Search for the Standard Model Production of a Single Top Quark in Association with a Z^0 Boson Using Machine Learning Techniques

A thesis submitted for the degree of Doctor of Philosophy

by

CORIN J. K. HOAD

Department of Electronic and Computer Engineering
Brunel University London

Colophon

This document was typeset using the LuaTeX distribution of LaTeX 2$_\varepsilon$ with the `memoir` package. References were managed with BibLaTeX and biber. The body is set in 12 pt Libertinus Serif, a fork of the Linux Libertine typeface by the Alif Type foundry. Sans serif is provided by Libertinus Sans and monospace by `Latin Modern Mono`. Mathematics is set in Libertinus Math with additional glyphs from the XITS font project. The initial on Page 1 is taken from the Royal Initialen font.

Abstract

This thesis presents a search for the production of a single top quark in association with a Z^0 boson in the dileptonic decay channel as predicted by the Standard Model. The search uses $77.8\,\mathrm{fb}^{-1}$ of data from $\sqrt{s} = 13\,\mathrm{TeV}$ proton–proton collisions collected by the Compact Muon Solenoid experiment at the Large Hadron Collider during the 2016–2017 data-taking period. The search identified events containing a Z^0 boson decay by requiring two opposite-sign same-flavour electrons or muons in the final state with invariant mass compatible with the nominal Z^0 boson mass. Products of the top quark decay were identified using techniques developed to identify jets originating from bottom quarks and searching for a jet pair with invariant mass compatible with the $W^{\pm}$ boson mass.

Machine learning techniques were used to further discriminate the signal process from background events. A study was carried out, comparing the performance of boosted decision trees with hyperparameters optimised using a Gaussian process and multi-layer perceptrons on this problem. The boosted decision trees were found to outperform the multi-layer perceptrons.

A signal strength of $\hat{r} = 6.52^{+2.30}_{-2.05}$ was observed, where $\hat{r} = 1.0$ corresponds to the Standard Model expectation. The corresponding observed (expected) significance is 3.12σ (0.48σ).

Declaration of authorship

The work contained in this thesis is solely that of the author except where collaboration with others occurred. The author primarily cooperated with other members of the High Energy Physics group at Brunel University London to create the analysis presented in this thesis. The primary contributions of the author were the adaptation of the analysis for data taken during the 2017 data-taking period and the application of the machine learning techniques described in Chapters 3 and 9.

No part of this thesis has been previously submitted to this or any other university as part of the requirement for a higher degree. When the published work of others has been consulted, it has been clearly attributed within the text.

The results presented in this thesis use data taken by the CMS experiment but are not officially endorsed by the CMS collaboration in any capacity.

Contents

List of Figures

List of Tables

List of Abbreviations

4FS Four-Flavour Scheme
5FS Five-Flavour Scheme

Adam Adaptive moment estimation
ALICE A Large Ion Collider Experiment
ANN Artificial neural network
aMC@NLO Automatic MC@NLO
ASIC Application-specific integrated circuit
ATLAS A Toroidal LHC Apparatus
AUROC Area under the ROC curve

BDT Boosted decision tree
BSM Beyond the Standard Model
BTV b tag and vertexing

CART Classification and regression tree
CERN European Organization for Nuclear Research
CKM Matrix Cabibbo–Kobayashi–Maskawa Matrix
CMS Compact Muon Solenoid
CPU Central processing unit
CR Control region
CSC Cathode strip chamber
CSVv2 Combined Secondary Vertex Version 2
CTF Combinatoral Track Finder

DeepCSV Deep Combined Secondary Vertex
DGLAP Equation Dokshitzer–Gribov–Lipatov–Altarelli–Parisi Equation
DT chamber Drift tube chamber

ECAL Electromagnetic calorimeter
ECDF Empirical cumulative distribution function
EWT Electroweak theory

FCNC Flavour changing neutral current
FPGA Field programmable gate array
FSR Final-state radiation

Geant4 Geometry and Tracking 4
GP Gaussian process
GSF Gaussian sum filter

HB HCAL Barrel
HCAL Hadronic calorimeter
HE HCAL End Cap
HF Forward HCAL

HLT High Level Trigger
HO Outer HCAL

IP Interaction point
ISR Initial-state radiation
IVF Inclusive Vertex Finder

JEC Jet energy correction
JERC Jet energy resolution and corrections

KF Kalman filter
KS Kolmogorov–Smirnov

L1 Trigger Level 1 Trigger
LASSO Least absolute shrinkage and selection operator
LEP Large Electron–Positron Collider
LHAPDF Les Houches Accord Parton Distribution Function
LHC Large Hadron Collider
LHCb Large Hadron Collider Beauty
LO Leading order
LS2 Long Shutdown 2

MC Monte Carlo
MC@NLO Monte Carlo at next-to-leading order
ME Matrix element
ML Machine learning
MLP Multilayer perceptron

NLO Next-to-leading order
NPL Nonprompt lepton

PDF Parton distribution function
PF Particle Flow
PMNS Matrix Pontecorvo–Maki–Nakagawa–Sakata Matrix
POG Physics Object Group
PS Proton Synchrotron

QCD Quantum chromodynamics
QED Quantum electrodynamics
QFD Quantum flavourdynamics
QFT Quantum field theory

RBF kernel Radial basis function kernel
ReLU Rectified linear unit
RFE Recursive feature elimination
ROC curve Receiver operating characteristic curve
RPC Resistive plate chamber

SL Superlayer
SM Standard Model
SMC Shower Monte Carlo

TEC Tracker End Cap
TIB Tracker Inner Barrel
TID Tracker Inner Disk
TOB Tracker Outer Barrel

WP Working point

XGBoost Extreme Gradient Boosting

Chapter 1

Introduction

> *In all things of nature there is something of the marvellous.*
>
> Aristotle

It was Leucippus, according to Aristotle, who first promoted atomism in the Western canon [1]. The sage Aruni had espoused similar ideas in India c. 800 BCE, three centuries prior [2]. The concept behind atomism is simple and, in the modern day, familiar: the macroscopic world with which we interact every day is underpinned by a microscopic world of imperceptibly small, indivisible atoms—the fundamental building blocks of everything. Of course, the atom of early atomism with its infinite possible shapes and sizes was not the atom of Dalton, where each element consisted of an atom of a unique type. Nor was, as Thomson would find in his investigations into cathode rays in 1897 [3], the atom itself fundamental. There existed a smaller unit: the electron. Rutherford would observe that these electrons orbited a dense nucleus containing a positively charged particle: the proton [4]. Later, the remaining component of the nucleus, the uncharged neutron, would be observed by Chadwick [5]. Over time, further fundamental particles were discovered. The muon in 1937 [6]. The electron neutrino in 1956 [7]. The muon neutrino in 1962 [8]. In 1969 the atom itself was further divided: not only did the nucleus consist of protons and neutrons, but protons and neutrons themselves contained indivisible point particles [9, 10]. These were later dubbed the up and down *quark*. By the discovery of the τ lepton in 1975 [11], a truly remarkable theory was in development. A theory that could unite the expanding particle zoo, codifying their relationships and interactions via the fundamental *electromagnetic*, *strong nuclear*, and *weak nuclear* forces. Today, we know this theory as the *Standard Model* (SM).

An impressive predictive power—such as the foretelling the discovery of the $W^{\pm}$ bosons (1983) [12], Z^{0} boson (1983) [12], and tau neutrino (2000) [13]—cemented the Standard Model as the standard model of modern particle physics. But, as the field developed, it became clear that the SM could not be a complete theory of matter. Where was the mechanism of gravity? Why is there so little antimatter in the Universe? What is dark matter? These are all open questions to which the SM provides no answer. Paradoxically, the predictive power of the SM never seemed to falter, its arguably crowing achievement arriving in 2012 when the *Compact Muon Solenoid* (CMS) and *A Toroidal LHC Apparatus* (ATLAS) experiments discovered the long-theorised Higgs Boson [14, 15]. An overarching goal of particle physics today is therefore to find where the predictions of the SM are amiss and in doing so understand the shape of the physics that lies beyond it. A physics in which the four fundamental forces can be unified, the matter–antimatter asymmetry is justified, and the ingredients of dark matter are known. Fortunately, in a 27 km tunnel below the Franco–Swiss countryside, particle physics has a powerful ally in this quest. The most powerful and luminous particle accelerator ever constructed: the *Large Hadron Collider* (LHC). The base of the aforementioned CMS and ATLAS experiments, the LHC allows the investigation of energy scales where divergence from the SM due to *beyond the Standard Model* (BSM) physics may be observable. Particle physics is a broad field, and so the question naturally arises of where efforts to discover these divergences are best served. A compelling candidate, upon which this thesis will focus, lies in an as-yet unmentioned element of the SM: the top quark.

The cynosure of the top quark is its mass; at 173.0 ± 0.4 GeV [16] it is the most massive particle in the SM. This grants the top quark some unique properties. Chiefly, unlike the other quarks, the top quark cannot form hadrons as its lifetime is shorter than the time scale of the hadronisation process. The properties of the top quark can, consequently, be probed more directly than other quarks. This thesis explores a rare process involving the top quark predicted by the SM: single top production in association with a Z^{0} boson (tZq). The tZq process is sensitive not only to the couplings of the top quark, but also the couplings between the Z^{0} and $W^{\pm}$ bosons, making an excellent probe of SM predictions in these sectors.

A search for the tZq process in the dilepton channel—i.e. following a leptonic decay of the Z^{0} boson and a hadronic decay of the $W^{\pm}$ boson produced by the decaying top quark—at the CMS experiment is presented. Chapter 2 explores in more detail the SM and top physics specifically. Chapter 3 provides background on the machine learning techniques used in the analysis to extract the tZq signal. The LHC and CMS detector are described in Chapter 4. Chapters 5 and 6 describe the reconstruction and simulation of proton–proton collisions at the CMS detector, and Chapter 7 the selection procedure applied to these events in the tZq analysis. Sources of systematic uncertainty are covered in Chapter 8. Chapter 9 presents the results of the analysis, these are discussed and summarised in Chapter 10.

Chapter 2

Physics Theory

> *There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable.*
>
> *There is another theory which states that this has already happened.*
>
> Douglas Adams

Modern particle physics theory today centres around the Standard Model. This chapter explores this model, both its practical effects and mathematical underpinnings, as well as its known limitations. This concludes with a more in-depth look at the physics of the top quark.

Throughout this thesis a combined system of natural and Heaviside–Lorentz units are used [17]. In this convention, the speed of light, c; the reduced Planck constant, $\hbar$; the Boltzmann constant, k_{B}; the permittivity of free space, ε_0; and the permeability of free space, μ_0, are all set to unity.

This chapter will make use of the notation defined in Appendices A.1 and A.2.

2.1 The Standard Model

Today, the Standard Model (SM) is the accepted theory describing particle physics. It describes of the known elementary particles in the Universe, their interactions via three of the four fundamental forces, and has repeatedly stood up to experiment.

The SM is not without its flaws—examined further in § 2.1.5—but still provides an unrivalled understanding of the fundamental nature of the Universe.

The fundamental particles in the SM are presented in Figure 2.1. There are six quarks and six leptons (plus their twelve corresponding antiparticles), divided into three generations of increasing mass[1]. The first generation are the building blocks of all stable matter in the known Universe. In addition to the fermions, there are the gauge bosons that mediate the fundamental forces: the electromagnetic force is mediated by photons (γ), the weak force by the $W^{\pm}$ and Z^0 bosons (hereafter referred to as the W and Z bosons), and the strong force by gluons (g). Finally, there is the Higgs boson (H), a minimal excitation of the Higgs field through which particles acquire mass (see § 2.1.3).

To each particle in the SM the charge conjugation transformation, C, can be applied, which replaces a particle with its antiparticle. Each fermion has an antiparticle partner[2], the W^+ and W^- bosons are antiparticle partners, and the remaining particles in the SM are their own antiparticles. If a particle meets it own antiparticle they will annihilate each other, releasing energy equal to their combined mass (within a window allowed by the Heisenberg uncertainty principle, discussed in § 2.1.3). Antiparticles have the same mass as their partners but opposite electric charge.

In this thesis, referring to a particle implicitly refers to both itself and its antiparticle partner; for example, the 't' in 'tZq' accounts for t and $\bar{t}$. The exceptions to this are where explicitly stated, where antiparticles are referred to explicitly in a formula (e.g. '$t\bar{t}$'), and where charge is specified (e.g. 'W^+', 'e^-'). Similarly, 'electron' will be used as a shorthand for 'electron or positron'. Finally, 'lepton' will be used to refer to electrons or muons but with the exclusion of the tau lepton and neutrinos (unless explicitly stated otherwise).

2.1.1 Fermions and Bosons

The difference between fermions and bosons is manifested in *spin*. Spin, S, is an intrinsic property of all elementary particles akin to classical angular momentum. The magnitude of S is described by the spin quantum number, s:

$$|S| = \sqrt{s(s+1)} \quad s \in \frac{\mathbb{N}}{2}; \tag{2.1}$$

where for bosons $s \in \mathbb{N}$ and for fermions $s \in \mathbb{N} + \frac{1}{2}$. The consequences of this can be understood through their representations in quantum theory. A system of

---

[1] In the neutrino sector it is not yet known if the generations follow a normal or inverted hierarchy.

[2] Neutrinos may be their own antiparticle partner [18].

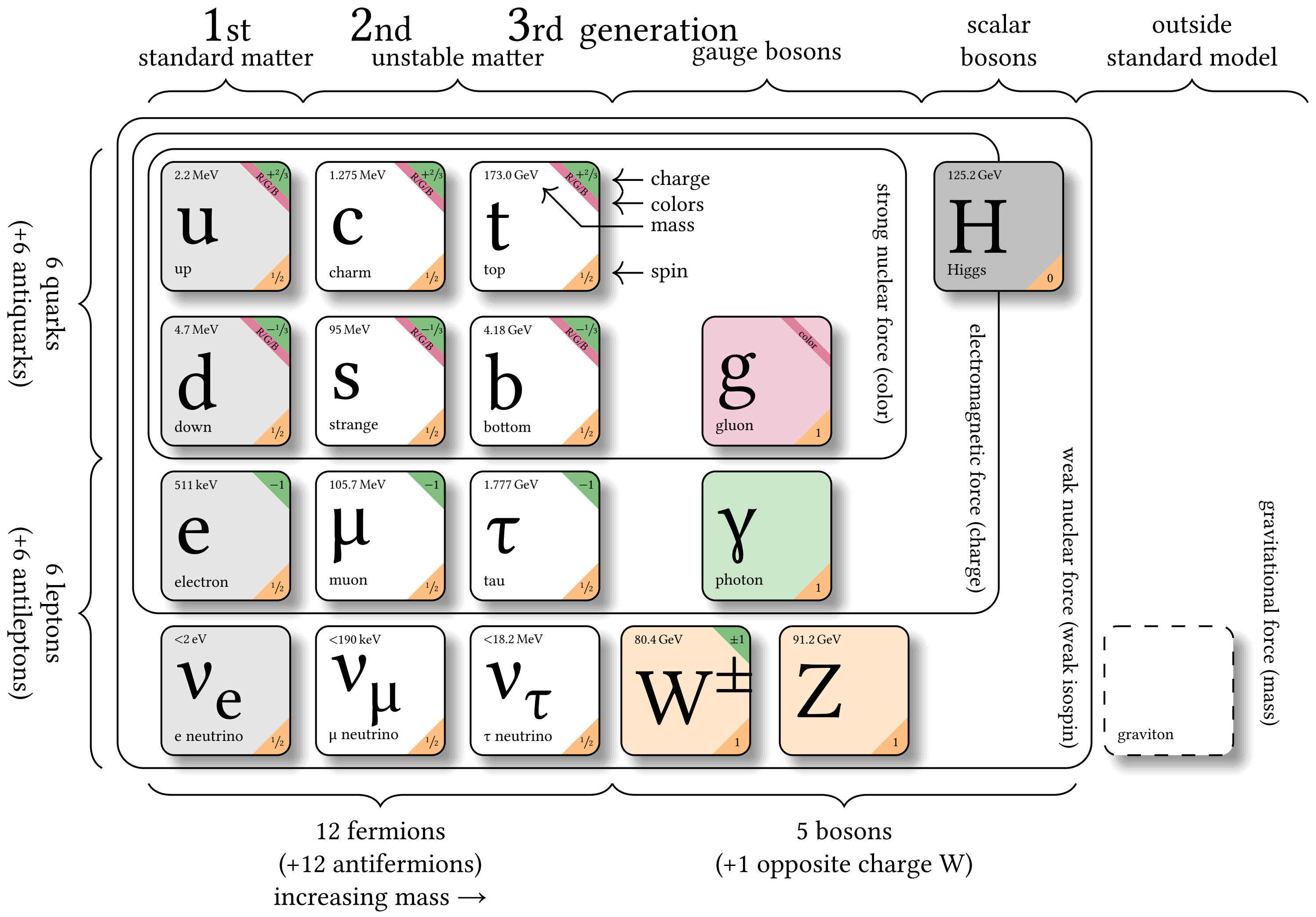


Figure 2.1: Fundamental particles in the SM. Mass values taken from the Particle Data Group [16]. Figure taken from [19].

n particles can be described by a wavefunction

$$\psi = \psi(x_1, x_2, ..., x_n). \tag{2.2}$$

The spin-statistics theorem [20] states that if x_i are particles with integer s (i.e. bosons), swapping two particles in the system will have no effect on the wavefunction:

$$\psi(x_1, x_2, ..., x_n) = \psi(x_2, x_1, ..., x_n). \tag{2.3}$$

However, if x_i are fermions

$$\psi(x_1, x_2, ..., x_n) = -\psi(x_2, x_1, ..., x_n). \tag{2.4}$$

From this it can be observed that if two fermions have the same quantum state, $x_1 = x_2 = x'$, then

$$\psi(x', x', x_3, ..., x_n) = -\psi(x', x', x_3, ..., x_n) \tag{2.5}$$

$$\Rightarrow \psi = 0. \tag{2.6}$$

This necessitates the *Pauli exclusion principle* [20], which forbids two fermions from possessing the same quantum state. Bosons suffer no such restriction and consequently the energy distributions of fermions and bosons in systems are described using two different statistical models: Bose–Einstein statistics for bosons and Fermi–Dirac statistics for fermions [21].

2.1.2 Gauge Theory

Gauge theories operate on the *Lagrangian* of a system,

$$L = T - V, \tag{2.7}$$

where T is the system's kinetic energy and V the system's potential energy [21]. A gauge theory is formed by requiring the Lagrangian to be invariant under local transformations (transformations with a dependence on the spacetime coordinate [22]).

A simple example is the *Lagrangian field density*, $\mathcal{L}$, of a fermion field [22]

$$\mathcal{L} = \bar{\psi}(x)(\mathrm{i}\partial\!\!\!/ - m)\psi(x) \tag{2.8}$$

when requiring that it is invariant under the introduction of a spacetime dependent arbitrary phase

$$\psi \mapsto \psi' = \mathrm{e}^{-\mathrm{i}f(x)}\psi. \tag{2.9}$$

This is consistent with our understanding of wavefunctions: such changes of phase should be unobservable as only alterations of the magnitude should alter the probability density, $|\psi|^2$.

A gauge theory is defined by a *Lie group* describing the transformations the Lagrangian should be invariant under, known as the *gauge group*. A *group* is defined by an underlying set, G, and a binary operator (the *group law*), $\circ$, denoted as $(G, \circ)$. These form a group if and only if the four *group axioms* are satisfied:

$$\textbf{Closure} \qquad \forall a, b \in G, a \circ b \in G \tag{2.10}$$

$$\textbf{Associativity} \qquad \forall a, b, c \in G, (a \circ b) \circ c = a \circ (b \circ c) \tag{2.11}$$

$$\textbf{Identity} \qquad \exists e \in G \, : \, \forall a \in G, e \circ a = a \circ e = a \tag{2.12}$$

$$\textbf{Inverse} \qquad \exists a^{-1} \in G \, : \, \forall a \in G, a^{-1} \circ a = a \circ a^{-1} = e. \tag{2.13}$$

If and only if a fifth axiom,

$$\textbf{Commutativity} \qquad \forall a, b \in G, a \circ b = b \circ a \tag{2.14}$$

is satisfied, the group is said to be *Abelian* [22]. Furthermore, to qualify as a Lie group G must be a *differentiable manifold*, satisfied if and only if the members of G can be described by a finite set of real parameters. An example of a Lie group is the set of all real invertible 2×2 matrices under multiplication: this satisfies the group axioms and every matrix can be described by four continuous real parameters. The arbitrary phase $\mathrm{e}^{-\mathrm{i}f(x)}$ introduced in Equation 2.9 is a group under multiplication, known as U(1) [23].

A group can defined by a *generating set*. generating set, S, of a group $(G, \circ)$ is a subset of G such that any element in G can be expressed as a finite series of combinations under the group law of elements in S and their inverses [24]. To demonstrate, a generating set of the group $(\mathbb{Z}, +)$ is $\{1\}$ as every integer can be represented as either a sum of 1s or -1s (where -1 is the inverse of 1 under addition). When discussing gauge theories, it is usually helpful to only study generating sets of minimal cardinality and so this will be assumed henceforth.

Two groups, e.g. $(G, \circ)$ and $(H, \bullet)$, can be combined into a new group through the direct product operation, $\times$:

$$G \times H = \big(\{(g, h) \, : \, g \in G, h \in H\}, *\big) \tag{2.15}$$

where

$$(g, h) * (g', h') = (g \circ g', h \bullet h'). \tag{2.16}$$

This can be used to represent the gauge group of a gauge theory in which $\mathcal{L}$ must be invariant under multiple transformations [22].

Unmodified, the Lagrangian field density in Equation 2.8 does not remain invariant under a phase shift:

$$\mathcal{L} \mapsto \mathcal{L}' = \mathcal{L} + \bar{\psi}\left(\mathrm{i}\not{\partial} f(x)\right)\psi(x) \tag{2.17}$$

To enforce invariance a *gauge field*, A_μ, must be added to $\mathcal{L}$, transforming as

$$A_\mu \mapsto A'_\mu = A_\mu(x) + \frac{1}{g}\partial_\mu f(x) \tag{2.18}$$

where g is a constant *gauge coupling parameter.* This field can be implanted into Equation 2.8 using the *gauge covariant derivative*, D_μ [22]:

$$D_\mu = \partial_\mu - \mathrm{i}gA_\mu \tag{2.19}$$

$$\mathcal{L} = \bar{\psi}(x)(\mathrm{i}\not{D} - m)\psi(x). \tag{2.20}$$

Modifying $\mathcal{L}$ in this way introduces a new interaction, which can be interpreted as the fermions represented by ψ interacting with minimal excitations of the A_μ vector field. These excitations are the force mediating bosons.

Finally, a term describing the field itself is added to $\mathcal{L}$. This takes the form $\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ where [22]

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu + \partial_\nu A^a_\mu + gf_{ab}{}^{c}A^b_\mu A^c_\nu \tag{2.21}$$

is the *field strength tensor.* The *structure constants*, $f_{ab}{}^{c}$, are defined by

$$T_aT_b - T_bT_a = \sum_c f_{ab}{}^{c}T_c \tag{2.22}$$

where T_i are members of the generating set of the Lie group. For each generator there is a gauge boson. The U(1) group has only one generator and is Abelian, ergo the single structure constant is zero. Equation 2.21 can therefore be rewritten as

$$F_{\mu\nu} = \partial_\mu A_\nu + \partial_\nu A_\mu. \tag{2.23}$$

The U(1) symmetry introduced here is the origin of the electromagnetic force; the introduced vector field, A_μ, is the electromagnetic four-potential $A^\mu = (V, \boldsymbol{A})$ associated with a single gauge boson, the photon; and the field strength tensor $F_{\mu\nu}$ encodes Maxwell's equations and correctly predicts that photons do not self-interact. Moreover, Noether's first theorem states that for every differentiable local symmetry there exists a conserved current [25]. In the case of the U(1) symmetry of electromagnetism, this is the electric charge, q.

The power of gauge theories should now be self-evident: electromagnetism has been composed purely from the requirement that $\mathcal{L}$ should be invariant under a spacetime dependent phase shift of ψ. Gauge theories applied to the description of quantum fields are known as *quantum field theories* (QFTs). The SM itself is a gauge theory with a gauge group of SU(3) × SU(2) × U(1) acting on:

- fermion fields, ψ
- electroweak boson fields, W_1, W_2, W_3, and B
- the gluon field, G_a
- the Higgs field, ϕ

where SU(n) is the Lie group of $n \times n$ matrices with unit determinant [22].

2.1.3 Electroweak Theory

Quantum electrodynamics

As previously described, the electromagnetic interaction is a gauge theory with a U(1) symmetry. This Abelian gauge theory is known as *quantum electrodynamics* (QED) and describes a force mediated by massless, chargeless, non–self-interacting bosons (photons) that preserves electric charge. As photons are massless, the electromagnetic force has an unlimited range. The intrinsic strength of the electromagnetic force is given by the *fine structure constant*, $\alpha = \frac{e^2}{4\pi} = 7.30 \times 10^{-3}$ [16] at zero energy. This is related to the gauge coupling parameter by

$$\alpha = \frac{g^2}{4\pi}. \tag{2.24}$$

Measures of intrinsic strength can be constructed similarly for other QFTs.

Heisenberg's uncertainty principle sets an insuperable limit on the precision to which certain properties of a particle can be simultaneously measured. One possible statement of this fundamental law is [26]

$$\Delta E \Delta t > \frac{1}{2}, \tag{2.25}$$

permitting the creation of ephemeral virtual particle–antiparticle pairs in the vacuum. The fields of virtual $\mathrm{e}^+\mathrm{e}^-$pairs created in this way interact with those of real particles, causing the vacuum to act as a dielectric medium. This effectively screens the *bare* charges of particles, with closer charges experiencing less screening. This is known as the *running* of α, where α increases with charge proximity and hence energy scale [18].

Quantum flavourdynamics

The QFT describing the weak force, sometimes referred to as *quantum flavourdynamics* (QFD), is based on an SU(2) symmetry. The weak force acts on fermions and couples to the projection of weak isospin along the z axis, T_3 [22].

Chirality determines the value of T_3 for a fermion. A spinor field can be decomposed into its *left-handed*, ψ^{L}, and *right-handed*, ψ^{R}, components by [22]

$$\psi^{\mathrm{L}} = \frac{1}{2}\left(1 - \gamma^5\right)\psi \qquad\qquad \psi^{\mathrm{R}} = \frac{1}{2}\left(1 + \gamma^5\right)\psi. \tag{2.26}$$

Left-handed fermions have $T_3 = \pm\frac{1}{2}$, while right-handed fermions have $T_3 = 0$ (and vice versa for antifermions), so only the former can couple to $\mathrm{W}^{\pm}$ bosons ($T_3 = \pm 1$). All fermions couple to the Z boson ($T_3 = 0$). The preference shown for certain chiralities by the weak nuclear force allows it to violate P symmetry: i.e. a system transformed under the parity transformation,

$$\mathrm{P} : \boldsymbol{r} \mapsto -\boldsymbol{r}, \tag{2.27}$$

behaves differently with respect to weak interactions as the chralities of the particles involved are inverted [27]. Uniquely among the fundamental forces, the weak force is capable—via couplings to the W boson only—of changing the flavour of a quark.

The mediating W and Z bosons of the weak force are massive, and so its range must be limited as a consequence of the uncertainty principle (Equation 2.25). The coupling constant of the weak nuclear force, α_{W}, is $\sim 10^{-6}$ at ~ 1 GeV [28].

Electroweak unification and the Higgs mechanism

Electroweak theory (EWT) is a gauge theory that describes both the electromagnetic and weak forces with a SU(2) × U(1) group symmetry [29, 30, 31]. It supposes that the weak and electromagnetic interactions become unified above some *electroweak unification energy*.

The U(1) group has a single generator and SU(2) has three generators, therefore there are four boson fields: W_1, W_2, W_3, and B. The B field acts on *weak hypercharge* [22],

$$Y = 2(q - T_3), \tag{2.28}$$

and is associated with the U(1) group. It is therefore non–self-interacting. The W_1, W_2, and W_3 fields couple to T_3 and self-interact.

The bosons described by the electroweak fields are massless, but in reality the W and Z bosons that mediate the weak force have mass. These bosons gain their mass via the *Higgs mechanism* [32, 33, 34, 35, 36, 37].

The Higgs mechanism presumes the existence of the *Higgs field*, a complex scalar field, ϕ, defined across all spacetime. This field's potential, $V(\phi)$, is constructed such that it has a local maximum at the vacuum expectation value of ϕ and an infinitely degenerate ground state forming the "Mexican hat" shape [38] seen in Figure 2.2. Initially in the Universe the Higgs field occupied the symmetric ground

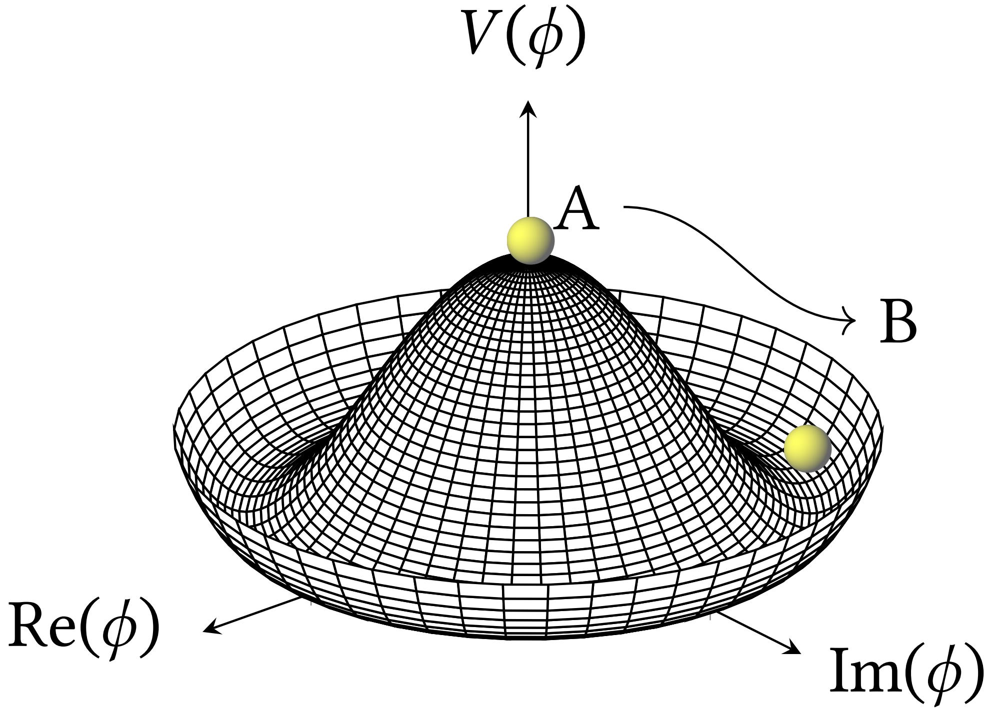


Figure 2.2: The variation of the Higgs field's potential with its imaginary and real components, exhibiting the "Mexican hat" shape.

state ('A' in Figure 2.2) and the electroweak force was mediated by massless bosons. As the Universe cooled, however, the Higgs field underwent *spontaneous symmetry breaking* when a ground state was chosen ('B' in Figure 2.2). When this occurred, three of the four degrees of freedom provided mass terms to the electroweak bosons with the remainder forming a scalar boson representing a minimal excitation of the Higgs field itself: the Higgs boson. The weak force and electromagnetism then manifested as distinct phenomena.

Consequently, the physically observable W bosons, Z boson, and photon are—because of the Higgs mechanism— mixtures of the electroweak boson fields defined by the Weinberg angle, $\cos\left(\theta_{\mathrm{W}}\right) = \frac{m_{\mathrm{W}}}{m_{\mathrm{Z}}}$ [22]:

$$A_\mu = \sin\left(\theta_{\mathrm{W}}\right) W_\mu^3 + \cos\left(\theta_{\mathrm{W}}\right) B_\mu \tag{2.29}$$

$$W_\mu^\pm = \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp \mathrm{i} W_\mu^2\right) \tag{2.30}$$

$$Z_\mu = \cos\left(\theta_{\mathrm{W}}\right) W_\mu^3 - \sin\left(\theta_{\mathrm{W}}\right) B_\mu \tag{2.31}$$

Finally, it is through interactions with the Higgs field that fermions in the SM acquire mass [22, 39].

Mass and flavour eigenstates

For fermions in the SM it is important to distinguish between mass eigenstates, which propagate through free space, and flavour eigenstates, which are involved in interactions. The mixing of these states for quarks is described by the *Cabibbo–Kobayashi–Maskawa Matrix* (CKM Matrix) [40, 41]

$$\begin{pmatrix} \mathrm{d}' \\ \mathrm{s}' \\ \mathrm{b}' \end{pmatrix} = V_{\mathrm{CKM}} \begin{pmatrix} \mathrm{d} \\ \mathrm{s} \\ \mathrm{b} \end{pmatrix} \tag{2.32}$$

$$V_{\mathrm{CKM}} = \begin{pmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{pmatrix} \tag{2.33}$$

where $\mathrm{d}', \mathrm{s}', \mathrm{b}'$ are the quark flavour eigenstates; $\mathrm{d}, \mathrm{s}, \mathrm{b}$ are the quark mass eigenstates; V_{CKM} is the CKM Matrix; and $\left|V_{\mathrm{qq}'}\right|^2$ can be interpreted as the branching ratio of $\mathrm{q} \to \mathrm{Wq}'$. It is common to represent V_{CKM} in terms of the *Wolfenstein parameters*

$$V_{\mathrm{CKM}} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3\left(\bar{\rho} - \mathrm{i}\bar{\eta}\right) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3\left(1 - \bar{\rho} - \mathrm{i}\bar{\eta}\right) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}\left(\lambda^4\right) \tag{2.34}$$

where, from experiment [16],

$$\begin{aligned} \lambda &= 0.22453 \pm 0.00044 & A &= 0.836 \pm 0.015 \\ \bar{\rho} &= 0.122^{+0.018}_{-0.017} & \bar{\eta} &= 0.355^{+0.012}_{-0.011}. \end{aligned} \tag{2.35}$$

These values mean that V_{CKM} is consistent with being a unitary matrix, as required by the SM. A similar matrix, the *Pontecorvo–Maki–Nakagawa–Sakata Matrix* (PMNS Matrix) [42, 43], exists to relate the flavour and mass eigenstates of neutrinos. Charged leptons (including the tau lepton), meanwhile, have mass and flavour eigenstates that coincide exactly.

2.1.4 The Strong Nuclear Force

The QFT describing the strong force is known as *quantum chromodynamics* (QCD); it is based on the SU(3) group and describes the gluon field, G_a [44]. The SU(3) group is non-Abelian and has a generating set of cardinality eight [45]. The structure constants are therefore nonzero and so, in accordance with Equation 2.21, G_a may self-interact.

Colour charge is the conserved charge in QCD. Unlike electric charge and the projection of weak isospin, colour charge is not positive or negative at some magnitude. Instead, an object carries some combination of red, green, blue, antired, antigreen, and antiblue colour charge. A quark can be either red, green, or blue and an antiquark can be antired, antigreen, or antiblue; a gluon carries a single colour and a single anticolour. These are the only objects in the SM with colour charge and consequently the only objects that experience strong interactions. Coloured objects can combine to form objects with no overall colour charge—known as *colourless* objects—in three ways: by combining a colour and its anticolour; by combining red, green, and blue; and by combining antired, antigreen, and antiblue.

A consequence of gluon self-coupling is that the coupling constant of the strong force, α_s, increases with distance. The cloud of virtual gluons separating two colour charges, unlike the cloud of virtual photons separating two electric charges, has the effect of increasing the effective colour charge. This effect is known as *asymptotic freedom* [18, 22].

Quarks bound via the strong nuclear force (mediated by gluons, with gluons and quarks referred to collectively as *partons*) form composite particles known as *hadrons*. Within a hadron the small distances—and hence low α_s—involved mean that quarks behave approximately as free particles. As the distance between quarks increases so does the potential energy, until it is more energetically favourable to produce quarks from the vacuum to create new confined states. This results in the phenomena of *colour confinement* [18, 22]: the increase of the potential with distance means that no object of net colour charge can exist. Traditionally, this restricted quarks to either exist within baryons (qqq), antibaryons ($\bar{\mathrm{q}}\bar{\mathrm{q}}\bar{\mathrm{q}}$), or mesons ($\mathrm{q}\bar{\mathrm{q}}$), but recent measurements have discovered tetraquark ($\mathrm{qq}\bar{\mathrm{q}}\bar{\mathrm{q}}$) [46, 47, 48, 49, 50] and pentaquark ($\mathrm{qqqq}\bar{\mathrm{q}}$ or $\mathrm{q}\bar{\mathrm{q}}\bar{\mathrm{q}}\bar{\mathrm{q}}\bar{\mathrm{q}}$) [51] states.

The concept of a three-quark baryon or two-quark meson is an oversimplification. When referring to a proton as consisting of two up quarks and one down quark, or a K^- meson as consisting of an up quark and an antistrange quark, this refers to the *valence* quarks. With a low-energy probe, a hadron will appear to consist of just its valence quarks [18]. Higher-energy probes reveal the existence of the so-called *sea* quarks and gluons from the hadron's G_a field. *Parton distribution functions* (PDFs) are used to describe the number density of parton flavours within a given hadron at a given energy scale. These number densities are presented as a function of *Bjorken x*, the fraction of the hadron's momentum carried by a specified parton. While the evolution of PDFs with energy scale can be predicted by the *Dokshitzer–Gribov–Lipatov–Altarelli–Parisi Equations* (DGLAP

Equations) [52, 53, 54], the dependence on Bjorken x must be determined experimentally; their accuracy is of foremost importance if hadron collisions are to be described correctly.

In high-energy collisions it is possible, and in fact desirable, to impart enough energy to a parton such that it is liberated from its confined state. As a consequence of colour confinement, the liberated parton forms a new confined state in a process known as *hadronisation*. This process can cascade to form showers of collimated particles known as *jets*. The behaviour of quarks inside a hadron can be approximated using perturbation theory but typically ignored higher-order terms become significant during hadronisation. No single model simultaneously describes the perturbative, high-energy *parton showering* directly following the hard scatter and the nonperturbative, low-energy hadronisation process. Rather, different models are used for each with a *matching algorithm* bridging the gap and ensuring the models are consistent (this is discussed further in § 6.1). The most prevalent models used to describe hadronisation are the *Cluster Model* [55] and *Lund String Model* [56].

2.1.5 Limitations of the Standard Model

Despite the SM's successes there remain phenomena that it is unable to explain. This section will briefly discuss some key outstanding issues of the SM that have motivated the search for BSM physics.

Gravity

Chief among the phenomena not covered by the SM is gravity, the fourth fundamental force. Gravity is currently understood through the lens of *general relativity*, which has so far been irreconcilable with QFT [57, 58]. The theoretical boson that mediates gravity has been dubbed the graviton, but it has never been experimentally observed and no widely accepted QFT describes it.

Mass is seemingly not quantised, and so the value of the gravitational coupling constant α_g depends on the masses used to define it. Using the proton mass [28],

$$\alpha_\mathrm{g} = G m_\mathrm{p}^2 \sim 10^{-39}. \tag{2.36}$$

The extreme discrepancy between the strength of Gravity and the weakest of the other fundamental forces (the weak force) is inauspicious to the unification of gravity with the other fundamental forces, and is known as the *hierarchy problem*.

Baryogenesis

It is assumed that in the creation of the Universe matter and antimatter were created in equal amounts. Some unknown process—known as *baryogenesis*—then occurred leading to the current state of *baryonic asymmetry* where more matter than antimatter exists in the observable Universe [59].

According to the *Sakharov conditions* [60], which describe the necessary circumstances for interactions to produce matter and antimatter at different rates, there must be interactions that violate *CP-symmetry*. The CP transform is the combination of the parity (Equation 2.27) and charge conjugation transforms. CP violation is allowed weak interactions, represented via the complex phase $i\bar{\eta}$ in Equation 2.32 [18]. Experimentally, CP violation has been observed in strange quark [61, 62], bottom quark [63, 64], and more recently charm quark decays [65].

Though CP violation is allowed and has been observed in the SM, it is not to a degree sufficient to describe the observed baryonic asymmetry in the Universe [18]. What is more, the Sakharov conditions require violations in baryon number conservation, which is upheld in SM interactions.

Dark matter

Most (but not all [66, 67]) galaxies contain *dark matter*, some form of matter that interacts gravitationally but is otherwise inert. Its existence can be deduced from galactic rotation curves that differ from those predicted by Newton's law of universal gravitation based solely on the observable matter in a galaxy [68, 69, 70]. This form of matter accounts for as much as 85% of matter in the Universe [18] and has no clear candidate for its constituents in the SM.

Dark energy

Since the Big Bang the Universe has been expanding; observations indicate that the rate of this expansion is accelerating [71]. A popular theory to explain this introduces *dark energy*: a form of energy permeating all space and contributing to the Universe's expansion [72]. This cosmological presence has no implementation in the SM.

2.2 Physics of the Top Quark

The first experimental evidence for a third generation of matter came in 1975 with the discovery of the tau lepton [11]. Maintaining symmetry in the quark and lepton sectors implied the existence of another quark pair, and the discovery of the bottom quark followed shortly after in 1977 [73]. The top quark, however, would remain elusive for nearly two decades.

The primary difficulty with discovering the top quark was its mass: the top quark is the most massive particle in the SM, with $m_{\mathrm{t}} = 173.0 \pm 0.4\,\mathrm{GeV}$ [16]. This formidable mass means that only the most powerful colliders are capable of creating top quarks, colliders that did not exist when the search for the top quark began in earnest. It was only after the Tevatron at Fermilab—the most powerful collider built until that point—was operational that the existence of the top quark was experimentally confirmed [74, 75].

The extreme mass of the top quark confers some novel properties. For instance, the large mass begets a short lifetime, approximately $5 \times 10^{-25}\,\mathrm{s}$ [76]. This is significantly shorter than the timescale of hadronisation and consequently a top quark will decay before it forms hadrons. Uniquely, this presents a window to examine the behaviour of a quark directly, rather than from within a confined state.

The mass of the top quark also grants it an important place within the SM. It has the largest coupling to the Higgs field—which couples to mass—and is an important parameter in fits determining the precise form of the Higgs field potential [77]. This includes determining whether there exists a lower-energy vacuum state the Universe could *topple* into, i.e. that we currently exist in a *false vacuum* [78].

Another exceptional property of the top quark is its decay. The branching ratio of $\mathrm{t} \to \mathrm{Wb}$ is, as can be seen in Equation 2.34, almost 100%. For the purposes of physics analysis it is usually taken as such (including within this thesis). Current measurements give the ratio $\mathscr{B}(\mathrm{t} \to \mathrm{Wb})/\mathscr{B}(\mathrm{t} \to \mathrm{Wq})$ where $\mathrm{q} = \mathrm{d}, \mathrm{s}, \mathrm{b}$ as, more precisely, 0.957 ± 0.034 [16].

2.2.1 Top Quark Pair Production

The leading top quark production method in high-energy collisions is through strong interactions; the strong force conserves quark flavour, so this must involve the creation of $\mathrm{t\bar{t}}$ pairs. Quark–antiquark annihilation and gluon fusion are the $\mathrm{t\bar{t}}$ production mechanisms at *leading order* (LO), with the associated Feynman diagrams shown in Figure 2.3. At the Tevatron, quark–antiquark annihilation was the dominant $\mathrm{t\bar{t}}$ production mode but this is not the case at the LHC. Collisions at the LHC take place at a greater centre-of-mass energy than at the Tevatron, and the PDFs demonstrate that the Bjorken x of gluons increases with energy scale. Furthermore, the LHC is a proton–proton rather than proton–antiproton collider, so antiquarks involved in annihilation must arise from the quark sea. Because of these factors, it is the gluon fusion mode that dominates at the LHC with ≈90% of $\mathrm{t\bar{t}}$ pairs produced in this manner at $\sqrt{s} = 13\,\mathrm{TeV}$ [16].

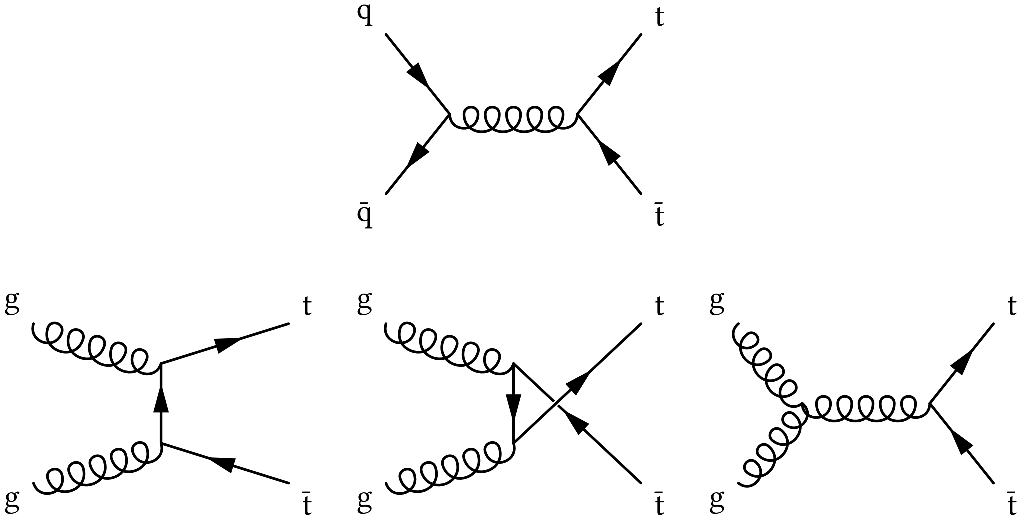


Figure 2.3: LO Feynman diagrams for $t\bar{t}$ production via quark–antiquark annihilation (top) and gluon fusion (bottom).

A W boson can either decay hadronically as $\mathscr{B}(W \to qq') = 32.72 \pm 0.30\%$ or leptonically as $\mathscr{B}(W \to \ell\nu) = 67.60 \pm 0.27\%$ [16]. Based on the W boson decay channel, three different $t\bar{t}$ decay modes are so defined:

Hadronic Both W bosons decay hadronically, probability $45.44 \pm 0.36\%$

Semileptonic One W boson decays hadronically and the other leptonically, probability $44.11 \pm 0.45\%$

Dileptonic Both W bosons decay leptonically, probability $10.71 \pm 0.20\%$

An alternative $t\bar{t}$ production method is in association with a vector or Higgs boson: $t\bar{t}W$, $t\bar{t}Z$, or $t\bar{t}H$. LO Feynman diagrams are shown in Figures 2.4 and 2.5. Inclusion of the small tW/tZ/tH couplings and additional particles in the final state means that $t\bar{t}$ production in association with a boson is orders of magnitude rarer than $t\bar{t}$ production. Despite this, measurements of the $t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}H$ cross sections have all been made at the LHC and in every case were found to be compatible with the SM with a confidence exceeding 5σ [79, 80].

2.2.2 Single Top Quark Production

Alongside $t\bar{t}$ production, top quarks can be produced on their own via weak interactions. As weak processes, the single top production modes naturally have smaller—but nevertheless detectable—cross sections. Three modes at LO exist:

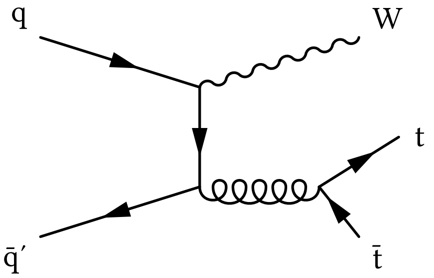


Figure 2.4: LO Feynman diagram for tt̄W production. The W boson must be produced as ISR.

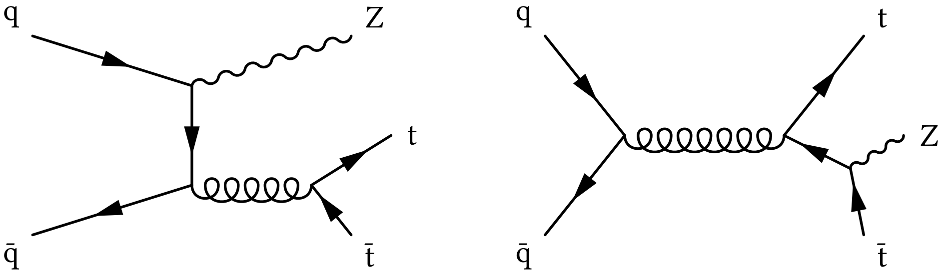


Figure 2.5: LO Feynman diagrams for tt̄Z production, showing how the Z boson can be produced as part of the ISR (left) or FSR (right). The case of Z boson production in the FSR is shown for quark–antiquark annihilation, but any tt̄ production diagram in Figure 2.3 may be substituted. The Feynman diagrams for tt̄H production are identical but for the radiation of a Higgs boson in place of the Z boson.

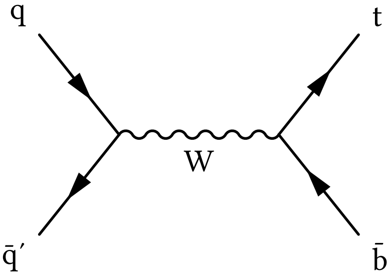


Figure 2.6: Feynman diagram for s-channel single top production.

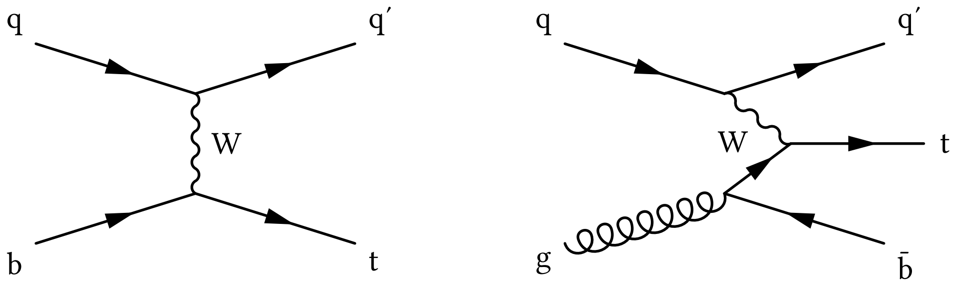


Figure 2.7: Feynman diagrams for t-channel single top production. The left diagram shows the process in the 5FS where the bottom quark has a PDF and can so exist in the initial state. The right diagram shows t-channel production in the case of the bottom quark originating from gluon splitting.

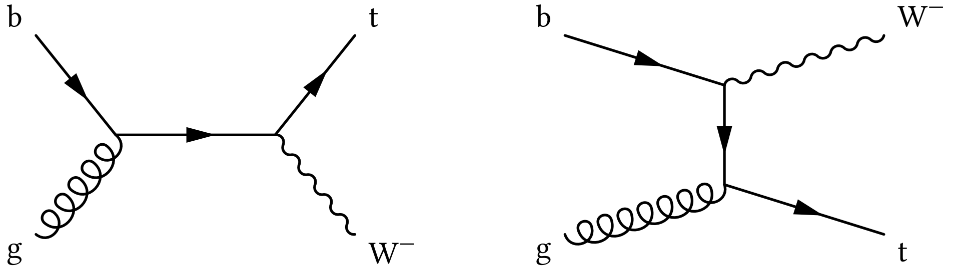


Figure 2.8: Feynman diagrams for single top production in association with a W boson.

s-channel production, t-channel production, and tW production. Feynman diagrams are shown for each in Figures 2.6 to 2.8.

The s-channel is the rarest production mode at the LHC. Like the quark–antiquark annihilation mode of $t\bar{t}$ production at the LHC, the s-channel requires an antiquark from the sea in the initial state. The requirement of an off-shell W boson reduces the cross section further and, combined with a final state difficult to distinguish from major backgrounds, leaves the s-channel as the only single top production mode yet unobserved at the LHC [81]. However, it has been observed at the Tevatron [82].

The predominant production mode at the LHC is the t-channel. It was initially observed at the Tevatron [83] and, subsequently, the LHC [84, 85]. In Figure 2.7 two distinct production methods are shown: the first in which the initial bottom quark arises from the quark sea, and the second where it originates from $g \rightarrow b\bar{b}$ splitting. When determining PDFs, it is assumed that quarks are massless entities. This is a reasonable approximation for most flavours, but begins to falter at energy scales of the order of the nominal bottom quark mass. This has given rise to two schools of PDF calculation: the *Five-Flavour Scheme* (5FS) where the bottom quark is treated as massless and so has its own PDF, and the *Four-Flavour Scheme* (4FS) where the bottom quark is treated as massive and so does not have an PDF [86]. In the 4FS the initial bottom quark in t-channel production must be the result of gluon splitting. In either case, the increased Bjorken x of gluons at the LHC allows the cross section of t-channel production to scale favourably.

Unlike the others, the final production mode, tW, was initially discovered at the LHC [87]. Its contribution at energy scales achieved at the Tevatron is negligible.

A firm understanding of single top production is important because:

- Single top processes allow the tWb vertex and hence V_{tb} to be probed, which is not possible in $\mathrm{t\bar{t}W}$ as the W boson is only produced in the *initial-state radiation* (ISR).
- Backgrounds for $\mathrm{t\bar{t}}$ and potential BSM processes include single top production as irreducible backgrounds.
- PDFs can be constrained by measurements of $\mathrm{t\bar{t}}$ and single top processes [88].

2.2.3 Single Top Quark Production in Association with a Z Boson

The study in this thesis concerns the production of a single top quark in association with a Z boson and an additional jet (tZq). The Feynman diagrams in Figure 2.9 show how the Z boson can be produced via radiation from any of the quark legs or through W boson fusion, making tZq sensitive to the tZ, tWb, and WWZ couplings. As such, tZq promises to be a sterling probe of SM predictions.

The more obvious choice to study the tZ vertex would be the $\mathrm{t\bar{t}Z}$ process, but at LHC collision energies tZq has a greater cross section [89]. Since $\mathrm{t\bar{t}Z}$ has been observed at the LHC [90, 91], the same may be possible for tZq.

The tZq process is an irreducible background to the yet-unobserved tH process, and so understanding it is of importance to those studies. It also forms a background for some potential BSM processes, including *flavour changing neutral currents* (FCNCs). An FCNC is an interaction that changes the flavour but not the charge of a fermion, e.g. t → Zc. This is forbidden at the tree level in the SM and so any evidence would pave the way for BSM studies [92].

Four different decay channels exist for tZq, defined by the decay modes of the W and Z bosons:

Trileptonic Both the W and Z bosons decay leptonically, giving three leptons in the final state.

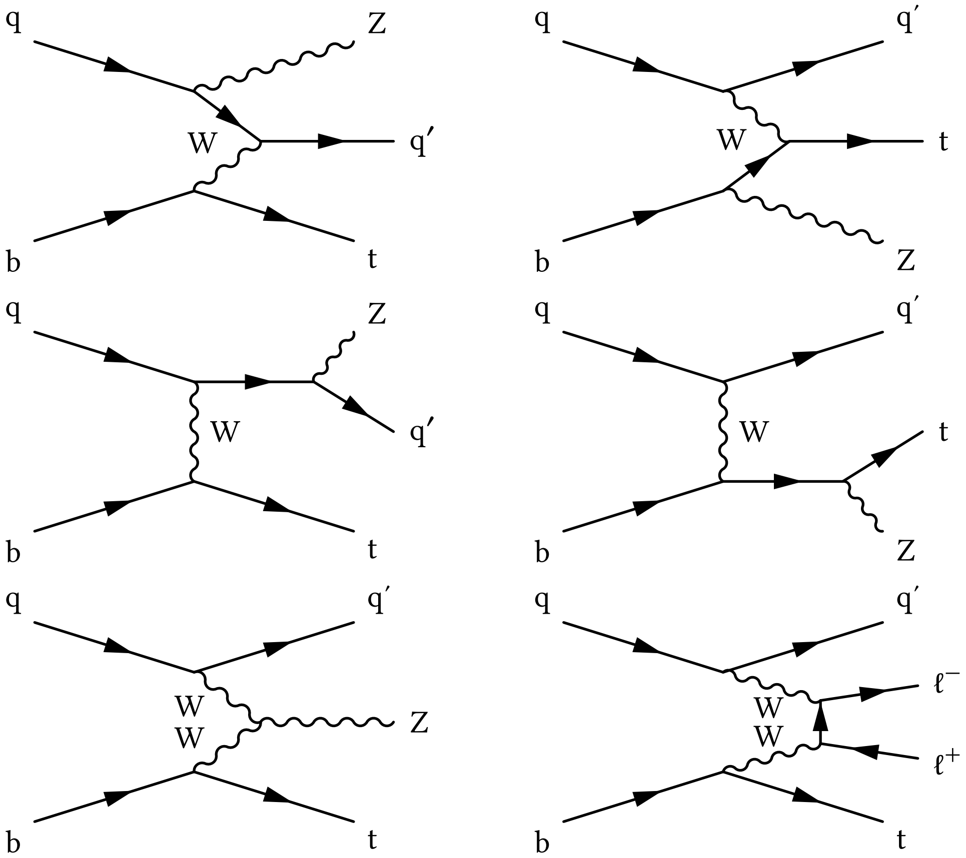


Figure 2.9: LO Feynman diagrams for tZq production. The nonresonant contribution is shown in the bottom right.

Dileptonic The Z boson decays into a $\ell^+\ell^-$ pair and the W boson decays hadronically.

Single lepton The W boson decays into a lepton and neutrino; the Z boson decays hadronically.

Hadronic Both the W and Z bosons decay hadronically ergo there are no leptons in the final state.

Despite having the smallest cross section, previous studies of tZq have sought the trilepton final state as it is the easiest to distinguish from background processes. These previous searches are discussed further in § 10.1. The analysis in this thesis searches for the dileptonic tZq channel. If successful, this would be the first observation of tZq outside the trilepton channel.

Chapter 3

Machine Learning Techniques

Trust the process

Tony Wroten

In its broadest sense, *machine learning* (ML) is a set of methods for data analysis that are able to identify patterns in data. It is an extremely wide and active field, with this chapter highlighting only a small part of the domain. The two problems tackled with ML techniques in this thesis—*classification* and *regression*—involve learning some mapping from a set of inputs X to a set of outputs $\boldsymbol{y}$ using *training input*. The learned mapping will then be used to make predictions about previously unseen input (input not included in the training set). ML methods used to solve problems of this type are known as *supervised learning* [93]. Three such techniques will be examined in this chapter: BDTs and *multilayer perceptrons* (MLPs) for classification, and *Gaussian processes* (GPs) for regression.

3.1 Classification

The task of a classifier is to assign category labels to individual observations. For this thesis, only the case of *binary classification* is considered, where the category can be positive or negative. In the parlance of particle physics—used for the rest of this chapter—observations are events, the positive category is signal events, and the negative category is background events.

A classifier does not typically assign a label directly but rather a probability that a given event is a signal event. This value is known as the *response* [93]. A *threshold* can then be freely set between 0 and 1, above which an event will be classified as signal. It is the distribution of the response, rather than class labels, that will ultimately be used to extract the tZq signal (see § 9.4).

To classify an event, a classifier is given a set of individual measurable properties describing each event. These are known as *features*. The classifier will then be *trained* using training input where the correct label is known, searching for patterns in these features that can be exploited to distinguish between signal and background. The power of ML is that these patterns are found automatically.

To train a classifier, an *objective function* is minimised [93]. By defining an objective function that is lower-valued for some definition of a "better" classifier—a so-called *cost function*—the minimisation improves the classifier. A cost function, C, takes the form:

$$C = L + \Omega \tag{3.1}$$

where L is the *loss term* and Ω the *regularisation term*.

The *loss* of an individual event describes how poorly it is described by a classifier and is defined by a *loss function*. In binary categorisation the loss function is typically the *binary cross entropy* [93, 94],

$$l(y, \hat{y}) = (y - 1)\log(1 - \hat{y}) - y\log(\hat{y}), \tag{3.2}$$

where $0 < \hat{y} < 1$ is the response of the classifier and y is the true label, that is,

$$y = \begin{cases} 1 & \text{for signal events} \\ 0 & \text{otherwise} \end{cases} . \tag{3.3}$$

The loss is smaller if the classifier makes a confident (i.e. the response is close to 1 or 0), correct prediction

$$\lim_{\hat{y}\to y} l = 0 \tag{3.4}$$

and greater if an incorrect, yet confident prediction is made

$$\lim_{\hat{y}\to 1} l(0, \hat{y}) = \lim_{\hat{y}\to 0} l(1, \hat{y}) = \infty. \tag{3.5}$$

The loss term is a measure of this loss across the training input.

The regularisation term stems from the principle that a simpler model is a better model: a complex model is more likely to be the result of a classifier modelling slight variations caused by noise in the training input. This reduces the generalisability of the classifier, harming its performance on unseen input. This phenomenon is known as *overtraining* (or *overfitting*) [93]. The regularisation term is defined to be greater for more complex models, and so encourages simpler models during training. The precise form of Ω depends on the ML method used and so will be addressed when specific classification techniques are discussed later in this chapter.

Before the training begins certain parameters that modify the objective function or otherwise influence the training process must be set. These are known as *hyperparameters*, represented by $\boldsymbol{\vartheta}$.

3.1.1 Evaluating a Classifier

In a dataset containing both signal and background events, a perfect classifier will label all signal events as signal and all background events as background. Events that are labelled correctly in this way are known as *true positives* and *true negatives*, respectively. In reality, a perfectly performing classifier is unlikely to be trained, so some events will be mislabelled. If a background event is labelled as signal, this is a *false positive* or *type I error*. Conversely, if a signal event is labelled as background, this is a *false negative* or *type II error* [93, 95].

Many different, but interrelated, metrics are used to evaluate the performance of a classifier. Some of the more prevalent are:

Precision Also known as the *positive predictive value*, precision is the fraction of events labelled as signal that are really signal events at a given threshold [96].

Recall Also known as the *sensitivity* or the *true positive rate*, recall is the fraction of signal events that are correctly identified at a given threshold [96].

Type I error rate Also known as the *false positive rate* and *fallout*, the type I error rate is the fraction of background events incorrectly identified as signal events at a given threshold [96].

Receiver operating characteristic curve The *receiver operating characteristic curve* (ROC curve) of a classifier is a plot of the false positive rate on the x axis, and the true positive rate on the y axis across all possible thresholds [93, 97].

Area under the ROC curve The *area under the ROC curve* (AUROC) can be interpreted as the probability that a classifier assigns a randomly chosen signal event a greater response than a randomly chosen background event [97]. For a perfect classifier, the AUROC is 1 and an AUROC of 0.5 indicates a classifier no better than random chance [93, 97]. Unlike precision, recall, and fallout, the AUROC is independent of the threshold.

To check for overtraining, the metrics defined above can be evaluated for both unseen and training input and checked for discrepancies. In addition, a metric specifically sensitive to overtraining is used: the two-sample *Kolmogorov–Smirnov* (KS) test.

The two-sample Kolmogorov–Smirnov test

The two-sample KS test compares the distribution of two independent samples, and calculates the probability that they were both drawn from the same distribution [98, 99]. It is performed by first constructing the *empirical cumulative distribution function* (ECDF),

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{[-\infty,x]}(X_i) \tag{3.6}$$

with n as the number of independent and identically distributed observations, X_i, and

$$I_{[-\infty,x]}(X_i) = \begin{cases} 1 & \text{for } X_i \leq x \\ 0 & \text{otherwise} \end{cases}, \tag{3.7}$$

for the two samples. An ECDF is therefore a step function starting at zero that increases by $\frac{1}{n}$ at every observed value. The two-sample KS statistic, $D_{n,m}$, is given by the greatest distance between the ECDF of the two samples:

$$D_{n,m} = \sup_x \left| F_{1\,n}(x) - F_{2\,m}(x) \right| \tag{3.8}$$

where n and m are the sizes of the samples, and $F_{1\,n}$ and $F_{2\,n}$ are their ECDFs. The probability that $\sqrt{\frac{mn}{m+n}} D_{n,m}$ will exceed a value z if the samples are drawn from the same underlying distribution is then given by the KS distribution

$$P(z) = 1 - 2 \sum_{j=1}^{\infty} \mathrm{e}^{-2z^2 j^2}. \tag{3.9}$$

So far, the form of the KS test presented assumes unweighted observations, X_i. In particle physics this assumption rarely holds, but [100] shows how the KS test can be adapted for the case of weighted observations. First, the ECDF is redefined as

$$F_n(x) = \frac{\sum_{i=1}^{n} W_i I_{[-\infty,x]}(X_i)}{\sum_{i=1}^{n} W_i} \tag{3.10}$$

where W_i is the weight associated with observation X_i. The ECDF is now a step function increasing by $\frac{W_i}{\sum_{i=1}^{n} W_i}$ at every observed value X_i. Then, instead of using the sample size n, the effective sample size, $\frac{1}{S_2}$, defined as

$$\frac{1}{S_2} = \frac{\left(\sum_{i=1}^{n} W_i\right)^2}{\sum_{i=1}^{n} W_i^2}, \tag{3.11}$$

is used in its place.

The two-sample KS test can be used as an overtraining check by performing it on the distribution of the response for the training sample and unseen input. If the classifier has not been overtrained the distribution of the two responses should be same, resulting in a high probability from the KS test.

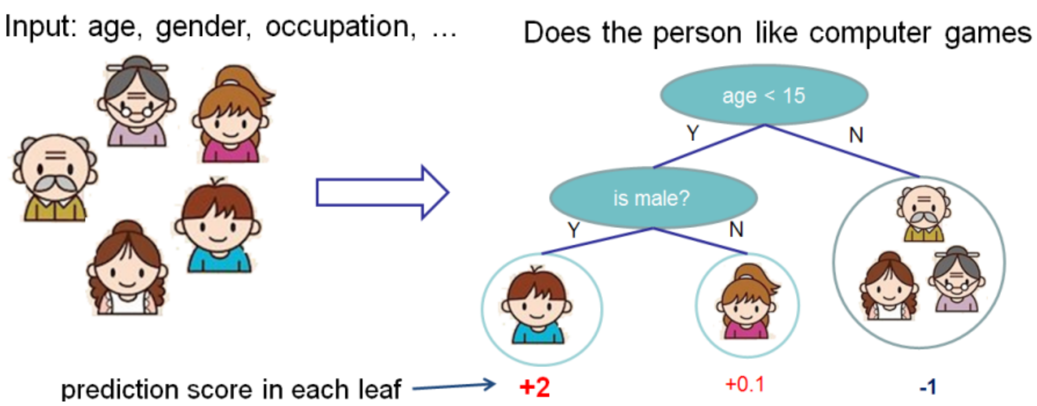


Figure 3.1: An example of a CART. Binary decisions are made about features, which sorts the input into leaves. Each input is assigned a score based on the leaf it is sorted into. Figure taken from [103].

3.2 Classification with a Boosted Decision Tree

BDTs belong to a class of ML techniques known as *ensemble methods* [101]. These methods work by combining multiple different base learners into an ensemble that performs better than any learner individually. Boosting is one method of creating an ensemble, in which base learners are created one after another. The performance of the previous learners informs the creation of the next learner in the sequence. The response of the final ensemble is the total response of each individual base learner.

3.2.1 Classification and Regression Trees

The base learners used in a BDT are *classification and regression trees* (CARTs). An example of a CART is shown in Figure 3.1: it is a tree of binary divisions on the available features that terminates in *leaves* with an associated *score*. A CART can be represented as a function, $f(\boldsymbol{x}_i)$, that maps the features of an event, $\boldsymbol{x}_i$, to a score, y_i [102]. The space of all possible CARTs is denoted by the set $\mathbb{T}$, where $f \in \mathbb{T}$.

3.2.2 Boosting with XGBoost

Many boosting methods have been developed, but the leading technique is known as *Extreme Gradient Boosting* (XGBoost) [102]. XGBoost sets itself apart from the competition through its performance, both in terms of computation time and predictive ability [104]. Given its position as the state-of-the-art boosting algorithm, XGBoost was chosen as the boosting algorithm for the BDTs used in this thesis.

The objective function used by XGBoost is given by

$$C^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{i=1}^{t} \Omega(f_i) \quad f_i \in \mathbb{T} \tag{3.12}$$

where n is the number of entries in the training set and t is the number CARTs in the ensemble. This equation follows the pattern of a cost function as defined in Equation 3.1: a loss term plus a regularisation term. The loss term is the sum of the values of the chosen loss function over all entries in the training input and the regularisation term, Ω, is some function of a CART's structure, f_i. When using Equation 3.2 as l, the score, w, of a leaf is converted into a response, $\hat{y}$, in the range $0 < \hat{y} < 1$ with a logistic function

$$\hat{y} = \frac{1}{1 + \mathrm{e}^{w}}. \tag{3.13}$$

While training, new CARTs f_i are added to the ensemble one at a time. Consider the response of the classifier as subsequent CARTs are added:

$$\hat{y}_i^{(1)} = f_1(\boldsymbol{x}_i) \tag{3.14}$$

$$\hat{y}_i^{(2)} = f_1(\boldsymbol{x}_i) + f_2(\boldsymbol{x}_i) = \hat{y}_i^{(1)} + f_2(\boldsymbol{x}_i). \tag{3.15}$$

Thus, more generally, the response at any given point in the sequence is

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(\boldsymbol{x}_i) = \hat{y}_i^{(t-1)} + f_t(\boldsymbol{x}_i). \tag{3.16}$$

Introducing this into Equation 3.12 gives

$$C^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}^{(t-1)} + f_t(\boldsymbol{x}_i)\right) + \sum_{i=1}^{t} \Omega(f_i) \quad f_i \in \mathbb{T}. \tag{3.17}$$

For each CART added to the ensemble, XGBoost finds the specific CART, $f_t \in \mathbb{T}$, that reduces the objective function by the largest amount. This process continues until a preordained number of CARTs (a hyperparameter) have been added.

It has been found that it is beneficial to modify Equation 3.17 to include a scaling factor, $0 < \nu \leq 1$, that is applied to each new CART:

$$C^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}^{(t-1)} + \nu f_t(\boldsymbol{x}_i)\right) + \sum_{i=1}^{t} \Omega(f_i) \quad f_i \in \mathbb{T}. \tag{3.18}$$

This term is known as the *shrinkage* or *learning rate* [102, 105]. It has been empirically determined that choosing $\nu < 0.1$ yields the greatest improvements in model generalisability [106].

3.2.3 Regularising a Boosted Decision Tree

In XGBoost the regularisation term, $\Omega(f)$, consists of three parts [102]:

$$\Omega(f) = \gamma T + \alpha \sum_{j=1}^{T} \left|w_j\right| + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{3.19}$$

where T is the number of leaves on a CART, w_j are the scores on each leaf, and γ, α, λ are tunable hyperparameters that control the influence of the respective regularisation terms on the objective function (*regularisation strength*).

The first term introduces a penalty for larger CARTs; the size of a CART is a simple measure of model complexity.

The second and third terms penalise large scores on a CART: the second term uses the modulus of the scores and is known as the *L1 term* and the third term uses the square of the scores and is known as the *L2 term*. Individually, these terms represent *least absolute shrinkage and selection operator* (LASSO) regression [107] and ridge regression [108], respectively, but together they represent L1L2 regularisation (also known as elastic net regression) [109]. Large scores on a leaf are punished because they could cause the CART to be sensitive to small fluctuations in a feature, a symptom of overtraining.

A final regularisation technique that can be used and that does not directly modify the objective function is *stochastic boosting*. Stochastic boosting trains each CART on a subset of the full training input. This reduces overtraining by preventing the classifier from adapting to statistically insignificant properties of the training input. It has been found that exposing each CART to 50–80% of the training input yields the best results [110].

3.3 Classification with a Multilayer Perceptron

Multilayer perceptrons are form of *artificial neural network* (ANN), a class of ML systems inspired by the architecture of the human brain [111]. An ANN consists of nodes, known as *neurons*, that are connected in a manner analogous to synapses. Neurons can be thought of as functions that receive input from other neurons, which then pass their output (*activation*) on to further neurons. What defines the type of ANN is the architecture of the connections between neurons.

The architecture of a typical MLP is shown in Figure 3.2. In an MLP the neurons are arranged into layers with each neuron in a given layer n feeding into every neuron in layer $n + 1$ [112]. A connection never skips a layer.

The first layer is the *input layer*. As there are no previous layers, neurons from this layer cannot take input from other neurons. Instead, the neurons in this layer

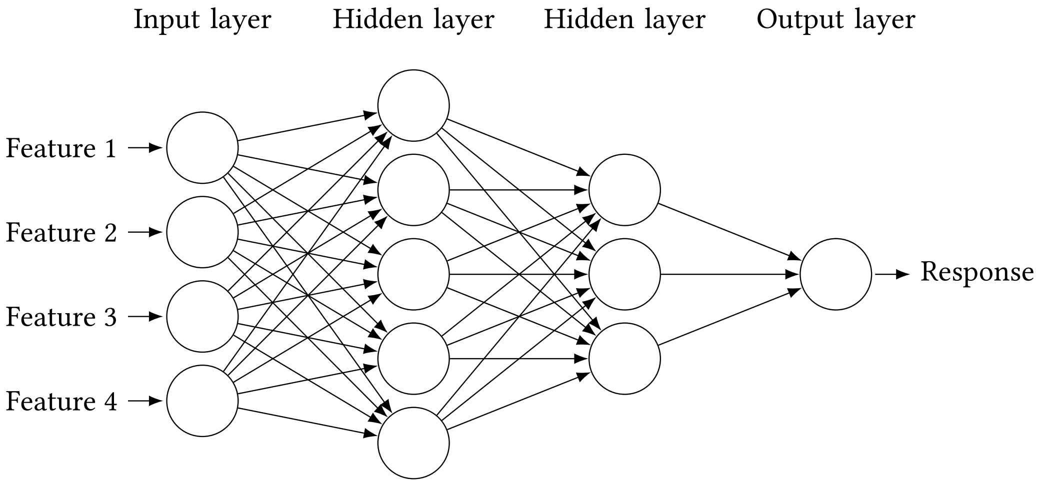


Figure 3.2: The basic structure of an MLP, showing the neurons (open circles) and connections in the input, hidden, and output layers. Features are propagated through the network to produce a response in the output layer.

have a 1 : 1 correspondence with the input features. Each neuron's activation is the value of an individual feature.

