Measurement of the top quark mass from top quarks produced in proton-proton collisions at $\sqrt{s} = 13$ TeV that decay to dimuons via a J/ψ

by

Kevin Nicholas Barends

Supervisor: Assoc. Prof. Sahal Yacoob

Co-supervisors: Assoc. Prof. Timothy Andeen Jr and Dr James Keaveney

A thesis submitted in fulfilment for the
degree of Doctor of Philosophy of Science

in the
UCT-ATLAS Group
Department of Physics
Faculty of Science
University of Cape Town

January 2023

CERN-THESIS-2023-008
//2023

Declaration of Authorship

I, Kevin Nicholas Barends , hereby certify that the work contained in this thesis titled, "Measurement of the top quark mass from top quarks produced in proton-proton collisions at $\sqrt{s} = 13$ TeV that decay to dimuons via a J/ψ" is original to me. I confirm that:

- This work was completed entirely or primarily while pursuing a research degree at this university.
- Where any portion of this thesis has been previously submitted for a degree or other qualification at this University or any other institution, it has been explicitly stated.
- Every time I have consulted someone else's published work, this has been done so with clear citation.
- I always cite the source when I quote from someone else's writing. This thesis is entirely my own work, with the exception of such quotations.
- All significant sources of assistance have been acknowledged.
- Where the thesis is based on work I did in collaboration with others, I have specified who did what and how much of it was mine.

Signed: Kevin Nicholas Barends

Date: January 2023

"Our greatest glory is not in never falling, but in rising every time we fall."

- Confucius

UNIVERSITY OF CAPE TOWN

Abstract

UCT-ATLAS Group
Department of Physics
Faculty of Science

Doctor of Philosophy of Science

by Kevin Nicholas Barends

The top quark mass is measured using $t\bar{t} \rightarrow$ lepton + $J/\psi(\rightarrow \mu^+\mu^-)$ events, using proton-proton collision data collected by the ATLAS detector at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, during 2015-2018. The data corresponds to a total integrated luminosity of 139.0 fb^{-1}. This lepton + J/ψ channel is statistically limited due to the low branching ratio of the $b \rightarrow J/\psi \rightarrow \mu^+\mu^-$. The top quark mass is measured from template fits over the m(lepton,J/ψ) distribution. The top quark mass is measured to be 172.03 $\pm$ 0.76 (stat) $\pm$ 2.14 (syst) GeV.

Acknowledgements

First and foremost, I would like to thank my Lord and Saviour Jesus Christ for providing me with the opportunity and strength to pursue and achieve one of my most important goals: obtaining a PhD in physics.

Second, I'd like to thank Shannon Seth Classen for always being there for me in any way during this degree. Without you, I would not have gotten this far.

Third, I'd like to thank my family and friends for their encouragement throughout this degree.

I'd also like to thank the National Research Foundation and Sahal Yacoob for their financial assistance. Your generosity has allowed me to accomplish one of my most important goals.

In addition, I'd like to thank my co-supervisor, Associate Professor Timothy Andeen Jr, for his willingness to help whenever I needed it and for the sacrifices he's made to ensure I get a better understanding.

Finally, I'd like to express my admiration and gratitude to my supervisor, Sahal Yacoob. You have no idea how many academic and personal lessons I have taken from you. I will never forget our journey.

Contents

List of Figures

List of Tables

Chapter 1

The Standard Model of particle physics

Unless otherwise stated, the following chapter contains information from "Modern Particle Physics" [1] and "Introduction to Particle Physics" [2].

1.1 Introduction

The Standard Model of particle physics [3–5] (SM) encapsulates the current understanding of the building blocks of matter and describes how these building blocks interact as fundamental particles. Fundamental particles have three inherent properties called spin, charge and mass. Spin is an intrinsic form of angular momentum, while charge generally refers to the electric charge of the particle. However, some particles possess other charged quantities called colour and weak isospin.

The interactions between the building blocks of matter are mediated by bosons, which carry the fundamental forces of nature. Three of the four fundamental forces of nature are described by the SM, except for the gravitational force. Thus far, the SM successfully passes all experimental tests [6] and provides an almost complete picture of the laws and characteristics of the universe. There are aspects about the universe that are not described by the SM such as the masses of the particles which are experimentally determined, the description of the gravitational force, dark matter and dark energy [7], the matter-antimatter imbalance in the universe [8], and more. There exist, however, other theories and models that go beyond the SM (known as BSM theories) [9, 10] which could describe these aspects of the universe, but are yet to be observed.

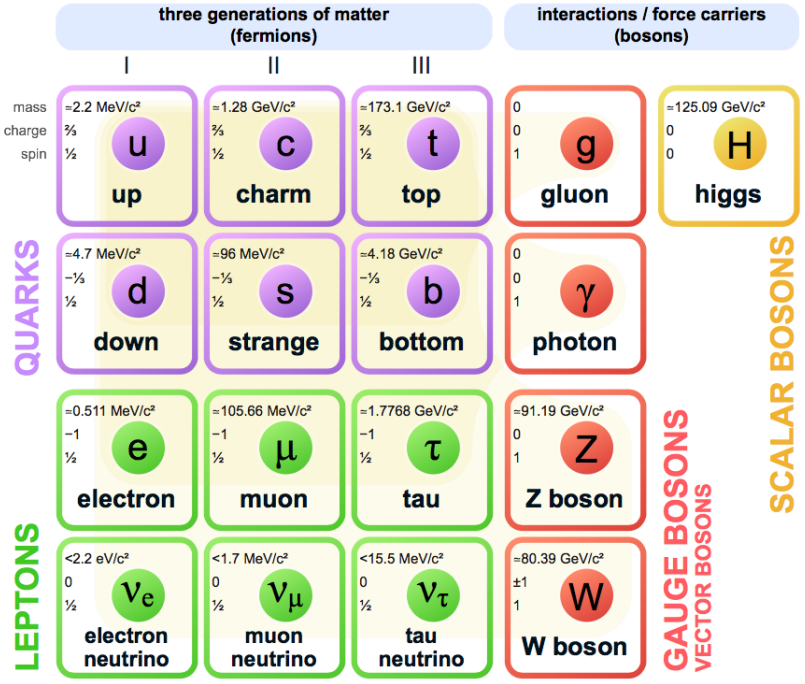


FIGURE 1.1: An illustration of the fundamental particles within the SM [11].

1.2 Fermions

Fermions are fundamental particles that have half integer spin. The SM consists of twelve fermions and twelve anti-fermions. Figure 1.1 shows the different types of fermions found in the SM, together with their mass, electrical charge, and spin which can be found in the top left corner of each fermion block. The anti-fermions, which are not depicted in Figure 1.1, are identical to their fermion counterparts except for having opposite quantum numbers, and thus any description of fermions applies to anti-fermions unless otherwise stated.

There are two categories of fermions found in the SM, namely quarks (see the purple section in Figure 1.1) and leptons (see the green section in Figure 1.1). Each category has six particles (or flavours), and each category is divided into three families (or generations). The three generations are made up of two quarks and two leptons that are similar to one another. The difference is in the masses of the particles in each generation: the quarks and leptons in each successive generation are heavier than those in previous generations. The first generation is in charge of all stable matter in the Universe, while the second and third generations appear only in highly energetic environments.

1.3 Bosons

Bosons are fundamental particles that have integer spin. The SM consists of four bosons with spin equal to one (known as vector bosons) and one boson with a spin of zero (known as a scalar boson). On the right side of Figure 1.1, these five bosons are shown. Additionally, the W boson is the only boson with an antiparticle, which can be seen in the same block in Figure 1.1. The only scalar boson is called the Higgs and the Higgs mechanism is responsible for the masses of all fundamental particles. The four vector bosons mediate the interactions between the fundamental particles, and these interactions are governed by three fundamental forces. The electromagnetic force is mediated by the photon (γ). The weak force is mediated by the W and Z bosons, while the strong force, is mediated by the gluon (g).

1.4 Theoretical overview of the Standard Model

The fundamental forces described by the SM are in-turn described by specific Quantum Field Theories (QFTs) [12]. A QFT is a mathematical model that describes a region in which each point in space and time has a physical quantity associated with it (or field) that is responsible for the interaction of separate systems (in our case, particles) and is manifested in a particle exchange between them. More technically, QFTs combine classical field theory, special relativity and quantum mechanics [12].

1.4.1 Electroweak Theory

The electroweak theory unifies [13] Quantum Electrodynamics (QED) [12, 14] and the Weak Theory or Quantum Flavourdynamics (QFD) [12]. QED was the first theory to produce full agreement between special relativity and quantum mechanics, and describes the electromagnetic force and the interactions between electrically charged particles through the exchange of photons. QED is symmetric under the transformation group $U(1)$. QFD describes the weak force and the interactions between flavour particles through the exchange of W and Z bosons, and is symmetric under $SU(2)$. The unification of the two theories showed that the two interactions can be described as a single theory with transformations described by the $SU(2) \times U(1)$ symmetric group.

1.4.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) [12] describes the strong force and the interactions of all quarks through the exchange of gluons, and is symmetric under $SU(3)$. QCD also describes the self interaction of gluons, specifically, three and four gluons interacting with each other.

In QCD, there are two phenomena that arise which are different from the other theories: colour confinement [15] and asymptotic freedom [16]. Colour confinement states that quarks and gluons cannot be found isolated in nature. Quarks possess another charge known as colour (or colour charge). There are three-colour charges: red, blue, and green. These names don't actually describe visible colour, but are a representation of the three different charges of QCD. Quarks are only allowed to be found in colour-neutral states called hadrons. Colour-neutral states or colourless states are when the hadrons have no net colour charge, or when the hadron has all three-colour charges combined. There are two common types of hadrons found: mesons, which consists of a quark and antiquark where the quark has one colour charge and the antiquark must contain the opposite colour charge to maintain a colour-neutral state, and baryons, which consists of three quarks (or antiquarks) that combine the three different colour charges to form a colour-neutral state. Asymptotic freedom describes the decrease of the interaction strength between quarks and gluons as the interaction energy increases. This means that quarks and gluons can be asymptotically free as the energy increases.

1.5 Standard Model Interactions

The SM interactions between fundamental particles can be described by Feynman diagrams. Feynman diagrams consist of symbols, which are used to describe fermions, bosons, and their interactions, where time runs from left to right. Additionally, the vertices in these diagrams can be rotated and combined. Particles and antiparticles are differentiated by the arrow on the solid lines. If the arrow is pointing forward (backward), the solid line represents the particle (antiparticle). The main interaction vertices of the SM between fundamental particles are described by the following Feynman diagrams.

- Electromagnetic Force: Interactions occur between electrically charged particles, p_{elec}, through the exchange of a photon. The photon does not change the flavour of the particle through the interaction.

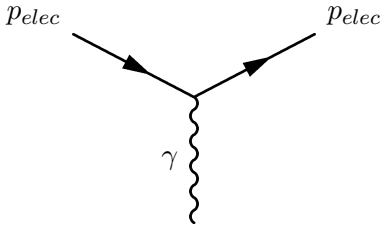


FIGURE 1.2: Feynman diagram showing the SM vertex for electromagnetic interactions.

- Strong Force: Interactions occur between quarks, q, through the exchange of a gluon. The gluon does not change the flavour of the quark through the interaction. Additionally, gluons can self-interact since they also carry the colour charge, and either three or four gluons can interact together.

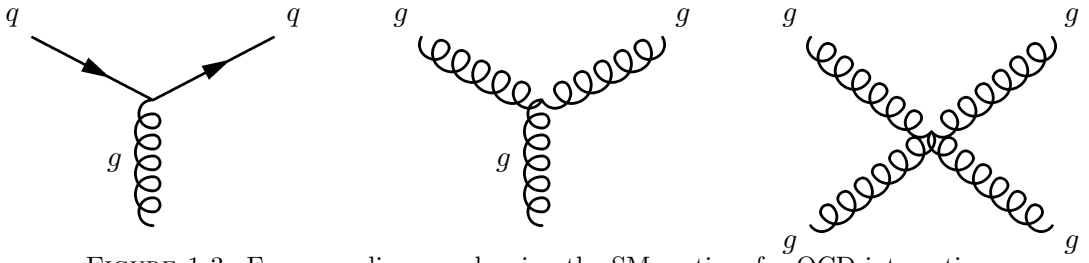


FIGURE 1.3: Feynman diagram showing the SM vertices for QCD interactions.

- Weak Force: Interactions occur between all fermions, f, through the exchange of either the W boson (which changes the flavour to f') or the Z boson (which does not change the flavour).

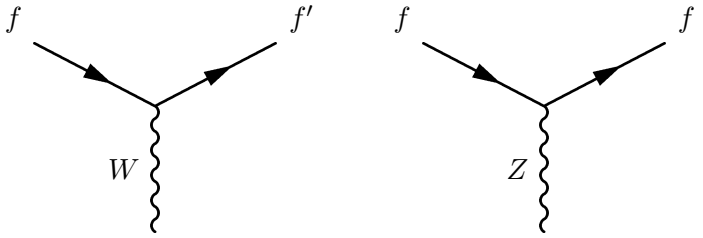


FIGURE 1.4: Feynman diagram showing the SM vertices for weak interactions.

- Higgs' interaction: The Higgs boson interacts with all massive fundamental particles. Since right-handed neutrinos are unable to be directly detected, it has not yet been verified that the Higgs boson couples with neutrinos.

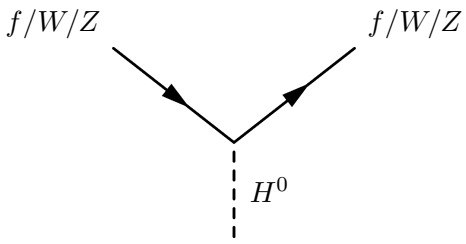


FIGURE 1.5: Feynman diagram showing the SM vertices for Higgs interactions.

1.6 Particle Decays in the Standard Model

The SM is made up of both stable particles that do not decay (such as the electron) and unstable particles that decay over time. For any particle to decay, the decay products must have a lower mass than the decaying particle and the process must be allowed to occur via one of the SM interaction vertices shown above.

In 1969, Richard Feynman proposed a parton model to describe the constituents inside hadrons [17], and today we know them as quarks and gluons. Most quarks live long enough to interact with the strong force. Quarks and gluons can only be found in colorless bound states known as hadrons due to colour confinement. Hadrons are created in high energy physics by combining quarks and gluons in a process known as hadronization [18]. In an experiment, a shower of hadrons and other particles are produced after a quark or gluon undergoes hadronization. This shower of particles usually presents itself in the shape of a cone, called a jet.

1.7 Particle masses in the Standard Model

The masses of the particles appear in the SM, but their values are not predicted by the model itself. As mentioned above, the mass of the particles in the SM come from their interaction with Higgs field [19]. Each particle has a different interaction strength with the Higgs field, which leads to the different masses. References [20, 21] are examples of how the top quark and W boson masses have been measured. See References within [6] for the different ways all the masses of the particles have been measured.

This thesis will describe a measurement of the top quark mass.

Chapter 2

The top quark

2.1 Introduction

The top quark is the heaviest fundamental particle as well as the heaviest quark within the SM. The heaviness of the top quark causes it to decay with an extremely short lifetime of 10^{-25} s [6]. The top quark will decay before being converted into a hadron because this timescale is shorter than the one required for this conversion ($\mathcal{O}(10^{-24}$ s) [1]). As a result, the top quark can only be studied experimentally through its decay products.

2.2 Production of the top quark

The top quark is the heaviest fundamental particle in the SM; producing top quarks requires extremely high energies. Top quark production can be divided into two categories: top and anti-top quark pair production ($t\bar{t}$) and single-top production. The mediating particles in these productions have masses inconsistent with their pole masses and are, therefore, considered off-shell particles, which are also known as virtual particles.

2.2.1 Single-top

According to the permitted vertices in the SM, there are three predominant methods for producing a single top quark, i.e. via a weak interaction with a W-boson (called t- or s-channel), or in associated production with a W-boson [6, 22, 23].

- t-channel:

  This process occurs with an initial b-quark interacting with a different flavour quark (denoted q in the following diagram) via a W-boson. This interaction produces a quark (denoted q') and a top quark.

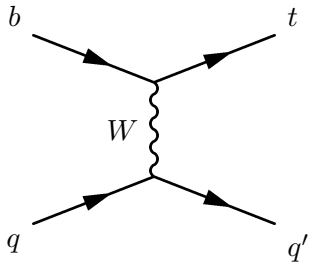


FIGURE 2.1: Feynman diagram showing the t-channel single top production process.

- s-channel:

  This process occurs with an initial anti-quark (denoted $\bar{q}'$) annihilating with a different flavour quark (denoted q) via a W boson. In order to produce a top quark, the W boson decays into a top and an anti-b quark.

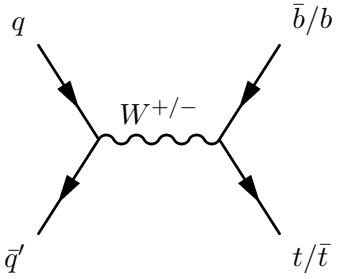


FIGURE 2.2: Feynman diagram showing the s-channel single top production process.

- W + top quark associated production:

  This process occurs with an initial gluon either being absorbed by a b quark or interacting with a b quark via a top quark. This interaction produces a W boson and a top quark.

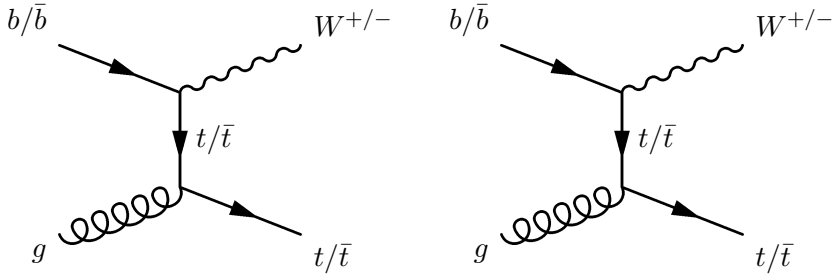


FIGURE 2.3: Feynman diagrams showing single top quark associated production with a W-boson.

2.2.2 Top quark pair production

According to the permitted vertices in the SM, there are two predominant methods for producing a top and an anti-top quark pair , i.e. via gluon-gluon fusion or quark annihilation [6, 23].

- gluon-gluon fusion:
  This process occurs when two gluons interact with each other, producing a top quark and an anti-top quark.

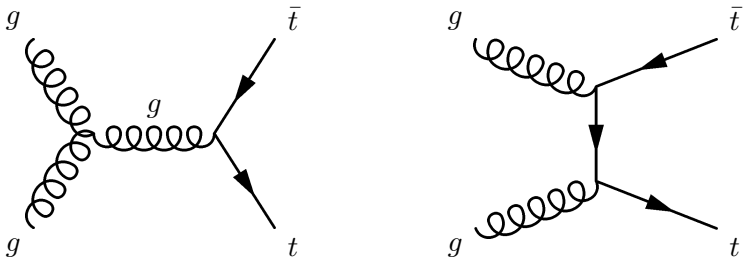


FIGURE 2.4: Feynman diagrams showing top quark pair production from gluon interactions.

- quark annihilation:
  This process occurs when a quark and its anti-quark annihilate, producing a gluon, which then produces a top quark pair.

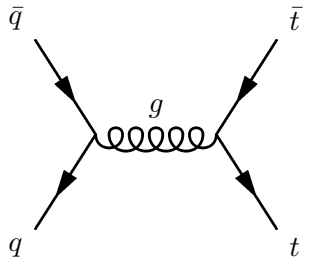


FIGURE 2.5: Feynman diagram showing top quark pair production from quark annihilation.

2.3 Decay of the top quark

After a top quark is produced, it will almost solely, $\sim 99.5\%$ of the time, decay into a W-boson and a b-quark [6]. The W-boson can decay either leptonically (into either a muon, an electron or a tau lepton, together with the corresponding neutrino) or hadronically (into two quarks) which occurs 33% and 67% of the time, respectively [6]. The b-quark lives long enough to undergo hadronization before decaying.

2.4 Top quark masses

The top quark mass can be determined from direct and indirect measurements. Direct measurements rely on the kinematic information of the decay products originating from the top quark [24], whereas indirect measurements rely on differential and absolute top and anti-top quark cross-section measurements to calculate the top quark mass [24].

Quarks generate their mass through the interaction with the Higgs field, and their masses appear as parameters in the SM. These masses will have different values depending on the theoretical formalism (also known as schemes) used to define their existence. Infinities arise in the calculations when determining quark masses due to the strong interaction strength, but these infinities can be removed (or absorbed) by introducing additional quantities into the calculations, producing finite and sensible results. This is known as renormalization. The modified minimal subtraction ($\overline{\text{MS}}$) and pole schemes are the two most well-known renormalization schemes [25]. The pole scheme describes long-distance quark mass parameters, while the $\overline{\text{MS}}$ describes short-distance quark mass parameters [6][26]. These two mass schemes are two common methods of quoting the top quark mass, and they are linked by non-perturbative effects known up to the third loop correction term (see references [6] and [25] for more information).

There is an alternative way to quote the top quark mass that corresponds to the top quark mass given to various Monte Carlo (MC) generators, often referred to as the Monte Carlo top quark mass (i.e. $m_{\text{top}}^{\text{MC}}$) [26]. The $m_{\text{top}}^{\text{MC}}$ is determined using a direct measurement approach.

The top quark mass mentioned for the rest of this thesis will correspond to the MC top quark mass, unless otherwise stated.

2.5 Discovery of the top quark

The top quark was first discovered by the CDF [27] and D0 [28] experiments in 1995. These experiments measured the mass of the top quark to be 176 $\pm$ 8 (statistical) $\pm$ 10 (systematic) GeV and 199^{+19}_{-21} (statistical) $\pm$ 22 (systematic) GeV, respectively. The relative uncertainty associated with these measurements were large, as they were $\sim$ 10% of the central value.

The discovery of the top quark led to a variety of different measurements trying to determine its mass [26, 29]. A precise measurement of the top quark mass impacts many areas in physics [24, 30, 31] with significant roles in determining the vacuum stability of the SM [32] (see Section 2.6), understanding the Higgs boson due to the interaction strength between the Higgs field and the top quark being on the order of unity [33], and by providing a good test for physics BSM with its unique link to the electroweak and Higgs sectors [33–36].

2.6 Vacuum Stability of the Standard Model

The following section is taken from my master's thesis [37].

"The SM involves quantum fields that can be in various configurations (also known as states), and one, or potentially more, of these states exhibit minimal energy. These minimal energy states are called vacuum states and can either be described as stable, unstable or meta-stable. If the state resides within a global minimum of energy, it is deemed stable. Unstable states are states within a local minimum which can decay into the global minimum by passing through the potential barrier (through a process known as quantum tunnelling) which exists between the local and global minima. Meta-stable states, however, reside within a local minimum which cannot decay into the global minimum due to the tunnelling probability being so small that the expected decay time far exceeds the expected lifetime of the universe [38].

To determine the vacuum stability of the SM, the effective potential of the SM needs to be examined. Examining the effective potential is ideally done non-perturbatively, but the top quark and Higgs boson interactions become so strong that the probabilities exceed 1 and therefore, the effective potential can only be computed perturbatively. After extensive calculations up to the next-to-next-to leading order of the effective potential in the perturbative regime (see [32, 39] and the references therein), the stability regions (as well as the non-perturbative region) are shown in a phase diagram in the top mass-Higgs mass plane, which can be seen in Figure 2.6. With the mass measurements at the time, the vacuum stability of the SM was found to be within the meta-stable region [32, 39]. However, due to the fact that the stability of the SM appears to be on the border between the stable and meta-stable region within the phase diagram, more precise measurements of the top quark mass and Higgs boson mass could put the vacuum stability within the stable region."

The most accurate W mass measurement was provided by the CDF collaboration [21]. This measurement shows deviations with the world average and the SM, which could be a sign of new physics, if the measurement is confirmed. Additionally, the mass of the W boson impacts the loop corrections for the Higgs boson mass [32], and therefore, could have an effect on the vacuum stability of the SM.

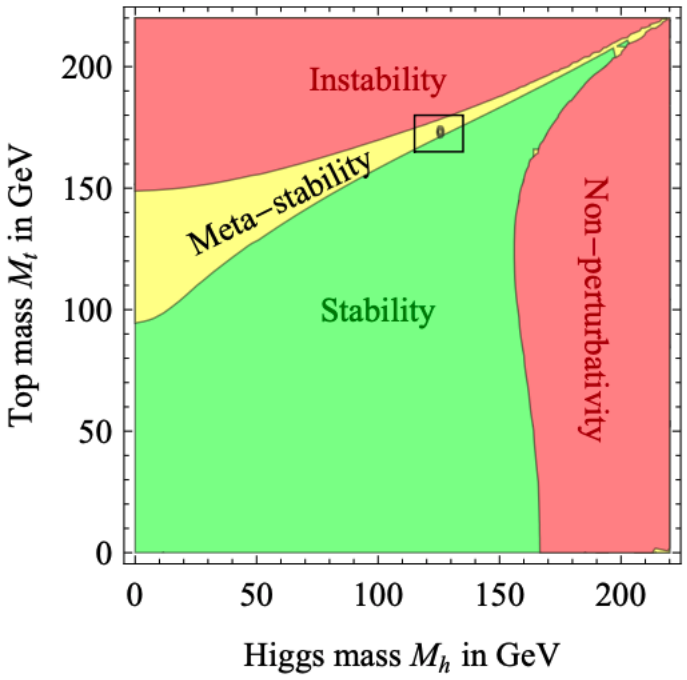


FIGURE 2.6: Stability regions of the SM vacuum in the top quark mass-Higgs boson mass plane. The current state of the SM vacuum is shown by the black dot [32].

2.7 A review on the different top quark mass measurements

As mentioned before, there are three different ways the top quark mass is generally quoted, namely: one common way to quote the top quark mass from direct measurements (i.e. MC mass) and two ways from indirect measurements (i.e. $\overline{\text{MS}}$ or pole mass). The leading precision measurement of the top quark mass is 171.77 ± 0.38 GeV [40]. However, this has yet to be published in a peer reviewed journal, and, therefore, the leading published precision measurement of the top quark mass is considered to be 172.69 ± 0.25 (statistical) ± 0.41 (systematic) GeV [41]. The leading precision for the pole mass is 171.1 ± 0.4 (statistical) ± 0.9 (systematic) $^{+0.7}_{-0.3}$ (theoretical) GeV [42], while for the $\overline{\text{MS}}$ mass it is 162.9 ± 0.5 (statistical) ± 1.0 (systematic) $^{+2.1}_{-1.2}$ (theoretical) GeV [42]. Since this thesis implemented a direct measurement approach, it will only be compared to the MC top quark mass measurement.

The top quark mass from direct measurements consists of combinations of measurements conducted by the Large Hadron Collider and Tevatron collaborations [41, 43–48]. Each of these measurements determined the top mass by either combining the kinematics of a lepton(s) with the kinematics of jets or just from the kinematics of jets (i.e. in the lepton+jets, dilepton+jets and/ or all jets decay modes). These decay modes consist of large amounts of events and therefore, the statistical uncertainties are quite small. The systematic uncertainty for each measurement is dominated by the experimental uncertainties associated with jet reconstruction, and is larger than the statistical uncertainty. The leading precision measurement of the

top quark mass was done by the ATLAS collaboration with a value of 172.69 ± 0.25 ± 0.41 GeV, where the jet energy scale contributed 0.28 GeV to the systematic uncertainty of 0.41 GeV [41].

The statistical uncertainty can only be reduced with more events, but the systematic uncertainty can be reduced by reducing the different systematic dependencies on the mass measurement. Additionally, the leading source of uncertainty can be reduced by improving the reconstruction of jets or using a less jet-dependent top quark decay mode.

2.8 Thesis outline

The different measurements made by the Large Hadron Collider (LHC) and Tevatron experiments all used a direct measurement approach [41, 43–48]. Each experiment extracted and combined the kinematic information of the jets and/ or leptons which originated from the top quark to determine the top quark mass. The reconstruction of jets impacted the precision of these measurements, as it proved to be the dominating uncertainty.

There are other decay modes of the top quark which are largely independent of the aforementioned jet uncertainty and could improve the precision of the top quark mass. However, these decay modes have relatively low branching ratios. One of these decay modes include a lepton and J/ψ, where the J/ψ decays into two oppositely charged muons. This decay mode relies on a b-quark undergoing hadronization and the B-hadron decaying to a J/ψ meson, which only occurs 1.16 ± 0.10% of the time [6]. Furthermore, only 5.961 ± 0.033% of J/ψ mesons decay into two oppositely charged muons [6]. Due to these low branching ratios, this decay mode wasn't used to measure the top quark mass, as enough data was not available until 2016. The CMS experiment made the first measurement of the top quark mass using this decay mode in 2016, using the data recorded in 2012, which corresponded to an integrated luminosity of 19.7 fb^{-1} [49]. They measured the top quark mass to be 173.5 ± 3.0 (statistical) ± 0.9 (systematic) GeV, which has a relative uncertainty of 1.8%. This is not the most precise measurement of the top quark mass, but shows the dominating systematic uncertainty from the jet energy scale and resolution has become negligible.

This thesis measures the top quark mass using the lepton and J/ψ decay mode of the top quark with data collected from proton-proton collisions at a centre-of-mass energy $\sqrt{s} = 13$ TeV by the ATLAS detector in 2015-2018, corresponding to an integrated luminosity of 139.0 fb^{-1}.

Chapter 3

Experimental setup

The experimental setup for measuring the top quark mass in this decay channel requires, first, a collider with sufficient energy to produce top-anti top quark pairs at a high rate, and, second, a detector able to reconstruct these events with high precision and efficiencies. These requirements are met by the Large Hadron Collider (LHC) and the ATLAS detector, both located at CERN in Geneva, Switzerland. The LHC is a proton-proton collider operating at 13 TeV and produces so many top-anti top quark pairs that it is sometimes colloquially referred to as a "top quark factory."

3.1 CERN and the Large Hadron Collider

CERN was established in 1954 after the CERN convention was drafted and signed by 12 European countries [50, 51]. Since then, CERN has expanded to 23 official Member States (countries that contribute financially and are represented on the CERN Council), as well as relationships and ties with many other countries worldwide, including South Africa [52]. These other countries have different privileges and duties than Member States, but they all contribute to CERN's research.

The first foundation was laid for the European laboratory, which is located across the French-Swiss border in Geneva, Switzerland, in 1955 [53], and the first accelerator (the Synchrocyclotron) was built at CERN in 1957 [54]. CERN has since built many new accelerators and detectors, which is shown in Figure 3.1, keeping it at the forefront of science and technology. Since its inception, many inventions and discoveries have been made: the W and Z bosons were discovered in 1983 [55]; a mere 6 years later, a British scientist while working at CERN invented the World Wide Web (WWW) [56]; anti-matter (i.e. anti-hydrogen) was first produced at CERN in 1995 [57]; and, most recently, the Higgs boson was discovered in 2012 [58]. The

most recent discovery was made with the Large Hadron Collider, the world's largest particle accelerator.

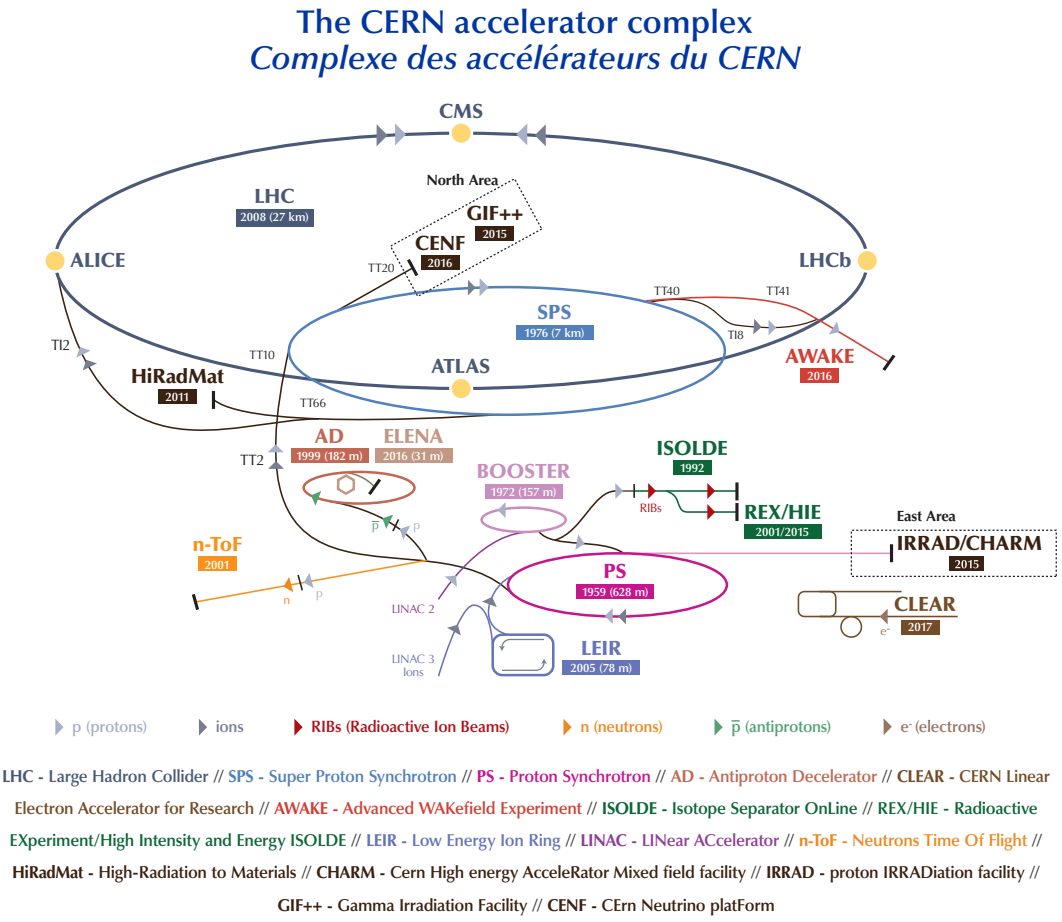


FIGURE 3.1: An illustration showing the accelerator complex at CERN [59].

3.1.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the most recent and largest particle collider among the other accelerators at CERN (see Figure 3.1).

Before the LHC existed, CERN constructed a 26.7 km tunnel, and collided electrons with positrons, called the Large Electron Positron collider (LEP), between 1984 and 1989 [60]. LEP was operational for 11 years and many key insights of electroweak interactions were obtained which were largely based on experimental evidence. In the year 2000, the LEP collider reached energies around 208 GeV, but was shutdown at the end of the same year to make way for the LHC [61].

The LHC was first approved by the CERN Council in 1994 and was built in the existing LEP tunnel between 1998 and 2008 [60]. The peak performance of the LHC was proposed to

be operating at a centre-of-mass energy, $\sqrt{s}$, of 14 TeV with an instantaneous luminosity of $10^{34}\text{cm}^2\text{s}^{-1}$ [60]. During 2015, 2016, 2017 and 2018, the LHC reached (and exceeded) the peak luminosity but only managed operating at $\sqrt{s} = 13$ TeV [62]. It was decided to start the LHC's second run at 13 TeV in order to maximize particle collision delivery for physics study and hasten the discovery of potential new physics. There are 1232 superconducting dipole magnets that direct the beams around the ring of the LHC. The magnetic field required to maintain a steady orbit must be greater the higher the beam energy, as must the electric current running through the superconducting coils of the magnet.

This paragraph is taken from my master's thesis [37]. "The LHC is not the only accelerator responsible for producing such high energies. There are successive accelerators all working together, feeding each other in order to produce and collide protons at extremely high energies (see Figure 3.1). The protons come from a bottle of hydrogen gas after removing the electrons with an electric field. These protons are injected into LINAC2, the first accelerator, accelerating protons up to an energy of 50 MeV [63]. After LINAC2, the protons travel through the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS), and are accelerated to an energy of 450 GeV [63]. The protons are then transferred into two beam pipes in the LHC tunnel and are accelerated to an energy of 6.5 TeV each [63]. After reaching 6.5 TeV of energy, the two proton beams collide at various points along the LHC ring, i.e. at ALICE, ATLAS, CMS and LHCb, where each point is accompanied by a detector to analyse the collisions. Each beam is intended to consist of 2808 bunches with 1.2×10^{11} protons per bunch and will collide 40 million times per second [64]."

Data-taking only happens after the proton beams are accelerated to the required energy and then aligned and focused by magnets to collide head on. Figure 3.2 shows the luminosity delivered by the LHC, the luminosity recorded by the ATLAS detector, and the luminosity recorded when data-taking occurred (also known as "Good for Physics") during Run 2. The data collected during Run 2 was used to perform the top quark mass measurement for this thesis.

The main objective of the LHC is to study the physics of the SM and beyond the SM by studying the particles that are produced from high energy collisions. These high energy collisions create an extremely dense and difficult environment to study fundamental particles. The high instantaneous luminosity of the LHC leads to multiple interactions per proton bunch crossing. Figure 3.3 shows the mean number of interactions per proton bunch crossing as recorded by the ATLAS detector during Run 2. This figure highlights the dense environments, as up to 70 collisions were taking place at the same time.

The ATLAS detector is one of the detectors along the LHC that was built to accurately study these extremely dense environments, and was used to study and analyse the data throughout this thesis.

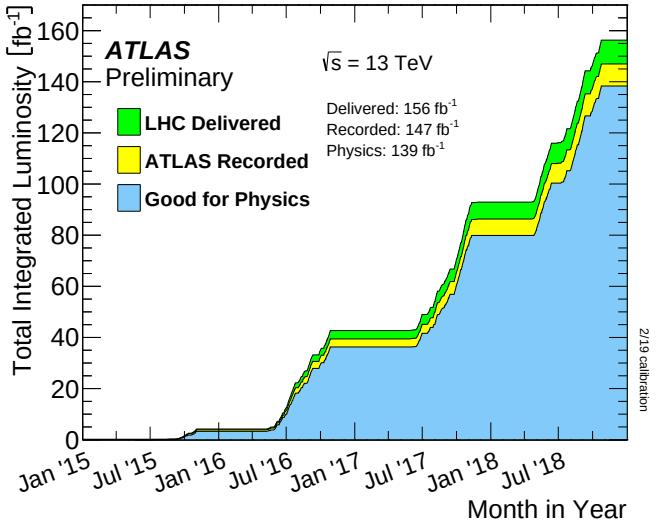


FIGURE 3.2: The total integrated luminosity of proton-proton collisions at $\sqrt{s} = 13$ TeV in the LHC during the data-taking period between 2015 and 2018. The delivered luminosity by the LHC (in green), the luminosity recorded by ATLAS (in yellow), and the luminosity where ATLAS was fully operational and stable data-taking conditions were maintained (in blue). The plot was taken from Reference [65].

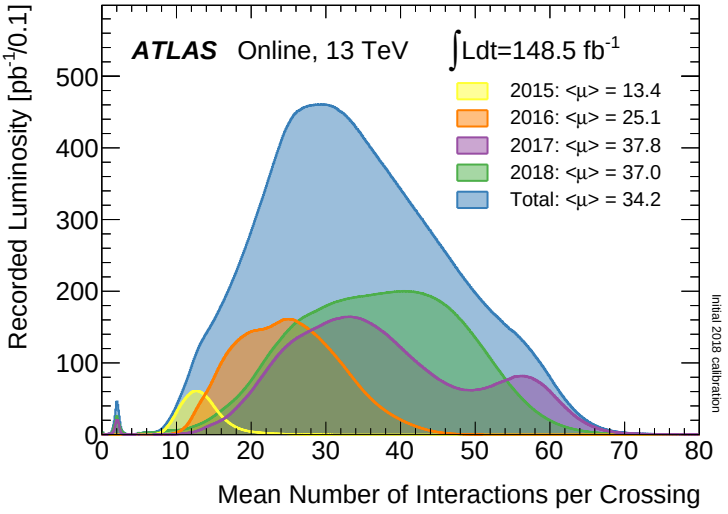


FIGURE 3.3: The recorded luminosity distribution of the mean number of interactions per bunch crossing in proton-proton collisions for the data-taking period of 2015-2018 at $\sqrt{s} = 13$ TeV by ATLAS [65].

3.2 The ATLAS Detector

Unless otherwise stated, the following section contains information from "The ATLAS Experiment at the CERN Large Hadron Collider" [66].

3.2.1 Overview

The ATLAS experiment was designed to fully exploit the LHC's discovery potential. The ATLAS collaboration consists of nearly 3000 scientific authors from 181 institutions around the world, that are not only scientists, but also engineers and technicians, who work together to create opportunities for success and discovery [67]. ATLAS conducts a wide range of studies to answer fundamental questions about the universe and whether there are new physics to investigate that go beyond the SM [68].

The ATLAS detector has a diameter of 25 m and a length of 44 m, and it is symmetric in the forward-backward direction about the detector's centre, as shown in Figure 3.4. It is housed in a cavern about 100 meters below ground and weighs a whopping 7000 tonnes [69]. The ATLAS detector is made up of many components (or subdetectors), but the Magnet system is its most prominent feature [64]. A thin superconducting solenoid surrounds the Inner Detector, and three large superconducting toroids, one barrel and two end-caps, are azimuthally symmetrically arranged around the Calorimeters. The ATLAS detector's other three major components are the Inner Detector, the Calorimeter, and the Muon Spectrometer. The Inner Detector is immersed in a 2 T solenoidal field, which is used for momentum resolution, pattern recognition, and collision point identification, while the Calorimeters measure particle energy and position. The Muon Spectrometer, which surrounds the Calorimeters, measures muon momentum. These subdetectors are linked to a Trigger and Data Acquisition system, as well as a computing system, which selects and stores physics events with specific properties of interest for further analysis [69].

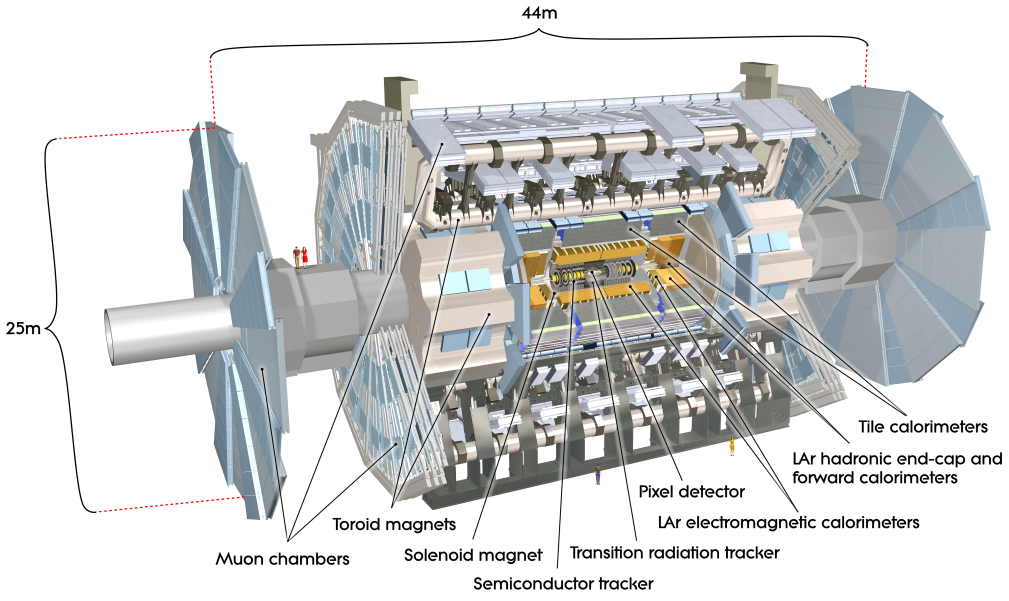


FIGURE 3.4: A computer generated image of the ATLAS detector [70].

3.2.1.1 The Coordinate System

The following section is taken from my master's thesis [37].

"The ATLAS detector uses a right-handed coordinate system, where the nominal interaction point is defined as the origin, and can be seen in Figure 3.5. The z-axis defines the beam direction and the x-y plane is, therefore, transverse to the beam direction. The positive x-axis points from the interaction point to the centre of the LHC ring, while the positive y-axis points upwards. The azimuthal angle ϕ is measured around the beam axis, and the polar angle θ is taken as the angle from the beam axis. The polar angle is more commonly replaced with pseudorapidity which is defined as $\eta = -\ln\tan(\theta/2)$. However, when dealing with objects that have a non-negligible mass, rapidity is used and defined as

$$y = \frac{1}{2}\ln\left[\frac{E + p_z}{E - p_z}\right]$$

where E is the energy of the object and p_z is the momentum of the object in the z-direction. Rapidity is used due to its difference being Lorentz invariant. Pseudorapidity is, however, preferred as it represents a detector quantity. A spatial quantity ΔR is defined in the η-ϕ space as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. There are other kinematic quantities which are commonly used and taken from the x-y plane: transverse momentum p_T, transverse energy E_T and missing transverse energy E_T^{miss}."

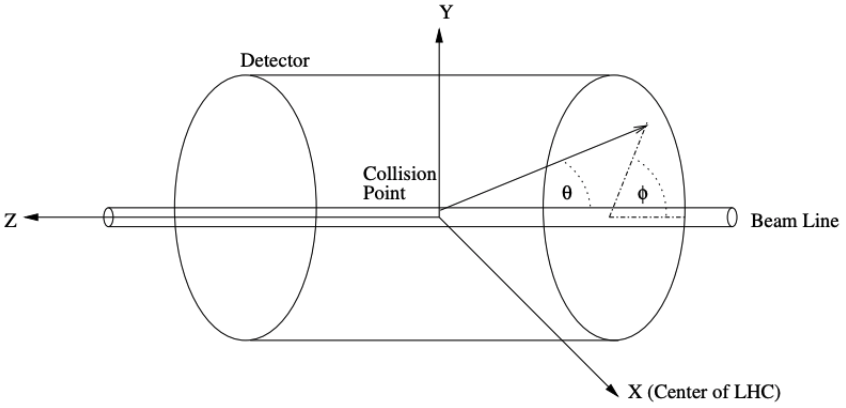


FIGURE 3.5: The ATLAS coordinate system [71].

3.2.2 The Magnet System

Within the magnet system, the ATLAS detector is made up of four large superconducting magnets. This system has a diameter of 22 meters and a length of 26 meters, and it stores 1.6 GJ of energy. The Inner Detector is surrounded by a solenoid magnet that generates a 2 T axial magnetic field, whereas the muon detectors are surrounded by three toroid magnets, one

barrel and two end-caps, which generate toroidal magnetic fields of approximately 0.5 T and 1 T, respectively. These magnets bend the trajectory of the particles as they pass through the ATLAS detector's subdetectors, allowing their momenta to be measured by the curvature of their trajectories. [72].

FIGURE 3.6: One of the solenoid magnets used in the ATLAS detector. The person can be used for scale [66].

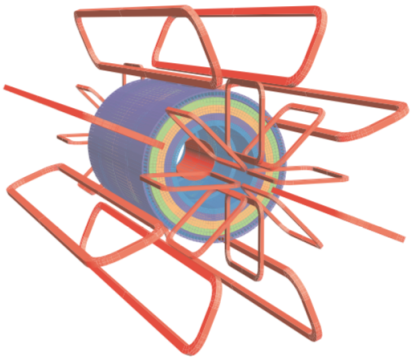


FIGURE 3.7: A computer generated image of the magnet system in the ATLAS detector [66].

3.2.3 Inner Detector

The Inner Detector (ID) surrounds the beam pipe, which is contained within a cylindrical envelope and is immersed in the large superconducting solenoid magnet. It is the first component of the ATLAS detector to detect decay products following the collision. The ID is made up of three parts that covers up to $|\eta|$ =2.5: the Pixel detector, the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT), all of which work together to identify primary and secondary vertices, track charged particles, and measure their momentum. Figure 3.8 shows the design of the ID.

The LHC was shut down in 2013 and 2014, and an extra pixel layer known as the insertable B-layer (IBL) [73] was inserted between the beam pipe and the pixel detector. The IBL was installed to deal with the increased luminosity during Run 2. The pixel detector is based on silicon semiconductor technology and consists of three barrel layers in the central region and three disks in the two forward regions. It's designed with high spatial resolution ($R-\phi$ resolution of 10 μm and a z resolution of 115 μm), which provides more than 100 million readout channels to track particles. The high resolution is critical for identifying primary interaction vertices, especially during Run 2 when the mean number of interactions was high. Secondary vertices of particle decays that occur within the detector volume are likewise identified using data from the pixel system. The SCT, a semiconductor strip detector with eight layers of strips in its barrel component and two end-caps, provides additional tracking information with a spatial resolution $R-\phi$ resolution of 17 μm and a z resolution of 580 μm. Half of the barrel strips are parallel to the beam axis and provide high-granularity data, while the other half is slanted by 40

mrad to provide tracking coordinates in the z direction. The TRT is a straw tube detector with tubes arranged parallel to the beam pipe in the barrel component and radially in the end-caps, and provides a spatial resolution of $R - \phi$ resolution of 170 μm. Apart from extra tracking information, the Lorentz factor of travelling particles impacts the TRT signal amplitude and as a result, amplitude differences can be utilized to distinguish light, ultra-relativistic electrons from pions and other hadrons.

The ID has a combined momentum resolution of $\frac{\sigma(p_T)}{p_T} = 0.05\% p_T \oplus 1\%$, where $\oplus$ signifies summing in quadrature, and provides a longitudinal and transverse vertex resolution of 30 μm and 20 μm, respectively [74].

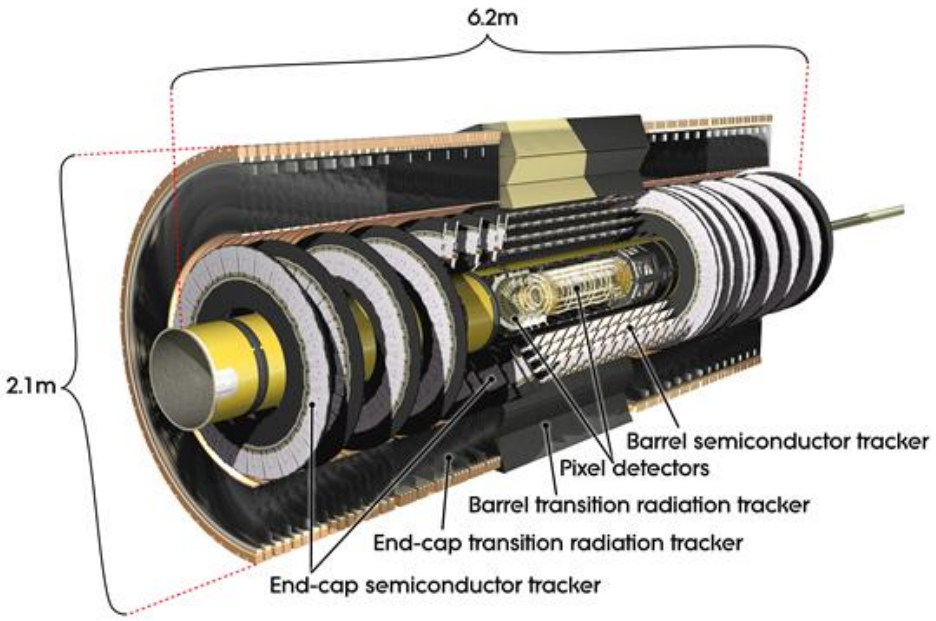


FIGURE 3.8: A computer generated image of the ATLAS inner detector and its components [75].

3.2.4 Calorimeters

The ATLAS detector consists of three calorimeters, the Electromagnetic Calorimeter (ECal), Hadronic Calorimeter (HCal), and Forward Calorimeter (FCal), situated between the ID and the Muon Spectrometer which covers $|\eta| < 4.9$. The ECal only covers $|\eta| < 3.2$ which includes the ID coverage, and consists of fine granularity components used to precisely measure the energy, timing information, and position of electrons and photons. On the other hand, the HCal has coarser granularity and is used to measure energy of hadrons and jets. The calorimeters are designed to provide good containment of both electromagnetic and hadronic showers, and minimises punch-through effects into the Muon Spectrometer system. The design of these calorimeters can be seen in Figure 3.9.

The ECal consists of a barrel structure and two end-cap structures which cover $|\eta| < 1.475$ and $1.375 < |\eta| < 3.2$, respectively. With lead absorber plates and active liquid argon (LAr) layers, all sections are built as a sampling calorimeter: the lead plates cause nuclear interactions with photons and electrons, which leads to the ionization of the LAr. The components that cover $|\eta| < 2.5$ have three calorimeter sections in depth, whilst the others have two sections in depth and coarser granularity. Upstream of the calorimeter, an extra presampler compensates for electron and photon energy losses up to $|\eta| < 1.8$. The HCal consists of a barrel region with steel absorber plates and scintillating tiles as the active material, and a LAr end-cap system. The tile barrels are three layers in depth extend up to $|\eta| < 1.7$ while the end-cap system consists of two wheels structures with two segments in depth and covers $1.5 < |\eta| < 3.2$. The FCal provides the coverage up to $|\eta| < 4.9$ and consists of three components, where the first is made of copper and measures electromagnetic showers, while the other two are made of tungsten and focus on hadronic interactions.

The energy resolution is an important parameter of calorimeters. This determines how precisely incoming particle energy can be measured. The resolution can be written as follows:

$$\frac{\sigma(E)}{E} = \frac{a}{\sigma E} \oplus \frac{b}{\sqrt{E}} \oplus c,$$

where a is the stochastic term resulting from statistical fluctuations in the amount of deposited energy, b is the noise term resulting from the electronics, and c is the constant term describing the variations in response across the calorimeter. The ECal has a resolution of 10%, 170 MeV and 0.7% for a, b and c, respectively, whereas the HCal has a lower precision of 52.9% and 5.7% for a and c, respectively, with negligible contributions from b [76].

3.2.5 Muon Spectrometer

The Muon Spectrometer (MS) is the outermost sub-detector of the ATLAS detector and is primarily used to measure muon momenta. Muons do not lose much energy as they traverse through the calorimeters as they are heavier than electrons and not affected as much by bremsstrahlung, and since they are leptons they won't undergo hadronic showers and be caught in the HCal. Therefore, muons will usually be the only particles to make it to the MS. Muon momenta are measured in the MS system, which is made up of many chamber systems, using the magnetic deflection of muon tracks in the air-core toroid magnets contained in the barrel and end-cap structures. The barrel toroid provides a magnetic field approximately orthogonal to the muon tracks in the range $|\eta| < 1.4$. It is made up of eight radially formed coils with a total length of around 25 m. The two smaller end-cap toroids, which are put into the ends of the barrel toroid and align with the inner solenoid magnet, bend muon trajectories for $1.6 < |\eta| < 2.7$. Each one

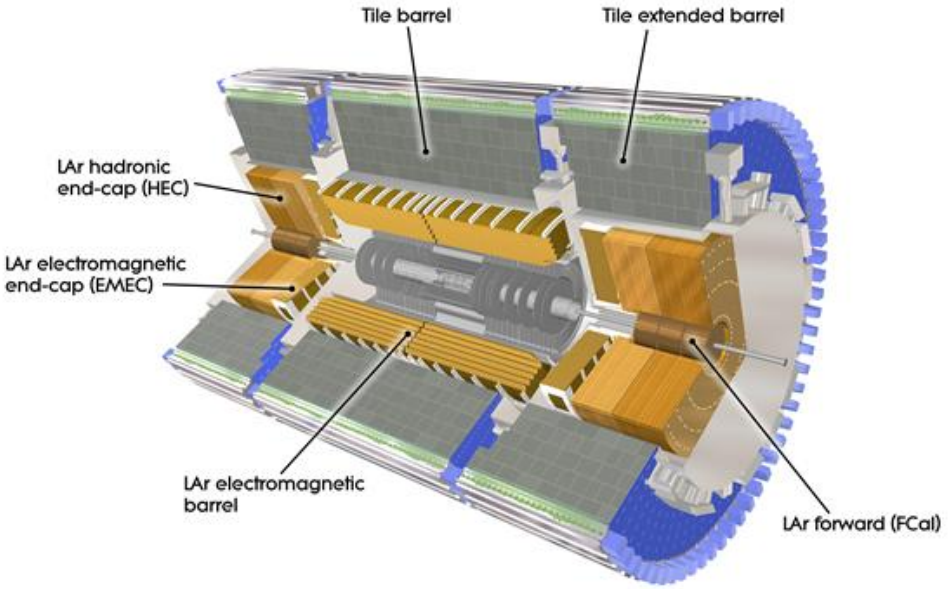


FIGURE 3.9: A computer generated image of the ATLAS calorimeters and its components [77].

is made up of eight racetrack-like coils. Both the barrel and end-cap toroids provide magnetic deflection in the transition region, in the range $1.4 < |\eta| < 1.6$. Figure 3.10 shows a diagram of the entire MS system, including the toroid magnets.

The muon system is intended for precise object tracking and triggering. In the muon system, two subsections are responsible for tracking, namely Monitored Drift Tube Chambers (MDT) and Cathode-Strip Chambers (CSC), and two subsections are responsible for triggering, namely Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC). In the barrel region, the MDT layers are aligned cylindrically around the beam pipe, whereas in the transition and end-cap regions, they are perpendicular to the beam, and provide a spatial resolution of 50 μm. For $2.0 < |\eta| < 2.7$, they are augmented with cathode strip chambers with increased granularity in the innermost layer and provide a spatial resolution of 60 μm. A separate chamber system with resistive plate chambers in the barrel region and narrow gap chambers in the forward regions operates in the range $|\eta| < 2.4$. This system's data is utilized to locate bunch-crossings and complement the tracking chambers with orthogonal muon coordinate measurements, in addition to providing trigger information on well-defined muon p_T thresholds.

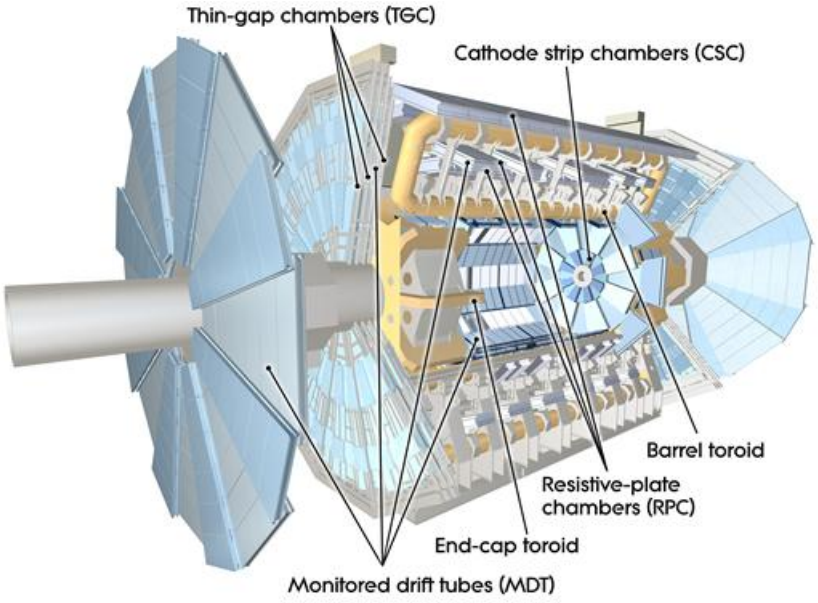


FIGURE 3.10: A computer generated image of the ATLAS muon spectrometer and its components [78].

3.2.6 Forward Detectors

Four smaller detectors sit alongside the ATLAS detector in the forward region. ALFA (Absolute Luminosity For ATLAS) and LUCID (LUminosity measurement using Cherenkov Integrating Detector) are used to measure the luminosity delivered to the ATLAS detector [79, 80], while the ZDC (Zero-Degree Calorimeter) studies heavy ion collisions [81], the AFP (ATLAS Forward Proton) studies the energy loss and momentum transfer of very forward protons [82]. Figure 3.11 shows a diagram of these smaller detectors alongside the ATLAS detector.

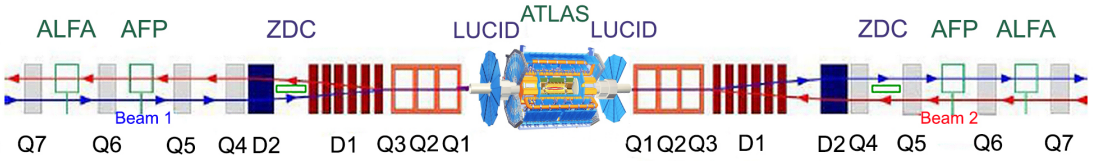


FIGURE 3.11: A computer generated image of the ATLAS forward detectors [83].

3.2.7 Trigger and Data Acquisition System

The ATLAS detector's Trigger and Data Acquisition (TDAQ) system is the central point for collecting event data from the many subsystems, processing and filtering it, and forwarding it to permanent storage [84, 85]. Figure 3.12 shows a schematic diagram of the TDAQ system, showing peak rates and band widths through each component. Only a small fraction of events can be read out and stored due to high bunch-crossing rates of up to 40 MHz during Run 2

and high levels of pile-up. Decisions concerning events and whether to save them require a fast-response system to minimize the dead time of individual components. In Run 2, the TDAQ system has two trigger levels: a low-level hardware trigger (L1) and a high-level software trigger (HLT).

The L1 trigger is made up of a central trigger decision unit that receives data from the calorimeters' low-granularity components and the muon-trigger system. This unit additionally establishes preventive dead times in the readout components to avoid overlapping readout windows and overflowing buffers. After the L1 trigger system accepts an event, it is briefly held in a separate buffer system off-detector called the Read-Out System. The L1 system also creates regions of interest in the η-ϕ plane to emphasize possible objects including muons, electromagnetic clusters, and substantial total transverse momentum. These regions of interest are conveyed to the HLT once the event has been L1-accepted and entered into the Read-Out System. The HLT uses a dedicated computing farm to reconstruct event data and then gathers data from all atlas components and uses software to make a more complex trigger choice on candidate objects. The HLT accepts events, which are subsequently written to disk and saved indefinitely. The L1 trigger reduces the original 40 MHz of collisions to a trigger rate of around 100 kHz, which is further decreased to the order of 1 kHz by the HLT.

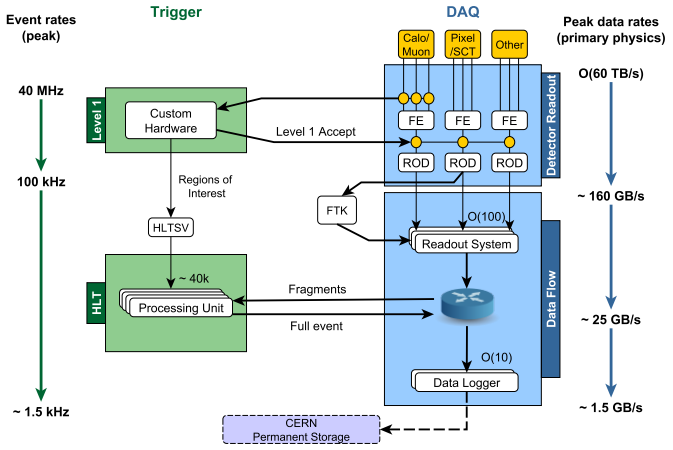


FIGURE 3.12: An illustration of the ATLAS TDAQ system with peak rates and bandwidths for each component [86].

3.2.8 ATLAS simulation

The majority of the MC samples used in this research are simulated using the full ATLAS detector simulation [87], which is based on GEANT4 [88]. GEANT4 provides the framework to build a simulation of the ATLAS detector and it's components. The full simulation provides high accuracy of the detector simulation and event reconstruction, but it is computationally

expensive. A small subset of samples, particularly those required to assess systematic uncertainties, are simulated using the faster AtlasFast-II (AFII) package [89, 90], which accelerates the detector response simulation by employing parametrized shower forms in the calorimeters. This is not as accurate as the full simulation, but it is computationally less expensive.

Due to the fact that the MC samples run through a simulation of the ATLAS detector, the response could be different with that of real data. Various corrections are applied to the MC samples to account for the difference responses. These corrections are applied as scale factors and described under Section 3.3.

3.3 Particle reconstruction within the ATLAS detector

The following section consists of information from "The ATLAS Experiment at the CERN Large Hadron Collider" [66] unless otherwise stated.

3.3.1 Overview

The protons can collide elastically, which results in uninteresting events because no new particles are produced, or inelastically, which results in a large number of particles that interact with the ATLAS detector—except for neutrinos, which do not interact with any detector material. The partons inside the protons interact in inelastic proton-proton collisions, and parton distribution functions (or PDFs) are required to describe their momenta. A large number of particles produced in inelastic collisions have significant transverse momentum, but decay quickly. Only a small fraction of particles from the collision survive long enough to be measured by the ATLAS detector. The other particles decay into lighter, more stable particles, which can be detected. Only fourteen different types of elementary and composite particles have long enough mean free paths to interact with the detector materials and allow detection. Muons, electrons, photons, pions, kaons, protons, and neutrons are these detectable particles (including antiparticles for fermions). The latter four are not elementary particles, but compound particles made from colour-charged elementary particles and created due to colour confinement. Jets are created when colour-charged particles produce complete showers of colour-neutral compound particles. Bottom quark jets have distinct features and are commonly referred to as b-jets. The different types of particles and interactions as seen by the ATLAS detector is shown in Figure 3.13.

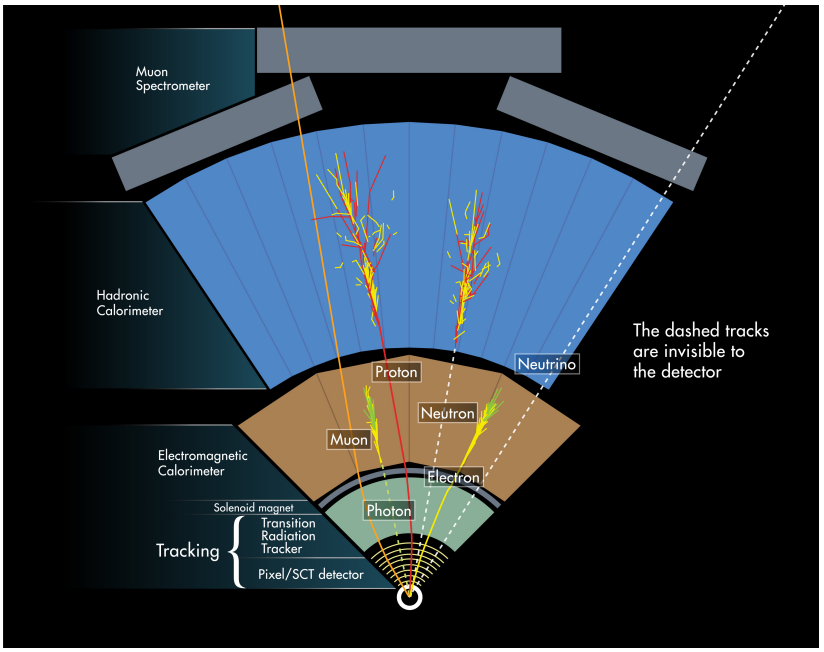


FIGURE 3.13: An illustration showing detector signatures for different particles in the ATLAS detector [91].

3.3.2 Track reconstruction

Track reconstruction in ATLAS comes from information collected by the ID. The initial step is to reconstruct clusters from deposits in the Pixel detector and SCT, while the raw timing information is obtained using the TRT. Track seeds are created by combining at least three points (clusters reconstructed in a 3D position) from the Pixel and SCT layers, which provide a preliminary estimate of the particle trajectory. The seeds are then extended to the remaining Pixel and SCT clusters using a global χ^2 and a combinatorial Kalman filter [92]. To identify and eliminate fake tracks, several criteria on the reconstructed tracks are used:

- $p_T > 500$ MeV and $|\eta| < 2.5$
- at least 7 Pixel and SCT clusters
- at most either one shared Pixel cluster or two shared SCT cluster on the same layer
- a maximum of two holes in the combined Pixel and SCT cluster
- at most one hole in the Pixel cluster

where a hole is defined as an expected cluster on a track but does not show up in the detector.

3.3.3 Vertex reconstruction

Vertex reconstruction is divided into two stages, vertex finding and vertex fitting [74]. Vertex finding involves a pattern recognition process, which is described as the association of reconstructed tracks to vertex candidates. The reconstruction of the real vertex position is part of the vertex fitting procedure.

A seed for the first vertex is used and based on the position of the beam spot in the transverse plane. The optimal vertex position is calculated using the seed and tracks with a fitting procedure. The tracks entering the fit must pass the previous track selection criteria, but can have a $p_T > 400$ MeV, must have at least nine hits in the Pixel and SCT detector, no Pixel holes, and at most one SCT hole [93]. Each time the fit is performed, the vertex position is recalculated and less suitable tracks are down-weighted. Tracks that are incompatible with the vertex are taken out of it once the vertex position has been established and can then be used to establish another vertex. The procedure is repeated with the remaining tracks in the event, until no unassociated tracks are left, or no additional vertex can be found. Only vertices with at least two related tracks are kept, and the primary vertex is defined as the one with the highest scalar sum of squares transverse momenta (Σp_{T}^2) of the associated tracks among all the vertices reconstructed in a bunch crossing.

3.3.4 Electrons

The electrons' are identified with a track in the ID that matches a narrow shower in the ECal. To reject the enormous backgrounds caused by jets, which result in misidentified (or fake) electrons, a sophisticated electron reconstruction is required. All the information provided in this section comes from reference [94], and should be consulted for more information.

3.3.4.1 Reconstruction

Electrons are reconstructed within the ID coverage up to $|\eta| < 2.47$ within ATLAS. The ECal has a transition region of between the barrel and end-caps that consists of cooling and support structures. Thus, electrons inside this region, i.e. $1.37 < |\eta| < 1.52$, are not reconstructed. Electron reconstruction is divided in three stages: clustering, cluster-track matching, and candidate reconstruction.

The clustering stage employs a topo-cluster reconstruction algorithm, which starts by constructing clusters in the ECAL and HCAL using a set of noise thresholds, with the cell originating the cluster having a significance $\zeta_{cell} = E_{cell}/E_{cell}^{exp.noise} > 4$, where E_{cell} is the energy of the cell and $E_{cell}^{exp.noise}$ is the energy due to the expected noise. The calorimeter noise can cause

negative-energy cells and therefore the algorithm uses $|\zeta_{cell}|$ instead of ζ_{cell} to avoid biasing the cluster energy upwards, which would occur if only positive-energy cells were collected. The clusters then aggregate cells that are close to each other and have a significance of $|\zeta_{cell}| > 2$. Two clusters are merged if they include the same cell with $|\zeta_{cell}| > 2$. Following the collection of all surrounding cells, a final set of bordering cells with $|\zeta_{cell}| \geq 0$ is added to the cluster. Clusters containing two or more local maxima are separated into independent clusters; a cell is deemed a local maximum if it has $E > 0.5$ GeV, at least four neighbours, and none of them has a larger signal. Topo-clusters are only permitted if they have $E_T > 400$ MeV, which prevents clusters from pile-up and $\pi^0 \to \gamma\gamma$ decays.

The cluster-track matching stage classifies a cluster as an electron if it has one or more connecting reconstructed tracks. The tracks must be within $|\eta| < 0.05$ of the cluster, and -0.20 $< q \cdot (\phi_{track} - \phi_{cluster}) < 0.05$ when using the track energy to extrapolate from the last ID hit, or -0.10 $< q \cdot (\phi_{track} - \phi_{cluster}) < 0.05$ when using the cluster energy to extrapolate from the track perigee; where q refers to the charge of the reconstructed track.

To be considered an electron candidate, the cluster must at least have $E_{\mathrm{T}} > 1$ GeV and matched to a track with at least four hits in the silicon tracking section of the ID. Photons undergo the same reconstruction as electrons, except the cluster must at least have $E_{\mathrm{T}} > 1$ GeV and no tracks matched.

The Track-To-Vertex-Association (TTVA), which requires the track to be compatible with the primary vertex to limit background from conversions and secondary particles, is an additional constraint on the electrons that is employed by many analyses. The TTVA selections for electrons are $|d_0/\sigma_{d_0}| < 5$ and $|\Delta z_0 \cdot \sin\theta| < 0.5$ mm, where d_0 is the transverse impact parameter, σ_{d_0} the associated uncertainty, Δz_0 is the distance between the track's longitudinal impact parameter (z_0) and the primary vertex, and θ is the polar angle of the track.

3.3.4.2 Identification

An identification technique that utilizes a likelihood-based (LH) strategy is used to distinguish between real and fake electrons. This technique employs a multivariate analysis (MVA) method that considers several properties of electron candidates and combines them into a discriminant, which is used to distinguish signal from background, such as fake tracks caused by converted photons, electrons produced in the decay of heavy-flavoured hadrons, and many more. Track quality, ECal shower shape, track-cluster matching related quantities, and information derived from the TRT are among the properties included in the likelihood. Depending on the information supplied to the discriminant, three possible working points (Tight, Medium, and Loose) are defined, corresponding to varying amounts of signal acceptance and background rejection, as shown in Table 3.1.

Working Point	Signal (%)	Non-Iso. Bkg (%)	Conversions (%)	Hadrons (%)	Unclassified Bkg (%)
Loose	93.027 ± 0.011	26.17 ± 0.33	2.502 ± 0.025	0.4327 ± 0.0035	1.368 ± 0.018
Medium	87.411 ± 0.011	19.56 ± 0.28	0.649 ± 0.013	0.2190 ± 0.0025	0.756 ± 0.014
Tight	78.587 ± 0.010	14.20 ± 0.23	0.406 ± 0.010	0.0950 ± 0.0016	0.451 ± 0.010

TABLE 3.1: A summary of the three electron working points showing signal and background identification efficiencies.

3.3.4.3 Isolation

Isolation constraints are applied to the leptons in order to further eliminate background electrons emanating from hadronic decays. To decide if a particle is prompt or not, the isolation uses the amount of momentum/energy surrounding the particle of interest. This momentum/energy can come from either charged or neutral particles; p_T is commonly employed for charged particles, while E_T is more commonly utilized for neutral particles. They are easily connected by the calculation $E^2 = m^2 + p^2$ (in natural units), and E_T and p_T are quite similar for low mass particles with relatively high momentum. When it comes to track-based isolation, p_T will be employed throughout this thesis, and E_T will only be used when isolation based on calorimeter cells is mentioned. For these reasons, an electron isolation requirement is used to improve separation from either fake or real electrons. This is dependent on three variables being defined:

- E_T^{cone20} is the sum of the E_T of ECal clusters within a cone of $\Delta R = 0.2$ around the electron candidate;
- $p_T^{\text{cone20(30)}}$ is the sum of the p_T of all tracks falling in a cone of $\Delta R = 0.2(0.3)$ around the electron candidate and excluding tracks associated to the electron or converted photons;
- $p_T^{\text{varcone20(30)}}$ is the sum of the p_T of all tracks with $\Delta R = \min(0.2(0.3),\ 10/E_T\ [\text{GeV}])$ around the electron track candidate and excluding the electron associated tracks.

The application of isolation criteria is specific to the needs of physics analysis, and therefore several working points have been developed. The isolation efficiencies are measured in data and determined through simulation, and typically range from approximately 90% for the tightest working points to nearly 100% for the loosest. The different isolation working points are shown in Table 3.2.

Working point	Calorimeter isolation	Track isolation
FCLoose	$E_T^{\text{cone20}}/p_T < 0.2$	$p_T^{\text{varcone20}}/p_T < 0.15$
FCTight	$E_T^{\text{cone20}}/p_T < 0.06$	$p_T^{\text{varcone20}}/p_T < 0.06$
Gradient	$\epsilon = 0.1143 \cdot p_T[\text{GeV}] + 92.14$ using E_T^{cone20}/p_T	$\epsilon = 0.1143 \cdot p_T[\text{GeV}] + 92.14$ using $p_T^{\text{varcone20}}/p_T$
HighPtCaloOnly	$E_T^{\text{cone20}} < \max(0.015 \cdot p_T, 3.5\ \text{GeV})$	-

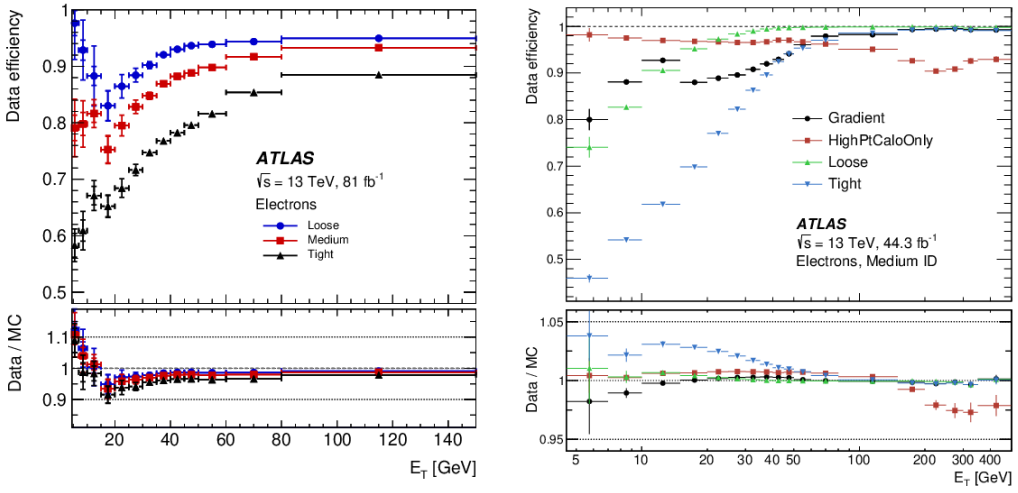
TABLE 3.2: A table showing different electron isolation working point definitions.

3.3.4.4 Simulation scale factors

Efficiency scale factors (SF), which are adjustments made to MC simulations, are used to match the efficiency of data's reconstruction, identification, and isolation. These efficiencies are determined using a tag-and-probe method, which is based on the decays of particles that produce pairs of particles, and in this case a pair of electrons. These electrons are identified as the tag if they pass a strict selection criteria, whereas the probe is considered an unbiased set of electrons that pass very loose selection criteria (either passing or failing the selection for which the efficiency is being measured). The efficiency is then the fraction of probe electrons that pass the selection criteria. The tag electron is used to ensure the decay of the two electrons originated from the same particle. These efficiencies were estimated from $J/\psi \rightarrow ee$ and $Z \rightarrow ee$ events, and the SF are derived from the ratio of efficiencies found in data with the ones in MC simulations. Most analyses select electrons based on the trigger it fired, the identification criteria, and the isolation requirement, while all analyses must consider the reconstruction. Each selection used on the electron has an efficiency attached to it and therefore has to be considered to correct the MC prediction to data. The following is a definition of the considered electron SF:

$$SF_e = SF_{\text{trigger}} \cdot SF_{\text{reconstruction}} \cdot SF_{\text{identification}} \cdot SF_{\text{isolation}} \tag{3.1}$$

The electron isolation and reconstruction efficiencies as a function of E_T are shown in Figure 3.14.



(a) (b)

FIGURE 3.14: The electron identification (a) and isolation (b) efficiencies in $Z \rightarrow ee$ events in data as a function of E_T for different identification operating points and isolation working points (where the electrons are required to fulfil the Medium selection from the likelihood-based electron identification). The lower panels show the ratio of the efficiencies measured in data and in MC simulations. The total uncertainties are shown, including the statistical and systematic components [94].

3.3.5 Muons

In the ID and MS, muon reconstruction and identification are carried out individually. The reconstruction of muons can be aided by small energy deposits found in the calorimetric systems. The muon tracks utilized in physics analyses are created by combining the information from the various sub-detectors. All the information provided in this section comes from reference [95], and should be consulted for more information.

3.3.5.1 Reconstruction

The reconstruction of muons with a momentum greater than 3 GeV is possible with the muon spectrometer. Muons with momenta below that are challenging to reconstruct because they either fail to reach the spectrometer due to excessive energy loss in the calorimeter or because they do not leave a discernible signal above the background noise in the muon spectrometer.

In order to build segments for the muon reconstruction, hit patterns are first looked for in each muon chamber. The Hough transform is employed to carry out the search for hits of the MS's MDT chambers, which are aligned on a trajectory on the detectors' bending plane. Measurement of the coordinate orthogonal to the bending plane is done by RPC or TGC hits. Segments are built for searching in the η and ϕ planes in the CSC chambers. The muon track candidates are reconstructed by fitting together hits of segments from several levels after the separated search in each component of the MS. Except at the barrel end-cap transition region, where a single high-quality segment containing η and ϕ information can be utilized to build a track, at least two matching segments must be present for the track to be accepted. The inner tracker reconstructs the tracks in the manner outlined in Section 3.3.2. The information from the ID, MS, and calorimeter are combined in five different methods to produce definitions of muons, each of which differs for the fake discrimination and the η coverage:

- Combined (CB) muons: these muons are distinguished by a track reconstruction that was carried out independently in the ID and MS. A combined track fit based on ID and MS hits are used to match the particle to the two reconstructed tracks while accounting for calorimeter energy loss. Starting with the MS track, the bigger portion of the muons is rebuilt, projected inward, and matched to a track in the inner detector. These are exclusively defined in the area when $|\eta| < 2.5$.
- Inside-out combined (IO) muons: without an MS reconstructed track, the algorithm to reconstruct IO muons is based on an ID track and hits in the MS, recovering some efficiency in the process. In order to look for MS hits to be used in a combined track fit, the trajectory is reconstructed by projecting the ID tracks to the MS.

- Segment-tagged (ST) muons: these muons have an ID track that matches at least one track segment in the MDT or CSC chambers when extended to the MS regions. Due to low p_T or because they fall in areas with limited MS acceptance, the ST muons are used when muons only pass through one layer of MS chambers.
- Calorimeter-tagged (CT) muons: by projecting ID tracks through the calorimeters to look for energy deposits consistent with a minimum-ionizing particle, these muons are detected.
- MS Extrapolated (ME): are muons where an MS track cannot be matched to an ID track. Only the MS track and a lax criterion on conformity with coming from the IP are used to reconstruct the muon trajectory. These muons are employed to expand the acceptance beyond the ID while taking use of the complete MS coverage up to $|\eta| < 2.7$.

