Higgs Boson and New Physics Searches in Events with Same-sign Lepton Pairs at the LHC

Francesco Nuti

(orcid.org/0000-0003-3491-7637)

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

Produced on archival quality paper

January 2017

School of Physics

The University of Melbourne

Abstract

The physics of elementary particles is well described by the Standard Model of particle physics. This theory predicts with great accuracy the electroweak and strong interactions of elementary particles. The introduction of the Brout-Englert-Higgs mechanism in Standard Model has also given justification to the non-null mass of the elementary particles. This mechanism has been validated by the discovery of the Higgs boson and the measurement of its properties. Despite the enormous and broad success of the Standard Model, a few observations cannot be explained by this theory. For instance, neutrino oscillations would require the neutrinos to have non-null mass in contrast with the Standard Model prediction. Several solutions to this problem have been proposed, but none has been validated yet. Some of these theories have additional modes to produce lepton pairs with same electric charge (same-sign lepton pairs). This production is also expected to be enhanced in other theories beyond the Standard Model.

The focus of this thesis is the inclusive search for anomalous production of same-sign lepton pairs in proton-proton collisions. The specific case of resonant production, typical of the Type-II see-saw and Zee-Babu models of neutrino mass generation, is also considered. 20.3 fb^{-1} of data collected by the ATLAS experiment at a centre-of-mass energy of 8 TeV is analysed. No significant production above the Standard Model prediction is found, thus the results are used to set upper limits on the cross-section of processes beyond the Standard Model. Limits are placed as a function of the di-lepton invariant mass within a fiducial region defined according to the event selection criteria. Exclusion limits are also derived for specific models of resonant production.

The production of same-sign lepton pairs from proton-proton collisions occurs rarely in the Standard Model. One process which generates same-sign lepton pairs is the associated production of a Higgs boson and a top anti-top quark pair ($t\bar{t}H$). The cross-section of this production mode is proportional to the Top Yukawa coupling squared, while other processes are only sensitive to the combination of this constant with several other couplings. The measurement of the $t\bar{t}H$ cross-section therefore constitutes an important test of the Standard Model and the Brout-Englert-Higgs mechanism. The $t\bar{t}H$ production has never been observed before, and is studied in this thesis. The analysis uses collisions at a centre-of-mass energy of 8 TeV collected by the ATLAS experiment containing a same-sign lepton pair and jets. The measured production cross-section is $2.8^{+2.1}_{-1.9}$ times the value expected in the Standard Model. This result has been combined with other analyses to obtain a more precise measurement of the $t\bar{t}H$ cross-section and the Top Yukawa coupling.

Declaration

This is to certify that:

- the thesis comprises only my original work towards the PhD except where indicated in the Preface,
- due acknowledgement has been made in the text to all other material used,
- the thesis is less than 100000 words in length, exclusive of tables, bibliographies and appendices.

Francesco Nuti

Preface

The study of particle physics at the Large Hadron Collider (LHC) requires a large collaborative effort, given the complexity of the experiments and the investigated physics. The ATLAS collaboration, which the author is part of, includes over 3000 scientists. Publications are never the work of an individual. Data analyses are developed in small teams (working groups), but the whole collaboration is involved in data acquisition, detector monitoring and development, performance studies, and simulations. Thus the entire collaboration of scientists are included as contributing authors on all ATLAS publications.

This thesis presents the work undertaken by the author within several working groups. The author production is described in Chapter 3, 4, 5 and 6, and led to the publication of three papers [1–3]. The study in Chapter 3 is the sole production of the author, excluding some of the results in Section 3.3.1 as explicitly stated. The studies in Chapter 4, 5 and 6 are mostly original work of the author, however in some cases the author had only marginal involvement. These cases are: Section 4.2.1, 4.2.2, 5.4.2.2, 5.4.2.3, 5.4.4.2, 5.4.4.3, 6.4.2.2, 6.4.4.1 and 6.7.2. The remaining content of these chapters is the original production of the author, unless explicitly stated. Please note that collaborators may have independently preformed similar studies for validation.

A Hilary, ai miei genitori e ai Broccoli

Acknowledgements

This work is the accomplishment of a long and extraordinary journey, which has substantially shaped me as a researcher as well as an individual. If this experience was successful, it is vastly because I was never alone. I have to thank plenty of people for their support, love, amusement, guidance and wisdom during these years. These people were extremely beneficial to my professional and personal growth.

I am immensely grateful to my supervisor, Prof. Elisabetta Barberio, who believed I had the talent to face this challenge. She has always guided and supported me along my PhD, especially in my first year when I needed more help. With time I have been able to develop my research more autonomously, but she has always followed my progress and has been of great help. I am also very thankful to her for granting me the opportunity to undertake plenty of activities such as schools of physics, classes, conferences and overseas research.

Special thanks to Dr. Kenji Hamano, who has led the search for non-Standard Model same-sign lepton production. He has been fantastic in managing the group and has always been very supportive with me. Thanks for being so professional and human at the same time.

I am very glad to have worked with Dr. Daniele Zanzi, who is probably the best young physicist I have ever met. His advices have been invaluable and he is an inspiration for everyone in term of work ethic. I owe him a lot.

I want to thank Prof. Geoffrey N. Taylor and Prof. Raymond R. Volkas who have been very accessible and beneficial to me along these years. I believe they are largely accountable for the pleasant atmosphere of the CoEPP work environment.

I thank all the people I have worked with in the ATLAS experiment and the whole CoEPP group. My best wishes go to Dr. David Jennens, Dr. Amelia Brennan, KG, Pere Rados and Laurence Spiller who have been my daily source of professional and technical knowledge, empathy, novelty, nerd culture and fun.

Grazie mille for so many years of good time to the Italian Clan: Dr. Matteo Volpi, Dr. Sara Diglio, Dr. Noel Dawe (Italian ad honorem), Dr. Francesca Ungaro, Dr. Francesco Tenchini, Dr. Federico Scutti, Marco Milesi, Giacomo Caria and the Baroncelli brothers. Our lunch fora are memorable. I am particularly grateful to Matteo and Sara to have made my integration in the local community so easy and quick.

Finally, I have immense gratitude for Hilary, my family and my friends, who have always given me full support and unlimited love.

Contents

Introduction

The Standard Model of particle physics is the most universally recognized theory that describes elementary particle interactions. Its success is due to the minimal number of particles and parameters used to describe strong and electroweak interactions, and to its great accuracy in predicting plenty of the particle physics observations. Nevertheless, the Standard Model has been, and is, challenged by few experimental observations.

The measurement of the non-null masses of the vector bosons $W^{\pm}$ and Z, which were originally massless in the theory, has led to the introduction of the Brout-Englert-Higgs mechanism in the Standard Model. This extension has recently been validated by the discovery of the Higgs boson and the measurement of its properties at the Large Hadron Collider (LHC). The associated production of a Higgs boson with a top anti-top quark pair ($t\bar{t}H$) has not been observed yet. This process is sensitive to the Top Yukawa coupling and can be investigated in data containing pairs of leptons with same electric charge (same-sign lepton pairs).

The observation of neutrino oscillations and, consequently, the acquisition of small masses by the neutrinos have not yet an explanation in Standard Model. The see-saw and the Zee-Babu models offer solutions to this problem introducing new massive fundamental particles. The physics of these models at hadron colliders leads to the production of same-sign lepton pairs at an enhanced rate. Further, some of these models predict doubly charged bosons which decay into same-sign lepton pairs. Other theories beyond the Standard Model predict additional same-sign lepton pair production and doubly charged bosons. A description of the Standard Model and extensions leading to same-sign lepton pair production is discussed in Chapter 1 of this thesis.

This thesis searches for evidence of anomalous production of same-sign lepton pairs and of $t\bar{t}H$ production in the data collected by the ATLAS detector at the LHC. The LHC has produced proton-proton collisions at a centre-of-mass energy of 8 TeV in 2012. This dataset is analysed in this thesis and corresponds to 20.3 fb^{-1}. The LHC, the ATLAS detector, and the experimental techniques of particle reconstruction are described in Chapter 2. Chapter 3 is dedicated to the calibration of the energy of hadronically decaying τ leptons, which has been developed by the author.

The analysis of data containing same-sign lepton pairs is the core of this thesis and is illustrated in Chapter 4, 5 and 6. The Standard Model processes, which can lead

to the detection of same-sign lepton pairs are characterised in Chapter 4. Chapter 5 searches for excesses of same-sign lepton pair production compatible with new physics, especially with generation of doubly charged bosons. The measurement of the $t\bar{t}H$ production cross-section and implications on the Top Yukawa coupling are illustrated in Chapter 6.

This thesis uses the convention of natural units for physics quantities, where the speed of light, the reduced Planck constant and the Boltzmann constant are normalised to unity, $c = \hbar = k_B = 1$. The energies, momenta, and masses are therefore expressed in units of electron volts (eV).

Chapter 1

Theoretical Overview

The understanding of the processes that rule the universe at the elementary scale is the goal of particle physics research. There are four types of fundamental forces: strong, electromagnetic, weak and gravitational. Relativistic quantum field theories are used to characterize the interactions of elementary particles via these forces. The Standard Model [4–6] is the most universally recognised of these theories and describes strong, weak and electromagnetic interactions. The gravitational force is not included in the Standard Model, however its effect on elementary particles is negligible compared to the other forces. In this chapter, the Standard Model is introduced along with some potential extensions, which propose solutions for otherwise unjustified observations. This thesis tests the predictions of these models in the production of leptons with same electric charge (same-sign leptons) in high energy proton collisions. Particular consideration is given to two processes with same-sign lepton pairs in the final state: the productions of doubly charged particles and the production of the Higgs boson in association with a pair of top anti-top quarks.

1.1 The Standard Model

The Standard Model describes the properties of elementary particles and their fundamental interactions in an unified formalism. The dynamics and interactions of the elementary particles are determined by the Standard Model Lagrangian and its symmetries. The invariance of the Lagrangian under local transformations of the gauge group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ results in the introduction of massless boson fields that mediate the interactions among the fermion fields associated to quarks and leptons. The symmetry under local transformations of the colour group $SU(3)_C$ demands eight spin-1 colour carriers, the gluons, which mediate the strong interaction among coloured particles such as the quarks and the gluons themselves. The local gauge invariance under $SU(2)_L \otimes U(1)_Y$ transformations generates four gauge fields: one for the hypercharge group, $U(1)_Y$, and three for weak isospin group, $SU(2)_L$. Linear combinations of these fields produce the spin-1 mediators of the electromagnetic and weak interactions: the photon (γ) and the

weak bosons (Z and $W^{\pm}$). The photon mediates the electromagnetic forces among electrically charged particles, like quarks, charged leptons and $W^{\pm}$. The weak bosons mediate the charged and neutral weak currents among fermions. Beside the gauge invariance, the Standard Model preserves also the baryon number, the lepton number, electric charge, weak isospin and is symmetric under CPT.

Experimental observations have led so far to the discovery of six types of quarks and six types of leptons. Each of these fermions has an anti-fermion partner with same spin and mass, but opposite charge. These fermions are obtained in the Standard Model as representations of the symmetry groups. Leptons are colourless singlets of $SU(3)_C$ and do not interact via strong force. Quarks are instead $SU(3)_C$ triplets and are strongly interacting. Since the weak force couples only to left-chiral fermions or right-chiral anti-fermions, i.e. violates parity, fermions have different representations in $SU(2)_L$ depending on their chirality. Right-chiral fermions are $SU(2)_L$ singlets and cannot weakly interact. There are right-chiral representations for quarks and charged leptons in Standard Model but not for neutrinos. Left-chiral fermions are represented by $SU(2)_L$ doublets and do interact via weak boson exchange. The leptonic doublets are formed by the fields associated to a charged lepton and a neutrino, while the quark doublets by an up-type quark and a down-type quark. The known leptons and quarks form therefore three generations each, which differs for the masses and flavours of the particles composing them.

The particles considered in the Standard Model and their properties are summarized in Table 1.1. The figure shows only fermions, but their respective anti-fermions are also accommodated into this theory. In the following, the name of elementary particles will be used to indicate either the particle or its anti-particle unless otherwise stated. For instance electrons and positrons will be both referred to as electrons. To further simplify the narrative, the charged leptons will be referred to simply as leptons while neutral leptons as neutrinos.

In addition to the particle described above, Table 1.1 shows also the Higgs boson (H). This particle was not part of the original Standard Model formulation but has been introduced afterwards to solve inconsistencies with the theory. The Standard Model was indeed able to classify the elementary particles in an elegant way but they could not have mass. This was in contrast with the experimental evidences and with the short range nature of the weak interactions.

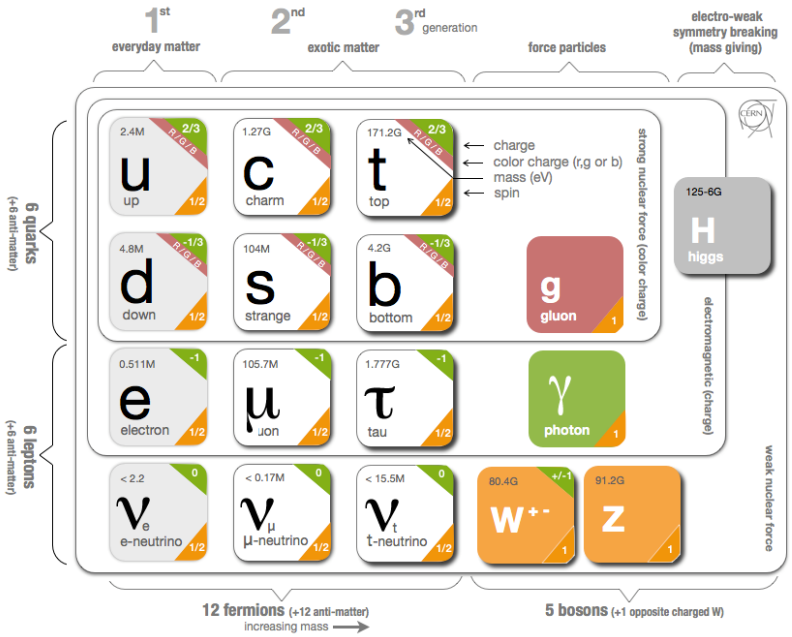


Table 1.1: The elementary particles of the Standard Model. The mass, the spin, the electric charge and the colour charge of each particle is given. Fermions are arranged in pairs and form three generations of quarks and three generations of leptons. Each fermion has an anti-fermion partner with opposite charge, but same mass and spin. The mediators of the weak, strong and electric forces and their domains of interaction are also shown. The Higgs boson and its mass are also given.

1.2 Electroweak symmetry breaking and the Higgs boson

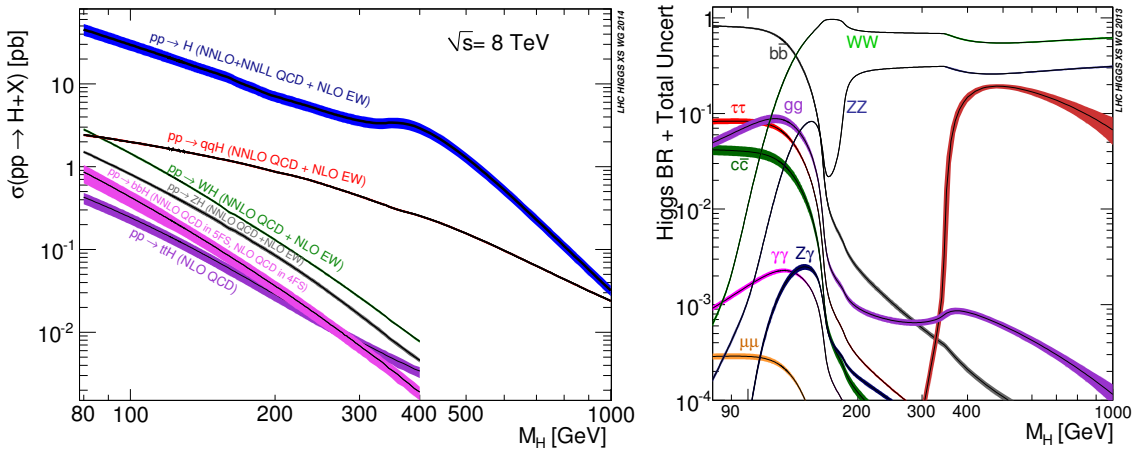
The addition to the Standard Model Lagrangian of a mass term for the fields associated to Standard Model particles, either fermions or weak bosons, would violate the gauge invariance. The acquisition of the mass by the weak bosons (V) occurs via the introduction a complex $SU(2)_L$ doublet scalar boson field, the Higgs field. This field is subject to a potential, whose minima occur at non-vanishing values of the field. The introduction of the Higgs field in the Lagrangian does not violate the gauge symmetry, however the field is forced into a ground state of the potential with a particular weak isospin causing the spontaneous breaking of the electroweak $SU(2)_L \otimes U(1)_Y$ symmetry. According to the Goldstone theorem [7], in a relativistic quantum field theory if the ground state is not invariant under the gauge symmetry of the Lagrangian, there are scalar particles with zero mass, the Goldstone bosons, corresponding to each broken symmetry generator (three in this case). Brout and Englert [8], Higgs [9, 10] and Kibble, Guralnik and Hagen [11] proposed that the massless Goldstone bosons could be combined with the massless fields of the weak bosons to form longitudinal polarization states of massive boson fields (BEH mechanism). This solution allows the weak bosons to acquire non-null mass preserving the gauge invariance. This mechanism also produces a massive neutral scalar, the Higgs boson, which couples directly to weak bosons but not to gluons and photons.

Fermions masses can also be generated via the Higgs field if Yukawa interaction terms [12, 13] between the Higgs and the fermion fields are introduced in the Lagrangian. This addition to the Lagrangian does not break the gauge symmetry. Since the Yukawa interaction term requires the existence of both chiral components of the fermion field, neutrinos are massless and do not couples with the Higgs field.

The BEH mechanism does not predict the vacuum expectation value v of the Higgs field, but determines the Higgs boson mass (m_H) and the couplings of the Higgs boson to fermions and vector bosons as a function of v. The couplings to fermions are proportional to the fermion masses and to the inverse of v, and the couplings to bosons are proportional to the squares of the boson masses and to the inverse of v (v^{-2} for quartic couplings).

The production cross-sections and the decay branching ratios (BRs) of the Higgs boson can be accurately calculated once the Higgs boson mass is known, as shown in Figure 1.1. In proton-proton collisions, the main Higgs boson production processes are in decreasing magnitude the gluon-gluon fusion (ggF), the vector boson fusion (VBF), the associated production with a vector boson (VH) and the associated production with a top anti-top quark pair ($t\bar{t}H$). The Feynman diagrams of these processes at leading order in the perturbative theory are illustrated in Figure 1.2. In the ggF production, the Higgs boson is produced from gluons via a loop of quarks: the heavier the quark, the higher the contribution to the ggF cross-section. The Top Yukawa interaction vertex has a role in the ggF and the $t\bar{t}H$ productions.

Since the Higgs boson directly couples with any massive Standard Model particle, the



(a) (b)

Figure 1.1: The cross-sections of the main Higgs boson production modes in proton-proton collisions for a centre-of-mass energy of 8 TeV (left) and the Higgs boson decay branching ratios (right) as a function of the Higgs boson mass [14] [15]. The coloured bands represent the theoretical uncertainties.

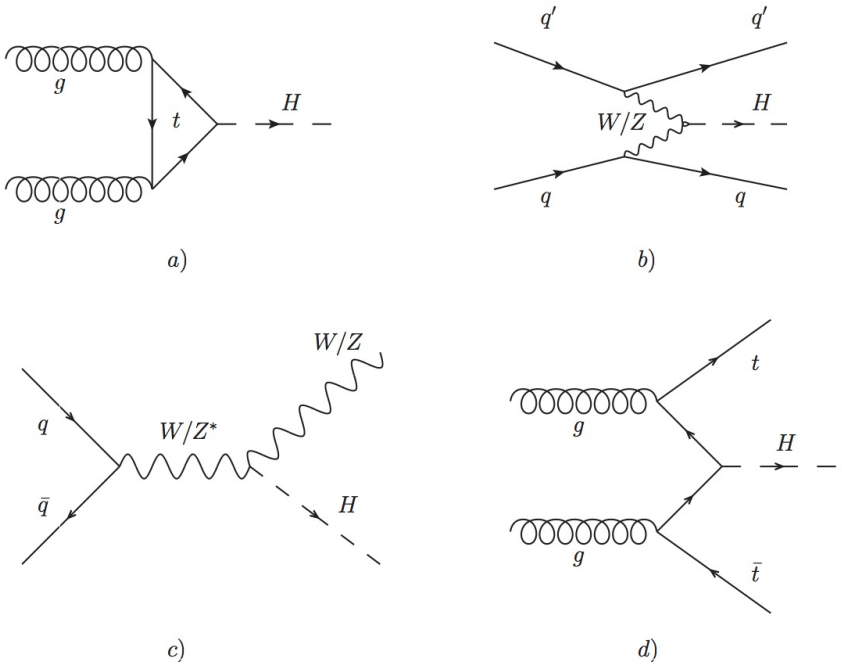


Figure 1.2: Leading-order Feynman diagrams of the main Standard Model Higgs boson production modes in proton-proton collisions: gluon-gluon fusion (a), vector boson fusion (b), associated production with a vector boson (c), associated production with a $t\bar{t}$ pair (d).

Higgs boson has many decay modes. The decays to a pair of gluons, $\gamma\gamma$ or $Z\gamma$ are also allowed via loops of heavy fermions and weak bosons. These decay modes are therefore influenced by the Top Yukawa coupling. For a Higgs boson mass of 125 GeV, the decay into $b\bar{b}$ and W^+W^- are dominant, and many other decay modes have substantial branching ratios. The branching ratio of the Higgs boson decay into photons is only of the order of 10^{-3}, however the detection of this process is experimentally easier compared to other decay modes.

1.3 Standard Model measurements and the Higgs boson discovery

The Standard Model of particle physics is an extraordinarily successful description of the fundamental constituents of matter and their interactions. Many experiments have verified the extremely precise predictions of the Standard Model in the past decades [16]. As an example of the predictive power of this theory, Figure 1.3 shows the measurements of the cross-sections for the most common production modes of elementary particles at the Large Hadron Collider (LHC) [17] with the ATLAS experiment [18]. The observations are well in agreement with the predictions of the Standard Model extended with the BEH mechanism, that is from now on just called the Standard Model for simplicity.

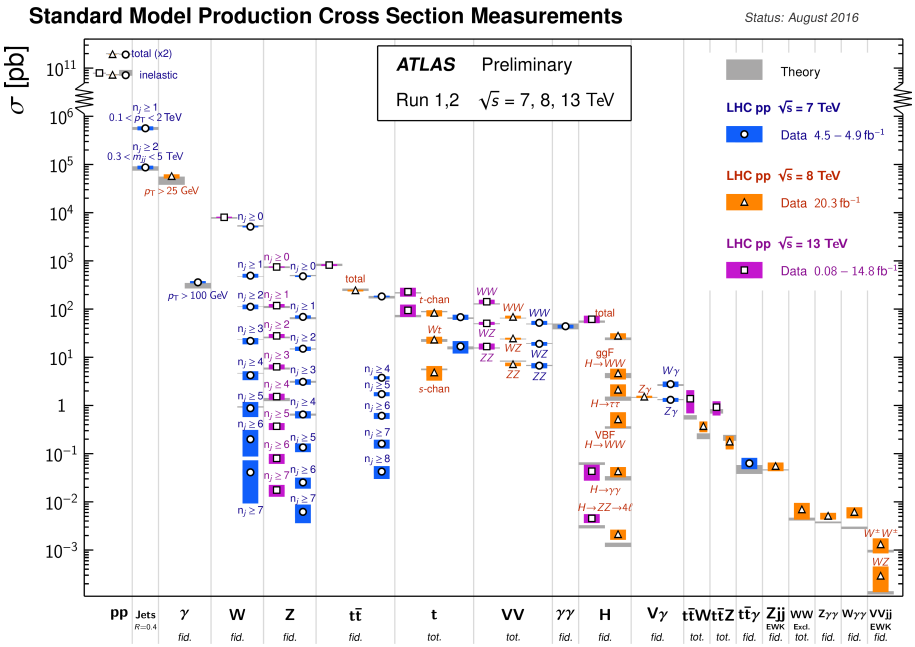


Figure 1.3: Detailed summary of several Standard Model cross-section measurements by ATLAS compared to the corresponding Standard Model expectations. Higgs boson boson production measurements are also included. All theoretical expectations were calculated at next-to-leading order in QCD perturbation theory or higher. Observations include statistical and systematic uncertainties.

For more then a decade, the Higgs boson has been the final missing piece of the Standard Model. Searches for the Higgs boson have been performed by the Large Electron Positron (LEP) experiments [19] and by the Tevatron experiments [20], which resulted in the exclusion of substantial m_H regions. The discovery of a new particle compatible with a Standard Model Higgs boson with mass of 125 GeV was announced by both ATLAS and CMS [21] in July 2012 [22, 23]. The Higgs boson boson was observed in the final states $H \to ZZ^{(*)}$, $H \to \gamma\gamma$ and $H \to WW^{(*)}$. Since then, stronger evidences have been established with additional data. The properties of the Higgs boson have been measured [24–26]. ATLAS and CMS have measured a mass for the boson of 125.09 ± 0.21 (stat.) ± 0.11 (syst.) GeV [26].

The first direct evidence of the Higgs boson Yukawa coupling to fermions was observed in the search for the $H \to \tau^+\tau^-$ decay [27]. From an experimental point of view, the calibration of the signal observed in the detectors due to the particles originating from the hadronic decay of the τ lepton plays an essential role for this analysis. The author was involved in this calibration and his studies are described in Chapter 3.

Confirmation of the coupling of the Higgs boson to quarks has been investigated in the searches for the $H \to b\bar{b}$ decay [28,29] and the $t\bar{t}H$ production [3,30–33]. The observations of the $H \to b\bar{b}$ analyses are compatible with Standard Model, however the sensitivity of the results is not high enough to constitute an evidence. The $t\bar{t}H$ production is a topic of this thesis and is studied in Chapter 6.

1.3.1 The $t\bar{t}H$ production

The $t\bar{t}H$ production has the smallest production cross-section among the Higgs boson production modes detectable at the Large Hadron Collider. For a proton-proton centre-of-mass energy of 8 TeV, its cross-section is $\sigma(t\bar{t}H) = 129^{+11}_{-16}$ fb [14]. The observation of $t\bar{t}H$ production is of particular importance because it has direct implications on the strength of the Top Yukawa coupling.

Table 1.2 shows the couplings involved in the production cross-sections and decay widths for a Higgs boson with 125 GeV of mass in Standard Model. The κ factors are called coupling modifiers and represent the coupling strengths measured in units of their expected values in Standard Model. The $t\bar{t}H$ production is the only process whose cross-section depends exclusively on the Top Yukawa coupling modifier (κ_t). The ggF production, the $H \to \gamma\gamma$ decay and other rarer processes, like the associated productions of the Higgs boson and a single t-quark, are sensitive to κ_t, but their amplitudes depend on a combination of several Higgs boson couplings. The diagrams of these processes indeed present loops that can be mediated by any massive particle. The heavier the particle, the larger the contribution to the amplitudes. Furthermore, if non-Standard Model particles exist and interact with the Higgs boson, they contribute to the amplitudes of such processes too. For instance, the ggF production could be affected by a forth generation of quarks.