Between the first input layer and the final *output layer* are the so-called *hidden layers*. Figure 3.2 shows a network with two hidden layers, but any number of hidden layers containing any number of neurons can be chosen. These choices are hyperparameters of an MLP. The function representing the activation of a hidden layer neuron is [113]

$$a_j^{(n+1)} = f\left(b + \sum_{i=1}^{k} w_i a_i^{(n)}\right). \tag{3.20}$$

Neuron jth in layer $n+1$ takes as input the sum over the activations of the k neurons in layer n, each multiplied by a weight w_i. A *bias*, b, is then added, and finally an *activation function*, f, is applied. Weights and biases are unique to each neuron, and are adjusted during the training of the MLP. The choice of activation function constitutes another hyperparameter.

The final layer is the output layer that, in the case of binary classification, consists of only one neuron. Its activation is determined by Equation 3.20, but f is chosen so that the activation is in the interval $(0, 1)$. In this way, the output can be interpreted as a probability that an event is signal [112].

The MLP is trained by adjusting the weights and biases to minimise the *objective function* [112]:

$$C(\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{b}) = \frac{1}{n}\sum_{i=1}^{n} l(y_i, \hat{y}_i) + \Omega(\boldsymbol{w}) \tag{3.21}$$

where $\boldsymbol{X}$ is the training input, $\boldsymbol{w}$ are the weights of the network, and $\boldsymbol{b}$ are the biases of the network. Unlike Equation 3.12, the average rather than total loss is used. The regularisation term, Ω, is discussed in § 3.3.1.

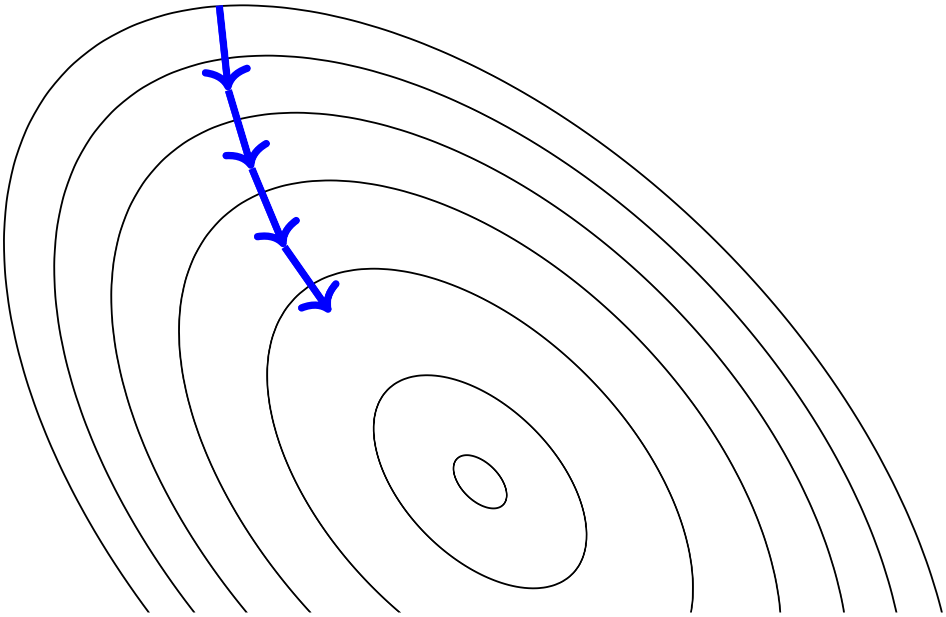


Figure 3.3: Demonstration of how the gradient of a two-dimensional function can be followed to approach a local minimum. The blue lines represent the progress of the descent, with the gradient recalculated after each step. Black lines are the isopleths of a 2D function.

Minimising this objective function is not trivial. Even the simple network displayed in Figure 3.2 contains 47 free parameters (38 weights and 9 biases). Many algorithms, known as *optimizers*, have been developed for this task, all with roots in *gradient descent*.

The core principle of gradient descent is expressed in Figure 3.3. The gradient (derivative) of a function is determined and then some distance, the *step size* (v), is travelled in the direction of negative gradient (direction of steepest descent). This process is then repeated until a local minimum is reached, where the gradient will be zero. When training a neural network the same principle applies, but in a higher-dimensional space. The gradient calculated is the partial gradient of the objective function with respect to $\boldsymbol{w}$ and $\boldsymbol{b}$ over the whole training sample. The step size is referred to as the *learning rate* [112]. A step defines a single *epoch* of an MLP's training. The process is repeated for fixed number of epochs, or until there is no decrease in the objective after a certain number of epochs (*early stopping* [112]).

Before the process can start, the weights and biases in the network must be assigned some initial value. The weights cannot all be assigned the same initial value throughout the network because the network cannot break this symmetry during training. A common initialisation scheme is known as *Glorot initialisation* [114]. Here, the weights are drawn at random from a Gaussian distribution truncated at two standard deviations, with zero mean and variance of $\frac{2}{n_{\text{in}}+n_{\text{out}}}$ where n_{in} is the number of neurons in the previous layer, and n_{out} the number of neurons in the next layer. Biases are typically initialised to one.

The training procedure is usually modified, particularly for large datasets, by the introduction of *minibatches* [112]. When implemented, the training input is split into minibatches of a certain *batch size*, and a gradient descent step is performed separately for each minibatch. An epoch is then defined as the propagation of every training event through the network. At the end of an epoch, the training input is shuffled.

Most optimisers are extensions of the basic gradient descent algorithm. One of the leading optimisers is *adaptive moment estimation* (Adam) [115], which incorporates the gradients from previous steps and adjusts the learning rate over time.

Once a network has been trained, it can be applied to input not seen in the training set. This can be done by feeding this novel input through the MLP using the final weights and biases obtained from the training.

3.3.1 Regularising a Multilayer Perceptron

As with BDTs, regularisation techniques can be applied to MLPs to reduce overtraining. This section examines how L1L2 regularisation can be applied to MLPs and discusses a regularisation technique unique to ANNs: *dropout* [116].

For an MLP, L1L2 regularisation takes the form [112]

$$\Omega = \frac{\alpha}{n}\sum_{w}|w| + \frac{\lambda}{2n}\sum_{w} w^2 \tag{3.22}$$

where $\sum_w$ represents the sum over all weights in the network, n is the number of events in the training sample (or batch size, if minibatches are used), and α, λ are the L1L2 regularisation strength parameters as before. The L1L2 terms in this instance penalise the network for having large weights with the same justification for penalising large scores on a CART (see § 3.2.3). It is possible to modify Equation 3.22 so that each layer has specific values for α and λ.

When using dropout, neurons in the hidden layers of an MLP have a probability p_{d} (another hyperparameter) of being *dropped out* every time the weights are updated (i.e. at the end of an epoch or minibatch) [112, 116]. A dropped out neuron is treated as if it is not part of the network: its activation is set to zero and any weights or biases associated with it cannot be modified. A version of the MLP shown in Figure 3.2 with dropout applied to the hidden layers is shown in Figure 3.4.

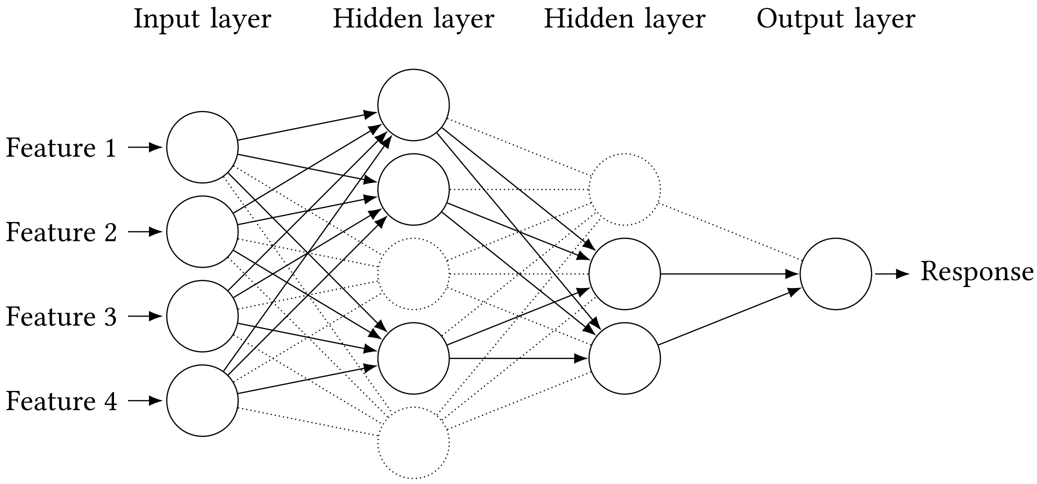


Figure 3.4: The MLP shown in Figure 3.2 with dropout applied to the hidden layers. The neurons and connections affected are dotted out. Inputs propagated through this network will behave as if the dropped out neurons do not exist.

Dropout prevents a network from becoming overreliant on specific neurons in the network. This prevents the training of a network in which only a few neurons have meaningful contributions to the response. Dropout also prevents co-adaption of neurons where one neuron will fix the mistakes of another. In other words, a neuron develops in such a way that it only performs well in a specific context provided by other neurons. These co-adaptions are not generalisable, and lead to overtraining [116].

Once the network has been trained, all neurons are used for classification. To account for the fact that during training the network will effectively have fewer neurons, the activation of each neuron is scaled by $\frac{1}{p_\mathrm{d}}$. The technique of applying this scaling at training time is known as *inverted dropout* [117].

3.4 Regression

Regression is similar to classification but rather than mapping features to a class label, the goal is to map features to the value of some unknown property [93]. An example of a regression problem, featured prominently in the ML literature, is the Boston Housing Dataset [118]. Using features such as the local crime rate, number of rooms, and building age, ML regression techniques are used to predict the value of a house.

Regression techniques are used in the tZq analysis to attempt to model the relationship between the hyperparameters of a BDT and its classification performance in order to find the values of the hyperparameters that give the most performant classifier. This is described further in § 9.2.3. The technique chosen for this regression is the GP [119], which excels at modelling complex functions with small datasets.

3.5 Regression with a Gaussian Process

This section introduces the GP and describes how it can be used to perform regression analysis. The notation defined in Appendix A.3 is used throughout.

The Gaussian (or normal) distribution is a continuous probability distribution described by

$$P\left(z|\mu_z,\sigma_z\right) = \frac{1}{\sigma_z\sqrt{2\pi}}\exp\left(-\frac{\left(z-\mu_z\right)^2}{2\sigma_z^2}\right) \tag{3.23}$$

or, alternatively,

$$Z \sim \mathcal{N}\left(\mu_z,\sigma_z\right). \tag{3.24}$$

It is entirely parametrised by the mean, μ_z, and standard deviation, $\sigma_z \geq 0$. The distribution of a length n vector in which each element is a normally-distributed random variable, $\boldsymbol{Z} = (Z_1,\dots,Z_n)$, is described by an n-dimensional Gaussian

$$P\left(\boldsymbol{z}|\boldsymbol{\mu}_z,\Sigma_{zz}\right) = \frac{1}{\sqrt{(2\pi)^n\,|\Sigma_{zz}|}}\exp\left(-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_z)^{\mathrm{T}}\Sigma_{zz}^{-1}\left(\boldsymbol{z}-\boldsymbol{\mu}_z\right)\right) \tag{3.25}$$

i.e.

$$\boldsymbol{Z} \sim \mathcal{N}\left(\boldsymbol{\mu}_z,\Sigma_{zz}\right), \tag{3.26}$$

where

$$\boldsymbol{\mu}_z = \mathrm{E}[\boldsymbol{Z}] = (\mathrm{E}[Z_1],\dots,\mathrm{E}[Z_n]) \tag{3.27}$$

and

$$\Sigma_{zz} = \begin{pmatrix} \mathrm{cov}\,(z_1,z_1) & \cdots & \mathrm{cov}\,(z_1,z_n) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}\,(z_n,z_1) & \cdots & \mathrm{cov}\,(z_n,z_n) \end{pmatrix}. \tag{3.28}$$

is a symmetric covariance matrix. This matrix satisfies the positive semidefinite condition,

$$\boldsymbol{v}^{\mathrm{T}}\Sigma_{zz}\boldsymbol{v} \geq 0 \quad \forall \boldsymbol{v} \in \mathbb{R}^n, \tag{3.29}$$

where $\mathbb{R}^n$ is the set of all real vectors of length n. An n-dimensional Gaussian, therefore, is entirely parametrised by the mean vector, $\boldsymbol{\mu}_z$, and covariance matrix, Σ_{zz}. If $\boldsymbol{Z}$ is partitioned as

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} = \left(X_1,\dots,X_{n_x},Y_1,\dots,Y_{n_y}\right) \tag{3.30}$$

then the resultant distribution—known as the *joint* distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$—is of the form [119]

$$Z = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{xy} \end{bmatrix}\right) \tag{3.31}$$

where Σ_{xy} is the *cross-covariance matrix* of $\boldsymbol{X}$ and $\boldsymbol{Y}$:

$$\begin{aligned} \left(\Sigma_{xy}\right)_{ij} &= \operatorname{cov}\left(X_i, Y_j\right) = \left(\Sigma_{yx}\right)_{ji} \\ \therefore \Sigma_{xy} &= \Sigma_{yx}^{\mathrm{T}}. \end{aligned} \tag{3.32}$$

The marginal distribution of $\boldsymbol{X}$,

$$P(\boldsymbol{x}) = \int P(\boldsymbol{x}, \boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}, \tag{3.33}$$

is $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_x, \Sigma_{xx})$ [119] and the conditional distribution of $\boldsymbol{Y}$ for a given $\boldsymbol{X}$ is [119]

$$Y|\boldsymbol{X} \sim \mathcal{N}\left(\mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}\left(\boldsymbol{X} - \mu_x\right), \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\right). \tag{3.34}$$

A GP is defined formally as follows [93]:

> *For any set S, a Gaussian process on S is a set of random variables $\{Z_t : t \in S\}$ such that $\forall k \in \mathbb{N}^*; \forall t_1, \ldots, t_k \in S$ the vector $\left(Z_{t_1}, \ldots, Z_{t_k}\right)$ is distributed as a k-dimensional Gaussian.*

Consider the case of a 10-dimensional Gaussian and the set $S = \{\mathbb{Z} \cap [1, 10]\}$ where $k = 4$ and $t_1, \ldots, t_k = 1, 3, 7, 9$. The resultant vector, $\left(Z_1, Z_3, Z_7, Z_9\right)$, is distributed as a kD (4D) Gaussian. This is true for any choices of $n, k \in \mathbb{N}^*$ and $t_1, \ldots, t_k \in S$ that could have been made. Therefore, a GP on S is a 10-dimensional Gaussian.

In the case where $S = \mathbb{Z}$, it is a small logical leap to observe that a GP in this case is an infinite-dimensional Gaussian, parametrised by an infinitely long $\boldsymbol{\mu}$ and a covariance matrix of infinite rows and columns. An alternative way of representing such objects is by using functions: an infinitely long vector, v, can be described as

$$v(i) = \boldsymbol{v} \cdot \hat{\boldsymbol{e}}_i \quad i \in \mathbb{Z} \tag{3.35}$$

where $\hat{\boldsymbol{e}}_i$ is a unit vector satisfying $\hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_j = \delta_{ij}$, and an infinite matrix, $\mathcal{M}$, can be described as

$$M(i, j) = \hat{\boldsymbol{e}}_i^{\mathrm{T}} \mathcal{M} \hat{\boldsymbol{e}}_j \quad i, j \in \mathbb{Z}. \tag{3.36}$$

This pivot to considering spaces of functions allows the possibility of a GP on the set $\mathbb{R}$. Rather than describing the distribution of vectors, Z, that contain Gaussian-distributed random variables, a GP on $\mathbb{R}$ describes a distribution of continuous functions f that return Gaussian-distributed random variables. This distribution is represented as

$$f \sim \mathcal{GP}(\mu(z), k(z, z')) \tag{3.37}$$

where $\mathcal{GP}$ represents a GP. The process is defined by a *mean function* $\mu(z)$ and a *kernel function* $k(z, z')$ [119]. The mean of a vector $Z = \left(f(z_1), \ldots, f(z_k)\right)$ is

$$\boldsymbol{\mu}_z = (\mu(z_1), \ldots, \mu(z_k)) \tag{3.38}$$

and its covariance is

$$\Sigma_{zz} = \begin{pmatrix} k(z_1, z_1) & \cdots & k(z_1, z_n) \\ \vdots & \ddots & \vdots \\ k(z_n, z_1) & \cdots & k(z_n, z_n) \end{pmatrix}. \tag{3.39}$$

Consequently, the kernel function must be symmetric and positive semidefinite, that is, Equation 3.39 must be a symmetric, positive semidefinite matrix $\forall z_i \in \mathbb{R}$.

It can be verified that this object still satisfies the definition of a GP. Choosing $k = 3$ and $t_1, \ldots, t_k = 0, 1, \pi$, forms a vector $Z = (Z_0, Z_1, Z_\pi)$. The mean vector of Z is $\boldsymbol{\mu}_z = \left(\mu(0), \mu(1), \mu(\pi)\right)$, the covariance matrix can be constructed as

$$\Sigma_{zz} = \begin{pmatrix} k(0,0) & k(0,1) & k(0,\pi) \\ k(1,0) & k(1,1) & k(1,1) \\ k(\pi,0) & k(\pi,1) & k(\pi,\pi) \end{pmatrix}, \tag{3.40}$$

and so $Z \sim \mathcal{N}\left(\boldsymbol{\mu}_z, \Sigma_{zz}\right)$. Such a distribution could have been constructed in the same manner for a vector Z of any length with any choices of $t \in \mathbb{R}$, and so the definition of a GP remains satisfied.

To use a GP to perform a regression, a GP is used to define a prior over functions: this distribution contains all functions that could *a priori* describe the input. For simplicity, it is common to assume the mean function of this GP is zero. The prior is therefore distributed as

$$f \sim \mathcal{GP}(0, k(x, x')). \tag{3.41}$$

The assumptions about the input are now entirely encoded by the choice of kernel function. Common choices for the kernel function are [120]:

The *radial basis function kernel* (RBF kernel)

$$k_{\text{RBF}}(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2l^2}\right) \tag{3.42}$$

Otherwise known as the squared-exponential, Gaussian, and exponentiated quadratic kernel, k_{RBF} has become a de-facto standard for GP regression.

Functions drawn from $f \sim \mathcal{GP}(\mu, k_{\mathrm{RBF}})$ are infinitely differentiable, and so it is a good choice when discontinuities are not expected. In a kernel function there are free parameters—in this case σ^2 and l—that form the hyperparameters $\boldsymbol{\vartheta}$ of the GP. The output variance σ^2 describes the average distance of a function from its mean. The characteristic length-scale l describes how proximate two points must be to be well-correlated i.e. using a kernel with a smaller l will produce functions that are "wigglier".

The periodic kernel

$$k_{\mathrm{Per}}(x, x') = \sigma^2 \exp\left(-\frac{2\sin^2\left(\pi\left|x - x'\right| p^{-1}\right)}{l^2}\right) \tag{3.43}$$

The periodic kernel [121] can be used to model functions that repeat themselves exactly with period p, useful when the input being modelled is expected to do the same. Like the RBF kernel it is parametrised, aside from p, by an output variance and characteristic length-scale.

The linear kernel

$$k_{\mathrm{Lin}}(x, x') = (x - a)(x' - a) \tag{3.44}$$

The linear kernel allows GPs to be used for the case of linear regression (in practice, however, using GPs for linear regression is computationally inefficient), as $f \sim \mathcal{GP}(0, k_{\mathrm{Lin}})$ is a distribution of linear functions that cross the x axis at $x = a$. This kernel differs from the other described here in that it is *nonstationary*: it cannot be expressed as $k(x - x')$ and so its value is determined by the absolute rather then relative positions of its arguments.

The Matérn kernels

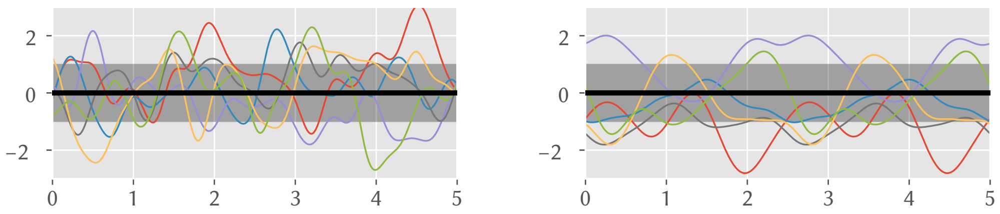
$$k_\nu(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\frac{x - x'}{l}\right)^\nu \mathrm{K}_\nu\left(\sqrt{2\nu}\frac{x - x'}{l}\right) \qquad \nu \in \mathbb{N}^* - \frac{1}{2} \tag{3.45}$$

The Matérn kernels [119, 122, 123] are a class of general-purpose kernels commonly used in GPs. The kernel k_ν models functions that are $\lfloor \nu \rfloor$ times differentiable. The Matérn kernels are defined in terms of the gamma function,

$$\Gamma(\nu) = \int_0^\infty z^{\nu-1} \mathrm{e}^{-z}\,\mathrm{d}z, \tag{3.46}$$

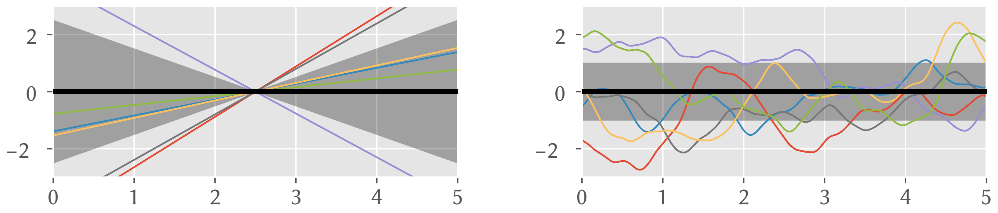
and modified Bessel functions of the second kind, K_ν. They have the same parameters as the RBF kernel, and

$$\lim_{\nu\to\infty} k_\nu = k_{\mathrm{RBF}}. \tag{3.47}$$



(a) Functions drawn from a GP using the RBF kernel with $\sigma^2 = 1, l = 0.2$.

(b) Functions drawn from a GP using the periodic kernel with $\sigma^2 = 1, l = 1, d = 2.5$.



(c) Functions drawn from a GP using the linear kernel with $c = 2.5$.

(d) Functions drawn from a GP using the Matérn kernel with $\sigma^2 = 1, l = 0.5, \nu = 2.5$.

Figure 3.5: Functions drawn from GPs with various kernel functions and $\mu(x) = 0$. The mean function is shown in black with a one standard deviation envelope shaded. Functions drawn from a GP with a kernel of a shorter characteristic length-scale, such as those in (a), show more variation over a given length.

Functions drawn from GPs using these kernels with $\mu(x) = 0$ are shown in Figure 3.5.

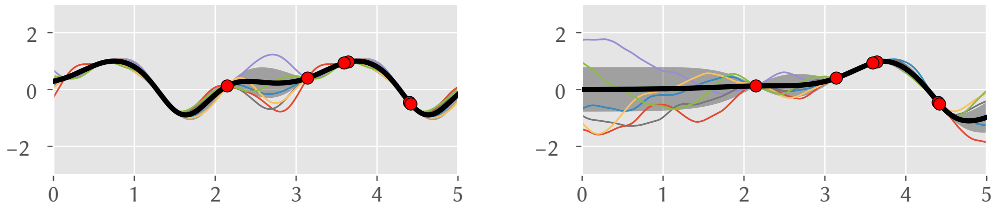
Suppose there is a set, $\boldsymbol{X}$, of observations corresponding to function outputs, $\boldsymbol{f}$. This forms the training input for regression. Assuming Equation 3.41 holds, it follows that

$$\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{0}, \Sigma) . \tag{3.48}$$

This prior GP can be conditioned to form a posterior GP that is restricted to those functions agreeing with the observations. The mean function of this posterior GP can then be used to find the expected value at a set of outputs $\boldsymbol{f}_* \sim \mathcal{N}(\boldsymbol{0}, \Sigma_{**})$. Equation 3.34 shows how to form this distribution of $\boldsymbol{f}_*$ conditioned by $\boldsymbol{f}$ [93, 119]:

$$\begin{aligned} \boldsymbol{f}_* | \boldsymbol{f} &\sim \mathcal{N}\left(\boldsymbol{\mu}_* + \Sigma_*^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{f} - \boldsymbol{\mu}), \Sigma_{**} - \Sigma_*^{\mathrm{T}} \Sigma^{-1} \Sigma_*\right) \\ &\sim \mathcal{N}\left(\boldsymbol{0} + \Sigma_*^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{f} - \boldsymbol{0}), \Sigma_{**} - \Sigma_*^{\mathrm{T}} \Sigma^{-1} \Sigma_*\right) \\ \Rightarrow \mathrm{E}[\boldsymbol{f} | \boldsymbol{f}_*] &= \Sigma_*^{\mathrm{T}} \Sigma^{-1} \boldsymbol{f} . \end{aligned} \tag{3.49}$$

To perform this regression values for $\boldsymbol{\vartheta}$ (the parameters of the kernel function) must be chosen. Using a GP allows for the automatic optimisation of $\boldsymbol{\vartheta}$ by minimising



(a) Plot showing functions drawn from a posterior GP defined using the periodic kernel. The optimised values of the kernel paramters are $\sigma^2 = 0.437$, $l = 0.764$, $d = 3$.

(b) Plot showing functions drawn from a posterior GP defined using the Matérn $\nu = \frac{5}{2}$ kernel. The optimised values of the kernel paramters are $\sigma^2 = 0.584$, $l = 0.639$. In the region with no training input, the GP reflects its prior state seen in Figure 3.5(d).

Figure 3.6: Functions drawn from GPs conditioned on training input, shown in red. The mean function is shown in black with a one standard deviation envelope shaded.

the log marginal likelihood [93, 119]:

$$P(\boldsymbol{f}_*|\boldsymbol{X}) = \int P(\boldsymbol{f}_*|\boldsymbol{f}, \boldsymbol{X})\, P(\boldsymbol{f}|\boldsymbol{X})\, \mathrm{d}\boldsymbol{f} \tag{3.50}$$

$$\Rightarrow \log P(\boldsymbol{f}_*|\boldsymbol{X}) = -\frac{1}{2}\boldsymbol{f}_*^{\mathrm{T}}\Sigma\boldsymbol{f} - \frac{1}{2}\log|\Sigma| + C \tag{3.51}$$

where C is a constant. Figure 3.6 shows the posterior GP for the periodic and Matérn kernels on some dummy input.

Thus far, it has been assumed that all observations are noiseless. It is possible to extend this framework to cover the case of noisy observations, but this lies outside the scope of this thesis. More information can be found in [93] and [119].

Chapter 4

The CMS Experiment at the Large Hadron Collider

> *Our judge is not God or governments, but Nature*
>
> Tejinder Virdee

4.1 The Large Hadron Collider

The largest and most energetic particle accelerator ever constructed is the Large Hadron Collider (LHC), operated by the *European Organization for Nuclear Research* (CERN). The LHC is a synchrotron of 27 km circumference, a record matched only by the *Large Electron–Positron Collider* (LEP) whose tunnel it now occupies. Unlike LEP, the LHC is primarily a proton–proton collider but is capable of performing lead ion collisions. Prior to the LHC, the most energetic particle collider was the Tevatron at Fermilab, a $p\bar{p}$ synchrotron that reached a centre-of-mass energy of $\sqrt{s} = 980\,\text{GeV}$ [124]. Centre-of-mass energy, $\sqrt{s}$, is defined as the total energy of a system as measured in the inertial frame in which the centre of mass of the system is at rest. Proton–proton collisions at the LHC outclass the Tevatron by an order of magnitude, reaching $\sqrt{s} = 13\,\text{TeV}$.

Such high centre-of-mass energies are achieved at the LHC by accelerating protons with 1600 superconducting magnets, cooled to 1.9 K with liquid helium. Protons enter the LHC ring at 450 GeV, an energy reached by utilising a series of older accelerators present at the CERN laboratory. The beam is initially produced by ionising hydrogen gas to produce protons, and accelerating them to 50 MeV using Linac 2. The beam is then injected into the *Proton Synchrotron* (PS) booster

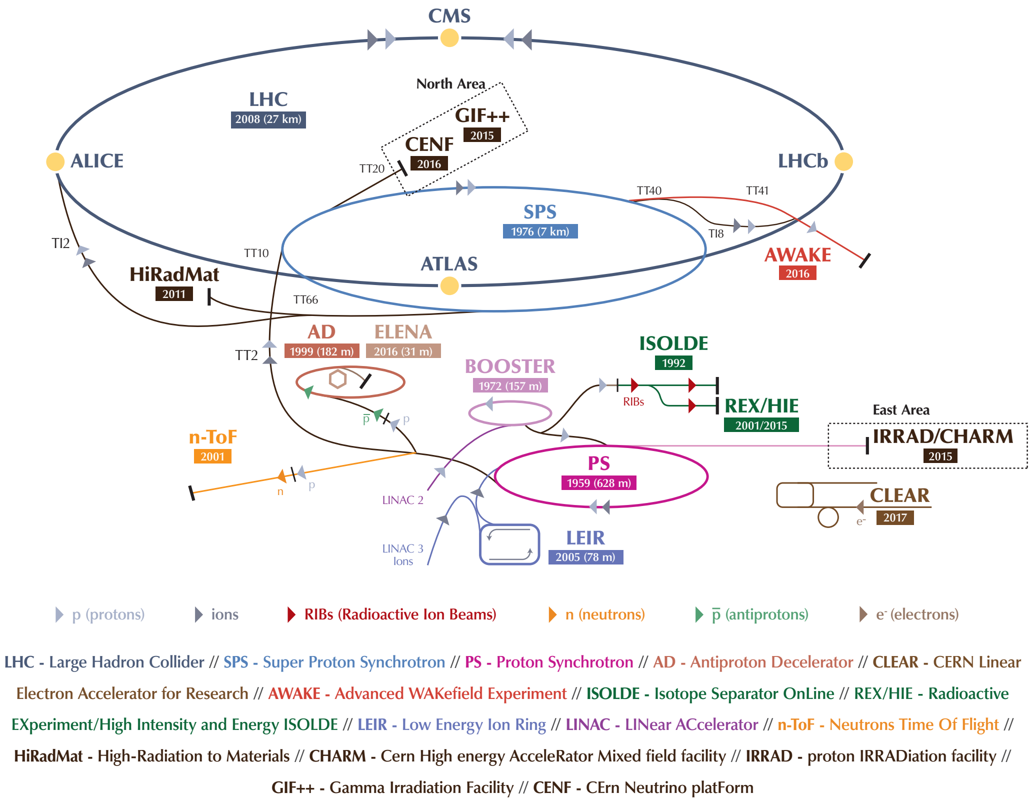


Figure 4.1: The interconnected system of accelerators at CERN. Figure taken from [125].

where it reaches 1.4 GeV, next the beam energy is raised 26 GeV in the PS itself, and finally the Super Proton Synchrotron increases the beam energy to the necessary 450 GeV. This accelerator chain is displayed in Figure 4.1.

Equally essential to the centre-of-mass-energy in the LHC's role as a window to new physics is its luminosity. The LHC has a design luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. This is not provided in one continuous beam, rather protons are collided every 25 ns in *bunches* [126].

The cross section for interactions at these energies, coupled with the high luminosity, result in a phenomenon known as *pileup*. This is the presence of multiple inelastic proton–proton collisions in a bunch crossing (*in-time pileup*) or in adjacent bunch crossings (*out-of-time pileup*).

In 2008 the LHC suffered a magnet quenching incident[1] [127], delaying the start of the first operational run (Run I) until 2009. The planned centre-of-mass energy during Run I was 10 TeV [128] but, because of safety concerns following the quench, collisions were initially performed at 7 TeV. This would increase to 8 TeV in 2012 [129], the greatest centre-of-mass energy achieved during Run I.

---

[1] A quench occurs in a superconducting magnet when it accrues a nonzero resistance; it leads to a rise in temperature and the explosive release of helium coolant.

After Run I a two-year shutdown began to prepare for 13 TeV collisions during the second operational run (Run II), which began in 2015 and concluded in 2018. In 2016 the design luminosity of the LHC was reached [130].

As seen in Figure 4.1, four major experiments are located at the four main beam crossing points on the LHC ring. Starting clockwise from the main CERN site at Meyrin, these are: A Toroidal LHC Apparatus (ATLAS), *A Large Ion Collider Experiment* (ALICE), Compact Muon Solenoid (CMS), and the *Large Hadron Collider Beauty* (LHCb) experiment. *A Large Ion Collider Experiment* (ALICE) and *Large Hadron Collider Beauty* (LHCb) are specialised to study heavy ion collisions and b hadrons, respectively. ATLAS and CMS, in contrast, are general-purpose detectors.

The high centre-of-mass energies at the LHC coupled with the available integrated luminosity allows the testing of theories and probing of processes outside the reach of earlier facilities. A primary goal was to determine the nature of electroweak symmetry breaking and validity of the Higgs mechanism. This was achieved in 2012 when CMS and ATLAS announced the discovery of a Higgs Boson [14, 15]. Beyond this, the LHC aims to search for new particles predicted by various BSM models, look for possible constituents of dark matter, investigate the behaviour of quark–gluon plasmas thought to be present in the early Universe (during heavy ion runs), and test the SM by probing rare processes such as tZq.

4.2 The CMS Detector

The CMS detector lies at intersection Point 5, on the opposite side of the LHC ring to the ATLAS detector. With a length of 21 m along the beam line and a 15 m diameter, CMS is the second largest experiment at the LHC (ATLAS being the largest). However, at 12 500 tonnes, CMS is the heaviest of the four main experiments.

Much of CMS's weight is accounted for by the steel *muon return yoke*. The *Muon System*, of which this yoke is part, forms the outermost layer of CMS's stratified detector system. Moving inwards, the *hadronic calorimeter* (HCAL)—designed to measure the energy of hadronic particles—forms the next detector layer. Components of the HCAL exist inside and outside the solenoidal magnet, which generates a uniform magnetic field of 3.8 T within it when operating [131]. Closer still to the beamline is the *electromagnetic calorimeter* (ECAL), which performs a similar function to the HCAL but for electrons and photons. Finally, closest to the beam, is the inner tracking system consisting of *silicon pixel* and *silicon strip* detectors. These detectors, in concert with the calorimeters, are used to reconstruct the trajectories

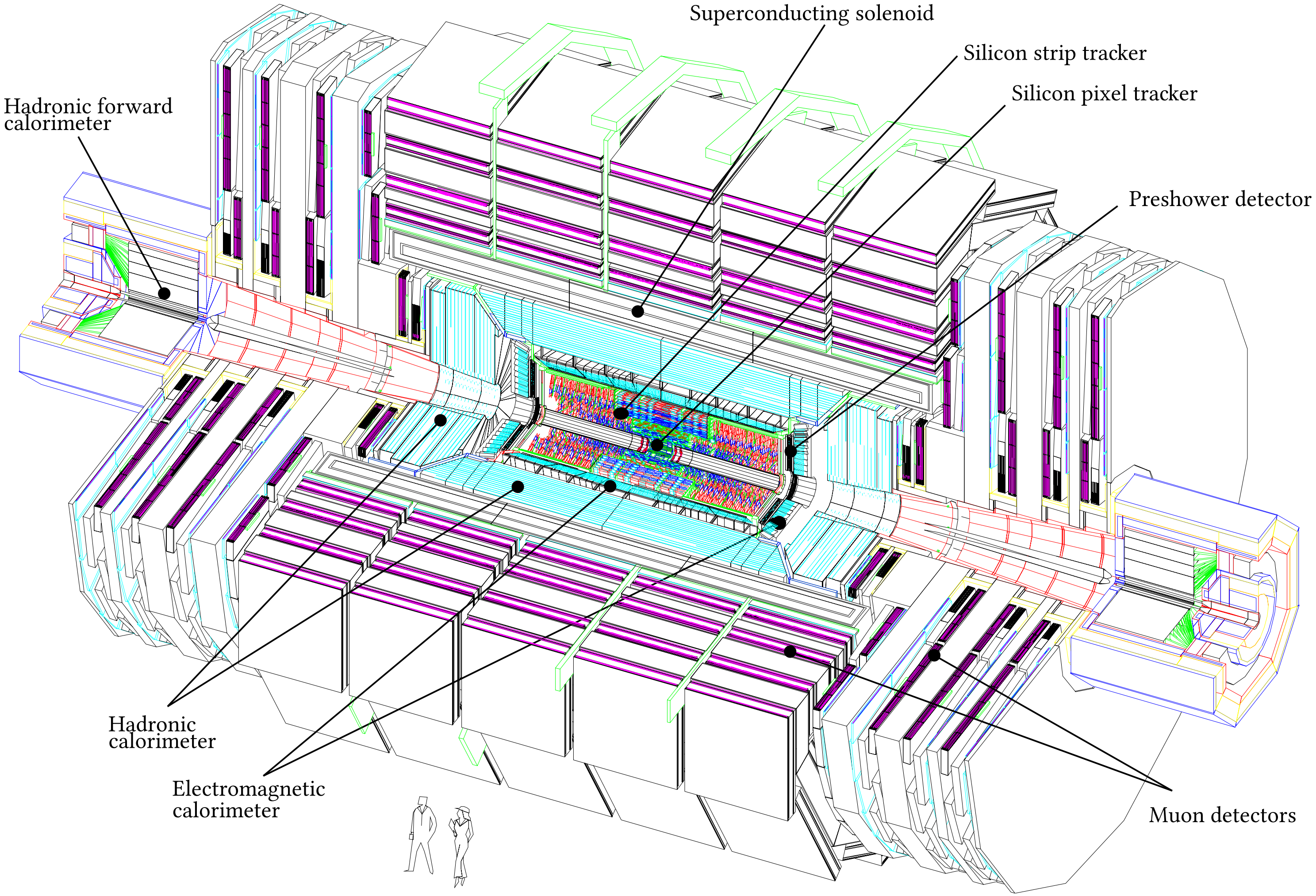


Figure 4.2: Cross section of the CMS detector. Figure taken from [132].

of collision debris that passes through them. Figure 4.2 shows a cross section of CMS with all these systems labelled.

In addition to these detectors around the beam pipe, CMS also features detectors exclusively present in the more forward region. Of relevance to the tZq analysis presented in this thesis is the preshower detector, again labelled in Figure 4.2.

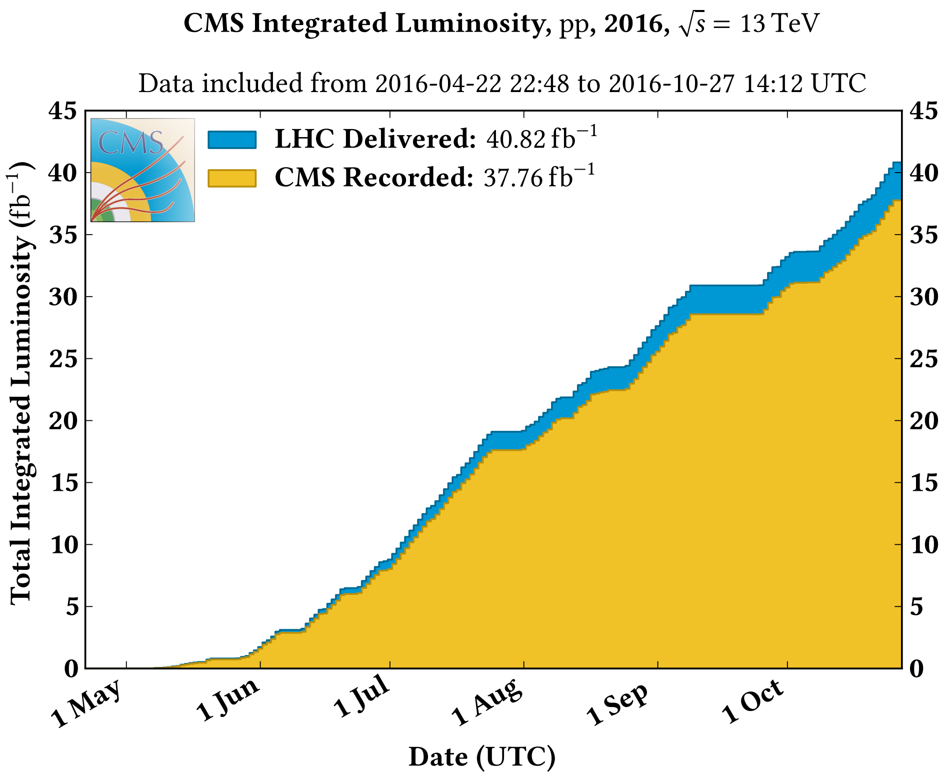
The design, construction, and installation of the detector systems at CMS presented a tremendous engineering challenge due to the high demands placed upon them. In accordance with its aim of being a general-purpose detector and discovering the electroweak symmetry breaking mechanism, a high-precision tracking system, ECAL system, and Muon System were required to cover a large solid angle from the *interaction point* (IP). The detectors and the associated electronics needed to be capable of reading out at the bunch crossing frequency of 40 MHz and withstand the high radiation environment close to the IP.

The period 2010–2020 comprises *Phase-1* of the LHC's operation. During this period, the observed instantaneous luminosity exceeded the design value, reaching $2.066 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ in 2017 [133, 134]. In order to prepare for the new challenges this luminosity will create, and in reaction to the performance of the CMS detector so far, a suite of upgrades known as the *Phase-1 upgrade* is planned for CMS during Phase-1. Most of these upgrades are intended for *Long Shutdown 2* (LS2), a two-year shutdown of the LHC that began in 2019. It was known, however, that radiation damage to the pixel detector (a component of the inner tracker) would necessitate a replacement before then. Accommodating this, the Phase-1 upgrade of the pixel detector was performed during the short technical top between LHC operation in 2016 and 2017. This is discussed further in § 4.2.6.

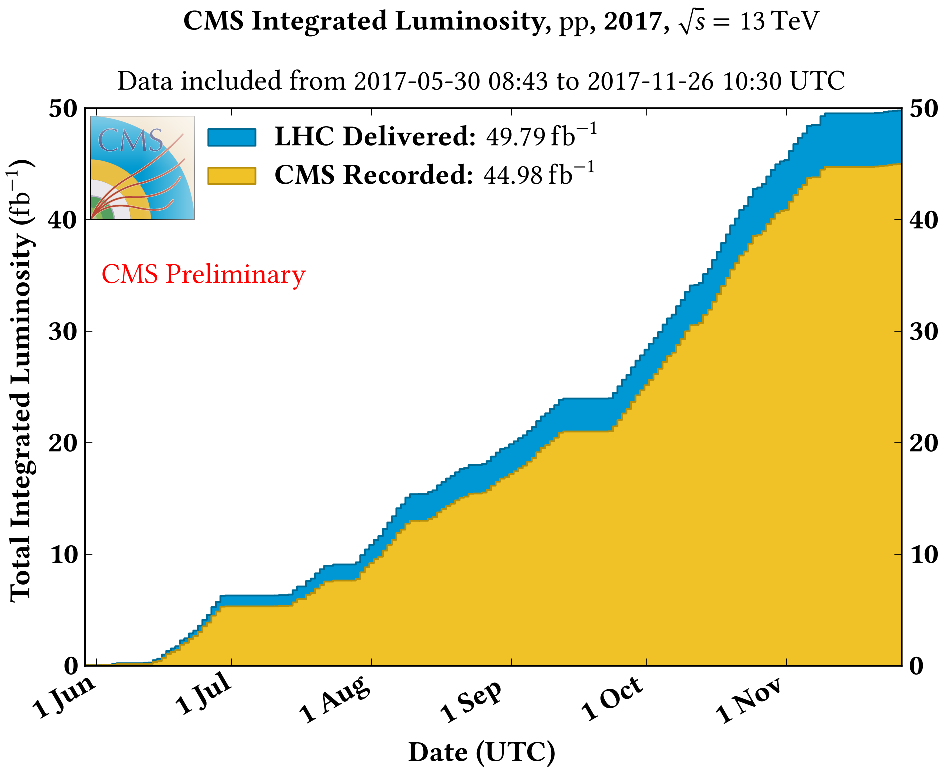
The integrated luminosity recorded by CMS for $\sqrt{s} = 13\,\mathrm{TeV}$ proton–proton collisions during 2017 and 2018 is shown in Figure 4.3. Both the total integrated luminosity delivered by the LHC and that recorded by CMS are displayed. The latter will gradually fall below the former because of transitory unavailability of detector systems or the data processing facilities being at capacity.

4.2.1 CMS Coordinate System

By taking the nominal IP as the origin, the x axis pointing directly towards the centre of the LHC ring, the y axis pointing directly upwards, and specifying a right-handed orientation, the 3D coordinate system used by CMS is constructed. This



(a)



(b)

Figure 4.3: Luminosity delivered and recorded by CMS; (a) shows 2016 and (b) shows 2017. Data-taking efficiency exceeds 90% in both years. Figures taken from [134].

results in a z axis parallel to the beam, heading approximately westward. Cylindrical and polar coordinates can be defined in the usual manner:

$$\begin{aligned} \varrho &= \sqrt{x^2 + y^2} & \varphi &= \arctan \left| \frac{y}{x} \right| \\ r &= \sqrt{x^2 + y^2 + z^2} & \theta &= \arccos \frac{z}{r}. \end{aligned} \tag{4.1}$$

The angle θ is often expressed in terms of *pseudorapidity*

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \tag{4.2}$$

$$= \operatorname{arctanh} \frac{p_z}{|\boldsymbol{p}|}. \tag{4.3}$$

In the *ultrarelativistic limit*, $|\boldsymbol{p}| \gg m$, η approaches the *rapidity*

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \tag{4.4}$$

Differences in rapidity are invariant for Lorentz transformations along the z axis, allowing angular separation to be defined as a quantity, $\Delta\eta$, independent of the net momentum of a collision.

Many of the detector systems in CMS are divided into *barrel* and *end cap* regions. The barrel refers to detectors that surround the IP at constant ϱ. End cap detectors cover higher rapidity regions and sit on planes of constant z.

4.2.2 Particle Tracking and Calorimetry

Charged particle tracking in CMS relies on two phenomena: a charged particle moving in a magnetic field experiences a force perpendicular to its direction of motion and ionises matter it comes into contact with.

The force experienced by a particle moving through a magnetic field is

$$F = q v_{\mathrm{T}} B \tag{4.5}$$

where q is the particle's charge, v_{T} its velocity perpendicular to the field, and B the strength of the magnetic field perpendicular to the direction of motion. The generated force, F, will be perpendicular to both the field and the direction of motion. Inside CMS's solenoid, the magnetic field is uniform and in a direction parallel to the z axis and so v_{T} represents motion transverse to the beam. Furthermore, in the absence of any other outside forces, the motion of a charged particle in a uniform magnetic field must be curved, which allows the above expression to be equated to that for the centripetal force:

$$\begin{aligned} F = qv_\mathrm{T}B &= \frac{mv_\mathrm{T}^2}{R} \\ \Rightarrow p_\mathrm{T} &= RqB \end{aligned} \tag{4.6}$$

where R is the radius of curvature, $m = \frac{p_\mathrm{T}}{v_\mathrm{T}}$ is the particle's mass, and p_T is momentum transverse to the beam (*transverse momentum*). In CMS, B is known and q will always be $\pm e$ as there is no known charged particle with $|q| \neq e$ possessing a lifetime long enough for it to enter the tracking system before decaying. The missing information needed to determine p_T can be found by tracking: the curvature of the track provides R and the direction of curvature the sign of q. The total momentum, $|\boldsymbol{p}|$, can then be obtained using

$$\begin{aligned} |\boldsymbol{p}| &= \sqrt{p_x^2 + p_y^2 + p_z^2} \\ p_x &= p_\mathrm{T}\cos\varphi \\ p_y &= p_\mathrm{T}\sin\varphi \\ p_z &= p_\mathrm{T}\cot\theta = p_\mathrm{T}\sinh\eta \\ \Rightarrow |\boldsymbol{p}| &= \sqrt{p_\mathrm{T}^2\left(1 + \sinh^2\eta\right)} \\ &= p_\mathrm{T}\cosh\eta. \end{aligned} \tag{4.7}$$

To acquire this information from charged particles, tracking detectors exploit the phenomenon of ionisation. Charged particles passing through matter deposit energy via various mechanisms, depending on the energies, materials, and particles involved. Identifying and measuring these deposits forms the operational basis of all charged particle detectors in CMS.

Those detectors at CMS that are not primarily concerned with tracking are, instead, primarily concerned with calorimetry. Besides providing a measure of energy for charged particles, calorimeters provide the only response from neutral particles in the CMS detector. Calorimeters function by exploiting the particle showers created when high energy particles enter a dense medium. Two types of shower can be initiated in this way: *electromagnetic showers*, initiated by an electron or photon; and *hadronic showers*, initiated by hadrons. The function of a calorimeter is to initiate such a shower and allow the energy of the daughter particles to be collected. This is achieved with a scintillator—transparent materials exhibiting luminescence upon the absorption of radiation. Light emitted by the scintillator can be collected, its energy measured, and the total ideally equalling the energy deposited. The scintillator can be paired with alternating layers of an absorber material tasked with initiating a shower, the resulting design known as a *sampling calorimeter.* Alternatively, the scintillating material can double as the absorber—a *homogeneous calorimeter.*

The characteristics of electromagnetic showers and hadronic showers are distinguished enough that constructing detectors that specialise in each is necessary. Hadronic showers are deeper and broader than their electromagnetic counterparts, making containing the shower more difficult and the resulting energy measurements less precise. The calorimeter systems used in CMS and the associated shower phenomenology are detailed in § 4.2.4 and § 4.2.5.

4.2.3 The Muon System

By virtue of its position as the outermost detector system at CMS, the Muon System is expected to experience far less particle flux than the rest of the detector. Only long-lived weakly ionising particles—namely muons and neutrinos—are expected to reach this component of the detector. Neutrinos react too weakly to be reliably detected even by dedicated neutrino experiments; their presence in an event is best inferred from missing energy and momentum, as explained in § 5.3.2. The Muon System then, as the name would imply, concerns itself only with muons.

Uniquely among the other major detector systems at CMS the Muon System resides entirely outside the solenoidal coil. The magnetic field is therefore not uniform. This motivates the presence of the steel return yoke, used to guide and confine the magnetic field.

Thanks to the reduced particle flux, the Muon System can employ detectors with longer response times than would be acceptable elsewhere. Three types of detector are used in the Muon System, from longest to shortest response time these are: *drift tube chambers* (DT chambers), *cathode strip chambers* (CSCs), and *resistive plate chambers* (RPCs). These three detectors are all used for tracking, with RPCs also used for triggering. Figure 4.4 shows a cross section of one quadrant of the muon system, demonstrating how the barrel region is split into five wheels along the z axis numbered from -2 to $+2$. Each wheel is split into 12 sectors running around the beamline.

The DT chambers have a response time of up to 380 ns [136] and so are located on the barrel, where the lowest particle flux is expected. They cover the $|\eta| > 1.2$ region. Every sector contains four DT chambers interposed with steel yoke, except for the uppermost and lowermost chambers of a given wheel. These are split in two to ease the assembly of CMS.

A schematic of a DT chamber is given in Figure 4.5. Each chamber contains 12 layers of DT chambers, grouped into three *superlayers* (SLs) of four layers each.

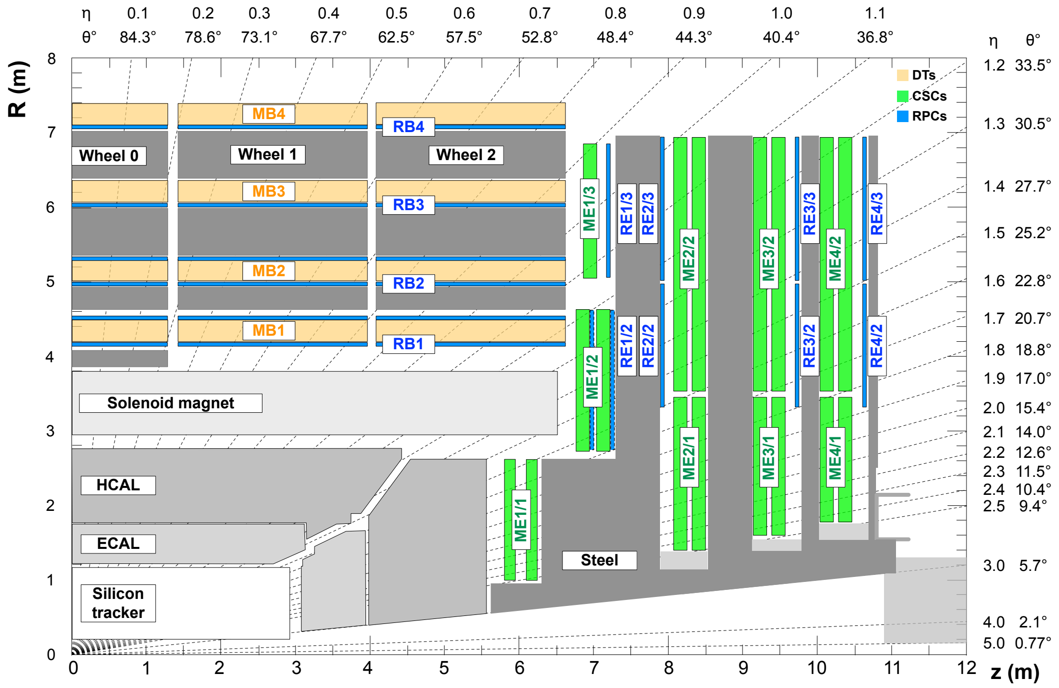


Figure 4.4: Diagram showing one quadrant of the CMS Muon System. The extent of the steel yoke is shown along with the location of DT chambers (labelled MB), CSCs (labelled ME), and RPCs (labelled RB in the barrel and RE in the end cap). The division into wheels is seen in the barrel. Figure taken from [135].

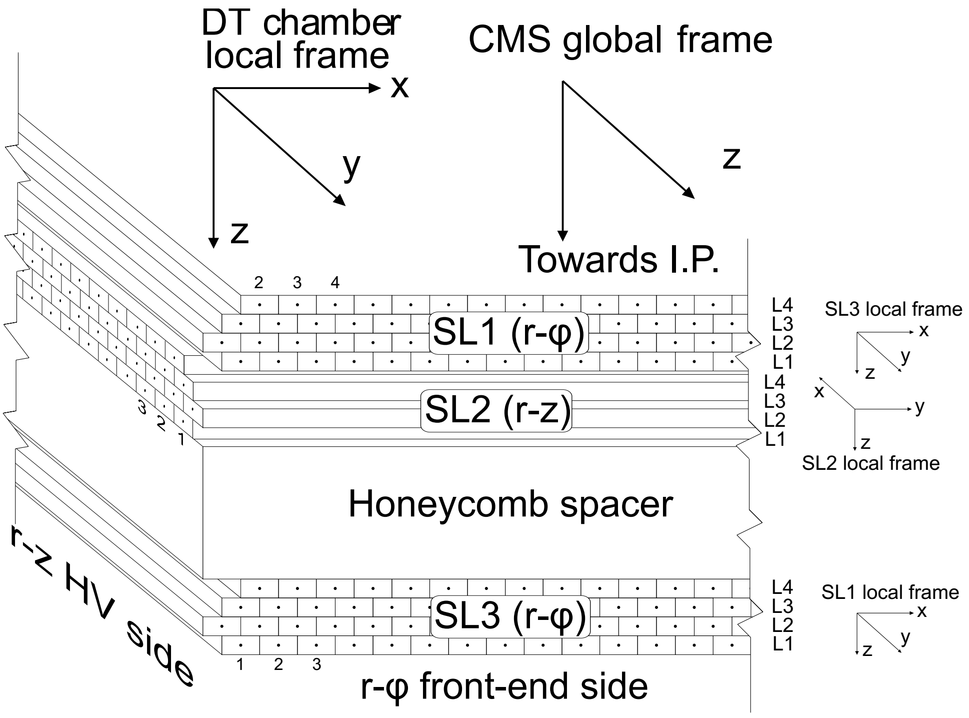


Figure 4.5: Layout of a CMS DT chamber.

SL1 (furthest from the beam) and SL3 (closest to the beam) contain DT chambers arranged in the $\hat{z}$ direction, while the DT chambers in SL2 run in the $\hat{\varphi}$ direction. This allows a chamber to provide two coordinates for a muon. Within each SL consecutive layers are offset half the width of a DT chamber in order to improve angular resolution.

A DT chamber contains an anodal wire running along the chamber's length. The chamber has a 13 mm × 42 mm cross sectional area along this wire, with the remaining volume filled with a gaseous mix of 85% Ar : 15% CO_2. When charged particles enter this gas, it is ionised and the freed electrons then drift toward the anodal wire. The drift velocity of the electrons in the gas is a known constant, so the time taken for the electrons to reach the wire provides the distance of the incident particle from the wire and hence a coordinate. The start time is provided by the RPCs, and an end time when a current is recorded in the wire.

Gas-based detectors like DT chambers are able to more cheaply cover a given volume than solid-state detectors, at the expense of response time. CSCs are also gas-based detectors but with a response time $\approx\frac{1}{85}$ of DT chambers', they can be used in areas of higher particle flux, such as the Muon System's end caps.

CSCs contain anodal wires running in the $\hat{\varphi}$ direction with perpendicular cathode strips in the $\hat{\varrho}$ direction. Charged particles passing through the 50% CO_2 : 40% Ar : 10% CF_4 gas within create a shower of positive ions and negative electrons, collected by the anode wires and cathode strips, respectively. The strips in which a current is observed provide the coordinates of the incident particle.

The fastest-responding detector used in the Muon System is the RPC. With a response time of 1.26 ns they can be used for triggering—unike the DT chambers and CSCs—and their measurements are used in track reconstruction.

The RPCs used in CMS are double-gap. Two 2 mm gaps between flat polyoxybenzylmethylenglycolanhydride (Bakelite) electrodes are filled with a mixture of 94.7% CH_2FCF_3 : 5% $(CH_3)_3CH$: 0.3% SF_6 gas. The external sides of the Bakelite electrodes are coated in a graphite paint and a copper readout plane is placed between the two gaps. A schematic is shown in Figure 4.6. An electric field is induced in the gas normal to the electrode plane at such a strength that electrons freed by incident particles will immediately cause further ionisations (*avalanche*). As the electrodes are resistive a signal is induced by the movement of charge in the electric field and not by the avalanched electrons directly, giving the RPC its excellent response time.

In $\sqrt{s} = 13$ TeV collisions, the p_{T} resolution for muons of $p_{\mathrm{T}} \lesssim 100$ GeV (where '$\lesssim$' indicates 'less than approximately') was found to be 1% in the barrel and 3% in the end cap when including information from the inner tracking system. For

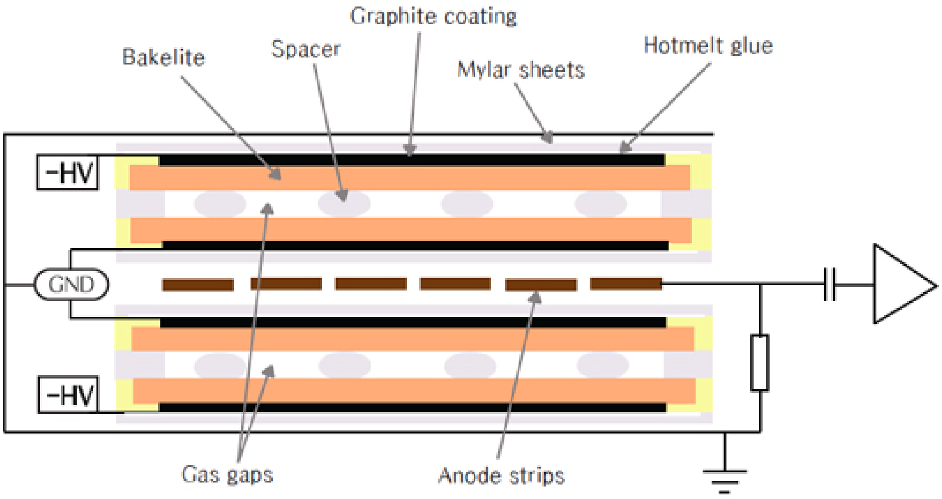


Figure 4.6: Cross section of a resistive plate chamber used in the CMS Muon System. Figure taken from [137].

muons in the range 100 GeV $< p_T <$ 1 TeV, the resolution in the barrel is better than 7%. The overall hit efficiency for the Muon System was 94–99% [138].

4.2.4 The Hadronic Calorimeter

The next layer of the CMS detector is the HCAL, a sampling calorimeter designed to measure the energy of hadronic particles and use energy deposit locations to assist with tracking. The HCAL is designed to tolerate significantly higher flux than the Muon System and absorb as much energy as possible to maximise the energy resolution.

Hadrons entering an absorber experience inelastic nuclear collisions. The mean distance travelled before an interaction takes place—the nuclear interaction length, λ_I—characterises the depth of a hadronic shower. Inelastic collisions produce secondary π^0, $\pi^\pm$, η, p, etc. that interact with the medium through a variety of phenomena. Approximately 20–30% of the energy deposited is used to overcome nuclear binding energies; losses like this are invisible to the detector, and provide a further challenge in hadronic calorimetry [18].

The HCAL consists of four subdetectors. First is the *HCAL Barrel* (HB), which covers the pseudorapidity range $|\eta| < 1.3$. The HB lies inside the solenoid but is supplemented by the *Outer HCAL* (HO), which lies outside. This is necessitated by the limited space inside the solenoid: the typically long λ_I of even the most effective absorbers means that a significant fraction of energy cannot be captured by the HB. Third is the *HCAL End Cap* (HE) covering $1.3 < |\eta| < 3$, and finally the *Forward HCAL* (HF) covers $3 < |\eta| < 5.2$ [136].

In the HB a brass absorber is used, except for the outermost and innermost layers. These are steel to ensure the structural integrity of CMS. The outermost steel layer is 75 mm thick, followed by six 50.5 mm thick brass layers, a further eight 56.55 mm thick brass layers, then a final 40 mm thick steel layer, for a total of 16 absorbing layers. The interceding plastic scintillator layers are 3.7 mm thick. These strata are arranged into 36 azimuthal wedges. In all, the HB has a minimum depth of $5.8\lambda_\text{I}$ at $\eta = 0$, increasing with $|\eta|$ to a maximum of $10.6\lambda_\text{I}$.

The HO is divided into five rings that match the five wheels of the muon system. It has only one absorbing layer: the solenoid itself, equal to $1.4 \csc\theta$ nuclear interaction lengths. This is then immediately followed by a single plastic scintillating layer. This design is modified in Ring 0 (the central ring), where the effective absorption length is shortest, by adding a second plastic scintillating layer with an interstitial 19.5 cm thick iron absorber.

The HE consists of plates containing 79 mm thick brass and 9 mm thick plastic scintillator layers on the end cap. In tandem with the HB and HO, the HE ensures that a particle must traverse $11.8\lambda_\text{I}$ in the $|\eta| < 3$ region, excepting a small section where the HB and HE meet.

The final component of the HCAL is the HF. This region experiences eight times the amount of deposited energy in a typical collision compared to the rest of the HCAL. Ensuring the HF could withstand this harsh environment was therefore paramount. As such, the HF uses a different design to the rest of the HCAL.

The HF relies on *Čerenkov radiation*, a form of radiation emitted by charged particles moving faster than the phase velocity of light in a medium. Čerenkov radiation peaks in the ultraviolet region [139] and is emitted at the Čerenkov angle [18]

$$\theta_\text{C} = \arccos\left(\frac{1}{n\beta}\right) \tag{4.8}$$

from the particle's direction of motion, where n is the refractive index of the medium and β is the particle's velocity as a fraction of the speed of light in a vacuum.

Steel is used as the absorber in the HF, with 5 mm thick plates layered to a total depth of 165 cm ($\approx 10\lambda_\text{I}$). Quartz fibres running in the $\hat{z}$ direction are fitted into grooves on these plates. They pick up Čerenkov radiation emitted by particles from showers triggered by incident high-energy particles. Half of these fibres do not extend the full length of the detector, starting at a depth of 22 cm. This allows the HF to identify photons and electrons in addition to hadrons, as the former deposit most of their energy in that length. The accumulated Čerenkov radiation is funnelled to photomultiplier tubes contained in readout boxes.

The energy resolution of the HCAL is parametrised as

$$\frac{\sigma_E}{E} = \sqrt{\frac{S^2}{E} + C^4}. \tag{4.9}$$

The first term represents the stochastic element, originating from intrinsic shower fluctuations, and the constant term arises from the physical limitations of the HCAL. This includes nonuniformity and radiation damage, but is dominated by energy leaking through the rear of the HCAL. A test beam of electrons, pions, protons, and muons [140] was used to determine the resolution. Where appropriate, this beam was first passed through a lead tungstate crystal to control for the effect of the ECAL. In the HB+HO barrel and HE, $S = 0.847 \pm 0.016\,\sqrt{\mathrm{GeV}}$ and $C = 0.074 \pm 0.008$ [141]. Meanwhile, in the HF, $S = 1.98\,\sqrt{\mathrm{GeV}}$ and $C = 0.09$ [141].

4.2.5 The Electromagnetic Calorimeter

Unlike the HCAL, the ECAL is a homogeneous rather than a sampling calorimeter. The material chosen for the ECAL is lead tungstate ($PbWO_4$) in crystalline form. Lead tungstate is radiation-hard, scintillating, and produces shallow, thin electromagnetic showers.

Two processes work together to create an electromagnetic shower: *bremsstrahlung radiation* ($e^\pm \rightarrow e^\pm\gamma$) and *pair production* ($\gamma \rightarrow e^-e^+$). Bremsstrahlung (‘breaking radiation’) is emitted when an electron (or any charged particle) decelerates, and a photon can undergo pair production when it has an energy greater than double the electron rest mass. Daughter electrons and photons can then produce further photons or electrons via the same processes, cascading to form an electromagnetic shower. This process ends when the energy loss of the electron is dominated by ionisation, occurring below the critical energy

$$E_\mathrm{C} \approx \frac{600}{\mathrm{Z}}\,\mathrm{MeV} \tag{4.10}$$

with Z as the average atomic number of the material [18]. A shower’s depth is described by the radiation length, X_0, defined as the distance over which an electron will lose $\frac{1}{\mathrm{e}}$ of its energy to bremsstrahlung radiation (the mean free path of a photon before undergoing pair production is $\frac{9}{7}X_0$). The Molière radius, R_M, describes the breadth of a shower—it is the radius of a cylinder containing 90% of a shower’s energy. In the ECAL, the lead tungstate crystals used have $X_0 = 0.89\,\mathrm{cm}$ and $R_\mathrm{M} = 2.2\,\mathrm{cm}$.

The ECAL contains 68 524 lead tungstate crystals. Most of these—61 200—reside in the barrel, which covers the pseudorapidity range $|\eta| < 1.4442$. The remaining 7524 crystals reside in the end cap, covering the range $1.566 < |\eta| < 2.5$ [142].

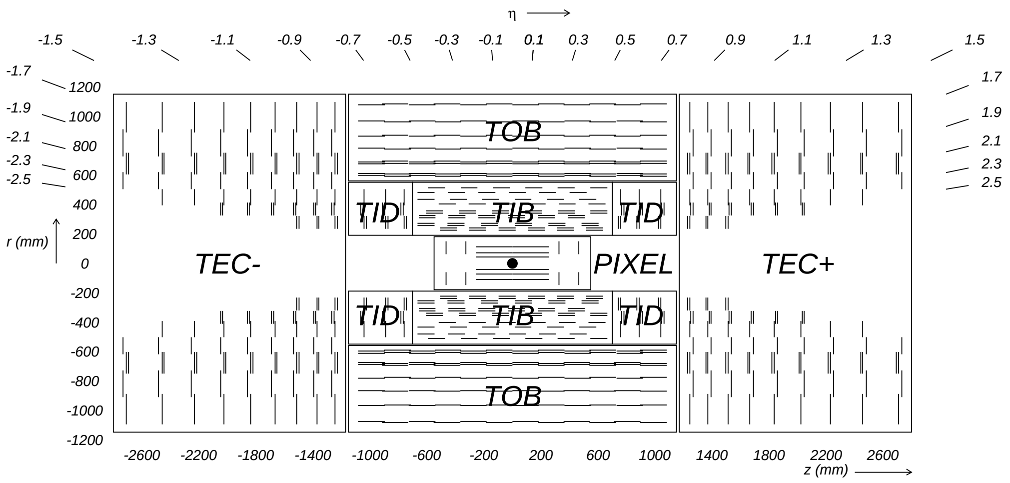


Figure 4.7: Layout of the Phase-0 CMS inner tracker, with each line representing a module an double lines back-to-back modules. The nominal IP is shown as a black dot in the centre of the pixel detector at $z = 0\,\text{mm}$. Figure taken from [136].

Barrel crystals are 23 cm ($25.8X_0$) long, while end cap crystals are 22 cm ($24.7X_0$) long.

In the $1.653 \leq |\eta| < 2.6$ end cap region, the ECAL system employs a preshower detector. Its design is a hybrid between a calorimeter and solid-state tracker, consisting of two layers of lead absorber plate backed by a silicon detector to a total depth of 20 cm. At $\eta = 1.653$ the first lead plate has a depth of $2X_0$ and the second $1X_0$. The enhanced resolution in the preshower allows electrons to be identified more reliably by examining the properties of their showers, and aids track reconstruction.

The energy resolution of the ECAL for electrons hitting the centre of a barrel crystal was measured in a test beam to be [143]

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{12\%}{E}\right)^2 + (0.3\%)^2} \tag{4.11}$$

in $\sqrt{s} = 7\,\text{TeV}$ pp collisions. As for the HCAL the stochastic and constant terms account for the shower fluctuations and physical limitations, respectively. Like the HCAL, the constant term is dominated by energy leakage but in the ECAL the non-uniformity of light collection is another significant factor.

4.2.6 The Inner Tracker

The inmost detector system at CMS is the aptly named *inner tracker*. The largest silicon tracker system ever constructed, it extends to a radius of 1.1 m and length

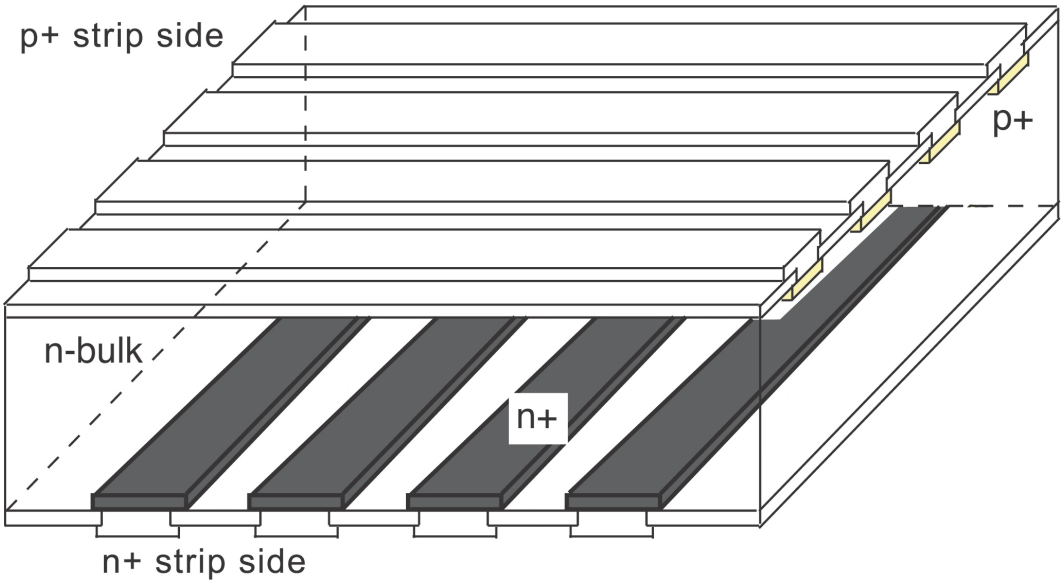


Figure 4.8: Cross section of a silicon-trip detector, showing the perpendicular n-type and p-type strips, with the intervening n-type depletion region (labelled n-bulk). Figure taken from [136].

of 5.8 m along the beam pipe, covering the range $|\eta| < 2.5$ [136]. It is split into five subdetectors—the *Tracker End Caps* (TECs), *Tracker Outer Barrel* (TOB), *Tracker Inner Disks* (TIDs), *Tracker Inner Barrel* (TIB), and *pixel detector*. The arrangement of the components is shown in Figure 4.7. Designed to be the primary source of tracking information in CMS, the inner tracker uses solid-state silicon detectors because of their fast response, high resolution, and radiation resistance.

