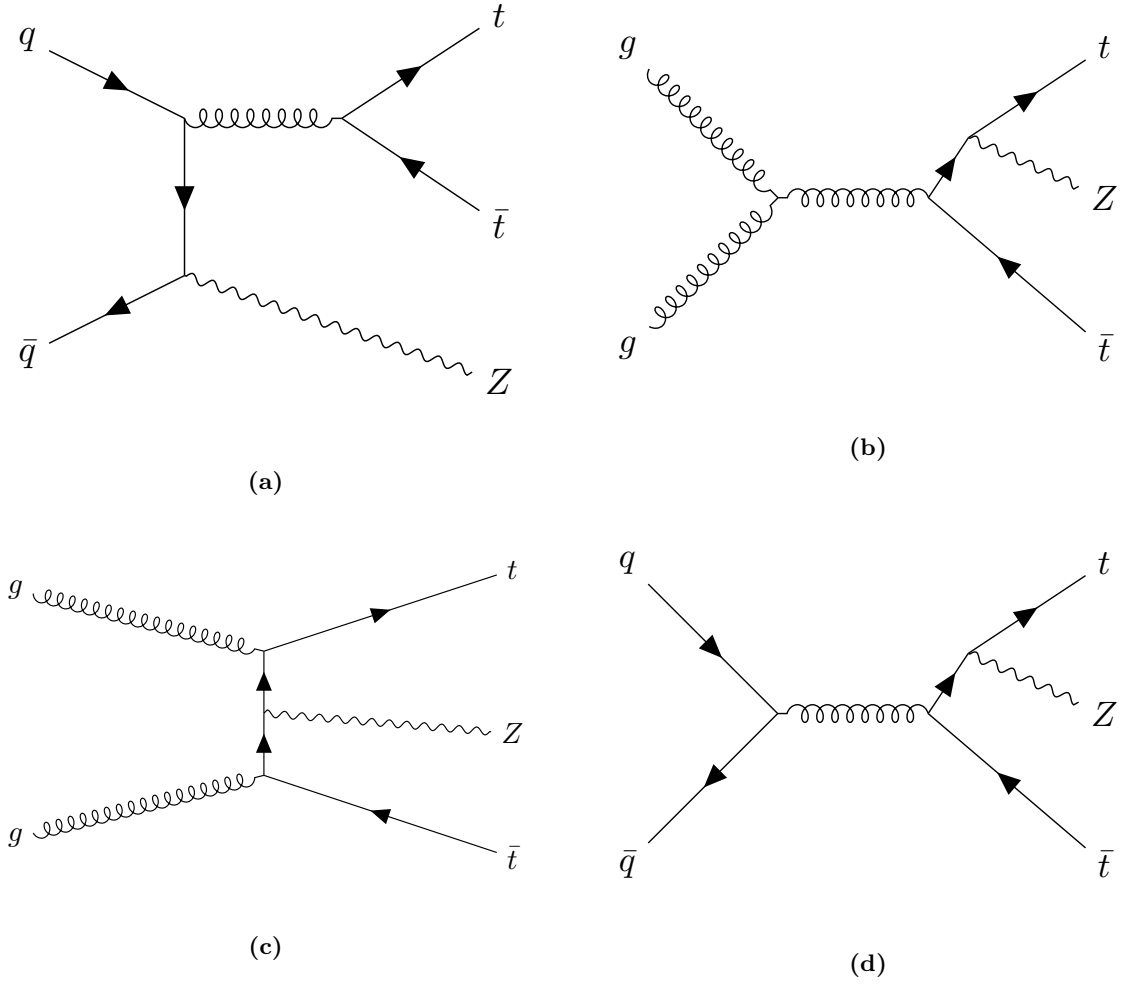


Figure 5.1: Leading order Feynman diagrams of the $t\bar{t}Z$ process. Figure (a) shows $t\bar{t}Z$ production from initial state radiation, while figures (b), (c) and (d) show final state radiation production mechanisms.

The top quark decays quasi exclusively to a W boson and a bottom quark [6]. The W boson can either decay leptonically or hadronically, while the b quark will always hadronize and form a b -jet. B-jets are particularly interesting in top quark decays, as a top quark will virtually always decay into a W boson and a b quark, and the presence of the latter can be exploited to discriminate top quark processes from uninteresting background processes. After jets have been reconstructed, an algorithm is used to determine whether a certain jet originated from a b quark or not. This algorithm will tell whether or not a jet is b-jet like, meaning that not all true b-jets will be tagged and some non-b-jets will be tagged mistakenly. In section 5.5 we will see that the working point for this algorithm used in this analysis will result in a mistag rate

of about 1%. As the discussion of these algorithms lies out of the scope of this thesis, we will not go more in-depth here, but more information can be found in Ref. [73]. An overview of the decay modes of the top quark, which are in turn determined from the decay modes of the W boson, can be seen in table 5.2.

5.1.3 Signature of the two lepton decay mode

As mentioned earlier, the analysis presented in this thesis focuses on the two lepton decay mode of the $t\bar{t}Z$ process, in which the Z boson decays to two opposite sign (OS) leptons and the top quarks¹ decay hadronically. Performing the analysis in the hadronic or invisible decay channels of the Z boson would in principle be possible, but these would be severely dominated by background, even more so than the two lepton case. They also require a completely different analysis and hence are out of the scope of this thesis. Taking into account the branching ratios from tables 5.1 and 5.2, the signal composition can be seen in figure 5.2. The channel of interest here is the two leptons OS which constitutes 10.2% of total $t\bar{t}Z$ events. Due to lepton universality, a third of the two OS lepton events will consist of tau leptons. A tau lepton is very short lived and will decay before the detector is able to detect it. It can decay both hadronically and leptonically, giving rise to a completely different signature than electrons or muons. This would require a separate analysis and hence is out of the scope of this thesis. The attentive reader might have noticed that there is a small contribution from a two SS lepton final state in figure 5.2, even though no combination of decay modes in tables 5.1 and 5.2 can give rise to such a final state. However, it is included in the pie chart as the pie chart actually represents the events as they would be observed by the detector, meaning that insufficiencies are also incorporated. The two SS lepton final state can then be caused by a number of mismeasurements: it is possible that the Z boson decays to two leptons, of which one goes undetected. An additional lepton, possibly of the SS as the detected lepton, can originate from the leptonic decay of the W boson or from a lepton contained within a b-tagged jet. Another possible artefact is a charge mismeasurement. The result achieved by combining the branching ratios in tables 5.1 and 5.2 with all detector insufficiencies is the pie chart in figure 5.2.

The two lepton channel, which will be considered in this thesis, has several key characteristics. In figure 5.3, the Feynman diagram for this particular process can be seen. Firstly, we can see that this channel is characterized by having exactly two leptons. The demand that the Z boson

¹Actually it is the W boson originating from the top quarks which decays hadronically, but this is often referred to as a hadronic top decay.

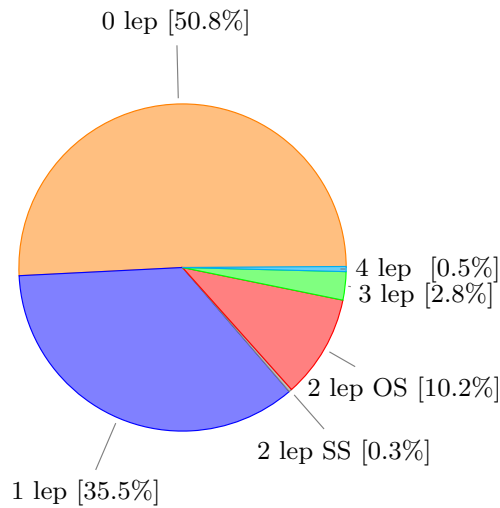


Figure 5.2: A pie chart showing the distribution of the different final states resulting from the $t\bar{t}Z$ process. OS indicates opposite sign leptons, while SS signifies same sign leptons.

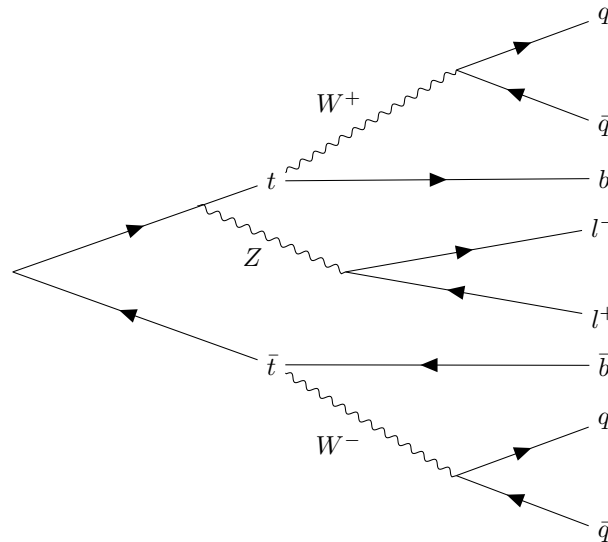


Figure 5.3: The Feynman diagram for the decay channel where the Z boson decays to two charged leptons and the W bosons both decay hadronically.

has to decay into two charged leptons requires these two leptons to be of opposite sign and same flavour (OSSF). The invariant mass of this dilepton pair has to lie close to the mass of the Z boson mass of 91.188 GeV. We do not require it to be exactly this mass, as the detector has some insufficiencies, which might lead to a mass slightly different than the Z boson mass. The hadronic decay of the top quark will give rise to three jets, of which one will be b-tagged. This

means that in total, this channel is expected to produce 6 jets, two of which are b-jets. However once again, the detector is not perfect, which necessitates us to loosen the requirements in this case as well. We will require a $t\bar{t}Z$ candidate event to have at least 5 jets, of which at least one is b-tagged. Lastly, as all end products of this process can be detected by the CMS detector, we expect the events to have no *true* MET. Any event, even if no MET is expected, will always lead to some MET. This MET is referred to as *fake* MET and originates from miscalibrations or mismeasurements of physics objects in the detector.

5.2 Dominant backgrounds

The previous section described the signal signature of the process which is being studied in this thesis. However, there are other processes which will yield a similar detector signature to the signal. Some of these processes have a generally much higher production cross section than the studied signal, which poses a real challenge for the presented analysis. Below we describe the two main backgrounds: $t\bar{t}$ production and the Drell-Yan (DY) process.

5.2.1 Drell-Yan

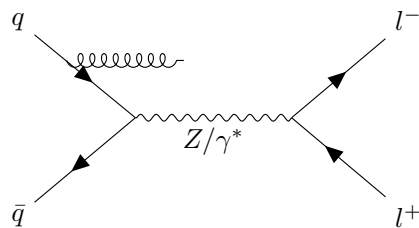


Figure 5.4: The Feynman diagram for the Drell-Yan background process.

The Feynman diagram for the DY process can be seen in figure 5.4. In the DY process, a quark-antiquark pair annihilates and forms a Z boson or a virtual photon γ^* . As can be seen in table 5.1, the Z boson decays to an OSSF pair 10.09% of the time, giving a background to the $t\bar{t}Z$ process. In case the event was very pure, one would not expect any jets arising from the DY process, let alone b-jets. However, the initial state may radiate gluons, which will hadronize and form jets, yielding a signature similar to that of the $t\bar{t}Z$ process. As the cross section for the DY process is very large, namely about four orders of magnitude larger than the predicted cross section of the $t\bar{t}Z$ process, and its decay products contains an OSSF lepton pair, it is especially

challenging to suppress this background source.

5.2.2 $t\bar{t}$

The second dominant background is top-antitop quark pair production. In order for $t\bar{t}$ production to be considered a background, both of the W bosons have to decay leptonically. As can be seen in the Feynman diagram for $t\bar{t}$ production in figure 5.5, the two leptons resulting from the leptonic W decays are of OS. It is also possible that they are of the SF, in which case they could fake a Z boson signature. By requiring the reconstructed dilepton mass to be near the Z boson mass, the background can already be suppressed significantly. Additionally, rejecting all events with opposite flavour (OF) leptons will also reduce this background considerably. The neutrinos from the leptonic W decays also cause some true MET to be present, which we do not expect in the $t\bar{t}Z$ process. In $t\bar{t}$ production, only two jets are expected, which should both be b -tagged. However, much as in the DY background, the hadronising of ISR gluons or FSR from top and bottom quarks can generate fake jets, making $t\bar{t}$ production one of the dominant backgrounds to the $t\bar{t}Z$ process.

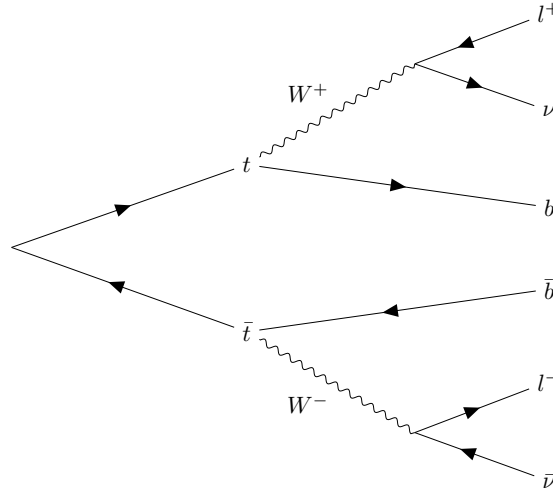


Figure 5.5: The Feynman diagram for the $t\bar{t}$ background process.

5.2.3 Other backgrounds

While the DY process and $t\bar{t}$ production are certainly the dominant backgrounds, they do not constitute the total background. There are a lot more processes contributing to the background, one of them being the previously mentioned $t\bar{t}W$ process. The main contributions to the back-

ground arise from processes including at least one top quark in association with another boson, for instance $t\bar{t}H$ or tZq . These background processes are collectively denoted by $t(\bar{t})X$, with the exception of $t\bar{t}W$ which will be treated as a separate background source. All other background processes will only contribute by a small amount and will be denoted as “other”.

5.3 Monte Carlo samples

All background and signal sources are estimated using simulated Monte-Carlo (MC) samples. Different processes are modelled using different generators. Table 5.3 gives an overview of which generator is used to model the considered signal and background processes. A list of all used MC samples is shown in Appendix A. The various backgrounds are grouped into the five different groups which were described in the previous sections: $t(\bar{t})X$, $t\bar{t}W$, $t\bar{t}$, DY and “other”.

Process	Event generator
$t\bar{t}Z$, WWZ , WZZ , ZZZ , $t\bar{t}WW$, $t\bar{t}WZ$, $t\bar{t}ZZ$	MadGraph5_aMC@NLO [74, 75]
$t\bar{t}W$ + jets	MadGraph_aMC@NLO Madspin [76, 77]
$t\bar{t}$ + jets, DY, W + jets	MadGraph_aMC@NLO MLM [74]
$t\bar{t}H$ (no $b\bar{b}$), tW , $\bar{t}W$, ZZ , WW	POWHEG v2 [78, 79]
tHq , tHW	MadGraph5 [74, 75]
tZq , Single t (s-channel), $t\bar{t}t\bar{t}$, Single t , WWW	aMC@NLO [80]
Single $t/\bar{t}$ (t-channel)	POWHEG v2-Madspin [76, 77, 81]
WZ	MadGraphFxFx [82]
$gg \rightarrow g \rightarrow ZZ$	MCFM v7.0.1 [83]
WH , ZH , $gg \rightarrow H \rightarrow ZZ$, $H \rightarrow ZZ$	POWHEG v2 JHUGen v7.0.11 [79, 84]

Table 5.3: Event generators used to simulate events for the various processes.

5.4 Lepton identification

In chapter 3, we discussed how the PF algorithm reconstructs physical objects like leptons. When studying a given event, it is necessary to distinguish prompt and non-prompt leptons. A prompt

lepton is a lepton originating from the main collision taking place in the detector. It is a direct result of the specific decay you are looking for. In this analysis, the prompt leptons are those originating from the leptonic decay of the Z boson or, in the case of some backgrounds, from the leptonic decays of top quarks. On the contrary, non-prompt leptons are leptons which arise from the hadronization of heavy flavor quarks or objects which have mistakenly been identified as a lepton. For instance, one of the most common decay products of b mesons is a lepton. Much like prompt leptons, these leptons will leave the same tracks and hits in relevant parts of the detector. However, by correlating these hits with nearby jets, in this case the b-tagged jet, it is possible to remove these non-prompt leptons. In order to have a high prompt lepton identification efficiency while keeping the mis-identification rate of non-prompt leptons as low as possible, additional quality cuts will be applied to the lepton objects. The next sections will cover these quality cuts and a summary can be found in table 5.4.

Criterion	Electrons	Muons
leptonMVA	> 0.4	
leading p_T	> 30 GeV	
sub-leading p_T	> 20 GeV	
$ \eta $	< 2.5	< 2.4
I_{rel}	< 0.4	
SIP _{3D}	< 8	
$ d_{xy} $	< 0.05 cm	
$ d_z $	< 0.1 cm	
IPOGLoose	-	✓

Table 5.4: Lepton selection criteria

5.4.1 Kinematic cuts

As the LHC's beam pipe runs through the CMS detector, the detector is unable to provide full angular coverage. Due to the detector not being hermetic, we will apply cuts on the accepted pseudorapidity range. The inner tracker provides us with measurements up to $|\eta| = 2.5$, which is why we require all leptons to have $|\eta| < 2.5$. Furthermore, the muon chambers reach up to $|\eta| = 2.4$, meaning we will only accept muons with $|\eta| < 2.4$. Additional to the η cuts, we also apply transverse momentum (p_T) cuts. These cuts are justified as in general, low p_T leptons will

not be of interest for the signal signature that is being studied in this analysis. We will require the leading lepton to always have $p_T > 30$ GeV, while the sub-leading lepton p_T must be above 20 GeV. The same p_T cuts are made for electrons and muons.

5.4.2 Isolation cuts

Leptons originating from the decay of a heavy particle like the Z boson are usually well separated from any hadronic activity in terms of angular coordinates. We can thus require the leptons to be isolated from jets in order to reduce the background. Usually, one would define a cone around the lepton and then require that there is not a lot of hadronic or photonic activity within this cone. However, the distinction becomes less evident for heavily boosted topologies. For this reason, a new variable I_{rel} is defined in equation 5.1, yielding a degree of isolation for leptons, even when considering a heavily boosted event [85]. In this equation, $\sum_{\text{charged}} p_T$, $\sum_{\text{neutral}} p_T$ and $\sum_{\gamma} p_T$ are respectively the scalar sums over the transverse momenta of charged hadrons, neutral hadrons and photons, located in a cone centered on the lepton direction in (η, ϕ) space [85]. The size of the cone used in this analysis is 0.3. The contribution of pileup photons and neutral hadrons is estimated by the product of an effective cone area A_{eff} with the neutral-particle p_T density ρ_{neutral} . The latter is obtained through the FASTJET algorithm, taking into account all PF photons and neutral hadrons in the event. This algorithm will not be discussed here, but more information can be found in [86]. We require I_{rel} to be smaller than 0.4 for both muons and electrons.

$$I_{rel} = \left(\sum_{\text{charged}} p_T + \max \left[0, \sum_{\text{neutral}} p_T + \sum_{\gamma} p_T - \rho_{\text{neutral}} \times A_{\text{eff}} \right] \right) / p_T^l \quad (5.1)$$

5.4.3 Vertex cuts

Not only are leptons originating from the decay of the Z boson usually well separated from hadronic activity, they are also created close to the primary vertex where the proton-proton collision occurred. The reason for this is that heavy particles generally have a short lifetime, and will decay almost instantaneously after being created. On the contrary, leptons originating from pileup events or mistakenly reconstructed tracks are usually not created near the primary interaction vertex. By demanding leptons to have their origin close to the primary vertex, non-prompt leptons can be rejected. Three variables are used to quantify this requirement: $|d_{xy}|$ and $|d_z|$, which are defined as the distance of closest approach of the track with respect to the primary

vertex respectively in the transverse plane and along the beam axis, and the significance of the impact parameter SIP_{3D} , defined as the impact parameter divided by its standard deviation. In this analysis we will require $|d_{xy}| < 0.05$ cm, $|d_z| < 0.1$ cm and $SIP_{3D} < 8$.

5.4.4 Lepton MVA

The quality of a lepton object is also assessed using a multivariate discriminator. Among the input variables to this MVA are the variables previously described in the sections regarding the kinematic, isolation and vertex cuts. A more precise description of all used input variables can be found in Ref. [87]. Because the lepton MVA combines a multitude of variables, it yields the most important quality criterion. In principle, using just the MVA cut could be enough, but having well defined kinematics is necessary for the measurement and a looser definition of the lepton criteria was kept for possible cross checks. The lepton MVA used in this analysis is the “Ghent lepton MVA” [88], of which the ROC curve can be seen in figure 5.6. It was preferred over the $t\bar{t}H$ MVA and the TOP MVA as the former is not optimized for $t\bar{t}Z$ analyses, while the latter is still under development. In this analysis, leptons are required to have a MVA output greater than 0.4, corresponding to the medium working point. This working point yields an efficiency of 81% and 87% for electrons and muons respectively, for all p_T ranges [89]. The electron fake rate is lower than 1%, while the muon fake rate is 1.5%.