Production	Loops	Interference	Multiplicative factor	
$\sigma(gg\mathrm{F})$	✓	$b-t$	$\kappa_g^2 \sim$	$1.06 \cdot \kappa_t^2 + 0.01 \cdot \kappa_b^2 - 0.07 \cdot \kappa_t\kappa_b$
$\sigma(V\mathrm{BF})$	–	–	$\sim$	$0.74 \cdot \kappa_W^2 + 0.26 \cdot \kappa_Z^2$
$\sigma(WH)$	–	–	$\sim$	κ_W^2
$\sigma(qq/qg \to ZH)$	–	–	$\sim$	κ_Z^2
$\sigma(gg \to ZH)$	✓	$Z-t$	$\sim$	$2.27 \cdot \kappa_Z^2 + 0.37 \cdot \kappa_t^2 - 1.64 \cdot \kappa_Z\kappa_t$
$\sigma(ttH)$	–	–	$\sim$	κ_t^2
$\sigma(gb \to WtH)$	–	$W-t$	$\sim$	$1.84 \cdot \kappa_t^2 + 1.57 \cdot \kappa_W^2 - 2.41 \cdot \kappa_t\kappa_W$
$\sigma(qb \to tHq)$	–	$W-t$	$\sim$	$3.4 \cdot \kappa_t^2 + 3.56 \cdot \kappa_W^2 - 5.96 \cdot \kappa_t\kappa_W$
$\sigma(bbH)$	–	–	$\sim$	κ_b^2
Partial decay width				
Γ^{ZZ}	–	–	$\sim$	κ_Z^2
Γ^{WW}	–	–	$\sim$	κ_W^2
$\Gamma^{\gamma\gamma}$	✓	$W-t$	$\kappa^2 \sim$	$1.59 \cdot \kappa_W^2 + 0.07 \cdot \kappa_t^2 - 0.66 \cdot \kappa_W\kappa_t$
$\Gamma^{\tau\tau}$	–	–	$\sim$	κ_τ^2
Γ^{bb}	–	–	$\sim$	κ_b^2
$\Gamma^{\mu\mu}$	–	–	$\sim$	κ_μ^2
Total width for $\mathrm{BR}_{\mathrm{BSM}} = 0$				
Γ_H	✓	–	$\kappa_H^2 \sim$	$0.57 \cdot \kappa_b^2 + 0.22 \cdot \kappa_W^2 + 0.09 \cdot \kappa_g^2 +$ $+ 0.06 \cdot \kappa^2 + 0.03 \cdot \kappa_Z^2 + 0.03 \cdot \kappa_c^2 +$ $+ 0.0023 \cdot \kappa^2 + 0.0016 \cdot \kappa_Z^2 +$ $+ 0.0001 \cdot \kappa_s^2 + 0.00022 \cdot \kappa^2$

Table 1.2: Higgs boson production cross-sections, partial decay widths and total decay width expressed in terms of the coupling modifiers (κ) in case of no beyond Standard Model particles [34].

This hypothesis has been widely excluded because heavier quarks would have a sizeable impact on the Higgs boson production cross-section [35]. However, there are beyond Standard Model theories, like Supersymmetry [36] and vector-like quark models [37], predicting new coloured particles which are allowed by present measurements [38–40]. Therefore, a direct measurement of the $t\bar{t}H$ cross-section has a dual purpose: to prove that the Top Yukawa coupling of the discovered boson is compatible with the Standard Model value; to investigate, in combination with other measurements, if non-Standard Model particles exist and contribute to certain production and decay processes of the Higgs boson.

Evidences of $t\bar{t}H$ production are searched for in proton-proton collisions producing leptons with same electric charge. This signature is the most sensitive to $t\bar{t}H$ production because the branching ratio is substantial and the background from other Standard Model processes is extremely low. The $t\bar{t}H$ production is the topic of Chapter 6.

1.4 Standard Model limitations and alternatives

Despite the numerous successes of the Standard Model at predicting the physics of elementary particles, this theory is not entirely flawless. There are several theoretical and experimental arguments that challenge it. In particular the Standard Model lacks of plausible explanations for issues such as the hierarchy problems [41], the matter-antimatter asymmetry in the universe [42, 43], the dark matter composition and the non-null mass of neutrinos. Plenty of models has been proposed to at least partially address these prob-

lems, however none of these theories has yet found compelling experimental validation. Several models, especially those addressing the neutrino mass generation problem, predict doubly charged particles and novel modes to produce pairs of leptons with same electric charge (same-sign lepton pairs). The study of the production of same-sign lepton pairs in proton-proton collision could show evidences of physics beyond Standard Model and can be used to test a wide range of theories.

Same-sign lepton pairs are produced at the LHC with a rate larger than that predicted by the Standard Model in Supersymmetry [44], models with universal extra dimensions [45], Left-right symmetric models [46–49], the little Higgs model [50], and models introducing vector-like quarks [51] or fourth generation quarks [52–54].

Supersymmetry [36] extends the Standard Model introducing a new gauge symmetry between bosons and fermions which leads to the creation of new set of fundamental particles. These particles own the same quantum numbers of the Standard Model particles with the exception of the spin which differs by one-half unit. The masses of particles and sparticles are not related. Supersymmetry introduces a mechanism to generate baryon asymmetry compatible with cosmological observations, provides candidate dark matter particles and addresses the hierarchy problem.

The production of sparticles can lead to the generation of same-sign lepton pairs in plenty of ways [44]. The free parameters of the model and especially the sparticle mass hierarchy determine these production modes. For instance, in one scenario, the gluino can decay into quarks and the lightest chargino. The chargino itself then produces a lepton, a neutrino and a neutralino via an intermediate decay of a $W^{\pm}$ or a slepton. Since the gluino is a Majorana fermion, the chargino and so the lepton produced in its decay can have either positive or negative charge. Therefore, the production of gluino pairs at the LHC would generate final states with two leptons whose electric charges coincide half of the times.

Extra dimensions models postulate that the gravitational field propagates into the additional dimensions. This would explain the relative weakness of gravity, hence would solve the hierarchy problem [41]. Because of the extra dimensions, the threshold of production of microscopic black holes at the LHC could be lowered to the TeV scale allowing their creation [55]. The process of black hole evaporation could then produce Same-sign pairs either directly or as decay products of other Standard Model particles [45].

Left-right symmetric models introduce a new symmetry: parity. The electroweak group becomes $SU(2)_L \otimes SU(2)_R \otimes U(1)_{Y_{L+R}}$ [56], but the symmetry is spontaneously broken into the Standard Model $SU(2)_L \otimes U(1)_Y$ at the energy scales explored so far. The model allows for right-handed Majorana neutrinos (N_R) and the right-handed fermions form doublets of $SU(2)_R$. The Higgs field expands into a bi-doublet of $SU(2)_L \otimes SU(2)_R$;

triplets of $SU(2)_L$ and $SU(2)_R$ are also introduced [57]. The breaking of the parity symmetry occurs spontaneously at higher energy scales due to higher vacuum expectation values of the additional Higgs fields compared to the Standard Model Higgs field. This produces heavy right-handed gauge fields in a similar fashion to the BEH mechanism: the right-handed $W^{\pm}$ boson (W_{R}) and the right-handed Z boson (Z_{R}). After the parity breaking, the electroweak symmetry collapses into the Standard Model $SU(2)_L \otimes U(1)_Y$, which, at lower energies, undergoes the traditional spontaneous breaking expected in Standard Model.

The introduction of the additional Higgs fields, beyond giving mass to the gauge bosons, produces four real scalars, two real pseudo-scalars, two singly charged complex scalars and two doubly charged complex scalars ($H_{\mathrm{L}}^{\pm\pm}$ and $H_{\mathrm{R}}^{\pm\pm}$) [58]. One scalar, the equivalent of the Higgs boson, has mass at the electroweak scale, while the other particles are predicted at the scale of the parity symmetry breaking. The Left-right symmetric model predicts physics at low energies similarly to the Standard Model, however restores the parity symmetry at higher energies and provides an explanation for the non-null neutrino masses and their smallness[1].

The Left-right symmetric model allows for several processes of production of same-sign lepton pairs [49]. The W_{R} boson can decay into a lepton and a N_R neutrino, whose decay produces another lepton and additional particles. The produced leptons can have same electric charge. Further, the doubly charged bosons, $H_{\mathrm{L}}^{\pm\pm}$ and $H_{\mathrm{R}}^{\pm\pm}$, can decay directly into same-sign lepton pairs.

Little Higgs models offer another approach to the hierarchy problem [59]. In these models the Higgs boson arises as a pseudo-Goldstone boson whose mass is stabilized by approximate global symmetries. In these models, the Standard Model is just an approximate effective theory and there are new states at energies of TeV which prevent the Higgs boson mass to diverge. The Little Higgs models present an extended Higgs sector including a triplet which leads to a doubly charged boson with mass at the TeV scale [50]. The doubly charged boson could be detected at the LHC as it would decay into resonant same-sign lepton pairs.

Fourth generation quarks and vector-like quarks. The existence of another generation of quarks would open to new ways of same-sign lepton pair production. For instance, a fourth generation down-type quark, b', is supposed to decay into tW^-, thus the production of a $b'\bar{b}'$ pair would lead to $t\bar{t}W^+W^-$ and ultimately to same-sign lepton pairs. The subsistence of a fourth generation of quarks has been highly undermined by the discovery of the Higgs boson and the precision measurement of its properties [60]. Since the fourth generation quark would be the heaviest among the Standard Model particles, they

[1]The Left-right symmetric model can produce neutrino masses via Type-I and Type-II see-saw mechanisms described in Section 1.4.1.

would significantly enhance the cross-section of ggF production and suppress the $H \to \gamma\gamma$ branching ratio, none of which seems to be observed.

A category of coloured fermions, which is not yet ruled out by experimental observations is constituted by vector-like quarks [37]. For these particles, the left-handed and right-handed chiral fields transform in the same way under the Standard Model gauge group, and both chiralities couple with the weak bosons. Vector-like quarks introduce new sources of matter-antimatter asymmetry and have effects on the coupling of quarks with the weak bosons and the Higgs boson [37]. Hadron colliders may produce pairs of same-sign vector-like quarks, like $u_- u_-$ or $u_- \bar{d}_-$, which would produce same-sign $W^{\pm}$ bosons and eventually a same-sign lepton pair [51].

Neutrino mass models propose mechanisms to give small masses to neutrinos which are necessary to explain the observation of neutrino oscillations [61,62]. Non-Standard Model production of same-sign lepton pairs is expected in the most notorious of these theories, like the see-saw models [63–70] and radiative models [71–74]. A brief overview of these theories is given in Section 1.4.1.

In conclusion, a wide range of theories beyond the Standard Model predict additional same-sign lepton pair production, which are therefore an interesting target for experiments at hadron colliders. The production rate of same-sign lepton pairs at the LHC is studied in this thesis. A search for doubly charged particles decaying to same-sign lepton pairs is also performed. The Standard Model physics is expected to give a low background due to the rarity of processes producing same-sign lepton pairs increasing the potential for new discoveries.

1.4.1 The neutrino mass problem

At the time of the Standard Model and the BEH mechanism formulations, neutrinos were considered massless. The evidences from the beta decay studies were in favour of neutrinos produced with singular chirality (left-handed for neutrinos and right-handed for anti-neutrinos) and massless. In Standard Model, right-handed neutrino (and left-handed anti-neutrino) fields do not exist, so there is no Yukawa interaction with the Higgs boson and as a consequence no mass.

The Super-Kamiokande and SNO collaborations published in the late 90s evidences that neutrinos could change flavour in-flight. Super-Kamiokande measured a reduced fraction of muon neutrinos from cosmic rays [61]. SNO found a deficit in the flux of electron neutrinos from the sun, as well as the presence of a non-electron flavour component in the flux [62]. This seems in contrast with the models of solar fusion where only electron flavoured neutrinos could be produced. In reality, the results from Super-Kamiokande and SNO embrace the concept of neutrino oscillations, where a neutrino has a periodical prob-

ability to be detected in a different flavour state as it propagates through space. In order to have non-null oscillations, the neutrino mass eigenvalues cannot be degenerate, and the mass eigenstates must be non-trivial combinations of the flavour eigenstates. This means that not all the neutrino mass states can be massless. A direct measurement of neutrino masses has not been achieved yet, however cosmological data constraint the sum of the neutrino masses to less than 0.3 eV [75, 76]. If neutrinos are Majorana particles, a limit on their mass of 61-165 meV has been derived from double beta decay experiments [77].

The simplest way to give mass to neutrinos is to suppose the existence a right-handed neutrino field. This field would only show up in the Standard Model to give the neutrino a-la Dirac mass, similarly to the other fermions, but otherwise would be inert because it does not couple to weak, electromagnetic and strong bosons[2]. This approach puts all the fermions on the same level, but is not experimentally verifiable and gives no explanation on the smallness of the neutrino mass. In this scenario, the Yukawa coupling of neutrinos would be at least 10^6 times smaller than that of the charged leptons, while, in the quark sector, the couplings do not differ by more than a factor hundred within the same generation. The smallness of the neutrino coupling, therefore mass, may be accidental but there are models where this emerges as a natural consequence of the theory.

If the neutrino is Majorana particle, i.e. is its own antiparticle, it is possible to define a neutrino mass term which does not violate the gauge symmetry introducing new physics at higher energy (TeV or higher). The simplest case consists in the introduction of a right-handed Majorana neutrino N_{R} (for each generation), which has a Yukawa coupling with the left-handed neutrino and a Majorana mass term. The presence of the interaction term between the right-handed and left-handed neutrino fields causes the neutrino mass eigenstates to be a linear combination of the chiral states. The two mass eigenvalues for the neutrino fields are $m_1 = m_{\mathrm{D}}^2/m_{\mathrm{R}}$ and $m_2 = m_{\mathrm{R}}\left(1 + m_{\mathrm{D}}^2/m_{\mathrm{R}}^2\right)$, where m_{R} and m_{D} are respectively the coefficients of the Majorana mass term and of the Yukawa term when the Higgs field assumes its vacuum expectation value. If $m_{\mathrm{R}} \gg m_{\mathrm{D}}$, m_1 becomes much smaller than m_2 and the associated eigenstate aligns with the left-handed chiral state. On the other hand, the right-handed state aligns with the heavy neutrino eigenstate assuming a mass of about m_{R}. Therefore in this model, the light mass of the observed neutrino comes from the presence of an heavy partner at some energy scale not yet observed. This mechanism goes under the name of Type-I see-saw model [63, 64].

The mass of the heavy Majorana neutrino is expected to be as large as 10^{15} GeV for m_{D} of the order of 10^2 GeV in the context of the Left-right symmetric model and in grand unification theories [78]. If m_{D} is taken of the order of the electron mass, a light neutrino mass of 0.1 eV could be obtained with a partner as heavy as few TeV, but it would not be possible to produce the heavy neutrino due to its tiny Yukawa coupling [79]. Heavy neutrinos with masses of the order of TeV may be generated at a significant rate only in

[2]The right-handed neutrinos would interact with the Higgs boson, however the coupling strength is so small that the contribution of this interaction to any process is experimentally unnoticeable.

case of structural cancellations [79–81]. Proton-proton collisions can generate a charged lepton in association with a heavy neutrino, which can consequently decay into a charged lepton and a $W^{\pm}$. This process would be observable as an enhancement of the rate of same-sign lepton pair production at the LHC.

There are alternative ways to introduce a neutrino mass in the Standard Model via new physics that may be explorable at the LHC. They are the Type-II [65–68] and Type-III [69,70] see-saw mechanisms. In the latter model, the right-handed Majorana neutrinos are replaced by Majorana fermion $SU(2)_L$ triplets Σ composed by charged and neutral fields: $\Sigma^{\pm}$ and Σ^0. The Lagrangian now contains the interaction terms between the Higgs boson, the left-handed lepton and the Σ fields (Similarly to a Yukawa term but with the fermion triplets in place of the fermion singlets) and mass terms for Σ. Also in this case the mass of the neutrinos are proportional to the inverse of the masses of the new heavy fermions $\Sigma^{\pm}$ and Σ^0, which are expected at the same scale within the same generation [82]. The associated production of $\Sigma^{\pm}$ and Σ^0 can lead to a final state containing a same-sign lepton pair [82].

The Type-II see-saw mechanism assumes the existence of a new scalar $SU(2)_L$ triplet field $\phi = \left(\phi^0, \phi^+, \phi^{++}\right)$ with its own vacuum expectation value $v_\phi \propto v^2/m_\phi^2$, where v is the vacuum expectation value of the Standard Model Higgs boson and m_ϕ is the mass scale of the new bosons emerged from the theory [83]. There are six massive bosons in addition to the Higgs boson, including two singly charged ($\phi^{\pm}$) and two doubly charged ($\phi^{\pm\pm}$). Neutrinos obtain mass via new Yukawa-like terms where two left-handed fermion doublets couple to ϕ. The neutrino masses are proportional to v_ϕ and the Yukawa couplings. With appropriate choices of the Yukawa couplings, a m_ϕ of the order of 1 TeV or smaller produces neutrino masses and mixing consistent with the experimental observation [83]. Therefore this mechanism could produce new physics observable at the LHC. The most peculiar aspect of this theory is the existence of doubly charged bosons which decay predominantly in same-sign lepton pairs if $v_\phi < 10^{-4}$ GeV, and $W^{\pm}W^{\pm}$ otherwise [83]. The $\phi^{\pm\pm}$ can be produced at the LHC in pairs or in association with a singly charged boson ϕ^+ or a $W^{\pm}$ boson. These processes generate one or more resonant same-sign lepton pairs, which is inconceivable for Standard Model processes.

The problem of the neutrino mass has been more recently approached using effective operators, and has led to the determination of 60 effective operators which may give small Majorana neutrino masses [84]. These operators can be expanded into ultraviolet-complete renormalizable models. The expansion of one of these operators (the Weinberg operator) leads to the see-saw models, which are tree level ultraviolet-complete models; the other operators lead to radiative mechanisms of neutrino mass generation [85]. In these models, the neutrinos acquire small masses via loop diagrams mediated by new physics particles. The neutrino mass is typically proportional to m_χ^{-i}, where m_χ is the mass of the new particle(s) and the power coefficient i can be greater than one allowing new physics at lower energies.

The Zee-Babu model [71, 72] is one of the most famous radiative neutrino mass models. It introduces a singly charged scalar $SU(2)_L$ singlet ($h^{\pm}$) and a doubly charged scalar $SU(2)_L$ singlet ($k^{\pm\pm}$). These fields contribute to a two-loop diagram of neutrino mass generation, which is shown in Figure 1.4(a). For masses of the new scalars above 100 GeV [73], this model provides neutrino masses compatible with experimental observations. The doubly charged scalars can be produced in pair at the LHC and would be detected as resonant pairs of same-sign leptons.

The Coloured Zee-Babu model proposes another way to generate tiny neutrino masses at the two-loop level by using scalar leptoquark ($S_{\rm LQ}$) and diquark ($S_{\rm DQ}$) multiplets [74]. Leptoquarks are a triplet under $SU(3)_C$ and singlets under $SU(2)_L$, while diquarks are a **6** representation of $SU(3)_C$ and singlets under $SU(2)_L$. The neutrino mass is generated via the loop shown in Figure 1.4(b). The diquark can be singly produced at the LHC, and it can decay into a pair of leptoquarks. The leptoquarks would then decay into leptons and quarks with the leptons forming a same-sign lepton pair.

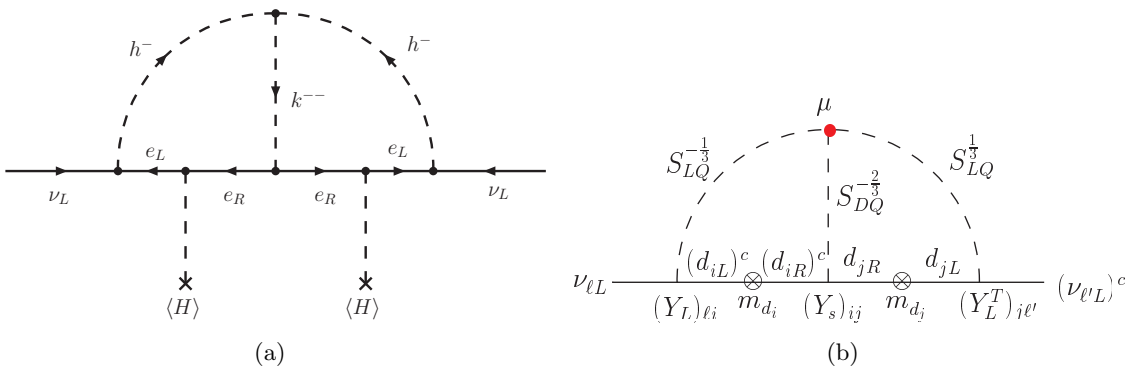


Figure 1.4: Loop diagrams of neutrino mass generation in the classic (left) and Coloured (right) Zee-Babu models [73, 74].

In summary, many neutrino mass generation models introduce new elementary particles, whose production at proton-proton colliders would lead to events with same-sign leptons in the final state. Non-Standard Model production of same-sign lepton pairs at LHC is studied in Chapter 5 of this thesis with a focus on doubly charged resonances typical of the Type-II see-saw and the Zee-Babu models.

Chapter 2

Experimental apparatus

The Large Hadron Collider (LHC) [17] is an particle accelerator that provides the most energetic proton-proton collisions in the world. As such, it allows investigation at energy scales never before explored, where the Standard Model has not been yet tested and new physics could emerge. This chapter gives a brief overview of the LHC and introduces the ATLAS detector [18]. The techniques used to simulate the physics at the LHC and the methods to reconstruct particle properties from detector signals are also described.

2.1 The Large Hadron Collider

The Large Hadron Collider at CERN (Geneva, Switzerland) is designed to accelerate proton beams and to produce proton-proton collisions at a centre-of-mass energy of 14 TeV and at a luminosity of 10^{34} cm^{-2}s^{-1}. The accelerator is also used to collide heavy ions. The LHC is installed in an underground ring with a circumference of 27 km.

Protons are obtained from hydrogen atoms which are stripped of their electrons. These protons are accelerated through several stages as shown in Fig. 2.1. They are first accelerated to 50 MeV by a linear accelerator (LINAC2) and then injected progressively into the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS). In these storage rings the proton energy increases to 11.4 GeV, 26 GeV and 450 GeV respectively. They are also compacted into bunches of about 10^{11} protons. These bunches are finally injected into the main LHC ring in both direction of motion, where they are accelerated to the final collision energy. There are four interaction points where the experiments ATLAS, CMS [21], LHCb [86] and ALICE [87] are located. The LHC has collided protons at centre-of-mass energies of: $\sqrt{s} = 7$ TeV in 2011, $\sqrt{s} = 8$ TeV in 2012 and $\sqrt{s} = 13$ TeV in 2016. The analyses presented in this thesis use the $\sqrt{s} = 8$ TeV collision data.

Beside the collision energy, an important characteristic of an accelerator is the luminosity ($\mathscr{L}$) which determines the collision rate. The luminosity is a measurement of the density of interaction centres expressed per unit area and unit time. The product of the

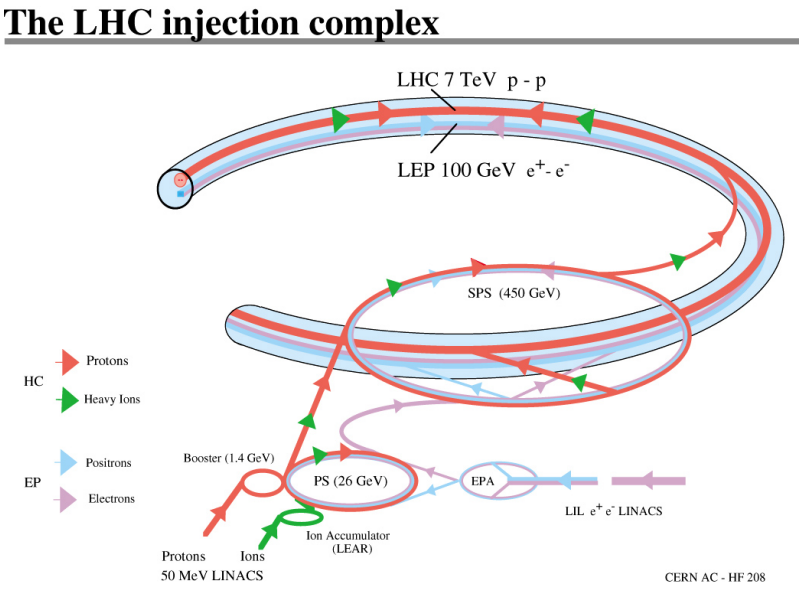


Figure 2.1: Schematic view of the LHC accelerator complex. [88]

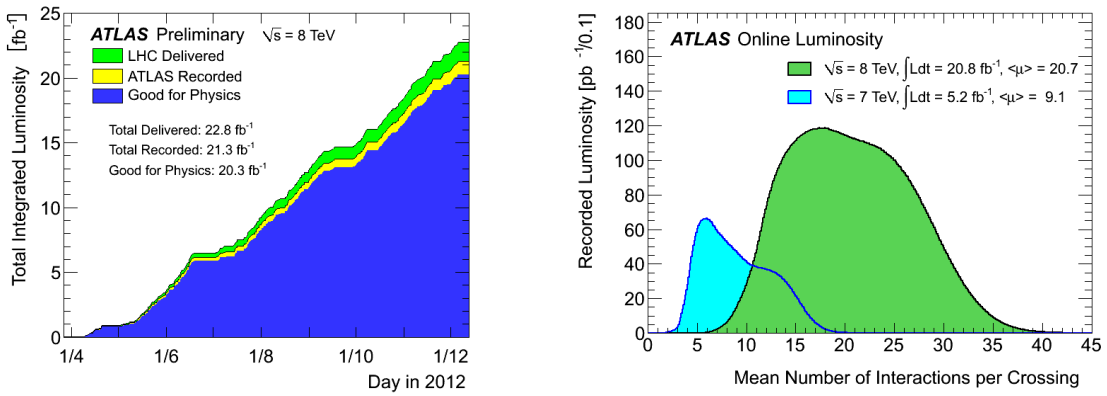
luminosity and the cross-section of a process provides the production rate of that specific process. To study rare processes, high luminosity and long acquisition times are necessary conditions.

At the LHC, the luminosity is a function of the number of bunches in each beam, the beam revolution frequency, the number of proton in each bunch, the spatial distribution of protons in the bunch and the crossing angle between the directions of interacting bunches [17]. During 2012, with standard acquisition conditions, the LHC was operating with 1380 bunches per beam, 50 ns time between consecutive bunch crossings and about 1.6×10^{11} protons per bunch, resulting in a luminosity of the order 10^{33} cm^{-2}s^{-1}. An accurate measurement of the luminosity at the ATLAS interaction point is performed using several luminosity-sensitive detectors and knowledge of the total inelastic proton-proton cross-section [89]. The maximum luminosity recorded in 2012 was 7.7×10^{33} cm^{-2}s^{-1}.

The measured luminosity for $\sqrt{s} = 8$ TeV collisions integrated over time is shown in Figure 2.2(a). Excluding the collisions that occurred during non-stable conditions and during the ATLAS detector downtime (less than 5%), the integrated luminosity of the $\sqrt{s} = 8$ TeV collisions recorded by ATLAS is 20.3 fb^{-1}. This is the dataset analysed in this thesis.

The number of protons per bunch and their spacial density are so high that multiple proton-proton inelastic interactions take place at each bunch crossing. In 2012, the average number of interactions per bunch crossing $\langle\mu\rangle$ was 20.7, however this condition was variable during data acquisition. The distribution of the measured number of mean[1] interactions per bunch crossing in 2012 data is shown in green in Figure 2.2(b). Since detectors have

[1]The measurement is generally the average of a few seconds of acquisition.



(a) (b)

Figure 2.2: (a): Integrated luminosity delivered over time (green), and recorded by ATLAS (yellow) during stable beams for proton-proton collisions at $\sqrt{s} = 8$ TeV. The integrated luminosity of the dataset acquired with no detector failures and used in thesis is also shown (blue). (b): Number of average interactions per bunch crossing for the ATLAS dataset collected in 2011 (light blue) and 2012 (green). [89]

natural limitations on spatial and temporal accuracy, there is a non-negligible probability that particles from independent interactions can be produced close enough in space and time to alter the performance of the detector in reconstructing the particles from the scattering event of interest. This effect is called pileup.

The processes studied in this thesis produce particles with large transverse momentum (p_{T}) and the proton-proton collisions from which they originate are referred as "hard" interactions. The vast majority of the proton-proton interactions however, generate particles with small p_{T} and small scattering angle ("soft" or pileup interactions) [90]. The collision of interest will typically have a hard interaction and tens of pileup interactions. The particles from pileup interactions generate noise overlaying the signal produced by the particles from the hard scattering in the detectors.