Similar to the electrons, muons have an additional criterion known as the Track-To-Vertex-Association (TTVA). For muons, the selections are $|d_0/\sigma_{d_0}| < 3$ and $|\Delta z_0 \cdot \sin\theta| < 0.5$ mm, where d_0 is the transverse impact parameter, σ_{d_0} the associated uncertainty, Δz_0 is the distance between the track's longitudinal impact parameter (z_0) and the primary vertex, and θ is the polar angle of the track.

3.3.5.2 Identification

Loose, Medium, Tight, High-p_T, and Low-p_T are the five sets of selection criteria that are specified with diminishing reconstruction efficiency but increasing purity. The rejection of light hadrons, which often produce muon tracks of inferior quality, is the focus of the selection working points. Good quality muon tracks are produced by charm and bottom decays, which can be separated from prompt muons by requiring isolation in the tracker and/or calorimeters and association to the primary vertex.

All muon types are included in the Loose identification criteria, which are tailored for Higgs searches. Except in the $|\eta| < 0.1$ area, where muons are considered but can only have one muon precision station, it allows CB and ME muons, needing at least two precision stations. To guarantee a loose comparability between the ID and MS measurements, the q/p significance must be smaller than 7; $\text{q/p} = \dfrac{|q/p_{ID} - q/p_{MS}|}{\sqrt{\sigma^2(q/p_{ID}) - \sigma^2(q/p_{MS})}}$ where q/p_{ID} and q/p_{MS} are measurements in the ID and MS of the ratio between the charge, q, and the momentum, p, of the muon expressed at the interaction point, while $\sigma^2(q/p_{ID})$ and $\sigma^2(q/p_{MS})$ are the corresponding uncertainties. By including ME muons, which need to have at least three precision stations, in the range $2.5< |\eta| <2.7$, the acceptance is expanded outside the ID coverage. The loose selection only permits CT and ST muons with $|\eta| <0.1$. IO muons with $p_T < 7$ GeV and a single precision station are accepted in the range $|\eta| < 1.3$ to improve the effectiveness of the Loose criterion for Low-p_T muons, provided they are independently reconstructed as ST muons

as well. Approximately 97% of the prompt muons passing the Loose WP in $t\bar{t}$ events are CB or IO muons.

Only CB and IO muons with an ID acceptance of $|\eta| < 2.5$ and ME muons with an ID acceptance of $2.5 < |\eta| < 2.7$ are accepted by the Medium working point. It is a subset of Loose muons except for CT, ST and barrel-bound low-p_T muons. Medium muons, as opposed to Loose WP, offer superior prompt muon purity, resulting in smaller systematics in the efficiency calibrations. It is employed in a wide range of investigations, particularly the precision measurements of the Standard Model. More than 98% of prompt muons passing the Medium WP in $t\bar{t}$ events are CB muons.

Only muons with at least two precision stations that are CB or IO are accepted by the Tight selection. The background rejection is significantly improved at the cost of a small efficiency loss, but it offers the maximum level of purity. Depending on the muon's p_T and η, requirements are put on the track's χ^2, q/p significance, and $\rho^{'}$; $\rho^{'} = \frac{|p_{\mathrm{T}}^{ID} - p_{\mathrm{T}}^{MS}|}{p_{\mathrm{T}}^{CB}}$, where p_{T}^{ID} and p_{T}^{MS} are the p_{T} of the muon measured in the ID and MS, respectively, while p_{T}^{CB} come from the combined track fit.

The High-p_{T} WP guarantees an ideal measurement of momentum for muons with p_{T} greater than 100 GeV. It is optimized for the $\mathrm{W}^{'}$ and $\mathrm{Z}^{'}$ BSM searches.

Low-p_{T} WP identification must be based on MS segments, since these muons are less likely to be independently reconstructed as entire tracks in the MS. In order to make the implementation of data-driven estimates simpler, two versions of the Low-p_{T} WP have been developed: one that uses a multivariate (MVA) technique to maximize overall performance and the other that uses a cut-based selection to minimize the kinematic dependencies of the background efficiencies. Measurements of Standard Model parameters in the quark-mixing region, searches for Supersymmetry with compressed mass spectra, and precision top mass measurements with rare decay modes including muons from hadron decays are a few examples of analyses that profit from the usage of the Low-p_{T} WP.

3.3.5.3 Isolation

To increase the distinction from fake muons, an isolation requirement is imposed on them, just as it was for the electrons. Similarly, muons have track-based and calorimeter-based isolation, but also combines the two, providing a third isolation requirement known as particle-flow-based isolation. Because the two isolation variables offer complementing information, combining choices on track-based and calorimeter-based isolation often outperforms using just one. Compared to calorimeter isolation, track-based isolation has superior resolution and less pileup dependence, and the tracker gives better transverse momentum scale and resolution for individual

soft hadrons than the calorimeter. On the other hand, track isolation ignores neutral particles and particles that are below the ID track p_T threshold, but calorimeter-based isolation takes them into account. However, since both the calorimeter and the tracker measure charged particles, track and calorimeter isolation measures hadronic activity twice. The track-based and calorimeter-based isolation's overlapping contributions can be eliminated using the particle-flow technique, which also reduces the correlation between the two variables.

- $\mathrm{E_T^{topocone20}}$ is the sum of the $\mathrm{E_T}$ of topological clusters within a cone of $\Delta\mathrm{R} = 0.2$ around the position of the muon, extrapolated to the calorimeter, after removing the muon's energy deposits.
- $p_T^{\mathrm{cone20(30)}}$ is the sum of the p_T of all tracks originating from the primary vertex falling in a cone of $\Delta\mathrm{R} = 0.2(0.3)$ around the muon track, but excluding muon track itself;
- $p_T^{\mathrm{varcone20(30)}}$ is the sum of the p_T of all tracks with $\Delta\mathrm{R} = \min(0.2(0.3),\ 10/p_T^{\mu}\ [\mathrm{GeV}])$ around the muon track and excluding the muon track itself.
- $\mathrm{E_T^{neflow20}}$ uses $p_T^{\mathrm{varcone20}}$ for $p_T^{\mu} < 50$ GeV and p_T^{cone20} for $p_T^{\mu} > 50$ GeV, and the transverse energy of neutral particle-flow objects in a cone of $\Delta\mathrm{R} = 0.2$ around the muon.

The isolation efficiencies are measured in data and determined through simulation, and typically range from approximately 87% for the tightest working points to nearly 100% for the loosest, when needing muons with a $p_T^{\mu} > 20$ GeV. When considering muons with $p_T^{\mu} < 20$ GeV, the efficiency can drop as low as 45%. Several working points are identified for muons as well, and they are presented in Table 3.3.

Working point	Definition	Track requirement
Loose	$p_T^{\mathrm{varcone30}} < 0.15 \cdot p_T^{\mu}$, $\mathrm{E_T^{topocone20}} < 0.3 \cdot p_T^{\mu}$	$p_T > 1$ GeV
Tight	$p_T^{\mathrm{varcone30}} < 0.04 \cdot p_T^{\mu}$, $\mathrm{E_T^{topocone20}} < 0.15 \cdot p_T^{\mu}$	$p_T > 1$ GeV
PflowLoose	$(p_T^{\mathrm{varcone30}} + 0.4 \cdot \mathrm{E_T^{neflow20}}) < 0.16 \cdot p_T^{\mu}$	$p_T > 500$ MeV
PflowTight	$(p_T^{\mathrm{varcone30}} + 0.4 \cdot \mathrm{E_T^{neflow20}}) < 0.045 \cdot p_T^{\mu}$	$p_T > 500$ MeV

TABLE 3.3: A table showing different muon isolation working point definitions.

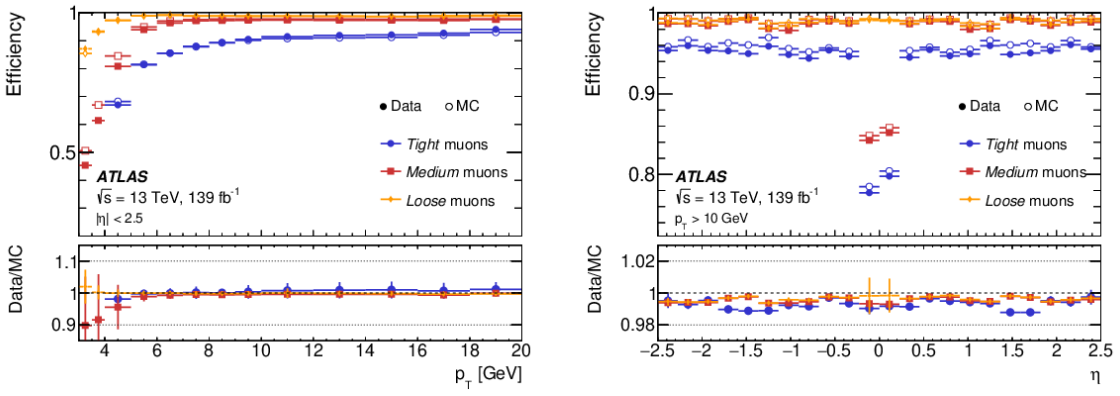
3.3.5.4 Simulation scale factors

In order to match the reconstruction, identification, and isolation efficiency in data, efficiency scale factors (SF) are used as corrections for the muons in Monte Carlo simulations. Depending on the $|\eta|$ area, two distinct approaches are utilized to quantify the high precision reconstruction, identification, isolation, and vertex association efficiency. Two independent detectors are provided in the region corresponding to the ID acceptance ($|\eta| < 2.5$), which are the ID and

MS, but only the MS detector is used to reconstruct muons in the $2.5 < |\eta| < 2.7$ region. A similar tag-and-probe method is used as mentioned for the electrons SF under Section 3.3.4.4, to measure the efficiencies but from $J/\psi \rightarrow \mu\mu$ and $Z \rightarrow \mu\mu$ events instead. The following is a definition of the muon scale factors:

$$SF_{\mu} = SF_{\text{trigger}} \cdot SF_{\text{reconstruction}} \cdot SF_{\text{identification}} \cdot SF_{\text{isolation}} \tag{3.2}$$

The muon reconstruction efficiencies as a function of transverse momentum and pseudo-rapidity are shown in Figure 3.15, and the isolation efficiencies as a function of transverse momentum are shown in Figure 3.16.



(a) (b)

FIGURE 3.15: Muon reconstruction and identification efficiencies for the Loose, Medium, and Tight criteria. The left plot shows the efficiencies measured in $J/\psi \rightarrow \mu\mu$ events as function of p_T. The right plot displays the efficiencies measured in $Z \rightarrow \mu\mu$ events as a function of η, for muons with $p_T > 10$ GeV. The predicted efficiencies are depicted as open circles, while filled dots illustrate the result of the measurement in collision data. The error bars on the efficiencies indicate the statistical uncertainty. The panel at the bottom shows the ratio of the measured to predicted efficiencies, with statistical and systematic uncertainties [95].

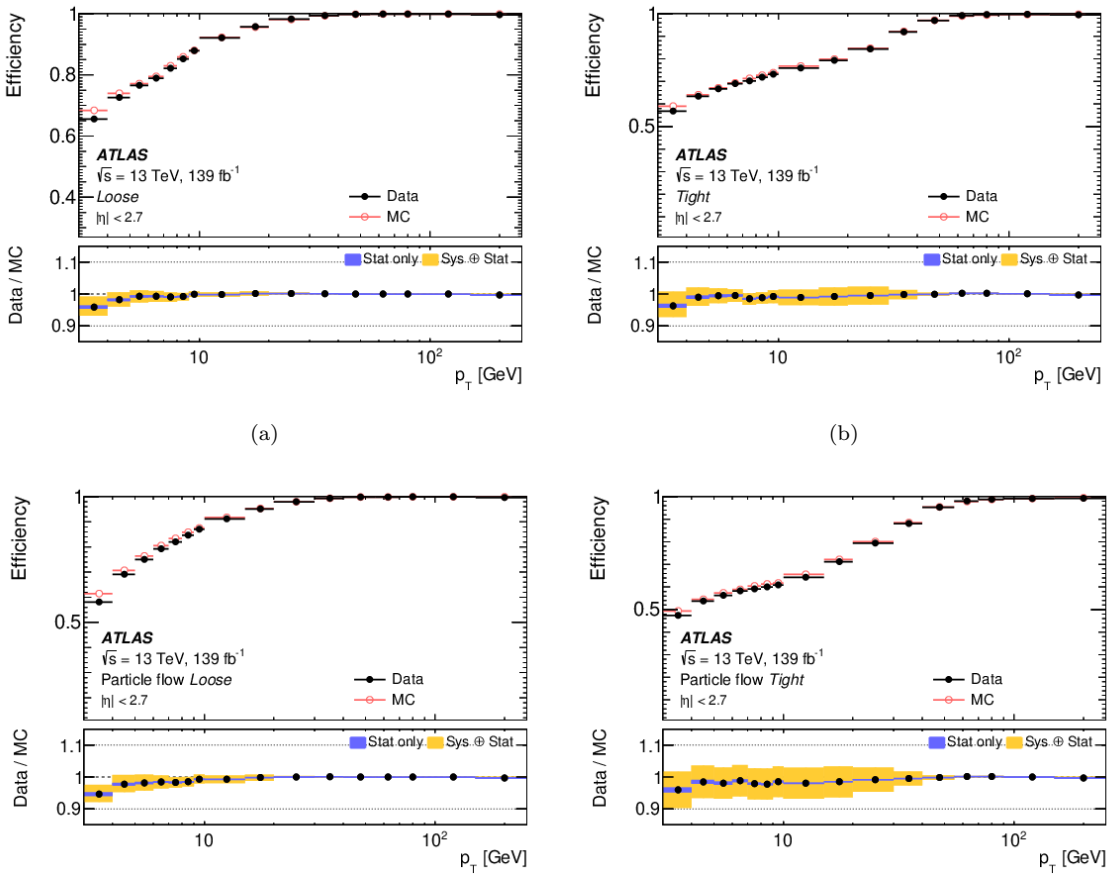


FIGURE 3.16: Muon isolation efficiency measured in Z→ $\mu\mu$ events for the Loose (a), Tight (b), PflowLoose (c) and PflowTight (d) criteria, as a function of p_T for muons with $p_T > 3$ GeV. The error bars on the efficiencies indicate the statistical uncertainty. The panel at the bottom shows the ratio of the measured to predicted efficiencies, with statistical and systematic uncertainties [95].

3.3.6 Jets

Jets are collimated hadron showers designed to capture all the by-products of hadronizing coloured particles (quarks or gluons). Since jets have the tendency to retain most of the momentum of the particles from which they originate, they can be used as a substitute for detecting quark or gluon momenta. Jets almost always contain both electrically charged and neutral objects due to the enormous number of hadrons they contain. The electromagnetic and hadronic calorimeters both contain energy deposits from sufficiently high energy jets. Either topo-cluster clustering (EMTopo jets) [96, 97] or the creation of so-called particle flow objects (PFlow jets) [98], which are made up of a combination of tracks and topo-clusters, are used to reconstruct jets. Since PFlow jets are the only ones employed in this thesis, EMTopo jets won't be mentioned in the following sections; instead, PFlow jets will simply be referred to as jets. A jet energy scale (JES) adjustment is used to calibrate reconstructed jets to the particle level after

they were first reconstructed at the EM scale of the calorimeter [99]. The jet-vertex-tagger (JVT) discriminant is then used to reduce pile-up jets.

3.3.6.1 Reconstruction

The particle flow (PFlow) algorithm [98] benefits from the fact that charged jet constituents leave behind tracks in the ID and energy deposits in the calorimeter. Combining measurements from these two components offer complementary energy measurements, as the calorimeter has superior resolution at high energy, the tracker has better resolution at low energies. Due to lower noise occupancy, the ID can also reconstruct significantly lower energy objects. Additionally, before entering the calorimeter, lower energy jet constituents are frequently redirected by the magnetic field from the initial jet cone. These occurrences can be discovered early and accurately connected to the jet cone thanks to the usage of ID information. Another advantage is that the ID has significantly better angular resolution than the calorimeter, but the calorimeter offers more forward coverage. The ability to reject signals coming from pile-up vertices is provided by the fact that tracks are linked to vertices.

The particle flow algorithm starts by choosing tracks with 0.5 GeV $< p_T <$ 40 GeV, are well measured, and have no matching previously reconstructed particles. The upper threshold is used to remove tracks that will not improve upon the current set of topo-clusters [96, 97], while the lower threshold is used to remove background and prevent spending compute time on tracks that have little impact on the final energy. Track positions are extrapolated to the calorimeter after track selection. The closest topo-cluster with an energy that is at least 10% of the track momentum and is located within ΔR < 1.64 is then matched to the tracks. Tracks that don't match are kept in the list of tracks and won't be used in further phases, while tracks with energies significantly greater than the matched cluster are assigned to other clusters until this is no longer the case, in order to ensure that showers split across many topo-clusters are correctly reconstructed. The energy contribution of those tracks that were successfully matched to a topo-cluster must then be deducted from it in order to prevent duplicate counting energy. If a track's projected energy deposit is more than its assigned topo-clusters, the topo-clusters are simply removed. Otherwise, this subtraction is carried out by removing rings of cells from each layer as it moves away from the projected energy density distribution's centre. The cell energies are scaled down to subtract the remaining track energy for the final ring, where the energy of the cells is greater than the remaining track energy. If the energy of the remaining topo-clusters and cells differs from the expected energy deposit of the track, they are eliminated. The chosen tracks and residual topo-clusters are regarded as fully reconstructed particle flow objects after this matching and subtraction process.

3.3.6.2 Anti-k_t algorithm

Before jet reconstruction occurs, pile-up is reduced by removing particle flow tracks that don't pass close enough to the primary vertex and by recalculating topo-cluster coordinates, mentioned under Section 3.3.4.1, relative to the primary vertex. The anti-kt jet clustering algorithm [100] is then used to recreate the jet after these changes.

The following section is taken from my master's thesis [37]. "This algorithm starts by comparing the distance between two clusters in the jet, d_{ij}, with the distance between one of the clusters and the beam, d_{iB}. These two variables are calculated as follows,

$$d_{ij} = min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right)\frac{\Delta R_{ij}^2}{R^2} \tag{3.3}$$

$$d_{iB} = \frac{1}{p_{T,i}^2} \tag{3.4}$$

where $p_{T,i}$ is the transverse momentum of the cluster i, R is the radial parameter used to define the size of the jet (usually set to 0.4) and ΔR_{ij} is the distance in the η-ϕ plane between the two clusters. The algorithm first determines all the d_{ij} and d_{iB} values and then selects the minimum between these distances. If d_{iB} is found to be the minimum, the cluster i is identified as a complete jet and is no longer considered in the algorithm. However, if d_{ij} is found to be the minimum, the four momenta of clusters i and j are combined and i and j are then identified as one cluster and the individual clusters are no longer considered in the algorithm. The process is reiterated until all the clusters are combined."

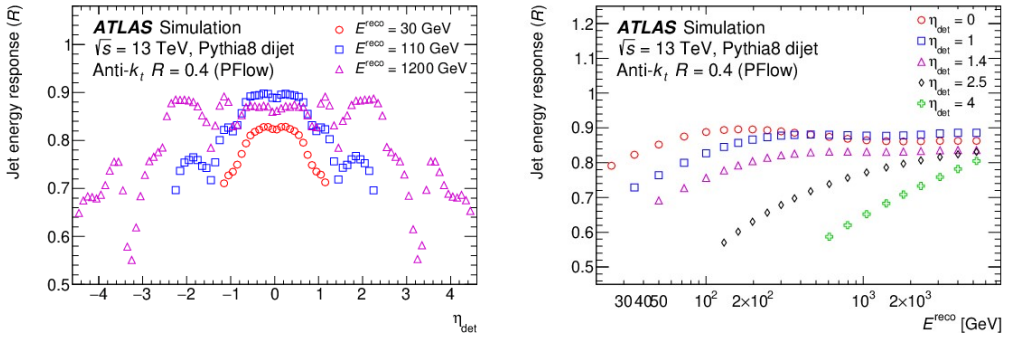
3.3.6.3 Jet Calibration

Calibrations were implemented to correct the jet energy scale of reconstructed jets to that of particle-level jets in MC simulations. There are various corrections implemented each correcting the four-momentum, energy, mass and the jet p_T scale. These variations must all be taken into consideration in order to determine the appropriate particle level energy for a particular jet. Jet calibration is the process used to do this [99].

- Pile-up correction: The calibration procedure's first two steps account for pile-up. In particular, this adjustment eliminates contributions from in-time and out-of-time pile-up as well as those from the underlying event, because pile-up from particles that left traces has already been eliminated. The k_t algorithm [101] with R = 0.4 is used to create a collection of k_t jets from positive energy topo-clusters in the range $|\eta| < 2$ in order to

calculate the pile-up contribution. This technique is employed because it reassembles a uniform background of pile-up jets naturally. Then, to determine the contributions from pile-up, the median momentum density of these k_t jets in the y–ϕ plane, ρ, is computed and multiplied by the area, A, of each anti–k_t jet. The p_T is then decreased by this amount by scaling the four-momentum of each anti-k_t jet. After that, a residual correction is used to account for variations in pile-up conditions between events. This adjustment depends on the event's NPV (number of vertices) and μ (mean number of interactions per bunch crossing) values. In order to take into consideration the various effects of these circumstances on various areas of the calorimeter as well as for jets with various p_T, separate residual adjustments are made for numerous bins of truth-level p_T and detector $|\eta|$.

- Absolute jet energy scale (JES) and η correction: The jet energy scale correction is used to account for the detector's fluctuating and imperfect response. A correction to the energy, plus a correction to the jet η, make-up this correction. The average jet energy response, as a function of the reconstructed jet energy and η, is depicted in Figure 3.17 as the average ratio of the reconstructed and true jet energies. The jet energy response varies greatly with energy and η. The greater fraction of jet energy present in the EM shower component for high energy particles is what causes the energy variance. The difference in η, however, is a result of the differing calorimeter and detector geometries.



(a) (b)

FIGURE 3.17: The average energy response as a function of reconstructed jet η and energy. Each value is obtained from the corresponding parametrized function derived with the Pythia8 MC sample, and only jets satisfying $p_T > 20$ GeV are shown [99].

- Global sequential correction (GSC): The Global Sequential Calibration (GSC), which aims to lessen the disparity in response between quark and gluon jets, consists of a track-based post-calibration adjustment. In actuality, a considerable portion of the jet's p_T is carried by the leading hadrons in a quark-initiated jet, as opposed to a gluon-initiated jet, which often contains much softer particles. To correct high-p_T jets whose energy is not entirely contained within the calorimeter, the GSC has a punch-through correction. The topology

of the calorimeter's energy deposits, the tracking data, and the MS data all influence the corrections that are made.

- Residual *in-situ* correction: Using in-situ measurements, a final residual calibration is created and applied solely to data. The goal of this last phase is to eliminate any remaining discrepancies between simulation- and data-based measurements of the jet energy response. The response of forward jets ($0.8 < |\eta| < 4.5$) is calibrated to the response of jets in the central region ($|\eta| < 0.8$) using an η inter-calibration, which is first derived from di-jet occurrences. In a subsequent phase, Z + jets events, in which the Z boson decays into an electron or a muon pair—are used to calibrate the jets p_T using the accurately reconstructed Z boson's recoil, up to 500 (950) GeV for electron (muon) pair, in the decay of the Z boson. Finally, a multi-jet approach is utilized to increase the calibration's p_T range up to 2 TeV by balancing a system of accurately calibrated lower p_T jets against a single very high-p_T jet.

3.3.6.4 Jet energy resolution

For accurate measurements of Standard Model jet production, characteristics of Standard Model particles that decay to jets, and searches for jet-related physics outside the Standard Model, it is critical to have a precise understanding of the jet energy resolution (JER). Three distinct contributions can be used to parametrize the JER-jet p_T dependence:

$$\frac{\sigma(p_T)}{p_T} = \frac{N}{p_T} \oplus \frac{S}{\sqrt{p_T}} \oplus C \tag{3.5}$$

where S is the stochastic term resulting from statistical fluctuations in the amount of deposited energy, N is the noise term resulting from the electronics as well as that resulting from pile-up, and C is the constant term containing depositions of energy in passive materials, the origin of hadron showers, and variations in response across the calorimeter. The JER is measured from di-jet events using the reconstructed energy imbalance between the two jets, and ranges from 0.25 to 0.04 for particle-flow jets as a function of jet p_T [99].

3.3.6.5 Jet-vertex-tagger

Both in-time pile-up and out-of-time pile-up have the potential to contaminate the reconstructed topo-clusters in the calorimeter. While the second results from pile-up interactions from the previous and following bunch crossings, the first pertains to multiple parton interactions inside the same bunch crossing. These fake jets that were created by pile-up fluctuations were rejected using the jet-vertex-tagger (JVT), a likelihood discriminant.

3.3.7 b-tagged jets

The extended decay time of hadrons including b-quarks ($\sim$ 1.6 ps [102]) makes it possible to identify a jet coming from a b-quark. As a result, there is a secondary vertex that is located a few millimetres away from the primary vertex where the B-hadron was produced. This secondary vertex can be used to determine that the jet is initiated by a b-quark. In addition, the b-tagging algorithms take into account the high B-hadron decay multiplicity as well as the characteristics of b-quark fragmentation.

Secondary vertices are determined using a single secondary vertex finding algorithm called the SSVF [103]. The SSVF can use any tracks that were either not included in the initial batch for primary vertex reconstruction or were not associated with a vertex at that time. It will only consider the tracks that fall within the cone of a jet. SSVF is based on the fact that for a n track vertex, all tracks must be sufficiently close to a common space point so that the vertex formed by any two of these tracks is also close to the n track vertex. In a high-density tracking environment, however, tracks that start at a primary vertex can occasionally get close enough to produce apparent secondary vertices that could be misidentified by such a basic algorithm. As a result, an additional step is added that rejects many two-track vertices that are incompatible with the vertex of interest. This rejection is performed by SSVF by requiring that two-track vertices have no tracks, with hits in layers closer to the detector's centre than the secondary vertex. The SSVF rejects two-track vertices that are consistent with background processes like long-lived particle decays, hadronic interaction with the detector, or photo-conversion because it is focused on flavour-tagging. The remaining two-track vertices are then combined into a multi-vertex by merging those that are close in space. If the resulting vertex has a > 6 GeV invariant mass or a high χ^2, the track with the highest χ^2 of the track-vertex association is removed from the set, and the process is repeated. This procedure is repeated until the vertex meets certain requirements, but if the final vertex only has two tracks, the vertex cleaning procedure is repeated.

A neural network called the DL1r tagger [104] produces a multidimensional output that corresponds to the likelihoods that a jet originated from a b-quark (b-jet), light-quark or gluon (light-jet). The DL1r tagger uses the impact parameter significance, invariant mass and multiplicity of the tracks matched to a jet as well as secondary vertices to distinguish the different flavour of the jet. The neural network is trained to distinguish b-jets from a background of light-jets and c-jets using simulated $t\bar{t}$ events.

Chapter 4

Experimental signature and event topology

Since the top quark factory has been built and the data obtained, the top quark mass can be measured.

The data consist of proton-proton collisions that produce many different interactions and particles. As mentioned in Section 2.2, the top quark can be produced alone or in pairs, and will predominantly decay into a W-boson and a b-quark. The W-boson can either decay leptonically into a lepton and its corresponding neutrino, or hadronically into two quarks. The b-quark undergoes hadronization and the B-hadrons can decay into a range of different particles.

This analysis selects events consisting of a W-boson, which decayed leptonically into either an electron or a muon and its corresponding lepton neutrino, and a b-quark that undergoes hadronization, in which the B-hadron decayed into a J/ψ meson which in turn decayed into two oppositely charged muons. This top quark decay mode of interest is represented in Equation 4.1.

$$t \to W(\to l\nu)b(\to J/\psi[\to \mu^+\mu^-] + X) \tag{4.1}$$

The final state particles in this selection, which will be called the lepton + J/ψ channel throughout the rest of this thesis, can be found in many SM processes. Therefore, in order to characterize the backgrounds associated within this channel, the theoretical predictions of signal and background have to be taken into account. There are also backgrounds which could have come from jets incorrectly identified as leptons (known as Fake Leptons, FL) or from leptons produced in jets which leave tracks in the detector (known as Non-Prompt leptons, NP). This contribution will be known as NPFL throughout the rest of this thesis.

4.1 Data and simulated samples

The analysis is performed using proton-proton collision data collected by the ATLAS detector between 2015 and 2018 at $\sqrt{s} = 13$ TeV corresponding to a total integrated luminosity of 139.0 fb^{-1}, where 3.2 fb^{-1} was produced in 2015, 33.0 fb^{-1} in 2016, 44.3 fb^{-1} in 2017, and 58.5 fb^{-1} in 2018. The uncertainty in the combined 2015-2018 integrated luminosity, obtained using the LUCID-2 detector [105] for the primary luminosity measurements, is 1.7% [106]. The data collected required all sub-detectors to be operational and stable beam conditions to be met.

4.1.1 Simulated samples

To assess and estimate the signal and background contributions, simulated events are used to develop the analysis and compared to data. These simulated events are generated by MC generators, which simulate proton-proton collisions, particle interactions, and particle kinematics. Detector simulations are then applied to the generator outputs. Each MC simulation is intended to generate and simulate different SM processes (also called MC simulated samples or simply MC samples). In all MC samples, the EvtGen [107] generator is used to model the decays of b- and c-flavored hadrons. Since this analysis focuses on the production of top quarks, both the production of top quark pairs and single top quarks will result in signal final state particles.

Pythia 8 [108] is used to generate additional proton-proton collisions in order to simulate the effect of pile-up. In MC simulations, the effect of pile-up is corrected to match the average number of interactions per bunch crossing observed in data. Data and all MC samples use the same algorithms and quality cuts used to reconstruct particles. Scale factors derived from dedicated data samples are also applied to MC samples to improve the agreement with data.

4.1.1.1 Signal processes

The $t\bar{t}$ and three single top quark processes mentioned in Section 2.2 were simulated using the next-to-leading order (NLO) matrix element (ME) event generator Powheg-Box v2 (r3026) [109–112] with the NNPDF3.0NLO parton distribution function (PDF) set [113]. These generated events were interfaced to Pythia 8 (v8.230) [108] with the NNPDF2.3LO PDF set [114] and the A14 tune [115] for the parton shower, hadronization, and underlying event modelling. Powheg-Box has a parameter which sets the cut-off scale for the first gluon emission called h_{damp}, and is set to $\frac{3}{2}m_{top} = 258.75$ GeV. A diagram removal scheme is employed to treat the interference between single top and $t\bar{t}$ production [116–118]. These samples were simulated with the full ATLAS detector simulation [87].

Simulated events are used to develop the method that extracts the top quark mass in data, and is described in Chapter 5. Nine different top quark mass samples of $t\bar{t}$ and single top events were generated, i.e. m_{top} = 169, 171, 172, 172.25, 172.5, 172.75, 173, 174, 176 GeV. These top quark mass samples were simulated with the fast ATLAS simulation [89, 90] using Powheg-Box v2 (r3026) matrix-element generator interfaced to Pythia 8 as the parton shower, hadronization, and underlying event modelling, with the same settings as mentioned in the previous paragraph.

4.1.1.2 Background processes

There are SM processes which produce the same final state particles within this lepton+J/ψ channel other than top quark production processes, known as background processes. These backgrounds could originate from W+jets, Z+jets, various diboson (i.e. WW, WZ and ZZ), top quark pair production in association with bosons (called $t\bar{t}V$), and $J/\psi + W$ processes. These samples were processed through the full ATLAS detector simulation [87].

The W+jets and Z+jets MC samples were generated using Sherpa v2.2.1 [119] where the matrix elements at NLO for up to two partons and the LO for up to four partons are calculated using OpenLoops [120] and Comix [121], respectively. The parton shower [122] and hadronization [123] are done within Sherpa using dedicated tuned parameters based on the NNPDF3.0NNLO PDF set.

The Diboson (VV) samples were generated using Sherpa v2.2.1 and v2.2.2, where two- and three-lepton decays were generated using the latter. The parton shower [122] and hadronization [123] are done within Sherpa using dedicated tuned parameters based on the NNPDF3.0NNLO PDF set.

The $t\bar{t}V$ samples were generated using MadGraph5_aMC@NLO (v2.3.3) [124] where the matrix elements were calculated at NLO using the NNPDF3.0NLO PDF set. The events were interfaced with Pythia 8 and uses the A14 set of tuned parameters and the NNPDF2.3LO PDF set [125].

The W+jets MC sample, however, does not account for events where a W-boson and a J/ψ are produced from one parton interaction (known as single parton scattering or SPS), nor where the two particles are produced from two different parton interactions (known as double parton scattering or DPS). The DPS MC sample was generated with Pythia 8 and the parton shower and hadronization was also through Pythia 8 using the AU2 tune [126] and CTEQ6L1 PDF set [127]. The SPS MC sample was generated using HelacOnia (v2.5.5) [128] with the CTEQ6L1 PDF set, and interfaced with Pythia 8 using the AU2 set of tuned parameters for the parton shower and hadronization.

4.1.1.3 MC cross-section values

Since each MC sample was produced with numerous events to void any dependence on the statistical uncertainty due to the amount of events generated, the number of events in the corresponding process in data will not match. Therefore, each MC sample will be normalized to their theoretical cross-section and integrated luminosity of the data.

The $t\bar{t}$ production cross-section is calculated at next-to-next-to-leading-order (NNLO), including the next-to-next-to-leading-logarithm (NNLL) soft gluon terms [129–133], and is set to $831.8^{+40.2}_{-45.6}$ pb^{-1} at $\sqrt{s} = 13$ TeV assuming a top quark mass of 172.5 GeV [134]. The single-top quark cross-sections are taken at NLO+NNLL QCD cross-sections, and are set to $216.99^{+9.04}_{-7.71}$ pb^{-1} for the t-channel, $10.32^{+0.40}_{-0.36}$ pb^{-1} for the s-channel, and 71.7±3.8 pb^{-1} for the tW process [135, 136].

The cross-sections of the background processes are normalized to their NNLO predictions [137–141], except the $J/\psi + W$ samples, which are normalized to the NLO predictions [142]. The uncertainties associated with these processes are taken from references [137–142] and described in Table 4.2.

<table>
<tr><td>Physics process</td><td>Uncertainty</td></tr>
<tr><td>$t\bar{t}V$</td><td>13% uncertainty on the cross-section values.</td></tr>
<tr><td>VV</td><td>50% uncertainty on the cross-section and additional jet production.</td></tr>
<tr><td>$J/\psi + W$</td><td>100% uncertainty on the cross-section values as they are not well-known.</td></tr>
<tr><td>W+jets</td><td>- 5% uncertainty on the cross-section values
- 30% normalization uncertainty to cover any data/MC discrepancies.</td></tr>
<tr><td>Z+jets</td><td>- 5% uncertainty on the cross-section
- normalization uncertainty that is dependent on the jet multiplicity that accounts for the scale, α_s and PDF variations. Table 4.1 show the uncertainty percentages per jet multiplicity.

<table>
<tr><td>Jet multiplicity</td><td>σ [%]</td></tr>
<tr><td>0</td><td>1.2</td></tr>
<tr><td>1</td><td>2.3</td></tr>
<tr><td>2</td><td>16.9</td></tr>
<tr><td>3</td><td>40.5</td></tr>
<tr><td>4</td><td>53.0</td></tr>
<tr><td>5</td><td>56.8</td></tr>
<tr><td>6</td><td>57.2</td></tr>
<tr><td>≥ 7</td><td>56.1</td></tr>
</table>
TABLE 4.1: Percentage of uncertainties based on jet multiplicity in Z+jets samples [137].</td></tr>
</table>

TABLE 4.2: Description of the various uncertainties associated to the different physics processes.

4.2 Preselection

Events are required to consist of at least one electron or muon with a p_T greater than the minimum p_T trigger threshold at the HLT level. The minimum trigger threshold requires electrons (muons) to at least have a $p_T > 25$ (21) GeV and an identification of medium (loose) in 2015, and $p_T > 27$ (27) GeV with an identification of tight (medium) in 2016-2018. Additionally, the electrons (muons) that fired the minimum triggers must have fired triggers at the L1 level with $p_T > 20$ (15) GeV. The identification requirement in the minimum trigger reduces the efficiency at much higher lepton p_T and therefore, additional triggers with higher p_T thresholds and either a looser or no identification requirements are used. These additional triggers require muons to have at least $p_T > 51$ GeV and no identification requirement in 2015-2018, whereas the electrons are required to have at least $p_T > 61$ GeV or 120 GeV, with identification requirements of medium and loose, respectively, in 2015-2018.

The data and simulated events that pass the triggers are then processed through the same preselection algorithm, which reduces the amount of events to analyse and selects events tailored to the decay mode of interest. These events have to consist of two oppositely charged muons, where the tracks of these muons were fitted to a common vertex and the invariant mass of the common vertex is within 2-3.6 GeV, corresponding to a window around the mass of a J/ψ particle ($\mathrm{m}_{J/\psi} = 3096.900 \pm 0.006$ MeV [6]). Additionally, the preselection process stores the fitted common vertex kinematic properties. Furthermore, it removes events with no muons or electrons with a $p_T > 20$ GeV within $|\eta| < 2.5$, and where the sum of the number of reconstructed muons and electrons is less than 3. Finally, events are rejected if jets arose from noise bursts in the ECal or if jet timing was incompatible with the event time.

4.3 Experimental Signature

The experimental signature involves final state particles from the decay mode of interest, shown in Equation 4.1. The selected events consisted of exactly one electron or muon that fired one of the lepton triggers, $\mathrm{E_T^{miss}} > 20$ GeV, an oppositely charged muon pair with an invariant mass around the mass of the J/ψ, i.e. within 2.9 and 3.3 GeV, at least two jets, and at least one b-tagged jet. The analysis included the additional selections of at least one b-quark and at least two jets to try and maximise the amount of top quarks by mainly selecting top quark pair events, with the other top quark decaying as $t \to W(\to qq')b$.

Additionally, the W-boson transverse mass is calculated by combining the kinematics of the lepton that fired the trigger and the neutrino (represented by E_T^{miss}, since neutrinos do not interact with the detector), using $m_T(\mathrm{lepton}, E_T^{\mathrm{miss}}) = \sqrt{2p_T(\mathrm{lepton})E_T^{\mathrm{miss}}(1 - \cos(\phi(\mathrm{lepton}) - \phi(E_T^{\mathrm{miss}})))}$. This quantity improves the selection of real W-bosons and therefore, an additional selection cut

of $m_T(\text{lepton}, E_T^{\text{miss}}) > 40$ GeV is applied. This experimental signature is also known as the signal region, and the reason for each selection can be found in Table 4.3.

Selection	Reason
Lepton that fired a trigger	Account for the lepton from the W-boson decay.
$E_T^{\text{miss}} > 20$ GeV	Account for the neutrino from the same W-boson decay mentioned above.
$2.9 < m(\mu^+\mu^-) < 3.3$ GeV	Account for the J/ψ from the b-quark
≥ 2 jets	Account for the W-boson decay from the other top quark
≥ 1 b-tagged jet	Account for the b-quark from the other top quark
$m_T(\text{lepton}, E_T^{\text{miss}}) > 40$ GeV	Try and ensure real W-bosons were selected

TABLE 4.3: The experimental signature and the reason for each selection.

4.3.1 W-boson: Electrons

Electron candidates are required to pass a tight-likelihood based selection criteria (which minimizes electrons reconstructed from photon pair-production of e^+e^- and suppresses hadrons identified as electrons), be found within the coverage of the ID, must have originated from the primary vertex, be isolated and have fired one of the electron triggers. Each event with an electron candidate, which passed all the selection criteria in Table 4.4, must not contain a muon candidate with $p_T > 25$ GeV and $|\eta| < 2.5$ to ensure only one of the W-bosons decayed leptonically.

Selection	Criteria
Reconstruction	TightLH
Isolation	FCTight
Transverse Momentum	$p_T > 25$ (27) GeV in 2015 (2016-2018)
Pseudorapidity	$\lvert\eta\rvert < 2.5$
Longitudinal impact parameter significance	$\lvert\Delta z_0 \sin\theta\rvert < 0.5$ mm
Transverse impact parameter significance	$\lvert d_0\rvert < 5\sigma_{d_0}$

TABLE 4.4: Electron candidates passing the above criteria were considered signal electrons.

4.3.2 W-boson: Muons

Muon candidates are required to pass a medium based selection criteria (which minimizes the uncertainty in muon reconstruction and suppresses hadrons being identified as muons), be identified as a combined muon, be found within the coverage of the ID, must have originated from the primary vertex, be isolated and have fired one of the muon triggers. Each event with a

muon candidate, which passed all the selection criteria in Table 4.5, must not contain an electron candidate with $p_T > 25$ GeV and $|\eta| < 2.5$ to ensure only one of the W-bosons decayed leptonically.

Selection	Criteria
Identification	Combined
Reconstruction	Medium
Isolation	PflowTight
Transverse Momentum	$p_T > 25$ (27) GeV in 2015 (2016-2018)
Pseudorapidity	$\|\eta\| < 2.5$
Longitudinal impact parameter significance	$\|\Delta z_0 \sin\theta\| < 0.5$ mm
Transverse impact parameter significance	$\|d_0\| < 3\sigma_{d_0}$

TABLE 4.5: Muon candidates passing the above criteria were considered signal muons.

4.3.3 Jets and b-tagged jets

Jet candidates must have had a minimum $p_T > 25$ GeV and be found within the coverage of the ID (i.e. $|\eta| < 2.5$). Jets are identified as b-tagged by the DL1r algorithm at a 77 % identification efficiency with a light-jet rejection, where rejection is defined as 1/efficiency, which rises quickly from 10 to 100 between 25 GeV and 50 GeV [104].

4.3.4 J/ψ

J/ψ candidates are reconstructed by combining the track and fit information from two oppositely charged muons in every event. The dimuon pair must have a post-vertexing invariant mass between 2.9 GeV and 3.3 GeV, $p_T > 8$ GeV and be found within a rapidity range of $|y| < 2.1$, to ensure high acceptance and efficiency of J/ψ candidates.

4.3.4.1 J/ψ: Muons

The muon candidates which decayed from the J/ψ meson must have had a track associated to a dimuon vertex, been reconstructed as either combined or segment-tagged muons, be found within the coverage of the ID (i.e. $|\eta| < 2.5$) and have had a minimum $p_T > 3.5$ GeV if $|\eta| < 1.3$ or $p_T > 3.0$ GeV if $|\eta| > 1.3$. If one of these muon candidates were reconstructed as the muon from the W-boson, the event was rejected. If there were more than two muon tracks associated to the same dimuon vertex, the event was rejected. The muon candidate from the W-boson and an oppositely charged muon candidate from the J/ψ decay were combined and, if the invariant mass of the combination was found to be within 81 GeV and 101 GeV, the events were rejected. This was to ensure the selected muons did not originate from a Z-boson decay.

4.3.4.2 Non-prompt J/ψ candidates

The following section is taken from my master's thesis [37].

"The selected J/ψ candidates could have originated from processes other than from top quark B-hadron decays. J/ψ candidates originating from top quark B-hadron decays and directly from the proton-proton collision are referred to as non-prompt signal and prompt J/ψ background mesons, respectively. There are, however, other background processes which include J/ψ candidates originating from B-hadron decays not coming from top quarks (i.e. non-prompt background) or from randomly combining two oppositely charged muons that have a mass near that of the J/ψ meson (i.e. prompt and non-prompt combinatorial background).

Since B-hadron decays exhibit displaced vertices from the primary vertex, prompt and non-prompt J/ψ candidates can be distinguished by studying their decay vertices. This can be done by using a pseudo-proper time variable which makes use of time dilation and length contraction, since the proton bunches collide at speeds near that of light. The proper decay time of B-hadrons is related to its proper decay length through $\tau = l/\mathrm{v}$, where l is the contracted length, v is the speed of the B-hadrons and τ is the dilated time variable. The proper decay length can be represented by its actual decay length undergoing time dilation through $l = \mathrm{L}/\gamma$, where L is the decay distance between the primary vertex and the B-hadrons decay vertex. Therefore, the proper decay time is related to the decay distance of the B-hadrons through $\tau = \mathrm{L}/(\gamma \mathrm{v})$. Using the relativistic momentum relation, i.e. $\mathrm{p} = \gamma \mathrm{mv}$, where m is the mass and p is the momentum of the B-hadrons, the proper decay time can be written as $\tau = \mathrm{Lm}/\mathrm{p}$. However, since the ATLAS detector cannot fully reconstruct the momentum of the B-hadrons, a good approximation would be to use the transverse momentum (and the mass) of the J/ψ candidates coming from B-hadrons. This will also aid in approximating the decay distance description, since the decay distance L can be projected onto the direction of the J/ψ candidates using the reconstructed transverse momentum. Thus, a "pseudo-proper time" variable τ can be used to represent the decay lifetime of B-hadrons, i.e.

$$\tau \equiv \frac{\vec{L} \cdot \vec{p_T}(J/\psi)}{p_T(J/\psi)} \frac{m_{\mu^+\mu^-}}{p_T(J/\psi)}, \tag{4.2}$$

where $\vec{L}$ is the displacement vector from the primary vertex to the J/ψ decay vertex and $\mathrm{m}_{\mu^+\mu^-}$ is the reconstructed mass of the J/ψ candidates (using the invariant mass of the dimuon pair). In theory, prompt J/ψ decay vertices should have a $\tau = 0$ ps (which represents a non-displaced vertex) whereas non-prompt J/ψ decay vertices should have a $\tau > 0$ ps (which represents a displaced vertex). Experimentally, however, these decay vertices will correspond to smeared τ values due to the vertex resolution of the detector. Therefore, an additional requirement of $\tau >$ 0 ps is applied to the J/ψ candidates."

4.3.5 Overlap removal

An overlap removal procedure is used to avoid counting the same energy clusters or tracks as different object types. This procedure is described below and is done consecutively. The last two criteria (5 and 6) are not applied to muons from the J/ψ, as these are expected to originate from a b-jet and will not pass.

1. Electron candidates are removed if they share a track with another electron candidate with a higher p_T.

2. Electron candidates are removed if they share a track with a muon candidate.

3. Jets are removed if they are found to be within $\Delta R < 0.2$ of an electron candidate.

4. Electron candidates are removed if they are found to be within $0.2 < \Delta R < 0.4$ of a jet.

5. Jets with more than two tracks are removed if they are found to be within $\Delta R < 0.2$ of a muon.

6. Muon candidates are removed if they are found to be within $\Delta R < 0.4$ of a jet.

4.4 Non-prompt and Fake Lepton Background

4.4.1 Overview

The identification of one or more charged isolated leptons from the decay of W-/Z-bosons, referred to as 'prompt' or'real' leptons in the following, is frequently used to select events with top quarks. These leptons are chosen based on detector acceptance, quality cuts, and isolation criteria.

However, due to the imperfect algorithms and detector signatures, non-prompt and fake leptons (NPFL) can be mis-reconstructed as prompt leptons. Non-prompt electrons and muons can arise from the decay of hadrons containing b- and c-quarks. Fake electrons can arise from photons pair-producing or from jets depositing large amounts of energy in the ECal. Fake muons, however, can arise from particles which emerge from highly energetic hadronic showers which leave tracks in the MS.

For events with one lepton, this background stems from multi-jet events, characterised by a cross-section several orders of magnitude larger than for W-boson or top quark events, and can

hardly be modelled with high precision. This background is instead estimated using data-driven techniques using a matrix-method technique. The number of background events in the signal region is evaluated by applying efficiency factors (fake and real lepton efficiencies) to the number of events passing a tighter (signal) and a looser lepton selection. The measurement presented in the following aims to measure these efficiencies in regions which are as close as possible to the signal region in order to avoid any possible differences.

4.4.2 Object Selection

For this analysis, the same selection criteria are applied for the lepton candidates (as described in Tables 4.4 and 4.5), b-tagged jets, and J/ψ candidates, except no pseudo-proper time selection is applied to the J/ψ candidates. If the electron (muon) candidates passed the entire selection criteria in Table 4.4 (4.5) then it is considered a "tight" electron (muon). A "loose" electron or muon candidate is defined as a lepton which passed the same selection criteria as the tight leptons, except the loose electron candidates have to pass an FCLoose isolation working point and a MediumLH identification, and the loose muon candidates have to pass a PFlowLoose isolation working point.

4.4.3 Matrix Method

The matrix method technique was adopted from reference [143] and is described below.

Since the signal experimental signature contains a single lepton, the number of events with one tight lepton (N^T) and one loose lepton (N^L) can be expressed as a linear combination of the number of events with a real or a non-prompt or fake lepton, i.e.

$$\begin{aligned} \mathrm{N^L} &= \mathrm{N^L_{real}} + \mathrm{N^L_{fake}}, \\ \mathrm{N^T} &= \epsilon_{\mathrm{real}}\mathrm{N^L_{real}} + \epsilon_{\mathrm{fake}}\mathrm{N^L_{fake}}, \end{aligned} \tag{4.3}$$

where ϵ_{real} and ϵ_{fake} are the fraction of real leptons in the loose selection which pass the tight selection and the fraction of non-prompt and fake leptons in the loose selection which pass the tight selection, respectively. Since $\mathrm{N^L}$ and $\mathrm{N^T}$ are measurable quantities, the number of events with a non-prompt or fake lepton can therefore be determined from Equation 4.3, if ϵ_{real} and ϵ_{fake} are known. These efficiencies are measured in data in control regions, where non-prompt and fake leptons are enhanced. The number of tight non-prompt and fake lepton background events can be determined using Equation 4.4, i.e.

$$\mathrm{N}^{\mathrm{T}}_{\mathrm{fake}} = \frac{\epsilon_{\mathrm{fake}}}{\epsilon_{\mathrm{real}} - \epsilon_{\mathrm{fake}}}(\epsilon_{\mathrm{real}}\mathrm{N}^{\mathrm{L}} - \mathrm{N}^{\mathrm{T}}). \tag{4.4}$$

Since both efficiencies depend on the number of loose and tight events, i.e. N^{T} and N^{L}, ϵ_{real} and ϵ_{fake} depend on the event signature and lepton kinematics. The efficiencies are parametrised as a function of the lepton's transverse momentum and pseudorapidity. To account for this non-prompt and fake lepton contribution, an event weight is computed using Equation 4.5, i.e.

$$\mathrm{w}_i = \frac{\epsilon_{\mathrm{fake}}}{\epsilon_{\mathrm{real}} - \epsilon_{\mathrm{fake}}}(\epsilon_{\mathrm{real}} - \delta_i), \tag{4.5}$$

where δ_i equals one if the loose event i passes the tight selection and 0 otherwise. The non-prompt and fake lepton contribution is estimated in each bin of the final observable by the sum of the w_i over all the events in that bin.

4.4.4 Measurement of the fake rates

The fake rates, ϵ_{fake}, are measured in a control region dominated by non-prompt and fake lepton background events, i.e. multi-jet events, but can still contain contributions from real leptons. This control region is defined to closely represent the signal region in order to estimate the NPFL more accurately, and the main difference is an orthogonal $\mathrm{E}_T^{\mathrm{miss}}$ selection. These events must satisfy the following selection criteria:

- At least one jet
- Exactly one loose electron (muon) that fired a lepton trigger
- Exactly one J/ψ candidate with no selection on τ
- $\mathrm{E}_T^{\mathrm{miss}} < 20$ GeV

This control region contains contributions from real leptons, which are modelled by different SM processes and therefore, to get a pure NPFL sample, the MC estimations of these contributions are subtracted from the data. The fake rate is determined using Equation 4.6.

$$\epsilon_{\mathrm{fake}} = \left(\frac{\mathrm{N}^{\mathrm{T}}_{\mathrm{fake}}}{\mathrm{N}^{\mathrm{L}}_{\mathrm{fake}}}\right) = \left(\frac{\mathrm{N}^{\mathrm{T}} - \mathrm{N}^{\mathrm{T}}_{\mathrm{real}}}{\mathrm{N}^{\mathrm{L}} - \mathrm{N}^{\mathrm{L}}_{\mathrm{real}}}\right). \tag{4.6}$$

The results are split per year and per lepton to account for the different pile-up distributions and triggers, respectively. Table 4.6 shows the number of events of the data and SM prediction

for loose and tight leptons in each channel split per year, as well as the total across all years. The yields are low for loose and tight leptons in both channels for every year, especially in 2015. Therefore, this control region is statistically limited across all years.

Channel	2015		2016		2017		2018		Total	
	data	model	data	model	data	model	data	model	data	model
Tight electron	135	64 ± 22	1025	685 ± 204	1434	1015 ± 285	1886	1359 ± 382	4480	3123 ± 519
Loose electron	239	66 ± 23	1339	711 ± 214	1921	1059 ± 302	2523	1425 ± 405	6022	3261 ± 549
Tight muon	86	66 ± 23	884	837 ± 233	1177	1190 ± 317	1556	1487 ± 417	3703	3580 ± 574
Loose muon	131	67 ± 24	1091	837 ± 244	1396	1235 ± 333	1856	1562 ± 444	4474	3701 ± 607

TABLE 4.6: The number of events of the data and model for each channel split per year when the lepton passes either the loose or tight selection criteria. The model uncertainties include the statistical and simulation uncertainties associated with the different SM processes.

The yields show deviations between the data and model within uncertainty in the electron channel, and agreement in the muon channel, for both loose and tight leptons. This indicates the presence of non-prompt and fake leptons (NPFL) in the electron channel and no NPFL in the muon channel.

The following results are shown for 2018, while the results for the other years can be found in Appendix A. The results for 2016 and 2017 show the same outcome, and therefore, the analysis for 2018 also applies to these years. The results for 2015 show slight differences when compared to 2018, but since it only accounts for $\sim$ 2-3% of the total yield, the same analysis for 2018 applies to 2015.

Figure 4.1 shows the comparison between the data and the SM prediction in the E_T^{miss} distribution when the lepton is either an electron (also called the electron channel) or a muon (also called the muon channel) for both loose and tight leptons, respectively, for 2018. These figures show large contributions originating from W/Z+jets and $J/\psi + W$ events, and small contributions from the other SM processes. The region between the top of the stacked simulated sources and the data is assumed to come from the NPFL background contribution. In the muon channel, the data and the SM prediction agree within uncertainty, indicating no presence of NPFL. In the electron channel, deviations between data and SM prediction indicate the presence of NPFL.

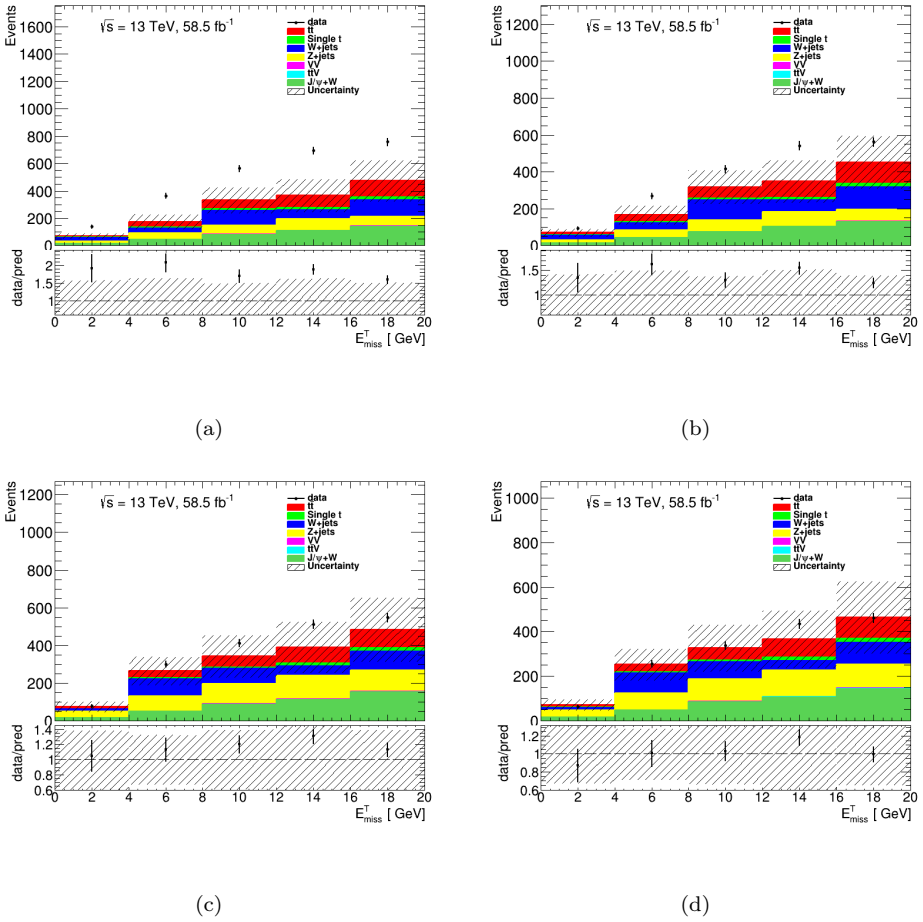


FIGURE 4.1: The upper panel shows the distribution of E_T^{miss} for data and the different SM processes for events passing the loose (a, c) and tight (b, d) selection criteria in the electron (top row) and muon (bottom row) channels for 2018, respectively. The statistical and luminosity uncertainties, together with the uncertainties associated with the SM processes, are shown. The lower panel shows the ratio between the data and the SM predictions.

Figures 4.2-4.3 show the lepton p_T and $|\eta|$ distributions for the data and the SM prediction for leptons passing the loose (left) and tight (right) selection criteria for 2018. The fake rate is determined by the difference between the number of tight leptons in the data subtracted by the number of tight leptons in the SM prediction, divided by the same differences in the loose lepton case. Since the tight lepton contribution represents the numerator and has "tighter" criteria, the difference in the data and SM prediction in the tight case is considered when analysing the results. In the electron channel, the presence of NPFL is clearly visible in the range $p_T < 35$ GeV and slightly visible between $35 < p_T < 40$ GeV, showing the presence of NPFL in that region. The data and SM prediction agree within uncertainty from $p_T > 40$ GeV and thus showing no presence of NPFL. The $|\eta|$ distribution shows the NPFL background is clearly visible between $1 < |\eta| < 2.5$, and no NPFL in $|\eta| < 1$. In the muon channel, the presence of NPFL is only visible in the range $p_T < 30$ GeV. The data and SM prediction agree within uncertainty from

$p_T > 30$ GeV, as well as the entire $|\eta|$ distribution, which shows no presence of NPFL in those regions.

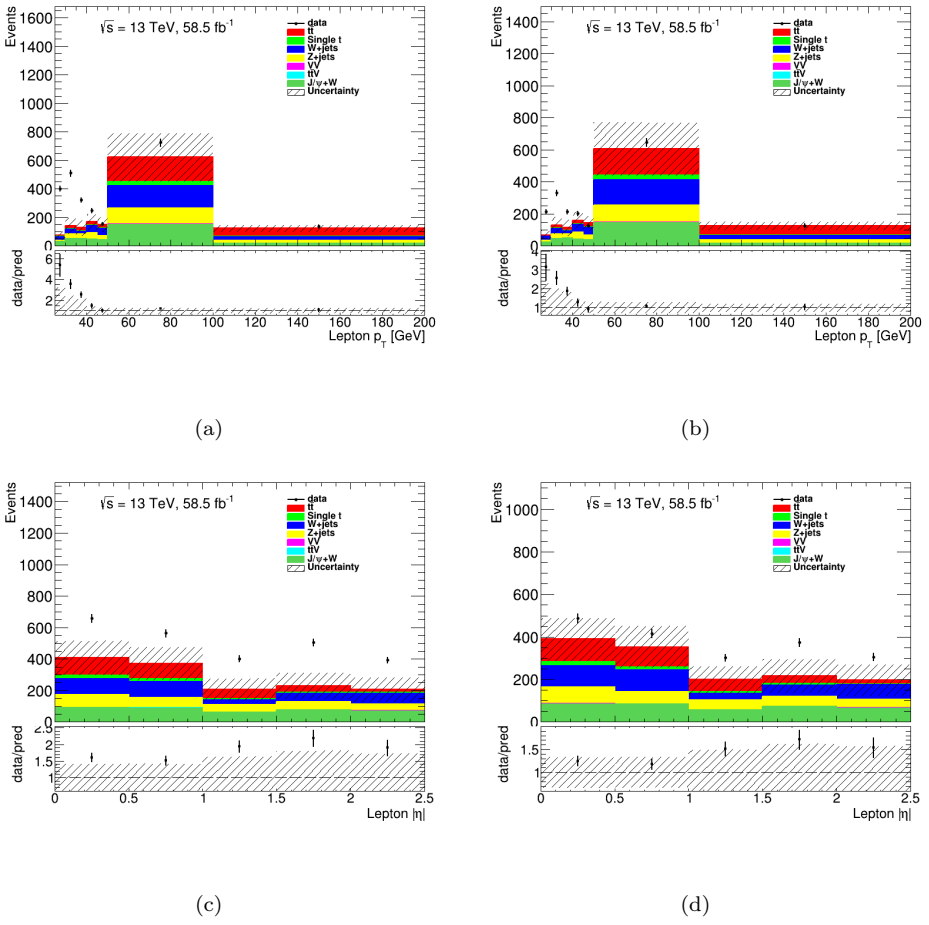


FIGURE 4.2: The upper panels show the distributions of p_T (a, b) and $|\eta|$ (c, d) for data and the different SM processes for events passing the loose (left) and tight (right) selection criteria in the electron channel for 2018. The statistical and luminosity uncertainties, together with the uncertainties associated with the SM processes, are shown. The lower panel shows the ratio between the data and the SM predictions.

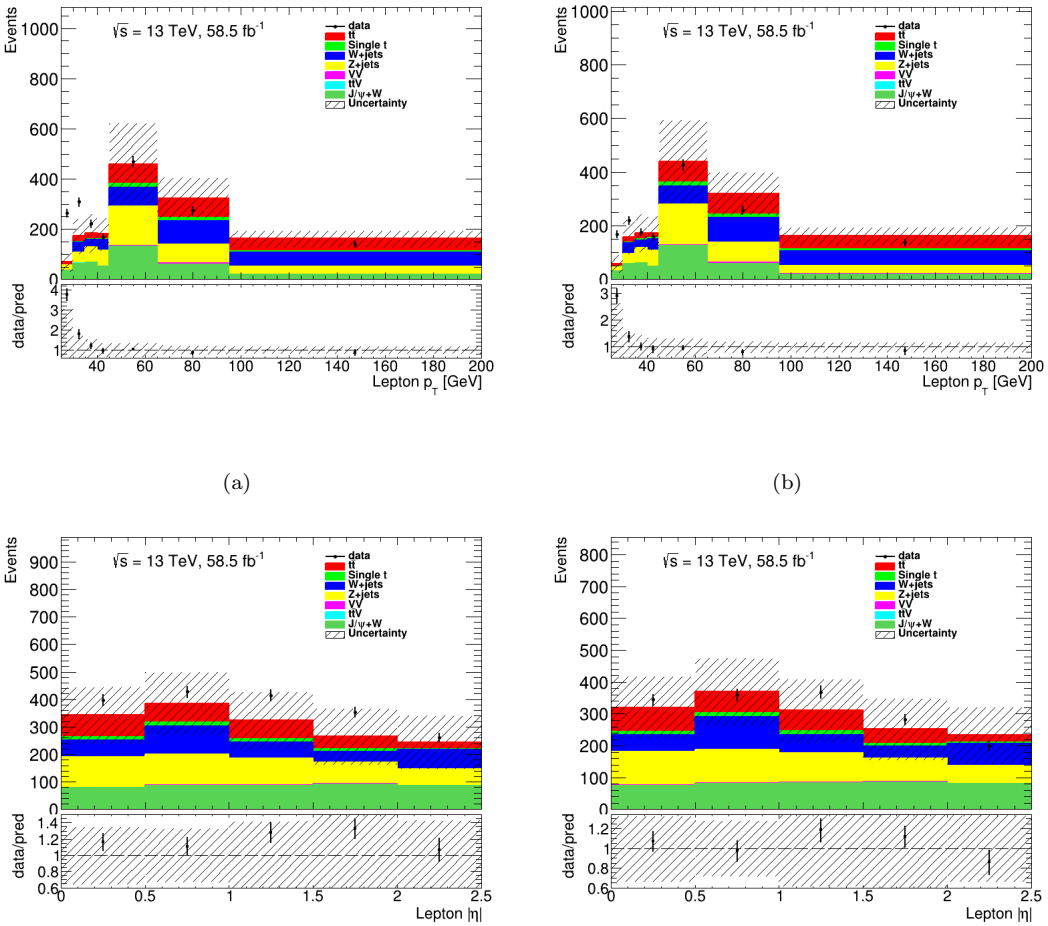


FIGURE 4.3: The upper panels show the distributions of p_T (a, b) and $|\eta|$ (c, d) for data and the different SM processes for events passing the loose (left) and tight (right) selection criteria in the muon channel for 2018. The statistical and luminosity uncertainties, together with the uncertainties associated with the SM processes, are shown. The lower panel shows the ratio between the data and the SM predictions.

4.4.5 Measurement of the efficiencies

The efficiencies, ϵ_{real}, are measured in a control region dominated by real leptons. The $t\bar{t}$ SM process is used to define the control region, where both W-bosons from the two different top quarks must have decayed leptonically. These events must satisfy the following selection criteria:

- At least one jet.
- Exactly two loose electrons (muons) with opposite signs, where both leptons must fire a lepton trigger.
- Exactly one J/ψ candidate with no selection on τ

The control region is defined for real leptons, and the efficiency is determined across the same object kinematics as the fake rates. The efficiencies are estimated using Equation 4.7.

$$\epsilon_{\text{real}} = \left(\frac{\mathrm{N}^{\mathrm{T}}}{\mathrm{N}^{\mathrm{L}}}\right). \tag{4.7}$$

The results are split per year and per lepton to account for the different pile-up distributions and triggers, respectively. Table 4.7 shows the number of $t\bar{t}$ events for loose and tight leptons in each channel split per year, as well as the total across all years. The yields are low for loose and tight leptons in both channels for every year, especially in 2015. Therefore, this control region is statistically limited across all years.

Channel	2015		2016		2017		2018		Total	
	Tight	Loose	Tight	Loose	Tight	Loose	Tight	Loose	Tight	Loose
Electron	16	16	132	136	175	184	255	267	578	603
Muon	10	10	107	110	140	144	190	197	447	461

TABLE 4.7: The number of $t\bar{t}$ events in each channel split per year when the leptons pass either the loose or tight selection criteria. The model uncertainties include the statistical and simulation uncertainties associated with the different SM processes.

The yields show negligible difference between loose and tight leptons in both channels and across all years. This indicates that loose and tight leptons are the same and the efficiencies should equal one.

The following results are shown for 2018, while the results for the other years can be found in Appendix A. The results for the other years show the same outcome as 2018 and, therefore, conclusions for 2018 applies to the other years.

Figure 4.4 shows the lepton p_T and $|\eta|$ distributions for leptons passing the loose (black) and tight (red) selection criteria in $t\bar{t}$ events for 2018. The results show the loose and tight contributions agree within uncertainty, however, since the comparison is made with the same MC sample, the tight contribution is systematically lower than the loose contribution.

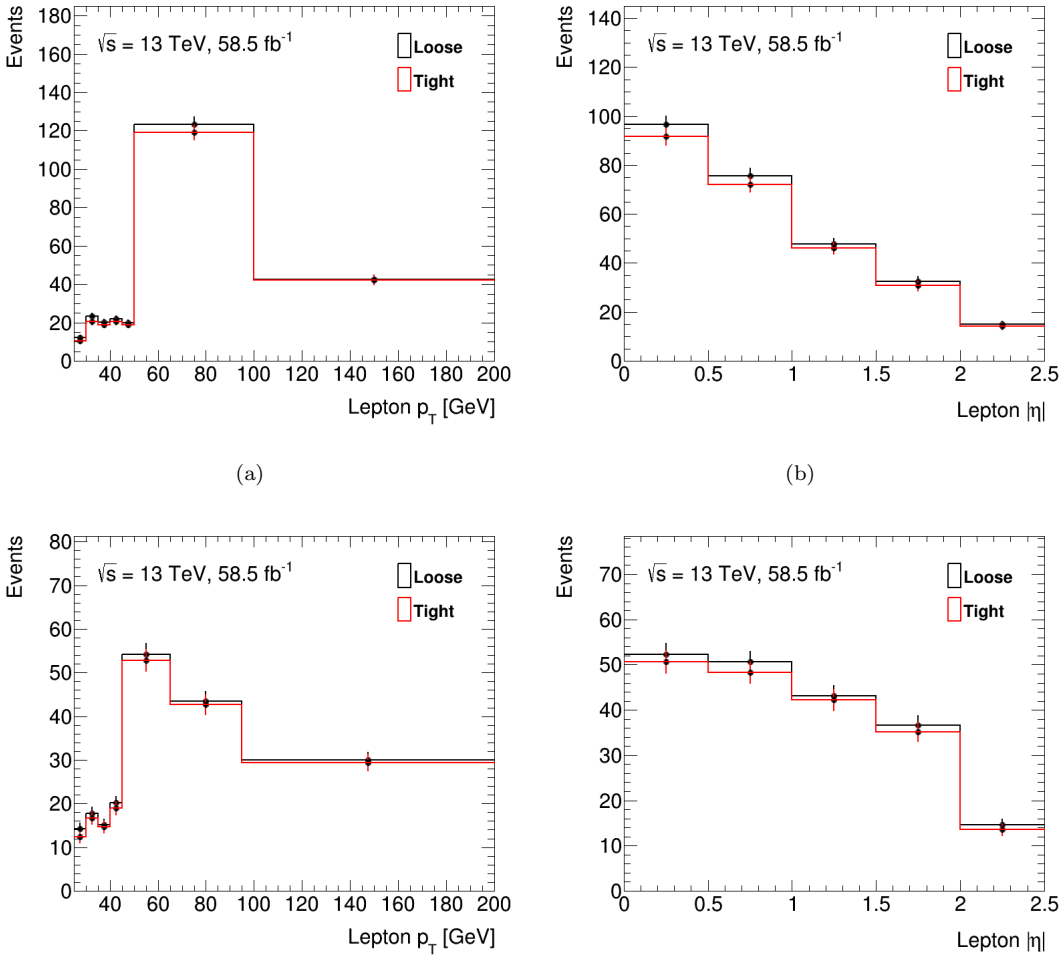


FIGURE 4.4: Distributions of p_T (a, c) and $|\eta|$ (b, d) for loose (black) and tight (red) leptons in the electron (top row) and muon (bottom row) channels for 2018. The error bars shows the statistical uncertainties.

4.4.6 Efficiencies and fake rates

Figure 4.5 shows the measured efficiencies and fake rates as a function of lepton p_T and $|\eta|$ in the electron and muon channels, respectively, for 2018. The results for the other years can be found in Appendix A. The results for the other years show the same outcome, and therefore, conclusions for 2018 also applies to the other years.

The fake rates and efficiencies have large statistical uncertainties due to the limited amount of events passing the selection criteria, which comes from the low branching ratio of having $J/\psi \to \mu\mu$ in every event. The efficiencies increase as a function of lepton p_T and are relatively constant across $|\eta|$. The fake rates, however, are relatively flat as a function of lepton p_T in the electron channel and decrease in the muon channel. Additionally, the fake rates vary across $|\eta|$

in both channels. The regions with no fake rate or efficiency are due to the fact that the data is smaller than the SM prediction. When estimating the NPFL contribution in the signal region, only the measured efficiencies and fake rates in regions where sizeable deviations between the data and the SM prediction are observed, will be used, while in the other regions, the amount of NPFL is considered to be zero.

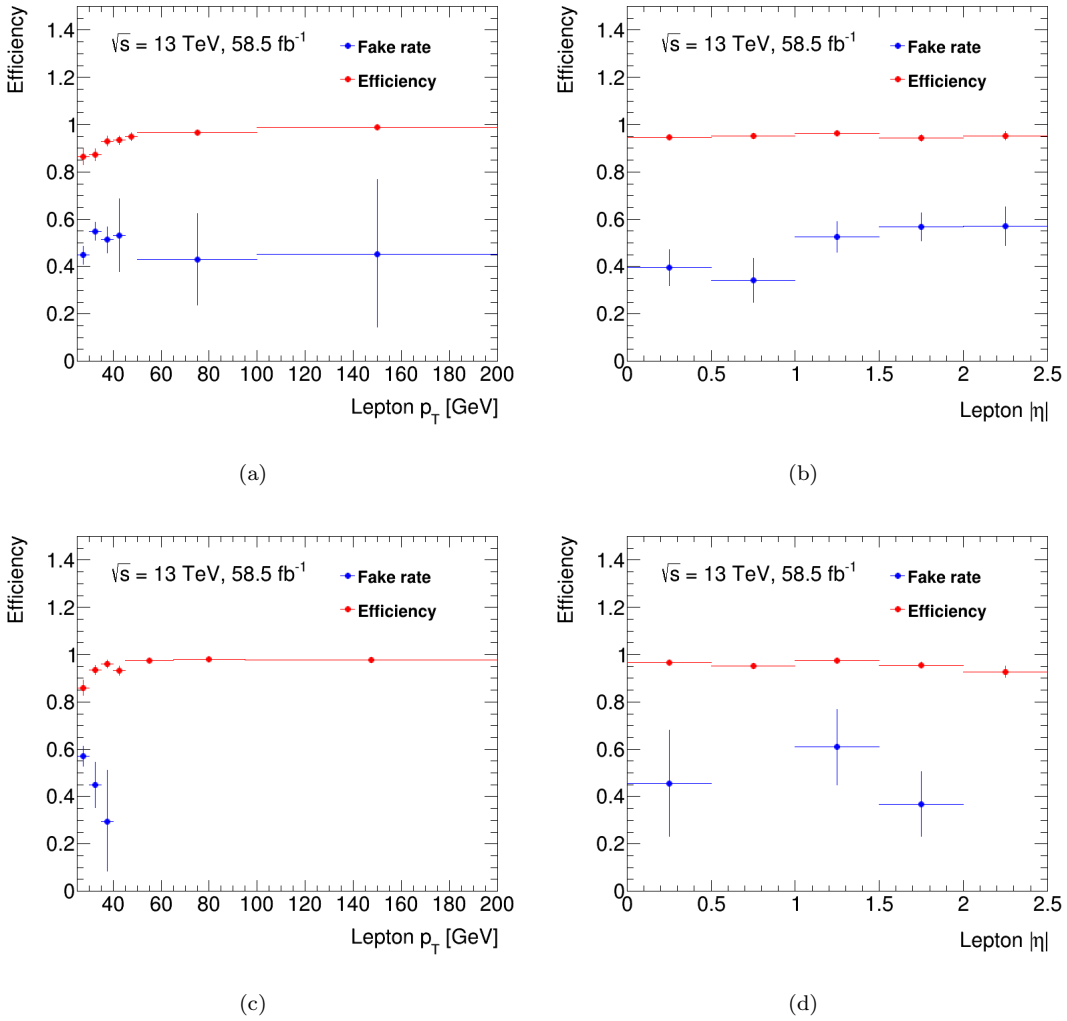


FIGURE 4.5: The measured efficiency (red) and fake rate (blue) in the transverse momentum (a, c) and absolute pseudorapidity (b, d) distributions in the electron and muon channels for 2018. The error bars correspond to the statistical uncertainty from the bin entries. No systematic uncertainties are shown.

4.4.7 Systematic uncertainties

The measured efficiencies and fake rates are shown with their statistical uncertainties and no systematic uncertainties.

The systematic uncertainties associated with the measured efficiencies can be derived from different control regions enriched in real leptons, i.e. using a tag-and-probe method with $Z \to \mu\mu$

and $Z \to ee$ events (as was done in reference [143]). The efficiencies in reference [143] were determined in data, whereas the efficiencies measured above are determined in MC simulation. The results are very similar despite this difference and, therefore, a negligible impact on the NPFL contribution is expected when compared to the large statistical uncertainties of the fake rates. At the time of writing, only the statistical uncertainty associated with the measured efficiencies are considered, and estimating the contribution from the different control region will be considered in future studies.

The systematic uncertainties associated with the measured fake rates comes from the simulation uncertainties of the different SM processes. Each process will be varied up and down by their respective uncertainties, and new fake rates will be determined. However, these SM processes are present in both the fake rate control region and the signal region. Therefore, the impact of varying each process on the top quark mass measurement is considered in both regions simultaneously, and is shown in Section 5.3.4.2.

4.5 Signal region object kinematics

The selected final state particles in the signal region are not unique to top quark processes and therefore, there are other SM processes which contribute to the background within the signal region.

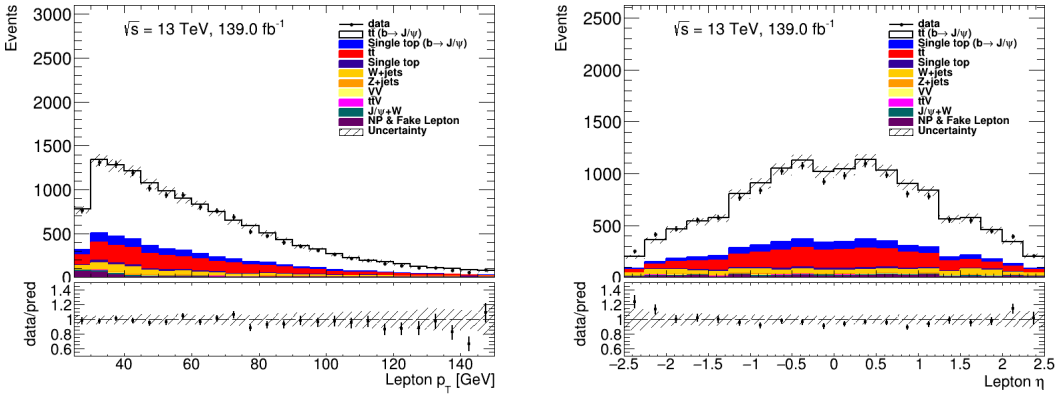
Table 4.8 shows the yields of the data and SM processes after all selections are applied. The statistical uncertainty and simulation uncertainties of the various SM processes are also shown. The signal region is dominated by top quark (signal) events, with a 91% contribution. These contributions are broken up into events that contain the top quark decay mode of interest, see Equation 4.1, at particle-level (73%) and those that don't (18%), which is referred to as $t\bar{t}$ in all figures and tables. There are minor contributions coming from the background processes (8%) and NPFL (1%). The model prediction agrees with the data within 1σ uncertainty.

Figures 4.6 - 4.8 show different object kinematic distributions for data and the individual contributions from signal and background SM processes, as well as the contribution from NPFL in the combined muon and electron channel. These object kinematics include the p_T and η of the lepton considered coming from the W-boson, $\mathrm{E}_T^{\mathrm{miss}}$, m_T(lepton, $\mathrm{E}_T^{\mathrm{miss}}$), the mass, p_T, rapidity and pseudo-proper time of the J/ψ, the p_T and η of the muons from the J/ψ, as well as the number of jets and b-tagged jets in the signal region. In each distribution, the data and prediction agree within 1σ uncertainty, except for the low $\mathrm{E}_T^{\mathrm{miss}}$ region, where the data and prediction agree within 2σ uncertainty. This shows a reasonably good agreement between the data and prediction in the signal region.

The data are well-described by the model, and the top quark mass in the data can be measured.

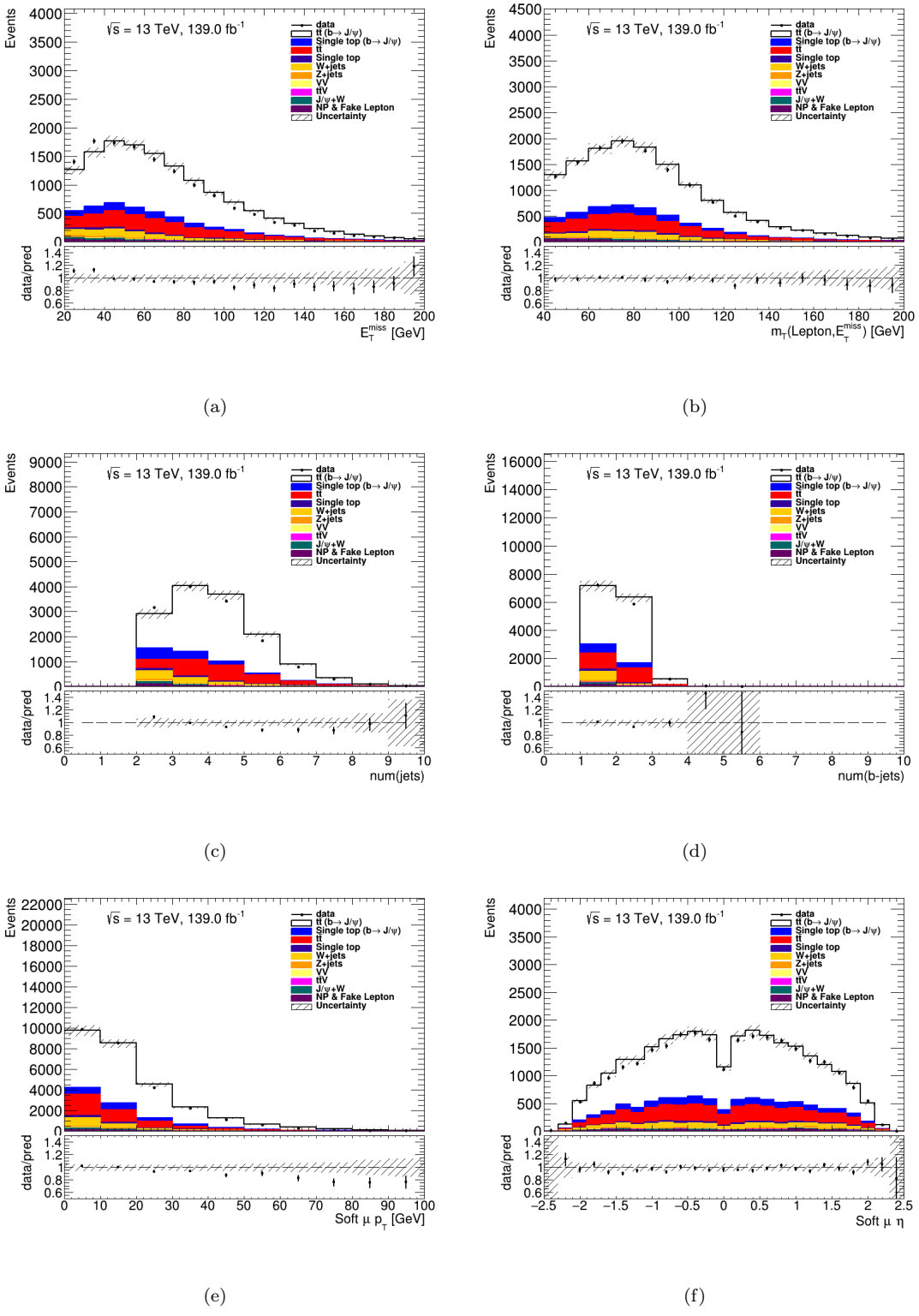
Physics process	Yields ± stat. unc. ± simulation unc.
$t\bar{t}$ with true top decay mode	9238 ± 96 ± 508
single top with true top decay mode	1052 ± 32 ± 82
$t\bar{t}$	2308 ± 48 ± 127
single top	245 ± 15 ± 19
W+jets	763 ± 28 ± 232
Z+jets	146 ± 12 ± 26
Diboson	38 ± 6 ± 19
$t\bar{t}V$	31 ± 6 ± 4
$J/\psi + W$	149 ± 12 ± 149
NPFL	184 ± 14
Model prediction	14152 ± 119 ± 599 (total = 611)
data	13712

TABLE 4.8: The yields found in the signal region for each different contribution with the statistical uncertainty, as well as the simulation uncertainty for each SM process.