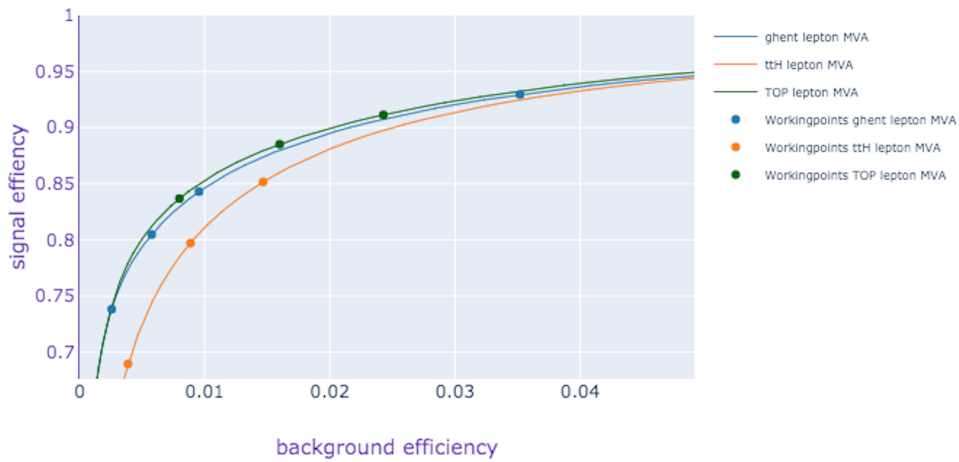


Figure 5.6: The ROC curve for three different lepton MVAs. This analysis made use of the Ghent lepton MVA. Figure produced using Ref. [89].

5.4.5 Extra muon identification cut

An additional way to identify muons is used, based on the recommendation of the physics object group (POG). In this analysis, the “POG Loose” ID will be used, yielding a muon identification efficiency above 95%. More details on how muons are identified can be found in Ref. [90].

5.5 Jet identification

Criterion	regular jet	b-jet
p_T	$> 30 \text{ GeV}$	
$ \eta $	< 2.4	
jetIsTight	✓	
ΔR_l	> 0.4	
DeepCsv_b	-	> 0.4941

Table 5.5: Jet selection criteria

Much like the lepton identification requirements, we also require jets to pass some extra cuts, yielding better quality jets overall. The kinematic jet cuts resemble those for leptons, as the motive for applying them is the same. In every event, there will be a large number of jets caused by pileup. These jets generally have a transverse momentum near the lower end of the spectrum, making employing a cut on the transverse momentum of the jets an effective way to get rid of a large part of these uninteresting pileup jets. We will require all jets to have at least a transverse momentum of 30 GeV. For this analysis, we limit the pseudorapidity range to $|\eta| < 2.4$. Jets are reconstructed up to $|\eta| = 5$, but the precision of their reconstruction drops significantly. To reject noisy and mismeasured jets, we require the jets to be tight. Whether or not a certain jet can be regarded as a tight jet is decided by the PF algorithm described in chapter 3. In order to avoid double counting leptons as jets, the separation ΔR_l between a certain jet and the closest lepton is defined by equation 5.2. We require ΔR_l to be at least 0.4. The process of removing possibly double counted leptons is also referred to as “jet cleaning”. Lastly, b-tagged jets have to be identified. As discussed in section 5.1.2, a MVA called deepCSV will be used for this purpose. More details on this algorithm can be found in Ref. [73]. In the analysis presented in this thesis, the medium working point for the deepCSV algorithm will be used. All jet identification requirements are summarized in table 5.5.

$$\Delta R_l = \sqrt{(\Delta\eta)_{j,l}^2 + (\Delta\phi)_{j,l}^2} \quad (5.2)$$

5.6 Event selection

Based on the requirements discussed in section 5.1.3, we can impose some event selections. These will be used to distinguish between signal and background events. For instance, the requirement of the dilepton mass to be close to the mass of the Z boson will get rid of a significant part of the $t\bar{t}$ background, as there is no reason why there would be a peak around the Z boson mass in the dilepton mass distribution for $t\bar{t}$ production. The event selection requirements based on the discussion from section 5.1.3 are summarized in table 5.6.

The effect of the cuts on the amount of jets can be seen in figure 5.7a. By requiring at least five jets to be apparent, the first bin will be rejected as it contains events with only four jets. It is clear that we will lose some signal due to this cut, but the amount of background events rejected severely outweighs this signal loss. One could argue that it might also be advantageous to reject events containing five jets, but this would reject only half of the amount of background events that are lost by rejecting the four jets bin, while rejecting an equal amount of signal. In order to retain at least some statistics, the cut on the amount of jets is placed at five and not six jets. The same story goes for the amount of b-tagged jets, which can be seen in figure 5.7b. The zero b-tagged jets bin contains some signal, but above all, it contains an enormous amount of DY background events, which is as expected. Removing the one b-tagged jet bin would once again punish the statistics severely, rejecting almost half of the remaining signal events, while the gain in signal-to-background ratio (SBR) is relatively small. This justifies performing the cut at one b-tagged jet instead of two.

Leptons	Jets
Prompt leptons == 2	≥ 5 jets
OSSF	≥ 1 bjet
$81 \text{ GeV} < m_{ll} < 101 \text{ GeV}$	

Table 5.6: Event selection criteria

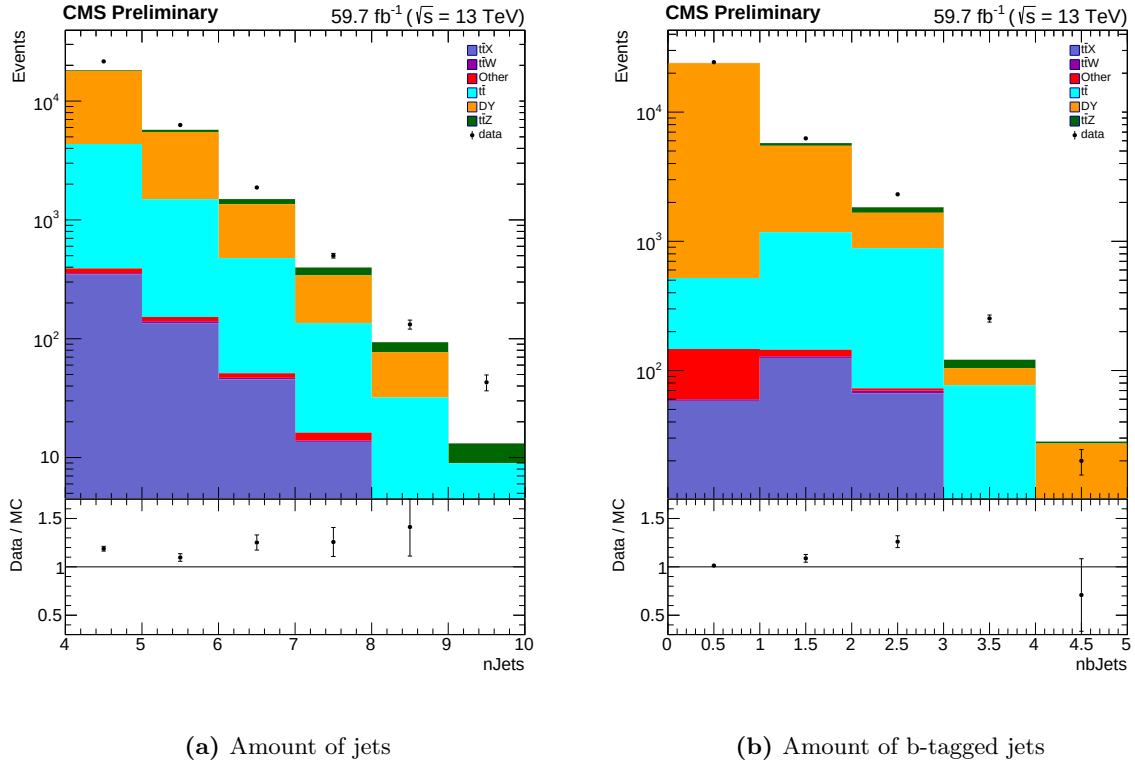


Figure 5.7: The amount of jets and the amount of b-tagged jets, achieved by removing the amount of jets and amount of b-tagged jets cuts respectively. In both cases, all remaining event selection criteria listed in table 5.6 are applied.

5.7 Triggers

In this analysis, various triggers will be used to select the data before applying the lepton and jet identification requirements and the event selection cuts. Data events were accepted if at least one of the following four triggers was fired: single muon, single electron, double muon or double electron.

5.8 Signal region

Using the object and event selections from the previous sections, some signal region (SR) plots can be generated. The invariant mass of the dilepton pair is presented in figure 5.8a. There is a clear peak around the Z boson mass in the distribution of the signal, $t\bar{t}Z$. Likewise, such a peak is also apparent for the DY background, which matches our expectations. For the second dominant background process, $t\bar{t}$, no peak is expected. This is also clearly visible, as the $t\bar{t}$

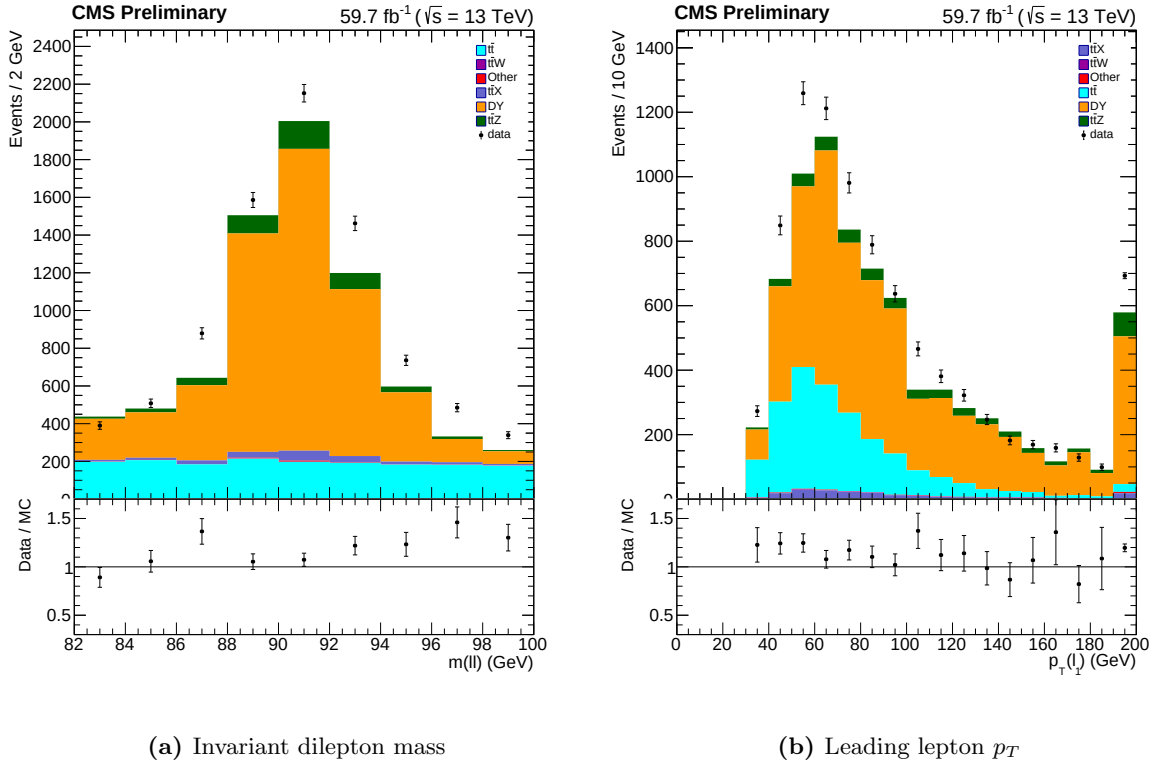


Figure 5.8: Reconstructed invariant dilepton mass and leading lepton transverse momentum with signal region selections applied.

spectrum shows a rather flat distribution. We can also see that the agreement between data and MC is not perfect, but tolerable. The MC underestimates the data, but the data does follow the same general shape as the MC. This same phenomenon is visible in figure 5.8b, which shows the transverse momentum of the leading lepton. The origin of the underestimation probably lies in the fact that the simulation software is not able to model events with more than four ISR jets accurately. This analysis requires each event to have at least five jets, so there will be a discrepancy between the data and MC. We will observe this effect in the next section, where control plots for the DY background are shown. More SR plots, showing for instance variables like the amount of jets or the pseudorapidity, can be found in appendix B. Possibly the most important observation which can be made from figure 5.8a is that the distribution is still heavily dominated by the background, even after applying the event selections. It is evident that the DY process is the leading source of background, followed by the $t\bar{t}$ background, just as we expected from the discussion in section 5.2. With a SBR of about 1/16 after applying all event selection criteria, measuring the cross section accurately and precisely will be a challenging endeavor.

5.9 Control regions

In order to check whether the backgrounds are correctly modelled, some control region (CR) plots will be made. For this purpose, we will take the event selections from section 5.6 and modify them to enhance the background source of which we want to assess the modelling quality. The object selections will remain completely the same, as we still want the leptons and jets to be of high quality. As figure 5.8a shows, DY and $t\bar{t}$ are the main backgrounds by far, so we will only make control plots for these two background sources.

5.9.1 Drell-Yan

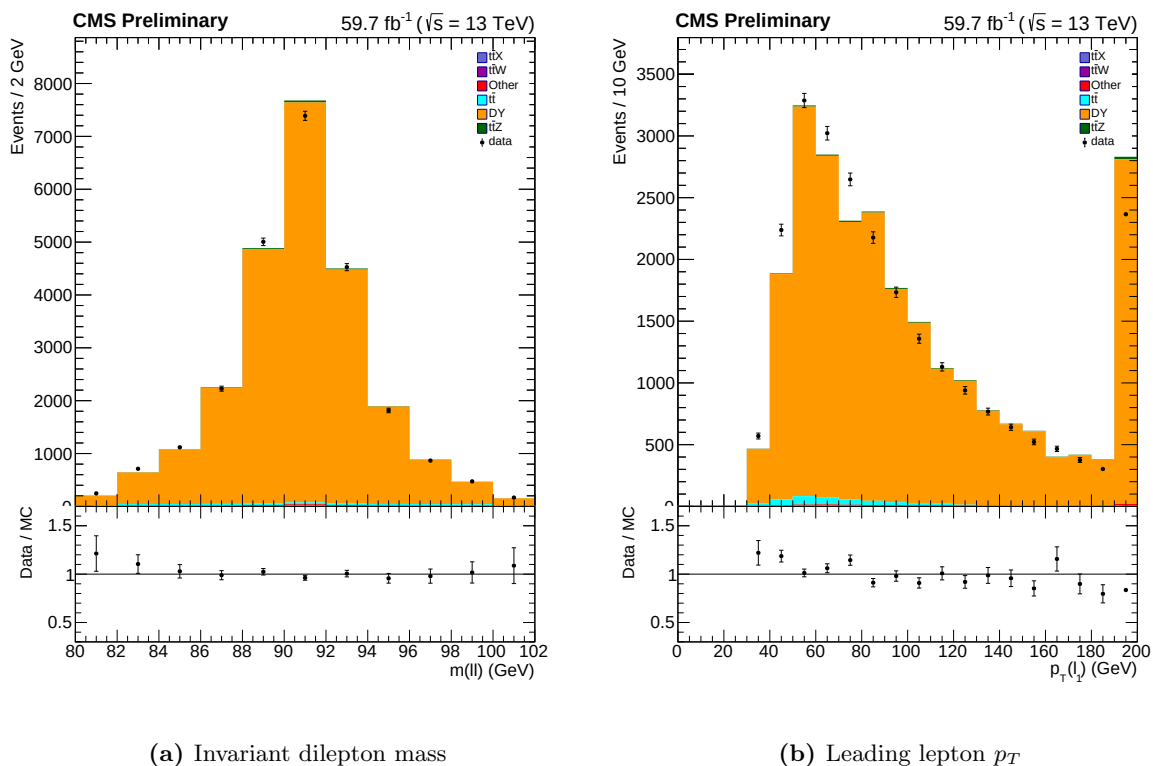


Figure 5.9: Reconstructed invariant dilepton mass and leading lepton transverse momentum with DY control region selections applied.

Recalling the characteristics of the DY background from section 5.2.1, we remember that in general, no b-tagged jets would be expected here. For this reason, the CR selection for DY will require exactly 0 b-tagged jets instead of at least one b-tagged jet in the SR event selections. In principle, no jets would be expected from the DY process on its own, but radiated gluons are likely to give a significant amount of jets, which is why we will not alter the number of jets

selection criterion. The zero b-jet requirement together with the OSSF leptons with a dilepton mass around the Z boson mass should give a CR which is very pure in DY backgrounds, rejecting almost all contributions from other backgrounds or signal. The dilepton mass plot for the DY CR can be seen in figure 5.9a. It is clear that this CR is indeed very pure in DY, rejecting all other sources almost entirely. The agreement between data and MC is not perfect, even though it is only off by a few percent. However, DY is the most dominant source of background and also the main source of events in the signal region, yielding just over 66% of the total amount of events passing the event selection criteria. Therefore, the error of a few percent is not negligible and will have to be accounted for when making the cross section measurement. The remaining DY CR plots can be found in appendix C.

5.9.2 $t\bar{t}$

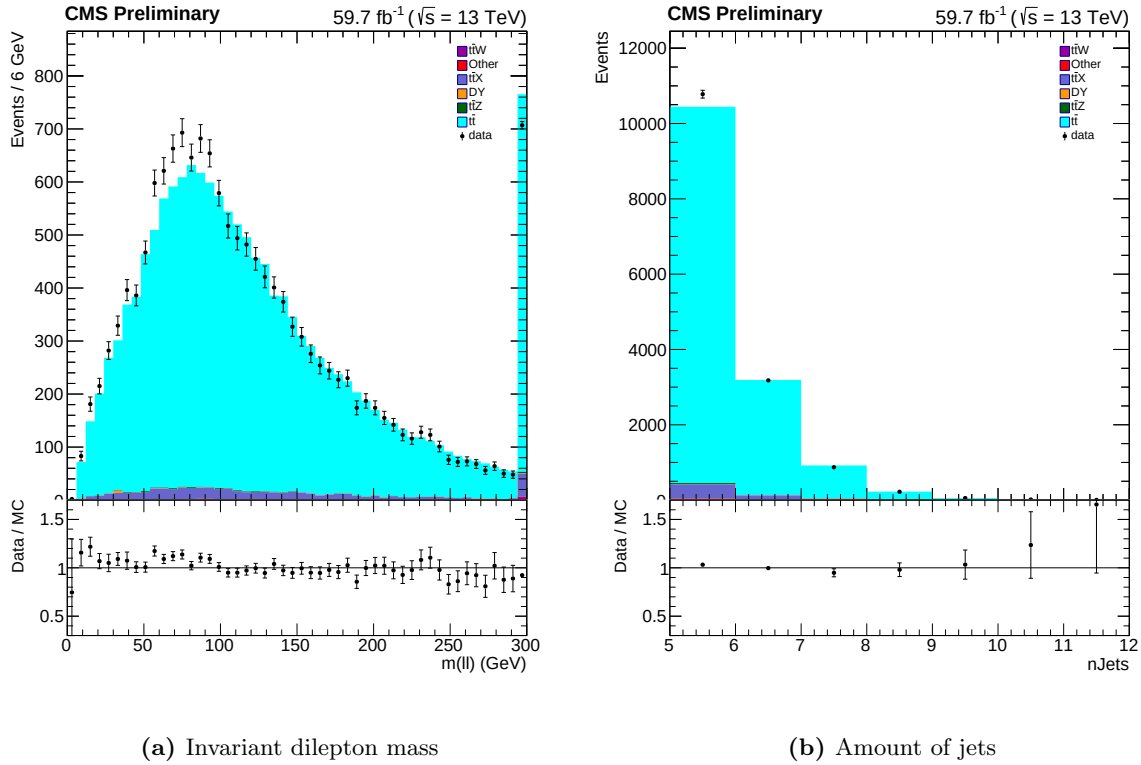


Figure 5.10: Reconstructed invariant dilepton mass and amount of jets with $t\bar{t}$ control region selections applied.

From the discussion of the $t\bar{t}$ background in section 5.2.2, we know that no Z boson is expected in this background source. We do however expect two b-jets along with two W bosons, both originating from the top quark decay. When both W bosons decay leptonically, a pair of leptons

with OS is created. While they can have the same flavour, contrary to the DY and $t\bar{t}Z$ case they are not obliged to be of SF. Therefore, we will require the leptons to be OF: this will get rid of a part of the $t\bar{t}$ background which we now want to study, but it will assuredly reject most of the DY and $t\bar{t}Z$ samples. As the two leptons are not originating from a Z boson in the $t\bar{t}$ case, we will also remove the requirement of having a dilepton mass close to the Z boson mass. We will however require the dilepton mass to be at least 15 GeV, as the modelling of dilepton pairs with a very low invariant mass is not accurate. The dilepton mass plot for the $t\bar{t}$ CR can be seen in figure 5.10a. Like in the DY CR, the agreement between data and MC is not perfect, but the overall error is only of the order of a few percent. This can be confirmed by figure 5.10b, which shows the amount of jets in the $t\bar{t}$ CR. However, like was the case for the DY process, we will take this error into consideration when calculating the cross section of the $t\bar{t}Z$ process. More $t\bar{t}$ CR plots can be found in appendix D.