Silicon is a semiconductor: a substance with a conductivity between an insulator and a conductor. By doping—introducing impurities into the silicon—the nature of the free charge carriers in silicon can be altered. *N-type* silicon contains an excess of electrons, which act as the charge carriers. *P-type* silicon, on the other hand, has an electron deficit. The missing electrons in the lattice behave as positively-charged charge carriers known as *holes*.

Joining p-type and n-type silicon creates a p–n junction [144]. Current will only travel from the p-type silicon to the n-type silicon i.e. the junction forms a diode. Connecting the p-type region to a negative terminal and the n-type to a positive terminal creates a reverse bias, which depletes the junction of charge carriers. The silicon trackers in the inner tracker all use an n-type depletion region. Ionising particles entering this depletion region create electron–hole pairs in the depleted region, which quickly flow to the n-type and p-type regions, respectively, resulting in a measurable current. The arrangement of the n-type depletion region, p-type region, anode, and cathode depends on the type of detector.

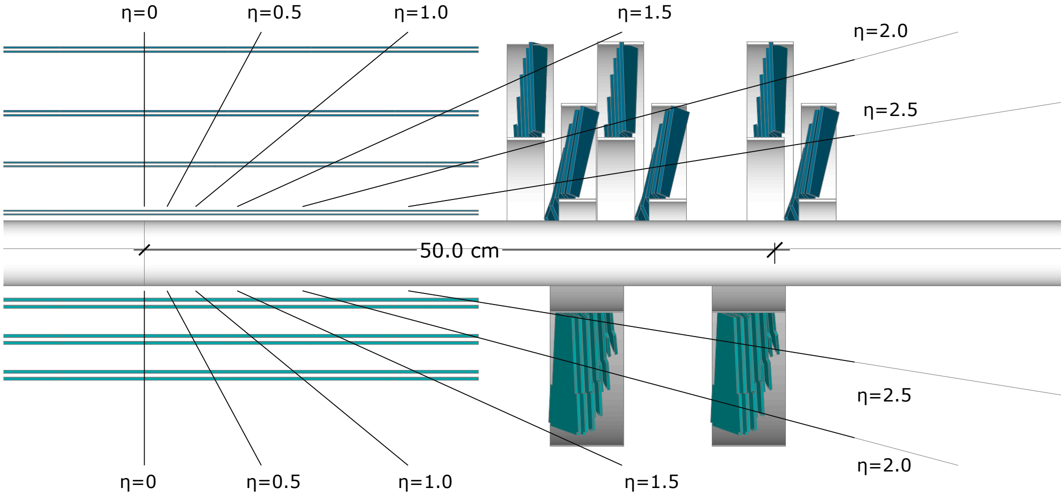


Figure 4.9: Comparison of the layout of the pixel Phase-1 upgrade detector (top) and original Phase-0 detector (bottom) on either side of the beamline. The nominal IP is marked to the left, and a schematic cross section of the barrel and end cap is shown. Figure taken from [145].

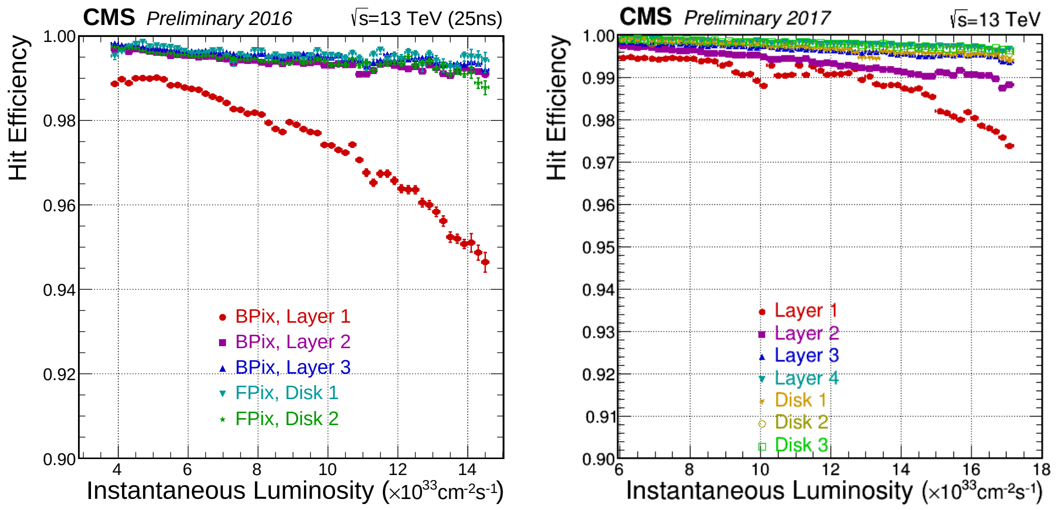
A diagram of a silicon strip module is shown in Figure 4.8. Charge carriers produced in the n-type depletion region are collected by the perpendicular n-type and p-type strips, giving position information in two axes.

The TOB contains six layers of 500 µm thick strip detectors complemented by four layers of 320 µm thick strip detectors in the TIB. In the TECs there are nine layers (*disks*) of strip detectors, with each disk split into concentric *rings*. Each disk contains a variable number of rings decreasing with $|z|$ as seen in Figure 4.7. The strip detectors in the three innermost rings on each disk are 320 µm thick, and 500 µm thick in the remaining rings. Finally, the TIDs contain three disks of three rings each, with all strip detectors 320 µm thick. Certain modules are *stereo*—formed of two back-to-back strip detectors—and are indicated on Figure 4.7.

The remaining inner tracker detectors are silicon pixel detectors. The appellative *pixel* refers to the construction: millions of tiny (100 µm × 150 µm) independent silicon detectors, or “pixels”. Each pixel is an n-type depletion region sandwiched between a cathode and an anode. In total, over 66 million individual pixels are used in the inner tracking system.

Radiation damage during operation demanded an upgrade to the pixel detector at the beginning of 2017. In this upgrade the original *Phase-0* pixel detector was replaced by the Phase-1 pixel detector.

The Phase-0 pixel detector consisted of three layers in the barrel and two discs in the end cap. The Phase-1 upgrade added a layer of pixel modules in the barrel [146].



(a) (b)

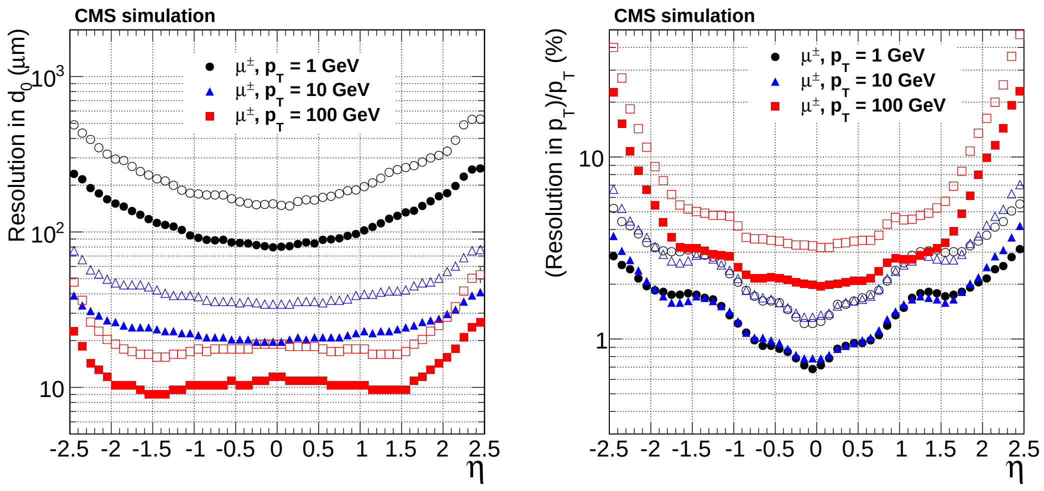
Figure 4.10: Plot of hit efficiency against luminosity for the Phase-0 detector (a) and the Phase-1 detector (b). BPix (FPix) refers to the modules in the barrel (end cap), with the layers (disks) therein enumerated from the inside out. Figures taken from [148].

In the end cap, the number of disks was increased from two to three and each disk now contains two partially overlapping rings. A comparison between the Phase-0 layout and Phase-1 layout is shown in Figure 4.9. The hit efficiency of the Phase-1 and original Phase-0 pixel detectors is shown in Figure 4.10. Without this upgrade a hit inefficiency of 16% in the innermost barrel layer was predicted [147], but now the Phase-1 detector shows greater efficiency in all layers compared to the Phase-0 detector. In particular the efficiency in the first layer of the barrel—which is located closest to the IP—shows a marked improvement.

In Figure 4.11 the p_T and impact parameter (d_0, the distance of the closest approach between a track and the IP) resolutions achieved in the Phase-0 tracker are shown. The study was performed on isolated muons of $p_T = 1\,\text{GeV}$, 10 GeV, and 100 GeV. The resolution worsens with $|\eta|$ as the distance between the IP and the first incidence observed in the tracker increases. Near $|\eta| = 1.5$ a small loss of resolution can be seen, caused by the gap between the barrel and end cap components [142].

4.3 The CMS Trigger System

The total amount of raw data generated by the CMS detector systems per event is 1–2 MB [150]. At a typical event rate of $\sim 10^6$ events per second, this creates $\sim 1\,\text{TB s}^{-1}$ of data to be processed and archived. This falls well outside the scope of modern computing hardware.



(a)

(b)

Figure 4.11: Resolution of d_0 (a) and p_{T} (b) versus η for tracks reconstructed using the Phase-0 inner tracker. Solid markers show the 1σ band and hollow markers the 2σ band. [149]

The total proton–proton cross section at $\sqrt{s} = 13\,\mathrm{TeV}$ is $110.6 \pm 3.4\,\mathrm{mb}$ [151], of which $28 \pm 2\%$ are elastic collisions. The inelastic cross section is dominated by QCD processes in which the momentum transfer, Q, is less than the energy at which non-perturbative effects are significant (Λ_{QCD}); these are known as *soft processes*. Consequently, the total inelastic proton–proton is many orders of magnitude greater than the majority of physics processes studied at CMS, which have cross sections at the picobarn order. Therefore, even at a theoretical CMS experiment capable of processing and recording all data from every single collision, most of the data would not be needed to further CMS's physics goals. Filtering these unnecessary events before full event reconstruction takes place would reduce the event rate to a manageable level without adversely affecting physics performance. This is the task taken up by the CMS trigger system.

The CMS trigger system consists of two stages: the *Level 1 Trigger* (L1 Trigger), and the *High Level Trigger* (HLT). The L1 Trigger consists of custom hardware—viz. *application-specific integrated circuits* (ASICs) and *field programmable gate arrays* (FPGAs)—built into the CMS detector. It makes a decision on whether to pass the event data to the HLT within $4\,\mu\mathrm{s}$; this extreme latency constraint means only information from the muon and calorimeter systems is available to the L1 Trigger.

Despite the limited time frame and available data, the L1 Trigger is able to perform some rudimentary physics object identification and $\not{E}_{\mathrm{T}}$ calculation [152] (see § 5.3.2 for the definition of $\not{E}_{\mathrm{T}}$). Event properties and objects determined by the

L1 Trigger will determine which, if any, L1 Trigger *flags* will be set for that event. These flags mark potentially "interesting" events. For example, a flag can be set for events that seemingly contain two muons, or exceed a certain $\not{E}_{\mathrm{T}}$ threshold. No flag will be set for most events and the associated data is discarded. On the other hand, should any flag be set the data is passed to the HLT for further processing. Ultimately, the L1 Trigger reduces the event rate to ~100 kHz.

In contrast to the L1 Trigger, the HLT uses consumer-grade general-purpose *central processing units* (CPUs) and, owing to an increased time budget of 175 ms per event, is able to perform more advanced reconstruction using data from all CMS detector systems. The time budget is still, however, the principal consideration: full event reconstruction is not performed at this stage, and the HLT computing hardware is positioned in a cavern adjacent to the CMS detector in order to minimise latency.

Event filtering at the HLT stage is constructed around HLT *paths*. A path is a succession of event processing steps used to classify an event as containing potentially interesting physics objects or properties. Conceptually, this is similar to the flags set by the L1 Trigger, but afforded a greater degree of granularity and sophistication by the increased processing time. HLT paths will be activated (or not) depending on the L1 Trigger flags set for that event: for instance, if an L1 Trigger flag indicating the presence of an isolated electron or photon is set, HLT paths designed to select events containing single, isolated electrons would be enabled. If an event is found, using the HLT's reconstruction process, to have the properties required by a given path it is said to have *passed* that path. Events that pass any activated HLT path are forwarded for complete reconstruction at a rate of ~100 Hz.

Chapter 5

Event Reconstruction

> *To achieve great things, two things are needed: a plan, and not quite enough time.*
>
> Leonard Bernstein

In order to make the enormous amount of data from the CMS detector systems meaningful, the particles created in collisions must be reconstructed. This process begins with the identification of tracks in the inner tracker, tracks in the Muon System, and clustered energy deposits in the calorimeters. From these, a list of particles present in the event is then constructed using the *Particle Flow* (PF) algorithm [153]. Finally, *high-level objects*—such as jets and $\not{E}_\mathrm{T}$—are determined.

5.1 Low-Level Physics Object Reconstruction

The first step in the physics object reconstruction involves the identification of *low-level objects* on a detector-by-detector basis; information from the different detector systems is not, at this stage, combined. These low-level objects are charged particle tracks, primary vertices, and energy clusters in the calorimeters.

5.1.1 Charged Particle Tracks

As discussed in § 4.2.2, the reconstruction of a charged particle track allows the particle's p_T to be determined. In addition, once a primary vertex has been found as described in § 5.1.2, the impact parameter of the charged particle can be found.

Two metrics are defined to describe the quality of tracking: the percentage of real tracks correctly reconstructed is the *tracking efficiency* and the percentage

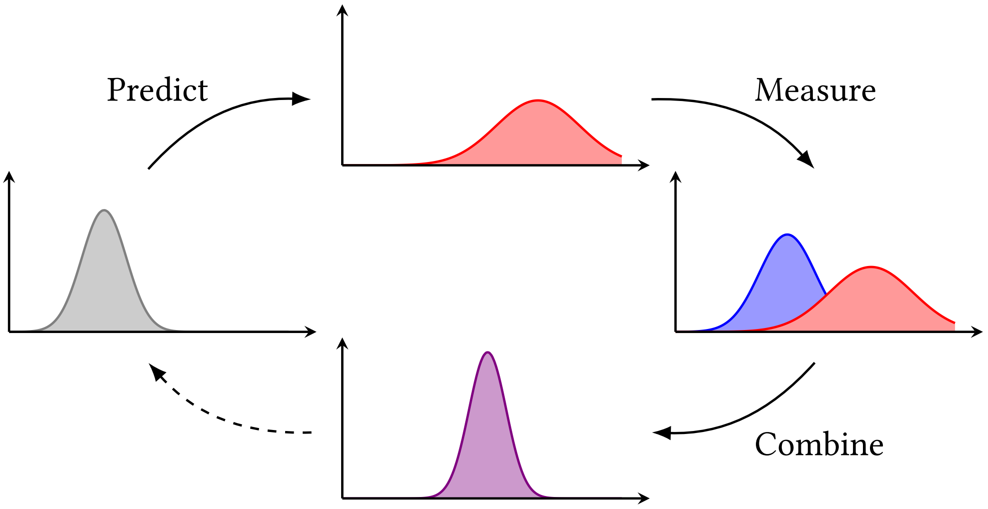


Figure 5.1: Diagram showing the basic operation of a KF. The initial state of the system, $\boldsymbol{x}(k-1)$, (grey) is used to make a prediction of $\boldsymbol{x}(k)$ (red). A KF is used to combine this prediction with a measurement of $\boldsymbol{x}(k)$ (blue) to create an estimate of $\boldsymbol{x}(k)$ more accurate than either the prediction or measurement (purple). This process can then be repeated, using the KF's estimate as the initial state to create a new estimate using a KF of $\boldsymbol{x}(k+1)$.

of reconstructed tracks with no real track analogue is the *fake rate*. To achieve the best performance, the *Combinatoral Track Finder* (CTF) algorithm is used [149]. The CTF utilises a *Kalman filter* (KF) to combine energy deposits in the inner tracker—known as *hits*—to form tracks.

Kalman filters

The KF is an algorithm allowing the creation of optimal predictions of dynamic systems based on uncertain information [154]. Consider a system that can be entirely specified by $\boldsymbol{x} = (x_1, x_2, x_3, \ldots)$ that evolves with k. Only noisy measurements of $\boldsymbol{x}$ can be made, and there exists a dynamic model of the system i.e. a method to predict the value of $\boldsymbol{x}$ at k based on the value of $\boldsymbol{x}(k-1)$. A KF combines this prediction with noisy measurements of $\boldsymbol{x}(k)$ to produce an estimate of $\boldsymbol{x}(k)$ that is more accurate than either individually. For linear systems, a KF will produce an optimal estimate: no other technique will result in a smaller mean squared error.

The estimate of $\boldsymbol{x}(k)$ obtained by the KF can then be used to predict $\boldsymbol{x}(k+1)$ using the dynamic model and, in combination with a noisy measurement of $\boldsymbol{x}(k+1)$, can create an estimate of $\boldsymbol{x}(k+1)$ using another KF. An overview of this process is presented in Figure 5.1. This estimate will be better than if a single KF had been used, starting at k. Despite this, the second KF only needed information from the preceding state. This gives the algorithm a low memory overhead and, combined with its relative simplicity, makes it ideal for track fitting at CMS. In this context,

$\boldsymbol{x}$ represents the coordinates of a hit in the detector; additional candidate hits for a track are added one at a time, with a KF used to combine the predicted motion of the particle with the hit. In this iterative manner, the uncertainty in a track candidate decreases with the inclusion of each additional hit [155].

The Combinatorial Track Finder algorithm

The CTF algorithm can be split broadly into four stages:

1. Pairs and triplets of hits in the inner tracker are used to create initial estimates of helical charged particle tracks. This is known as *track seeding*.

2. These *proto-tracks* are extrapolated and used to identify further candidate hits in other tracker layers, which are then combined with the proto-track using a KF [156]. The algorithm searches for hits in each tracker layer, allowing both zero and multiple hits in a layer. In the latter case, a track candidate is created for each hit, which can lead to track candidates sharing many of the same hits.

3. Each track candidate is refit twice using a KF: first starting from the innermost hit outwards and second starting from the outermost hit inwards (the KF in the second fit is initialised with the result of the first). An average of the two resultant tracks is used to provide the optimal track parameters.

4. Low-quality track candidates are discarded. The quality of a track candidate is determined by the quality of the fit, the number of intersecting layers that contain hits, and its compatibility with originating from a primary interaction vertex (see § 5.1.2). The fraction of shared hits between two track candidates is calculated as

$$f_{\text{shared}} = \frac{N_{\text{shared}}}{\min(N_1, N_2)} \tag{5.1}$$

   where N_{shared} is the number of shared hits and N_1 (N_2) is the number of hits used in forming the first (second) track candidate. If, for any pair of track candidates, f_{shared} exceeds 19% the track candidate with fewer hits is discarded (if $N_1 = N_2$ the track with the lowest quality fit is discarded).

Application of the CTF is performed six times per event. Each iteration is tuned differently: the location and quantity of hits needed to form proto-tracks is changed as are the p_{T} and impact parameter requirements for high-quality track candidates. Finally, the selection between track candidate pairs with $f_{\text{shared}} > 19\%$ is performed on the collection of all track candidates from every iteration.

5.1.2 Primary Vertex Reconstruction

Reconstructed tracks that are well-fit, have two or more hits in the pixel detector, five or more hits in the tracker, and a small impact parameter with respect to the beam line are used to reconstruct *primary interaction vertices* [149]. A primary interaction vertex is a point where a proton–proton collision took place. In any given event there will be multiple such vertices, because of pileup. The primary interaction vertex with the greatest total p_{T} is chosen as the *primary vertex*, with those remaining considered to be pileup vertices.

Primary interaction vertices are found by using a deterministic annealing algorithm [157] to cluster tracks according to the z coordinate of their closest approach to the beam axis. This is followed by an adaptive vertex fit [158] to determine the position of each possible vertex from its associated tracks.

5.1.3 Calorimeter Energy Clusters

The HCAL and ECAL detectors are divided into *cells*. Incident particles create deposits in groups of neighbouring cells (*clusters*) that must be identified to correctly reconstruct the location and energy of the particle that caused the deposit.

The large cell size of the HF renders clustering unnecessary: each individual cell is considered a cluster. In the remaining calorimeters the clustering algorithm begins by identifying *seed cells*. Seed cells must have deposited energy in excess of a threshold that varies by subdetector and must contain more deposited energy than any neighbouring cells. Clusters are grown from these seeds by adding cells that share at least one corner with a cell already in the cluster and contain deposited energy in excess of two standard deviations from the expected electronics noise. This equates to 800 MeV in the HCAL, 80 MeV in the ECAL barrel, and 300 MeV in the ECAL end cap [153].

5.2 Particle Flow

The PF algorithm [153] uses information from all CMS detector systems to reconstruct the particles present in an event. It begins with a linking step, connecting tracks from the inner tracker (and potentially the Muon System) to energy clusters in the calorimeter systems. The resulting linked objects are known as *blocks*. The PF algorithm then attempts to identify the particle responsible for each block in the following order: muons, electrons, neutral hadrons, photons, and finally charged hadrons. Once a block has been associated with a particular particle type it is removed from further consideration.

5.2.1 The Linking Algorithm

Charged particle tracks found by the CTF are projected into the calorimeter systems. If the projected track lies within an identified calorimeter cluster, the track and cluster are linked as a block. Clusters in the ECAL and HCAL can be linked too, provided the former lies within the boundary of the latter.

This process is complicated slightly by photons emitted as bremsstrahlung radiation. Photon trajectories are not bent by the magnetic field and so will propagate at a tangent to charged particle tracks. A charged particle will release bremsstrahlung radiation when interacting with material in the inner tracking system, hence a tangent is extrapolated from every hit in every track. Any ECAL cluster whose location is consistent with one of these tangents is flagged as the possible result of a bremsstrahlung photon.

5.2.2 Muon Reconstruction

Muon tracks, like any other, are reconstructed in the inner tracker using the CTF. An independent track reconstruction process is performed in parallel using the muon system alone. This process is similar to that performed in the inner tracker, using small seed track segments (much the same as proto-tracks) constructed using hits in the DT chambers and CSCs. Full tracks are then built using a KF that combines candidate hits in the DT chambers, the CSCs, and the RPCs.

The existence of two sets of independently reconstructed muon tracks allows for two separate methods of muon reconstruction:

Global (outside-in) All tracks reconstructed in the Muon System are projected inward, towards the inner tracker. They are then combined with the closest-matching track in the inner tracker using a KF. The resulting complete block is a *global muon*.

Tracker (inside-out) All tracks in the inner tracker that satisfy $p_{\mathrm{T}} > 0.5$ GeV and $|\boldsymbol{p}| > 2.5$ GeV are extrapolated to the Muon System. If a matching muon track segment is found the track is considered a *tracker muon*.

A muon could be reconstructed as both a tracker muon and a global muon. To account for this, any pairs of tracker muons and global muons that share the same track in the inner tracking system are merged. Approximately 99% of muons produced in a proton–proton collision within the Muon System η acceptance are identified by PF as either a global or tracker muon [138].

Rochester Corrections

Beyond the PF muon reconstruction, corrections that apply a scale factor to muon four-momentum known as the Rochester Corrections [159] are recommended for all CMS analyses with muons in the final state. The scale factor is derived from a muon's charge, p_{T}, η, and φ. These corrections correct for biases arising from detector misalignment and errors in the measurement of the magnetic field and alter the muon's p_{T} by $\lesssim 2\%$ in the tZq analysis.

5.2.3 Electron Reconstruction

Electrons are particularly susceptible to bremsstrahlung radiation, losing 33–86% of their energy to it in the inner tracker [160]. This poses a problem for electron reconstruction using a KF. While the KF formalism does not assume Gaussian noise, it only performs optimally in this case. Bremsstrahlung radiation, however, produces uncertainties that are not and cannot be reasonably approximated by a Gaussian, to the extent that electron reconstruction performance is significantly degraded when using a KF alone. To resolve this, a KF is used for an initial fit and is then refined with the more comprehensive—and computationally expensive—*Gaussian sum filter* (GSF) algorithm [161]. Bremsstrahlung radiation presents a further problem in the ECAL, namely the association of all bremsstrahlung energy deposits from an electron to the correct source. This is essential to fully reconstruct an electron's initial energy. The fact that the electron has a curved trajectory while that of the photon's is straight results in the energy deposits associated with a single electron potentially covering a range over φ. Clusters that fall into this pattern are thus arranged into *superclusters* through the use of two different algorithms: the *hybrid algorithm* in the ECAL barrel and the *multi algorithm* in the ECAL end cap [160].

Two different methods are used by the PF algorithm to identify electron track seeds [160]:

ECAL-driven The energy and layout of a supercluster is used to calculate the associated electron's hits in the inner tracker. This is most effective for isolated, high-p_{T} electrons: nonisolated electrons (e.g. those within jets) can have other tracker hits incorrectly associated with them, and the bremsstrahlung distribution of a low-p_{T} electron can exceed the bounds of a single supercluster.

Tracker-driven A KF is used to fit tracks in the inner tracker, followed by a GSF for tracks indicating significant bremsstrahlung energy loss. A discriminant from a BDT is used to select electron tracks, using quantities such as the number and quality of tracks as reconstructed by both the KF and GSF as features. This method complements the shortcomings of the ECAL-driven procedure, as it is more suited to low-p_{T}, nonisolated electrons.

Once identified, the seeds created by each method are consolidated. Subsequently, a final fit with a GSF is performed utilising the additional track information afforded by the merged seed collection, allowing for an improved reconstruction. An overall efficiency of more than 95% was estimated using simulated $\mathrm{Z} \rightarrow \mathrm{ee}$ decays [160].

5.2.4 Photon and Hadron Reconstruction

Reconstruction of isolated high-energy photons is performed simultaneously with the electron reconstruction procedure described above. Superclusters in the ECAL with $E_{\mathrm{T}} > 10\,\mathrm{GeV}$ (where $E_{\mathrm{T}} = E \sin\theta$), no linked track, and a photon-compatible energy ratio between the energy deposited in the ECAL and matching deposits in the HCAL are reconstructed as photons.

After all tracker hits and calorimeter deposits associated with electrons, muons, and isolated photons have been removed from further consideration, those remaining are considered by PF to be the product of either nonisolated photons or hadrons. Within the tracker acceptance, leftover clusters in the ECAL or HCAL with no associated tracks are classed as photons or neutral hadrons, respectively. Remaining tracks with associated clusters in the HCAL are considered to be charged hadrons, and can be linked with compatible ECAL clusters.

Outside the tracker acceptance it is impossible to distinguish between charged and neutral hadrons [153]. In this region, compatible ECAL and HCAL clusters are considered to originate from hadrons, while ECAL clusters not linked with HCAL clusters are considered the result of photons.

5.3 High-Level Physics Object Reconstruction

Once the PF algorithm has successfully reconstructed the particles in an event, higher-level objects can be constructed. These objects depend on the full particle listing and include composite objects, such as jets, and properties of the event as a whole, such as missing transverse energy.

5.3.1 Jets

Jets are cone-like sprays of hadronic particles originating from a partons [16], described previously in § 2.1.4. A sequential recombination algorithm is used to identify jets, evaluating entities (e.g. particles, pseudojets, labelled i, j herein) based on two metrics:

$$d_{\mathrm{B}\,i} = p_{\mathrm{T}\,i}^{2k} \tag{5.2}$$

and

$$d_{ij} = \min\left(p_{\mathrm{T}\,i}^{2k}, p_{\mathrm{T}\,j}^{2k}\right) \frac{\Delta R_{ij}^2}{R^2} \tag{5.3}$$

where

$$\Delta R_{ij}^2 = \left(y_i - y_j\right)^2 + \left(\varphi_i - \varphi_j\right)^2 , \tag{5.4}$$

y is rapidity, and R is the *jet size parameter*. The metrics d_{ij} and $d_{\mathrm{B}\,i}$ are distance metrics between entity pairs and an entity and the beam (B), respectively. The choice of k distinguishes between three different jet-finding algorithms: the k_{T} algorithm corresponds to $k = 1$ [162], the anti-k_{T} algorithm to $k = -1$ [163], and setting $k = 0$ defines the Cambridge–Aachen algorithm [164].

Irrespective of the choice of k, the algorithm proceeds by finding d_{ij} and $d_{\mathrm{B}\,i}$ for every pair of particles identified by the PF algorithm. If and only if $d_{ij} < d_{\mathrm{B}\,i}$, then i and j are combined and the process starts anew, otherwise i is classified as a "jet" and removed from the set of objects under consideration (a "jet" in this case could be a single, isolated particle). When i and j are combined, the resultant object has the following properties:

$$p_{\mathrm{T}} = p_{\mathrm{T}\,i} + p_{\mathrm{T}\,j} \qquad y = \frac{p_{\mathrm{T}\,i}\, y_i + p_{\mathrm{T}\,j}\, y_j}{p_{\mathrm{T}}} \qquad \varphi = \frac{p_{\mathrm{T}\,i}\, \varphi_i + p_{\mathrm{T}\,j}\, \varphi_j}{p_{\mathrm{T}}}. \tag{5.5}$$

Jets identified with the anti-k_{T} algorithm using PF objects are called *PF jets*. Making use of PF objects provides a marked improvement over relying exclusively on calorimeter energy clusters. Within a jet, the typical energy distribution is 65% charged particles : 25% photons : 10% neutral hadrons, and so 90% of the particles present will not be observed in the HCAL. This, combined with the poor resolution of the HCAL, leads to the improvement seen when using anti-k_{T}.

At CMS the anti-k_{T} algorithm is used with a jet size parameter of $R = 0.4$. Jets found using this configuration are referred to as *AK4 jets*.

Jet Energy Corrections

Once the jets have been identified, they are subject to various *jet energy corrections* (JECs) in order to account for nonuniformity in the detector response in η and p_{T}. These take the form of scale factors to a jet's four-momentum and are derived using simulated data samples (see Chapter 6). In order of application, the corrections are [165]:

L1 Pileup Correction Corrects for the presence of additional energy arising from pileup interactions and electronic noise.

L2 Relative Jet Correction An η-dependent scale factor aiming to make the detector response η-independent. Derived from dijet events with one jet in the barrel region, the p_T of the barrel jet is used to find the scale factor needed for the other jet.

L3 Absolute Jet Correction A p_T-dependent scale factor aiming to make the detector response p_T-independent. Derived using events containing a leptonic Z boson decay ($Z \rightarrow \ell^+\ell^-$) in association with one or more jets ($Z/\gamma^* +$ jets events), with the goal to set the scale factors such that a flat response in p_T is observed.

L2L3 Residual Final corrections, applied to data to reconcile any remaining discrepancies between simulated events and data.

The uncertainties associated with these corrections form a systematic uncertainty in the tZq analysis discussed further in § 8.2.1.

Identification of jets from b hadron decays

The structure of the CKM Matrix in Equation 2.32 means that the $b \rightarrow cW$ and $b \rightarrow uW$ decays are suppressed, resulting in a longer lifetime for hadrons containing bottom quarks. This can be exploited in order to specifically identify (*tag*) jets originating from bottom quarks (known as b jets). The longer lifetime of b hadrons often allows the identification of a *secondary vertex* a measurable distance from the primary vertex where the b hadron decayed. These secondary vertices are identified by the *Inclusive Vertex Finder* (IVF) algorithm [166, 167].

CMS has developed multiple b tagging algorithms to classify b jets. The flagship b taggging algorithm for data taken in 2016 and 2017 was *Combined Secondary Vertex Version 2* (CSVv2) [167]. CSVv2 makes use of ML techniques, leveraging an MLP and using observables associated with both the secondary and primary vertices as features. It is trained separately on three categories of jet: those for which a secondary vertex has been fully reconstructed; those for which a secondary vertex could not be reconstructed but have a pseudo-secondary vertex identified from tracks with a significant, positive and compatible impact parameter; and jets with neither a pseudo- nor fully reconstructed secondary vertex. Two discriminants are produced—one between b jets and c jets (jets originating from charm quarks in charm hadron decays) and the other between b jets and all other jets (i.e. *light jets*)—and combined to give an overall CSVv2 discriminant.

Table 5.1: The efficiency (i.e. the recall, see § 3.1.1), ε_{b}; type I error rate for c jets, α_{c}; and type I error rate for light jets, α_{udsg}, for $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ jets achieved by CSVv2 in simulated $\mathrm{t\bar{t}}$ events. Values taken from [167].

CSVv2 Working Point	ε_{b}	α_{c}	α_{udsg}
Loose	84%	39%	8.3%
Medium	66%	13%	0.8%
Tight	46%	2.6%	0.1%

In analyses the requirements placed on a b tagger's discriminant are set at one of three possible *working points* (WPs), each defined by the associated type I error rate: 10% at the *loose WP*, 1% at the *medium WP*, and 0.1% at the *tight WP* [167]. A summary of CSVv2's performance on simulated $\mathrm{t\bar{t}}$ samples is presented in Table 5.1.

5.3.2 Missing Transverse Energy

Weakly interacting particles like neutrinos, including potential but as-yet unobserved particles such as those that may compose dark matter, do not sufficiently interact with the CMS detector to be observable. Their presence can, however, be inferred through conservation of momentum: any significant difference in the total momentum in the final state when compared to the initial state can be attributed to undetected particles. A substantial fraction of a collision's debris will escape along the beamline, making reliable reconstruction of the missing p_z impossible. This is not the case for p_{T}, the total of which must be zero in the initial and, therefore, final state. The missing transverse energy, $\not{E}_{\mathrm{T}}$, is thus defined as the component transverse to the beam of the total momentum of all particles in the final state (while this technically constitutes the missing transverse momentum, this is considered to be $\approx \not{E}_{\mathrm{T}}$ in the ultrarelativistic limit).

PF objects can be used to determine the $\not{E}_{\mathrm{T}}$ on an event-by-event basis, in which case it is bestowed the moniker *PF* $\not{E}_{\mathrm{T}}$. Energy inside unclustered energy deposits—energy deposits in the calorimeters excluded from clusters due to low p_{T} or isolation—are also included in this calculation, counting against the $\not{E}_{\mathrm{T}}$ in an event. Studies performed on $\mathrm{Z}/\gamma^* +$ jets events, which should contain zero $\not{E}_{\mathrm{T}}$, have been used to demonstrate that this method is the most accurate available to the CMS experiment [168]. These studies have also identified the need for the *type I corrections* to the PF $\not{E}_{\mathrm{T}}$, where the JECs are propagated to the $\not{E}_{\mathrm{T}}$ calculation. The uncertainty associated with these and other corrections to the $\not{E}_{\mathrm{T}}$ in the tZq analysis are discussed in § 8.2.3. It is assumed henceforth that any mention of $\not{E}_{\mathrm{T}}$ refers to PF $\not{E}_{\mathrm{T}}$.

Chapter 6

Event Simulation

> *Essentially, all models are wrong,*
> *but some are useful.*
>
> George Box

Analyses at the CMS experiment routinely incorporate simulated data generated using *Monte Carlo* (MC) techniques [169]. This not only allows the optimisation of a search while blinded to the real data, but also allows access to the *MC truth*. Truth values describe events at a generator level (i.e. the level of fundamental particles) rather than at detector level (i.e. following CMS's reconstruction). This information is essential, for example, when evaluating the performance of a ML classifier. Simulated samples are often central to an analysis' methodology, so it is crucial they reflect reality as accurately and in as much detail as possible. This chapter discusses the techniques and software used to generate the simulated samples used at CMS.

6.1 Event Generation

The generation of a simulated event can be split, in general terms, into three stages: the initial hard proton–proton interaction, the subsequent parton shower and hadronisation, and the CMS detector system's response. Typically, different specialist software packages that interface with one another are used for each step, the final result being a simulated event with generator-level truth information and simulated response at the detector level.

Across all the simulated samples used in the tZq analysis (listed later in § 7.1), three different *event generators* are used to replicate the proton–proton interaction: MadGraph5 [170], *Automatic MC@NLO* (aMC@NLO) [170], and Powheg

Box V2 [171, 172, 173]. MadGraph5 and aMC@NLO refer to the same generator framework, known collectively as MadGraph5_aMC@NLO, but operating to the tree and one-loop levels, respectively. Both MadGraph5_aMC@NLO and Powheg Box V2 sample parton momenta from a proton PDF and use perturbative methods to determine the interactions that take place. This involves the calculation of the so-called *matrix elements* (MEs), which describe the likelihood of transitions to specific final states.

The ensuing parton showering and hadronisation is simulated by *shower Monte Carlo* (SMC) generators; Pythia8 [174] fills this role for all simulated samples used in the tZq analysis[1]. This includes the simulation of gluon initial-state radiation (ISR) at the chosen renormalisation scale, μ_{R}, and gluon *final-state radiation* (FSR) at the chosen factorisation scale, μ_{F}. PDFs are functions of these energy scales in practice but not in principle: the dependence exists only at finite levels of perturbation theory. As μ_{F} and μ_{R} are chosen, a systematic error is associated with this choice (see §§ 8.2.7 and 8.2.8).

The matching of the particles described in the perturbative QCD hard interaction simulation to those described in the nonperturbative QCD hadronisation simulation is also performed by Pythia8. Depending on the event generator, a different matching algorithm is used: MLM [175, 176] for MadGraph5, FxFx [177] for aMC@NLO, whilst Powheg Box V2 interfaces with SMC generators directly. This matching step introduces another systematic uncertainty, covered in § 8.2.9.

In a simulated sample, additional minimum-bias events (events that are not selected to contain certain physics objects but still exceed some minimum amount of detector activity) generated by Pythia8 are superimposed with events from the process of interest to imitate the effects of pileup. This does not exactly replicate the pileup seen in data, and so an additional reweighting detailed in § 6.2.1 is applied.

To view simulated events though the lens of the CMS detector, a simulation of the CMS detector itself was created with the *Geometry and Tracking 4* (Geant4) software package [178, 179, 180, 181]. The simulation attempts to replicate interactions of particles with the CMS detector's magnetic field and material. A simulated detector readout is created, used to reconstruct the generated events in the same way as real data, as described in Chapter 5.

Any number of events can be generated for a given process, and so a given simulated sample must be renormalised to reflect the number of events expected

---

[1] Pythia8 itself is a full event generator like MadGraph5_aMC@NLO and Powheg Box V2, but only its parton shower simulation capabilities are used in this case.

of that process in a data sample corresponding to a given integrated luminosity. This normalisation factor, w, is given by

$$w = \frac{\mathcal{L}\sigma}{N} \tag{6.1}$$

where $\mathcal{L}$ is the integrated luminosity, σ the cross section of the process, and N the number of simulated events.

6.2 Simulation Corrections

Corrections are applied to simulated samples to account for known problems and rectify observed discrepancies with data. These corrections may alter the properties of only a subset of observable quantities in an event, e.g. jet p_{T}, or may apply a corrective scale factor on a per-event basis. Most applied corrections introduce associated systematic uncertainties, covered in Chapter 8.

6.2.1 Pileup Modelling

The pileup observed in data is difficult to both predict and model, resulting in considerable disparity between the number of pileup interactions observed in data and simulated events. To correct for this, a scale factor is applied to each simulated event. This scale factor is a function of the number of primary vertices in an event, n_{PV}, and uses the n_{PV} distribution in minimum-bias data events as the baseline.

6.2.2 Jet Energy Smearing

In data, the observed jet energy resolution is ≈10% poorer than in simulated samples. To account for this a scale factor, c_{JER}, is applied to the four-momentum of jet objects in simulation in a process known as *jet smearing*.

If a matching generator-level jet can be found, the scale factor is determined by

$$c_{\mathrm{JER}} = 1 + \left(s_{\mathrm{JER}} - 1\right) \frac{p_{\mathrm{T}} - p_{\mathrm{T}}^{\mathrm{gen}}}{p_{\mathrm{T}}} \tag{6.2}$$

where $p_{\mathrm{T}}^{\mathrm{gen}}$ is the p_{T} of the matching generator-level jet and s_{JER} is a η-dependant scale factor provided by the CMS *jet energy resolution and corrections* (JERC) subgroup. A detector-level jet is considered to match with a generator-level jet if and only if

$$\Delta R < \frac{R}{2} \quad \wedge \quad \left|p_{\mathrm{T}} - p_{\mathrm{T}}^{\mathrm{gen}}\right| < 3\sigma_{p_{\mathrm{T}}} p_{\mathrm{T}} \tag{6.3}$$

where σ_{p_T} is the relative p_T uncertainty in simulation, $R = 0.4$ in AK4 jets (see § 5.3.1), and ΔR is defined in Equation 5.4. When a matching generator-level jet is not found, *stochastic smearing* is used. In this case,

$$c_\mathrm{JER} = 1 + \mathcal{N}\left(0, \sigma_{p_\mathrm{T}}\right) \sqrt{\max(s_\mathrm{JER}^2 - 1, 0)}. \tag{6.4}$$

The chosen value of $\mathcal{N}\left(0, \sigma_{p_\mathrm{T}}\right)$ is fixed for each jet i.e. it will remain the same when calculating systematic variations described in Chapter 8.

An error was found in the calculation of c_JER in the 2016 analysis where the uncertainty in s_JER was used in place of σ_{p_T} to determine c_JER. This has been fully remedied in the 2017 analysis but still affects the μμ channel of the 2016 analysis. An attempt was made to mitigate the impact of this error; its effectiveness is assessed in § 9.1.

6.2.3 b Tagging Efficiency

The CMS b *tag and vertexing* (BTV) group is responsible for measuring and comparing the efficiency and misidentification rates of b tagging algorithms in data and simulation. This is performed using multijet and $\mathrm{t\bar{t}}$ samples. Scale factors are applied on a per-event basis as functions of jet flavour, p_T, and pseudorapidity in order to mitigate any differences seen in b tagging efficiency between data and simulation.

Chapter 7

Event Selection

> *That's the cup of a carpenter.*
>
> INDIANA JONES

Event selection in the tZq analysis is motivated by the dileptonic final state of the tZq process. In this channel, we are restricted to those events where the Z boson decays into two opposite-charge same-flavour leptons, and the top quark can be assumed to decay as $\mathrm{t} \rightarrow \mathrm{Wb}$. This results in four jets in the final state: one from the recoil quark, two from the decay of the W (which must decay hadronically in the dileptonic channel), and a final b jet from the decay of the top quark. These attributes inform both the event selection criteria and the background processes that must be considered. As tZq is such a rare process, even processes with small cross sections may have contributions after the full event selection comparable to or exceeding that of tZq. A thorough consideration of potential background processes is therefore required.

7.1 Data and Simulation Samples

In 2016, $35.92\,\mathrm{fb}^{-1}$ of data was certified by the CMS collaboration for use in physics analyses. In 2017 this increased to $41.86\,\mathrm{fb}^{-1}$, giving a total of $77.78\,\mathrm{fb}^{-1}$ of data to be considered in this thesis [134].

The simulated samples used in the tZq analysis are listed in Tables 7.1 and 7.2, along with the respective cross sections used for normalisation and generators used to create the sample. The same samples are used in the *double-electron* (ee) and *double-muon* ($\mu\mu$) channels. Backgrounds considered in the tZq analysis can be arranged into six categories:

Table 7.1: MC samples used for each process in the 2016 analysis, with the associated cross section used for normalisation. All samples are generated at NLO unless specified otherwise. In this table, the shorthand allowing ‘t’ to refer either to the top or antitop quark is suspended.

Process	Events ($\times 10^6$)	Cross section (pb)	Generator
tZq†	14.5	0.0758	MadGraph5_aMC@NLO
tHq†‡	3.50	0.07462	MadGraph5_aMC@NLO
tWZ/tW$\ell\ell^{\dagger\ddagger}$	0.0500	0.01104	MadGraph5_aMC@NLO
tW	7.95	35.85	Powheg Box V1
$\bar{\text{t}}$W	6.93	35.85	Powheg Box V1
tq (s-channel)†	2.99	10.32	MadGraph5_aMC@NLO
tq (t-channel)	67.2	136.0	Powheg Box V2
$\bar{\text{t}}$q (t-channel)	38.8	80.95	Powheg Box V2
t$\bar{\text{t}}$	77.1	831.8	Powheg Box V2
t$\bar{\text{t}}$H → b$\bar{\text{b}}$	3.85	0.2942	Powheg Box V2
t$\bar{\text{t}}$H → non-b$\bar{\text{b}}$	3.98	0.2123	Powheg Box V2
t$\bar{\text{t}}$W → $\ell\nu$	5.28	0.2001	MadGraph5_aMC@NLO
t$\bar{\text{t}}$W → qq	0.833	0.405	MadGraph5_aMC@NLO
t$\bar{\text{t}}$Z → $\ell\ell/\nu\nu$	13.9	0.2529	MadGraph5_aMC@NLO
t$\bar{\text{t}}$Z → qq	0.749	0.5297	MadGraph5_aMC@NLO
W + jets	24.1	61520	MadGraph5_aMC@NLO
Z/γ^* + jets (Z p_T 0–50 GeV)	32.7	5352	MadGraph5_aMC@NLO
Z/γ^* + jets (Z p_T 50–100 GeV)	130	363.8	MadGraph5_aMC@NLO
Z/γ^* + jets (Z p_T 100–250 GeV)	83.7	84.01	MadGraph5_aMC@NLO
Z/γ^* + jets (Z p_T 250–400 GeV)	21.2	3.228	MadGraph5_aMC@NLO
Z/γ^* + jets (Z p_T 400–650 GeV)	1.63	0.4360	MadGraph5_aMC@NLO
Z/γ^* + jets (Z p_T $\geq$ 650 GeV)	1.63	0.04098	MadGraph5_aMC@NLO
WW → $\ell\nu$qq	9.00	50.00	Powheg Box V2
WW → $\ell\nu\ell\nu$	2.00	12.18	Powheg Box V2
WZ → $\ell\nu$qq	24.2	10.73	MadGraph5_aMC@NLO
WZ → qq$\ell\ell$	26.5	5.606	MadGraph5_aMC@NLO
WZ → $\ell\nu\ell\ell$	1.93	5.26	MadGraph5_aMC@NLO
ZZ → $\ell\ell\nu\nu$	8.84	0.5644	Powheg Box V2
ZZ → $\ell\ell$qq	15.3	3.222	MadGraph5_aMC@NLO
ZZ → $\ell\ell\ell\ell$	10.7	1.204	MadGraph5_aMC@NLO
WWW	0.240	0.2086	MadGraph5_aMC@NLO
WWZ	0.250	0.1651	MadGraph5_aMC@NLO
WZZ	0.247	0.5565	MadGraph5_aMC@NLO
ZZZ	0.249	0.01500	MadGraph5_aMC@NLO

† Including both t and $\bar{\text{t}}$ decays

‡ Generated at LO

Table 7.2: MC samples used for each process in the 2017 analysis, with the associated cross section used for normalisation. All samples are generated at NLO unless specified otherwise. In this table, the shorthand allowing ‘t’ to refer either to the top or antitop quark is suspended.

Process	Events ($\times 10^6$)	Cross section (pb)	Generator
tZq[†§]	13.3	0.07358	MadGraph5_aMC@NLO
tHq[†]	3.38	0.3184	MadGraph5_aMC@NLO
tWZ/tWℓℓ[†‡]	0.986	0.01103	MadGraph5_aMC@NLO
tW[§]	7.95	34.91	Powheg Box V2
$\bar{\mathrm{t}}$W[§]	7.75	34.97	Powheg Box V2
tq (s-channel)	9.88	3.74	MadGraph5_aMC@NLO
tq (t-channel)[§]	122	136.02	Powheg Box V2
$\bar{\mathrm{t}}$q (t-channel)[§]	38.8	80.95	Powheg Box V2
$\mathrm{t\bar{t}} \to \ell\nu\ell\nu$[§]	69.2	88.29	Powheg Box V2
$\mathrm{t\bar{t}} \to \ell\nu\mathrm{qq}$[§]	110	365.34	Powheg Box V2
$\mathrm{t\bar{t}} \to \mathrm{qqqq}$[§]	130	377.96	Powheg Box V2
$\mathrm{t\bar{t}Z} \to \mathrm{qq}$[§]	0.750	0.5104	MadGraph5_aMC@NLO
$\mathrm{t\bar{t}W} \to \ell\nu$[§]	4.92	0.2198	MadGraph5_aMC@NLO
$\mathrm{t\bar{t}W} \to \mathrm{qq}$	0.811	0.4316	MadGraph5_aMC@NLO
$\mathrm{t\bar{t}H} \to \mathrm{b\bar{b}}$	8.00	0.5269	Powheg Box V2
$\mathrm{t\bar{t}H} \to \text{non-}\mathrm{b\bar{b}}$	15.9	0.5638	Powheg Box V2
$\mathrm{t\bar{t}}\gamma$[‡§]	4.64	0.5804	MadGraph5_aMC@NLO
W + jets	77.8	52940.0	MadGraph5_aMC@NLO
Z/γ^* + jets (m_Z 10–50 GeV)	39.5	16000.0	MadGraph5_aMC@NLO
Z/γ^* + jets ($m_\mathrm{Z} \geq 50$ GeV)	210	6529.0	MadGraph5_aMC@NLO
$\mathrm{WW} \to \ell\nu\mathrm{qq}$	8.79	45.99	Powheg Box V2
$\mathrm{WW} \to \ell\nu\ell\nu$[§]	2.00	11.08	Powheg Box V2
$\mathrm{WZ} \to \ell\nu\mathrm{qq}$	19.1	11.66	MadGraph5_aMC@NLO
$\mathrm{WZ} \to \mathrm{qq}\ell\ell$	27.6	6.331	MadGraph5_aMC@NLO
$\mathrm{WZ} \to \ell\nu\ell\ell$	11.0	5.052	MadGraph5_aMC@NLO
$\mathrm{ZZ} \to \ell\ell\nu\nu$	8.74	0.5644	Powheg Box V2
$\mathrm{ZZ} \to \ell\ell\mathrm{qq}$	62.2	3.222	MadGraph5_aMC@NLO
$\mathrm{ZZ} \to \ell\ell\ell\ell$	6.96	1.325	Powheg Box V2
$\mathrm{W}\gamma \to \ell\nu\gamma$[‡]	6.28	405.27	MadGraph5_aMC@NLO
$\mathrm{Z}\gamma \to \ell\ell\gamma$[‡]	30.49	51.50	MadGraph5_aMC@NLO
WWW	0.232	0.2086	MadGraph5_aMC@NLO
WWZ	0.250	0.1651	MadGraph5_aMC@NLO
WZZ	0.250	0.5565	MadGraph5_aMC@NLO
ZZZ	0.250	0.01398	MadGraph5_aMC@NLO

[†] Including both t and $\bar{\mathrm{t}}$ decays

[‡] Generated at LO

[§] Contains parton shower weights (see § 8.2.8)

Single top This incorporates processes in which a single top quark is produced via a weak process e.g. single top quark production in the t-channel, single top quark production in the s-channel, and single top quark production in association with a Higgs boson (tHq). The largest contributor in this category is tW, which has a similar final state to tZq when both W bosons decay leptonically (the second W boson arising from the $\mathrm{t} \rightarrow \mathrm{Wb}$ decay).

$\mathrm{t\bar{t}}$ **production** The fully leptonic decay channel of the $\mathrm{t\bar{t}}$ process forms, by dint of its high cross section and similar final state, the second largest contributor to the background in the analysis. The semileptonic and fully hadronic decay channels are minor contributors as they lack the required number of leptons in the final state.

$\mathrm{t\bar{t}}$ **production in association with a boson (**$\mathrm{t\bar{t}V/t\bar{t}H}$**)** Production of $\mathrm{t\bar{t}}$ in association with a boson can have a similar final state as tZq, but has a significantly reduced cross section compared to $\mathrm{t\bar{t}}$ production. Consequently, $\mathrm{t\bar{t}V/t\bar{t}H}$ processes are a minor background. In this thesis, ‘V’ is used as a shorthand to represent either Z or W boson.

Vector boson + jets The W + jets and Z/γ^* + jets processes have the two largest cross sections of all considered backgrounds. The prompt leptons from the Z boson decay in the so-called *Drell–Yan* Z/γ^* + jets process makes it particularly difficult to distinguish from tZq, and thus forms the largest background contribution to the tZq analysis.

Diboson (VV**)** Includes the electroweak processes containing two vector bosons and no additional jets.

Triboson (VVV**)** Includes the electroweak WWW, WWZ, WZZ, and ZZZ processes. Due to their small cross sections, these form a minor contribution to the background.

The most notable difference in the simulation samples between 2016 and 2017 is the splitting of the $\mathrm{t\bar{t}}$ decay channels in 2017, where in 2016 an inclusive sample was used. This was necessary as no inclusive $\mathrm{t\bar{t}}$ simulated sample was available for use in 2017 analyses.

Top quark p_{T} reweigthing

It has been previously established [182] that the top p_{T} spectra in data does not match with what is predicted in simulation. Consequently, events in $\mathrm{t\bar{t}}$ simulation samples have an additional weight applied per-event to close the discrepancy.

7.2 Event Selection in the Signal Region

The event selection is performed in steps, each step examining different physics objects in the reconstructed event. In this section, each selection step is described in order of application.

7.2.1 Trigger Selection

The HLT paths passed by an event form the first step of the event selection in the tZq analysis. An electron pair or muon pair is expected in the final state, hence events that have passed HLT paths selecting for this are used. Furthermore, the sample size is increased by allowing events passing HLT paths requiring the presence of a single electron or muon, with the expectation that the missing lepton will, in some cases, be recovered following the full event reconstruction. A complete accounting of the HLT paths used is given in Table 7.3. To pass this selection step an event must pass at least one of the chosen HLT paths for the appropriate year and channel, and must not pass any path used exclusively in another channel in that year. For example, in the ee channel an event must pass a double- or single-electron HLT path and must not pass a single-, double-, or electron–muon HLT path. The HLT paths chosen were those with the lowest usable p_T thresholds for the leptons in order to further maximise the amount of tZq signal retained. For the double-lepton HLT paths these thresholds were 23 GeV (12 GeV) for the leading[1] (sub-leading) electron in the ee channel and 17 GeV (8 GeV) for the leading (sub-leading) muon in the μμ channel. HLT paths with lower p_T thresholds do exist, but are *prescaled*. Prescaled HLT paths are only activated for a fraction of eligible events in order to stay within timing constraints, and are not recommended for general use in analysis by the CMS collaboration. The total luminosity available for in each channel in each year was determined by summing the luminosities of the runs for which the chosen HLT paths were activated. This is 35.86 fb^{-1} in 2016 and 41.53 fb^{-1} in 2017.

Single- and double-lepton datasets are provided by CMS for use in physics analyses, with some overlap between the samples. The HLT path selection requirements eliminate double-counting issues between channels, but in order to avoid double-counting within each channel events in a double-lepton data sample that also feature a single lepton data sample were removed from consideration.

Trigger efficiency

The efficiency of the selected HLT paths was determined with the *cross-trigger method* used previously in the measurement of the $\mathrm{t\bar{t}}$ cross section in the eμ final

---

[1] greatest p_T

Table 7.3: The single- and double-lepton HLT paths required for selection in the ee channel, μμ channel, and $t\bar{t}$ CR (eμ). Single lepton paths are displayed in *`italics`*. The names of trigger paths describe the objects they select: `Ele` refers to electrons, `Mu` to global muons, and `TkMu` to tracker muons. The number immediately following one of these identifiers is a minimum p_{T} threshold in GeV. In some cases, additional criteria are in place such as particle identification (`Id`) or isolation (`Iso`) requirements using data from the calorimeters (`Calo`) or tracker (`Trk`). These can be loose (`L`), very loose, (`VL`) or very very loose (`VVL`), in descending order of stringency. Some HLT paths reject objects not originating from the IP and are marked `DZ`.

Channel	2016 HLT trigger paths	2017 HLT trigger paths
ee	*HLT_Ele32_eta2p1_WPTight_Gsf* HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	*HLT_Ele35_WPTight_Gsf* *HLT_Ele32_WPTight_Gsf_L1DoubleEG* HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
μμ	*HLT_IsoMu24* HLT_Mu17_TrkIsoVVL_Mu8 HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ	*HLT_IsoMu27* HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8 HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
eμ	*HLT_Ele32_eta2p1_WPTight_Gsf* *HLT_IsoMu24* HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	*HLT_Ele35_WPTight_Gsf* *HLT_Ele32_WPTight_Gsf_L1DoubleEG* *HLT_IsoMu27* HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ

Table 7.4: The lepton trigger scale factors determined by the cross-trigger method with their statistical uncertainty.

(a) Trigger scale factors in 2016

Channel	Run	Scale factor
ee	B–H	0.969 ± 0.001
μμ	B–F	0.977 ± 0.001
	G–H	0.989 ± 0.001
eμ	B–H	0.987 ± 0.001

(b) Trigger scale factors in 2017

Channel	Scale factor
ee	0.931 ± 0.001
μμ	0.967 ± 0.001
eμ	0.949 ± 0.001

state at CMS [183, 184]. In the cross-trigger method the efficiency of a given set of HLT paths is estimated by finding the ratio of the number of events that pass the *cross-triggers* and the HLT paths to the number of events that pass the cross-triggers only. The cross-triggers are a set of HLT paths weakly correlated to those used in the analysis, specifically $\not{E}_{\mathrm{T}}$ triggers. A scale factor to be applied to simulated samples to match the efficiencies observed in simulation and data is then computed as the ratio of efficiency in data to efficiency in simulation.

The inclusive $\mathrm{t\bar{t}}$ sample in 2016 and the $\mathrm{t\bar{t}} \to \ell\nu\ell\nu$ sample in 2017 were used to estimate the trigger efficiency. Table 7.4 gives the scale factors obtained in both years. Because of a tracking inefficiency during the first half of data-taking in 2016 (Runs B–F) [184, 185], the scale factors in the μμ channel were calculated separately for the first and second half (Runs G–H) of data taking in 2016. The implementation of the cross-trigger methodology was verified by reproducing the trigger efficiencies reported by the $\mathrm{t\bar{t}}$ analysis [184].

The systematic uncertainty associated with the lepton trigger efficiency is discussed in § 8.1.4.

7.2.2 Event Cleaning

After passing the trigger selection, various filters provided and recommended by the JetMET *Physics Object Group* (POG)[2] are applied to remove events affected by known anomalies:

Primary vertex filter Removes events in which the primary vertex is located at either $|z| > 24$ cm, $\varrho > 2$ cm, or was found using a fit with less than 5 degrees of freedom.

---

[2] POGs are groups within CMS that provide algorithms and recommendations for the reconstruction, identification, and measurement of physics objects.

Beam halo filter The LHC beams can interact with residual gas in the beam pipe or the edges of the beam pipe itself, potentially creating a *halo* of particles travelling with the beam. This filter removes events with an energy deposit signature compatible with what is produced by halo particles.

HBHE noise filter Removes events with anomalous noise in the HB and HE hybrid photodiodes and readout boxes.

HBHEIso noise filter Complements the HBHE noise filter by removing events with large, isolated energy deposits in the HCAL that are likely to be the result of noise.

ECAL Trigger Primitive filter To offset a deficiency in the ECAL data links, the L1 Trigger readout is used to estimate some of the energy deposited in the ECAL. The L1 Trigger readout is more easily saturated and may therefore underreport the deposited energy in some cases, artificially inflating $\not{E}_{\mathrm{T}}$. This filter removes these cases.

Bad PF Muon Filter Removes events containing muons of $p_{\mathrm{T}} > 100\,\mathrm{GeV}$ with track segments of poor compatibility.

Bad Charged Hadron Filter Removes events containing muon candidates rejected by PF but included as a charged hadron in $\not{E}_{\mathrm{T}}$ calculation.

Altogether, these affect $\lesssim$0.1% of events selected by the triggers.

7.2.3 Lepton Selection

The lepton selection requires four objects to be defined: loose electrons, tight electrons, loose muons, and tight muons. Tight and loose electrons must satisfy the following requirements:

- identified as an electron by PF
- identified as an electron by the GSF
- $p_{\mathrm{T}} > 15\,\mathrm{GeV}$
- pseudorapidity within the ECAL acceptance, that is, $|\eta| \leq 2.5$
- pseudorapidity outside the gap between the ECAL barrel and end cap, that is, $1.4442 \leq |\eta| \leq 1.566$ [142]
- $d_z < 0.1\,\mathrm{cm}$ in the ECAL barrel and $<0.2\,\mathrm{cm}$ in the ECAL end cap

- projection of d_0 transverse to the beam <0.05 cm in the ECAL barrel and <0.1 cm in the ECAL end cap.

Tight and loose muons, meanwhile, must be:

- identified as a muon by PF
- of $p_{\mathrm{T}} > 20\,\mathrm{GeV}$
- of $|\eta| \leq 2.5$ (i.e. within the muon system acceptance).

What distinguishes tight electrons from loose electrons and tight muons from loose muons are the identification and isolation requirements provided by the CMS Egamma POG and Muon POG, respectively [186, 187]. These requirements are designed to select leptons that originate from W and Z boson decays, and reject leptons originating from τ lepton decays, decays within jets, and incorrect reconstruction. Akin to b tagging algorithms, different *working points* (WPs) are defined. In the tZq analysis two WPs are used: tight and loose, where the tight WP exchanges lower efficiency for greater purity. Per-event scale factors associated with each WP are provided centrally by CMS to correct for observed discrepancies between simulated samples and data. These are subject to a systematic uncertainty described in § 8.1.4.

A tight electron must pass the electron identification requirements operating at the tight WP, a loose electron the identification requirements operating at the loose WP. For muons the same logic applies: tight muons must pass the muon identification requirements at the tight WP and loose muons must pass the loose WP. In both cases, the loose WPs are strictly less stringent requirements than the tight WP, ergo a tight electron (muon) is by definition also a loose electron (muon). An additional requirement exists for muons: the isolation requirement.

The relative isolation, I^{rel}, of a lepton is the total energy of all PF particles in a $\Delta R = 0.3$ cone (defined in Equation 5.4) surrounding the lepton divided by the lepton's p_{T}. The definition of this cone is shown in Figure 7.1. Requirements on the I^{rel} of electrons is included in the WPs described above, but for muons isolation and identification requirements are specified by the Muon POG separately. Hence, it is required that a tight muon passes both tight identification and isolation requirements and a loose muon passes both loose identification and isolation requirements.

Once tight and loose leptons are identified, events in the ee channel are required to contain exactly two tight electrons and no additional loose leptons. Likewise, in the μμ channel events must contain exactly two tight muons and no additional loose

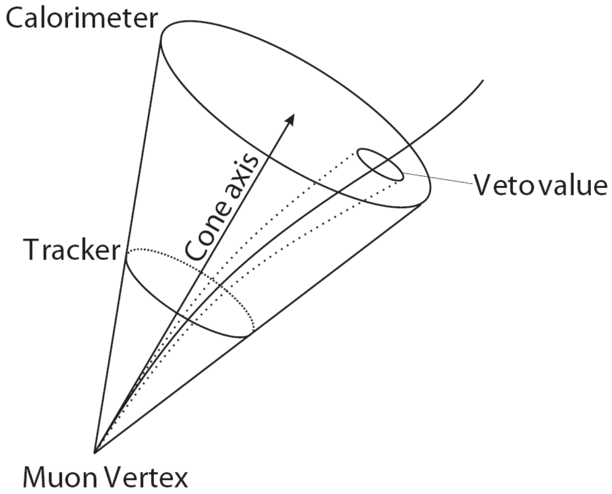


Figure 7.1: Definition of cone used to calculate I^{rel} for muons. The cone has a $\Delta R = 0.3$ radius and the axis of the cone is placed at a tangent to the initial muon momentum. Energy deposits in a small radius around the muon—the veto value—are excluded from the I^{rel} calculation to avoid the inclusion of p_{T} leaking from the muon. The same definition for the cone is used to calculate I^{rel} for electrons. Figure taken from [188].

leptons. Additionally, at least one of the selected electrons in the ee channel must have $p_{\mathrm{T}} > 35\,\mathrm{GeV}$ in 2016 and $p_{\mathrm{T}} > 38\,\mathrm{GeV}$ in 2017, and at least one of the selected muons in the μμ channel must have $p_{\mathrm{T}} > 26\,\mathrm{GeV}$ in 2016 and $p_{\mathrm{T}} > 29\,\mathrm{GeV}$ in 2017. The increased p_{T} requirements in 2017 arise from the increased p_{T} thresholds of the single lepton HLT paths used in 2017 (refer to Table 7.3).

The final element of the lepton selection is motivated by the Z boson. In a tZq event both of the leptons are expected to originate from the same Z boson parent; accordingly, the total invariant mass of the two selected leptons is required to be within 20 GeV of the nominal Z boson mass.

The lepton requirements are designed to reject events from processes with the incorrect number of prompt leptons in the final state, such as W + jets and single top production. The Z boson mass requirements further reduce backgrounds that do not contain a real Z boson, such as $\mathrm{t\bar{t}}$ production.

7.2.4 Jet Selection

In the tZq analysis a jet is considered to be any object identified by the PF algorithm as a jet of $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 4.7$, and $\Delta R > 0.4$ between itself and both selected leptons.

Further identification requirements follow recommendations from the JetMET POG [189]. These are designed to remove “jets” constructed from fake tracks or electronics noise by requiring jets to be composed of more than one particle and be linked to energy deposits in the ECAL and HCAL. Different WPs are provided for these requirements. In 2016 the loose WP is used, but in 2017 the loose WP was officially deprecated in favour of the tight WP. For an event to pass the selection requirements in the tZq analysis it must contain between four and six jets, inclusive. Although only four jets are expected in a LO tZq event, additional jets can result from jet splitting and higher-order processes.

The jet originating from the top quark decay is assumed to be a b jet as the branching ratio is ≈100%. The recoil quark will provide a second b jet, if it is a bottom quark (or a single b jet may split). As such, 1–2 of the jets are required to be b tagged using the CSVv2 algorithm at the medium WP. These jets must fall within the tracker acceptance of $|\eta| < 2.4$ in 2016 and $|\eta| < 2.5$ in 2017.

The final selection requirement is motivated by the hadronic decay of the W boson: each selected event must contain a pair of jets with an invariant mass within 20 GeV of the nominal W boson mass. The jet pair with invariant mass closest to the nominal W boson mass is assumed to be the decay products of the W boson and so is used to reconstruct all quantities associated with the W boson. When selecting this jet pair the leading b tagged jet is excluded from consideration as it is assumed that this b jet is from the top quark decay. The leading b jet and W boson candidate jet pair are, therefore, used to reconstruct all quantities associated with the top quark.

7.2.5 Experimental Blinding and the Side-Band Region

When constructing and optimising the analysis procedure, it is possible to make choices that unintentionally bias the analysis. This will typically take the form of artificially increasing the significance, or bringing the signal strength in line with what is expected (i.e. the SM prediction). To avoid this, access to the signal region data is restricted until the analysis strategy is finalized. This process is known as *blinding*.

To compare simulated samples and data in the signal region without considering the actual tZq signal the quantity

$$\chi^2 = \left(\frac{m_{\mathrm{W}}^{\mathrm{rec}} - m_{\mathrm{W}}}{\sigma_{\mathrm{W}}}\right)^2 + \left(\frac{m_{\mathrm{t}}^{\mathrm{rec}} - m_{\mathrm{t}}}{\sigma_{\mathrm{t}}}\right)^2 \tag{7.1}$$

is defined, where m_{W} ($m_{\mathrm{W}}^{\mathrm{rec}}$) is the nominal (reconstructed) W boson mass, m_{t} ($m_{\mathrm{t}}^{\mathrm{rec}}$) is the nominal (reconstructed) top quark mass, and σ_{W} (σ_{t}) is the resolution of the reconstructed W boson (top quark) mass. These resolutions were found by fitting a normal distribution to the respective mass peaks in simulation (shown in Figure 7.2). The χ^2 variable should be small for tZq events, so it is possible to define

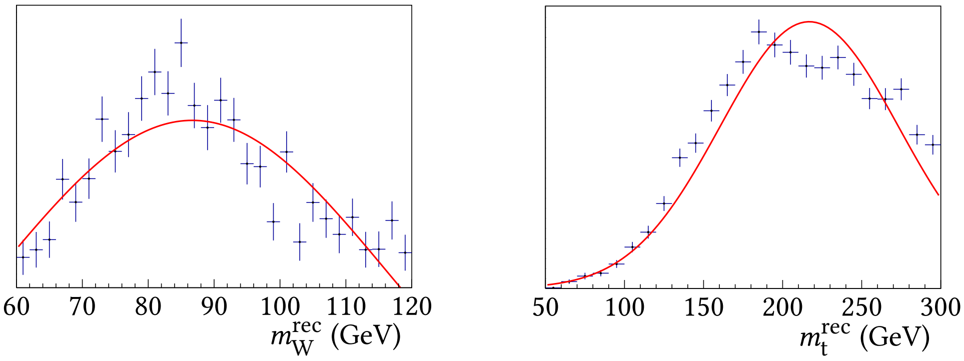


Figure 7.2: The result of fitting a normal distribution to the reconstructed W boson mass and top quark mass in 2016 data.

a range of χ^2 values that excludes the tZq signal but retains all major backgrounds. This is the *side-band* region. Comparisons between data and simulated samples can be made in this side-band region while the analysis is blinded. This technique is inspired by the ATLAS and CMS $\mathrm{H} \rightarrow \mathrm{b\bar{b}b\bar{b}}$ analyses [190, 191].

The χ^2 requirements for the side-band region were chosen in order to ensure that most of the signal was excluded, but enough events remained to make meaningful comparisons. It has been found previously [184] that a requirement of $5 < \chi^2 < 30$ defines a suitable region.

7.3 Data-Driven Nonprompt Lepton Estimation

Nonprompt leptons (NPLs) are leptons originating from outside the initial hard interaction in a proton–proton collision. This includes both real leptons from e.g. subsequent heavy quark decay and leptons incorrectly reconstructed from e.g. jets or photons. A significant source of events containing NPLs is QCD processes, which are difficult to model in simulation. In response to this, a data-driven methodology is used to estimate the NPL contribution to the tZq background.

The technique used is adapted from previous analyses concerning top quark pair production [192] and searches for particles predicted by the supersymmetric extension to the SM [193]. To use this technique a region dominated by NPL events is defined. By observing that few processes contain a final state with two leptons of the same flavour and charge, such a region can be created by simply inverting the opposite-sign lepton requirement imposed on the signal region (i.e. by requiring same-sign leptons). The resulting selection creates a region dominated by events with nonprompt or charge misidentified leptons, with a small contribution of prompt leptons from processes such as $\mathrm{t\bar{t}W}$. As a shorthand, events with two leptons of the same flavour and opposite charge will be referred to as *opposite-sign events*, and events with two leptons of the same flavour and charge will be referred to as *same-sign events*.

The number of opposite-sign nonprompt events in data, $X^{\mathrm{OS}}_{\mathrm{data}}$, can be estimated as

$$X^{\mathrm{OS}}_{\mathrm{data}} = \frac{X^{\mathrm{OS}}_{\mathrm{MC}}}{X^{\mathrm{SS}}_{\mathrm{MC}}} \left(N^{\mathrm{SS}}_{\mathrm{data}} - N^{\mathrm{SS}}_{\mathrm{real}} - N^{\mathrm{SS}}_{\text{mis-ID}} \right) \tag{7.2}$$

where $N^{\mathrm{SS}}_{\mathrm{data}}$ is the number of same-sign events observed in data, $N^{\mathrm{SS}}_{\mathrm{real}}$ is the number of same-sign events expected to contain two prompt leptons (from contributions such as $\mathrm{t\bar{t}W}$), and $N^{\mathrm{SS}}_{\text{mis-ID}}$ is the number of same-sign events expected due to the misidentification of one of the lepton's charges. The total $N^{\mathrm{SS}}_{\mathrm{real}} + N^{\mathrm{SS}}_{\text{mis-ID}}$ can be estimated with the yield from applying the same-sign event selection to simulated samples. In simulated samples, the ratio of opposite-sign and same-sign nonprompt events, $\frac{X^{\mathrm{OS}}_{\mathrm{MC}}}{X^{\mathrm{SS}}_{\mathrm{MC}}}$, is determined using generator-level information. Further details of how this NPL estimate was used in the tZq analysis can be found in [184].

7.4 Control Regions

In order to assess the description in simulation of the two largest background contributions to the tZq process—Z/γ^* + jets and $\mathrm{t\bar{t}}$ production—control regions enriched in these contributions were constructed. A comparison of data and simulated samples can then be performed in these CRs, as in the signal (or side-band) region. In addition to being enriched in the relevant background, the CRs are required to be topologically similar yet orthogonal to the signal region.

The Z/γ^* + jets CR is defined based on the event selection described in § 7.2, except the W boson mass requirement is inverted (i.e. no pair of jets may have an invariant mass within 20 GeV of the nominal W boson mass, excluding the leading b tagged jet) and the event must have less than 50 GeV of $E\!\!\!/_{\mathrm{T}}$. The inversion of the W boson mass requirement ensures orthogonality with the signal region while still retaining Z/γ^* + jets events, and the $E\!\!\!/_{\mathrm{T}}$ requirement reduces the $\mathrm{t\bar{t}}$ contribution, which contains $E\!\!\!/_{\mathrm{T}}$ when the W bosons decay leptonically.