There are two types of pileup: *in-time* and *out-of-time*. The *in-time* pileup is generated by the collisions occurring in the same bunch crossing, also called "event", of the hard interaction and depends on the size and proton density of the bunches. The *out-of-time* pileup is due to collisions that occurred in different bunch crossings. Some of the ATLAS detector components have integration times and readout latencies larger than 50 ns. Collisions generated in the events preceding and following the process of interest can therefore generate additional noise in these components.

Pileup affects the kinematic and identification efficiency of the reconstructed particles. Corrections are studied and applied to accurately measure the properties of the hard scattering.

2.2 The ATLAS detector

ATLAS [18] is a multi-purpose detector designed to identify and reconstruct the kinematic behavior of a large variety of particles. This allows the study of a wide range of physics processes including new physics. The detector features radiation-hard electronics and sensor elements to cope with the challenging experimental conditions. Detector elements are designed to have high granularity and fast acquisition timing to reduce the superimposition of signals from the multitude of particles generated in the collisions. The ATLAS detector, shown in Fig. 2.3, has a barrel shape with the axis along the beam direction; it has a diameter of 25 m and a length of 44 m. It is equipped with high-resolution complementary detectors and magnets to allow the identification of charged leptons, photons, jets of hadrons and non-interacting particles. ATLAS is particularly accurate in the measurements of charged leptons and photons. Moving from the interaction point outwards, ATLAS consists of tracking detectors, a superconducting solenoid, calorimeters and muon detectors surrounded by toroid magnets. The sensitive area of the detector system almost completely covers the solid angle around the interaction point, excluding small areas around the beam direction. The detector is divided into five sections of pseudorapidity[2]: a central part (small η) called the *barrel* region, two lateral disk-shaped parts, the *end-cap* regions; and two annular parts close to the beam axis that form the *forward* regions.

The analyses described in this thesis are based on signatures from electrons, muons, tau leptons, jets and missing transverse momentum that are reconstructed combining information from all the detector sub-systems. A brief description of each element is provided in the following, after the coordinate system is introduced.

2.2.1 Coordinate system

The beam direction defines the Z-axis of a right-handed coordinate system whose origin is the nominal interaction point within the detector. The X-Y plane is transverse to the beam direction with the X-axis directed from the origin to the centre of the LHC ring, and the Y-axis pointing upwards. It is convenient to introduce the coordinates $\phi \in [0, 2\pi)$ and $\eta \in (-\infty, +\infty)$ which are respectively the azimuthal angle in the plane transverse to the beam direction and the pseudorapidity $\eta = -\ln[\tan(\theta/2)]$, where θ is the polar angle. For a relativistic particle, the pseudorapidity coincides with the rapidity, and together with ϕ and p_{T} completely defines the kinematic of the particle. Pseudorapidity is preferable to the polar angle because differences in pseudorapidity $\Delta\eta$ (as well as differences in the azimuthal angle $\Delta\phi$) for relativistic particles are Lorentz invariant under boosts along the beam axis. This also allows to define an angular distance ΔR between two particles a and

[2]See Section 2.2.1 for the definition of pseudorapidity.

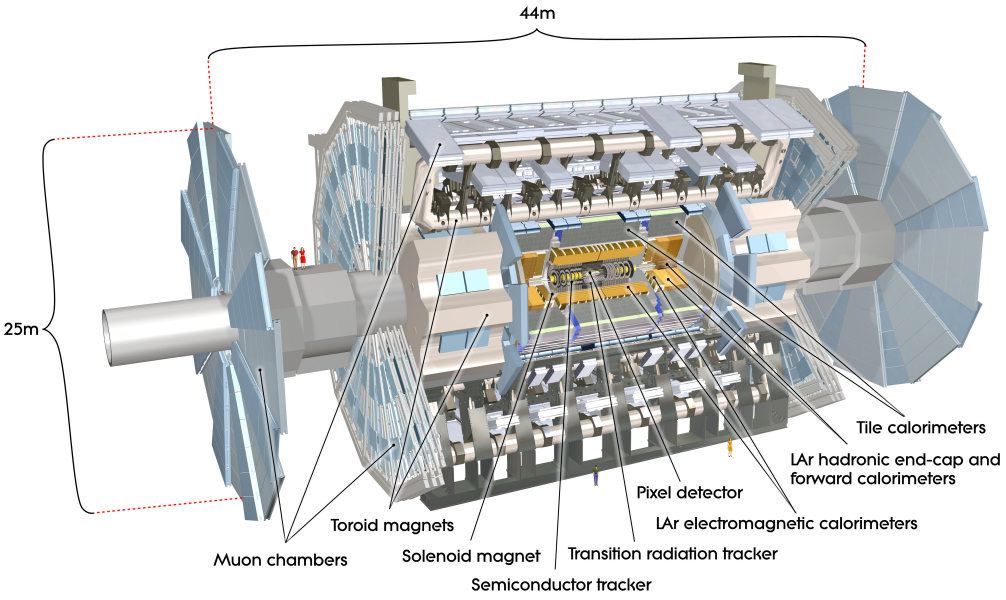


Figure 2.3: Section of the ATLAS detector.

b, which is Lorentz invariant under longitudinal boosts:

$$\Delta R\left(a,b\right)=\sqrt{\Delta\eta\left(a,b\right)^{2}+\Delta\phi\left(a,b\right)^{2}} \tag{2.1}$$

2.2.2 Magnet system

The magnet system is fundamental for an accurate measurement of charged particle properties. Since charged particles are subject to the Lorentz force, their trajectory is bent by the magnetic field and depends on their momentum and charge. An accurate measurement of the trajectory of a charged particle, from now on referred to as its "track", combined with the knowledge of the magnetic field allows to reconstruct the charge[3] and the momentum of the particle. The stronger the magnetic field, the more pronounced the curvature and the better the accuracy of the momentum measurement.

The magnetic system used by ATLAS is show in Figure 2.4. ATLAS is equipped with a central solenoid surrounding the inner tracking detectors that provides a 2 T axial magnetic field with retuning flux in the calorimeters. The inner diameter of the solenoid is 2.46 m, and it is 10 cm thick to minimize the material in front of the calorimeters.

Additional magnetic fields are generated outside the calorimeters by three toroids consisting of eight coils each arranged radially and symmetrically around the beam axis. This magnet configuration provides a field which is mostly orthogonal to the particles escaping the calorimeter allowing a total momentum measurement. The *barrel* toroid has 25.3 m long coils covering $|\eta| < 1.4$ and providing a variable bending power between 2 and 6 Tm.

[3] A charged particle traversing the detector is assumed to have charge $+1$ or -1 times the electron charge, therefore only the sign of the charge is measured.

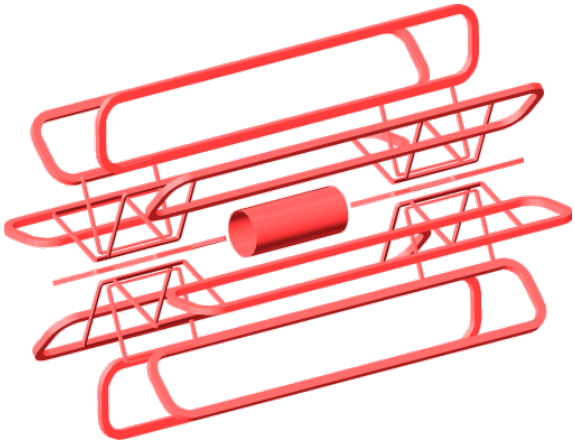
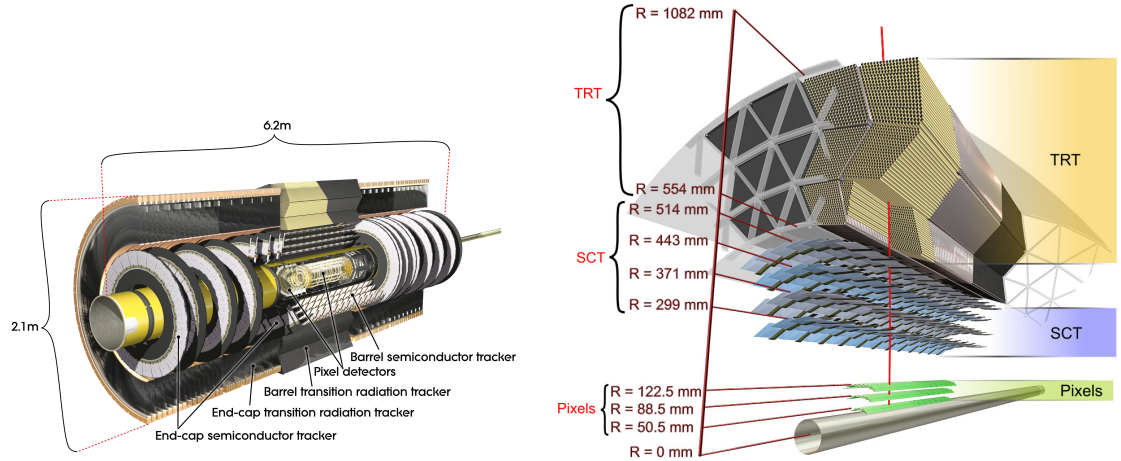


Figure 2.4: ATLAS magnet system composed by three toroids formed by eight coils each surrounding the central solenoid.



(a) (b)

Figure 2.5: (a): Cutaway view of the Inner Detector. (b): Section of the *barrel* region showing the distances of the PIXEL, the SCT and the TRT layers from the beams.

The two smaller *end-cap* toroids provide a bending power between 4 and 8 Tm in the region $1.6 < |\eta| < 2.7$. The magnetic fields of *barrel* and *end-cap* toroids combine in the region $1.4 < |\eta| < 1.6$ but result in a smaller bending power.

2.2.3 Inner Detector

The Inner Detector consists of three complementary tracking devices that give full coverage in ϕ and up to $|\eta| < 2.5$ in pseudorapidity. The closest detector to the interaction point is the Pixel Detector (PIXEL) followed by the Silicon Microstrip Tracker (SCT) and the Transition Radiation Tracker (TRT). The positions of the PIXEL, the SCT and the TRT detectors are shown in Fig. 2.5. The PIXEL consists of three *barrel* layers and three *end-cap* discs containing 80 million pixels with granularity $\sigma_{R-\phi} \times \sigma_Z = 10 \times 115\ \mu\text{m}^2$ and

low occupancy time. The Silicon Microstrip Tracker (SCT) is built of four double-layer cylinders with crossing strips in the *barrel* region and nine double-layer disks at each *end-cap* with a position resolution of $\sigma_{R-\phi} \times \sigma_Z = 17 \times 580\ \mu\text{m}^2$. The TRT is made of 351000 gas-filled (70% Xe, 27% CO_2 and 3% O_2) drift straw tubes forming 73 cylindrical layers in the *barrel* region and 160 wheels in the *end-cap* regions. The location of a transiting charged particle is obtained from the time of collection (drift time) of electrons produced by gas ionization on the anode. The resolution on the passage position is 130 μm. The design of the TRT is such that the passage of electrons between the gas and the straw tube wall produces transition radiation that generates additional Xe ionization. Therefore, the signal measured at the anode when an electron traverses the tube is higher. This property is used to discriminate electrons from other charged particles.

The Inner Detector (IDet) provides the position of transit (hit) of charged particles through its sensitive components. Hits in the PIXEL, SCT and TRT are then combined using a pattern recognition algorithm to reconstruct particle tracks. The IDet modules are characterized by high granularity and precise alignment in order to provide an excellent measurement of the track of a charged particle. From the curvature of the track, the transverse momentum of the particle is obtained. Charge and direction are also measured. Interaction vertices are reconstructed combining tracks. The impact parameter of a track to the interaction vertex is also measured. This information is used to determine whether the particle was produced in the hard interaction, in a pileup interaction or in the decay of a particle with a measurable decay length. IDet tracks are used in the reconstruction of electrons, muons and tau leptons and in the determination of the origin of jets of hadrons as described in detail in Section 2.4.

All the detectors are designed to have small radiation lengths to minimize energy loss, multiple scatterings and the production of secondary particles. Each detector contributes similarly to the accuracy of the momentum measurement, however PIXEL hits count more in the measurement of vertex positions and impact parameters. Combining the information from all the detectors, the IDet measures the transverse momentum (p_{T}) of charged particles with a relative uncertainty of $\frac{\sigma_{p_{\text{T}}}}{p_{\text{T}}} = 0.05\% \cdot p_{\text{T}} \oplus 1\%$.

2.2.4 The Calorimeter System

The primary role of the ATLAS calorimetry is to measure the energy of hadrons, electrons and photons. The calorimeters also provide directional information thanks to segmented modules. The ATLAS calorimeter system surrounds the IDet and solenoid magnet and extends up to $|\eta| < 4.9$. The ATLAS calorimeters are sampling devices, i.e. they alternate layers of active material with dense layers of passive material that slow down particles. Different techniques are used depending on the physics requirements and intensity of the radiation in the covered region. Fig. 2.6 shows a section of the ATLAS calorimeter system that is composed of an Electromagnetic Calorimeter (ECal) enclosed

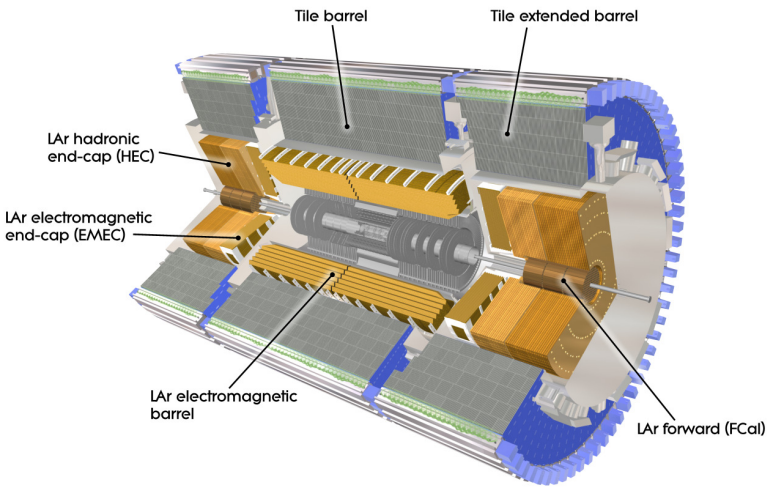


Figure 2.6: Section of the ATLAS calorimeter system.

in Hadronic Calorimeters (TileCal and HEP) and Forward Calorimeters (FCal) covering $3.1 < |\eta| < 4.9$.

The electromagnetic calorimeters are lead-liquid argon (LAr) detectors with accordion-shaped absorbers interleaved with readout electrodes. The ECal is composed of a barrel covering $|\eta| < 1.475$ and two coaxial wheels at each *end-cap* covering $1.375 < |\eta| < 3.2$. For $|\eta| < 2.5$, the same region covered by the IDet, the ECal is divided in three longitudinal layers with the first having a finer granularity to distinguish signals generated by close particles. A "presampler" is placed between the solenoid and calorimeter for $|\eta| < 1.8$ to evaluate the energy loss by particles in the passive material in front of the calorimeter; the thickness of this passive material varies between 2 and 6 radiation lengths (X_0). The total depth of the electromagnetic calorimeter varies with η, but is generally greater than 22 X_0 as shown in Figure 2.7(a-b).

The main purpose of the ECal is the detection of electrons and photons and the measurement of their energy and direction. Electrons and photons form electromagnetic showers in the calorimeters. The characteristics of the energy deposit (cluster) provide information about the type and energy of the interacting particle. The energy resolution for electrons and photons is $\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\%$ in both *barrel* and *end-cap* calorimeters, but worsens in the transition region $1.37 < |\eta| < 1.52$. Electrons falling in this region are excluded from the analyses presented in this thesis.

Hadrons interact with calorimeters differently to electron and photons. Their energy deposit is generally wider and extends more in dept. The ECal is about two interaction lengths (λ) thick, which is not enough to completely absorb high energy hadrons. The hadronic calorimeters are located outside the Ecal to contain hadrons and to measure their energy and position.

The *barrel* part of the hadronic calorimeter is a sampling calorimeter made of steel and

plastic scintillator tiles (TileCal). It covers $|\eta| < 1.7$ and is segmented in three layers. The Hadronic End-cap Calorimeter (HEC) covers $1.5 < |\eta| < 3.2$ and is a LAr-copper-sampling calorimeter. It sustains more radiation than the TileCal, but has the same design energy resolution for hadrons of $\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\%$. The total thickness of the calorimeter system is shown in Figure 2.7(c) as a function of the pseudorapidity and is about 10 λ.

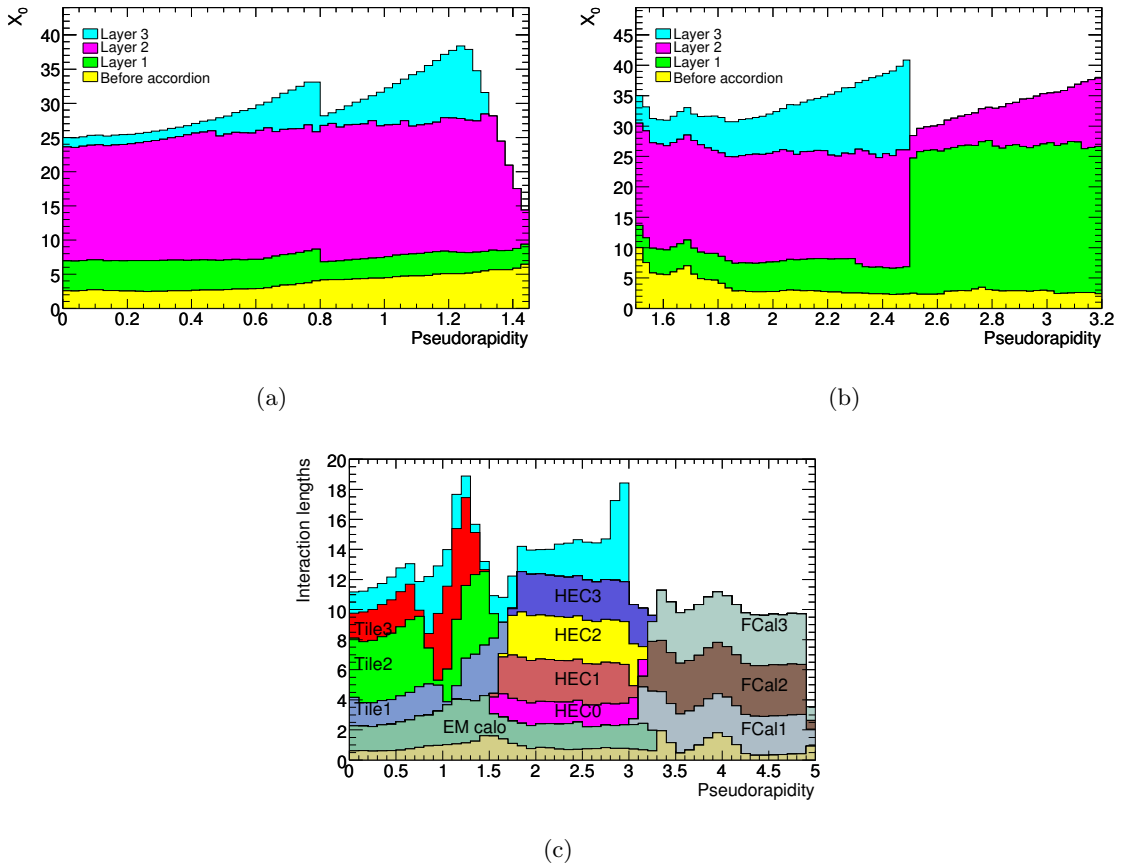


Figure 2.7: For the *barrel* (a) and *end-cap* (b) regions, the cumulative amount of material, in units of X_0, as a function of $|\eta|$ in front of the electromagnetic calorimeters (yellow), and in the three layers of the ECal (green, purple, cyan). (c): the cumulative amount of material, in units of λ, as a function of $|\eta|$, in front of the ECal (light gold), in the ECal (EM calo) and in each layer of the hadronic calorimeters (Tile$n$, HECn, FCal$n$). The total amount of material in front of the first active layer of the muon spectrometer (cyan) is also shown.

The Forward Calorimeter (FCal) measures the energy and direction of particles in the region $3.1 < |\eta| < 4.9$. It is made of LAr as the active medium interposed with an internal layer of copper and two layers of tungsten, resulting in an enhanced radiation strength and heat dissipation. In this thesis, the FCal is only used to calculate the missing transverse energy (E_{T}^{miss}) defined in Section 2.4.9.

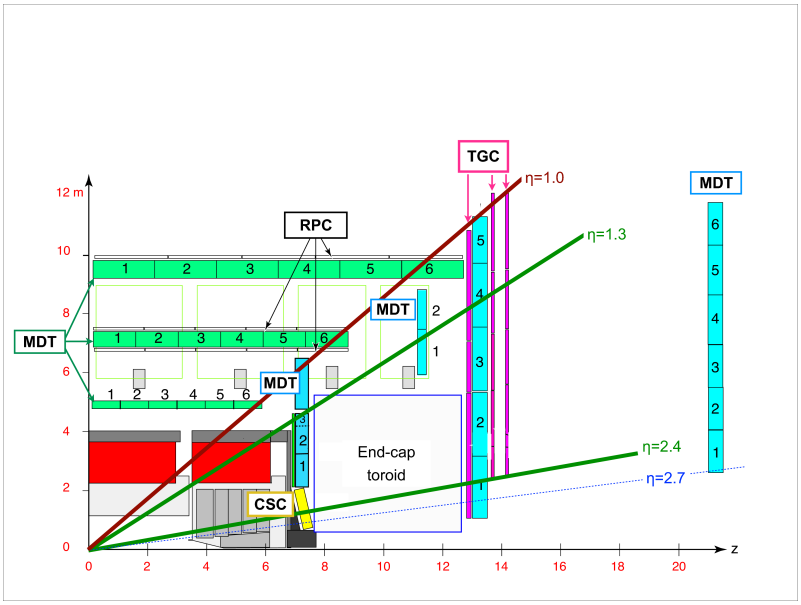


Figure 2.8: A schematic picture showing a quarter-section of the muon spectrometer. [92]

2.2.5 Muon Spectrometer

Muons produced at the LHC typically have energies of tens to occasionally hundreds of GeV. While electrons, photons, and hadrons of a similar energy are stopped by the calorimeter system, muons escape. A 10 GeV muon loses about 3 GeV of energy in the IDet and calorimeters mainly via ionization [91]. Neutrinos also escape the calorimeters, however they are not directly detectable at collider experiments. The track of a muon that passes through the calorimeters is measured by the Muon Spectrometer (MS), which is embedded within the toroidal superconducting magnets surrounding the calorimeters.

The MS consists of trigger systems and high precision tracking chambers, as shown in Figure 2.8. Precision measurements are taken by the Monitored Drift Tubes (MDTs), for those which measure up to $|\eta| < 2$. Meanwhile measurements between $2 < |\eta| < 2.7$ are collected by a combination of MTDs and Cathode Strip Chambers (CSCs). The MDTs are made up of gas-filled tubes, which collect avalanche amplified charges from ionization due to the muon transit. The CSCs are multi-wire proportional chambers with segmented cathodes. In order to cope with the higher particle rate in the *end-cap* region, the CSC technology has a lower occupancy time and a higher time resolution than the MDT technology. The hit resolution is approximately 40 μm in the orthogonal plane to the magnetic field direction, however coarser in the parallel direction. The tracks reconstructed from the MDTs and CSCs are bent by the toroidal magnetic field allowing the measurement of momentum with a resolution of $\frac{\sigma_p}{p} < 10\%$ at $p = 1$ TeV and with higher resolution at lower momentum.

The detection in the $|\eta| < 2.4$ region is completed by fast response detectors used to activate the data acquisition. Resistive Plate Chambers (RPC) are gaseous parallel

electrode plate detectors covering the region $|\eta| < 1.05$. Thin Gap Chambers (TGC) are multi-wire proportional chambers installed at $1.05 < |\eta| < 2.7$, providing triggering up to $|\eta| < 2.4$. RPCs and TGCs time resolution is 1.5 ns and 4 ns, respectively.

2.2.6 Trigger and Data Acquisition (DAQ)

The ATLAS trigger and data acquisition system (TDAQ) is designed to select and store collisions of interest for off-line analysis. During the 2012 data collection, proton bunches were colliding every 50 ns, i.e. approximately 20 million times per second. It is not possible to record the information of every event at this rate of collision with today's technology. However most of the collisions do not contain processes of interest. The ATLAS data acquisition in 2012 was able to record on average 700 events per second [93]. The trigger system has an essential role in the acquisition approval, rapidly deciding whether a given event contained an interesting signature for collection. The trigger only allows the collection of one event in every tens of thousands, all other events are permanently lost.

The acquisition decision is taken by three consecutive level of triggers: the Level 1 (L1), the Level 2 (L2), and the Event Filter (EF). Each trigger level refines the event selection of the preceding trigger, accessing further detailed information from the detector. From L1 to EF, the rate of accepted events reduces, allowing for longer latencies to process the more detailed event information. Events must pass through all of the three levels to be stored.

The L1 trigger is required to make rapid decisions, since it has to processes every event to look for specific signatures. The triggers which have been used in this thesis, look for signals which are compatible with the presence of electrons or muons with large transverse momentum. At L1, muons are reconstructed using signals from the RPCs and TGCs, while the electron energy deposits are searched in the calorimeters, using reduced granularity information. The L1 trigger is required to make a decision within 2.5 μs. Data are temporarily stored in the detector's front-end electronic buffers, until a decision is taken. The output rate from L1 is designed to reach 75 kHz. The geometric location of the electron or muon candidate is recorded as a Region of Interest (RoI), and is communicated to the L2 trigger for further selection.

The L2 trigger uses the complete detector information in the RoI to make a decision. For electron signals passed from L1, L2 has access to the full calorimeter granularity in the RoI; for L1 muon signals, the L2 analyses the hits in the MDTs and CSCs, located in the RoI. In both cases, the information from the IDet in the direction of the RoI is also used to proceed with a fast reconstruction of muons and electrons. In approximately 40 ms L2 makes a decision on which events to pass, based on the reconstructed electron and muon properties. The output rate of L2 is reduced to 6.5 kHz.

For the events which fulfil the L2 selection criteria, the full detector information is assembled and passed through to the Event Filter. Muon and electron are reconstructed by

algorithms similar to those used off-line which have greater accuracy but slower processing time. In about four seconds the Event Filter makes a final decision on which events to record. The event rate is further reduced to a final 700 Hz. The accepted events are recorded to be used for off-line analysis.

2.3 Simulation of collisions and detection

To predict the signal that specific processes would produce in the ATLAS detector, simulated events are produced with Monte Carlo generators. These generators establish the non-deterministic properties of events by randomly sampling known probability density functions. Simulations are divided in three stages: event generation; simulation of the detector and the interactions of particles within it; transformation of the activity in the sensitive regions of the detector into readout voltages and currents.

The event generators reflect the theoretical and experimental understanding of the physics involved in the production and decay of processes. The calculation of the process physics is approximated, due to the use of perturbation theory. The order of the perturbation theory calculation for a given process is chosen based on the precision required by the study. Every interaction and decay that leads from proton-proton collisions to the final state of the event is simulated. The final state includes stable particles, along with any particle that has a lifetime long enough to interact with or escape the detector before decaying.

The generation of events proceeds through consecutive steps: initial state parton generation, final state particle generation, parton radiation, parton showering and hadronization. For each step, numerous generators are developed to model the underlying physics. The choice of the generator is based on its suitability to replicate the required physics process. This is typically assessed as a result of performance studies. In some cases, the analyses presented in this thesis will use different generators to model the same process, because the targeted final state content and kinematics are different. Therefore, the chosen combination of generators is considered more appropriate for the type of analysis required. Table 2.1 lists the set of simulations used in this thesis.

The first stage of the generation consists in the determination of the momentum carried by the partons which produce the hard collision. The parton momentum is distributed according to parton density functions (PDFs), which depend on the parton type and the resolution scale $\mathcal{Q}^2$. The sets of PDFs used in this thesis are CTEQ6L1 [94,95], CT10 [96] and CT10/MRST LO [97].