5.10 Cutflow

In this section, we will have a look at the cutflow table. In such a table, the evolution of the amount of remaining events after applying a number of cuts is given. This can be used as a way to confirm that our choices for the event selection criteria were justified. The cutflow table is presented in table 5.7.

The object selections contain all lepton and jet quality selections mentioned in tables 5.4 and 5.5. If, at any point, less than two leptons which meet the lepton quality selections remain, the event is rejected. From the cutflow table, it is apparent that the requirement on the amount of jets reduces the $t\bar{t}$ and DY backgrounds considerably. Requiring at least one b-tagged jet dramatically reduces the DY background, while only rejecting a small fraction of signal events. The last two cuts, which require that the leptons are of the SF and have an invariant dilepton mass near the mass of the Z boson, leave the DY process mostly untouched, but greatly decrease the amount of $t\bar{t}$ background events. The other three background sources are also reduced to only a few percent of their original yields, justifying the event selections we proposed. In the end, a SBR of about 1/16 is reached, whereas we started with a SBR of 1/200.

Cut	Process	$t\bar{t}Z$	fraction	$t(\bar{t})X$	fraction	$t\bar{t}W$	fraction	Other	fraction	$t\bar{t}$	fraction	DY	fraction	SBR
Object selections	Object selections	1358.72	1	5383.99	1	610.24	1	1504.23	1	99418.43	1	165370.65	1	0.0050
	twoOS	1205.83	0.887	5122.84	0.951	322.49	0.528	1234.88	0.821	99211.82	0.998	165014.94	0.998	0.0045
	njets	803.75	0.592	1609.77	0.299	143.45	0.235	256.63	0.171	34539.52	0.347	33295.63	0.201	0.0115
	nbjets	647.43	0.476	1291.06	0.240	117.89	0.193	43.72	0.029	28986.88	0.292	5916.23	0.036	0.0178
	twoSF	603.07	0.444	742.21	0.138	61.35	0.101	34.09	0.023	14745.55	0.148	5908.04	0.036	0.0281
onZ	459.19	0.338	198.93	0.037	7.13	0.012	19.35	0.013	1923.58	0.019	5131.19	0.031	0.0631	

Table 5.7: The cutflow table. The cutflow of each group of backgrounds is represented, along with the fractions of events which remain after applying the event selection criteria, where the reference point is the amount of events left after the basic object selections. In the rightmost column, the signal-to-background ratio is visible.

5.11 Discriminating variables

In section 5.8 we saw that even after applying all event cuts, the background still severely outweighs the signal. In order to be able to better distinguish signal from background, ML techniques will be used. The implementation of these ML models will be covered in the next section. In this section, we will look into some variables which might help in distinguishing signal from background. We will not be making any extra cuts based on these variables, but we will give them as an input to the ML algorithms.

The invariant mass of the dilepton pair is quite an obvious choice for a discriminating variable, as not all background processes contain a Z boson, but the invariant dilepton mass for the $t\bar{t}Z$ signal will lie around the Z boson mass. The plot of the dilepton mass was already shown when discussing the SR, but for clarity and comparison to the other discriminant variables it is also shown in figure 5.11a. As discussed earlier, the dilepton mass will be an important discriminant to reduce the $t\bar{t}$ background, but it will not reduce the DY background due to the apparent Z boson in this background.

Secondly, the MET will also be an important variable. In $t\bar{t}Z$ and DY processes, no true MET is expected, but true MET is expected in the $t\bar{t}$ background due to the apparent neutrinos. This will manifest itself as a longer tail in the MET distribution for the $t\bar{t}$ process, which can be seen in figure 5.11b.

In section 5.2.1 we discussed that in general, no b-tagged jets are expected in the DY background. For this reason, we required all events to have at least one b-tagged jet, while we actually expect two b-tagged jets in the $t\bar{t}Z$ process. The amount of b-tagged jets could therefore be used to discriminate between the DY and $t\bar{t}Z$ process. This can be confirmed by figure 5.11c, in which the amount of b-tagged jets is shown. The first bin, holding events with only one b-tagged jet, is dominated by the DY background. We can also see that, if we would have required events to have at least two b-tagged jets, we would have rejected a lot of signal. Even though the background would have significantly decreased, the statistics for the signal might have become too low.

Additionally to the invariant dilepton pair mass, the reconstructed invariant mass of the W boson is also considered. The W mass will be reconstructed by searching for its decay products and then combining their four-vectors to get the invariant mass. In the topology considered in this analysis, the W boson will decay to two jets. The invariant mass of the W boson will be reconstructed as follows. First, the invariant mass of every jet-jet pair will be reconstructed.

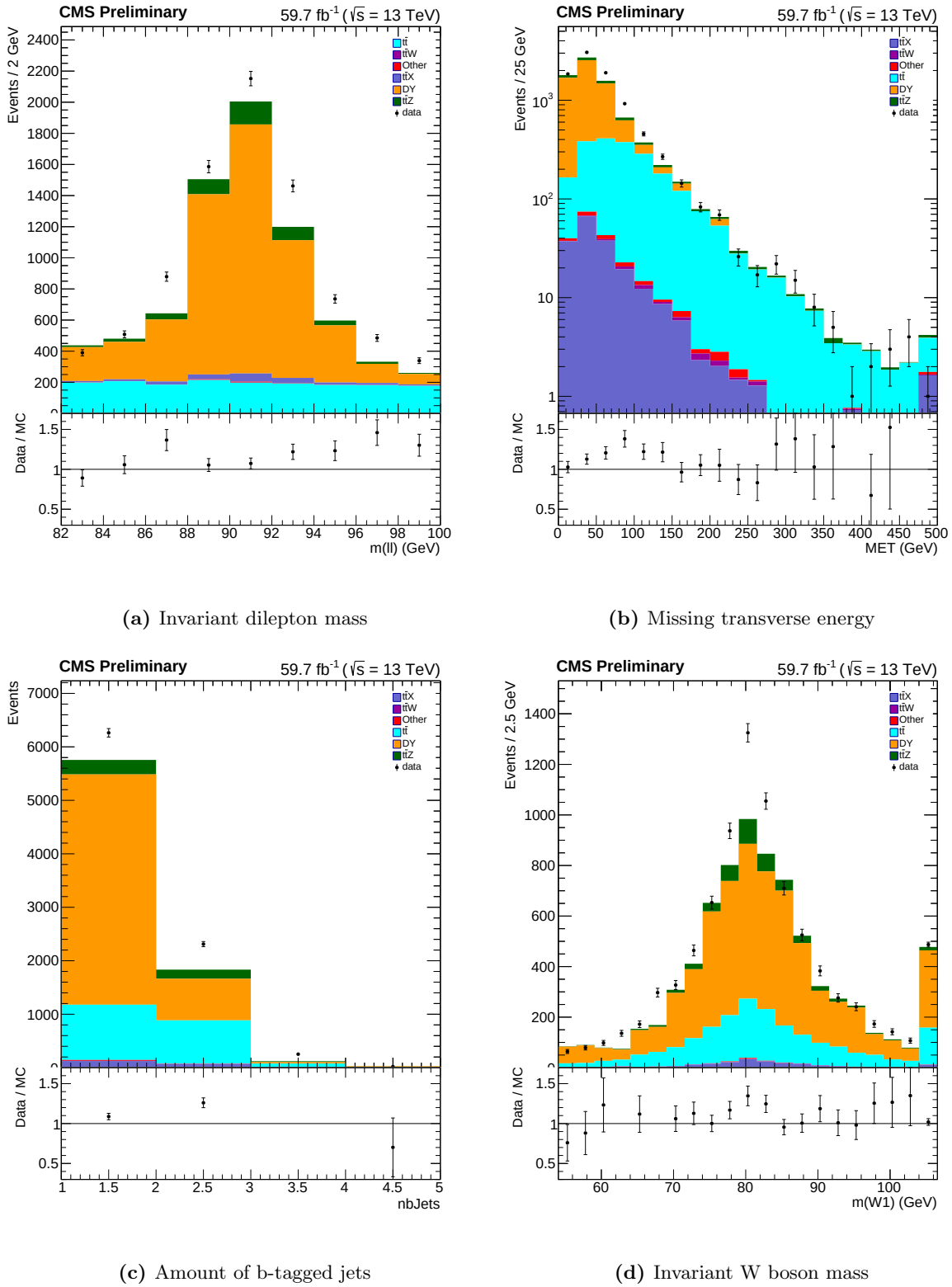


Figure 5.11: Plots of the discriminant variables used in this analysis: the invariant dilepton mass, missing transverse energy, amount of b-tagged jets and reconstructed W boson mass.

Note that we include b-tagged jets in this as well, even though we do not expect the jets originating from the W boson decay to be b-tagged. As we require each event to have at least five jets passing the selection criteria, we will reconstruct at least twenty W boson candidate masses per event. When all masses are reconstructed, a loop over all masses is conducted and the reconstructed W mass will be the one which lies the closest to the actual W mass of 80.379 GeV. A second reconstructed W mass is found in the same way, but excluding jets which were already used to reconstruct the first one. The reconstruction for this second W boson mass is quite a lot worse than the first one and thus will not be used as an input variable for the ML algorithms. The distribution of the best matching² W masses can be seen in figure 5.11d. It is clear that there is a peak around the W boson mass for the $t\bar{t}Z$ signal. However, in the DY process no W boson is apparent, while in the $t\bar{t}$ process, the W bosons must decay leptonically for $t\bar{t}$ to be considered a background. And still, there is an unambiguous peak around the W boson mass for both DY and $t\bar{t}$. This can be explained by the bias which is inherent to the reconstruction algorithm used here: as the reconstructed W mass is always the one which lies the closest to the actual W mass, a peak in the distribution for the backgrounds is expected.

Lastly, the reconstructed invariant top mass could also be an interesting discriminant variable. As the decay mode of the top quark is decided by the decay of the W boson, the justification for the interest in a well reconstructed top mass is the same as the one mentioned in the previous paragraph for the W boson. In this analysis, the top quark is reconstructed through a very simple algorithm, which combines the invariant W masses reconstructed using the algorithm described above with a b-tagged jet. One of the main problems of this approach is that we will not be able to reconstruct a top mass 100% off the time. The results of this reconstruction could certainly be greatly improved upon, which will be discussed in more detail in section 6.1. Nonetheless, the top mass will be given as an input to the ML algorithms. In the case that no top mass is reconstructed, a value of 1000 GeV will be assigned to it, in an attempt to let the ML algorithms know that this variable, having a value which would be considered as a very severe outlier, should not be used when predicting an output for this event.

5.12 Machine learning algorithms

The analysis presented in this thesis will make use of ML algorithms to enhance the signal-to-background ratio. More precisely, a BDT and a NN will be trained with the goal of being able

²For each event, the reconstructed W mass that lies the closest to the actual W mass is plotted.

to predict whether a given event is more likely to be signal or background. In the ideal scenario, the distribution of the model output of the signal will be very different from the distribution of the background, yielding a clear way to discriminate between background and signal. The goal of this thesis is to calculate a cross section estimate for the $t\bar{t}Z$ process and we will calculate this by fitting the data to MC using the model output distribution. This will be discussed in section 5.13.

The model will be trained on a subset of the MC data available. In order to prevent introducing an unphysical bias in the fitting process, the MC data will be split in two equal parts: one part will be used for training and optimizing the ML algorithms, while the other part will be used for fitting. As discussed in chapter 4, the part of the dataset used for training will be split into a training and validation set. In this analysis, 80% of the training data, which equates to 40% of the total data, will be used as the training set. The remaining 20% will be used as the validation set. It should be noted that in this case, the 50% of the total dataset being used for fitting can be considered the test set, meaning the other 50% can indeed be used entirely for training and hyperparameter optimization.

In this section, a list of all used input variables will be presented. Additionally, the optimization, performance and results of the BDT and NN will be discussed.

5.12.1 Input variables

Along with the discriminant variables described in section 5.11, some other variables are given as an input to the BDT and NN. While it is expected that the discriminant variables will yield the largest contribution towards enhancing the signal, the ML algorithms could gain a better understanding of the considered event by supplying more relevant input variables. For this purpose, all variables regarding the lepton and jet kinematics, p_T , η , and ϕ , are provided to the MVA. They are provided for both the leading and sub-leading lepton, but only for the five jets with the highest p_T , as we require events to have at least five jets and the input to the ML algorithms used in this analysis always has to have the same shape. The isolation of the leptons is also considered through the variable I_{rel} . The closest normal jet and b-tagged jet to both the leading and sub-leading lepton is provided through the angular separations $\Delta R_{j/b,l1/2}$, adding four more input variables. Additionally to the amount of b-tagged jets, which was already discussed in the previous section, the total amount of jets is also used in the training of the models. The output of the b-tagging algorithm for the leading (in most cases the only) b-tagged

jet is provided to include the degree of confidence with which we state that this b-tagged jet is in fact a b-jet. Lastly, the H_T^{miss} is used, which is defined as the amount of missing energy when only considering the jets which pass all object and event selection cuts. In order to calculate the H_T^{miss} , one takes the four-vector sum over all momenta and the H_T^{miss} will then be the momentum of the resulting four-vector. All input values are summarised in table 5.8.

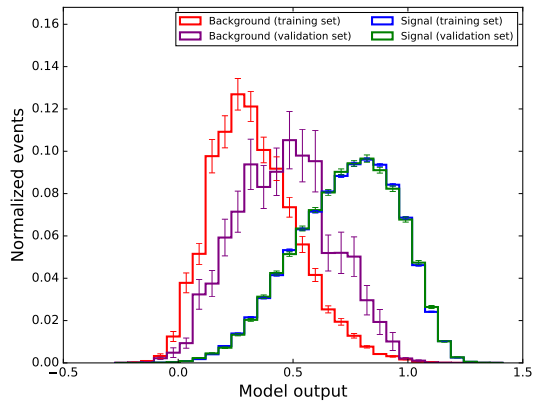
Category	Variables
Leptons	p_T, η, ϕ, I_{rel}
Jets	$p_T, \eta, \phi, \Delta R_{j,l1/2}, \Delta R_{b,l1/2}$
Event variables	nJets, nbJets
Invariant masses	m_{ll}, m_W, m_{top}
Other	MET, H_T^{miss} , jetDeepCsv(b1)

Table 5.8: All input variables given to the BDT and NN. The lepton variables are given for both the leading and sub-leading lepton, while jet variables are provided for the five jets with highest p_T . An exception is made for $\Delta R_{l1/2}$ and $\Delta R(b)_{l1/2}$, as only the $\Delta R_{j/b}$ of the (b)jet which minimizes $\Delta R_{j/b}$ is used, both for the leading and sub-leading lepton.

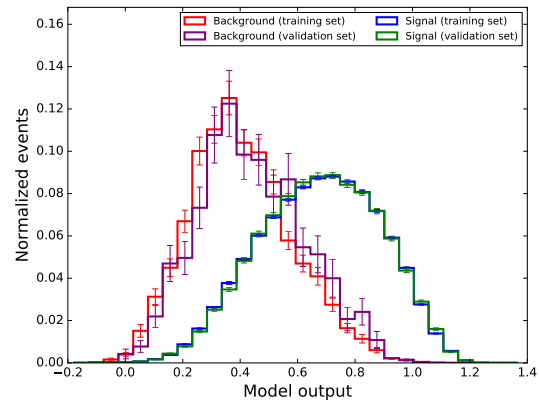
5.12.2 Optimization of hyperparameters

In chapter 4, genetic algorithms for hyperparameter optimization were discussed. For both the BDT and NN, a genetic algorithm was used, which tried to maximize the AUC as defined in chapter 4. However, the genetic algorithm tended to overfit the BDTs very easily. For this reason, a genetic algorithm was used to narrow down the hyperparameter space, after which a grid search (GS) was conducted. In a GS, a multidimensional grid of hyperparameters is made. A model will be trained using each unique combination of hyperparameters on this grid, after which the models are ranked by performance, with the performance metric of choice being the AUC. We also ranked the models based on a metric which quantified the “degree of overfitting” of the considered model. This metric was calculated by creating histograms such as the ones in figure 5.12, which show the output shapes for background and signal, both for the training and validation set. By calculating the sum of the differences in bin content for all bins, a quantitative metric for the overfitting of a model was defined. The final set of hyperparameters was chosen by combining the AUC value with this degree of overfitting in order to get a model with a high performance and good generalization to previously unseen samples. An example is given in

figure 5.12. While the BDT in figure 5.12a has a similar AUC to the BDT in figure 5.12b, the former is severely overfitted. It will not perform well on unseen data, which is why the BDT in figure 5.12b is the better choice.



(a) Overfitting parameter: 0.538



(b) Overfitting parameter: 0.181

Figure 5.12: A comparison between a severely overfitted BDT and a BDT with a lot less overfitting. For both BDTs, their “overfitting parameter” is given, for which lower means better.

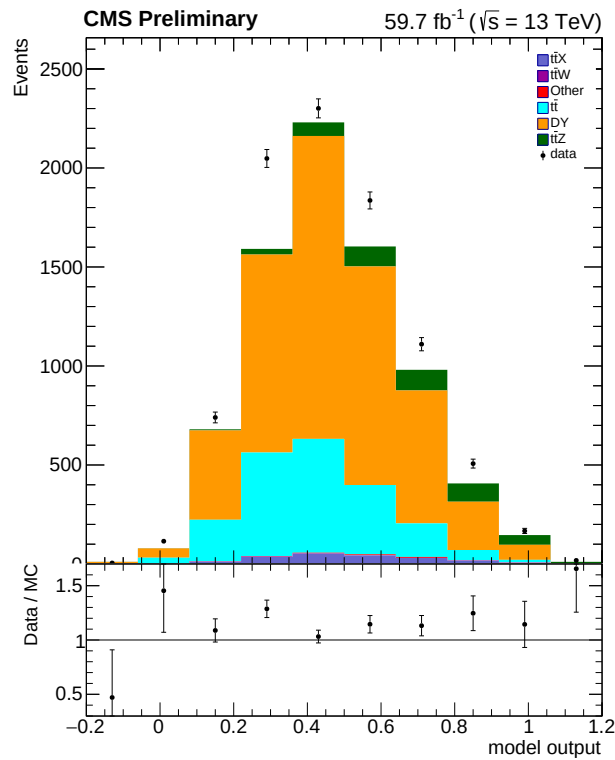


Figure 5.13: The output shape of the boosted decision tree.

5.12.3 Boosted decision tree

The final set of hyperparameters used to train the BDT can be found in appendix E. This set of hyperparameters gives rise to the BDT output shape which can be seen in figure 5.13. It is immediately clear that the shapes of the dominant backgrounds, DY and $t\bar{t}$, are different from the signal BDT output shape: the peak for the DY and $t\bar{t}$ backgrounds occurs at lower BDT output values than the peak for the signal distribution. This is exactly the purpose of using ML techniques, as the goal was to find some way to further distinguish background from signal. The data follows roughly the same shape of the total MC, but evidently, the MC somewhat underestimates the data as was already observed in the SR plots.

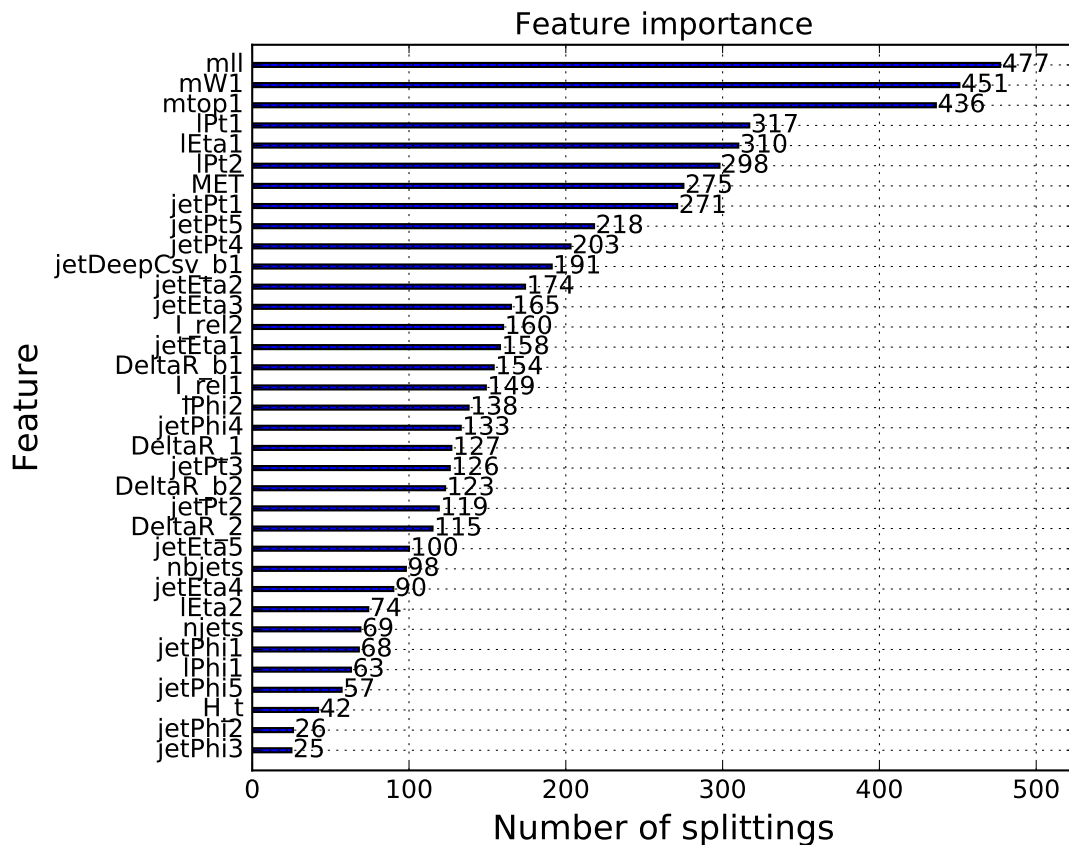


Figure 5.14: The variables that were used as an input to the BDT and NN, ranked by importance. This feature importance plot was obtained from training the BDT.