(a) (b)

FIGURE 4.6: The upper panel shows the distributions of the transverse momentum (a) and pseudorapidity (b) of the lepton candidates in the signal region. These distributions contain the data, the contributions from the different signal and background SM processes and the contribution from NPFL. The statistical and luminosity uncertainties, together with the uncertainties associated with the SM processes, are shown. The lower panel shows the ratio between the data and the predictions of the various processes and NPFL.

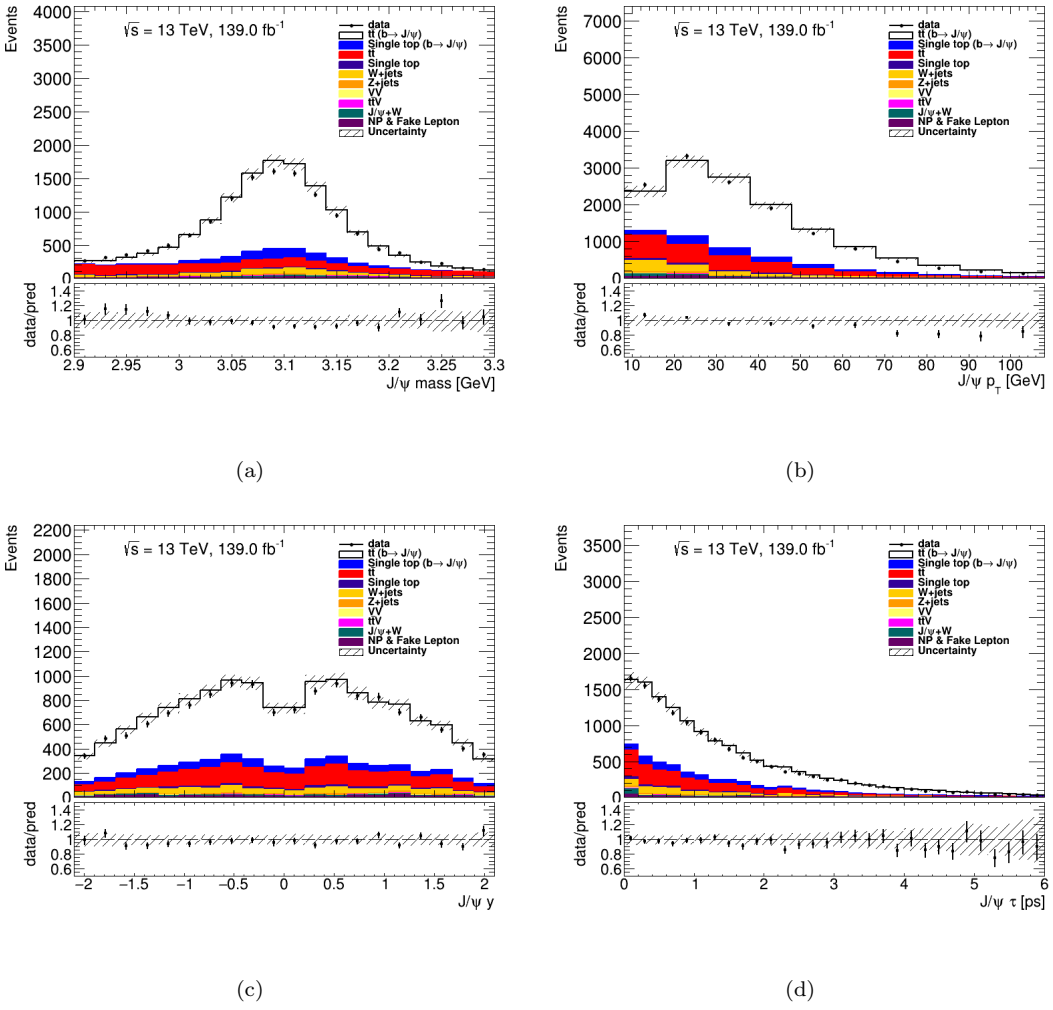


(a) (b)

(c) (d)

(e) (f)

FIGURE 4.7: The upper panel shows the distributions of the missing transverse energy (a), transverse mass of the W-boson (b), the number of jets (c) and b-tagged jets (d), as well as the transverse momentum (e) and pseudorapidity (f) of the muons originated from the J/ψ in the signal region. These distributions contain the data, the contributions from the different signal and background SM processes and the contribution from NPFL. The statistical and luminosity uncertainties, together with the uncertainties associated with the SM processes, are shown. The lower panel shows the ratio between the data and the predictions of the various processes and NPFL.



(a) (b)

(c) (d)

FIGURE 4.8: The upper panel shows the distributions of the invariant mass (a) and transverse momentum (b), rapidity (c) and pseudo-proper time (d) of the dimuon candidates (i.e. J/ψ). These distributions contain the data, the contributions from the different signal and background SM processes and the contribution from NPFL. The statistical and luminosity uncertainties, together with the uncertainties associated with the SM processes, are shown. The lower panel shows the ratio between the data and the predictions of the various processes and NPFL.

Chapter 5

Measurement of the top quark mass

Direct top quark mass (m_{top}) measurements are obtained by utilizing information from the kinematic reconstruction of the measured top quark decay products, and their corresponding combinations. This information is derived from Monte Carlo (MC) simulated events using several assumed values for the top quark mass in the simulation. These results are related to measurements of the various input parameters of MC simulations, and the differences between MC simulations are covered by systematic uncertainties.

The template method is used, which is based on a partial reconstruction of the kinematics underlying the top-quark(s) decay, with probability density functions (templates) for observables that are sensitive to the underlying m_{top}. These templates are fitted to functions interpolating between the different input values of the m_{top}. An unbinned likelihood fit to the observed data distribution of the observable is used to determine the m_{top} value that best describes the data. The invariant mass of the lepton (originating from the W-boson decay) + $J/\psi \rightarrow \mu^+\mu^-$ (originating from the B-hadron decay) distribution is the observable sensitive to the underlying m_{top} and therefore can be used to determine the m_{top} in data [144].

The following sections describe the method and measurement of the top quark mass, with its statistical and systematic uncertainties.

5.1 Extracting the top quark mass

Figure 5.1 shows the invariant mass of the lepton + J/ψ distribution in the signal region using nine different generated top quark mass samples, which includes both top quark pair and single top events. Each distribution can be distinguished from the next, and thus a template method will be developed using these samples.

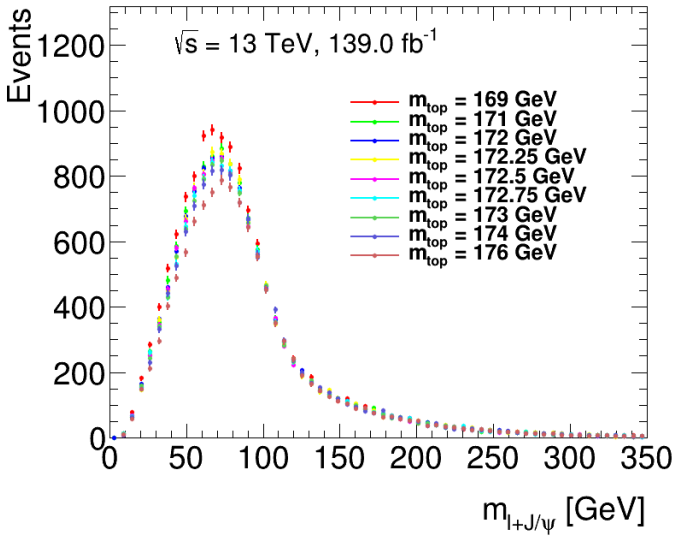


FIGURE 5.1: Invariant mass distribution of the lepton + J/ψ in the signal region over simulated top quark samples using nine different generated top quark masses.

5.1.1 Signal templates

The signal region contains a lepton from a W-boson decay and J/ψ meson from the hadronization of a b-quark. Experimentally, these particles could have originated from either the same (also known as correct pairing) or different (also known as incorrect pairing) top quarks, and therefore the invariant mass distribution of the lepton + J/ψ distribution in the signal region contains contributions from correct and incorrect pairings. Figure 5.2 shows these contributions at the particle-level over the $t\bar{t}$ and single top MC samples with top quark mass set to 172.5 GeV.

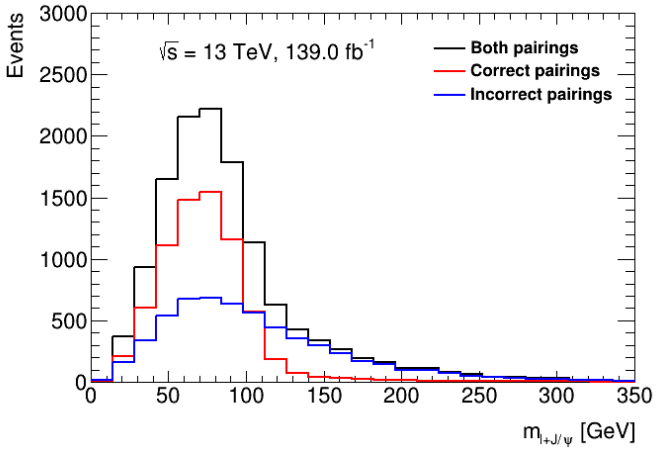
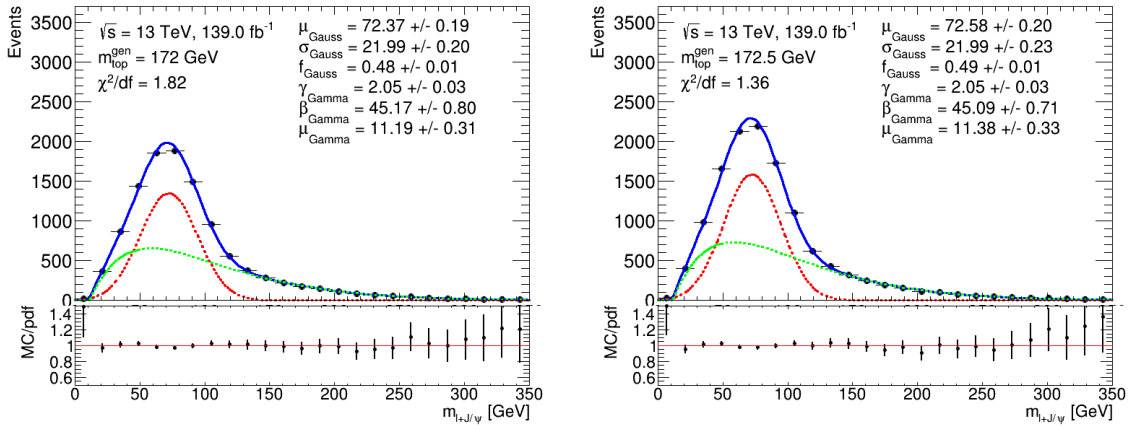


FIGURE 5.2: The invariant mass distribution of the lepton + J/ψ in the signal region from the m_{top} = 172.5 GeV $t\bar{t}$ and single top MC samples, broken up into contributions from correct (red) and incorrect (blue) pairings, as well as the total (black).

Figure 5.2 shows that the correct pairings produce a Gaussian shape distribution, while the incorrect pairings produce a Gamma shape distribution. Therefore, the probability density function (or pdf) describing the invariant mass of the lepton + J/ψ distribution in the signal samples is defined as the sum of a Gaussian and a Gamma distribution. This pdf consists of six parameters: i.e. the Gaussian mean (μ_{Gauss}) and width (σ_{Gauss}), the normalization fraction for the Gaussian (f_{Gauss}), and the Gamma function's gamma (γ_{Gamma}), beta (β_{Gamma}) and mu (μ_{Gamma}) parameters. The μ_{Gauss} parameter describes the position of the Gaussian peak, whereas the σ_{Gauss} parameter describes the Gaussian's width. The γ_{Gamma} and β_{Gamma} parameters together describe the shape of the Gamma function, while the μ_{Gamma} parameter describes the Gamma function's starting point.

The pdf is fitted over the invariant mass of the lepton + J/ψ distribution for each of the nine different top quark mass samples. Figure 5.3 shows the invariant mass of the lepton + J/ψ distribution with the pdf fit and the individual contributions from the Gaussian (red) and Gamma (green) functions, when the top quark mass is set to 172 and 172.5 GeV, as well as the best fit values for the different parameters of the pdf. These fit results are called signal templates. The fits over the other generated top quark masses can be found in Appendix B. The ratio between the pdf and the simulation is consistent with 1 across the invariant mass of the lepton + J/ψ distributions for the different generated top quark masses, except in the first bin. However, the pdf describes the distribution reasonably well, since the $\chi^2/\mathrm{df} \sim 1$, and therefore should not affect the results.



(a) (b)

FIGURE 5.3: The upper panel shows the invariant mass of the lepton + J/ψ distribution with the total pdf (blue) fitted over the m_{top} = 172 GeV (a) and 172.5 GeV (b) GeV signal MC samples, respectively. The individual contributions of the Gaussian (red) and Gamma (green) functions and the best fit values are shown. The lower panel shows the ratio between the signal MC samples and pdf.

5.1.2 The romance between the parameters from the pdf and top quark mass

Figure 5.3 shows the best fit parameter values for two different top quark mass samples, while the others can be found in Appendix B. A priori is that the top quark mass is expected to be more sensitive to the Gaussian parameters as they describe the correct pairings with the top quark decay products, whereas the Gamma parameters are expected to be insensitive as they describe the incorrect pairings. Some parameters can be seen to be consistent within 1σ uncertainty, while others are not. This indicates some parameters are either dependent or independent on the top quark mass. The relationship between the pdf's parameters and the top quark mass can be determined by fitting a straight line. However, each parameter is correlated (or anti-correlated) with the others to some degree, and therefore the correlations have to be considered when determining the relationship between the parameters and the top quark mass. Table 5.1 shows the strength of the correlations between each parameter (i.e. the correlation matrix). The correlation matrix for the other top quark masses can be in Appendix C.

The Gaussian's peak position has medium correlations with the other Gaussian parameters, and weak correlations with the Gamma function's parameters. The Gaussian's width and normalization fraction have medium to strong correlations with the shape and starting point of the Gamma function. The Gamma function's parameters have strong correlations with each other.

Parameter	μ_{Gauss}	σ_{Gauss}	f_{Gauss}	γ_{Gamma}	β_{Gauss}	μ_{Gamma}
μ_{Gauss}	1	-0.36	-0.35	-0.17	-0.04	-0.28
σ_{Gauss}	-0.36	1	0.82	-0.37	0.55	0.69
f_{gauss}	-0.35	0.82	1	-0.50	0.73	0.63
γ_{Gamma}	-0.17	-0.37	-0.50	1	-0.91	-0.66
β_{Gamma}	-0.04	0.55	0.73	-0.91	1	0.65
μ_{Gamma}	-0.28	0.69	0.63	-0.66	0.65	1

TABLE 5.1: Correlation matrix of each parameter in the pdf describing the invariant mass of the three lepton system fitted over the $m_{top} = 172.5$ GeV MC sample.

A χ^2 function can be developed to take into account the correlations between parameters, according to G. Cowan - *Statistical data analysis* textbook [145]. This is briefly described below.

5.1.2.1 Building a χ^2 taking the parameter correlations into account

"According to the central limit theorem, the measured value, y, can be considered to follow a normal distribution with mean value λ when taking a large number of random samples of a given dataset. Consider a set of N independent (or uncorrelated) Gaussian random variables y_i,

each related to another variable x_i, which is assumed to be known without error. Assume that each value y_i has a different unknown mean, λ_i, and a different but known variance, σ_i^2. The N measurements of y_i can be equivalently regarded as a single measurement of an N-dimensional random vector, for which the joint pdf is the product of N Gaussians,

$$g(y_1, ..., y_N; \lambda_1, ..., \lambda_N, \sigma_1^2, ...\sigma_N^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} exp\left(\frac{-(y_i - \lambda_i)^2}{2\sigma_i^2}\right) \tag{5.1}$$

Suppose further that the true value is given as a function of x, $\lambda = \lambda(x;\theta)$, which depends on unknown parameters $\theta = (\theta_1,...,\theta_m)$. The aim of the method of least squares is to estimate the parameters θ. In addition, the method allows for a simple evaluation of the goodness-of-fit of the hypothesized function $\lambda(x;\theta)$. Taking the logarithm of the joint pdf and dropping additive terms that do not depend on the parameters gives the log-likelihood function,

$$log\mathcal{L}(\theta) = -\frac{1}{2}\sum_{i=1}^{N} \frac{(y_i - \lambda(x_i;\theta))^2}{\sigma_i^2} \tag{5.2}$$

This is maximized by finding the values of the parameters θ that minimize the quantity, namely the quadratic sum of the differences between measure and hypothesized values, weighted by the inverse of the variances. This is the basis of the method of least squares (LS), and is used to define the procedure even in the cases where the individual measurements y_i are not Gaussian, but as long as they are independent.

$$\chi^2(\theta) = \sum_{i=1}^{N} \frac{(y_i - \lambda(x_i;\theta))^2}{\sigma_i^2} \tag{5.3}$$

If the measurements are not independent (or correlated) but described by an N-dimensional Gaussian pdf with known covariance matrix V but unknown mean values, the corresponding log-likelihood function is obtained from the logarithm of the joint pdf (5.4) given by the following equation

$$f(\mathbf{x};\mu,V) = \frac{1}{(2\pi)^{N/2}|V|^{1/2}} exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^T V^{-1}(\mathbf{x}-\mu)\right] \tag{5.4}$$

$$log\mathcal{L}(\theta) = -\frac{1}{2}\sum_{i,j=1}^{N} (y_i - \lambda(x_i;\theta))(V^{-1})_{ij}(y_i - \lambda(x_j;\theta)) \tag{5.5}$$

where the additive terms not depending on the parameters have been dropped. This is maximized by minimizing the following

$$\chi^2(\theta) = \sum_{i,j=1}^{N} ((y_i - \lambda(x_i;\theta))(V^{-1})_{ij}(y_i - \lambda(x_j;\theta)).\text{"} \tag{5.6}$$

To determine the relationship between the pdf's parameters and the top quark mass, equation 5.6 was minimized, where y_i are the measured parameter values, $\lambda(x_i, \theta)$ are the relationships functions (i.e. straight line equations of the form a(m_{top}-172.5)+b), and V_{ij}^{-1} is the inverse of the covariance matrix.

5.1.2.2 The relationship between the parameters and the top quark mass

The χ^2 function above was minimized, and the relationship between the parameters and the top quark mass can be seen in Figures 5.4-5.5.

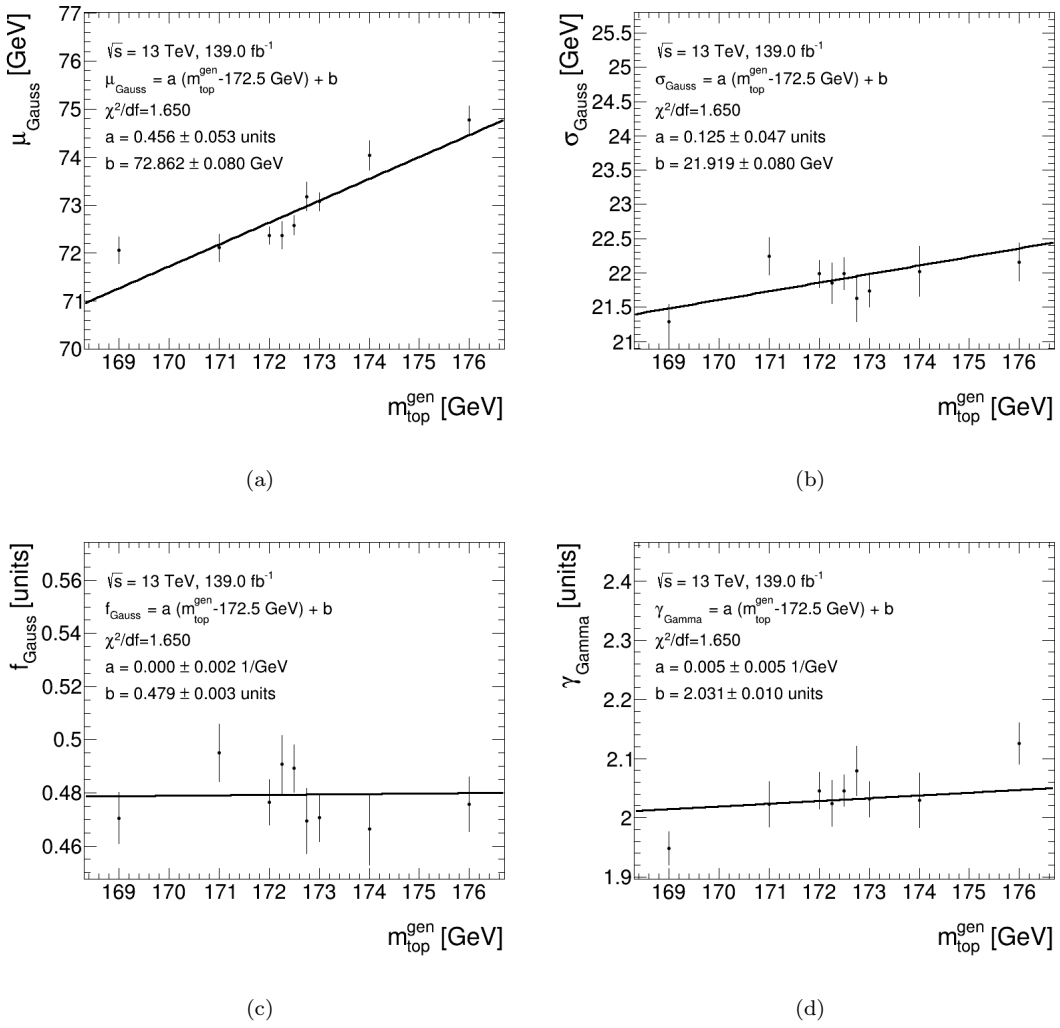
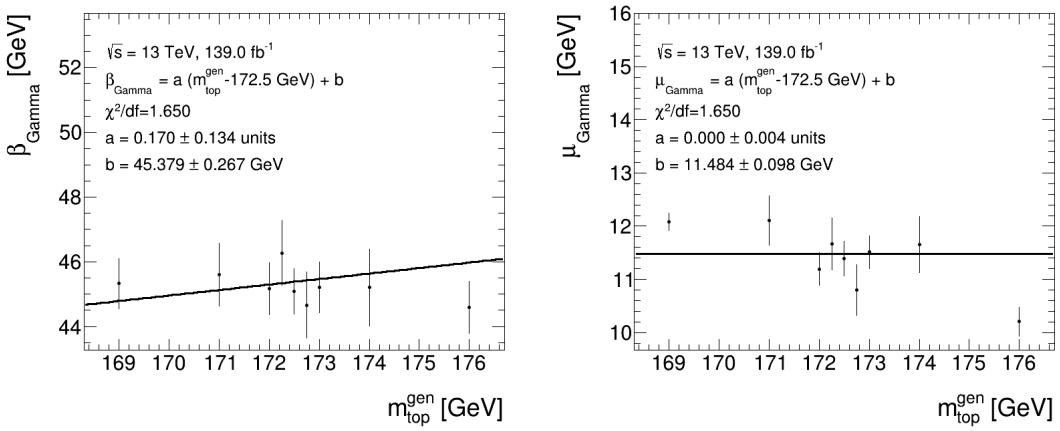


FIGURE 5.4: The relationship between the Gaussian's mean (a) and width (b), fraction of Gaussian (c) and Gamma's gamma (d), and the top quark mass after minimizing the χ^2 function that takes into account the correlations between the parameters.



(a) (b)

FIGURE 5.5: The relationship between the Gamma's beta (a) and mu (b), and the top quark mass after minimizing the χ^2 function that takes into account the correlations between the parameters.

The Gamma's γ and μ parameters and the Gaussian fraction have no dependence on the top quark mass, since their slopes are consistent with 0 within 1σ uncertainty. The Gamma's β parameter is considered to also have no dependence on the top quark mass, since its slope is consistent 0 within 1.3σ uncertainty. The Gaussian mean parameter shows the largest dependency on the top quark, while the Gaussian width parameter shows some dependence.

These relationships can now be substituted into the pdf describing the signal sample's invariant mass of the lepton + J/ψ distribution, producing a pdf that is solely dependent on the top quark mass of the form

$$pdf_{sig}(m_{top}) = f * Gaussian(m_{top}) + (1 - f) * Gamma(m_{top}). \tag{5.7}$$

5.1.3 Background template

The pdf above, however, does not describe the contributions from the background SM processes and the NPFL contribution. These contributions are top quark independent, and therefore only one pdf is needed. However, the $t\bar{t}V$ is top quark dependent, but the contribution accounts for 0.2% of the signal region, and therefore is treated as part of the background contributions. A Novosibirsk function is used to describe these top quark independent samples. Figure 5.6 shows the invariant mass of the lepton + J/ψ distribution for the top quark independent contributions fitted with the Novosibirsk pdf (blue) and the best fit values for the different parameters of the pdf. This fit result is called the background template and is considered a constant function, since it is m_{top} independent. The ratio between the pdf and the simulation is consistent with

1 across the invariant mass of the lepton + J/ψ distribution, except in the first bin. However, the pdf describes the distribution reasonably well, since the $\chi^2/\mathrm{df} \sim 1$, and therefore should not affect the results.

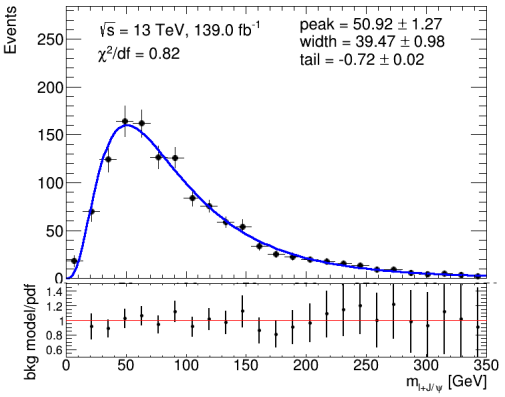


FIGURE 5.6: The upper panel shows the invariant mass of the lepton + J/ψ distribution from the background SM processes (W/Z+jets, VV, $t\bar{t}V$, and J/ψ+W) and the NPFL contributions, fitted with a Novosibirsk function (blue). The best fit values for the Novosibirsk parameters together with its uncertainty is shown in the top right. The lower panel shows the ratio between the top quark independent samples and the pdf.

5.1.4 Total pdf

Combining the top quark dependent and independent pdfs can now describe the top quark mass responsible for producing the invariant mass of the lepton + J/ψ distribution in data. The combined pdf has two parameters, 1) the top quark mass, and 2) the fraction of the top quark dependent contribution (g), and is of the form

$$pdf(m_{top}, g) = g * pdf_{sig}(m_{top}) + (1 - g) * Novosibirsk(constant). \tag{5.8}$$

5.2 Performance of the top quark mass extraction method

The total pdf developed above can now be used to determine the top quark mass in data. However, the performance of the method has to be estimated in order to understand the results. The full model is considered to be the contributions from signal top quark and background SM processes, and the NPFL contribution.

It is important to understand the relationship between the reconstructed and generated top quark masses. The previous parameter dependence with top quark mass, depicted in Figures 5.4-5.5, incorporates information from all nine top quark masses. In order to understand the

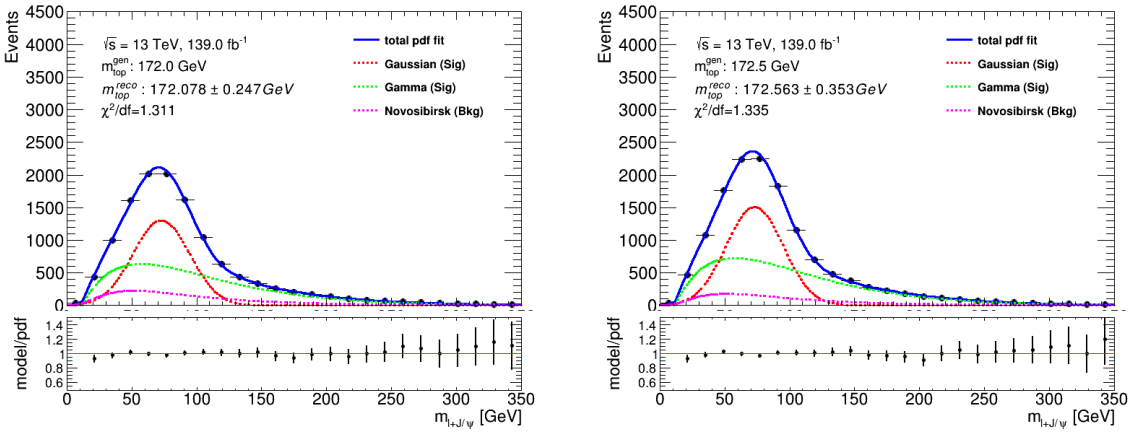
relationship between the reconstructed and generated top quark masses in an unbiased manner, one top quark mass point has to be removed (also called the testing sample), and the parameter relationships will be redetermined without the testing sample's information. Table 5.2 shows the parameter relationships with the top quark mass either using all nine top quark masses information or after removing the m_{top} = 172.5 GeV information from the parameter fits. All slopes and intercepts are consistent within uncertainty, and thus the same parameter dependency with the top quark mass is observed. Appendix D contains the parameter relationships with the top quark mass after removing each top quark mass's information from the parameter fits.

Parameter	All nine m_{top} information	Removing m_{top} = 172.5 GeV information
μ_{Gauss} slope	0.46 ± 0.05	0.51 ± 0.05
μ_{Gauss} intercept	72.86 ± 0.08	72.94 ± 0.09
σ_{Gauss} slope	0.13 ± 0.05	0.08 ± 0.04
σ_{Gauss} intercept	21.92 ± 0.08	21.86 ± 0.09
f_{Gauss} slope	0.00 ± 0.00	0.00 ± 0.00
f_{Gauss} intercept	0.48 ± 0.00	0.48 ± 0.00
γ_{Gamma} slope	0.01 ± 0.01	0.00 ± 0.00
γ_{Gamma} intercept	2.03 ± 0.01	2.00 ± 0.02
β_{Gamma} slope	0.17 ± 0.13	0.00 ± 0.00
β_{Gamma} intercept	45.38 ± 0.27	45.66 ± 0.39
μ_{Gamma} slope	0.00 ± 0.00	0.00 ± 0.00
μ_{Gamma} intercept	11.48 ± 0.10	11.73 ± 0.13

TABLE 5.2: The slopes and intercepts of the relationships between the parameters and the top quark mass, either using all nine m_{top} information or after removing m_{top} = 172.5 GeV information from the fits. The uncertainty values quoted as 0.00 are smaller than 0.005. The relationships are of the form parameter = slope(m_{top}^{gen}-172.5)+intercept.

The resulting pdf's are used to fit over the full model using the tested sample as the signal top quark contribution to determine the reconstructed top quark mass. Figure 5.7 shows the total pdf fit (blue) over the full model for two different top quark mass samples after removing the testing sample's information from the parameter fits. The Gaussian (red), Gamma (green) and Novosibirsk (purple) are shown as well. The reconstructed and generated top quark masses can be found in the top left corner of the upper panel, while the ratio between the total pdf and the model is shown in the lower panel. The other six fits can be found in Appendix E.

Figure 5.8 shows the relationship between the generated and reconstructed top quark masses fitted with a straight line. The best fit values for the straight line show a slope consistent with unity and an intercept consistent with 0. Therefore, the reconstructed top quark mass obtained from the fit can be expected to be the same as the true top quark mass that produced the associated invariant mass of the lepton + J/ψ distribution. In other words, the method is well calibrated and can be used to measure the top quark mass in the data.



(a) (b)

FIGURE 5.7: The upper panel shows the total pdf fit (blue) over the invariant mass of the lepton + J/ψ distribution of the full model when using $m_{top} = 172$ (a) and 172.5 (b) GeV as the signal top quark processes, after removing the testing sample's information from the parameter fits. The individual contributions of the Gaussian (red), Gamma (green) and Novosibirsk (purple) functions are shown. The generated and reconstructed top quark masses are shown in the top left corner, together with the fit quality. The lower panel shows the ratio between the full model and the pdf.

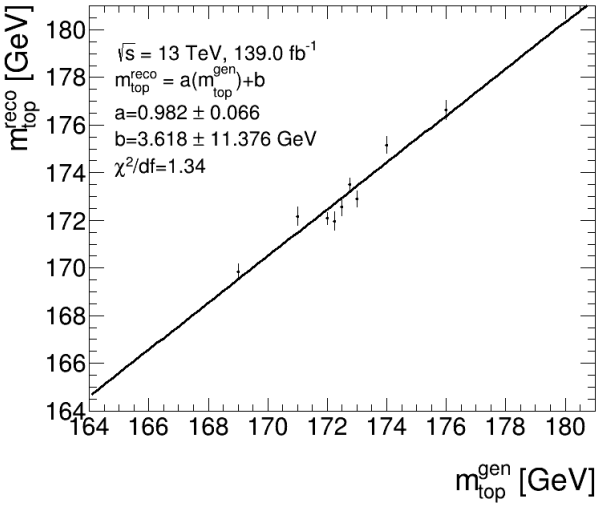


FIGURE 5.8: The relationship between the generated and reconstructed top quark mass obtained from the extraction method, fitted with a straight line (solid black line). The best fit parameter values and the fit quality for the straight line can be found in the top left corner. The uncertainty in the data points include the uncertainty associated with the fit and the method uncertainty.

An additional performance test was done to test the bias of the fit and verify the uncertainty associated with the fit. This can be estimated by creating a pull distribution, where the bias is defined as the pull mean and uncertainty the pull width. A model which is unbiased and

produces an accurate uncertainty should have a pull mean and width consistent with 0 and 1, respectively. The pull is calculated by using

$$\text{Pull} = \frac{\text{measured value} - \text{true value}}{\text{uncertainty in measured value}}. \tag{5.9}$$

Since there are nine different top quark mass MC samples (i.e. 9 true values), nine pull distributions can be made by generating pseudo-data using the total pdf. The same procedure is applied by removing the testing sample's information. Figure 5.9 shows one set of generated pseudo-data using $m_{top} = 172.5$ GeV with the same number of events found in data (i.e. 13712), in order to be statistically comparable with data, together with the pdf fit and the measured top quark mass value and its uncertainty.

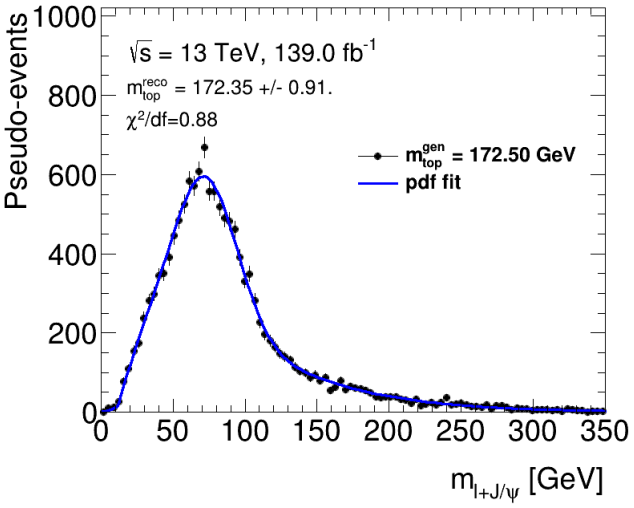
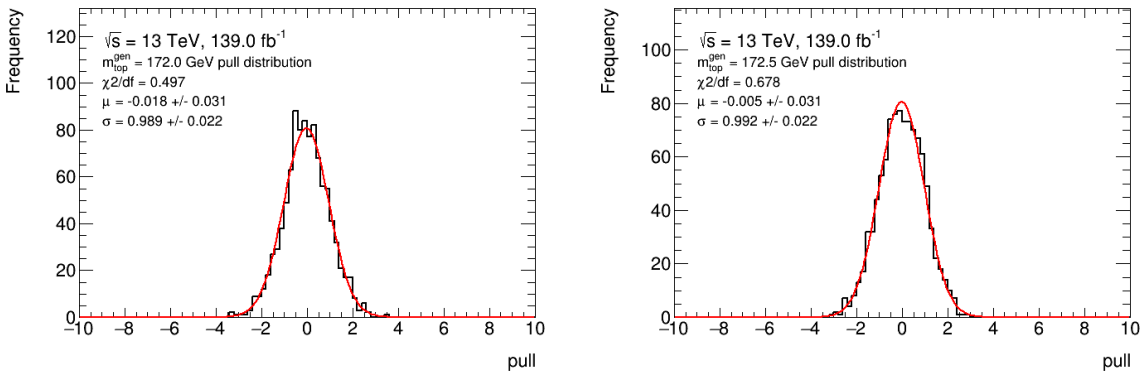


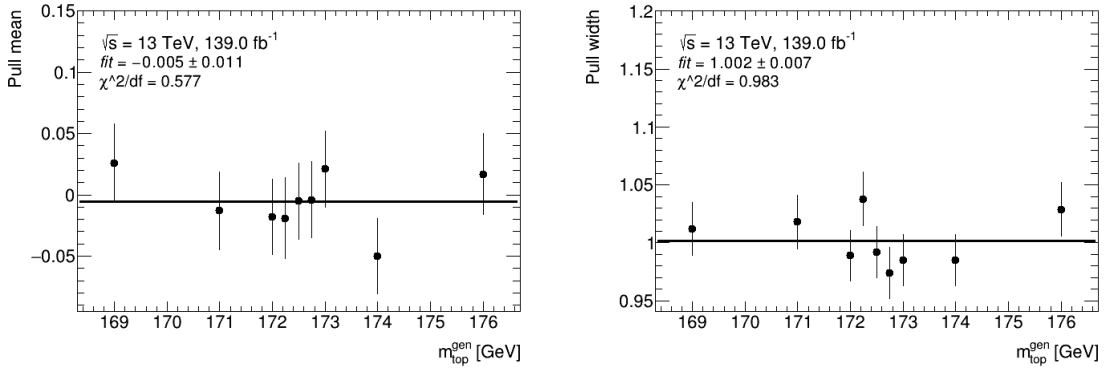
FIGURE 5.9: The invariant mass of the lepton + J/ψ distribution from one set of pseudo-data generated from the total pdf with the total pdf fitted over the pseudo-data. The result of the top quark mass and its uncertainty is shown on the top right, together with the mass used to generate the pseudo-data.

In order to produce a pull distribution, 1000 sets of pseudo-data were generated for each top quark mass. A Gaussian was fitted over the pull distribution to determine the pull mean and width. Figure 5.10 shows the pull distribution when $m_{top} = 169$, 172.5 and 176 GeV with the Gaussian fit to obtain the pull mean and width. The pull distribution for the other top quark mass values can be found in Appendix F. The pull mean and width is a function of m_{top} and the relationship between them can be obtained by fitting a constant line over both distributions. Figure 5.11 shows the relationship between the top quark mass and the pull mean, as well as the pull width. The pull mean, i.e. -0.003 $\pm$ 0.005 is consistent with 0 within 1σ, which shows the method is unbiased in determining the top quark mass value. The pull width, i.e. 1.003 $\pm$ 0.003 is consistent with 1 within 1σ, which shows the method accurately estimates the uncertainty of the top quark mass.



(a) (b)

FIGURE 5.10: Pull distributions from 5000 pseudo-data sets with the Gaussian fit (red) when $m_{top} = 172$ (a) and 172.5 (b) GeV. The pull mean and width is shown in the top left corner.



(a) (b)

FIGURE 5.11: Pull mean (a) and width (b) as a function of the top quark mass. The solid black line shows the constant line fit over the distributions, with the fit results shown in the top left corner.

5.3 Systematic uncertainties

The method to extract the top quark mass has been developed and tested, and thus the top quark mass in data can be measured. However, because the extraction method is based on simulation, there are several sources of systematic uncertainties based on the modelling of the signal and background processes, and even the extraction method itself, which impacts the measurement. Additionally, there are reconstruction-related (or detector) uncertainties that impact the measurement. Table 5.3 displays a list of all the systematic sources impacting the measurement, with the number of independent components per source.

Source of uncertainty	Number of components
Method	1
Modelling of signal processes	
NLO generator	1
Parton shower and hadronization	1
b-quark fragmentation	1
B-hadron production fractions	4
B-hadron decay branching ratios	3
h_{damp} tuning	1
Initial-state radiation	3
Final-state radiation	1
Colour reconnection	1
Choice of PDF	30
Modelling of background processes	
W+jets	2
Z+jets	2
VV	1
$t\bar{t}V$	1
$J/\psi + W$	1
NPFL	2
Detector Response	
Electrons	6
Muons	16
Jet energy scale	29
b-jet energy scale	1
Jet energy resolution	8
Jet vertex fraction	1
Flavour tagging	16
Missing transverse momentum	3
Pile-up	1

TABLE 5.3: A list of all the systematic uncertainty sources included in the analysis, together with the number of independent components taken into account.

The full model shown above consists of the background SM processes simulated with the full ATLAS simulation, the NPFL contribution and the signal top quark SM processes simulated with the fast ATLAS simulation. This model will be used to estimate the systematic uncertainties associated with this measurement because the samples used to estimate the signal modelling

systematics were generated with the fast ATLAS simulation, which allows for consistent estimations. This will be called the nominal model going forward. All the samples used to estimate the systematics were generated with a top quark mass set to 172.5 GeV, unless otherwise stated. Figure 5.12 shows the fit over the invariant mass of the lepton + J/ψ distribution from the nominal model in the upper panel and the ratio between the model and pdf in the lower panel. The reconstructed top quark mass of 172.563 ± 0.353 GeV is consistent with the generated top quark mass of 172.5 GeV. The uncertainty quoted in this distribution takes into account the full statistical power of the model, which is not consistent with data. The full statistical power of the model results from a simulated data set consisting of approximately ten times the number of events as in the measured data set. The statistical uncertainty in the model that is consistent with the statistical power of the data, and therefore, the expected uncertainty in the data, is 0.90 GeV.

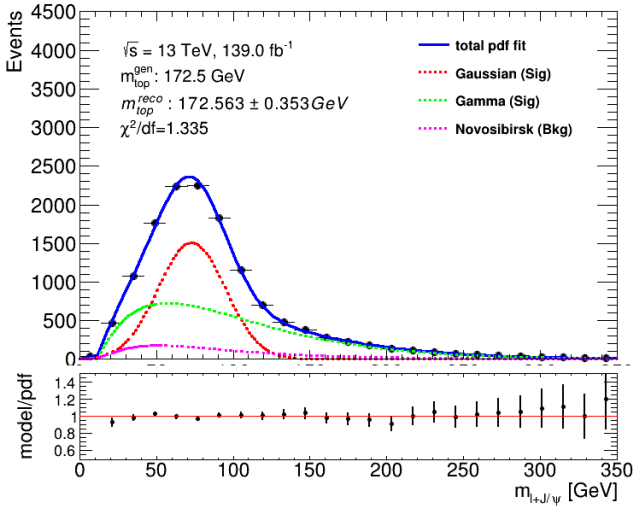


FIGURE 5.12: Distribution of the invariant mass of the lepton + J/ψ system using the ATLAS-Fast $t\bar{t}$ and single top SM processes with $m_{top} = 172.5$ GeV, together with full ATLAS simulated W/Z+jets, VV, $t\bar{t}V$, and $J/\psi + W$ SM processes with $m_{top} = 172.5$ GeV, as well as the NPFL in the upper panel. The total pdf fit (blue) is shown, and the reconstructed top quark mass and fit quality can be seen in the top left corner. The lower panel represents the ratio between simulation and pdf.

5.3.1 Systematic uncertainty calculation

The extraction method in Section 5.1 was developed using templates for the signal and background processes. These templates will remain unchanged when varying the different sources of uncertainty to estimate the impact on the top quark mass. Each source of uncertainty is evaluated using simulated events and is constructed to be uncorrelated with one other.

Some sources of uncertainty consist of an up-down variation or a single variation. For systematics with an up-down variation, the uncertainty will be taken as half the difference whereas, for systematics with a single variation, the difference between the nominal top quark mass and the one obtained from the variation is taken as the systematic uncertainty.

To facilitate comparison with other results, each systematic uncertainty is assigned a statistical uncertainty, which takes into account the statistical correlation of the samples under consideration [146]. Using the individual uncertainties $u_i \pm s_i$, the total uncertainty is calculated as $u \pm s$, where $u = \sqrt{\sum_i u_i^2}$ and $s = (1/u) \times \sqrt{\sum_i (u_i^2 \times s_i^2)}$. The statistical precision from a comparison of two samples σ_{12} is determined for each source of uncertainty based on the correlation, ρ, of the underlying samples, using $\sigma_{12}^2 = \sigma_1^2 + \sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2$. The statistical correlation is expressed as $\rho_{12} = w_{12}/\sqrt{w_1 w_2}$ with $w_1 = \sum_{i,1} w_i^2$ and $w_2 = \sum_{i,2} w_i^2$ is the sum of the squared root of weights in the two samples, and $w_{12} = \sum_{i,12} w_i^2$ is the sum of the squared root of weights of the events present in both samples. Most estimations are based on the same sample with only a change in a single parameter, such as lepton energy scale uncertainties. This leads to a high ρ value and a correspondingly low statistical uncertainty in their difference. Others, which do not share the same generated events or exhibit other significant differences, have a lower ρ value, and the corresponding statistical uncertainty is higher, such as in the case of some of the signal modelling uncertainties. The statistical uncertainty in the total systematic uncertainty is calculated from the individual statistical uncertainties by the propagation of uncertainties mentioned above.

5.3.2 Method uncertainty

The pdf_{sig} described before was determined from the relationships between the parameters of the Gaussian and Gamma functions and the top quark mass. These relationships are in the form of a straight line and consist of two parameters, i.e. a slope and an intercept. These slopes and intercepts come with an uncertainty, which impacts the extracted top quark mass, and are correlated with each other. Additionally, the Novosibirsk parameters have uncertainties and are correlated, and thus also has to be considered. The slopes and intercepts used to measure the top quark mass with generated $m_{top} = 172.5$ GeV, which can be found in Table 5.2, produces the correlation matrix found in Table 5.4. The parameters from the Novosibirsk pdf produces the correlation matrix found in Table 5.5.

According to Evaluation of Measurement Data - *Guide to the Expression of Uncertainty in Measurement* [147], the impact of correlated parameters on the top quark mass can be estimated with the following equation

$$u^2(y) = \left(\frac{\partial y}{\partial x_1}\right)^2 u^2(x_1) + \left(\frac{\partial y}{\partial x_2}\right)^2 u^2(x_2) + ... + \left(\frac{\partial y}{\partial x_n}\right)^2 u^2(x_n) +$$
$$2r_{1,2}\left(\frac{\partial y}{\partial x_1}\frac{\partial y}{\partial x_2}\right) u(x_1)u(x_2) + 2r_{1,3}\left(\frac{\partial y}{\partial x_1}\frac{\partial y}{\partial x_3}\right) u(x_1)u(x_3) + ...$$
$$2r_{2,3}\left(\frac{\partial y}{\partial x_2}\frac{\partial y}{\partial x_3}\right) u(x_2)u(x_3) + ...2r_{n-1,n}\left(\frac{\partial y}{\partial x_{n-1}}\frac{\partial y}{\partial x_n}\right) u(x_{n-1})u(x_n)$$

where $\frac{\partial y}{\partial x_i}$ is the change in m_{top} as the variable changes, $u^2(x_i)$ is the uncertainty of the variable, and $r_{i,j}$ is the correlation between variable i and variable j.

The uncertainty due to the slopes and intercepts from pdf_{sig} and the parameters from the Novosibirsk, is determined by varying each variable individually up and down by 1 and 2σ, respectively. The reconstructed top quark mass is measured for each variation, and the relationship with each variable is then determined. Figures 5.13-5.15 show the relationship between the variable and the reconstructed top quark mass fitted with a straight line. The gradient from the straight line fit is considered the change in m_{top} as each variable changes. All variables except μ_{Gauss}, β_{Gamma} and γ_{Gamma} produce gradients consistent with 0 within uncertainty. The gradients of μ_{Gauss}, β_{Gamma} and γ_{Gamma} are not consistent with 0 within uncertainty, but have large uncertainties. The statistical uncertainty is determined by varying each gradient within $\pm 1\sigma$ uncertainty individually, and half the difference between the up and down variations are added in quadrature, and considered the total statistical uncertainty. The systematic uncertainty associated with the extraction method is 0.18 ± 0.06 GeV.

Variable	slope(μ_{Gauss})	int(μ_{Gauss})	slope(σ_{Gauss})	int(σ_{Gauss})	slope(f_{Gauss})	int(f_{Gauss})	slope(β_{Gamma})	int(β_{Gamma})	slope(μ_{Gamma})	int(μ_{Gamma})	slope(γ_{Gamma})	int(γ_{Gamma})
slope(μ_{Gauss})	1	0.01	-0.03	-0.02	0.07	0.04	-0.13	-0.10	-0.04	-0.04	0.24	0.19
int(μ_{Gauss})	0.01	1	-0.03	-0.20	0.01	-0.16	0.02	0.10	0.03	-0.01	-0.02	-0.26
slope(σ_{Gauss})	-0.03	0.03	1	0.09	-0.51	-0.26	-0.26	-0.14	-0.25	-0.12	0.16	0.07
int(σ_{Gauss})	-0.02	-0.20	0.09	1	-0.05	0.57	-0.07	0.28	-0.14	0.28	0.10	-0.11
slope(f_{Gauss})	0.07	0.01	-0.51	-0.05	1	0.59	0.63	0.40	0.23	0.11	-0.41	-0.24
int(f_{Gauss})	0.04	-0.16	-0.26	0.57	0.59	1	0.34	0.63	0.04	0.27	-0.17	-0.35
slope(β_{Gamma})	-0.13	0.02	-0.27	-0.07	0.63	0.34	1	0.65	0.51	0.28	-0.91	-0.59
int(β_{Gamma})	-0.10	0.10	-0.14	0.28	0.40	0.63	0.65	1	0.24	0.48	-0.55	-0.89
slope(μ_{Gamma})	-0.04	0.03	-0.25	-0.14	0.23	0.04	0.51	0.24	1	0.58	-0.67	-0.36
int(μ_{Gamma})	-0.04	-0.01	-0.12	0.28	0.11	0.27	0.28	0.48	0.58	1	-0.37	-0.61
slope(γ_{Gamma})	0.24	-0.02	0.16	0.10	-0.41	-0.17	-0.91	-0.55	-0.67	-0.37	1	0.64
int(γ_{Gamma})	0.19	-0.26	0.07	-0.12	-0.24	-0.35	-0.59	-0.89	-0.36	-0.61	0.64	1

TABLE 5.4: Correlation matrix of each variable in the signal pdf when removing the m_{top} = 172.5 GeV top quark mass information from the parameter relationships with the top quark mass.

Variable	peak	tail	width
peak	1	0.60	0.50
tail	0.60	1	0.73
width	0.50	0.73	1

TABLE 5.5: Correlation matrix of each variable in the background pdf.

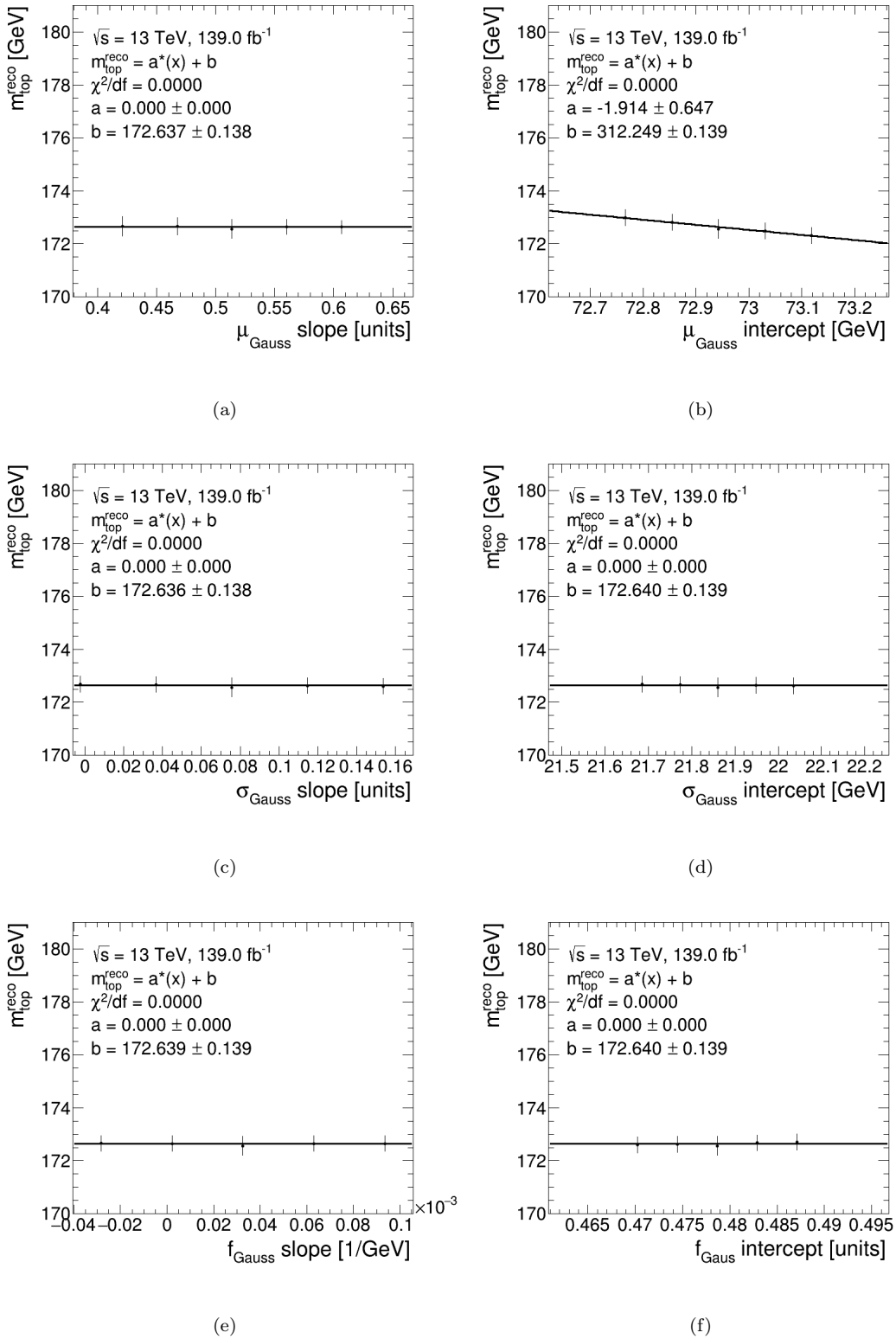


FIGURE 5.13: The relationship between the Gaussian's mean slope (a) and intercept (b), Gaussian's width slope (c) and intercept (d), the fraction of Gaussian's slope (e) and intercept (f), and the top quark mass after removing the information from $m_{top} = 172.5$ GeV.

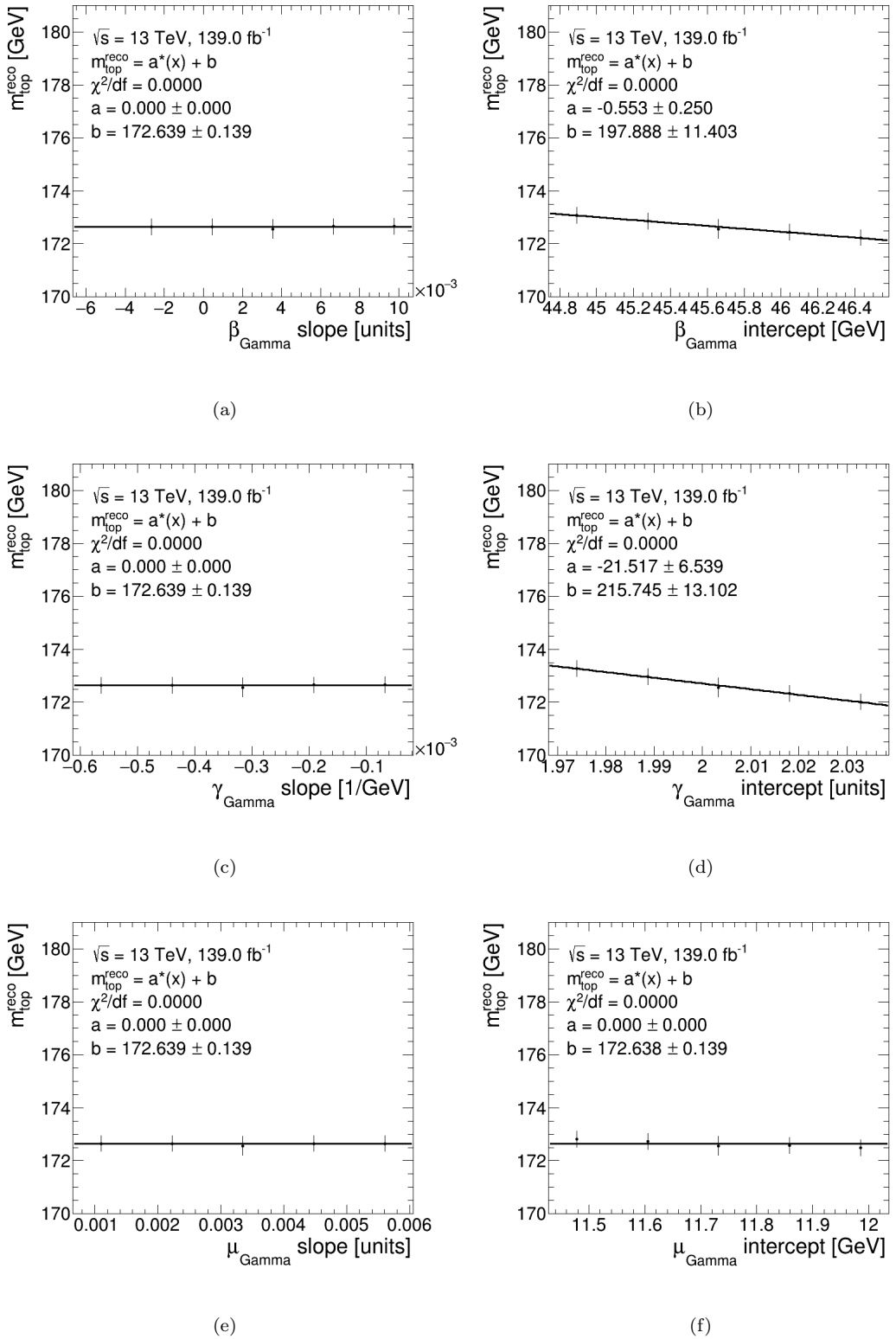


FIGURE 5.14: The relationship between the Gamma's β slope (a) and intercept (b), Gamma's γ slope (c) and intercept (d), the Gamma's μ slope (e) and intercept (f), and the top quark mass after removing the information from $m_{top} = 172.5$ GeV.

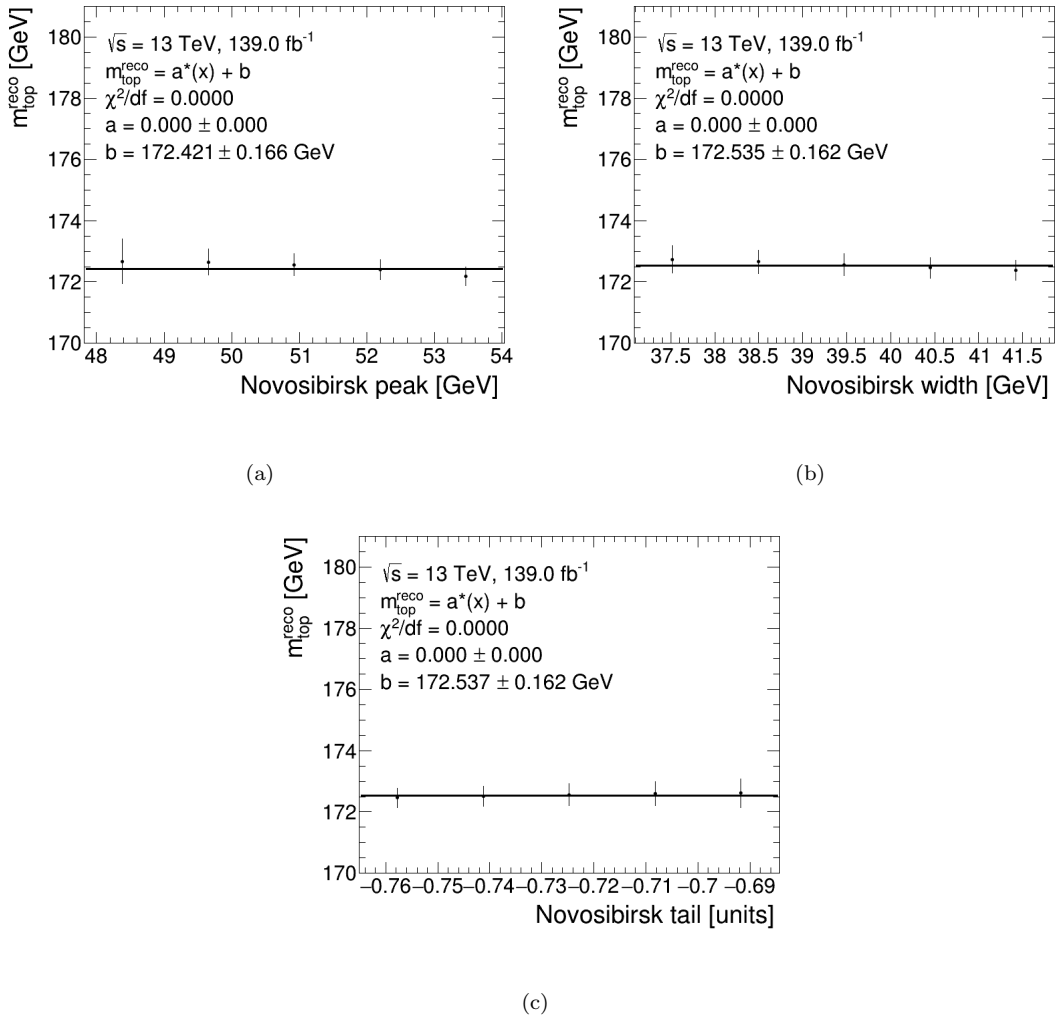


FIGURE 5.15: The relationship between the Novisibirsk's peak (a), width (b), and tail (c) parameters, and the top quark mass after removing the information from $m_{top} = 172.5$ GeV.

5.3.3 Modelling of signal processes

The Powheg-Box NLO matrix element generator was used to create the nominal top quark samples, while Pythia 8 models the hadronization and parton shower using the ATLAS A14 tune. The parton shower generator, which implements matrix-element corrections with accuracy comparable to the NLO level, fully handles radiation in top quark decays. The p_T of the first gluon emission is controlled by the h_{damp} parameter, which is set to 1.5 times the top quark mass of $m_{top} = 172.5$ GeV. The h_{damp} setting's primary impact is to control the high-p_T emission from which the $t\bar{t}$ system recoils.

The modelling of the momentum transfer between the b-quark and the B-hadron, and consequently the description of the b-quark fragmentation function, are particularly important to this

analysis. The fragmentation and hadronization parameters of the A14 tune are inherited from the Monash tune [148], which corresponds to LEP and Tevatron data. When combined with ATLAS 7 TeV measurements of the underlying event, jet generation, Z-boson production, and top quark production, the A14 tune further constrains shower, multiple parton interactions, and colour reconnection parameters. According to reference [149], the A14 tune discovers a weaker coupling than Monash in the initial- and final-state radiation shower, which results in a better description of the Z-boson p_T distribution and the jet multiplicity distribution in the generation of top quark pairs. These parameters in the A14 tune shows deviations with the data in e^+e^- colliders. However, the description of the b-fragmentation in $t\bar{t}$ events was constrained using LEP (DELPHI, ALEPH, and OPAL experiments) and SLD data from $Z \to bb$ events [149]. In Pythia 8, the A14 tune description of the fragmentation contains a parameter that is specific to b-quark fragmentation, called r_b, which is set to 0.855. The nominal $t\bar{t}$ sample mentioned before, and the ones used to produce the signal templates, were generated with r_b = 0.855, but an additional $t\bar{t}$ sample was generated using a r_b value of 1.05, which comes from the best combined fit to LEP+SLD data, as seen in reference [149]. The sample with r_b = 1.05 will take the place of the nominal $t\bar{t}$ sample for systematic uncertainties related to parton shower, hadronization, and b-fragmentation, as those samples produced a better description of the fragmentation derived from the LEP data mentioned above. See Section 5.3.3.3 for more information.

The signal region contains signal contributions from top quark pair events ($\sim 82\%$) and single top quark events ($\sim 9\%$). Top quark pair events dominate and the associated modelling systematics will be estimated. All signal systematic uncertainties going forward are estimated from the modelling of top quark pair events. At the time of writing, the single top quark events modelling systematics have not been estimated. Based on previous ATLAS measurements, see reference [41], the contribution from single top events are expected to be negligible, but are in the process of being investigated.

The pdf fit over each signal modelling systematic component can be found in Appendix G.

5.3.3.1 NLO generator

The uncertainty due to the NLO matrix element matching scheme is estimated by comparing a sample generated with Powheg-Box + Pythia 8 and one with MadGraph5_aMC@NLO + Pythia 8. To maintain the NLO accuracy, the MadGraph5_aMC@NLO matching requires special settings for the Pythia 8 shower, where the global-recoil settings, utilized for final-state radiation emissions, and the matrix-element corrections are turned off for both initial- and final-state radiation. The settings used to shower the nominal Powheg-Box sample in Pythia 8 are different from the ones used in MadGraph5_aMC@NLO, and the description of the final-state

radiation will be inconsistent. Therefore, comparing the nominal Powheg-Box + Pythia 8 with the MadGraph5_aMC@NLO would produce an artificially high uncertainty that is unrelated to the NLO emission. A new Powheg-Box + Pythia 8 sample is created using the same Pythia 8 setup that was used to shower MadGraph5_aMC@NLO events in order to ensure consistent parameters for the radiation given off by the b-quark while enabling comparisons with the top quark's first emission's NLO matching. The systematic uncertainty associated with the NLO generator is 1.08 ± 0.61 GeV. The statistical component is ∼ 60% of systematic uncertainty and thus with additional simulation, this systematic could be improved.

5.3.3.2 Parton shower and hadronization

The impact of the parton shower and hadronization model is evaluated by comparing a sample generated using the default model in Pythia 8 and one using the default model in Herwig 7. The Pythia 8 model uses r_b=1.05 and the A14 set of tuned parameters, while the Herwig 7 model uses the H7-UE set of tuned parameters. The systematic uncertainty associated with parton shower and hadronization model is 0.09 ± 0.68 GeV. This systematic uncertainty is largely dominated by its statistical component. Additional simulation could reduce the sensitivity to possible statistical fluctuations.

5.3.3.3 Modelling of heavy-quark fragmentation

An essential component of this analysis is the modelling of the momentum transfer between the b-quark and the B-hadron. The string and cluster models, which contain parameters that are adapted to data, are the phenomenological models that Monte Carlo event generators like the Pythia and Herwig packages use to represent this transition. While Herwig7 utilizes a non-parametric model to account for the entire parton shower evolution, Pythia 8 permits the use of many parametrizations for the b-quark fragmentation function. This analysis assumes that the b-quark fragmentation properties at a reference q^2 scale are the same in e^+e^- and hadron collisions because the free parameters in both models are commonly fitted to measurements from e^+e^- colliders.

The Lund-Bowler parametrisation [150, 151] used to model the b-quark fragmentation function in the Pythia 8 samples is:

$$f(z) = \frac{1}{z^{1+br_b m_b^2}}(1-z)^a \exp(-bm_T^2/z)$$

where a, b and r_b are the function parameters, m_b is the b-quark mass, m_T is the B-hadron transverse mass and z is the fraction of the longitudinal energy of the B-hadron with respect to the b-quark, in the light cone reference frame.

The b-fragmentation uncertainty is taken as the difference between the top quark masses obtained from the sample generated with r_b=1.05 and r_b=0.855. The systematic uncertainty associated with b-fragmentation is 1.69 ± 0.61 GeV. The two r_b values used to estimate this systematic are far apart and a large uncertainty is expected. However, comparing samples with closer r_b values could improve this systematic. Additionally, the statistical component is around 36% of the systematic uncertainty, and therefore additional simulation could also improve this systematic.

5.3.3.4 Modelling of hadron production

The branching fractions of the decays of B-hadrons to mesons, which are expected to alter the kinematics of the mesons, as well as the effect of the uncertainties on the production fractions for various species of B-hadrons, all contribute to the uncertainty in the production of B-hadrons.

The PDG provides documentation for the computation of the production fractions, which are shown in the second column of Table 5.6 and based on an unbiased sample of weakly decaying b-hadrons. The world average of the time-integrated mixing probability averaged across an unbiased sample of semi-leptonic B-hadron decays was combined with direct rate data from the Delphi [152], CDF [153–156], LHCb [157–159], and ATLAS [160] experiments to estimate these values. The HFLAV [161] group combined the data on the presumption that the production fractions are constant across all the centre-of-mass energies considered. HFLAV also offers correlations between the various B-hadrons, which are shown below. The underlying presumption is that production fractions are constant across all experiments and add up to 1 for top quark decays, as the p_T is relatively high.

$$\begin{aligned} cor(B_S^0, B - baryon) &= -0.260, \\ cor(B_S^0, B^{\pm} = B^0) &= -0.136, \\ cor(B - baryon, B^{\pm} = B^0) &= -0.922. \end{aligned}$$

The maximal variations after accounting for these correlations are displayed in Table 5.6. The quadratic sum of the four variations listed in this table yields the overall B-hadron production uncertainty.

Hadron	PDG	$B^{\pm 0}$ up	$B^{\pm 0}$ down	B_S^0 up	B_S^0 down	B-baryon up	B-baryon down
B^0	0.404 ± 0.006	0.410	0.398	0.403	0.405	0.398	0.410
$B^{\pm}$	0.404 ± 0.006	0.410	0.398	0.403	0.405	0.398	0.410
B_S^0	0.103 ± 0.006	0.102	0.108	0.108	0.098	0.102	0.104
B-baryon	0.088 ± 0.012	0.077	0.099	0.085	0.091	0.100	0.076

TABLE 5.6: Production fractions for weakly decaying B-hadrons at high energy, taken from the PDG [6].

The individual systematic uncertainties can be seen in Table 5.7, and the total systematic uncertainty associated with the modelling of hadron production is 0.08 ± 0.01 GeV.

Source of uncertainty	$\sigma(m_{top})$ [GeV]
B^0	0.047 ± 0.006
$B^{\pm}$	0.047 ± 0.006
B_s^0	0.006 ± 0.002
B-baryon	0.040 ± 0.007
B-hadron production fraction	**0.08 ± 0.01**

TABLE 5.7: The contribution from the different B-hadron production fractions affecting the top quark mass measurement.

5.3.3.5 Modelling of hadron decays

The B-hadrons can directly or indirectly decay into a J/ψ. The direct decays come from charged and neutral B-mesons (B^+, B^-, B_c^+, B_c^i, B_d^0, $\overline{B_d^0}$, B_s^0, $\overline{B_s^0}$) and B-baryons (Λ_b^0, χ_b^0, χ_b^+, χ_b^-, Ω_b^-), with BR($b \to J/\psi(direct) + X = (7.8 \pm 0.4)\cdot 10^{-3}$. Indirect decays come from mesons and baryons in an excited state, which then decay into a J/ψ with an emission of a photon. The dominant indirect decays come from ψ(2S) and χ_{c1}(1P) with branching fractions of $(1.76 \pm 0.19)\cdot 10^{-3}$ and $(4.8 \pm 1.5)\cdot 10^{-3}$. The total branching fraction of both direct and indirect decays is 1.16 ± 0.1%. All of these values come from the PDG [6] and the uncertainty is obtained by the quadratic sum of these three decay modes.

The individual systematic uncertainties can be seen in Table 5.8, and the total systematic uncertainty associated with the modelling of B-hadron decay ratios is 0.08 ± 0.01 GeV.

Source of uncertainty	$\sigma(m_{top})$ [GeV]
direct J/ψ	0.080 ± 0.007
indirect ψ(2S)	0.026 ± 0.024
indirect χ_{c1}(1P)	0.004 ± 0.031
B-hadron decay ratios	**0.08 ± 0.01**

TABLE 5.8: The contribution from the B-hadron decay ratios affecting the top quark mass measurement.

5.3.3.6 Modelling of QCD radiation

A $t\bar{t}$ sample with h_{damp} set to m_{top} and another $t\bar{t}$ sample with h_{damp} set to $3m_{top}$ are compared in order to assess the effects of the choice of the parameter, which regulates the p_T of the first additional gluon emission in $t\bar{t}$ recoils [162, 163]. The systematic uncertainty associated with the h_{damp} parameter is 0.08 ± 0.42 GeV. This systematic uncertainty is largely dominated by its statistical component. Additional simulation is expected to reduce the statistical uncertainty on this estimate, as is discussed for other systematics in Section 5.4.

The uncertainty on the modelling of initial-state radiation (ISR) is evaluated using varied weights on the matrix element. This uncertainty is a combination of three different variations, i.e. vary the α_s^{ISR} parameter higher and lower, and adjust the renormalization (μ_R) and factorisation (μ_F) scales independently by a factor of two up and down. The individual systematic uncertainties can be seen in Table 5.9, and the total systematic uncertainty associated with the modelling of initial-state radiation is 0.13 ± 0.08 GeV. The statistical component is ∼ 60% of the systematic uncertainty, and therefore additional simulation could improve this systematic.

Source of uncertainty	$\sigma(m_{top})$ [GeV]
α_s^{ISR}	0.057 ± 0.017
μ_R	0.090 ± 0.100
μ_F	0.070 ± 0.049
Initial-state radiation	**0.13 ± 0.08**

TABLE 5.9: The contributions from the modelling of initial-state radiation affecting the top quark mass measurement.

Final-state radiation (FSR) considers the impact of QCD emissions and is evaluated by adjusting the renormalization scale up and down by a factor of two (i.e. μ_R=2.0 and μ_R=0.5), while keeping the factorisation scale constant. The systematic uncertainty associated with the final-state radiation is 1.76 ± 0.57 GeV. The statistical component is ∼ 33% of the systematic uncertainty, and therefore additional simulation could improve this systematic.

There are two components that could be causing the large uncertainty due to the FSR. One is the interplay of the sensitivity between FSR and b-fragmentation, which is discussed in Section 5.4.3and two is the effect of evaluating the systematic due to FSR using the Pythia internal reweighting. Bear in mind, the authors' advise caution when extremely large weights are found [164], which is the case in this analysis and discussed further below. However, there is an expectation that this issue can be remedied with a larger sample size.

The way the final-state radiation systematic is estimated could be the reason for the large result. Pythia 8 provides event weights to account for different renormalization scales in the nominal sample. Figure 5.16 shows the event weights for the two different renormalization scales. The event weights for μ_R=2.0 are up to 25000, and produce a top quark mass of 173.459 ± 0.274, which is 0.896 GeV away from the nominal value. The event weights for μ_R=0.5 are up to 390000, and produce a top quark mass of 169.942 ± 0.846, which is -2.621 GeV away from the nominal value. The large weights from Pythia 8 results in a large systematic uncertainty associated with the final-state radiation. Reference [164] shows that Pythia 8's event weight variations of μ_R=2.0 and μ_R=0.5 roughly correspond to samples generated with set renormalization scale values of $\mu_R = \sqrt{2}$ and $\mu_R = 1/\sqrt{2}$, respectively. These new samples should not consist of large event weights, and could produce a more reasonable estimation of the final-state radiation systematic.

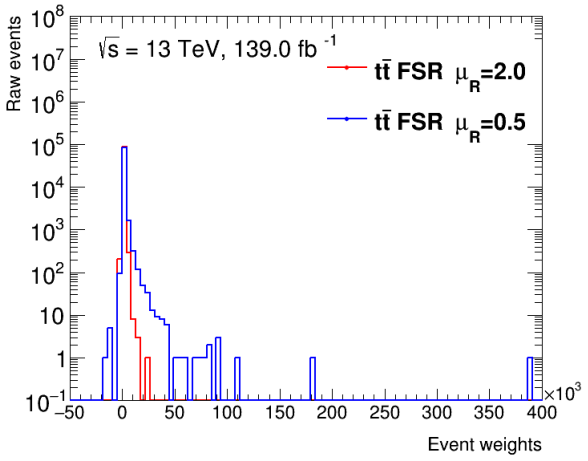


FIGURE 5.16: Distribution of event weights in the nominal top quark pair sample when using a renormalization scale of 2.0 (blue) and 0.5 (red), respectively.

5.3.3.7 Colour Reconnection

The underlying event and colour reconnection can vary the kinematic distribution of the B-hadron. This indirectly affects the invariant mass of the lepton + J/ψ distribution and the top quark mass distribution.

The uncertainty due to the modelling of colour reconnection and multiple parton interactions is estimated by comparing the default beam-remnant model [165] in the nominal Pythia 8 sample with a Pythia 8 sample using a QCD base model [166]. For more information on different colour reconnection models in Pythia 8, see reference [167]. The systematic uncertainty associated with colour reconnection is 0.40 ± 0.44 GeV. This systematic uncertainty is dominated by its

statistical component. Additional simulation could reduce the sensitivity to possible statistical fluctuations.

5.3.3.8 Choice of parton distribution function

The nominal sample uses the NNPDF3.0 PDF set, which is a component of PDF4LHC15 [168]. The uncertainty due to the choice of PDF is evaluated using the PDF4LHC15 error set applied to the NNPDF3.0 set by reweighting each event for 31 different PDFs. The one reweighting corresponds to a nominal variation, and the other 30 is compared to this variation. The sum in quadrature of the individual variations is used to calculate the total systematic uncertainty. The individual systematic uncertainties can be seen in Table 5.10, and the total systematic uncertainty associated with the choice of PDF is 0.05 ± 0.01 GeV.

5.3.4 Modelling of background processes

The background SM physics processes and NPFL contribute to the model prediction that describes the data in the signal region shown in Section 4.5.

The pdf fit over each background modelling systematic component can be found in Appendix H.

5.3.4.1 NPFL

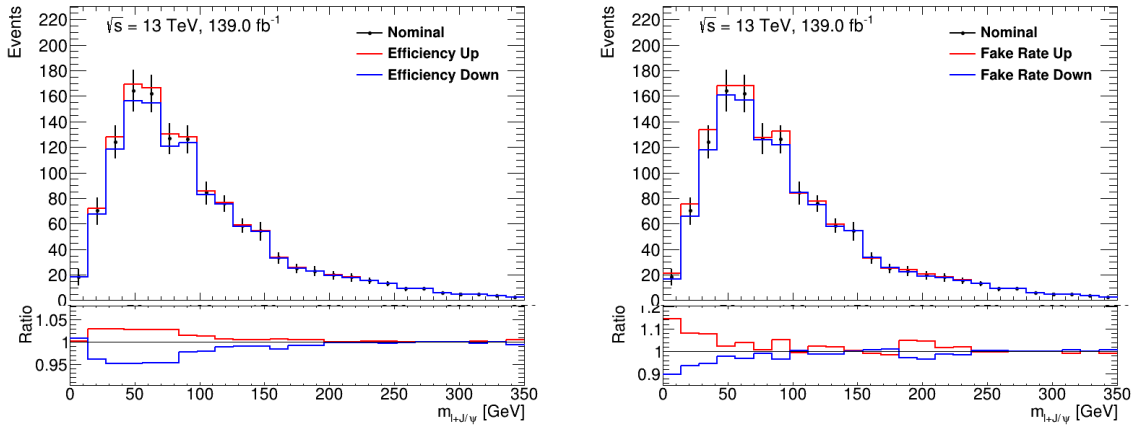
The NPFL contribution is estimated from the efficiencies derived in Section 4.4.6. As mentioned in Section 4.4.7, the systematic uncertainties associated with measuring the efficiencies is not included, as a different control region has not been investigated. This systematic uncertainty is estimated by varying the efficiencies within their statistical uncertainties. The systematic uncertainties associated with measuring the fake rates come from the different MC simulation uncertainties. However, these MC simulations are used in both the fake rate control region and the signal region, respectively, and thus the variations have to be considered simultaneously. This is done in Section 5.3.4.2. Therefore, the systematic uncertainty associated with measuring the fake rates, is estimated by varying the fake rates within their statistical uncertainties.