The calculation of the matrix element physics of the process and the final state generation are performed with Alpgen [98], gg2ZZ [99], gg2WW [100], MC@NLO [101], Herwig [102], MG5_aMC@NLO [103], MadGraph [104], Powheg-BOX [105–111], AcerMC [112], HELAC-Oneloop [113, 114], Pythia8 [115] and Sherpa [116]. The decays of particles which travel a negligible distance from the collision point due to their

short lifetime, are also simulated with these generators. The decay of τ leptons is instead modeled with TAUOLA [117]. Electromagnetic radiation is treated using PHOTOS [118].

Due to the nature of colour interaction, partons in the initial and final state can radiate additional partons, forming showers of partons with progressively lower energy. When the energy of the resulting partons is low enough (non-perturbative scale), they bound together to form colourless hadrons. In simulations, the energy at which parton showering stops and the hadronization starts is determined by the factorisation scale. The processes of parton showering and hadronizations, which lead to the colourless observable final state, are simulated with PYTHIA6 [119], HERWIG [102], HERWIG++ [120], PYTHIA8 and SHERPA. These simulations also take into account the underlying event partons, i.e. the partons in the proton-proton, which do not contribute to the hard scattering. When the parton showering and hadronization are simulated by HERWIG, the underlying event is treated with JIMMY [121]. The underlying event simulation is also tuned using ATLAS data. The tunes which have been considered in this thesis are PERUGIA2011C [122], AUET2 [123], AUET2B [124], AU2 [125], UE-EE-4 [126] and the default tune for SHERPA.

The event generation is completed by simulating a variable number of additional soft collisions with PYTHIA8 [115] to account for the *in-time* and *out-of-time* pileup. The simulated events are scaled with event-specific correction factors in order to reproduce the profile of the average number of interactions per bunch crossing in the analysed dataset shown in Figure 2.2(b).

The next step of the simulation models the expected propagation and interaction of the final state particles with the ATLAS detector. A detailed model of the ATLAS detector has been implemented in the ATLAS detector simulation package [127], which is based on the GEANT4 libraries [128]. The emulated detector features near-real detector setup and operational conditions such as module alignments, calibration and dead sensors which have been determined in previous studies. The particles generated in the collision are propagated through the detector and interact with matter according to interaction models. In-flight decays are also accounted. GEANT4 contains several models of interaction with matter and the ATLAS simulation uses those designed for high energy physics experiments. In particular, for interacting hadrons, the Quark-Gluon String Precompound model is used at high energies and the Bertini Intranuclear Cascade model at low energies [129].

The complex detector geometry and the detailed physics description used by the ATLAS simulation are computationally demanding. Limits on the computational power available to run simulations in some cases have required the use of a faster simulation software [130]. The simulation of the detection is tens of times faster than the default one, however it does contain some approximations. The most relevant approximation is the parametrisation of the calorimeter response. This thesis mostly uses simulations with full detector description. Fast simulations are used only when explicitly indicated.

As in the real detector, the energy released due to particle interactions in the sensitive detectors is converted into readout signals. Noise, crosstalk, and channel dependent

responses are modeled in the simulations according to performance studies. The signals which feed the trigger system and the trigger decision process are also emulated.

The detected signals in simulation and real data are processed with the same algorithms to reconstruct the particles produced in the collision, as described in the following section. In simulation, the set of generated particles and interactions is also kept. Such information constitutes the generator or "truth" information of the event and is used in the analyses to assess the nature of the reconstructed particles.

The prediction of a process estimated with simulations is obtained by scaling the simulated events according to the cross-section of the process and the data luminosity. To improve the accuracy of the simulation, data-driven corrections are applied to the reconstructed particles, based on detector performance measurements.

Reference name	ME Generator	Parton Shower + underlying event	PDF	Tune
Alpgen+Pythia6	Alpgen	Pythia6	CTEQ6L1	Perugia2011C
Alpgen+Herwig	Alpgen	Herwig + Jimmy	CTEQ6L1	AUET2
gg2ZZ+Herwig	gg2ZZ	Herwig + Jimmy	CT10	AUET2
gg2WW+Herwig	gg2WW	Herwig + Jimmy	CT10	AUET2
MC@NLO	MC@NLO	Herwig + Jimmy	CT10	AUET2
Herwig	Herwig	Herwig	CTEQ6L1	AUET2
MG5+Herwig	MG5_aMC@NLO	Herwig++	CT10/MRST LO	UE-EE-4
MadGraph+Pythia6	MadGraph	Pythia6	CTEQ6L1	AUET2B
MadGraph+Pythia8	MadGraph	Pythia8	CTEQ6L1	AU2
Powheg+Pythia6	Powheg-BOX	Pythia6	CT10/CTEQ6L1	Perugia2011C
Powheg+Pythia8	Powheg-BOX	Pythia8	CT10	AU2
AcerMC+Pythia6	AcerMC	Pythia6	CTEQ6L1	AUET2B
PowHel+Pythia8	HELAC-Oneloop + Powheg-BOX	Pythia8	CT10	AU2
Pythia8	Pythia8	Pythia8	CTEQ6L1	AU2
Sherpa	Sherpa	Sherpa	CT10	Sherpa default

Table 2.1: Configuration of the simulations used in the thesis. The software used for the calculation of the matrix element interaction (ME), the parton shower and underlying event, the parton density functions (PDFs) and the tunes for the underlying event are specified.

2.4 Particle reconstruction

The raw data collected by the ATLAS detector consists of an ensemble of signals from approximately 100 million readout channels. This information is combined by several algorithms to extract the nature and four-momentum of the particles that generated these signals. The algorithms used to reconstruct and identify electrons, muons, tau leptons and jets are discussed in this chapter, together with the triggers used to collect events. Uncertainties associated with the particle reconstruction and trigger are also described. The production of neutrinos and other undetectable particles in the collisions is inferred using the missing transverse energy (E_{T}^{miss}), discussed in the last section of this chapter.

2.4.1 Electrons

Electrons within $|\eta| < 2.47$ are reconstructed with an algorithm that combines the information from the ECal and the IDet. The procedure is described in detail in [131, 132]. Fixed-size energy deposits [133] in the ECal ($\Delta\eta \times \Delta\phi = 0.075 \times 0.175$ in the *barrel* and 0.1×0.1 in the *end-cap*), compatible with the interaction of electrons with $p_{\mathrm{T}} > 2.5$ GeV, are matched to tracks in the IDet. The track associated to the cluster needs to be within a distance from the cluster of $\Delta\eta < 0.05$, and within $\Delta\phi < 0.05$ of the bending direction, or $\Delta\phi < 0.1$ away from that direction, to account for bremsstrahlung.

The energy of the reconstructed electron is the sum of the measured cluster energy, the estimated energy loss in the material in front of the ECal, and the estimated lateral and longitudinal energy leakage outside the cluster. The direction of the electron is obtained from the reconstructed track.

2.4.1.1 Electron identification

Electrons are not the only particles that can deposit energy in the ECal and form tracks in the IDet. Charged hadrons, produced in abundance in proton-proton collisions, can in some cases be reconstructed as electrons. Moreover, electrons produced from particles interacting with the detectors and in the decay of hadrons are not of interest to the analyses in this thesis and therefore must be identified and removed. These types of electrons are referred to as "non-prompt" and are rejected by imposing identification criteria on the reconstructed electrons.

Several levels of identification have been developed as result of performance studies [132]. The analyses presented here use those levels that provide higher background rejection: *tight++* and *very tight likelihood*. The former imposes constraints on key variables of the reconstructed electron, while the latter selects electrons based on the value of a likelihood function built from a similar set of variables.

The identification key variables describe the properties of the track and the energy deposit of the reconstructed electron. Electron energy deposits are narrower than those

of hadrons and have little to no leakage in the hadronic calorimeter. The distribution of the energy inside the cluster also provides discriminating information. Requirements are imposed on the minimum number of hits in the track, on the number of higher (electron-like) signal hits in the TRTs and on the ratio of the cluster energy to the track momentum. A more accurate matching between track and energy cluster is required. Electrons are also checked against having come from photon conversions.

The *tight++* and *very tight Likelihood* requirements select real electrons from the primary interaction (prompt) with similar efficiencies that vary between 70 and 80% in η [132]. The background is suppressed by several orders of magnitude with a better rejection power better by a factor of two for the *very tight likelihood* identification. Since this identification was only developed recently, it is used in just one of the analyses presented in this thesis. The looser identification criteria *medium++* and *loose++* provide higher efficiency and lower background rejection. They are used in this thesis for background estimates and validation.

2.4.1.2 Electron baseline selection

The analyses presented in this thesis utilize optimized selection criteria for reconstructed electrons. There is, however, a minimum set of requirements common to all the analyses. Any reconstructed electron in this thesis should satisfy:

- $p_{\mathrm{T}} > 10$ GeV;
- position of the energy cluster within the acceptance of the IDet tracker excluding the transition region, that is, $|\eta_{cl}| < 1.37$ or $1.52 < |\eta_{cl}| < 2.47$;
- *loose++* identification.

2.4.2 Muons

Muons are reconstructed using information from the muon spectrometer (MS), the IDet and the calorimeters. The method of reconstruction is called *combined Chain 1* reconstruction and is described in [92]. Tracks in the muon spectrometer are reconstructed starting from semi-aligned hits that form a segment in single modules of the MDT or CSC. An MS track is constructed as a combination of two or more segments in different MS modules using a least-square fitting method and taking into account the magnetic field displacement. The reconstruction algorithm then matches an MS track to an IDet track , after the former is corrected for the energy lost in the calorimeters and dead material. A statistical combination of the parameters of the two tracks is performed, and used to reconstruct the muon momentum at the point of closest approach to the hard interaction vertex. The IDet track measurement dominates the momentum reconstruction up to $p_{\mathrm{T}} \sim 80$ GeV in the *barrel* and $p_{\mathrm{T}} \sim 20$ GeV in the *end-caps*; the MS track measurement dominates for $p_{\mathrm{T}} > 100$ GeV.

To reduce accidental matches of IDet and MS tracks from different particles, and reduce muons produced far from the beam interaction point, constraints are applied to the type and number of hits comprising the IDet track. The muon reconstruction efficiency is about 97% according to studies performed with a high purity $Z \to \mu^+\mu^-$ dataset [92].

2.4.2.1 Muon baseline selection

The minimum requirements for muons selected in this thesis are:

- $p_{\mathrm{T}} > 10$ GeV;
- pseudorapidity within the acceptance of the IDet tracker, $|\eta| < 2.5$;
- IDet optimized track quality requirements from dedicated studies [92]:
  - a hit in the first layer of PIXEL, unless the pixel crossed is inactive;
  - at least five hits in the SCT accounting for non-active strips crossed;
  - at most two missed hits in PIXEL and SCT;
  - at least nine TRT hits if the track is in the TRT acceptance region ($0.1 < |\eta| < 1.9$).

Further criteria are applied depending on the analysis strategy.

2.4.3 Lepton impact parameter and isolation

In order to reject non-prompt electrons or muons, requirements are set on the minimum distance of the track of the reconstructed lepton from the hard interaction vertex, and on the activity around the lepton direction.

The events studied in this thesis have at least one interaction vertex reconstructed from tracks. The hard collision is identified by the vertex with the highest value of $\sum p_{\mathrm{T}}^2\,(\mathrm{track})$, referred to as the primary vertex, where $p_{\mathrm{T}}\,(\mathrm{track})$ is the transverse momentum of the tracks associated to the vertex. The impact parameters of a track, d_0 and z_0, are the distances of closest approach of the track to the primary vertex projected in the transverse plane, and along the beam line, respectively. The lepton track itself is removed from the primary vertex fit to avoid introducing a bias. Requirements on the impact parameters help to reduce leptons produced in the decay of hadrons and in interactions with the detector material.

The production of many particles collinear with the lepton is generally a hint of a reconstructed lepton originating in a jet of hadrons. IDet tracks and calorimeter energy deposits are studied to set isolation requirements. The track-based isolation parameter is called $ptconeXX$ and is defined as the sum of the p_{T} of reconstructed tracks with angular distance ΔR from the lepton track smaller than $0.XX$. The distance limit XX is commonly 20, 30 or 40, depending on the analysis. The tracks considered in the electron (muon)

track isolation have $p_{\mathrm{T}} > 0.4$ (1) GeV, are associated to the primary vertex and satisfy requirements on hits in the IDet. The lepton track is excluded from the calculation.

The activity in the calorimeter is quantified with the *etconeXX* isolation. *etconeXX* is the sum of the energies deposited in the calorimeter cells projected in the transverse plane for cells with angular distance $\Delta R < 0.XX$ from the lepton. The *etconeXX* is corrected for pileup effects and the energy due to the lepton.

2.4.4 Lepton Triggers

The acquisition of the data analysed in this thesis is triggered by electrons or muons. The analyses presented in this thesis have triggers that target a single high p_{T} lepton or a pair of high p_{T} leptons, as specified later.

Events with at least one muon can be collected using the single muon triggers (collectively known as SMTrig). The candidate trigger muon at L1 must have at least 15 GeV of transverse momentum. At Event Filter level (EF), the trigger muon needs to satisfy either $p_{\mathrm{T}} > 36$ GeV, or $p_{\mathrm{T}} > 24$ GeV and isolation $ptcone20/p_{\mathrm{T}}(\mu) < 0.12$. The SMTrig efficiency reaches its maximum for muons with $p_{\mathrm{T}} > 25$ GeV: 70% for a single muon with $|\eta| < 1.05$ and 85% for one with $1.05 < |\eta| < 2.5$ [93].

The di-muon trigger (DMTrig) is activated by events with at least two muons. At L1, it searches for a muon with $p_{\mathrm{T}} > 15$ GeV. The event is then recorded if the trigger muon has $p_{\mathrm{T}} > 18$ GeV at EF and if an additional muon with $p_{\mathrm{T}} > 8$ GeV is found in the event. The maximum triggering efficiency is similar to that of SMTrig, however it is reached at $p_{\mathrm{T}} > 20$ GeV for the leading p_{T} muon and $p_{\mathrm{T}} > 8$ GeV for the other muon.

Single electron triggers (SETrig) are used to selected events with at least one high p_{T} electron. The trigger electron at L1 can satisfy $p_{\mathrm{T}} > 30$ GeV, or $p_{\mathrm{T}} > 18$ GeV, with less than 1 GeV of energy deposited in the hadronic calorimeters. Depending on which of the two conditions is satisfied, the trigger electron is required at EF to satisfy $p_{\mathrm{T}} > 60$ GeV, or $p_{\mathrm{T}} > 24$ GeV and $ptcone20/p_{\mathrm{T}}(e) < 0.1$, respectively. Identification criteria similar to that imposed by the *medium++* requirement on the reconstructed electrons are applied. The SETrig efficiency is above 70% for single electrons with $p_{\mathrm{T}} > 25$ GeV and increases with p_{T} up to 98% [134].

Events with multiple electrons can be selected with the di-electron trigger (DETrig). At L1, two trigger electrons with $p_{\mathrm{T}} > 10$ GeV and hadronic energy deposit smaller than 1 GeV are required. The two electrons then must satisfy $p_{\mathrm{T}} > 12$ GeV at EF, as well as the identification criteria similar to the *loose++* condition. The high efficiency plateau of about 95% is reached for $p_{\mathrm{T}} > 15$ GeV for both electrons [135].

To collect events with one muon and one electron, an electron-muon trigger (EMTrig) can be used. The trigger searches at L1 for an electron with $p_{\mathrm{T}} > 10$ GeV and low hadronic energy deposit (1 GeV at most). If found, the EF electron is required to have greater than 12 GeV and to pass an identification criterion similar to the *medium++* requirement for

the electron. In addition, at EF, a muon with $p_{\mathrm{T}} > 8$ GeV must be reconstructed. The trigger turn-on point is reached for an electron with $p_{\mathrm{T}} > 15$ GeV and a muon with p_{T} greater than 8 GeV. The maximum efficiency is approximately 95% [93,135]. The EMTrig is used in combination with the high p_{T} no isolation single electron trigger ($p_{\mathrm{T}} > 30$ GeV at L1 and $p_{\mathrm{T}} > 60$ GeV plus *medium++* identification at EF) to keep the efficiency high for very energetic electrons, explained further in Section 5.3.2.

2.4.5 Lepton corrections and uncertainties

The lepton reconstruction and triggering performances have been tested using $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ data and simulations [92, 93, 131, 132, 134, 135].

Efficiencies (ϵ) of reconstruction, identification and trigger selection are measured with a *tag-and-probe* method. In each $Z \to \ell^+\ell^-$ event, one lepton, the *tag* lepton, is reconstructed with the tightest selection criteria, while the other lepton, the *probe* lepton, is selected by requiring basic acceptance conditions (including that the invariant mass of the *tag-and-probe* leptons is compatible with the Z mass). The *probe* lepton is then tested against a specific condition to obtain the efficiency of the lepton passing that condition, as a function of the lepton kinematic variables and pileup conditions.

Efficiencies for lepton reconstruction, identification and trigger selection are found to be similar in data and simulation, indicating good modeling of the interaction of leptons with the detector in simulation. Differences are of the order of a few percent, and vary with the lepton p_{T} and η. The simulations are scaled with correction factors to bridge the gap in efficiency. The uncertainty on each correction factor affects the predictions based on simulation, and is taken into account in the analyses. The efficiency of the isolation and impact parameter requirements depends on their definitions, so it is analysis-specific. The validation of the isolation requirements, and potential corrections and uncertainties, are therefore investigated in the analyses.

The calibration and resolution of the reconstructed lepton momentum in data and simulation are also compared. Corrections are applied to the reconstructed momentum of the simulated leptons to reproduce the resolution and the calibration measured in data. The associated uncertainties are also considered. Uncertainties on the direction of reconstructed leptons are found to be negligible.

The types of uncertainties associated to reconstructed leptons are summarized in Table 2.2. Corrections and uncertainties have been calculated for simulation with full detector modeling, as well as for fast simulation, and applied to the respective simulation type in the analyses.

2.4.6 Jets

The detectable result of the parton showering and hadronization of an energetic final-state parton produced in a proton-proton collision is a spray of close-by hadrons known as a jet.

Type of uncertainty	Origin
Electrons	
Efficiency	Reconstruction
	Identification
	Isolation
	Trigger
Momentum	Calibration
	Resolution
Muons	
Efficiency	Reconstruction
	Isolation
	Trigger
Momentum	Calibration
	IDet Resolution
	MS Resolution

Table 2.2: List of experimental uncertainties affecting simulated predictions associated with muon, electron and trigger selection.

Jets can also contain photons and leptons due to hadron decays.

A jet is expected to produce several clusters of energy in the calorimeters and, if it contains charged particles, tracks in the IDet. The jet reconstruction is based on calorimeter signals and proceeds through three steps: reconstruction of the energy clusters, energy calibration of the clusters according to local properties, and combination of calibrated clusters into jets.

The reconstruction begins by merging together the signals of calorimeter cells into three-dimensional topological clusters [133, 136]. Clusters are constructed starting from cells with high signal-to-noise ratio; Neighbouring cells are aggregated based on this ratio to form a cluster of variable size and shape. This type of reconstruction reduces the impact of electronic noise and pileup.

The topological clusters are calibrated according to the local hadron calibration [137, 138]. This calibration is based on local properties of the reconstructed clusters, such as shape and energy density. Clusters are classified as electromagnetic (typical of electron and photon interactions) or hadronic. The calibration applied in each case is different since hadrons can lose their energy in non-detectable processes, like nuclear excitation, nuclear breakup or neutron capture, and escaping particles. The calibration is completed by compensating for the energy deposited outside the calorimeters, estimated with simulation.

Jet reconstruction algorithms aim to group the calibrated topological clusters generated from the same parton shower fragmentation. In this thesis, jets are reconstructed combining topological clusters using the *anti-kt* algorithm with distance parameter $R =$

0.4 [139, 140]. The topological clusters are assumed to be massless and have angular position given by the direction connecting the interaction point to the cluster centre.

The *anti-kt* algorithm uses the following definition of distance between two objects:

$$d_{ij} = \begin{cases} min\left(p_{\mathrm{T}i}^{-2}, p_{\mathrm{T}j}^{-2}\right) \Delta R_{ij}^2 / R^2 & \text{if } i \neq j \\ p_{\mathrm{T}i}^{-2} & \text{if } i = j \end{cases} \tag{2.2}$$

where $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, and y_i (y_j) and $p_{\mathrm{T}i}$ ($p_{\mathrm{T}j}$) are the rapidity[4] and the transverse momentum of the i-th (j-th) object, respectively. The algorithm searches for the combination of indices i, j which minimise d_{ij}. If $i = j$, the i-th object becomes a jet and is no longer considered by the algorithm. If $i \neq j$, the i-th and j-th objects are combined, and substituted with an object that has the sum of their four-momenta. The algorithm is recursively applied until all remaining objects are separated by at least distance R.

The reconstruction efficiency of jets from topological clusters is greater than 99% [141].

2.4.6.1 Jet baseline selection

The common requirements for jets selected in this thesis are:

- transverse momentum $p_{\mathrm{T}} > 25$ GeV;
- pseudorapidity $|\eta| < 2.8$;
- Jet Vertex Fraction (JVF) > 0.5 unless the jet has no associated tracks or $p_{\mathrm{T}} > 50$ GeV or $|\eta| > 2.4$.

The Jet Vertex Fraction (JVF) is a variable calculated from the tracks associated with the reconstructed jet. It is the sum of the p_{T} of tracks from the primary vertex divided by the sum of the p_{T} of all tracks matched to the jet. The JVF requirement suppresses jets whose energy is highly affected by pileup.

Sometimes jets are reconstructed as a result of noise bursts, beam background and cosmic ray showers. Jets of these types are distinguished from jets originated in proton bunch collisions by looking at the jet properties such as the quality of the energy reconstruction at the calorimeter cell level, jet energy deposits in the direction of the shower development and reconstructed tracks matched to the jets. Events containing jets of non-collisional origin are rejected.

[4] In this case, rapidity is preferred to pseudorapidity in order to keep the definition of distance Lorentz invariant under boosts along the beam axis. The combined objects obtained from the algorithm can indeed have non-negligible mass, hence their pseudorapidities cannot be considered a good approximation of their rapidities any more.

2.4.6.2 Jet corrections and uncertainties

The four-momentum of a reconstructed jet is further calibrated with event-by-event and jet-by-jet corrections. Each jet is corrected to point back to the interaction vertex, subtract the pileup energy contribution, calibrate the energy according to data and simulation studies, and adapt the response to the type and flavour of the parton initiating the jet. A dedicated calibration is applied to simulations with fast detector modeling. The calibration methods are explained in detail in [142].

The uncertainties associated with jet reconstruction have been investigated in [142] and are summarised in Table 2.3. Uncertainties on jet calibration, jet resolution and JVF selection efficiency are the most relevant for the analyses presented.

Type of uncertainty	Origin
	Jets
Efficiency	Jet Vertex Fraction
Four-momentum	Calibration (JES) - *In-situ* Method: statistical(0-3), modeling(0-4), detector(0-3), mixed(0-2) - η inter-calibration: statistical, modeling - High p_{T} jets - Pileup: NPV,$\langle\mu\rangle$, p_{T}, ρ - Non-Closure - Flavour: response, composition, b-jet
	Resolution (JER)

Table 2.3: Summary of experimental uncertainties affecting the jet reconstruction in simulation. The uncertainty on the jet energy calibration is divided into six components: calibration method with data, η inter-calibration, high p_{T} jets pileup, non-closure and flavour response. The intervals (0-n) indicate that there are $n+1$ independent uncertainties associated to that source.

2.4.7 b-jets

The properties of a jet can depend on the type and the flavour of the parton from which it originated. When the parent parton is a b-quark, the produced heavy flavour hadron has a long enough lifetime that its decay point can be distinguished from its production point. Furthermore, due to the mass of b-quark hadrons, there is a significant angular distance between the direction of its decay products in the detector frame. This decay process would be recorded in the IDet as tracks associated to the jet with non-null impact

parameters from the primary vertex, or as forming a secondary vertex.

The identification of jets originating from b-quarks is performed by the MV1 b-tagging algorithm [143]. The algorithm establishes if the properties of a reconstructed jet are compatible with a b-quark origin using a neural network that combines the results of the IP3D, SV1 and JetFitter algorithms [144]. These algorithms consider the tracks associated to the reconstructed jet and search for large impact parameter tracks (IP3D), secondary vertices (SV1), and topologies compatible with the weak decay of heavy flavour hadrons (JetFitter).

The performance of the b-tagging algorithm is studied in data and simulation [143, 145]. The identification criterion is tuned to select jets produced from b-quarks with 70% efficiency. The average selection efficiencies for jets from c-quarks or light quarks/gluons are 20% and less than 1%, respectively [143]. Jets from c-quarks have a higher probability of being selected since they also contain heavy hadrons with a detectable lifetime. The reconstructed jets tagged as originating by b-quarks will here on be referred to as b-jets.

The b-tagging efficiencies in simulation and data are found to be compatible within few percent [145]. Simulations are scaled with correction factors depending on the p_{T} and η of the b-jet, the distance between the primary and secondary vertices and the activity surrounding the b-jet. Uncertainties associated to the b-tagging efficiency are accounted and listed in Table 2.4.

The identification of b-jets is important in the selection of processes with b-quarks in the final state, such as $t\bar{t}H$ production, where b-quarks are produced in the decay of the top quarks. Requiring b-jet in the event suppresses backgrounds containing light flavour jets only.

Type of uncertainty	Origin
	b-tagging
Efficiency	MV1 b-tag for b-quark originated jets, η dependent (0-5)
	MV1 b-tag for c-quark originated jets, η dependent (0-3)
	MV1 b-tag for τ_{had} originated jets, η dependent (0-3)
	MV1 b-tag for light quark and gluon originated jets, p_{T} and η dependent (0-11)

Table 2.4: Summary of the experimental uncertainties associated to the MV1 b-tagging algorithm. The intervals (0-n) indicate that there are $n+1$ independent uncertainties associated to that source.

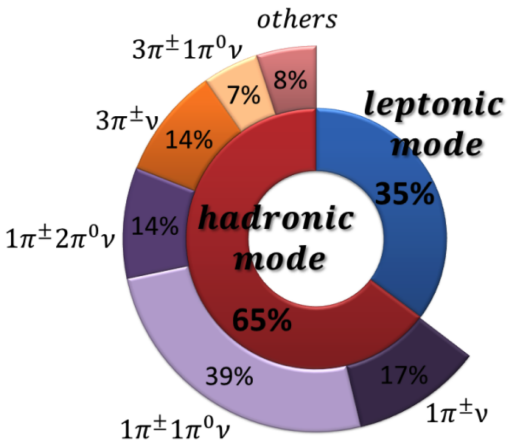


Figure 2.9: The τ lepton decay modes.

2.4.8 Tau leptons

The τ lepton is the heaviest known charged lepton. It has a mass of 1776.82±0.16 MeV and a decay length of 87.03 μm [146]. The τ lepton decay modes and their decay probabilities, or branching ratios (BR), are shown in Figure 2.9.

The branching ratio of the τ lepton decay into two neutrinos and a lighter charged lepton (the leptonic decay mode) is 35%. The τ leptons decaying this way are detected as isolated electrons or muons. A lepton from a τ decay would differ from a prompt lepton by the small but non-zero impact parameter of the track associated to the former. This is due to the τ decay length, which is not negligible but still about five times smaller than that of heavy flavour hadrons.

The remaining 65% of τ decays are into a neutrino and hadrons (the hadronic decay mode). The visible particles from hadronic decay are detected as a jet of particles with distinctive properties, referred to in the following as τ_{had}. The τ_{had} is generally composed of one or three charged pions plus a few neutral pions. The charged pions produce tracks in the IDet and hadronic showers in the calorimeters; the neutral pions decay into photon pairs that produce electromagnetic showers with no associated tracks. The particles produced in the hadronic decay of a τ are usually more collimated than those produced in a parton-initiated jet. The τ_{had} has a specific hadron content and kinematic behaviour which allows them to be distinguished from parton-initiated jets and calibrated by them taking advantage of their known composition [1].