BDTs, but also normal DTs, can be used as a way to find the most important features, which are the event variables in the case of high energy particle physics, used to generate a prediction. By counting the amount of times a feature is used to split a node, we can rank the features by

importance. This can be explained as follows: as the training algorithm decides which feature maximally decreases the impurity for each node, the feature that is used most will be the most informative or decisive feature. All input variables are ranked by importance in figure 5.14. In this figure, we can see that nearly all of the discriminating variables discussed in section 5.11 are among the most important input variables. It is apparent that reconstructing invariant masses of the particles which are expected to be present in a $t\bar{t}Z$ event is important and assists the BDT in its decision making significantly.

5.12.4 Deep neural network

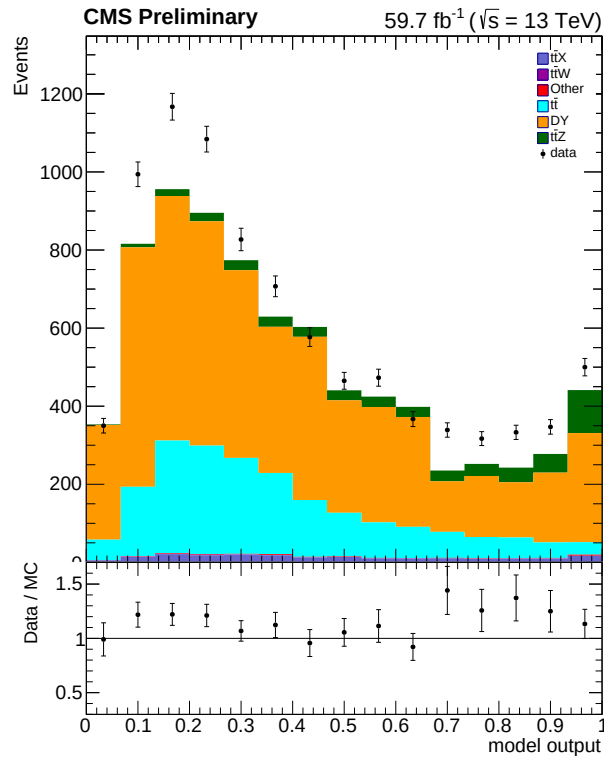


Figure 5.15: The output shape of the neural network.

Figure 5.15 shows the output shape for the NN. This NN model was achieved by using the set of hyperparameters listed in appendix E. It is immediately apparent that the output shapes of the BDT and NN are very different, but we did not expect them to be similar. The signal peaks at maximal output values, while the peak for the dominant background sources lies close to the minimal output values. The NN certainly enables us to distinguish background from signal better compared to using purely event selections, but even in the high NN output region,

quite a lot of background remains. This makes sense, because accurately identifying the $t\bar{t}Z$ signal is like finding a needle in a haystack due to the immense backgrounds present in the studied signal topology. In section 5.10 we saw that the event selections already decrease the SBR dramatically, which leaves the challenging part of further distinguishing the signal from the background to the ML algorithms. As was also the case for the BDT, the data roughly follows the shape of the MC.

5.13 Cross section calculation using the *Higgs combine tool*

The final section of this chapter will deal with the cross section calculation. Firstly, the combine tool is briefly discussed, after which a list of all employed uncertainties will be given. Finally, the cross section will be calculated and compared with state-of-the-art results.

5.13.1 The combine tool

The combine tool provides a command line interface to many statistical analysis techniques [91, 92]. It is created by the Higgs Physics Analysis Group and is widely used within CMS. In this analysis, we will be using combine to extract signal strengths by performing maximum likelihood fits. The extracted signal strength μ enables us to calculate the observed cross section for the process which is being studied, in this case the $t\bar{t}Z$ process. The measured cross section can be calculated through formula 5.3, where σ_{MC} is the theoretical estimate used to generate the MC samples.

$$\sigma = \mu \cdot \sigma_{MC} \tag{5.3}$$

The input to combine is a datacard, which contains the histograms used for the fit along with the different sources of systematic uncertainty. The histograms used in this analysis are the outputs of the BDT and the NN. Combine will then perform the fit using a certain statistical method chosen by the user, taking into account the statistical and systematic uncertainties. The different systematic uncertainties will be described in the next section. As already mentioned, this analysis makes use of maximum likelihood fits. In practice, combine will maximize the likelihood by finding the scaling of the signal and background that describes the data best.

5.13.2 Systematic uncertainties

Systematic uncertainties can be divided into two categories: rate and shape uncertainties. The former expresses the systematic deviation of a source by a percentage which is the same for all bins. The latter is more complex. To implement a shape uncertainty, for example the effect of pileup interactions, the relevant parameters for that source of uncertainty are shifted up and down by one standard deviation. This gives rise to two additional histograms which have to be given to the combine program. Combine will then compare the nominal shape to the shape obtained from varying relevant parameters one standard deviation up or down. This way, a bin-per-bin systematic error can be applied. More information on how the combine tool incorporates shape uncertainties can be found in Ref. [93]. The following sections list all rate and shape uncertainties used in this analysis, along with a brief explanation why these corrections are needed.

5.13.2.1 Rate uncertainties

All MC samples used to model the signal and background processes are normalised to data using the luminosity of the corresponding data taking period. Therefore, the uncertainty on the luminosity has to be included in the measurement. In this analysis, 2018 data was used, for which the systematic uncertainty on the luminosity was 2.5% [94].

Much like the uncertainty on the luminosity, there is also an uncertainty on the trigger efficiency. A systematic uncertainty on the trigger efficiency of 2% will be used in this analysis, based on the value used in Ref. [70].

Apart from the uncertainties on the luminosity and trigger efficiency, five other rate uncertainties will be used. These five sources of uncertainty are an estimation of the uncertainty on the cross sections of all background processes in this analysis. In the CR plots for the DY and $t\bar{t}$ process, we observed a deviation of a few percent between data and MC. In order to include this uncertainty in the $t\bar{t}Z$ cross section measurement, a conservative systematic uncertainty of 10% will be applied to both these processes. In a similar manner, an uncertainty on the cross section of the $t\bar{t}X$, $t\bar{t}W$ and other background processes of respectively 11%, 11% and 50% is used. The numeric values of these last three uncertainties are based on those in Ref. [70]. All systematic rate uncertainties are summarised in table 5.9.

Uncertainty source	Affected processes	Rate
Luminosity	all	2.5%
Trigger efficiency	all	2%
$t\bar{t}X$ cross section	$t\bar{t}X$	11%
$t\bar{t}W$ cross section	$t\bar{t}W$	11%
other processes cross section	other	50%
$t\bar{t}$ cross section	$t\bar{t}$	10%
DY cross section	DY	10%

Table 5.9: All included rate systematic uncertainties.

5.13.2.2 Shape uncertainties

We will now briefly discuss all sources of shape uncertainties, as an in-depth discussion of these uncertainties lies out of the scope of this thesis. Each source will give rise to two additional input shapes (input histograms): one will correspond to a shift up of the relevant quantity by one standard deviation, while the other will correspond to a shift down of the relevant quantity by one standard deviation. We include four sources of shape uncertainty in total, meaning that we will give nine³ input histograms to the combine tool. Plots visualizing the effect of the shape uncertainties on the model output distributions can be found in appendix F.

Jet energy scale corrections The jet energy scale corrections (JEC) map the measured jet energy deposition to the particle-level jet energy [95]. To estimate the systematic uncertainties due to the JEC, these correction factors are shifted up and down by their uncertainties. More details on JEC can be found in Refs. [95, 96].

Pileup The pileup (PU) profile of the MC samples has to match the one seen in data. However, the agreement is often not perfect. For this reason, the MC primary vertex (PV) distribution has to be reweighted, which is a technique called *pileup-reweighting* [97]. The PV distribution for data is supplied by the Physics Validation Group for each data taking period. From this, up and down corrections are calculated, which are subsequently used to calculate the PU reweighting factors. It should be noted that this procedure not only provides weights for the up and down corrections, but also reweights the nominal case.

³Four shape uncertainties with two histograms each, plus the nominal histogram.

b-tagging efficiency The performance of the b-tagging algorithm differs between data and MC and this effect has to be corrected for. This is done by reweighting the events using the b-tag scale factors provided by CMS BTAG group [98]. It is important to note that the method described below only changes the event weight and not the amount of b-tagged jets contained within this event. The b-tag reweighting factor is calculated by dividing the probability of a given configuration of jets in data by the probability of this same configuration in MC [98]. These probabilities are defined as

$$P(\text{MC}) = \prod_{i = \text{tagged}} \epsilon_i \prod_{j = \text{not tagged}} (1 - \epsilon_j) \quad (5.4)$$

$$P(\text{DATA}) = \prod_{i = \text{tagged}} \text{SF}_i \epsilon_i \prod_{j = \text{not tagged}} (1 - \text{SF}_j \epsilon_j) \quad (5.5)$$

where ϵ_i is the MC b-tagging efficiency and both SF_i and ϵ_i are functions of the jet flavor, jet p_T and jet η . Using these probabilities, the b-tag reweighting factor w is given by

$$w = \frac{P(\text{DATA})}{P(\text{MC})} \quad (5.6)$$

The b-tagging scale factors have associated uncertainties, which are used to calculate an up and down deviation of the b-tag reweighting factor. Similar to the PU reweighting procedure, both the up, down and nominal cases are reweighted, and it are the up and down uncertainties which give rise to the shape uncertainties.

QCD scale The QCD scale uncertainty tries to accommodate for the uncertainty caused by the factorization and renormalization scales in QCD. These two scales are arbitrary parameters, of which the values are chosen to give a good description of observations. As they are arbitrarily chosen, the uncertainty is estimated by varying these scales by a factor of two up and down. An in-depth discussion of QCD scales is out of the scope of this thesis, a more detailed discussion can be found in Ref. [99].

5.13.3 Cross section calculation

Using the combine tool with the input histograms described in the previous sections, along with all considered sources of systematic uncertainty, we can now calculate the cross section of the $t\bar{t}Z$ process. We will get two separate results, one for the BDT and one for the NN. The cross section used to generate the MC samples is $\sigma_{MC} = 0.7974$ pb. Using equation 5.3 we get:

$$\sigma = \mu \cdot \sigma_{MC} = \mu \cdot 0.7974 \text{ pb} \quad (5.7)$$

From equation 5.7 we can instantly see that a signal strength of zero corresponds to the background-only hypothesis, which would predict that the $t\bar{t}Z$ process does not occur, while a signal strength of one corresponds to the SM hypothesis. The results for the signal strengths are presented in table 5.10 along with their respective cross section measurements. Also given are the uncertainties which provide the 68% confidence level (CL) interval. Figures 5.16 and 5.17 show the impact plots for the BDT and NN respectively. These plots show the impact of the nuisance parameters (NP) θ , for instance the uncertainty on the luminosity, on the parameter of interest $\hat{r}$, which represents the signal strength μ in this analysis. The impact of a NP on $\hat{r}$ is defined as the shift $\Delta\hat{r}$ that is induced as θ is fixed and brought to its $+1\sigma$ or -1σ post-fit values⁴, with all other parameters profiled as normal [100]. In an ideal scenario, one has so much data collected that you do not have to worry about statistical uncertainties on the data. In this case, the total uncertainty on the signal strength will be dominated by sources of systematic uncertainties. The systematic uncertainties can be split in two groups: those discussed in section 5.13.2 and the statistical uncertainties on the MC samples. Ideally, the statistical uncertainty on the MC samples would be very small, such that the total systematic uncertainty is dominated by the uncertainties described in section 5.13.2. However, as can be seen in figures 5.16 and 5.17, this is not the case: the uncertainty on $\hat{r}$ is dominated by statistical uncertainties on the bin contents of the ML model output histograms. It will be important in future analyses to reduce all statistical uncertainties on the MC as much as possible in order to achieve a precise cross section measurement. The combine tool also provides us with the post-fit yields for each signal or background source separately. A scale factor for each source can be calculated by dividing the post-fit yields by the pre-fit yields. Post-fit plots can be achieved by multiplying the pre-fit yields in each bin by these scale factors. A comparison of the pre- and post-fit yields for both the BDT and NN can be found in figures 5.18 and 5.19.

⁴The σ used here corresponds to one standard deviation and not the production cross section.

	signal strength	cross section (pb)
BDT	$1.609^{+0.134}_{-0.131}(\text{stat})^{+0.339}_{-0.303}(\text{syst}) = 1.609^{+0.364}_{-0.330}$	$1.283^{+0.107}_{-0.105}(\text{stat})^{+0.271}_{-0.242}(\text{syst}) = 1.283^{+0.291}_{-0.264}$
NN	$1.436^{+0.136}_{-0.138}(\text{stat})^{+0.464}_{-0.467}(\text{syst}) = 1.436^{+0.484}_{-0.487}$	$1.145^{+0.109}_{-0.110}(\text{stat})^{+0.370}_{-0.373}(\text{syst}) = 1.145^{+0.386}_{-0.389}$

Table 5.10: The signal strength measurements, together with the cross section calculated using this signal strength. For both quantities, 68% CL errors are provided.

In the SM, a $t\bar{t}Z$ production cross section of $\sigma_{SM} = 0.84 \pm 0.10$ pb is predicted [101, 102, 103]. The cross section measurement obtained using the NN agrees with the SM within uncertainties, while the BDT measurement deviates from the SM prediction by 1.68 standard deviations.

The results in table 5.10 can be directly compared with the most recent CMS result, which measures a cross section $\sigma = 0.95 \pm 0.05(\text{stat}) \pm 0.06(\text{syst})$ pb by studying the three and four lepton channels [70]. The cross sections found in this analysis are compatible with this CMS result, however, the precision on the CMS result is significantly better than the precision obtained on the results presented in this thesis. This can partially be explained by the fact that a larger integrated luminosity was used in the most recent CMS result, but the systematic uncertainties were also significantly lower.

We can also compare our results to a measurement conducted by the ATLAS collaboration. They report a signal strength $\mu = 0.73 \pm 0.28$ in the OS dilepton channel, giving rise to a cross section $\sigma = 0.642 \pm 0.247$ pb [104]. The results from this thesis are not compatible with the ATLAS results. It can be observed that in both the analysis presented in this thesis and the measurement conducted by ATLAS, the relative errors of the result are rather large. The ATLAS measurement uncertainty is dominated by the b-tagging efficiency and the modelling of background and signal samples. The latter is also the dominant source of uncertainty in the analysis presented in this thesis. It is important to note that the uncertainties of both the ATLAS dilepton analysis and the analysis presented in this thesis are far bigger than those of the state of the art theory predictions.

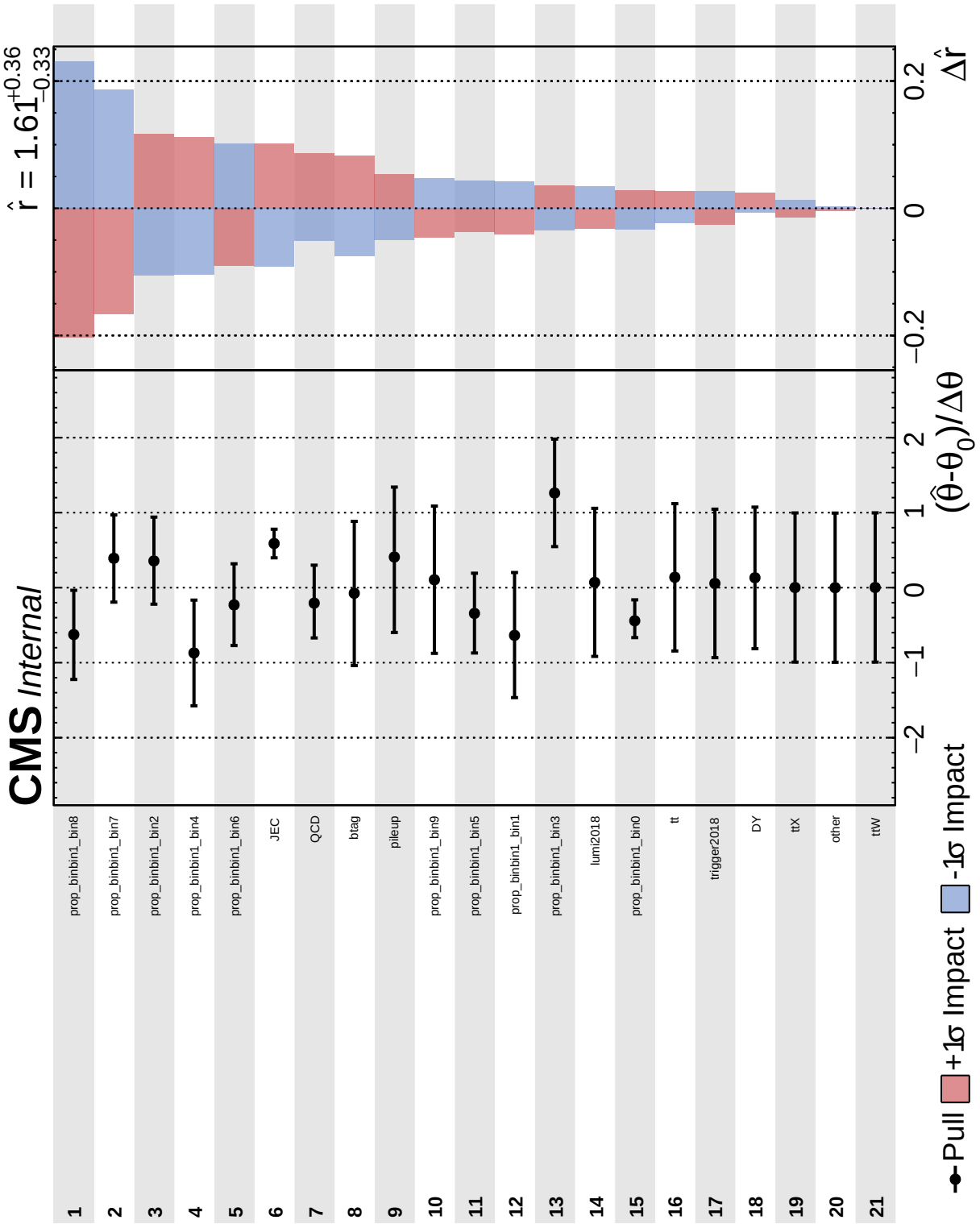


Figure 5.16: The impact plot for the BDT.

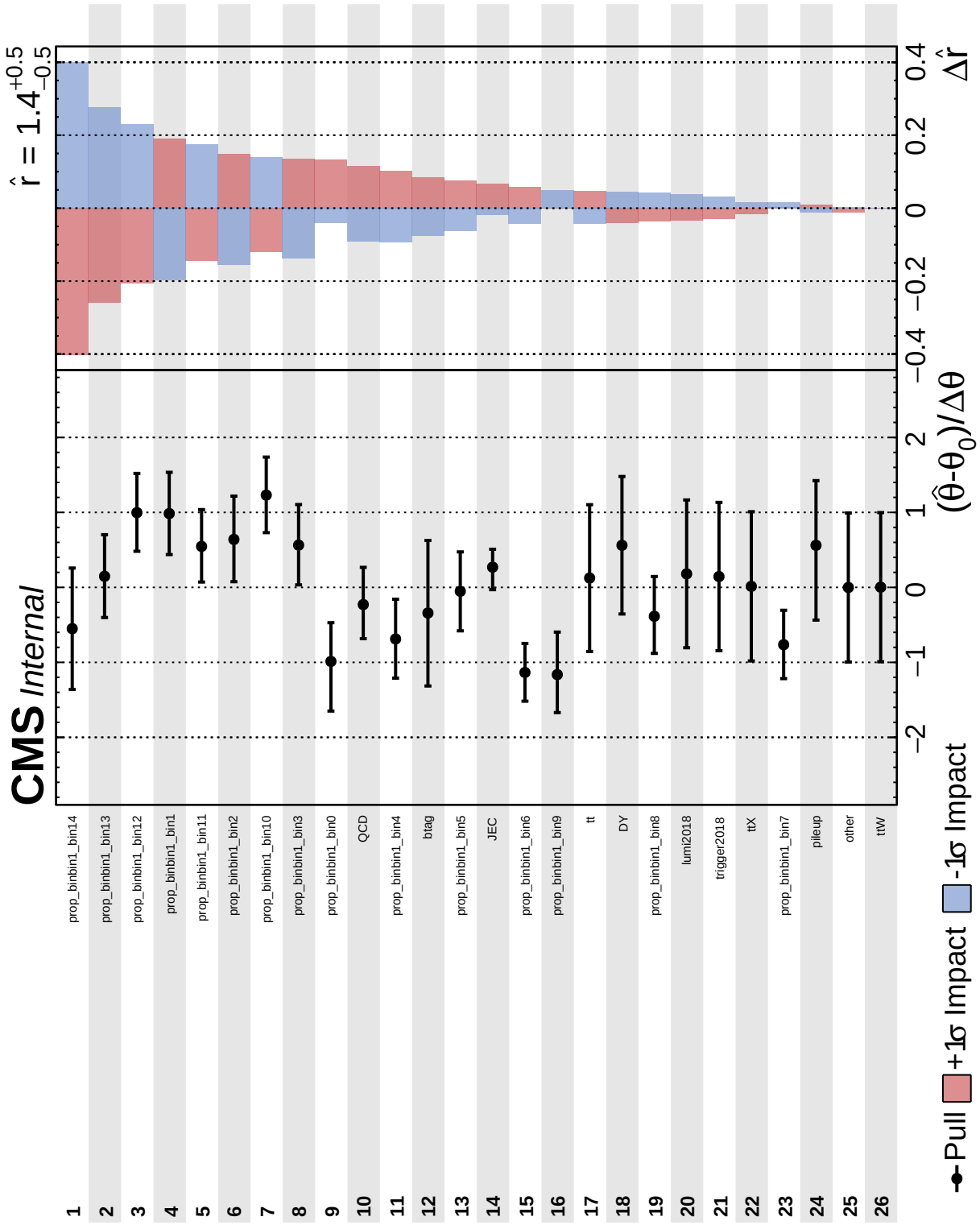


Figure 5.17: The impact plot for the NN.