Figure 5.17 shows the invariant mass of the lepton + J/ψ distribution of the nominal (black) background model, and the background model when the efficiencies (a) and fake rates (b) are varied up (red) and down (blue) within their statistical uncertainties. The lower panel shows the ratio between the variations and the nominal. The largest variation for the efficiencies is $\sim$ 5% and for the fake rates it's under 15%, predominantly at low m(lepton,J/ψ). The individual

Source of uncertainty	$\sigma(m_{top})$ [GeV]
PDF 1	0.006 ± 0.000
PDF 2	0.014 ± 0.000
PDF 3	0.000 ± 0.000
PDF 4	0.019 ± 0.000
PDF 5	0.027 ± 0.000
PDF 6	0.006 ± 0.000
PDF 7	0.002 ± 0.000
PDF 8	0.003 ± 0.000
PDF 9	0.005 ± 0.000
PDF 10	0.004 ± 0.000
PDF 11	0.005 ± 0.000
PDF 12	0.001 ± 0.000
PDF 13	0.002 ± 0.000
PDF 14	0.001 ± 0.000
PDF 15	0.003 ± 0.000
PDF 16	0.004 ± 0.000
PDF 17	0.019 ± 0.000
PDF 18	0.001 ± 0.000
PDF 19	0.007 ± 0.000
PDF 20	0.007 ± 0.000
PDF 21	0.000 ± 0.000
PDF 22	0.008 ± 0.000
PDF 23	0.000 ± 0.000
PDF 24	0.000 ± 0.000
PDF 25	0.001 ± 0.000
PDF 26	0.008 ± 0.000
PDF 27	0.002 ± 0.000
PDF 28	0.017 ± 0.000
PDF 29	0.001 ± 0.000
PDF 30	0.003 ± 0.000
PDF	**0.05 ± 0.01**

TABLE 5.10: The contributions from the various lepton systematic sources affecting the top quark mass measurement. The uncertainty values quoted as 0.000 are smaller than 0.0005.

systematic uncertainties can be seen in Table 5.11, and the total systematic uncertainty associated with the NPFL contribution is 0.16 ± 0.05 GeV. The statistical uncertainty is ∼ 28% of the systematic uncertainty, and therefore additional data and simulation could improve this systematic.



(a) (b)

FIGURE 5.17: The upper panel shows the invariant mass of the lepton + J/ψ distribution of the nominal (black) background model, and the background model when the efficiencies (a) and fake rates (b) are varied up (red) and down (blue) within their statistical uncertainties. The lower panel shows the ratio between the variations and the nominal.

Source of uncertainty	$\sigma(m_{top})$ [GeV]
Efficiency	0.104 ± 0.009
Fake rate	0.118 ± 0.060
NPFL	**0.16 ± 0.05**

TABLE 5.11: The contributions from the NPFL estimation affecting the top quark mass measurement.

5.3.4.2 Background SM processes

The uncertainties from the background physics processes come from the simulation uncertainties, which are shown in Table 4.2. These physics processes occur in both the fake rate control region and the signal region, and thus has to be varied simultaneously in both regions. When varying the simulation uncertainties in the fake rate control region, the fake rates change, and thus a new NPFL contribution will be produced in the signal region. The change in the top quark mass that results from the change in the NPFL and physics process contributions in the signal region, are considered as the systematic uncertainty.

As mentioned under Section 4.4.4, the NPFL is only present in the range $p_T < 40$ GeV and $p_T < 30$ GeV in the electron and muon channels, respectively, since the data and SM prediction agree within uncertainty in the higher p_T regions in both channels. This observation came from the tight lepton distributions, as the tight lepton contribution is the numerator of the fake rate

calculations. Similarly, the data will be compared to the total SM processes before and after varying each SM processes by the relevant uncertainties, when the lepton is identified as tight.

- **W+jets**

The W+jets SM processes have a cross-section uncertainty of 5% and a normalization uncertainty of 30% [137]. Both variations will be applied individually in both the fake rate control region and the signal region.

The contribution from W+jets processes is one of the largest in the fake rate control region. Figure 5.18 shows the data and total SM contributions before and after varying the W+jets contribution by the cross-section (top row) and normalization (bottom row) uncertainties, in the transverse momentum distribution of the electron (left) and muon (right) channels, respectively, for 2018. The other years' results can be found in Appendix H. The same analysis made for 2018 applies to the other years.

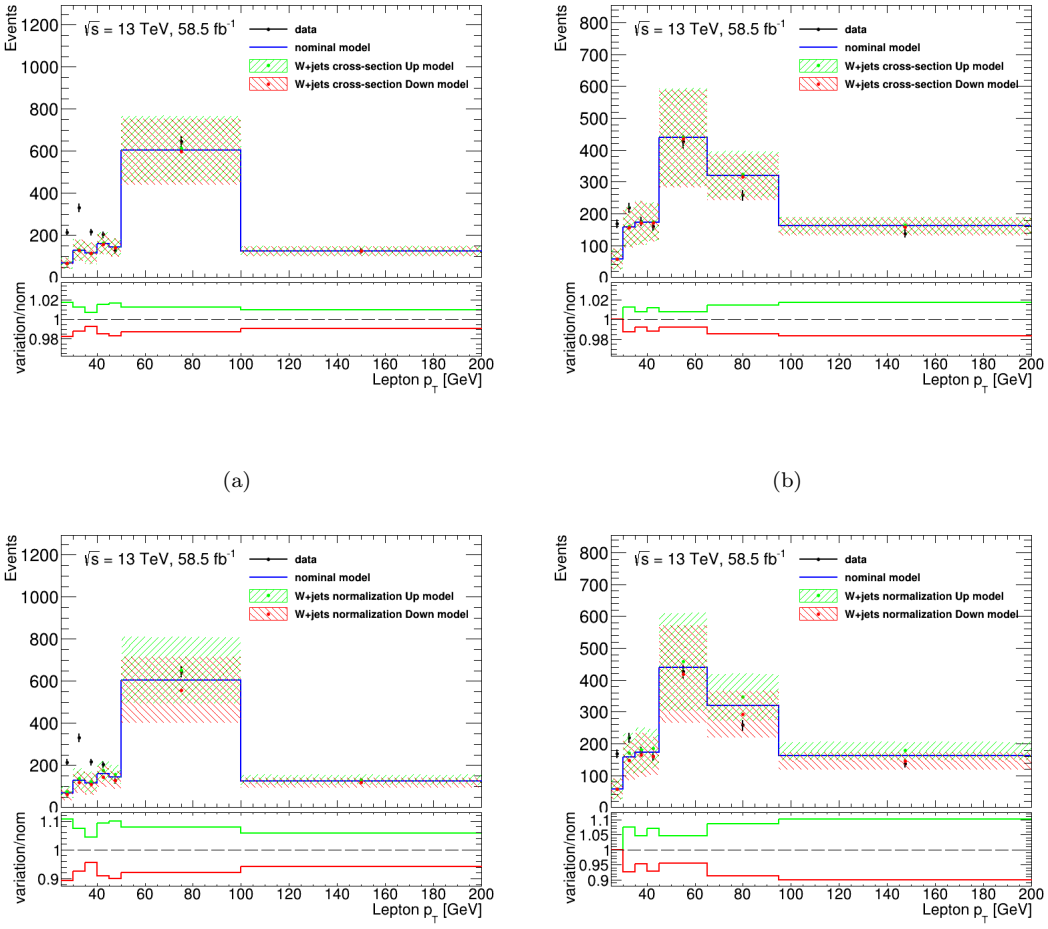


FIGURE 5.18: The upper panels show the data and total SM contribution in the fake rates control region in the transverse momentum distribution in the tight electron (left) and muon (right) channels for 2018, when varying the W+jets SM processes up and down by its cross-section (top row) and normalization (bottom row) uncertainties. The black points show the data and the blue line shows the nominal model contribution, while the green and red show the model contributions after the variations. The hashed areas show the statistical, luminosity and simulation uncertainties on the model contributions. The lower panel plot shows the ratio between the model after the variations and the nominal model.

The cross-section uncertainty variations produce differences up to 2% when compared to the nominal model, and therefore the same outcome as the nominal is observed, which shows the presence of NPFL when $p_T < 40$ GeV in the electron and $p_T < 30$ GeV in the muon channels, respectively. The normalization uncertainty variations produce differences up to 10% when compared to the nominal model, however, the same outcome as the nominal is observed. Only the fake rates in these p_T regions will be used to estimate the NPFL in the signal region. Figure 5.19 shows the nominal efficiency and fake rate, and the fake rates after the cross-section (top row) and normalization (bottom row) uncertainty variations in the transverse momentum distribution for 2018.

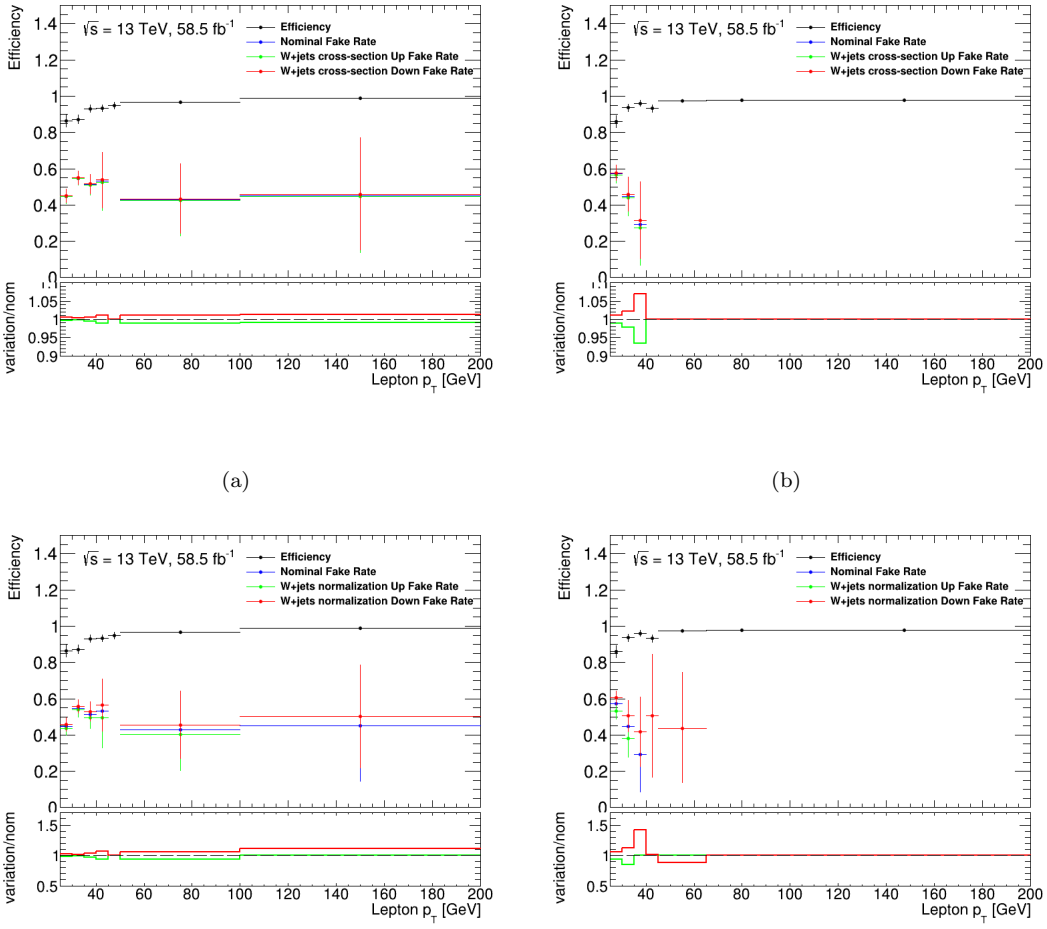
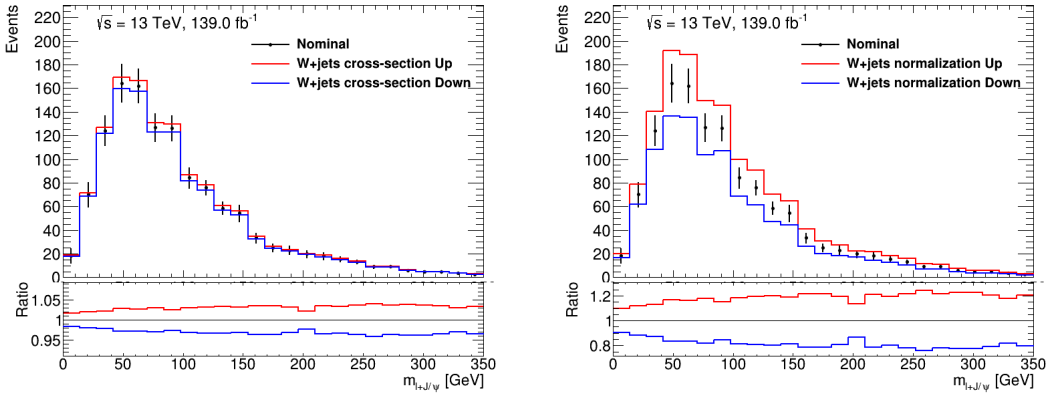


FIGURE 5.19: The upper panels show the nominal efficiencies and fake rates, and after the cross-section (top row) and normalization (bottom row) uncertainty variations have been applied to the W+jets SM processes in the transverse momentum distribution, in the electron (left) and muon (right) channels for 2018. The black and blue points show the nominal efficiencies and fake rates, while the green and red show the up and down variations, respectively. The error bars show the statistical uncertainty. The lower panel plot shows the ratio between the fake rate after variation and the nominal.

Figure 5.20 shows the invariant mass of the lepton + J/ψ distribution of the nominal (black) background model, and the background model when the W+jets SM processes are varied up (red) and down (blue) within their cross-section (a) and normalization (b) uncertainties. The lower panel shows the ratio between the variations and the nominal. The largest difference is around 4% for the cross-section variation and 20% for the normalization variation. The individual systematic uncertainties can be seen in Table 5.12, and the total systematic uncertainty associated with the W+jets contribution is 0.10 ± 0.12 GeV. This systematic uncertainty is dominated by its statistical component. Additional simulation could reduce the sensitivity to possible statistical fluctuations.



(a) (b)

FIGURE 5.20: The upper panel shows the invariant mass of the lepton + J/ψ distribution of the nominal (black) background model, and the background model when the W+jets SM processes are varied up (red) and down (blue) within their cross-section (a) and normalization (b) uncertainties. The lower panel shows the ratio between the variations and the nominal.

Source of uncertainty	$\sigma(m_{top})$ [GeV]
Cross-section	0.015 ± 0.013
normalization	0.094 ± 0.125
W+jets	**0.10 ± 0.12**

TABLE 5.12: The contributions from the W+jets SM processes affecting the top quark mass measurement.

- **Z+jets**

The Z+jets SM processes have a cross-section uncertainty of 5% and a normalization uncertainty that is dependent on the jet multiplicity [137]. Both variations will be applied individually in both the fake rate control region and the signal region.

The contribution from Z+jets processes is reasonably large in the fake rate control region. Figure 5.21 shows the data and total SM contributions before and after varying the Z+jets contribution by the cross-section (top row) and normalization (bottom row) uncertainties in the transverse momentum distribution in the electron (left) and muon (right) channels, respectively, for 2018. The other years' results can be found in Appendix H. The same anlaysis made for 2018 applies to the other years.

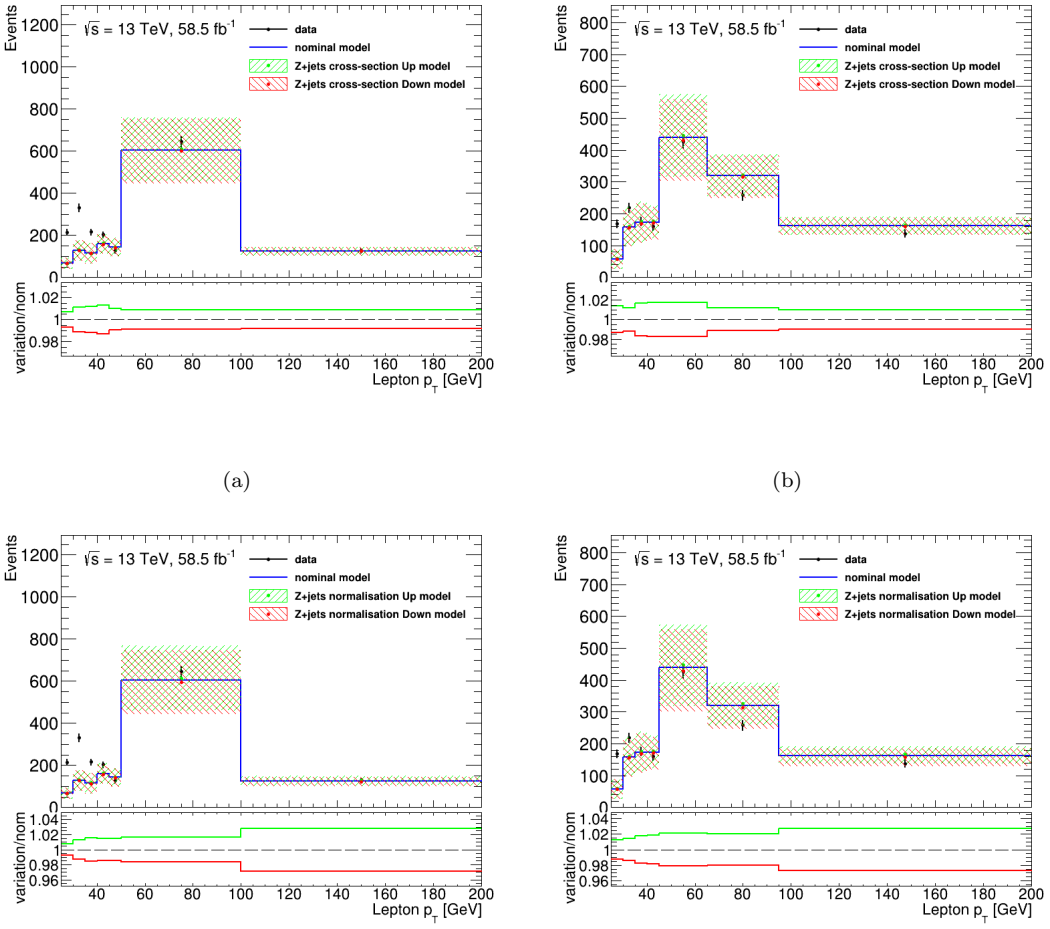


FIGURE 5.21: The upper panels show the data and total SM contribution in the fake rates control region in the transverse momentum distribution in the tight electron (left) and muon (right) channels for 2018, when varying the Z+jets SM processes up and down by its cross-section (top row) and normalization (bottom row) uncertainties. The black points show the data and the blue line shows the nominal model contribution, while the green and red show the model contributions after the variations. The hashed areas show the statistical, luminosity and simulation uncertainties on the model contributions. The lower panel plot shows the ratio between the model after the variations and the nominal model.

The cross-section uncertainty variations produce differences up to 2% when compared to the nominal model, and therefore the same outcome as the nominal is observed, which shows the presence of NPFL when $p_T < 40$ GeV in the electron and $p_T < 30$ GeV in the muon channels, respectively. The normalization uncertainty variations produce differences up to 3% when compared to the nominal model, however the same outcome as the nominal is observed. Only the fake rates in these p_T regions will be used to estimate the NPFL in the signal region. Figure 5.22 shows the nominal efficiency and fake rate, and the fake rates after the cross-section (top row) and normalization (bottom row) uncertainty variations in the transverse momentum distribution for 2018.

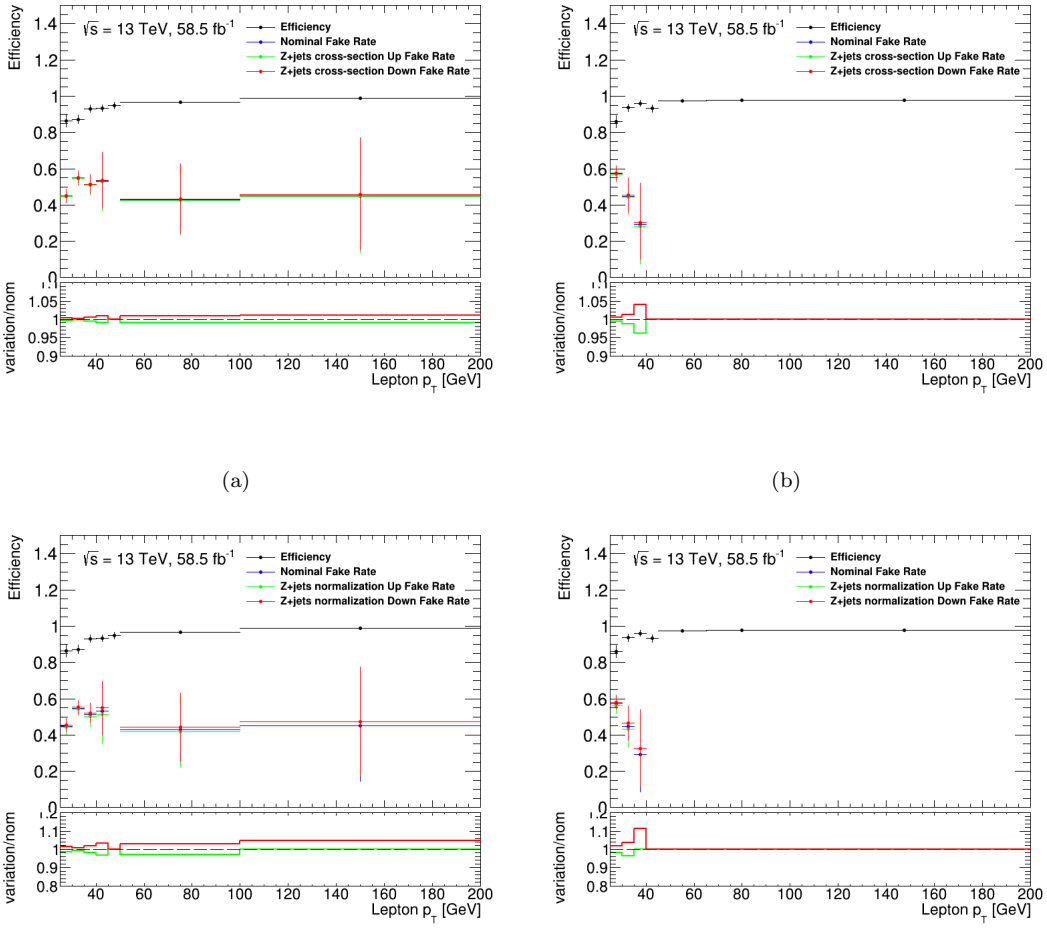
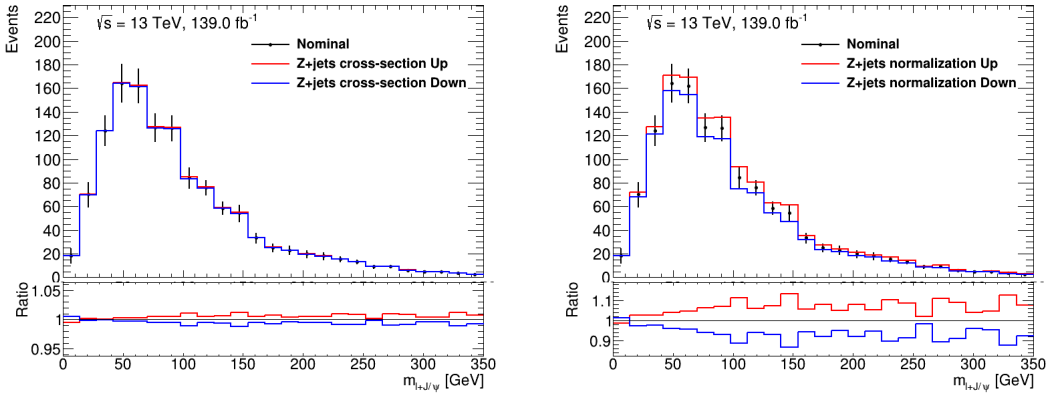


FIGURE 5.22: The upper panels show the nominal efficiencies and fake rates, and after the cross-section (top row) and normalization (bottom row) uncertainty variations have been applied to the Z+jets SM processes in the transverse momentum distribution, in the electron (left) and muon (right) channels for 2018. The black and blue points show the nominal efficiencies and fake rates, while the green and red show the up and down variations, respectively. The error bars show the statistical uncertainty. The lower panel plot shows the ratio between the fake rate after variation and the nominal.

Figure 5.23 shows the invariant mass of the lepton + J/ψ distribution of the nominal (black) background model, and the background model when the Z+jets SM processes are varied up (red) and down (blue) within their cross-section (a) and normalization (b) uncertainties. The lower panel shows the ratio between the variations and the nominal. The largest difference is around 2% for the cross-section variation and 15% for the normalization variation. The individual systematic uncertainties can be seen in Table 5.12, and the total systematic uncertainty associated with the Z+jets contribution is 0.12 ± 0.10 GeV. The statistical component is $\sim 81\%$ of the systematic uncertainty, and therefore additional simulation could improve this systematic.



(a) (b)

FIGURE 5.23: The upper panel shows the invariant mass of the lepton + J/ψ distribution of the nominal (black) background model, and the background model when the Z+jets SM processes are varied up (red) and down (blue) within their cross-section (a) and normalization (b) uncertainties. The lower panel shows the ratio between the variations and the nominal.

Source of uncertainty	$\sigma(m_{top})$ [GeV]
Cross-section	0.022 ± 0.013
normalization	0.122 ± 0.101
Z+jets	$\mathbf{0.12 \pm 0.10}$

TABLE 5.13: The contributions from the Z+jets SM processes affecting the top quark mass measurement.

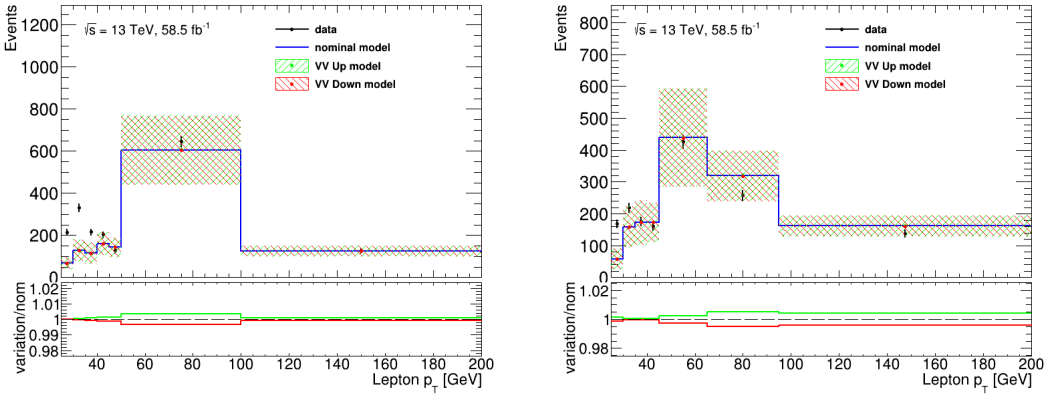
- VV

The diboson (VV) SM processes have a cross-section uncertainty of 50% [138], which will be applied in both the fake rate control region and the signal region.

The contribution from VV processes is reasonably small in the fake rate control region. Figure 5.24 shows the data and total SM contributions before and after varying the VV contribution by its cross-section uncertainty in the transverse momentum distribution in the electron (a) and muon (b) channels, respectively, for 2018. The other years' results can be found in Appendix H. The same anlaysis made for 2018 applies to the other years.

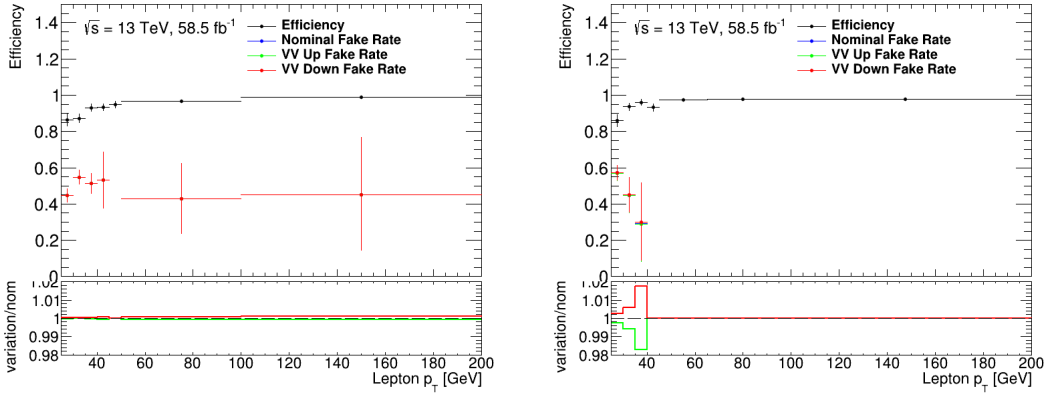
The cross-section uncertainty variations produce differences up to 0.5% when compared to the nominal model, and therefore the same outcome as the nominal is observed, which shows the presence of NPFL when $p_T < 40$ GeV in the electron and $p_T < 30$ GeV in the muon channels, respectively. Only the fake rates in these p_T regions will be used to estimate the NPFL in the

signal region. Figure 5.25 shows the nominal efficiency and fake rate, and the fake rates after the cross-section uncertainty variations in the transverse momentum distribution for 2018.



(a) (b)

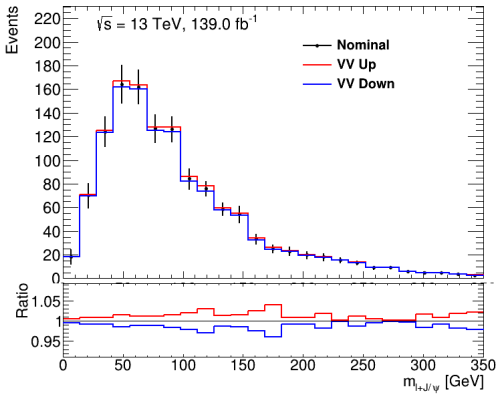
FIGURE 5.24: The upper panels show the data and total SM contribution in the fake rates control region in the transverse momentum distribution in the tight electron (a) and muon (b) channels for 2018, when varying the VV SM processes up and down by their cross-section uncertainties. The black points show the data and the blue line shows the nominal model contribution, while the green and red show the model contributions after the variations. The hashed areas show the statistical, luminosity and simulation uncertainties on the model contributions. The lower panel plot shows the ratio between the model after the variations and the nominal model.



(a) (b)

FIGURE 5.25: The upper panels show the nominal efficiencies and fake rates, and after the cross-section uncertainty variations have been applied to the VV SM processes in the transverse momentum distribution, in the electron (a) and muon (b) channels for 2018. The black and blue points show the nominal efficiencies and fake rates, while the green and red show the up and down variations, respectively. The error bars show the statistical uncertainty. The lower panel plot shows the ratio between the fake rate after variation and the nominal.

Figure 5.26 shows the invariant mass of the lepton + J/ψ distribution of the nominal (black) background model, and the background model when the VV SM processes are varied up (red) and down (blue) within their cross-section uncertainties. The lower panel shows the ratio between the variations and the nominal. The largest difference is around 4%, and thus a small impact is expected. The systematic uncertainty associated with the VV contribution is 0.03 $\pm$ 0.02 GeV. The statistical uncertainty is $\sim$ 55% of the systematic uncertainty, and therefore additional simulation could improve this systematic.



(a)

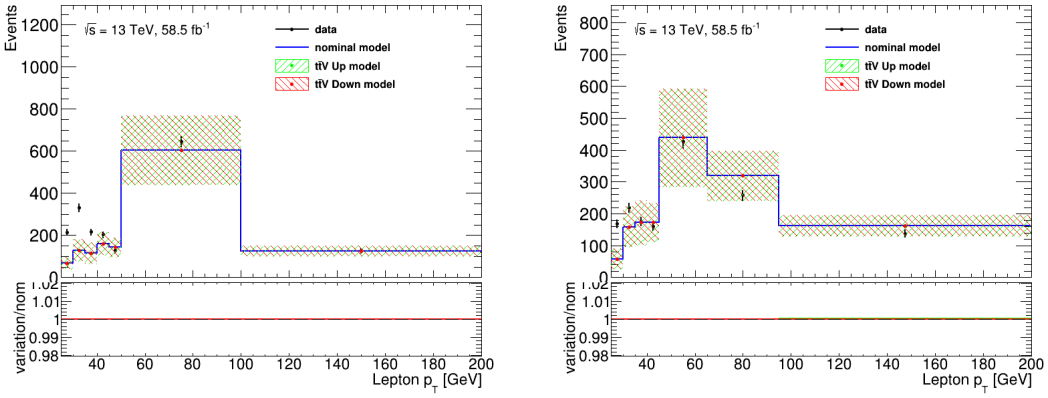
FIGURE 5.26: The upper panel shows the invariant mass of the lepton + J/ψ distribution of the nominal (black) background model, and the background model when the VV SM processes are varied up (red) and down (blue) within their uncertainties. The lower panel shows the ratio between the variations and the nominal.

- $t\bar{t}V$

The $t\bar{t}V$ SM processes have a cross-section uncertainty of 13% [139–141], which will be applied in both the fake rate control region and the signal region.

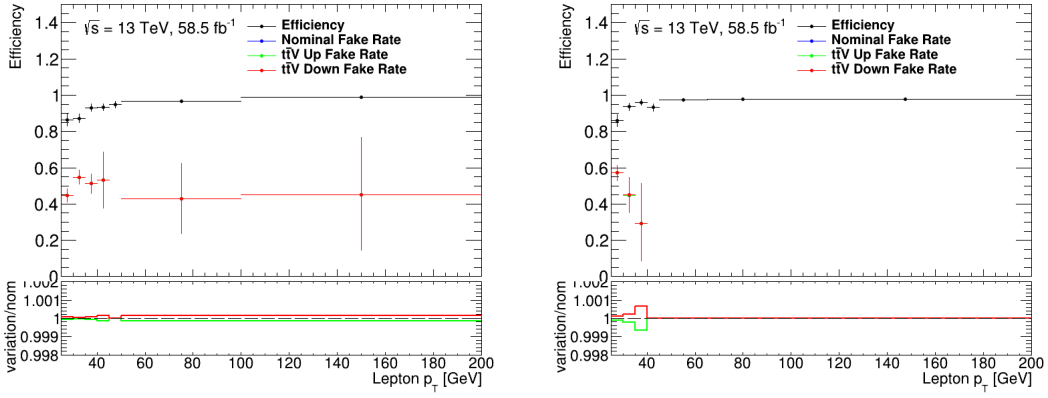
The contribution from $t\bar{t}V$ processes is negligibly small in the fake rate control region. Figure 5.27 shows the data and total SM contributions before and after varying the $t\bar{t}V$ contribution by its cross-section uncertainty in the transverse momentum distribution in the electron (a) and muon (b) channels, respectively, for 2018. The other years' results can be found in Appendix H. The same anlaysis made for 2018 applies to the other years.

The cross-section uncertainty variations produce negligible differences when compared to the nominal model, and therefore the same outcome as the nominal is observed, which shows the presence of NPFL when $p_T < 40$ GeV in the electron and $p_T < 30$ GeV in the muon channels, respectively. Only the fake rates in these p_T regions will be used to estimate the NPFL in the signal region. Figure 5.28 shows the nominal efficiency and fake rate, and the fake rates after the cross-section uncertainty variations in the transverse momentum distribution for 2018.



(a) (b)

FIGURE 5.27: The upper panels show the data and total SM contribution in the fake rates control region in the transverse momentum distribution in the tight electron (a) and muon (b) channels for 2018, when varying the $t\bar{t}V$ SM processes up and down by their cross-section uncertainties. The black points show the data and the blue line shows the nominal model contribution, while the green and red show the model contributions after the variations. The hashed areas show the statistical, luminosity and simulation uncertainties on the model contributions. The lower panel plot shows the ratio between the model after the variations and the nominal model.

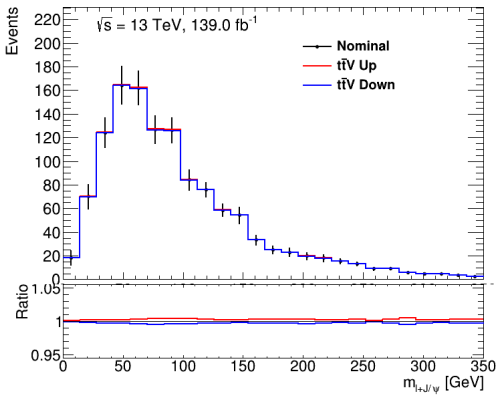


(a) (b)

FIGURE 5.28: The upper panels show the nominal efficiencies and fake rates, and after the cross-section uncertainty variations have been applied to the $t\bar{t}V$ SM processes in the transverse momentum distribution, in the electron (a) and muon (b) channels for 2018. The black and blue points show the nominal efficiencies and fake rates, while the green and red show the up and down variations, respectively. The error bars show the statistical uncertainty. The lower panel plot shows the ratio between the fake rate after variation and the nominal.

Figure 5.29 shows the invariant mass of the lepton + J/ψ distribution of the nominal (black) background model, and the background model when the $t\bar{t}V$ SM processes are varied up (red) and down (blue) within their cross-section uncertainties. The lower panel shows the ratio

between the variations and the nominal. The largest difference is around 0.5%, and thus a small impact is expected. The systematic uncertainty associated with the $t\bar{t}V$ contribution is 0.00 ± 0.00 GeV.



(a)

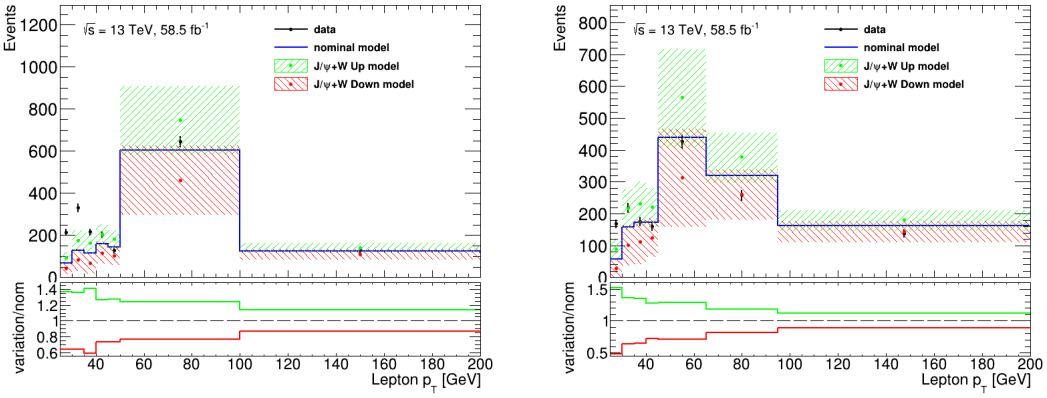
FIGURE 5.29: The upper panel shows the invariant mass of the lepton + J/ψ distribution of the nominal (black) background model, and the background model when the $t\bar{t}V$ SM processes are varied up (red) and down (blue) within their uncertainties. The lower panel shows the ratio between the variations and the nominal.

- $J/\psi + W$

The $J/\psi + W$ SM processes have a cross-section uncertainty of 100% [142], which will be applied in both the fake rate control region and the signal region.

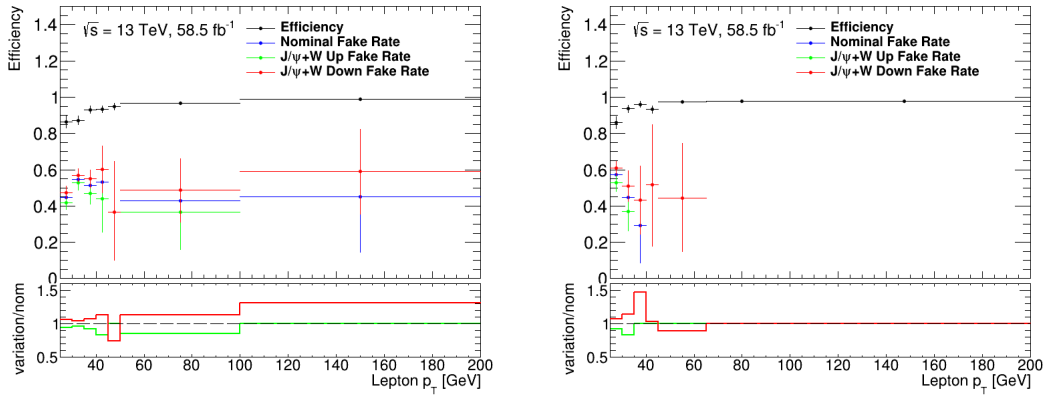
The contribution from $J/\psi + W$ processes is arguably the largest in the fake rate control region, and with 100% uncertainty on the cross-section, large changes are expected. Figure 5.30 shows the data and total SM contributions before and after varying the $J/\psi + W$ contribution by its cross-section uncertainty in the transverse momentum distribution in the electron (a) and muon (b) channels, respectively, for 2018. The other years' results can be found in Appendix H. The same anlaysis made for 2018 applies to the other years.

The cross-section uncertainty variations produce large differences, up to 50% when compared to the nominal model. In both channels, the up variations results in the same outcome as the nominal, which shows the presence of NPFL when $p_T < 40$ GeV in the electron and $p_T < 30$ GeV in the muon channels, respectively. However, the down variations result in an extra bin where the data and SM processes disagree, and therefore the presence of NPFL exists when $p_T < 45$ GeV in the electron and $p_T < 35$ GeV in the muon channels, respectively. Only the fake rates in these p_T regions will be used to estimate the NPFL in the signal region. Figure 5.31 shows the nominal efficiency and fake rate, and the fake rates after the cross-section uncertainty variations in the transverse momentum distribution for 2018.



(a) (b)

FIGURE 5.30: The upper panels show the data and total SM contribution in the fake rates control region in the transverse momentum distribution in the tight electron (a) and muon (b) channels for 2018, when varying the $J/\psi + W$ SM processes up and down by their cross-section uncertainties. The black points show the data and the blue line shows the nominal model contribution, while the green and red show the model contributions after the variations. The hashed areas show the statistical, luminosity and simulation uncertainties on the model contributions. The lower panel plot shows the ratio between the model after the variations and the nominal model.

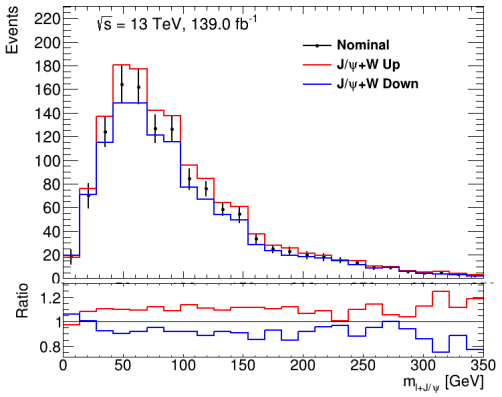


(a) (b)

FIGURE 5.31: The upper panels show the nominal efficiencies and fake rates, and after the cross-section uncertainty variations have been applied to the $J/\psi + W$ SM processes in the transverse momentum distribution, in the electron (a) and muon (b) channels for 2018. The black and blue points show the nominal efficiencies and fake rates, while the green and red show the up and down variations, respectively. The error bars show the statistical uncertainty. The lower panel plot shows the ratio between the fake rate after variation and the nominal.

Figure 5.32 shows the invariant mass of the lepton + J/ψ distribution of the nominal (black) background model, and the background model when the $J/\psi + W$ SM processes are varied up (red) and down (blue) within their cross-section uncertainties. The lower panel shows the ratio

between the variations and the nominal. The largest difference is around 30%. The systematic uncertainty associated with the $J/\psi + W$ contribution is 0.04 ± 0.15 GeV. This systematic uncertainty is dominated by its statistical component. Additional simulation could reduce the sensitivity to possible statistical fluctuations. Furthermore, the cross-section is not well-known, and therefore, if more accurate measurements of the cross-section are produced, the impact of this systematic could improve.



(a)

FIGURE 5.32: The upper panel shows the invariant mass of the lepton + J/ψ distribution of the nominal (black) background model, and the background model when the $J/\psi + W$ SM processes are varied up (red) and down (blue) within their uncertainties. The lower panel shows the ratio between the variations and the nominal.

5.3.5 Experimental uncertainties

The uncertainties associated to the reconstructed objects used in the analysis affect the measured top quark mass. There are a wide range of uncertainties tied to each object based on a number of different aspects, which could due to the efficiencies arising from the various selections criteria, parameters within algorithms, and more.

All experimental (or detector) systematic uncertainties going forward are estimated from top quark pair events. At the time of writing, the single top quark events detector systematics have not been estimated, and will be considered in future studies.

The pdf fit over each experimental systematic component can be found in Appendix I.

5.3.5.1 Leptons

As mentioned under Section 3.3.5.4 and 3.3.4.4, there are efficiency corrections (or scale factors, SF) applied to the simulated events to match data. These scale factors have a total uncertainty

of $< 0.5\%$ for muons with $p_T > 15$ GeV across the muon p_T spectrum [95], and for electrons with $p_T > 30$ GeV [94]. The muons from the J/ψ can have a momentum down to $p_T = 3.0$ GeV. The scale factors for muons with $3.0 < p_T < 15$ GeV were estimated from $J/\psi \to \mu\mu$ events, while scale factors for muons with $p_T > 15$ GeV came from $Z \to \mu\mu$ events. The uncertainties from these two different selections are treated as uncorrelated, as mentioned in reference [95]. This results in a discontinuity between the low and high p_T muon scale factors, and thus leads to significant uncertainties. All muon scale factors consist of separate statistical and systematic variations.

Additional sources of uncertainty come from the corrections made to the simulation's lepton momentum scale and resolution to match those in the data. These corrections were made using reconstructed distributions of the $Z \to \mu\mu/ee$ and $J/\psi \to \mu\mu/ee$ masses as well as the measured E/p in $W \to e\nu$ events, where E and p are the electron energy and momentum as determined by the calorimeter and the tracker, respectively [94, 95]. The event selection is performed with the lepton momentum changed by $\pm 1\sigma$ in order to assess the impact of momentum scale uncertainty, and smeared to account for the momentum resolution uncertainties.

The individual systematic uncertainties can be seen in Table 5.14, and the total systematic uncertainty associated with the selected leptons is 0.36 ± 0.09 GeV. The statistical component is $\sim 26\%$ of the systematic uncertainty, and therefore additional simulation could improve this systematic.

However, this systematic is dominated by the identification scale factor for muons from the J/ψ, especially the systematic variation for muons with $p_T < 15$ GeV. These muons must satisfy the Low-p_T identification criteria. Figure 5.33 shows the efficiency of the Low-p_T selection criteria in different η regions for muons with $3 < p_T < 15$ GeV. The total uncertainty of muons satisfying the Low-p_T selection with $3.0 < p_T < 5.0$ GeV is quite large ($> 10\%$) and therefore, a large uncertainty should be expected for these muons. Selecting J/ψ with muons that have $p_T > 5$ GeV could reduce the systematic uncertainties, however, the statistics is already limited and throwing away these events would reduce the precision by a factor of three. Additional data and simulation would allow for an increase in the p_T selection of the muons from the J/ψ and thus could improve this systematic uncertainty.

Source of uncertainty	$\sigma(m_{top})$ [GeV]
Electron momenta systematics	
Momentum scale	0.040 ± 0.015
Momentum resolution	0.020 ± 0.004
W-boson electron systematics	
Trigger SF	0.000 ± 0.001
Reconstruction SF	0.003 ± 0.002
Identification SF	0.007 ± 0.007
Isolation SF	0.003 ± 0.002
Muon momenta systematics	
Momentum scale	0.006 ± 0.001
Momentum ID resolution	0.012 ± 0.009
Momentum MS resolution	0.009 ± 0.004
Charge-dependent momentum scale	0.008 ± 0.004
W-boson muons systematics	
Trigger SF statistical	0.003 ± 0.002
Trigger SF systematic	0.001 ± 0.001
Identification SF statistical	0.000 ± 0.000
Identification SF systematic	0.009 ± 0.003
Isolation SF statistical	0.010 ± 0.004
Isolation SF systematic	0.108 ± 0.023
TTVA SF statistical	0.002 ± 0.001
TTVA SF systematic	0.003 ± 0.001
J/ψ muons systematics	
Identification SF statistical $p_T > 15$ GeV	0.007 ± 0.002
Identification SF systematic $p_T > 15$ GeV	0.019 ± 0.008
Identification SF statistical $p_T < 15$ GeV	0.074 ± 0.021
Identification SF systematic $p_T < 15$ GeV	0.328 ± 0.102
Leptons	**0.36 ± 0.09**

TABLE 5.14: The contributions from the various lepton systematic sources affecting the top quark mass measurement. The uncertainty values quoted as 0.000 are smaller than 0.0005.

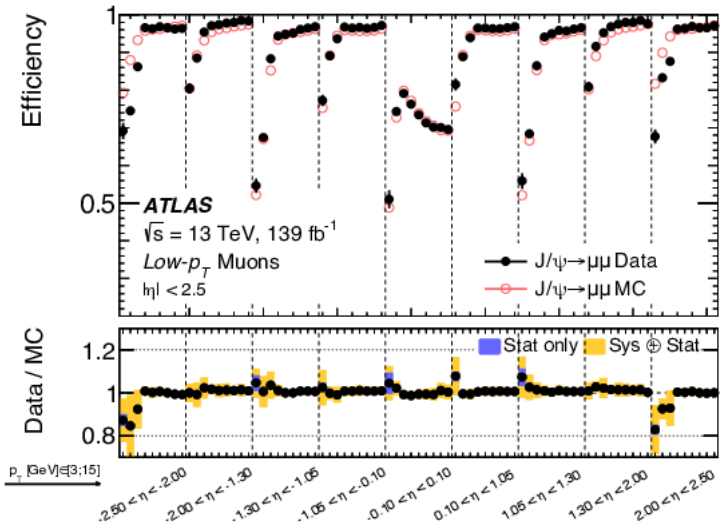


FIGURE 5.33: Muon reconstruction and identification efficiencies for the Low-p_T selection criteria. The plot shows the efficiencies measured in J$\psi \to \mu\mu$ events for different η regions in nine p_T bins (3-3.5, 3.5-4, 4-5, 5-6, 6-7, 7-8, 8-10, 10-12, 12-15 GeV). The resulting values are plotted as distinct measurements in each η bin, with p_T increasing from 3 to 15 GeV going from left to right. When not negligible, the statistical uncertainty in the efficiency measurement is indicated by the error bars. The panel at the bottom shows the ratio of the measured to predicted efficiencies, with statistical and systematic uncertainties [95].

5.3.5.2 Jets

The accuracy of jet reconstruction and identification using the JVT variable, as well as the jet energy scale (JES) and resolution (JER), all contribute to the uncertainty surrounding jets.

The jet-vertex-tagger (JVT) uncertainty includes a systematic uncertainty determined by utilizing various MC generators for simulation of $Z \to \mu\mu$ and $t\bar{t}$ events, as well as uncertainty on the assessment of the residual contamination from pile-up jets after pile-up suppression [169]. The systematic uncertainty associated with the JVT is 0.01 $\pm$ 0.01 GeV.

The uncertainty due to the JES is estimated by combining LHC collision data and simulation and test-beam data [170], and consist of 29 decorrelated nuisance parameters with contributions from flavour of the jet, pile-up, punch-through and single-particle response. The JES uncertainty rapidly decreases with increasing jet p_T and is around 5.5% for jets with p_T = 25 GeV [171]. For central jets with p_T between 100 GeV and 1.5 TeV, it is less than 1.5%. It is one of the main causes of uncertainty for reconstructed objects, having an impact on how the signal and backgrounds are normalized in relation to one another across different bins of the number of jets. The individual systematic uncertainties can be seen in Table 5.15, and the total systematic uncertainty associated with the jet energy scale is 0.14 $\pm$ 0.01 GeV. Additionally, the systematic uncertainty associated with the b-jet energy scale is 0.06 $\pm$ 0.01 GeV.

Source of uncertainty	$\sigma(m_{top})$ [GeV]
Pile-up parameters	
Number of vertices	0.012 ± 0.005
Mean number of interactions	0.001 ± 0.004
p_T term	0.016 ± 0.007
Density	0.038 ± 0.002
***in-situ* parameters**	
Detector 1	0.001 ± 0.000
Detector 2	0.001 ± 0.001
Mixed 1	0.012 ± 0.004
Mixed 2	0.009 ± 0.001
Mixed 3	0.001 ± 0.000
Modelling 1	0.004 ± 0.002
Modelling 2	0.001 ± 0.002
Modelling 3	0.002 ± 0.001
Modelling 4	0.000 ± 0.001
Statistical 1	0.004 ± 0.000
Statistical 2	0.007 ± 0.000
Statistical 3	0.000 ± 0.001
Statistical 4	0.001 ± 0.000
Statistical 5	0.001 ± 0.000
Statistical 6	0.002 ± 0.000
η-calibration parameters	
Method	0.012 ± 0.007
Modelling	0.060 ± 0.017
High energy non-closure	0.000 ± 0.000
Positive η non-closure	0.001 ± 0.001
Negative η non-closure	0.001 ± 0.001
Additional parameters	
Flavour composition	0.105 ± 0.015
Flavour response	0.058 ± 0.003
High-p_T jets	0.000 ± 0.000
Punch-through jets	0.000 ± 0.000
Fast simulation non-closure	0.001 ± 0.003
JES	**0.14 ± 0.01**

TABLE 5.15: The contributions from the various JES systematic sources affecting the top quark mass measurement. The uncertainty values quoted as 0.000 are smaller than 0.0005.

The JER uncertainty was estimated using dijet events in Run 1 data and simulation, i.e. collected between 2010 and 2012, as a function of jet p_T and rapidity. They were discovered to be within 10% of each other [172]. The extrapolation from the conditions of Run 1 to Run 2 has been analysed with additional uncertainties [170]. The combined uncertainty is propagated in the simulation by smearing the jet p_T [171]. It is determined through an eigenvector decomposition approach comparable to the systematic uncertainty in the jet energy scale. Numerous effects are examined through simulation-to-data comparisons using eight nuisance parameters. The individual systematic uncertainties can be seen in Table 5.16, and the total systematic uncertainty associated with the jet energy resolution is 0.12 ± 0.02 GeV.

Source of uncertainty	$\sigma(m_{top})$ [GeV]
Data vs MC	0.045 ± 0.018
Parameter 1	0.022 ± 0.037
Parameter 2	0.038 ± 0.020
Parameter 3	0.037 ± 0.001
Parameter 4	0.058 ± 0.011
Parameter 5	0.010 ± 0.004
Parameter 6	0.050 ± 0.021
Parameter 7	0.047 ± 0.004
JER	**0.12 ± 0.02**

TABLE 5.16: The contributions from the various JER systematic sources affecting the top quark mass measurement.

5.3.5.3 Flavour tagging

The efficiencies of b-tagging in simulated samples are rectified to match the efficiencies in data. Separate calibration analyses are used to produce correction scale factors (SF) for jets emanating from b, c, and light quarks [173–175]. SF are calculated as a function of p_T for jets originating from b- and c-quarks, whereas light-jet efficiency is scaled by p_T- and η-dependent parameters. The minimal spacing angle between the target jet and its nearby jets, as well as the hadron p_T over jet p_T ratio, are all factors that influence the tagging efficiency. The simulation's parton shower and hadronization model establishes these dependencies. In order to take these impacts into account, MC-to-MC correction factors are put in place. These factors correct the data-to-MC SF for the difference in the parton shower and hadronization model's tagging effectiveness and the sample in question.

By adjusting each source of uncertainty up and down by one standard deviation, uncertainties on the data-to-MC SF are estimated. These uncertainties are included into an eigen-variation

model with a reduction strategy that treats only significant variations separately while combining all other small variations into a single variation. A total of eight, three and three independent variations are considered for b-, c- and light-jets, respectively, and are taken as uncorrelated between the different tagged jets. Two additional uncertainties are included due to the extrapolation of SF for jets with p_T beyond the kinematic reach of the data calibration samples used. The uncertainty due to flavour tagging is the sum in quadrature of the these variations. The individual systematic uncertainties can be seen in Table 5.17, and the total systematic uncertainty associated with the flavour tagging is 0.02 ± 0.00 GeV.

Source of uncertainty	$\sigma(m_{top})$ [GeV]
b-tagging	
B variation 1	0.010 ± 0.004
B variation 2	0.015 ± 0.002
B variation 3	0.006 ± 0.002
B variation 4	0.001 ± 0.001
B variation 5	0.000 ± 0.000
B variation 6	0.000 ± 0.000
B variation 7	0.000 ± 0.000
B variation 8	0.000 ± 0.000
c-tagging	
C variation 1	0.001 ± 0.000
C variation 2	0.002 ± 0.000
C variation 3	0.000 ± 0.000
light jet-tagging	
L variation 1	0.000 ± 0.000
L variation 2	0.001 ± 0.000
L variation 3	0.000 ± 0.000
High-p_T extrapolation	
Extrapolation b-tagging	0.000 ± 0.000
Extrapolation c-tagging	0.014 ± 0.002
Flavour tagging	**0.02 ± 0.00**

TABLE 5.17: The contributions from the various flavour tagging systematic sources affecting the top quark mass measurement. The uncertainty values quoted as 0.000 are smaller than 0.0005.

5.3.5.4 Missing transverse momentum

The missing transverse momentum (or E_T^{miss}, or even just MET) uncertainty comes from the uncertainties associated with leptons and jet, since they affect the reconstruction of E_T^{miss}. Additional uncertainties related to the modelling of the underlying event are taken into consideration, including their effects on the p_T scale and resolution of unclustered energy. The individual systematic uncertainties can be seen in Table 5.18, and the total systematic uncertainty associated with the MET is 0.10 ± 0.01 GeV.

Source of uncertainty	$\sigma(m_{top})$ [GeV]
Scale	0.046 ± 0.012
Parallel resolution	0.089 ± 0.012
Perpendicular resolution	0.018 ± 0.006
MET	**0.102 ± 0.012**

TABLE 5.18: The contributions from the various MET systematic sources affecting the top quark mass measurement.

5.3.5.5 Pile-up

To match the conditions in the data, the distribution of the average number of interactions per bunch crossing in MC samples is reweighted. The uncertainty is based on the difference in instantaneous luminosity between data and simulation [176]. The systematic uncertainty associated with pile-up is 0.100 ± 0.04 GeV. The statistical component is ∼ 40% of the systematic uncertainty, and therefore additional simulation could improve this systematic.

5.3.6 Summary of the systematic uncertainties

There are a wide range of systematic uncertainties affecting this top quark mass measurement. Table 5.19 shows the various sources of systematic uncertainties and the impact each one has on the top quark mass.

The top quark mass measured in MC is 172.56 ± 0.90 (statistical) ± 2.75 (systematic) GeV. This has a precision of 1.67% on the total uncertainty and 1.59% on the total systematic uncertainty. The modelling of signal processes is the largest source of uncertainty, with the final-state radiation, b-quark fragmentation and choice of NLO generator dominating the contribution. The second-largest source of uncertainty comes from the statistical uncertainty in the determination of the top quark mass. The systematic uncertainties due to the modelling of background processes and detector response systematics are small in comparison, with the largest contributions coming from NPFL and W/Z+jets, as well as leptons, respectively.

This analysis searched for a top quark decay mode with a low branching ratio, and is expected to be statistical limited. This statistics limitation is observed in the statistical uncertainty of the measurement, as well as in the MC simulation statistics. The limited MC statistics leads to large systematic uncertainties, which dominate every area of the analysis, and therefore, most of the systematics could be improved with additional simulation.

Source of uncertainty	$\sigma(m_{top})$ [GeV]
Method	0.18 ± 0.06
NLO generator	1.08 ± 0.61
Parton shower and hadronization	0.09 ± 0.68
b-quark fragmentation	1.69 ± 0.61
B-hadron production fractions	0.08 ± 0.01
B-hadron decay branching ratios	0.08 ± 0.01
h_{damp} tuning	0.08 ± 0.42
Initial-state radiation	0.13 ± 0.08
Final-state radiation	1.76 ± 0.57
Colour reconnection	0.40 ± 0.44
Choice of PDF	0.05 ± 0.01
Modelling of signal processes	**2.71 ± 0.54**
W+jets	0.10 ± 0.12
Z+jets	0.12 ± 0.10
VV	0.03 ± 0.02
$t\bar{t}V$	0.00 ± 0.00
$J/\psi + W$	0.04 ± 0.15
NPFL	0.16 ± 0.05
Modelling of background processes	**0.23 ± 0.09**
Leptons	0.36 ± 0.09
Jet energy scale	0.14 ± 0.01
b-jet energy scale	0.06 ± 0.01
Jet energy resolution	0.12 ± 0.02
Jet vertex fraction	0.01 ± 0.01
Flavour tagging	0.02 ± 0.00
Missing transverse momentum	0.10 ± 0.01
Pile-up	0.10 ± 0.04
Detector Response	**0.43 ± 0.08**
Total systematic	**2.75 ± 0.53**
Statistics	0.90
Total	**2.89 ± 0.53**

TABLE 5.19: The contributions from the various sources affecting the top quark mass measurement. The uncertainty values quoted as 0.00 are smaller than 0.005.

5.4 The effect of larger samples on systematic uncertainties

As mentioned above, this analysis is heavily statistically limited due to the low branching ratio of the top quark decay mode of interest. In order to reduce the uncertainties, more events would need to be generated, which requires a lot of computational resources.

An event filter has been developed in the generation of simulated events to try and increase the statistical power. The event filter generates top quark pair events and only keeps events where one of the top quarks decayed to a b-quark and the b-quark produced a J/ψ that decayed into two oppositely charged muons. This filter reduced the amount of computational resources required to produce far more events than the previous simulated samples.

The dominant source of uncertainty comes from the modelling of the signal processes, and thus new samples for various signal modelling systematics can be generated. These new samples were generated with the same settings used before, and increased the statistical power at least seven-fold. However, these new filtered samples only account for events that have the top quark decay mode at the particle-level. Since the signal region consists of top quark events that do not have the top quark decay mode at the particle-level ($\sim$ 18 %), this contribution is taken from the non-filtered samples to ensure the model with these new filtered samples describe the data and is consistent with the nominal model.

At the time of writing, only some filtered samples have been produced and simulated using the fast ATLAS simulation. Filtered samples associated with the largest three signal modelling systematics are available and will be estimated. The pdf fit over these new samples can be found in Appendix J.

5.4.1 NLO generator

The uncertainty due to the NLO matrix element matching scheme is estimated by comparing a sample generated with Powheg-Box + Pythia 8 and one with MadGraph5_aMC@NLO + Pythia 8. Only the Powheg-Box sample is available with the filter at the time of writing, as simulation issues arose when producing the MadGraph5_aMC@NLO. The systematic uncertainty associated with the NLO generator was 1.08 $\pm$ 0.61 GeV and is now 0.05 $\pm$ 0.56 GeV. The reduction in the systematic uncertainty seems like a statistical fluctuation, since the statistical component has not improved by much. The statistical uncertainty is large and originates from the non-filtered MadGraph5_aMC@NLO sample, which has much fewer statistics than the new filtered Powheg-Box sample.

5.4.2 *b*-quark fragmentation

The *b*-quark fragmentation estimated above compared the reconstructed top quark masses from a sample with r_b=0.855 and one with r_b=1.05. These two r_b values are far apart and a large uncertainty was expected.

According to a previous top quark mass measurement [149], the best combined fit to LEP+SLD data produces an r_b=1.05 ± 0.02. Two samples were generated with the filter using r_b values of 1.03 and 1.07, respectively. Unfortunately, the events that do not contain the top quark decay mode of interest at the particle-level were not generated for the r_b values of 1.03 and 1.07. This contribution will be taken from r_b=1.05 to ensure the model is consistent and complete. The systematic uncertainty associated with the *b*-quark fragmentation was 1.69 ± 0.61 GeV and is now 0.18 ± 0.42 GeV.

The 0.18 ± 0.42 GeV is a more accurate representation of the *b*-quark fragmentation systematic, however, the nominal sample, as well as the templates, were generated with an r_b value of 0.855. At the time of writing, there are top quark samples generated with a m_{top} = 172.5 GeV and an r_b = 1.05 available, but not for the other m_{top} samples used to develop the extraction method. The r_b = 1.05 is the more accurate *b*-quark fragmentation parameter, and new samples with this r_b value will be considered in future studies.

Since the nominal sample is generated with an r_b = 0.855, 1.69 ± 0.61 GeV will be considered as the systematic uncertainty associated with *b*-quark fragmentation.

5.4.3 Final-state radiation

Final-state radiation (FSR) considers the impact of QCD emissions and is evaluated using varied PS weights in the Pythia 8 sample with r_b=1.05, which adjusts the renormalization scale up and down by a factor of two while keeping the factorisation scale constant. The systematic uncertainty associated with the final-state radiation was 1.76 ± 0.57 GeV and is now 1.12 ± 0.40 GeV.

This statistical component and systematic uncertainty are large, despite the increase in statistics. The same analysis as before is drawn, which is that the large event weights from Pythia 8 causes the large systematic uncertainty. Figure 5.34 shows the event weights for the renormalization scales 2.0 and 0.5, respectively. The event weights for μ_R=2.0 produce weights up to 170000, and a top quark mass of 170.32 ± 0.17 GeV, which is -2.243 GeV away from the nominal value. The event weights for μ_R=0.5 produce weights up to 40000, and a top quark mass of 172.56 ± 0.57 GeV, which is 0.001 GeV away from the nominal value. The same suggestion as before is given: reference [164] shows that Pythia 8's event weight variations of μ_R=2.0 and μ_R=0.5

roughly correspond to samples generated with set renormalization scale values of $\mu_R = \sqrt{2}$ and $\mu_R = 1/\sqrt{2}$, respectively. These new samples should not consist of large event weights, and could produce a more reasonable estimation of the final-state radiation systematic.

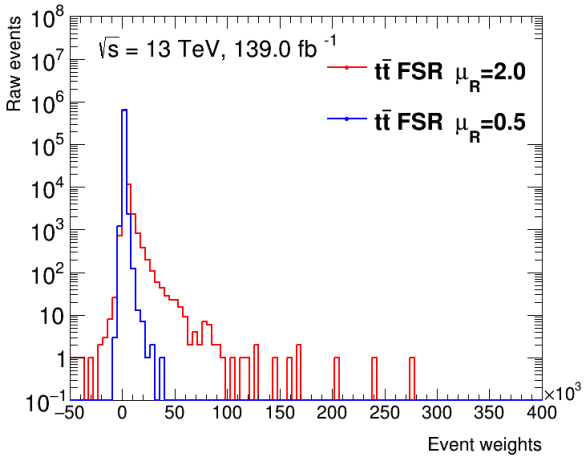


FIGURE 5.34: Distribution of event weights in the nominal top quark pair sample using the new filter when using a renormalization scale of 2.0 (blue) and 0.5 (red), respectively.

5.4.4 Summary of systematic uncertainties

Table 5.20 shows the comparison of two of the three largest signal modelling systematics estimated with and without the filtered samples, and the impact on the total systematic uncertainty. The signal modelling systematics improves by $\sim$ 24% after re-estimating the previous two of the three largest sources of uncertainty. The samples with more events improved the precision of the systematic uncertainty from 1.59% to 1.2%, however, the systematic due to b-quark fragmentation and FSR is large and dominates this analysis. Suggestions have been made to try and improve these systematic uncertainties, and should be considered in future studies.

Source of uncertainty	Without filter $\sigma(m_{top})$ [GeV]	With filter $\sigma(m_{top})$ [GeV]
NLO generator	1.08 ± 0.61	0.05 ± 0.56
Final-state radiation	1.76 ± 0.57	1.12 ± 0.40
Modelling of signal processes	**2.71 ± 0.54**	**2.08 ± 0.45**
Total systematic	**2.75 ± 0.53**	**2.14 ± 0.44**
Total	**2.89 ± 0.53**	**2.32 ± 0.44**

TABLE 5.20: The contributions from the three signal modelling systematics with and without using the filtered samples, and its impact on the total systematic uncertainty.

5.5 Extracting the top quark mass in data

The extraction method has been developed, tested, and the various sources of uncertainties has been estimated, and now the top quark mass in data can finally be measured.

Figure 5.35 shows the data and SM prediction in the signal region where the top quark is described by Pythia 8 with a mass of 172.5 GeV, in the invariant mass of the lepton + J/ψ distribution in the upper panel. The statistical and systematic uncertainties are shown. The systematic uncertainties are used as a visual representation and is not completely accurate. The systematic uncertainty comes from generating the invariant mass of the lepton + J/ψ distribution with a top quark mass $\pm 1\sigma$ systematic from the nominal value, using the total pdf. The lower panel shows the ratio between the data and SM prediction. The data and SM prediction agree within uncertainty over the variable sensitive to the top quark mass.

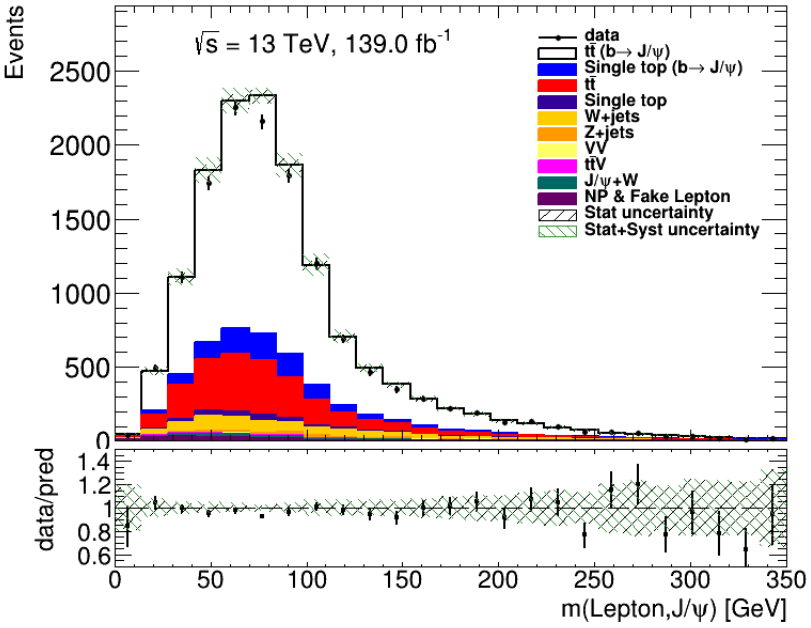


FIGURE 5.35: The upper panel shows the data and SM prediction in the signal region where the top quark is described by Pythia 8 with a mass of 172.5 GeV, in the invariant mass of the lepton + J/ψ distribution in the signal region. The statistical and systematic uncertainties are shown. The lower panel shows the ratio between the data and SM prediction.

Figure 5.36 shows the data and total pdf fit (blue) in the invariant mass of the lepton + J/ψ distribution in the upper panel. The individual contributions from the Gaussian (red), Gamma (green), and Novosibirsk (purple) are shown. The fit quality can be found in the top left corner of the upper panel. The lower panel shows the ratio between the data and the total pdf fit. The pdf fit over the data produces a reasonable fit quality of 0.76. Figure 5.37 shows the -log$\mathcal{L}$ of the fit to data. The curve is smooth around the minimum at least up to 5σ of the statistical

uncertainty, which indicates an accurate estimation of the statistical uncertainty. The top quark mass in data is, therefore, measured to be

$$m_{top} = 172.03 \pm 0.76 \text{ (stat)} \pm 2.14 \text{ (syst) GeV}.$$

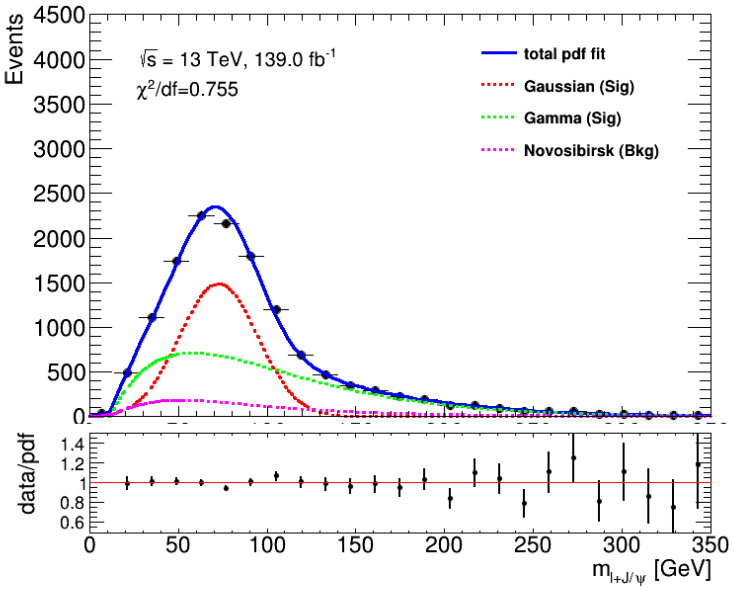


FIGURE 5.36: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution in data. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The lower panel shows the ratio between the model and the pdf.

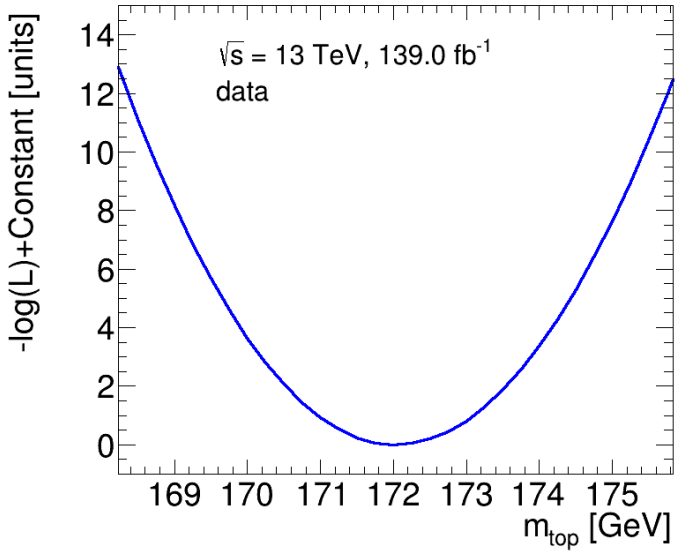


FIGURE 5.37: The -log$\mathcal{L}$ of the fit to data as a function of the top quark mass. The y-axis was manually shifted to zero by adding a constant.

Chapter 6

Conclusion

The top quark is not called the top without reason. It's the heaviest quark and fundamental particle in the SM. The top quark mass plays vital roles in a number of areas of physics, with the vacuum stability of the SM being one of the most crucial. The precision of the top quark mass is a treasure many physicists seek.

The leading precision measurement of the top quark mass is 171.77 ± 0.38 GeV. However, this has yet to be published in a peer reviewed journal, and, therefore, the leading published precision measurement of the top quark mass is considered to be 172.69 ± 0.25 (statistical) ± 0.41 (systematic) GeV. The most precise measurement comes from a combination of ATLAS measurements that used the kinematics of jets with either leptons or more jets. The dominant source of uncertainty comes from the jet energy scale, which constitutes 0.22 GeV. Furthermore, the uncertainties associated with jets plagues this measurement.

Measuring the top quark mass using the kinematics of the decay products of the top quark that do not consist of jets have low branching ratios. Huge amounts of data are required to start measuring the top quark mass with these non-jet decay modes, and even then, the measurements will be tormented by the limited statistics available.

However, you don't have to be great to start, but you have to start to be great. This analysis performed a top quark mass measurement using proton-proton collision data collected by the ATLAS detector during 2015-2018, which corresponded to a total integrated luminosity of 139.0 fb^{-1}. The top quark decay mode used to perform this measurement is $t \to W(\to l\nu)b(\to J/\psi[\to \mu^+\mu^-] + X)$, which has a branching ratio on the order of 10^{-4}. This analysis is expected to be statistically limited and not yet be competitive, but is a stepping stone for future measurements when more data becomes available.

The top quark mass is measured to be 172.03 ± 0.76 (statistical) ± 2.14 (systematic) GeV. The systematic uncertainties originating from modelling the signal processes were estimated only

from top quark pair processes and not from single-top quark process, which will need to be evaluated in the future. The statistical uncertainty is large, and the limited statistics has been shown to be a consistent occurrence throughout this measurement. The b-quark fragmentation and final-state radiation dominates the systematic uncertainty, and without them, the systematic uncertainty comes to 0.45 GeV, which would be more competitive. The large systematic from b-quark fragmentation is due to comparing far apart r_b values, however, comparing closer r_b values has been shown to drastically improve this systematic. The large systematic from FSR is affected by the large event weights in Pythia 8 when using renormalization scales of 2.0 and 0.5, respectively. Producing dedicated samples with renormalization scales of $\sqrt{2}$ and $1/\sqrt{2}$ could remove the large event weights problem and provide an improved estimate of the systematic uncertainty.

The statistical uncertainty associated with this measurement is larger than the combination of the statistical and systematic uncertainties in the most precise top quark measurement. This measurement becomes more inspiring when comparing specific systematic uncertainties between this and the leading precision measurement. The jet energy scale uncertainty in the leading precision measurement comes to 0.22 GeV, whereas in this measurement it is 0.14 GeV. The main objective of using a top quark decay mode without jets, is to reduce the dominating source of uncertainty, and this objective has been met.

There is light at the end of the tunnel, however, there are obstacles in the way. These obstacles require effort and time to overcome. Suggestions have been presented to try and reduce the dominant sources of systematic uncertainties, and should be investigated. However, the top quark decay mode used in this analysis will only become competitive, and potentially be the most precise top quark mass measurement, when the curse of limited statistics is broken, as almost every source of systematic uncertainty could be improved with more data and simulation.

THE END

Appendix A

Non-prompt and fake leptons background

The following chapter shows the non-prompt and fake lepton background results for 2015, 2016, and 2017, respectively.

A.1 Measurement of the fake rates

The fake rates, ϵ_{fake}, are measured in a control region dominated by non-prompt and fake lepton background events, i.e. multi-jet events, but can still contain contributions from real leptons. This control region was defined to closely represent the signal region in order to estimate the fake leptons more accurately, and the main difference is an orthogonal $\mathrm{E}_T^{\text{miss}}$ selection. These events must satisfy the following selection criteria:

- At least one jet
- Exactly one loose electron (muon) that fired a lepton trigger
- Exactly one J/ψ candidate with no selection on τ
- $\mathrm{E}_T^{\text{miss}} < 20$ GeV

This control region contains contributions from real leptons, which are modelled by different SM processes and therefore, to get a pure NPFL sample, these contributions are subtracted from the data. The fake rate is determined using Equation A.1.

$$\epsilon_{\text{fake}} = \left(\frac{\mathrm{N}^{\mathrm{T}}_{\text{fake}}}{\mathrm{N}^{\mathrm{L}}_{\text{fake}}}\right) = \left(\frac{\mathrm{N}^{\mathrm{T}} - \mathrm{N}^{\mathrm{T}}_{\text{real}}}{\mathrm{N}^{\mathrm{L}} - \mathrm{N}^{\mathrm{L}}_{\text{real}}}\right). \tag{A.1}$$

The results are split per year and per lepton to account for the different pile-up distributions and triggers, respectively. Table A.1 is the same table shown in the main body of the thesis, but is shown again for convenience. The yields are low for loose and tight leptons in both channels for every year, especially in 2015. Therefore, this control region is statistically limited across all years.

Channel	2015		2016		2017		2018		Total	
	data	model	data	model	data	model	data	model	data	model
Tight electron	135	64 ± 22	1025	685 ± 204	1434	1015 ± 285	1886	1359 ± 382	4480	3123 ± 519
Loose electron	239	66 ± 23	1339	711 ± 214	1921	1059 ± 302	2523	1425 ± 405	6022	3261 ± 549
Tight muon	86	66 ± 23	884	837 ± 233	1177	1190 ± 317	1556	1487 ± 417	3703	3580 ± 574
Loose muon	131	67 ± 24	1091	837 ± 244	1396	1235 ± 333	1856	1562 ± 444	4474	3701 ± 607

TABLE A.1: The number of events of the data and model for each channel split per year when the lepton passes either the loose or tight selection criteria.

The yields show deviations between the data and model within uncertainty in the electron channel, and agreement in the muon channel, for both loose and tight leptons. This indicates the presence of non-prompt and fake leptons (NPFL) in the electron channel and no NPFL in the muon channel.

The following results are shown for 2015, 2016 and 2017, while the results for 2018 can be found in the main body of the thesis in Section 4.4.4. Figures A.1 and A.2 show the distribution of E_T^{miss} for 2015, 2016, and 2017 in the electron and muon channels, respectively. These figures show large contributions originating from W/Z+jets and $J/\psi + W$ events in 2016 and 2017, with small contributions from the other SM processes. This is the same outcome found in 2018. In 2015, however, the largest contribution originates from $J/\psi + W$ events, with small contributions from W/Z+jets events and the other SM processes. The different W/Z+jets contributions between 2015 and 2016-2018 is due to the fact that 2015 has at least a factor of 10 fewer events than the other years. This control region is already statistically limited and therefore, the minimal contributions from W/Z+jets in 2015 are understandable.

The region between the top of the stacked simulated sources and the data is assumed to come from the NPFL background contribution. In the muon channel, the data and the SM prediction agree within uncertainty, indicating no presence of NPFL. In the electron channel, deviations between data and SM prediction indicate the presence of NPFL.