For the $\mathrm{t\bar{t}}$ CR the lepton selection is altered, now requiring exactly one tight electron and exactly one tight muon with a minimum p_{T} of 25 GeV for both (and, as before, no additional loose leptons). As the lepton selection has changed, different HLT paths are used. These are listed in Table 7.3.

Table 7.5: The yield in the signal regions, side-band regions, Z/γ^* + jets CRs, and $t\bar{t}$ CR of each process considered in the 2016 analysis.

Process	Signal region		Side-band region		Z/γ^* + jets CR		$t\bar{t}$ CR
	ee	μμ	ee	μμ	ee	μμ	eμ
tZq	23.66 ± 0.23	57.25 ± 0.40	5.28 ± 0.11	12.22 ± 0.18	10.41 ± 0.16	25.29 ± 0.26	0.59 ± 0.04
W + jets	nil	nil	nil	nil	nil	nil	nil
Z/γ^* + jets	2668.08 ± 143.80	6934.56 ± 244.51	884.40 ± 79.96	2027.02 ± 131.39	2495.91 ± 126.48	5606.96 ± 217.97	1.41 ± 1.03
$t\bar{t}$	1113.53 ± 17.97	3079.68 ± 32.48	360.29 ± 10.21	978.74 ± 18.30	275.88 ± 9.05	779.85 ± 16.41	3450.59 ± 34.07
WW	1.03 ± 0.44	2.49 ± 0.75	0.43 ± 0.29	0.67 ± 0.39	0.01 ± 0.01	nil	2.75 ± 0.75
WZ	57.38 ± 1.15	124.59 ± 1.77	15.31 ± 0.60	34.39 ± 0.94	31.76 ± 0.85	70.22 ± 1.31	−0.04 ± 0.05
ZZ	44.04 ± 0.99	101.05 ± 1.60	12.64 ± 0.53	27.97 ± 0.84	28.01 ± 0.77	64.46 ± 1.25	0.12 ± 0.06
$t\bar{t}H$	3.41 ± 0.08	9.69 ± 0.15	0.97 ± 0.04	2.90 ± 0.08	0.39 ± 0.03	1.22 ± 0.05	10.14 ± 0.15
$t\bar{t}W$	5.09 ± 0.37	11.32 ± 0.65	1.73 ± 0.21	3.35 ± 0.35	0.69 ± 0.14	1.30 ± 0.22	14.68 ± 0.72
$t\bar{t}Z$	47.71 ± 0.61	111.96 ± 1.08	12.42 ± 0.37	27.83 ± 0.60	7.83 ± 0.23	18.20 ± 0.36	12.93 ± 0.96
tW	25.53 ± 2.06	59.10 ± 3.33	9.72 ± 1.23	20.04 ± 1.95	2.01 ± 0.53	11.41 ± 1.44	70.80 ± 3.61
$\bar{t}W$	23.70 ± 2.04	60.29 ± 3.35	6.94 ± 1.10	19.01 ± 1.88	6.64 ± 1.09	16.07 ± 1.74	67.09 ± 3.50
tq (s-channel)	nil	0.20 ± 0.20	nil	0.20 ± 0.20	nil	−0.06 ± 0.06	0.22 ± 0.22
tq (t-channel)	0.33 ± 0.15	0.70 ± 0.22	0.13 ± 0.09	0.27 ± 0.14	0.01 ± 0.01	0.42 ± 0.18	0.34 ± 0.15
$\bar{t}q$ (t-channel)	0.08 ± 0.08	0.92 ± 0.26	0.08 ± 0.08	0.38 ± 0.17	0.08 ± 0.08	0.32 ± 0.16	0.91 ± 0.26
tHq	0.12 ± 0.01	0.41 ± 0.02	0.04 ± 0.01	0.12 ± 0.01	0.03 ± 0.00	0.10 ± 0.01	0.41 ± 0.02
$tWZ/tW\ell\ell$	5.06 ± 0.19	11.12 ± 0.30	1.18 ± 0.09	2.75 ± 0.15	0.81 ± 0.08	2.03 ± 0.13	0.25 ± 0.04
WWW	0.07 ± 0.05	0.32 ± 0.11	0.01 ± 0.01	0.12 ± 0.07	nil	0.04 ± 0.04	0.21 ± 0.09
WWZ	1.10 ± 0.17	2.27 ± 0.27	0.27 ± 0.09	0.62 ± 0.14	0.21 ± 0.08	0.73 ± 0.15	0.13 ± 0.06
WZZ	1.22 ± 0.11	2.67 ± 0.17	0.31 ± 0.05	0.66 ± 0.08	0.40 ± 0.06	0.73 ± 0.09	0.02 ± 0.01
ZZZ	0.55 ± 0.04	1.15 ± 0.06	0.16 ± 0.02	0.31 ± 0.03	0.16 ± 0.02	0.26 ± 0.03	0.00 ± 0.00
NPL	13.36 ± 18.56	42.48 ± 12.34	7.94 ± 9.34	12.01 ± 6.25	—	—	—
Total	4035.04 ± 146.14	10 614.22 ± 247.02	1320.24 ± 81.18	3171.58 ± 132.84	2861.16 ± 126.82	6599.54 ± 218.61	3633.58 ± 34.49
Data	4375 ± 66	9753 ± 99	1363 ± 37	3003 ± 55	2661 ± 52	5785 ± 76	3407 ± 58

Table 7.6: The yield in the signal regions, side-band regions, Z/γ^* + jets CRs, and $t\bar{t}$ CR of each process considered in the 2017 analysis.

Process	Signal region		Side-band region		Z/γ^* + jets CR		$t\bar{t}$ CR
	ee	μμ	ee	μμ	ee	μμ	eμ
tZq	27.81 ± 0.29	66.87 ± 0.48	5.61 ± 0.13	14.10 ± 0.22	11.10 ± 0.18	26.66 ± 0.30	0.49 ± 0.04
W + jets	nil	nil	nil	nil	nil	nil	nil
Z/γ^* + jets	4026.30 ± 136.76	9424.47 ± 233.00	1113.28 ± 73.37	2794.43 ± 125.67	2945.15 ± 115.69	6623.10 ± 192.53	4.18 ± 2.99
$t\bar{t}$	1698.17 ± 9.19	4918.81 ± 81.00	551.29 ± 5.22	1561.04 ± 9.56	422.93 ± 4.61	1208.61 ± 8.48	5491.39 ± 17.32
WW	2.25 ± 0.70	2.67 ± 0.75	0.31 ± 0.22	1.36 ± 0.54	0.18 ± 0.16	0.68 ± 0.41	1.89 ± 0.66
WZ	189.18 ± 2.87	424.49 ± 4.56	54.45 ± 1.52	120.43 ± 2.39	99.80 ± 2.03	209.87 ± 3.17	1.30 ± 0.28
Wγ	nil	nil	nil	nil	nil	nil	nil
ZZ	5.26 ± 0.24	12.17 ± 0.40	1.38 ± 0.12	3.74 ± 0.22	0.99 ± 0.10	2.58 ± 0.17	0.02 ± 0.01
Zγ	24.68 ± 2.05	59.35 ± 3.36	7.63 ± 1.11	15.93 ± 1.72	18.74 ± 1.67	41.20 ± 2.75	0.31 ± 18.00
$t\bar{t}H$	9.49 ± 0.19	27.82 ± 0.35	2.68 ± 0.10	8.04 ± 0.18	1.07 ± 0.06	3.07 ± 0.11	29.43 ± 0.34
$t\bar{t}W$	6.61 ± 0.48	15.93 ± 0.75	1.96 ± 0.26	4.41 ± 0.42	0.48 ± 0.14	1.52 ± 0.25	18.36 ± 0.82
$t\bar{t}Z$	55.46 ± 0.46	127.77 ± 0.75	13.74 ± 0.23	31.66 ± 0.39	8.77 ± 0.78	19.05 ± 0.28	14.10 ± 0.34
$t\bar{t}\gamma$	11.50 ± 0.23	32.16 ± 0.42	3.55 ± 0.13	10.33 ± 0.23	2.69 ± 0.11	6.91 ± 0.19	38.21 ± 0.44
tW	26.67 ± 2.07	76.95 ± 3.80	7.75 ± 1.11	24.65 ± 2.17	4.83 ± 0.89	15.82 ± 1.78	73.05 ± 3.55
$\bar{t}W$	24.12 ± 2.01	70.66 ± 3.70	8.16 ± 1.17	23.65 ± 2.15	5.61 ± 0.96	14.23 ± 1.62	86.97 ± 3.97
tq (s-channel)	0.23 ± 0.09	0.20 ± 0.09	nil	0.06 ± 0.05	0.05 ± 0.04	nil	0.13 ± 0.06
tq (t-channel)	0.23 ± 0.09	2.60 ± 0.34	nil	0.88 ± 0.20	0.13 ± 0.06	0.76 ± 0.19	2.10 ± 0.29
$\bar{t}q$ (t-channel)	0.17 ± 0.08	1.35 ± 0.23	0.06 ± 0.04	0.18 ± 0.07	0.08 ± 0.06	0.49 ± 0.15	0.70 ± 0.16
tHq	0.44 ± 0.04	1.43 ± 0.08	0.13 ± 0.02	0.39 ± 0.04	0.16 ± 0.02	0.45 ± 0.04	1.48 ± 0.08
$tWZ/tW\ell\ell$	5.83 ± 0.05	12.98 ± 0.08	1.45 ± 0.02	3.17 ± 0.04	0.92 ± 0.02	2.07 ± 0.03	0.34 ± 0.01
WWW	0.00 ± 0.00	0.02 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	nil	0.05 ± 0.08	0.22 ± 0.11
WWZ	0.73 ± 0.16	3.21 ± 0.33	0.13 ± 0.07	0.85 ± 0.17	0.21 ± 0.09	0.40 ± 0.12	0.25 ± 0.10
WZZ	1.52 ± 0.13	2.90 ± 0.18	0.46 ± 0.07	0.91 ± 0.10	0.47 ± 0.07	0.89 ± 0.11	0.01 ± 0.02
ZZZ	0.53 ± 0.04	1.30 ± 0.07	0.13 ± 0.02	0.35 ± 0.03	0.12 ± 0.02	0.31 ± 0.03	0.00 ± 0.00
NPL	71.77 ± 23.98	−24.96 ± 18.63	9.03 ± 11.18	−13.91 ± 9.47	—	—	—
Total	6188.73 ± 138.24	15 261.16 ± 233.96	1783.18 ± 74.45	4606.66 ± 126.47	3524.46 ± 115.82	8178.72 ± 192.78	5764.91 ± 18.41
Data	6407 ± 80	16 406 ± 128	1931 ± 44	4931 ± 70	2974 ± 55	7726 ± 88	5155 ± 72

7.5 Event Yield

Tables 7.5 and 7.6 show the event yield after the full event selection for each process in the signal, side-band, and CRs as well as a comparison to the event yield in data. Corresponding cut-flow plots demonstrating the evolution of the yield through the event selection process are given in Figures 7.3 and 7.4.

In the signal regions, it can be seen that backgrounds containing NPLs, such as the data-driven NPL estimate and W + jets, are greatly reduced at when the jet requirements are applied. This indicates that the $\Delta R > 0.4$ between a jet and any lepton requirement is effective at removing NPLs. The Z/γ^* + jets background contribution is significantly diminished at the jet and b tag cut stages, but remains the greatest background in the signal region. The W boson mass requirement is less effective at reducing backgrounds not containing a real W boson than the Z mass requirement is at reducing backgrounds that do not contain a real Z boson. This is because when considering the W mass many pairs of jets are considered, making it more likely a jet pair will meet the requirement by chance.

The NPL contribution in the 2017 μμ channel is negative. The discrepancy between the 2016 and 2017 contributions seen here originates from the change in $t\bar{t}$ samples, with the $t\bar{t} \rightarrow \ell\nu qq$ sample containing more NPLs than the inclusive $t\bar{t}$ sample used in 2016. This requires further verification, but for the results in this thesis this is interpreted as the data-driven NPL estimate having no contribution in the 2017 μμ signal region.

Most of the tZq signal contribution is removed in the side-band regions and events from the most significant backgrounds are retained. Therefore, examining agreement between data and simulated samples in this region should be an effective proxy for the full signal region without potentially biasing the final result.

The Z/γ^* + jets CRs are dominated by the Z/γ^* + jets process, with a purity between 76.1% in the 2017 μμ channel and 87.2% in the 2016 ee channel. As in the signal region, $t\bar{t}$ remains the second largest contributor but has been reduced approximately fourfold by the inverted W mass and $\not{E}_T$ requirements.

The $t\bar{t}$ CR is purer than the Z/γ^* + jets CRs, consisting of ≈95% of the process of interest in both years. Requiring two leptons of different flavours eliminates the Z/γ^* + jets process almost entirely, leaving the tW and $\bar{t}$W processes as the next-largest contributors.

Across all signal and CRs, there is a residual disagreement in the final yield between data and simulated samples, as much as 11% in the μμ Z/γ^* + jets CR. While this exceeds any deviation expected by statistical uncertainty, Tables 8.3 and 8.4 shows that data and simulation agree within a 1σ envelope when systematic uncertainties are considered.

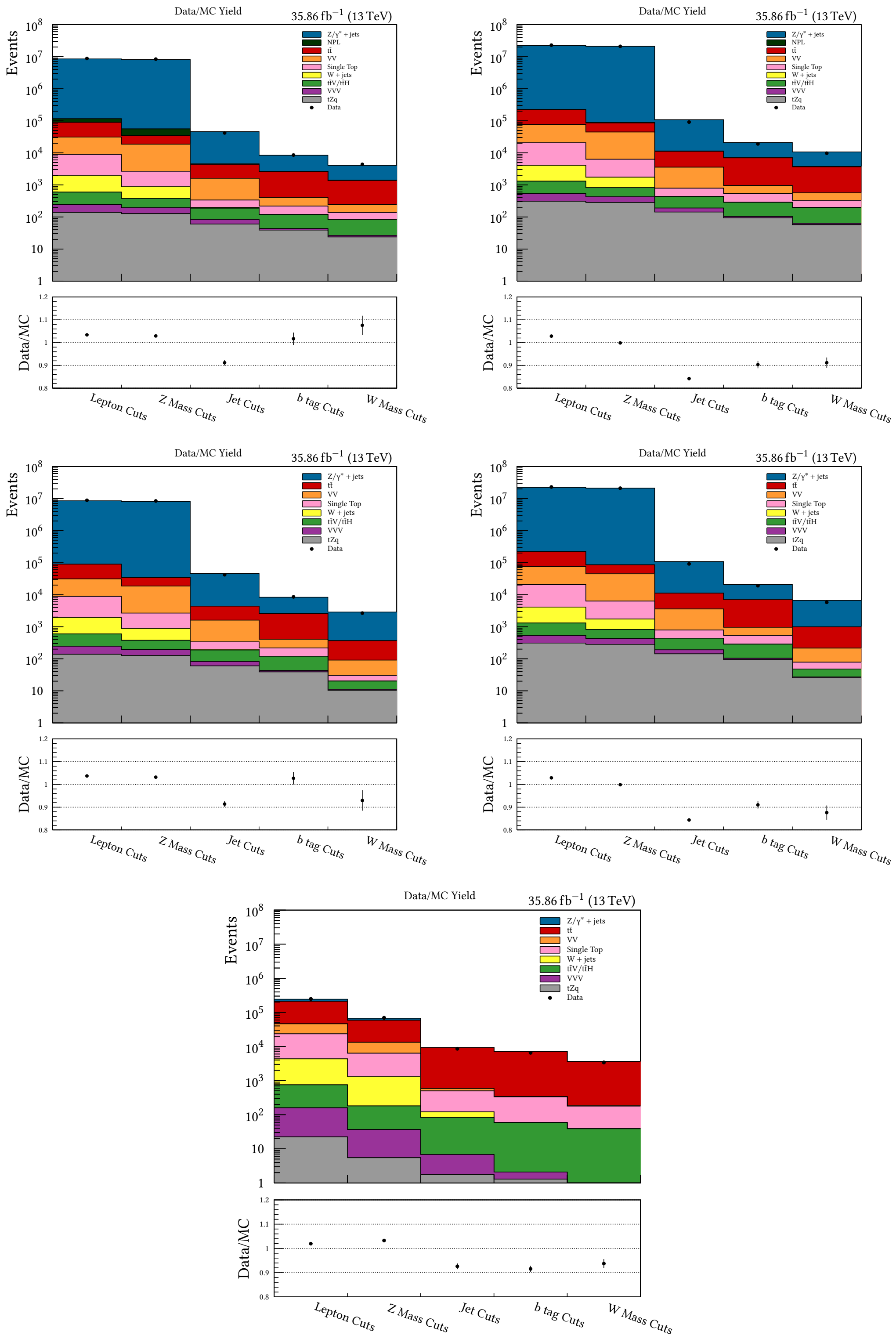


Figure 7.3: Cut-flow plots showing the yields in the 2016 analysis at each cut step in the ee signal region (top left), μμ signal region (top right), ee Z/γ^* + jets CR (middle left), μμ Z/γ^* + jets CR (middle right), and $t\bar{t}$ CR (bottom). The $\not{E}_T$ cut applied in the Z/γ^* + jets CR is included in the ‘W Mass Cuts’ step.

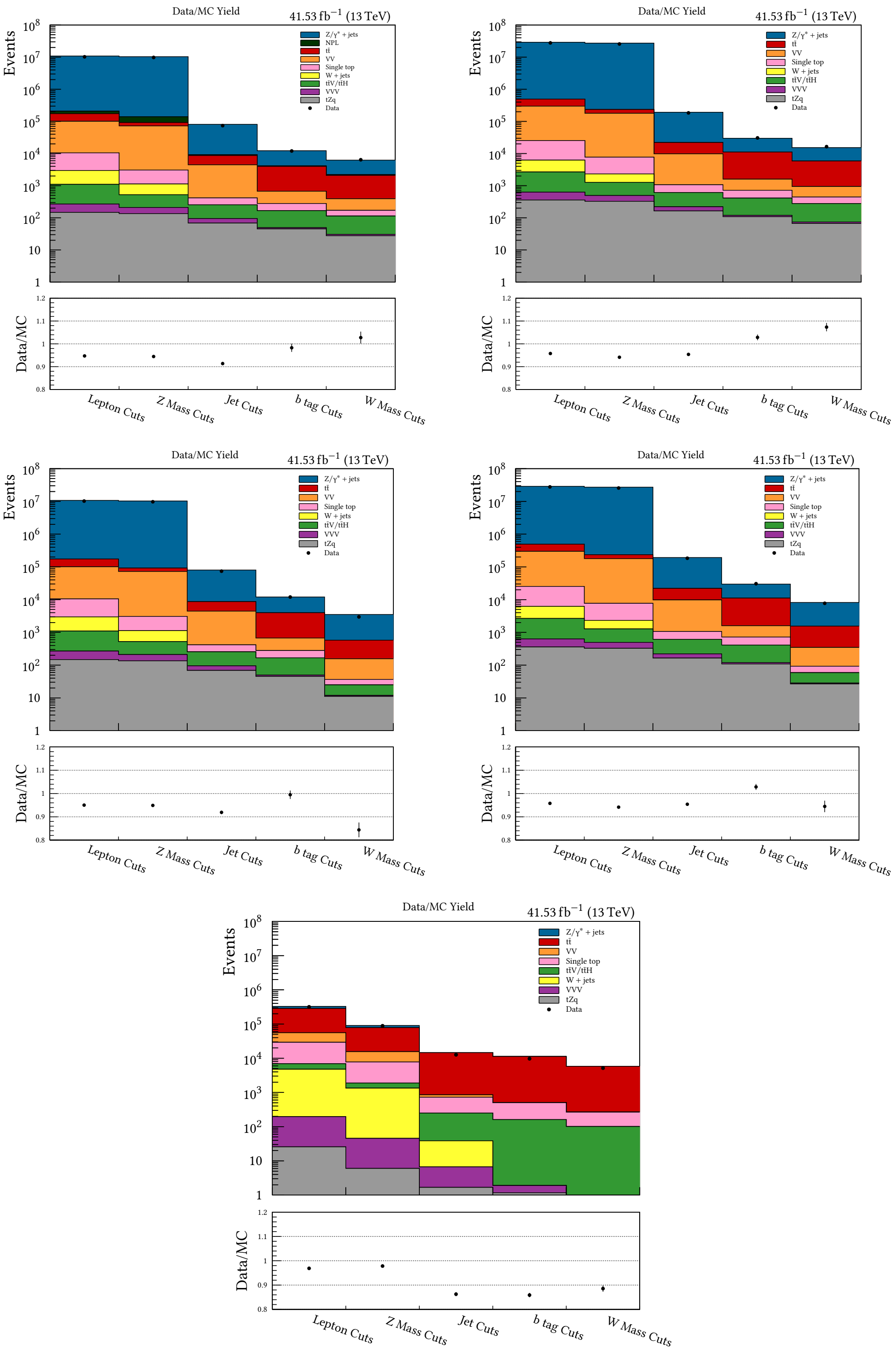


Figure 7.4: Cut-flow plots showing the yields in the 2017 analysis at each cut step in the ee signal region (top left), $\mu\mu$ signal region (top right), ee $Z/\gamma^* + $jets CR (middle left), $\mu\mu$ $Z/\gamma^* + $jets CR (middle right), and $t\bar{t}$ CR (bottom). The $\not{E}_T$ cut applied in the $Z/\gamma^* + $jets CR is included in the ‘W Mass Cuts’ step.

Chapter 8

Systematic Uncertainties

> *I beseech you, in the bowels of Christ, think it possible you may be mistaken.*
>
> Oliver Cromwell

When performing a physics analysis, attentive account must be kept of the sources of systematic uncertainty. These can originate from e.g. insufficient understanding of the underlying physics, imperfect modelling in simulated samples, or limitations of the CMS detector itself. Careful consideration of systematic uncertainties is of particular importance in searches for rare processes such as tZq, as even systematic uncertainties with a minor impact may be significant compared to the tZq cross section.

Systematic uncertainties can be split into two categories, depending on their effects on the underlying distributions of observables. *Rate uncertainties* affect the overall normalisation of distributions, but do not otherwise change distributions' shapes. *Shape uncertainties*, in contrast, change distributions' shapes and may, as a result, affect the overall normalisation.

The systematic uncertainties described in this chapter are included as nuisance parameters in the signal extraction fit, described in § 9.4.1. To do this, the impact of a $\pm 1\sigma$ change in each systematic uncertainty on the distributions of observables is defined. In many cases, this involves the use of *per-event weights*, which reweight events on an individual basis so the overall distributions of observables (shapes) reflect what is expected from a $\pm 1\sigma$ variation for a given source of systematic uncertainty.

8.1 Rate Uncertainties

8.1.1 Integrated Luminosity

The relative uncertainty in the integrated luminosity collected by CMS in 2016 and 2017 is ±2.5% and ±2.3%, respectively [194, 195]. Every simulated sample is scaled to this luminosity as per Equation 6.1, and so the effect of this uncertainty is an overall normalisation uncertainty applied equally to all simulated samples.

8.1.2 Cross Section Normalisation

To account for the uncertainty in the cross sections of simulated samples, a normalisation uncertainty is applied individually to each background process. Aping the previous tZq analysis at CMS in the trilepton channel, this uncertainty is set at ±30% for each process [196].

8.1.3 Data-Driven Nonprompt Lepton Estimate

Following the $\mathrm{t\bar{t}}$ analysis on which the NPL estimation strategy was based [193], a 30% uncertainty in the normalisation of the data-driven NPL contribution (see § 7.3) is applied. This is intended to cover all systematic uncertainties acting on the simulated samples used in the determination of data-driven NPL contribution. The only other systematic uncertainty the NPL estimate is subject to in the fit is the uncertainty in luminosity described above.

8.1.4 Lepton Efficiencies

Centrally-provided $\pm 1\sigma$ variations to the scale factors associated with lepton identification (see § 7.2.3) forms one component of the lepton efficiency systematic uncertainty. The other component incorporates the uncertainty in the trigger efficiencies (described in § 7.2.1), which are varied by ±1% in the ee and μμ channels, and ±2% in the eμ channel to form the $\pm 1\sigma$ variations. This alteration to the trigger efficiencies has been previously found to account for residual differences between data and simulation [184].

8.2 Shape Uncertainties

8.2.1 Jet Energy Corrections

As described in § 5.3.1, JECs are applied to simulated events so they more closely reflect data. The uncertainties associated with these corrections are provided by the CMS JEC group. In order to evaluate the impact of this systematic uncertainty, jet properties are recalculated after varying the JECs by ± 1 standard deviation.

8.2.2 Jet Smearing

A nuisance parameter associated with the jet smearing procedure described in § 6.2.2 is formed by altering s_{JER} by $\pm 1\sigma$ in the calculation of c_{JER} (Equations 6.2 and 6.4). This uncertainty in s_{JER} is centrally provided.

8.2.3 Missing Transverse Energy

As $\not{E}_{\mathrm{T}}$ is determined using the total p_{T} of all PF objects, the JECs and jet smearing have a knock-on effect on its value. This is accounted for by propagating the $\pm 1\sigma$ variations of JECs and jet smearing to the $\not{E}_{\mathrm{T}}$ calculation when those systematic uncertainties are evaluated (the so-called type I corrections).

An additional systematic uncertainty affecting the $\not{E}_{\mathrm{T}}$ alone arises from the contribution of unclustered energy deposits to the $\not{E}_{\mathrm{T}}$. This is estimated by varying the energy in each unclustered deposit by its resolution, and propagating the result to the $\not{E}_{\mathrm{T}}$ calculation.

8.2.4 Pileup Reweighting

A systematic uncertainty associated with the pileup reweighting applied to simulated data (see § 6.2.1) is created by altering the expected minimum-bias cross section by $\pm 4.6\%$. This alters the distribution of the number of primary vertices in simulated samples, allowing the effect of having more or fewer pileup interactions to be evaluated.

8.2.5 b Tagging Scale Factors

The scale factors introduced in § 6.2.3 are varied by 1σ uncertainty intervals provided by the CMS BTV group in order to estimate the impact of b tagging uncertainty on the analysis.

8.2.6 Parton Density Functions

Uncertainties on the PDFs are propagated through the tZq analysis by creating a nuisance parameter in the signal extraction step representing a $\pm 1\sigma$ change to the PDFs. This is in accordance with PDF4LHC recommendations[1] [199].

8.2.7 Perturbative Factorisation and Renormalisation Scales

A systematic uncertainty was introduced to account for the choice of the factorisation energy scale, μ_{F}, and renormalisation energy scale, μ_{R}, when performing the ME calculations under perturbative QCD (see § 6.1). The $+1\sigma$ variation is created by simultaneously halving the choices of μ_{F} and μ_{R} from the nominal value, and the -1σ variation from doubling the choices of μ_{F} and μ_{R} from the nominal value.

8.2.8 Non-Perturbative Factorisation and Renormalisation Scales

The effect of the choice of μ_{F} and μ_{R} must also be considered in the nonperturbative QCD regime used when simulating parton showers. In the 2016 analysis this was achieved through dedicated systematic samples listed in Table 8.1. The 'ISR up' and 'FSR down' samples consider the effects of changing μ_{F} and μ_{R} in the ISR and FSR separately, while for the 'scale up' and 'scale down' samples the ISR and FSR are considered together. In the 2017 analysis dedicated samples are no longer used, instead some nominal samples contain per-event weights used to create the shapes required evaluate this systematic uncertainty. Samples containing these weights are indicated in Table 7.2 and these weights were used when available. The $+1\sigma$ (-1σ) variation was created by doubling (halving) the choices of μ_{F} and μ_{R}.

8.2.9 Matching Threshold Energy

Dedicated simulated samples were used to estimate the impact of the choice of *matching threshold energy*. This is the energy scale at which the matching of particles described by perturbative and nonperturbative QCD is performed (see § 6.1). These samples exist only for certain Powheg Box V2 samples: $\mathrm{t\bar{t}}$ and single top t-channel in 2016 and the three $\mathrm{t\bar{t}}$ decay modes in 2017. They are listed as 'matching up' and 'matching down' in Tables 8.1 and 8.2.

---

[1] In most cases, per-event weights provided with the sample were used to create the required shapes. However, for the tW and $\mathrm{\bar{t}W}$ samples in 2016, which were generated with Powheg Box V1, these weights were not included and were instead generated by the *Les Houches Accord Parton Distribution Function* (LHAPDF) library [197] using the NNPDF3.0 PDF set [184, 198].

Table 8.1: Dedicated systematic samples used to estimate uncertainties in the 2016 tZq analysis. In this table, the shorthand allowing ‘t’ to refer either to the top or antitop quark is suspended.

Process	Systematic	Events ($\times 10^6$)	Generator
tZq[*]	scale up	6.89	MadGraph5_aMC@NLO
tZq[*]	scale down	6.98	MadGraph5_aMC@NLO
tW	scale up	0.998	Powheg Box V2
tW	scale down	0.994	Powheg Box V2
$\bar{\text{t}}$W	scale up	1.00	Powheg Box V2
$\bar{\text{t}}$W	scale down	0.999	Powheg Box V2
tq (t-channel)	matching up	6.00	Powheg Box V2
tq (t-channel)	matching down	6.00	Powheg Box V2
tq (t-channel)	scale up	5.71	Powheg Box V2
tq (t-channel)	scale down	5.95	Powheg Box V2
$\bar{\text{t}}$q (t-channel)	matching up	4.00	Powheg Box V2
$\bar{\text{t}}$q (t-channel)	matching down	4.00	Powheg Box V2
$\bar{\text{t}}$q (t-channel)	scale up	3.97	Powheg Box V2
$\bar{\text{t}}$q (t-channel)	scale down	3.89	Powheg Box V2
t$\bar{\text{t}}$	ISR up	156	Powheg Box V2
t$\bar{\text{t}}$	ISR down	150	Powheg Box V2
t$\bar{\text{t}}$	FSR up	153	Powheg Box V2
t$\bar{\text{t}}$	FSR down	156	Powheg Box V2
t$\bar{\text{t}}$	matching up	58.9	Powheg Box V2
t$\bar{\text{t}}$	matching down	58.2	Powheg Box V2

[*] Including $\bar{\text{t}}$ decays

Table 8.2: Dedicated systematic samples used to estimate uncertainties in the 2017 tZq analysis.

Process	Systematic	Events ($\times 10^6$)	Generator
t$\bar{\text{t}} \to \ell\nu\ell\nu$	matching up	3.29	Powheg Box V2
t$\bar{\text{t}} \to \ell\nu\ell\nu$	matching down	5.48	Powheg Box V2
t$\bar{\text{t}} \to \ell\nu$qq	matching up	24.0	Powheg Box V2
t$\bar{\text{t}} \to \ell\nu$qq	matching down	26.8	Powheg Box V2
t$\bar{\text{t}} \to$ qqqq	matching up	27.3	Powheg Box V2
t$\bar{\text{t}} \to$ qqqq	matching down	27.1	Powheg Box V2

The matching energy threshold in Powheg Box V2 is defined as

$$E_{\text{matching}} = \frac{h_{\text{damp}}^2}{h_{\text{damp}}^2 + p_{\text{T}}^2} \tag{8.1}$$

where p_{T} is the transverse momentum of the object being matched and nominally $h_{\text{damp}} = 1.58 m_{\text{t}}$. The 'matching down' and 'matching up' samples are created by reducing and increasing h_{damp} by one standard deviation, respectively [200].

8.3 Pre-Fit Impact of Systematic Uncertainties

The influence on the overall yield of simulated samples in the signal region for each systematic uncertainty featured in this chapter is shown in Table 8.3 for 2016 and Table 8.4 for 2017. In the case of shape uncertainties it should be noted that this measure may belie the magnitude of their influence on the underlying shapes. For example, the effect on the overall normalisation of the systematic uncertainty in $\not{E}_{\text{T}}$ is listed as zero in Tables 8.3 and 8.4 (as no requirements are placed on $\not{E}_{\text{T}}$ in the signal region). That is not to say that this uncertainty has no effect on the $\not{E}_{\text{T}}$ distribution, only that the overall normalisation remains unchanged.

8.4 Correlation of systematic uncertainties

In the tZq analysis the sources of systematic uncertainty are assumed to be 100% uncorrelated with each other. This will be used to build the likelihood model used in signal extraction, as described in § 9.4.

Across years, the same source of systematic uncertainty is either treated as 100% correlated or 100% uncorrelated. Those treated as 100% correlated are:

- the uncertainty in cross-section normalisation for each background process
- the uncertainty associated with the pileup reweighting applied to simulated data
- all uncertainties stemming from the choice of μ_{F} and μ_{R}
- the uncertainty from the choice of matching threshold energy (h_{damp}).

Treating these uncertainties as correlated between data-taking eras is the follows the latest recommendations of the CMS collaboration.

Table 8.3: Impact of systematic uncertainties on the normalisation of simulated samples in the signal region in 2016.

Uncertainty (ee / μμ)	Total (%)	tZq (%)	Z/γ* + jets (%)	$t\bar{t}$ (%)	$t\bar{t}V/t\bar{t}H$ (%)	Single top (%)	VV (%)	VVV (%)
Luminosity	±2.5 / ±2.5	±2.5 / ±2.5	±2.5 / ±2.5	±2.5 / ±2.5	±2.5 / ±2.5	±2.5 / ±2.5	±2.5 / ±2.5	±2.5 / ±2.5
Normalisation	±7.16 / ±7.14	—	±10.00 / ±10.00	±10.00 / ±10.00	±8.56 / ±8.49	±6.42 / ±6.42	±7.10 / ±7.06	±5.90 / ±5.78
Lepton efficiency	${}^{+5.14}_{-4.94} / {}^{+5.26}_{+0.89}$	${}^{+5.33}_{-5.11} / {}^{+5.19}_{+0.98}$	${}^{+5.24}_{-5.03} / {}^{+5.28}_{+0.90}$	${}^{+4.93}_{-4.74} / {}^{+5.30}_{+0.88}$	${}^{+5.51}_{-5.28} / {}^{+5.17}_{+0.99}$	${}^{+5.20}_{-5.00} / {}^{+5.32}_{+0.86}$	${}^{+5.46}_{-5.22} / {}^{+5.18}_{+0.98}$	${}^{+5.44}_{-5.44} / {}^{+5.15}_{-1.09}$
Jet smearing	${}^{+16.27}_{-12.86} / {}^{+17.56}_{-12.83}$	${}^{+3.72}_{-3.93} / {}^{+4.05}_{-4.19}$	${}^{+20.76}_{-15.64} / {}^{+21.84}_{-15.39}$	${}^{+7.87}_{-7.84} / {}^{+9.90}_{-8.40}$	${}^{-1.42}_{+1.12} / {}^{+0.01}_{-0.38}$	${}^{+10.80}_{-8.01} / {}^{+14.19}_{-11.01}$	${}^{+9.11}_{-9.58} / {}^{+10.76}_{-8.14}$	${}^{+1.02}_{-1.02} / {}^{+0.16}_{-2.96}$
JECs	${}^{+9.72}_{-4.96} / {}^{+8.72}_{-4.72}$	${}^{+0.30}_{-0.59} / {}^{+0.66}_{-0.75}$	${}^{+12.90}_{-5.73} / {}^{+11.20}_{-5.79}$	${}^{+3.93}_{-3.83} / {}^{+4.34}_{-2.92}$	${}^{-1.65}_{+1.25} / {}^{-0.93}_{+0.61}$	${}^{+3.99}_{-3.70} / {}^{+6.50}_{-4.08}$	${}^{+3.26}_{-3.18} / {}^{+3.60}_{-2.09}$	${}^{+1.36}_{-0.68} / {}^{+0.31}_{-2.18}$
Pileup reweighting	${}^{+1.59}_{-1.31} / {}^{+2.51}_{-2.47}$	${}^{+0.63}_{-0.59} / {}^{+0.24}_{-0.26}$	${}^{+2.26}_{-1.83} / {}^{+3.45}_{-3.42}$	${}^{+0.25}_{-0.26} / {}^{+0.74}_{-0.66}$	${}^{+0.50}_{-0.53} / {}^{-0.29}_{+0.29}$	${}^{+1.26}_{-1.26} / {}^{+1.11}_{-1.18}$	${}^{+1.02}_{-0.97} / {}^{+1.60}_{-1.43}$	${}^{+0.00}_{-0.34} / {}^{-0.47}_{+0.00}$
b tagging	${}^{+7.59}_{-7.59} / {}^{+4.81}_{-4.81}$	${}^{+12.43}_{-12.43} / {}^{+3.60}_{-3.60}$	${}^{+6.41}_{-6.41} / {}^{+5.34}_{-5.34}$	${}^{+10.10}_{-10.10} / {}^{+3.69}_{-3.69}$	${}^{+15.26}_{-15.26} / {}^{+3.84}_{-3.84}$	${}^{+6.24}_{-6.24} / {}^{+3.42}_{-3.42}$	${}^{+7.33}_{-7.33} / {}^{+6.28}_{-6.28}$	${}^{+11.22}_{-11.22} / {}^{+0.00}_{-0.00}$
PDF	${}^{+7.98}_{-7.96} / {}^{+7.39}_{-7.38}$	${}^{+15.93}_{-15.93} / {}^{+16.14}_{-16.14}$	${}^{+10.30}_{-10.30} / {}^{+9.52}_{-9.52}$	${}^{+2.38}_{-2.38} / {}^{+2.43}_{-2.43}$	${}^{+6.90}_{-6.90} / {}^{+6.85}_{-6.85}$	${}^{+8.87}_{-8.04} / {}^{+9.04}_{-8.07}$	${}^{+7.72}_{-7.72} / {}^{+8.27}_{-8.27}$	${}^{+4.42}_{-4.42} / {}^{+0.00}_{-0.00}$
μ_F and μ_R scale (ME)	${}^{+10.66}_{-9.48} / {}^{+9.95}_{-8.99}$	${}^{+7.31}_{-7.40} / {}^{+6.76}_{-7.06}$	${}^{+8.86}_{-8.05} / {}^{+7.91}_{-7.32}$	${}^{+15.73}_{-13.56} / {}^{+15.11}_{-13.27}$	${}^{+11.37}_{-11.85} / {}^{+10.62}_{-11.50}$	${}^{+1.02}_{-0.97} / {}^{+1.00}_{-0.95}$	${}^{+9.79}_{-7.76} / {}^{+9.90}_{-7.75}$	${}^{+8.84}_{-6.80} / {}^{+0.00}_{-0.00}$
μ_F and μ_R scale (ISR)	${}^{+4.19}_{+1.20} / {}^{-1.88}_{-4.04}$	—	—	${}^{+15.22}_{+4.34} / {}^{-6.47}_{-13.95}$	—	—	—	—
μ_F and μ_R scale (FSR)	${}^{-0.90}_{+4.03} / {}^{-5.66}_{-1.19}$	—	—	${}^{-3.28}_{+14.63} / {}^{-19.54}_{-4.11}$	—	—	—	—
μ_F and μ_R scale (ISR and FSR)	${}^{-0.08}_{-0.29} / {}^{+0.04}_{-0.13}$	${}^{+19.19}_{-0.80} / {}^{+25.87}_{-13.38}$	—	—	—	${}^{-14.12}_{-20.83} / {}^{-8.04}_{-16.39}$	—	—
ME–PS matching scale (h_{damp})	${}^{+2.03}_{+0.23} / {}^{-1.89}_{-4.40}$	—	—	${}^{+7.41}_{+0.74} / {}^{-6.54}_{-15.16}$	—	${}^{-0.75}_{+1.62} / {}^{+0.66}_{-0.25}$	—	—
$\not{E}_T$	${}^{+0.00}_{-0.00} / {}^{+0.00}_{+0.00}$	${}^{+0.00}_{-0.00} / {}^{+0.00}_{+0.00}$	${}^{+0.00}_{-0.00} / {}^{+0.00}_{+0.00}$	${}^{+0.00}_{-0.00} / {}^{+0.00}_{+0.00}$	${}^{+0.00}_{-0.00} / {}^{+0.00}_{+0.00}$	${}^{+0.00}_{-0.00} / {}^{+0.00}_{+0.00}$	${}^{+0.00}_{-0.00} / {}^{+0.00}_{+0.00}$	${}^{+0.00}_{-0.00} / {}^{+0.00}_{-0.00}$
NPL estimate	${}^{+0.11}_{-0.11} / {}^{+0.13}_{-0.13}$	—	—	—	—	—	—	—
Total systematic uncertainty	${}^{+26.51}_{-22.41} / {}^{+26.30}_{-21.13}$	${}^{+29.65}_{-22.63} / {}^{+32.22}_{-22.97}$	${}^{+31.02}_{-24.99} / {}^{+30.50}_{-23.71}$	${}^{+29.33}_{-27.02} / {}^{+31.08}_{-28.70}$	${}^{+22.90}_{-23.05} / {}^{+16.75}_{-16.55}$	${}^{+22.96}_{-26.33} / {}^{+21.91}_{-23.11}$	${}^{+19.76}_{-18.98} / {}^{+20.50}_{-17.25}$	${}^{+17.24}_{-16.25} / {}^{+13.36}_{-12.10}$

Table 8.4: Impact of systematic uncertainties on the normalisation of simulated samples in the signal region 2017.

Uncertainty (ee / μμ)	Total (%)	tZq (%)	Z/γ* + jets (%)	$t\bar{t}$ (%)	$t\bar{t}V/t\bar{t}H$ (%)	Single top (%)	VV (%)	VVV (%)
Luminosity	±2.3 / ±2.3	±2.3 / ±2.3	±2.3 / ±2.3	±2.3 / ±2.3	±2.3 / ±2.3	±2.3 / ±2.3	±2.3 / ±2.3	±2.3 / ±2.3
Normalisation	±7.07 / ±7.04	—	±10.00 / ±10.00	±10.00 / ±10.00	±7.20 / ±7.17	±6.34 / ±6.34	±8.62 / ±8.59	±6.36 / ±6.28
Lepton efficiency	${}^{+4.45}_{-4.29}$ / ${}^{+1.68}_{-1.67}$	${}^{+4.75}_{-4.57}$ / ${}^{+1.70}_{-1.67}$	${}^{+4.58}_{-4.42}$ / ${}^{+1.68}_{-1.67}$	${}^{+4.20}_{-4.06}$ / ${}^{+1.69}_{-1.67}$	${}^{+4.90}_{-4.71}$ / ${}^{+1.66}_{-1.47}$	${}^{+4.40}_{-4.26}$ / ${}^{+1.67}_{-1.66}$	${}^{+5.05}_{-4.86}$ / ${}^{+1.69}_{-1.67}$	${}^{+5.40}_{-5.04}$ / ${}^{+1.62}_{-1.62}$
Jet smearing	${}^{+20.39}_{-16.28}$ / ${}^{+21.69}_{-17.91}$	${}^{+4.85}_{-4.60}$ / ${}^{+4.59}_{-5.34}$	${}^{+25.31}_{-19.25}$ / ${}^{+27.82}_{-22.06}$	${}^{+11.69}_{-11.66}$ / ${}^{+12.03}_{-11.31}$	${}^{+1.17}_{-1.60}$ / ${}^{+2.68}_{-3.36}$	${}^{+14.84}_{-13.19}$ / ${}^{+18.34}_{-18.98}$	${}^{+15.07}_{-10.90}$ / ${}^{+12.59}_{-11.98}$	${}^{+6.47}_{-9.35}$ / ${}^{+4.71}_{-6.19}$
JECs	${}^{+10.86}_{-13.06}$ / ${}^{+11.46}_{-14.06}$	${}^{+0.25}_{-1.29}$ / ${}^{+0.81}_{-1.75}$	${}^{+14.05}_{-16.31}$ / ${}^{+15.22}_{-18.37}$	${}^{+5.11}_{-7.59}$ / ${}^{+5.44}_{-7.21}$	${}^{+0.05}_{-0.17}$ / ${}^{+1.63}_{-1.22}$	${}^{+8.53}_{-11.43}$ / ${}^{+9.06}_{-13.31}$	${}^{+6.59}_{-6.88}$ / ${}^{+6.24}_{-7.41}$	${}^{+3.24}_{-5.40}$ / ${}^{+0.81}_{-2.56}$
Pileup reweighting	${}^{+0.12}_{-0.12}$ / ${}^{+0.20}_{-0.20}$	${}^{-0.04}_{+0.04}$ / ${}^{+0.00}_{-0.00}$	${}^{+0.17}_{-0.17}$ / ${}^{+0.27}_{-0.27}$	${}^{+0.02}_{-0.02}$ / ${}^{+0.08}_{-0.08}$	${}^{-0.01}_{+0.01}$ / ${}^{+0.03}_{-0.03}$	${}^{+0.00}_{-0.00}$ / ${}^{+0.16}_{-0.16}$	${}^{+0.10}_{-0.10}$ / ${}^{+0.22}_{-0.22}$	${}^{+0.00}_{-0.00}$ / ${}^{+0.00}_{-0.00}$
b tagging	${}^{+5.03}_{-5.03}$ / ${}^{+5.03}_{-5.03}$	${}^{+3.70}_{-3.70}$ / ${}^{+3.71}_{-3.71}$	${}^{+5.48}_{-5.48}$ / ${}^{+5.51}_{-5.51}$	${}^{+3.86}_{-3.86}$ / ${}^{+3.86}_{-3.86}$	${}^{+4.35}_{-4.35}$ / ${}^{+4.25}_{-4.25}$	${}^{+4.33}_{-4.33}$ / ${}^{+3.75}_{-3.75}$	${}^{+8.07}_{-8.07}$ / ${}^{+8.36}_{-8.36}$	${}^{+6.12}_{-6.12}$ / ${}^{+7.40}_{-7.40}$
PDF	${}^{+0.82}_{-0.82}$ / ${}^{+0.78}_{-0.78}$	${}^{+20.64}_{-20.64}$ / ${}^{+20.28}_{-20.28}$	${}^{+0.95}_{-0.95}$ / ${}^{+0.94}_{-0.94}$	${}^{+0.19}_{-0.19}$ / ${}^{+0.19}_{-0.19}$	${}^{+0.79}_{-0.79}$ / ${}^{+0.78}_{-0.78}$	${}^{+0.19}_{-0.19}$ / ${}^{+0.26}_{-0.26}$	${}^{+1.05}_{-1.05}$ / ${}^{+0.93}_{-0.93}$	${}^{+4.68}_{-4.68}$ / ${}^{+3.50}_{-3.50}$
μ_F and μ_R scale (ME)	${}^{+9.40}_{-8.78}$ / ${}^{+9.10}_{-8.69}$	${}^{+7.62}_{-7.95}$ / ${}^{+7.31}_{-7.73}$	${}^{+7.18}_{-7.12}$ / ${}^{+5.95}_{-6.35}$	${}^{+14.97}_{-13.20}$ / ${}^{+14.90}_{-13.20}$	${}^{+10.33}_{-10.37}$ / ${}^{+9.75}_{-9.81}$	${}^{+5.57}_{-4.04}$ / ${}^{+6.72}_{-4.87}$	${}^{+11.06}_{-8.65}$ / ${}^{+12.17}_{-9.37}$	${}^{+7.19}_{-5.76}$ / ${}^{+7.00}_{-5.52}$
μ_F and μ_R scale (ISR)	${}^{-0.72}_{+0.87}$ / ${}^{-0.81}_{+0.98}$	${}^{-0.47}_{+0.36}$ / ${}^{-0.58}_{+0.55}$	—	${}^{-2.52}_{+3.05}$ / ${}^{-2.33}_{+2.79}$	${}^{+0.83}_{-1.23}$ / ${}^{+0.69}_{-1.07}$	${}^{+3.85}_{-4.59}$ / ${}^{-4.12}_{+5.13}$	${}^{-0.08}_{+0.09}$ / ${}^{-0.81}_{+1.02}$	—
μ_F and μ_R scale (FSR)	${}^{+0.26}_{-0.48}$ / ${}^{+0.40}_{-0.72}$	${}^{+0.51}_{-0.65}$ / ${}^{+1.93}_{-3.13}$	—	${}^{+0.98}_{-1.66}$ / ${}^{+1.17}_{-2.13}$	${}^{+0.28}_{-0.12}$ / ${}^{+0.87}_{-1.79}$	${}^{-2.23}_{-1.22}$ / ${}^{+0.93}_{+0.40}$	${}^{-0.03}_{-0.17}$ / ${}^{-0.15}_{+0.06}$	—
ME–PS matching scale (h_{damp})	${}^{+0.80}_{-1.92}$ / ${}^{+1.22}_{-1.53}$	—	—	${}^{+2.93}_{-6.98}$ / ${}^{+3.81}_{-4.74}$	—	—	—	—
$\not{E}_T$	${}^{+0.00}_{-0.00}$ / ${}^{+0.00}_{+0.00}$	${}^{+0.00}_{-0.00}$ / ${}^{+0.00}_{+0.00}$	${}^{+0.00}_{-0.00}$ / ${}^{+0.00}_{+0.00}$	${}^{+0.00}_{-0.00}$ / ${}^{+0.00}_{+0.00}$	${}^{+0.00}_{-0.00}$ / ${}^{+0.00}_{+0.00}$	${}^{+0.00}_{-0.00}$ / ${}^{+0.00}_{+0.00}$	${}^{+0.00}_{-0.00}$ / ${}^{+0.00}_{+0.00}$	${}^{+0.00}_{-0.00}$ / ${}^{+0.00}_{-0.00}$
NPL estimate	${}^{+0.35}_{-0.35}$ / —	—	—	—	—	—	—	—
Total systematic uncertainty	${}^{+26.92}_{-24.84}$ / ${}^{+27.74}_{-26.08}$	${}^{+23.51}_{-23.61}$ / ${}^{+22.64}_{-23.12}$	${}^{+32.35}_{-29.03}$ / ${}^{+34.36}_{-31.69}$	${}^{+23.25}_{-23.78}$ / ${}^{+23.25}_{-22.63}$	${}^{+14.47}_{-14.50}$ / ${}^{+13.37}_{-13.62}$	${}^{+20.67}_{-20.64}$ / ${}^{+23.32}_{-25.50}$	${}^{+23.76}_{-20.26}$ / ${}^{+22.34}_{-20.98}$	${}^{+15.44}_{-16.73}$ / ${}^{+13.55}_{-13.68}$

8.5 Unresolved Issues in Leading Order Samples

In 2017 simulated samples generated at LO, there are unresolved issues in the per-event weights used to determine the effect of some systematic uncertainties. The affected systematic uncertainties are those associated with

- the choice of factorisation and renormalisation scales for perturbative QCD simulation in the tHq, tWZ/tW$\ell\ell$, and $\mathrm{t\bar{t}\gamma}$ samples
- the PDFs in tWZ/tW$\ell\ell$ and $\mathrm{t\bar{t}\gamma}$ samples
- the choice of factorisation and renormalisation scales for nonperturbative QCD simulation in the $\mathrm{t\bar{t}\gamma}$ sample.

When using these weights, the effect on the underlying distributions of a $\pm1\sigma$ variation in the systematic uncertainty is increased by $>$1000×. The reason for this was not identified before the completion of this thesis. These systematic uncertainties are, therefore, excluded from further consideration; the excluded systematics are expected have a $\lesssim \pm\ 0.01\%$ effect on the overall normalisation in the signal region, judging by their effect in 2016 data.

Chapter 9

Results

Forty-two.

DEEP THOUGHT

This chapter presents the results of the tZq analysis. First, the effectiveness of the jet smearing mitigation used to alleviate the issues introduced in § 6.2.2 is examined. The details and performance of ML classifier training is then discussed, followed by the statistical methodology used to extract the signal strength and significance from the distribution of the classifier response, along with the signal strength and significance themselves.

9.1 Effectiveness of Incorrect Jet Smearing Mitigation

As stated in § 6.2.2, the jet smearing in the 2016 analysis was not implemented correctly in the μμ channel; however, a mitigation was applied. In order to judge the effectiveness of this mitigation, the 2017 μμ channel samples were processed both with the corrected smearing and the incorrect smearing with the mitigation applied. Figure 9.1 shows jet p_T distributions for both cases. Little difference is seen between the corrected and mitigated distributions, thus it is concluded that results from the 2016 μμ channel are not invalidated by this unresolved jet smearing issue.

9.2 Classification with a Boosted Decision Tree

As explored in Chapter 3, ML classifiers can be used to create a response representing the probability a given event is part of the tZq signal or from a background

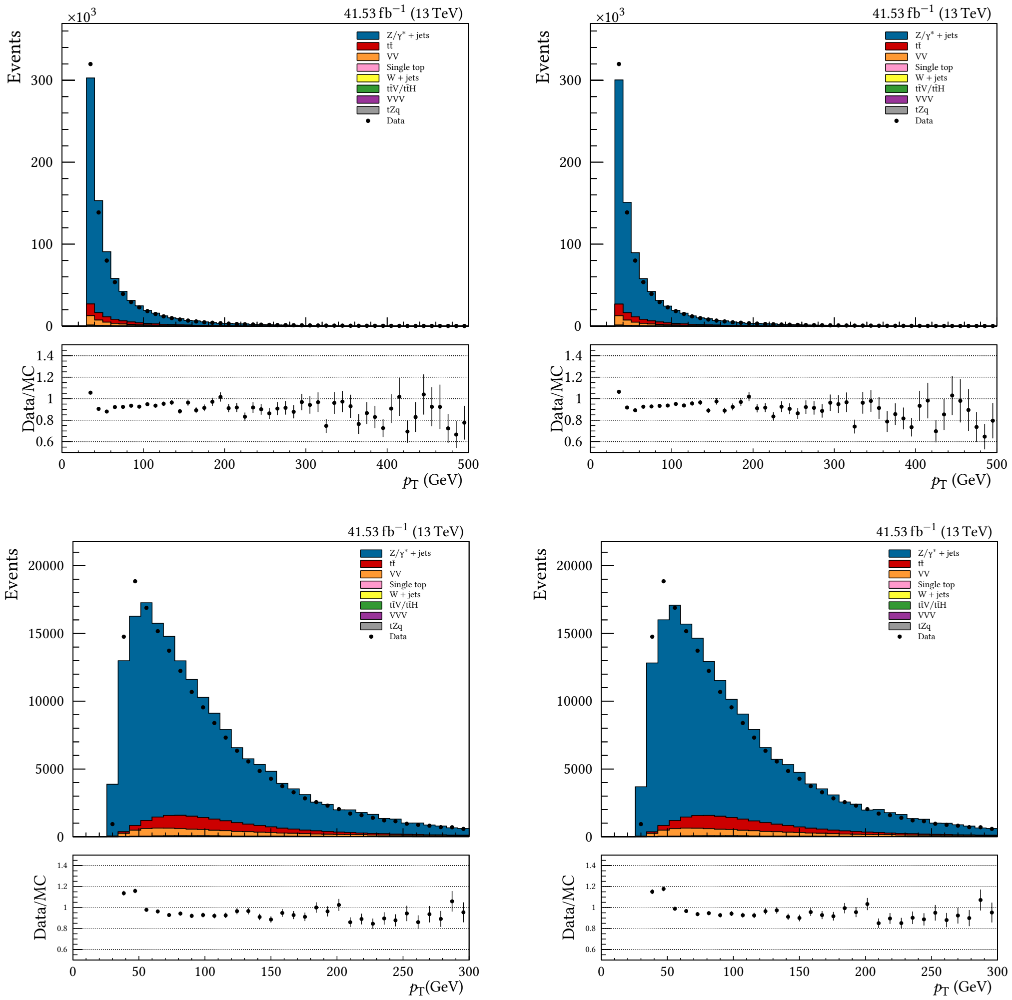


Figure 9.1: Comparison of the distribution of the p_T of all jets (top) and the leading jet only (bottom) between corrected (left) and mitigated (right) jet smearing in the 2017 μμ channel directly following the jet cuts.

process. Once the full event selection has been performed, a classifier is trained on the surviving simulated events in order to create a response distribution from which the signal strength and significance are extracted. Using the response of a classifier allows this extraction to be performed on a distribution that distinguishes tZq better than any existing observable individually.

Classifiers were trained on the nominal shapes of each process. The response distribution for shape uncertainties were obtained by applying the trained classifier to the shapes representing $\pm 1\sigma$ variations.

Two ML classification techniques were considered: MLPs, discussed in § 9.3, and BDTs, discussed in this section.

9.2.1 Input Preprocessing

Often it is necessary to perform some form of preprocessing on the input of a classifier in order to aid or enable training. In the tZq analysis, preprocessing is used to solve a data imbalance problem.

From Tables 7.5 and 7.6, it can be seen that tZq consists of ≈0.5% of the sample in the signal region (classifier input). This is an example of an imbalanced dataset: in terms of a binary classifier one class (the tZq signal) is far less represented than the other (the background). Training classifiers with such datasets can result in unintended and suboptimal results [201]. For example, a classifier that labelled every event as background would very accurate (low type I error rate), but useless for the purposes of identifying the tZq signal. To solve this, signal and background events are separately reweighted with a constant scale factor such that the normalisation for both is the same. These weights are used only in classifier training. A classifier that always predicts background would now have a type I error rate of 50%.

9.2.2 Feature Selection

The first step in optimising a classifier is *feature selection.* A naïve approach would be to perform no such selection and provide any classifier with as many features about a dataset as can be conceived, but in practice such an approach has two drawbacks:

- models with more features take longer to train
- in the same vein as the logic behind regularisation, a model with fewer features is considered a simpler model and so less prone to overtraining.

Considering this, an effort is made to select features that can effectively discriminate between signal and background i.e. have *discriminating power.* Furthermore, the selected features should be uncorrelated, as more correlated features provide a classifier with less information.

Over 100 different observables associated with reconstructed physics objects were created for use as features in the tZq analysis. A complete list is given in Table B.1. To aid in selecting the features with the most discriminative power, a *recursive feature elimination* (RFE) strategy [202] was employed to rank the features. This is performed by training a BDT with the default hyperparameters used by the XGBoost library—listed in Table 9.1—using all potential features. The least important feature, where importance is defined as the fraction of times the feature was used to split a tree, is then removed. Ties are broken at random. This

Table 9.1: The default hyperparameter values used by the XGBoost library. The hyperparameters are defined in § 9.2.6.

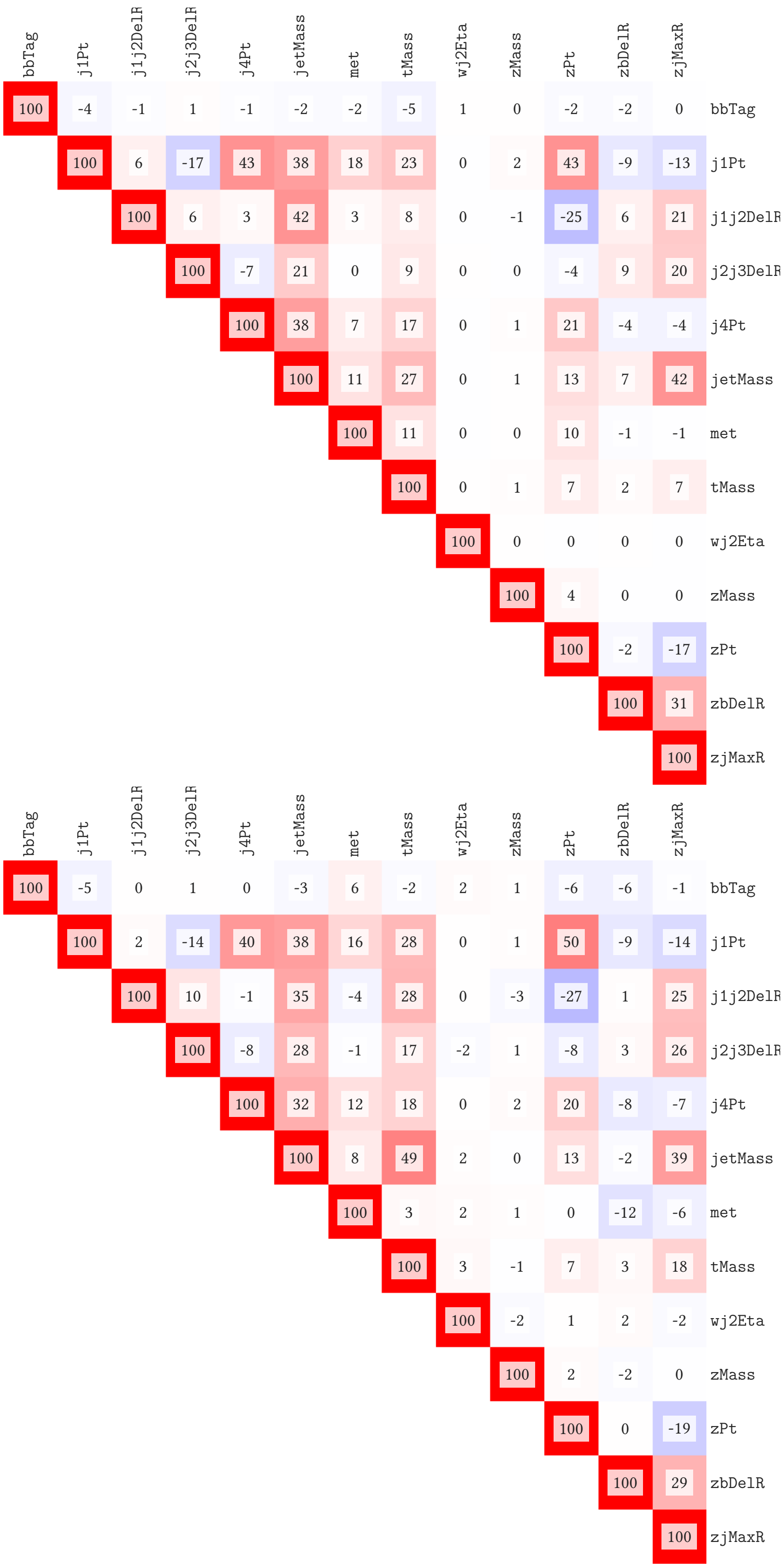
Hyperparameter	Default
`gamma`	0.0
`learning_rate`	0.1
`max_depth`	3
`min_child_weight`	1.0
`n_estimators`	100
`reg_alpha`	0.0
`reg_lambda`	1.0
`subsample`	1.0

process is repeated until one feature remains, which receives the top rank. The remaining features are ranked in the inverse order in which they were removed from consideration.

An advantage to the RFE method is that it will naturally rank uncorrelated features more highly. The mechanism behind this is that correlated features will "share" importance: the choice between using two highly correlated features to split a branch is effectively random, sensitive to statistical fluctuations. If one of a correlated set of features is removed, however, the remaining correlated features will "inherit" some of the importance of the removed feature. It is because of this that features are removed individually, rather than obtaining a ranking by training a single BDT with all possible features: the latter approach would undervalue the discriminating power of correlated features.

It is still possible, in the case of highly discriminating features, for correlated features to rank highly. After the ranking is performed, the correlation matrix of the highest-ranked features in signal and background is checked and highly correlated or anticorrelated features are pruned. Additionally, a feature is disqualified if it demonstrates poor agreement between data and simulation in the side-band or CRs.

The final set of selected features is given in Table 9.2 and the corresponding correlation matrices in Figures 9.2 to 9.5. Comparison of the distributions of the selected features in signal and background are displayed in Figures 9.6 to 9.9 and between data and simulation for the side-band region in Figures 9.10 to 9.13. Additional comparisons between data and simulation in the control and unblinded signal regions can be found in Appendix B.



bbTag	j1Pt	j1j2DelR	j2j3DelR	j4Pt	jetMass	met	tMass	wj2Eta	zMass	zPt	zbDelR	zjMaxR	
100	-4	-1	1	-1	-2	-2	-5	1	0	-2	-2	0	bbTag
	100	6	-17	43	38	18	23	0	2	43	-9	-13	j1Pt
		100	6	3	42	3	8	0	-1	-25	6	21	j1j2DelR
			100	-7	21	0	9	0	0	-4	9	20	j2j3DelR
				100	38	7	17	0	1	21	-4	-4	j4Pt
					100	11	27	0	1	13	7	42	jetMass
						100	11	0	0	10	-1	-1	met
							100	0	1	7	2	7	tMass
								100	0	0	0	0	wj2Eta
									100	4	0	0	zMass
										100	-2	-17	zPt
											100	31	zbDelR
												100	zjMaxR

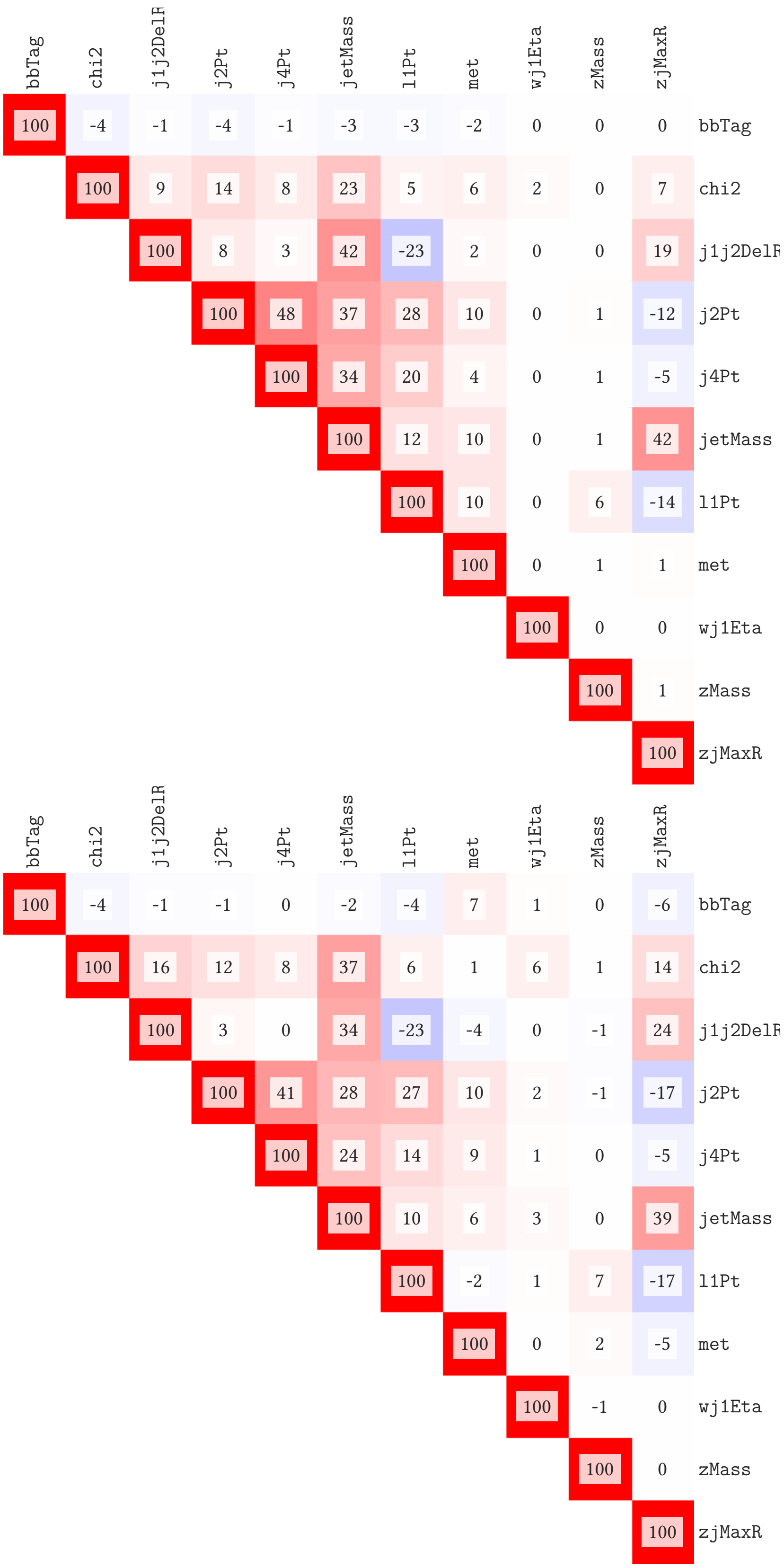
bbTag	j1Pt	j1j2DelR	j2j3DelR	j4Pt	jetMass	met	tMass	wj2Eta	zMass	zPt	zbDelR	zjMaxR	
100	-5	0	1	0	-3	6	-2	2	1	-6	-6	-1	bbTag
	100	2	-14	40	38	16	28	0	1	50	-9	-14	j1Pt
		100	10	-1	35	-4	28	0	-3	-27	1	25	j1j2DelR
			100	-8	28	-1	17	-2	1	-8	3	26	j2j3DelR
				100	32	12	18	0	2	20	-8	-7	j4Pt
					100	8	49	2	0	13	-2	39	jetMass
						100	3	2	1	0	-12	-6	met
							100	3	-1	7	3	18	tMass
								100	-2	1	2	-2	wj2Eta
									100	2	-2	0	zMass
										100	0	-19	zPt
											100	29	zbDelR
												100	zjMaxR

Figure 9.2: The feature correlation matrix for the ee signal region in 2016 for signal (top) and background (bottom) samples.

	bbTag	j1Eta	j1j2DelR	j2Pt	jetMass	l1Pt	met	tMass	wEta	zEta	zMass	zbDelR	zjMaxR
bbTag	100	0	-1	-3	-2	-2	-2	-5	1	0	0	-2	0
j1Eta		100	0	0	0	0	0	0	13	36	0	0	0
j1j2DelR			100	8	42	-21	3	8	0	0	0	7	21
j2Pt				100	38	26	13	24	0	0	1	-6	-10
jetMass					100	12	11	27	0	0	1	7	43
l1Pt						100	11	7	0	0	7	-2	-14
met							100	12	0	0	1	-1	0
tMass								100	0	0	1	3	7
wEta									100	11	0	0	0
zEta										100	0	0	1
zMass											100	0	0
zbDelR												100	32
zjMaxR													100

	bbTag	j1Eta	j1j2DelR	j2Pt	jetMass	l1Pt	met	tMass	wEta	zEta	zMass	zbDelR	zjMaxR
bbTag	100	2	-2	-2	-1	-3	8	-1	0	-1	-3	-11	-6
j1Eta		100	4	0	3	0	1	1	33	17	-1	0	0
j1j2DelR			100	5	34	-22	-5	27	1	1	2	3	26
j2Pt				100	37	23	11	25	2	-1	0	-12	-13
jetMass					100	12	9	49	-1	0	1	0	38
l1Pt						100	-2	8	2	0	17	0	-14
met							100	3	1	0	-8	-12	-8
tMass								100	1	2	2	0	18
wEta									100	4	1	-1	-4
zEta										100	0	3	5
zMass											100	6	6
zbDelR												100	35
zjMaxR													100

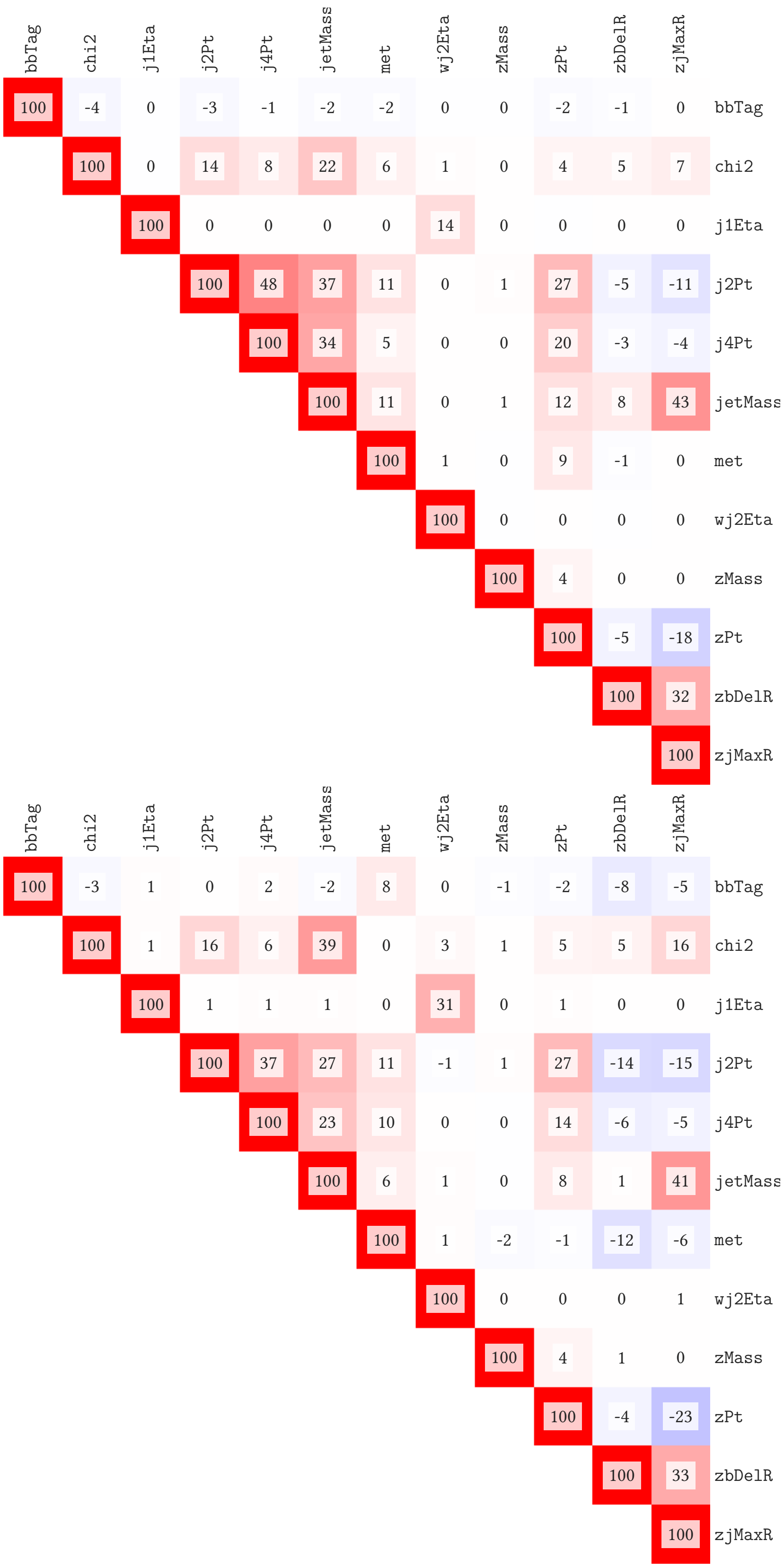
Figure 9.3: The feature correlation matrix for the μμ signal region in 2016 for signal (top) and background (bottom) samples.



	bbTag	chi2	j1j2DelR	j2Pt	j4Pt	jetMass	l1Pt	met	wj1Eta	zMass	zjMaxR
bbTag	100	-4	-1	-4	-1	-3	-3	-2	0	0	0
chi2		100	9	14	8	23	5	6	2	0	7
j1j2DelR			100	8	3	42	-23	2	0	0	19
j2Pt				100	48	37	28	10	0	1	-12
j4Pt					100	34	20	4	0	1	-5
jetMass						100	12	10	0	1	42
l1Pt							100	10	0	6	-14
met								100	0	1	1
wj1Eta									100	0	0
zMass										100	1
zjMaxR											100

	bbTag	chi2	j1j2DelR	j2Pt	j4Pt	jetMass	l1Pt	met	wj1Eta	zMass	zjMaxR
bbTag	100	-4	-1	-1	0	-2	-4	7	1	0	-6
chi2		100	16	12	8	37	6	1	6	1	14
j1j2DelR			100	3	0	34	-23	-4	0	-1	24
j2Pt				100	41	28	27	10	2	-1	-17
j4Pt					100	24	14	9	1	0	-5
jetMass						100	10	6	3	0	39
l1Pt							100	-2	1	7	-17
met								100	0	2	-5
wj1Eta									100	-1	0
zMass										100	0
zjMaxR											100

Figure 9.4: The feature correlation matrix for the ee signal region in 2017 for signal (top) and background (bottom) samples.



	bbTag	chi2	j1Eta	j2Pt	j4Pt	jetMass	met	wj2Eta	zMass	zPt	zbDelR	zjMaxR
bbTag	100	-4	0	-3	-1	-2	-2	0	0	-2	-1	0
chi2		100	0	14	8	22	6	1	0	4	5	7
j1Eta			100	0	0	0	0	14	0	0	0	0
j2Pt				100	48	37	11	0	1	27	-5	-11
j4Pt					100	34	5	0	0	20	-3	-4
jetMass						100	11	0	1	12	8	43
met							100	1	0	9	-1	0
wj2Eta								100	0	0	0	0
zMass									100	4	0	0
zPt										100	-5	-18
zbDelR											100	32
zjMaxR												100

	bbTag	chi2	j1Eta	j2Pt	j4Pt	jetMass	met	wj2Eta	zMass	zPt	zbDelR	zjMaxR
bbTag	100	-3	1	0	2	-2	8	0	-1	-2	-8	-5
chi2		100	1	16	6	39	0	3	1	5	5	16
j1Eta			100	1	1	1	0	31	0	1	0	0
j2Pt				100	37	27	11	-1	1	27	-14	-15
j4Pt					100	23	10	0	0	14	-6	-5
jetMass						100	6	1	0	8	1	41
met							100	1	-2	-1	-12	-6
wj2Eta								100	0	0	0	1
zMass									100	4	1	0
zPt										100	-4	-23
zbDelR											100	33
zjMaxR												100

Figure 9.5: The feature correlation matrix for the μμ signal region in 2017 for signal (top) and background (bottom) samples.