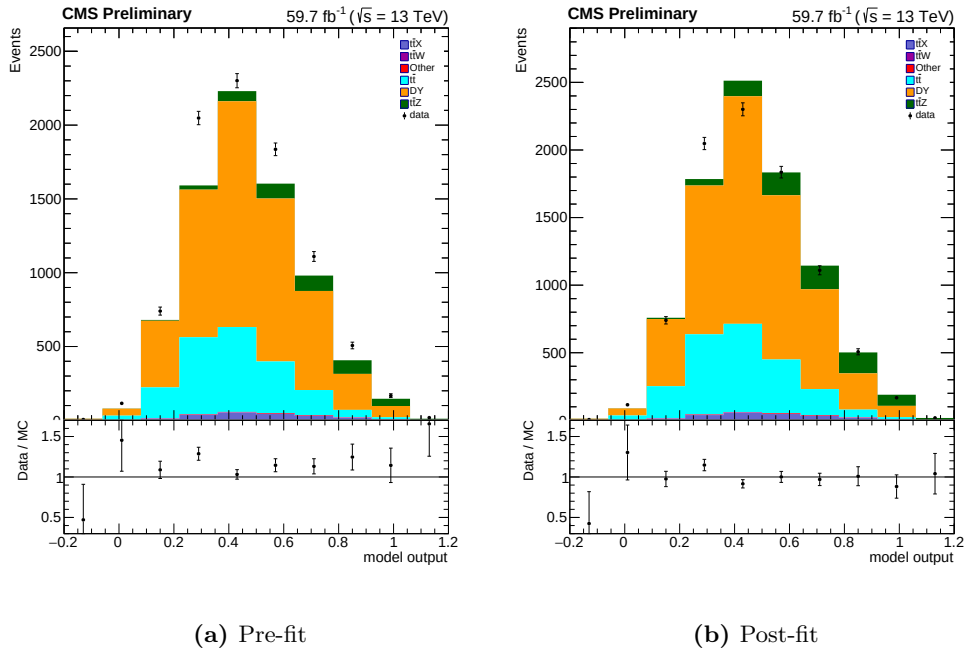


Figure 5.18: Pre- and post-fit plots of the BDT output showing the observed (points) and predicted (histograms) yields.

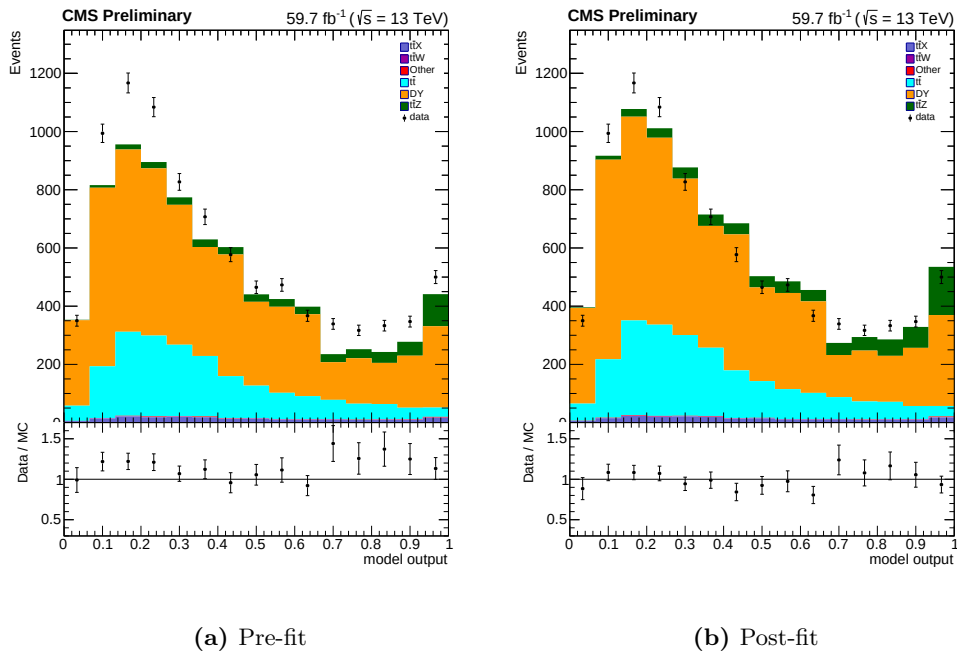


Figure 5.19: Pre- and post-fit plots of the NN output showing the observed (points) and predicted (histograms) yields.

Chapter 6

Conclusions and outlook

In this thesis, a measurement of the production cross section of a top quark pair in association with a Z boson has been presented. The analysis used data from proton-proton collisions at a COM energy of 13 TeV collected by the CMS detector at the LHC in 2018, corresponding to an integrated luminosity of 59.7 fb^{-1} . In the analysis presented in this thesis, the two-lepton final state was studied, in which the leptons are of opposite sign and same flavour. As the event selection was not sufficient to yield a good distinction between signal and background, two multivariate techniques were used, namely a neural network and a boosted decision tree. The usage of a BDT and NN resulted in measurements of the production cross section of $\sigma_{BDT} = 1.283^{+0.107}_{-0.105}(\text{stat})^{+0.271}_{-0.242}(\text{syst}) \text{ pb} = 1.283^{+0.291}_{-0.264} \text{ pb}$ and $\sigma_{NN} = 1.145^{+0.109}_{-0.110}(\text{stat})^{+0.370}_{-0.373}(\text{syst}) \text{ pb} = 1.145^{+0.386}_{-0.389} \text{ pb}$ respectively. Both of these results are compatible with the most recent cross section measurement by the CMS collaboration, which was performed in the three- and four-lepton final states [70, 105]. However, the precision of the CMS result is significantly better than those achieved in this thesis. The cross section measurement obtained using the NN is in agreement with the standard model prediction $\sigma_{SM} = 0.84 \pm 0.10 \text{ pb}$, while the measurement using the BDT deviates from the SM prediction by 1.68 standard deviations.

6.1 Future prospects

From the results achieved in this analysis, we can be optimistic that a measurement of the production cross section of the $t\bar{t}Z$ process in the two-lepton final state with an uncertainty of comparable order to the theory predictions is possible. However, such an analysis will have to incorporate various improvements with respect to the analysis presented in this thesis. An

obvious next step would be to make use of the full Run 2 dataset instead of exclusively using 2018 data. A larger amount of data gives rise to better statistics and should improve the performance of the fit. This could be beneficial for both the statistical uncertainty and the part of the systematic uncertainty caused by statistical uncertainties on the Monte-Carlo samples, which dominated the uncertainty in this analysis.

As mentioned in section 5.11, this analysis would benefit from a better algorithm to reconstruct the invariant mass of the top quark. This would mainly affect the performance of the BDT, as the NN has more tuneable parameters which might be able to reconstruct the top quark mass by itself.

From the signal region plots in section 5.8, it became apparent that a better modelling of the background and signal samples is desired. A possible way of improving this modelling could be to infer correction factors by studying the discrepancies between data and MC in well-defined control regions.

With the refined inclusive cross section measurement conducted, the next step could be an attempt to measure the differential cross sections. Such differential cross sections yield a very good test of the validity of the assumed theory. In order to be able to measure the differential cross section, the top quark kinematics might have to be considered in more detail than in this thesis, as possible differential cross section measurements could be a function of the top quark momentum or the mass of the top quark pair system. Differential cross section measurements have already been conducted by the CMS collaboration, which had a big impact on the community of theoretical particle physicists [70, 105, 106].

Appendices

Appendix A

Used MC samples

Listed below is an overview of all used MC samples in the analysis presented in this thesis. They are grouped into six categories: the signal, $t\bar{t}Z$, and the five backgrounds, $t\bar{t}X$, $t\bar{t}W$, $t\bar{t}$, DY and “other”.

Process	MC sample	cross section (pb)
$t\bar{t}Z \rightarrow ll\nu\nu$	TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8_MiniAOD2018	0.2728
$t\bar{t}Z \rightarrow ll$	TTZToLL_M-1to10_TuneCP5_13TeV-amcatnlo-pythia8_MiniAOD2018	0.0493

Table A.1: List of all used $t\bar{t}Z$ datasets along with their respective cross sections.

Process	MC sample	cross section (pb)
$WZ \rightarrow lll\nu$	WZTo3LNU_TuneCP5_13TeV-amcatnloFXFX-pythia8_MiniAOD2018	4.666
$ZZ \rightarrow llll$	ZZTo4L_TuneCP5_13TeV_powheg-pythia8_MiniAOD2018	1.3816
$gg \rightarrow g \rightarrow ZZ \rightarrow eeee$	GluGluToContinToZZTo4e_13TeV_MCFM701_pythia8_MiniAOD2018	0.0027
$gg \rightarrow g \rightarrow ZZ \rightarrow \mu\mu\mu\mu$	GluGluToContinToZZTo4mu_13TeV_MCFM701_pythia8_MiniAOD2018	0.0027
$gg \rightarrow g \rightarrow ZZ \rightarrow \tau\tau\tau\tau$	GluGluToContinToZZTo4tau_13TeV_MCFM701_pythia8_MiniAOD2018	0.0027

Table A.2 – continued from previous page

Process	MC sample	cross section (pb)
$gg \rightarrow H \rightarrow ZZ \rightarrow ll$	GluGluHToZZTo4L_M125_13TeV_ powheg2_JHUGenV7011_pythia8_ MiniAOD2018	0.01212
$H \rightarrow ZZ \rightarrow llll$	VBF_HToZZTo4L_M125_13TeV_ powheg2_JHUGenV7011_pythia8_ MiniAOD2018	0.001034
$W^- H, H \rightarrow ZZ \rightarrow llll$	WminusH_HToZZTo4L_M125_13TeV_ powheg2-minlo-HWJ_JHUGenV7011_ pythia8_MiniAOD2018	0.0002339
$W^+ H, H \rightarrow ZZ \rightarrow llll$	WplusH_HToZZTo4L_M125_13TeV_ powheg2-minlo-HWJ_JHUGenV7011_ pythia8_MiniAOD2018	0.0001471
$ZH \rightarrow H \rightarrow ZZ$	ZH_HToZZ_4LFilter_M125_13TeV_ powheg2-minlo-HZJ_JHUGenV7011_ pythia8_MiniAOD2018	0.000652
$WW \rightarrow ll\nu\nu$	WWTo2L2Nu_NNPdf31_TuneCP5_ 13TeV-powheg-pythia8_MiniAOD2018	10.481
$WW \rightarrow ll\nu\nu(\text{DS})$	WWTo2L2Nu_DoubleScattering_13TeV- pythia8_MiniAOD2018	0.1729
WWW	WWW_4F_TuneCP5_13TeV-amcatnlo- pythia8_MiniAOD2018	0.2086
WWZ	WWZ_TuneCP5_13TeV-amcatnlo- pythia8_MiniAOD2018	0.1651
WZZ	WZZ_TuneCP5_13TeV-amcatnlo- pythia8_MiniAOD2018	0.05565
ZZZ	ZZZ_TuneCP5_13TeV-amcatnlo-pythia8_ MiniAOD2018	0.01398
$W + \text{jets} \rightarrow l\nu$	WJetsToLNu_TuneCP5_13TeV- madgraphMLM-pythia8_MiniAOD2018	61526.7

Table A.2: All “other” used datasets along with their respective cross sections.

Process	MC sample	cross section (pb)
$t\bar{t}H$ (no $b\bar{b}$)	ttHTToNonbb_M125_TuneCP5_13TeV-powheg-pythia8_MiniAOD2018	0.2151
tZq	tZq_ll_4f_ckm_NLO_TuneCP5_13TeV-madgraph-pythia8_MiniAOD2018	0.09418
$t\bar{t}t\bar{t}$	TTTT_TuneCP5_13TeV-amcatnlo-pythia8_MiniAOD2018	0.009103
Single t	ST_tWll_5f_LO_TuneCP5_PSweights_13TeV-madgraph-pythia8_MiniAOD2018	0.01123
tHq	THQ_4f_Hincl_13TeV_madgraph-pythia8_MiniAOD2018	0.261
tHW	THW_5f_Hincl_13TeV_madgraph-pythia8_MiniAOD2018	0.01561
$t\bar{t}WW$	TTWW_TuneCP5_13TeV-madgraph-pythia8_MiniAOD2018	0.007829
$t\bar{t}WZ$	TTWZ_TuneCP5_13TeV-madgraph-pythia8_MiniAOD2018	0.002919
$t\bar{t}ZZ$	TTZZ_TuneCP5_13TeV-madgraph-pythia8_MiniAOD2018	0.001573
tW	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8_MiniAOD2018	35.6
$\bar{t}W$	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8_MiniAOD2018	35.6
Single t (t-channel)	ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8_MiniAOD2018	136.02
Single $\bar{t}$ (t-channel)	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8_MiniAOD2018	80.95
Single t (s-channel)	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8_MiniAOD2018	3.68

Table A.3: List of all used $t\bar{t}X$ datasets along with their respective cross sections.

Process	MC sample	cross section (pb)
$t\bar{t}W + \text{jets} \rightarrow l\nu$	TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8_MiniAOD2018	0.2043

Table A.4: List of all used $t\bar{t}W$ datasets along with their respective cross sections.

Process	MC sample	cross section (pb)
$t\bar{t} + \text{jets} \rightarrow ll$	TTJets_DiLept_TuneCP5_13TeV-madgraphMLM-pythia8_MiniAOD2018	54.23

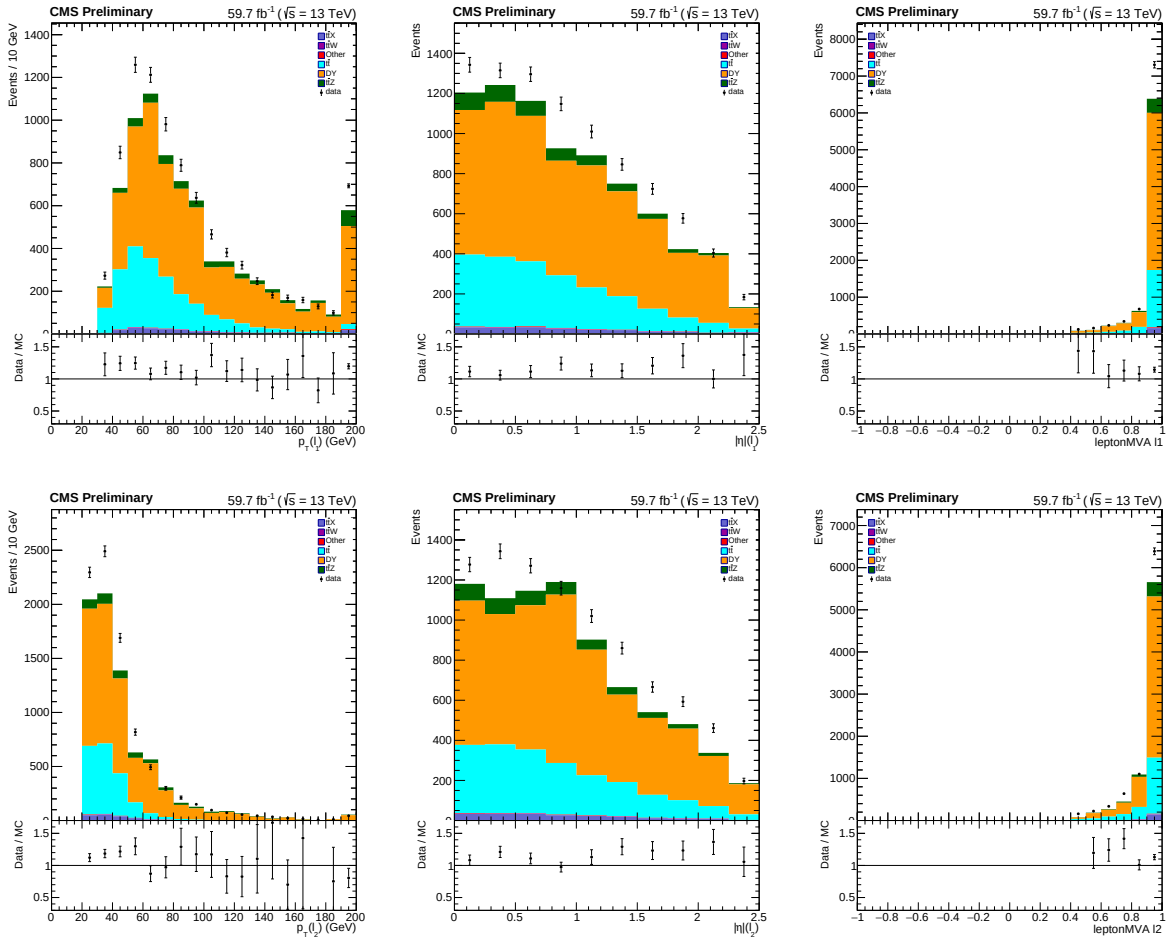
Table A.5: List of all used $t\bar{t}$ datasets along with their respective cross sections.