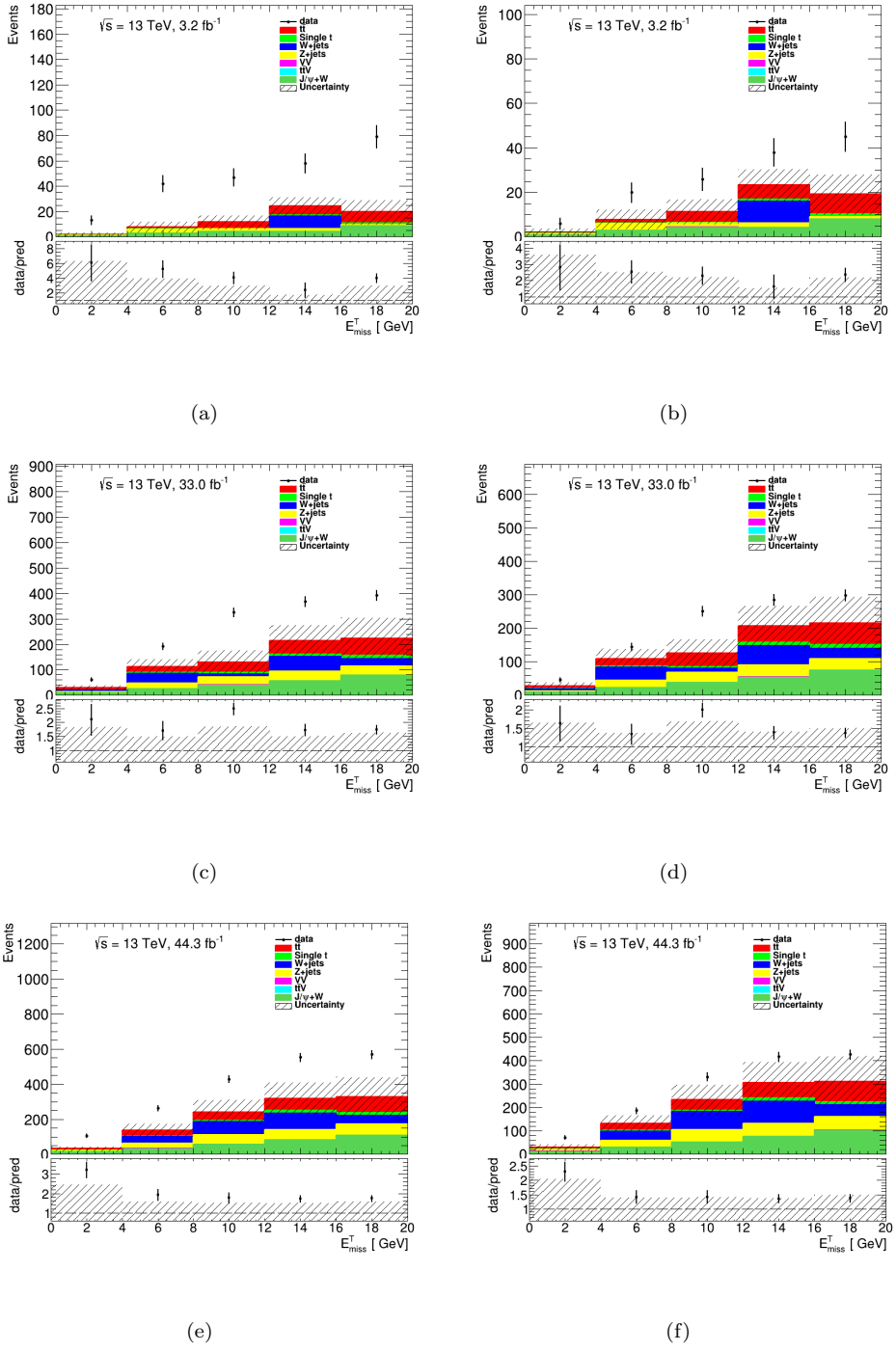


FIGURE A.1: The upper panel shows the distribution of E_T^{miss} for data and the different SM processes for events passing the loose (left) and tight (right) selection criteria in the electron channel for 2015 (a, b), 2016 (c, d), and 2017 (e, f), respectively. Only the statistical, cross-section and luminosity uncertainties are shown. The lower panel shows the ratio between the data and the SM predictions.

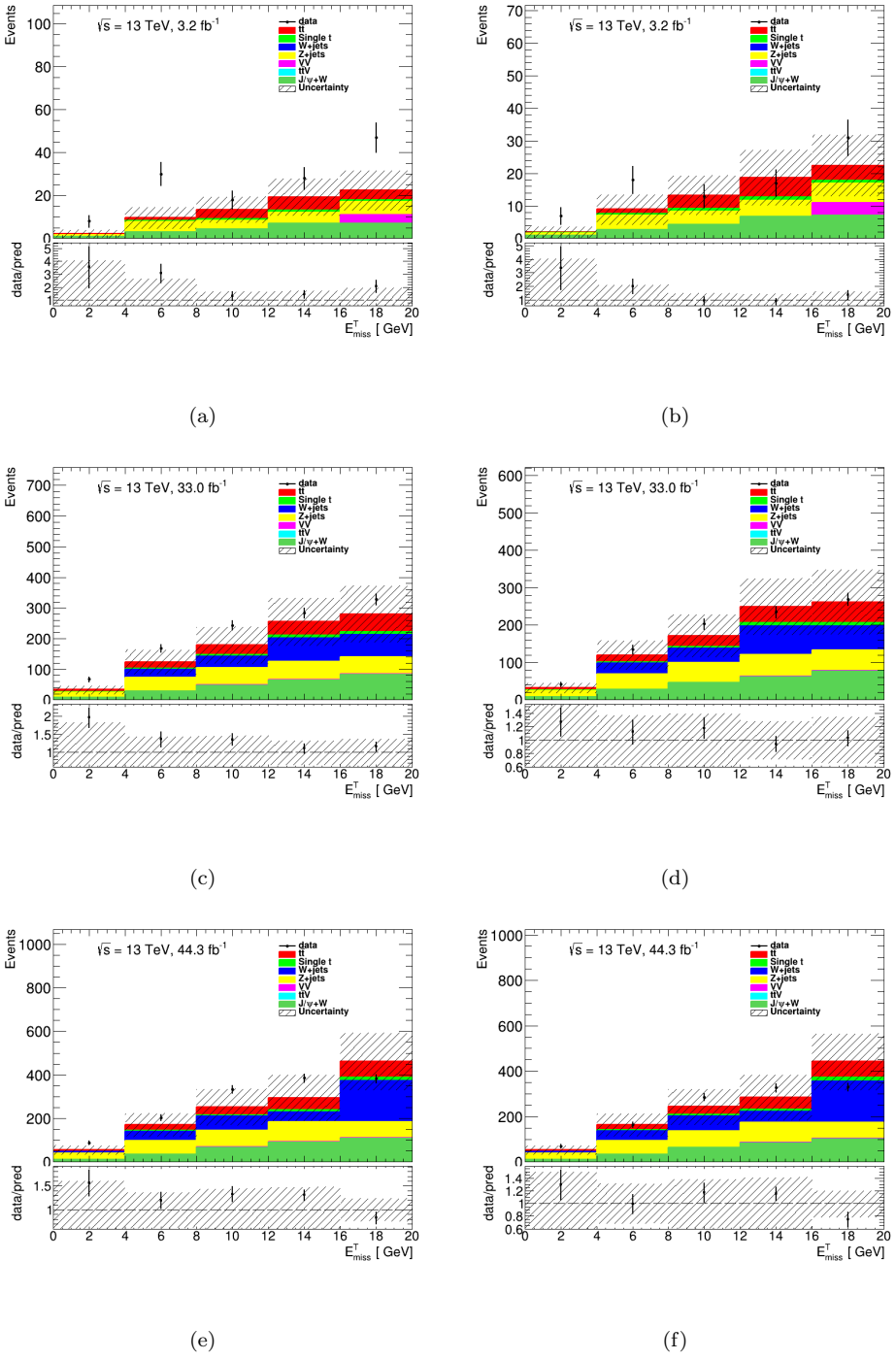


FIGURE A.2: The upper panel shows the distribution of E_T^{miss} for data and the different SM processes for events passing the loose (left) and tight (right) selection criteria in the muon channel for 2015 (a, b), 2016 (c, d), and 2017 (e, f), respectively. Only the statistical, cross-section and luminosity uncertainties are shown. The lower panel shows the ratio between the data and the SM predictions.

Figures A.3-A.8 show the lepton p_T and $|\eta|$ distributions for the data and the SM prediction for leptons passing the loose (left) and tight (right) selection criteria for 2015, 2016 and 2017, respectively. The fake rate is determined by the difference between the number of tight leptons

in the data subtracted by the number of tight leptons in the SM prediction, divided by the same differences in the loose lepton case. Since the tight lepton contribution represents the numerator and has "tighter" criteria, the difference in the data and SM prediction in the tight case is considered when analysing the results.

Across all years in the electron channel, the presence of NPFL is clearly visible in the range $p_T < 35$ GeV and slightly visible between $35 < p_T < 40$ GeV, showing the presence of NPFL in that region. The data and SM prediction agree within uncertainty from $p_T > 40$ GeV and thus showing no presence of NPFL. The $|\eta|$ distribution shows the NPFL background is clearly visible in the range $|\eta| < 0.5$ and $1.5 < |\eta| < 2.5$ in 2015, $1.5 < |\eta| < 2.5$ in 2016, and $1 < |\eta| < 1.5$ and $2 < |\eta| < 2.5$ in 2017, and no NPFL in the other ranges. Across all years in the muon channel, the presence of NPFL is only visible in the range $p_T < 30$ GeV. The data and SM prediction agree within uncertainty from $p_T > 30$ GeV, as well as the entire $|\eta|$ distribution, which shows no presence of NPFL in those regions.

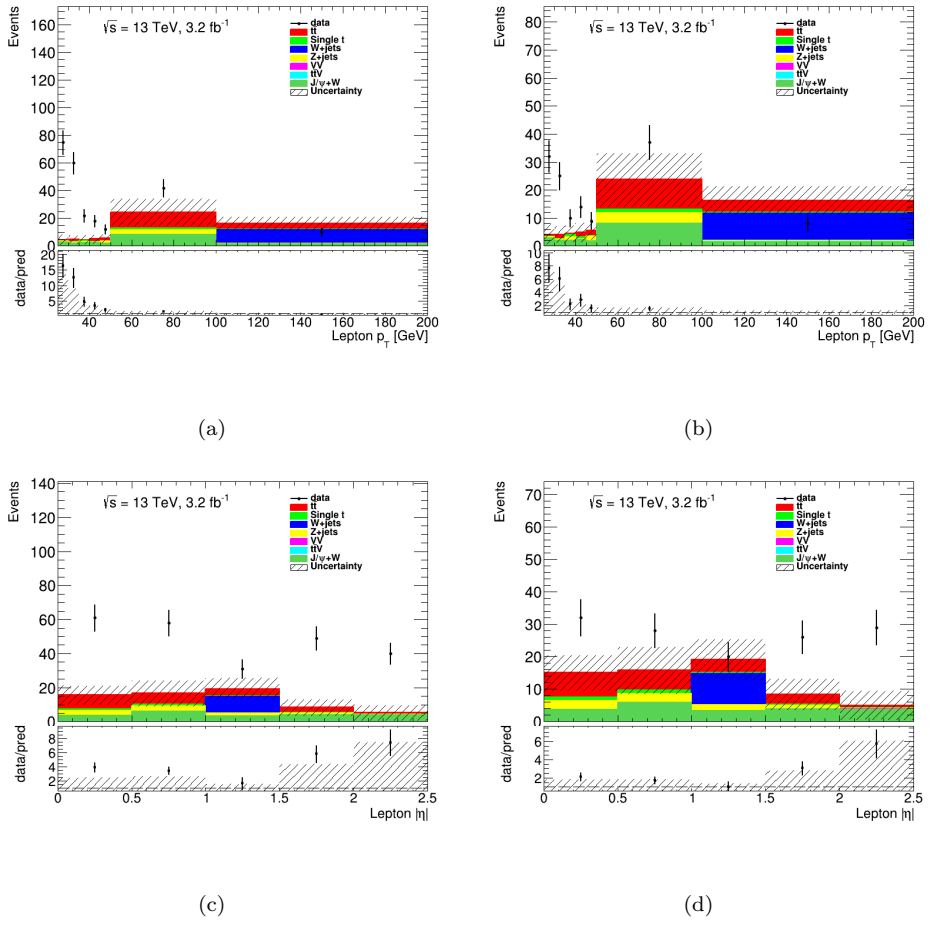
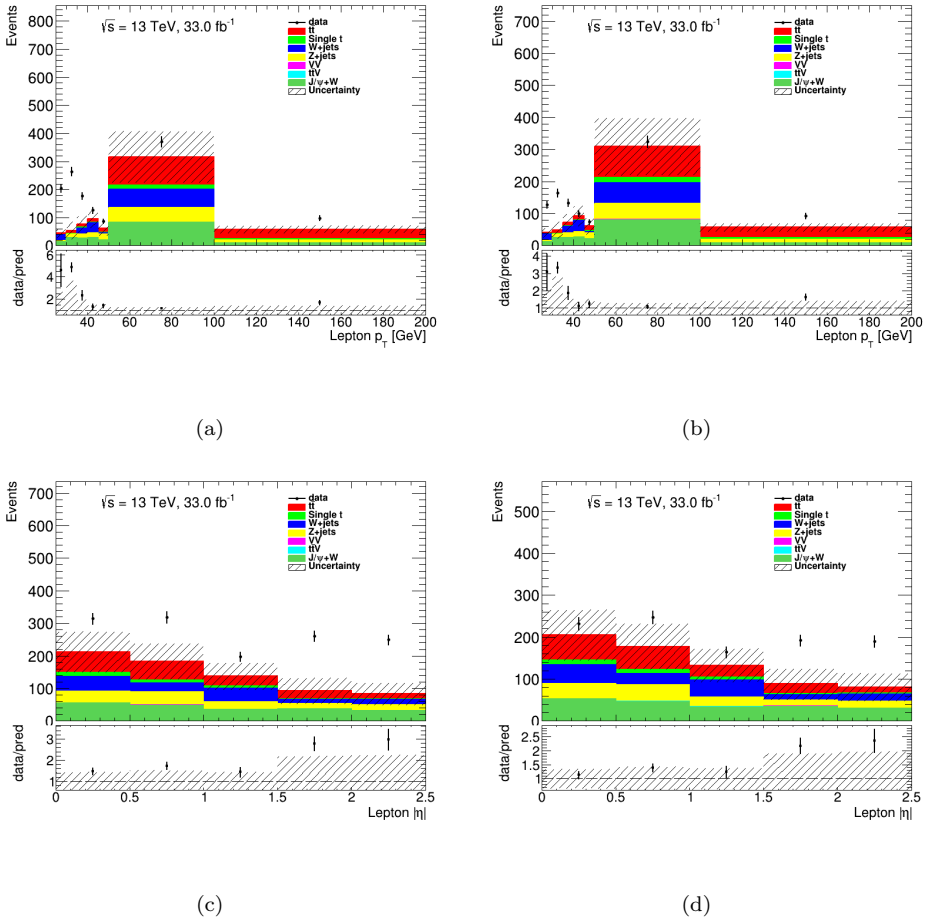


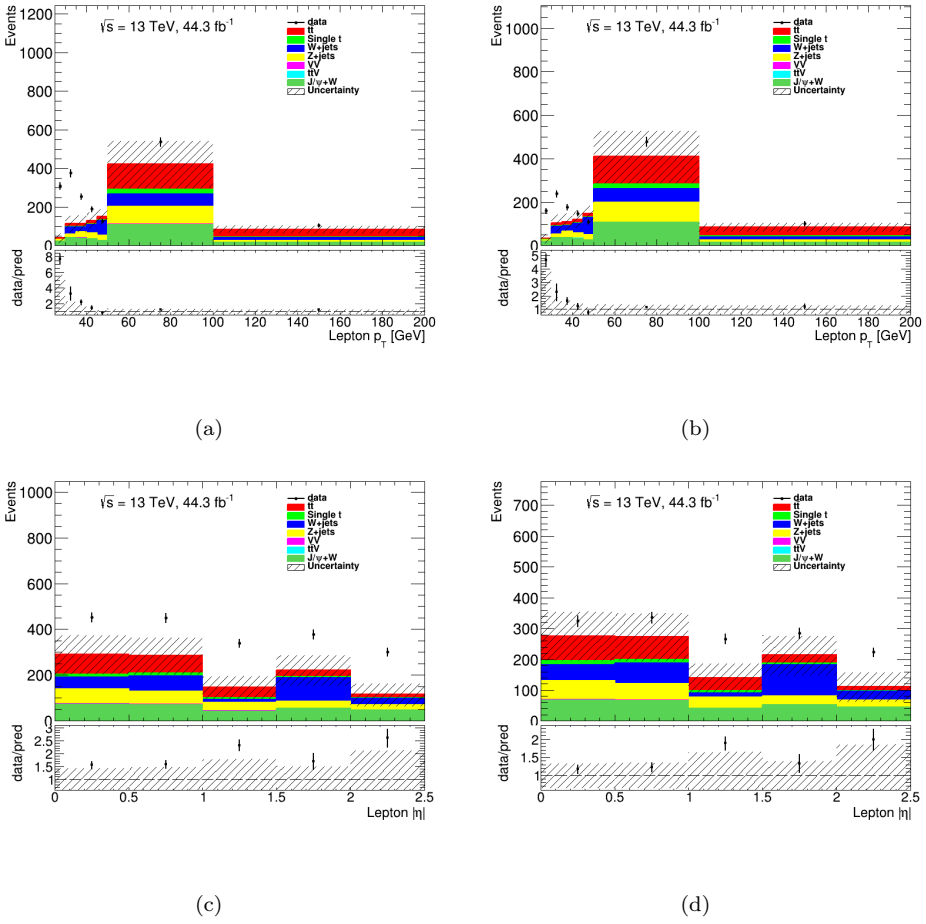
FIGURE A.3: The upper panels show the distributions of p_T (a, b) and $|\eta|$ (c, d) for data and the different SM processes for events passing the loose (left) and tight (right) selection criteria in the electron channel for 2015. Only the statistical, cross-section and luminosity uncertainties are shown. The lower panel shows the ratio between the data and the SM predictions.



(a) (b)

(c) (d)

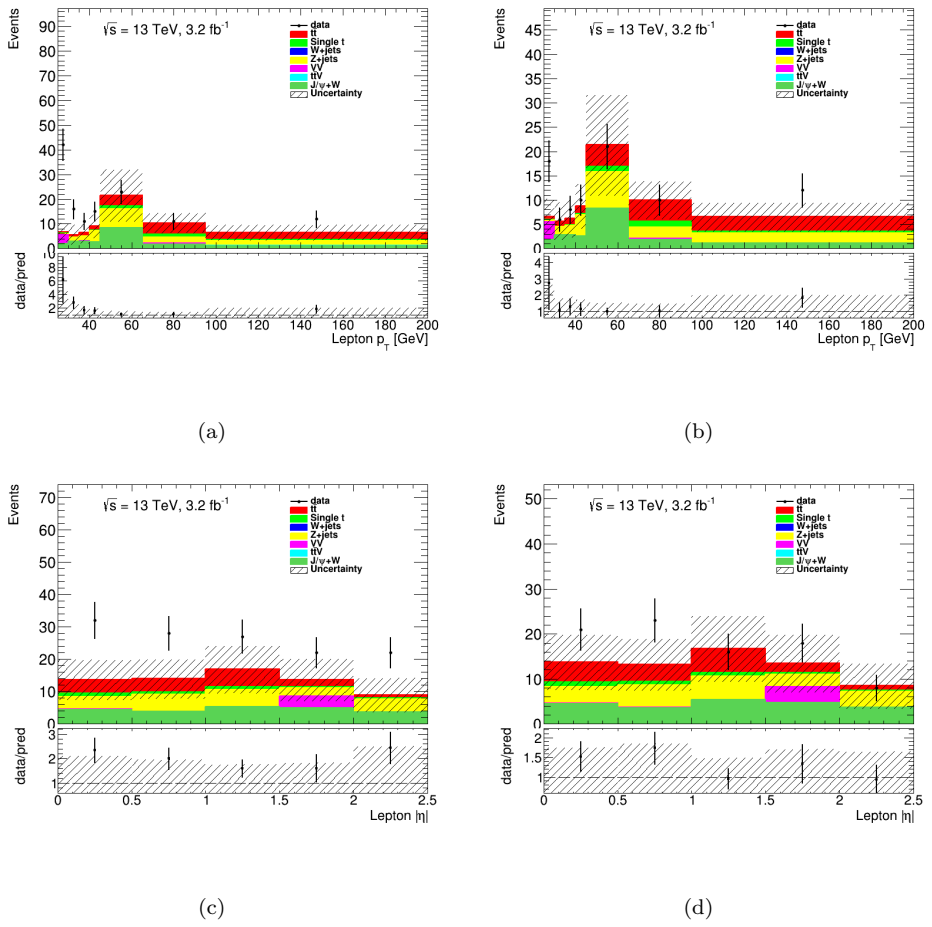
FIGURE A.4: The upper panels show the distributions of p_T (a, b) and $|\eta|$ (c, d) for data and the different SM processes for events passing the loose (left) and tight (right) selection criteria in the electron channel for 2016. Only the statistical, cross-section and luminosity uncertainties are shown. The lower panel shows the ratio between the data and the SM predictions.



(a) (b)

(c) (d)

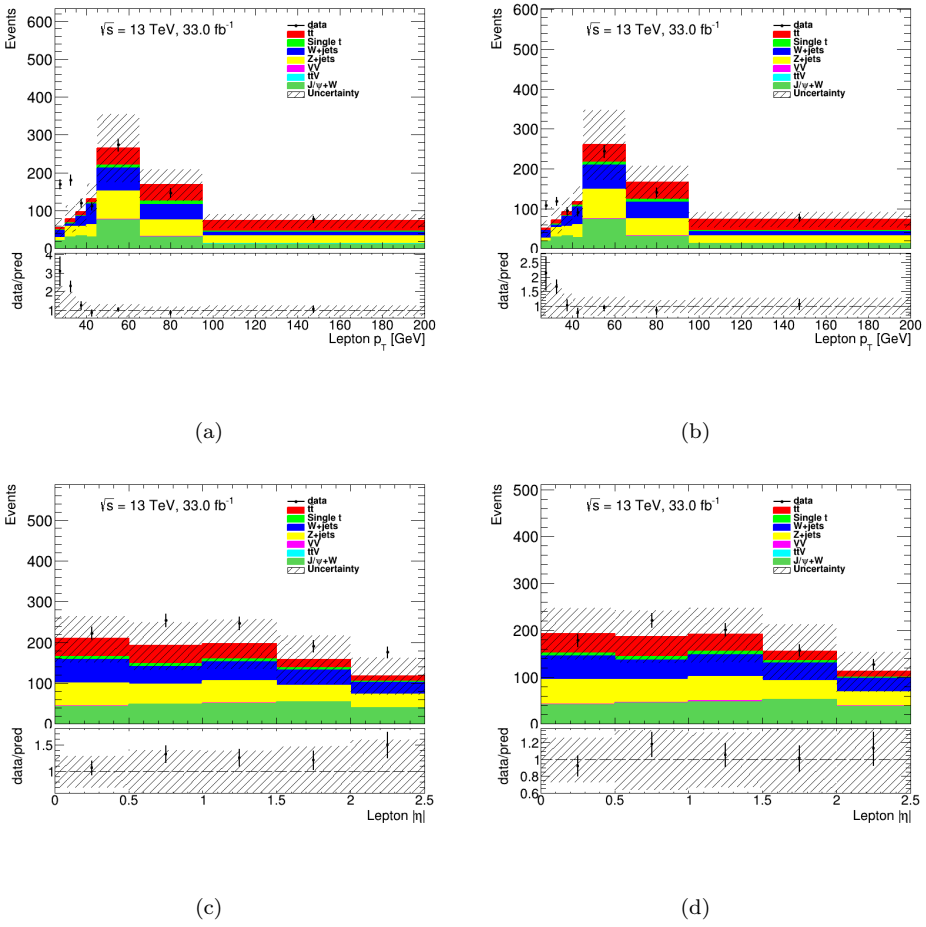
FIGURE A.5: The upper panels show the distributions of p_T (a, b) and $|\eta|$ (c, d) for data and the different SM processes for events passing the loose (left) and tight (right) selection criteria in the electron channel for 2017. Only the statistical, cross-section and luminosity uncertainties are shown. The lower panel shows the ratio between the data and the SM predictions.



(a) (b)

(c) (d)

FIGURE A.6: The upper panels show the distributions of p_T (a, b) and $|\eta|$ (c, d) for data and the different SM processes for events passing the loose (left) and tight (right) selection criteria in the muon channel for 2015. Only the statistical, cross-section and luminosity uncertainties are shown. The lower panel shows the ratio between the data and the SM predictions.



(a) (b)

(c) (d)

FIGURE A.7: The upper panels show the distributions of p_T (a, b) and $|\eta|$ (c, d) for data and the different SM processes for events passing the loose (left) and tight (right) selection criteria in the muon channel for 2016. Only the statistical, cross-section and luminosity uncertainties are shown. The lower panel shows the ratio between the data and the SM predictions.

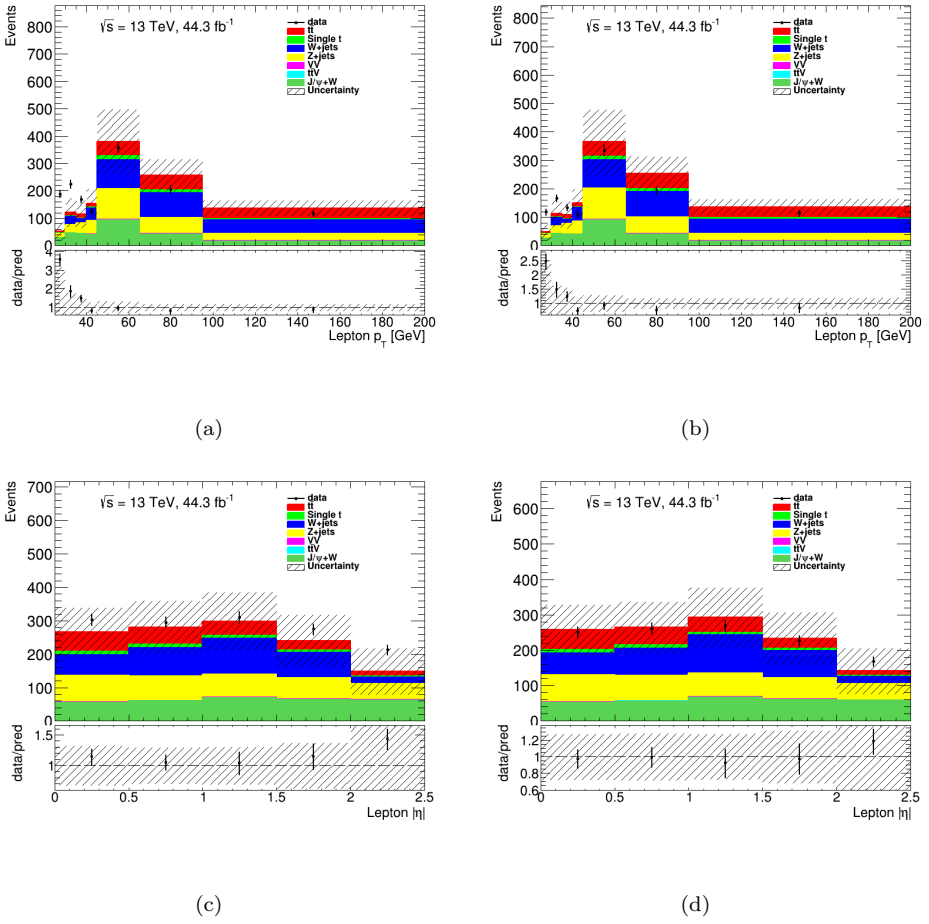


FIGURE A.8: The upper panels show the distributions of p_T (a, b) and $|\eta|$ (c, d) for data and the different SM processes for events passing the loose (left) and tight (right) selection criteria in the muon channel for 2017. Only the statistical, cross-section and luminosity uncertainties are shown. The lower panel shows the ratio between the data and the SM predictions.

A.1.1 Measurement of the efficiencies

The efficiencies, ϵ_{real}, are measured in a control region dominated by real leptons. The $t\bar{t}$ SM process is used to define the control region, where both W-bosons from the two different top quarks must have decayed leptonically. These events must satisfy the following selection criteria:

- At least one jet.
- Exactly two loose electrons (muons) with opposite signs, where both leptons must fire a lepton trigger.

The control region is defined for real leptons, and the efficiency is determined across the same object kinematics as the fake rates. The efficiencies are estimated using Equation A.2.

$$\epsilon_{\text{real}} = \left(\frac{\mathrm{N}^{\mathrm{T}}}{\mathrm{N}^{\mathrm{L}}}\right). \tag{A.2}$$

The results are split per year and per lepton to account for the different pile-up distributions and triggers, respectively. Table A.2 is the same table shown in the main body of the thesis, but is shown again for convenience. The yields are low for loose and tight leptons in both channels for every year, especially in 2015. Therefore, this control region is statistically limited across all years.

Channel	2015		2016		2017		2018		Total	
	Tight	Loose	Tight	Loose	Tight	Loose	Tight	Loose	Tight	Loose
Electron	16	16	132	136	175	184	255	267	578	603
Muon	10	10	107	110	140	144	190	197	447	461

TABLE A.2: The number of $t\bar{t}$ events in each channel split per year when the leptons pass either the loose or tight selection criteria.

The yields show negligible difference between loose and tight leptons in both channels and across all years. This indicates that loose and tight leptons are the same and the efficiencies should equal one.

Similarly to the fake rate measurements, the results are split per year and per lepton, and shown for 2015, 2016 and 2017, respectively. The results for 2018 can be found in the main body of the thesis in Section 4.4.5. Figures A.9 and A.10 show the lepton p_T and $|\eta|$ distributions for leptons passing the loose (black) and tight (red) selection criteria in $t\bar{t}$ events for 2015, 2016 and 2017, respectively. The results show the loose and tight contributions agree within uncertainty, and therefore should indicate the efficiencies are 1. However, since the comparison is made with the same SM process, if the loose contribution is reduced by the uncertainty, so is the tight contribution. This means the central value between the bins will be different regardless of the variation, and therefore the central values will be used to determine the efficiencies in all bins.

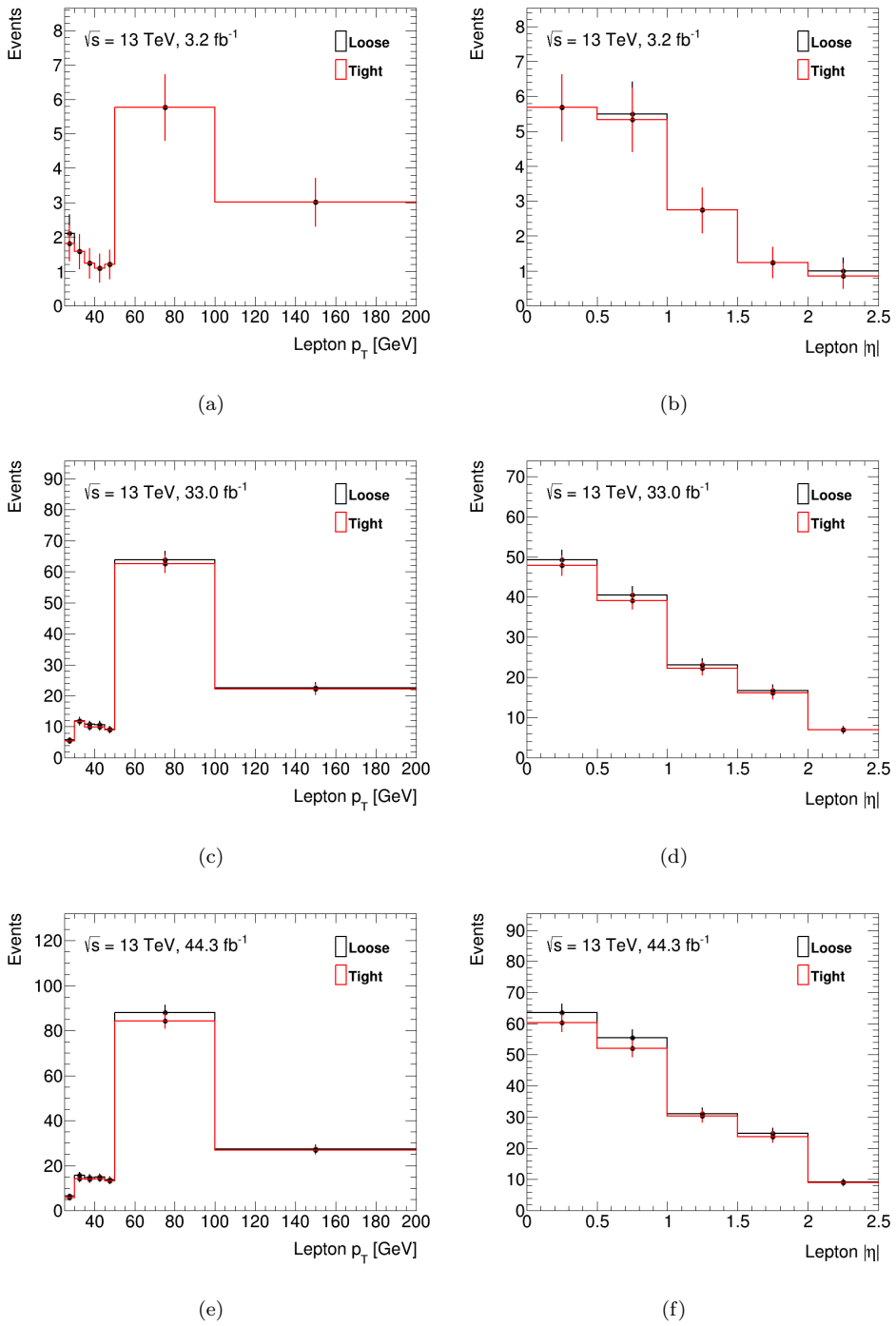


FIGURE A.9: Distributions of p_T (left) and $|\eta|$ (right) for loose (black) and tight (red) leptons in the electron channel for 2015 (a, b), 2016 (c, d), and 2017 (e, f), respectively. The error bars show the statistical uncertainties.

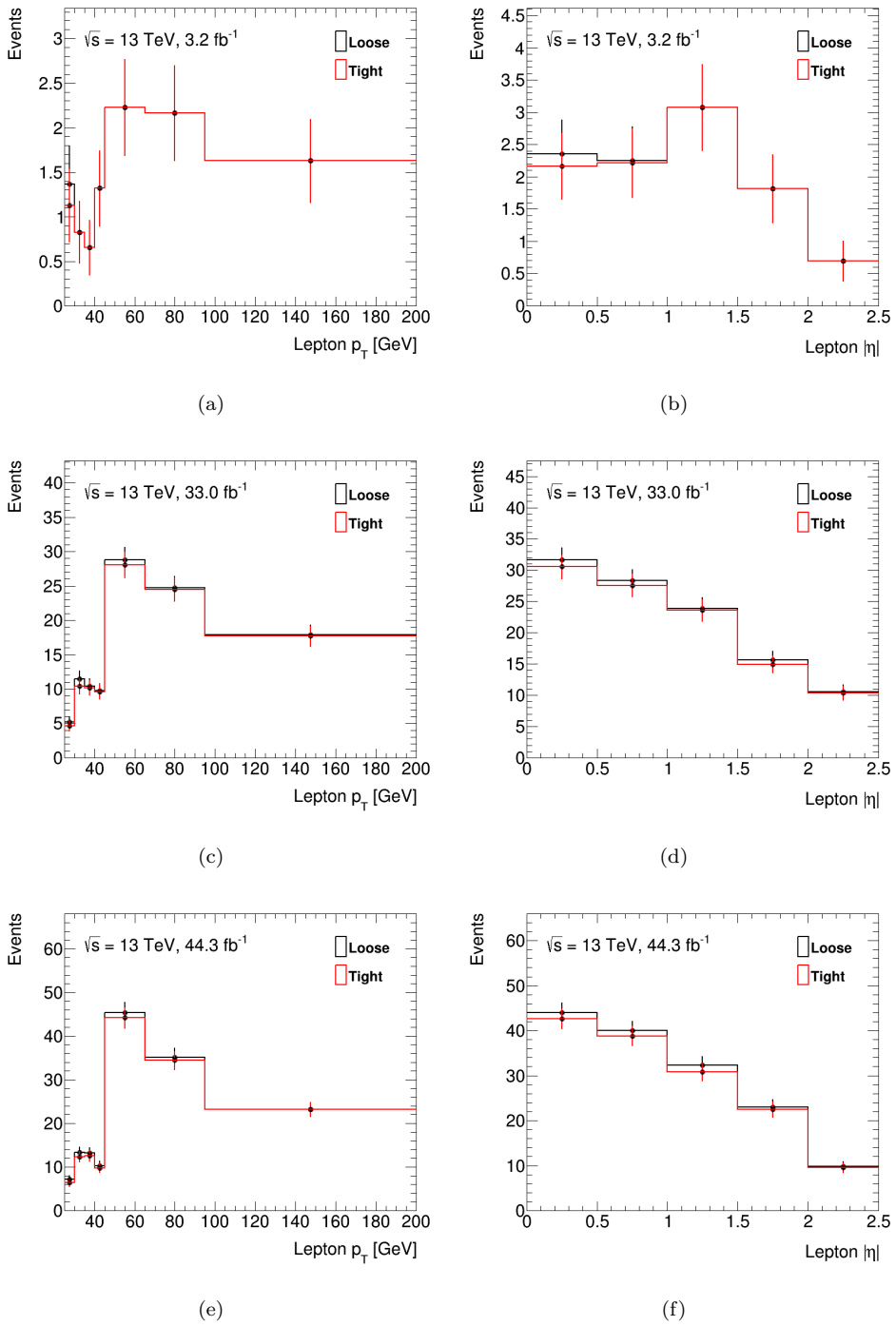


FIGURE A.10: Distributions of p_T (left) and $|\eta|$ (right) for loose (black) and tight (red) leptons in the muon channel for 2015 (a, b), 2016 (c, d), and 2017 (e, f), respectively. The error bars show the statistical uncertainties.

A.2 Efficiencies and fake rates

Figures A.11 and A.13 show the measured efficiencies and fake rates as a function of lepton p_T and $|\eta|$ in both electron and muon channels for 2015, 2016 and 2017, respectively. The results

for 2018 can be found in the main body of the thesis in Section 4.4.6.

The fake rates and efficiencies have large statistical uncertainties due to the limited amount of events passing the selection criteria, which comes from the low branching ratio of having $J/\psi \to \mu\mu$ in every event. The efficiencies increase as a function of lepton p_T and are relatively constant across $|\eta|$ in 2016 and 2017. In 2015, the efficiencies also increase as a function of lepton p_T, but vary across $|\eta|$. The fake rates, on the hand, are relatively flat at low p_T and increase at higher p_T in the electron channel, and decrease at low p_T in the muon channel. Additionally, the fake rates vary across $|\eta|$ in both channels for 2015, 2016 and 2017, respectively. The regions with no fake rate or efficiency are due to the fact that the data is smaller than the SM prediction. When estimating the NPFL contribution in the signal region, only the measured efficiencies and fake rates in regions where sizeable deviations between the data and SM prediction are observed, will be used, while in the other regions, the amount of NPFL is considered to be zero.

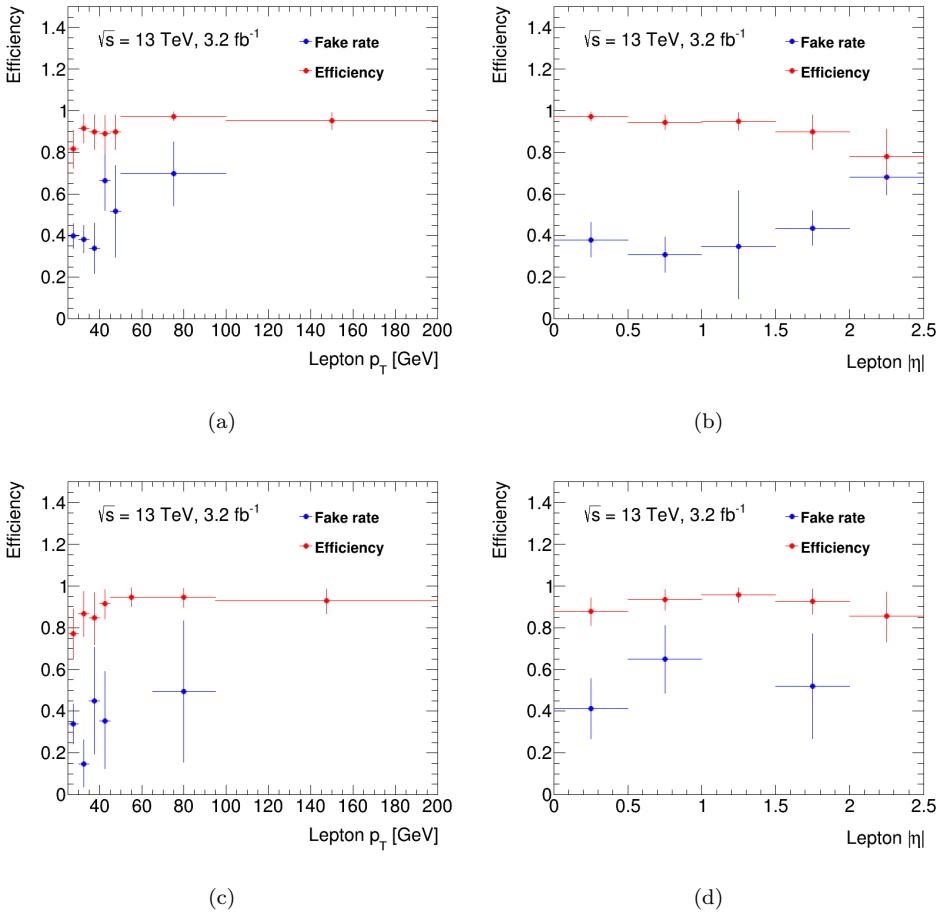


FIGURE A.11: The measured efficiency (red) and fake rate (blue) in the transverse momentum (a, c) and absolute pseudorapidity (b, d) distributions in the electron (top row) and muon (bottom row) channels for 2015. The error bars correspond to the statistical uncertainty from the bin entries. No systematic uncertainties are shown.

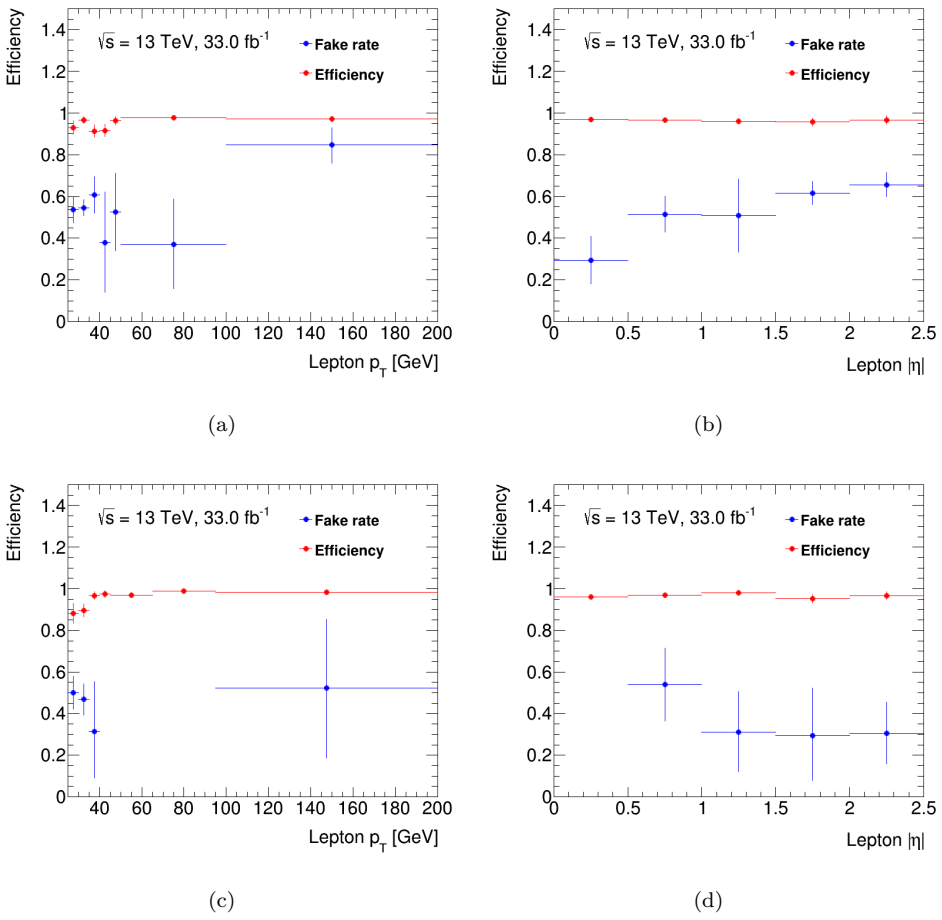
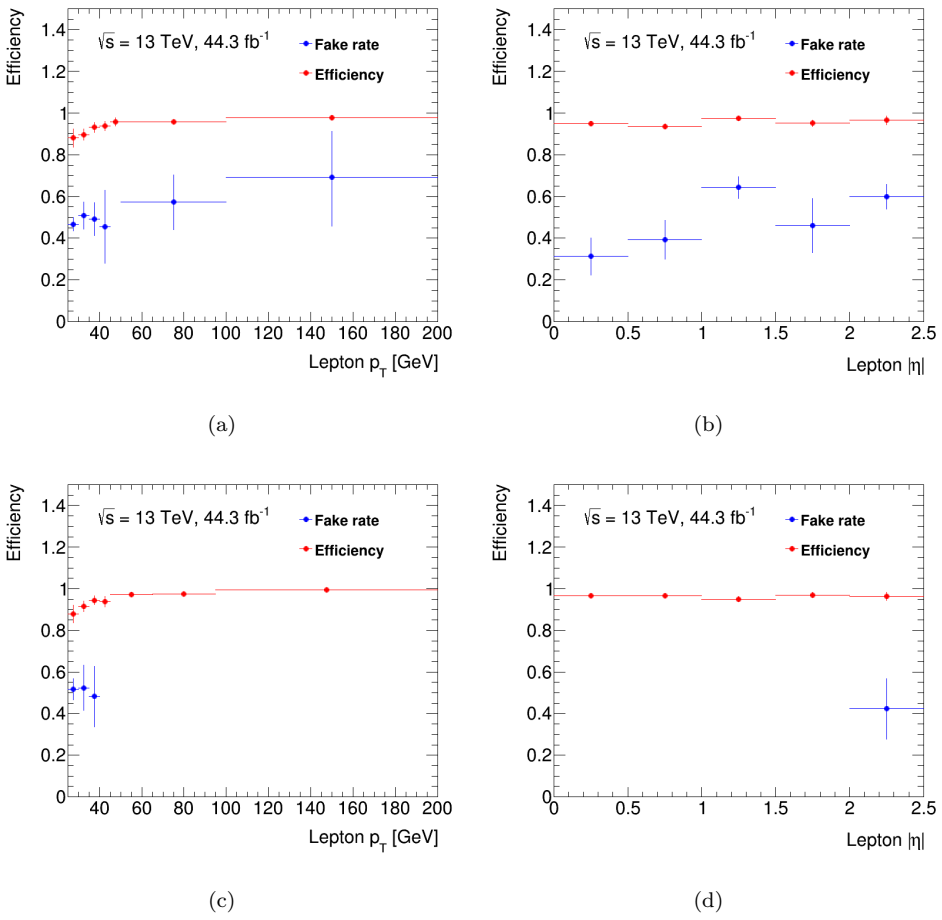


FIGURE A.12: The measured efficiency (red) and fake rate (blue) in the transverse momentum (a, c) and absolute pseudorapidity (b, d) distributions in the electron (top row) and muon (bottom row) channels for 2016. The error bars correspond to the statistical uncertainty from the bin entries. No systematic uncertainties are shown.



(a) (b)

(c) (d)

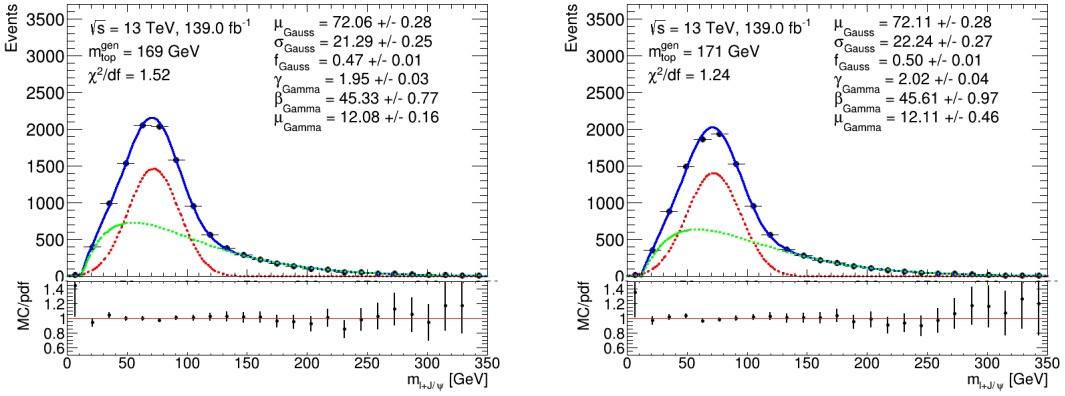
FIGURE A.13: The measured efficiency (red) and fake rate (blue) in the transverse momentum (a, c) and absolute pseudorapidity (b, d) distributions in the electron (top row) and muon (bottom row) channels for 2017. The error bars correspond to the statistical uncertainty from the bin entries. No systematic uncertainties are shown.

Appendix B

Signal templates fits

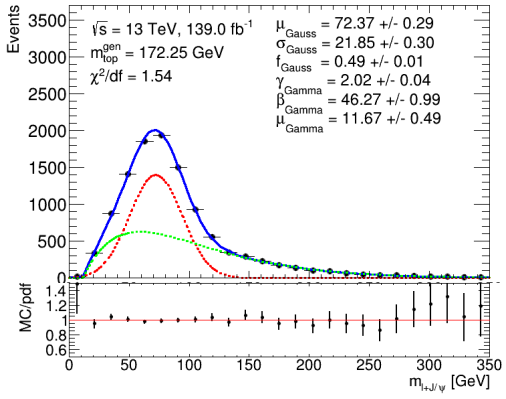
The following chapter shows the signal pdf fit results over the invariant mass of the lepton + J/ψ distribution in the signal region, for the seven different top quark mass samples used to develop the extraction method to obtain the top quark mass in data, that are not shown in the main body of the thesis. The other two top quark mass distributions can be found in Section 5.1.1.

Figures B.1-B.2 shows the invariant mass of the lepton + J/ψ distribution with the pdf fit (blue) and the individual contributions from the correct (red) and incorrect (green) pairing components when the top quark mass is set to 169, 171, 172.25, 172.75, 173, 174 and 176 GeV, as well as the best fit values for the different parameters of the pdf in the upper panel. The lower panel shows the ratio between the MC sample and pdf.

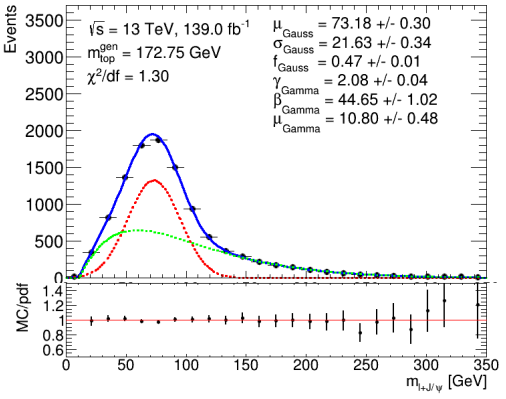


(a) (b)

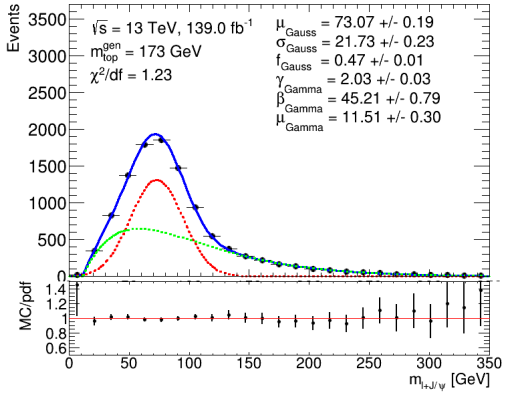
FIGURE B.1: The upper panel shows the invariant mass of the lepton+J/ψ distribution with the total pdf (blue) fitted over the m_{top} = 169 GeV (a) and 171 GeV (b) signal MC samples, respectively. The individual contributions of the Gaussian (red) and Gamma (green) functions and the best fit values are shown. The lower panel shows the ratio between the signal MC samples and pdf.



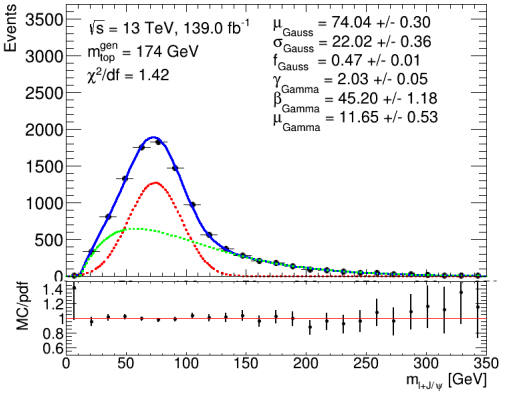
(a)



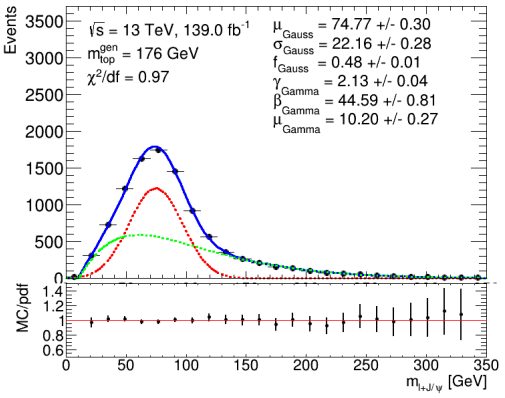
(b)



(c)



(d)



(e)

FIGURE B.2: The upper panel shows the invariant mass of the lepton+J/ψ distribution with the total pdf (blue) fitted over the m_{top} = 172.25 GeV (a), 172.75 (b), 173 (c), 174 (d) and 176 (e) GeV signal MC samples, respectively. The individual contributions of the Gaussian (red) and Gamma (green) functions and the best fit values are shown. The lower panel shows the ratio between the signal MC samples and pdf.

Appendix C

Parameter correlation matrices

The following chapter shows the correlation matrices for the parameters in the signal template pdf for the eight different top quark mass samples used to develop the extraction method to obtain the top quark mass in data. The ninth top quark mass sample's correlation matrix can be found in Section 5.1.2.

Tables C.1-C.8 shows the strength of the correlations between each parameter in the signal template pdf for eight different top quark mass MC samples. The μ_{Gauss} parameter describes the position of the Gaussian peak, whereas the σ_{Gauss} parameter describes the Gaussian's width. The γ_{Gamma} and β_{Gamma} parameters together describe the shape of the Gamma function, while the μ_{Gamma} parameter describes the Gamma function's starting point.

Parameter	μ_{Gauss}	σ_{Gauss}	f_{Gauss}	γ_{Gamma}	β_{Gauss}	μ_{Gamma}
μ_{Gauss}	1	-0.18	-0.20	-0.39	-0.14	0.06
σ_{Gauss}	-0.18	1	0.67	0.00	0.25	0.13
f_{gauss}	-0.20	0.67	1	-0.23	0.57	0.09
γ_{Gamma}	-0.39	0.00	-0.23	1	-0.86	-0.44
β_{Gamma}	0.14	0.25	0.57	-0.86	1	0.30
μ_{Gamma}	0.06	0.13	0.09	-0.44	0.30	1

TABLE C.1: Correlation matrix of each parameter in the pdf describing the invariant mass of the three lepton system fitted over the $m_{top} = 169$ GeV MC sample.

Parameter	μ_{Gauss}	σ_{Gauss}	f_{Gauss}	γ_{Gamma}	β_{Gauss}	μ_{Gamma}
μ_{Gauss}	1	-0.15	-0.17	-0.30	0.13	-0.08
σ_{Gauss}	-0.15	1	0.72	-0.29	0.44	0.53
f_{gauss}	-0.17	0.72	1	-0.40	0.65	0.45
γ_{Gamma}	-0.30	-0.29	-0.40	1	-0.90	-0.68
β_{Gamma}	0.13	0.44	0.65	-0.90	1	0.59
μ_{Gamma}	-0.08	0.53	0.45	-0.68	0.59	1

TABLE C.2: Correlation matrix of each parameter in the pdf describing the invariant mass of the three lepton system fitted over the m_{top} = 171 GeV MC sample.

Parameter	μ_{Gauss}	σ_{Gauss}	f_{Gauss}	γ_{Gamma}	β_{Gauss}	μ_{Gamma}
μ_{Gauss}	1	-0.17	-0.17	-0.26	0.10	-0.01
σ_{Gauss}	-0.17	1	0.80	-0.45	0.60	0.58
f_{gauss}	-0.17	0.80	1	-0.57	0.77	0.57
γ_{Gamma}	-0.26	-0.45	-0.57	1	-0.93	-0.80
β_{Gamma}	0.10	0.60	0.77	-0.93	1	0.74
μ_{Gamma}	-0.01	0.58	0.57	-0.80	0.74	1

TABLE C.3: Correlation matrix of each parameter in the pdf describing the invariant mass of the three lepton system fitted over the m_{top} = 172 GeV MC sample.

Parameter	μ_{Gauss}	σ_{Gauss}	f_{Gauss}	γ_{Gamma}	β_{Gauss}	μ_{Gamma}
μ_{Gauss}	1	-0.23	-0.23	-0.27	0.09	-0.14
σ_{Gauss}	-0.23	1	0.75	-0.29	0.45	0.58
f_{gauss}	-0.23	0.75	1	-0.40	0.65	0.48
γ_{Gamma}	-0.27	-0.29	-0.40	1	-0.90	-0.68
β_{Gamma}	0.10	0.45	0.65	-0.90	1	0.60
μ_{Gamma}	-0.14	0.58	0.48	-0.68	0.60	1

TABLE C.4: Correlation matrix of each parameter in the pdf describing the invariant mass of the three lepton system fitted over the m_{top} = 172.25 GeV MC sample.

Parameter	μ_{Gauss}	σ_{Gauss}	f_{Gauss}	γ_{Gamma}	β_{Gauss}	μ_{Gamma}
μ_{Gauss}	1	-0.29	-0.27	-0.18	0.01	-0.15
σ_{Gauss}	-0.29	1	0.81	-0.40	0.56	0.62
f_{gauss}	-0.27	0.81	1	-0.51	0.73	0.56
γ_{Gamma}	-0.18	-0.40	-0.51	1	-0.92	-0.74
β_{Gamma}	0.01	0.56	0.73	-0.92	1	0.68
μ_{Gamma}	-0.15	0.62	0.56	-0.74	0.68	1

TABLE C.5: Correlation matrix of each parameter in the pdf describing the invariant mass of the three lepton system fitted over the m_{top} = 172.75 GeV MC sample.

Parameter	μ_{Gauss}	σ_{Gauss}	f_{Gauss}	γ_{Gamma}	β_{Gauss}	μ_{Gamma}
μ_{Gauss}	1	-0.26	-0.25	-0.18	0.01	-0.12
σ_{Gauss}	-0.26	1	0.83	-0.48	0.62	0.66
f_{gauss}	-0.25	0.83	1	-0.61	0.78	0.64
γ_{Gamma}	-0.18	-0.48	-0.61	1	-0.94	-0.78
β_{Gamma}	0.01	0.62	0.78	-0.94	1	0.74
μ_{Gamma}	-0.12	0.66	0.64	-0.78	0.74	1

TABLE C.6: Correlation matrix of each parameter in the pdf describing the invariant mass of the three lepton system fitted over the m_{top} = 173 GeV MC sample.

Parameter	μ_{Gauss}	σ_{Gauss}	f_{Gauss}	γ_{Gamma}	β_{Gauss}	μ_{Gamma}
μ_{Gauss}	1	-0.26	-0.26	-0.11	-0.04	-0.17
σ_{Gauss}	-0.26	1	0.83	-0.54	0.64	0.70
f_{gauss}	-0.26	0.83	1	-0.65	0.80	0.67
γ_{Gamma}	-0.11	-0.54	-0.65	1	-0.94	-0.81
β_{Gamma}	-0.04	0.64	0.80	-0.94	1	0.76
μ_{Gamma}	-0.17	0.70	0.67	-0.81	0.76	1

TABLE C.7: Correlation matrix of each parameter in the pdf describing the invariant mass of the three lepton system fitted over the m_{top} = 174 GeV MC sample.

Parameter	μ_{Gauss}	σ_{Gauss}	f_{Gauss}	γ_{Gamma}	β_{Gauss}	μ_{Gamma}
μ_{Gauss}	1	-0.27	-0.26	-0.42	0.16	0.09
σ_{Gauss}	-0.27	1	0.72	0.02	0.26	0.16
f_{gauss}	-0.26	0.72	1	-0.18	0.54	0.11
γ_{Gamma}	-0.42	0.02	-0.18	1	-0.86	-0.58
β_{Gamma}	0.16	0.26	0.54	-0.86	1	0.42
μ_{Gamma}	0.09	0.16	0.11	-0.58	0.42	1

TABLE C.8: Correlation matrix of each parameter in the pdf describing the invariant mass of the three lepton system fitted over the m_{top} = 176 GeV MC sample.

Appendix D

Parameter relationship with the top quark mass

The following chapter shows the signal template's parameter relationship with the top quark mass after removing each top quark mass sample's information from the fits. This is done for each top quark mass individually.

Table D.1 shows the slope and intercept values between the fraction of Gaussian and the Gaussian mean and width parameters with the top quark mass, after removing one top quark mass sample's information from the fits. The relationship between each parameter and the top quark mass is of the form parameter = slope(m_{top}-172.5) + intercept. The fraction of Gaussian has slopes consistent with 0 within uncertainty, and therefore is independent of the top quark mass. Additionally, the slopes and intercepts are consistent with each other within uncertainty after removing each top quark mass sample's information individually. This shows the stability of the independence of the fraction of Gaussian and the top quark mass. The Gaussian width has slopes not consistent with 0 within uncertainty, and shows dependence of the top quark mass. Additionally, the slopes and intercepts are consistent with each other within uncertainty after removing each top quark mass sample's information individually. This shows the stability of the dependence of the Gaussian width and the top quark mass. The Gaussian mean has slopes not consistent with 0 within uncertainty, and the size of the slope is relatively large, which shows large dependence with the top quark mass. Additionally, the intercepts are consistent with each other within uncertainty after removing each top quark mass sample's information individually. However, the slopes are only consistent with each other within uncertainty for top quark masses between 171-174 GeV. The slopes, after removing the 169 GeV and 176 GeV top quark mass sample's information from the fits, show consistency with other slopes within 2σ uncertainty. The method uncertainty systematic includes variations up to 2σ, and therefore accounts for these possible fluctuations.

Removed m_{top} information from parameter fits [GeV]	μ_{Gauss}		σ_{Gauss}		f_{Gauss}	
	slope	intercept	slope	intercept	slope	intercept
169	0.67 ± 0.06	72.76 ± 0.09	0.07 ± 0.05	21.86 ± 0.09	0.00 ± 0.00	0.48 ± 0.01
171	0.52 ± 0.05	72.95 ± 0.08	0.09 ± 0.04	21.78 ± 0.09	0.00 ± 0.00	0.48 ± 0.01
172	0.51 ± 0.05	72.89 ± 0.09	0.10 ± 0.04	21.79 ± 0.09	0.00 ± 0.00	0.48 ± 0.01
172.25	0.51 ± 0.05	72.91 ± 0.08	0.10 ± 0.04	21.79 ± 0.09	0.00 ± 0.00	0.48 ± 0.01
172.5	0.51 ± 0.05	72.94 ± 0.09	0.08 ± 0.04	21.86 ± 0.09	0.00 ± 0.00	0.48 ± 0.00
172.75	0.52 ± 0.05	72.86 ± 0.08	0.07 ± 0.04	21.91 ± 0.08	0.00 ± 0.00	0.48 ± 0.00
173	0.51 ± 0.05	72.88 ± 0.09	0.06 ± 0.04	21.96 ± 0.09	0.00 ± 0.00	0.48 ± 0.00
174	0.47 ± 0.05	72.81 ± 0.08	0.08 ± 0.04	21.97 ± 0.08	0.00 ± 0.00	0.48 ± 0.00
176	0.35 ± 0.07	72.78 ± 0.09	0.13 ± 0.06	22.00 ± 0.09	0.00 ± 0.00	0.48 ± 0.00

TABLE D.1: The slopes and intercepts of the relationships between the fraction of Gaussian and Gaussian's mean and width parameters with the top quark mass, after removing each top quark mass sample's information from the fits.

Table D.2 shows the slope and intercept values between the Gamma's parameters and the top quark mass after removing one top quark mass sample's information from the fits. The relationship between each parameter and the top quark mass is of the form parameter = slope(m_{top}-172.5) + intercept. The Gamma's parameters have slopes consistent with 0 within uncertainty and therefore, is independent of the top quark mass. Additionally, these slopes and intercepts are consistent with each other within uncertainty after removing each top quark mass sample's information individually. This shows the stability of the independence of the Gamma's parameters and the top quark mass.

Removed m_{top} information from parameter fits [GeV]	γ_{Gamma}		β_{Gamma}		μ_{Gamma}	
	slope	intercept	slope	intercept	slope	intercept
169	0.00 ± 0.00	2.06 ± 0.04	0.00 ± 0.01	44.57 ± 0.86	0.00 ± 0.00	11.66 ± 0.40
171	0.00 ± 0.00	1.99 ± 0.03	0.00 ± 0.01	44.96 ± 0.74	0.01 ± 0.00	12.02 ± 0.16
172	0.00 ± 0.00	2.00 ± 0.02	0.00 ± 0.00	44.99 ± 0.56	0.00 ± 0.00	11.95 ± 0.15
172.25	0.00 ± 0.00	2.01 ± 0.02	0.00 ± 0.00	45.43 ± 0.42	0.00 ± 0.00	11.74 ± 0.13
172.5	0.00 ± 0.00	2.00 ± 0.02	0.00 ± 0.00	45.66 ± 0.39	0.00 ± 0.00	11.73 ± 0.13
172.75	0.00 ± 0.00	2.01 ± 0.01	0.00 ± 0.00	45.54 ± 0.33	0.00 ± 0.00	11.65 ± 0.12
173	0.00 ± 0.00	2.01 ± 0.01	0.00 ± 0.00	45.66 ± 0.32	0.00 ± 0.00	11.64 ± 0.11
174	0.00 ± 0.00	2.02 ± 0.01	0.01 ± 0.01	45.66 ± 0.29	0.01 ± 0.00	11.66 ± 0.11
176	0.01 ± 0.01	2.02 ± 0.01	0.15 ± 0.18	45.68 ± 0.29	0.00 ± 0.01	11.69 ± 0.11

TABLE D.2: The slopes and intercepts of the relationships between the Gamma's parameters and the top quark mass, after removing each top quark mass sample's information from the fits.

Appendix E

Extraction method calibration

The following chapter shows the individual fits in order to understand the relationship between the reconstructed and generated top quark masses in an unbiased manner. One top quark mass point is removed from the fit (also called the testing sample), and the parameter relationships are re-determined without the testing sample's information. Appendix D shows the parameter relationships with the top quark mass after removing the information from each top quark mass individually. Every parameter dependency with the top quark mass is consistent no matter what top quark mass is used as the testing sample.

The resulting pdf's are fit over the full model, with the testing sample used as the signal top quark contribution, to determine the reconstructed top quark mass. Figures E.1 and E.2 show the total pdf fit (blue) over the full model for all nine different top quark masses. The Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed and generated top quark masses can be found in the top left corner of the upper panel, while the ratio between the total pdf and the model is shown in the lower panel.

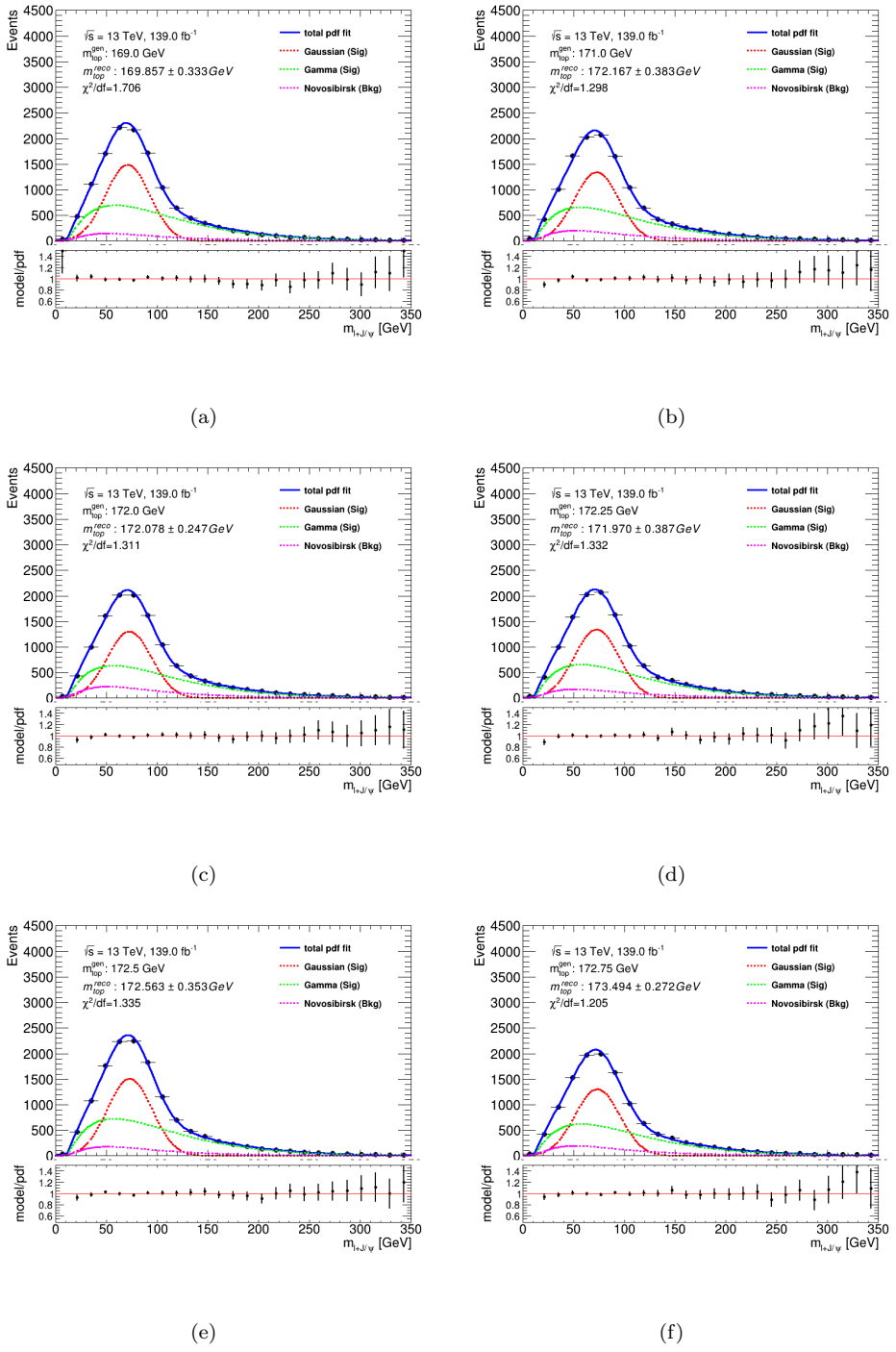


FIGURE E.1: The upper panel shows the invariant mass of the lepton + J/ψ distribution of the full model when using m_{top} = 169 (a), 171 (b), 172 (c), 172.25 (d), 172.5 (e), and 172.75 (f) GeV as the signal top quark processes. The total (blue), Gaussian (red), Gamma (green) and Novosibirsk (purple) pdf fits are shown. The generated and reconstructed top quark masses are shown in the top left corner, together with the fit quality. The lower panel shows the ratio between the full model and the pdf.

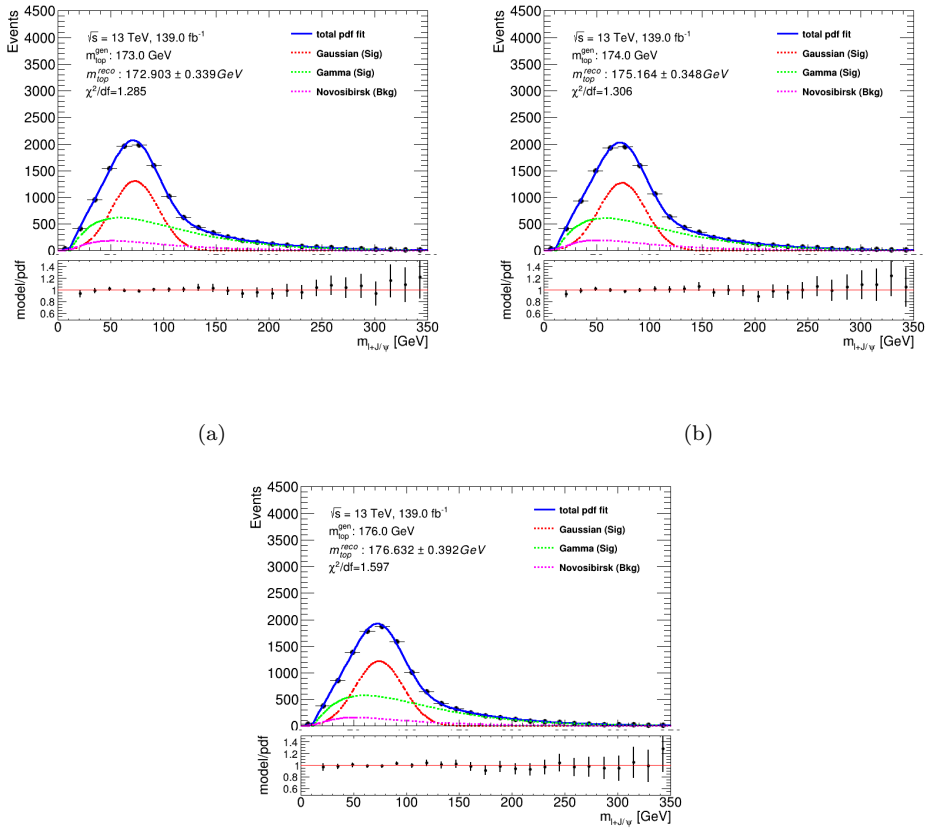


FIGURE E.2: The upper panel shows the invariant mass of the lepton + J/ψ distribution of the full model when using m_{top} = 173 (a), 174 (b), and 176 (c) GeV as the signal top quark processes. The total (blue), Gaussian (red), Gamma (green) and Novosibirsk (purple) pdf fits are shown. The generated and reconstructed top quark masses are shown in the top left corner, together with the fit quality. The lower panel shows the ratio between the full model and the pdf.

Appendix F

Pull study

The following chapter shows the pull distributions for the seven different m_{top} values not shown in the main body of the thesis in Section 5.2.

Figures F.1-F.2 show the pull distributions for m_{top} = 169, 171, 172.25, 172.75, 173, 174 and 176 GeV, fitted with a Gaussian where the pull mean and width, and fit quality can be seen in the top left corner.

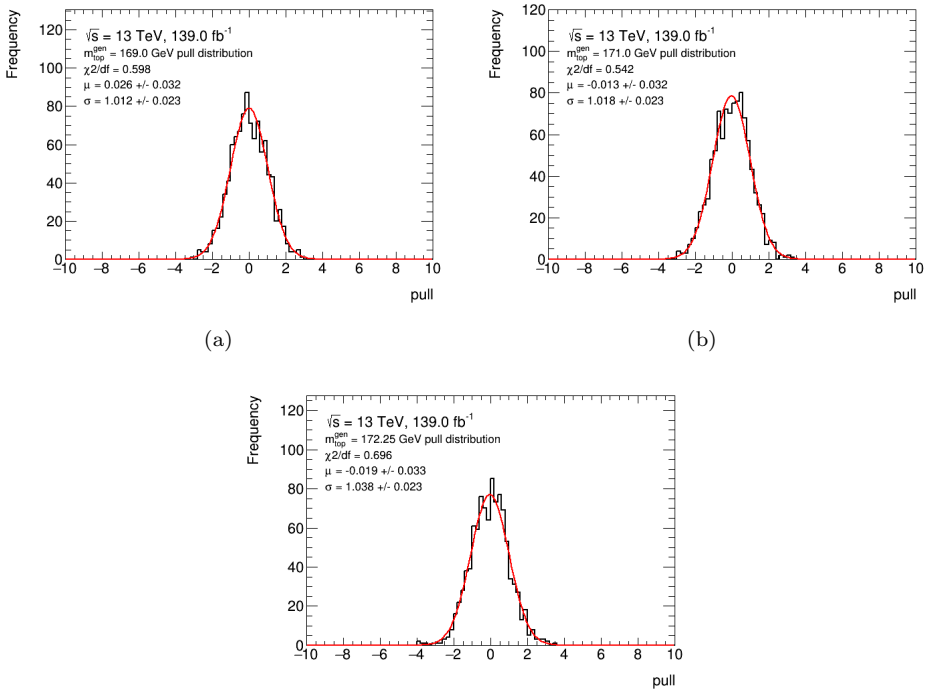


FIGURE F.1: Pull distributions from 1000 pseudo-data sets with the Gaussian fit when m_{top} = 169 (a), 171 (b), and 172.25 (c) GeV. The pull mean and width is shown in the top left corner.

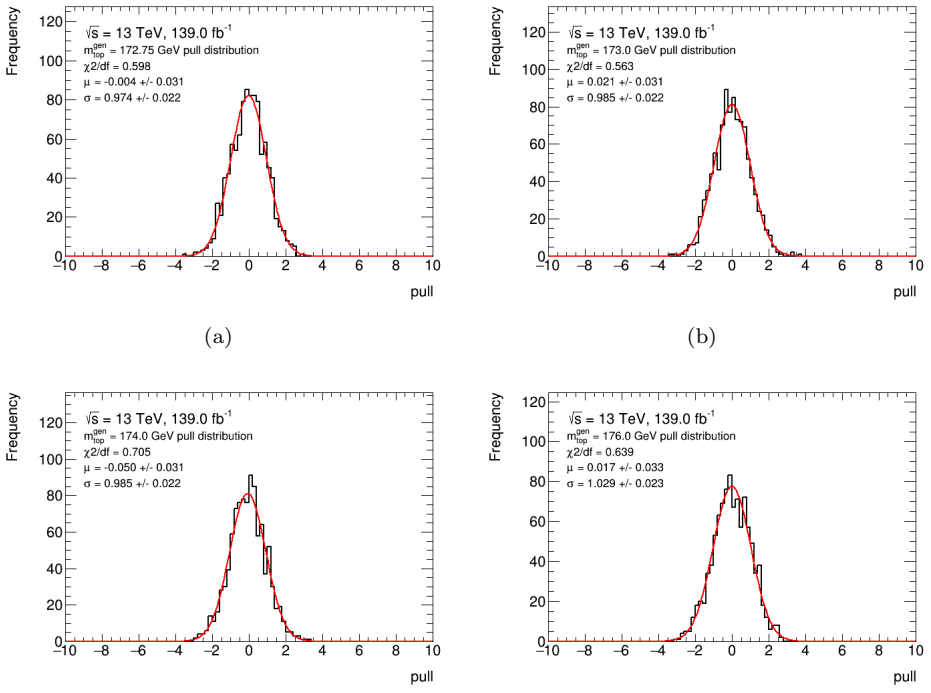


FIGURE F.2: Pull distributions from 1000 pseudo-data sets with the Gaussian fit when m_{top} = 172.75 (a), 173 (b), 174 and 176 (d) GeV. The pull mean and width is shown in the top left corner.

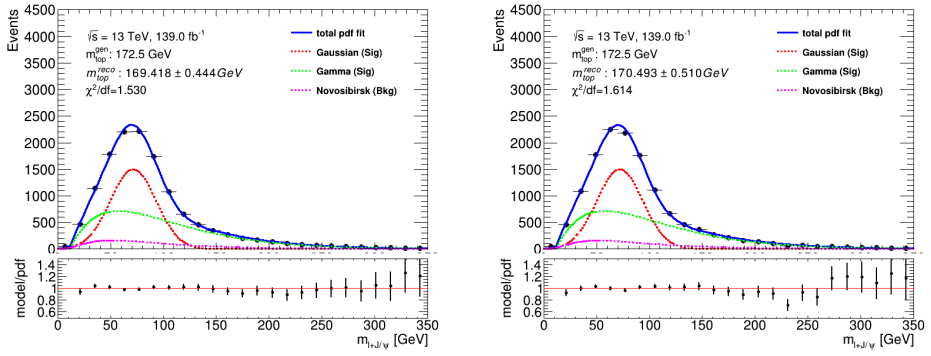
Appendix G

Signal modelling systematic fit results

The following chapter shows the fit results for the different signal modelling systematic uncertainties.

Figures G.1-G.15 shows the fit results for the different signal systematic uncertainties with the reconstructed top quark mass and its statistical uncertainty, as well as the fit quality in the top left corner of the upper panels. The lower panels consist of the ratio between the model and the pdf. The blue shows the total pdf while the red, green and purple shows the Gaussian, Gamma and Novosibirsk contributions, respectively. Each signal systematic fit shows a good fit with a $\chi^2/\mathrm{df} \sim 1$.