The τ_{had} reconstruction begins with a jet obtained using the algorithm described in Section 2.4.6. The jet considered must satisfy $p_{\text{T}} > 10$ GeV and $|\eta| < 2.5$. In addition, the jet must have one track (1-track) or three tracks (3-tracks) associated to it with $p_{\text{T}} > 1$ GeV, $\Delta R < 0.2$ from the jet direction, and constrained impact parameters relative to the primary vertex, to exclude pileup tracks [1]. The reconstructed τ_{had} is obtained by combining only the topological clusters within $\Delta R < 0.2$ or the underlying jet direction.

The direction of each topological cluster is calculated with respect to the primary vertex. The momentum and direction of a reconstructed τ_{had} are obtained from the vector sum of the topological cluster momenta, and its mass is set to zero.

The τ_{had} energy, at this stage, is not optimised for the specific mix of hadrons produced in the τ decay, and is not corrected for underlying event and pileup contributions. Thus, an additional calibration is applied to bring the reconstructed energy to the level of the true visible energy of the τ decay products. The τ_{had} energy resolution improves.

The calibration is first derived from simulation and then additional small corrections, derived through a data-driven method, are applied. The simulation-based calibration compares the initial reconstructed energy, obtained with the local hadron calibration, to the energy of the visible τ decay products in simulated $Z \to \tau^+\tau^-$ and $W^\pm \to \tau^\pm\nu$ events. $Z' \to \tau^+\tau^-$ simulated events with Z' masses between 250 GeV and 1250 GeV are also used to obtain the calibration for high energy τ_{had}. This calibration, known as the tau energy scale (TES), brings the reconstructed τ_{had} energy into agreement with the true visible τ_{had} energy scale at the level of a few percent and removes any significant dependencies of the energy scale on pseudorapidity, energy, track multiplicity and pileup [1]. The simulation-based calibration is applied to reconstructed τ_{had} both in simulation and in data.

The data-driven calibration procedure corrects the residual difference observed in the energy scale of reconstructed τ_{had} in data and in simulation. The procedure is described in detail in Chapter 3 and the correction is applied to data, leading to an improved modeling of τ_{had} properties in data.

2.4.8.1 τ_{had} identification

Parton-initiated jets, electrons and muons can also be reconstructed as τ_{had}. Identification criteria are therefore studied to reduce the chance of this occurring.

Since jets are copiously produced in proton-proton collisions, the effectiveness of the identification criteria that reject parton initiated jets (QCD jets) reconstructed as τ_{had} is critical in analyses with τ leptons. The identification is provided by Boosted Decision Trees (BDT) algorithms [147, 148] and is based on the fact that for hadronically decaying τ leptons:

- the associated energy clusters are more collimated than those from QCD jets;
- the associated tracks are more collimated than those from QCD jets;
- the fraction of total energy carried by the highest momentum track is generally large, while for QCD jets the energy is more evenly distributed among particles;
- tracks can have a significant impact parameter relative to the primary vertex or form a secondary vertex, while tracks from light QCD jets originate in the primary vertex.
- the number of neutral pions is typically smaller than in QCD jets;

- the invariant mass of the track system, assuming that all the particles are pions, is smaller than the τ lepton mass.

Separate BDT algorithms are trained for 1-track and 3-tracks τ_{had}. The analyses presented here uses the Jet BDT *medium* criterion which has efficiencies for hadronically-decaying τ leptons of 60% for 1-track real τ_{had} and 40% for 3-tracks ones [1]. The rejection power for a parton-initiated jet is more than 20 when reconstructed as a 1-track τ_{had} and of the order of hundreds when reconstructed as a 3-tracks τ_{had}.

A smaller but non-negligible contamination of τ_{had} reconstructed from electrons and muons is suppressed with dedicated requirements. Electrons can reproduce the signature of a 1-track τ_{had} and are discriminated against by using another BDT algorithm that takes into account:

- the transition radiation produced in the TRT;
- the fraction of energy deposited in the Ecal compared to the total energy deposited in the calorimeters;
- the longitudinal and transverse shower shapes;
- the angular distance between the τ_{had} associated tracks and its direction.

The Electron BDT *medium* criterion is used in this thesis. The efficiency for hadronically-decaying τ leptons is 85% and the rejection power for electrons is in the hundreds [1].

Muons can be reconstructed as a τ_{had} if they deposit a large amount of energy in the calorimeter or if they are wrongly associated to unrelated energy clusters. The former case generally occurs with energetic muons causing the τ_{had} to have a small energy fraction in the ECal and a large ratio of track p_{T} to calorimeter energy. The latter case is characteristic of soft muons and the resulting τ_{had} would have a large energy fraction in the ECal and a small ratio of track p_{T} to calorimeter energy. Based on this information, a muon veto requirement is developed to reduce the muon contamination to a negligible level. The efficiency of the veto is 96% for hadronically decaying τ leptons and about 60% for muons.

2.4.8.2 τ_{had} baseline selection

The τ_{had} considered in this thesis have:

- $p_{\text{T}} > 20$ GeV,
- pseudorapidity within the IDet acceptance $|\eta| < 2.5$,
- one or three associated tracks,
- electric charge of ± 1,
- Jet BDT *medium* identification,

- Electron BDT *medium* identification,
- muon veto.

When a τ_{had} is selected, the jet from which the τ_{had} reconstruction began is removed from the event for consistency.

2.4.8.3 τ_{had} corrections and uncertainties

The τ_{had} calibration procedure is affected by uncertainties due to the calorimeter response to the hadrons composing the τ_{had}, the modeling of pileup and underlying event energy contributions, the detector description, the uncertainties of the data-driven measurement, and the remaining difference between the calibrated reconstructed energy and the true visible energy [1]. Additional uncertainties are considered for the resolution of the τ_{had} energy and for the efficiency of the identification requirements [1]. The sources of uncertainty associated to the reconstruction of τ_{had} are summarised in Table 2.5.

Type of uncertainty	Origin
τ_{had}	
Efficiency	Jets BDT Electron BDT
Energy	Calibration (TES) Resolution (TER)

Table 2.5: Summary of experimental uncertainties related to simulated τ_{had}.

2.4.9 Missing Transverse Energy (E_{T}^{miss})

Neutrinos and other long-living weakly interacting particles are not detected by the ATLAS detector. The transverse momentum of these invisible particles, however, can be recovered. Indeed, events producing high-p_{T} invisible particles are characterised by a large amount of missing transverse energy (E_{T}^{miss}).

The E_{T}^{miss} is defined as the negative vectorial sum of the transverse momenta of all detected particles. Since the total transverse momentum is conserved and is null in the initial state of the proton-proton collisions, the net p_{T} of the detectable particles has to balance the net p_{T} of the invisible particles in each event. Therefore, neglecting experimental uncertainties, the E_{T}^{miss} has direction and magnitude equal to the transverse momentum of the system formed by the invisible particles.

Fake E_{T}^{miss} can be also created by the finite resolution of the p_{T} of reconstructed particles and by particles outside the detector acceptance. For this reason, the detector coverage of the solid angle must be as close as possible to 4π.

The E_{T}^{miss} reconstruction uses energy deposits in the calorimeters, including the FCal, and reconstructed muons in the muon spectrometer [149]. The calorimeter signals are calibrated according to whether the underlying particle is recognised as an electron, photon, τ_{had}, jet or muon. Deposits not associated with any reconstructed particle, mostly due to soft emissions, are also taken into account.

Procedures to remove pileup energy contributions are applied to reconstructed jets and isolated energy clusters by means of the tracking information [149]. Tracks matched to energy deposits are associated to one of the interaction vertices of the event. Corrections are applied based on the fraction of the transverse momentum carried by the tracks associated to the primary vertex relative to the transverse momentum of all the tracks.

Chapter 3

Calibration of the Energy of Hadronically Decaying Tau Leptons with $Z \to \tau^+\tau^-$ events

The reconstruction of the visible decay products of the hadronically decaying tau leptons (τ_{had}) with the ATLAS experiment has been described in details in Section 2.4.8. The τ_{had} energy calibration is derived from simulation by scaling the energy of the reconstructed τ_{had} to the energy of the true τ_{had}. This procedure ensures the correct calibration of the energy of simulated τ_{had}, however this may not be true for τ_{had} in data. For instance, properties of calorimeter clusters and pileup conditions in data could not be properly modeled in simulation. Consequently, the application of the simulation-based τ_{had} energy calibration (TES) to τ_{had} in data may not be accurate. To correct for such potential differences, $Z \to \tau^+\tau^-$ events have been used to compare the energies of τ_{had} in data and in simulation. The analysis distinguishes between τ_{had} with one associated charged particle (1-track) and three associated charged particles (3-tracks), because the simulation-based calibration is different for the two types of τ_{had}.

The $Z \to \tau^+\tau^-$ analysis targets a specific decay mode where one tau lepton decays into a muon and two neutrinos (leptonically) and the other one decays into a neutrino and hadrons (hadronically). This final state allows to select a high statistics and high purity sample of events with true τ_{had} with known kinematics. A sample of $Z \to \tau^+\tau^-$ events where an electron is produced in the decay of the tau lepton in place of a muon has also been studied. Although the production rate and the kinematic of the two processes are identical, the study of final states with muons are more accurate due to experimental reasons. The probability of an electron to be reconstructed as a τ_{had} is larger than that of a muon and this reduces the purity of the sample. In addition, the uncertainty on the electron energy reconstruction is higher than that on the muon energy at the energies expected for $Z \to \tau^+\tau^-$ events, and the effect on the τ_{had} energy measurement is not

negligible. The final state with an electron is therefore not used and only events with exactly one muon and one τ_{had} consistent with a $Z \to \tau^+\tau^-$ topology are selected.

The visible mass (M_{vis}) is used to compare the τ_{had} energy calibration (TES) in data and simulation. The M_{vis} is the invariant mass of the μ-τ_{had} system, or in other words, the invariant mass of the reconstructed decay products of the $Z \to \tau^+\tau^-$ decay:

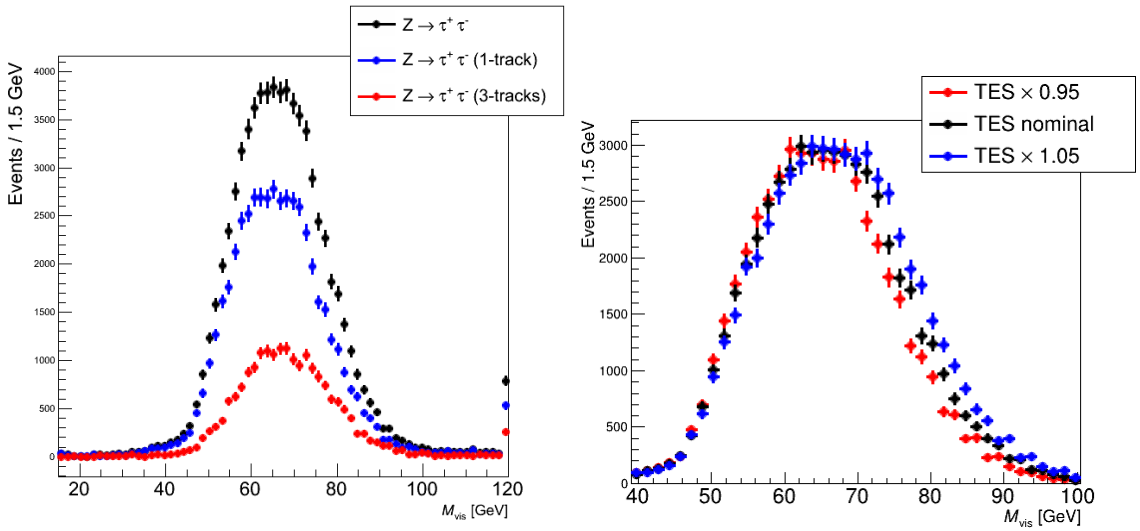
$$M_{vis} = \sqrt{\left(p^{\nu}_{\tau_{\text{had}}} + p^{\nu}_{\mu}\right)^2} \simeq \sqrt{2E^{\tau_{\text{had}}}E^{\mu}\left(1-\cos\theta\right)} = \\ = \sqrt{2p_{\text{T}}^{\tau_{\text{had}}}p_{\text{T}}^{\mu}\left(\cosh\left(\eta^{\tau_{\text{had}}}-\eta^{\mu}\right)-\cos\left(\phi^{\tau_{\text{had}}}-\phi^{\mu}\right)\right)} \quad (3.1)$$

where E^{particle}, $p_{\text{T}}^{\text{particle}}$, η^{particle} and ϕ^{particle} are respectively the energy, the transverse momentum, the pseudorapidity and the azimuthal angle of the *particle* specified; θ is the angle between the direction of the two particles. The approximation in Equation 3.1 is valid in the limit of relativistic particles which is a good assumption in this measurement[1] The M_{vis} is therefore proportional to the square root of the τ_{had} energy (and p_{T}).

Figure 3.1(a) shows the expected distributions of the visible mass for $Z \to \tau^+\tau^-$ events in the $\mu\tau_{\text{had}}$ final state, including also the distributions for 1-track and 3-tracks τ_{had}. The narrow peak-like distribution of the $Z \to \tau^+\tau^-$ events is very sensitive to the TES. Figure 3.1(b) compares the expected $Z \to \tau^+\tau^-$ M_{vis} distribution with the distributions obtained by scaling of 5% and -5% the energies of the reconstructed τ_{had} in the sample. If the TES is increased, the M_{vis} distribution shifts towards higher mass values and spreads, while if it is decreased, the M_{vis} distribution shifts towards lower mass values and shrinks. For small variations of TES the M_{vis} distribution drifts linearly, as it will be shown in Section 3.4. Therefore, an estimator of the M_{vis} peak "position" in data and simulation can be used to determine any difference in TES in data and simulation and to correct for it. In principle, the τ_{had} energy and p_{T} distributions could also be used to measure a difference in TES in data and simulation. The expected accuracy is worse though. The main reasons are that the energy and p_{T} distributions are more sensitive to the τ_{had} resolution uncertainty and to backgrounds (see Figure A.1). The background is distributed similarly to the $Z \to \tau^+\tau^-$ events making its subtraction less effective. In the M_{vis} distribution instead, the $Z \to \tau^+\tau^-$ have the distinctive peak-like shape and the uncertainty due to the background determination is relatively small.

This chapter describes the measurement of the TES from the M_{vis} distribution in data and simulation. The data sample and the simulations used in the measurement are illustrated in Section 3.1. Section 3.2 explains the event selection used to increase the $Z \to \tau^+\tau^-$ purity and Section 3.3 describes the measurement of the background events. The method to extract the TES is described in details in Section 3.4, and the uncertainties on the measurement and the results are reported in Section 3.5 and Section 3.6, respectively.

[1] The reconstructed τ_{had} is by construction massless, while the muons considered in this study have a $\beta\gamma$ of hundreds.



(a) (b)

Figure 3.1: (a): M_{vis} distributions in simulated $Z \to \mu\tau_{\text{had}}$ events. The distributions for 1-track (blue) and 3-tracks (red) τ_{had} are also shown. (b): M_{vis} distributions in $Z \to \mu\tau_{\text{had}}$ simulated events obtained with the nominal TES and with TES scaled by 5% and −5% for all the reconstructed τ_{had}. These distributions are obtained after the event selection detailed in Section 3.2.

3.1 Data sample and simulation

The proton-proton collisions collected in 2012 by ATLAS are processed to select the $Z \to \tau^+\tau^-$ sample. The integrated luminosity recorded corresponds to 20.3 fb^{-1} and about 23 millions of $Z \to \tau^+\tau^-$ events are expected to be produced. Since the final state considered contains a muon and a τ_{had}, the single muon triggers, described in Section 2.4.4, are used to select the events for this analysis. τ_{had} triggers are not used because they generally have lower efficiencies then muon triggers and require the τ_{had} to have a large p_{T}.

Simulated data are used to model the $Z \to \tau^+\tau^-$ production and other Standard Model processes, that may be selected as background in this analysis. The simulation of collisional events is a multi-step process that uses dedicated programs to model specific aspects of the interaction. In addition to the process generation, the simulation takes into account the interaction of particles with the ATLAS detector and the resulting readout signals generated. A detailed description of the simulation procedure has been given in Section 2.3.

The processes that are estimated with simulation in this analysis are: single boson production processes (Z, $W^{\pm}$), Drell-Yan ($\gamma^* \to \ell^+\ell^-$), diboson production ($W^+W^-$, $W^{\pm}Z$, ZZ) and production of top quarks ($t\bar{t}$, $tW^{\pm}$ and *single-top* in the $s-t$ channels). The target of the analysis are $Z \to \tau^+\tau^-$ events but $\gamma^* \to \tau^+\tau^-$ events are also considered as part of the signal. The contribution from Higgs boson production has been also considered, but was found negligible for this analysis.

The simulated Standard Model processes are summarized in Table 3.1. For each process

the generator and the cross-section used to scale the simulated yields are given. The complete set of simulations used in the generation of these processes are specified in Table 2.1. For some processes, event filters are applied to select specific decay modes, to require a certain number of radiated partons or to apply kinematic constraints. The cross-sections provided take into account the efficiency of these event filters. The simulated cross-sections are also corrected with K-factors to account for higher order corrections in the perturbation theory. The K-factors are also given in Table 3.1.

All the simulations are corrected for the pileup conditions as observed in data, and have data-driven corrections for particle detection efficiencies, energy calibration and energy resolution, as specified in Section 2.4.

Table 3.1: Simulated processes used for signal and background modeling. For each process, the generator, the cross-section times the generator filter efficiency and the K-factor used to scale the simulated cross-section value to the latest theoretical calculation are given. The full simulation setup is specified in Table 2.1. $X+np$ with n integer and X a generic process indicates the number of partons (quarks or gluons) produced at matrix element level in addition to the process X.

Process	Generator	$\sigma \times \varepsilon_{filter}$ [pb]	K-factor
Signal processes			
$Z/\gamma^* \to \tau^+\tau^-$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	3477.1	1.19
$Z/\gamma^* \to \tau^+\tau^- + 1p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	108.74	1.19
$Z/\gamma^* \to \tau^+\tau^- + 2p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	52.73	1.19
$Z/\gamma^* \to \tau^+\tau^- + 3p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	11.33	1.19
$Z/\gamma^* \to \tau^+\tau^- + 4p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	2.59	1.19
$Z/\gamma^* \to \tau^+\tau^- + \geq 5p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	0.69	1.19
$Z/\gamma^* \to \tau^+\tau^-$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	719.18	1.18
$Z/\gamma^* \to \tau^+\tau^- + 1p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	175.72	1.18
$Z/\gamma^* \to \tau^+\tau^- + 2p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	58.86	1.18
$Z/\gamma^* \to \tau^+\tau^- + 3p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	15.66	1.18
$Z/\gamma^* \to \tau^+\tau^- + 4p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	4.01	1.18
$Z/\gamma^* \to \tau^+\tau^- + \geq 5p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	1.26	1.18
$Z/\gamma^* \to \ell^+\ell^-$ processes			
$Z/\gamma^* \to e^+e^-$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	3477.2	1.19
$Z/\gamma^* \to e^+e^- + 1p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	108.80	1.19
$Z/\gamma^* \to e^+e^- + 2p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	52.77	1.19
$Z/\gamma^* \to e^+e^- + 3p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	11.30	1.19
$Z/\gamma^* \to e^+e^- + 4p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	2.58	1.19
$Z/\gamma^* \to e^+e^- + \geq 5p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	0.69	1.19
$Z/\gamma^* \to e^+e^-$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	718.97	1.18
$Z/\gamma^* \to e^+e^- + 1p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	175.70	1.18

Table 3.1: Simulated processes used for signal and background modeling. For each process, the generator, the cross-section times the generator filter efficiency and the K-factor used to scale the simulated cross-section value to the latest theoretical calculation are given. The full simulation setup is specified in Table 2.1. $X+np$ with n integer and X a generic process indicates the number of partons (quarks or gluons) produced at matrix element level in addition to the process X.

Process	Generator	$\sigma \times \varepsilon_{filter}$ [pb]	K-factor
$Z/\gamma^* \to e^+e^- + 2p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	58.88	1.18
$Z/\gamma^* \to e^+e^- + 3p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	1.56	1.18
$Z/\gamma^* \to e^+e^- + 4p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	4.01	1.18
$Z/\gamma^* \to e^+e^- + \geq 5p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	1.26	1.18
$Z/\gamma^* \to \mu^+\mu^-$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	3477.1	1.19
$Z/\gamma^* \to \mu^+\mu^- + 1p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	108.75	1.19
$Z/\gamma^* \to \mu^+\mu^- + 2p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	52.74	1.19
$Z/\gamma^* \to \mu^+\mu^- + 3p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	11.24	1.19
$Z/\gamma^* \to \mu^+\mu^- + 4p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	2.60	1.19
$Z/\gamma^* \to \mu^+\mu^- + \geq 5p$ ($M_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Herwig	0.69	1.19
$Z/\gamma^* \to \mu^+\mu^-$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	719.16	1.18
$Z/\gamma^* \to \mu^+\mu^- + 1p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	175.74	1.18
$Z/\gamma^* \to \mu^+\mu^- + 2p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	58.88	1.18
$Z/\gamma^* \to \mu^+\mu^- + 3p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	15.67	1.18
$Z/\gamma^* \to \mu^+\mu^- + 4p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	4.01	1.18
$Z/\gamma^* \to \mu^+\mu^- + \geq 5p$ ($M_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	1.25	1.18
$W^\pm$ + jets processes			
$W^\pm \to e^\pm\nu$	Alpgen+Pythia6	8136.8	1.15
$W^\pm \to e^\pm\nu + 1p$	Alpgen+Pythia6	1791.5	1.15
$W^\pm \to e^\pm\nu + 2p$	Alpgen+Pythia6	541.60	1.15
$W^\pm \to e^\pm\nu + 3p$	Alpgen+Pythia6	146.65	1.15
$W^\pm \to e^\pm\nu + 4p$	Alpgen+Pythia6	37.30	1.15
$W^\pm \to e^\pm\nu + \geq 5p$	Alpgen+Pythia6	11.37	1.15
$W^\pm \to \mu^\pm\nu$	Alpgen+Pythia6	8133.4	1.15
$W^\pm \to \mu^\pm\nu + 1p$	Alpgen+Pythia6	1792.7	1.15
$W^\pm \to \mu^\pm\nu + 2p$	Alpgen+Pythia6	541.27	1.15
$W^\pm \to \mu^\pm\nu + 3p$	Alpgen+Pythia6	146.49	1.15
$W^\pm \to \mu^\pm\nu + 4p$	Alpgen+Pythia6	37.33	1.15
$W^\pm \to \mu^\pm\nu + \geq 5p$	Alpgen+Pythia6	11.41	1.15
$W^\pm \to \tau^\pm\nu$	Alpgen+Pythia6	8135.7	1.15
$W^\pm \to \tau^\pm\nu + 1p$	Alpgen+Pythia6	1793.7	1.15
$W^\pm \to \tau^\pm\nu + 2p$	Alpgen+Pythia6	541.24	1.15
$W^\pm \to \tau^\pm\nu + 3p$	Alpgen+Pythia6	146.48	1.15
$W^\pm \to \tau^\pm\nu + 4p$	Alpgen+Pythia6	37.26	1.15

Table 3.1: Simulated processes used for signal and background modeling. For each process, the generator, the cross-section times the generator filter efficiency and the K-factor used to scale the simulated cross-section value to the latest theoretical calculation are given. The full simulation setup is specified in Table 2.1. $X+np$ with n integer and X a generic process indicates the number of partons (quarks or gluons) produced at matrix element level in addition to the process X.

Process	Generator	$\sigma \times \varepsilon_{filter}$ [pb]	K-factor
$W^\pm \to \tau^\pm\nu + \geq 5p$	Alpgen+Pythia6	11.54	1.15
Top quark processes			
$t\bar{t}\,(\geq 1W^\pm \to \ell\nu_\ell)$	MC@NLO	129.27	1
$t\bar{t}\,(W^\pm \to qq')$	MC@NLO	108.79	1
single-top (s channel, $W^\pm \to e\nu_e$)	MC@NLO	0.61	1
single-top (s channel, $W^\pm \to \mu\nu_\mu$)	MC@NLO	0.61	1
single-top (s channel, $W^\pm \to \tau\nu_\tau$)	MC@NLO	0.61	1
$tW^\pm$	MC@NLO	22.37	1
single-top (t channel, $W^\pm \to e\nu_e$)	AcerMC+Pythia6	9.48	1
single-top (t channel, $W^\pm \to \mu\nu_\mu$)	AcerMC+Pythia6	9.48	1
single-top (t channel, $W^\pm \to \tau\nu_\tau$)	AcerMC+Pythia6	9.48	1
VV processes			
$W^\pm Z$	Herwig	6.80	1
ZZ	Herwig	1.55	1
$W^+W^- \to \ell\nu\ell\nu$	Alpgen+Herwig	2.50	1.21
$W^+W^- \to \ell\nu\ell\nu + 1p$	Alpgen+Herwig	1.25	1.21
$W^+W^- \to \ell\nu\ell\nu + 2p$	Alpgen+Herwig	0.59	1.21
$W^+W^- \to \ell\nu\ell\nu + 3p$	Alpgen+Herwig	0.33	1.21
$W^+W^- \to qq'\ell\nu$	Alpgen+Herwig	9.98	1.26
$W^+W^- \to qq'\ell\nu + 1p$	Alpgen+Herwig	5.01	1.26
$W^+W^- \to qq'\ell\nu + 2p$	Alpgen+Herwig	2.37	1.26
$W^+W^- \to qq'\ell\nu + 3p$	Alpgen+Herwig	1.31	1.26
$gg \to W^+W^- \to e^+\nu e^-\nu$	gg2WW+Herwig	0.024	1
$gg \to W^+W^- \to e^+\nu\mu^-\nu$	gg2WW+Herwig	0.021	1
$gg \to W^+W^- \to e^+\nu\tau^-\nu$	gg2WW+Herwig	0.021	1
$gg \to W^+W^- \to \mu^+\nu\mu^-\nu$	gg2WW+Herwig	0.024	1
$gg \to W^+W^- \to \mu^+\nu e^-\nu$	gg2WW+Herwig	0.021	1
$gg \to W^+W^- \to \mu^+\nu\tau^-\nu$	gg2WW+Herwig	0.021	1
$gg \to W^+W^- \to \tau^+\nu\tau^-\nu$	gg2WW+Herwig	0.024	1
$gg \to W^+W^- \to \tau^+\nu e^-\nu$	gg2WW+Herwig	0.021	1
$gg \to W^+W^- \to \tau^+\nu\mu^-\nu$	gg2WW+Herwig	0.021	1

3.2 $Z \to \mu\tau_{\mathrm{had}}$ event selection

The signature of the production a $Z \to \mu\tau_{\mathrm{had}}$ event is the presence of a reconstructed muon and hadrons reconstructed as a τ_{had}. E_{T}^{miss} can also be generated due to the neutrinos produced in the tau lepton decays. Since the muon and the τ_{had} come from the Z, their invariant mass can be as large as the Z mass and they must have opposite electric charge. Due to the energy carried away by neutrinos, the μ-τ_{had} invariant mass is generally smaller (30% less on average). $Z \to \tau^+\tau^-$ events can also contain additional activity in the final state like hadronic jets and radiated photons.

To avoid non-collisional events, at least one interaction vertex with four tracks must be reconstructed. The vertex with higher $\sum p_{\mathrm{T}}^2$ (track) is considered the point of production of the Z. Quality requirements are applied to discard incomplete events or events with detector failures or containing background from beam interactions and cosmic rays.

3.2.1 Lepton selection

Muons, electrons and τ_{had} are reconstructed as explained in Section 2.4. The sets of general requirements on the acceptance and the quality of the reconstructed particles have been introduced in Section 2.4.2.1 for muons, in Section 2.4.1.2 for electrons and in Section 2.4.8.2 for τ_{had}. Minimal p_{T} requirements are also applied to avoid that high rate low energy particles (produced for instance in the underlying event and pileup events) are erroneously reconstructed as μ, τ_{had} or e. The reconstructed leptons need to pass additional criteria that are listed in Table 3.2.