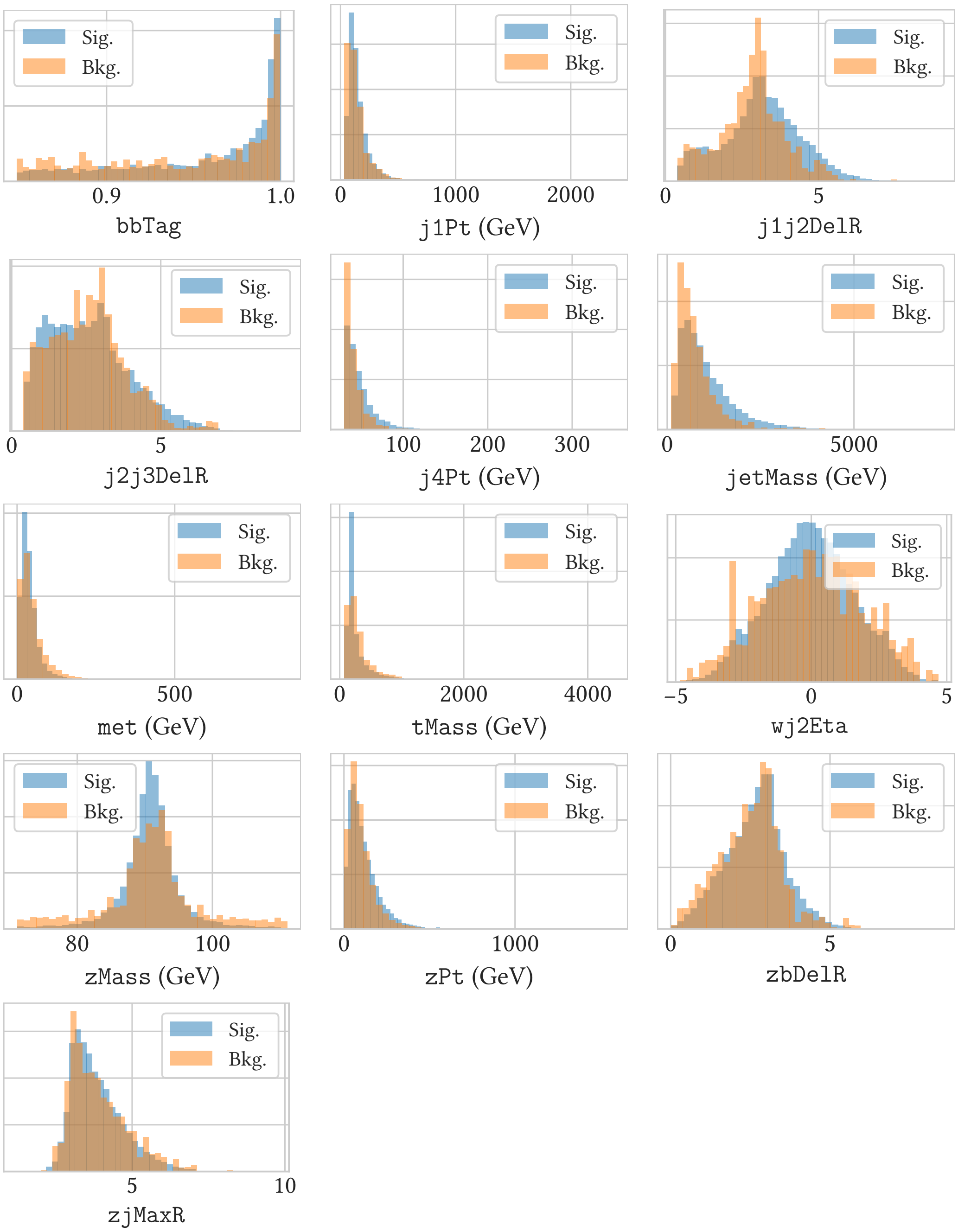


Figure 9.6: A comparison of the distribution in signal and background of the selected features for the BDT in the ee signal region in 2016. The signal and background distributions have been normalised to the same area. See Table 9.2 for a definition of each feature.

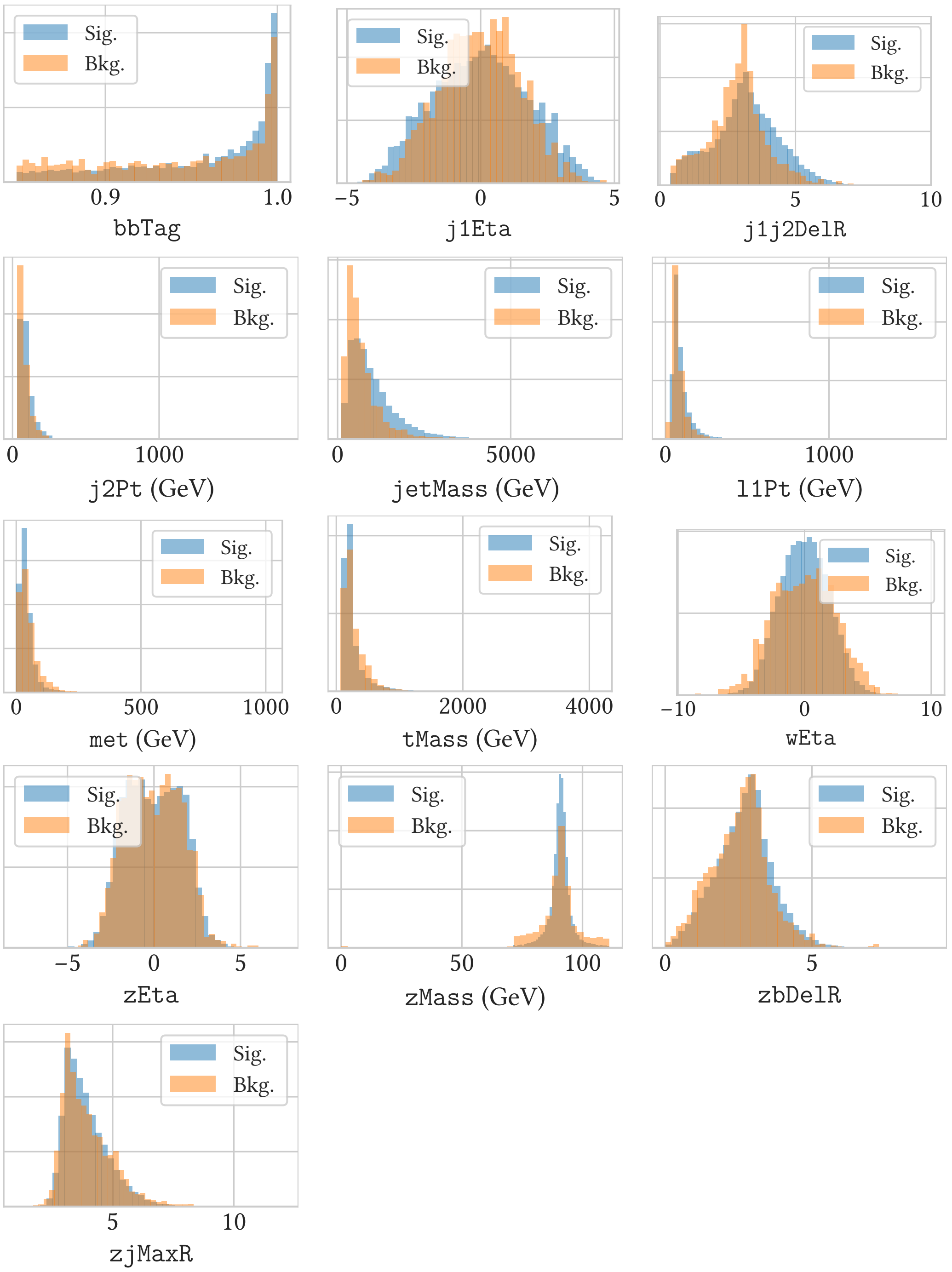


Figure 9.7: A comparison of the distribution in signal and background of the selected features for the BDT in the μμ signal region in 2016. The signal and background distributions have been normalised to the same area. See Table 9.2 for a definition of each feature.

Table 9.2: The features used in the BDTs. If a feature was not used in a BDT's training a '—' is given in the corresponding column. Otherwise, a number representing its rank in the RFE process is displayed.

Feature	Definition	2016		2017	
		ee	μμ	ee	μμ
`bbTag`	b tag discriminator of leading b tagged jet	11	12	11	15
`chi2`	χ^2 as used to define side-band region	—	—	5	3
`j1Eta`	η of the leading jet	—	11	—	12
`j1Pt`	p_{T} of the leading jet	3	—	—	—
`j1j2DelR`	ΔR between leading and second-leading jet	14	14	2	—
`j2Pt`	p_{T} of the second-leading jet	—	2	3	1
`j2j3DelR`	ΔR between second-leading and third-leading jet	17	—	—	—
`j4Pt`	p_{T} of the fourth-leading jet	13	—	14	19
`jetMass`	Total mass of all jets	4	3	4	2
`l1Pt`	p_{T} of the leading lepton	—	9	10	—
`met`	PF $\not{E}_{\mathrm{T}}$ in event	8	7	6	4
`tMass`	Mass of reconstructed top quark	2	1	—	—
`wEta`	η of the reconstructed W boson	—	6	—	—
`wj1Eta`	η of the leading W boson candidate jet	—	—	18	—
`wj2Eta`	η of the second-leading W boson candidate jet	7	—	—	13
`zEta`	η of reconstructed Z boson	—	5	—	—
`zMass`	Mass of reconstructed Z boson	9	8	7	5
`zPt`	p_{T} of reconstructed Z boson	6	—	—	14
`zbDelR`	ΔR between reconstructed Z boson and leading b tagged jet	12	19	—	6
`zjMaxR`	Greatest ΔR between the reconstructed Z boson and any jet	5	4	9	7

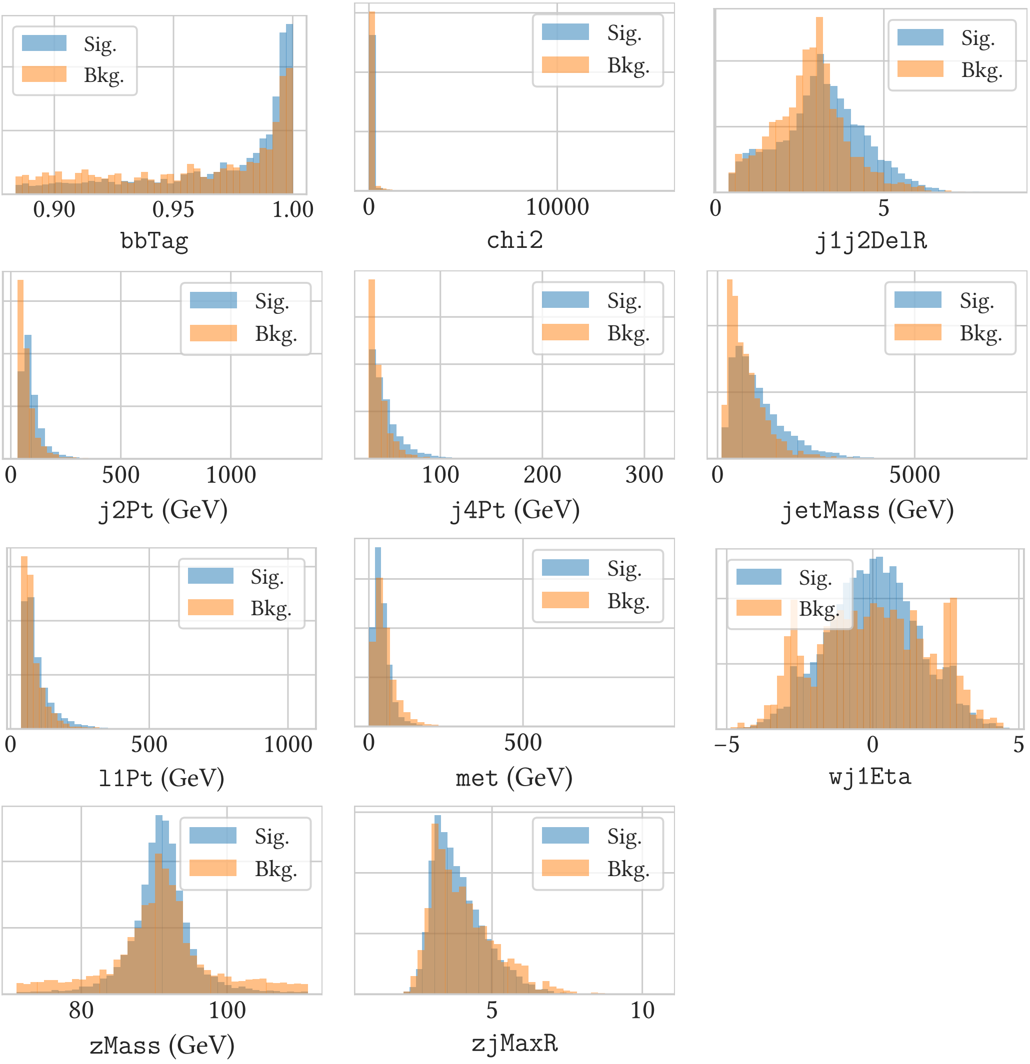


Figure 9.8: A comparison of the distribution in signal and background of the selected features for the BDT in the ee signal region in 2017. The signal and background distributions have been normalised to the same area. See Table 9.2 for a definition of each feature.

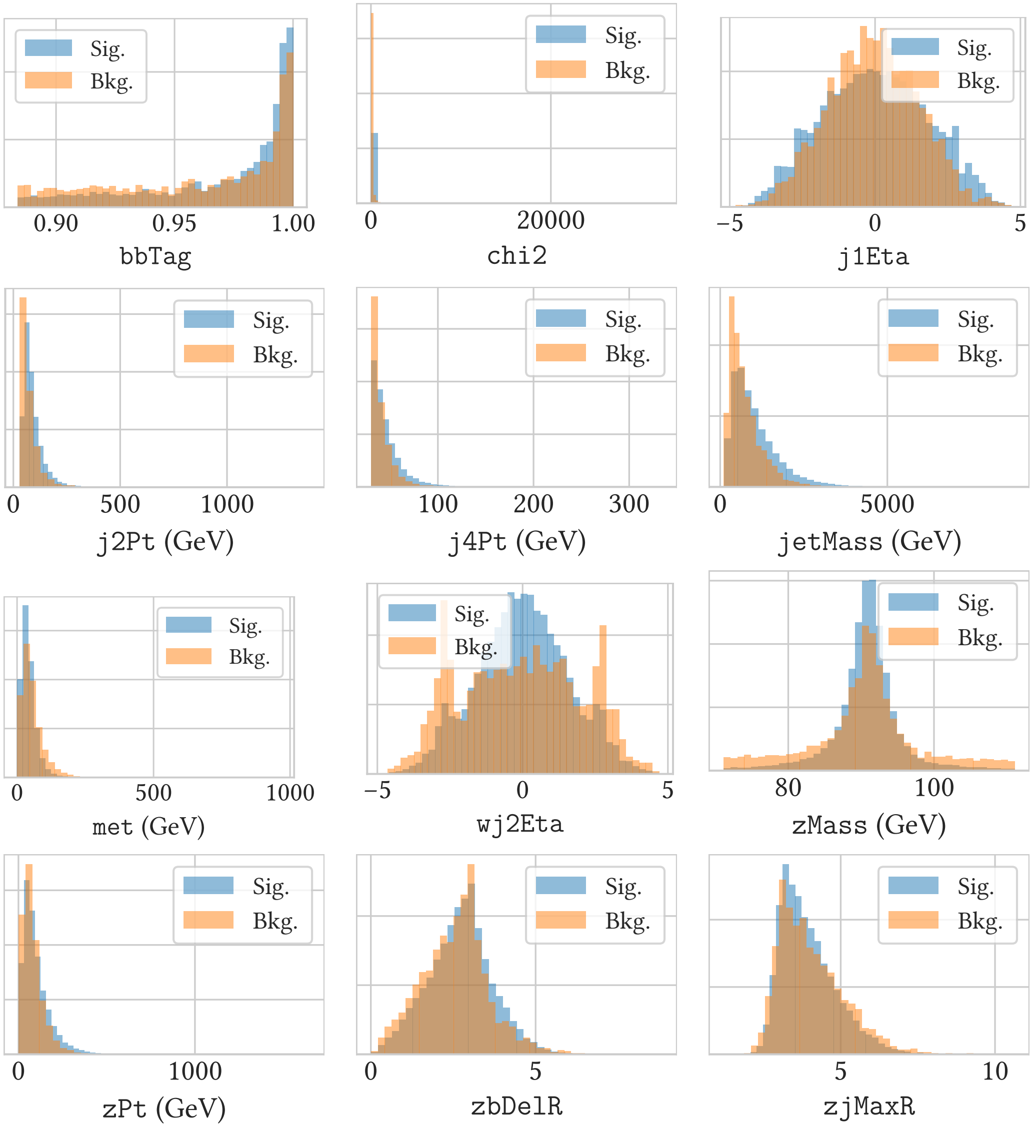


Figure 9.9: A comparison of the distribution in signal and background of the selected features for the BDT in the μμ signal region in 2017. The signal and background distributions have been normalised to the same area. See Table 9.2 for a definition of each feature.

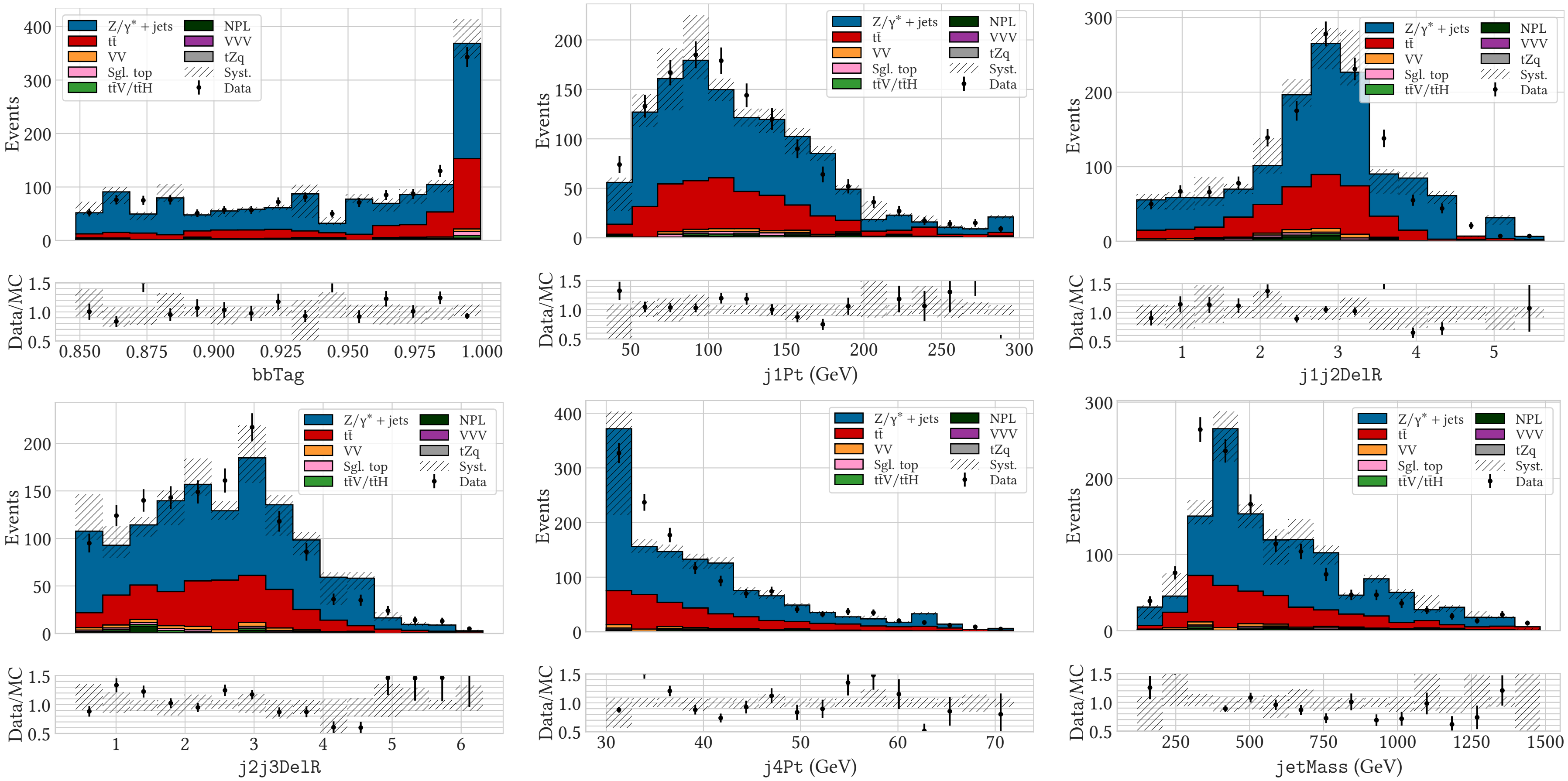


Figure 9.10: Distribution of selected features in simulated samples and data for the BDT in the 2016 ee side-band region. See Table 9.2 for a definition of each feature.

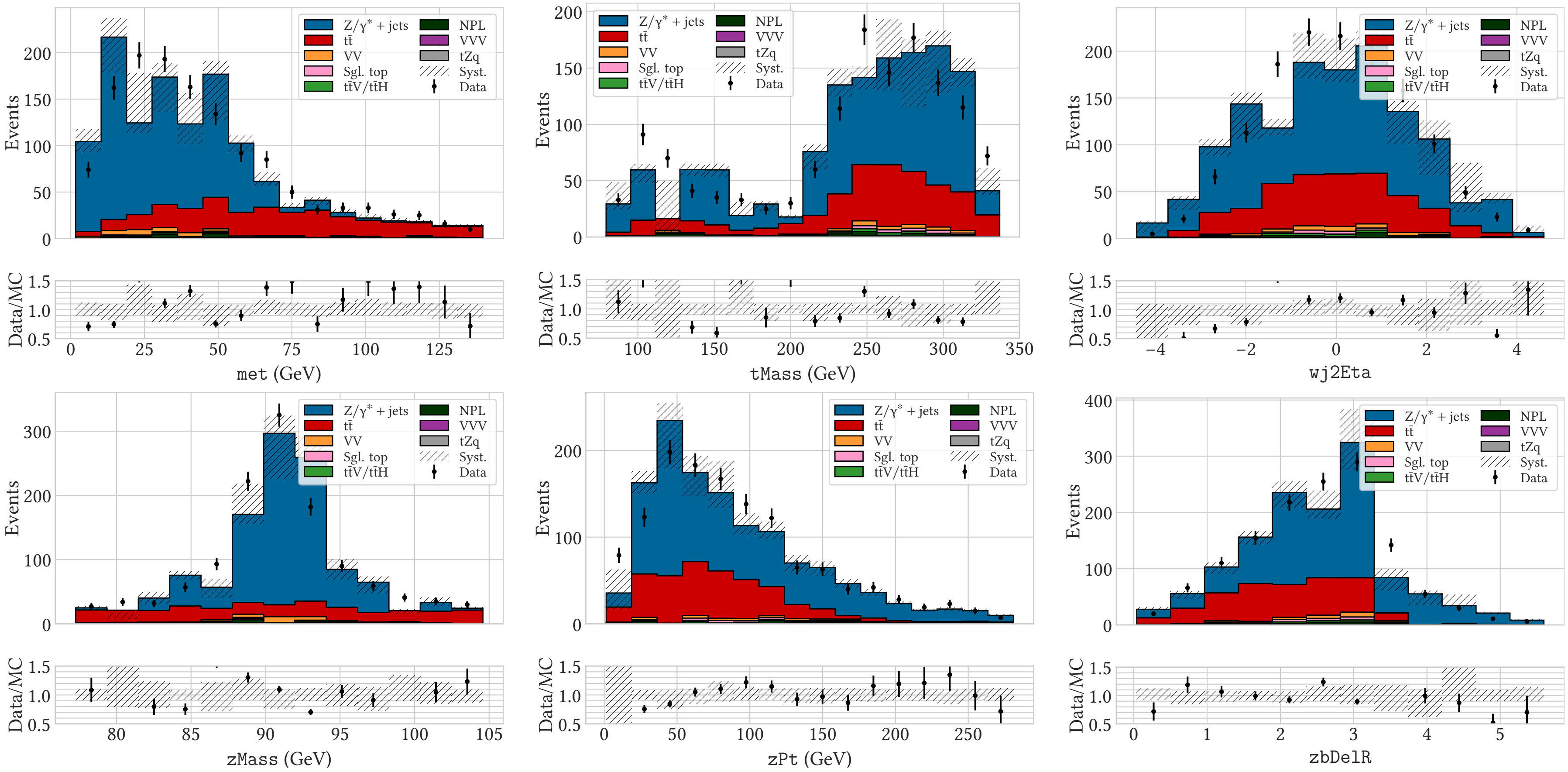


Figure 9.10: Distribution of selected features in simulated samples and data for the BDT in the 2016 ee side-band region. See Table 9.2 for a definition of each feature.

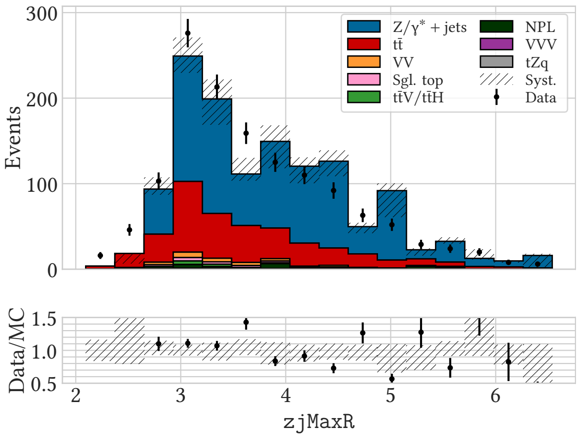


Figure 9.10: Distribution of selected features in simulated samples and data for the BDT in the 2016 ee side-band region. See Table 9.2 for a definition of each feature.

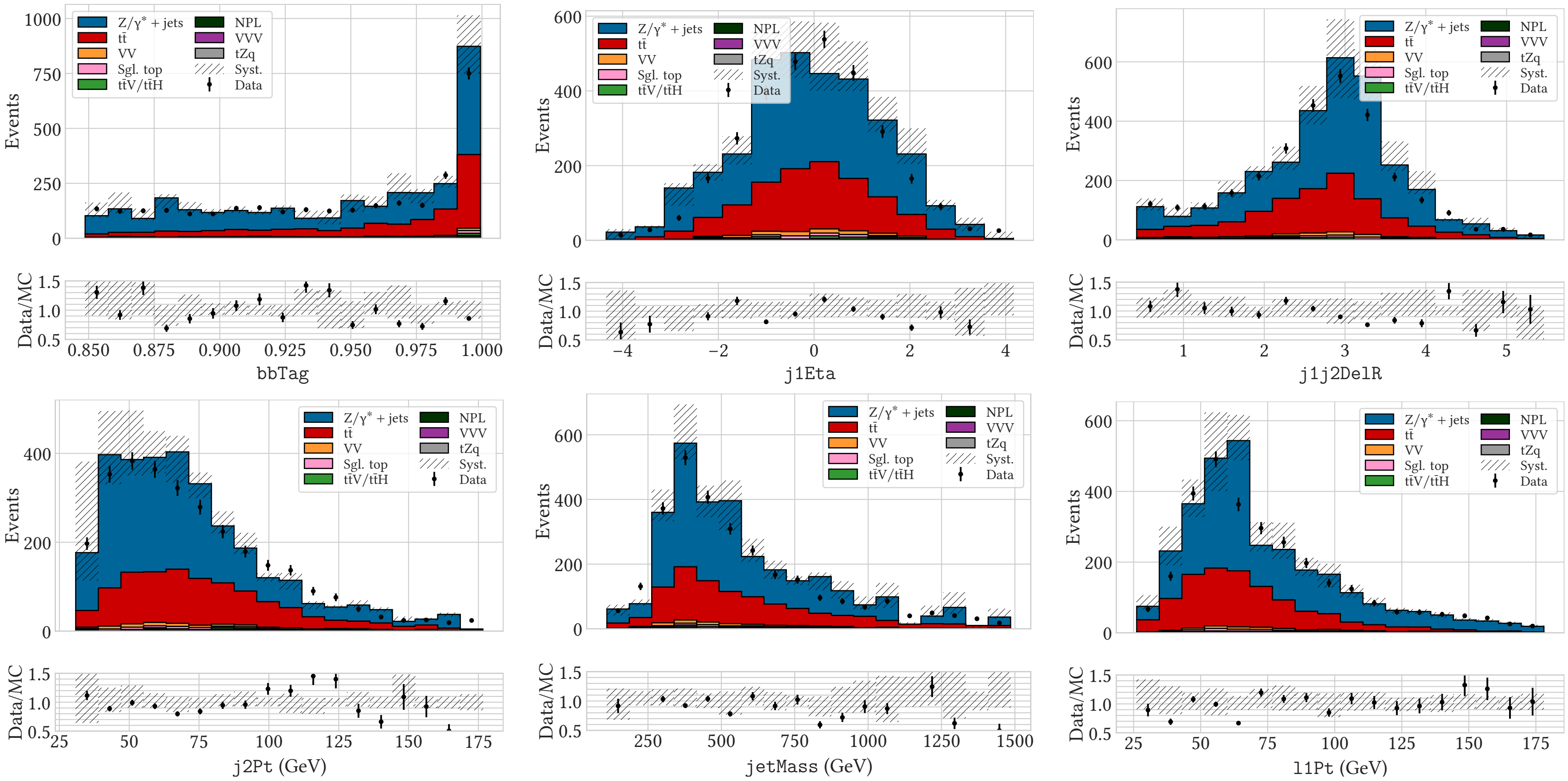


Figure 9.11: Distribution of selected features in simulated samples and data for the BDT in the 2016 μμ side-band region. See Table 9.2 for a definition of each feature.

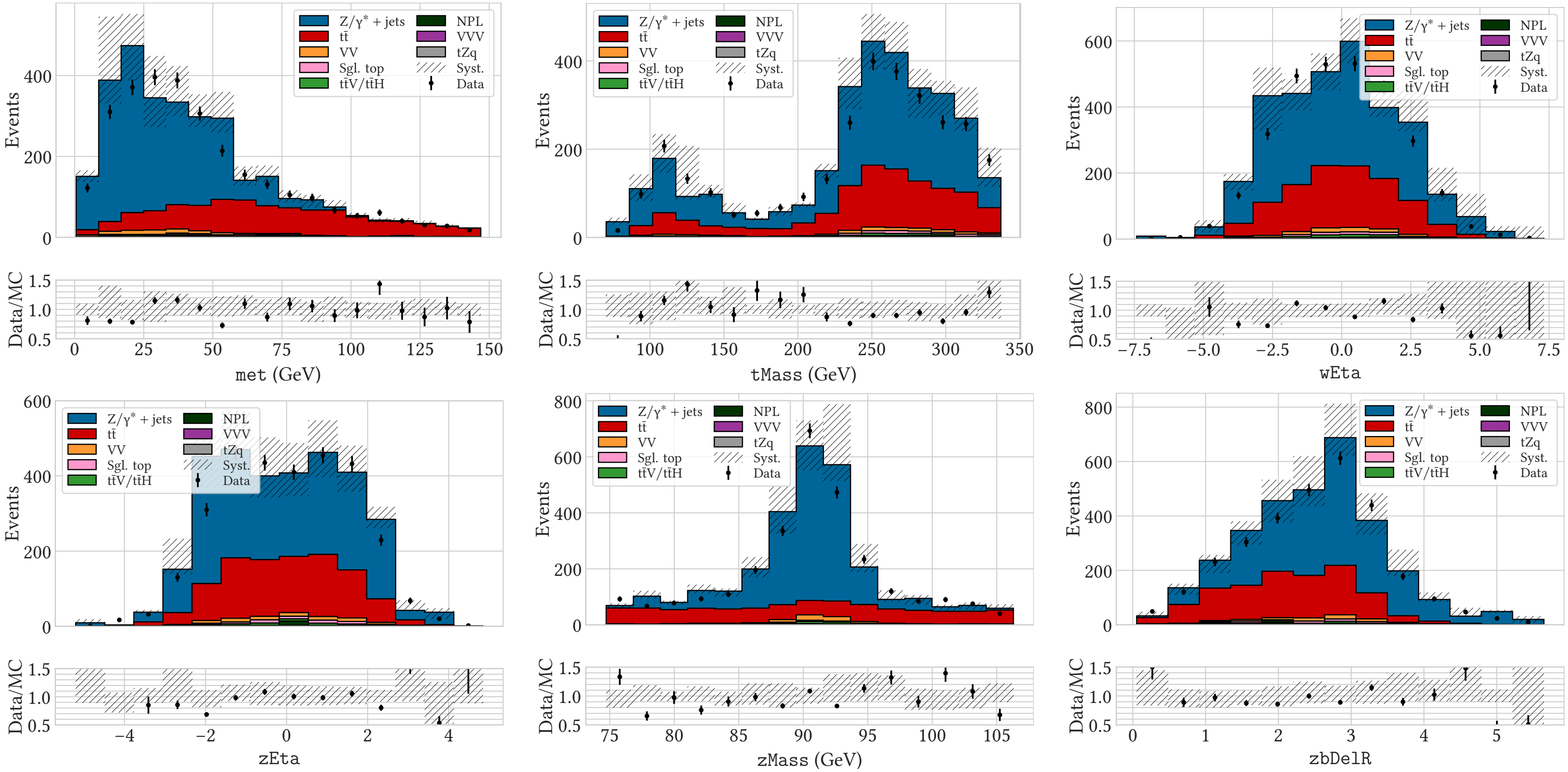


Figure 9.11: Distribution of selected features in simulated samples and data for the BDT in the 2016 μμ side-band region. See Table 9.2 for a definition of each feature.

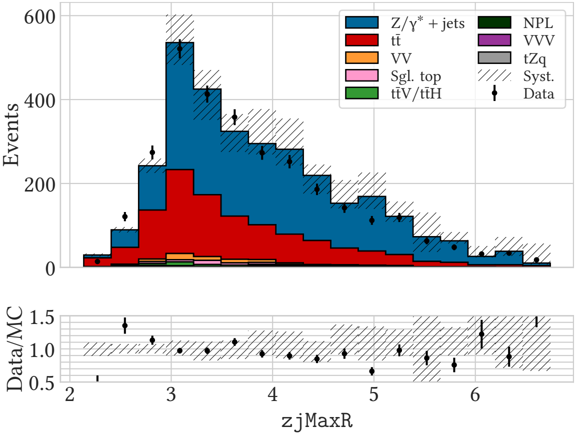


Figure 9.11: Distribution of selected features in simulated samples and data for the BDT in the 2016 μμ side-band region. See Table 9.2 for a definition of each feature.

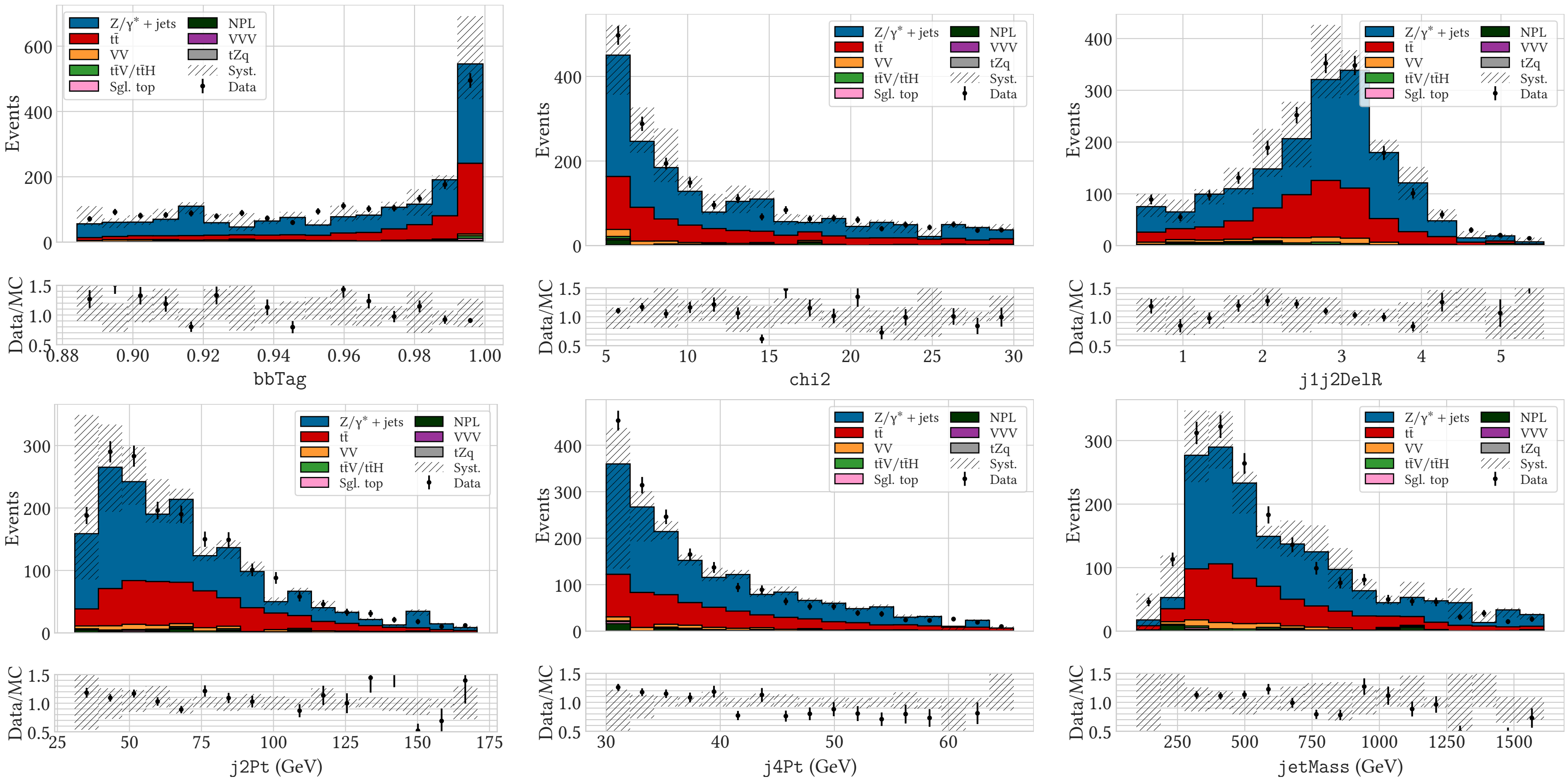


Figure 9.12: Distribution of selected features in simulated samples and data for the BDT in the 2017 ee side-band region. See Table 9.2 for a definition of each feature.

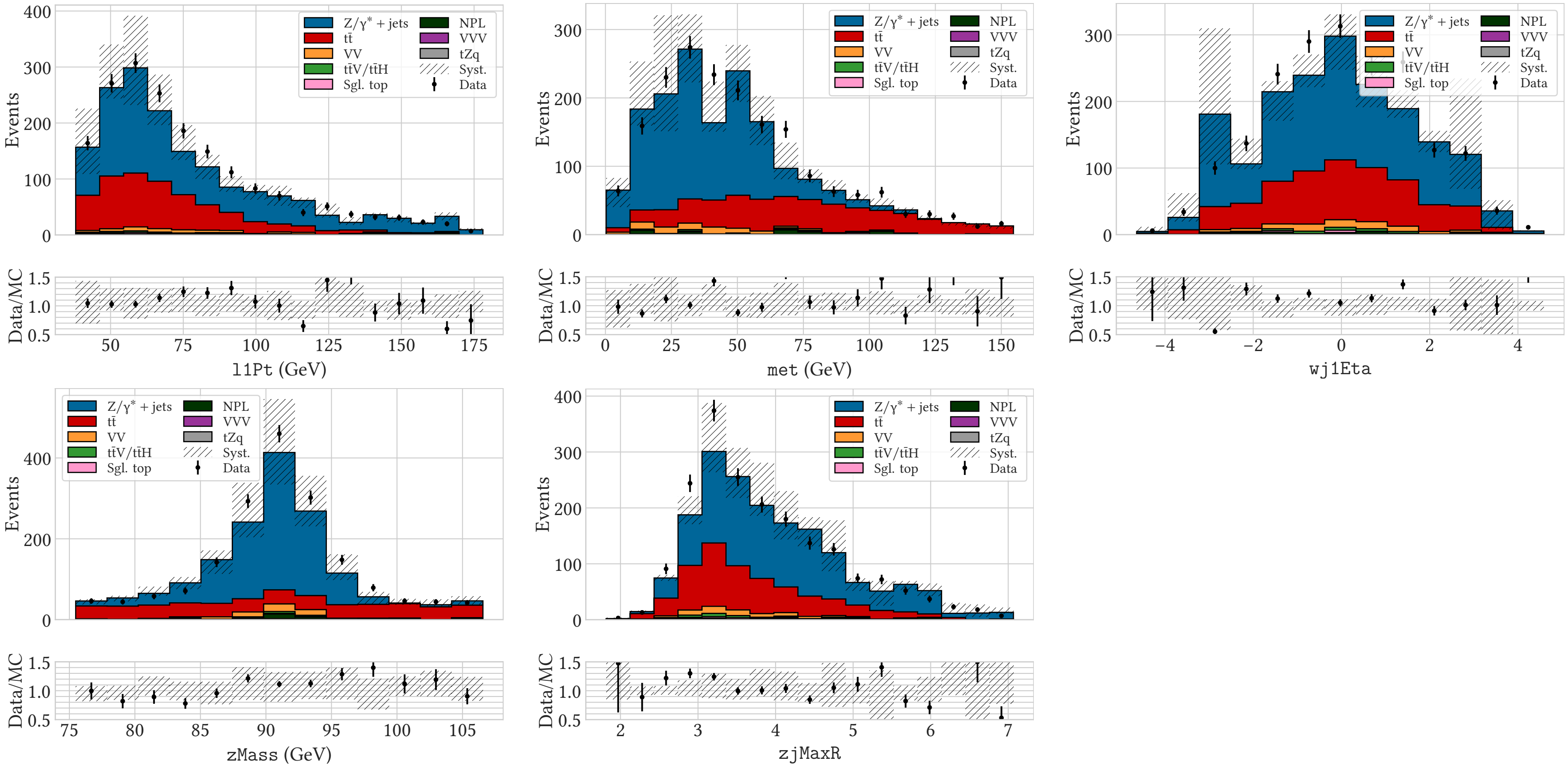


Figure 9.12: Distribution of selected features in simulated samples and data for the BDT in the 2017 ee side-band region. See Table 9.2 for a definition of each feature.

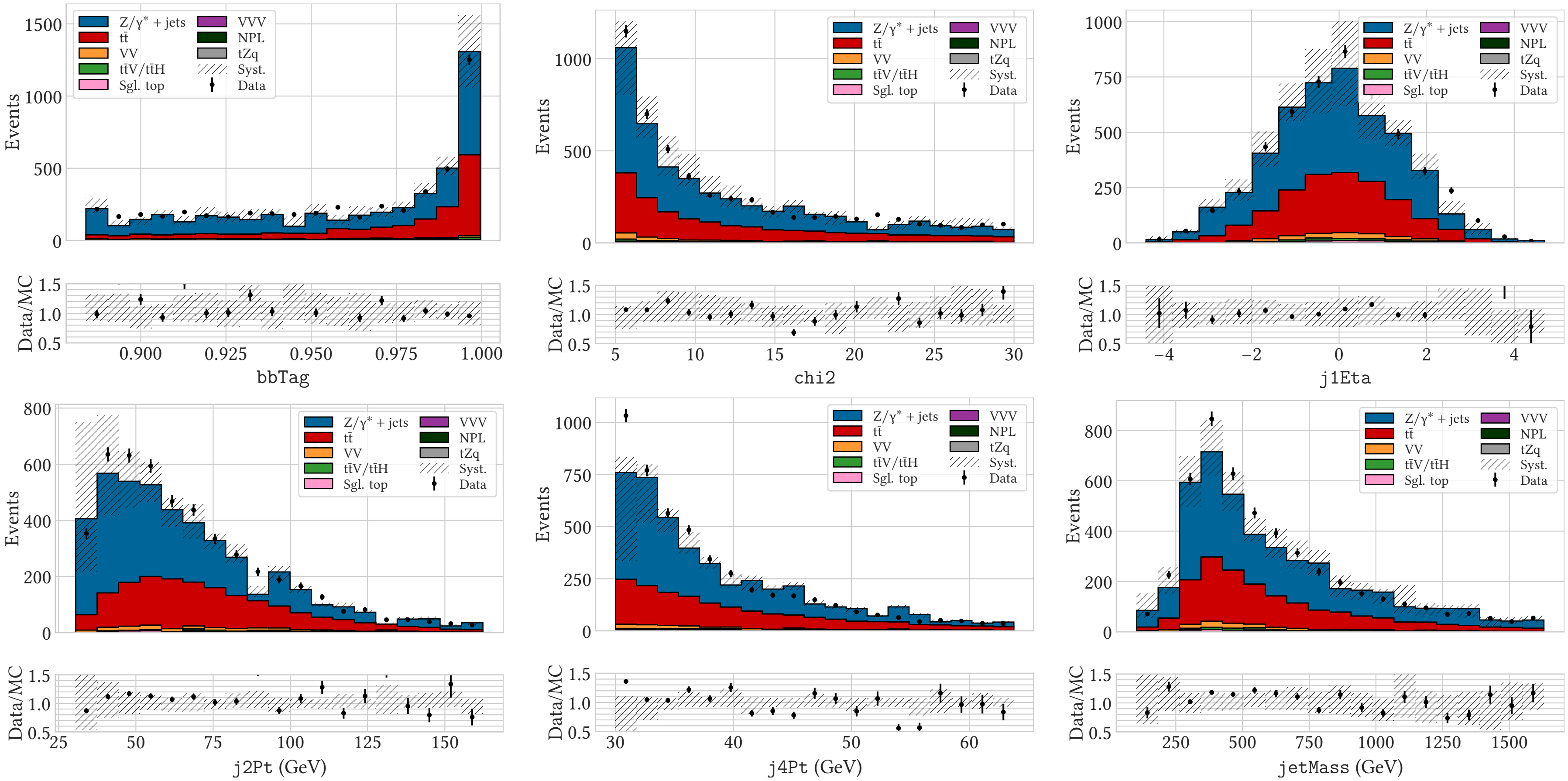


Figure 9.13: Distribution of selected features in simulated samples and data for the BDT in the 2017 μμ side-band region. See Table 9.2 for a definition of each feature.

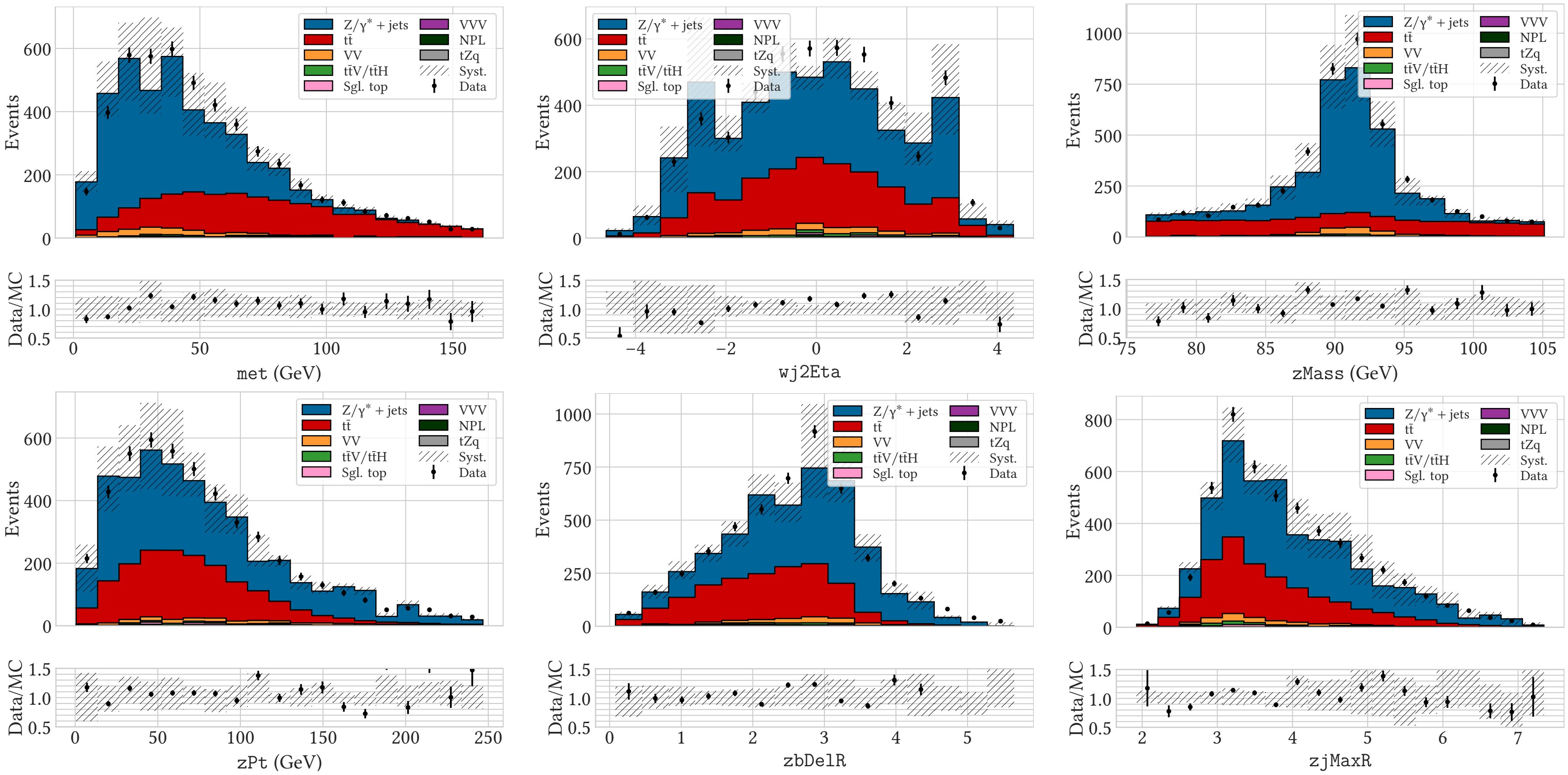


Figure 9.13: Distribution of selected features in simulated samples and data for the BDT in the 2017 μμ side-band region. See Table 9.2 for a definition of each feature.

9.2.3 Hyperparameter Optimisation with a Gaussian Process

The optimisation of hyperparameters has historically been a by-hand process relying on trial-and-error, intuition, and rules of thumb. For the BDTs in the tZq analysis a more adaptable, formal approach was desired and so methods of automatic hyperparameter optimisation were explored.

A widely-used technique for hyperparameter optimisation is the *exhaustive grid search* [203]. For each hyperparameter under consideration, a set of "sensible" values is chosen e.g. for a BDTs this could be $\{50, 100, 200, 500, 1000\}$ CARTs and $\nu = \{0.001, 0.01, 0.1\}$. The search is then performed by training a classifier with every possible combination of hyperparameter values, and evaluating each according to some predefined metric. The optimal hyperparameter choice is that which gives the best score according to this metric. This is a fairly primitive approach with three significant drawbacks:

1. Many hyperparameters exist on a continuum, but only the preprogrammed values can be considered.

2. Adding a hyperparameter increases the size of the space to be searched exponentially; it is possible to design a search that will take an infeasibly long time to complete with a handful of hyperparameters.

3. The search is not guided by information from previously-trained classifiers. For example, if during the search a BDT with 500 CARTs and $\nu = 0.01$ exhibited overtraining, an exhaustive grid search would still evaluate BDTs with more CARTs or an increased learning rate despite the fact they would likely also exhibit overtraining.

A refinement to the exhaustive grid search is the *random grid search* [203]. In this approach, a random subset of the possible combinations are tested. However, this only addresses the second concern and introduces the possibility that more optimal solutions are missed.

In the tZq analysis the hyperparameter optimisation approach models a predefined *classifier quality metric*, f, as a function of the input features, X, and the hyperparameters, $\boldsymbol{\vartheta}$. The input features, X, are decided by the event and feature selection process and so $f(X, \boldsymbol{\vartheta}) = f(\boldsymbol{\vartheta})$. If the quality metric is defined such that a lower value corresponds to a "better" classifier then this becomes a minimisation problem: find the hyperparameters ($\boldsymbol{\vartheta}$) that give the best classifier (minimum $f(\boldsymbol{\vartheta})$).

Evaluations of $f(\boldsymbol{\vartheta})$ are expensive, as they represent the training and evaluation of a classifier. Furthermore, $f(\boldsymbol{\vartheta})$ is a black box: no information on its analytic form or that of its gradients is available. This rules out or renders impractical many minimisation algorithms, such as stochastic gradient descent. An alternative, however, exists in regression: by building a regression model of $f(\boldsymbol{\vartheta})$, its minima can be approximated.

The Scikit-Optimize [204, 205] library, designed with this purpose in mind, is used to build the regression model and find the minimum of $f(\boldsymbol{\vartheta})$. This model is created using a GP with the Matérn $\nu = \frac{5}{2}$ kernel [122, 123],

$$k(x, x') = \sigma^2 \left(1 + \frac{\sqrt{5}\,(x - x')}{\rho} + \frac{5\,(x - x')^2}{3\rho^2} \right) \mathrm{e}^{-\frac{\sqrt{5}(x - x')}{\rho}}, \tag{9.1}$$

and is seeded by evaluating $f(\boldsymbol{\vartheta})$ at random values of $\boldsymbol{\vartheta}$. After this, future choices of $\boldsymbol{\vartheta}$ are motivated by the approximated form of $f(\boldsymbol{\vartheta})$: regions of the hyperparameter space where $f(\boldsymbol{\vartheta})$ has been found to be minimal or unexplored regions where the uncertainty in $f(\boldsymbol{\vartheta})$ is high are prioritised [206]. The model of $f(\boldsymbol{\vartheta})$ is updated following each iteration (choice of $\boldsymbol{\vartheta}$). After a predetermined number of iterations are performed, the choice of $\boldsymbol{\vartheta}$ that returned the smallest value of $f(\boldsymbol{\vartheta})$ is used as the hyperparameters for the BDT.

9.2.4 *k*-Fold Cross-Validation

When training the BDT, 20% of X is excluded from the training. This forms a *test set* of unseen input, X_{test}, and the remaining 80% of the available input forms the *training set*, X_{train}, used during the training of the classifier. After training, the KS test is performed between the classifier response in X_{test} and X_{train} in order to provide a measure of overtraining (as described previously in § 3.1.1). The ROC curve curves and AUROC of X_{test} and X_{train} are also compared. As both the test set, training set, and data from the CMS detector will ultimately be used in the signal extraction, it is essential that performance between seen and unseen input is as consistent as possible.

In order to further ensure this, k-fold cross-validation [208] is used during the hyperparameter search. With this technique X_{train} is randomly split into k *partitions*. Then, k classifiers are trained, with a different partition—dubbed the *validation partition*—reserved from the training each time. The metric, f, is then evaluated for each classifier on the corresponding validation partition. The final score as the average f over the k classifiers. A visual representation is shown in Figure 9.14. Once the best classifier is selected, it is retrained on all of X_{train}.

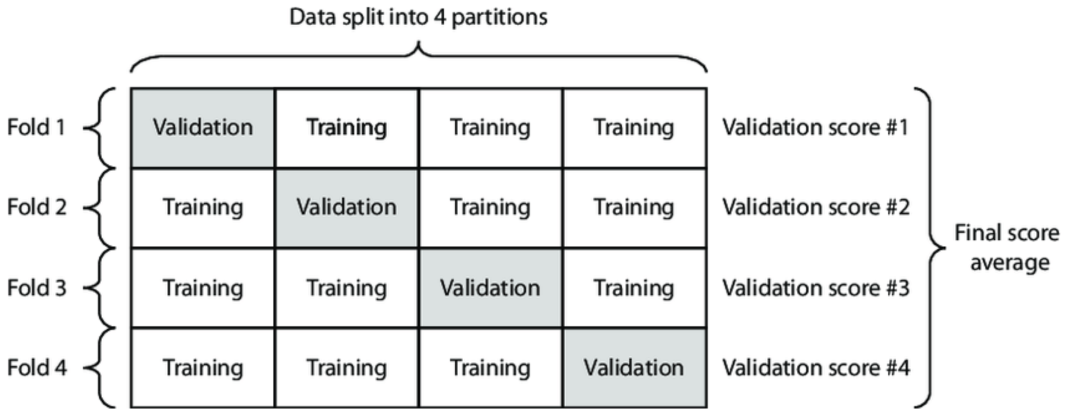


Figure 9.14: Representation of the k-fold partitioning process with $k = 4$ folds. Figure taken from [207].

9.2.5 The Classifier Quality Metric

The most obvious choice for f is the AUROC on the validation partition,

$$f = -\,\mathrm{AUROC}\left(P_{\mathrm{val}}\right), \tag{9.2}$$

because the AUROC provides an overall description of how effectively the classifier is able to separate signal and background (the negative AUROC is used as f is minimised). However, this metric was not found to be sufficient and would lead to the training of classifiers that, while performing well on unseen data, would perform significantly better on X_{test}. In response, Equation 9.2 was modified to introduce a term penalising inconsistent performance between training partitions (P_{train}) and validation partitions:

$$f = \mathrm{AUROC}\left(P_{\mathrm{val}}\right) - \left|\mathrm{AUROC}\left(P_{\mathrm{val}}\right) - \mathrm{AUROC}\left(P_{\mathrm{train}}\right)\right|, \tag{9.3}$$

which was used as the final classifier quality metric.

9.2.6 Results of the Hyperparameter Search

The hyperparameter search was run for 100 iterations, the first 10 of which used random values of $\boldsymbol{\vartheta}$. The hyperparameters considered in the search were:

`gamma` The regularisation strength hyperparameter, γ, in Equation 3.19

`learning_rate` The learning rate, ν, defined in Equation 3.18

`max_depth` The maximum depth of a CART i.e. the maximum number of edges between the root node and a leaf node

Table 9.3: Results of optimal BDT hyperparameter search. The allowed ranges and selected values in ee and μμ channels in 2016 and 2017 are shown. Hyperparameters with a uniform prior were sampled uniformly between the minimum and maximum, hyperparameters with a log-uniform prior were sampled uniformly between log(minimum) and log(maximum).

Hyperparameter	Prior	Min	Max	2016		2017	
				ee	μμ	ee	μμ
`gamma`	log-uniform	1×10^{-5}	10	10	7.81	8.22×10^{-5}	10
`learning_rate`	log-uniform	1×10^{-5}	0.1	0.1	3.81×10^{-3}	1.15×10^{-2}	2.270×10^{-3}
`max_depth`	uniform	2	8	8	2	2	4
`min_child_weight`	log-uniform	1×10^{-5}	1000	146	277	27.0	190
`n_estimators`	uniform	32	5000	456	3095	1307	4923
`reg_alpha`	log-uniform	1×10^{-5}	1×10^{5}	1×10^{-5}	1.66	2.01	16.5
`reg_lambda`	log-uniform	1×10^{-5}	1×10^{5}	167	427	2.81×10^{-5}	8.21
`subsample`	uniform	0.5	0.8	0.720	0.522	0.500	0.551

`min_child_weight` The minimum number of entries a branch must contain in order for a partition to occur

`n_estimators` The number of CARTs in the final model.

`reg_alpha` The regularisation strength hyperparameter α in Equation 3.19

`reg_lambda` The regularisation strength hyperparameter λ in Equation 3.19

`subsample` The fraction of the training input used to train each CART (stochastic boosting).

Table 9.3 shows the best hyperparameters found for the ee and μμ channels in the 2016 and 2017 analyses.

9.2.7 Performance

The response of the BDT in 2016 and 2017 is shown in Figure 9.15, with the corresponding ROC curves in Figure 9.16. Performance is favourable overall, with a typical AUROC of 0.81 but, as seen in Figures 9.19 and 9.20, the classifier does struggle to recognize the $Z/\gamma^* + \text{jets}$ background. Performance in the test and training samples is similar; results from the KS test are presented in Table 9.5. Table 9.4 lists the importance of each selected feature in the final model.

9.3 Classification with a Multilayer Perceptron

MLPs were considered as an alternative to BDTs to distinguish the tZq signal from background. This section describes the procedure used to optimise and train the MLPs and compares their performance to what is seen in BDTs. The networks described herein were implemented using the KERAS [209] library with the THEANO [210] back end.

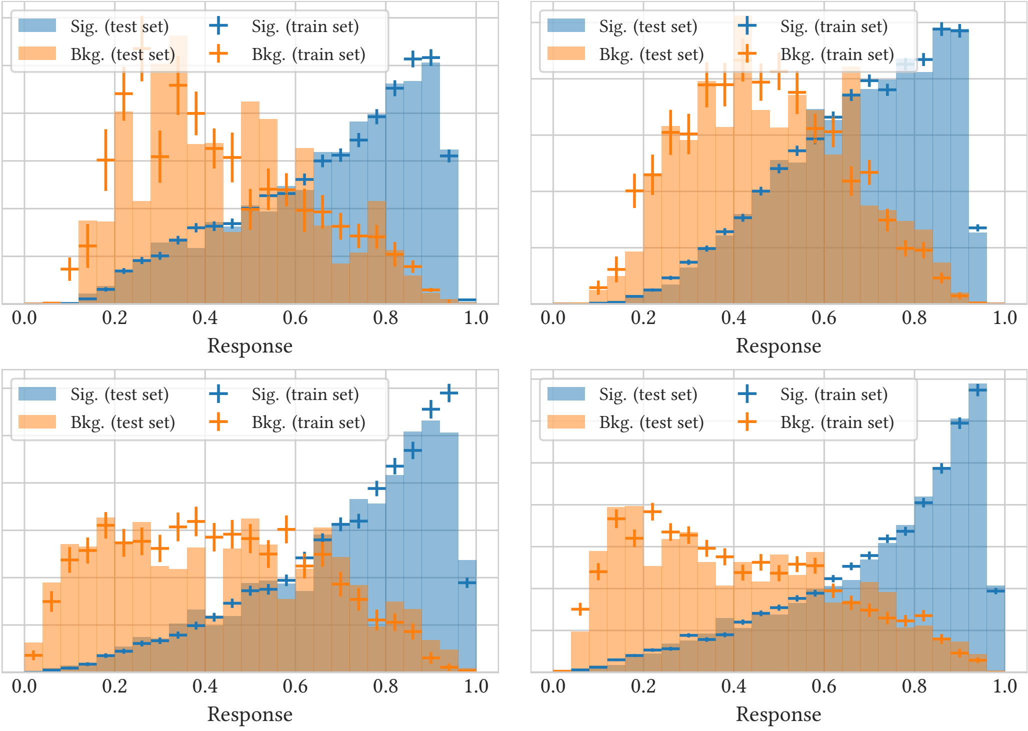


Figure 9.15: The distribution of the response of the trained BDT on the test and training sets in 2016 (top) and 2017 (bottom) in the ee (left) and μμ (right) channels.

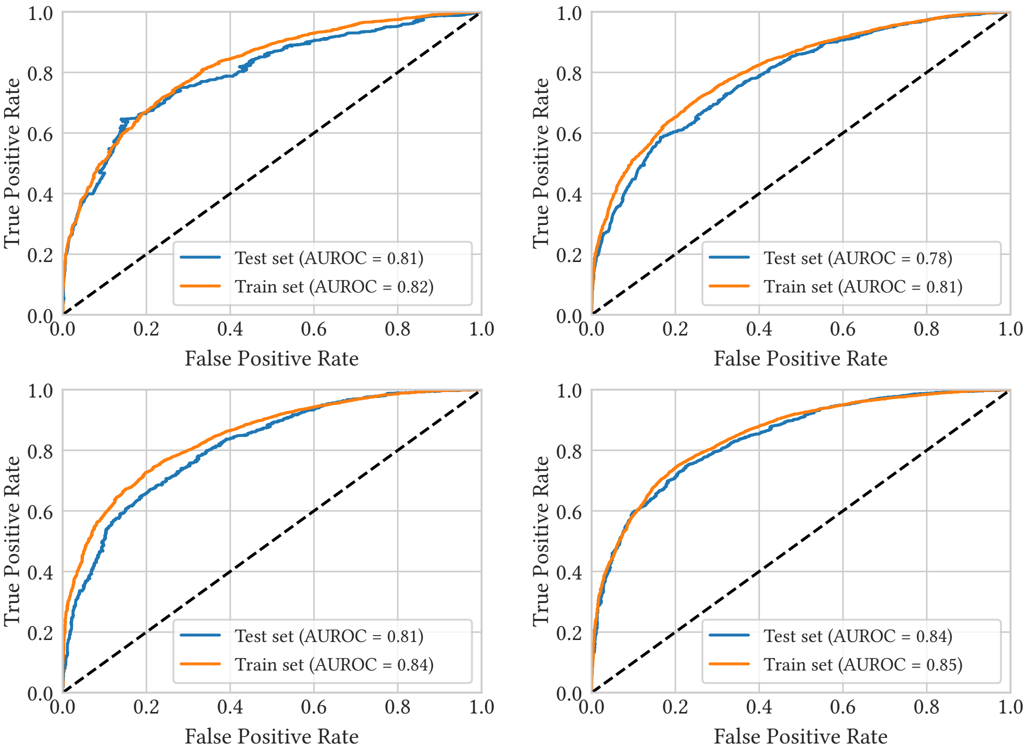


Figure 9.16: The ROC curve of the trained BDT on the test and training sets in 2016 (top) and 2017 (bottom) in the ee (left) and μμ (right) channels.

Table 9.4: The importance of the selected features in the trained BDT

(a) Feature importances for the ee channel in 2016

Feature	Importance
`tMass`	1.766×10^{-1}
`jetMass`	1.460×10^{-1}
`j1Pt`	9.847×10^{-2}
`zMass`	9.847×10^{-2}
`zjMaxR`	7.980×10^{-2}
`bbTag`	7.640×10^{-2}
`j4Pt`	7.640×10^{-2}
`j1j2DelR`	6.112×10^{-2}
`wj2Eta`	6.112×10^{-2}
`met`	4.924×10^{-2}
`zPt`	3.735×10^{-2}
`zbDelR`	2.886×10^{-2}
`j2j3DelR`	1.019×10^{-2}

(b) Feature importances for the μμ channel in 2016

Feature	Importance
`tMass`	1.947×10^{-1}
`jetMass`	1.707×10^{-1}
`j2Pt`	1.170×10^{-1}
`zMass`	9.631×10^{-2}
`bbTag`	9.169×10^{-2}
`zjMaxR`	8.759×10^{-2}
`j1j2DelR`	5.283×10^{-2}
`wEta`	4.591×10^{-2}
`met`	3.924×10^{-2}
`j1Eta`	3.668×10^{-2}
`l1Pt`	3.668×10^{-2}
`zbDelR`	1.774×10^{-2}
`zEta`	1.064×10^{-2}

(c) Feature importances for the ee channel in 2017

Feature	Importance
`chi2`	1.395×10^{-1}
`jetMass`	1.319×10^{-1}
`j2Pt`	1.280×10^{-1}
`zMass`	1.118×10^{-1}
`met`	8.856×10^{-2}
`j1j2DelR`	8.859×10^{-2}
`zjMaxR`	8.307×10^{-2}
`wj1Eta`	7.785×10^{-2}
`bbTag`	6.870×10^{-2}
`j4Pt`	6.113×10^{-2}
`l1Pt`	2.351×10^{-2}

(d) Feature importances for the μμ channel in 2017

Feature	Importance
`jetMass`	1.456×10^{-1}
`chi2`	1.300×10^{-1}
`zMass`	1.011×10^{-1}
`j2Pt`	9.739×10^{-2}
`wj2Eta`	9.521×10^{-2}
`met`	8.933×10^{-2}
`zjMaxR`	8.105×10^{-2}
`bbTag`	8.009×10^{-2}
`j1Eta`	6.608×10^{-2}
`zbDelR`	5.608×10^{-2}
`j4Pt`	4.479×10^{-2}
`zPt`	1.322×10^{-2}

Table 9.5: The p-values from the KS test between the distribution of the BDT response in test and training sets.

Channel	2016		2016	
	Signal	Background	Signal	Background
ee	0.887	0.158	0.581	0.051
μμ	0.951	0.143	0.196	0.161

9.3.1 Data Preprocessing

As with the BDT, before samples are presented to the MLP for training the tZq signal and background processes are reweighted such that their normalisations are the same. Following this, an additional preprocessing step is performed: every feature is translated and scaled to have zero mean and unit variance. This transform allows the training of a neural network to converge sooner, as it no longer needs to account for relative scales of the provided features [211].

9.3.2 Feature Selection

When selecting features for the MLP, an emphasis was placed on low-level features (i.e. jet and lepton properties) from which higher-level features (e.g. reconstructed W mass, Z pseudorapidity) can be constructed. By the nature of how an ANN operates, MLPs given access to low-level features can theoretically reconstruct its own useful high-level features. Less emphasis was placed on the removal of correlated features, as MLPs should learn to reduce the corresponding activations if necessary during training.