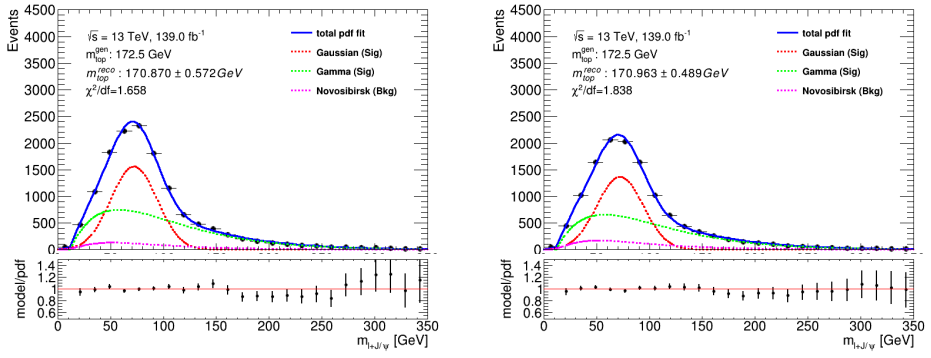
G.1 NLO Generator



(a) (b)

FIGURE G.1: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution using Powheg-Box (a) and MG5_aMC (b) to generate the NLO matrix element matching scheme. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

G.2 Parton shower and hadronisation model



(a) (b)

FIGURE G.2: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution using Pythia 8 (a) and Herwig 7 (b) to model parton showering and hadronisation. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

G.3 Modelling of heavy-quark fragmentation

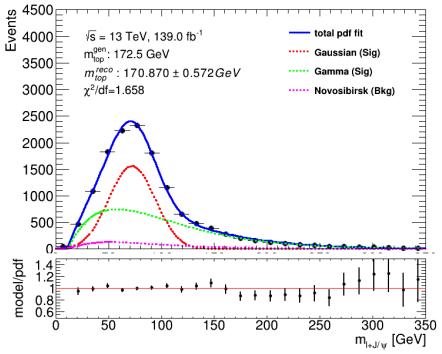


FIGURE G.3: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution, setting the fragmentation parameter r_b to 1.05 in Pythia 8. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

G.4 Colour Reconnection

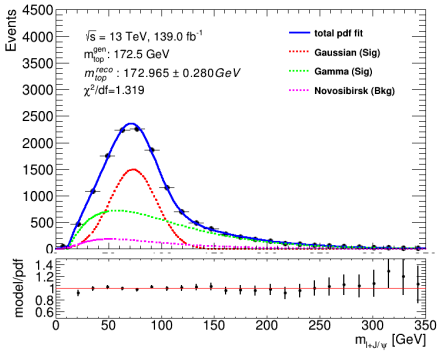


FIGURE G.4: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when using a QCD base model to describe multiple parton interactions and colour reconnection. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

G.5 Modelling of hadron production

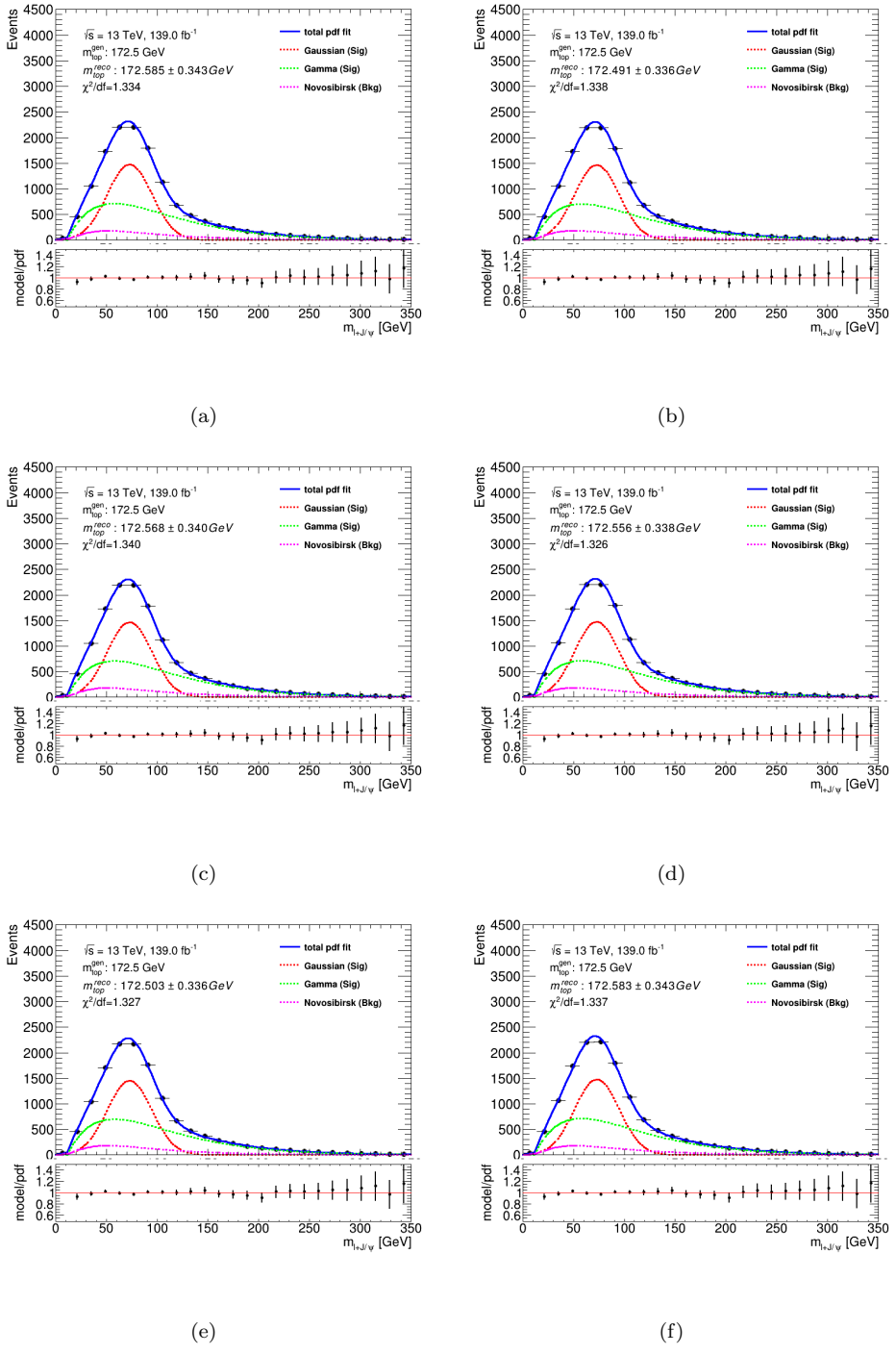


FIGURE G.5: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution, when varying the production fractions of $B^0/B^{\pm}$ up (a) and down (b), B_s^0 up (c) and down (d), and B-baryon up (e) and down (f). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

G.6 Modelling of hadron decays

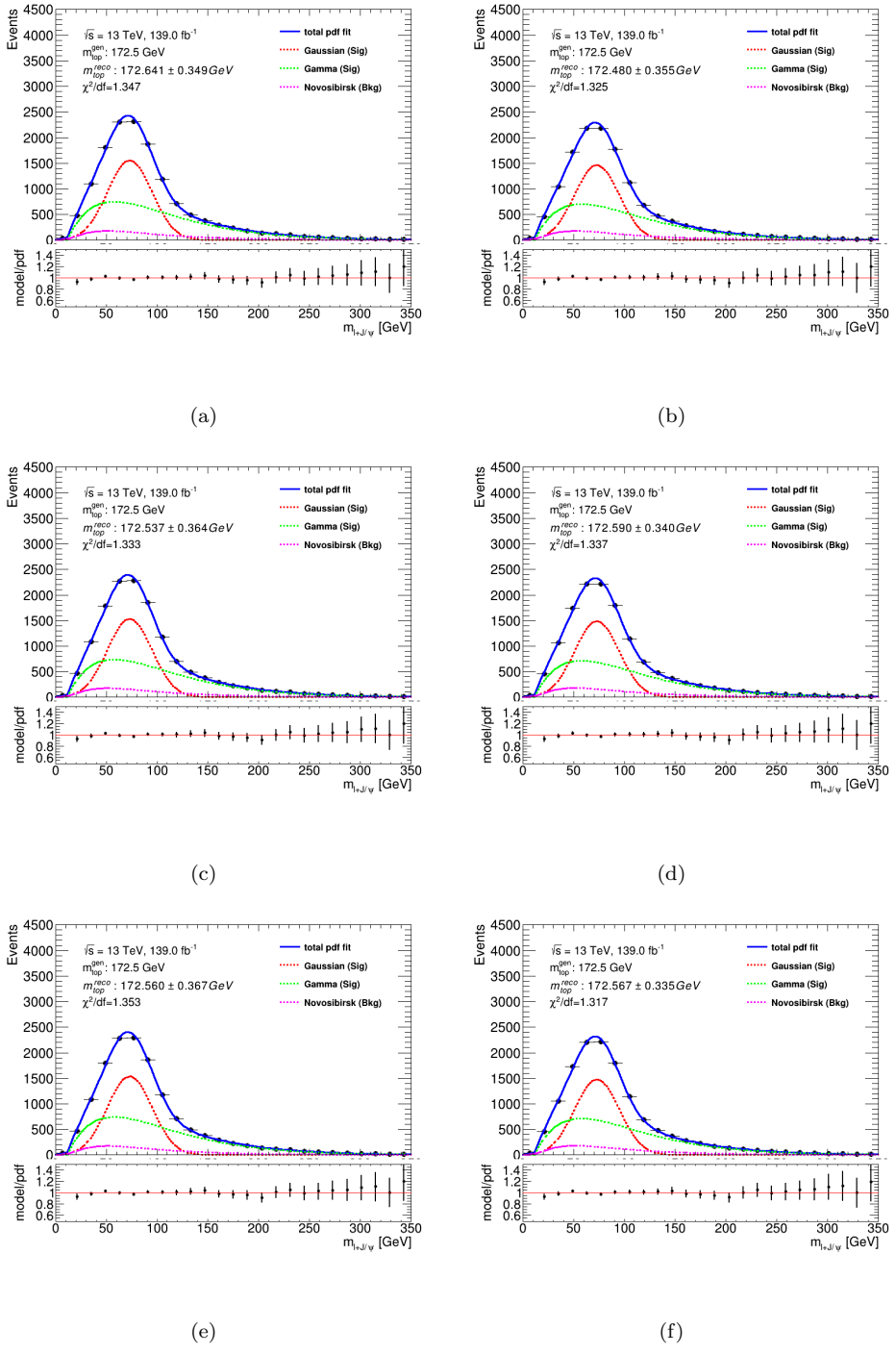


FIGURE G.6: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution, when varying the direct J/ψ decay fractions up (a) and down (b), indirect $\psi(2S)$ decay fractions up (c) and down (d), and indirect $\chi_{c1}(1P)$ decay fractions up (e) and down (f). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

G.7 Modelling of radiation

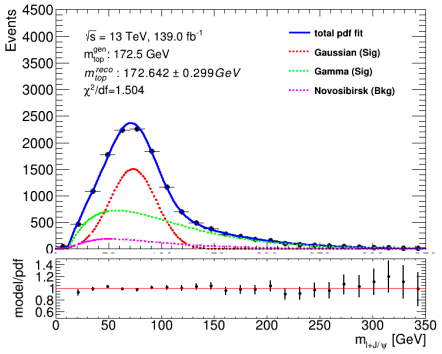
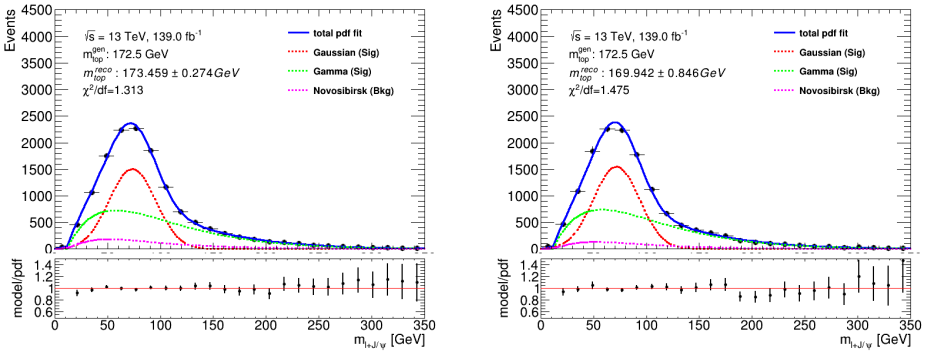
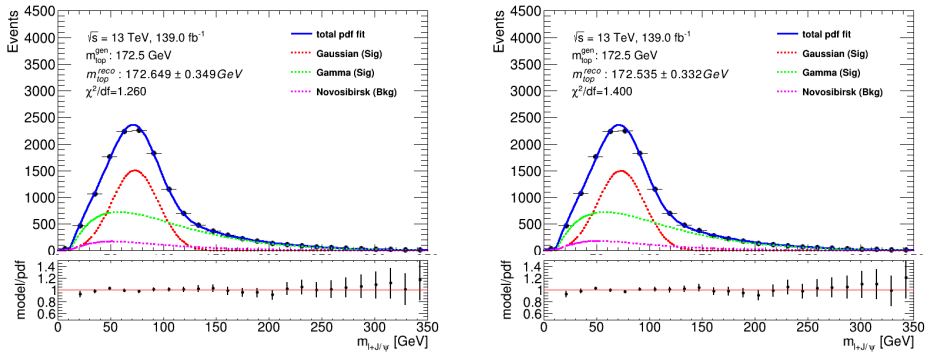


FIGURE G.7: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution, setting the h_{damp} parameter to $3m_{top}$ in Pythia 8. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.



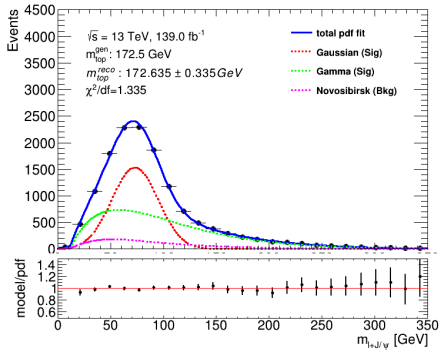
(a) (b)

FIGURE G.8: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution, when varying the final-state radiation μ_R up to 2.0 (a) and down to 0.5 (b). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

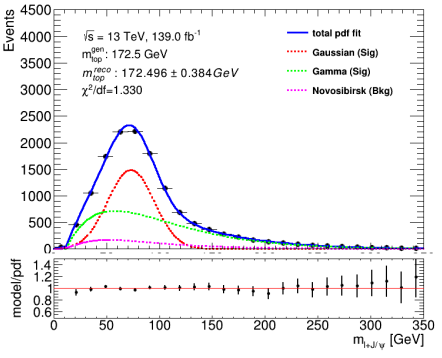


(a)

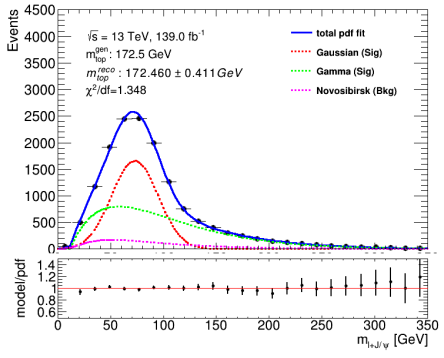
(b)



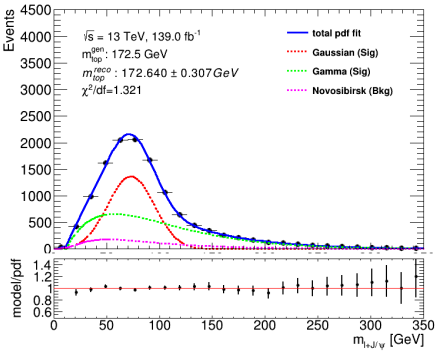
(c)



(d)



(e)



(f)

FIGURE G.9: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution, when varying the initial-state radiation α_s^{ISR} up (a) and down (b), and μ_F up (c) and down (d), and μ_R up (e) and down (f). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

G.8 Choice of parton distribution function

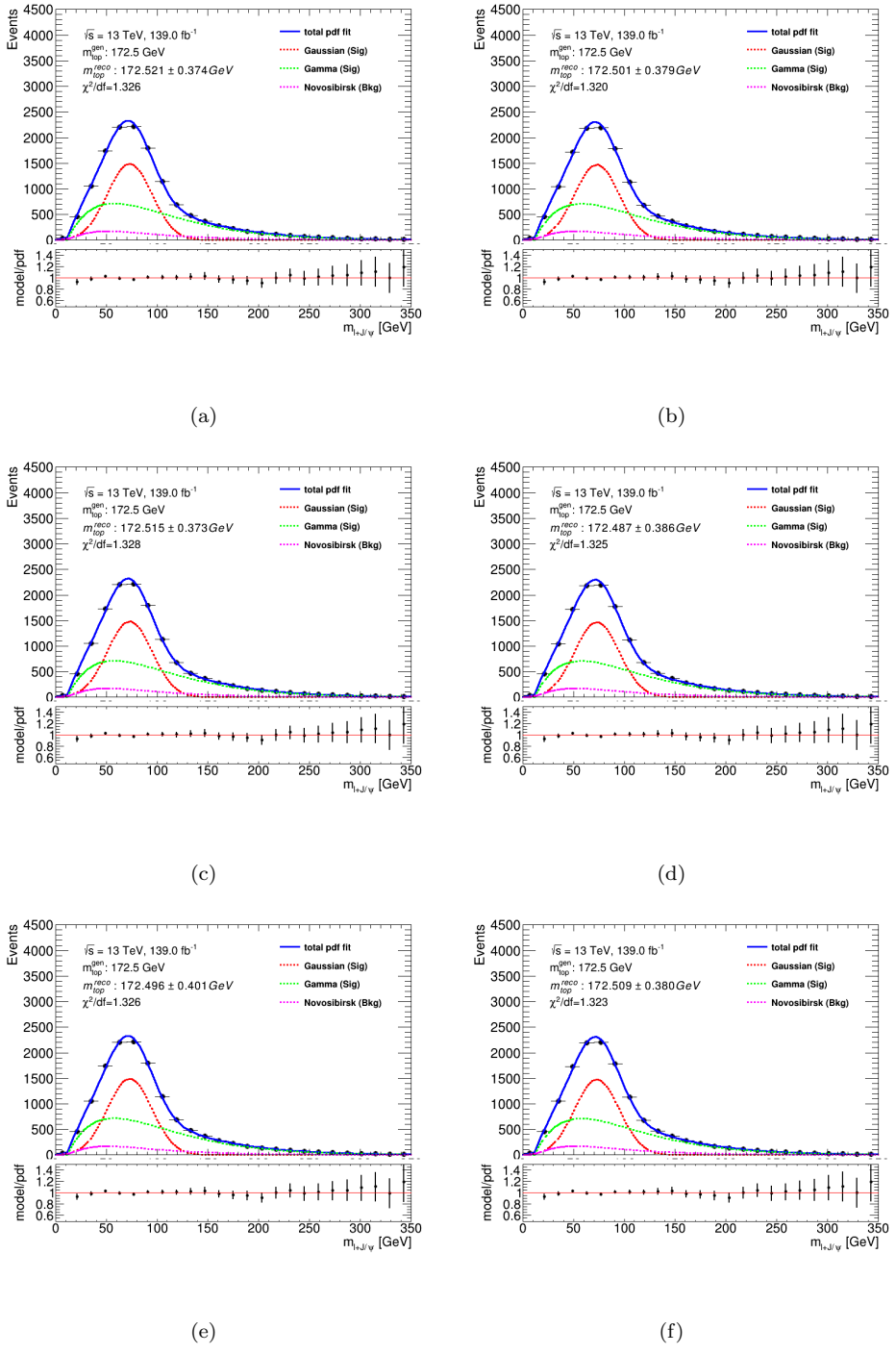


FIGURE G.10: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution of the 1 (a), 2 (b), 3 (c), 4 (d), 5 (e), and 6 (f) PDF replicas of the PDF4LHC15 error set. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

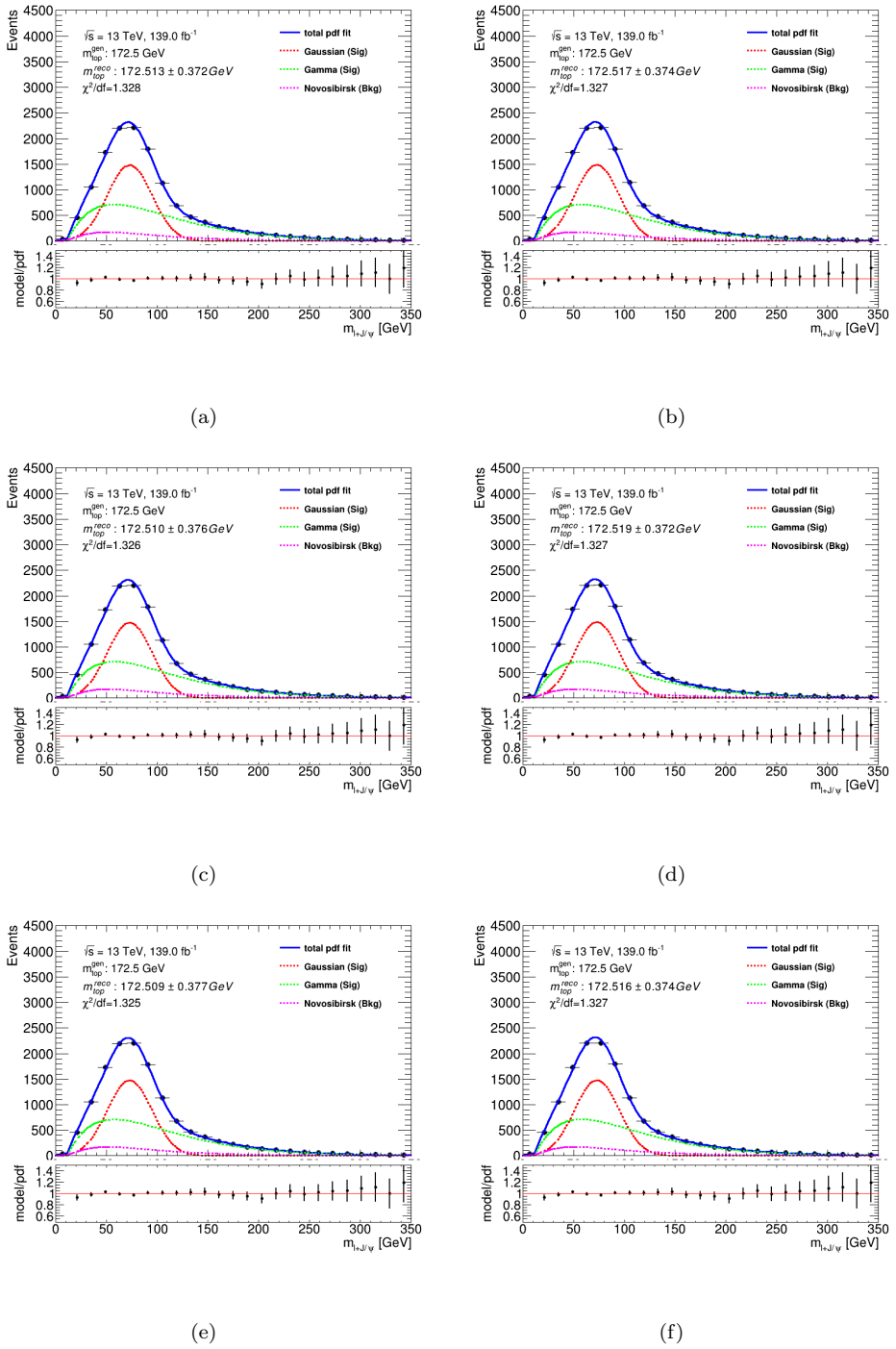


FIGURE G.11: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution of the 7 (a), 8 (b), 9 (c), 10 (d), 11 (e), and 12 (f) PDF replicas of the PDF4LHC15 error set. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

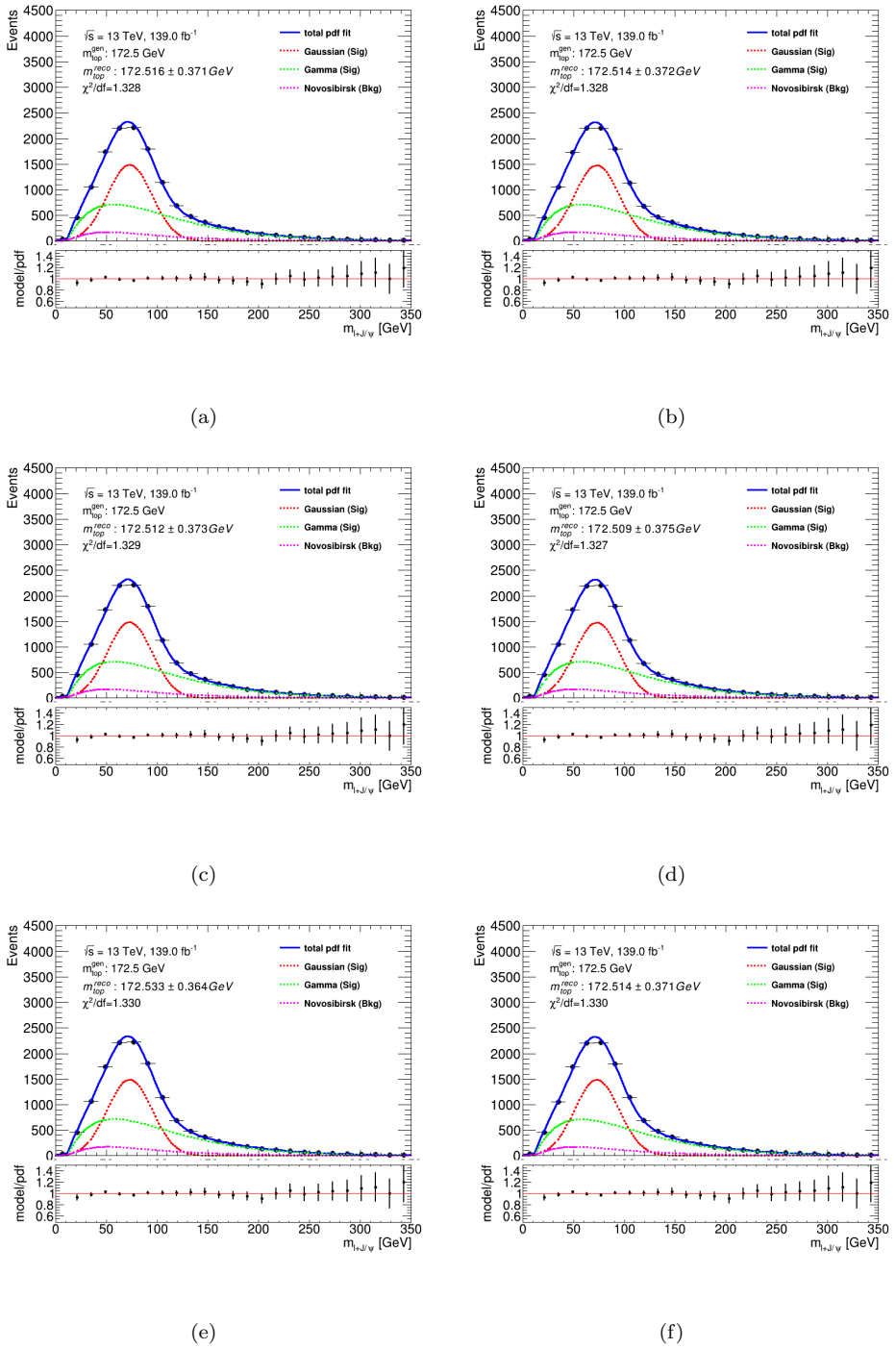


FIGURE G.12: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution of the 13 (a), 14 (b), 15 (c), 16 (d), 17 (e), and 18 (f) PDF replicas of the PDF4LHC15 error set. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

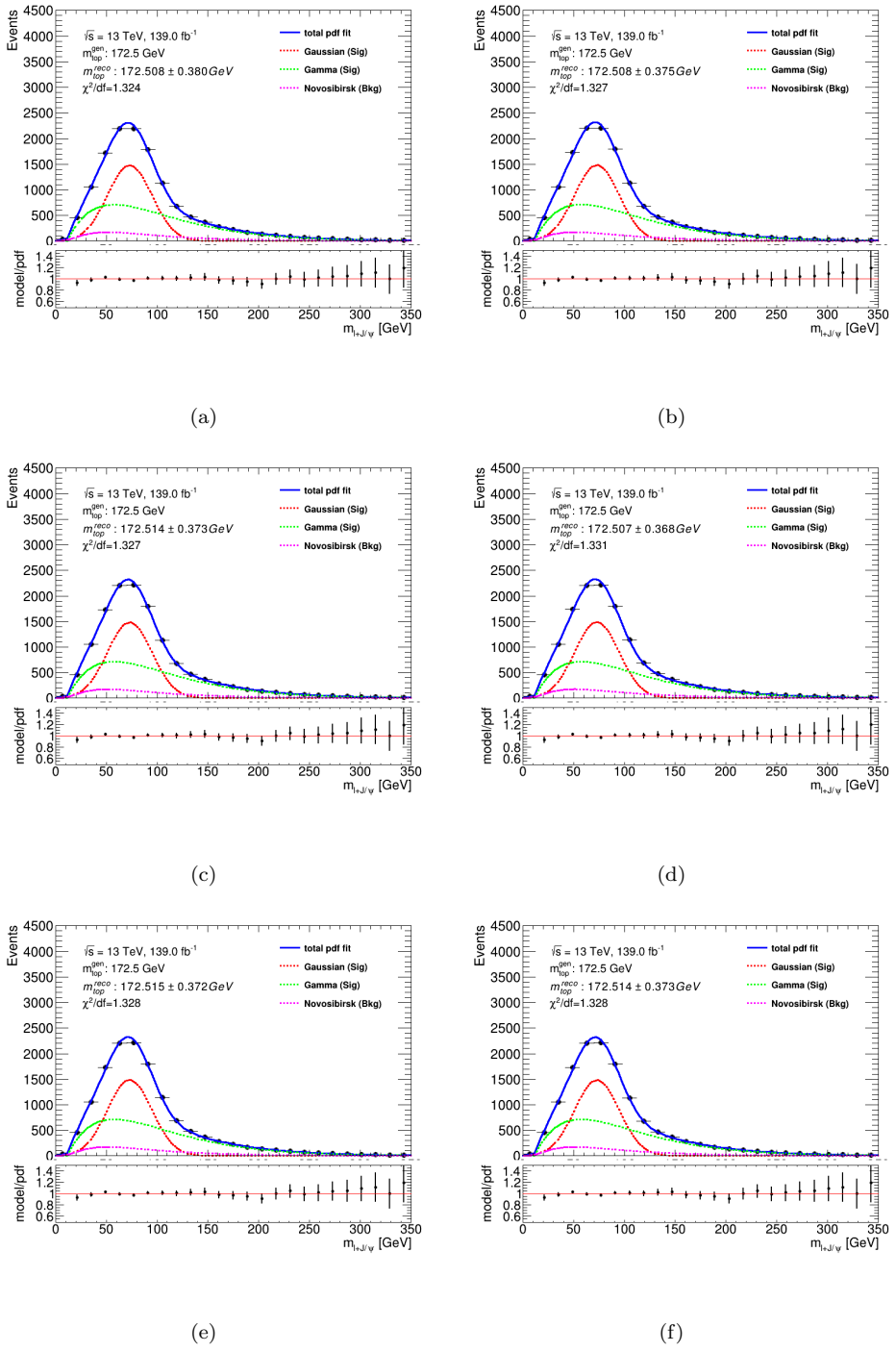
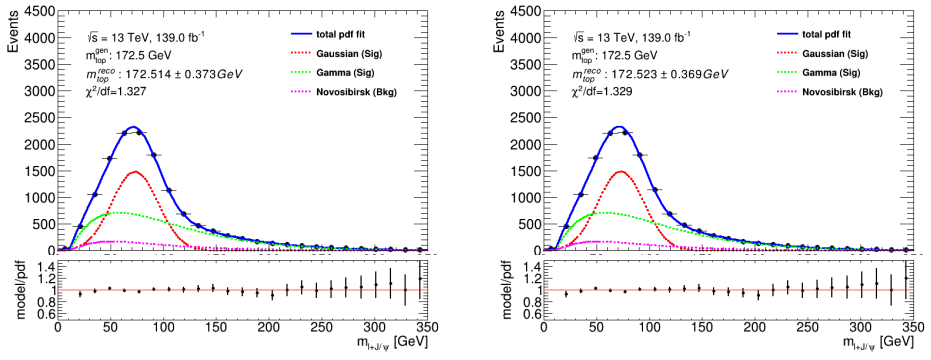
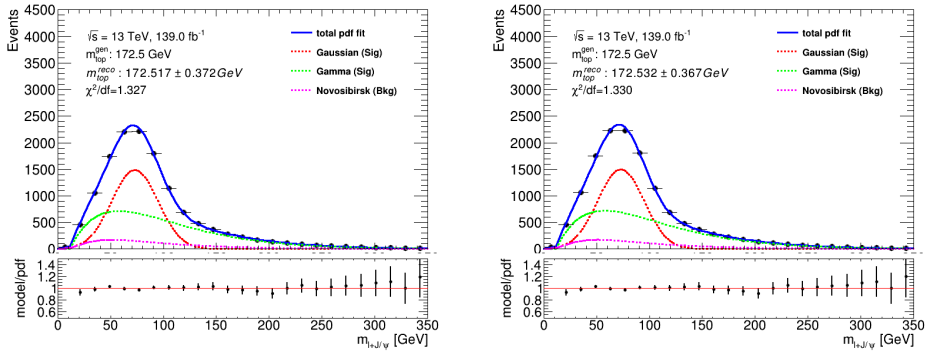


FIGURE G.13: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution of the 19 (a), 20 (b), 21 (c), 22 (d), 23 (e), and 24 (f) PDF replicas of the PDF4LHC15 error set. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.



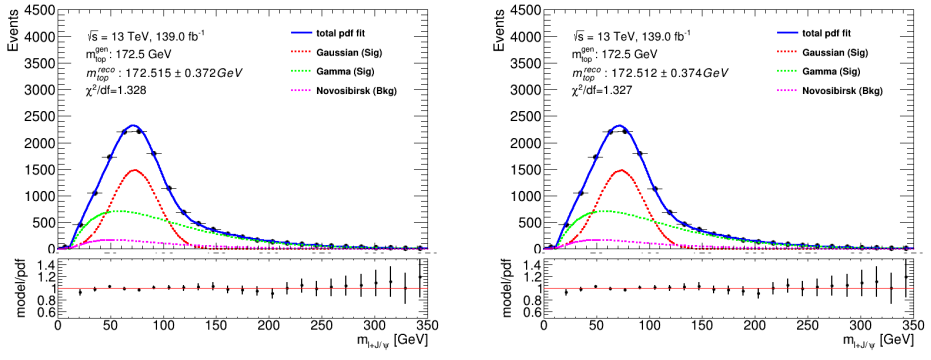
(a)

(b)



(c)

(d)



(e)

(f)

FIGURE G.14: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution of the 25 (a), 26 (b), 27 (c), 28 (d), 29 (e), and 30 (f) PDF replicas of the PDF4LHC15 error set. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

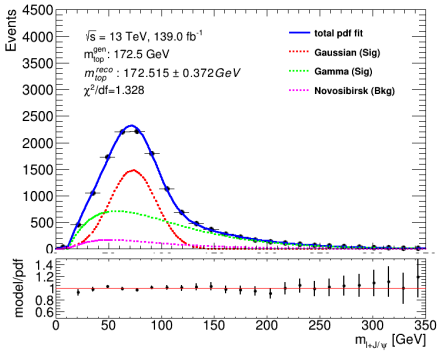


FIGURE G.15: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution of the nominal PDF4LHC15 PDF error set. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

Appendix H

Background modelling systematic fit results

The following chapter shows the fit results for the different background modelling systematic uncertainties. Additionally, the control region results for measuring the fake rates are shown after varying the background SM processes accordingly.

H.1 NPFL

Figure H.1 shows the fit results for the different NPFL systematic uncertainties with the reconstructed top quark mass and its statistical uncertainty, as well as the fit quality in the top left corner of the upper panels. The lower panels consist of the ratio between the model and the pdf. The blue shows the total pdf while the red, green and purple shows the Gaussian, Gamma and Novosibirsk contributions, respectively. Each signal systematic fit shows a good fit with a $\chi^2/\text{df} \sim 1$.

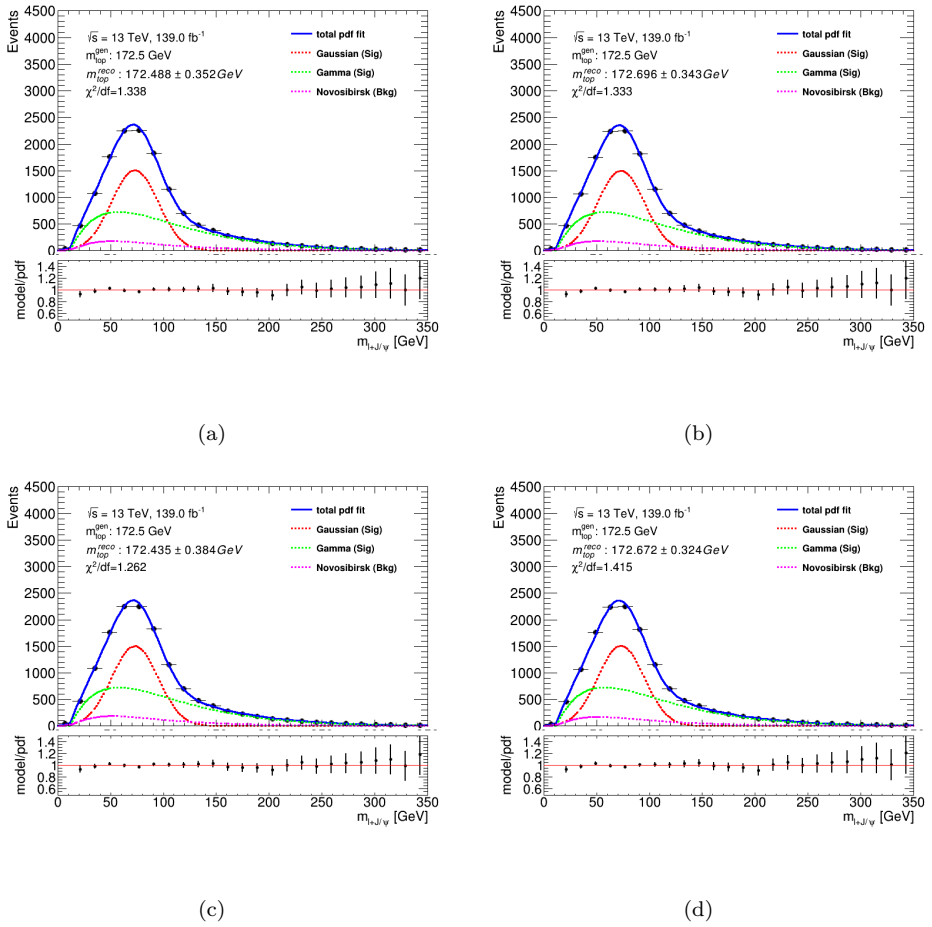


FIGURE H.1: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution after varying the efficiencies up (a) and down (b) and the fake rates up (c) and down (d) within their statistical uncertainties. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

H.2 Background SM processes

Figures H.2-H.4, H.9-H.11, H.16, H.19, and H.22 show the results in the control region used to measure the fake rates before and after varying the different background SM processes accordingly, for 2015, 2016, and 2017, respectively. The results for 2018 is found under Section 5.3.4.2 in main body of the thesis.

Figures H.5-H.7, H.12-H.14, H.17, H.20, and H.23 show the nominal efficiencies and fake rates, and the fake rates after varying the different background SM processes accordingly for 2015, 2016, and 2017, respectively. The results for 2018 is found under Section 5.3.4.2 in main body of the thesis.

Figures H.8, H.15, H.18, H.21, and H.24 show the fit results for the background SM processes systematic uncertainties with the reconstructed top quark mass and its statistical uncertainty, as well as the fit quality in the top left corner of the upper panels. The lower panels consist of the ratio between the model and the pdf. The blue shows the total pdf while the red, green and purple shows the Gaussian, Gamma and Novosibirsk contributions, respectively. Each signal systematic fit shows a good fit with a $\chi^2/\mathrm{df} \sim 1$.

H.2.1 W+jets

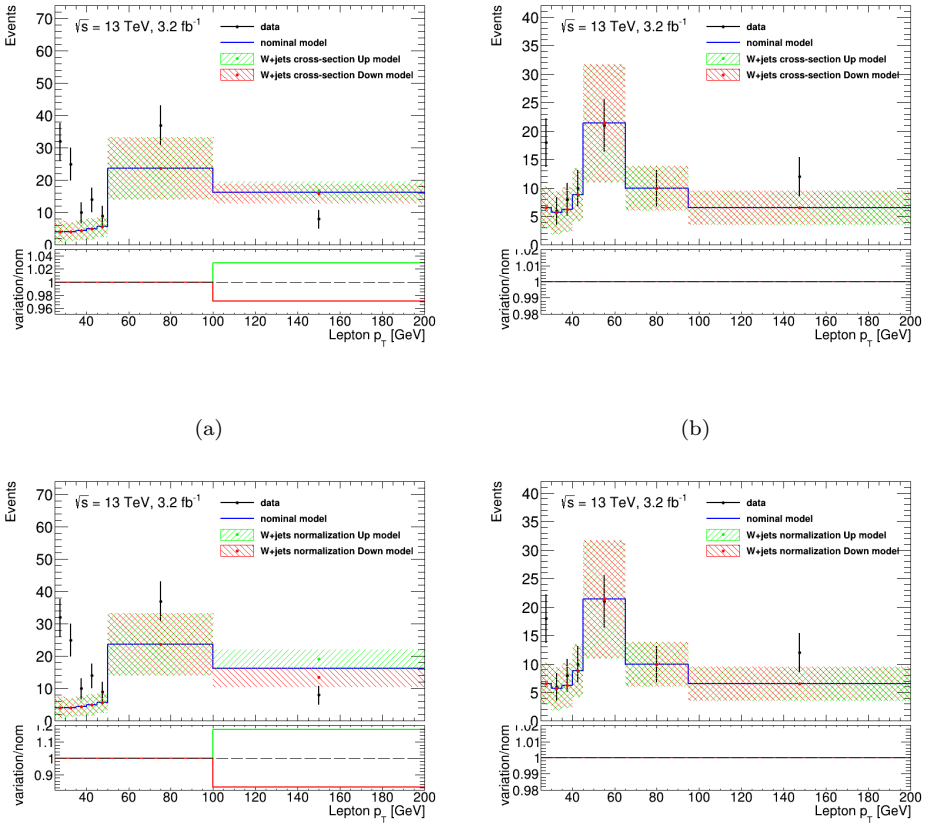


FIGURE H.2: The upper panels show the data and total SM contribution in the fake rates control region in the transverse momentum distribution in the tight electron (left) and muon (right) channels for 2015, when varying the W+jets SM processes up and down by its cross-section (top row) and normalisation (bottom row) uncertainties. The black points show the data and the blue line shows the nominal model contribution, while the green and red show the model contributions after the variations. The hashed area shows the statistical, cross-section and luminosity uncertainties on the model contributions. The lower panel plot shows the ratio between the model after the variations and the nominal model.

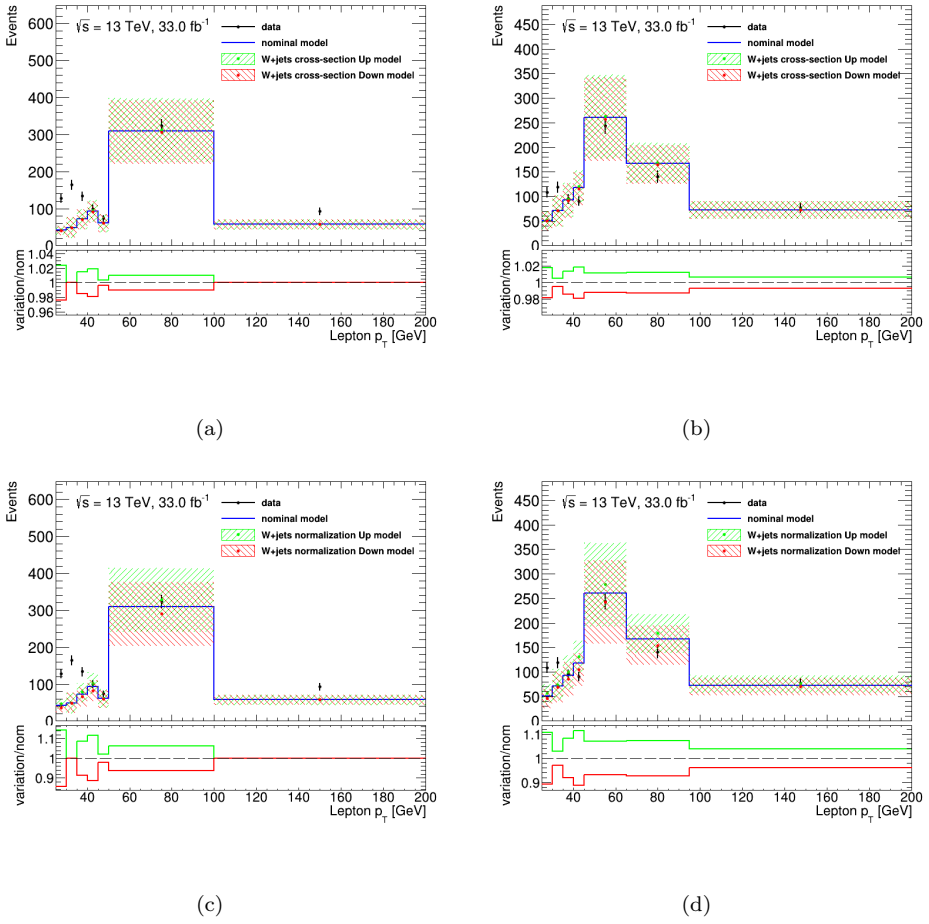
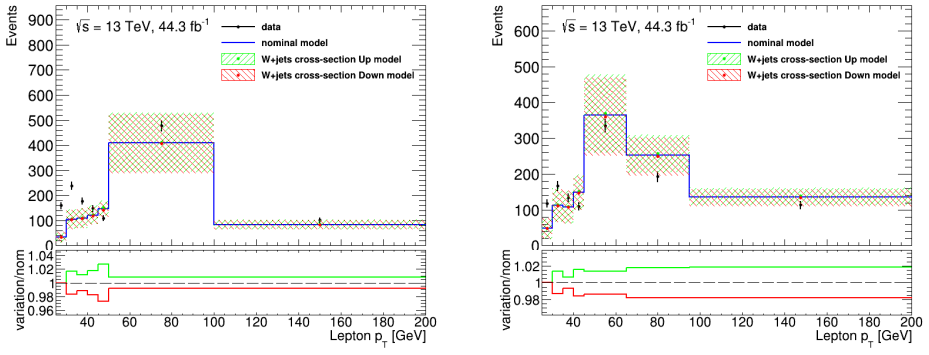
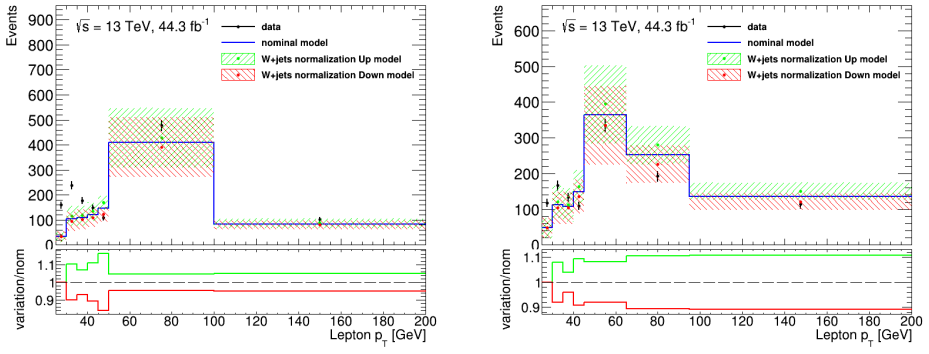


FIGURE H.3: The upper panels show the data and total SM contribution in the fake rates control region in the transverse momentum distribution in the tight electron (left) and muon (right) channels for 2016, when varying the W+jets SM processes up and down by its cross-section (top row) and normalisation (bottom row) uncertainties. The black points show the data and the blue line shows the nominal model contribution, while the green and red show the model contributions after the variations. The hashed area shows the statistical, cross-section and luminosity uncertainties on the model contributions. The lower panel plot shows the ratio between the model after the variations and the nominal model.

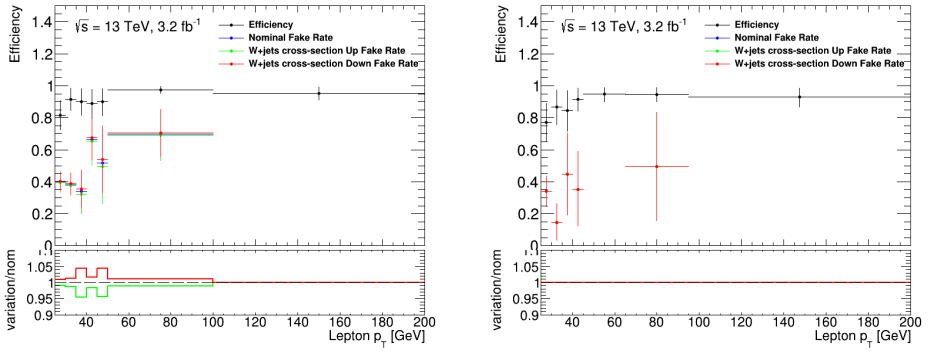


(a) (b)

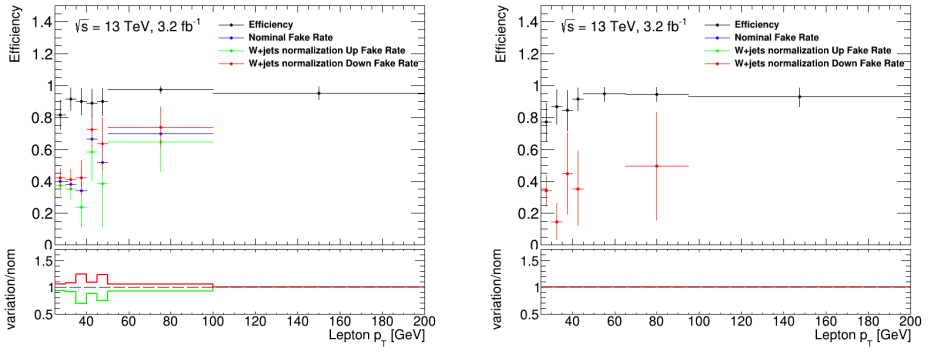


(c) (d)

FIGURE H.4: The upper panels show the data and total SM contribution in the fake rates control region in the transverse momentum distribution in the tight electron (left) and muon (right) channels for 2017, when varying the W+jets SM processes up and down by its cross-section (top row) and normalisation (bottom row) uncertainties. The black points show the data and the blue line shows the nominal model contribution, while the green and red show the model contributions after the variations. The hashed area shows the statistical, cross-section and luminosity uncertainties on the model contributions. The lower panel plot shows the ratio between the model after the variations and the nominal model.

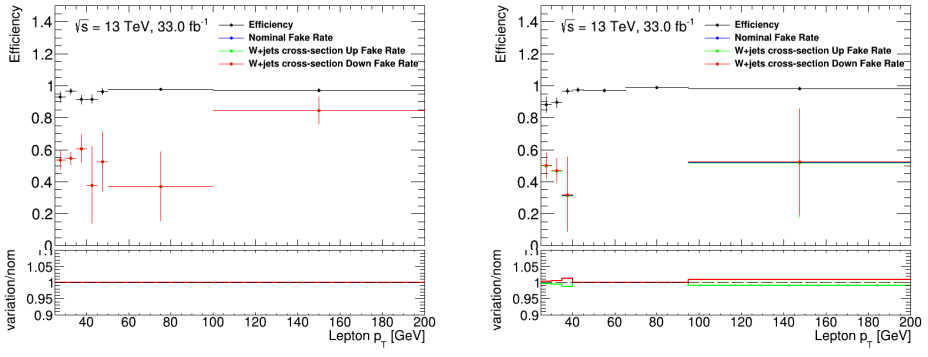


(a) (b)

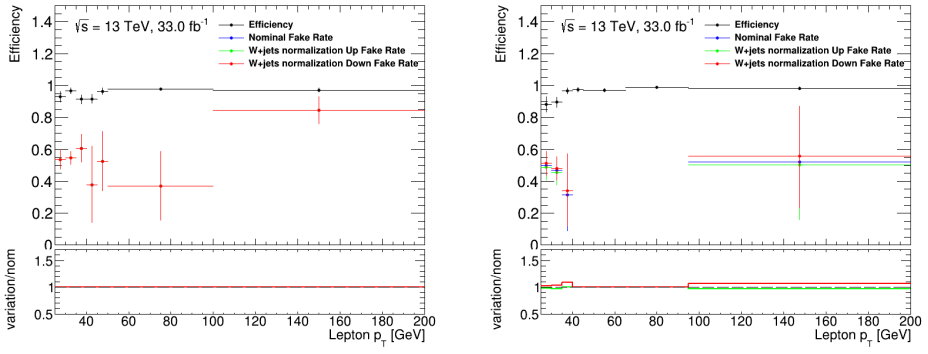


(c) (d)

FIGURE H.5: The upper panels show the nominal efficiencies and fake rates, and after the cross-section (top row) and normalisation (bottom row) uncertainty variations have been applied to the W+jets SM processes in the transverse momentum distribution, in the electron (left) and muon (right) channels for 2015. The black and blue points show the nominal efficiencies and fake rates, while the green and red show the up and down variations, respectively. The error bars show the statistical uncertainty. The lower panel plot shows the ratio between the fake rate after variation and the nominal.

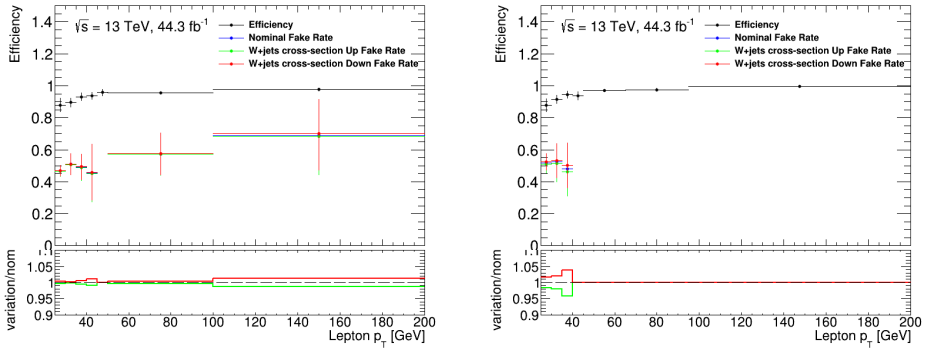


(a) (b)

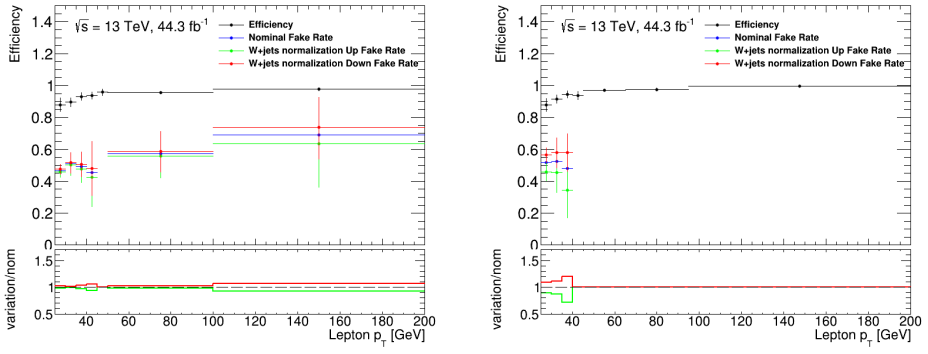


(c) (d)

FIGURE H.6: The upper panels show the nominal efficiencies and fake rates, and after the cross-section (top row) and normalisation (bottom row) uncertainty variations have been applied to the W+jets SM processes in the transverse momentum distribution, in the electron (left) and muon (right) channels for 2016. The black and blue points show the nominal efficiencies and fake rates, while the green and red show the up and down variations, respectively. The error bars show the statistical uncertainty. The lower panel plot shows the ratio between the fake rate after variation and the nominal.



(a) (b)



(c) (d)

FIGURE H.7: The upper panels show the nominal efficiencies and fake rates, and after the cross-section (top row) and normalisation (bottom row) uncertainty variations have been applied to the W+jets SM processes in the transverse momentum distribution, in the electron (left) and muon (right) channels for 2017. The black and blue points show the nominal efficiencies and fake rates, while the green and red show the up and down variations, respectively. The error bars show the statistical uncertainty. The lower panel plot shows the ratio between the fake rate after variation and the nominal.

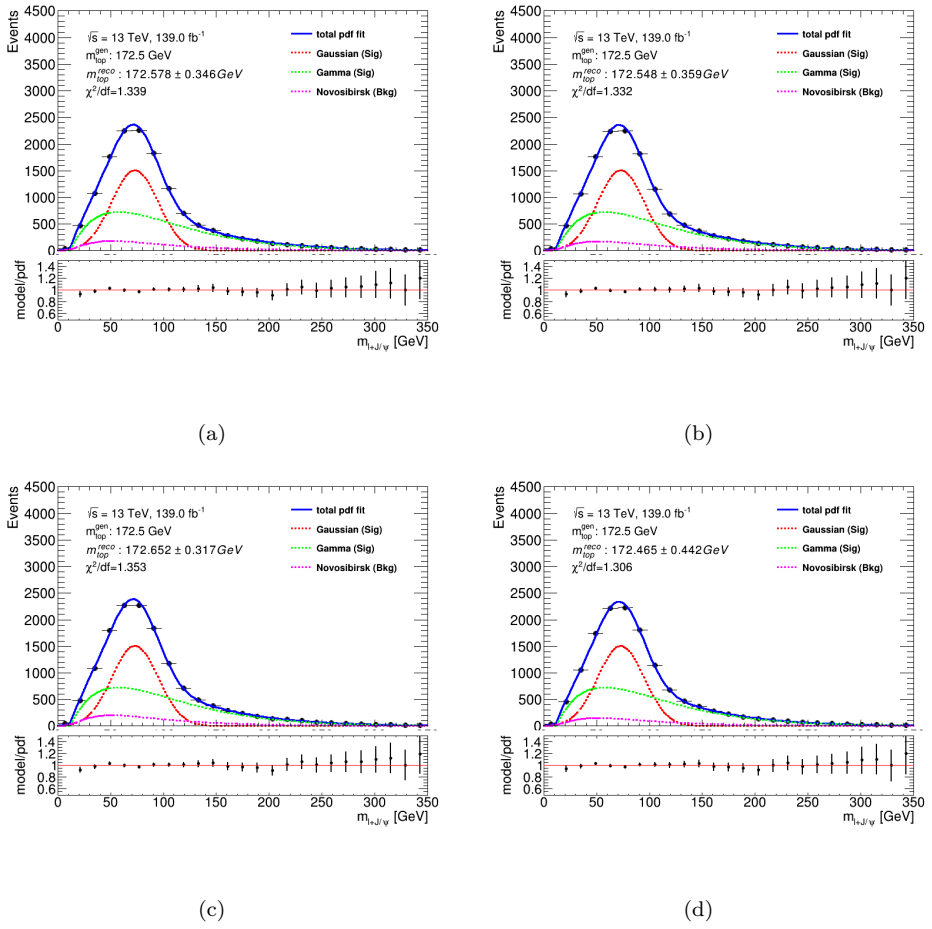


FIGURE H.8: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution after varying the W+jets SM processes by its cross-section, up (a) and down (b), and normalisation, up (c) and down (d), uncertainties. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

H.2.2 Z+jets

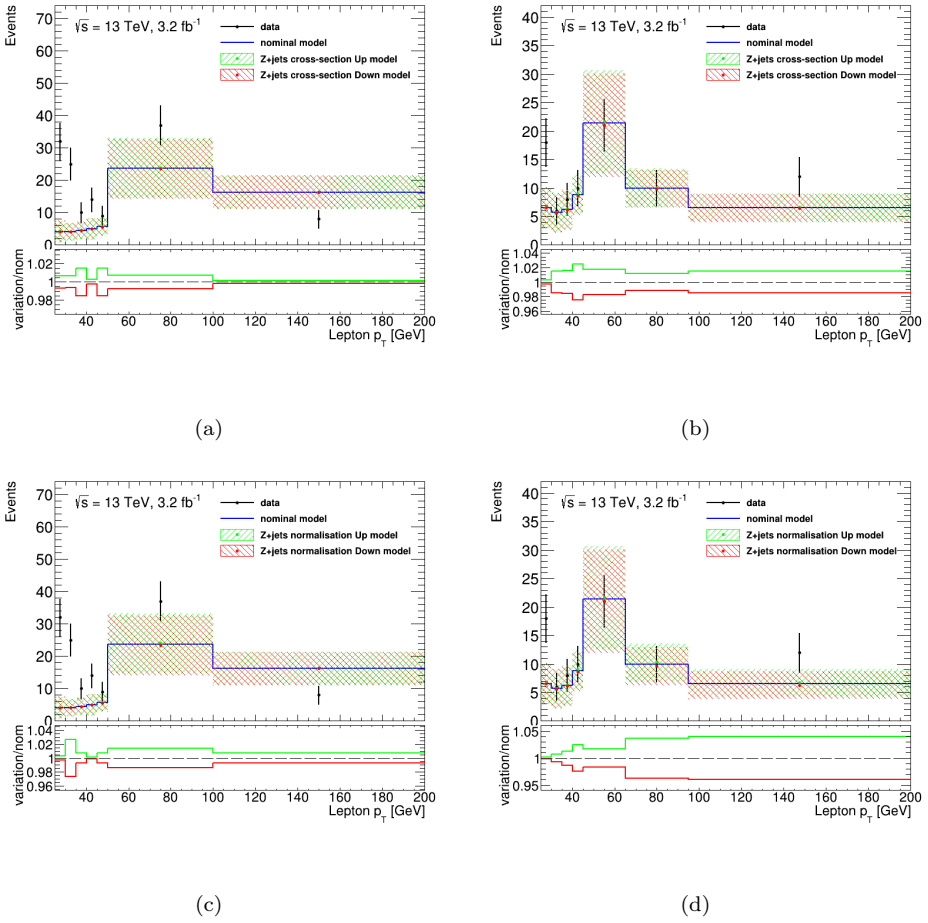
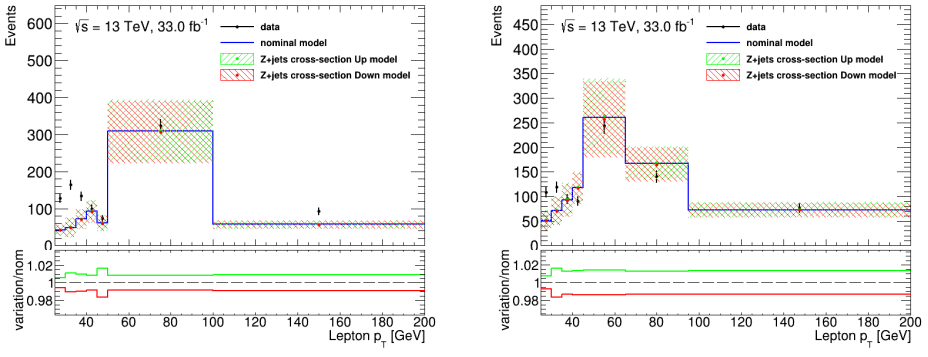
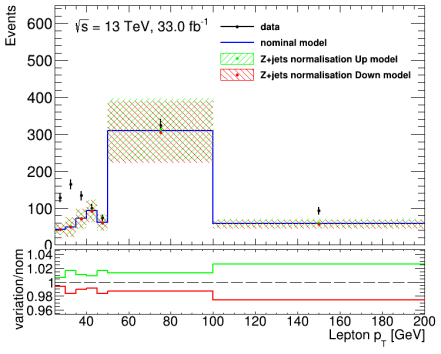


FIGURE H.9: The upper panels show the data and total SM contribution in the fake rates control region in the transverse momentum distribution in the tight electron (left) and muon (right) channels for 2015, when varying the Z+jets SM processes up and down by its cross-section (top row) and normalisation (bottom row) uncertainties. The black points show the data and the blue line shows the nominal model contribution, while the green and red show the model contributions after the variations. The hashed area shows the statistical, cross-section and luminosity uncertainties on the model contributions. The lower panel plot shows the ratio between the model after the variations and the nominal model.

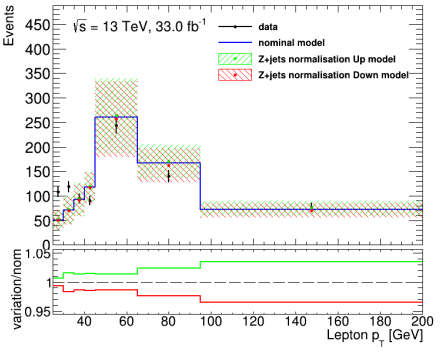


(a)

(b)



(c)



(d)

FIGURE H.10: The upper panels show the data and total SM contribution in the fake rates control region in the transverse momentum distribution in the tight electron (left) and muon (right) channels for 2016, when varying the Z+jets SM processes up and down by its cross-section (top row) and normalisation (bottom row) uncertainties. The black points show the data and the blue line shows the nominal model contribution, while the green and red show the model contributions after the variations. The hashed area shows the statistical, cross-section and luminosity uncertainties on the model contributions. The lower panel plot shows the ratio between the model after the variations and the nominal model.

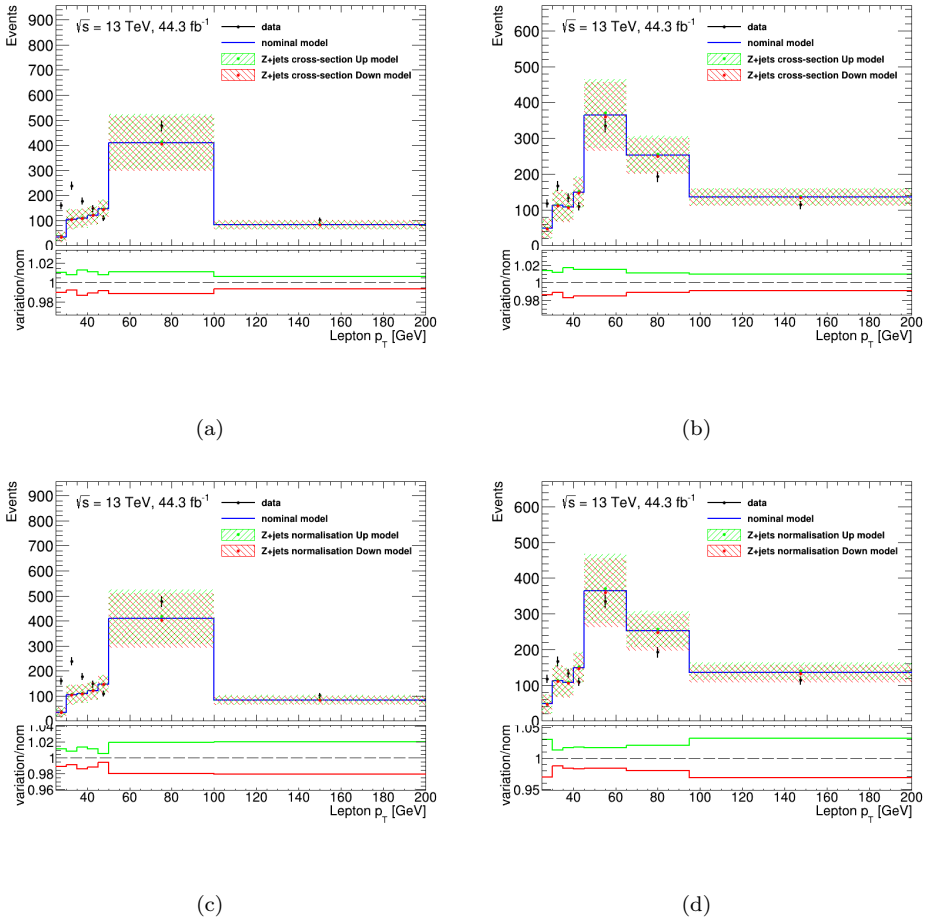
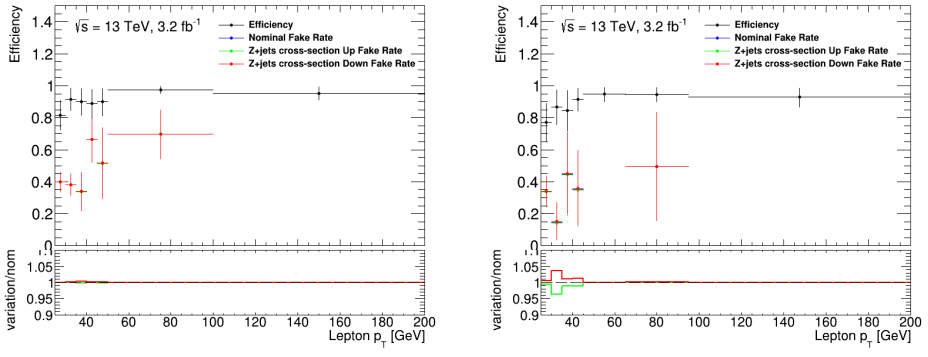
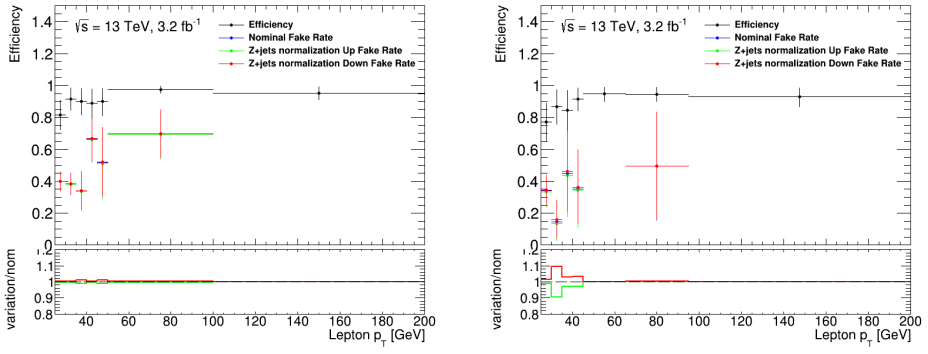


FIGURE H.11: The upper panels show the data and total SM contribution in the fake rates control region in the transverse momentum distribution in the tight electron (left) and muon (right) channels for 2017, when varying the Z+jets SM processes up and down by its cross-section (top row) and normalisation (bottom row) uncertainties. The black points show the data and the blue line shows the nominal model contribution, while the green and red show the model contributions after the variations. The hashed area shows the statistical, cross-section and luminosity uncertainties on the model contributions. The lower panel plot shows the ratio between the model after the variations and the nominal model.

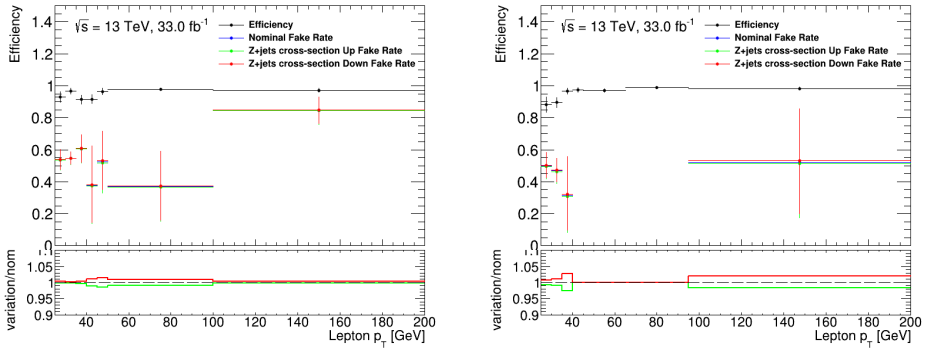


(a) (b)

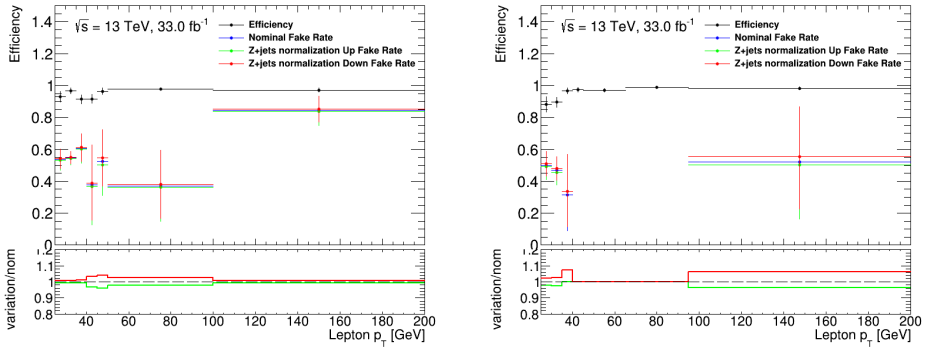


(c) (d)

FIGURE H.12: The upper panels show the nominal efficiencies and fake rates, and after the cross-section (top row) and normalisation (bottom row) uncertainty variations have been applied to the Z+jets SM processes in the transverse momentum distribution, in the electron (left) and muon (right) channels for 2015. The black and blue points show the nominal efficiencies and fake rates, while the green and red show the up and down variations, respectively. The error bars show the statistical uncertainty. The lower panel plot shows the ratio between the fake rate after variation and the nominal.

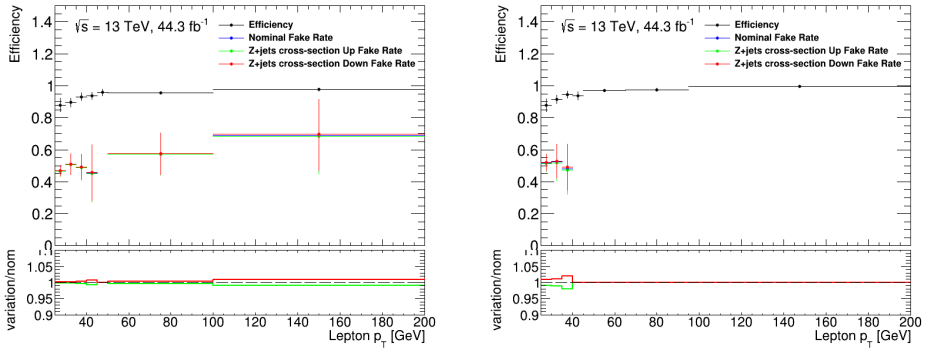


(a) (b)

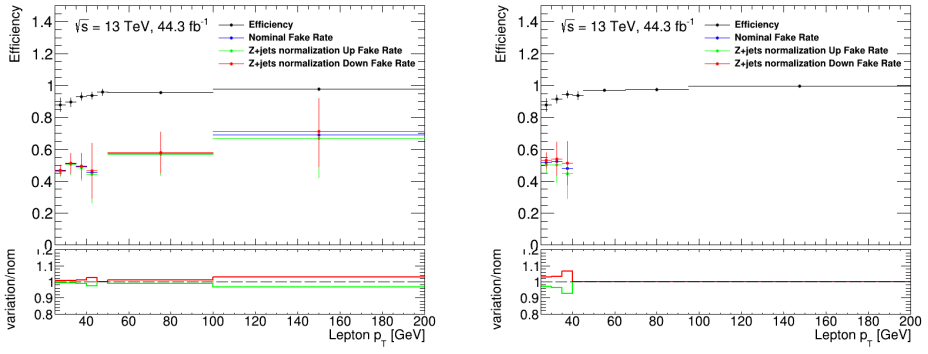


(c) (d)

FIGURE H.13: The upper panels show the nominal efficiencies and fake rates, and after the cross-section (top row) and normalisation (bottom row) uncertainty variations have been applied to the Z+jets SM processes in the transverse momentum distribution, in the electron (left) and muon (right) channels for 2016. The black and blue points show the nominal efficiencies and fake rates, while the green and red show the up and down variations, respectively. The error bars show the statistical uncertainty. The lower panel plot shows the ratio between the fake rate after variation and the nominal.



(a) (b)



(c) (d)

FIGURE H.14: The upper panels show the nominal efficiencies and fake rates, and after the cross-section (top row) and normalisation (bottom row) uncertainty variations have been applied to the Z+jets SM processes in the transverse momentum distribution, in the electron (left) and muon (right) channels for 2017. The black and blue points show the nominal efficiencies and fake rates, while the green and red show the up and down variations, respectively. The error bars show the statistical uncertainty. The lower panel plot shows the ratio between the fake rate after variation and the nominal.