Particle	Muon	τ_{had}	Electron
Particle count selection	$N(\mu) = 1$	$\Delta R(\tau_{\mathrm{had}}, \mu) > 0.4$ $N(\tau_{\mathrm{had}}) = 1$	$\Delta R(e, \mu) > 0.2$ $N(e) = 0$
Final selection	p_{T} >26 GeV activates the SMTrig $ptcone40/p_{\mathrm{T}} < 0.06$ $etcone20/p_{\mathrm{T}} < 0.06$	$Q(\tau_{\mathrm{had}}) \cdot Q(\mu)$ =-1	

Table 3.2: Selection applied on muons, τ_{had} and electrons to isolate events with one τ_{had} and one muon. Additional requirements on the muons and τ_{had} are then applied (Final selection).

The ΔR requirements are applied to ensure that the reconstructed τ_{had} or electrons are not due to mis-reconstructed muons. The reconstructed events must contain exactly one muon and one τ_{had} and no electrons in order to reduce backgrounds with multiple leptons. The charge of the muon and the τ_{had} must be opposite.

The muon selection is further refined to suppress backgrounds with non-prompt muons. Isolation criteria are required: $ptcone40/p_{\mathrm{T}} < 0.06$ and $etcone20/p_{\mathrm{T}} < 0.06$, where p_{T} is the transverse momentum of the muon itself. Since data events are collected with the

single muon triggers (SMTrig), the selected muon must be the one firing the SMTrig and its p_{T} is at least 26 GeV to guarantee maximum triggering efficiency.

3.2.2 Event selection optimization

The fraction of background events from processes other than $Z \to \tau^+\tau^-$ in the sample selected as in Section 3.2.1 can be as large as 75%. The purity of the sample can be increased by applying additional topological selections. This event selection has been tuned considering its effect on:

- the fraction of expected signal and background events;
- the statistical significance of the selected sample by using $S/\sqrt{(S+B)}$ as a figure of merit, where S and B are the signal and background events, respectively;
- the distortion or exclusion of parts of the τ_{had} p_{T} spectrum as the analysis aims at testing the overall agreement between the TES in data and in simulations;
- the sensitivity of the TES measurements to variations of the selection thresholds.

The kinematic variables used in the selection are listed below and their distributions are shown in Figure 3.2.

Transverse mass, $M_{\mathrm{T}}\left(\mu, E_{\mathrm{T}}^{miss}\right)$. The transverse mass is required to be smaller than 50 GeV to reduce $W^{\pm}$+jets background. Since the longitudinal momentum of the neutrino is not retrievable, the $W^{\pm}$ mass cannot be reconstructed, but the transverse mass M_{T} can be. The transverse mass of a system is the invariant mass of the system after having set the momentum along the z-axis to zero. $M_{\mathrm{T}}\left(\mu, E_{\mathrm{T}}^{miss}\right)$ is the transverse mass of the E_{T}^{miss}-μ system. For relativistic particles like muons and neutrinos, assumed to be the responsible for the E_{T}^{miss} in $W^{\pm}+$jets events, the transverse mass assumes the form:

$$M_{\mathrm{T}}\left(\mu, E_{\mathrm{T}}^{miss}\right) \simeq \sqrt{2p_{\mathrm{T}}\left(\mu\right) E_{\mathrm{T}}^{miss}\left(1-\cos\left(\Delta\phi\left(\mu, E_{\mathrm{T}}^{miss}\right)\right)\right)} \tag{3.2}$$

where $\Delta\phi\left(\mu, E_{\mathrm{T}}^{miss}\right)$ is the angle between E_{T}^{miss} and μ in the transverse plane. $Z \to \tau^+\tau^-$ events populate mainly the low $M_{\mathrm{T}}\left(\mu, E_{\mathrm{T}}^{miss}\right)$ region, while $W^{\pm}+$jets events have larger $M_{\mathrm{T}}\left(\mu, E_{\mathrm{T}}^{miss}\right)$ values peaking at about 75 GeV. Therefore, the $M_{\mathrm{T}}\left(\mu, E_{\mathrm{T}}^{miss}\right) < 50$ GeV requirement reduces the $W^{\pm}+$jets background. Since $W^{\pm}$ also produced in the *top* decay, the *top* quark background is reduced too.

Sum of cosines of azimuthal angles between μ-E_{T}^{miss} and τ_{had}-E_{T}^{miss}, $\Sigma\cos\Delta\phi$. $\Sigma\cos\Delta\phi = \cos\Delta\phi\left(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{miss}\right) + \cos\Delta\phi\left(\mu, E_{\mathrm{T}}^{miss}\right)$ is sensitive to how collinear is the E_{T}^{miss} direction with the reconstructed particles. In $Z \to \tau^+\tau^-$ events, E_{T}^{miss} is mainly due to the emission of neutrinos from the tau lepton decays, so it is generally small and aligned

with the τ_{had} or the μ for relativistic taus. In such events, either $\cos\Delta\phi\left(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{miss}\right)$ or $\cos\Delta\phi\left(\mu, E_{\mathrm{T}}^{miss}\right)$ is about one and the sum of the two is usually positive. When the E_{T}^{miss} is due to experimental energy resolution effects or to neutrinos produced by heavy particles, e.g. $W^{\pm}$, the E_{T}^{miss} does not align with τ_{had} and μ and $\Sigma\cos\Delta\phi$ is preferentially negative. The events with $\Sigma\cos\Delta\phi < -0.5$ are therefore excluded from the measurement.

Difference in pseudo-rapidity between τ_{had} and μ, $\Delta\eta\left(\tau_{\mathrm{had}}, \mu\right)$. The $\Delta\eta\left(\tau_{\mathrm{had}}, \mu\right)$ is generally small in $Z \to \tau^{+}\tau^{-}$ events because the τ_{had} and μ are produced in the same resonant decay. When the τ_{had} and the muon have different origin, as for instance in $W^{\pm}$ + jets, $t\bar{t}$ and $W^{+}W^{-}$ events, the difference in pseudorapidity can be larger. The maximum value allowed in the analysis for $\Delta\eta\left(\tau_{\mathrm{had}}, \mu\right)$ is 1.5.

Difference in p_{T} between τ_{had} and μ, $p_{\mathrm{T}}\left(\tau_{\mathrm{had}}\right) - p_{\mathrm{T}}\left(\mu\right)$. In background events the τ_{had} tends to have p_{T} smaller than the muon especially when the τ_{had} is a mis-identified jet. Jets have indeed a different energy calibration and are produced with lower p_{T} unless they come from the decay of a heavy resonances. For $Z \to \tau^{+}\tau^{-}$ events the leptonic tau decay produces two neutrinos, while the hadronic decay only one. The τ_{had} on average has therefore a larger p_{T} then the muon. Events with $p_{\mathrm{T}}\left(\tau_{\mathrm{had}}\right) - p_{\mathrm{T}}\left(\mu\right) < -15$ GeV are mostly backgrounds and therefore are excluded from the analysis.

Other variables have been considered but they were found less effective in suppressing the backgrounds.

The topological selection described above is applied to τ_{had}-μ events and defines the sample of events used for the TES measurement.

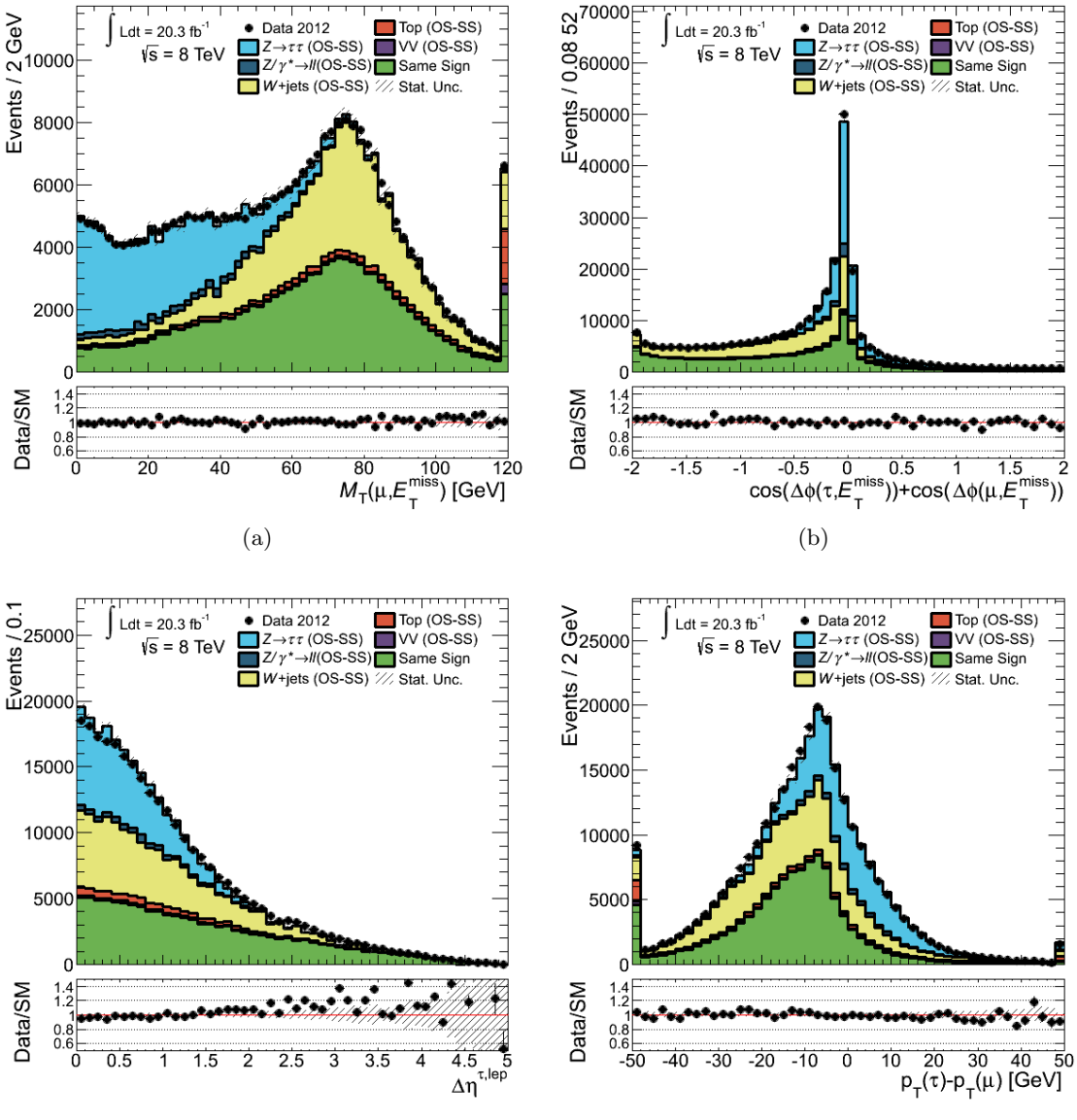


Figure 3.2: Distributions of the transverse mass $M_{\mathrm{T}}\left(\mu, E_{\mathrm{T}}^{miss}\right)$ (a), the $\Sigma \cos \Delta\phi$ (b), the $\Delta\eta\left(\tau,\mu\right)$ (c) and the $p_{\mathrm{T}}\left(\tau\right) - p_{\mathrm{T}}\left(\mu\right)$ (d) for $\mu\tau_{\mathrm{had}}$ events as selected in Section 3.2.1. Events in data (black dots) are compared to the expected stacked contributions from $Z \to \tau^+\tau^-$ (light blue) and background processes determined as described in Section 3.3. Only statistical uncertainties are shown. Underflow and overflow events are added to the first and last bin, respectively. The lower panel shows the ratio of observed events in data to the sum of expected signal and background events.

3.3 Background estimate

The dominant sources of background events for this measurement can be classified in two categories: prompt and non-prompt.

Prompt backgrounds produce a muon and a τ_{had} directly in the primary interaction or via the decay of short lived particles like $W^{\pm}$, Z, H and tau lepton. The muon and the τ_{had} produced in this way are called prompt. The prompt processes relevant for this analysis are $t\bar{t}$, VV ($= W^{+}W^{-}$, $W^{\pm}Z$, ZZ), $tW^{\pm}$ productions, with $t\bar{t}$ being the dominant. These backgrounds have been reduced by rejecting events with additional leptons or τ_{had} and certain topological properties. A small fraction of events with multiple leptons is nonetheless accepted due to leptons outside the detector acceptance or not reconstructed. The prompt background estimate is based on simulations, as explained in Section 3.1.

Non-prompt backgrounds are events with at least one non-prompt muon or one non-prompt τ_{had}. A reconstructed muon or τ_{had} is considered non-prompt when the particle they are originated from, is either a different particle, or produced in the decay of hadrons or produced via interaction with matter. The probability for a non-prompt particle to pass the muon or τ_{had} selection requirements is very small, however the production rate of non-prompt particle sources is so high that non-prompt backgrounds cannot be neglected. The dominant sources of non-prompt background events are QCD multi-jet and $W^{\pm}+\text{jets}$, but smaller contributions come also from $Z \to \mu^{+}\mu^{-}+\text{jets}$, $t\bar{t}$ and *single-top*. In QCD multi-jet events both the muon and the τ_{had} are non-prompt while in the other processes there is usually only one non-prompt particle, most likely the τ_{had}. Non-prompt particles can be modeled by simulation although with worse accuracy compared to prompt particles. In particular the probability for non-prompt particles to be selected is difficult to model. The backgrounds with one non-prompt particle and one prompt particle are estimated with simulations, and data-driven scale factors are used to correct their rate. For backgrounds with two non-prompt particles, simulations are avoided. The modeling of these processes in simulation is inaccurate and the statistics required to produce sensitive predictions is computationally unsustainable. The background with two non-prompt particles is obtained entirely from data using events where the muon and the τ_{had} have same electric charge (same-sign particles). The method to estimate non-prompt backgrounds is called the $OS-SS$ method and it is described in the following section.

3.3.1 The $OS-SS$ method

The $OS-SS$ method is based on the idea that in non-prompt background events the electric charge of the reconstructed muon and τ_{had} can be either the same (same-sign or SS) or opposite (opposite-sign or OS), while for signal and prompt backgrounds the charge is mostly opposite. Therefore same-sign μ-τ_{had} events contain mainly non-prompt backgrounds and can be used to estimate the contribution of non-prompt events passing the selection with opposite-sign particles (OS region).

Backgrounds with at least one prompt particle are evaluated using simulations. If the background has also a non-prompt particle, its simulation is corrected with scale factors (k) in order to address possible mis-modeling of the $\mu \to \tau_{\text{had}}$ and $jet \to \tau_{\text{had}}$ mis-identification rates and to reduce systematic uncertainties on the background normalization.

Backgrounds with no prompt particles (mainly QCD multi-jet) are obtained from SS data after subtraction of events with prompt particles. Scale factors k are applied to backgrounds with one non-prompt particle before subtraction. The scale factors k are calculated independently for $Z \to \mu^+\mu^-$, $W^\pm$ + jets, VV and events with *top* quarks in dedicated control regions. The scale factors can be different in case of SS and OS particles. The $Z \to \mu^+\mu^-$ background is divided into events where the non-prompt τ_{had} is a muon ($Z \to \mu^+\mu^- (\to \tau_{\text{had}})$) or a jet ($Z \to \mu^+\mu^-$ + jets $(\to \tau_{\text{had}})$) and each has its own scale factor.

The number of QCD multi-jet events in the SS region is

$$N_{SS}^{QCD} = N_{SS}^{Data} - \left(\sum_{i \in R} k_{SS}^i \cdot N_{SS}^i \right) \tag{3.3}$$

where k_{SS}^i and N_{SS}^i are the scale factors and the number of expected events for the i-th process with at least one prompt particle in the SS region and N_{SS}^{Data} is the observed number of events in the same region. R is the ensemble of the processes with at least one prompt particle. The shape of the distribution for QCD multi-jet is obtained analogously with the Equation 3.3 by replacing the number of events N_{SS}^i with the corresponding distribution h_{SS}^i.

The total number of expected background events in the OS region is given by the Equation 3.4.

$$N_{OS}^{bkg} = r_{QCD} \cdot N_{SS}^{QCD} + \left(\sum_{i \in R} k_{OS}^i \cdot N_{OS}^i \right) \tag{3.4}$$

where k_{OS}^i and N_{OS}^i are the scale factors and the number of expected Standard Model events with at least one prompt particle in the OS region, and r_{QCD} accounts for potential differences in $jet \to \mu$ and $jet \to \tau_{\text{had}}$ mis-identification rates introduced by the same or opposite charge requirements. This is due to differences in flavour composition of final state jets in the two regions. The method of non-prompt background estimate assumes that the QCD multi-jet distributions, relevant to the analysis, in the SS and OS region have the same shape.

Equation 3.4 can be rewritten as:

$$N_{OS}^{bkg} = r_{QCD} \cdot N_{SS}^{Data} + \left(\sum_{i \in R} N_{OS-SS}^i \right) \tag{3.5}$$

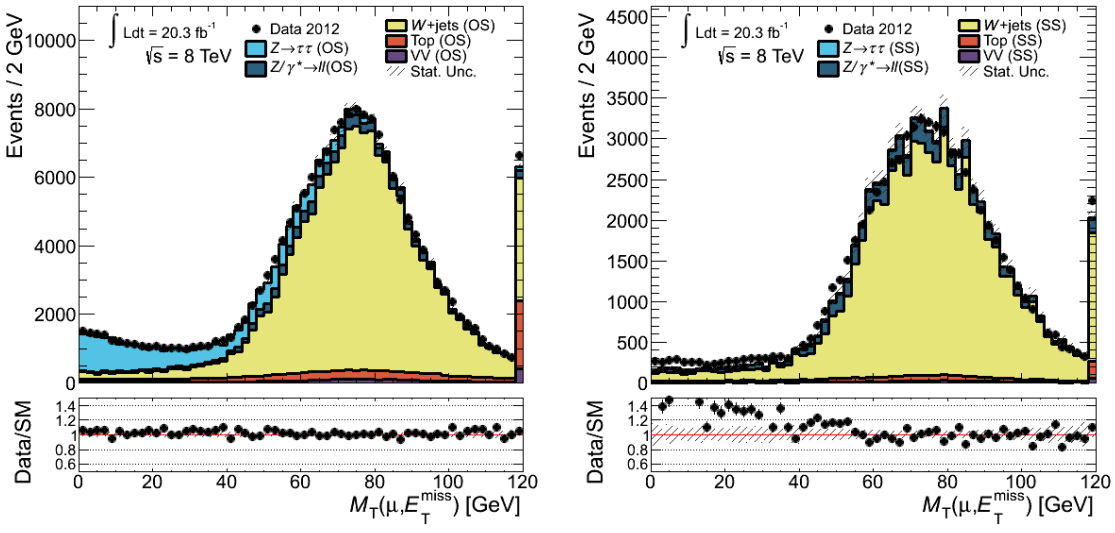
with $N_{OS-SS}^i = k_{OS}^i \cdot N_{OS}^i - r_{QCD} \cdot k_{SS}^i \cdot N_{SS}^i$.

3.3.1.1 Scale factor measurements

The factors r_{QCD}, k^i_{OS} and k^i_{SS} are evaluated in dedicated control regions. The selection described in Section 3.2.2 does not apply to the control regions but other conditions must be satisfied. The measurements of $k^{W^\pm+\text{jets}}_{OS}$ and $k^{W^\pm+\text{jets}}_{SS}$ have been done by the author, while the other scale factors have been evaluated in another analysis that shares the same final state and uses the same background estimate method [27].

The r_{QCD} is derived from QCD multi-jet enriched control region in data. At leading order, QCD di-jet events include the following parton pairs in the final state: $q\bar{q}$, qq', $q\bar{q'}$, qg and gg. Relative fractions of these parton pairs are different for OS and SS non-prompt events. No electric charge correlation between a non-prompt muon and a non-prompt τ_{had} is expected in events with jets from qg and gg parton pairs. However, charge correlation is expected in events with jets from $q\bar{q}$, qq', and $q\bar{q'}$ pairs, due to the proton-proton initial state. The QCD multi-jet-enriched data control region is defined by requiring $E^{miss}_{\text{T}} < 15$ GeV and $M_{\text{T}} < 30$ GeV, allowing the muon to have any $etcone20$ isolation and the τ_{had} to pass a looser identification criterion. Contributions from electroweak and top backgrounds in this control region are around 27% and subtracted based on simulated predictions. A linear fit of r_{QCD} as a function of the $etcone20/p_{\text{T}}$ isolation requirement is made for $0.06 < etcone20/p_{\text{T}} < 0.38$ to obtain the value for the $etcone20/p_{\text{T}} < 0.06$ requirement applied in this analysis. The measured r_{QCD} values is 1.11 ± 0.08.

The $W^\pm$ + jets control regions are defined separately for opposite-sign and same-sign particles and requiring $E^{miss}_{\text{T}} > 20$ GeV and $M_{\text{T}}\left(\mu, E^{miss}_{\text{T}}\right) > 70$ GeV. The large E^{miss}_{T} requirement is due to the neutrino produced in the decay of the $W^\pm$. For events with $W^\pm$ decays the $M_{\text{T}}\left(\mu, E^{miss}_{\text{T}}\right)$ tends to be large and close to 90 GeV, while a smaller $M_{\text{T}}\left(\mu, E^{miss}_{\text{T}}\right)$ is expected in other processes. The $k^{W^\pm+\text{jets}}$ are calculated separately for OS and SS events, since the $jet \to \tau_{\text{had}}$ mis-identification rate is different for jets from quark or gluon hadronization. The proportion of quark initiated jets is higher in the opposite-sign selection with respect to the gluon component, that is dominant in the same-sign selection. Figure 3.3 shows the $M_{\text{T}}\left(\mu, E^{miss}_{\text{T}}\right)$ distribution for events in the $W^\pm$+jets control regions without the $M_{\text{T}}\left(\mu, E^{miss}_{\text{T}}\right) > 70$ GeV requirements. The $k^{W^\pm+\text{jets}}$ are calculated considering only the events with $M_{\text{T}} > 70$ GeV where the contribution of other processes including $Z \to \tau^+\tau^-$ is very small. The background from QCD multi-jet is negligible in that interval. In QCD multi-jet events, the E^{miss}_{T} is created by the finite energy resolution of the detector or by wrong energy calibration of the reconstructed particles. When a jet is reconstructed as an isolated and well identified muon, some of the jet energy is usually lost and E^{miss}_{T} tends to point in the same direction of the lepton. As a consequence, the $M_{\text{T}}\left(\mu, E^{miss}_{\text{T}}\right)$ is small. The QCD multi-jet contribution would appear as an excess of data over the estimate from simulation. This is visible in the SS region for $M_{\text{T}} < 20$ GeV in Figure 3.3(b), but such contribution is negligible above 70 GeV of transverse mass. The measured scale factors in SS and OS events are $k^{W^\pm+\text{jets}}_{OS} = 0.80 \pm 0.05$



(a) (b)

Figure 3.3: The M_{T} distribution in the regions used to determine $k_{OS}^{W^{\pm}+\mathrm{jets}}$ (a) and the $k_{SS}^{W^{\pm}+\mathrm{jets}}$ (b). Events in data (black dots) are compared to the expected stacked contributions from $Z \to \tau^+\tau^-$ (light blue) and background processes determined as described in Section 3.3. Only statistical uncertainties are shown. Underflow and overflow events are added to the first and last bin, respectively. The lower panel shows the ratio of observed events in data to the sum of expected signal and background events.

and $k_{SS}^{W^{\pm}+\mathrm{jets}} = 0.93 \pm 0.08$, respectively.

The prediction of $Z \to \mu^+\mu^-$ backgrounds where one of the muons is reconstructed as τ_{had} are validated in two control regions with opposite-sign and same-sign leptons. The invariant mass of the muon and the τ_{had} leading p_{T} track, thought to be the other muon in $Z \to \mu^+\mu^-$ events, must be in 80 GeV $< M(\mu, \mathrm{track}) <$ 100 GeV in the OS and SS regions. The prediction from simulations agrees with data within the statistical uncertainty in both regions, hence no scaling is applied to simulations.

In $Z \to \mu^+\mu^-$ + jets events with a jet passing the τ_{had} selection, there is no charge correlation between the τ_{had} candidate and the selected muon. Therefore the number of expected events in the OS and SS regions for this process is the same ($N_{SS}^{Z\to\mu^+\mu^-+\mathrm{jets}} \equiv N_{OS}^{Z\to\mu^+\mu^-+\mathrm{jets}}$). The scale factor $k_{OS}^{Z\to\mu^+\mu^-+\mathrm{jets}(\to\tau_{\mathrm{had}})} \equiv k_{SS}^{Z\to\mu^+\mu^-+\mathrm{jets}(\to\tau_{\mathrm{had}})}$, is calculated from the control region with events with two oppositely charged leptons (e^+e^-, $\mu^+\mu^-$) and a reconstructed τ_{had}. Events with two electrons are included to increase the statistics since the $jet \to \tau_{\mathrm{had}}$ mis-identification probability is independent from the lepton flavour. The simulation overestimates the mis-identification probability, so a scale factor of 0.56 ± 0.03 is obtained.

The $t\bar{t}$ and *single-top* processes can produce, in addition to a muon and a τ_{had}, multiple high-p_{T} jets, including those originated from b-quarks, and high E_{T}^{miss}. The control region for the measurement of k_{OS}^{top} and k_{SS}^{top} is obtained by requiring $M_{\mathrm{T}} > 50$ GeV, $E_{\mathrm{T}}^{miss} >$

20 GeV, at least two jets with $p_{\mathrm{T}} > 30$ GeV and the highest p_{T} jet to be b-tagged (see Section 2.4.7). Scale factors are calculated separately for opposite-sign and same-sign events: $k_{OS}^{top} = 0.84 \pm 0.16$ and $k_{SS}^{top} = 0.95 \pm 0.22$.

Multi boson production generates mostly prompt muon and τ_{had} and constitutes the smallest background considered in the analysis. Multi boson predictions are therefore based only on simulations.

The scale factors used to estimate the background processes in this analysis are summarized in Table 3.3.

Scale factor	Value	Scale factor	Value
$k_{OS}^{W^{\pm}+\mathrm{jets}}$	0.80 ± 0.05	$k_{SS}^{W^{\pm}+\mathrm{jets}}$	0.93 ± 0.08
$k_{OS}^{Z\to\mu^+\mu^-(\to\tau_{\mathrm{had}})}$	1	$k_{SS}^{Z\to\mu^+\mu^-(\to\tau_{\mathrm{had}})}$	1
$k_{OS}^{Z\to\mu^+\mu^-+\mathrm{jets}(\to\tau_{\mathrm{had}})}$	0.56 ± 0.03	$k_{SS}^{Z\to\mu^+\mu^-+\mathrm{jets}(\to\tau_{\mathrm{had}})}$	0.56 ± 0.03
k_{OS}^{top}	0.84 ± 0.16	k_{SS}^{top}	0.95 ± 0.22
k_{OS}^{VV}	1	k_{SS}^{VV}	1
r_{QCD}	1.11 ± 0.08		

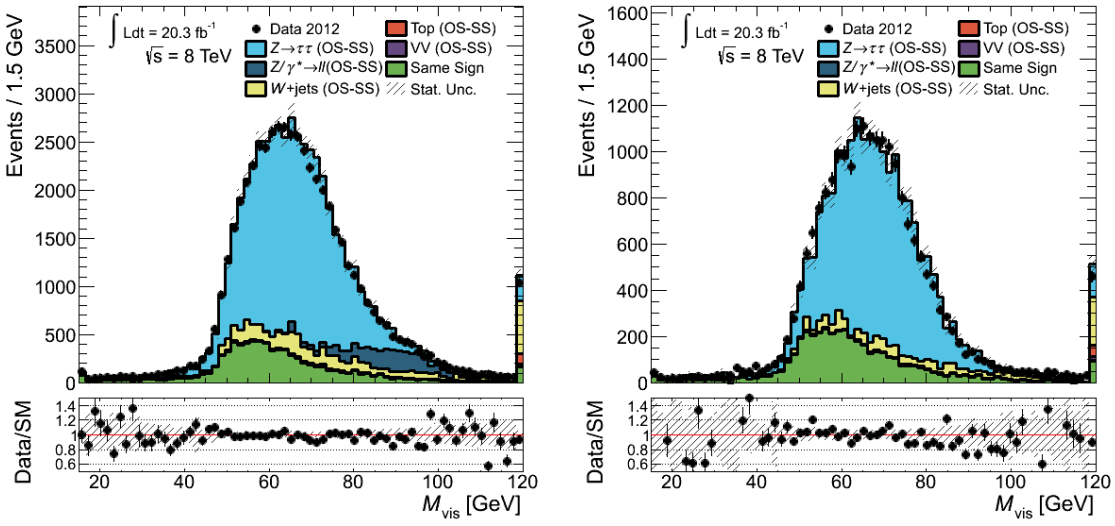
Table 3.3: Scale factors measured in the control regions and used in the background estimate with the $OS - SS$ method.