The transverse momentums, pseudorapidities, and φ coordinates of the two leptons and four leading jets were used as features. Additionally, the b tag discriminator of each jet and the relative isolation of each lepton were included. Finally, some higher-level features ranked highly in the BDT training were included, viz. `bbTag`, `jetMass`, `met`, `zjMaxR`, `tMass` (2016 only), and `chi2` (2017 only).

9.3.3 Hyperparameter Selection

The parameter space of MLPs makes it less conducive to hyperparameter optimisation with a GP. Some parameters, such as the choice of activation function and optimiser, are qualitative. Beyond this, the parameter space of the qualitative parameters does not form a hypercube; for example, the regularisation strength of the nth hidden layer cannot be chosen if only $n-1$ hidden layers exist. Due to these additional complications introduced by MLPs, hyperparameter optimisation was performed by hand. Extending the automatic hyperparameter selection methodology used for the BDT to MLPs is a possible direction for future work, discussed in § 10.3.

In the 2016 ee channel, MLPs with a single hidden layer of 32 neurons with a tanh activation function were used. Regularisation was applied to the hidden layer in the form of $p_\mathrm{d} = 20\%$ dropout and L1L2 regularisation with $\alpha = 2 \times 10^{-4}$, $\lambda = 1 \times 10^{-3}$.

Table 9.6: The p-values from the KS test between the distribution of the MLP response in test and training sets.

Channel	2016		2017	
	Signal	Background	Signal	Background
ee	0.759	0.337	0.471	0.277
μμ	0.245	0.350	0.099	0.468

In the 2017 ee channel a single hidden layer with tanh activation was again used, but at a reduced size of 16 neurons. Strong regularisation was required to prevent overtraining, with $p_{\mathrm{d}} = 33\%$ dropout, and L1L2 regularisation with $\alpha = 5 \times 10^{-4}$, $\lambda = 0.02$ applied to the hidden layer.

In the 2016 and 2017 μμ channels two hidden layers of 32 and 8 neurons were used, with *rectified linear unit* (ReLU) activations [212]. The ReLU is defined as

$$\mathrm{ReLU}(x) = \max(x, 0). \tag{9.4}$$

In 2016, $p_{\mathrm{d}} = 20\%$ dropout was applied to the first hidden layer and L1L2 regularisation with $\alpha = 1 \times 10^{-4}$ and $\lambda = 5 \times 10^{-4}$ to both hidden layers. In 2017 dropout of $p_{\mathrm{d}} = 20\%$ and L1L2 regularisation with $\alpha = 1 \times 10^{-4}$ and $\lambda = 5 \times 10^{-4}$ were applied to the first hidden layer only.

For all MLPs, a batch size of 4096 was used and the network was optimised using Adam with a learning rate of 1×10^{-3}. Biases were initialised to unity and Glorot initialisation was used to initialise the weights. The output layer had a sigmoidal activation, restricting the response to $(0, 1)$ and thereby allowing it to be interpreted as a probability.

9.3.4 Test, Train, and Validation Sets

As with the BDT, 20% of the available input was excluded from the training, forming X_{test}. Of the remaining input, one quarter was used as a validation sample, X_{val}. This is not used when calculating the gradient during training, but is used to determine when to trigger *early stopping*: in all networks trained, training was stopped if there was no reduction in the objective function as calculated on X_{val} in 15 successive epochs (or after 10 000 epochs, a deliberately high limit chosen to force a reliance on early stopping behaviour).

9.3.5 Performance

The MLP training lasted for 175 epochs in the 2016 ee channel, 75 epochs in the 2016 μμ channel, 232 epochs in the 2017 ee channel, and 55 epochs in the 2017 μμ

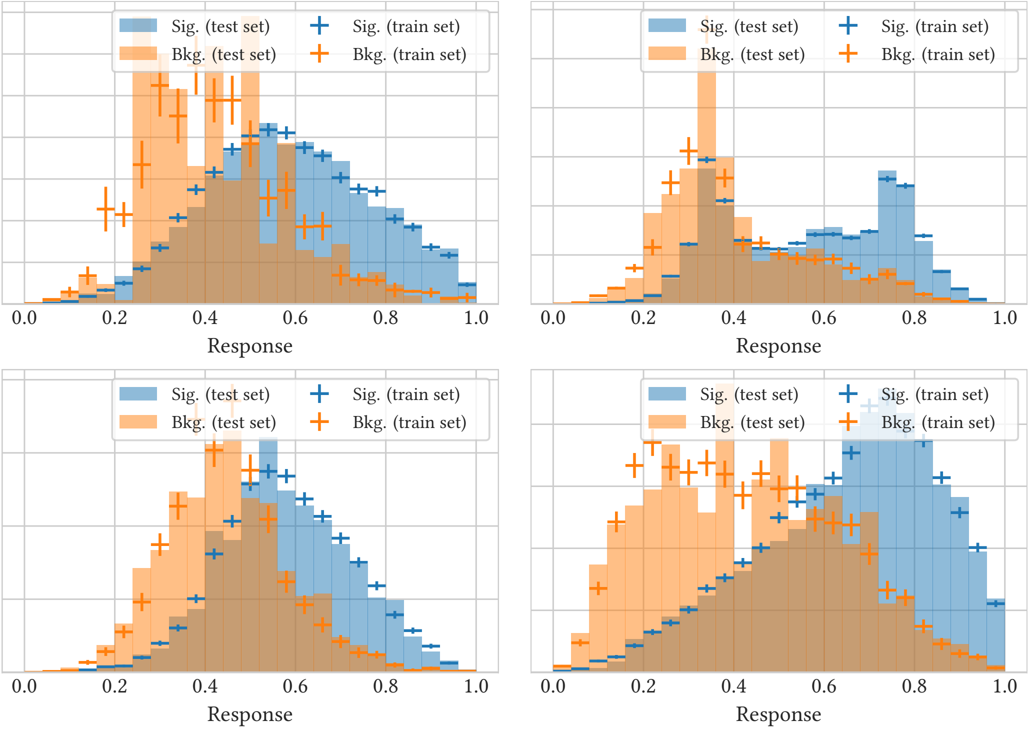


Figure 9.17: The distribution of the response of the trained MLP on the test and training sets in 2016 (top) and 2017 (bottom) in the ee (left) and μμ (right) channels.

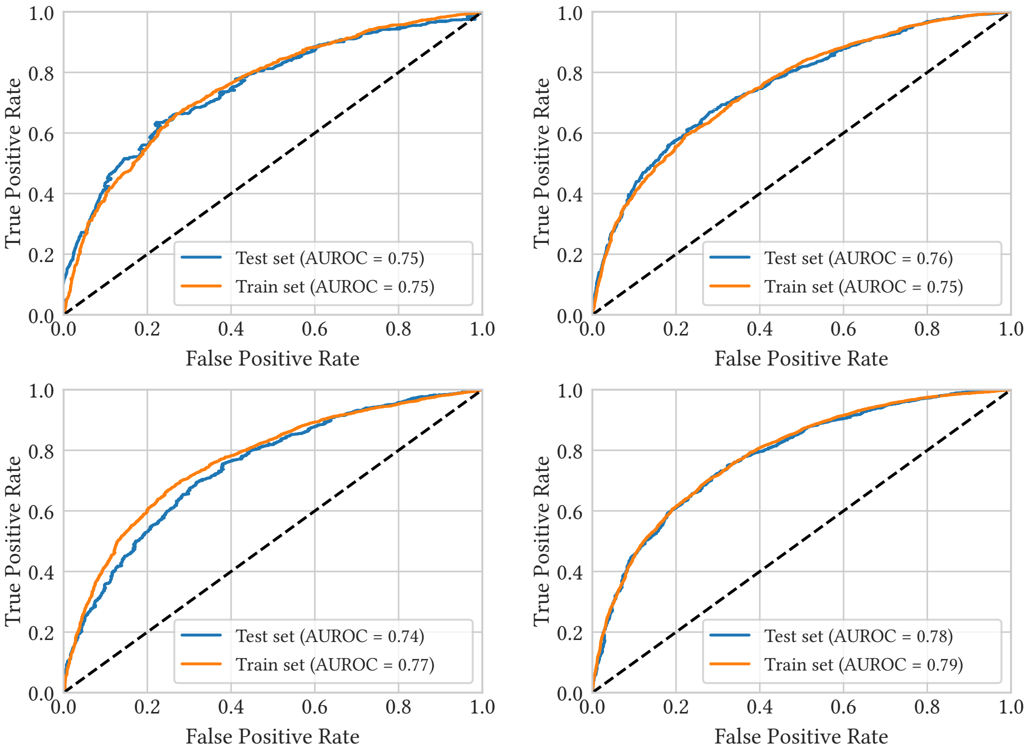


Figure 9.18: The ROC curve of the trained MLP on the test and training sets in 2016 (top) and 2017 (bottom) in the ee (left) and μμ (right) channels.

channel. The distributions of the MLP response are given in Figure 9.17 and the corresponding ROC curves in Figure 9.18.

Agreement between the test and training sets is generally excellent, except for the ee channel in 2017 where the corresponding ROC curves show some inconsistency. That being said, the KS test p-value in this region is still above 5%, as can be seen in Table 9.6.

Compared to the BDT, the MLP is not as effective at distinguishing tZq events from background. Comparing Figures 9.16 and 9.18 shows that in all regions the BDTs have a greater AUROCs than the corresponding MLPs. For this reason, the BDTs response distribution was chosen for use in the final step: signal extraction via a maximum likelihood fit.

9.4 Signal Extraction via a Maximum Likelihood Fit

A binned maximum likelihood fit performed on the classifier response was used to extract the signal strength and significance. Within each year, the fit was performed simultaneously on the ee and μμ channel. The fit was performed with the Higgs Analysis Combined Limit Tool (COMBINE), based on the RooStats project [213].

9.4.1 Likelihood Model

In a binned distribution of the BDT discriminator, the expected yield in a given bin, λ_i, can be described by

$$\lambda_i = rs_i + \sum_{j}^{n_{\text{bkg}}} b_{ij} \tag{9.5}$$

where s_i is the expected number of signal events in bin i, b_{ij} is the expected number of background events in background j in bin i, and n_{bkg} is the number of backgrounds. The *signal strength*,

$$r = \frac{\sigma_{\text{obs}}}{\sigma_{\text{s}}}, \tag{9.6}$$

is ratio of the observed cross section σ_{obs} and expected cross section σ_{s}. Uncertainties, both statistical and systematic, can be introduced as a set of nuisance parameters, $\boldsymbol{\theta}$, which modify the signal and background contributions of a bin:

$$s_i \rightarrow s_i(\boldsymbol{\theta}) \tag{9.7}$$

$$b_i \rightarrow b_i(\boldsymbol{\theta}). \tag{9.8}$$

According to Poisson statistics, the probability of observing n_i events in a given bin is

$$P(n_i) = \frac{\lambda_i^{n_i}}{n_i!} \mathrm{e}^{-\lambda_i} \tag{9.9}$$

$$= \frac{\left(rs_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})\right)^{n_i}}{n_i!} \mathrm{e}^{-rs_i(\boldsymbol{\theta})-b_i(\boldsymbol{\theta})} \tag{9.10}$$

and so by calculating the product of this across all bins the probability of observing a given dataset (i.e. set of bins $\boldsymbol{n}$),

$$\mathfrak{P}(\boldsymbol{n}|r,\boldsymbol{\theta}) = \prod_{i=1}^{N} P(n_i), \tag{9.11}$$

is obtained.

As it is assumed that all sources of systematic uncertainty are 100% uncorrelated (see § 8.4), they can be described by a probability density function $\rho(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}})$ where $\tilde{\boldsymbol{\theta}}$ is the nominal value of $\boldsymbol{\theta}$. Rate-based systematic uncertainties are modelled with a log-normal distribution [214] whilst shape-based systematic uncertainties are modelled by morphing the nominal shape by $\pm 1\sigma$ [215]. Statistical uncertainties were modelled using the Barlow–Beeston-lite approach, creating one nuisance parameter per bin rather than one per bin per sample [215, 216]. This reduces the fitting time and increases the fit stability.

The likelihood for the entire dataset, $\mathfrak{L}$, is the probability of observing a given dataset multiplied by the probability density of the nuisance parameters:

$$\mathfrak{L}(\boldsymbol{n}|r,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}) = \mathfrak{P}(r,\boldsymbol{\theta}) \cdot \rho(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}). \tag{9.12}$$

The significance can be determined by creating a test statistic, q_r, of the form [217]

$$q_r = \begin{cases} -2\ln\left(\frac{\mathfrak{L}(\text{data}|0,\hat{\boldsymbol{\theta}}_0)}{\mathfrak{L}(\text{data}|\hat{r},\hat{\boldsymbol{\theta}})}\right) & \hat{r} \geq 0 \\ 0 & \hat{r} < 0 \end{cases} \tag{9.13}$$

with $r = \hat{r}$ and $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ corresponding to the global maximum of $\mathfrak{L}(\text{data}|r,\hat{\boldsymbol{\theta}}_r)$, and $\hat{\boldsymbol{\theta}}_0$ the value of $\boldsymbol{\theta}$ that gives the greatest likelihood for $r = 0$. The distribution of the test statistic can be used to find the p-value for the background-only hypothesis by

$$p_0 = \int_{q_{0\,\text{obs}}}^{\infty} P(q_0|0)\,\mathrm{d}q_0 \tag{9.14}$$

in which $P(q_0|0)$ is the probability density function of q_0 and $q_{0\,\text{obs}}$ is the value of q_0 observed in data. This p-value can then be converted into a Z-score (significance) with the relationship

$$Z = \Phi^{-1}(1-p) \tag{9.15}$$

where Φ^{-1} is the quantile function of the normal distribution.

In order to calculate an expected i.e. blinded significance, $q_{0\,\mathrm{obs}}$ is determined from *pseudodata.* A single, representative pseudodata sample known as the *Asimov dataset* [217] is constructed in which all observable quantities and nuisances are set to their expectation values.

The observed signal strength is determined by performing a maximum likelihood fit, i.e. finding $\hat{r}$ and $\hat{\boldsymbol{\theta}}$. This is performed with COMBINE using the MINUIT2 minimisation library [218].

9.4.2 Binning Strategy

As COMBINE performs binned fits, a binning strategy must be chosen for the BDT response. In private communication with the COMBINE contact of the CMS top particle analysis group it was recommended that each bin should have:

- less than 30% statistical uncertainty in s_i and b_i
- $b_i \geq 1$.

To ensure this, a recursive binning strategy was employed. Initially, the BDT response was placed into a single bin. This was then partitioned at the unweighed median (i.e. the median after setting all event weights to unity) of the bin contents. The two resultant “daughter” bins were partitioned in the same manner, until doing so would create a bin violating one of the above limits. The resulting binned distributions of the response are given in Figures 9.19 and 9.20.

9.4.3 Fit Results

The expected significances and signal strengths calculated by combine are presented in Table 9.7. Diagrams showing the impact of each source of systematic uncertainty are shown in Figures 9.21 to 9.26.

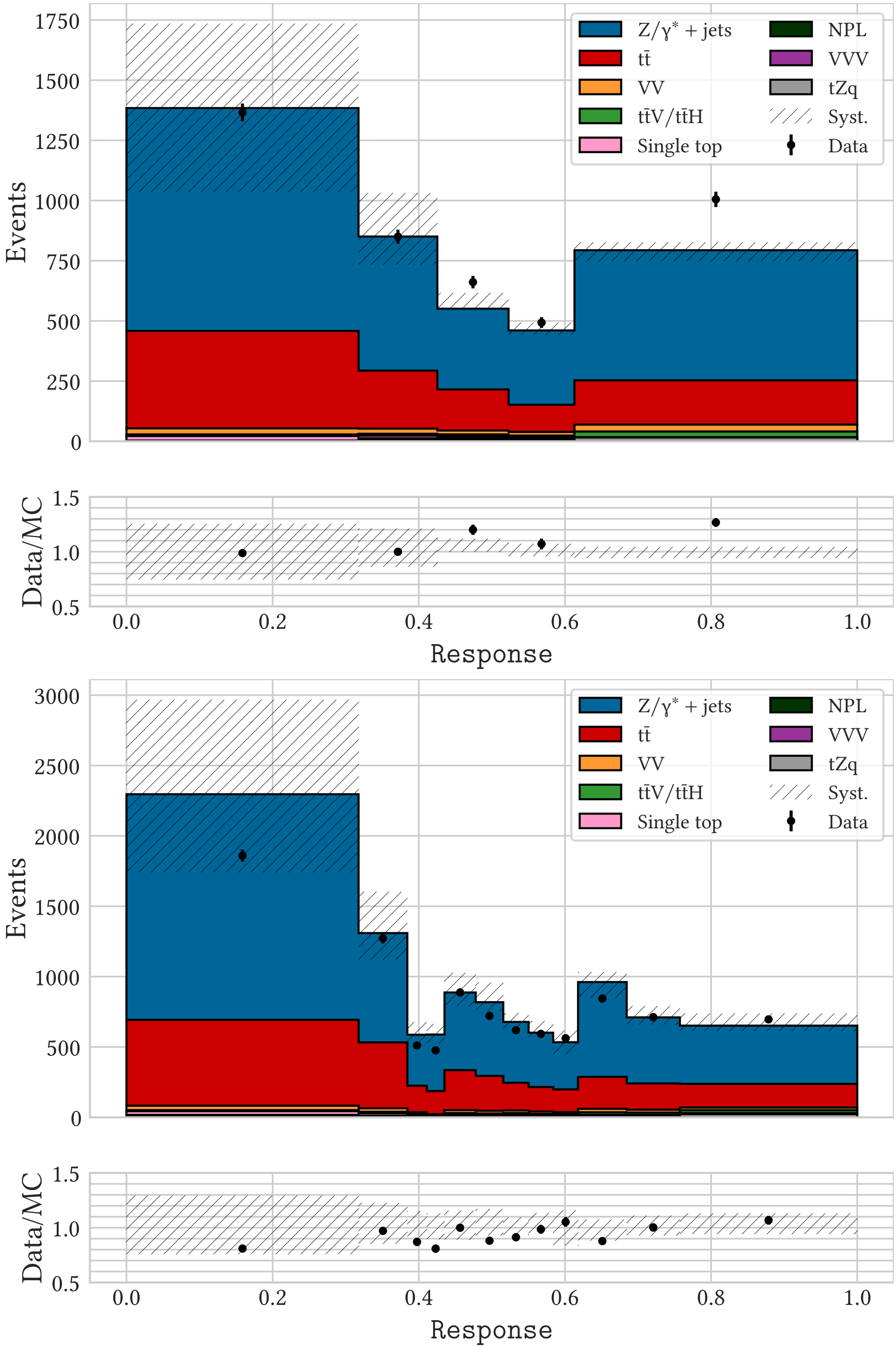


Figure 9.19: The binned BDT response in the 2016 ee (top) and μμ (bottom) signal regions as used by COMBINE to extract the signal strength and significance.

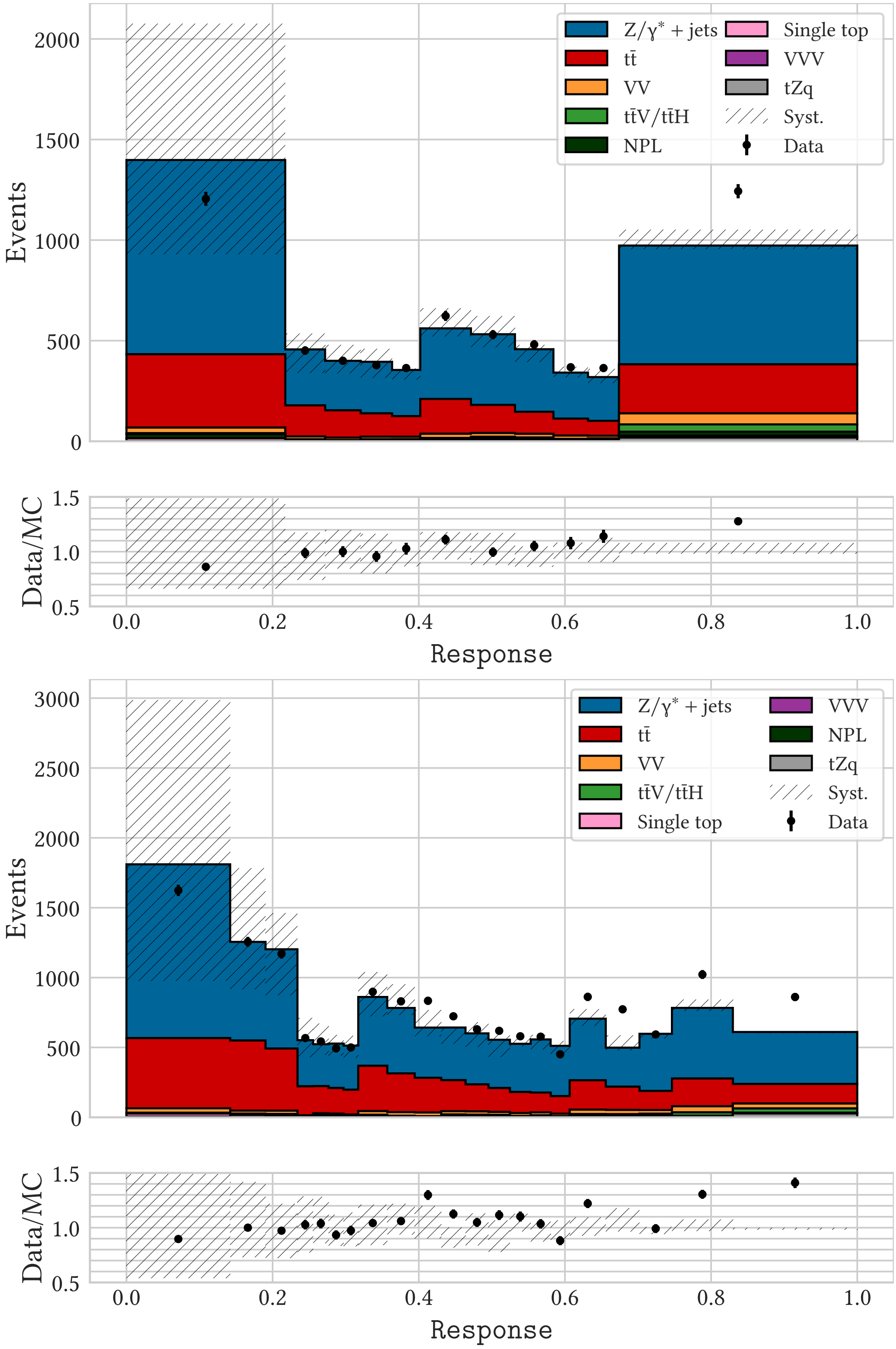


Figure 9.20: The binned BDT response in the 2017 ee (top) and μμ (bottom) signal regions as used by COMBINE to extract the signal strength and significance.

Table 9.7: The signal strength, expected significance an observed significance calculated by COMBINE. Expected significance is calculated using an Asimov dataset with $\hat{r} = 1$. In each year, results for fits in the ee and μμ channel are given individually, and from a combined simultaneous fit in both channels. The final row shows the combined result across both years and channels.

Era	Channel	Expected σ	Observed σ	$\hat{r}$
2016	ee	0.03	1.50	$9.96^{+8.03}_{-6.51}$
	μμ	0.27	1.22	$4.34^{+3.47}_{-3.07}$
	ee + μμ	0.29	1.67	$5.06^{+3.52}_{-3.00}$
2017	ee	0.16	1.98	$11.95^{+7.16}_{-5.69}$
	μμ	0.34	2.50	$7.37^{+3.77}_{-2.90}$
	ee + μμ	0.38	2.97	$8.19^{+3.80}_{-2.80}$
2016 + 2017	ee + μμ	0.48	3.12	$6.52^{+2.30}_{-2.05}$

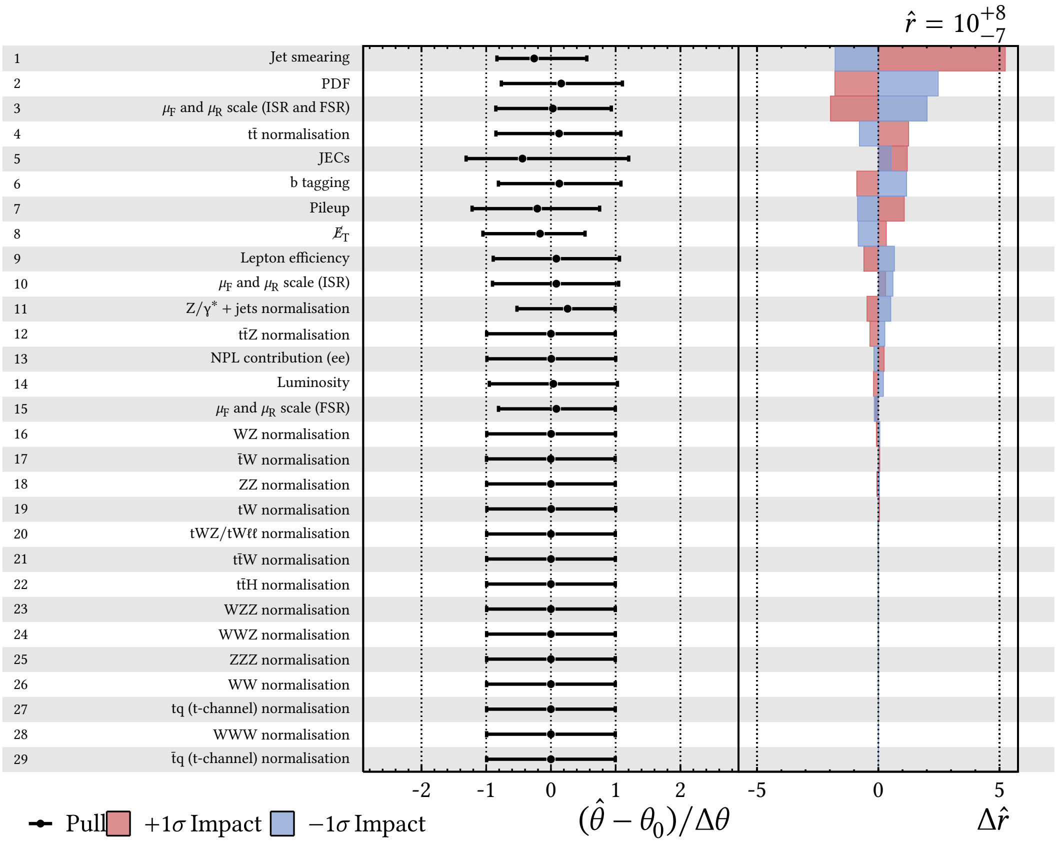


Figure 9.21: The best fit values and impacts on the signal strength for each systematic uncertainty in the 2016 ee channel fit. The pre fit uncertainty of a given nuisance parameter is $\Delta\theta$.

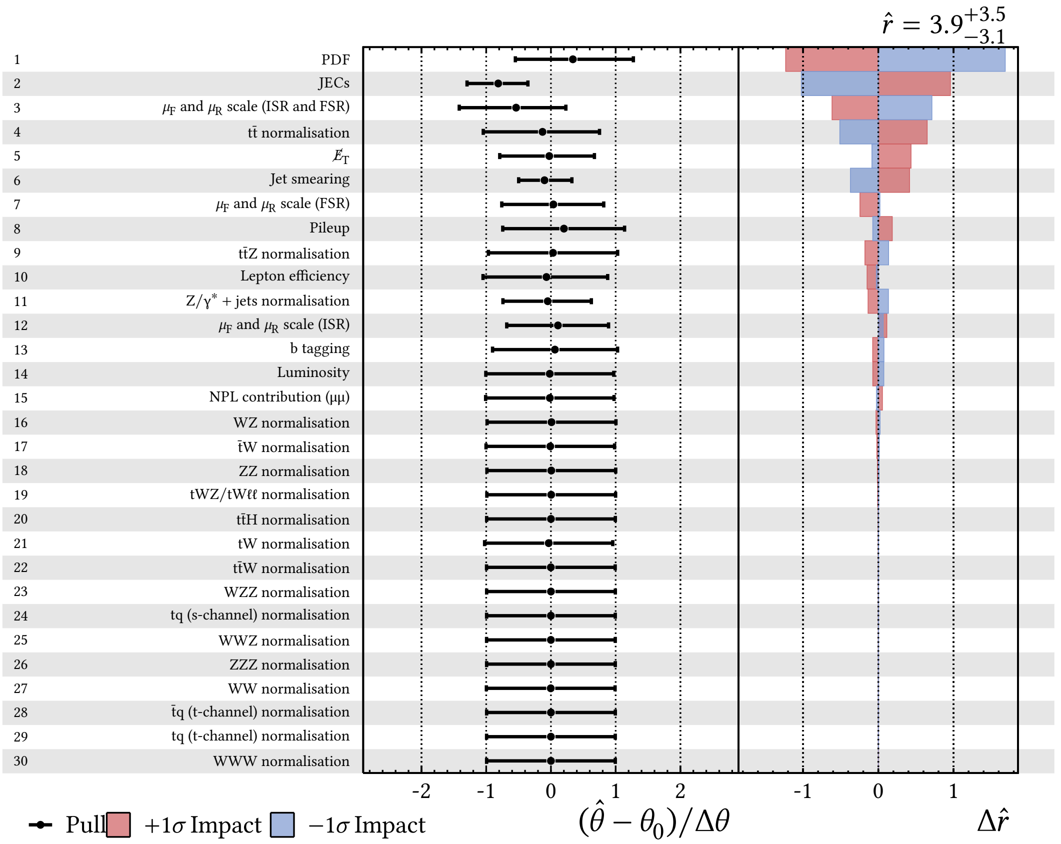


Figure 9.22: The best fit values and impacts on the signal strength for each systematic uncertainty in the 2016 μμ channel fit. The pre fit uncertainty of a given nuisance parameter is $\Delta\theta$.

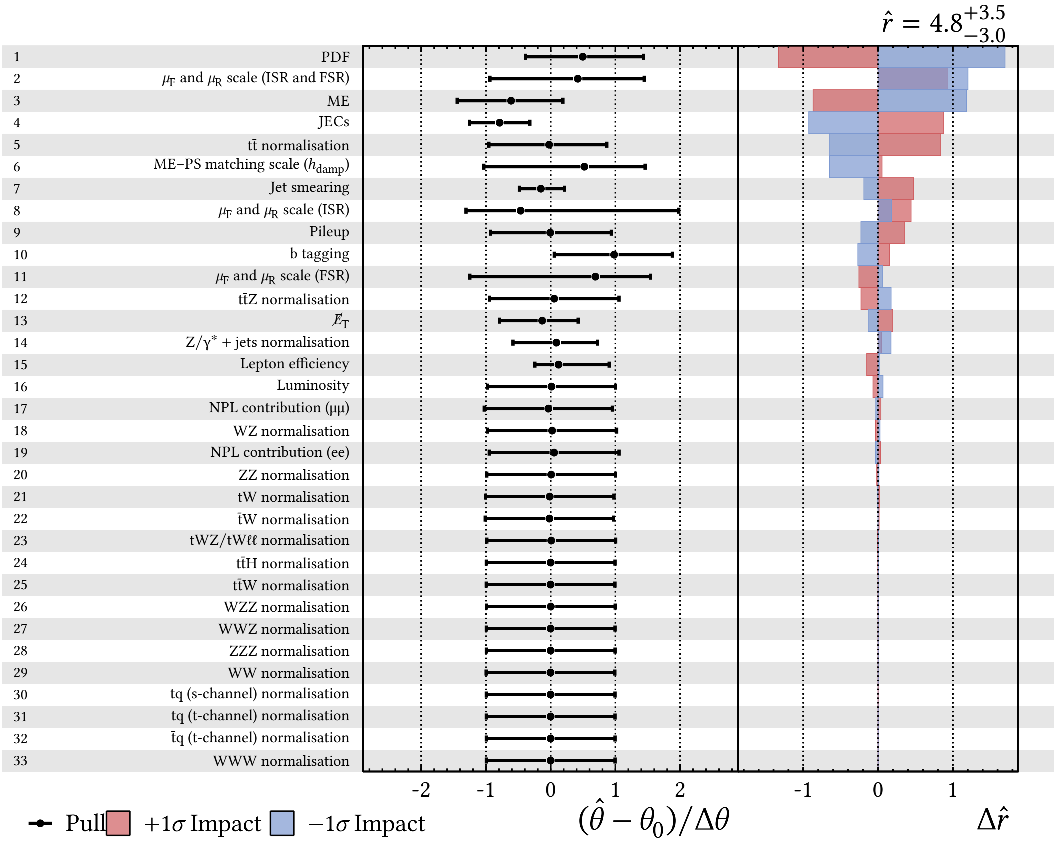


Figure 9.23: The best fit values and impacts on the signal strength for each systematic uncertainty in the 2016 ee + μμ channel fit. The pre fit uncertainty of a given nuisance parameter is $\Delta\theta$.

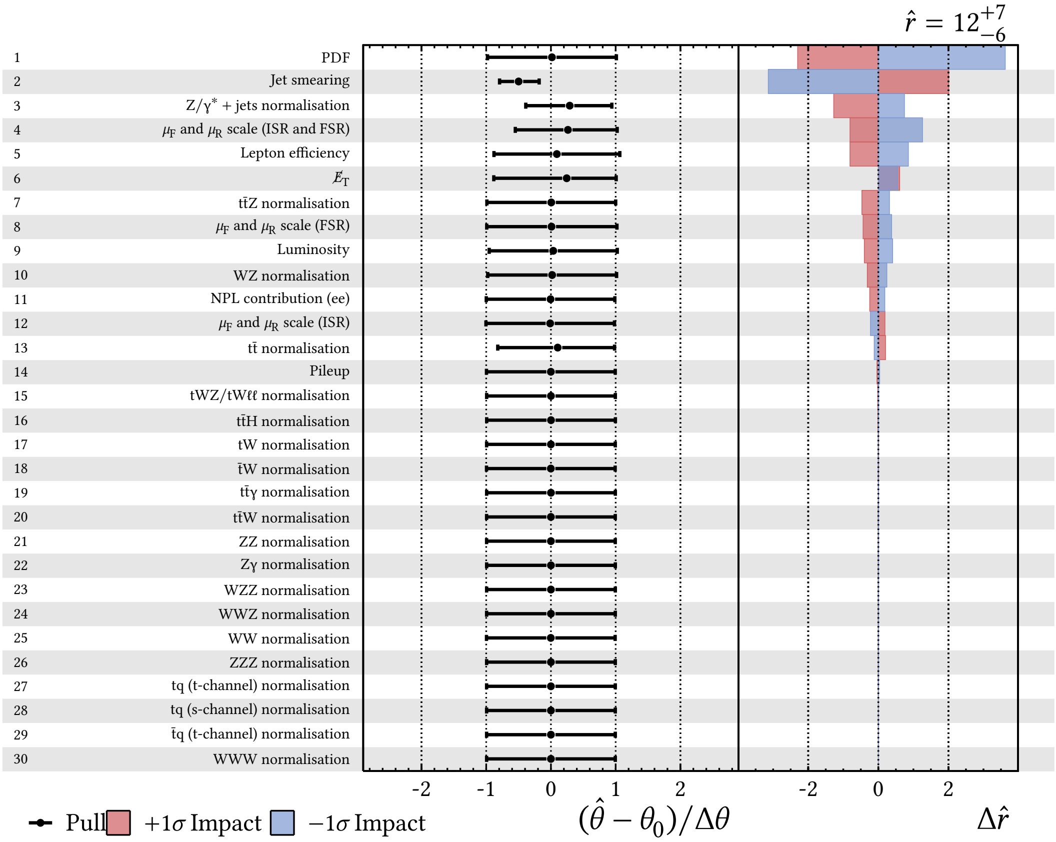


Figure 9.24: The best fit values and impacts on the signal strength for each systematic uncertainty in the 2017 ee channel fit. The pre fit uncertainty of a given nuisance parameter is $\Delta\theta$.

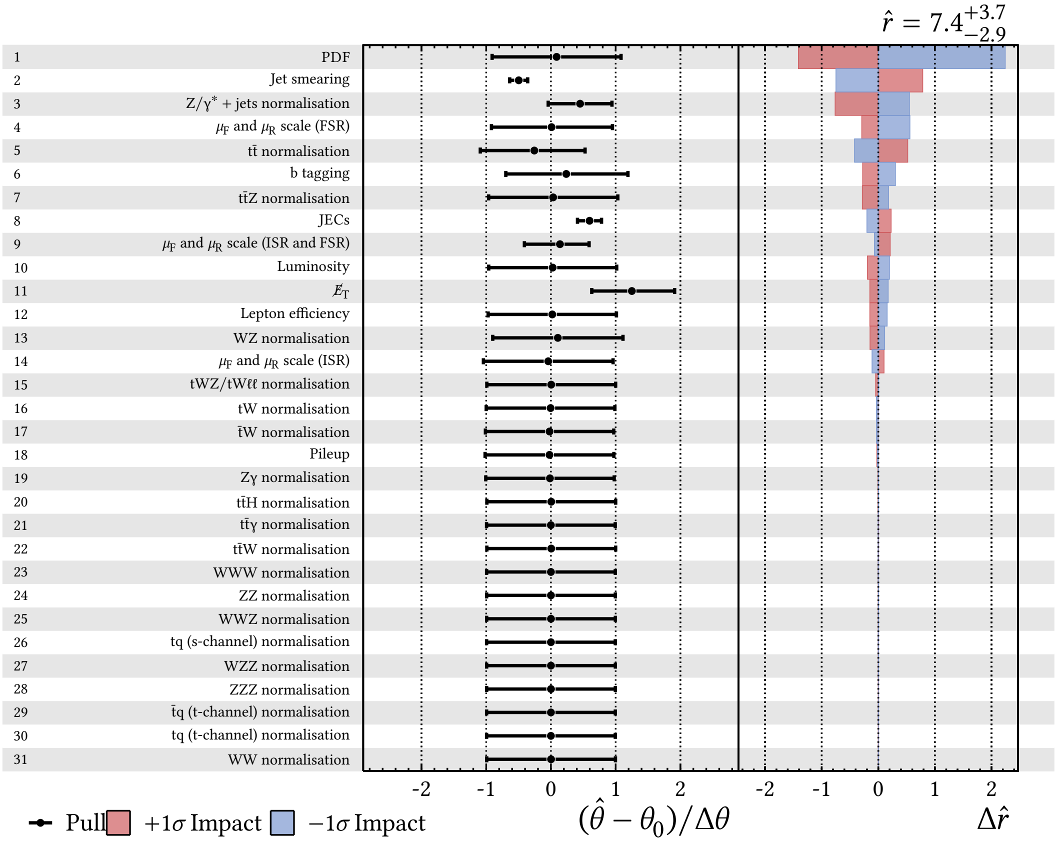


Figure 9.25: The best fit values and impacts on the signal strength for each systematic uncertainty in the 2017 μμ channel fit. The pre fit uncertainty of a given nuisance parameter is $\Delta\theta$.

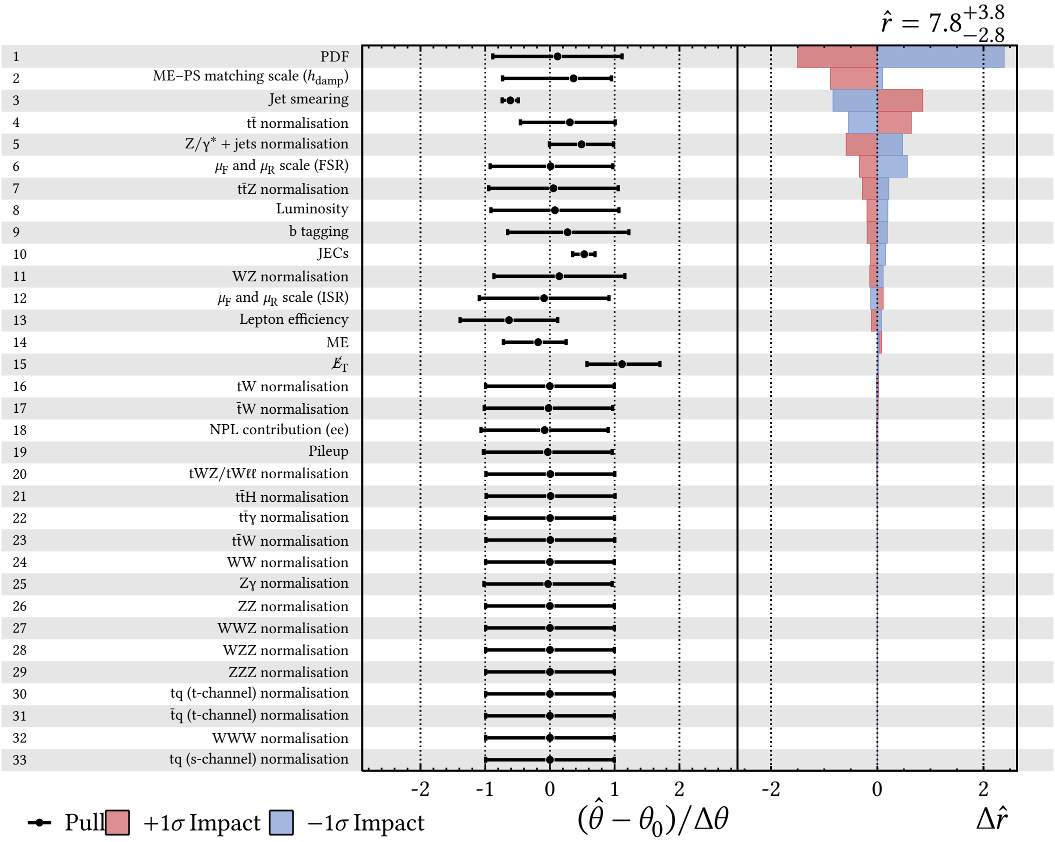


Figure 9.26: The best fit values and impacts on the signal strength for each systematic uncertainty in the 2017 ee + μμ channel fit. The pre fit uncertainty of a given nuisance parameter is $\Delta\theta$.

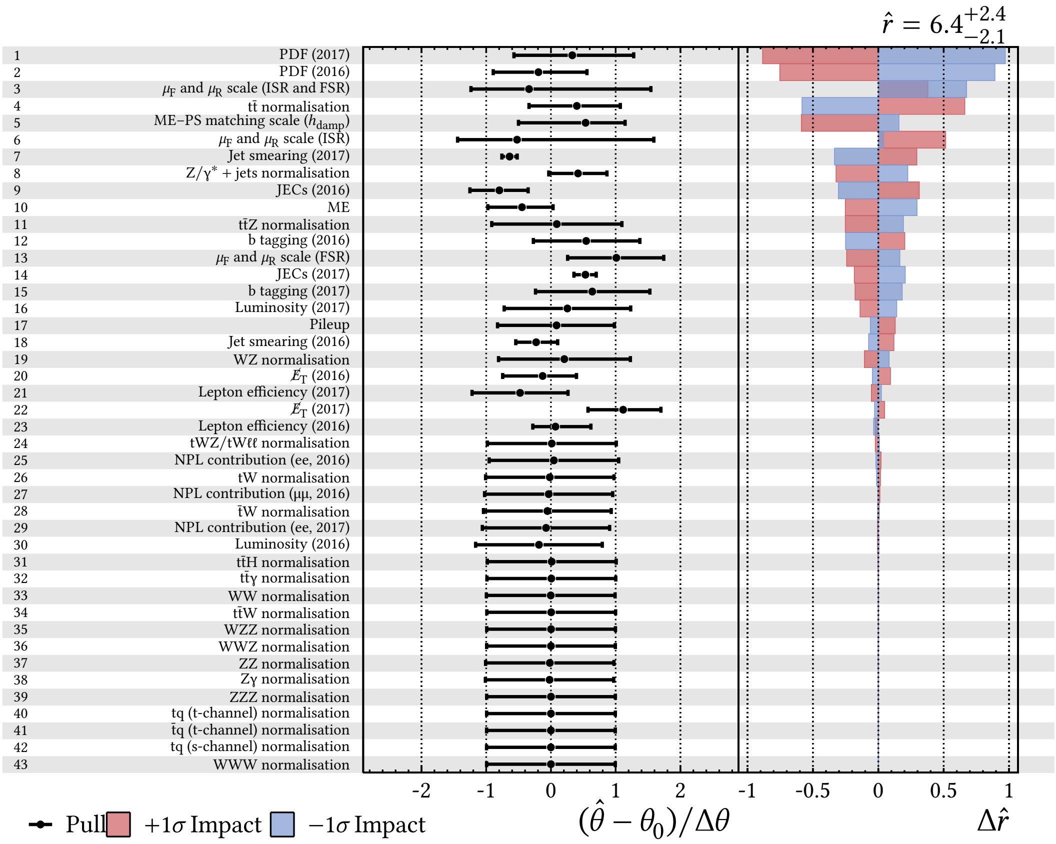


Figure 9.27: The best fit values and impacts on the signal strength for each systematic uncertainty in the combined 2016 and 2017 fit in the ee and μμ channel. The pre fit uncertainty of a given nuisance parameter is $\Delta\theta$. Systematic uncertainties from the same source that were not treated as correlated between 2016 and 2017 data are listed separately for each year.

Chapter 10

Discussion and Conclusion

> *A witty saying proves nothing*
>
> VOLTAIRE

10.1 Other Searches for tZq

No search for tZq in the dilepton final state has been published by CMS or any other physics experiment. Four previous searches for tZq in the trileptonic final state have been published, all from experiments situated on the LHC. Three of these searches were published by the CMS Collaboration. The first of these was conducted on $19\,\mathrm{fb}^{-1}$ of $\sqrt{s} = 8\,\mathrm{TeV}$ collision data from the 2012 dataset, a signal strength of $1.22^{+0.98}_{-0.85}$ was measured with a corresponding observed (expected) significance of $2.4\sigma\,(1.8\sigma)$ [219]. The second search at CMS used $36.9\,\mathrm{fb}^{-1}$ of $\sqrt{s} = 13\,\mathrm{TeV}$ collision data from the 2016 data-taking period and observed a signal strength of $1.31^{+0.35}_{-0.33}(\mathrm{stat})^{+0.31}_{-0.25}(\mathrm{syst})$ with an observed (expected) significance of $3.7\sigma\,(3.1\sigma)$ [196]. The third published search for tZq at CMS used $77.4\,\mathrm{fb}^{-1}$ of $\sqrt{s} = 13\,\mathrm{TeV}$ collision data from the 2016–2017 data-taking period; it measured a signal strength of $1.18^{+0.14}_{-0.13}(\mathrm{stat})^{+0.11}_{-0.10}(\mathrm{syst})^{+0.04}_{-0.04}(\mathrm{theo})$ with an observed (expected) significance of $8.2\sigma\,(7.7\sigma)$ [220]. The final published tZq analysis comes from the ATLAS collaboration, performed on $36.1\,\mathrm{fb}^{-1}$ of $\sqrt{s} = 13\,\mathrm{TeV}$ collision data from 2015–2016. A signal strength of 0.75 ± 0.28 was observed, with corresponding observed (expected) significance of $4.2\sigma\,(5.4\sigma)$ [221].

All measurements of the tZq cross section in the trilepton final state are compatible with the SM prediction. The observed and expected cross sections reported by the trileptonic tZq analyses are greater than the expected significance in the dileptonic tZq analysis presented in this thesis, despite the dileptonic decay channel having a greater predicted cross section. This can be attributed the profile of the backgrounds in the two channels. In the trilepton channel, the greatest contributors to the background are the WZ and $\mathrm{t\bar{t}Z}$ processes. The contribution of the two largest backgrounds to the dileptonic search, the Z/γ^{*} + jets and $\mathrm{t\bar{t}}$ processes, is greatly reduced due to the three lepton requirement: an additional NPL is required in these processes in order to have the required number of leptons in the final state. The WZ and $\mathrm{t\bar{t}Z}$ processes have smaller cross sections relative to tZq than Z/γ^{*} + jets and $\mathrm{t\bar{t}}$, allowing tZq events to be more easily isolated. This allows for the improved performance over the dilepton analysis.

10.2 Summary of the tZq Analysis

A search was performed for the production of a Z boson in association with a single top quark. Following from previous observations in the trilepton final state, the analysis presented in this thesis attempted to observe the tZq process in the dilepton final state using $78\,\mathrm{fb}^{-1}$ of data obtained by the CMS experiment in the 2016–2017 data-taking period.

A selection procedure was devised in order to isolate tZq events, selecting events with exactly two opposite-sign same-flavour leptons compatible with the nominal Z boson mass. In addition, these events required 4–6 jets, 1–2 of which must be b tagged, and a pair of which must be compatible with the nominal W boson mass (with a veto on the leading b tagged jet). CRs enriched in Z/γ^{*} + jets and $\mathrm{t\bar{t}}$ production were created. Agreement between data and simulated samples in these regions lay within the combined statistical and systematic uncertainties.

The purity of the signal region is low, and so ML techniques were used to identify tZq events. GPs were used to automate the hyperparameter selection procedure for BDTs; hyperparameter choices that resulted in well-performing classifiers were found within 100 iterations. The performance of these automatically optimised BDTs compared favourably to MLPs optimised by hand.

The expected signal strength across the full dataset is $\hat{r} = 6.52^{+2.30}_{-2.05}$ with an observed (expected) significance of 3.12σ (0.48σ). This may indicate excess of tZq process in the dilepton channel over the SM prediction, however $\hat{r} = 1$ still lies within a 3σ envelope. The observed significance far exceeds the expected significance because of the large value of $\hat{r}$: the expected significance was determined assuming $\hat{r} = 1$.

10.3 Future Work

As mentioned in § 9.3.3, one improvement that can be made to the analysis is the application of the automatic hyperparameter selection to the MLP classifiers. A possible workaround to the known challenges would to be exclude qualitative hyperparameters and the number of hidden layers from the optimisation procedure. Multiple optimisations could then be performed with the excluded hyperparameters at fixed values. Alternatively, an emerging method of hyperparameter optimisation is to use genetic programming [222]. In this method, multiple different classifiers are evaluated. The best-performing are then mutated and bred (i.e. properties from multiple classifiers are combined) to form a new generation, where the process repeats. The advantage of such a method is not only the ability to easily include qualitative hyperparameters, but also allowing both feature preprocessing and feature selection to be incorporated into the automatic optimisation procedure.

The statistical uncertainty in the tZq analysis can, of course, be reduced by including more data. Work is currently under way to encompass the full Run II dataset (i.e. include data from the 2018 data-taking period). Ideally, including 2018 data will resolve the question of whether an excess in the tZq process exists by reducing or increasing the tension between the observed signal strength and SM prediction.

One change that must be made to the analysis as part of utilising the full Run II dataset is a change of b tagging algorithm. CSVv2 has been retired for use in data taken in 2018 and beyond, replaced by the *Deep Combined Secondary Vertex* (DeepCSV) b tagging algorithm [167, 224]. DeepCSV uses deep neural networks to achieve an improved classification performance; the ROC curves of CSVv2 and DeepCSV are compared in Figure 10.1. Replacing CSVv2 with DeepCSV should help further distinguish tZq events backgrounds without a real b jet across the whole Run II dataset, as it can be applied retroactively.

An area that could be improved with respect to reducing the dominant Z/γ^* + jets background is the W boson reconstruction. As can be seen in Figures 7.3 and 7.4, the W mass cut retains much of the Z/γ^* + jets background, despite it not containing a real W boson. A possible improvement would be to closely investigate the properties of jets that originate from W boson decay and determine if they can be distinguished from jets that do not. With this information, events that do not contain two jets identified as potentially originating from a W boson decay within a W boson mass window could be rejected. Properties of jet pairs outside of their invariant mass could also be considered. If W boson reconstruction can be improved, a widening of the W boson mass cut window could be considered, as

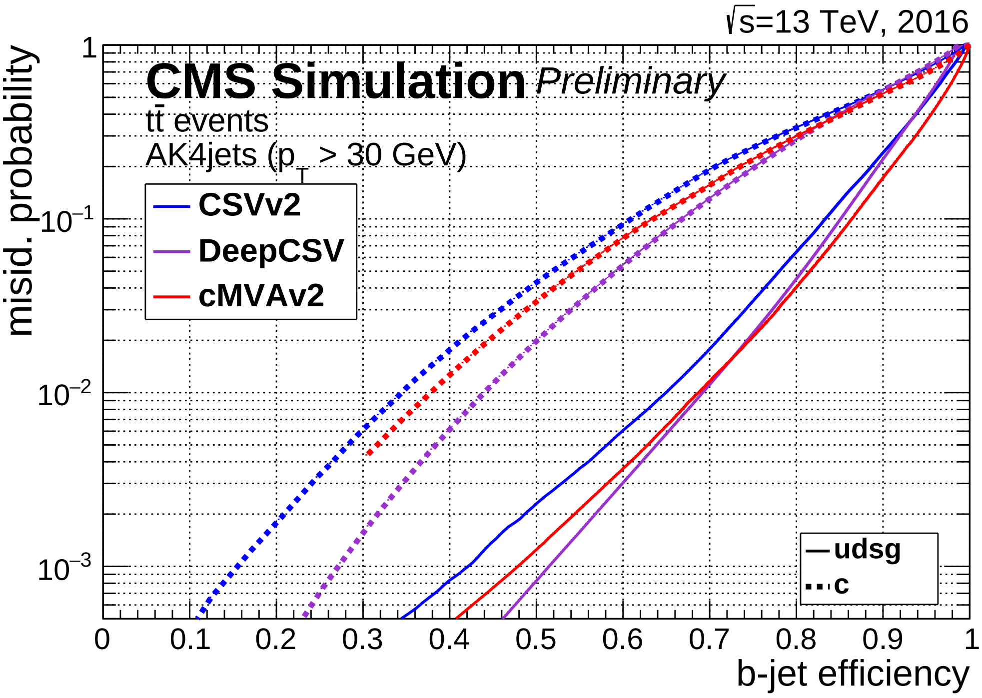


Figure 10.1: ROC curve comparing the performance of the DeepCSV b tagging algorithm with the older CSVv2 and cMVAv2 b tagging algorithms. Performance was measured on AK4 jets of $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ in 2016 $\mathrm{t\bar{t}}$ events and is shown separately for charm and light (udsg) jets. Figure taken from [223].

Figures 7.3 and 7.4 suggest that tZq events could be recovered by doing so (currently, due to the Z/γ^*+jets background, no advantage was found in widening the W mass window). This must be done with care to avoid starving the Z/γ^* + jets CR, or a new Z/γ^* + jets CR may need to be defined.

The latest tZq trilepton analysis [220] made use of ML classification to identify NPLs. NPLs are a much larger contributor to the background in the trilepton analysis but nevertheless implementing this in the tZq analysis should improve the purity of the signal region, albeit only slightly.

The tZq analysis suffers from large systematic uncertainties, and efforts made to constrain these may prove fruitful. A simultaneous fit in the signal and CRs is a possible way to achieve this—it would help to constrain the uncertainty associated with the Z/γ^* + jets and $\mathrm{t\bar{t}}$ normalisation. It could also be possible to exclude regions in which systematic uncertainties are particularly large.

Acknowledgements

Despite the near-certainty of missing somebody out, it would be remiss of me not to thank those who made this thesis possible. To those who I forget to mention here I can only offer my sincerest apologies and a promise that it reflects more poorly on myself than on them.

I must first thank my supervisor, Joanne Cole, who has deftly guided me these last four years. I am certain without her guidance, support, and fastidious feedback this thesis would be a shadow of itself, if it even existed at all. At Brunel, I have been lucky enough to work with a superlative team. The previous PhD cohort, Alexander Morton and Duncan Legatt, have been immeasurably patient and charitable with their knowledge and time; the latter also deserves special thanks for making me feel more at home during my time at CERN. I am indebted to Ivan Reid for his vigilant management of the computing facilities at Brunel, I can only apologise for the occasional technical hiccup created by some of my more ill-conceived code and my substandard attendance at the weekly group meeting. My development advisor, Paul Kyberd, is deserving of thanks for his counsel and patience regarding my… we will say *somewhat overdue* progress meetings. I am also grateful to David Smith for stepping as my secondary advisor in these last few months after Peter Hobson left us for Queen Mary University. Peter needs be thanked not only as my secondary supervisor for the bulk of my study, but also for his work as group leader during that time—I wish him the best of luck at his new position. Good luck and thanks goes to Kathryn Coldham, Shirin Golpayegani, and Maziar Ghorbani who continue to work on their PhDs, it was a pleasure to work with each of you. Last and—as cliché as it might be—certainly not least, I must thank Catherine Mackay for her tireless support, positive attitude, and generosity with her time despite an interminably busy schedule.

I also need to thank my friends and family for their support and kindness during my studies. It has been invaluable. A special mention to the new friends I made at CERN, many of whom were part of the UK CERN long-term attachment group.

Much of the analysis in this thesis was performed and developed using the Root framework [225], with Root and Matplotlib [226] used to generate the plots included in this thesis. The ML aspects of the research were made possible thanks to the Scikit-learn API [227, 228] and a smörgåsbord of high-quality data analysis libraries written for the Python programming language [229, 230] including NumPy [231, 232], SciPy [233], and Pandas [234]. The interoperation between these libraries and Root was achieved thanks to the excellent tools developed by the Scikit-HEP project [235]. Feynman diagrams were created using the `feynmp` package [236]. Other graphics were drawn with PGF/Ti*k*z [237], with Figure 5.1 created using pgfplots [238].

Finally, I would like to thank the Science and Technologies Facilities Council whose funding made my work possible.

Appendix A

Notes on Notation

In general, the notation used in this thesis follows the ISO 80000–2:2009 standard [239]. Additional notation used is defined in the sections below.

A.1 Index Notation

In this thesis the notation of the Ricci calculus [240] is used. A covariant four-vector is represented as a majuscule Latin letter with a minuscule Greek lower index: A_μ. The Greek letter represents the four indexes of spacetime, $\{0, 1, 2, 3\}$, where 0 is the time component and $1, 2, 3$ the spatial components. A contravariant vector is represented by a raised index, e.g. A^μ, and tensors by multiple indices, e.g. $F^{\mu\nu}$.

Indices can be raised or lowered by the Minkowski metric [23]:

$$\begin{aligned} A^{\xi}{}_{\nu\ldots} &= \eta^{\xi\mu} A_{\mu\nu\ldots} \\ A_{\mu\nu\ldots} &= \eta_{\mu\xi} A^{\xi}{}_{\nu\ldots} \end{aligned} \qquad \eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{A.1}$$

The same symbol appearing as an upper index and lower index in a single term implies a summation

$$A_\mu B^\mu = \sum_\mu A_\mu B^\mu \qquad A^\mu B_\mu = \sum_\mu A^\mu B_\mu \tag{A.2}$$

and other combinations of repeated indices are ill-formed. The four-gradient, ∂^μ, is defined [22]

$$\partial^\mu = (\partial_t, -\nabla) = \left(\frac{\partial}{\partial t}, -\frac{\partial}{\partial x}, -\frac{\partial}{\partial y}, -\frac{\partial}{\partial z}\right). \tag{A.3}$$

A.2 Dirac Matrices

The Dirac matrices, γ^{μ}, are defined as [23]

$$\gamma_0 = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix} \qquad \gamma_i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix} \tag{A.4}$$

where σ^i are the Pauli matrices [18]

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{A.5}$$

and I_2 is the 2×2 identity matrix. The fifth gamma matrix, γ_5, is [22]

$$\gamma_5 = \mathrm{i}\gamma_0\gamma_1\gamma_2\gamma_3 = \begin{pmatrix} 0 & I_2 \\ I_2 & 0 \end{pmatrix}. \tag{A.6}$$

Dirac matrices can be used to transform a four-vector into a linear operator on spinor fields, ψ. This is represented using the *Feynman slash* notation [241],

$$\not{\partial} = \gamma^{\mu}\partial_{\mu}. \tag{A.7}$$

The matrix γ^0 is also used to relate the barred an unbarred forms of spinor fields [23]:

$$\bar{\psi} = \psi^{\dagger}\gamma^{0} \tag{A.8}$$

where $\psi^{\dagger}$ is the Hermitian adjoint of ψ [22].

A.3 Probability

When discussing probability, random variables are set in Latin majuscule: X, Y, Z, etc. with specific realisations set in the corresponding minuscule: x, y, z, etc. This extends to vectors, that is, $\boldsymbol{x}$ represents a specific realisation of the random vector $\boldsymbol{X}$.

The distribution of a random variable, Z, can be defined by a continuous probability distribution of its realisations, $P(z)$. In the case of the Gaussian distribution,

$$P\left(z|\mu_z, \sigma_z\right) = \frac{1}{\sigma_z\sqrt{2\pi}} \exp\left(-\frac{\left(z-\mu_z\right)^2}{2\sigma_z^2}\right) \tag{A.9}$$

where μ_z is the mean of Z (i.e. $\mu_z = \mathrm{E}[Z]$), and σ_z is the standard deviation of Z. A shorthand for this is

$$Z \sim \mathcal{N}\left(\mu_z, \sigma_z\right) \tag{A.10}$$

where the tilde denotes ‘is distributed as’ and $\mathcal{N}\left(\mu_z, \sigma_z\right)$ represents a Gaussian distribution of mean μ_z and standard deviation σ_z.

Appendix B

Supplementary Material for Chapter 9

This appendix contains supplementary material referenced in Chapter 9.

Table B.1: Definition of all features available for use in ML classification.

Feature	Definition
`bEta`	η of leading b tagged jet
`bPhi`	φ of leading b tagged jet
`bPt`	p_{T} of leading b tagged jet
`bbTag`	b tag discriminator of leading b tagged jet
`chi2`	χ^2 as used to define side-band region
`j1Eta`	η of leading jet
`j1Phi`	φ of leading jet
`j1Pt`	p_{T} of leading jet
`j1bTag`	b tag discriminator of leading jet
`j1j2DelR`	ΔR between leading jet and second-leading jet
`j1j3DelR`	ΔR between leading jet and third-leading jet
`j1j4DelR`	ΔR between leading jet and fourth-leading jet
`j1l1DelR`	ΔR between leading jet and leading lepton
`j1l2DelR`	ΔR between leading jet and second-leading lepton
`j1tDelR`	ΔR between leading jet and reconstructed top quark
`j1wDelR`	ΔR between leading jet and reconstructed W boson
`j1wj1DelR`	ΔR between leading jet and leading W boson candidate jet
`j1wj2DelR`	ΔR between leading jet and second-leading W boson candidate jet
`j1zDelR`	ΔR between leading jet and reconstructed Z boson
`j2Eta`	η of second-leading jet
`j2Phi`	φ of second-leading jet
`j2Pt`	p_{T} of second-leading jet
`j2bTag`	b tag discriminator of second-leading jet
`j2j3DelR`	ΔR between second-leading jet and third-leading jet

`j2j4DelR`	ΔR between second-leading jet and fourth-leading jet
`j2l1DelR`	ΔR between second-leading jet and leading lepton
`j2l2DelR`	ΔR between second-leading jet and second-leading lepton
`j2tDelR`	ΔR between second-leading jet and reconstructed top quark
`j2wDelR`	ΔR between second-leading jet and reconstructed W boson
`j2wj1DelR`	ΔR between second-leading jet and leading W boson candidate jet
`j2wj2DelR`	ΔR between second-leading jet and second-leading W boson candidate jet
`j2zDelR`	ΔR between second-leading jet and reconstructed Z boson
`j3Eta`	η of third-leading jet
`j3Phi`	φ of third-leading jet
`j3Pt`	p_T of third-leading jet
`j3bTag`	b tag discriminator of third-leading jet
`j3j4DelR`	ΔR between third-leading jet and fourth-leading jet
`j3l1DelR`	ΔR between third-leading jet and leading lepton
`j3l2DelR`	ΔR between third-leading jet and second-leading lepton
`j3tDelR`	ΔR between third-leading jet and reconstructed top quark
`j3wDelR`	ΔR between third-leading jet and reconstructed W boson
`j3wj1DelR`	ΔR between third-leading jet and leading W boson candidate jet
`j3wj2DelR`	ΔR between third-leading jet and second-leading W boson candidate jet
`j3zDelR`	ΔR between third-leading jet and reconstructed Z boson
`j4Eta`	η of fourth-leading jet
`j4Phi`	φ of fourth-leading jet
`j4Pt`	p_T of fourth-leading jet
`j4bTag`	b tag discriminator of fourth-leading jet
`j4l1DelR`	ΔR between fourth-leading jet and leading lepton
`j4l2DelR`	ΔR between fourth-leading jet and second-leading lepton
`j4tDelR`	ΔR between fourth-leading jet and reconstructed top quark
`j4wDelR`	ΔR between fourth-leading jet and reconstructed W boson
`j4wj1DelR`	ΔR between fourth-leading jet and leading W boson candidate jet
`j4wj2DelR`	ΔR between fourth-leading jet and second-leading W boson candidate jet
`j4zDelR`	ΔR between fourth-leading jet and reconstructed Z boson
`jetMass`	Total mass of all jets
`jetMass3`	Total mass of three most leading jets
`jetMt`	Total transverse mass of all jets
`jetPt`	Total p_T of all jets
`l1Eta`	η of leading lepton
`l1Phi`	φ of leading lepton
`l1Pt`	p_T of leading lepton
`l1RelIso`	I^{rel} of leading lepton
`l1bDelR`	ΔR between leading lepton and leading b tagged jet
`l1tDelR`	ΔR between leading lepton and reconstructed top quark
`l1wj1DelR`	ΔR between leading lepton and leading W boson candidate jet
`l1wj2DelR`	ΔR between leading lepton and second-leading W boson candidate jet
`l2Eta`	η of second-leading lepton
`l2Phi`	φ of second-leading lepton
`l2Pt`	p_T of second-leading lepton

`l2RelIso`	I^{rel} of second-leading lepton
`l2bDelR`	ΔR between second-leading lepton and leading b tagged jet
`l2tDelR`	ΔR between second-leading lepton and reconstructed top quark
`l2wj1DelR`	ΔR between second-leading lepton and leading W boson candidate jet
`l2wj2DelR`	ΔR between second-leading lepton and second-leading W boson candidate jet
`met`	PF $\not{E}_{\mathrm{T}}$ in event
`nBjets`	Number of b tagged jets
`nJets`	Number of jets
`tEta`	η of reconstructed top quark
`tMass`	Mass of reconstructed top quark
`tMt`	Transverse mass of reconstructed top quark
`tPhi`	φ of reconstructed top quark
`tPt`	p_{T} of reconstructed top quark
`tbDelR`	ΔR between reconstructed top quark and leading b tagged jet
`totMass`	Total mass of all objects in the event
`totMt`	Total transverse mass of all objects in the event
`totPt`	Total p_{T} of all objects in the event
`wEta`	η of reconstructed W boson
`wMass`	Mass of reconstructed W boson
`wMt`	Transverse mass of reconstructed W boson
`wPhi`	φ of reconstructed W boson
`wPt`	p_{T} of reconstructed W boson
`wbDelR`	ΔR between reconstructed W boson and leading b tagged jet
`wj1Eta`	η of leading W boson candidate jet
`wj1Mass`	Mass of leading W boson candidate jet
`wj1Mt`	Transverse mass of leading W boson candidate jet
`wj1Phi`	φ of leading W boson candidate jet
`wj1Pt`	p_{T} of leading W boson candidate jet
`wj1bDelR`	ΔR of leading W boson candidate jet and leading b tagged jet
`wj1tDelR`	ΔR of leading W boson candidate jet and reconstructed top quark
`wj2Eta`	η of second-leading W boson candidate jet
`wj2Mass`	Mass of second-leading W boson candidate jet
`wj2Mt`	Transverse mass of second-leading W boson candidate jet
`wj2Phi`	φ of second-leading W boson candidate jet
`wj2Pt`	p_{T} of second-leading W boson candidate jet
`wj2bDelR`	ΔR of second-leading W boson candidate jet and leading b tagged jet
`wj2tDelR`	ΔR of second-leading W boson candidate jet and reconstructed top quark
`wtDelR`	ΔR between reconstructed W boson and reconstructed top quark
`wwDelR`	ΔR between W boson candidate jets
`wzDelR`	ΔR between reconstructed W boson and reconstructed Z boson
`zEta`	η of reconstructed Z boson
`zMass`	Mass of reconstructed Z boson
`zMt`	Transverse mass of reconstructed Z boson
`zPhi`	φ of reconstructed Z boson
`zbDelR`	ΔR between reconstructed Z boson and leading b tagged jet
`zjMaxR`	Greatest ΔR between the reconstructed Z boson and any jet

`zjMinR`	Smallest ΔR between the reconstructed Z boson and any jet
`ztDelR`	ΔR between reconstructed Z boson and reconstructed top quark
`zwj1DelR`	ΔR between reconstructed Z boson and leading W boson candidate jet
`zwj2DelR`	ΔR between reconstructed Z boson and second-leading W boson candidate jet
`zzDelR`	ΔR between leptons

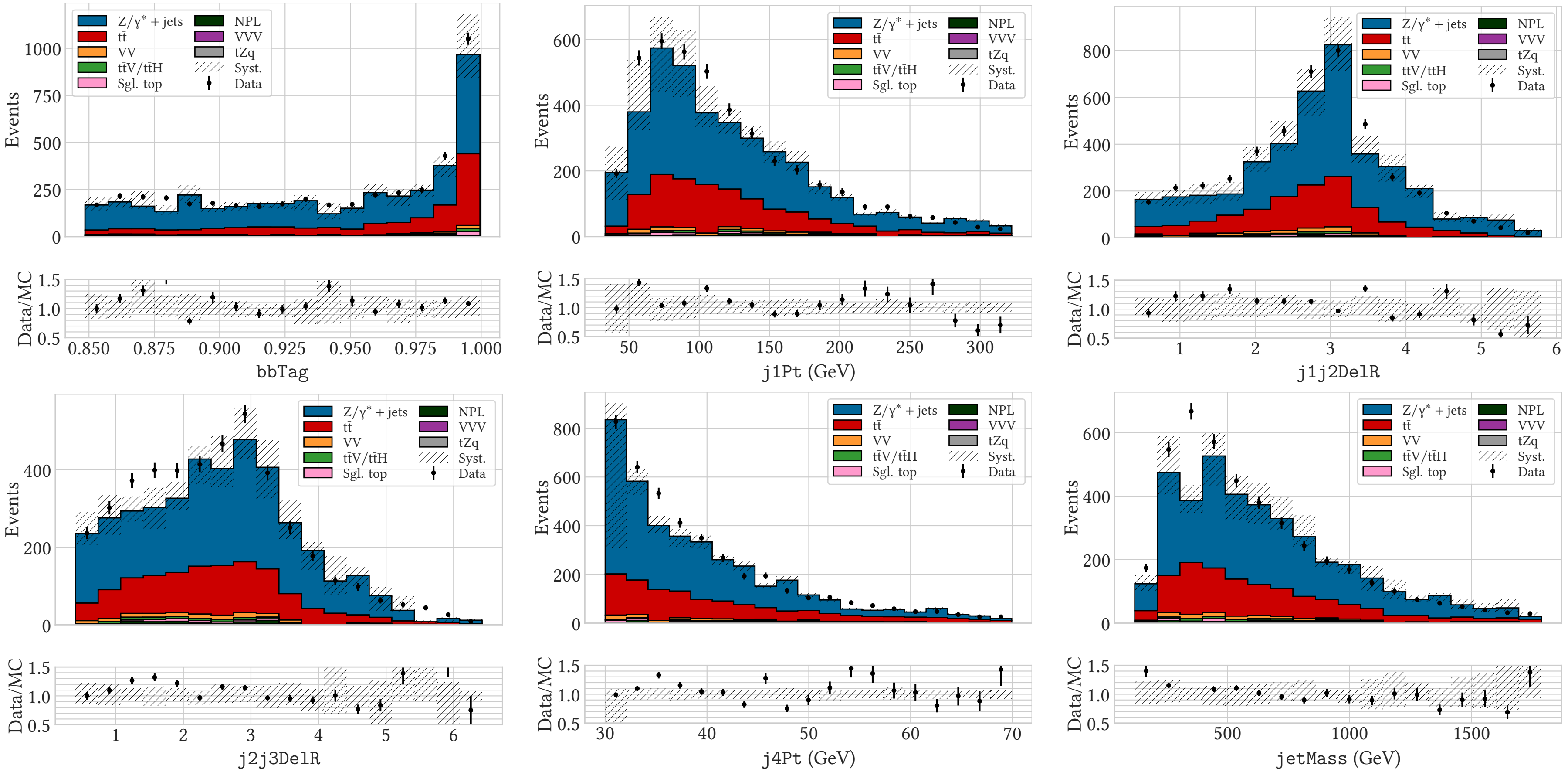


Figure B.1: Distribution of selected features in simulated samples and data for the BDT in the 2016 ee signal region. See Table B.1 for a definition of each feature.