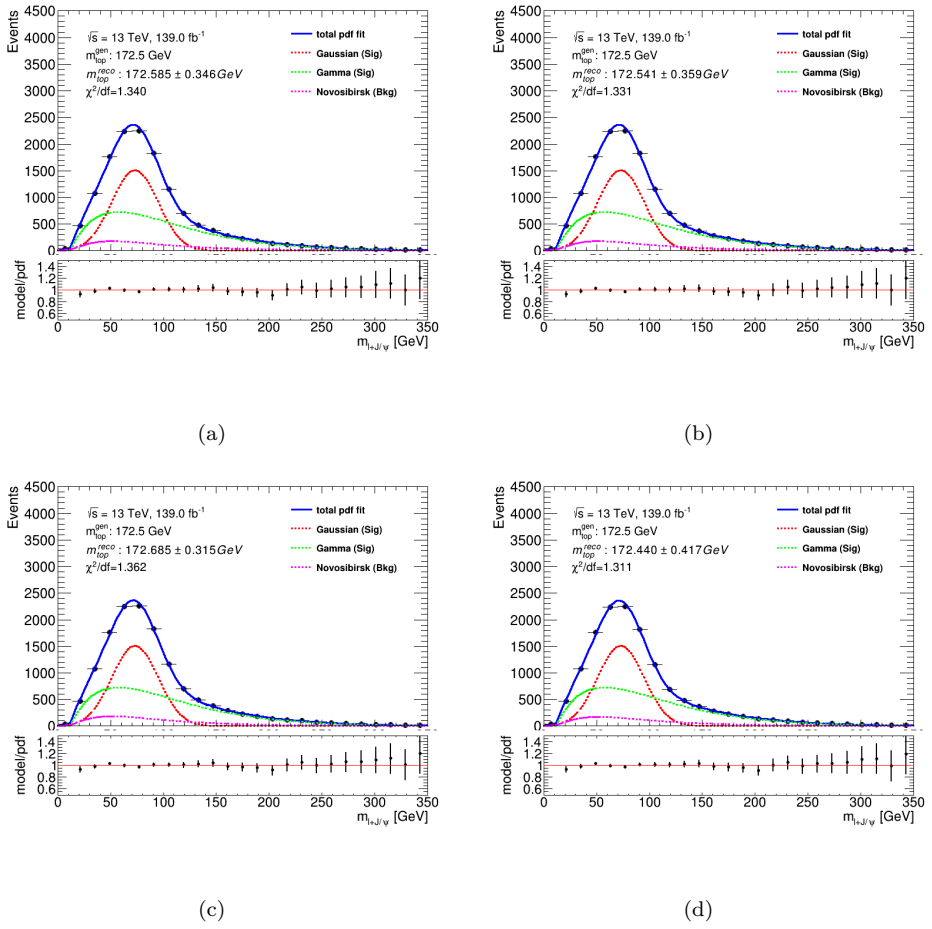


FIGURE H.15: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution after varying the Z+jets SM processes by its cross-section, up (a) and down (b), and normalisation, up (c) and down (d), uncertainties. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

H.2.3 Diboson (VV)

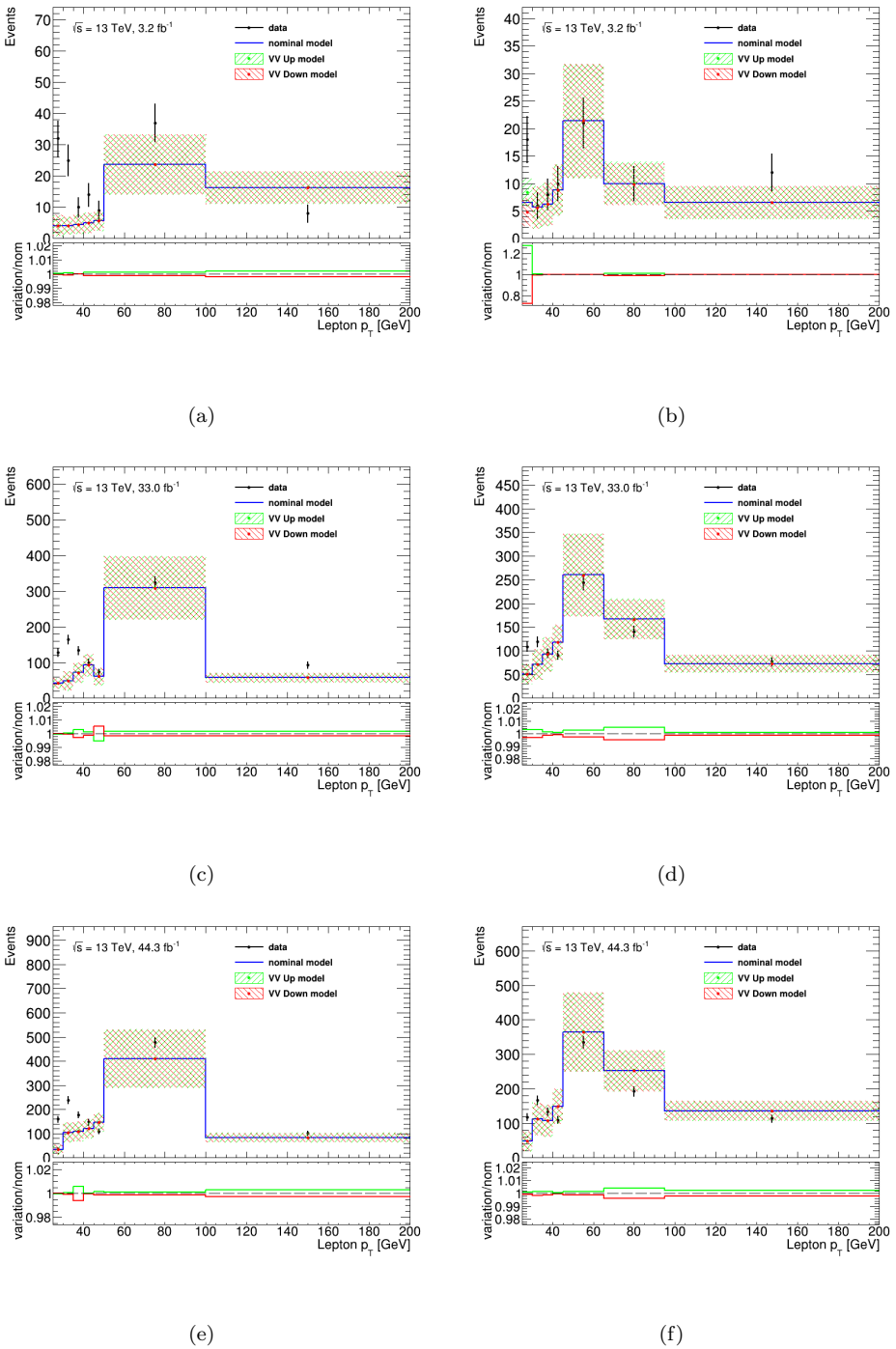
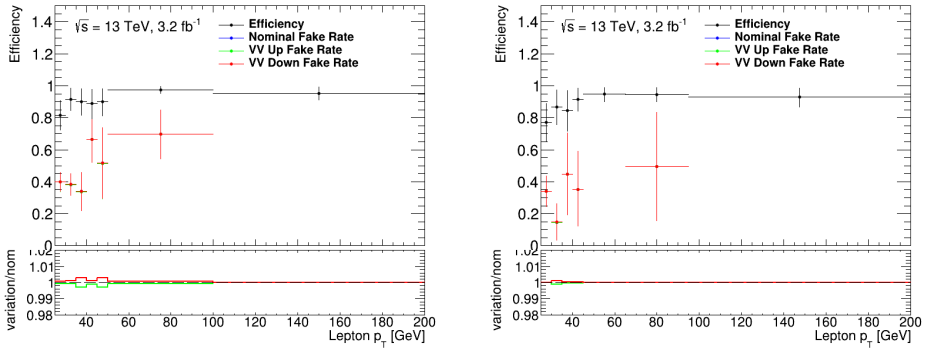
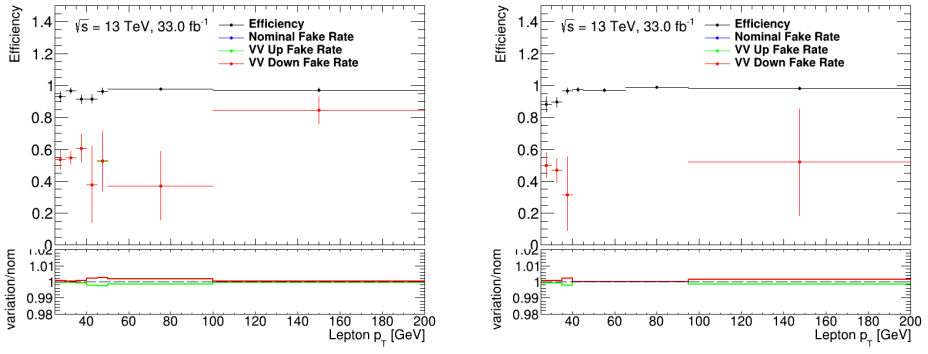


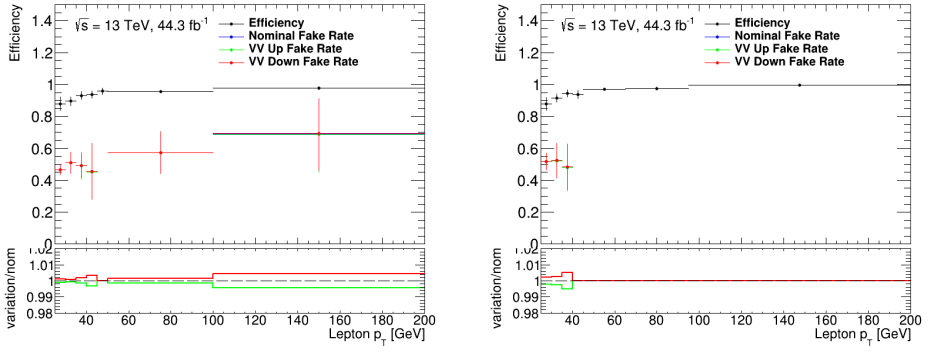
FIGURE H.16: The upper panels show the data and total SM contribution in the fake rates control region in the transverse momentum distribution in the tight electron (left) and muon (right) channels for 2015 (a, b), 2016 (c, d) and 2017 (e, f), when varying the VV SM processes up and down by their cross-section uncertainties. The black points show the data and the blue line shows the nominal model contribution, while the green and red show the model contributions after the variations. The hashed area shows the statistical, cross-section and luminosity uncertainties on the model contributions. The lower panel plot shows the ratio between the model after the variations and the nominal model.



(a) (b)

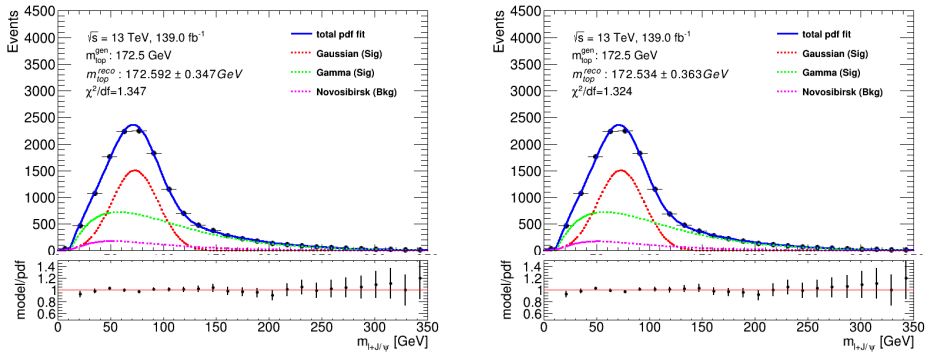


(c) (d)



(e) (f)

FIGURE H.17: The upper panels show the nominal efficiencies and fake rates, and after the cross-section uncertainty variations have been applied to the VV SM processes in the transverse momentum distribution, in the electron (left) and muon (right) channels for 2015 (a, b), 2016 (c, d) and 2017 (e, f), respectively. The black and blue points show the nominal efficiencies and fake rates, while the green and red show the up and down variations, respectively. The error bars show the statistical uncertainty. The lower panel plot shows the ratio between the fake rate after variation and the nominal.



(a) (b)

FIGURE H.18: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution after varying the VV SM processes by their cross-section uncertainties, up (a) and down (b), respectively. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

H.2.4 $t\bar{t}V$

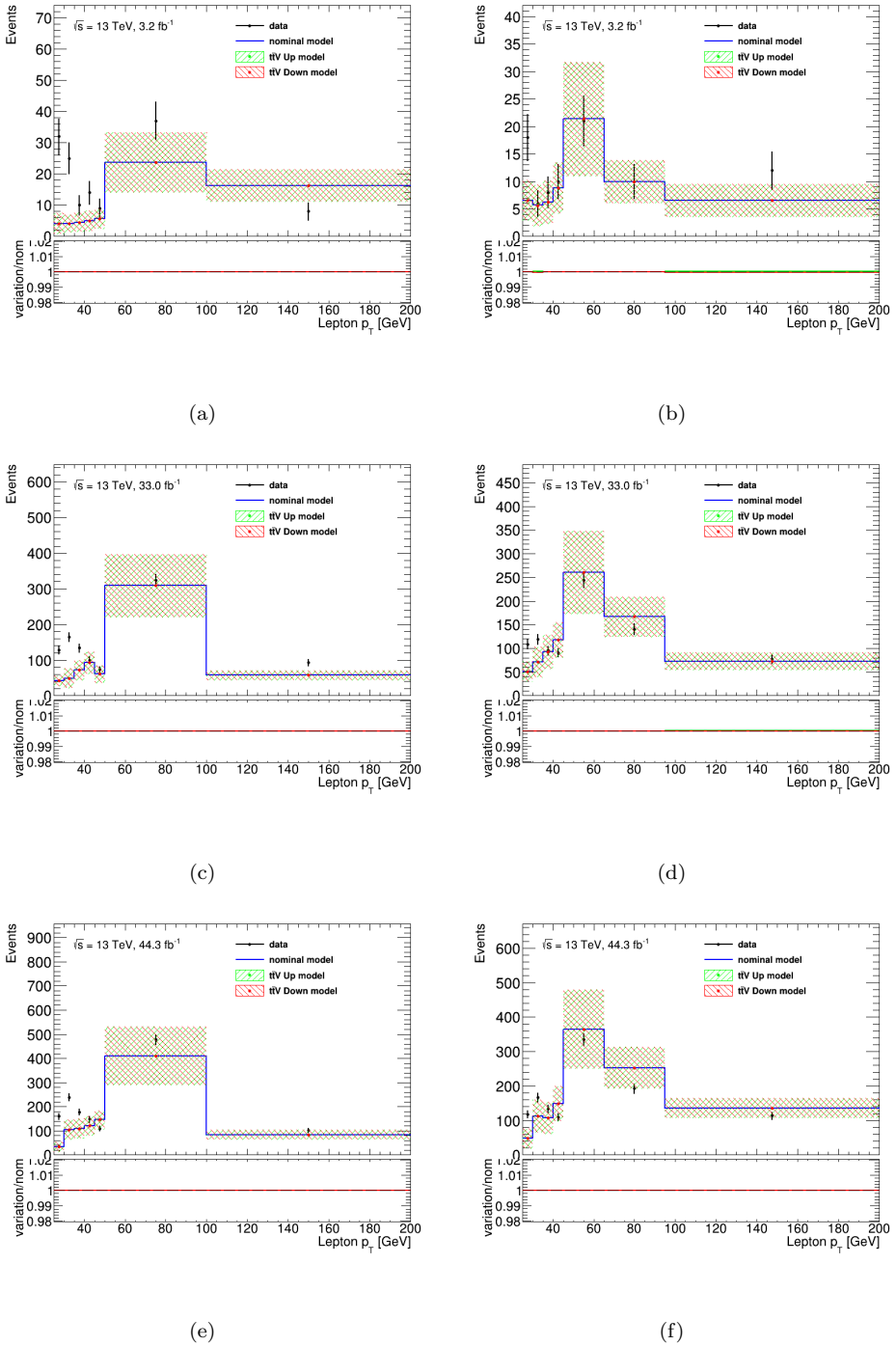
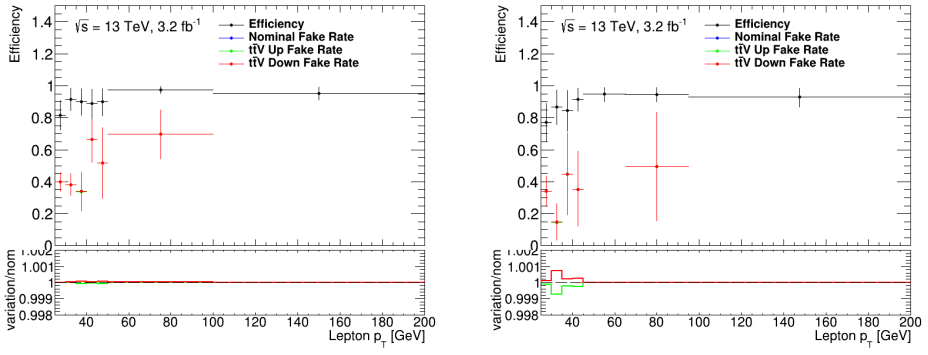
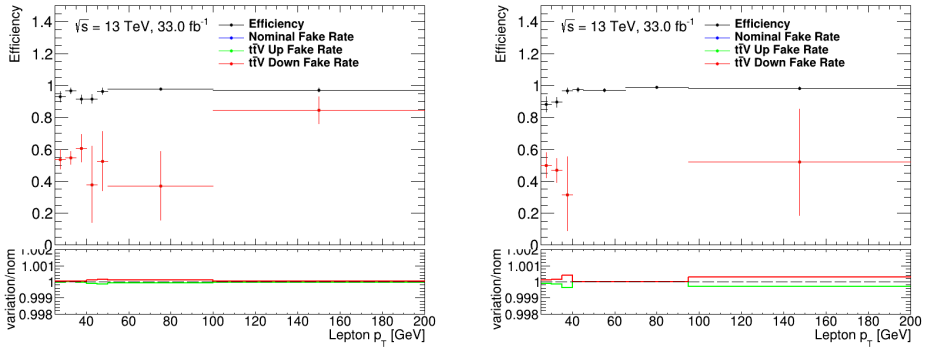


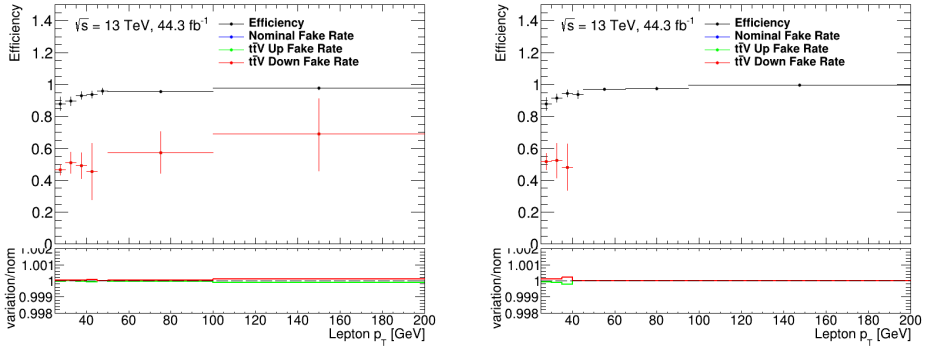
FIGURE H.19: The upper panels show the data and total SM contribution in the fake rates control region in the transverse momentum distribution in the tight electron (left) and muon (right) channels for 2015 (a, b), 2016 (c, d) and 2017 (e, f), respectively, when varying the $t\bar{t}V$ SM processes up and down by their cross-section uncertainties. The black points show the data and the blue line shows the nominal model contribution, while the green and red show the model contributions after the variations. The hashed area shows the statistical, cross-section and luminosity uncertainties on the model contributions. The lower panel plot shows the ratio between the model after the variations and the nominal model.



(a) (b)

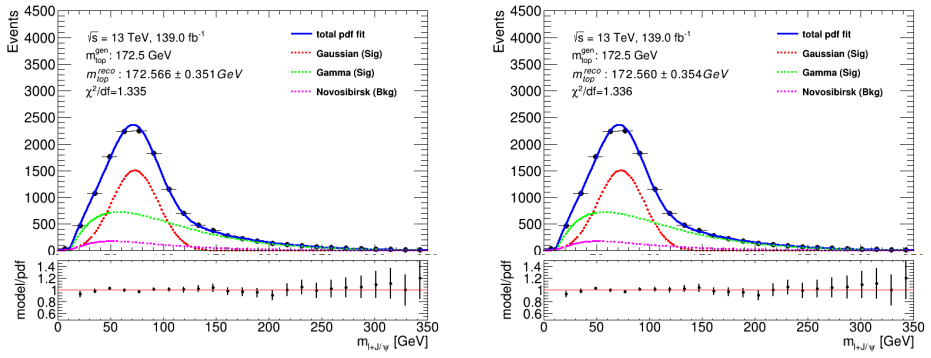


(c) (d)



(e) (f)

FIGURE H.20: The upper panels show the nominal efficiencies and fake rates, and after the cross-section uncertainty variations have been applied to the $t\bar{t}V$ SM processes in the transverse momentum distribution, in the electron (a) and muon (b) channels for 2015 (a, b), 2016 (c, d) and 2017 (e, f), respectively. The black and blue points show the nominal efficiencies and fake rates, while the green and red show the up and down variations, respectively. The error bars show the statistical uncertainty. The lower panel plot shows the ratio between the fake rate after variation and the nominal.



(a) (b)

FIGURE H.21: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution after varying the $t\bar{t}V$ SM processes by their cross-section uncertainties, up (a) and down (b), respectively. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

H.2.5 $J/\psi + W$

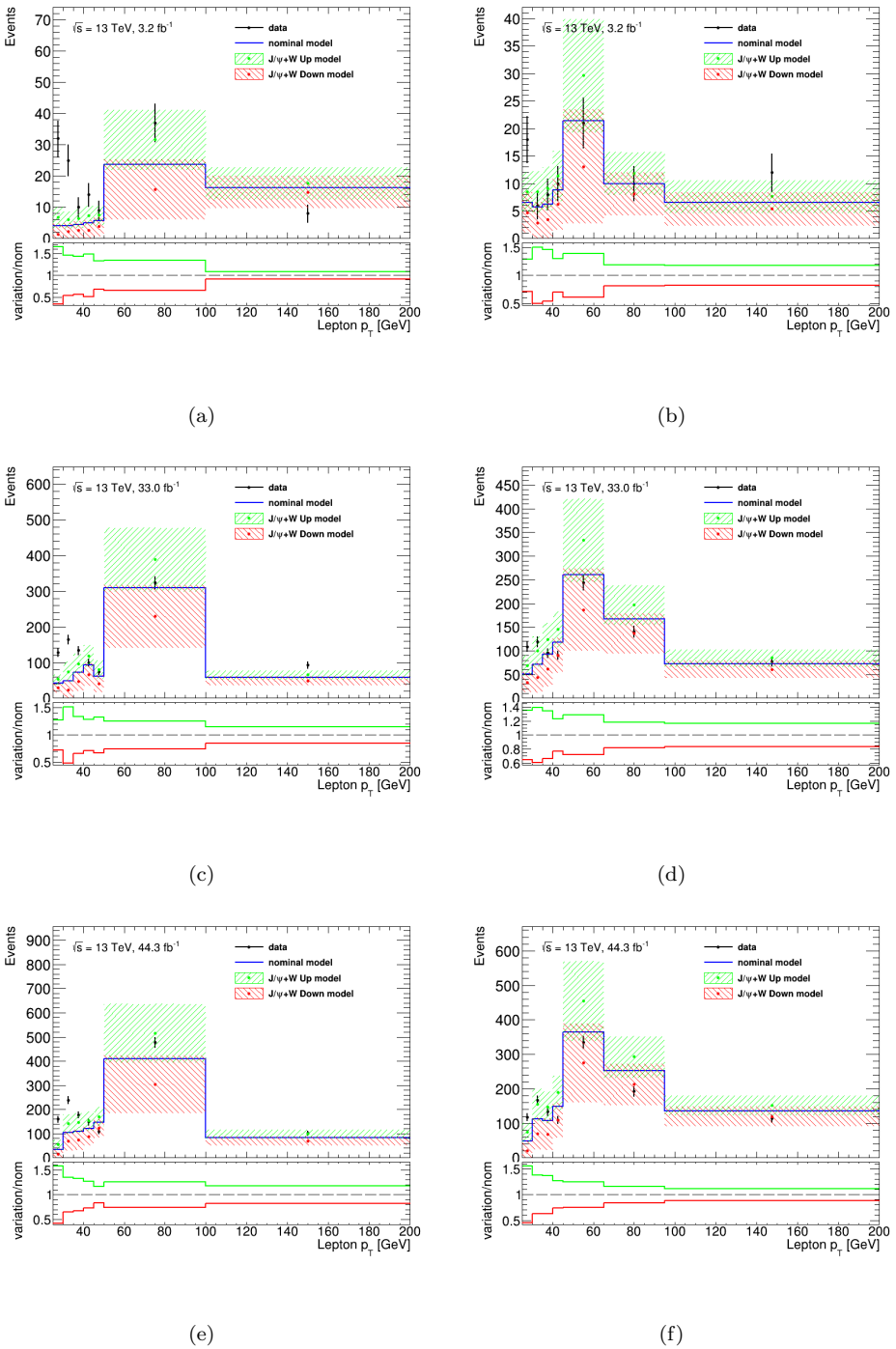
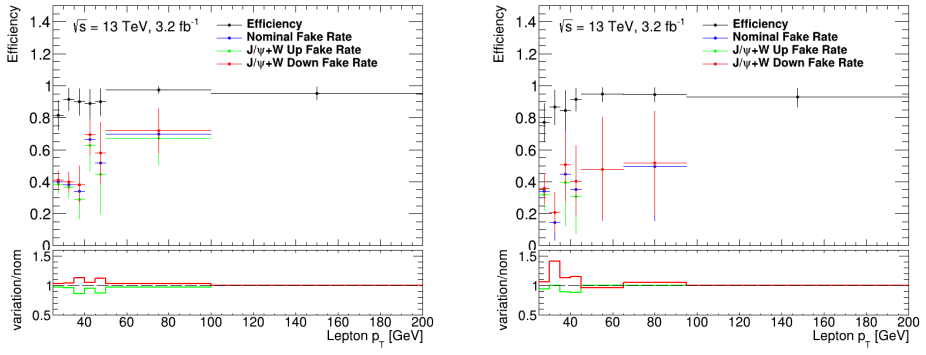
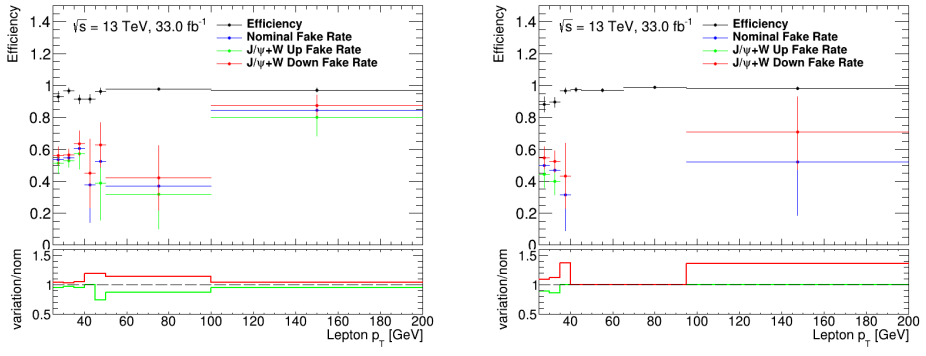


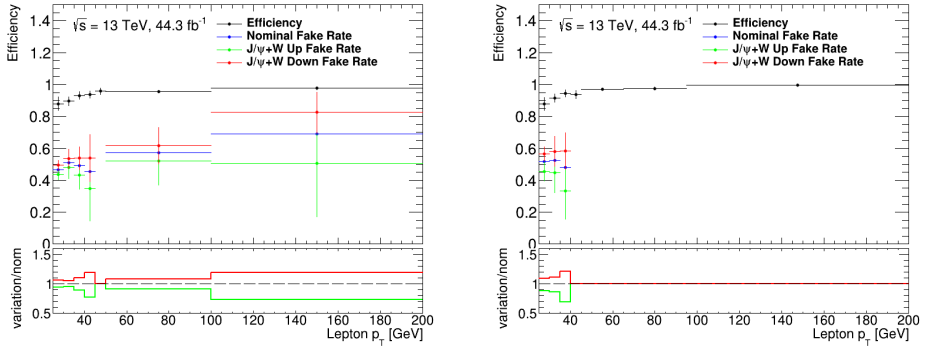
FIGURE H.22: The upper panels show the data and total SM contribution in the fake rates control region in the transverse momentum distribution in the tight electron (left) and muon (right) channels for 2015 (a, b), 2016 (c, d) and 2017 (e, f), respectively, when varying the $J/\psi + W$ SM processes up and down by their cross-section uncertainties. The black points show the data and the blue line shows the nominal model contribution, while the green and red show the model contributions after the variations. The hashed area shows the statistical, cross-section and luminosity uncertainties on the model contributions. The lower panel plot shows the ratio between the model after the variations and the nominal model.



(a) (b)

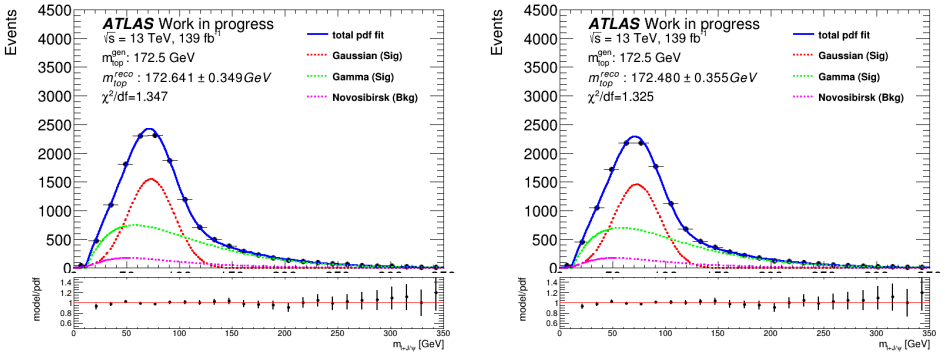


(c) (d)



(e) (f)

FIGURE H.23: The upper panels show the nominal efficiencies and fake rates, and after the cross-section uncertainty variations have been applied to the $J/\psi + W$ SM processes in the transverse momentum distribution, in the electron (left) and muon (right) channels for 2015 (a, b), 2016 (c, d) and 2017 (e, f), respectively. The black and blue points show the nominal efficiencies and fake rates, while the green and red show the up and down variations, respectively. The error bars show the statistical uncertainty. The lower panel plot shows the ratio between the fake rate after variation and the nominal.



(a) (b)

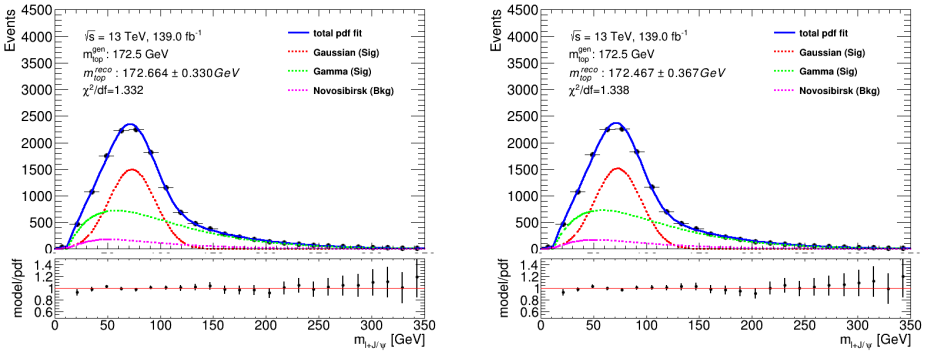
FIGURE H.24: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution after varying the $J/\psi + W$ SM processes by their cross-section uncertainties, up (a) and down (b), respectively. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

Appendix I

Experimental systematic fit results

The following chapter shows the fit results for the different experimental systematic uncertainties. Figures I.1-I.34 shows the fit results for the different experimental systematic uncertainties with the reconstructed top quark mass and its statistical uncertainty, as well as the fit quality in the top left corner of the upper panels. The lower panels consist of the ratio between the model and the pdf. The blue line shows the total pdf while the red, green and purple lines show the Gaussian, Gamma and Novosibirsk contributions, respectively. Each experimental systematic fit shows a good fit with a $\chi^2/\mathrm{df} \sim 1$.

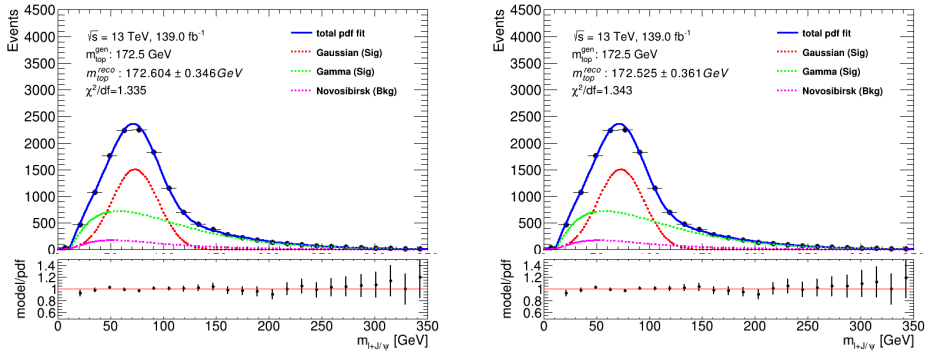
I.1 Pile-up



(a) (b)

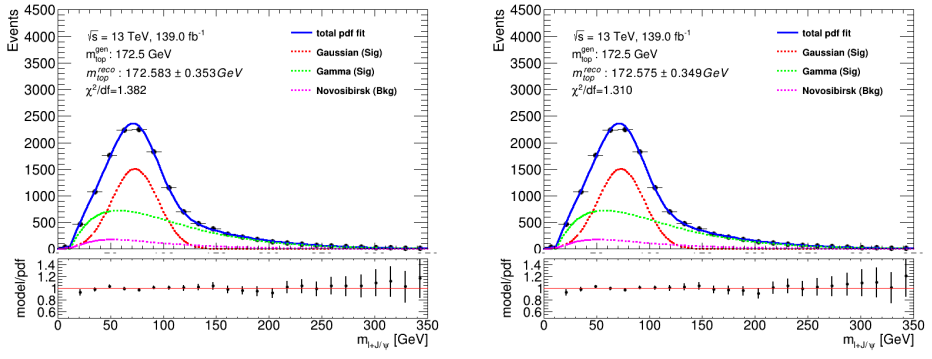
FIGURE I.1: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the pile-up reweighting factors up (a) and down (b). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

I.2 Leptons



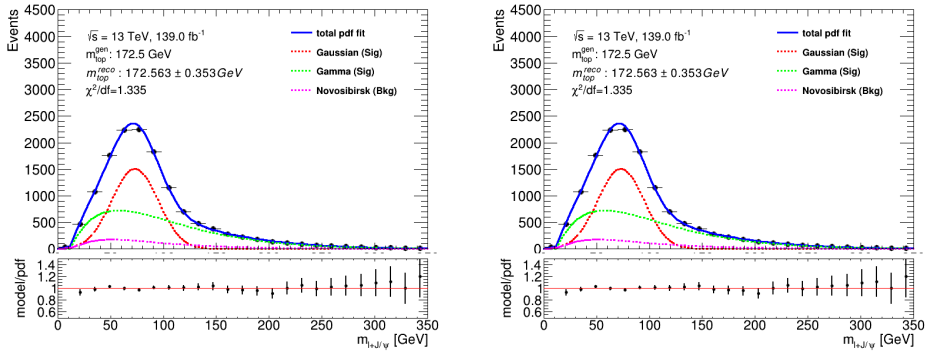
(a)

(b)



(c)

(d)



(e)

(f)

FIGURE I.2: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the electron's momentum scale up (a) and down (b), momentum resolution up (c) and down (d), and the trigger scale factors up (e) and down (f). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

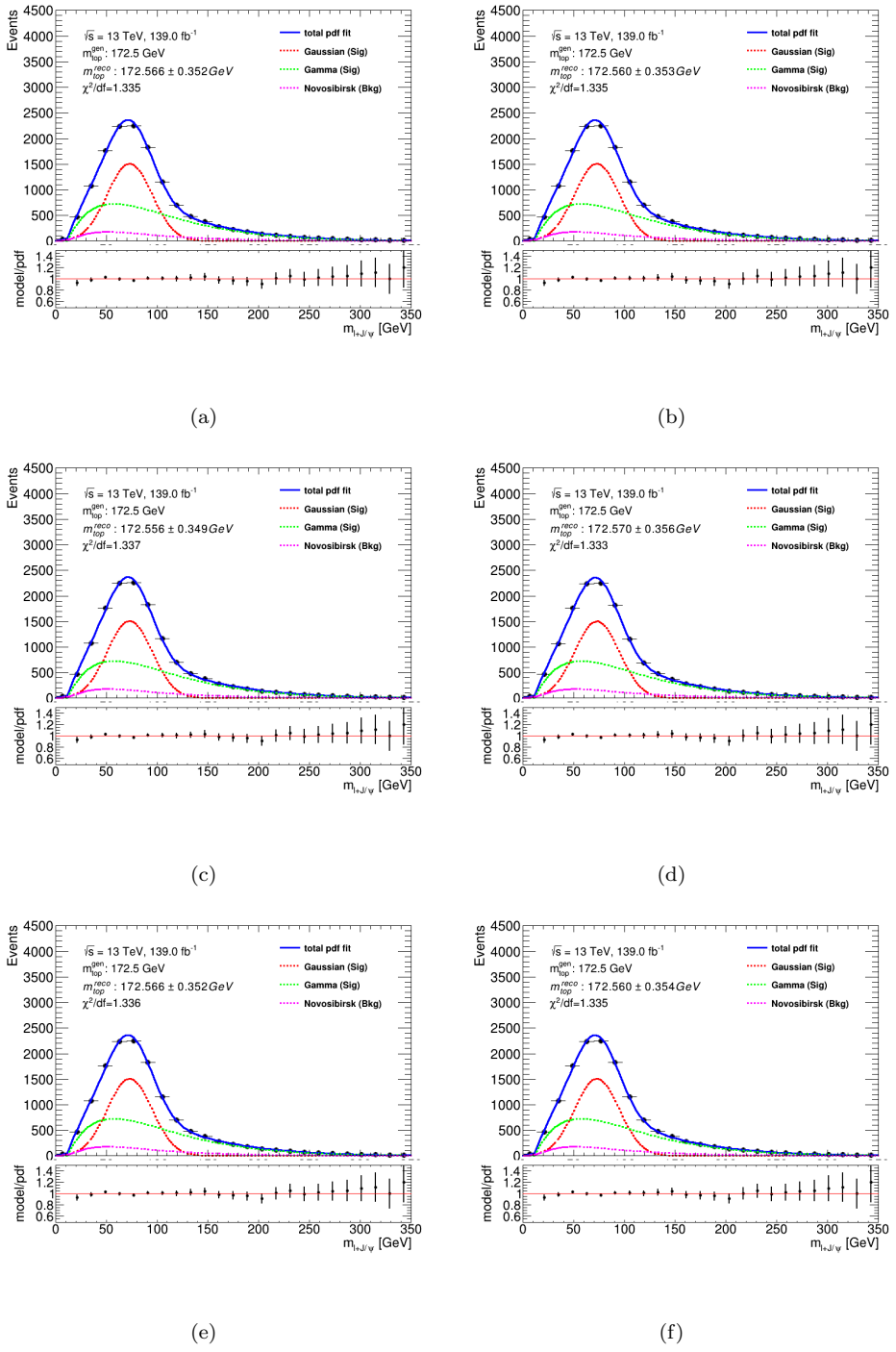


FIGURE I.3: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the electron's reconstruction scale factors up (a) and down (b), identification scale factors up (c) and down (d), and isolation scale factors up (e) and down (f). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

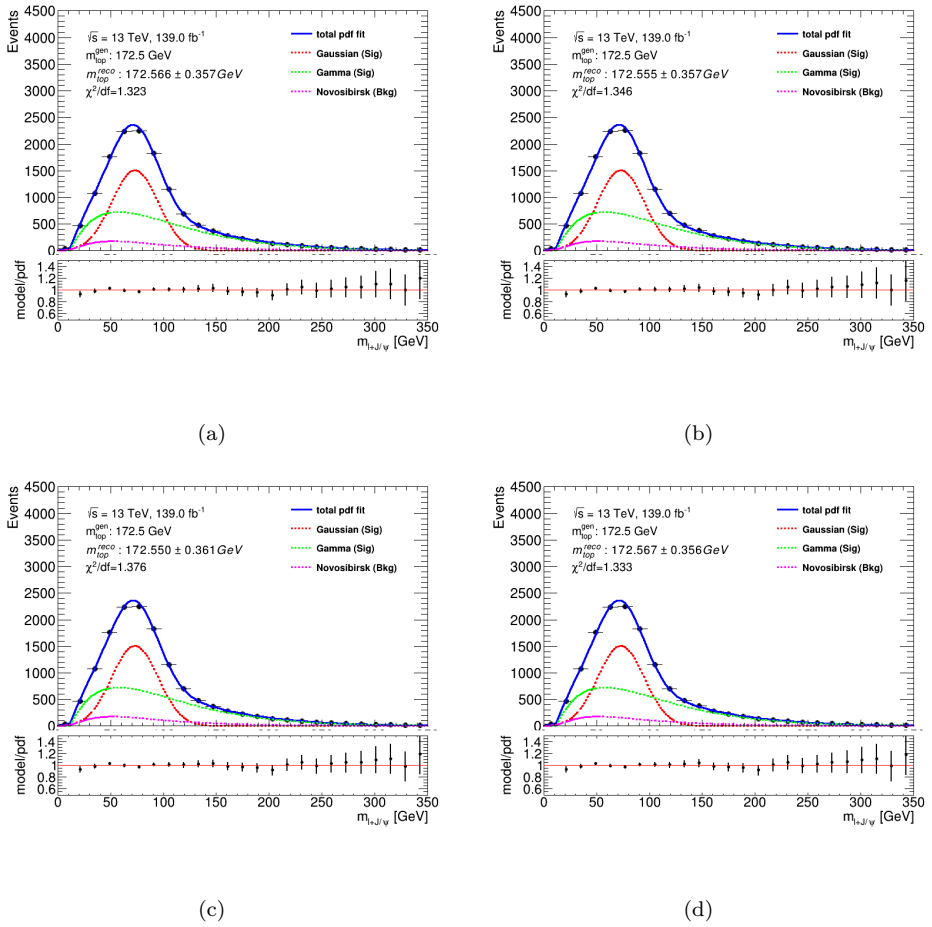


FIGURE I.4: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the muon's momentum scale up (a) and down (b) and charge-dependent momentum scale up (c) and down (d). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

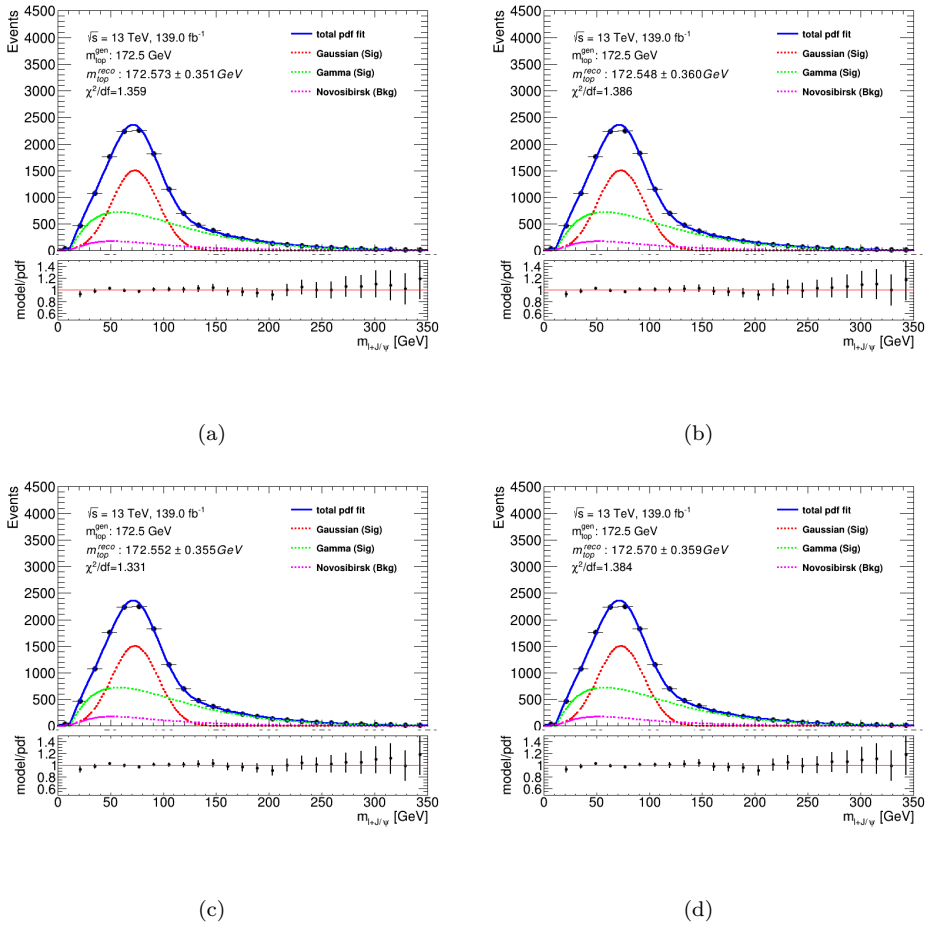
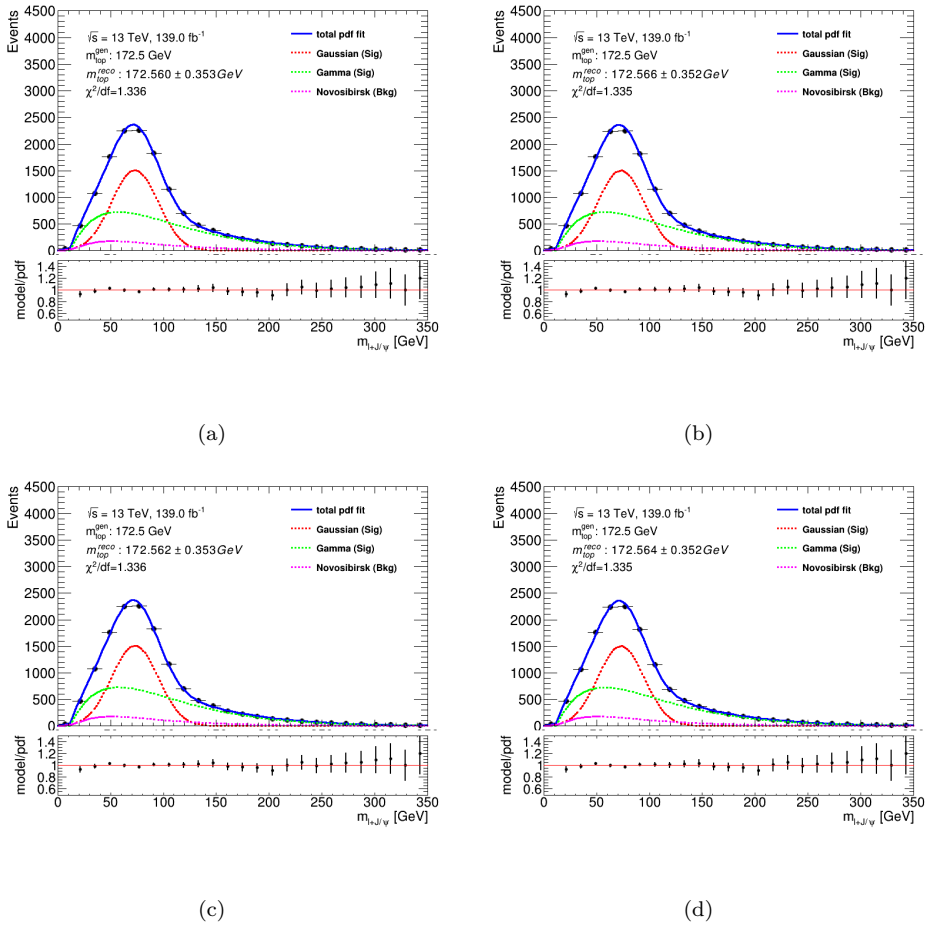


FIGURE I.5: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the muon's momentum inner detector resolution up (a) and down (b) and momentum muon spectrometer resolution up (c) and down (d). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.



(a) (b)

(c) (d)

FIGURE I.6: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the muon's trigger scale factors up (a) and down (b) by its statistical and up (c) and down (d) by its systematic uncertainties. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

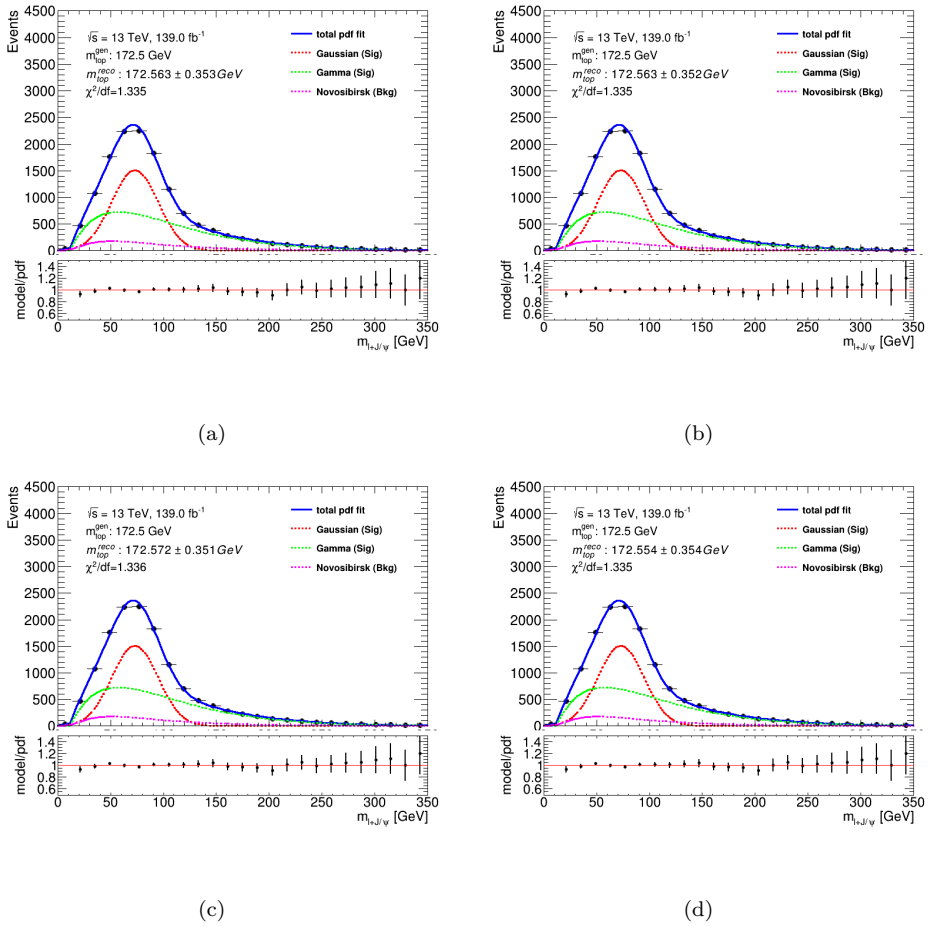


FIGURE I.7: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the muon's identification scale factors up (a) and down (b) by its statistical and up (c) and down (d) by its systematic uncertainties. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

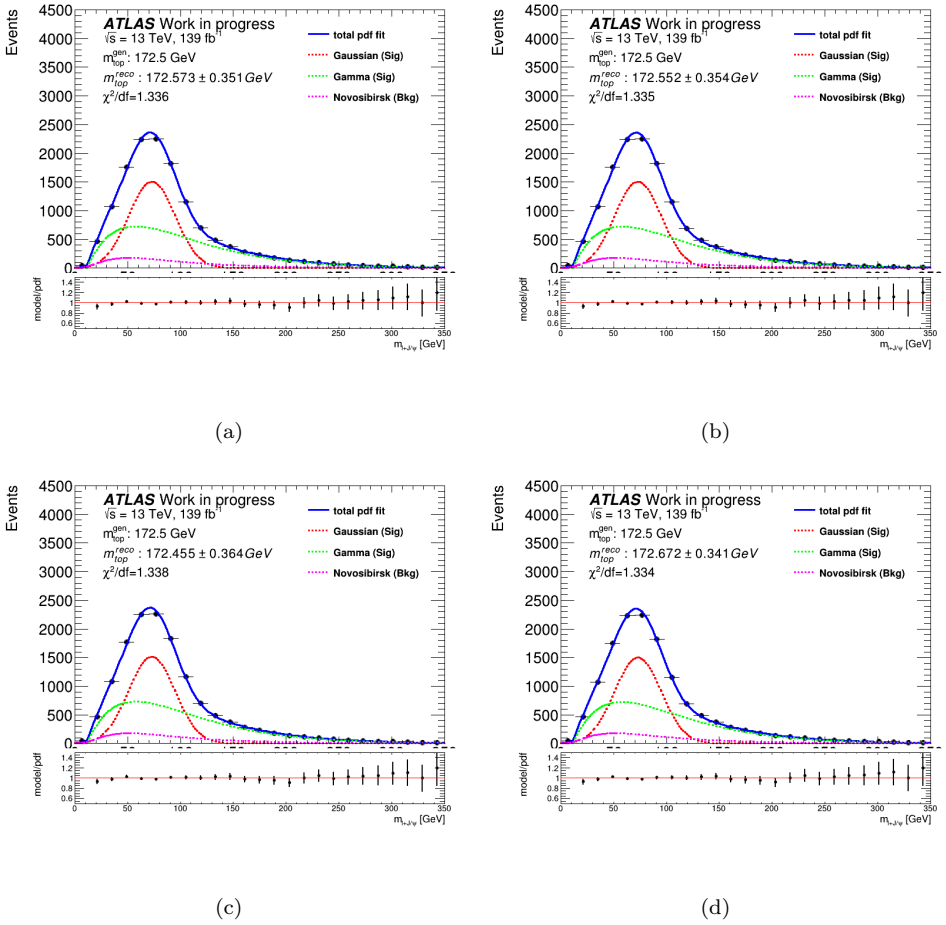


FIGURE I.8: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the muon's isolation scale factors up (a) and down (b) by its statistical and up (c) and down (d) by its systematic uncertainties. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

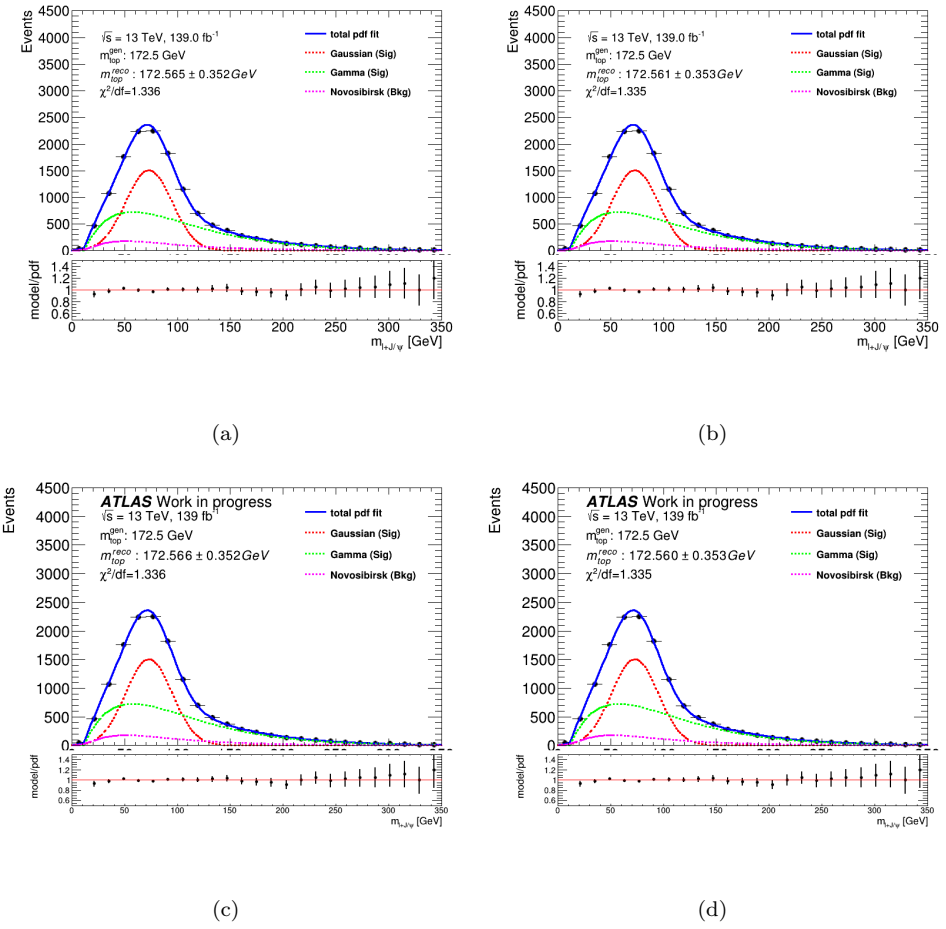


FIGURE I.9: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the muon's track-to-vertex association scale factors up (a) and down (b) by its statistical and up (c) and down (d) by its systematic uncertainties. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

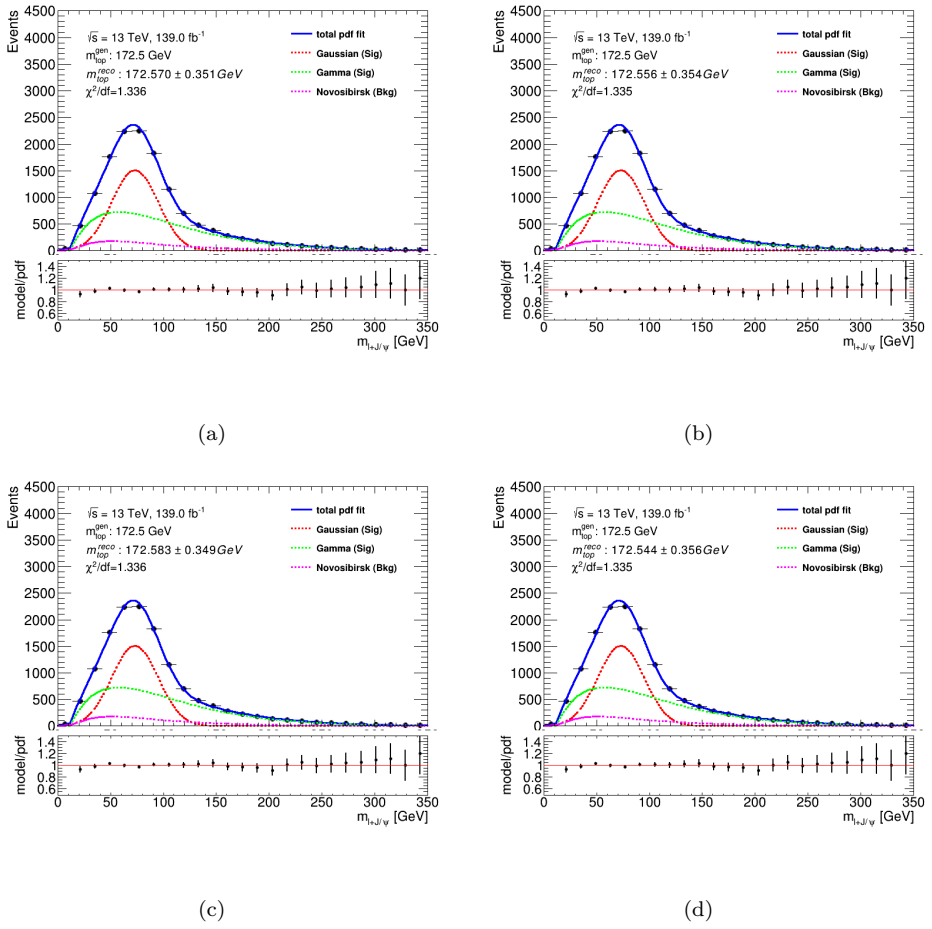


FIGURE I.10: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the muon's identification scale factors up (a) and down (b) by its statistical and up (c) and down (d) by its systematic uncertainties, for muons with $p_T > 15$ GeV. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

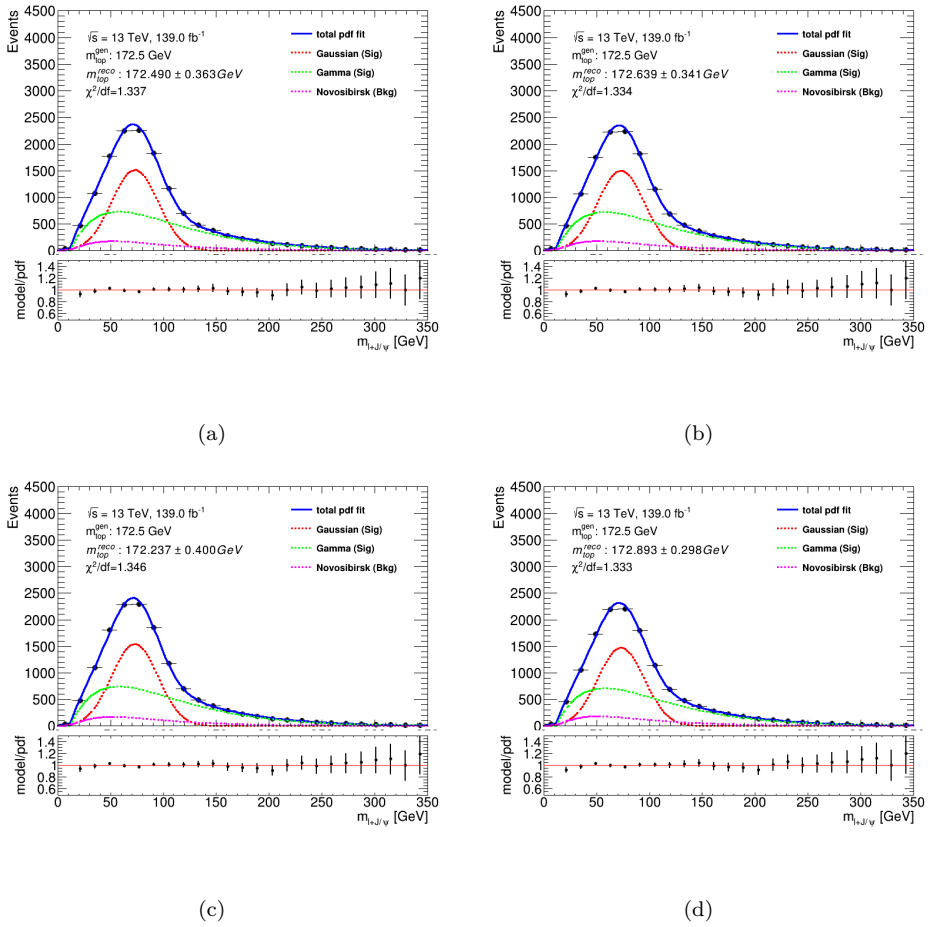


FIGURE I.11: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the muon's identification scale factors up (a) and down (b) by its statistical and up (c) and down (d) by its systematic uncertainties, for muons with $p_T < 15$ GeV. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

I.3 Jet energy scale (JES)

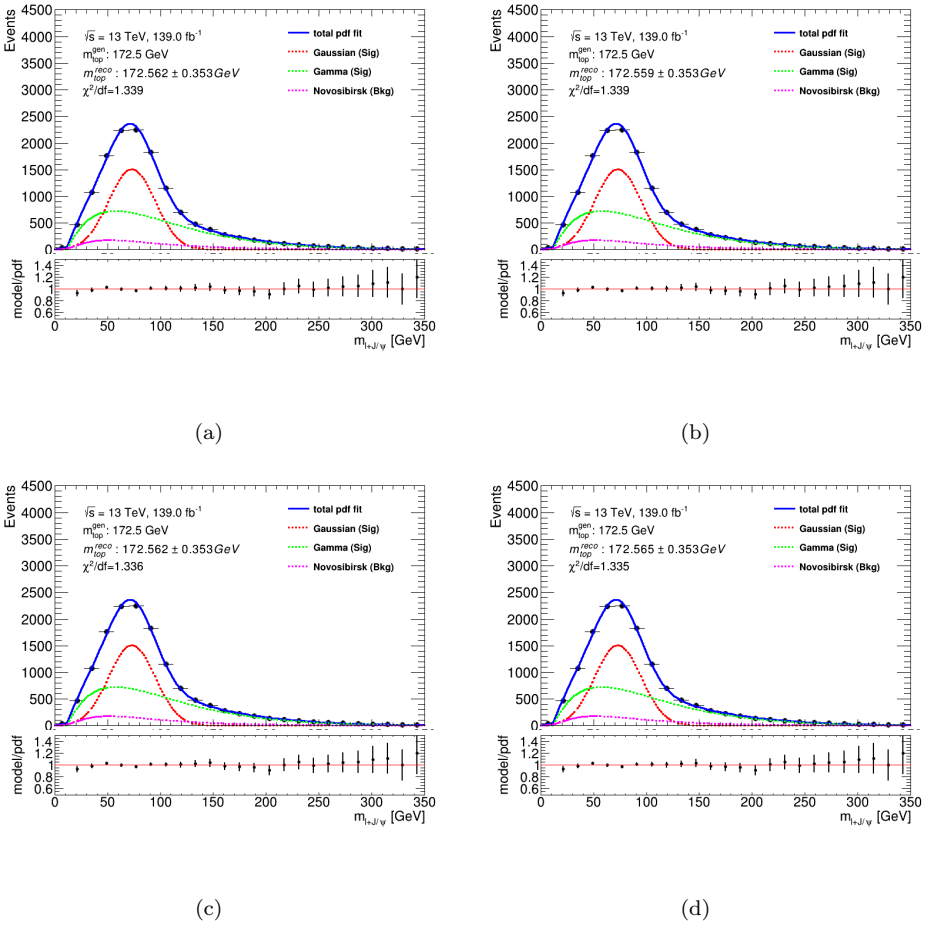
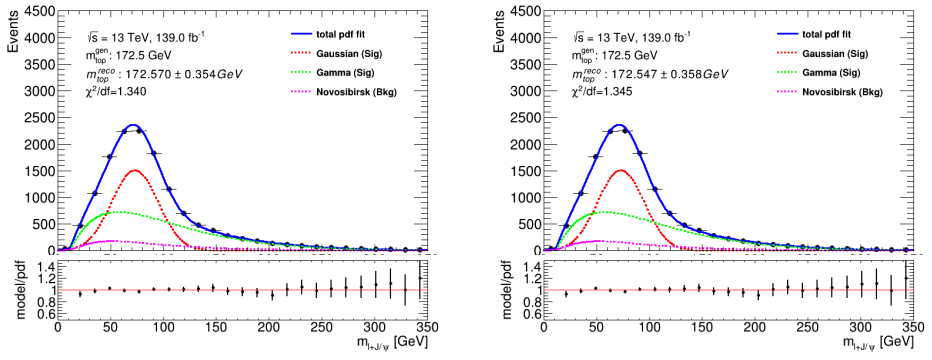
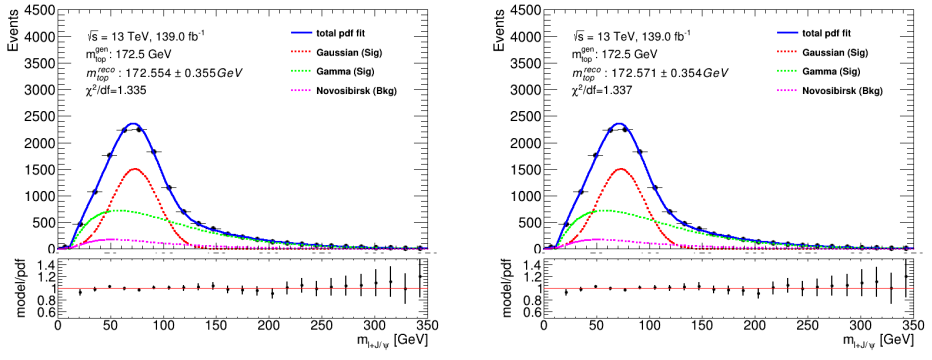


FIGURE I.12: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the *in-situ* corrections' detector parameter 1 up (a) and down (b) and 2 up (c) and down (d). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.



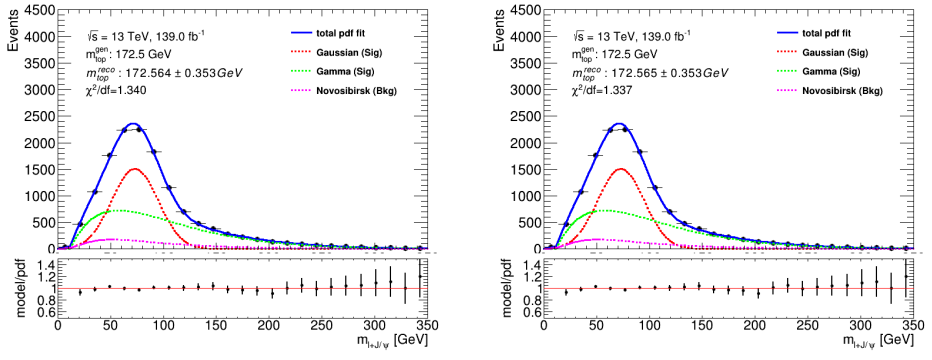
(a)

(b)



(c)

(d)



(e)

(f)

FIGURE I.13: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the *in-situ* corrections' mixed parameter 1 up (a) and down (b), 2 up (c) and down (d), and 3 up (e) and down (f). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

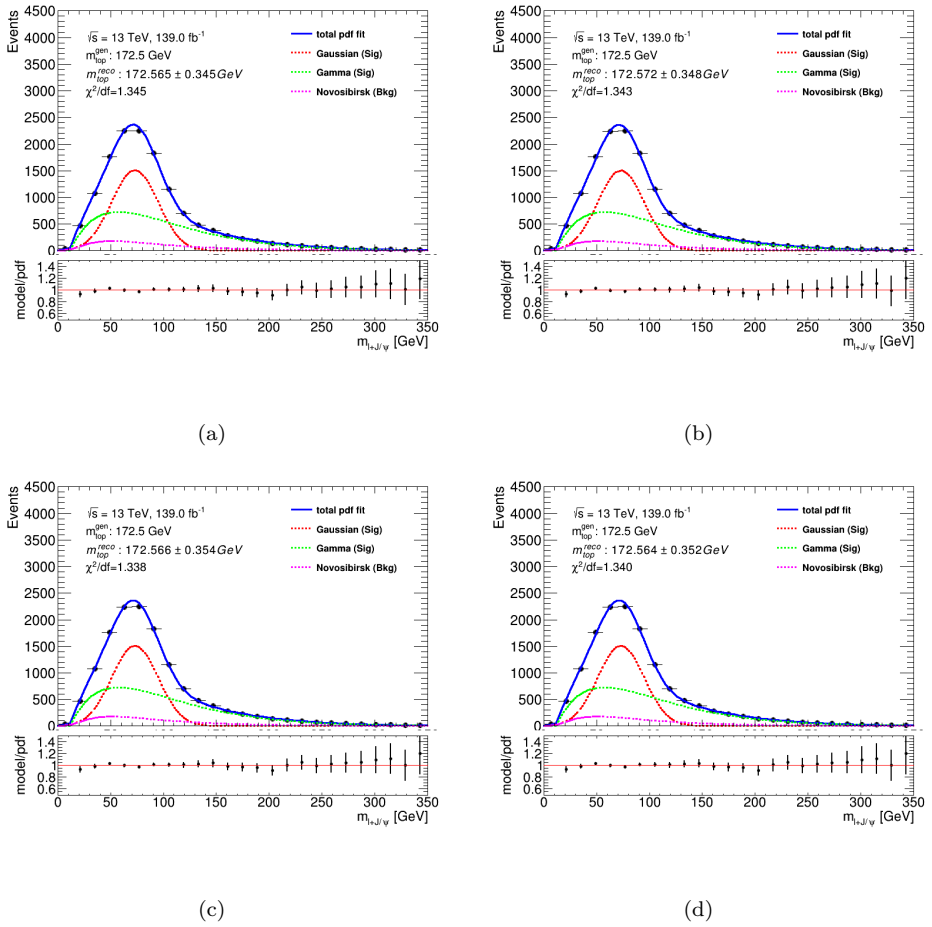


FIGURE I.14: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the *in-situ* corrections' modelling parameter 1 up (a) and down (b) and 2 up (c) and down (d). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

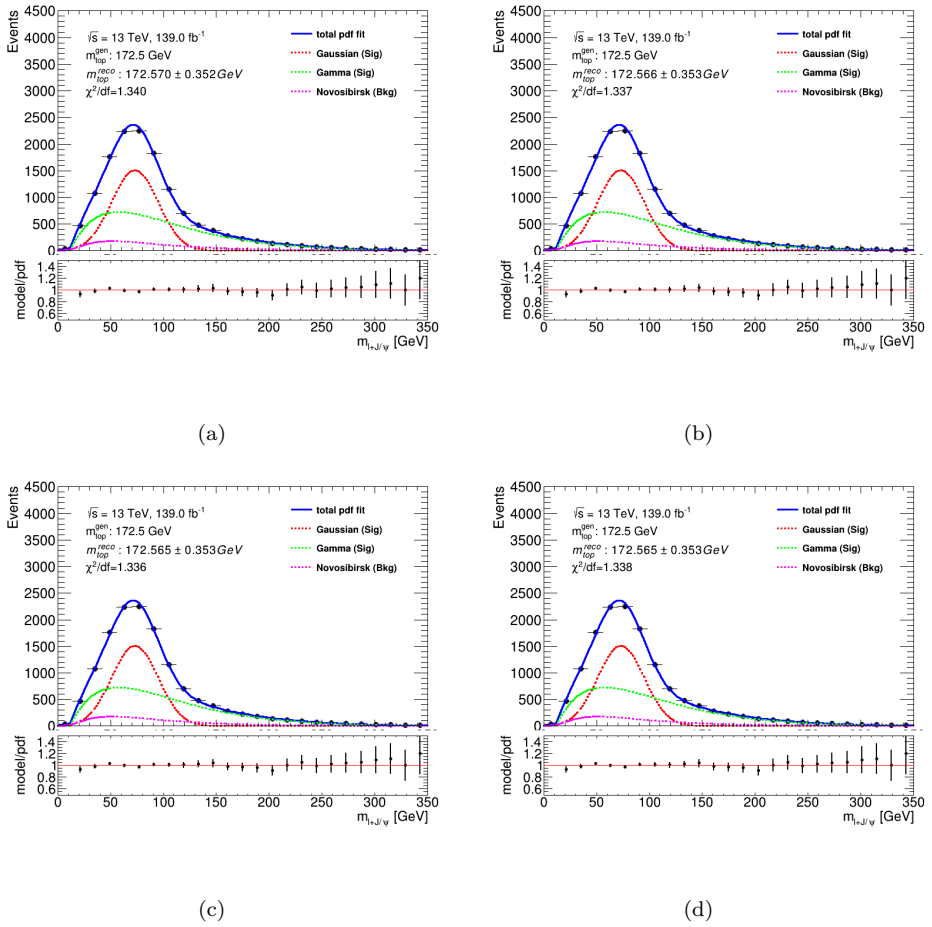
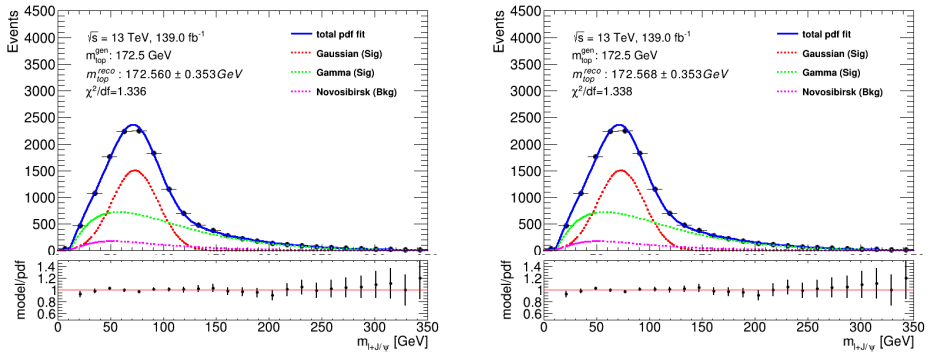
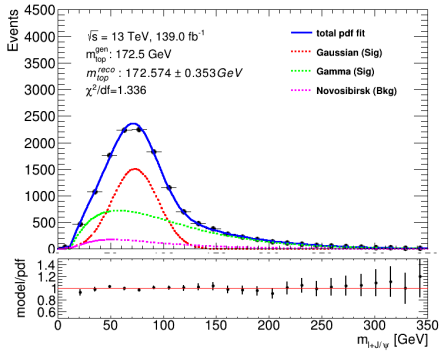


FIGURE I.15: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the *in-situ* corrections' modelling parameter 3 up (a) and down (b) and 4 up (c) and down (d). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

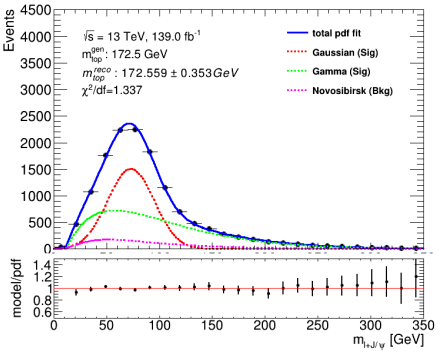


(a)

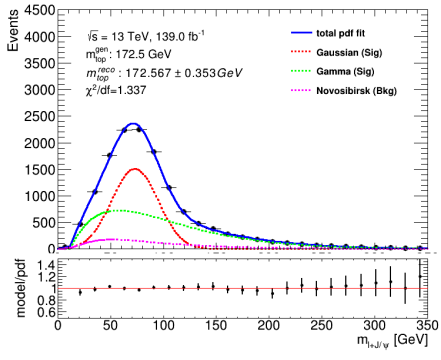
(b)



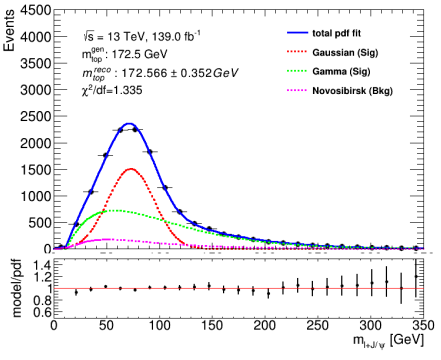
(c)



(d)



(e)



(f)

FIGURE I.16: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the *in-situ* corrections' statistical parameter 1 up (a) and down (b), 2 up (c) and down (d), and 3 up (e) and down (f). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

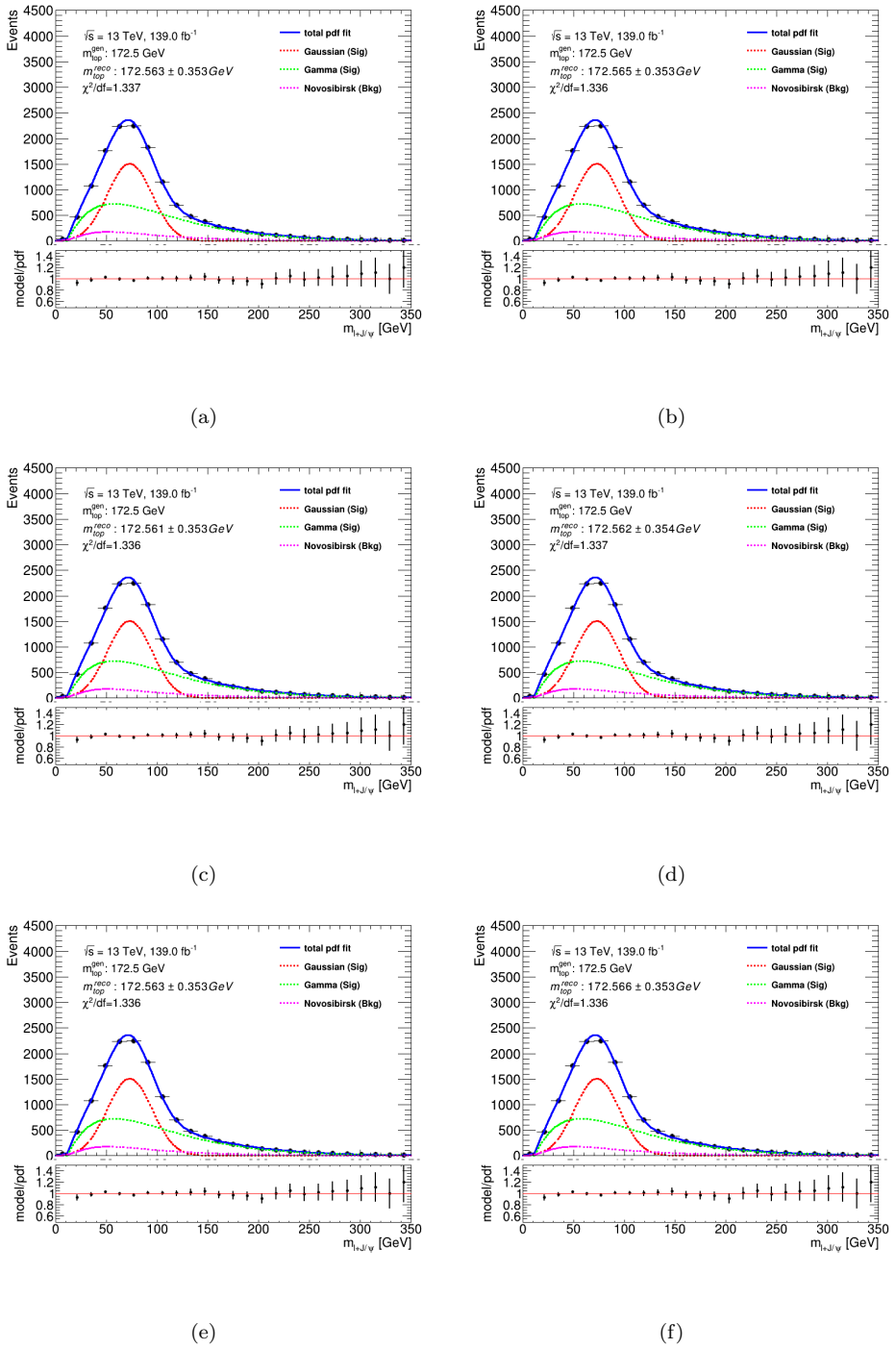


FIGURE I.17: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the *in-situ* corrections' statistical parameter 4 up (a) and down (b), 5 up (c) and down (d), and 6 up (e) and down (f). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

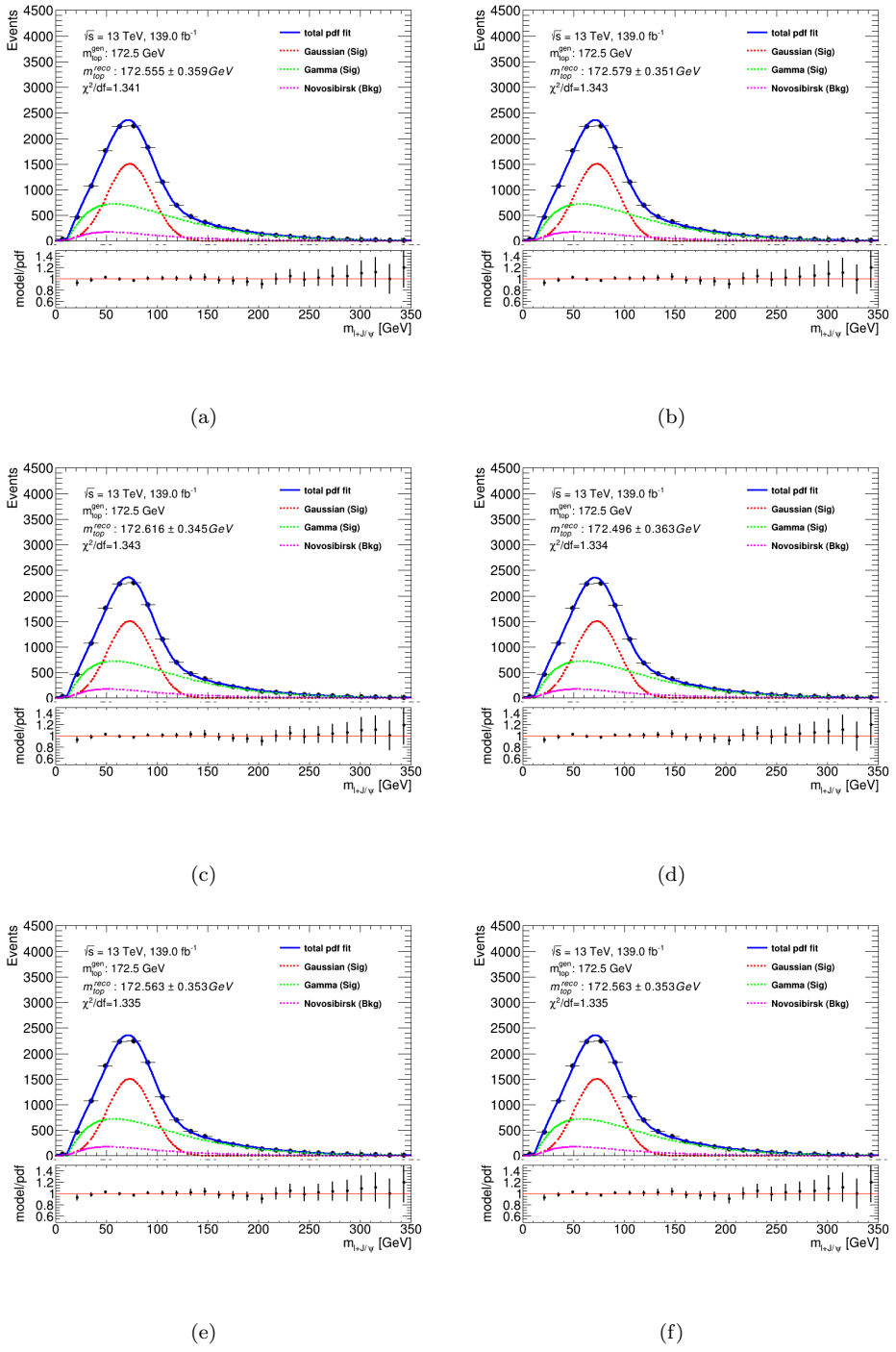


FIGURE I.18: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the η-calibration parameters, i.e. method up (a) and down (b), modelling up (c) and down (d), and high energy non-closure up (e) and down (f). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

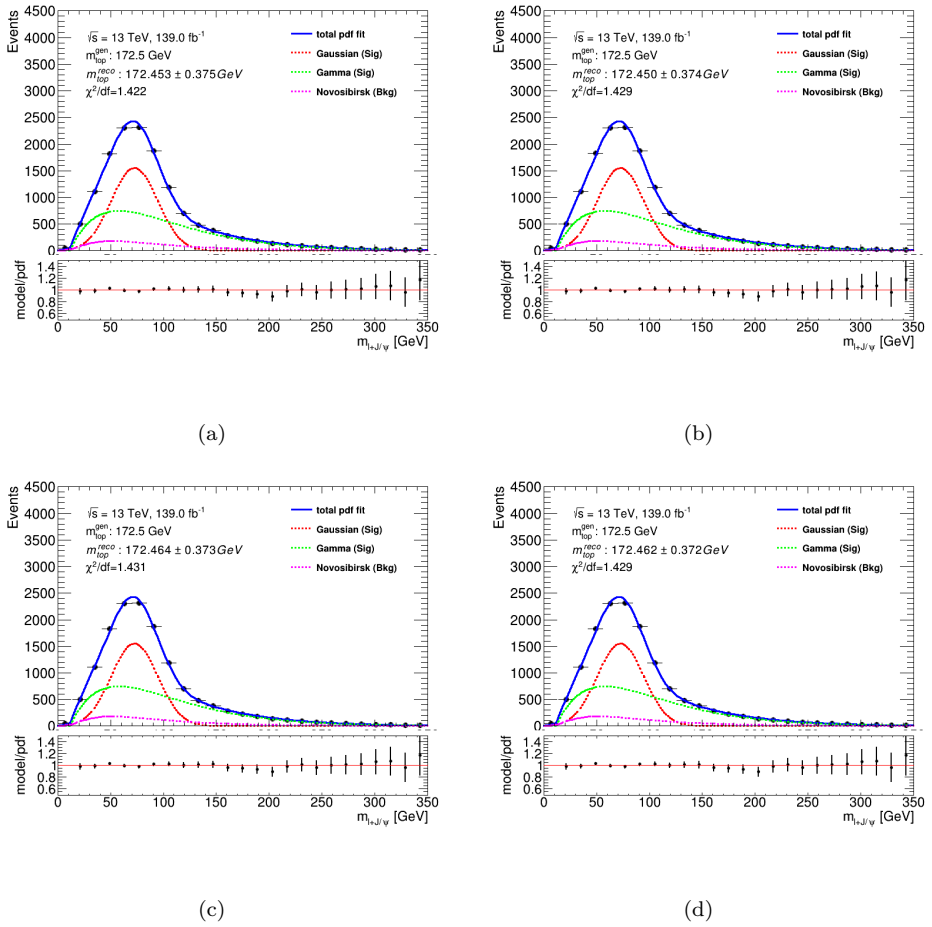


FIGURE I.19: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the η-calibration parameters, i.e. positive η non-closure up (a) and down (b), and negative η non-closure up (c) and down (d). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

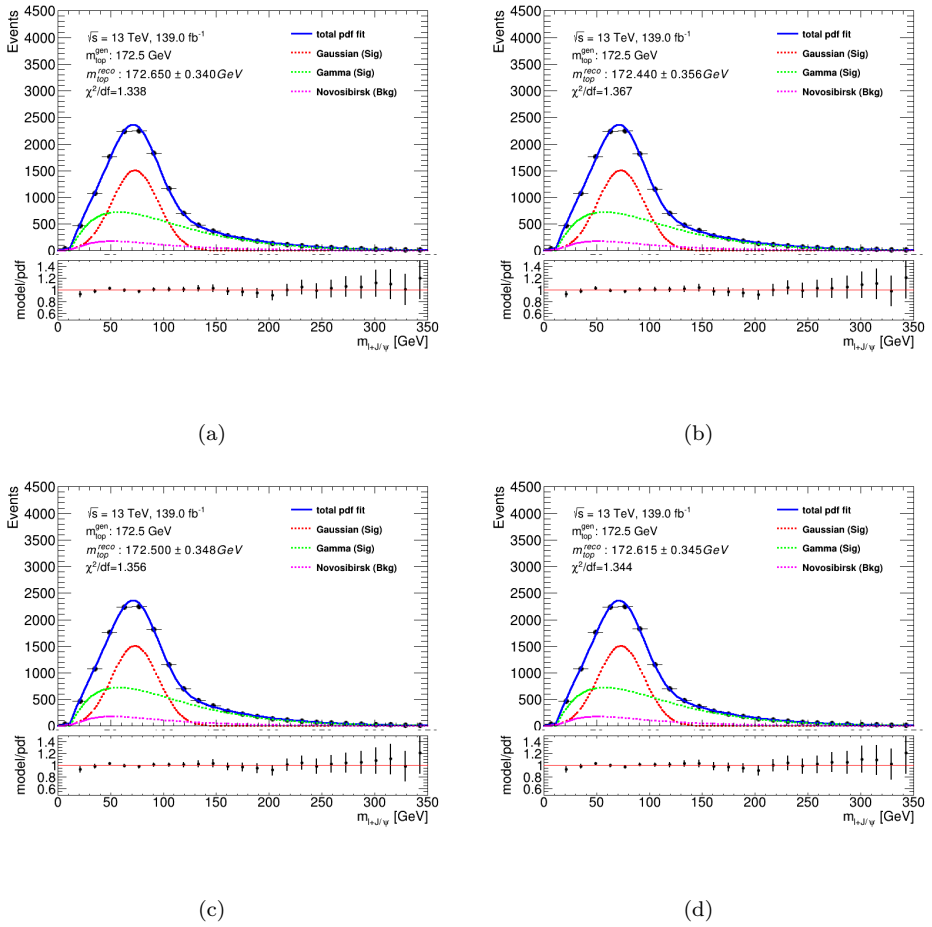


FIGURE I.20: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the jet's flavour composition up (a) and down (b), and flavour response up (c) and down (d). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

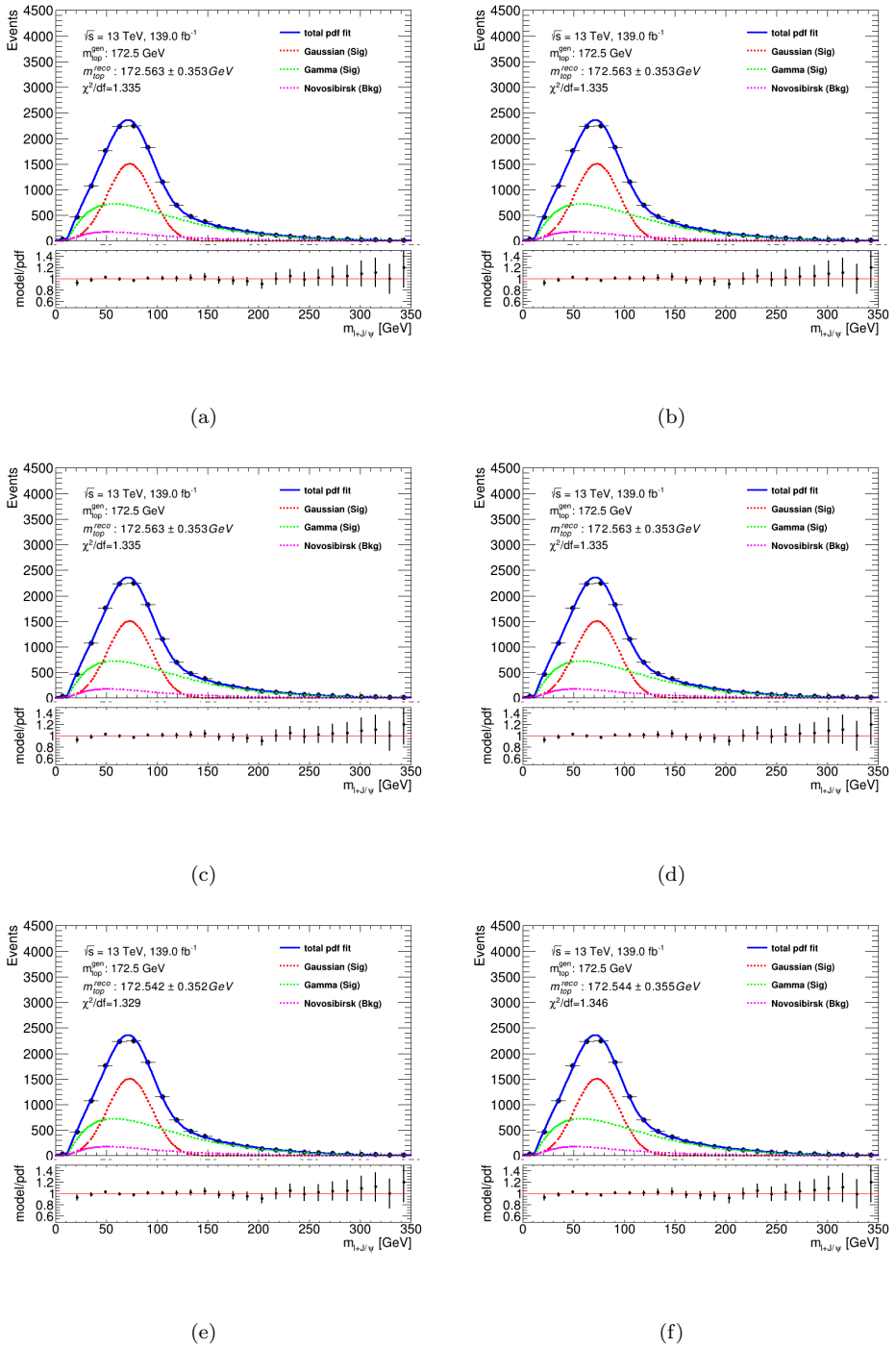
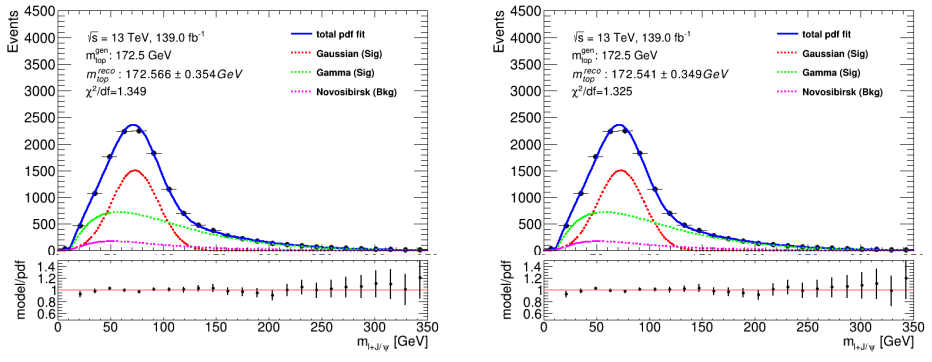
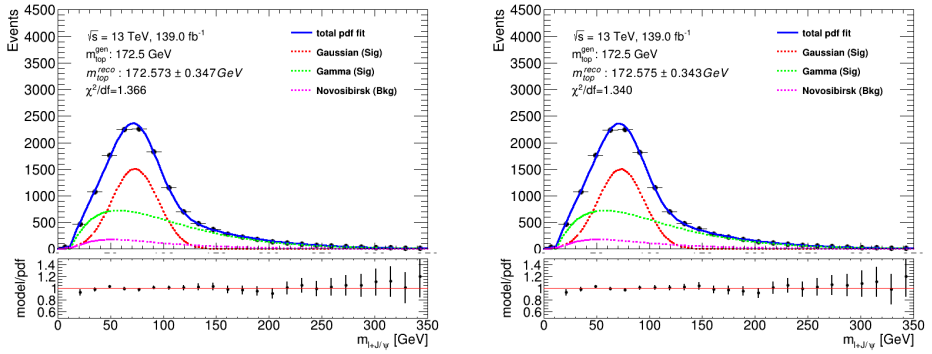


FIGURE I.21: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the high-p_T jets up (a) and down (b), punch-through jets up (c) and down (d), and relative non-closure up (e) and (f). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.