3.4 TES measurement

The selection described in Section 3.2 defines the signal region of the analysis. The M_{vis} distribution of the events in the signal region is shown in Figure 3.4, separately for 1-track and 3-tracks τ_{had}. The sources of backgrounds are estimated in Section 3.3 and are grouped in the following categories:

- $Z \to \ell^+\ell^- = Z \to \mu^+\mu^- \,(\to \tau_{\mathrm{had}}) + Z \to \mu^+\mu^- + \mathrm{jets}\,(\to \tau_{\mathrm{had}})$ $(OS - SS)$
- $W^{\pm} + \mathrm{jets}$ $(OS - SS)$
- $top = t\bar{t} + \textit{single-top}$ $(OS - SS)$
- $VV = W^+W^- + W^{\pm}Z + ZZ$ $(OS - SS)$
- Same Sign $= r_{QCD} \cdot N_{SS}^{Data}$

For $Z \to \ell^+\ell^-$, $W^{\pm} + \mathrm{jets}$, top and VV, the $OS - SS$ contributions are shown as obtained from Equation 3.5. Contributions from other processes, e.g. the Higgs boson production, are found negligible. The background contamination in the selected events is the 31% (28%) for the 1-track (3-tracks) τ_{had}. The intervals $M_{vis} < 45$ GeV and $M_{vis} > 90$ GeV, however, is dominated by background events and contain a small fraction of $Z \to \tau^+\tau^-$ events. The interval 45 GeV $< M_{vis} <$ 90 GeV is therefore used in the TES



(a) 1-track τ_{had} (b) 3-tracks τ_{had}

Figure 3.4: The M_{vis} distribution after the optimized selection for τ_{had} with one (a) or three (b) associated tracks. Events in data (black dots) are compared to the expected stacked contributions from $Z \to \tau^+\tau^-$ (light blue) and background processes determined as described in Section 3.3. Only statistical uncertainties are shown. Underflow and overflow events are added to the first and last bin, respectively. The lower panel shows the ratio of observed events in data to the sum of expected signal and background events.

measurement. Table 3.4 contains the number of data events and the expected events from signal and backgrounds processes in this M_{vis} interval. The $Z \to \mu\tau_{\text{had}}$ events in the M_{vis}

Process	Any τ_{had}	1-track τ_{had}	3-tracks τ_{had}
$Z \to \tau^+\tau^-$ $(OS-SS)$	52100 ± 400	37200 ± 350	14860 ± 220
$Z \to \mu^+\mu^-$ $(\to \tau_{\text{had}})$ $(OS-SS)$	2170 ± 130	2170 ± 130	3 ± 4
$Z \to \mu^+\mu^-$ + jets $(\to \tau_{\text{had}})$ $(OS-SS)$	-120 ± 100	-90 ± 90	-40 ± 60
$W^\pm$ + jets $(OS-SS)$	4800 ± 220	3700 ± 180	1080 ± 130
top $(OS-SS)$	444 ± 15	330 ± 12	114 ± 9
VV $(OS-SS)$	188 ± 9	139 ± 7	49 ± 5
Same sign	10340 ± 110	6650 ± 90	3690 ± 60
Total SM	69900 ± 500	50100 ± 400	19760 ± 270
Data	68620 ± 260	49080 ± 220	19540 ± 140

Table 3.4: Expected and observed numbers of events in the M_{vis} interval 45 GeV $< M_{vis} <$ 90 GeV and passing the analysis selection. Statistical uncertainties are shown.

peak contribute to the 74% and 75% of the total expected events with 1-track and 3-tracks τ_{had}, respectively. The expected number of $Z \to \tau^+\tau^-$ events with same-sign μ-τ_{had} pairs is negligible, therefore $N_{OS}^{Z\to\tau^+\tau^-} \simeq N_{OS-SS}^{Z\to\tau^+\tau^-}$. The largest backgrounds is the Same sign and is followed by $W^\pm$+jets $(OS-SS)$ and $Z \to \mu^+\mu^-$ $(\to \tau_{\text{had}})$ $(OS-SS)$. The expected $Z \to \mu^+\mu^-$ + jets $(\to \tau_{\text{had}})$ $(OS-SS)$ contribution is negative by construction given that

$r_{QCD} > 1$ and that $k_{SS} = k_{OS}$ and $N_{SS} = N_{OS}$ for this process. Note that this is physically acceptable, as this contribution corrects for the excess of $Z \to \mu^+\mu^- + \text{jets}\,(\to \tau_{\text{had}})$ events predicted by the Same sign background.

Predictions are in good agreement with data, even though systematic uncertainties have not been considered at this stage. The M_{vis} distributions obtained in $Z \to \tau^+\tau^-$ simulation and in data after subtraction of the residual expected background are investigated in search for potential differences attributable to biases in the τ_{had} energy calibration. First, the effect of small variations of the TES on the M_{vis} distribution is investigated. Then, a method to measure differences in M_{vis} distributions and to translate these into relative difference in TES is developed.

The M_{vis} to the τ_{had} energy are connected by Equation 3.1. For small variations α of the τ_{had} energy $E^{\tau_{\text{had}}} \to (1+\alpha)E^{\tau_{\text{had}}}$, the visible mass would shift as

$$M_{vis} \to M_{vis}\sqrt{1+\alpha} \simeq M_{vis}(1 + \frac{1}{2}\alpha - \frac{1}{8}\alpha^2 + \cdots) \tag{3.6}$$

As a consequence, the M_{vis} distribution changes in position and width. Neglecting event selection effects, the mean value and the standard deviation of the distribution would scaled by $\sqrt{1+\alpha}$ too. Skewness and kurtosis are not expected to vary significantly. For $\alpha = 5\%$ the Taylor expansion at first order of M_{vis} is less than 0.03% larger than the exact M_{vis} value. Since the expected sensitivity to TES variations is of the order of few percents, the linear dependence of the M_{vis} from α is a good approximation. Variations in M_{vis} are therefore proportional to $\alpha/2$.

To study the relation between TES and the M_{vis} distribution, clones of the simulated samples have been made including processes with non-prompt τ_{had}. In the cloned samples, the energy of the τ_{had} is shifted of a fraction α from the nominal value. α ranges between $[-5\%, 5\%]$ in 1% steps. The clones contain therefore the same events, but with a different TES. Any variable related to the TES (e.g. $p_{\text{T}}\,(\tau_{\text{had}})$, E_{T}^{miss}, M_{T}, M_{vis}) is recalculated and the event selection is applied afterwards. Figure 3.1(b) shows the M_{vis} distribution in $Z \to \tau^+\tau^-$ simulated for $\alpha = 0$, 5% and -5%. The M_{vis} distribution shifts with α, as expected.

In order to measure variations in TES, an estimator that quantifies variations of the M_{vis} distribution position has to be defined. The challenge lies in the definition of a quantity that is sensitive to TES variations and insensitive to systematic effects as it will be described in the following. Ideally, the value of the estimator should depend linearly in α as this is the dependence expected on the M_{vis} distribution. Given an estimator ϵ, its value is measured on the M_{vis} distributions in data after background subtraction ($\epsilon_{Data}\,(\alpha)$) and in $Z \to \tau^+\tau^-$ simulation ($\epsilon_{Sim}\,(\alpha)$) for each cloned sample with $\alpha \in [-5\%, 5\%]$. The values for $\alpha = 0$ of the estimator measured in data and in simulation should be the same if the original τ_{had} energy calibration is identical in the two cases. On the other hand, a difference in the estimator values $\Delta\epsilon = \epsilon_{Data} - \epsilon_{Sim}$ would be proportional to the difference in TES

in data and simulation. The value α_0 for which $\Delta\epsilon = 0$ gives the relative shift in TES and is obtained from a linear fit of the $\Delta\epsilon$ measured in each cloned sample as a function of α.

The position of the peak formed by the $Z \to \tau^+\tau^-$ M_{vis} distribution is measured with several estimators. Both statistical properties of the distribution and parametrisations of the distributions with peak-like functions have been investigated.

The statistical properties considered are the mean and the median of the M_{vis} distribution. They are sensitive to TES variations linearly. However, two major issues were found with these estimators: median and mean are highly affected by TES shift in backgrounds (i.e. the shift of non-prompt τ_{had}) and by migrations of event in and out of the signal region. Since the background M_{vis} distribution is not flat in the interval 45 GeV $< M_{vis} <$ 90 GeV (see Figure 3.4), the mean and median of the background distribution change with α. The prompt and non-prompt τ_{had} are due to unrelated processes which can generate different signals in the detector. The signals for the two types of τ_{had} may not be modeled with the same accuracy in simulation. Thus, calibration differences for prompt τ_{had} in data and simulations are not related to calibration differences for non-prompt τ_{had}. The goal of the analysis is to compare the TES of prompt τ_{had} in data and simulations, hence the background influence must be minimal. The simulated background (i.e. Same sign background excluded) is larger at $M_{vis} = 90$ GeV than at $M_{vis} = 45$ GeV for 1-track τ_{had}. TES variations cause asymmetric migration of the background in and out the two boundaries of the signal region M_{vis} interval and this has an impact on the TES measurement. The change in the median and the mean due to the TES variation in the background simulation only are shown in Figure A.3. The changes are also evaluated for the estimates obtained with parametric functions, described later, The calibration of the background has an effect on the value of α_0 as big 40% for the median, 25% for the mean, and less than 10% for parametric functions. Other M_{vis} intervals have been considered for the median and the mean estimate, but similar background dependences were found. The median, that was used in a previous calibration study [150], and the mean have been discarded.

In order to reduce the sensitivity to the background calibration, parametric functions are instead used to fit the $Z \to \tau^+\tau^-$ M_{vis} distribution in the interval 45 GeV $< M_{vis} <$ 90 GeV and to obtain the position of the maximum of the distribution. The maximum position depends linearly on TES variations and, contrary to the mean and the median, the dependence from the background calibration and on the migration of events in and out the M_{vis} interval is limited. The simulated background indeed is small and rather uniform at the maximum position of the M_{vis} distribution (around 65 GeV), as shown in Figure 3.4. Thus, the relative position of the maximum position in data and simulation is sensitive to TES differences and minimally biased by the background calibration.

The definition of a parametric function able to describe the M_{vis} distribution is quite challenging. The shape of the $Z \to \tau^+\tau^-$ M_{vis} distribution is the result of a combination of effects. In principle, if all the particles produced in the $Z \to \tau^+\tau^-$ decay could be detected with perfect accuracy, the distribution would look like a Breit-Wigner function

with the width of the Z boson (about 2.5 GeV). However, neutrinos are not detected and take away energy from the system. The visible particles form a wider distribution peaked at lower masses. The finite resolution on the muon and τ_{had} energies and the event selection additionally alter the shape of the distribution, which is therefore extremely hard to parametrise analytically.

A set of continuous and derivable peak-like functions has been tested:

Bifurcated Gaussian: a Gaussian function with different widths on either sides of the peak. This is used to model peak-like distributions with asymmetric sides and where the resolution uncertainty is dominant;

Gaussian $\otimes$ Breit-Wigner: a convolution of a Gaussian function and a Breit-Wigner, used to model resonances with non-null decay width and resolution uncertainty with symmetric sides;

Composed Bifurcated Gaussian and Landau: sums a Bifurcated Gaussian function and a Landau function which acts on one of the distribution tails;

Crystal Ball: a function with a Gaussian peak and a power-law falling tail on one side that accounts for non-Gaussian reconstruction effects;

Bifurcated Crystal Ball: a Crystal Ball function with additional asymmetry thanks to a Bifurcated Gaussian core and power-law tails on both sides;

Polynomial (5th order): a simple polynomial function with a number of degrees of freedom tuned on the M_{vis} distribution itself.

It is found that the maximum position of the Composed Bifurcated Gaussian and Landau function fitted to the M_{vis} do not scale linearly with α in the range $[-5\%, 5\%]$, and there are two values of α for which the maximum positions in data and simulation match. This is due to the fact that the TES variation is absorbed in other parameters of the fit rather than the maximum position. The Composed Bifurcated Gaussian and Landau function has no parameters suitable to measure TES differences, as found in Section A.2, and is therefore no further considered.

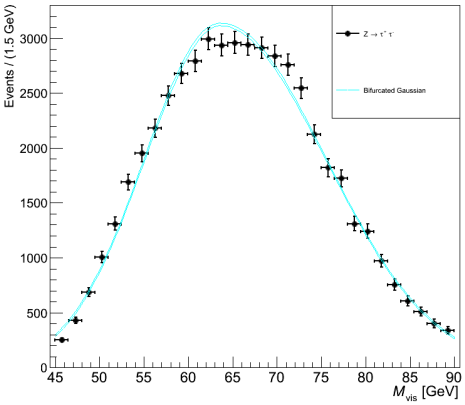
The fits to the $Z \to \tau^+\tau^-$ M_{vis} distribution in simulation and data with the remaining parametric functions are shown in Figure 3.5 and Figure 3.6, respectively. The Bifurcated Gaussian and the Bifurcated Crystal Ball functions fit well the sides of the distribution but show a bias in the modeling of the peak. The maximum position is biased in the same direction in data and simulation and the bias would, at least partially, cancel out when calculating $\Delta\epsilon$. The Crystal Ball and Breit-Wigner functions better model the maximum position compared to the bifurcated functions. However, the fitted peak shape is sharper than the M_{vis} distribution, and the low mass side of the distribution seems also to be more curvy than the line-shape obtained in both fits.

The best modeling is obtained with the Polynomial (5th order) fit which has the best χ^2 normalized to the number of degrees of freedom. Although there is no physical reason justifying the choice this function, it is enough flexible to model the visible mass distribution peak in the interval 45 GeV $< M_{vis} <$ 90 GeV. Polynomial functions of different orders have also been tested, but a higher order polynomial functions adapt to fine features of the distribution due to statistical fluctuations, while a lower order one is not able to model the line-shape adequately.

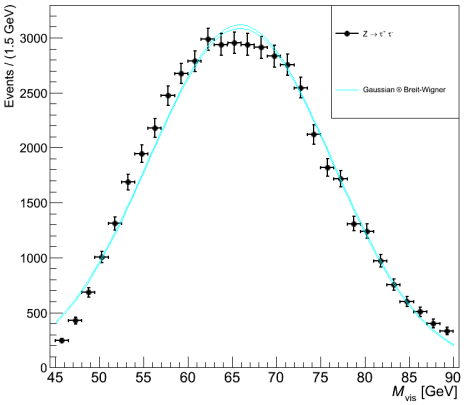
The differences of the estimator value in data and simulation, $\Delta\epsilon$, measured as a function of the TES variation α in simulation are shown in Figure 3.7 for each parametric function. The 1-track and 3-tracks τ_{had} are treated separately.

The Bifurcated Gaussian and Bifurcated Crystal Ball functions show some large fluctuations between consecutive measurements, while the Crystal Ball, Breit-Wigner and Polynomial (5th order) fits give more stable results with enhanced linearity. The relative shift in TES (α_0) measured with Crystal Ball, Breit-Wigner and Polynomial (5th order) is similar and compatible within uncertainties estimated according to the procedure explained in the next section. Because of the better modeling of the M_{vis} distribution, the stability of its maximum position estimator and the compatibility with other estimators, the Polynomial (5th order) fit is the method chosen to estimate TES corrections.

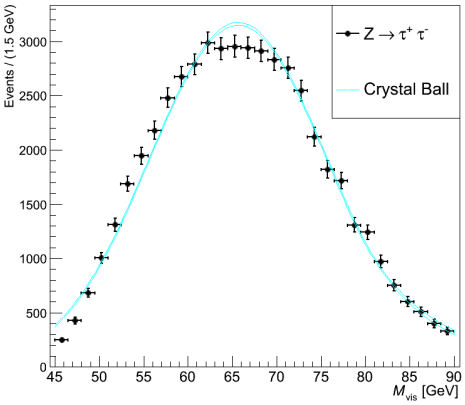
Studies to test the possibility to include additional information in the TES measurement have also been performed. The shape of the M_{vis} distributions is investigated through the use of statistical higher order moments of the distribution: the standard deviation (to measure the the spread or width), the skewness (to measure the asymmetry) and the kurtosis (to measure the "sharpness"). The standard deviation in first approximation should vary of $\alpha/2$ as for the maximum position, while the skewness and kurtosis should not be affected. However, the migration of events in the signal region as a result of a change in α may influence the moments of the M_{vis} distribution. The moments of the M_{vis} distribution measured in data, in the predictions, in $Z \to \tau^+\tau^-$ simulations and in data after background subtraction are shown in Figure 3.8 for 1-track and 3-tracks τ_{had}. The standard deviation of the simulated distributions increases linearly with α as expected, while the kurtosis and the skewness decreases as result of the event migration. The measurements of the skewness and kurtosis are affected by an higher statistical error but their dependence from α seems to be linear. The values of α for which the moments in data and simulation match, α_0, are in the interval $[0\%, +1\%]$ for 1-track τ_{had} and in $[-2\%, -1\%]$ for 3-tracks τ_{had}. These values are similar to the α_0 obtained from the Polynomial (5th order) fit: -0.8% for 1-track τ_{had} and -1.1% for 3-tracks τ_{had}. Although this study is just preliminary and uncertainties on the measured α_0 should be evaluated properly, the study of the moments of the M_{vis} distribution has the potential to improve the measurement of the TES in future.



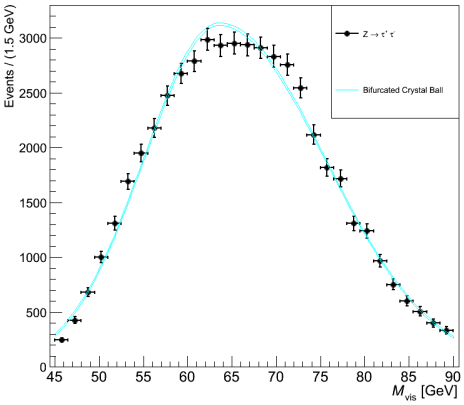
(a) Bifurcated Gaussian



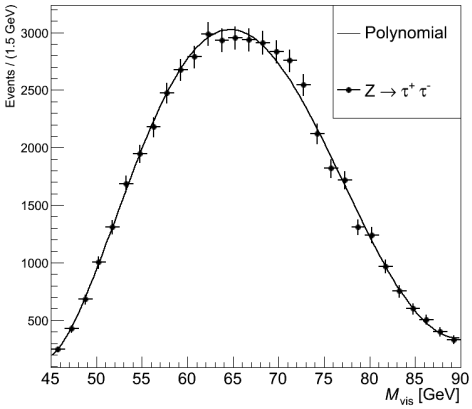
(b) Gaussian ⊗ Breit-Wigner



(c) Crystal Ball

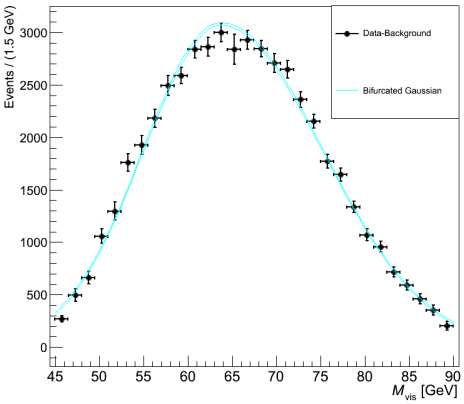


(d) Bifurcated Crystal Ball

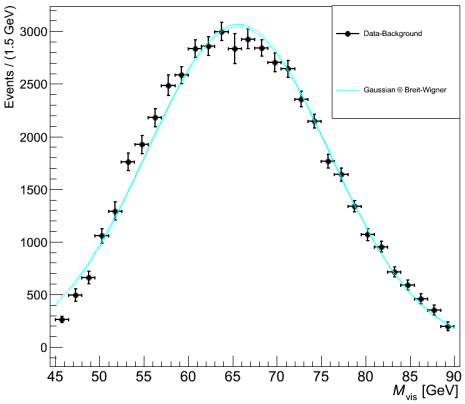


(e) Polynomial (5th order)

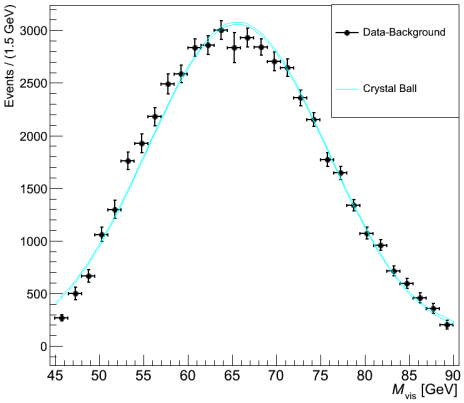
Figure 3.5: Fits to the M_{vis} distribution in $Z \to \tau^+\tau^-$ simulated events in the range 45 GeV $< M_{vis} <$ 90 GeV with the parametric functions: Bifurcated Gaussian (a), Gaussian $\otimes$ Breit-Wigner (b), Crystal Ball (c), Bifurcated Crystal Ball (d), Polynomial (5th order) (e).



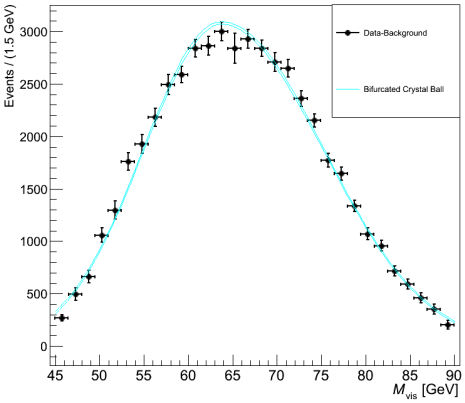
(a) Bifurcated Gaussian



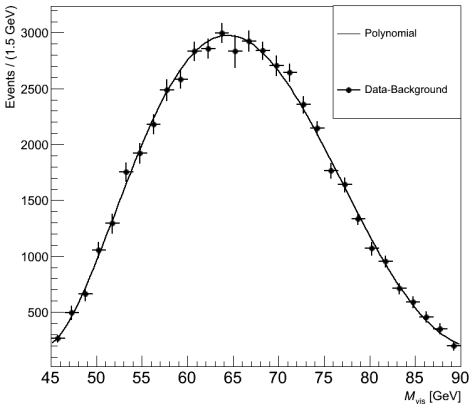
(b) Gaussian $\otimes$ Breit-Wigner



(c) Crystal Ball

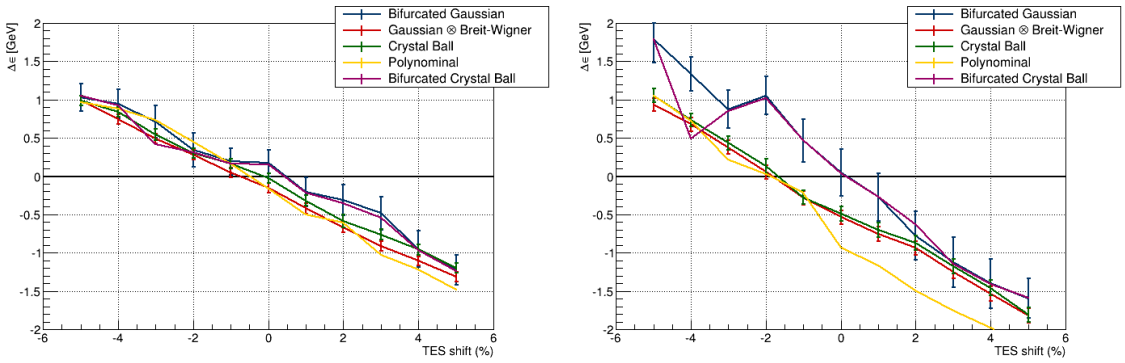


(d) Bifurcated Crystal Ball



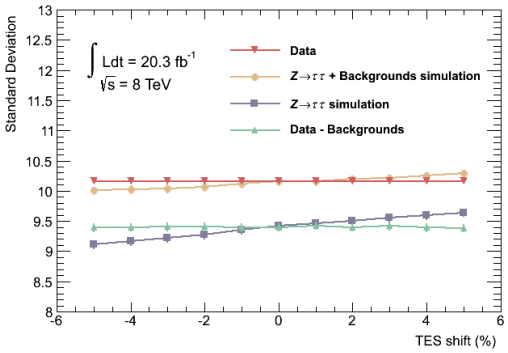
(e) Polynomial (5th order)

Figure 3.6: Fits to the M_{vis} distribution in data events after background subtraction in the range 45 GeV $< M_{vis} <$ 90 GeV with the parametric functions: Bifurcated Gaussian (a), Gaussian $\otimes$ Breit-Wigner (b), Crystal Ball (c), Bifurcated Crystal Ball (d), Polynomial (5th order) (e).

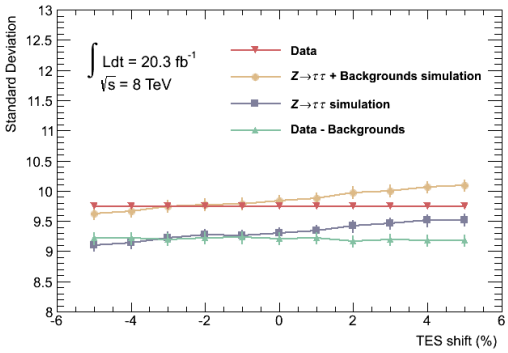


(a) 1-track τ_{had} (b) 3-tracks τ_{had}

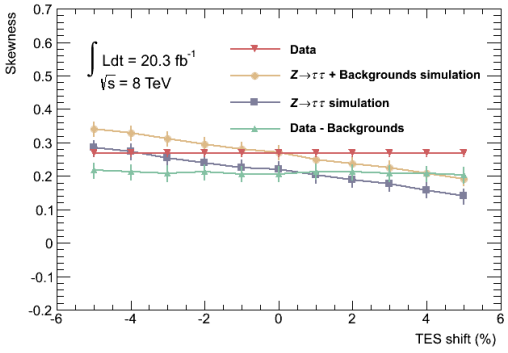
Figure 3.7: The value of $\Delta\epsilon$ as a function of the TES variation α for 1-track (a) and 3-tracks (b) τ_{had}. The estimator considered are the maximum positions of the parametrized functions: Bifurcated Gaussian, Bifurcated Crystal Ball, Gaussian $\otimes$ Breit-Wigner, Crystal Ball and Polynomial (5th order).