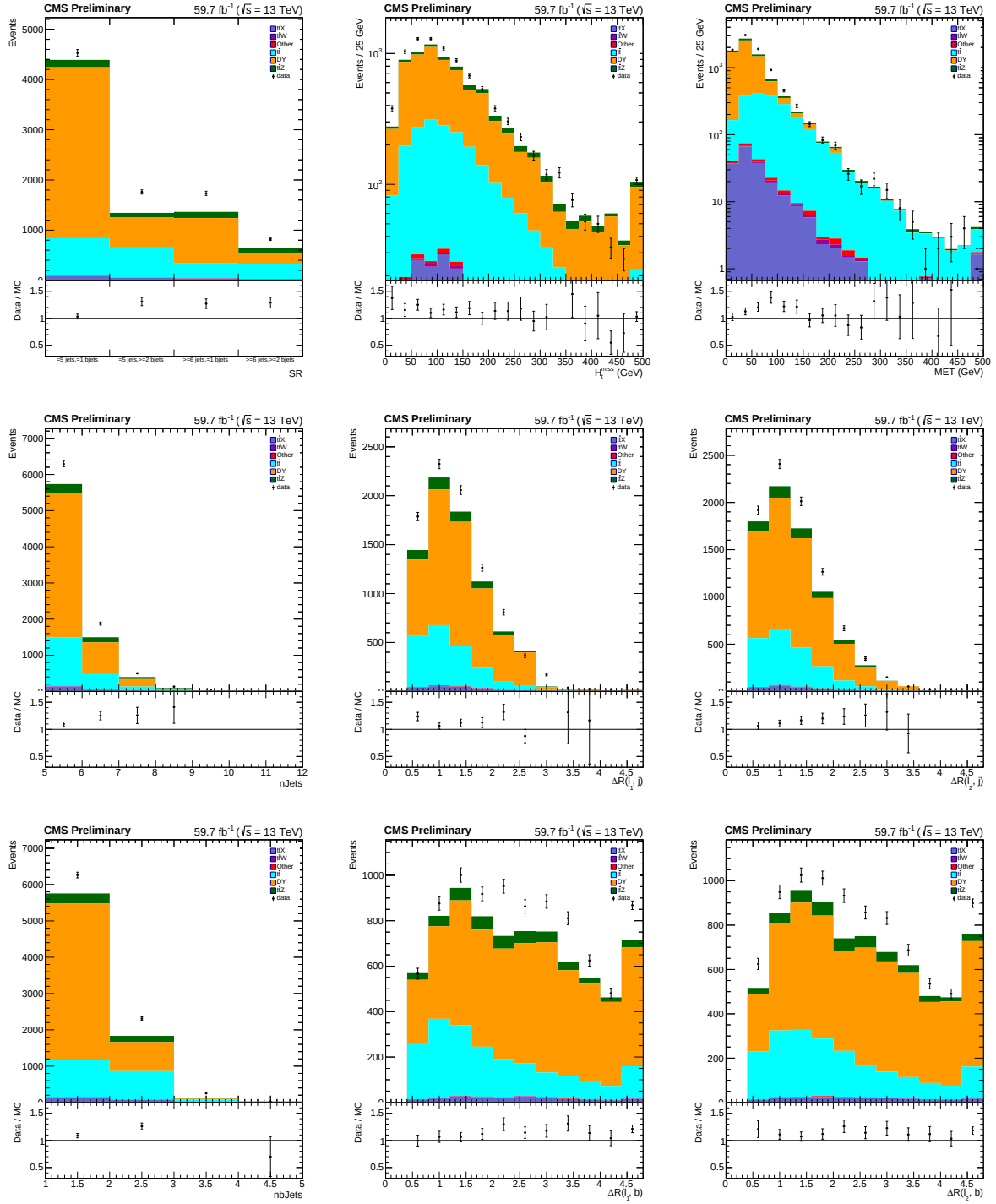
Process	MC sample	cross section (pb)
DY	DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8_MiniAOD2018	6225.42

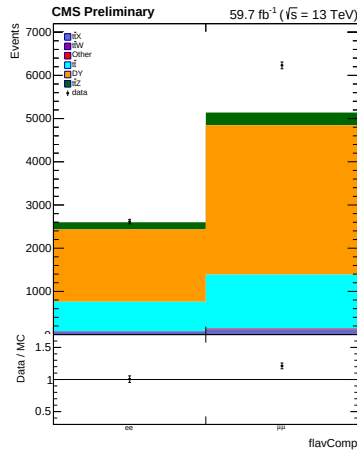
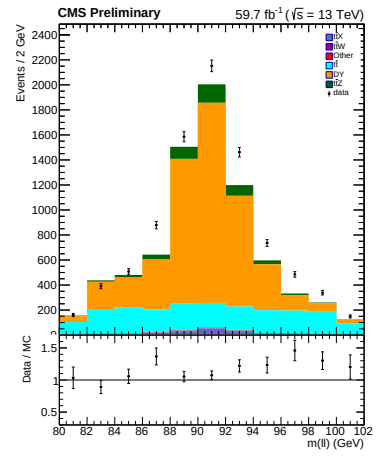
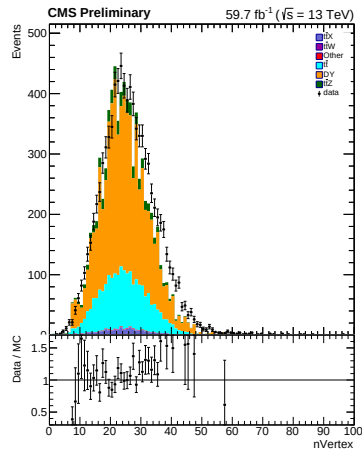
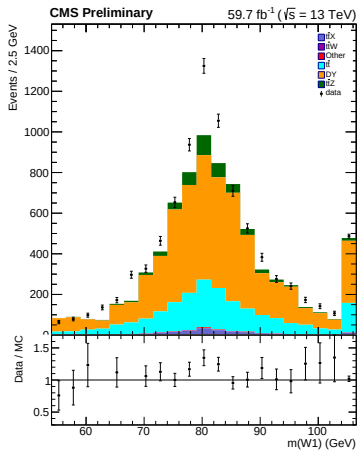
Table A.6: List of all used Drell-Yan datasets along with their respective cross sections.

Appendix B

Signal region plots

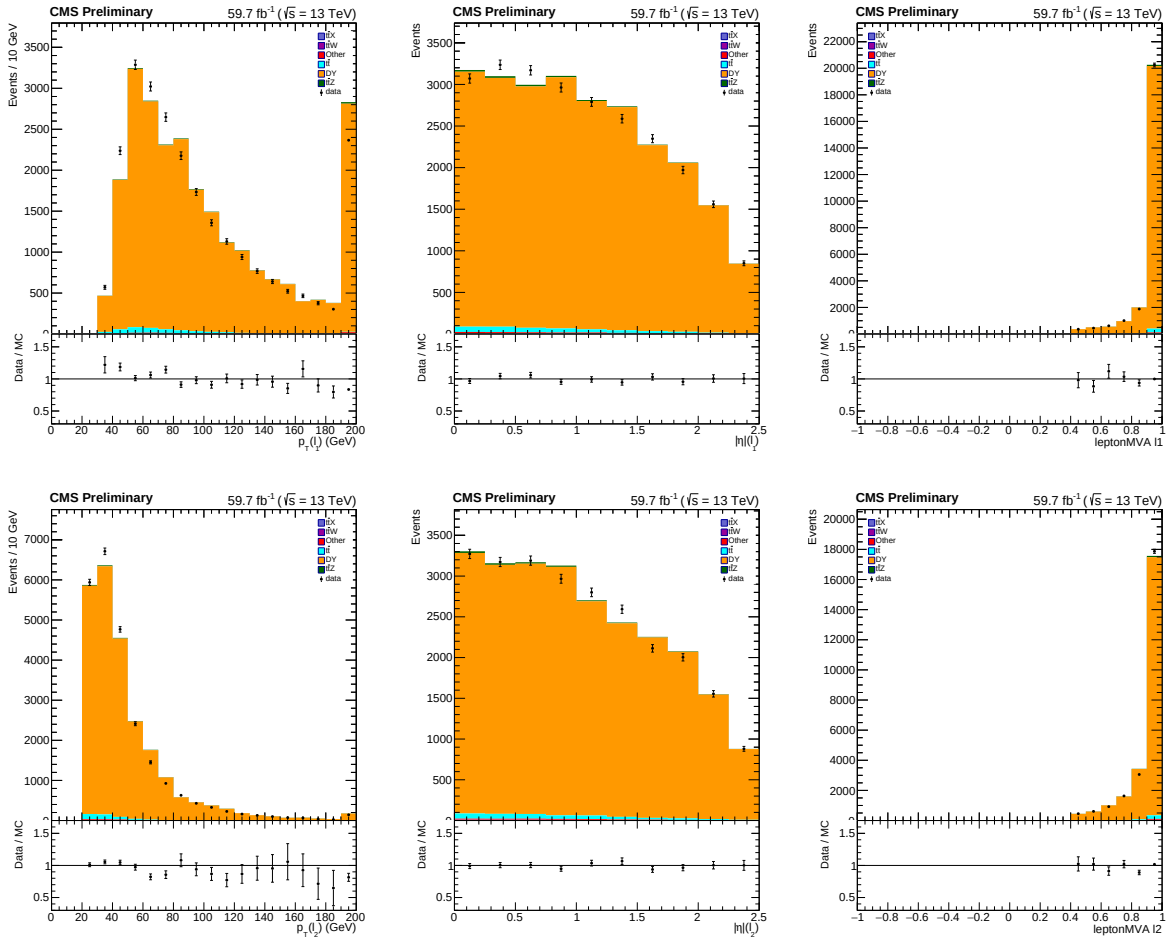


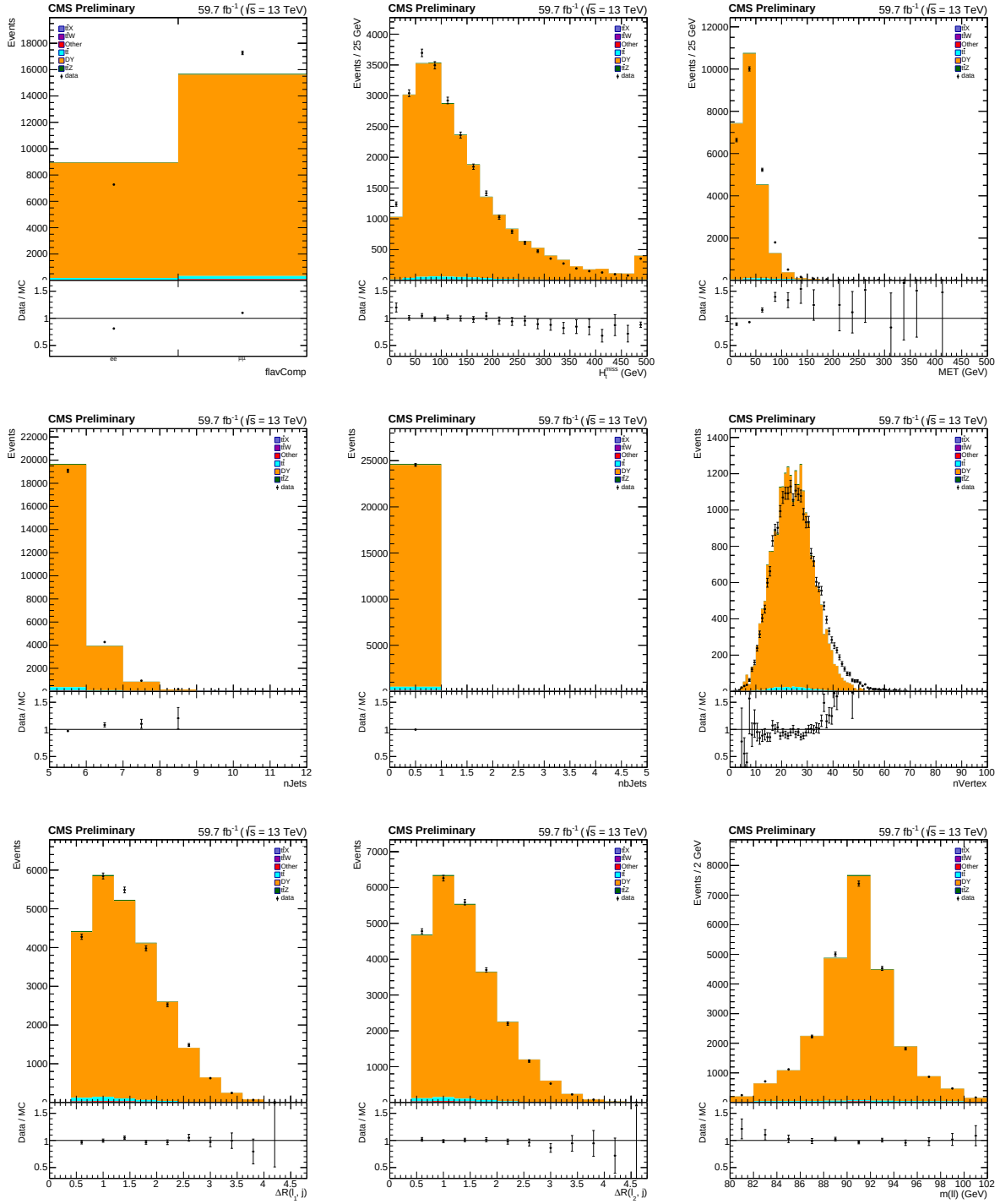




Appendix C

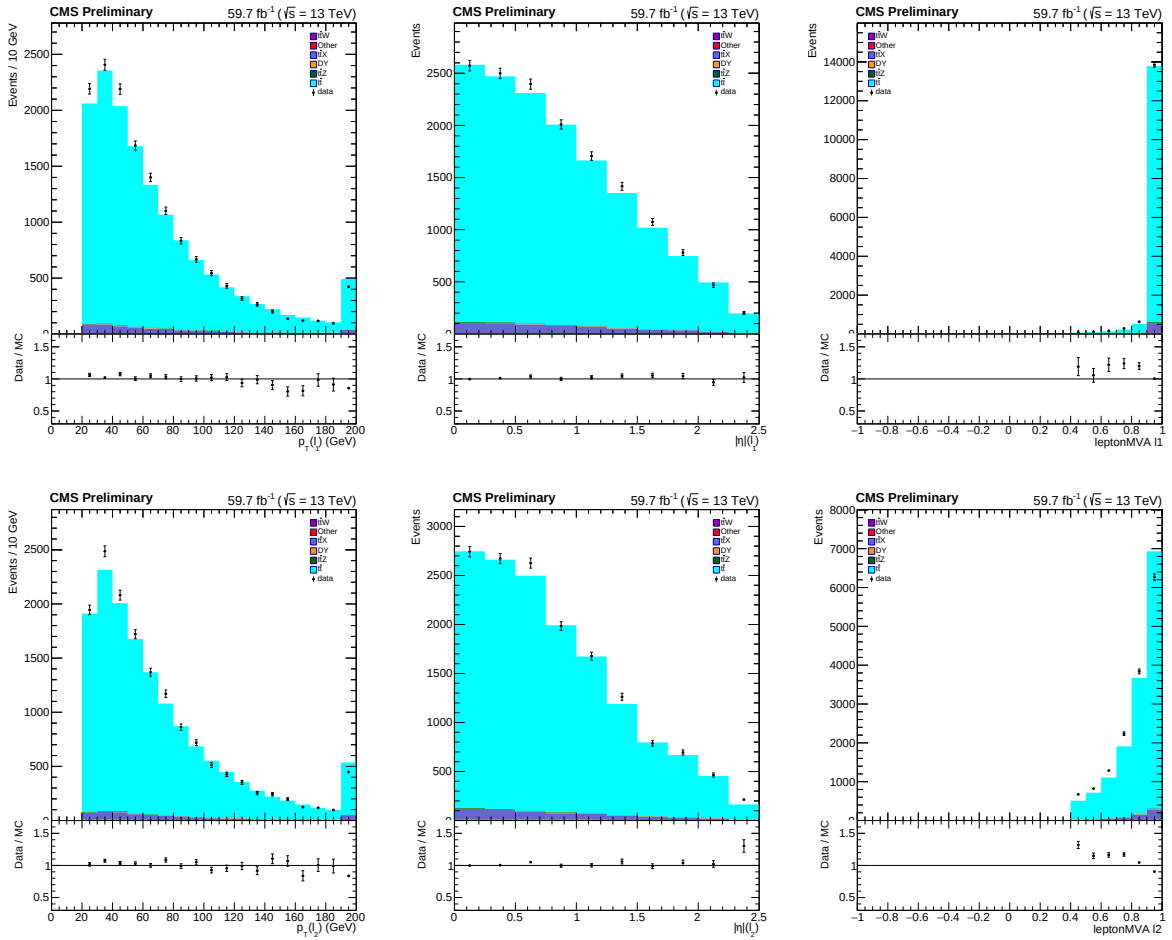
Drell-Yan control region plots

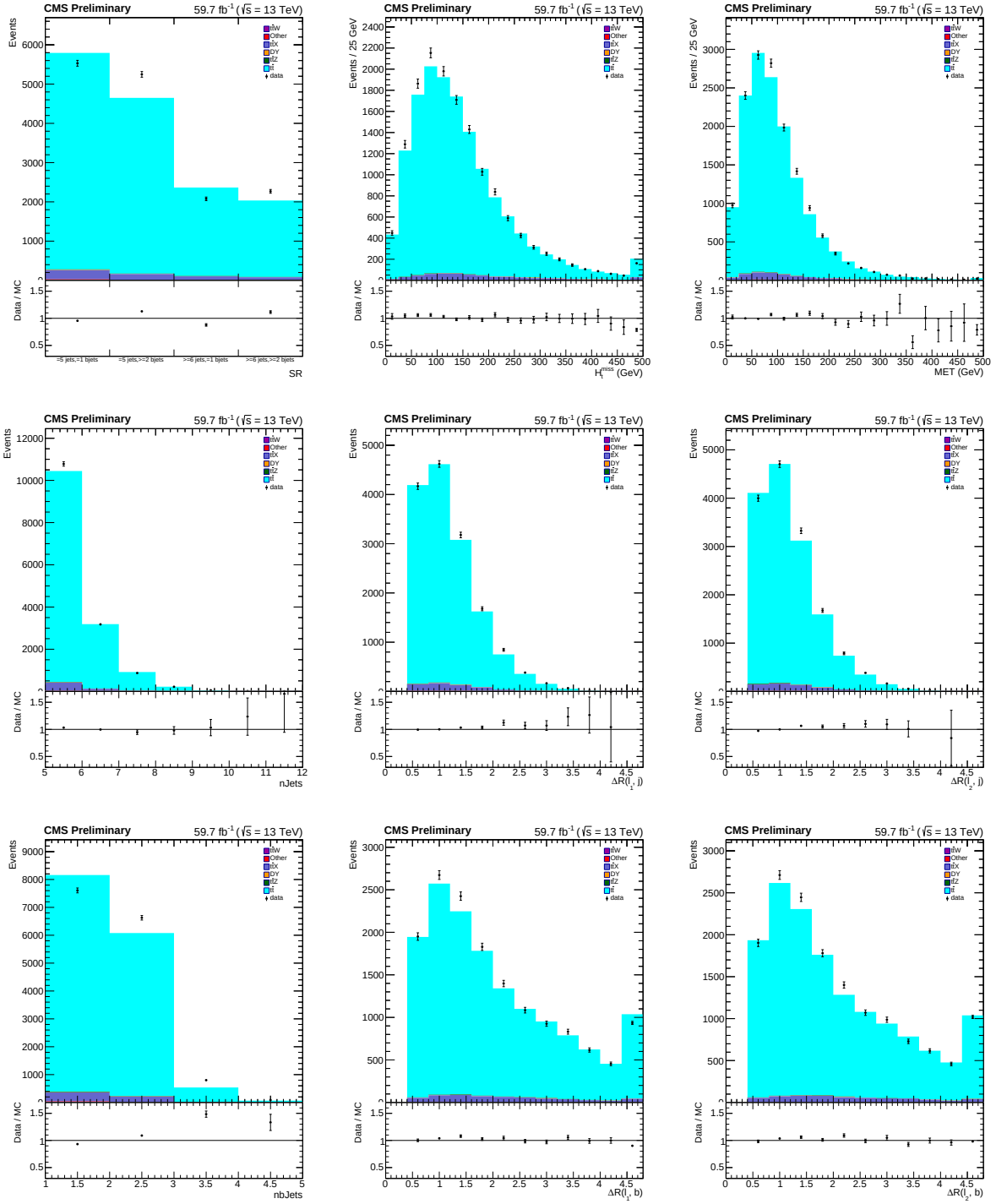


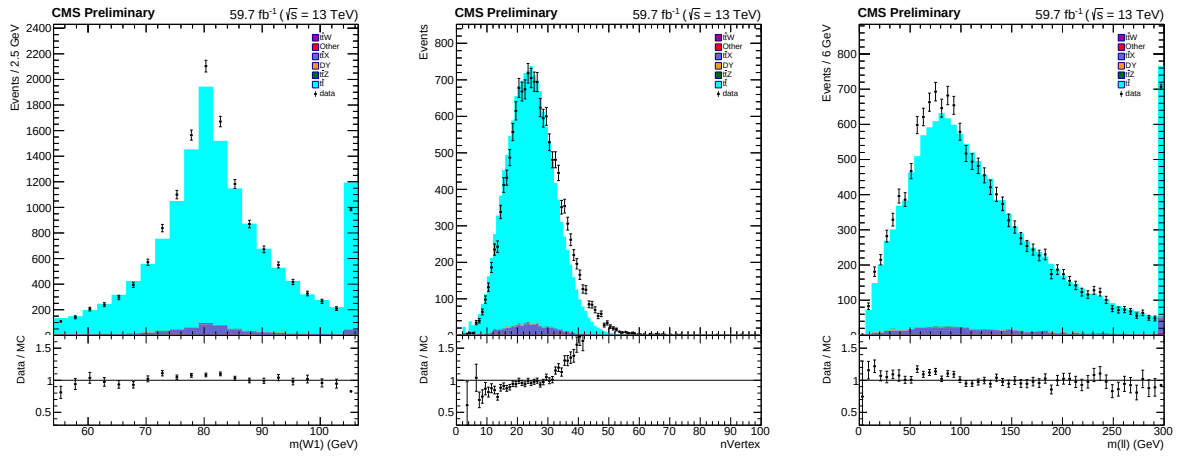


Appendix D

$t\bar{t}$ control region plots







Appendix E

Used hyperparameters for the BDT and NN

Hyperparameter	Explanation	Used value
α	L1 regularization term on the weights. A higher α results in a more conservative model.	0.5
γ	Minimum loss reduction required to make a further partition on a leaf node of the tree.	0
learning_rate	The step size used in each boosting step. Choosing a low step size will prevent overfitting.	0.01
number_of_trees	The number of trees trained during boosting.	2000
max_depth	Maximum amount of nodes the same event is allowed to go through. Increasing this value makes the BDT more likely to overfit.	2

Table E.1 – continued from previous page

Hyperparameter	Explanation	Used value
min_child_weight	The minimum sum of instance weights needed in a node. If the tree training process results in a node with the sum of instance weights less than min_child_weight, this node will become a leaf and will not split any further.	1
subsample	Subsample ratio of the training instances. Setting this to one means every boosting iteration will use all training events.	1
colsample_bytree	Subsample ratio of columns when constructing each tree.	1

Table E.1: The set of hyperparameters used to train the BDT. The explanation of the hyperparameters is largely based on Ref. [64].