(a)

(b)



(c)

(d)

FIGURE I.22: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the JES pile-up parameters: mean number of interactions up (a) and down (b), and number of primary vertices up (c) and down (d). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

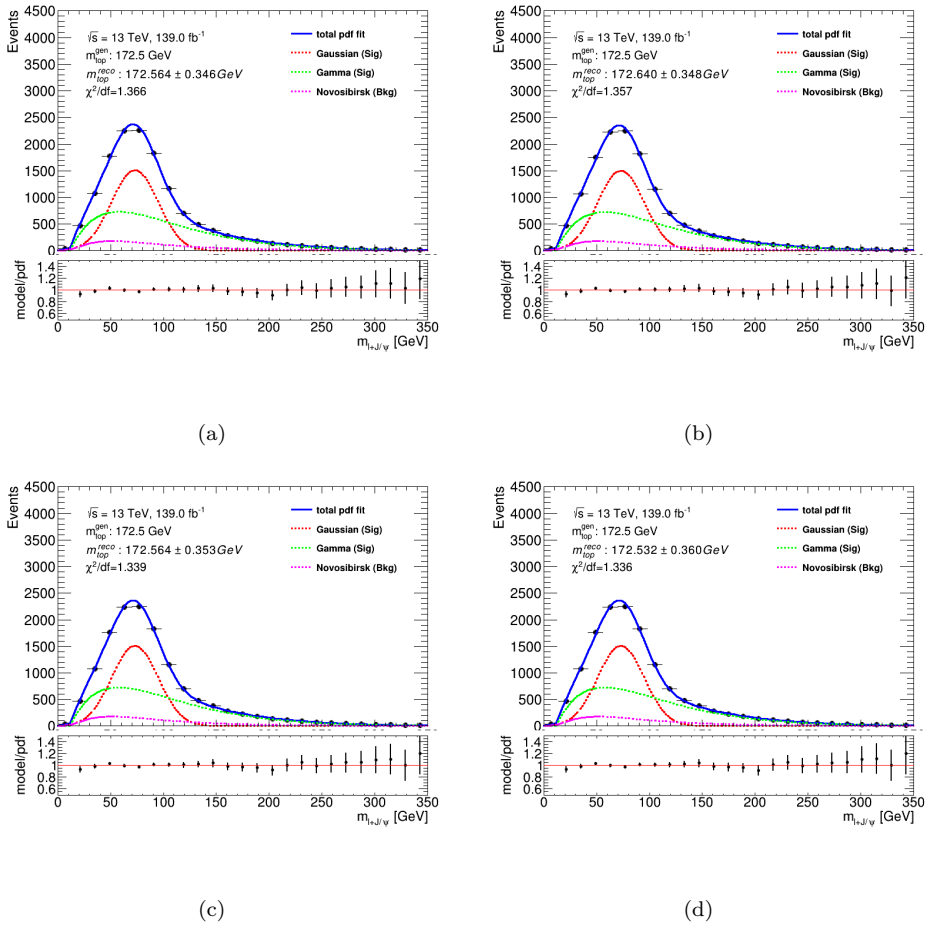
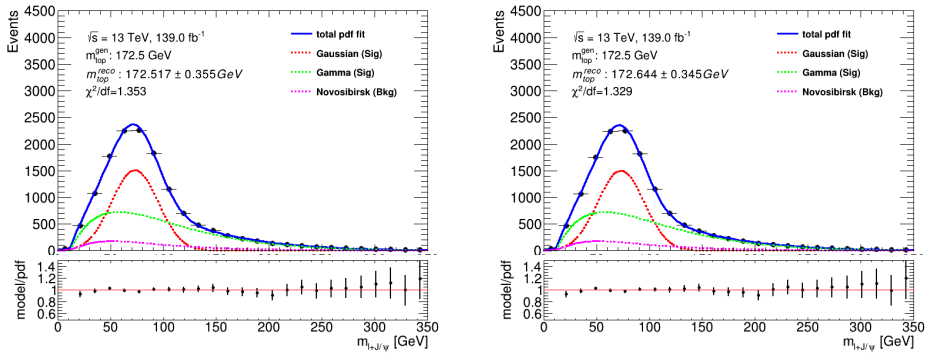


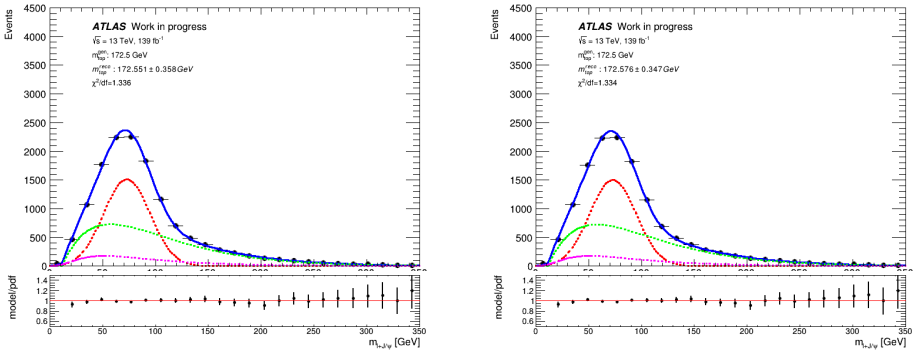
FIGURE I.23: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the JES pile-up parameters: density up (a) and down (b), and p_T term up (c) and down (d). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.



(a) (b)

FIGURE I.24: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the JES for b-tagged jets response up (a) and down (b). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

I.4 Jet-vertex-tagger (JVT)



(a) (b)

FIGURE I.25: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the JVT up (a) and down (b). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

I.5 Jet energy resolution (JER)

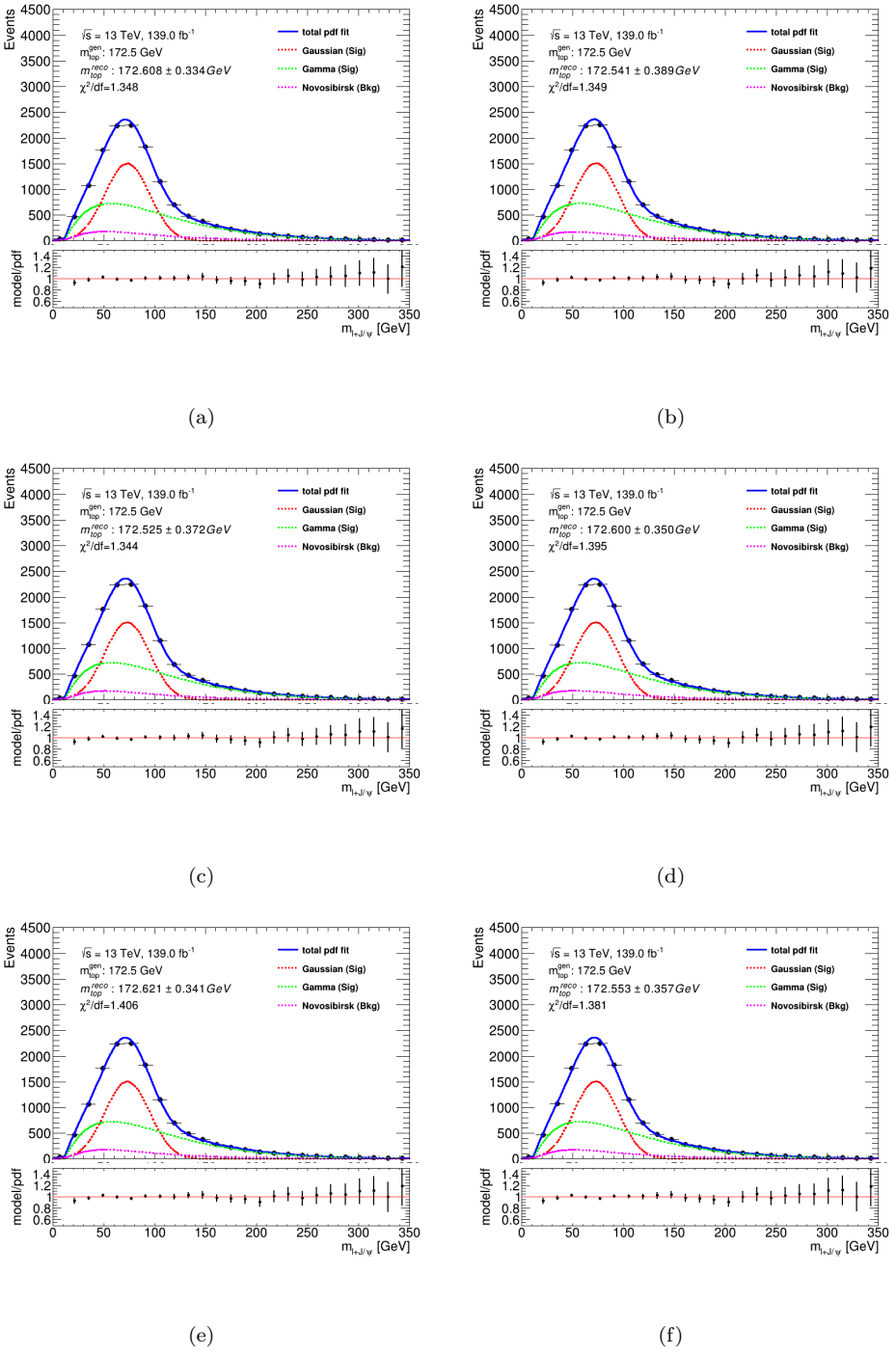
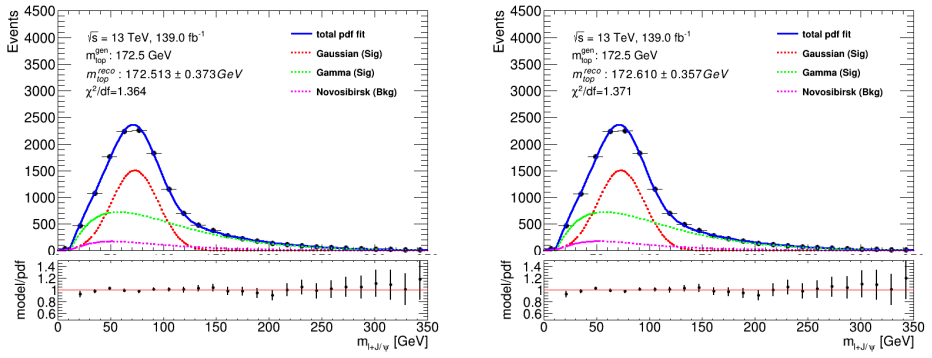


FIGURE I.26: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the JER data versus MC parameter (a), and the JER 1 (b), 2 (c), 3 (d), 4 (e), and 5 (f) nuisance parameters. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.



(a) (b)

FIGURE I.27: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the JER 6 (a) and 7 (b) nuisance parameters. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

I.6 Flavour tagging

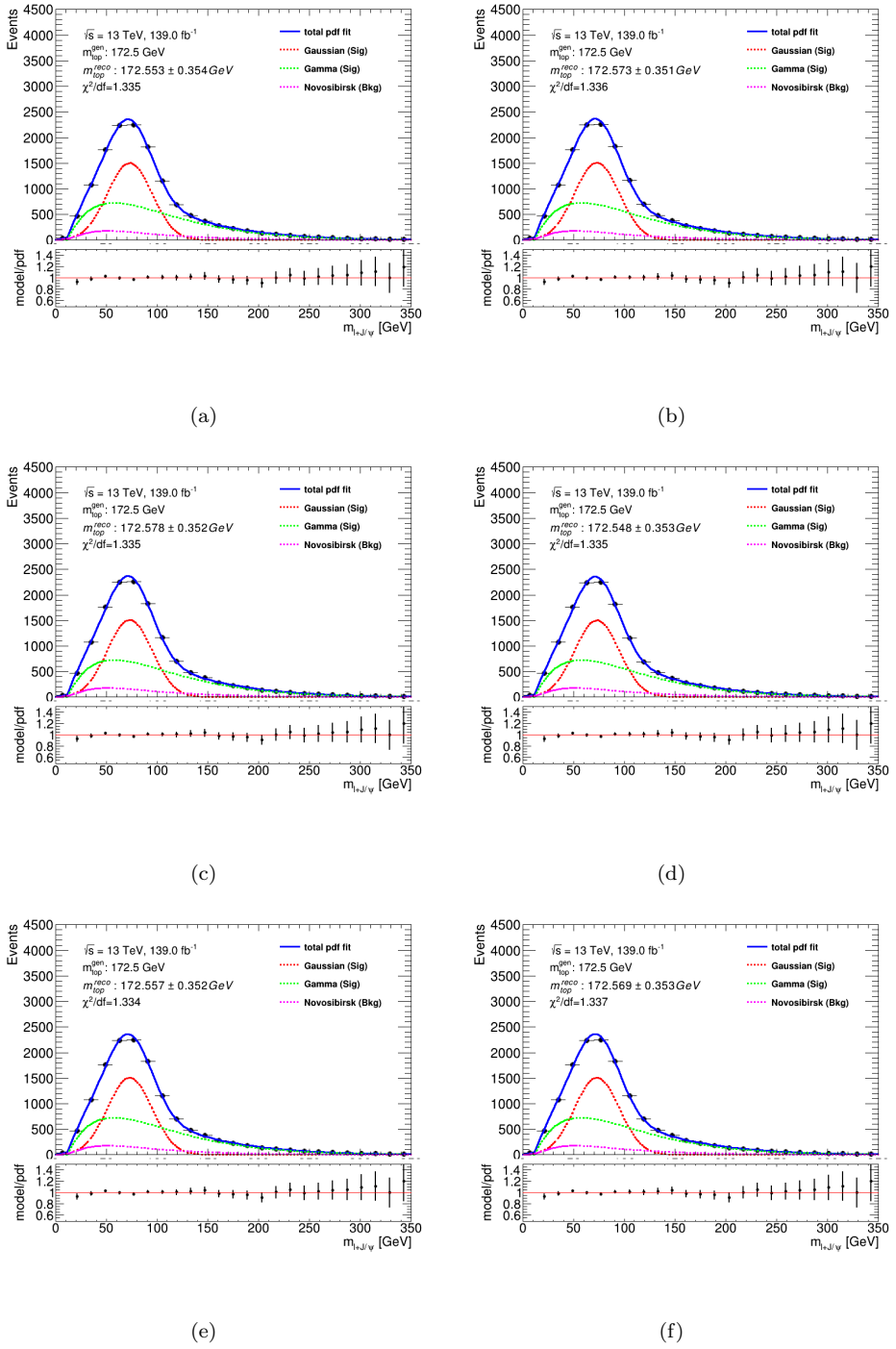


FIGURE I.28: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the flavour tagging calibration eigen-vectors 1 up (a) and down (b), 2 up (c) and down (d), and 3 up (e) and down (f) for b-tagged jets. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

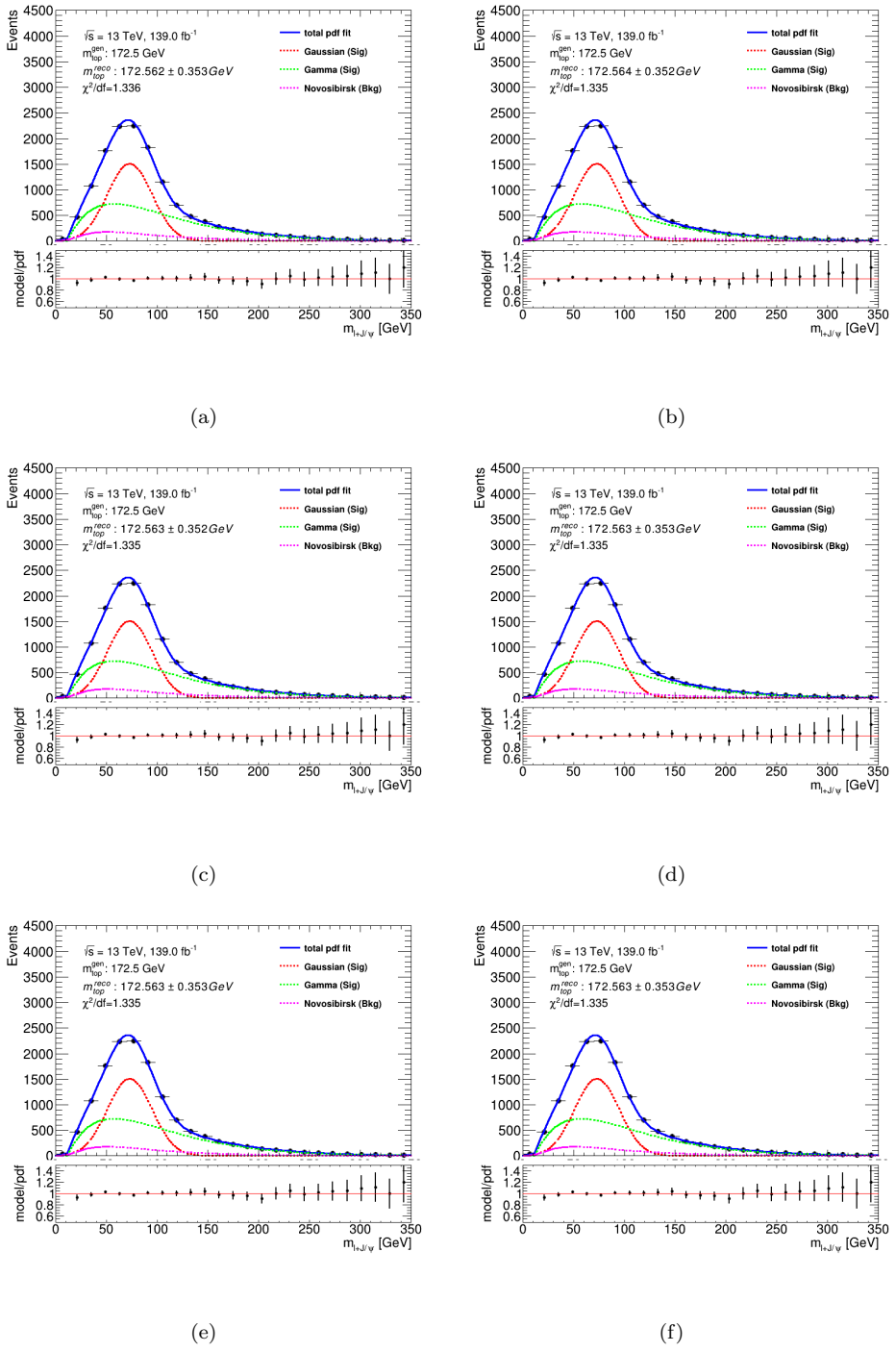
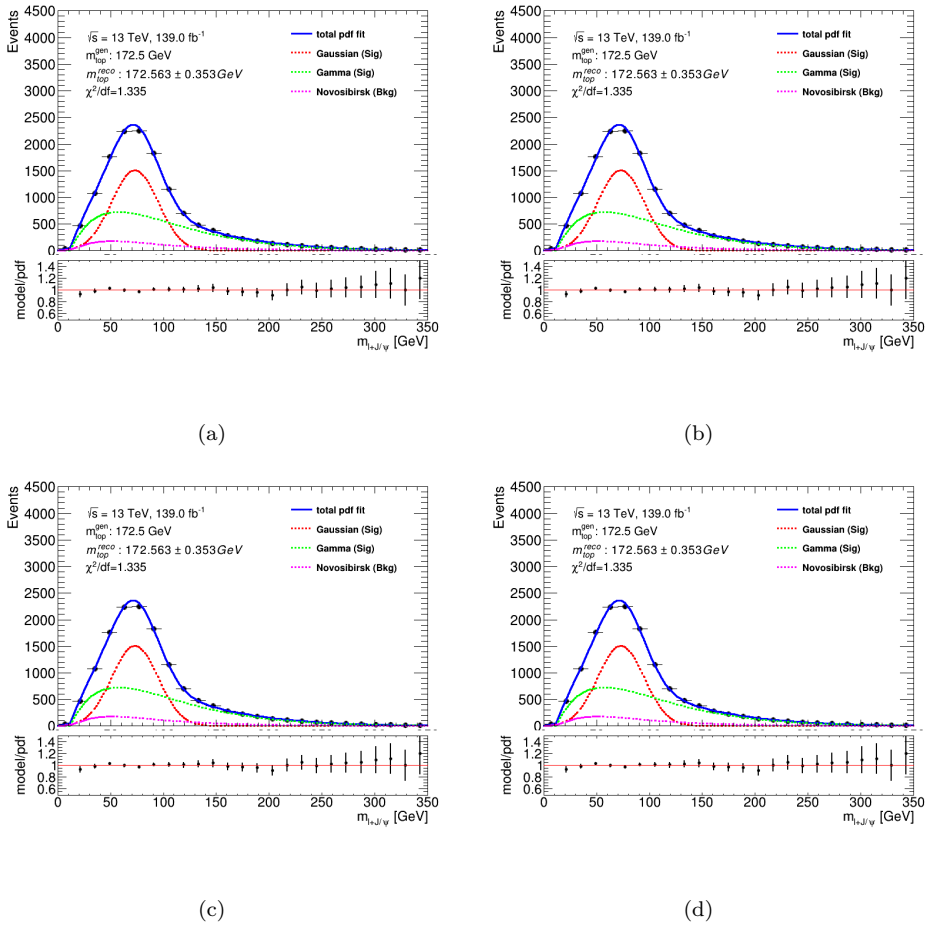


FIGURE I.29: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the flavour tagging calibration eigen-vectors 4 up (a) and down (b), 5 up (c) and down (d), and 6 up (e) and down (f) for b-tagged jets. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.



(a) (b)

(c) (d)

FIGURE I.30: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the flavour tagging calibration eigen-vectors 7 up (a) and down (b), and 8 up (c) and down (d) for b-tagged jets. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

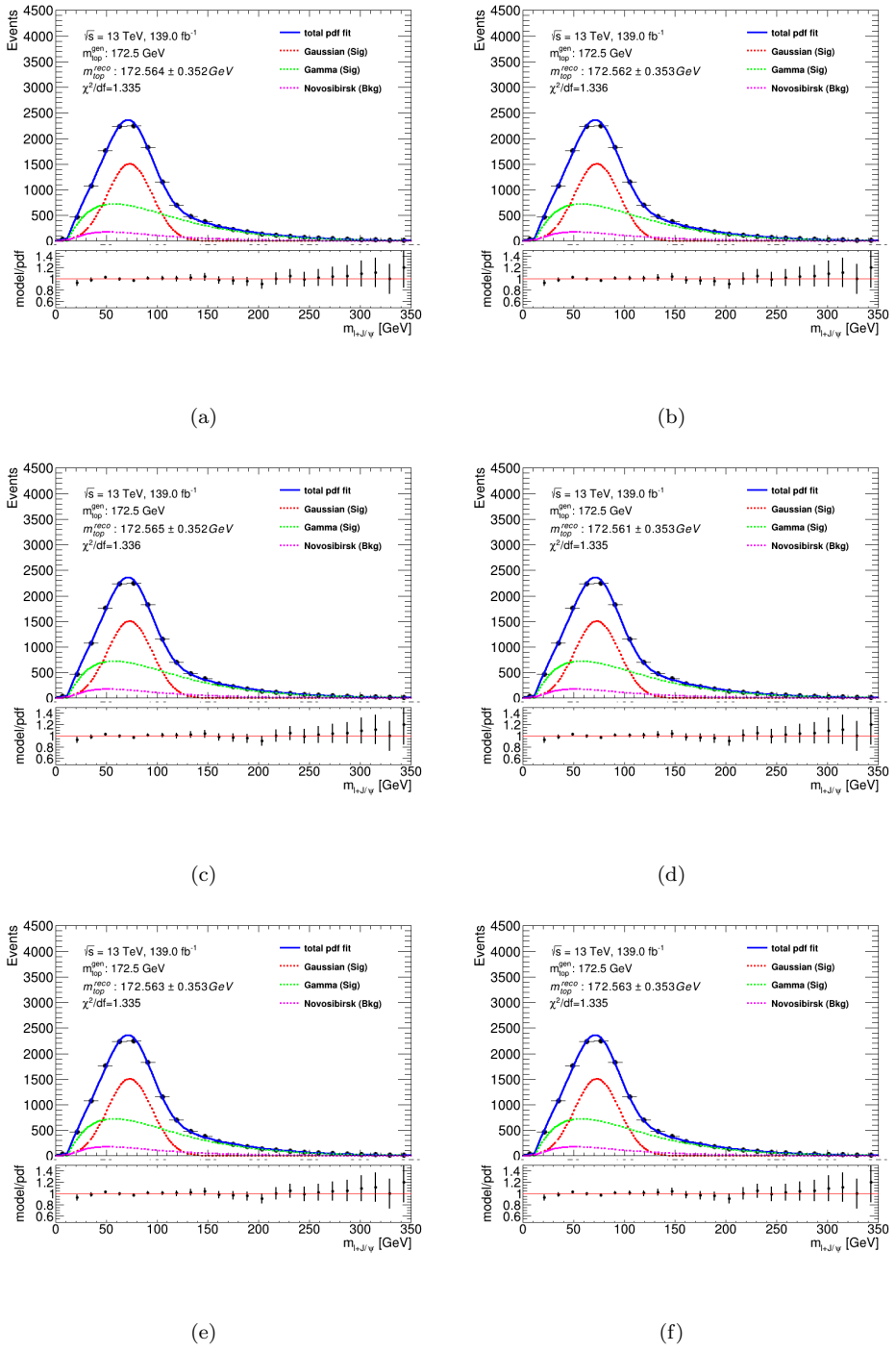


FIGURE I.31: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the flavour tagging calibration eigen-vectors 1 up (a) and down (b), 2 up (c) and down (d), and 3 up (e) and down (f) for c-tagged jets. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

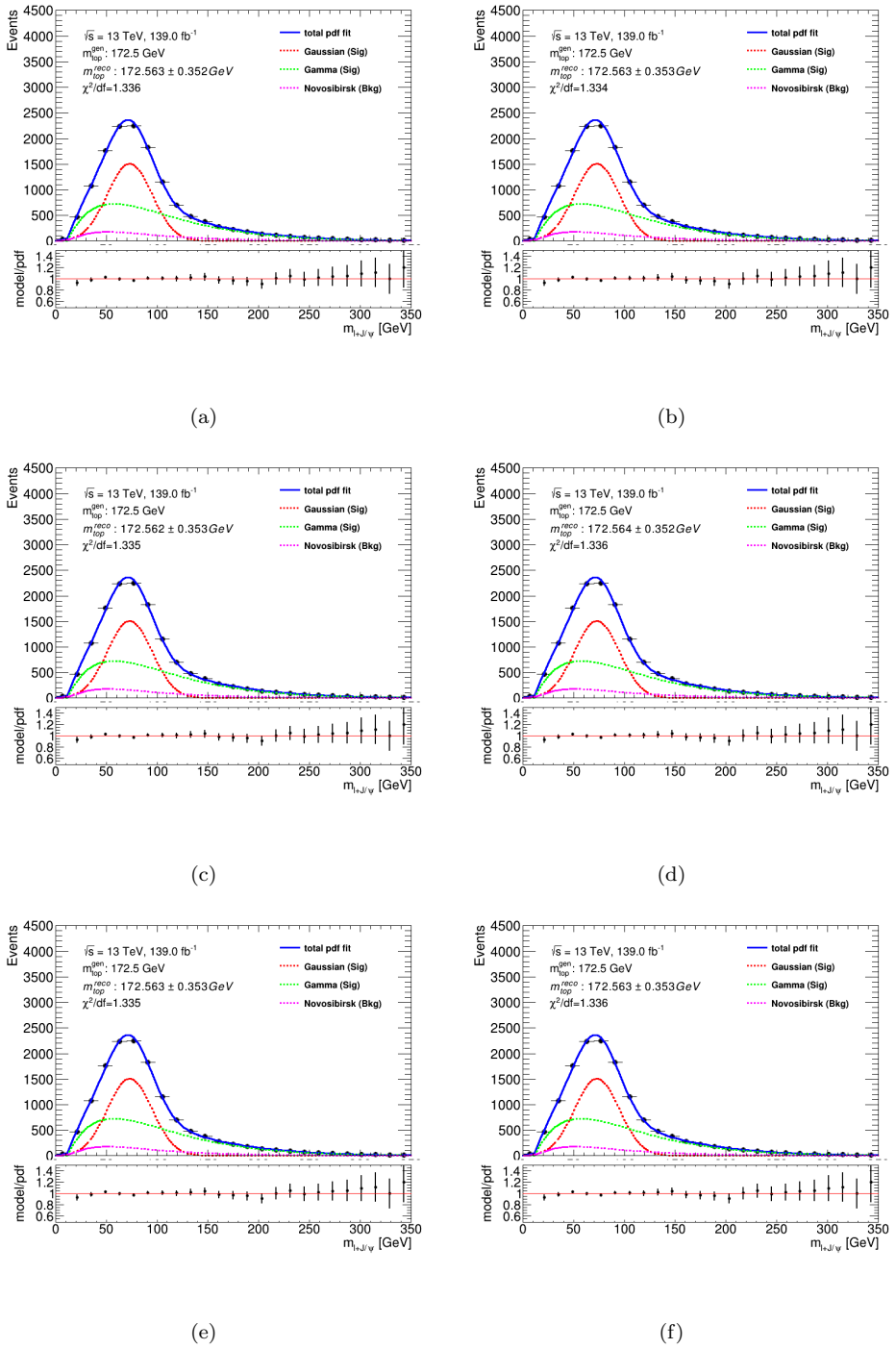


FIGURE I.32: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the flavour tagging calibration eigen-vectors 1 up (a) and down (b), 2 up (c) and down (d), and 3 up (e) and down (f) for light-tagged jets. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

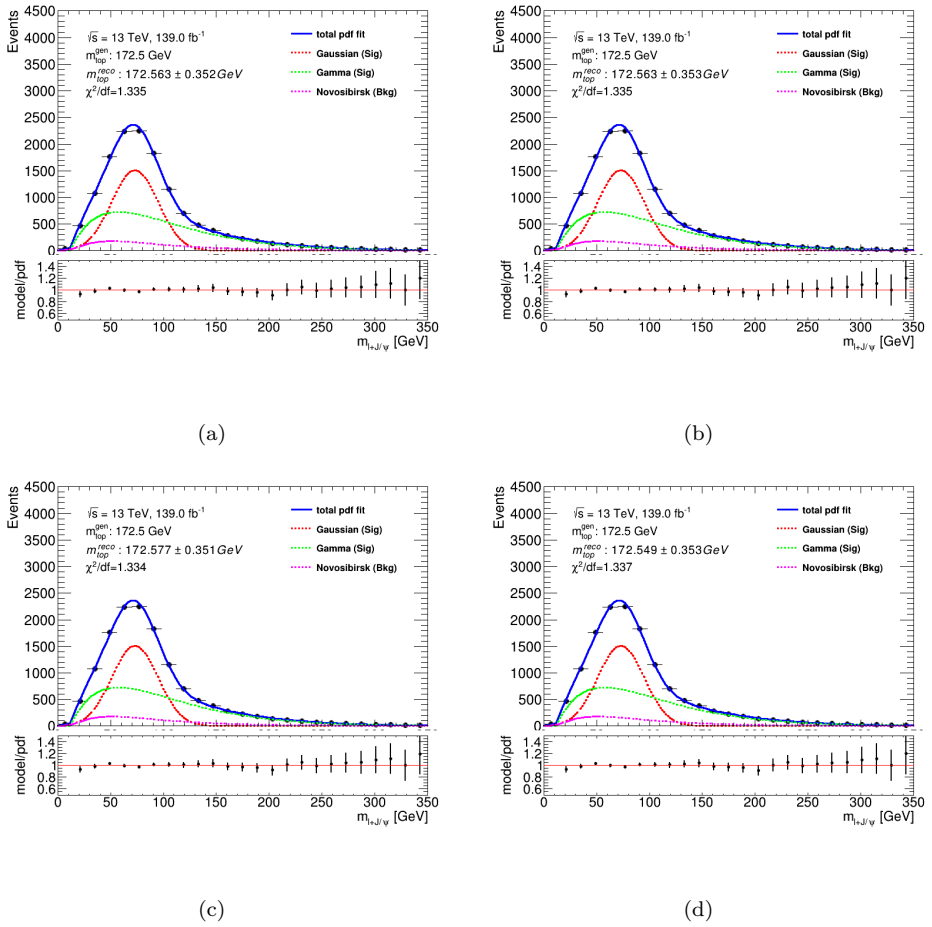


FIGURE I.33: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the flavour tagging calibration extrapolation of high-p_T b-tagged jets up (a) and down (b), and c-tagged jets up (c) and down (d). The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

I.7 Missing transverse momentum (MET)

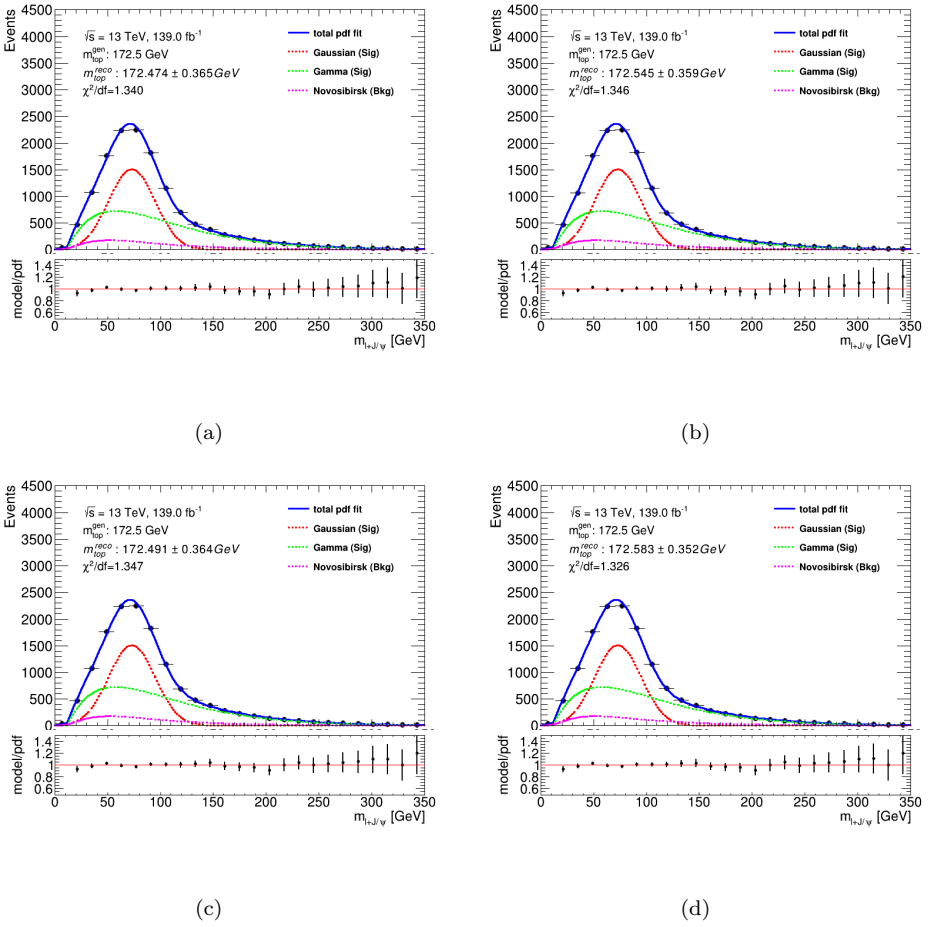


FIGURE I.34: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution when varying the effects of the p_T resolution in both the parallel (a) and perpendicular (b) directions of unclustered energy, and the p_T scale up (c) and down (d) of unclustered energy, in MET. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

Appendix J

Signal modelling systematic fit results with filtered samples

The following chapter shows the fit results for the different signal modelling systematic uncertainties produced with the $t \to b \to J/\psi \to \mu\mu$ filter.

Figures J.1-J.3 shows the fit results for the different signal systematic uncertainties produced with the new filter, with the reconstructed top quark mass and its statistical uncertainty, as well as the fit quality in the top left corner of the upper panels. The lower panels consist of the ratio between the model and the pdf. The blue shows the total pdf while the red, green and purple shows the Gaussian, Gamma and Novosibirsk contributions, respectively. Each signal systematic fit shows a reasonable fit with a χ^2/df between 1 and 2.

J.1 NLO Generator

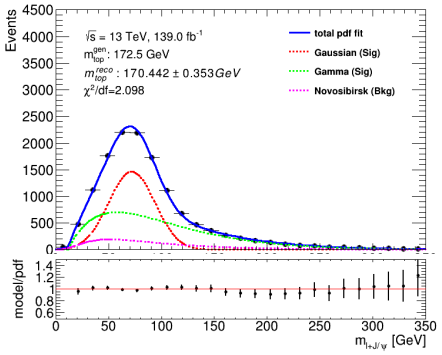
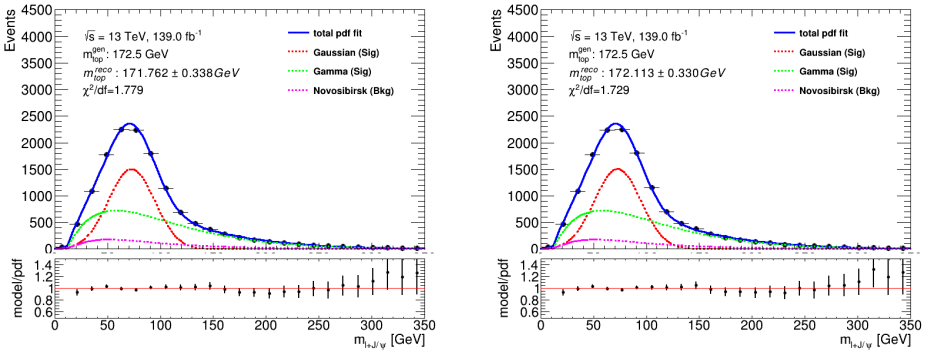


FIGURE J.1: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution using Powheg-Box to generate the NLO matrix element matching scheme produced with the new filter. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

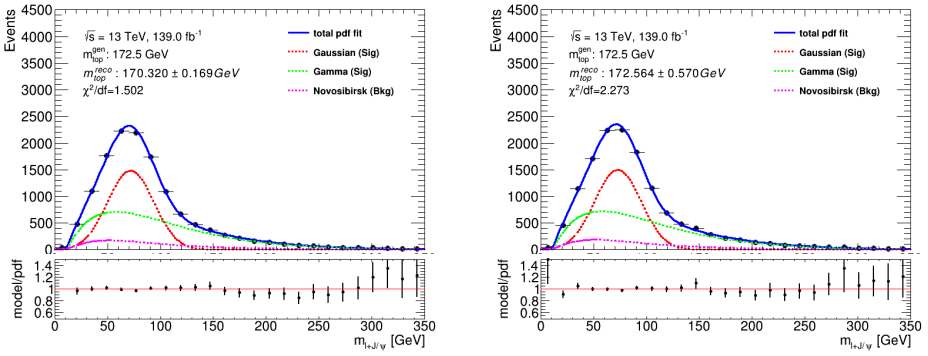
J.2 Modelling of heavy-quark fragmentation



(a) (b)

FIGURE J.2: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution, setting the fragmentation parameter r_b to 1.07 (a) and 1.03 (b) in Pythia 8 produced with the new filter. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

J.3 Modelling of final-state radiation



(a) (b)

FIGURE J.3: The upper panel shows the total (blue) pdf fit over the invariant mass of the lepton + J/ψ distribution, when varying the final-state radiation μ_R up to 2.0 (a) and down to 0.5 (b), produced with the new filter. The individual contributions from the Gaussian (red), Gamma (green) and Novosibirsk (purple) are also shown. The reconstructed top quark masses with the statistical uncertainty can be found in the top left corner. The lower panel shows the ratio between the model and the pdf.

Bibliography

[1] Mark Thomson. *Modern particle physics*. Cambridge University Press, New York, 2013. ISBN 9781107034266. URL http://www-spires.fnal.gov/spires/find/books/www?cl=QC793.2.T46::2013.

[2] David Griffiths. *Introduction to elementary particles*. 2008. ISBN 978-3-527-40601-2.

[3] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961. doi: 10.1016/0029-5582(61)90469-2.

[4] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967. doi: 10.1103/PhysRevLett.19.1264.

[5] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964. doi: 10.1103/PhysRevLett.13.321. [,157(1964)].

[6] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022. doi: 10.1093/ptep/ptac097.

[7] A. Arbey and F. Mahmoudi. Dark matter and the early Universe: a review. *Prog. Part. Nucl. Phys.*, 119:103865, 2021. doi: 10.1016/j.ppnp.2021.103865.

[8] Michael Dine and Alexander Kusenko. The Origin of the matter - antimatter asymmetry. *Rev. Mod. Phys.*, 76:1, 2003. doi: 10.1103/RevModPhys.76.1.

[9] Joseph D Lykken. Beyond the standard model. *arXiv preprint arXiv:1005.1676*, 2010.

[10] Michael Dine. *Supersymmetry and String Theory: Beyond the Standard Model*. Cambridge University Press, 1 2016. ISBN 978-1-107-04838-6, 978-1-316-40530-7.

[11] Wikipedia. Standard Model, 2018. URL https://en.wikipedia.org/wiki/Standard_Model. Accessed: 2018-10-15.

[12] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995. ISBN 978-0-201-50397-5.

[13] Abdus Salam and John Clive Ward. Electromagnetic and weak interactions. *Phys. Lett.*, 13:168–171, 1964. doi: 10.1016/0031-9163(64)90711-5.

[14] R. P. Feynman. *QED: The Strange Theory of Light and Matter*. 1986. ISBN 978-0-691-02417-2.

[15] Jeff Greensite. *An introduction to the confinement problem*, volume 821. 2011. doi: 10.1007/978-3-642-14382-3.

[16] David J. Gross and Frank Wilczek. Ultraviolet Behavior of Nonabelian Gauge Theories. *Phys. Rev. Lett.*, 30:1343–1346, 1973. doi: 10.1103/PhysRevLett.30.1343.

[17] R. P. Feynman. The behavior of hadron collisions at extreme energies. *Conf. Proc. C*, 690905:237–258, 1969.

[18] B. R. Webber. Fragmentation and hadronization. *Int. J. Mod. Phys. A*, 15S1:577–606, 2000. doi: 10.1142/S0217751X00005334.

[19] Abdelhak Djouadi. The Higgs mechanism and the origin of mass. *Fundam. Theor. Phys.*, 162:1–23, 2011. doi: 10.1007/978-90-481-3015-3_1.

[20] ATLAS Collaboration. Measurement of the top quark mass in the $t\bar{t} \to$ lepton+jets channel from $\sqrt{s} = 8$ TeV ATLAS data and combination with previous results. *Eur. Phys. J. C*, 79(4):290, 2019. doi: 10.1140/epjc/s10052-019-6757-9.

[21] CDF Collaboration. High-precision measurement of the W boson mass with the CDF II detector. *Science*, 376(6589):170–176, 2022. doi: 10.1126/science.abk1781.

[22] Andrea Giammanco. Single top quark production at the LHC. *Rev. Phys.*, 1:1–12, 2016. doi: 10.1016/j.revip.2015.12.001.

[23] Kevin Kröninger, Andreas B. Meyer, and Peter Uwer. Top-Quark Physics at the LHC. In Thomas Schörner-Sadenius, editor, *The Large Hadron Collider: Harvest of Run 1*, pages 259–300. 2015. doi: 10.1007/978-3-319-15001-7_7.

[24] Giorgio Cortiana. Top-quark mass measurements: review and perspectives. *Rev. Phys.*, 1:60–76, 2016. doi: 10.1016/j.revip.2016.04.001.

[25] Kirill Melnikov and Timo van Ritbergen. The Three loop relation between the MS-bar and the pole quark masses. *Phys. Lett.*, B482:99–108, 2000. doi: 10.1016/S0370-2693(00)00507-4.

[26] S. Moch. Precision determination of the top-quark mass. *PoS*, LL2014:054, 2014. doi: 10.22323/1.211.0054.

[27] CDF Collaboration. Observation of top quark production in $\bar{p}p$ collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995. doi: 10.1103/PhysRevLett.74.2626.

[28] D0 Collaboration. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995. doi: 10.1103/PhysRevLett.74.2632.

[29] D0 Collaboration. Determination of the pole and $\overline{MS}$ masses of the top quark from the $t\bar{t}$ cross section. *Phys. Lett.*, B703:422–427, 2011. doi: 10.1016/j.physletb.2011.08.015.

[30] S. Heinemeyer, S. Kraml, W. Porod, and G. Weiglein. Physics impact of a precise determination of the top quark mass at an e^+e^- linear collider. *JHEP*, 09:075, 2003. doi: 10.1088/1126-6708/2003/09/075.

[31] Michele Gallinaro. Top quark physics: A tool for discoveries. *Journal of Physics: Conference Series*, 447(1):012012, 2013. URL http://stacks.iop.org/1742-6596/447/i=1/a=012012.

[32] Giuseppe Degrassi, Stefano Di Vita, Joan Elias-Miro, Jose R. Espinosa, Gian F. Giudice, Gino Isidori, and Alessandro Strumia. Higgs mass and vacuum stability in the Standard Model at NNLO. *JHEP*, 08:098, 2012. doi: 10.1007/JHEP08(2012)098.

[33] Fedor Bezrukov and Mikhail Shaposhnikov. Why should we care about the top quark Yukawa coupling? *J. Exp. Theor. Phys.*, 120:335–343, 2015. doi: 10.1134/S1063776115030152. [Zh. Eksp. Teor. Fiz.147,389(2015)].

[34] M. Baak, J. Cúth, J. Haller, A. Hoecker, R. Kogler, K. Mönig, M. Schott, and J. Stelzer. The global electroweak fit at NNLO and prospects for the LHC and ILC. *Eur. Phys. J.*, C74:3046, 2014. doi: 10.1140/epjc/s10052-014-3046-5.

[35] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012. doi: 10.1016/j.physletb.2012.08.020.

[36] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012. doi: 10.1016/j.physletb.2012.08.021.

[37] K.N. Barends. Characterizing the background in $t\bar{t}$ events with a $J/\psi \to \mu\mu$ in proton-proton collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. 2019. URL http://hdl.handle.net/11427/31431.

[38] Yong Tang. Vacuum Stability in the Standard Model. *Mod. Phys. Lett.*, A28:1330002, 2013. doi: 10.1142/S0217732313300024.

[39] Gino Isidori, Giovanni Ridolfi, and Alessandro Strumia. On the metastability of the standard model vacuum. *Nucl. Phys.*, B609:387–409, 2001. doi: 10.1016/S0550-3213(01)00302-9.

[40] CMS Collaboration. A profile likelihood approach to measure the top quark mass in the lepton+jets channel at $\sqrt{s} = 13$ TeV. (CMS-PAS-TOP-20-008), 2022.

[41] ATLAS Collaboration. Measurement of the top quark mass in the $t\bar{t} \rightarrow$ lepton+jets channel from $\sqrt{s} = 8$ TeV ATLAS data and combination with previous results. *Eur. Phys. J. C*, 79(4):290, 2019. doi: 10.1140/epjc/s10052-019-6757-9.

[42] ATLAS Collaboration. Measurement of the top-quark mass in $t\bar{t} + 1$-jet events collected with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 11:150, 2019. doi: 10.1007/JHEP11(2019)150.

[43] CMS Collaboration. Measurement of the top quark mass using events with a single reconstructed top quark in pp collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 12:161, 2021. doi: 10.1007/JHEP12(2021)161.

[44] CMS Collaboration. Measurement of the Jet Mass Distribution and Top Quark Mass in Hadronic Decays of Boosted Top Quarks in pp Collisions at $\sqrt{s} =$ TeV. *Phys. Rev. Lett.*, 124(20):202001, 2020. doi: 10.1103/PhysRevLett.124.202001.

[45] CMS Collaboration. Measurement of the top quark mass in the all-jets final state at $\sqrt{s} =$ 13 TeV and combination with the lepton+jets channel. *Eur. Phys. J. C*, 79(4):313, 2019. doi: 10.1140/epjc/s10052-019-6788-2.

[46] CMS Collaboration. Measurement of the $\mathrm{t\bar{t}}$ production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79(5):368, 2019. doi: 10.1140/epjc/s10052-019-6863-8.

[47] CMS Collaboration. Measurement of the top quark mass using proton-proton data at $\sqrt{(s)}$ = 7 and 8 TeV. *Phys. Rev. D*, 93(7):072004, 2016. doi: 10.1103/PhysRevD.93.072004.

[48] Tevatron Electroweak Working Group et al. Combination of cdf and d0 results on the mass of the top quark using up 9.7 fb−1 at the tevatron. *arXiv preprint arXiv:1608.01881*, 2016.

[49] CMS Collaboration. Measurement of the mass of the top quark in decays with a J/ψ meson in pp collisions at 8 TeV. *JHEP*, 12:123, 2016. doi: 10.1007/JHEP12(2016)123.

[50] CERN. Origins, 2018. URL https://timeline.web.cern.ch/events/origins. Accessed: 2018-12-10.

[51] CERN. The European Organization for Nuclear Research is born, 2018. URL https://timeline.web.cern.ch/events/the-european-organization-for-nuclear-research-is-born. Accessed: 2018-12-10.

[52] CERN. Our Member States, 2018. URL https://home.cern/about/who-we-are/our-governance/member-states. Accessed: 2018-12-10.

[53] CERN. Where did it all begin?, 2018. URL https://home.cern/about/who-we-are/our-history. Accessed: 2018-12-10.

[54] CERN. CERN's first accelerator - the Synchrocyclotron - starts up, 2018. URL https://timeline.web.cern.ch/cerns-first-accelerator-synchrocyclotron-starts. Accessed: 2018-12-10.

[55] CERN. W and Z particles discovered, 2018. URL https://timeline.web.cern.ch/events/w-and-z-particles-discovered. Accessed: 2018-12-10.

[56] CERN. The birth of the Web, 2018. URL https://home.cern/science/computing/birth-web. Accessed: 2018-12-10.

[57] CERN. First antiatoms produced: antihydrogen, at CERN, 2018. URL https://timeline.web.cern.ch/events/first-antiatoms-produced-antihydrogen-at-cern. Accessed: 2018-12-10.

[58] CERN. The Higgs boson, 2018. URL https://home.cern/science/physics/higgs-boson. Accessed: 2018-12-10.

[59] Esma Mobs. The CERN accelerator complex. August 2018. URL https://cds.cern.ch/record/2636343. General Photo.

[60] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008. doi: 10.1088/1748-0221/3/08/S08001.

[61] CERN. Large Electron-Positron collider: First injection, 2018. URL https://timeline.web.cern.ch/events/large-electron-positron-collider-first-injection. Accessed: 2018-12-11.

[62] CERN. Restarting the LHC: Why 13 TeV?, 2015. URL https://home.cern/science/engineering/restarting-lhc-why-13-tev. Accessed: 2022-08-01.

[63] CERN. The accelerator complex, 2018. URL https://home.cern/science/accelerators/accelerator-complex. Accessed: 2018-12-11.

[64] LHC Guide. Mar 2017. URL https://cds.cern.ch/record/2255762.

[65] ATLAS Collaboration. LuminosityPublicResultsRun2, 2018. URL https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2. Accessed: 2022-08-01.

[66] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008. doi: 10.1088/1748-0221/3/08/S08003.

[67] ATLAS Collaboration. The Collaboration, 2018. URL https://atlas.cern/discover/collaboration. Accessed: 2018-12-13.

[68] ATLAS Collaboration. The Physics, 2018. URL https://atlas.cern/discover/physics. Accessed: 2018-12-13.

[69] ATLAS Collaboration. Detector & Technology, 2018. URL https://atlas.cern/discover/detector. Accessed: 2018-12-13.

[70] Joao Pequenao. Computer generated image of the whole ATLAS detector. Mar 2008. URL https://cds.cern.ch/record/1095924.

[71] Matthias Schott and Monica Dunford. Review of single vector boson production in pp collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J.*, C74:2916, 2014. doi: 10.1140/epjc/s10052-014-2916-1.

[72] ATLAS Collaboration. Magnet System, 2019. URL https://atlas.cern/discover/detector/magnet-system. Accessed: 2019-01-07.

[73] ATLAS Collaboration. ATLAS Insertable B-Layer Technical Design Report. Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, Sep 2010. URL https://cds.cern.ch/record/1291633.

[74] ATLAS Collaboration. Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC. *Eur. Phys. J. C*, 77(5):332, 2017. doi: 10.1140/epjc/s10052-017-4887-5.

[75] Joao Pequenao. Computer generated image of the ATLAS inner detector. Mar 2008. URL https://cds.cern.ch/record/1095926.

[76] Francesca Cavallari. Performance of calorimeters at the LHC. *J. Phys. Conf. Ser.*, 293: 012001, 2011. doi: 10.1088/1742-6596/293/1/012001.

[77] Joao Pequenao. Computer generated image of the ATLAS calorimeter. Mar 2008. URL https://cds.cern.ch/record/1095927.

[78] Joao Pequenao. Computer generated image of the ATLAS Muons subsystem. Mar 2008. URL https://cds.cern.ch/record/1095929.

[79] Peter Jenni, Markus Nordberg, Marzio Nessi, and Kerstin Jon-And. *ATLAS Forward Detectors for Measurement of Elastic Scattering and Luminosity*. Technical Design Report ATLAS. CERN, Geneva, 2008. URL https://cds.cern.ch/record/1095847.

[80] Peter Jenni and Marzio Nessi. ATLAS Forward Detectors for Luminosity Measurement and Monitoring. Technical Report CERN-LHCC-2004-010. LHCC-I-014, CERN, Geneva, Mar 2004. URL https://cds.cern.ch/record/721908. revised version number 1 submitted on 2004-03-22 14:56:11.

[81] Peter Jenni, Marzio Nessi, and Markus Nordberg. Zero Degree Calorimeters for ATLAS. Technical Report CERN-LHCC-2007-001. LHCC-I-016, CERN, Geneva, Jan 2007. URL http://cds.cern.ch/record/1009649.

[82] L Adamczyk, E Banaś, A Brandt, M Bruschi, S Grinstein, J Lange, M Rijssenbeek, P Sicho, R Staszewski, T Sykora, M Trzebiński, J Chwastowski, and K Korcyl. Technical Design Report for the ATLAS Forward Proton Detector. Technical Report CERN-LHCC-2015-009. ATLAS-TDR-024, May 2015. URL https://cds.cern.ch/record/2017378.

[83] ATLAS Collaboration. ATLAS Forward Detectors. General Photo, Jun 2018. URL http://cds.cern.ch/record/2627582.

[84] ATLAS Collaboration. Performance of the ATLAS Trigger System in 2015. *Eur. Phys. J. C*, 77(5):317, 2017. doi: 10.1140/epjc/s10052-017-4852-3.

[85] ATLAS Collaboration. The ATLAS Data Acquisition and High-Level Trigger: Concept, Design and Status. (ATL-DAQ-CONF-2006-016, ATL-COM-DAQ-2006-030), 5 2006.

[86] ATLAS Collaboration. ApprovedPlotsDAQ, 2018. URL https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsDAQ. Accessed: 2022-08-01.

[87] ATLAS Collaboration. The ATLAS Simulation Infrastructure. *Eur. Phys. J.*, C70:823–874, 2010. doi: 10.1140/epjc/s10052-010-1429-9.

[88] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003. doi: 10.1016/S0168-9002(03)01368-8.

[89] M Beckingham, M. Duehrssen, E. Schmidt, M. Shapiro, M. Venturi, J. Virzi, I Vivarelli, M. Werner, S. Yamamoto, and T. Yamanaka. The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim. 10 2010.

[90] W. Lukas. Fast Simulation for ATLAS: Atlfast-II and ISF. *J. Phys. Conf. Ser.*, 396: 022031, 2012. doi: 10.1088/1742-6596/396/2/022031.

[91] Joao Pequenao and Paul Schaffner. An computer generated image representing how ATLAS detects particles. Jan 2013. URL https://cds.cern.ch/record/1505342.

[92] R. Frühwirth. Application of kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2):444–450, 1987. ISSN 0168-9002. doi: https://doi.

org/10.1016/0168-9002(87)90887-4. URL https://www.sciencedirect.com/science/article/pii/0168900287908874.

[93] ATLAS Collaboration. Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV. (ATL-PHYS-PUB-2015-026), 2015. URL https://cds.cern.ch/record/2037717.

[94] ATLAS Collaboration. Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data. *JINST*, 14(12):P12006, 2019. doi: 10.1088/1748-0221/14/12/P12006.

[95] ATLAS Collaboration. Muon reconstruction and identification efficiency in ATLAS using the full Run 2 *pp* collision data set at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 81(7):578, 2021. doi: 10.1140/epjc/s10052-021-09233-2.

[96] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal. Calorimeter clustering algorithms: Description and performance. (ATL-LARG-PUB-2008-002, ATL-COM-LARG-2008-003), 5 2008.

[97] ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J. C*, 77:490, 2017. doi: 10.1140/epjc/s10052-017-5004-5.

[98] ATLAS Collaboration. Jet reconstruction and performance using particle flow with the ATLAS Detector. *Eur. Phys. J. C*, 77(7):466, 2017. doi: 10.1140/epjc/s10052-017-5031-2.

[99] ATLAS Collaboration. Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 81(8):689, 2021. doi: 10.1140/epjc/s10052-021-09402-3.

[100] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008. doi: 10.1088/1126-6708/2008/04/063.

[101] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160–3166, 1993. doi: 10.1103/PhysRevD.48.3160.

[102] HFLAV Collaboration. Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018. *Eur. Phys. J. C*, 81(3):226, 2021. doi: 10.1140/epjc/s10052-020-8156-7.

[103] Sebastian Heer. The secondary vertex finding algorithm with the ATLAS detector. *PoS*, EPS-HEP2017:762, 2017. doi: 10.22323/1.314.0762.

[104] ATLAS Collaboration. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79(11):970, 2019. doi: 10.1140/epjc/s10052-019-7450-8.

[105] G. Avoni et al. The new lucid-2 detector for luminosity measurement and monitoring in atlas. *JINST*, 13(07):P07017, 2018. doi: 10.1088/1748-0221/13/07/P07017.

[106] ATLAS Collaboration. Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. *arXiv preprint arXiv:2212.09379*, 2022.

[107] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462: 152–155, 2001. doi: 10.1016/S0168-9002(01)00089-4.

[108] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008. doi: 10.1016/j.cpc.2008.01.036.

[109] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004. doi: 10.1088/1126-6708/2004/11/040.

[110] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007. doi: 10.1088/1126-6708/2007/11/070.

[111] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010. doi: 10.1007/JHEP06(2010)043.

[112] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. *JHEP*, 09:126, 2007. doi: 10.1088/1126-6708/2007/09/126.

[113] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015. doi: 10.1007/JHEP04(2015)040.

[114] Richard D. Ball et al. Parton distributions with LHC data. *Nucl. Phys. B*, 867:244–289, 2013. doi: 10.1016/j.nuclphysb.2012.10.003.

[115] ATLAS Collaboration. ATLAS Run 1 Pythia8 tunes. Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov 2014. URL https://cds.cern.ch/record/1966419.

[116] Stefano Frixione, Eric Laenen, Patrick Motylinski, and Bryan R. Webber. Single-top production in MC@NLO. *JHEP*, 03:092, 2006. doi: 10.1088/1126-6708/2006/03/092.

[117] Stefano Frixione, Eric Laenen, Patrick Motylinski, Bryan R. Webber, and Chris D. White. Single-top hadroproduction in association with a W boson. *JHEP*, 07:029, 2008. doi: 10.1088/1126-6708/2008/07/029.

[118] Rikkert Frederix, Emanuele Re, and Paolo Torrielli. Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO. *JHEP*, 09:130, 2012. doi: 10.1007/JHEP09(2012)130.

[119] Enrico Bothmann et al. Event Generation with Sherpa 2.2. *SciPost Phys.*, 7(3):034, 2019. doi: 10.21468/SciPostPhys.7.3.034.

[120] Fabio Cascioli, Philipp Maierhofer, and Stefano Pozzorini. Scattering Amplitudes with Open Loops. *Phys. Rev. Lett.*, 108:111601, 2012. doi: 10.1103/PhysRevLett.108.111601.

[121] Tanju Gleisberg and Stefan Hoeche. Comix, a new matrix element generator. *JHEP*, 12: 039, 2008. doi: 10.1088/1126-6708/2008/12/039.

[122] Steffen Schumann and Frank Krauss. A Parton shower algorithm based on Catani-Seymour dipole factorisation. *JHEP*, 03:038, 2008. doi: 10.1088/1126-6708/2008/03/038.

[123] Jan-Christopher Winter, Frank Krauss, and Gerhard Soff. A Modified cluster hadronization model. *Eur. Phys. J. C*, 36:381–395, 2004. doi: 10.1140/epjc/s2004-01960-8.

[124] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: 10.1007/JHEP07(2014)079.

[125] ATLAS Collaboration. Modelling of the $t\bar{t}H$ and $t\bar{t}V$ ($V = W, Z$) processes for $\sqrt{s} = 13$ TeV ATLAS analyses. (ATL-PHYS-PUB-2016-005), 2016.

[126] ATLAS Collaboration. Summary of ATLAS Pythia 8 tunes. (ATL-PHYS-PUB-2012-003), 2012.

[127] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, Pavel M. Nadolsky, and W. K. Tung. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 07:012, 2002. doi: 10.1088/1126-6708/2002/07/012.

[128] Hua-Sheng Shao. HELAC-Onia 2.0: an upgraded matrix-element and event generator for heavy quarkonium physics. *Comput. Phys. Commun.*, 198:238–259, 2016. doi: 10.1016/j.cpc.2015.09.011.

[129] Matteo Cacciari, Michal Czakon, Michelangelo Mangano, Alexander Mitov, and Paolo Nason. Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation. *Phys. Lett. B*, 710:612–622, 2012. doi: 10.1016/j.physletb.2012.03.013.

[130] Peter Bärnreuther, Michal Czakon, and Alexander Mitov. Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$. *Phys. Rev. Lett.*, 109:132001, 2012. doi: 10.1103/PhysRevLett.109.132001.

[131] Michal Czakon and Alexander Mitov. NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction. *JHEP*, 01:080, 2013. doi: 10.1007/JHEP01(2013)080.

[132] Michal Czakon and Alexander Mitov. NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels. *JHEP*, 12:054, 2012. doi: 10.1007/JHEP12(2012)054.

[133] Michał Czakon, Paul Fiedler, and Alexander Mitov. Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$. *Phys. Rev. Lett.*, 110:252004, 2013. doi: 10.1103/PhysRevLett.110.252004.

[134] J Ferrando, D Wendland, and Francesco Giuli. Reference $t\bar{t}$ production cross sections for use in ATLAS analyses. Technical report, CERN, Geneva, Feb 2014. URL https://cds.cern.ch/record/1662536.

[135] Nikolaos Kidonakis. Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-. *Phys. Rev. D*, 82:054018, 2010. doi: 10.1103/PhysRevD.82.054018.

[136] Nikolaos Kidonakis. Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production. *Phys. Rev. D*, 83:091503, 2011. doi: 10.1103/PhysRevD.83.091503.

[137] ATLAS Collaboration. ATLAS simulation of boson plus jets processes in Run 2. (ATL-PHYS-PUB-2017-006), 5 2017.

[138] ATLAS Collaboration. Multi-Boson Simulation for 13 TeV ATLAS Analyses. (ATL-PHYS-PUB-2016-002), 2016.

[139] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: 10.1007/JHEP07(2014)079.

[140] S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro. Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons. *JHEP*, 06:184, 2015. doi: 10.1007/JHEP06(2015)184.

[141] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2/2017, 10 2016. doi: 10.23731/CYRM-2017-002.

[142] ATLAS Collaboration. Measurement of J/ψ production in association with a $W^{\pm}$ boson with pp data at 8 TeV. *JHEP*, 01:095, 2020. doi: 10.1007/JHEP01(2020)095.

[143] ATLAS Collaboration. Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. (ATLAS-CONF-2014-058, ATLAS-CONF-2014-058), 10 2014.

[144] Avto Kharchilava. Top mass determination in leptonic final states with J/ψ. *Phys. Lett. B*, 476:73–78, 2000. doi: 10.1016/S0370-2693(00)00120-9.

[145] G. Cowan. *Statistical data analysis*. 1998. ISBN 978-0-19-850156-5.

[146] ATLAS Collaboration. Measurement of the top quark mass in the $t\bar{t} \to$ dilepton channel from $\sqrt{s} = 8$ TeV ATLAS data. *Phys. Lett. B*, 761:350–371, 2016. doi: 10.1016/j.physletb.2016.08.042.

[147] Joint Committee for Guides in Metrology. Jcgm 100: Evaluation of measurement data - guide to the expression of uncertainty in measurement. Technical report, JCGM, 2008.

[148] Peter Skands, Stefano Carrazza, and Juan Rojo. Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J. C*, 74(8):3024, 2014. doi: 10.1140/epjc/s10052-014-3024-y.

[149] ATLAS Collaboration. Measurement of the top-quark mass using a leptonic invariant mass in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *arXiv preprint arXiv:2209.00583*, 9 2022.

[150] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys. Rept.*, 97:31–145, 1983. doi: 10.1016/0370-1573(83)90080-7.

[151] M. G. Bowler. e+ e- Production of Heavy Quarks in the String Model. *Z. Phys. C*, 11: 169, 1981. doi: 10.1007/BF01574001.

[152] DELPHI Collaboration. A Measurement of the branching fractions of the b quark into charged and neutral b hadrons. *Phys. Lett. B*, 576:29–42, 2003. doi: 10.1016/j.physletb.2003.09.070.

[153] ACDF Collaboration. Measurement of Ratios of Fragmentation Fractions for Bottom Hadrons in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$-TeV. *Phys. Rev. D*, 77:072003, 2008. doi: 10.1103/PhysRevD.77.072003.

[154] CDF Collaboration. Measurement of b quark fragmentation fractions in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev. Lett.*, 84:1663–1668, 2000. doi: 10.1103/PhysRevLett.84.1663.

[155] CDF Collaboration. Measurement of b quark fragmentation fractions in the production of strange and light B mesons in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev. D*, 60:092005, 1999. doi: 10.1103/PhysRevD.60.092005.

[156] CDF Collaboration. First Measurement of the Ratio of Branching Fractions $B(\Lambda_b^0 \to \Lambda_c^+ \mu^- \bar{\nu}_\mu / B(\Lambda_b^0 \to \Lambda_c^+ \pi^-)$. *Phys. Rev. D*, 79:032001, 2009. doi: 10.1103/PhysRevD.79.032001.

[157] LHCb Collaboration. Measurement of b-hadron production fractions in 7 TeV pp collisions. *Phys. Rev. D*, 85:032008, 2012. doi: 10.1103/PhysRevD.85.032008.

[158] LHCb Collaboration. Determination of f_s/f_d for 7 TeV pp collisions and a measurement of the branching fraction of the decay $B_d \to D^- K^+$. *Phys. Rev. Lett.*, 107:211801, 2011. doi: 10.1103/PhysRevLett.107.211801.

[159] LHCb Collaboration. Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics. *JHEP*, 04:001, 2013. doi: 10.1007/JHEP04(2013)001.

[160] ATLAS Collaboration. Determination of the ratio of b-quark fragmentation fractions f_s/f_d in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys. Rev. Lett.*, 115(26): 262001, 2015. doi: 10.1103/PhysRevLett.115.262001.

[161] HFLAV Collaboration. Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018. *Eur. Phys. J. C*, 81(3):226, 2021. doi: 10.1140/epjc/s10052-020-8156-7.

[162] ATLAS Collaboration. Simulation of top quark production for the ATLAS experiment at sqrt(s) = 13 TeV. (ATL-PHYS-PUB-2016-004), 1 2016.

[163] ATLAS Collaboration. Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO. (ATL-PHYS-PUB-2017-007), 2017.

[164] S. Mrenna and P. Skands. Automated Parton-Shower Variations in Pythia 8. *Phys. Rev. D*, 94(7):074005, 2016. doi: 10.1103/PhysRevD.94.074005.

[165] T. Sjostrand and Peter Z. Skands. Multiple interactions and the structure of beam remnants. *JHEP*, 03:053, 2004. doi: 10.1088/1126-6708/2004/03/053.

[166] Jesper R. Christiansen and Peter Z. Skands. String Formation Beyond Leading Colour. *JHEP*, 08:003, 2015. doi: 10.1007/JHEP08(2015)003.

[167] ATLAS Collaboration. A study of different colour reconnection settings for Pythia8 generator using underlying event observables. (ATL-PHYS-PUB-2017-008), 5 2017.

[168] Zahari Kassabov, Emanuele R. Nocera, and Michael Wilson. Regularising experimental correlations in LHC data: theory and application to a global analysis of parton distributions. *Eur. Phys. J. C*, 82(10):956, 2022. doi: 10.1140/epjc/s10052-022-10932-7.

[169] ATLAS Collaboration. Tagging and suppression of pileup jets. (ATLAS-CONF-2014-018), 5 2014.

[170] ATLAS Collaboration. Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 96 (7):072002, 2017. doi: 10.1103/PhysRevD.96.072002.

[171] ATLAS Collaboration. Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV. (ATL-PHYS-PUB-2015-015), 6 2015.

[172] ATLAS Collaboration. Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C*, 73(3):2304, 2013. doi: 10.1140/epjc/s10052-013-2304-2.

[173] ATLAS Collaboration. Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV. *JHEP*, 08:089, 2018. doi: 10.1007/JHEP08(2018)089.

[174] ATLAS Collaboration. Measurement of b-tagging Efficiency of c-jets in $t\bar{t}$ Events Using a Likelihood Approach with the ATLAS Detector. (ATLAS-CONF-2018-001, ATLAS-CONF-2018-001), 3 2018.

[175] ATLAS Collaboration. Calibration of light-flavour jet b-tagging rates on ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV. (ATLAS-CONF-2018-006), 4 2018.

[176] ATLAS Collaboration. Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC. *Eur. Phys. J. C*, 76(12):653, 2016. doi: 10.1140/epjc/s10052-016-4466-1.