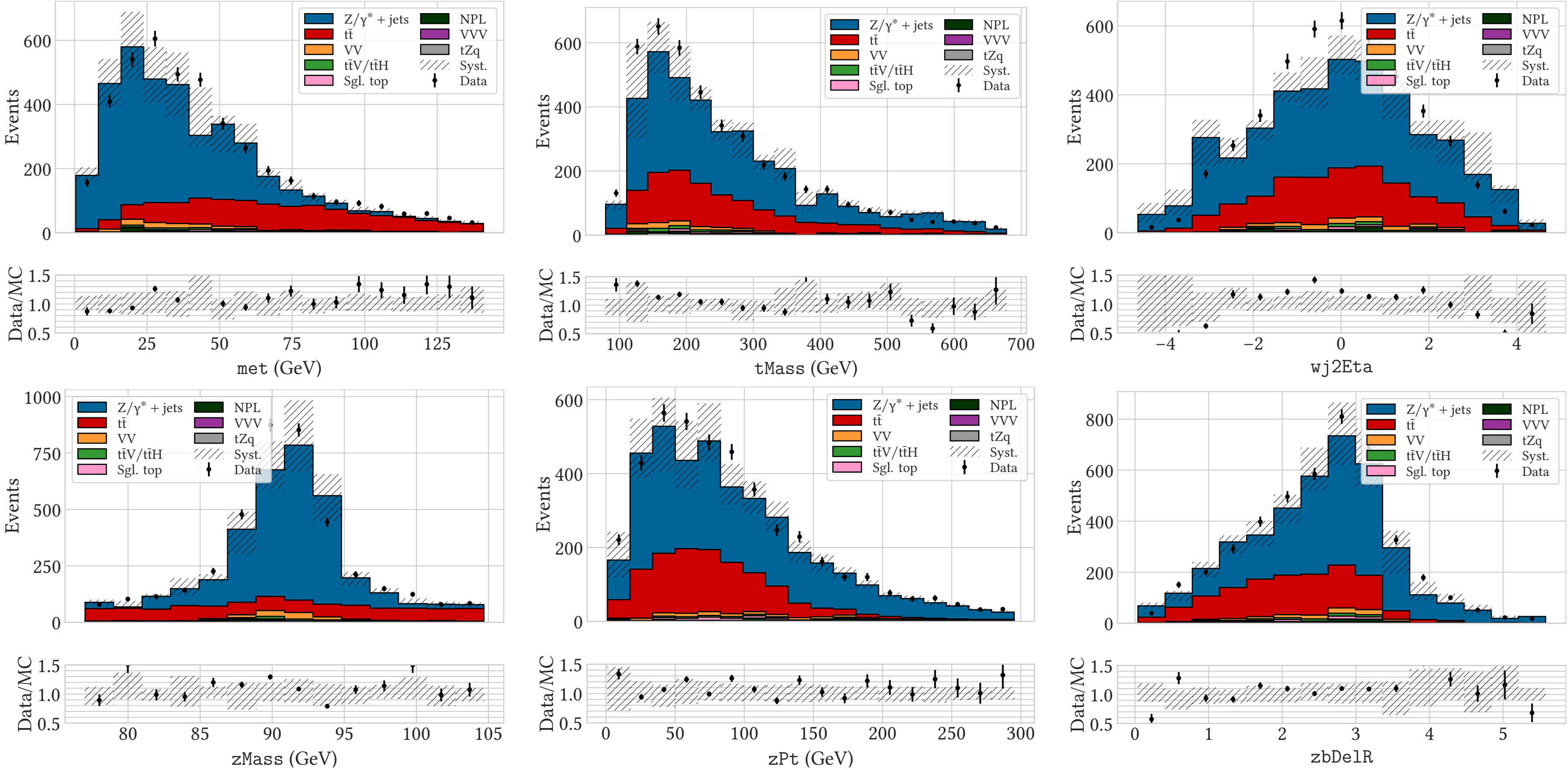


Figure B.1: Distribution of selected features in simulated samples and data for the BDT in the 2016 ee signal region. See Table B.1 for a definition of each feature.

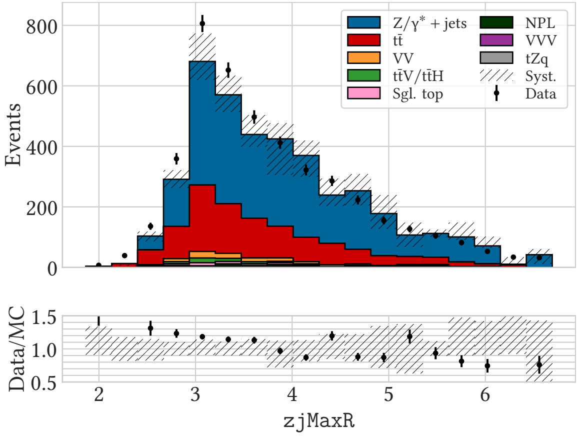


Figure B.1: Distribution of selected features in simulated samples and data for the BDT in the 2016 ee signal region. See Table B.1 for a definition of each feature.

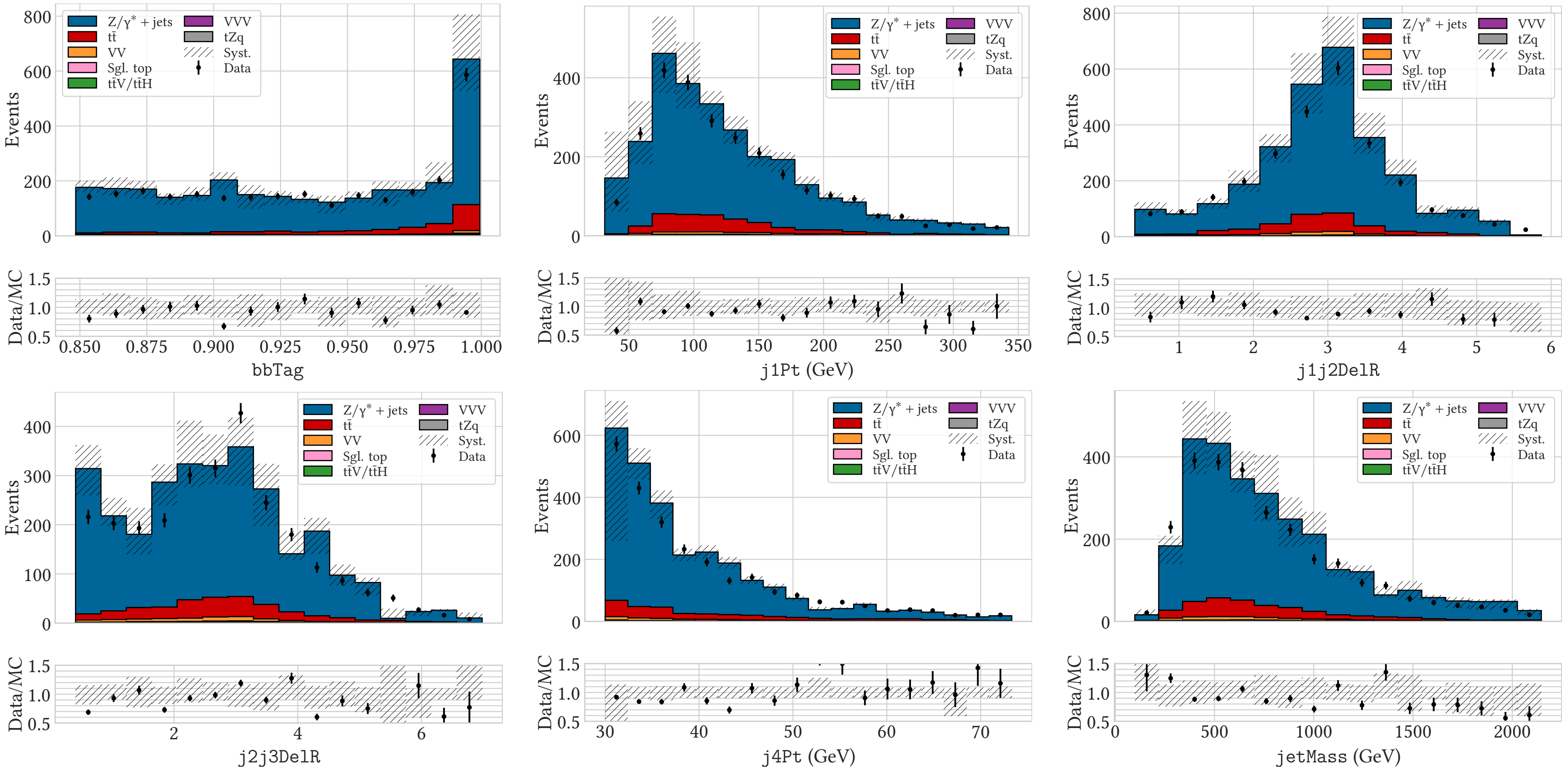


Figure B.2: Distribution of selected features in simulated samples and data for the BDT in the 2016 ee Z/γ^* + jets CR. See Table B.1 for a definition of each feature.

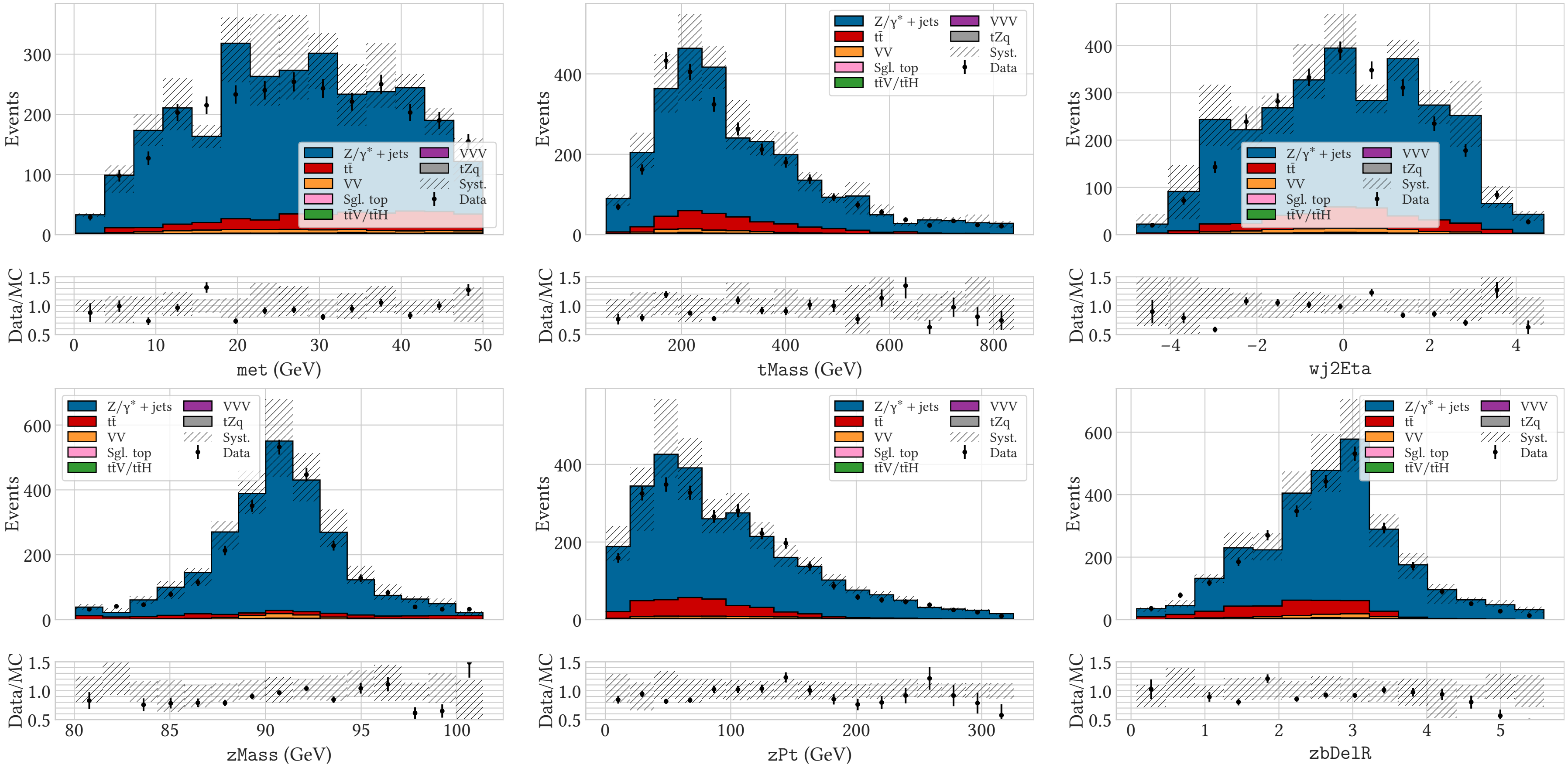


Figure B.2: Distribution of selected features in simulated samples and data for the BDT in the 2016 ee Z/γ^* + jets CR. See Table B.1 for a definition of each feature.

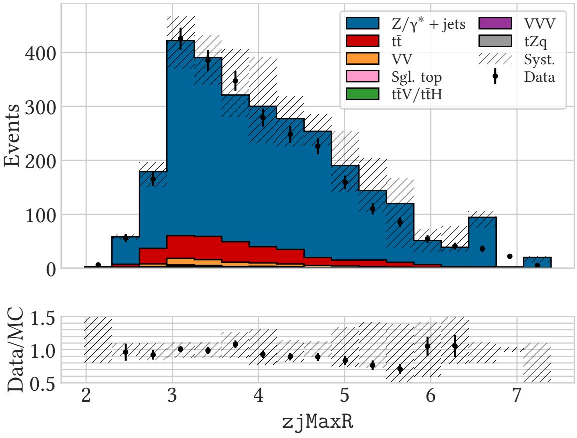


Figure B.2: Distribution of selected features in simulated samples and data for the BDT in the 2016 ee Z/γ^* + jets CR. See Table B.1 for a definition of each feature.

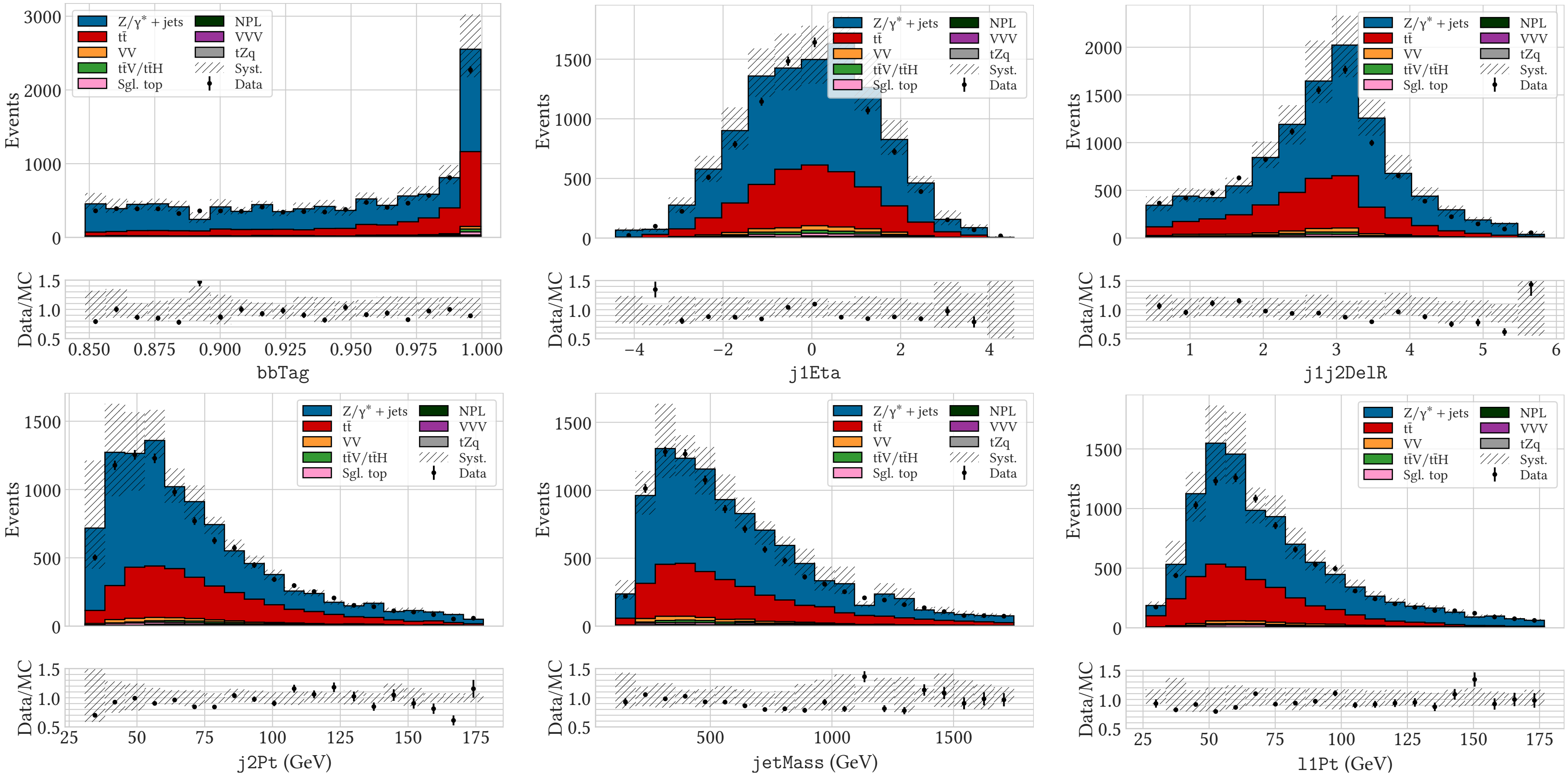


Figure B.3: Distribution of selected features in simulated samples and data for the BDT in the 2016 μμ signal region. See Table B.1 for a definition of each feature.

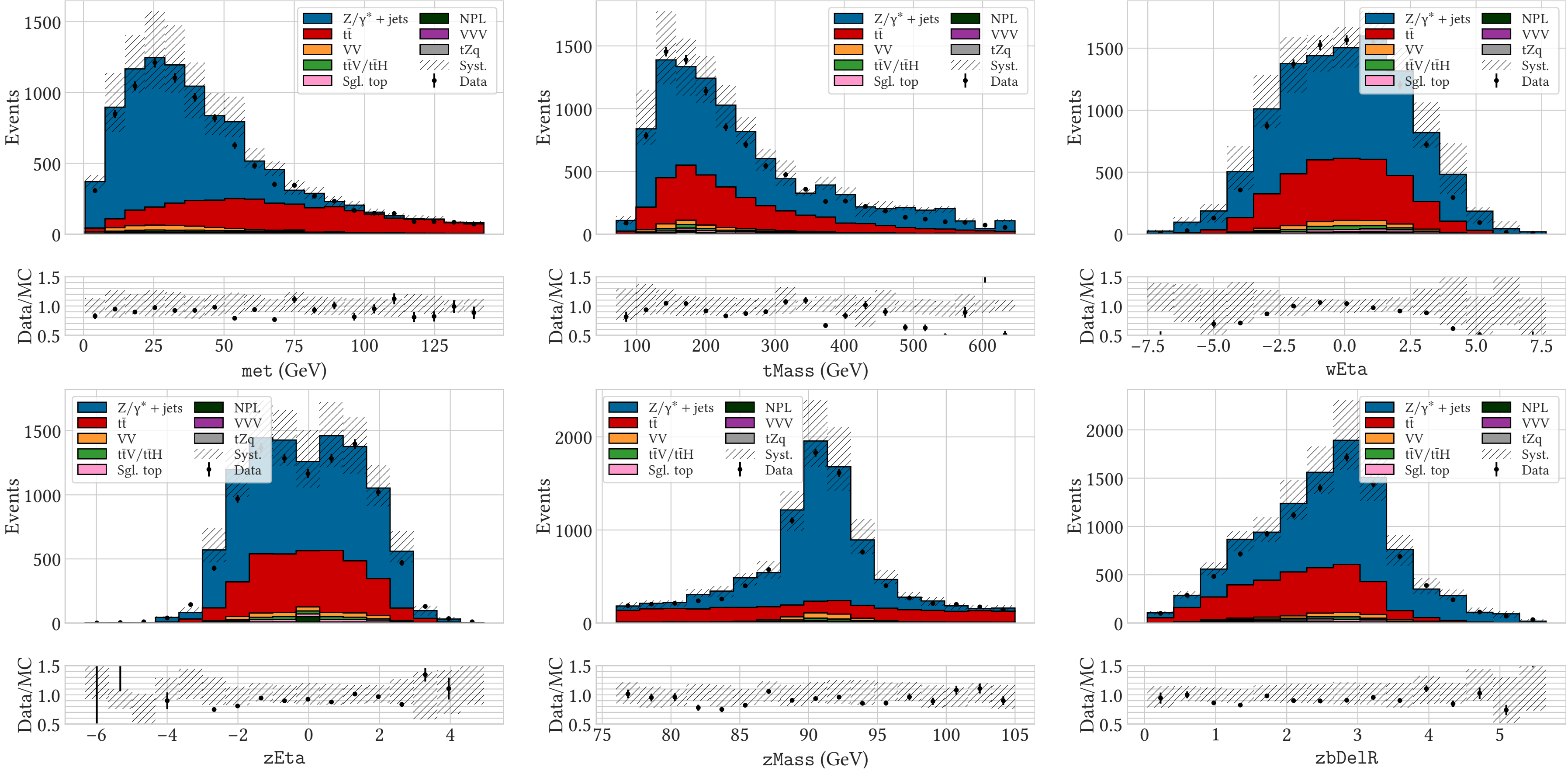


Figure B.3: Distribution of selected features in simulated samples and data for the BDT in the 2016 μμ signal region. See Table B.1 for a definition of each feature.

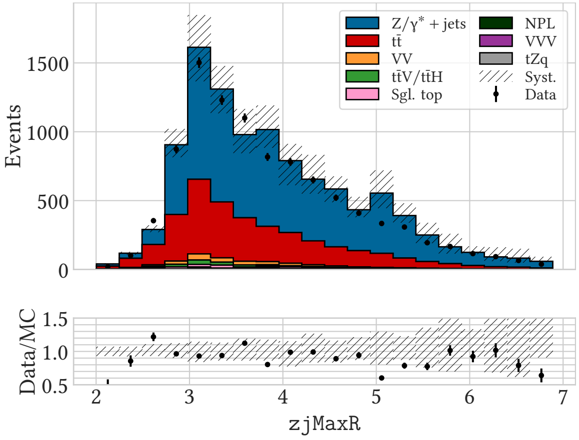


Figure B.3: Distribution of selected features in simulated samples and data for the BDT in the 2016 μμ signal region. See Table B.1 for a definition of each feature.

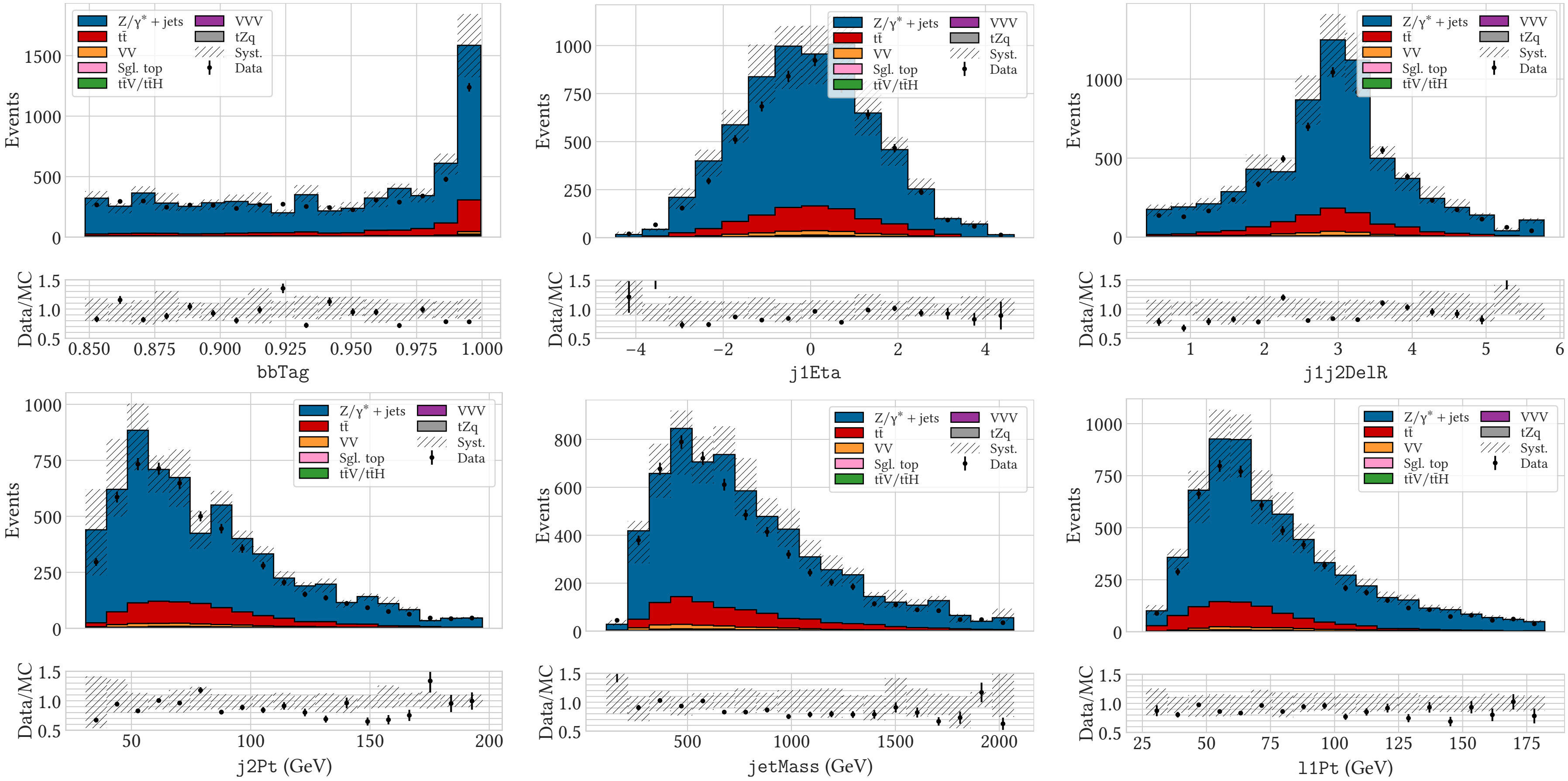


Figure B.4: Distribution of selected features in simulated samples and data for the BDT in the 2016 μμ Z/γ* + jets CR. See Table B.1 for a definition of each feature.

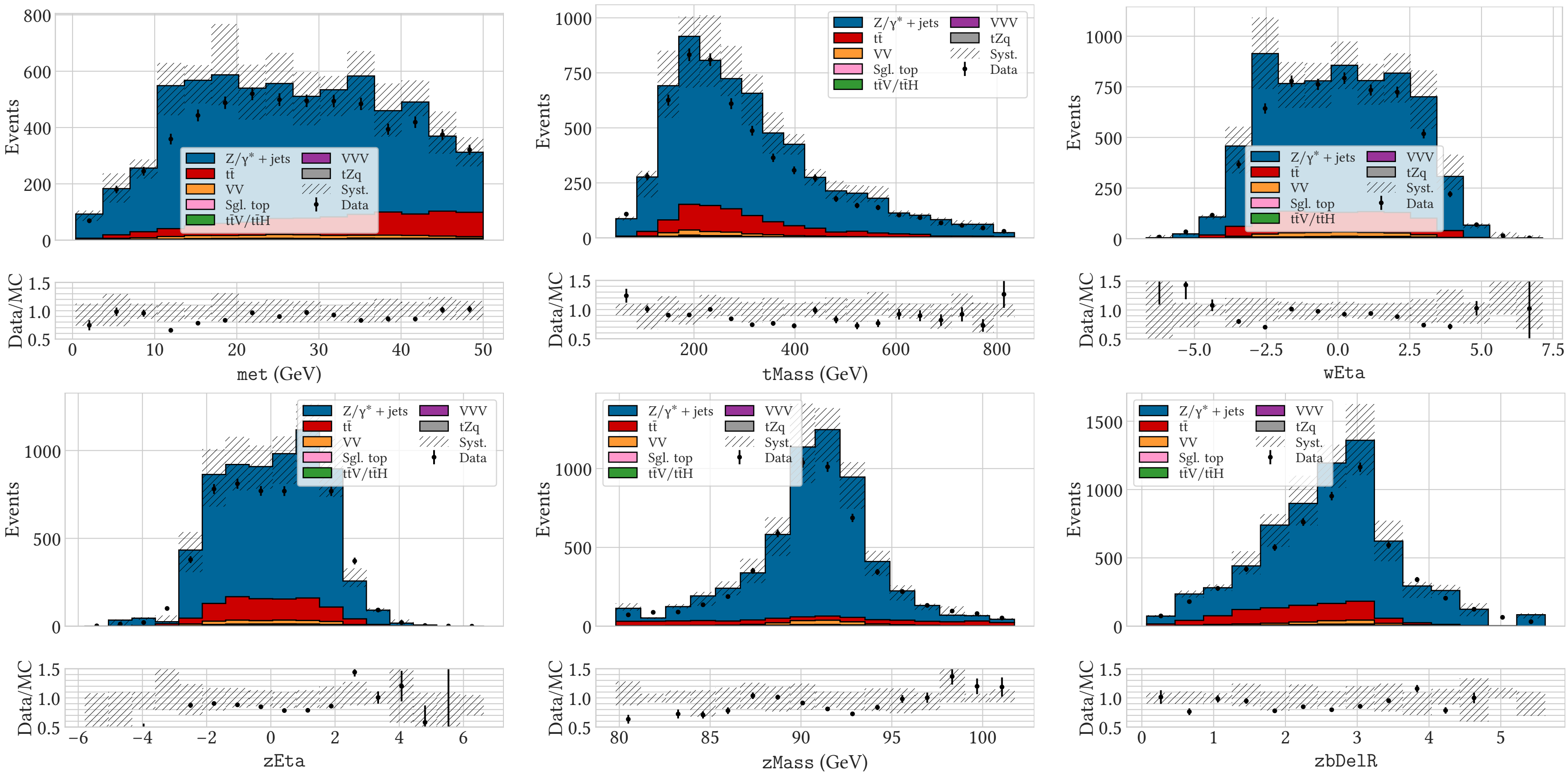


Figure B.4: Distribution of selected features in simulated samples and data for the BDT in the 2016 μμ Z/γ* + jets CR. See Table B.1 for a definition of each feature.

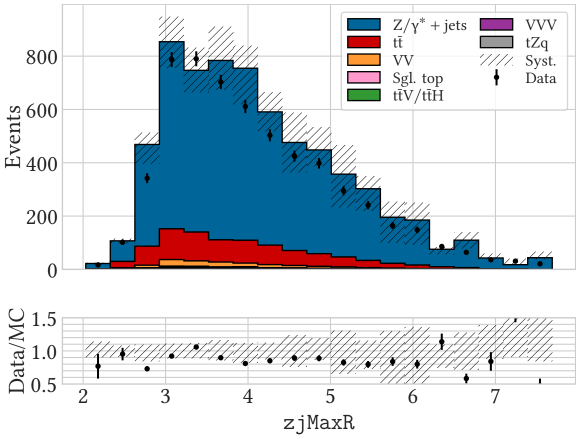


Figure B.4: Distribution of selected features in simulated samples and data for the BDT in the 2016 μμ Z/γ^* + jets CR. See Table B.1 for a definition of each feature.

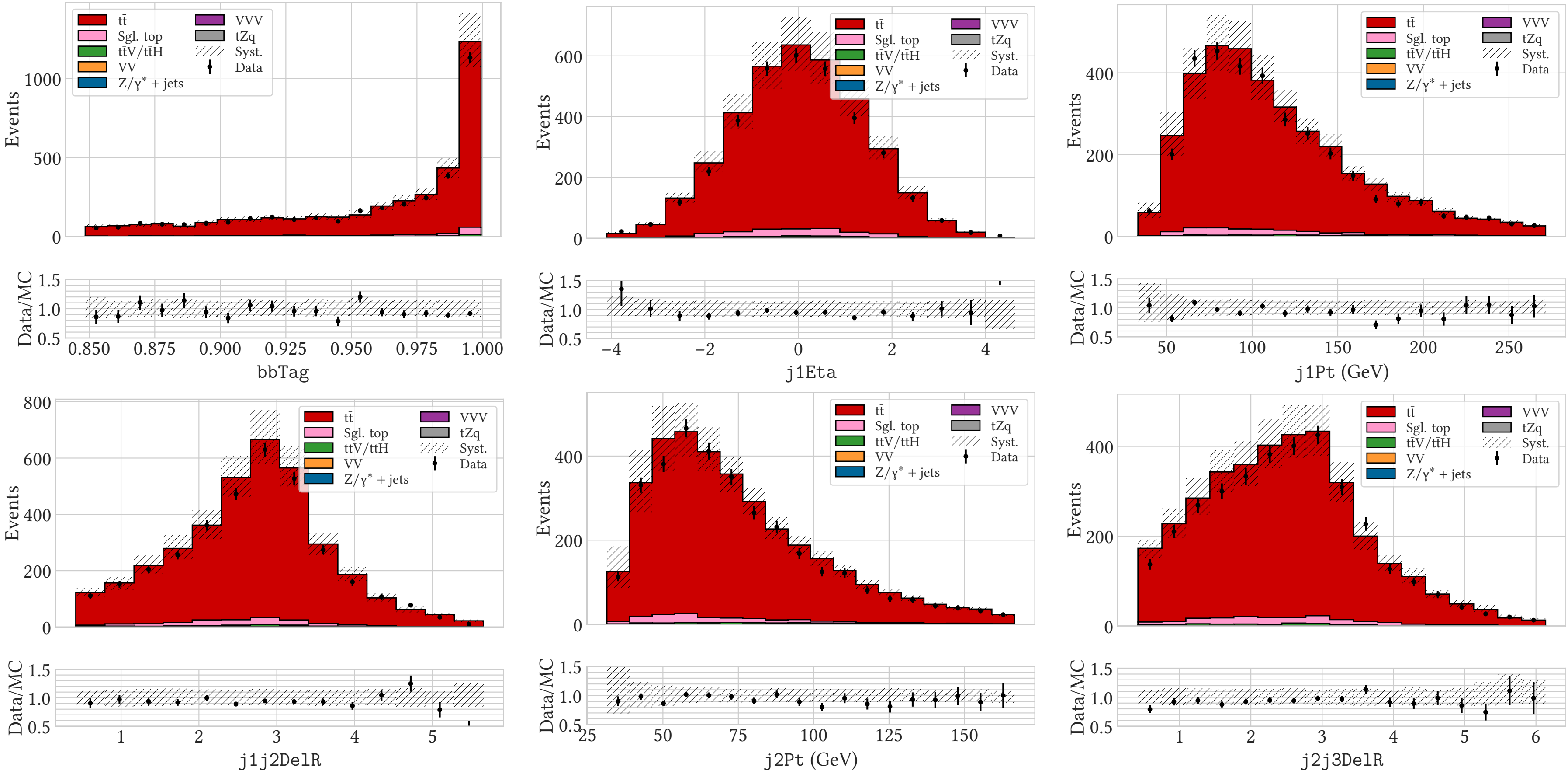


Figure B.5: Distribution of features selected for use in the ee or μμ BDT in the $t\bar{t}$ CR in 2016. See Table B.1 for a definition of each feature.

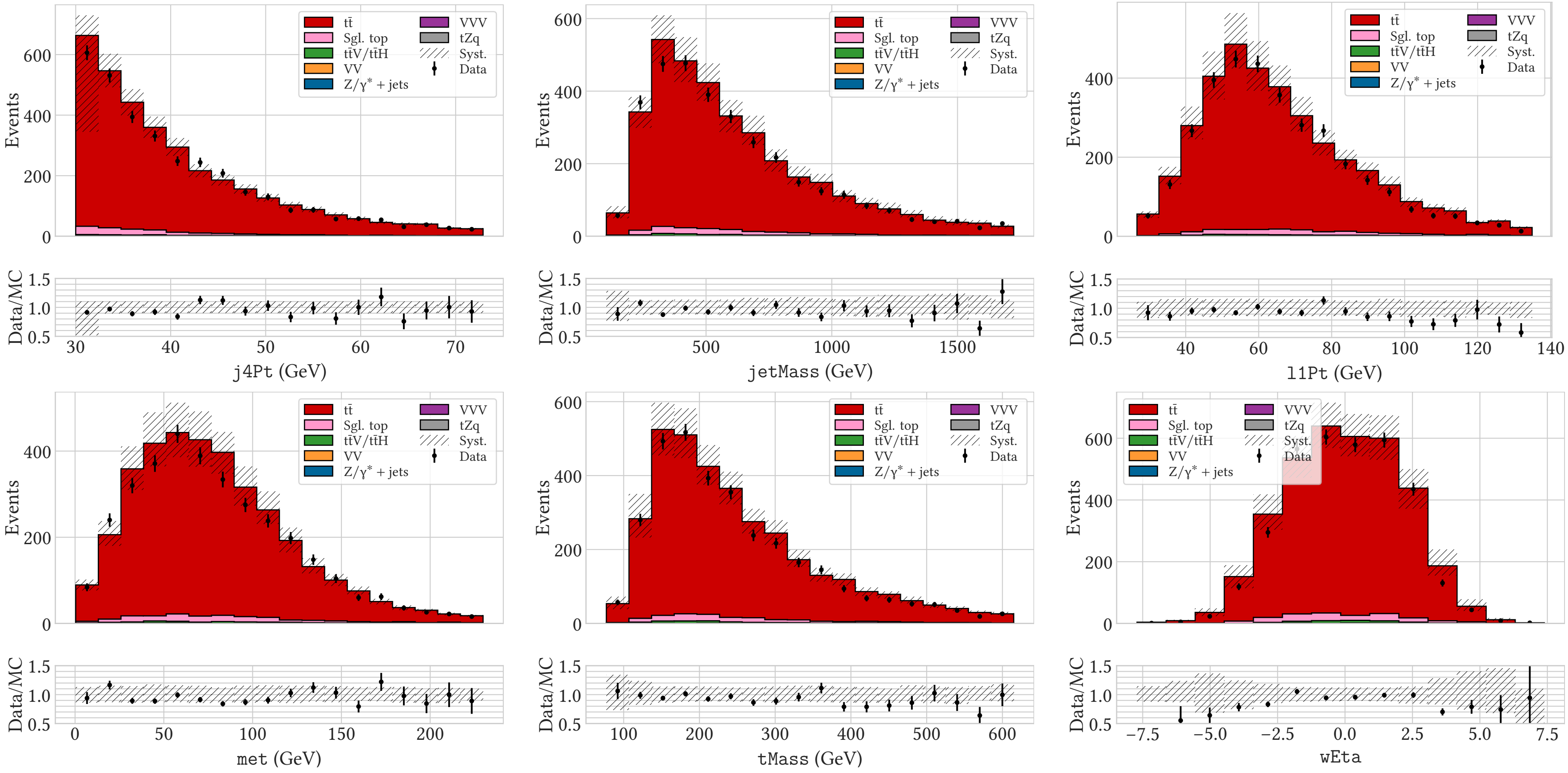


Figure B.5: Distribution of features selected for use in the ee or μμ BDT in the tt̄ CR in 2016. See Table B.1 for a definition of each feature.

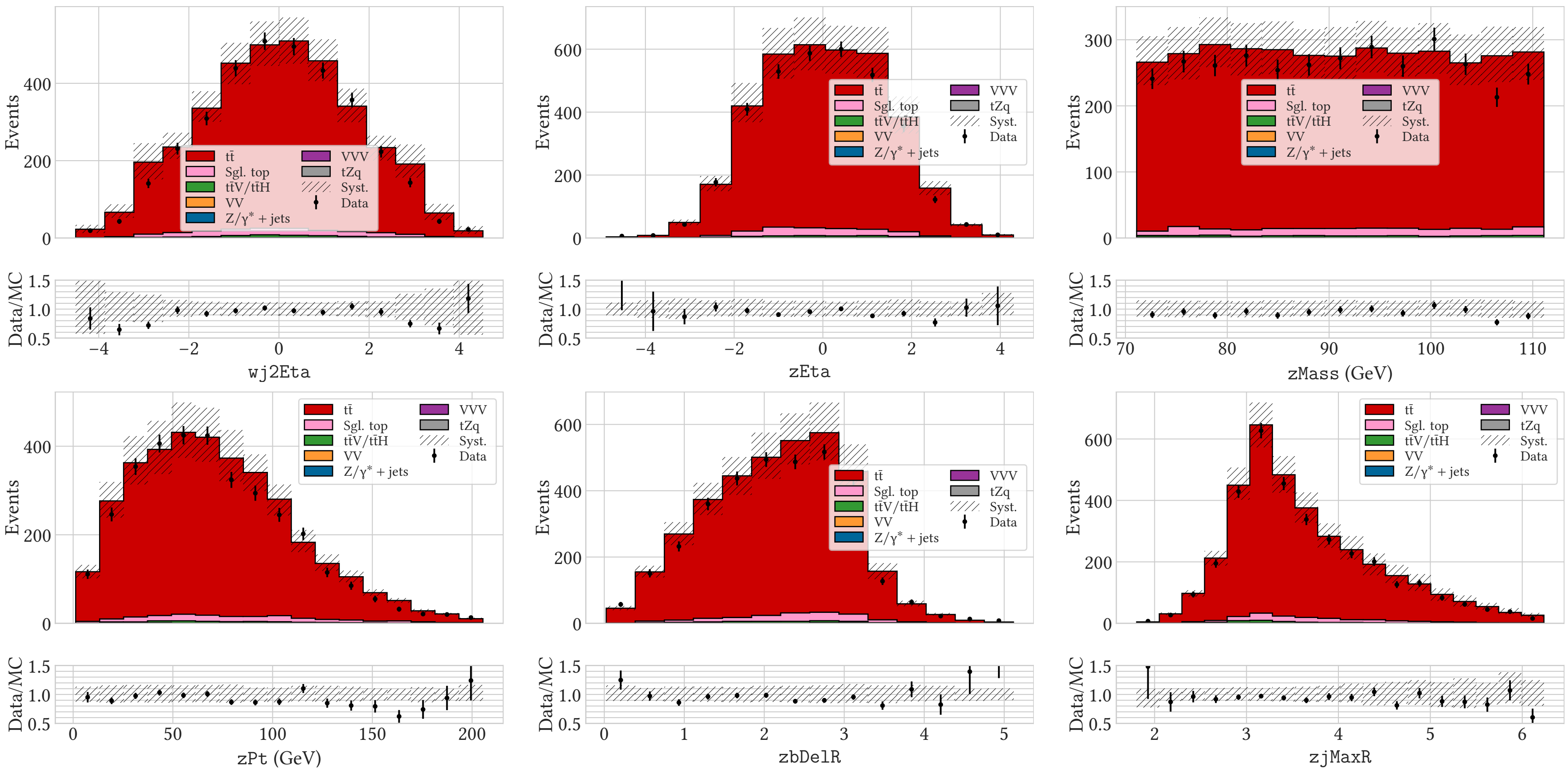


Figure B.5: Distribution of features selected for use in the ee or μμ BDT in the tt̄ CR in 2016. See Table B.1 for a definition of each feature.

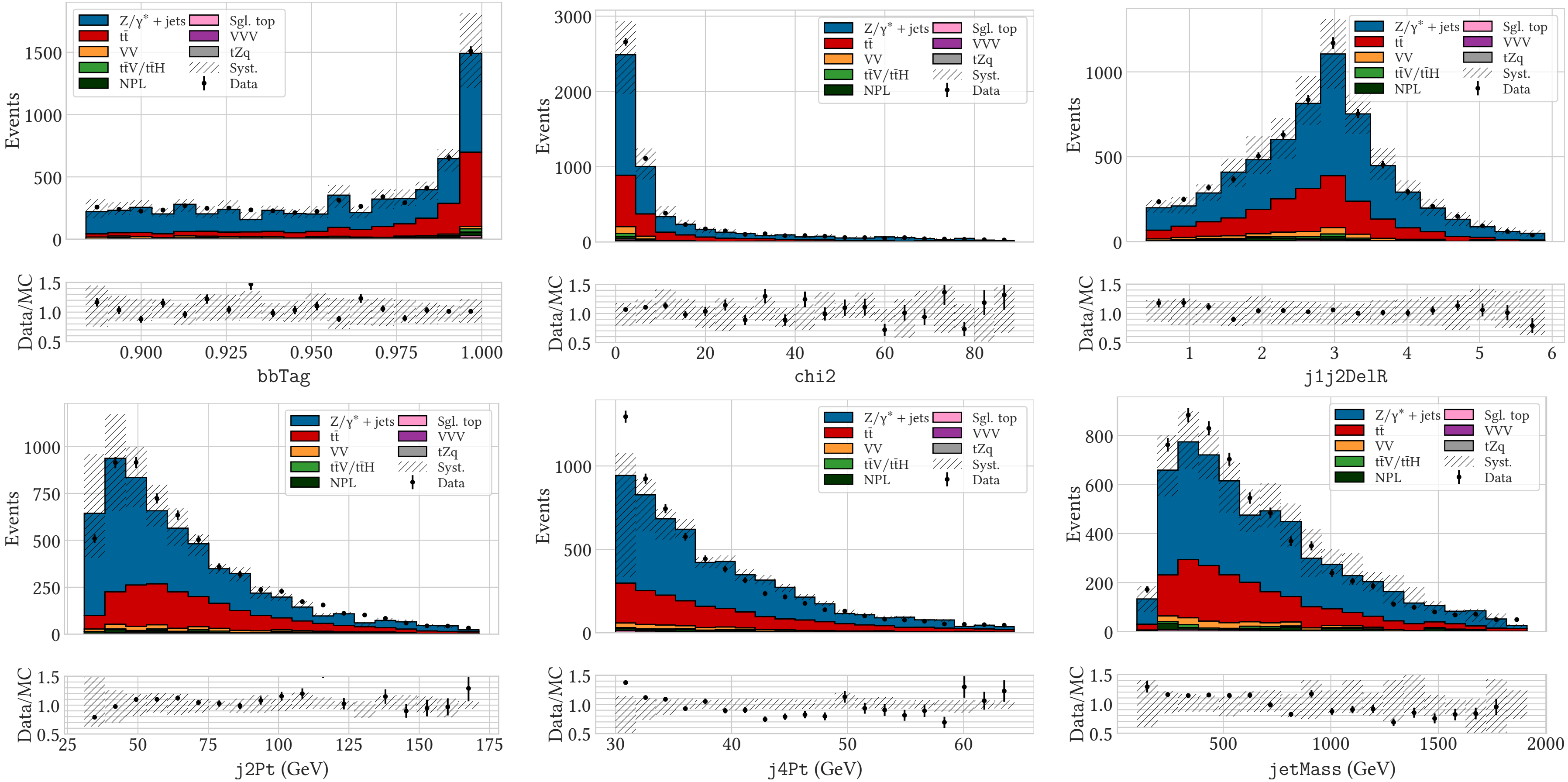


Figure B.6: Distribution of selected features in simulated samples and data for the BDT in the 2017 ee signal region. See Table B.1 for a definition of each feature.

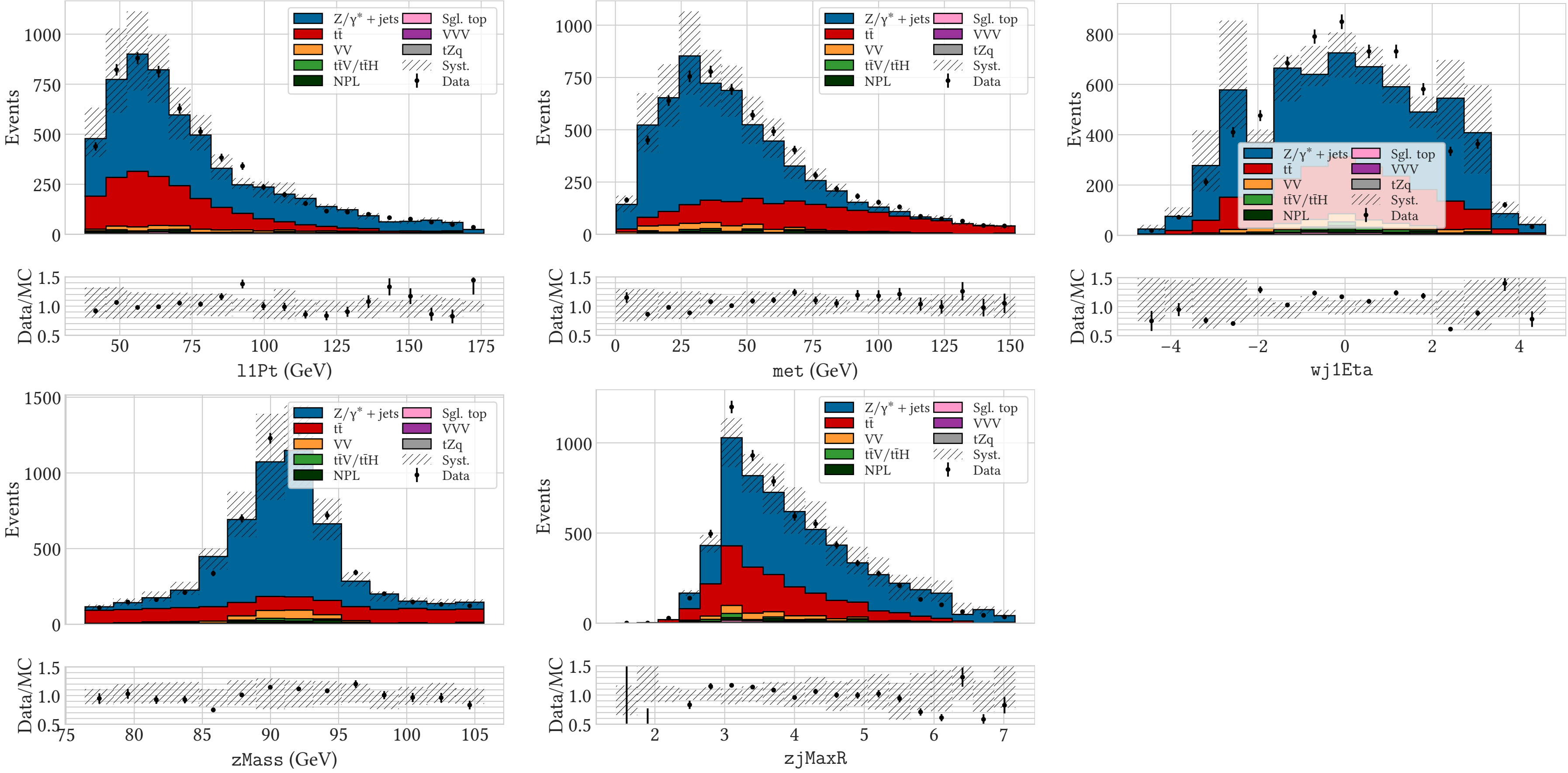


Figure B.6: Distribution of selected features in simulated samples and data for the BDT in the 2017 ee signal region. See Table B.1 for a definition of each feature.

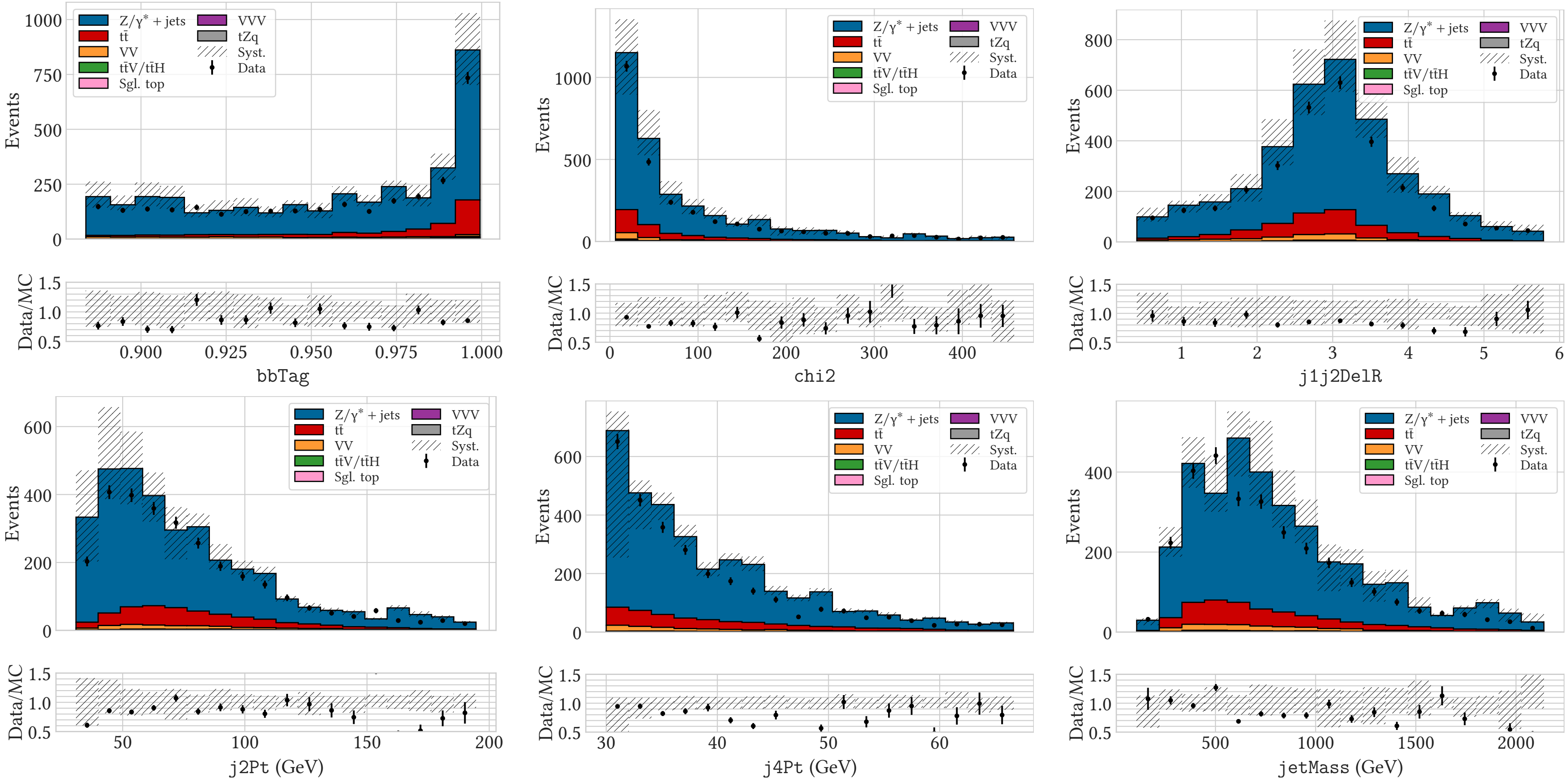


Figure B.7: Distribution of selected features in simulated samples and data for the BDT in the 2017 ee $Z/\gamma^* + \text{jets}$ CR. See Table B.1 for a definition of each feature.

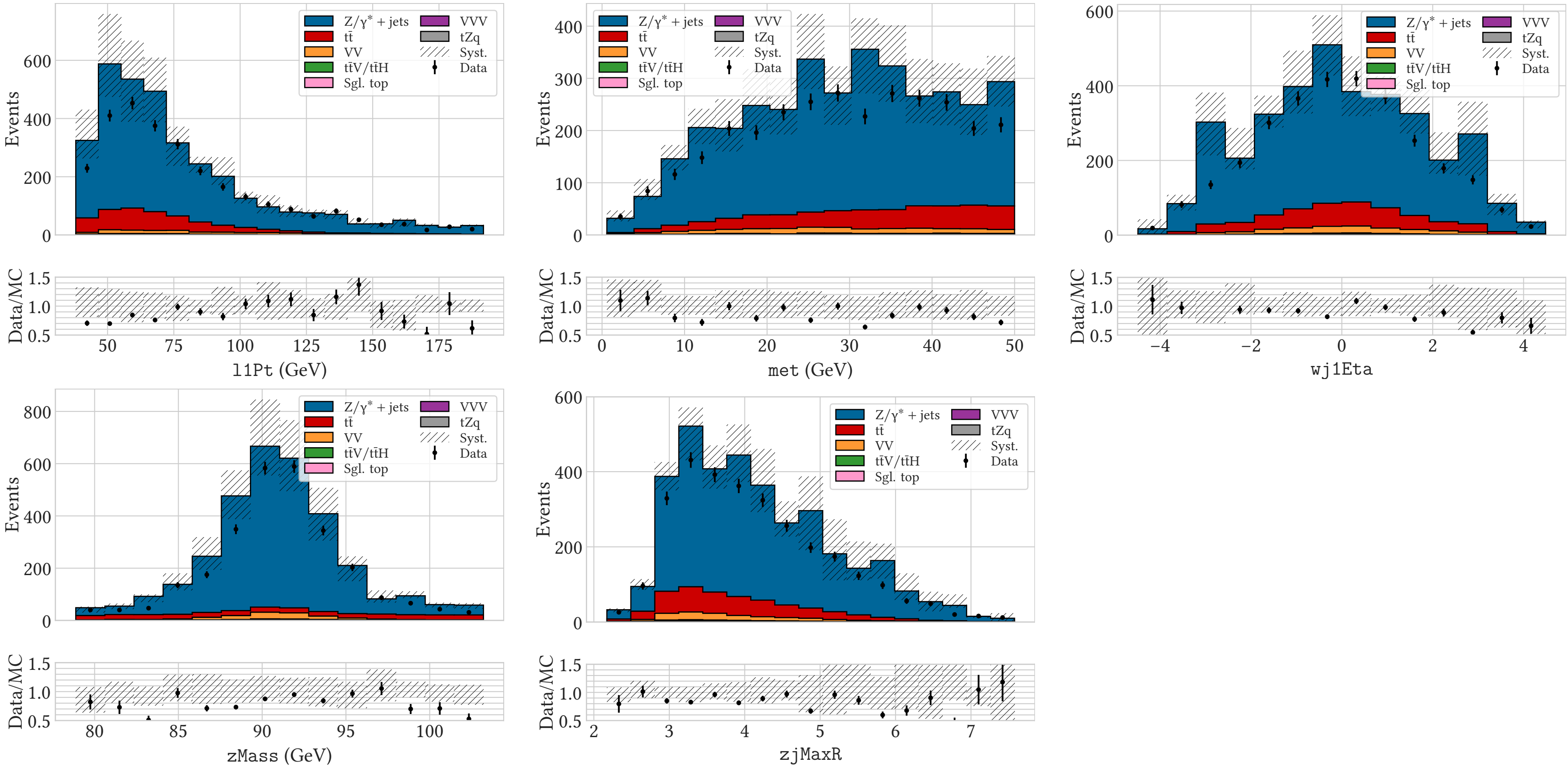


Figure B.7: Distribution of selected features in simulated samples and data for the BDT in the 2017 ee Z/γ^* + jets CR. See Table B.1 for a definition of each feature.

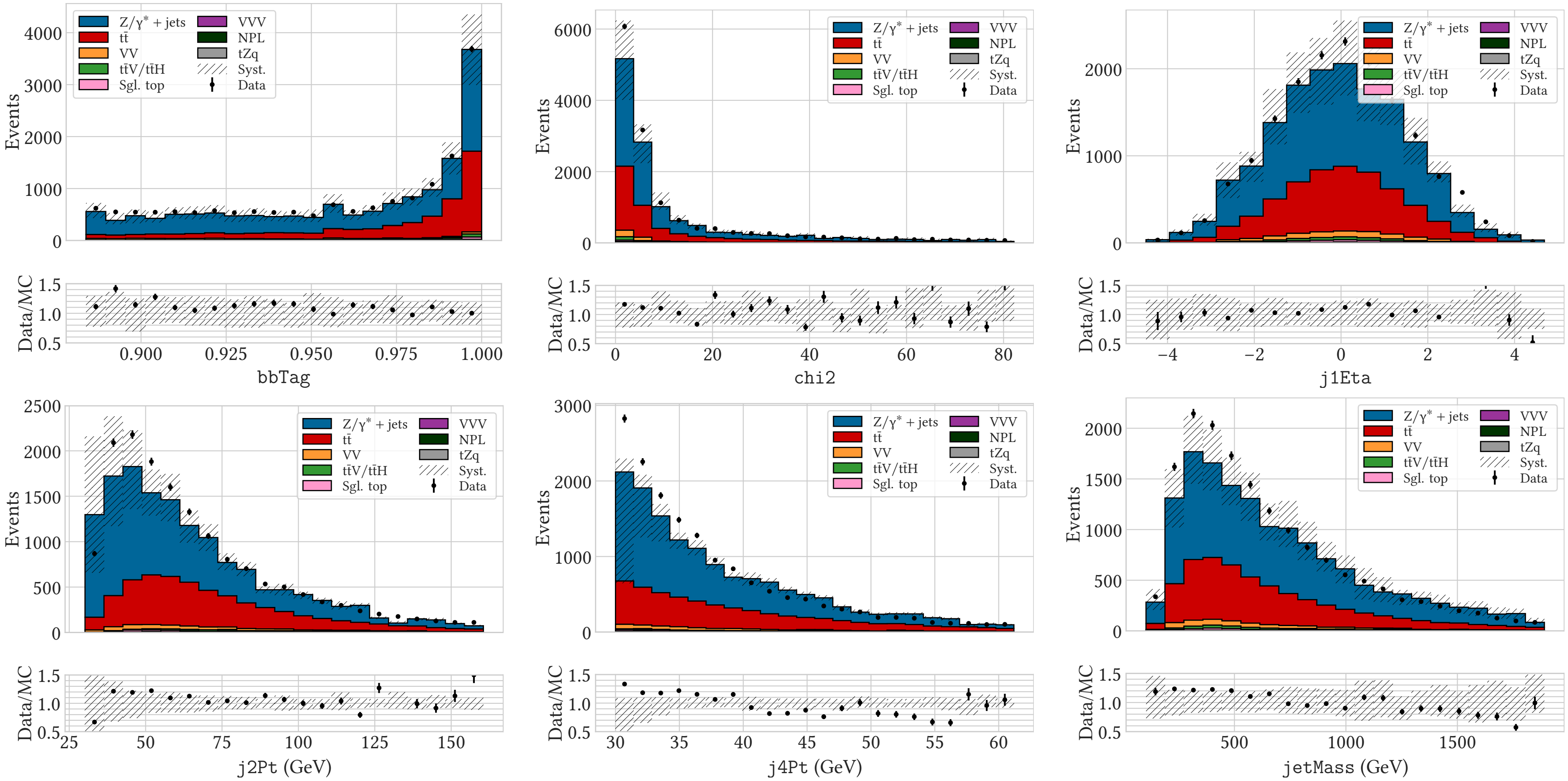


Figure B.8: Distribution of selected features in simulated samples and data for the BDT in the 2017 μμ signal region. See Table B.1 for a definition of each feature.

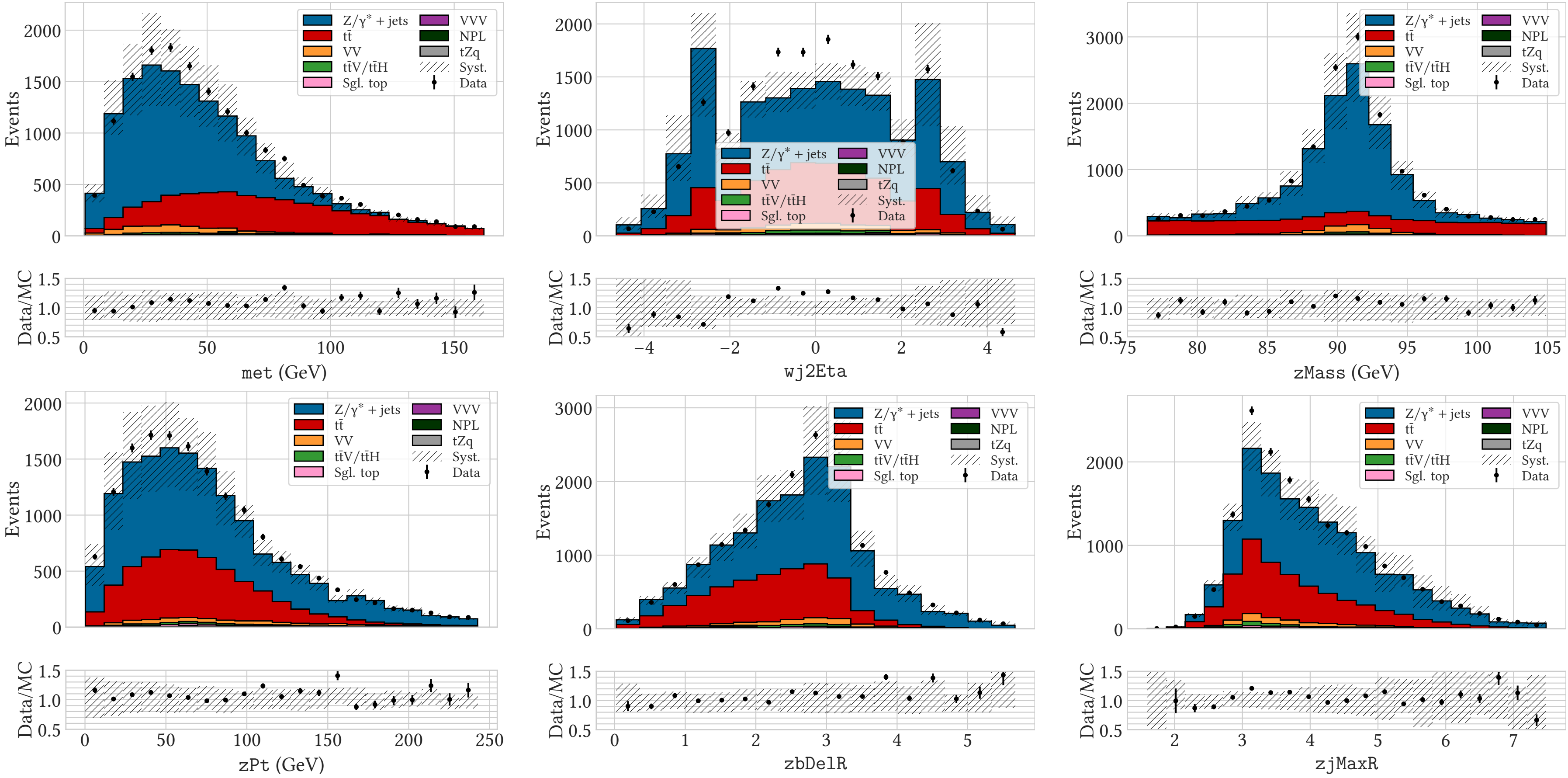


Figure B.8: Distribution of selected features in simulated samples and data for the BDT in the 2017 μμ signal region. See Table B.1 for a definition of each feature.

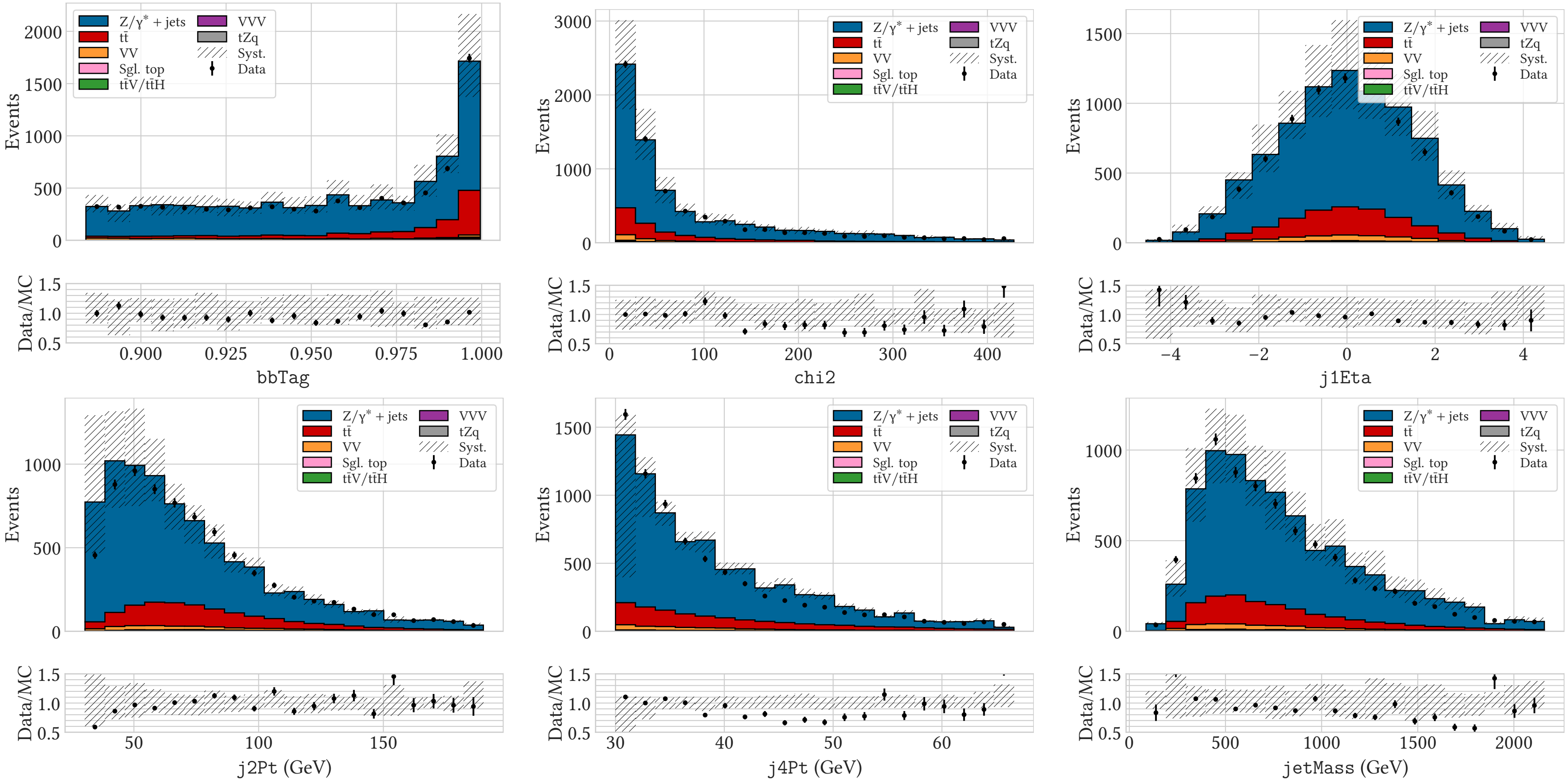


Figure B.9: Distribution of selected features in simulated samples and data for the BDT in the 2017 μμ Z/γ* + jets CR. See Table B.1 for a definition of each feature.

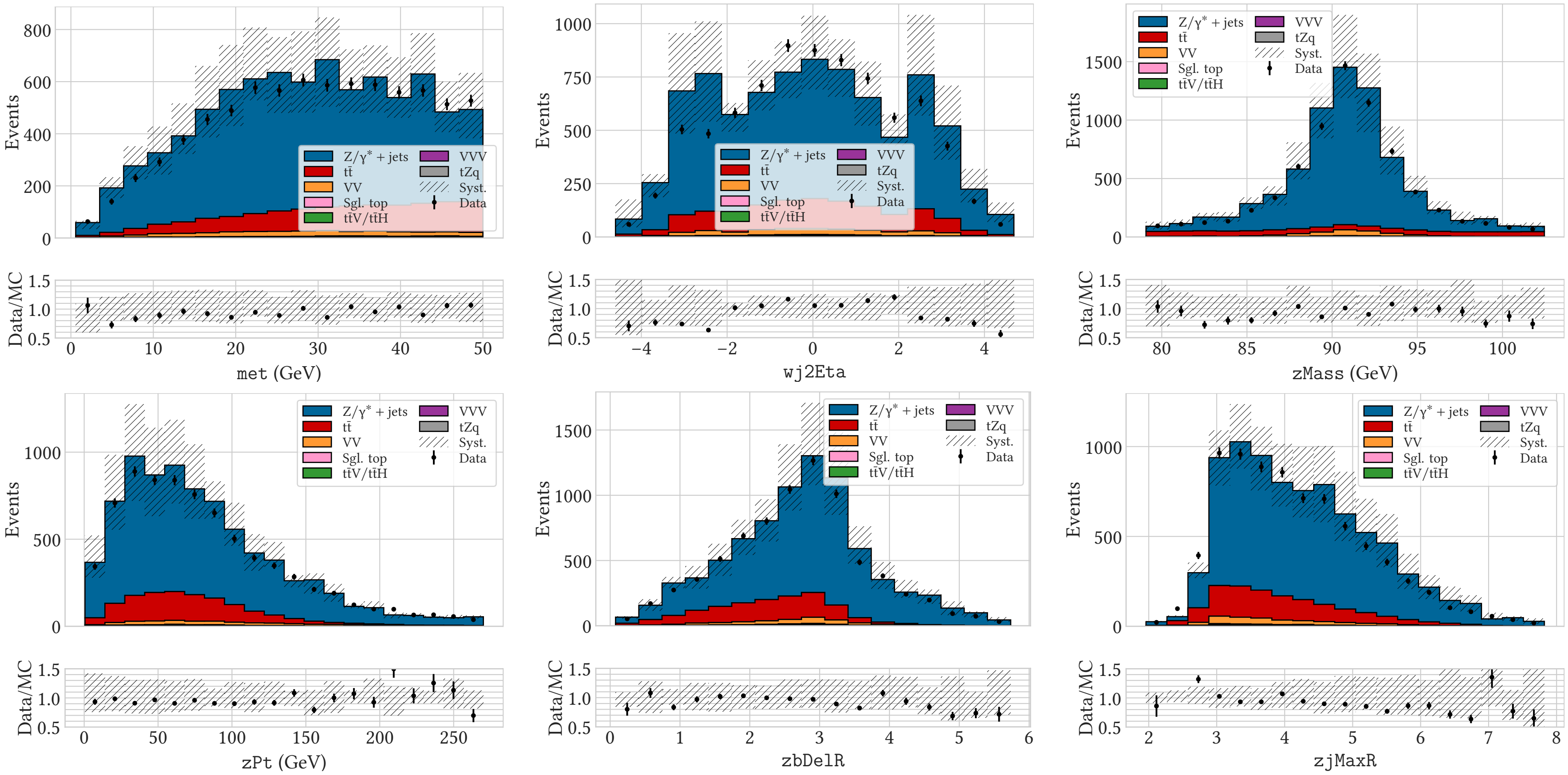


Figure B.9: Distribution of selected features in simulated samples and data for the BDT in the 2017 μμ Z/γ^* + jets CR. See Table B.1 for a definition of each feature.