Hyperparameter	Explanation	Used value
Activation function	Function used to add non-linearity to the NN.	leakyrelu
batch_size	Amount of training samples evaluated by the NN before the model's internal parameters are updated.	234
batchnorm_before_activation	Normalization of the batch before the activation function is applied.	True
batchnorm_first	Normalization of the batch before the batch is given to the input layer.	True
batchnorm_hidden	Normalization of the batch before the hidden layers.	True
dropout_all	Add a dropout layer before all hidden layers.	True

Table E.2 – continued from previous page

Hyperparameter	Explanation	Used value
dropout_first	Add a dropout layer after the first layer. If dropout_all is True, this hyperparameter is redundant.	True
dropout_rate	The fraction of nodes that will have their values randomly set to zero.	0.167
learning_rate	Step size used to update the weights of the model.	0.636
learning_rate_decay	Factor used to quantify the decay of the learning rate. By letting the learning rate decay as the training process occurs, one can start by making large changes to the weights and fine-tune the weights with small steps near the end of the training process.	0.920
number_of_epochs	The amount of times the NN will evaluate each training sample. The amount of weight updates is equal to batch_size $\times$ number_of_epochs.	0.636
number_of_hidden_layers	The amount of hidden layers.	3
optimizer	Optimizer used to update the weights.	Nadam
loss function	Function used to quantify the deviation of the prediction from the desired output.	binary_crossentropy
units_per_layer	Amount of neurons per hidden layer.	618

Table E.2: The set of hyperparameters used to train the NN.

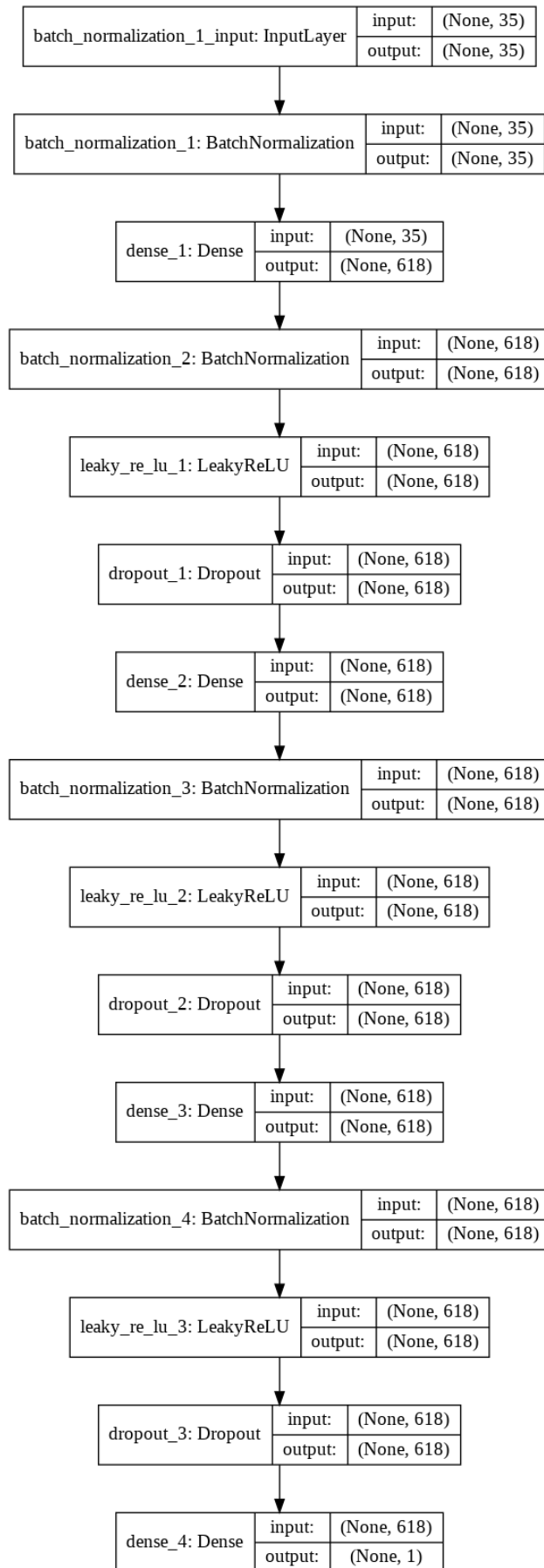


Figure E.1: A schematic overview of the structure of the NN. Figure made using [68].

Appendix F

Shape uncertainty corrections

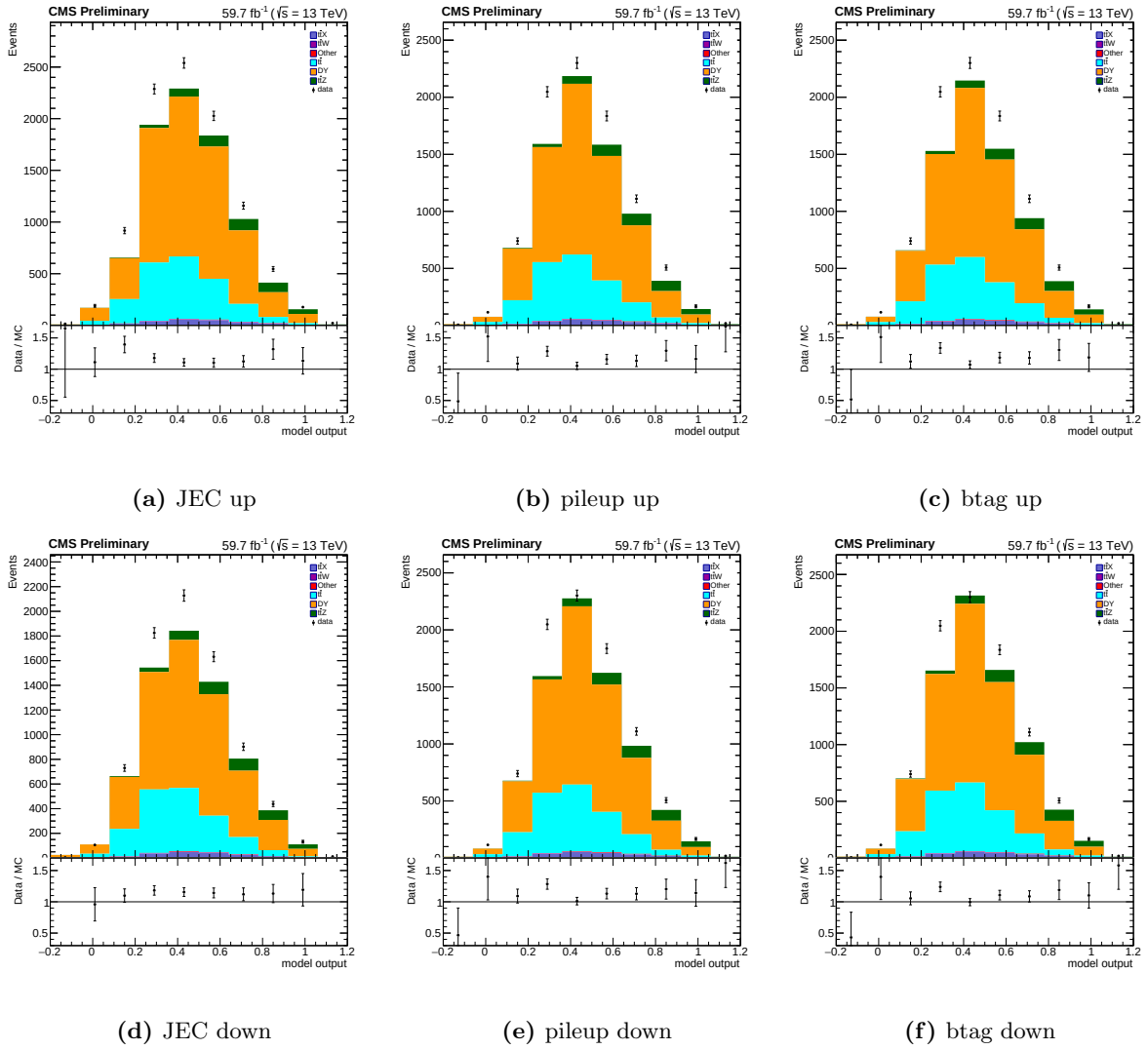


Figure F.1: The effect on the BDT output distribution of the different shape uncertainties varied one standard deviation up and down.

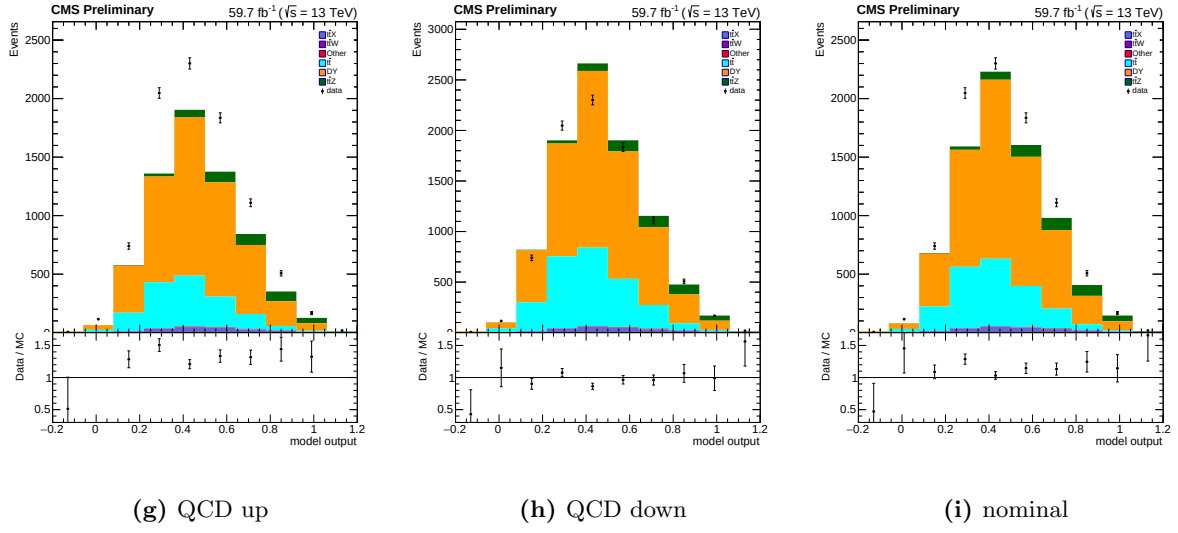


Figure F.1: The effect on the BDT output distribution of the different shape uncertainties varied one standard deviation up and down.

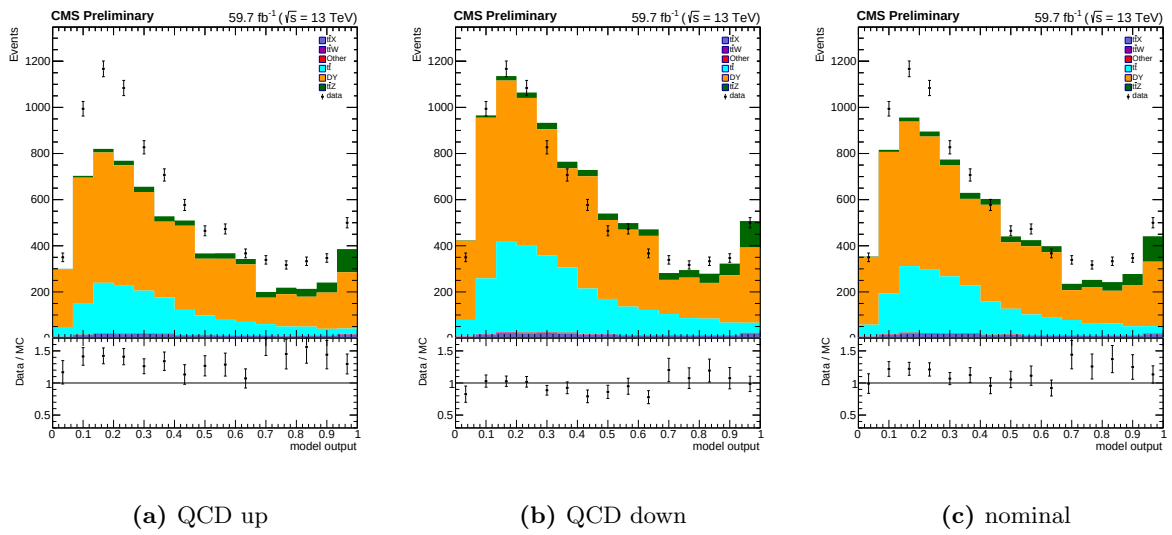


Figure F.2: The effect on the NN output distribution of the different shape uncertainties varied one standard deviation up and down.

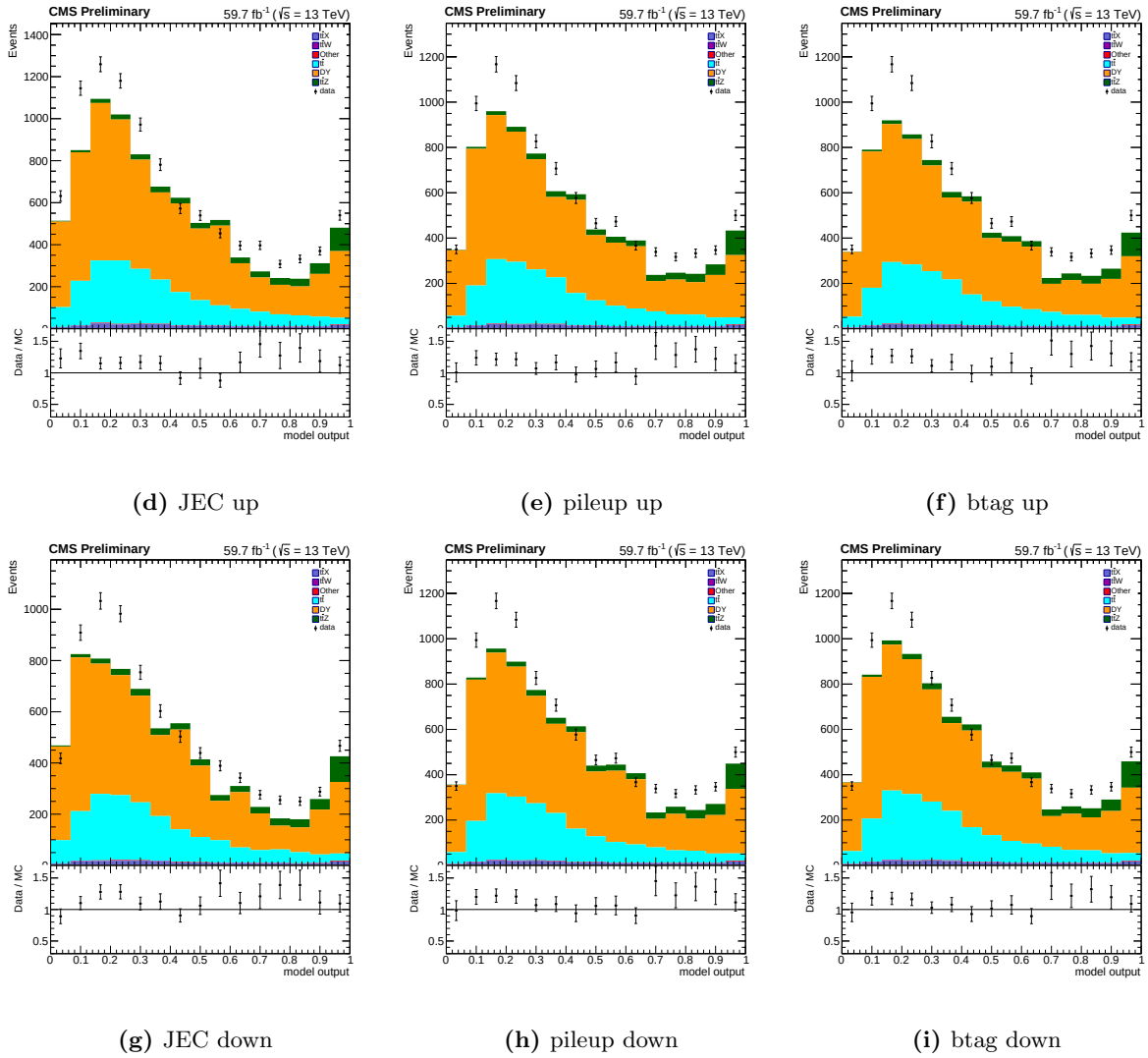


Figure F.2: The effect on the NN output distribution of the different shape uncertainties varied one standard deviation up and down.

Appendix G

Science popularization

All theses at the faculty of sciences at Ghent University are required to contain a science popularization component. For this purpose, a short introduction into particle physics was given, in Dutch. The introduction can be found below and is also published at <https://vvn.ugent.be/blog/introductie-tot-de-deeltjesfysica/>.

Elementaire deeltjes en het CERN

De mensheid is altijd zeer geïnteresseerd geweest in de fundamentele bouwstenen van alledaagse materie. Zo dachten de oude Grieken reeds dat materie uit een eindig aantal onsplitsbare deeltjes bestond. Zij hadden echter nog geen gespecialiseerde detectoren ter beschikking, waardoor hun ideeën eerder voortkwamen uit metafysische redenering en niet uit empirische bewijzen. In de 19de eeuw stelde John Dalton een theorie voor waarin alle elementen bestonden uit fundamentele deeltjes: atomen. Deze benaming vindt zijn origine in het Oudgriekse woord *ατομος* (atomos), wat “niet deelbaar” betekent.

Fermionen en fundamentele interacties

Een eerste grote doorbraak vond plaats in 1897, toen Thomson het elektron ontdekte. Aangezien dit elektron ongeveer 1800 keer kleiner was dan de toenmalig gekende atomen, moest de theorie dat atomen onsplitsbaar zijn overboord gegooid worden. Tegenwoordig weten we dat een atoom bestaat uit een nucleus met daarrond een elektronenwolk. Het elektron is een fundamenteel onsplitsbaar deeltje met een negatieve elektrische lading. De nucleus daarentegen bestaat uit andere deeltjes: positief geladen protonen en neutrale neutronen. Met een massa van ongeveer

2000 keer de massa van een elektron zijn deze deeltjes ook veel zwaarder. Elektronen worden rond de kern gehouden door een fundamentele interactie: de **elektromagnetische kracht**, die gelijk geladen deeltjes van elkaar afstoot en deeltjes met een tegengestelde lading tot elkaar aantrekt. Maar als gelijk geladen deeltjes elkaar afstoten, hoe kan het dan dat meerdere positief geladen protonen in dezelfde kern gebonden zitten? Het antwoord op deze vraag is de **sterke kernkracht**, die de elektromagnetische afstoting tussen de protonen onderling zal overwinnen om een gebonden nucleus te vormen, bestaande uit protonen en neutronen.

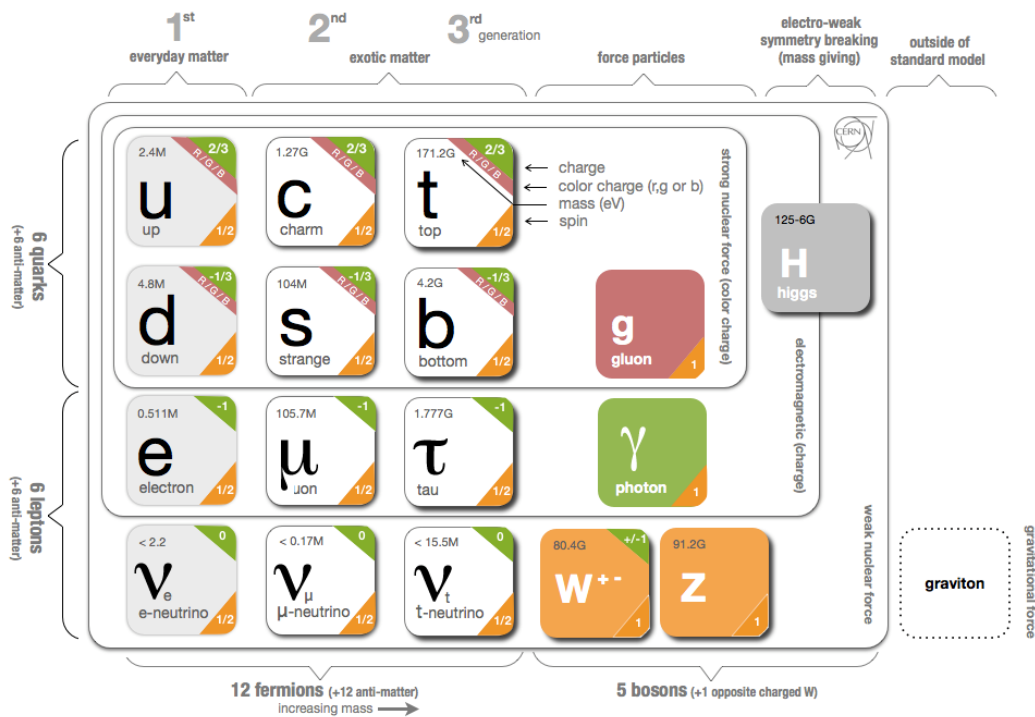


Figure G.1: De deeltjes van het standaardmodel. Figuur genomen uit [2].