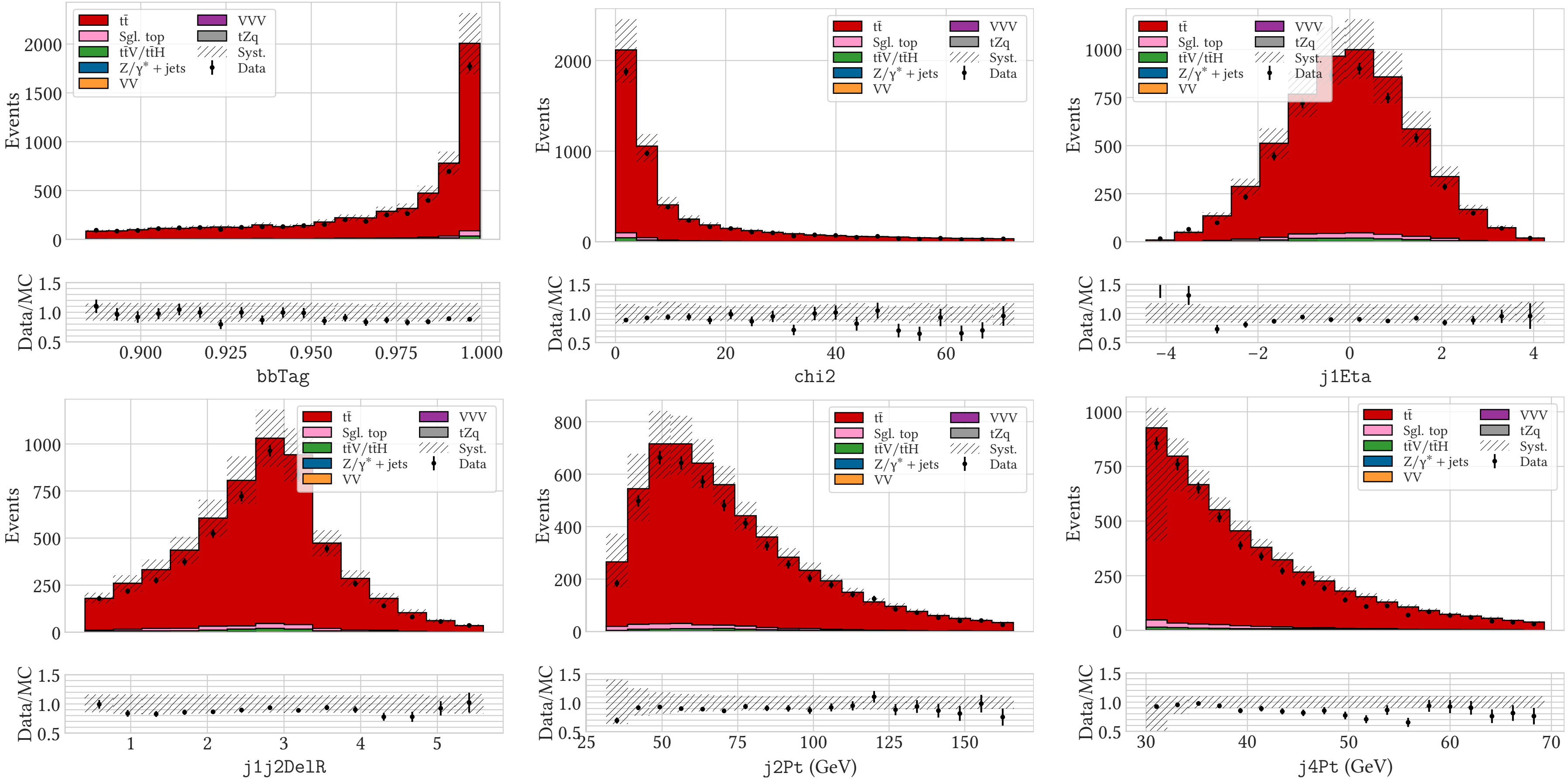


Figure B.10: Distribution of features selected for use in the ee or μμ BDT in the tt̄ CR in 2017. See Table B.1 for a definition of each feature.

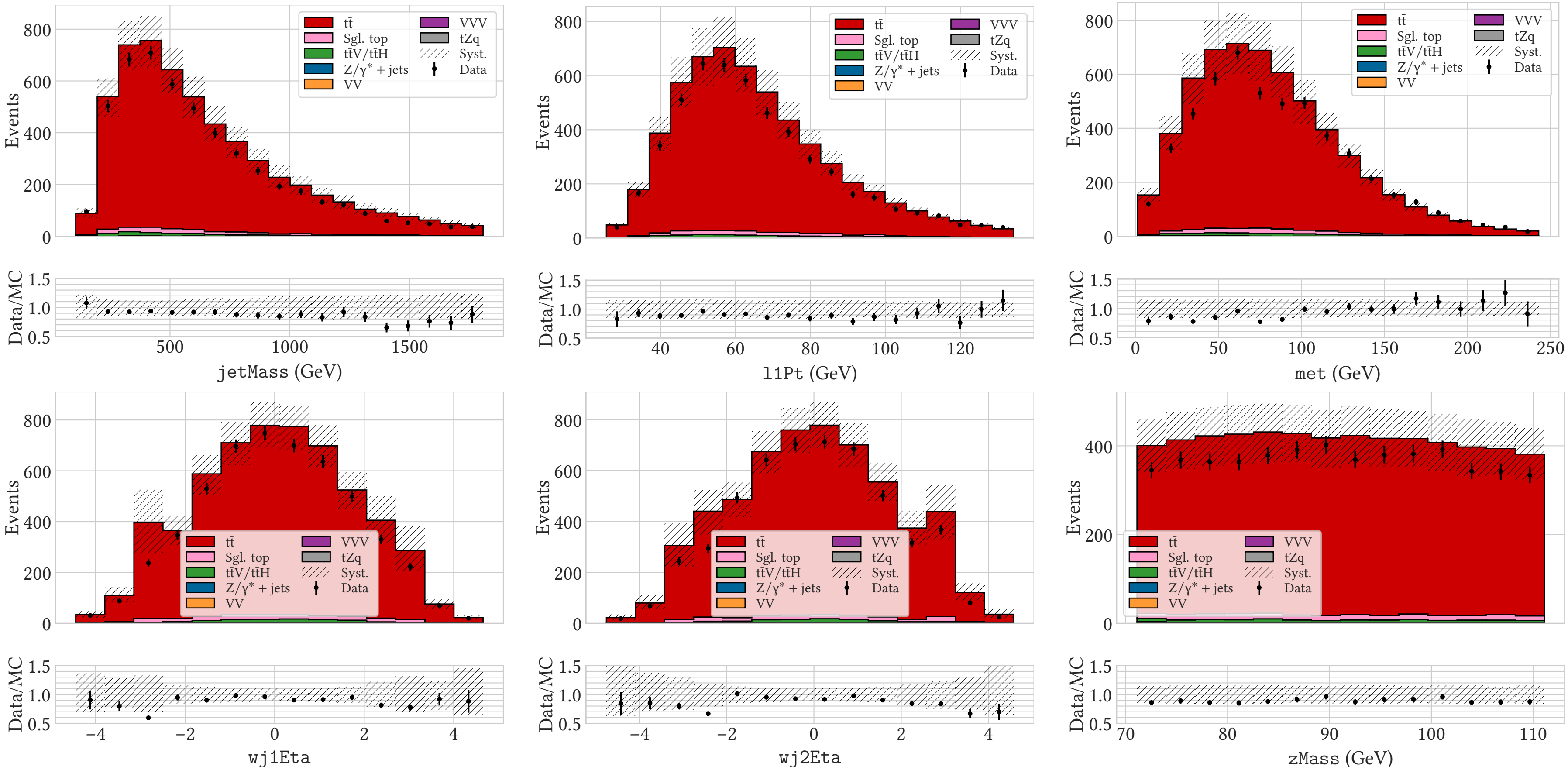


Figure B.10: Distribution of features selected for use in the ee or μμ BDT in the tt̄ CR in 2017. See Table B.1 for a definition of each feature.

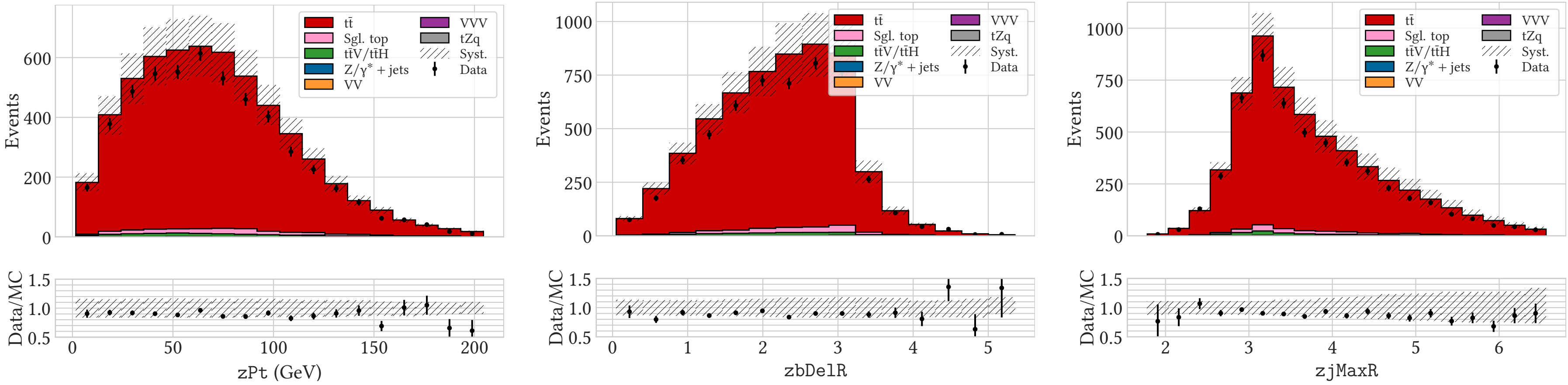


Figure B.10: Distribution of features selected for use in the ee or μμ BDT in the tt̄ CR in 2017. See Table B.1 for a definition of each feature.

Bibliography

[1] J. Burnet. *Early Greek Philosophy.* 4th ed. A & C Black, 1948. isbn: 978-1-199-63270-8.

[2] T. McEvilley. *The Shape of Ancient Thought: Comparative Studied in Greek and Indian Philosophies.* Allworth Press, 2002. isbn: 978-1-58115-203-6.

[3] J. J. Thomson. “Cathode rays”. In: *Philosophical Magazine.* 5th ser. 44 (Oct. 1897), p. 293. issn: 1478-6435. doi: `10.1080/14786449708621070`.

[4] E. Rutherford. “Nuclear Constitution of Atoms”. In: *Nature* 105 (17th June 1920), p. 500. issn: 0028-0836. doi: `10.1038/105500b0`.

[5] J. Chadwick. “Possible Existence of a Neutron”. In: *Nature* 129 (27th Feb. 1932), p. 312. issn: 0028-0836. doi: `10.1038/129312a0`.

[6] J. C. Street and E. C. Stevenson. “New Evidence for the Existence of a Particle of Mass Intermediate Between the Proton and Electron”. In: *Physical Review* 52 (1st Nov. 1937), p. 1003. issn: 0031-899X. doi: `10.1103/PhysRev.52.1003`.

[7] F. Reines and C. L. Cowan. “The Neutrino”. In: *Nature* 178 (1st Sept. 1956), p. 446. issn: 0028-0836. doi: `10.1038/178446a0`.

[8] G. Danby et al. “Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos”. In: *Physical Review Letters* 9 (1st July 1962), p. 36. issn: 0031-9007. doi: `10.1103/PhysRevLett.9.36`.

[9] E. D. Bloom et al. “High-Energy Inelastic e–p Scattering at 6° and 10°”. In: *Physical Review Letters* 23 (20th Oct. 1969), p. 930. issn: 0031-9007. doi: `10.1103/PhysRevLett.23.930`.

[10] M. Breidenbach et al. “Observed Behavior of Highly Inelastic Electron–Proton Scattering”. In: *Physical Review Letters* 23 (1969), p. 935. issn: 0031-9007. doi: `10.1103/PhysRevLett.23.935`.

[11] M. L. Perl et al. “Evidence for Anomalous Lepton Production in e^+–e^- Annihilation”. In: *Physical Review Letters* 35 (1st Dec. 1975), p. 1489. issn: 0031-9007. doi: `10.1103/PhysRevLett.35.1489`.

[12] G. Arnison et al. “Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c^2 at the CERN SPS Collider”. In: *Physics Letters B* 126 (7th July 1983), p. 398. issn: 0370-2693. doi: `10.1016/0370-2693(83)90188-0`.

[13] K. Kodama et al. “Observation of tau neutrino interactions”. In: *Physics Letters B* 504 (12th Apr. 2001), p. 218. issn: 0370-2693. doi: `10.1016/S0370-2693(01)00307-0`. arXiv: `hep-ex/0012035 [hep-ex]`.

[14] G. Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Physics Letters B* 716 (17th Sept. 2012), p. 1. issn: 0370-2693. doi: `10.1016/j.physletb.2012.08.020`. arXiv: `1207.7214 [hep-ex]`.

[15] S. Chatrchyan et al. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Physics Letters B* 716 (17th Sept. 2012), p. 30. issn: 0370-2693. doi: `10.1016/j.physletb.2012.08.021`. arXiv: `1207.7235 [hep-ex]`.

[16] M. Tanabashi et al. “Review of Particle Physics”. In: *Physical Review D: Particles, fields, gravitation, and cosmology* 98.3 (3 Aug. 2018), p. 030001. issn: 1550-2368. doi: `10.1103/PhysRevD.98.030001`.

[17] N. Carron. “Babel of Units: The Evolution of Units Systems in Classical Electromagnetism”. 21st May 2015. arXiv: `1506.01951 [physics.hist-ph]`.

[18] B. Martin and G. Shaw. *Particle Physics.* 3rd ed. Manchester Physics Series. John Wiley & Sons, Oct. 2008. isbn: 978-0-470-03294-7.

[19] C. Burgard. “Standard model of physics”. Modified. 31st Dec. 2016. url: `http://www.texample.net/tikz/examples/model-physics/` (visited on 15/10/2018).

[20] A. C. Phillips. *Introduction to Quantum Mechanics.* 1st ed. Manchester Physics Series. John Wiley & Sons, 16th July 2003. isbn: 978-0-470-85323-8.

[21] R. Shankar. *Principles of Quantum Mechanics.* 2nd ed. Plenum Press, 31st Aug. 1994. isbn: 978-0-306-44790-7.

[22] T.-P. Cheng and L.-F. Li. *Gauge theory of elementary particle physics.* Clarendon Press, 10th Aug. 1995. isbn: 978-0-19-851961-4.

[23] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*. 1st ed. Perseus Books, 2nd Oct. 1995. isbn: 978-0-201-50397-5.

[24] R. C. James. *Mathematics Dictionary*. 5th ed. Chapman & Hall, 31st July 1992. isbn: 978-0-412-99031-1.

[25] E. Noether. "Invariant Variation Problems". Trans. by M. A. Tavel. In: *Transport Theory and Statistical Physics* 1.3 (1971), p. 186. issn: 0041-1450. doi: 10.1080/00411457108231446. arXiv: physics/0503066 [physics]. Trans. of "Invariante Variationsprobleme". German. In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen: Mathematisch-Physikalische Klasse* 1918 (1918), p. 235. issn: 0369-6650.

[26] I. J. R. Aitchison and A. J. G. Hey. *Gauge Theories in Particle Physics: A Practical Introduction*. Vol. 1: *From Relativistic Quantum Mechanics to QED*. Ed. by D. F. Brewer. 3rd ed. 2 vols. Graduate Student Series in Physics. Institute of Physics Publishing, Sept. 2002. isbn: 978-0-7503-0864-9.

[27] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay". In: *Physical Review* 105 (15th Feb. 1957), p. 1413. issn: 0031-899X. doi: 10.1103/PhysRev.105.1413.

[28] J. W. Rohlf. *Modern Physics from α to Z^0*. 1st ed. John Wiley & Sons, Mar. 1994. isbn: 978-0-471-57270-1.

[29] S. L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nuclear Physics* 22 (Feb. 1961), p. 579. issn: 0029-5582. doi: 10.1016/0029-5582(61)90469-2.

[30] A. Salam and J. C. Ward. "Electromagnetic and weak interactions". In: *Physics Letters* 13 (15th Nov. 1964), p. 168. issn: 0031-9163. doi: 10.1016/0031-9163(64)90711-5.

[31] S. Weinberg. "A Model of Leptons". In: *Physical Review Letters* 19 (17th Oct. 1967), p. 1264. issn: 0031-9007. doi: 10.1103/PhysRevLett.19.1264.

[32] P. W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Physics Letters* 12 (15th Sept. 1964), p. 132. issn: 0031-9163. doi: 10.1016/0031-9163(64)91136-9.

[33] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Physical Review Letters* 13 (31st Aug. 1964), p. 321. issn: 0031-9007. doi: 10.1103/PhysRevLett.13.321.

[34] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Physical Review Letters* 13 (19th Oct. 1964), p. 508. issn: 0031-9007. doi: 10.1103/PhysRevLett.13.508.

[35] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Physical Review Letters* 13 (16th Nov. 1964), p. 585. issn: 0031-9007. doi: `10.1103/PhysRevLett.13.585`.

[36] P. W. Higgs. “Spontaneous Symmetry Breakdown without Massless Bosons”. In: *Physical Review* 145 (27th May 1966), p. 1156. issn: 0031-899X. doi: `10.1103/PhysRev.145.1156`.

[37] T. W. B. Kibble. “Symmetry breaking in non-Abelian gauge theories”. In: *Physical Review* 155 (25th Mar. 1967), p. 1554. issn: 0031-899X. doi: `10.1103/PhysRev.155.1554`.

[38] J. Ellis, M. K. Gaillard, and D. V. Nanopoulos. “A Historical Profile of the Higgs Boson”. In: *The standard theory of particle physics: Essays to celebrate CERN’s 60th anniversary.* Ed. by L. Maiani and L. Rolandi. Advanced series on directions in high energy physics 26. World Scientific, 18th Oct. 2016, p. 255. isbn: 978-981-4733-50-2. doi: `10.1142/9789814733519_0014`. arXiv: `1504.07217 [hep-ph]`.

[39] H. Yukawa. “On the Interaction of Elementary Particles I”. In: *Proceedings of the Physico-Mathematical Society of Japan* 17 (1935), p. 48. issn: 0370-1239. doi: `10.1143/PTPS.1.1`.

[40] N. Cabibbo. “Unitary Symmetry and Leptonic Decays”. In: *Physical Review Letters* 10 (15th June 1963), p. 531. issn: 0031-9007. doi: `10.1103/PhysRevLett.10.531`.

[41] M. Kobayashi and T. Maskawa. “CP Violation in the Renormalizable Theory of Weak Interaction”. In: *Progress of Theoretical Physics* 49 (1st Feb. 1973), p. 652. issn: 0033-068X. doi: `10.1143/PTP.49.652`.

[42] B. Pontecorvo. “Inverse beta processes and nonconservation of lepton charge”. In: *Journal of Experimental and Theoretical Physics* 7 (Jan. 1958), p. 172. issn: 1063-7761.

[43] Z. Maki, M. Nakagawa, and S. Sakata. “Remarks on the unified model of elementary particles”. In: *Progress of Theoretical Physics* 28 (1st Nov. 1962), p. 870. issn: 0033-068X. doi: `10.1143/PTP.28.870`.

[44] M. Gell-Mann. “A Schematic Model of Baryons and Mesons”. In: *Physics Letters* 8 (1st Feb. 1964), p. 214. issn: 0031-9163. doi: `10.1016/S0031-9163(64)92001-3`.

[45] M. Gell-Mann. “Symmetries of baryons and mesons”. In: *Physical Review* 125 (1st Feb. 1962), p. 1067. issn: 0031-899X. doi: `10.1103/PhysRev.125.1067`.

[46] M. Ablikim et al. "Observation of a Charged Charmoniumlike Structure in $e^+e^- \to \pi^+\pi^- J/\psi$ at $\sqrt{s} = 4.26\,\mathrm{GeV}$". In: *Physical Review Letters* 110 (17th June 2013), p. 252001. ISSN: 0031-9007. DOI: `10.1103/PhysRevLett.110.252001`. arXiv: `1303.5949 [hep-ex]`.

[47] Z. Q. Liu et al. "Study of $e^+e^- \to \pi^+\pi^- J/\psi$ and Observation of a Charged Charmoniumlike State at Belle". In: *Physical Review Letters* 110 (17th June 2013), p. 252002. ISSN: 0031-9007. DOI: `10.1103/PhysRevLett.110.252002`. arXiv: `1304.0121 [hep-ex]`.

[48] R. Aaij et al. "Observation of the resonant character of the $Z(4430)^-$ state". In: *Physical Review Letters* 112.22 (4th June 2014), p. 222002. ISSN: 0031-9007. DOI: `10.1103/PhysRevLett.112.222002`. arXiv: `1404.1903 [hep-ex]`.

[49] R. Aaij et al. "Observation of $J/\psi\phi$ structures consistent with exotic states from amplitude analysis of $B^+ \to J/\psi\phi K^+$ decays". In: *Physical Review Letters* 118.2 (11th Jan. 2017), p. 022003. ISSN: 0031-9007. DOI: `10.1103/PhysRevLett.118.022003`. arXiv: `1606.07895 [hep-ex]`.

[50] R. Aaij et al. "Amplitude analysis of $B^+ \to J/\psi\phi K^+$ decays". In: *Physical Review D: Particles, fields, gravitation, and cosmology* 95.1 (11th Jan. 2017), p. 012002. ISSN: 1550-2368. DOI: `10.1103/PhysRevD.95.012002`. arXiv: `1606.07898 [hep-ex]`.

[51] R. Aaij et al. "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays". In: *Physical Review Letters* 115 (12th Aug. 2015), p. 072001. ISSN: 0031-9007. DOI: `10.1103/PhysRevLett.115.072001`. arXiv: `1507.03414 [hep-ex]`.

[52] V. N. Gribov and L. N. Lipatov. "Deep inelastic ep scattering in perturbation theory". In: *Soviet Jornal of Nuclear Physics* 15 (1st Nov. 1972), p. 438. ISSN: 0038-5506.

[53] G. Altarelli and G. Parisi. "Asymptotic Freedom in Parton Language". In: *Nuclear Physics B: Proceedings Supplements* 126 (8th Aug. 1977), p. 298. ISSN: 0550-3213. DOI: `10.1016/0550-3213(77)90384-4`.

[54] Y. L. Dokshitzer. "Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics". In: *Journal of Experimental and Theoretical Physics* 46 (Oct. 1977), p. 641. ISSN: 1063-7761.

[55] J.-C. Winter, F. Krauss, and G. Soff. "A Modified cluster hadronization model". In: *European Physical Journal C: Particles and Fields* 36 (Aug. 2004), p. 381. ISSN: 1434-6044. DOI: `10.1140/epjc/s2004-01960-8`. arXiv: `hep-ph/0311085 [hep-ph]`.

[56] B. Andersson et al. “Parton Fragmentation and String Dynamics”. In: *Physics Reports* 97 (July 1983), p. 31. issn: 0370-1573. doi: `10.1016/0370-1573(83)90080-7`.

[57] E. Alvarez. “Quantum gravity: an introduction to some recent results”. In: *Reviews of Modern Physics* 61 (1st July 1989), p. 561. issn: 0034-6861. doi: `10.1103/RevModPhys.61.561`.

[58] A. Khrennikov. “The present situation in quantum theory and its merging with general relativity”. In: *Foundations of Physics* 47.8 (Aug. 2017), p. 1077. issn: 0015-9018. doi: `10.1007/s10701-017-0089-0`. arXiv: `1704.04679 [quant-ph]`.

[59] G. Steigman. “Observational tests of antimatter cosmologies”. In: *Annual Review of Astronomy and Astrophysics* 14 (1976), p. 339. issn: 0066-4146. doi: `10.1146/annurev.aa.14.090176.002011`.

[60] A. D. Sakharov. “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe”. In: *JETP Letters* 5 (10th Mar. 1967), p. 32. issn: 0021-3640. doi: `10.1070/PU1991v034n05ABEH002497`.

[61] A. Alavi-Harati et al. “Observation of direct CP violation in $K_{S,L} \to \pi\pi$ decays”. In: *Physical Review Letters* 83 (5th July 1999), p. 22. issn: 0031-9007. doi: `10.1103/PhysRevLett.83.22`. arXiv: `hep-ex/9905060 [hep-ex]`.

[62] V. Fanti et al. “A new measurement of direct CP violation in two pion decays of the neutral kaon”. In: *Physics Letters B* 465 (21st Oct. 1999), p. 335. issn: 0370-2693. doi: `10.1016/S0370-2693(99)01030-8`. arXiv: `hep-ex/9909022 [hep-ex]`.

[63] B. Aubert et al. “Measurement of CP violating asymmetries in B^0 decays to CP eigenstates”. In: *Physical Review Letters* 86 (19th Mar. 2001), p. 2515. issn: 0031-9007. doi: `10.1103/PhysRevLett.86.2515`. arXiv: `hep-ex/0102030 [hep-ex]`.

[64] K. Abe et al. “Observation of large CP violation in the neutral B meson system”. In: *Physical Review Letters* 87 (14th Aug. 2001), p. 091802. issn: 0031-9007. doi: `10.1103/PhysRevLett.87.091802`. arXiv: `hep-ex/0107061 [hep-ex]`.

[65] R. Aaij et al. “Observation of CP Violation in Charm Decays”. In: *Physical Review Letters* 122.21 (29th May 2019), p. 211803. issn: 0031-9007. doi: `10.1103/PhysRevLett.122.211803`. arXiv: `1903.08726 [hep-ex]`.

[66] P. van Dokkum et al. “A galaxy lacking dark matter”. In: *Nature* 555.7698 (29th Mar. 2018), p. 629. ISSN: 0028-0836. DOI: `10.1038/nature25767`. arXiv: `1803.10237 [astro-ph.GA]`.

[67] P. van Dokkum et al. “A Second Galaxy Missing Dark Matter in the NGC 1052 Group”. In: *The Astrophysical Journal Letters* 874.1 (20th Mar. 2019), p. L5. ISSN: 2041-8205. DOI: `10.3847/2041-8213/ab0d92`. arXiv: `1901.05973 [astro-ph.GA]`.

[68] V. C. Rubin and W. K. Ford Jr. “Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions”. In: *The Astrophysical Journal* 159 (Feb. 1970), p. 379. ISSN: 0004-637X. DOI: `10.1086/150317`.

[69] V. C. Rubin, N. Thonnard, and W. K. Ford Jr. “Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 ($R = 4\,\mathrm{kpc}$) to UGC 2885 ($R = 122\,\mathrm{kpc}$)”. In: *The Astrophysical Journal* 238 (June 1980), p. 471. ISSN: 0004-637X. DOI: `10.1086/158003`.

[70] G. Bertone and D. Hooper. “History of dark matter”. In: *Reviews of Modern Physics* 90.4 (15th Oct. 2018), p. 045002. ISSN: 0034-6861. DOI: `10.1103/RevModPhys.90.045002`. arXiv: `1605.04909 [astro-ph.CO]`.

[71] S. Perlmutter et al. “Measurements of Ω and Λ from 42 high redshift supernovae”. In: *The Astrophysical Journal* 517 (June 1999), p. 565. ISSN: 0004-637X. DOI: `10.1086/307221`. arXiv: `astro-ph/9812133 [astro-ph]`.

[72] P. J. E. Peebles and B. Ratra. “The Cosmological constant and dark energy”. In: *Reviews of Modern Physics* 75 (22nd Aug. 2003), p. 559. ISSN: 0034-6861. DOI: `10.1103/RevModPhys.75.559`. arXiv: `astro-ph/0207347 [astro-ph]`.

[73] S. W. Herb et al. “Observation of a Dimuon Resonance at 9.5 GeV in 400 GeV Proton–Nucleus Collisions”. In: *Physical Review Letters* 39 (1st Aug. 1977), p. 252. ISSN: 0031-9007. DOI: `10.1103/PhysRevLett.39.252`.

[74] F. Abe et al. “Observation of top quark production in $p\bar{p}$ collisions”. In: *Physical Review Letters* 74 (3rd Apr. 1995), p. 2626. ISSN: 0031-9007. DOI: `10.1103/PhysRevLett.74.2626`. arXiv: `hep-ex/9503002 [hep-ex]`.

[75] S. Abachi et al. “Search for high mass top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8\,\mathrm{TeV}$”. In: *Physical Review Letters* 74 (3rd Apr. 1995), p. 2422. ISSN: 0031-9007. DOI: `10.1103/PhysRevLett.74.2422`. arXiv: `hep-ex/9411001 [hep-ex]`.

[76] A. Quadt. *Top Quark Physics at Hadron Colliders*. Ed. by D. H. Beck, D. Haidt, and J. W. Negele. Advances in the Physics of Particles and Nuclei 28. Springer-Verlag, 2007. isbn: 978-3-540-71059-2.

[77] G. Degrassi et al. "Higgs mass and vacuum stability in the Standard Model at NNLO". In: *Journal of High Energy Physics* 08 (Aug. 2012), p. 098. issn: 1029-8479. doi: `10.1007/JHEP08(2012)098`. arXiv: `1205.6497 [hep-ph]`.

[78] M. S. Turner and F. Wilczek. "Might our vacuum be metastable?" In: *Nature* 298 (1982), p. 633. issn: 0028-0836. doi: `10.1038/298633a0`.

[79] A. M. Sirunyan 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 $\sqrt{s} = 13\,\text{TeV}$". In: *Journal of High Energy Physics* 2018.8 (3rd Aug. 2018), p. 011. issn: 1029-8479. doi: `10.1007/JHEP08(2018)011`. arXiv: `1711.02547 [hep-ex]`.

[80] A. M. Sirunyan et al. "Observation of $\mathrm{t\bar{t}H}$ production". In: *Physical Review Letters* 120.23 (4th June 2018), p. 231801. issn: 0031-9007. doi: `10.1103/PhysRevLett.120.231801`. arXiv: `1804.02610 [hep-ex]`.

[81] V. Khachatryan et al. "Search for s channel single top quark production in pp collisions at $\sqrt{s} = 7$ and $8\,\text{TeV}$". In: *Journal of High Energy Physics* 09 (Sept. 2016), p. 027. issn: 1029-8479. doi: `10.1007/JHEP09(2016)027`. arXiv: `1603.02555 [hep-ex]`.

[82] T. A. Aaltonen et al. "Observation of s-channel production of single top quarks at the Tevatron". In: *Physical Review Letters* 112 (9th June 2014), p. 231803. issn: 0031-9007. doi: `10.1103/PhysRevLett.112.231803`. arXiv: `1402.5126 [hep-ex]`.

[83] V. M. Abazov et al. "Model-independent measurement of t-channel single top quark production in $\mathrm{p\bar{p}}$ collisions at $\sqrt{s} = 1.96\,\text{TeV}$". In: *Physics Letters B* (17th Nov. 2011), p. 313. issn: 0370-2693. doi: `10.1016/j.physletb.2011.10.035`. arXiv: `1105.2788 [hep-ex]`.

[84] G. Aad et al. "Measurement of the t-channel single top-quark production cross section in ppcollisions at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector". In: *Physics Letters B* 717 (31st Oct. 2012), p. 330. issn: 0370-2693. doi: `10.1016/j.physletb.2012.09.031`. arXiv: `1205.3130 [hep-ex]`.

[85] S. Chatrchyan et al. "Measurement of the single-top-quark t-channel cross section in ppcollisions at $\sqrt{s} = 7\,\text{TeV}$". In: *Journal of High Energy Physics* 12 (Dec. 2012), p. 035. issn: 1029-8479. doi: `10.1007/JHEP12(2012)035`. arXiv: `1209.4533 [hep-ex]`.

[86] F. Krauss and D. Napoletano. “Towards a fully massive five-flavor scheme”. In: *Physical Review D: Particles, fields, gravitation, and cosmology* 98.9 (5th Nov. 2018), p. 096002. issn: 1550-2368. doi: `10.1103/PhysRevD.98.096002`. arXiv: `1712.06832 [hep-ph]`.

[87] S. Chatrchyan et al. “Observation of the associated production of a single top quark and a W boson in pp collisions at $\sqrt{s} = 8\,\text{TeV}$”. In: *Physical Review Letters* 112.23 (9th June 2014), p. 231802. issn: 0031-9007. doi: `10.1103/PhysRevLett.112.231802`. arXiv: `1401.2942 [hep-ex]`.

[88] F. Maltoni et al. “Associated production of Higgs and single top at hadron colliders”. In: *Physical Review D: Particles, fields, gravitation, and cosmology* 64 (11th Oct. 2001), p. 094023. issn: 1550-2368. doi: `10.1103/PhysRevD.64.094023`. arXiv: `hep-ph/0106293 [hep-ph]`.

[89] J. Campbell, R. K. Ellis, and R. Röntsch. “Single top production in association with a Z boson at the LHC”. In: *Physical Review D: Particles, fields, gravitation, and cosmology* 87 (5th June 2013), p. 114006. issn: 1550-2368. doi: `10.1103/PhysRevD.87.114006`. arXiv: `1302.3856 [hep-ph]`.

[90] CMS Collaboration. *Measurement of the Associated Production of Vector Bosons with Top–Antitop Pairs at* 7 TeV. Tech. rep. CMS-PAS-TOP-12-014. Geneva: CERN, 6th July 2012. url: `https://cds.cern.ch/record/1460101`.

[91] ATLAS Collaboration. *Search for* $\mathrm{t\bar{t}Z}$ *production in the three lepton final state with* $4.7\,\text{fb}^{-1}$ *of* $\sqrt{s} = 7\,\text{TeV}$ pp *collision data collected by the ATLAS detector*. Tech. rep. ATLAS-CONF-2012-126. Geneva: CERN, 27th Aug. 2012. url: `https://cds.cern.ch/record/1474643`.

[92] M. S. Sozzi. *Discrete Symmetries and CP Violation: From Experiment to Theory*. Oxford Graduate Texts. Oxford University Press, 1st Dec. 2007. isbn: 978-0-19-929666-8.

[93] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning. The MIT Press, Sept. 2012. isbn: 978-0-262-01802-9.

[94] S. Kullback and R. A. Leibler. “On Information and Sufficiency”. In: *The Annals of Mathematical Statistics* 22 (Sept. 1951), p. 79. issn: 0003-4851. doi: `10.1214/aoms/1177729694`.

[95] J. Neyman and E. S. Pearson. “The testing of statistical hypotheses in relation to probabilities a priori”. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 29 (Oct. 1933), p. 492. issn: 0305-0041. doi: `10.1017/S030500410001152X`.

[96] C. J. van Rijsbergen. *Information Retrieval.* 2nd ed. Butterworth–Heinemann, Mar. 1979. isbn: 978-0-408-70929-3.

[97] T. Fawcett. “An introduction to ROC analysis”. In: *Pattern Recognition Letters* 27.8 (June 2006), p. 861. issn: 0167-8655. doi: `10.1016/j.patrec.2005.10.010`.

[98] A. N. Kolmogorov. “On The Empirical Determination of A Distribution Law”. In: *Selected Works of A. N. Kolmogorov.* Vol. 2: *Probability Theory and Mathematical Statistics.* Ed. by A. N. Shiryaev. Trans. by G. Lindquist. Springer Science+Business Media, 1992, p. 139. Trans. of “Об Эмпирическом Определении Закона Распределения”. Russian. In: *Теория Вероятностей И Математическая Статистика.* Ed. by H. H. Bogolyubov et al. Comp. by A. N. Shiryaev. Наука, 1986, p. 134. Trans. of “Sulla Determinazione Empirica di una Legge di Distribuzione”. Italian. In: *Giornale dell’Istituto Italiano degli Attuari* 4 (Oct. 1933), p. 83. issn: 0390-5780.

[99] N. V. Smirnov. “Sur la distribution de ω^2 (Critérium de M.R. v. Mises)”. French. In: *Comptes rendus hebdomadaires des séances de l’Académie des Sciences* 202 (Jan. 1936), p. 449. issn: 0001-4036.

[100] J. F. Monahan. *Numerical Methods of Statistics.* 2nd ed. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Apr. 2011. isbn: 978-0-521-13951-9.

[101] R. Polikar. “Ensemble based systems in decision making”. In: *IEEE Circuits and Systems Magazine* 6 (Sept. 2006), p. 21. issn: 1531-636X. doi: `10.1109/MCAS.2006.1688199`.

[102] T. Chen and C. Guestrin. “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA). ACM, 13th Aug. 2016, p. 785. isbn: 978-1-4503-4232-2. doi: `10.1145/2939672.2939785`. arXiv: `1603.02754 [cs.LG]`.

[103] XGBoost Developers. *XGBoost Documentation.* 2016.

[104] D. Nielsen. “Tree Boosting With XGBoost: Why Does XGBoost Win “Every” Machine Learning Competition?” Thesis. Norwegian University of Science and Technology, Dec. 2016.

[105] J. H. Friedman. “Greedy function approximation: A gradient boosting machine”. In: *The Annals of Statistics* 29 (Nov. 2001). issn: 0090-5364. doi: `10.1214/aos/1013203451`.

[106] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed. Springer Series in Statistics. Springer, 1st Oct. 2009. isbn: 978-0-387-84857-0.

[107] R. Tibshirani. “Regression Shrinkage and Selection via the Lasso”. In: *Journal of the Royal Statistical Society. Series B: (Methodological)* 58 (1996), p. 267. issn: 0035-9246.

[108] A. E. Hoerl and R. W. Kennard. “Ridge Regression: Biased Estimation for Nonorthogonal Problems”. In: *Technometrics: a journal of statistics for the physical, chemical, and engineering sciences* 12 (Feb. 1970), p. 55. issn: 0040-1706. doi: `10.1080/00401706.1970.10488634`.

[109] H. Zou and T. Hastie. “Regularization and Variable Selection via the Elastic Net”. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 67.2 (2005), p. 301. issn: 1369-7412.

[110] J. H. Friedman. “Stochastic Gradient Boosting”. In: *Computational Statistics & Data Analysis* 38 (Feb. 2002), p. 367. issn: 0167-9473. doi: `10.1016/S0167-9473(01)00065-2`.

[111] W. S. McCulloch and W. Pitts. “A logical calculus of the ideas immanent in nervous activity”. In: *The Bulletin of Mathematical Biophysics* 5 (1st Dec. 1943), p. 115. issn: 0007-4985. doi: `10.1007/BF02478259`.

[112] M. A. Nielsen. *Neural Networks and Deep Learning.* Determination Press, 2015. url: `http://neuralnetworksanddeeplearning.com/` (visited on 09/02/2019).

[113] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning.* MIT Press, 2016. url: `http://www.deeplearningbook.org` (visited on 11/02/2019).

[114] X. Glorot and Y. Bengio. “Understanding the difficulty of training deep feedforward neural networks”. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Chia Laguna Resort, Sardinia, Italy, 13th May 2010). Ed. by Y. W. Teh and M. Titterington. Vol. 9. Proceedings of Machine Learning Research. PMLR, 31st Mar. 2010, p. 249.

[115] D. P. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization”. In: *Computing Research Repository* (Dec. 2014). arXiv: `1412.6980 [cs.LG]`. Pre-published.

[116] N. Srivastava et al. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *Journal of Machine Learning Research* 15 (June 2014), p. 1929. issn: 1532-4435.

[117] N. Buduma and N. Locascio. *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*. 1st ed. O'Reilly Media, Nov. 2015. isbn: 978-1-4919-2561-4.

[118] D. Harrison and D. L. Rubinfeld. "Hedonic housing prices and the demand for clean air". In: *Journal of Environmental Economics and Management* 5 (1978), p. 81. issn: 0095-0696. doi: `10.1016/0095-0696(78)90006-2`.

[119] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. The MIT Press, Jan. 2006. isbn: 978-0-262-18253-9.

[120] D. K. Duvenaud. "Automatic Model Constructionwith Gaussian Processes". Thesis. University of Cambridge, June 2014. doi: `10.17863/CAM.14087`.

[121] D. J. C. MacKay. "Introduction to Gaussian processes". In: *NATO ASI Series F: Computer and Systems Sciences* 168 (1998), p. 133. issn: 0258-1248.

[122] B. Matérn. *Spatial Variation*. Ed. by D. Brillinger et al. 2nd ed. Lecture Notes in Statistics. Springer-Verlag, 12th Dec. 1986. isbn: 978-0-387-96365-5. doi: `10.1007/978-1-4615-7892-5`.

[123] M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. 1st ed. Springer Series in Statistics. Springer-Verlag, 22nd June 1999. isbn: 978-0-387-98629-6.

[124] S. Holmes, R. S. Moore, and V. Shiltsev. "Overview of the Tevatron Collider Complex: Goals, Operations and Performance". In: *Journal of Instrumentation* 6 (5th June 2011), p. T08001. issn: 1748-0221. doi: `10.1088/1748-0221/6/08/T08001`. arXiv: `1106.0909 [physics.acc-ph]`.

[125] E. Mobs. "The CERN accelerator complex – August 2018. Complexe des accélérateurs du CERN – Août 2018". 28th Aug. 2018. url: `https://cds.cern.ch/record/2636343` (visited on 10/01/2018).

[126] O. S. Brüning et al. *LHC Design Report*. Vol. 1: *The LHC Main Ring*. 3 vols. CERN Yellow Reports: Monographs. CERN, 4th June 2004. isbn: 978-92-9083-224-9. doi: `10.5170/CERN-2004-003-V-1`. url: `https://cds.cern.ch/record/782076`.

[127] M. Bajko et al. *Report of the task force on the incident of 19th September 2008 at the LHC*. Project report. CERN, 31st Mar. 2009.

[128] *CERN announces start-up date for LHC*. CERN. 11th Aug. 2008. url: `https://cds.cern.ch/record/1194443` (visited on 07/09/2018).

[129] CERN. *LHC physics data taking gets underway at new record collision energy of* 8 TeV. 4th May 2012. URL: `https://cds.cern.ch/record/1989553` (visited on 14/02/2019).

[130] C. Pralavorio. *LHC performance reaches new highs.* 2016. URL: `https://home.cern/about/updates/2016/07/lhc-performance-reaches-new-highs` (visited on 06/08/2018).

[131] S. Chatrchyan et al. "Commissioning of the CMS Experiment and the Cosmic Run at Four Tesla". In: *Journal of Instrumentation* 5 (26th Nov. 2010), p. T03001. ISSN: 1748-0221. DOI: `10.1088/1748-0221/5/03/T03001`. arXiv: `0911.4845 [physics.ins-det]`.

[132] CMS Collaboration. "Detector Drawings". CMS Collection. Mar. 2012. URL: `https://cds.cern.ch/record/1433717`.

[133] S. Chatrchyan et al. *Technical proposal for the upgrade of the CMS detector through 2020.* Tech. rep. CERN-LHCC-2011-006. LHCC-P-004. CERN, 2nd June 2011. URL: `https://cds.cern.ch/record/1355706`.

[134] P. J. Lujan. *CMS Luminosity – Public Results.* 10th Sept. 2018. URL: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults` (visited on 30/09/2018).

[135] S. Chatrchyan et al. "The performance of the CMS muon detector in proton–proton collisions at $\sqrt{s} = 7$ TeV at the LHC". In: *Journal of Instrumentation* 8 (23rd June 2013), p. P11002. ISSN: 1748-0221. DOI: `10.1088/1748-0221/8/11/P11002`. arXiv: `1306.6905 [physics.ins-det]`.

[136] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC". In: *Journal of Instrumentation* 3 (Aug. 2008), p. S08004. ISSN: 1748-0221. DOI: `10.1088/1748-0221/3/08/S08004`.

[137] S. K. Park et al. "CMS end cap RPC gas gap production for upgrade". In: *Journal of Instrumentation* 7 (15th Nov. 2012), p. P11013. ISSN: 1748-0221. DOI: `10.1088/1748-0221/7/11/P11013`.

[138] A. M. Sirunyan et al. "Performance of the CMS muon detector and muon reconstruction with proton–proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of Instrumentation* 13.06 (19th June 2018), p. P06015. ISSN: 1748-0221. DOI: `10.1088/1748-0221/13/06/P06015`. arXiv: `1804.04528 [physics.ins-det]`.

[139] I. M. Frank and I. E. Tamm. “Coherent Visible Radiation of Fast Electrons Passing Through Matter”. In: *Selected Papers*. Ed. by B. M. Bolotovskii and V. Y. Frenkel. In collab. with R. Peierls. Springer-Verlag, 1991, p. 29. isbn: 978-3-642-74626-0. doi: `10.1007/978-3-642-74626-0_2`. Trans. of “Когерентное излучение быстрого электрона в среде”. Russian. In: *Доклады Академии Наук СССР* 14.3 (1937), p. 107. issn: 0002-3264. doi: `10.3367/UFNr.0093.196710o.0388`.

[140] S. Chatrchyan et al. “Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data”. In: *Journal of Instrumentation* 5 (9th Mar. 2010), p. T03012. issn: 1748-0221. doi: `10.1088/1748-0221/5/03/T03012`. arXiv: `0911.4991 [physics.ins-det]`.

[141] S. Chatrchyan et al. “Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data”. In: *Journal of Instrumentation* 5 (19th Mar. 2010), p. T03012. issn: 1748-0221. doi: `10.1088/1748-0221/5/03/T03012`. arXiv: `0911.4991 [physics.ins-det]`.

[142] V. Khachatryan et al. “Performance of Photon Reconstruction and Identification with the CMS Detector in Proton–Proton Collisions at $\sqrt{s} = 8$ TeV”. In: *Journal of Instrumentation* 10.08 (15th Aug. 2015), p. P08010. issn: 1748-0221. doi: `10.1088/1748-0221/10/08/P08010`. arXiv: `1502.02702 [physics.ins-det]`.

[143] S. Chatrchyan et al. “Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV”. In: *Journal of Instrumentation* 8 (2013), p. P09009. issn: 1748-0221. doi: `10.1088/1748-0221/8/09/P09009`. arXiv: `1306.2016 [hep-ex]`.

[144] J. R. Hook and E. Hall. *Solid State Physics*. 2nd ed. Manchester Physics Series. John Wiley & Sons, July 1991. isbn: 978-0-471-92805-8.

[145] T. Sakuma and T. McCauley. “Detector and Event Visualization with SketchUp at the CMS Experiment”. In: *Journal of Physics: Conference Series* 513.2 (19th Nov. 2014): *20th International Conference on Computing in High Energy and Nuclear Physics (CHEP2013)*, p. 022032. issn: 1742-6588. doi: `10.1088/1742-6596/513/2/022032`. arXiv: `1311.4942 [physics.ins-det]`.

[146] A. Dominguez et al. *CMS Technical Design Report for the Pixel Detector Upgrade*. Tech. rep. CMS-TDR-011. CERN, 26th Sept. 2012. doi: `10.2172/1151650`.

[147] K. Klein. “The Phase-1 upgrade of the CMS pixel detector”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 845 (11th Mar. 2016): *Proceedings of the 14th Vienna Conference on Instrumentation*, p. 101. issn: 0168-9002. doi: `10.1016/j.nima.2016.06.039`.

[148] L. Caldas. *Commissioning and first results from the CMS phase-1 upgrade pixel detector*. Tech. rep. arXiv:1807.08987. CERN, 20th Nov. 2017. url: `https://cds.cern.ch/record/2297549`.

[149] S. Chatrchyan et al. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (2014), p. P10009. issn: 1748-0221. doi: `10.1088/1748-0221/9/10/P10009`. arXiv: `1405.6569 [physics.ins-det]`.

[150] G. L. Bayatian et al. *CMS Physics*. Techincal Design Report. Vol. 1: *Detector Performance and Software*. 2 vols. CERN, 2006. isbn: 978-92-9083-268-3. url: `https://cds.cern.ch/record/922757`.

[151] G. Antchev et al. “First measurement of elastic, inelastic and total cross-section at $\sqrt{s} = 13\,\text{TeV}$ by TOTEM and overview of cross-section data at LHC energies”. In: *European Physical Journal C: Particles and Fields* 79.2 (2019), p. 103. issn: 1434-6044. doi: `10.1140/epjc/s10052-019-6567-0`. arXiv: `1712.06153 [hep-ex]`.

[152] V. Khachatryan et al. “The CMS trigger system”. In: *Journal of Instrumentation* 12.01 (24th Jan. 2017), p. P01020. issn: 1748-0221. doi: `10.1088/1748-0221/12/01/P01020`. arXiv: `1609.02366 [physics.ins-det]`.

[153] A. M. Sirunyan et al. “Particle-flow reconstruction and global event description with the CMS detector”. In: *Journal of Instrumentation* 12.10 (6th Oct. 2017), p. P10003. issn: 1748-0221. doi: `10.1088/1748-0221/12/10/P10003`. arXiv: `1706.04965 [physics.ins-det]`.

[154] R. E. Kálmán. “A New Approach to Linear Filtering and Prediction Problems”. In: *Journal of Basic Engineering* 82.1 (1st Mar. 1980), p. 34. issn: 0021-9223. doi: `10.1115/1.3662552`.

[155] R. Frühwirth. “Application of Kalman filtering to track and vertex fitting”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262 (15th Dec. 1987), p. 444. issn: 0168-9002. doi: `10.1016/0168-9002(87)90887-4`.

[156] P. Billoir. “Progressive track recognition with a Kalman-like fitting procedure”. In: *Computer Physics Communications* 57 (2nd Dec. 1989), p. 390. issn: 0010-4655. doi: `10.1016/0010-4655(89)90249-X`.

[157] K. Rose. "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems". In: *Proceedings of the IEEE* 86.11 (11th Nov. 1998), p. 2210. ISSN: 0018-9219. DOI: `10.1109/5.726788`.

[158] R. Frühwirth, W. Waltenberger, and P. Vanlaer. "Adaptive vertex fitting". In: *Journal of Physics G: Nuclear and Particle Physics* 34 (16th Nov. 2007), p. N343. ISSN: 0954-3899. DOI: `10.1088/0954-3899/34/12/N01`.

[159] A. Bodek et al. "Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments". In: *European Physical Journal C: Particles and Fields* 72 (Oct. 2012), p. 2194. ISSN: 1434-6044. DOI: `10.1140/epjc/s10052-012-2194-8`. arXiv: `1208.3710 [hep-ex]`.

[160] V. Khachatryan et al. "Performance of electron reconstruction and selection with the CMS detector in proton–proton collisions at $\sqrt{s} = 8\,\mathrm{TeV}$". In: *Journal of Instrumentation* 10 (10th June 2015), p. P06005. ISSN: 1748-0221. DOI: `10.1088/1748-0221/10/06/P06005`. arXiv: `1502.02701 [physics.ins-det]`.

[161] W. Adam et al. "Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC". In: *Proceedings, 13th International Conference on Computing in High-Enery and Nuclear Physics.* CHEP 2003 (La Jolla, California). 24th Mar. 2003. DOI: `10.1088/0954-3899/31/9/N01`. arXiv: `physics/0306087 [physics.data-an]`.

[162] S. D. Ellis and D. E. Soper. "Successive combination jet algorithm for hadron collisions". In: *Physical Review D: Particles, fields, gravitation, and cosmology* 48 (1st Oct. 1993), p. 3160. ISSN: 1550-2368. DOI: `10.1103/PhysRevD.48.3160`. arXiv: `hep-ph/9305266 [hep-ph]`.

[163] M. Cacciari, G. P. Salam, and G. Soyez. "The anti-k_t jet clustering algorithm". In: *Journal of High Energy Physics* 04 (16th Apr. 2008), p. 063. ISSN: 1029-8479. DOI: `10.1088/1126-6708/2008/04/063`. arXiv: `0802.1189 [hep-ph]`.

[164] M. Wobisch and T. Wengler. "Hadronization corrections to jet cross-sections in deep inelastic scattering". In: *Monte Carlo generators for HERA physics.* Proceedings of the Workshop 1998–1999. Workshop on Monte Carlo Generators for HERA Physics (DESY, Hamburg, Germany, 27th Apr. 1998). Ed. by A. Doyle et al. 25th Oct. 1999, p. 270. arXiv: `hep-ph/9907280 [hep-ph]`.

[165] V. Khachatryan et al. “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”. In: *Journal of Instrumentation* 12.02 (22nd Feb. 2017), p. P02014. issn: 1748-0221. doi: `10.1088/1748-0221/12/02/P02014`. arXiv: `1607.03663 [hep-ex]`.

[166] V. Khachatryan et al. “Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV”. In: *Journal of High Energy Physics* 3 (28th Mar. 2011), p. 136. issn: 1029-8479. doi: `10.1007/JHEP03(2011)136`. arXiv: `1102.3194 [hep-ex]`.

[167] A. M. Sirunyan et al. “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”. In: *Journal of Instrumentation* 13.05 (8th May 2018), p. P05011. issn: 1748-0221. doi: `10.1088/1748-0221/13/05/P05011`. arXiv: `1712.07158 [physics.ins-det]`.

[168] S. Chatrchyan et al. “Missing transverse energy performance of the CMS detector”. In: *Journal of Instrumentation* 6 (9th Sept. 2011), p. P09001. issn: 1748-0221. doi: `10.1088/1748-0221/6/09/P09001`. arXiv: `1106.5048 [physics.ins-det]`.

[169] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo Method.* 3rd ed. Wiley Series in Probability and Statistics. John Wiley & Sons, Nov. 2016. isbn: 978-1-118-63216-1.

[170] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *Journal of High Energy Physics* 07 (17th July 2014), p. 079. issn: 1029-8479. doi: `10.1007/JHEP07(2014)079`. arXiv: `1405.0301 [hep-ph]`.

[171] S. Frixione, P. Nason, and C. Oleari. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”. In: *Journal of High Energy Physics* 11 (23rd Nov. 2007), p. 070. issn: 1029-8479. doi: `10.1088/1126-6708/2007/11/070`. arXiv: `0709.2092 [hep-ph]`.

[172] S. Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the Powheg Box”. In: *Journal of High Energy Physics* 06 (9th June 2010), p. 043. issn: 1029-8479. doi: `10.1007/JHEP06(2010)043`. arXiv: `1002.2581 [hep-ph]`.

[173] P. Nason. “A New method for combining NLO QCD with shower Monte Carlo algorithms”. In: *Journal of High Energy Physics* 11 (12th Oct. 2004), p. 040. issn: 1029-8479. doi: `10.1088/1126-6708/2004/11/040`. arXiv: `hep-ph/0409146 [hep-ph]`.

[174] T. Sjöstrand et al. "An Introduction to Pythia 8.2". In: *Computer Physics Communications* 191 (11th Feb. 2015), p. 159. issn: 0010-4655. doi: `10.1016/j.cpc.2015.01.024`. arXiv: `1410.3012 [hep-ph]`.

[175] J. Alwall et al. "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions". In: *European Physical Journal C: Particles and Fields* 53 (15th Dec. 2007), p. 473. issn: 1434-6044. doi: `10.1140/epjc/s10052-007-0490-5`. arXiv: `0706.2569 [hep-ph]`.

[176] J. Alwall, S. de Visscher, and F. Maltoni. "QCD radiation in the production of heavy colored particles at the LHC". In: *Journal of High Energy Physics* 02 (5th Feb. 2009), p. 017. issn: 1029-8479. doi: `10.1088/1126-6708/2009/02/017`. arXiv: `0810.5350 [hep-ph]`.

[177] R. Frederix and S. Frixione. "Merging meets matching in MC@NLO". In: *Journal of High Energy Physics* 12 (12th Dec. 2012), p. 061. issn: 1029-8479. doi: `10.1007/JHEP12(2012)061`. arXiv: `1209.6215 [hep-ph]`.

[178] S. Agostinelli et al. "Geant4—A Simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506 (1st July 2003), p. 250. issn: 0168-9002. doi: `10.1016/S0168-9002(03)01368-8`.

[179] J. Allison et al. "Geant4 developments and applications". In: *IEEE Transactions on Nuclear Science* 53 (27th Mar. 2006), p. 270. issn: 0018-9499. doi: `10.1109/TNS.2006.869826`.

[180] J. Allison et al. "Recent developments in Geant4". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 835 (1st Nov. 2016), p. 186. issn: 0168-9002. doi: `10.1016/j.nima.2016.06.125`.

[181] S. Banerjee. "CMS simulation software". In: *Journal of Physics: Conference Series* 396 (19th Dec. 2012): *19th International Conference on Computing in High Energy and Nuclear Physics (CHEP2012)*, p. 022003. issn: 1742-6588. doi: `10.1088/1742-6596/396/2/022003`.

[182] M. Czakon, D. Heymes, and A. Mitov. "High-precision differential predictions for top-quark pairs at the LHC". In: *Physical Review Letters* 116.8 (25th Feb. 2016), p. 082003. issn: 0031-9007. doi: `10.1103/PhysRevLett.116.082003`. arXiv: `1511.00549 [hep-ph]`.

[183] V. Khachatryan et al. “Measurement of the $\mathrm{t\bar{t}}$ production cross section using events in the eμ final state in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$”. In: *European Physical Journal C: Particles and Fields* 77 (2017), p. 172. ISSN: 1434-6044. DOI: `10.1140/epjc/s10052-017-4718-8`. arXiv: `1611.04040 [hep-ex]`.

[184] A. D. J. Morton. “A search for the production of a single top quark in association with a Z Boson at the LHC”. Thesis. Brunel University London, 2019. In press.

[185] F. Fiori. “CMS Tracker operational experience”. In: *Proceedings of Science* 287 (3rd Aug. 2017): *Proceedings, 25th International Workshop on Vertex Detectors (Vertex 2016)*, p. 6. ISSN: 1824-8039. DOI: `10.22323/1.287.0006`.

[186] V. Khachatryan et al. “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton–Proton Collisions at $\sqrt{s} = 8\,\mathrm{TeV}$”. In: *Journal of Instrumentation* 10.06 (10th June 2015), p. P06005. ISSN: 1748-0221. DOI: `10.1088/1748-0221/10/06/P06005`. arXiv: `1502.02701 [physics.ins-det]`.

[187] S. Chatrchyan et al. “Performance of CMS Muon Reconstruction in pp Collision Events at $\sqrt{s} = 8\,\mathrm{TeV}$”. In: *Journal of Instrumentation* 7 (5th Oct. 2012), p. P10002. ISSN: 1748-0221. DOI: `10.1088/1748-0221/7/10/P10002`. arXiv: `1206.4071 [physics.ins-det]`.

[188] H. Jansen. “Study of Unparticle plus Lepton Signatures at CMS”. Thesis. Rheinisch-Westfälische Technische Hochschule Aachen, Oct. 2009.

[189] CMS Collaboration. *Jet algorithms performance in* 13 TeV *data*. Tech. rep. CMS-PAS-JME-16-003. CERN, 24th Mar. 2017. URL: `https://cds.cern.ch/record/2256875`.

[190] M. Aaboud et al. “Search for pair production of Higgs bosons in the $\mathrm{b\bar{b}b\bar{b}}$ final state using proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector”. In: *Physical Review D: Particles, fields, gravitation, and cosmology* 94.5 (2nd Sept. 2016), p. 052002. ISSN: 1550-2368. DOI: `10.1103/PhysRevD.94.052002`. arXiv: `1606.04782 [hep-ex]`.

[191] V. Khachatryan et al. “Search for resonant pair production of Higgs bosons decaying to two bottom quark–antiquark pairs in proton–proton collisions at 8 TeV”. In: *Physics Letters B* 749 (7th Oct. 2015), p. 560. ISSN: 0370-2693. DOI: `10.1016/j.physletb.2015.08.047`. arXiv: `1503.04114 [hep-ex]`.

[192] CMS Collaboration. *Measurement of the top quark pair production cross section using* eμ *events in proton–proton collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the CMS detector.* Tech. rep. CMS-PAS-TOP-16-005. CERN, 26th Mar. 2016. URL: `https://cds.cern.ch/record/2141738`.

[193] CMS Collaboration. *Search for SUSY in same-sign dilepton events at* $\sqrt{s} = 13\,\mathrm{TeV}$. Tech. rep. CMS-PAS-SUS-15-008. CERN, 15th Dec. 2015. URL: `https://cds.cern.ch/record/2114813`.

[194] CMS Collaboration. *CMS Luminosity Measurements for the 2016 Data Taking Period.* Tech. rep. CMS-PAS-LUM-17-001. CERN, 26th Mar. 2017. URL: `https://cds.cern.ch/record/2257069`.

[195] CMS Collaboration. *CMS luminosity measurement for the 2017 data-taking period at* $\sqrt{s} = 13\,\mathrm{TeV}$. Tech. rep. CMS-PAS-LUM-17-004. CERN, 6th June 2018. URL: `https://cds.cern.ch/record/2621960`.

[196] A. M. Sirunyan et al. “Measurement of the associated production of a single top quark and a Z boson in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$”. In: *Physics Letters B* 779 (10th Apr. 2018), p. 358. ISSN: 0370-2693. DOI: `10.1016/j.physletb.2018.02.025`. arXiv: `1712.02825 [hep-ex]`.

[197] A. Buckley et al. “LHAPDF6: parton density access in the LHC precision era”. In: *European Physical Journal C: Particles and Fields* 75 (Mar. 2015), p. 132. ISSN: 1434-6044. DOI: `10.1140/epjc/s10052-015-3318-8`. arXiv: `1412.7420 [hep-ph]`.

[198] R. D. Ball et al. “Parton distributions for the LHC Run II”. In: *Journal of High Energy Physics* 04 (Apr. 2015), p. 040. ISSN: 1029-8479. DOI: `10.1007/JHEP04(2015)040`. arXiv: `1410.8849 [hep-ph]`.

[199] J. Butterworth et al. “PDF4LHC recommendations for LHC Run II”. In: *Journal of Physics G: Nuclear and Particle Physics* (6th Jan. 2016), p. 023001. ISSN: 0954-3899. DOI: `10.1088/0954-3899/43/2/023001`. arXiv: `1510.03865 [hep-ph]`.

[200] CMS Collaboration. *Investigations of the impact of the parton shower tuning in Pythia8 in the modelling of* $\mathrm{t\bar{t}}$ *at* $\sqrt{s} = 8$ *and* $13\,\mathrm{TeV}$. Tech. rep. CMS-PAS-TOP-16-021. CERN, 24th Nov. 2016. URL: `https://cds.cern.ch/record/2235192`.

[201] B. Krawczyk. “Learning from imbalanced data: open challenges and future directions”. In: *Progress in Artificial Intelligence* 5.4 (1st Nov. 2016), p. 221. ISSN: 2192-6352. DOI: `10.1007/s13748-016-0094-0`.

[202] I. Guyon et al. "Gene Selection for Cancer Classification using Support Vector Machines". In: *Machine Learning* 46.1 (1st Jan. 2002), p. 389. issn: 0885-6125. doi: `10.1023/A:1012487302797`.

[203] J. Bergstra and Y. Bengio. "Random Search for Hyper-Parameter Optimization". In: *Journal of Machine Learning Research* 13 (Feb. 2012), p. 281. issn: 1532-4435.

[204] T. Head et al. "scikit-optimize/scikit-optimize: v0.5.2". 25th Mar. 2018. doi: `10.5281/zenodo.1207017`.

[205] G. Louppe. "Bayesian optimisation with Scikit-Optimize". Talk Presented at PyData 2017, Amsterdam. 8th Apr. 2017. url: `https://youtu.be/DGJTEBt0d-s`.

[206] E. Brochu, V. M. Cora, and N. de Freitas. *A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning.* 14th Dec. 2010. arXiv: `1012.2599 [cs.LG]`.

[207] J. Toutouh, J. Arellano, and E. Alba. "BiPred: A Bilevel Evolutionary Algorithm for Prediction in Smart Mobility". In: *Sensors* 18.12 (2018). issn: 1424-8220. doi: `10.3390/s18124123`. url: `https://www.mdpi.com/1424-8220/18/12/4123`.

[208] F. Mosteller and J. Tukey. "Data analysis, including statistics". In: *Handbook of Social Psychology.* Ed. by G. Lindzey and E. Aronson. Vol. 2. 5 vols. Addison-Wesley, 1968. isbn: 978-0-201-04263-4.

[209] F. Chollet et al. "Keras". `https://keras.io`. 2015.

[210] R. Al-Rfou et al. *Theano: A Python framework for fast computation of mathematical expressions.* May 2016. arXiv: `1605.02688 [cs.SC]`.

[211] T. Li et al. "Adaptive Scaling". In: *The Annals of Applied Statistics* (2nd Sept. 2017). issn: 1932-6157. arXiv: `1709.00566 [stat.ML]`. Submitted.

[212] X. Glorot, A. Bordes, and Y. Bengio. "Deep Sparse Rectifier Neural Networks". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics.* Ed. by G. Gordon, D. Dunson, and M. Dudík. Vol. 15. Proceedings of Machine Learning Research. PMLR, 11th Apr. 2011, p. 315. url: `http://proceedings.mlr.press/v15/glorot11a.html`.

[213] L. Moneta et al. “The RooStats Project”. In: *Proceedings of Science* 93 (24th Feb. 2011): *Proceedings, 13th International Workshop on Advanced computing and analysis techniques in physics research (ACAT2010)*, p. 057. issn: 1824-8039. doi: `10.22323/1.093.0057`. arXiv: `1009.1003 [physics.data-an]`.

[214] The ATLAS Collaboration, The CMS Collaboration, and The LHC Higgs Combination Group. *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. CMS-NOTE-2011-005. CERN, 16th Aug. 2011. url: `https://cds.cern.ch/record/1379837`.

[215] J. S. Conway. “Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra”. In: *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding* (CERN, Geneva, Switzerland, 17th Jan. 2011). 3rd Apr. 2011, p. 115. doi: `10.5170/CERN-2011-006.115`. arXiv: `1103.0354 [physics.data-an]`.

[216] R. J. Barlow and C. Beeston. “Fitting using finite Monte Carlo samples”. In: *Computer Physics Communications* 77 (Oct. 1993), p. 219. issn: 0010-4655. doi: `10.1016/0010-4655(93)90005-W`.

[217] G. Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *European Physical Journal C: Particles and Fields* 71 (Feb. 2011), p. 1554. issn: 1434-6044. doi: `10.1140/epjc/s10052-011-1554-0`. Corrected in “Erratum to: Asymptotic formulae for likelihood-based tests of new physics”. In: *European Physical Journal C: Particles and Fields* 73 (July 2013), p. 2501. issn: 1434-6044. doi: `10.1140/epjc/s10052-013-2501-z`. arXiv: `1007.1727 [physics.data-an]`.

[218] M. Hatlo et al. “Developments of mathematical software libraries for the LHC experiments”. In: *IEEE Transactions on Nuclear Science* 52 (Dec. 2005), p. 2818. issn: 0018-9499. doi: `10.1109/TNS.2005.860152`.

[219] A. M. Sirunyan et al. “Search for associated production of a Z boson with a single top quark and for tZ flavour-changing interactions in pp collisions at $\sqrt{s} = 8\,\text{TeV}$”. In: *Journal of High Energy Physics* 07 (July 2017), p. 003. issn: 1029-8479. doi: `10.1007/JHEP07(2017)003`. arXiv: `1702.01404 [hep-ex]`.

[220] A. M. Sirunyan et al. “Observation of single top quark production in association with a Z boson in pp collisions at $\sqrt{s} = 13\,\text{TeV}$”. In: *Physical Review Letters* 122.13 (5th Apr. 2019), p. 132003. issn: 0031-9007. doi: `10.1103/PhysRevLett.122.132003`. arXiv: `1812.05900 [hep-ex]`.

[221] M. Aaboud et al. "Measurement of the production cross-section of a single top quark in association with a Z boson in pp collisions at 13 TeV with the ATLAS detector". In: *Physics Letters B* 780 (10th May 2018), p. 557. issn: 0370-2693. doi: `10.1016/j.physletb.2018.03.023`. arXiv: `1710.03659 [hep-ex]`.

[222] R. S. Olson et al. "Automating Biomedical Data Science Through Tree-Based Pipeline Optimization". In: *Applications of Evolutionary Computation.* EvoApplications. 19th European Conference (Porto, Portugal, 30th Mar. 2016). Ed. by G. Squillero and P. Burelli. Springer International Publishing, 2016, p. 123. isbn: 978-3-319-31204-0. doi: `10.1007/978-3-319-31204-0_9`. arXiv: `1601.07925 [cs.NE]`.

[223] The CMS Collaboration. *Heavy flavor identification at CMS with deep neural networks.* Tech. rep. CMS-DP-2017-005. CERN, 13th Mar. 2017. url: `https://cds.cern.ch/record/2255736`.

[224] D. Guest et al. "Jet Flavor Classification in High-Energy Physics with Deep Neural Networks". In: *Physical Review D: Particles, fields, gravitation, and cosmology* 94.11 (2nd Dec. 2016), p. 112002. issn: 1550-2368. doi: `10.1103/PhysRevD.94.112002`. arXiv: `1607.08633 [hep-ex]`.

[225] R. Brun and F. Rademakers. "Root: An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* (11th Apr. 1997): *New Computing Techniques in Physics Research V*, p. 81. issn: 0168-9002. doi: `10.1016/S0168-9002(97)00048-X`.

[226] J. D. Hunter. "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3 (18th June 2007), p. 90. issn: 1521-9615. doi: `10.1109/MCSE.2007.55`.

[227] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (12th Oct. 2011), p. 2825. issn: 1532-4435. arXiv: `1201.0490 [cs.LG]`.

[228] L. Buitinck et al. "API design for machine learning software: Experiences from the Scikit-learn project". In: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning.* 1st Sept. 2013, p. 108. arXiv: `1309.0238 [cs.MS]`.

[229] T. E. Oliphant. "Python for Scientific Computing". In: *Computing in Science & Engineering* 9.3 (May 2007), p. 10. issn: 1521-9615. doi: `10.1109/MCSE.2007.58`.

[230] K. J. MILLMAN and M. AIVAZIS. “Python for Scientists and Engineers”. In: *Computing in Science & Engineering* 13 (Mar. 2011), p. 9. ISSN: 1521-9615. DOI: `10.1109/MCSE.2011.36`.

[231] S. van der WALT, S. C. COLBERT, and G. VAROQUAUX. “The NUMPY Array: A Structure for Efficient Numerical Computation”. In: *Comput. Sci. Eng.* 13.2 (7th Mar. 2011), p. 22. DOI: `10.1109/MCSE.2011.37`. arXiv: `1102.1523 [cs.MS]`.

[232] T. E. OLIPHANT. *Guide to NUMPY*. 2nd ed. Manchester Physics Series. On-Demand Publishing, 15th Aug. 2008. ISBN: 978-0-470-03294-7.

[233] E. JONES, T. OLIPHANT, P. PETERSON, et al. *SCIPY: Open source scientific tools for Python*. 2001. URL: `http://www.scipy.org/` (visited on 22/09/2019).

[234] W. MCKINNEY. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. Ed. by S. van der WALT and J. MILLMAN. 2010, p. 51.

[235] E. RODRIGUES. “The SCIKIT-HEP Project”. In: *Proceedings of the 23rd International Conference on Computing in High Energy and Nuclear Physics* (Sofia, Bulgaria, 13th July 2018). 2019. arXiv: `1905.00002 [physics.comp-ph]`.

[236] T. OHL. *feynMF: Drawing Feynman Diagrams with LaTeX and METAFONT*. Version 1.08. Technische Universität Darmstadt. 2nd Dec. 1996.

[237] T. TANTAU. *The Tikz and PGF Packages: Manual for version 3.1.4b*. Version 3.1.4b. Institut für Theoretische Informatik. 2nd Dec. 1996.

[238] C. FEUERSÄNGER. *Manual for Package PGFPLOTS: 2D/3D Plots in LaTeX*. Version 1.16. 31st Aug. 2018.

[239] International Organization for Standardization. *Quantities and units – Part 2: Mathematical signs and symbols to be used in the natural sciences and technology*. Standard ISO 80000–2:2009(E). International Organization for Standardization, Aug. 2009.

[240] J. A. SCHOUTEN. *Der Ricci-Kalkül: Eine Einführung in die Neueren Methoden und Probleme der Mehrdimensionalen Differentialgeometrie*. German. Ed. by R. COURANT. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, 1924. ISBN: 978-3-540-08565-2.

[241] P. W. MILONNI. *The Quantum Vacuum: An Introduction to Quantum Electrodynamics*. Academic Press, 14th Jan. 1994. ISBN: 978-0-12-498080-8.