Protonen en neutronen zijn echter geen fundamentele deeltjes: ze bestaan uit **quarks**. Een proton bestaat uit twee “up” quarks en één “down” quark, terwijl een neutron uit één up quark en twee down quarks bestaat. In totaal bestaan er zes verschillende quarks, die wél fundamentele deeltjes zijn. Tot nu toe kennen we dus al zeven fundamentele deeltjes: het elektron en zes quarks. Het elektron (e) heeft echter nog twee grotere broers: het muon (μ) en het tau deeltje (τ). Ze hebben exact dezelfde eigenschappen als het elektron, buiten hun massa: ze zijn een stuk zwaarder dan het elektron. Elk van deze drie deeltjes heeft een elektrisch neutrale partner, een neutrino. Tot op heden zijn de massa’s van deze neutrino’s niet gekend, we weten alleen dat ze enorm licht zijn (maar niet gewichtloos). Het elektron, muon en tau deeltje vormen samen met hun bijhorende neutrino’s de **leptonen**. De leptonen worden op hun beurt samen met de

quarks gegroepeerd als **fermionen**, zo genoemd omdat al deze deeltjes onderhevig zijn aan de Fermi-Dirac statistiek. Elk fermion heeft ook een antideeltje. Deze antideeltjes hebben exact dezelfde eigenschappen als de “normale” deeltjes, behalve hun elektrische lading. Die is namelijk omgekeerd.

Tot nu toe hebben we twee fundamentele interacties besproken: de sterke kernkracht en de elektromagnetische kracht. Er zijn echter nog twee andere fundamentele interacties, namelijk de **zwaartekracht** en de **zwakke kernkracht**. De sterke kernkracht is, zoals de naam reeds doet vermoeden, de sterkste van deze interacties. Daarna volgen de elektromagnetische interactie en de zwakke kernkracht, respectievelijk ongeveer twee en zeven grootteordes zwakker dan de sterke kernkracht. In subatomaire fysica is het effect van de zwaartekracht te verwaarlozen, aangezien deze ongeveer 38 (!) grootteordes zwakker is dan de sterke kernkracht. Op macroscopische schaal daarentegen, bijvoorbeeld ons zonnestelsel, is de zwaartekracht enorm belangrijk. De zwakke kernkracht is vooral bekend door zijn rol in radioactief β -verval, waarbij een neutron omgevormd wordt in een proton en daardoor een elektron en elektron-neutrino uitstraalt.

Bosonen

Nu we de vier fundamentele interacties en de fermionen hebben besproken, kunnen we overgaan naar de laatste groep van elementaire deeltjes: de **bosonen**. Deze groep deeltjes wordt zo genoemd omdat ze onderhevig zijn aan de Bose-Einstein statistiek. Binnen de groep van fundamentele bosonen maken we een onderscheid tussen ijkbosonen en het Higgs-boson. De ijkbosonen zijn verantwoordelijk voor het dragen van de fundamentele interacties: het massaloze foton draagt de elektromagnetische kracht, terwijl de zwakke kernkracht gedragen wordt door de massieve Z , W^+ en W^- bosonen. De sterke kernkracht wordt gedragen door massaloze gluonen, waarvan er acht verschillende soorten zijn. Maar wat met de zwaartekracht? Voor deze interactie is er tot op heden nog geen krachtdragend boson geobserveerd, maar het bestaan van het graviton als krachtdrager van de zwaartekracht is wel gepostuleerd. Het Higgs boson is een buitenbeentje onder de fundamentele bosonen, aangezien het geen drager is van een fundamentele interactie. Via het Brout-Englert-Higgs mechanisme is het Higgs boson verantwoordelijk voor het massief zijn van de meeste fundamentele deeltjes. Met zijn ontdekking in 2012 aan het CERN is het Higgs boson tot nu toe het meest recent ontdekte fundamentele deeltje. In 2013 ontving Peter Higgs samen met Belgisch fysicus François Englert de Nobelprijs in de fysica voor hun ontdekking van het Brout-Englert-Higgs mechanisme. De andere Belgische

fysicus die meegewerkt had aan de ontdekking van dit mechanisme, Robert Brout, was toen helaas reeds overleden. Hij ontving de Nobelprijs niet aangezien deze normaliter niet postuum wordt uitgereikt.

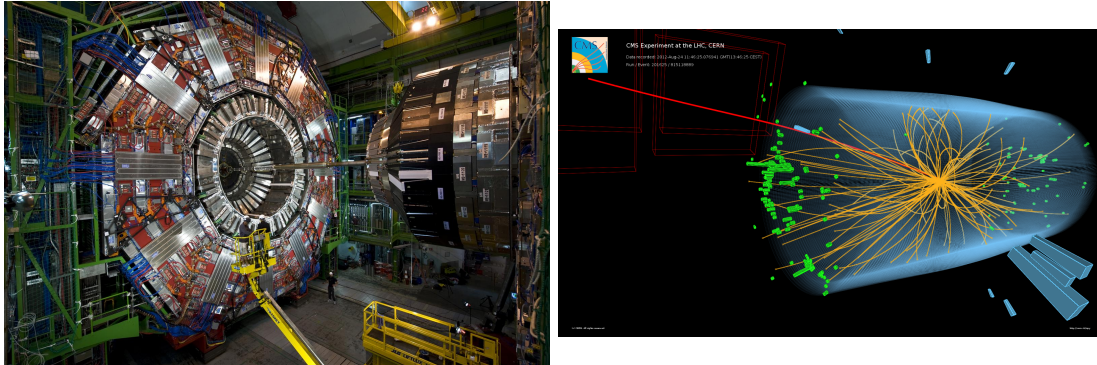


Figure G.2: Links: Foto van de CMS detector tijdens onderhoud. Rechts: Een typisch event waargenomen door de CMS detector. In dit event vervalt een Higgs boson naar twee τ leptonen. Figuren uit Refs. [107, 108].

Het standaardmodel

Al deze fermionen en bosonen, samen met alle fundamentele interacties behalve de zwaartekracht, worden beschreven in een wiskundige theorie genaamd het **standaardmodel van de deeltjesfysica** (SM). Het SM is een enorm succesvolle theorie, maar faalt toch bij de beschrijving van sommige fenomenen. Er is een groot aanbod aan theorieën die deze onregelmatigheden trachten op te lossen, tot op heden zonder succes. Voorgestelde theorieën moeten uiteraard experimenteel getest worden. We kunnen deze theorieën testen met behulp van de Large Hadron Collider (LHC). De LHC bestaat uit een ring met een omtrek van 27 km, die zich gemiddeld zo'n 100 meter onder de grond bevindt. In de LHC vinden er elke seconde ongeveer 40 miljoen botsingen plaats tussen protonen, die zullen interageren met elkaar en zo andere deeltjes maken. Op deze ring bevinden zich vier detectoren die de geproduceerde deeltjes zullen waarnemen.

Mijn thesis

Eén van deze detectoren is de Compact Muon Solenoid (CMS). Mijn thesis maakt gebruik van data verzameld door de CMS detector in 2018. Aan de hand van deze data trachtte ik de koppeling van het zwaarste deeltje in het SM, de top quark, met het Z boson te bestuderen. Aangezien

de top quark het zwaarste deeltje in het SM is, is het een heel erg interessant deeltje: veel theorieën die buiten het SM liggen zouden deeltjes bevatten die een grote koppeling met de top quark hebben. Door de top quark beter te begrijpen, kunnen deze nieuwe theorieën soms uitgesloten worden, of tot een bepaald geldigheidsinterval gelimiteerd worden. Om deze reden wordt de top quark wel eens als “een portaal naar nieuwe theorieën die voorbij het standaardmodel gaan” gezien.

Bibliography

- [1] CERN. The standard model. <https://home.cern/science/physics/standard-model>. [Online; Accessed: 10-06-2020].
- [2] Andrew Purcell. Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules. (BUL-NA-2012-269. 35/2012):10, Aug 2012.
- [3] Richard M. Weiner. Spin-statistics-quantum number connection and supersymmetry. *Physical Review D*, 87(5), Mar 2013.
- [4] S. Chatrchyan, V. Khachatryan, A.M. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan, and et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, Sep 2012.
- [5] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.
- [6] M. Tanabashi et al. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.
- [7] C.S. Wu, E. Ambler, R.W. Hayward, D.D. Hoppes, and R.P. Hudson. Experimental Test of Parity Conservation in β Decay. *Phys. Rev.*, 105:1413–1414, 1957.
- [8] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay. Evidence for the 2π decay of the k_2^0 meson. *Phys. Rev. Lett.*, 13:138–140, Jul 1964.
- [9] Christine Sutton. Electroweak theory. <https://www.britannica.com/science/electroweak-theory>, 2006. [Online; Accessed: 28-05-2020].
- [10] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.

- [11] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc.*, C680519:367–377, 1968.
- [12] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.
- [13] Henri Verschelde. Quantum field theory course. Ghent University, 2018.
- [14] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A.A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins, and et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, Sep 2012.
- [15] Daniel Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara. Progress toward a theory of supergravity. *Physical Review D*, 13(12):3214–3218, June 1976.
- [16] Planck Collaboration. Planck 2018 results. VI. Cosmological parameters, 2018.
- [17] P. J. E. Peebles and Bharat Ratra. The cosmological constant and dark energy. *Reviews of Modern Physics*, 75(2):559–606, Apr 2003.
- [18] Sean M. Carroll. The cosmological constant. *Living Reviews in Relativity*, 4(1), Feb 2001.
- [19] CERN. The Matter-Antimatter problem. <https://home.cern/science/physics/matter-antimatter-asymmetry-problem>. [Online; Accessed: 28-05-2020].
- [20] A.D. Sakharov. Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. *Sov. Phys. Usp.*, 34(5):392–393, 1991.
- [21] S. Abe, T. Ebihara, S. Enomoto, K. Furuno, Y. Gando, K. Ichimura, H. Ikeda, K. Inoue, Y. Kibe, Y. Kishimoto, and et al. Precision measurement of neutrino oscillation parameters with kamland. *Physical Review Letters*, 100(22), Jun 2008.
- [22] Susanne Mertens. Direct neutrino mass experiments. *Journal of Physics: Conference Series*, 718:022013, May 2016.
- [23] F. Abe, H. Akimoto, A. Akopian, M. G. Albrow, S. R. Amendolia, D. Amidei, J. Antos, C. Anway-Wiese, S. Aota, G. Apollinari, and et al. Observation of top quark production in $p\bar{p}$ collisions with the collider detector at fermilab. *Physical Review Letters*, 74(14):2626–2631, Apr 1995.

- [24] S. Abachi, B. Abbott, M. Abolins, B. S. Acharya, I. Adam, D. L. Adams, M. Adams, S. Ahn, H. Aihara, J. Alitti, and et al. Observation of the top quark. *Physical Review Letters*, 74(14):2632–2637, Apr 1995.
- [25] Johannes Haller, Andreas Hoecker, Roman Kogler, Klaus Mönig, Thomas Peiffer, and Jörg Stelzer. Update of the global electroweak fit and constraints on two-Higgs-doublet models. *Eur. Phys. J. C*, 78(arXiv:1803.01853. 8):675. 33 p, Mar 2018. 33 pages, 13 figures, submitted to EPJC.
- [26] Michele Gallinaro. Top quark physics: A tool for discoveries. *Journal of Physics: Conference Series*, 447:012012, jul 2013.
- [27] CERN. The Large Hadron Collider. <https://home.cern/science/accelerators/large-hadron-collider>. [Online; Accessed: 22-05-2020].
- [28] Oliver Bruning and Lucio Rossi. The High-Luminosity Large Hadron Collider. *Nature Rev. Phys.*, 1(CERN-ACC-2019-0039):1. 3 p, Mar 2019.
- [29] Fabienne Marcastel. CERN’s Accelerator Complex. La chaîne des accélérateurs du CERN. Oct 2013. General Photo.
- [30] CMS Collaboration. The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment. *JINST*, 3:S08004. 361 p, 2008. Also published by CERN Geneva in 2010.
- [31] CMS Collaboration. CMS detector design. <http://cms.web.cern.ch/news/cms-detector-design>. [Online; Accessed: 23-05-2020].
- [32] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, 2006.
- [33] https://wiki.physik.uzh.ch/cms/_detail/latex:cms_coordinate_system.png?id=latex%3Aexample_spherical_coordinates. [Online; Accessed: 23-05-2020].
- [34] CMS Collaboration. Interactive slice of the CMS detector. <https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=4172>. [Online; Accessed: 23-05-2020].
- [35] CMS Collaboration. Alignment of the CMS Silicon Tracker during Commissioning with Cosmic Rays. *JINST*, 5(arXiv:0910.2505):T03009. 39 p, Oct 2009.

-
- [36] Jorge Molina, CMS Warsaw, Krzysztof Pozniak, Ryszard Romaniuk, and Wojciech Zabolotny. Performance of CMS hadron calorimeter timing. *Journal of Instrumentation*, 5:T03013, 03 2010.
- [37] O. Poeth. *The CMS Silicon Strip Tracker: Concept, Production and Commissioning*. Vieweg+Teubner research. Vieweg+Teubner Verlag, 2010.
- [38] CMS Collaboration. Cathode strip chambers. <http://cms.web.cern.ch/news/cathode-strip-chambers>. [Online; Accessed: 23-05-2020].
- [39] D Acosta et al. Large CMS cathode strip chambers: design and performance. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 453(1):182 – 187, 2000. Proc. 7th Int. Conf on Instrumentation for colliding Beam Physics.
- [40] CMS Collaboration. Resistive plate chambers. <https://cms.cern/detector/detecting-muons/resistive-plate-chambers>. [Online; Accessed: 23-05-2020].
- [41] CMS Collaboration. Triggering and data acquisition. <https://cms.cern/detector/triggering-and-data-acquisition>. [Online; Accessed: 24-05-2020].
- [42] CERN. Facts and figures about the LHC. <https://home.cern/resources/faqs/facts-and-figures-about-lhc>. [Online; Accessed: 02-06-2020].
- [43] A.M. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, M. Flechl, and et al. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003–P10003, Oct 2017.
- [44] Nadja Catharina Strobbe. The Razor Boost analysis - Another step in the hunt for new physics at CMS, 2015.
- [45] Andrea Perrotta. CMS event reconstruction status in Run 2. Technical Report CMS-CR-2018-269, CERN, Geneva, Oct 2018.
- [46] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, Apr 2009.

- [47] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, Apr 2008.
- [48] Tom M Mitchell et al. Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, 45(37):870–877, 1997.
- [49] Taylor Fogarty. Regression or Classification? Linear or Logistic? <https://towardsdatascience.com/regression-or-classification-linear-or-logistic-f093e8757b9c>. [Online; Accessed: 16-05-2020].
- [50] Tarang Shah. About train, validation and test sets in machine learning. <https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7>. [Online; Accessed: 16-05-2020].
- [51] Anas Al-Masri. What are overfitting and underfitting in machine learning? <https://towardsdatascience.com/what-are-overfitting-and-underfitting-in-machine-learning-a96b30864690>. [Online; Accessed: 16-05-2020].
- [52] Mallikarjun. ML Fundamentals: Bias Variance Trade-off. <https://medium.com/@arjunm2305/ml-fundamentals-bias-variance-trade-off-ab58bc7d3d90>. [Online; Accessed: 17-05-2020].
- [53] Jason Brownlee. Gentle introduction to the bias-variance trade-off in machine learning. <https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/>. [Online; Accessed: 16-05-2020].
- [54] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.
- [55] What is machine learning? A definition. <https://expertsystem.com/machine-learning-definition/>. [Online; Accessed: 17-05-2020].
- [56] Thomas Plapinger. What is a decision tree? <https://towardsdatascience.com/what-is-a-decision-tree-22975f00f3e1>. [Online; Accessed: 17-05-2020].
- [57] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282 vol.1, 1995.

- [58] Prashant Gupta. Decision trees in machine learning. <https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052>. [Online; Accessed: 17-05-2020].
- [59] Jerome Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29, 11 2000.
- [60] Soner Yıldırım. Gradient boosted decision trees-explained. <https://towardsdatascience.com/gradient-boosted-decision-trees-explained-9259bd8205af>. [Online; Accessed: 20-05-2020].
- [61] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.
- [62] Josh Starmer. Statquest. <https://statquest.org/video-index/>. [Online; Accessed: 20-05-2020].
- [63] Jerome Friedman. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38:367–378, 02 2002.
- [64] Tianqi Chen and Carlos Guestrin. XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug 2016.
- [65] Myroslava Zelenska. Neural networks — relation to human brain and cognition. <https://becominghuman.ai/neural-networks-relation-to-human-brain-and-cognition-b45575359f64>. [Online; Accessed: 21-05-2020].
- [66] Yvan Saeys. Machine learning course. Ghent University, 2019.
- [67] Jason Brownlee. A gentle introduction to the rectified linear unit (relu). <https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/>. [Online; Accessed: 21-05-2020].
- [68] François Chollet et al. Keras. <https://keras.io>, 2015.
- [69] Vijini Mallawaarachchi. Introduction to genetic algorithms. <https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3>. [Online; Accessed: 21-05-2020].

- [70] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogio, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, A. Escalante Del Valle, and et al. Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 2020(3), Mar 2020.
- [71] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogio, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, and et al. Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 2018(8), Aug 2018.
- [72] Thomas G. McCarthy. Measurements of the production of single top quarks or of top quark pairs in association with a W boson, a Z boson, or a photon with ATLAS & CMS, 2020.
- [73] A.M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogio, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, and et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *Journal of Instrumentation*, 13(05):P05011–P05011, May 2018.
- [74] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5: going beyond. *Journal of High Energy Physics*, 2011(6):128, 2011.
- [75] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), Jul 2014.
- [76] Stefano Frixione, Eric Laenen, Patrick Motylinski, and Bryan R. Webber. Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations. *JHEP*, 04:081, 2007.
- [77] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *JHEP*, 03:015, 2013.
- [78] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *Journal of High Energy Physics*, 2010(6), Jun 2010.

- [79] Paolo Nason and Giulia Zanderighi. W^+W^- WZ and ZZ production in the POWHEG-BOX-V2. *The European Physical Journal C*, 74(1), Jan 2014.
- [80] Stefano Frixione and Bryan R Webber. Matching NLO QCD computations and parton shower simulations. *Journal of High Energy Physics*, 2002(06):029–029, Jun 2002.
- [81] Rikkert Frederix, Emanuele Re, and Paolo Torrielli. Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO. *Journal of High Energy Physics*, 2012(9), Sep 2012.
- [82] Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *Journal of High Energy Physics*, 2012(12), Dec 2012.
- [83] John M. Campbell and R.K. Ellis. MCFM for the Tevatron and the LHC. *Nuclear Physics B - Proceedings Supplements*, 205-206:10–15, Aug 2010.
- [84] Sara Bolognesi, Yanyan Gao, Andrei V. Gritsan, Kirill Melnikov, Markus Schulze, Nhan V. Tran, and Andrew Whitbeck. Spin and parity of a single-produced resonance at the LHC. *Physical Review D*, 86(9), Nov 2012.
- [85] S. Chatrchyan, V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan, and et al. Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV. *Journal of High Energy Physics*, 2013(6), Jun 2013.
- [86] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. Fastjet user manual. *The European Physical Journal C*, 72(3), Mar 2012.
- [87] https://twiki.cern.ch/twiki/bin/viewauth/CMS/LeptonMVA#Input_Variables.
- [88] Albert M Sirunyan et al. Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 122(13):132003, 2019.
- [89] Tom Cornelis, Kirill Skovpen and Willem Verbeke. <https://homepage.iihe.ac.be/~tomc/mvaRoc/rocs-2018-TTZ/>. [Online; Accessed: 8-6-2020].
- [90] <https://twiki.cern.ch/twiki/bin/view/CMS/MuonPOG>.

- [91] CMS Collaboration. Higgsanalysis-combinedlimit. <http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/>. [Online; Accessed: 7-6-2020].
- [92] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), Feb 2011.
- [93] J. S. Conway. Incorporating nuisance parameters in likelihoods for multisource spectra, 2011.
- [94] CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-LUM-18-002, CERN, Geneva, 2019.
- [95] <https://twiki.cern.ch/twiki/bin/view/CMS/IntroToJEC>.
- [96] https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC#2018_Data.
- [97] <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupInformation>.
- [98] <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMethods>.
- [99] R.Keith Ellis, W.James Stirling, and B.R. Webber. *QCD and collider physics*, volume 8. Cambridge University Press, 2 2011.
- [100] <https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/part3/nonstandard/#nuisance-parameter-impacts>.
- [101] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H.-S. Shao, and M. Zaro. The automation of next-to-leading order electroweak calculations. *Journal of High Energy Physics*, 2018(7), Jul 2018.
- [102] S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro. Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons, 2015.
- [103] D de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector, 2016.
- [104] M. Aaboud, G. Aad, B. Abbott, D.C. Abbott, O. Abdinov, B. Abeloos, D.K. Abhayasinghe, S.H. Abidi, O.S. AbouZeid, N.L. Abraham, and et al. Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $s=13$ TeV with the ATLAS detector. *Physical Review D*, 99(7), Apr 2019.

-
- [105] Anna Kulesza, Leszek Motyka, Daniel Schwartländer, Tomasz Stebel, and Vincent Theeuwes. Associated top quark pair production with a heavy boson: differential cross sections at NLO+NNLL accuracy. *Eur. Phys. J. C*, 80(5):428, 2020.
- [106] Alessandro Broggio, Andrea Ferroglia, Rikkert Frederix, Davide Pagani, Benjamin D. Pecjak, and Ioannis Tsinikos. Top-quark pair hadroproduction in association with a heavy boson at NLO+NNLL including EW corrections. *JHEP*, 08:039, 2019.
- [107] <https://home.cern/resources/image/experiments/cms-images-gallery>.
- [108] CMS Collaboration and Thomas McCauley. CMS Higgs Search in 2011 and 2012 data: candidate tau-tau event (8 TeV). CMS Collection., Dec 2013.