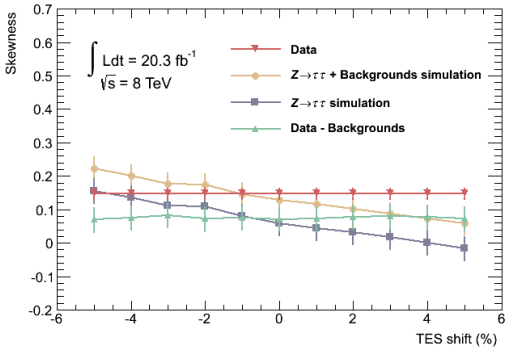
(a) 1-track τ_{had}



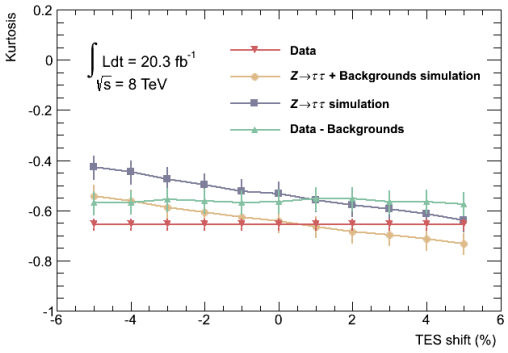
(b) 3-tracks τ_{had}



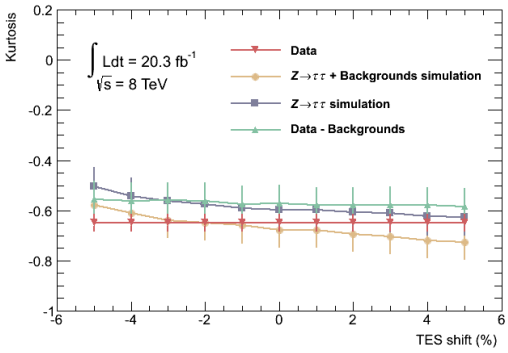
(c) 1-track τ_{had}



(d) 3-tracks τ_{had}



(e) 1-track τ_{had}



(f) 3-tracks τ_{had}

Figure 3.8: The standard deviation (top), the skewness (centre) and the kurtosis (bottom) of the M_{vis} distribution in data, Standard Model predictions, data with subtracted background and simulated $Z \to \tau^+\tau^-$ as a function of α for 1-track τ_{had} (left) and 3-tracks τ_{had} (right).

3.5 Statistical and systematic uncertainties

The uncertainties that affect the measurement of α_0 can be classified in: statistical, experimental, theoretical, $OS - SS$ method and procedural uncertainties.

Statistical uncertainties. The uncertainty due to the statistics of data and of the simulations is evaluated by generating ensembles of pseudo-experiments. Pseudo-distributions of the visible mass are obtained from the observed M_{vis} distribution and the predicted distributions of both $Z \to \tau^+\tau^-$ and background processes.

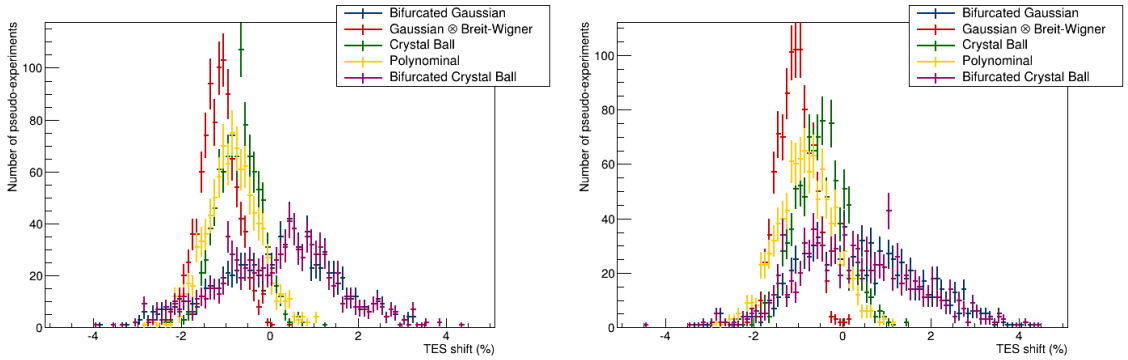
The number of pseudo-data events is randomly picked from a Poisson distribution with mean equal to the number of observed data events. The value of the M_{vis} for each pseudo-event is generated randomly using as probability density function the data M_{vis} distribution.

Since simulated events are scaled with cross-section and event-by-event correction factors, the expected number of events is not distributed according to the Poisson probability. The generation of the pseudo-distributions for $Z \to \tau^+\tau^-$ and background processes consists in casting the content of each bin of the M_{vis} distribution from a Gaussian distribution with mean equal to expected background in that bin, and standard deviation equal to the expected statistical error. This is calculated as the sum in quadrature of the scaled events in that bin.

The generation of pseudo-distributions is done separately for data, $Z \to \tau^+\tau^-$ and the backgrounds (including also the Same sign background). Pseudo-experiments have been generated only for $\alpha = 0$ and it is assumed that the statistical uncertainty does not change significantly with the choice of α. This is a reasonable assumption since the statistics of expected events is weakly affected by α.

The position of maximum of the pseudo-distribution is obtained from the fit with the parametric functions as described in Section 3.4. The pseudo-background is subtracted from the pseudo-data before the fit. α_0 is calculated by replacing data and predictions with pseudo-data or pseudo-predictions, respectively. Figure 3.9 shows the distribution of α_0 for 1000 pseudo-data measurements and 1000 pseudo-prediction measurements. For the Polynomial (5th order), Crystal Ball and Gaussian $\otimes$ Breit-Wigner fits, the distributions obtained are narrow and symmetric around the maximum. The values obtained with bifurcated functions fluctuate more. The standard deviations of the α_0 distributions for pseudo-data and pseudo-predictions are taken as statistical uncertainties. The sum in quadrature of the two uncertainties gives the statistical error on α_0. For the Polynomial (5th order) fit this uncertainty is 1.3% and 1.4% for events with for 1-track and 3-tracks τ_{had}, respectively.

Experimental uncertainties. These uncertainties are related to accuracy in the modeling of the collisional conditions, the event acquisition and the particle detection in sim-



(a) Data-background (b) $Z \to \tau^+\tau^-$ simulation

Figure 3.9: Distribution of the α_0 measured from 1000 pseudo-experiments. The measurements in pseudo-data (left) and pseudo-predictions (right) are investigated separately.

ulations. The uncertainties associated to the lepton trigger and reconstruction have been described in Section 2.4.5. An uncertainty on the efficiency of the muon isolation requirement of 2% is also applied according to dedicated studies [27]. The E_{T}^{miss} is used in the selection of signal region events. Although the uncertainty on the E_{T}^{miss} does not influence directly the M_{vis} calculation, the effect on the selection efficiency and consequently on the TES measurement is not negligible. The E_{T}^{miss} scale and resolution uncertainties are considered. These uncertainties are rather large because are obtained combining the error associated to the large number of detector inputs that are used in the E_{T}^{miss} calculation.

$OS - SS$ method uncertainties. These are the uncertainties associated with the method of background estimate. The uncertainties on the scale factors measured to correct the backgrounds predictions in the $OS - SS$ method are given in Table 3.3. The TES corrections are recalculated by scaling the predictions with scale factors varied up or down within their uncertainty.

Theoretical uncertainties. These uncertainties are associated to the modeling of the simulated processes. Since the normalizations of the main simulated backgrounds have been checked in dedicated control regions and scale factors have been applied when needed (see Section 3.3.1.1), the production cross-section uncertainties for simulated processes are not considered. The modeling the parton showering and the underlying event is however investigated for $Z \to \tau^+\tau^-$, $Z \to \mu^+\mu^-$ and $W^{\pm}$ + jets events. Samples of these processes have been simulated with an alternative generator (ALPGEN+HERWIG rather than ALPGEN+PYTHIA6). The value of α_0 measured with the alternative simulation is compatible with the original one within its statistical uncertainty.

An additional test has been performed using τ-embedded $Z \to \mu^+\mu^-$ data [151] in place of the $Z \to \tau^+\tau^-$ simulation. The embedded sample consists of $Z \to \mu^+\mu^-$ data

events where the muons are replaced by simulated τ leptons with same momentum. The advantage of the embedded sample is that there are no uncertainties on the production mechanism of the Z, because it is taken from data. However, systematic uncertainties on the replacement of muons with τ leptons affect significantly the TES measurement making the overall uncertainty bigger. The results obtained with the simulation and the embedded sample are compatible confirming that uncertainties associated with the process generation have small impact on the TES measurement.

No theoretical uncertainty is therefore significant for this measurement.

Procedural uncertainties. The uncertainties on the procedure used to extract the TES corrections are evaluated by studying:

- differences between the average α_0 in pseudo-experiments and the one which is measured in data;
- biases from the definition of the M_{vis} interval considered assessed by changing the range of the fit to 50 GeV $< M_{vis} <$ 85 GeV and 40 GeV $< M_{vis} <$ 95 GeV;
- the value of α_0 obtained without removing the expected background from data and fitting the whole total expected and observed M_{vis} distribution;
- biases in the linearity dependence between $\Delta\epsilon$ and α exploited by calculating α_0 via linear interpolation rather then a linear fit.

Summary of the systematic uncertainties The contributions to the uncertainty on the α_0 measurement due to the sources of systematic uncertainties considered is shown in Table 3.5 for events with 1-track and 3-tracks τ_{had}. The largest contribution is due to the procedural uncertainties: 0.5% in both cases. Muon resolution and E_{T}^{miss} resolution are also significant, especially for 3-tracks τ_{had}. The largest of the $OS - SS$ method uncertainties is due to r_{QCD} since this influences the normalization of not only the Same sign background, but also the other backgrounds. The total systematic uncertainty is 0.6% on the TES correction for 1-track τ_{had} and 0.7% on that for 3-tracks τ_{had}.

Source	Uncertainty (%)	
	1-track	3-tracks
Trigger efficiency	0.12	0.06
Muon reconstruction efficiency	<0.01	<0.01
Muon isolation efficiency	0.01	<0.01
Muon momentum resolution	0.11	0.32
E_{T}^{miss} resolution	0.19	0.25
E_{T}^{miss} calibration	0.07	0.1
r_{QCD}	0.28	0.09
$k_{OS}^{W^{\pm}+\mathrm{jets}}$	0.01	0.04
$k_{SS}^{W^{\pm}+\mathrm{jets}}$	0.02	0.06
$k^{Z\to\mu^+\mu^-+\mathrm{jets}(\to\tau_{\mathrm{had}})}$	<0.01	0.01
k_{OS}^{top}	0.03	0.06
k_{SS}^{top}	0.01	<0.01
Fit procedure	0.5	0.5
Total	0.6	0.7

Table 3.5: The contributions to the systematic uncertainty on the TES correction measured with the the Polynomial (5th order) estimator for 1-track and 3-tracks τ_{had}.

3.6 Results

The method described in this chapter measures the differences of the TES between data and simulation. This difference α_0 has been estimated with the procedure explained in Section 3.4 by means of Polynomial (5th order) fits to the M_{vis} distribution. Corrections for the calibration of τ_{had} in data are obtained. These corrections are calculated as $(-\alpha_0/(1+\alpha_0))$ and are given in Table 3.6 for 1-track and 3-tracks τ_{had}. This calibration is combined

τ_{had} selection	TES correction
1-track	0.8% ± 1.3% (stat.) ± 0.6% (syst.)
3-tracks	1.1% ± 1.4% (stat.) ± 0.7% (syst.)

Table 3.6: Corrections for the TES of 1-track and 3-tracks τ_{had} reconstructed in data.

with the simulation-based one to form the official calibration applied to the reconstructed τ_{had} in data events collected by the ATLAS at $\sqrt{s} = 8$ TeV [1]. The calibration has been applied to the data where the $H \to \tau^+\tau^-$ decay was observed [27] and is used in the $t\bar{t}H$ analysis in the final state with leptons and τ_{had}.

Additional studies of TES have been performed to evaluate differences in data and simulation due to detector effects and collisional conditions. TES corrections are measured considering τ_{had} reconstructed in the *barrel* region ($|\eta| < 0.8$) and in the *end-cap* regions ($0.8 < |\eta| < 2.5$), or comparing TES corrections for events with few or many pileup collisions. The corrections obtained are compatible within uncertainties, therefore no

further corrections are applied to the reconstructed τ_{had} in data. The TES corrections measured as a function of $|\eta|$ can be found in Section A.3.

Chapter 4

Analysis of events with same-sign lepton pairs

Chapter 1 has introduced the Standard Model and has given motivations for studies of $t\bar{t}H$ production and searches for processes beyond Standard Model that can produce pairs of same-sign leptons. These analyses are illustrated in the following chapters of this thesis: searches for new physics in Chapter 5 and measurement of $t\bar{t}H$ production in Chapter 6. Both studies focus on the analysis of proton-proton collisions data at $\sqrt{s} = 8$ TeV where same-sign lepton pairs are produced in the final state.

Events with less than two leptons or with opposite-sign leptons are produced in large quantity by QCD multi-jet, $W^{\pm}+\text{jets}$, Z/γ^* and $t\bar{t}$ processes. Excluding for the moment leptons that are produced in hadron decays, same-sign leptons are not produced by any of these processes. Only rare Standard Model processes can originate same-sign lepton, like $W^{\pm}Z$ and ZZ production. The Standard Model background in events with this signature is therefore expected to be small making these data ideal to measure the rare $t\bar{t}H$ production for the first time and to search for new physics.

The two analyses are generally different. The main difference is that, contrary to the new physics search, the $t\bar{t}H$ study requires a large number of jets in the final state in addition to the same-sign leptons. This makes the background compositions different: processes with jets are enhanced in the $t\bar{t}H$ analysis compared to the new physics search. The techniques used to estimate backgrounds are specific to each analysis and are described in the correspondent chapters. Despite these differences, in both analyses the backgrounds are classified in the same categories depending on the type and origin of the reconstructed leptons. This chapter describes these categories and the processes contributing to each of them.

Every Standard Model process that can lead to the reconstruction of same-sign leptons is a background for the analyses. These background processes do not necessarily have same-sign leptons produced in the proton-proton collision, but rather mis-reconstructed leptons. Backgrounds can have for instance particles that mimic the interaction of lepton

with the detectors and are reconstructed as such. Backgrounds can also have leptons with electric charges that are not correctly reconstructed or leptons produced in the interaction with the detector. The backgrounds for these analyses are divided into four categories: prompt, charge misID, photon conversion and non-prompt backgrounds. These categories are defined in the following sections.

4.1 Sources of prompt background

A prompt lepton is a lepton produced at the interaction point as direct result of the proton-proton collision or in the decay of vector bosons $W^{\pm}$ and Z or the Higgs boson. Leptons produced in the decay of τ leptons are also considered prompt as well as leptons from the new physics particles described in Chapter 1. Leptons from hadron decays are considered non-prompt due to the relatively large hadron lifetime. A reconstructed lepton is classified as prompt if its signal has been generated by a prompt lepton and its charge is correctly reconstructed. Prompt backgrounds have two of these leptons with same charge.

The Standard Model processes that constitute prompt backgrounds with largest cross-section are the diboson production $W^{\pm}Z \rightarrow \ell^{\pm}\nu\ell^{\pm}\ell^{\mp}$ and $ZZ \rightarrow \ell^{\pm}\ell^{\mp}\ell^{\pm}\ell^{\mp}$ and the production of same-sign $W^{\pm}$ ($W^{\pm}W^{\pm}$+2 jets$\rightarrow \ell^{\pm}\ell^{\pm}\nu\nu$+2 jets). Prompt backgrounds with large jet multiplicity are due to $t\bar{t}$ production in association with a W or Z boson ($t\bar{t}W^{\pm}$ and $t\bar{t}Z$ respectively), the dominant prompt backgrounds in the $t\bar{t}H$ analysis. Diboson production is instead the dominant prompt background in the new physics search. The estimate of the prompt background is obtained from simulations and described in the analysis chapters.

4.2 Sources of charge misID background

The charge misID background is composed by processes producing a pair of prompt opposite-sign leptons that are reconstructed as a same-sign lepton pair. This can occur when the charge of one of the prompt leptons is wrongly reconstructed or when the lepton interacts with the detector producing an additional lepton with opposite-sign that is subsequently reconstructed. In both cases the reconstructed lepton is classified as charge misID lepton. Charge misID backgrounds have one prompt and one charge misID lepton. The main processes producing opposite-sign leptons and consequently constituting a source of charge misID background are Z/γ^*, W^+W^- and $t\bar{t}$. For high jet multiplicity, $t\bar{t}$ is the dominant charge misID background. Since the mis-identification of the lepton charge has to do with the interaction of the lepton with the detector, the charge misID backgrounds are different for muon and electrons and they are studied separately. The causes of charge misID for electrons and muons are investigated in Section 4.2.1 and Section 4.2.2, respectively. Given the large cross-section for processes producing opposite-sign leptons, it is vital for these analyses that the probability for a prompt lepton to be reconstructed as charge misID lepton is very small. As described in the following, the background with

charge misID muons is found to be negligibly small for both analyses, while the backgrounds with charge misID electrons have to be accounted for. The methods used to determine these backgrounds are described in the analysis chapters.

4.2.1 Electron charge misID background

The charge of electrons is reconstructed from the bending direction of the track associated with the energy deposit in the electromagnetic calorimeters that seeds the electron reconstruction. There are three ways in which an electron can be reconstructed with charge opposite to its real one:

- The reconstructed electron is not the one produced in the primary interaction. The prompt electron interacts with the detector material producing a hard photon via Bremsstrahlung; the photon subsequently interacts with the detector forming a pair of opposite-sign electrons ($e^{\pm} \rightarrow e^{\pm}\gamma^{*} \rightarrow e^{\pm}e^{+}e^{-}$). This process is called a trident event and a representation of it is given in Figure 4.1. The electron in the triplet with opposite-sign if reconstructed becomes the charge misID electron. The other electrons may or may not be reconstructed, they can be superimposed to the charge misID electron or can lay inside the isolation cone around the charge misID electron. The probability to have electrons forming trident events is proportional to the amount of material traversed by the electrons. The larger $|\eta|$, the ticker the material in the inner detector that the electron has to traverse before reaching the calorimeter. Therefore, the production rate of charge misID electron from trident events has a strong dependence on $|\eta|$.

- The bending direction of the reconstructed track is wrongly measured. This occurs when the electron p_{T} is very large leading to a track with small curvature. In this case the resolution on the hit points and the alignment of the IDet modules play an important role. The algorithm of reconstruction could find that a track with bending direction opposite to the real one fits the hits best. The probability of this type of charge misID electrons to occur has a dependence on p_{T} and on the local alignment.

- The electron energy deposit in the calorimeter is associated to the wrong reconstructed track. Additional charged particle can be produced in any of the collisions in the event. The more produced particles, the higher the probability to accidentally associate one of them to an electromagnetic energy cluster. These particles typically have low p_{T} and are produced more frequently at large $|\eta|$.

The dominant source of charge misID electron arises from electrons interacting with the material and forming trident events. Figure 4.2 shows the probability to reconstruct a charge misID electron as a function of the electron p_{T} and the probability that this has occurred via trident event for simulated $Z/\gamma^{*} \rightarrow e^{+}e^{-}$ events. The requirements

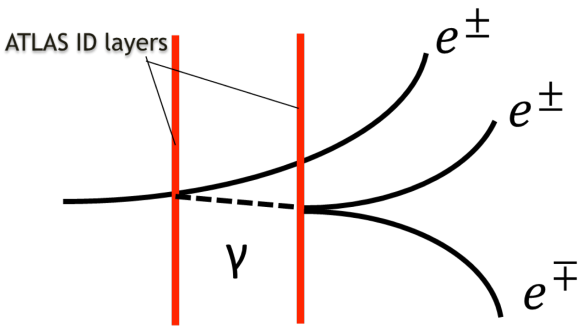


Figure 4.1: A visualisation of the production of a trident event. The electron interacts with the material of the inner detector and radiates a photon via Bremsstrahlung; the Bremsstrahlung photon interacts again with the detector and creates an e^+e^- pair. One of the leptons in the pair has charge opposite to the initially interacting electron. The initial electron and the photon may or not interact with the material of the same IDet module.

used to select electrons are the same used in the new physics search. The probability is obtained from the generator information. For p_{T} below few hundred GeV, about 80%

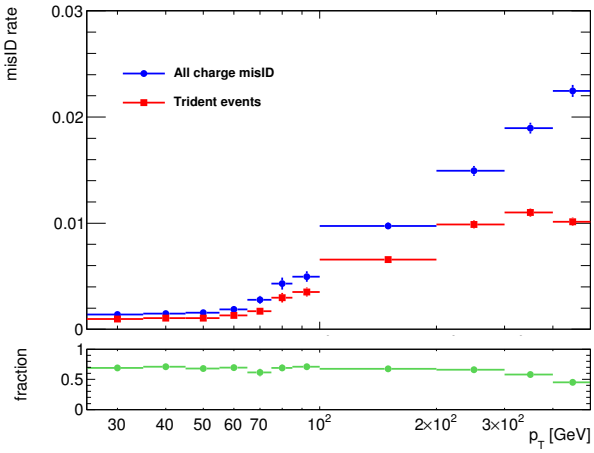


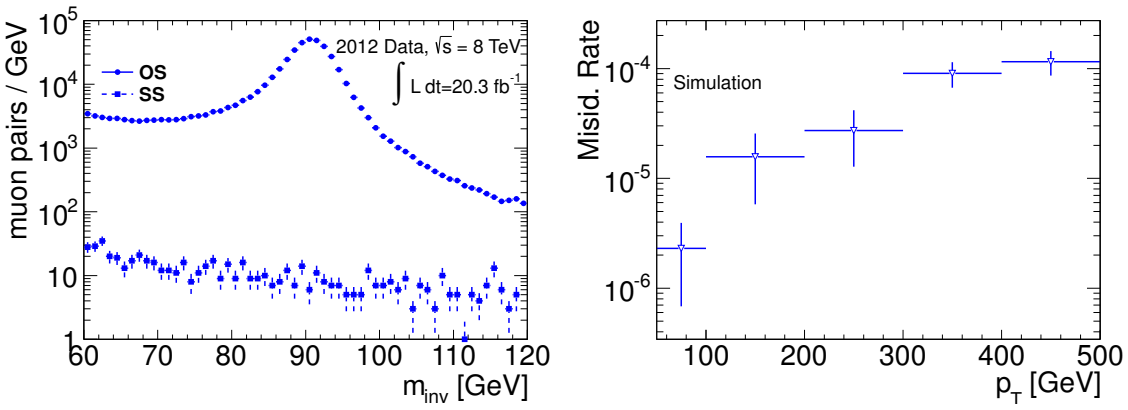
Figure 4.2: Probability to reconstruct an electron with wrong charge as a function of p_{T} in simulated Z/γ^* events (blue). The charge misID rate due to trident events is also shown (red). The bottom panel shows the fraction of charge misID electrons due to trident events with respect to all processes of charge mis-measurement.

of the charge misID electrons arises from trident events; for $p_{\mathrm{T}} > 400$ GeV the track curvature becomes so small that wrong measurements of the bending direction become the main source of charge misID electrons. The expected total charge misID rate is large enough to make the background from charge misID not negligible. The determination of the background events with charge misID electrons is described in the analysis chapters.

4.2.2 Muon charge misID background

The muon reconstruction is based on the statistical combination of a track in the IDet with a track in the MS. The charge of the muon is determined from the bending direction of the two tracks. The direction of the magnetic fields in the toroids and in the solenoid is orthogonal offering complementary measurements of the curvature, and therefore of the p_{T} and the charge. If the charge from the IDet and the MS measurements are opposite, the charge of the track which has the higher weight in the statistical combination is used. Because of the long lever arm of the MS, at high p_{T} (> 100 GeV) the MS measurement dominates and its charge is used. A charge misID muon is a reconstructed muon whose track with higher weight has been measured with wrong bending direction. This is expected to occur with very small probability since the best of two independent charge measurements is taken. The muon charge is more likely to be misidentified at very large momentum because the muon trajectory would bend very little and the hit resolution and the alignment precision would substantially affect the bending direction measurement. Bremsstrahlung occurs for muons at very high energies ($O(\mathrm{TeV})$) and pair production of $\mu^+\mu^-$ is largely suppressed compared to e^+e^-. Therefore muon trident events are expected to give a negligible contribution in the two analyses. Track pairing errors are also negligible due to the request of compatibility of the momentum measurements in the IDet and MS.

The rate of charge misID muons has been investigated in di-muon events in data at low $m(\mu,\mu)$ and in $Z/\gamma^* \to \mu^+\mu^-$ simulation at large $m(\mu,\mu)$. The muon pairs are selected as described in the new physics search chapter with the exception that one of the muons must have $10 < p_{\mathrm{T}} < 20$ GeV to be orthogonal to the signal regions of the analysis. The invariant mass of opposite-sign and same-sign di-muon pair with $60 < m(\mu,\mu) < 120$ GeV is shown in Figure 4.3(a). If the probability of reconstructing charge misID muons is significantly large, the same-sign di-muon invariant mass distribution would show a peak around the Z mass (≈ 90 GeV) since many opposite-sign leptons are produced in that interval. This is not observed meaning that the muon charge misID background is negligible for the muon momentum range considered in the test. The pair considered have mainly muons with $p_{\mathrm{T}} < 100$ GeV. At larger p_{T}, the muon charge misID rate is investigated using off-shell $Z/\gamma^* \to \mu^+\mu^-$ simulations. The charge of the generated muon from $Z/\gamma^* \to \mu^+\mu^-$ is compared to the reconstructed one for $p_{\mathrm{T}} > 100$ GeV and the probability of wrong charge measurement as a function of p_{T} is obtained. The charge misID rate for muons is shown in Figure 4.3(b) as a function of the muon p_{T}. The muon charge misID probability at large p_{T} is still relatively small (about 10^{-4}). The electron charge misID rate is hundreds of times larger for the same p_{T}. Similar tests have been performed with the muons selected for the $t\bar{t}H$ analysis obtaining similar results. The muon charge misID background is therefore considered negligible in both analyses.



(a) (b)

Figure 4.3: (a): Invariant mass distributions of opposite-sign (OS) and same-sign (SS) muon pairs in data events with $60 < m(\mu,\mu) < 120$ GeV. (b): The muon charge misID probability as a function of p_{T} from $Z/\gamma^* \to \mu^+\mu^-$ simulations.

4.3 Sources of photon conversion background

The photon conversion background is due to processes producing at least one prompt lepton and a photon with the photon being reconstructed as an electron with the same charge of the prompt lepton. Photons can convert in a e^+e^- pair by interaction with the detector. The selection applied to electrons in the analysis signal regions excludes electrons that form a pair compatible with a photon conversion. However, if only one of the two electrons from photon conversion is reconstructed, it can pass the selection and be paired with another lepton in the event to form a same-sign pair. Photons can be reconstructed as electrons also when the energy deposited by the photon in the electromagnetic calorimeter is paired with an unrelated track in the IDet. The $\mu^+\mu^-$ pair production is negligible compared to e^+e^- and electromagnetic calorimeter energy deposits only marginally affect the muon reconstruction algorithm. Therefore the photon conversion background affects only analyses with reconstructed electrons. The main processes contributing to the photon conversion background are $W^{\pm}\gamma$ and $Z\gamma$. The estimate of these backgrounds are described in the analyses chapters.

4.4 Sources of non-prompt background

Non-prompt leptons are hadrons erroneously reconstructed as leptons or decay product of hadrons reconstructed as leptons. Non-prompt leptons are often due to true leptons produced in the decay of heavy flavour hadrons (containing b or c-quarks). The Dalitz decay of π^0 can also produce true electrons. Charged pions and kaons stopped in the electromagnetic calorimeters can sometimes be reconstructed as electrons too. Although these particles are not electrons and can have prompt origin, they are considered non-

prompt electrons. Very high energy hadrons can pierce the calorimeters and make tracks in the MS that are paired by the muon reconstruction algorithm to IDet tracks to form muons. Leptons created by the interaction of hadrons or of the hadron decay products with the detector are also considered non-prompt. In the processes considered above, the reconstructed lepton is always the result of a sequence of events initiated by a quark or gluon. As such these leptons are grouped in the same category and are referred to globally as non-prompt leptons.

The non-prompt background includes all the processes where hadronic jets are generated and at least one produces a non-prompt lepton. In di-lepton analyses the fake background can have one prompt and one non-prompt lepton or two non-prompt leptons. Sources can be $W^{\pm}$+jets, QCD multi-jet, Z/γ^{*}+jets, $t\bar{t}$, $tW^{\pm}$ and *single-top* productions. $W^{\pm}$ + jets and QCD multi-jet processes are the main backgrounds in events with low jet and b-jet multiplicity; $t\bar{t}$ is dominant among the non-prompt backgrounds with b-jets and high number of jets.

The charge of non-prompt leptons is barely related to the charge of the parton that originated the jet. Therefore, there is no correlation between the charge of a non-prompt lepton and any other lepton in the event meaning that the non-prompt background are expected to produce reconstructed same-sign leptons as frequently as opposite-sign leptons. Given that the cross-section for the processes responsible for the non-prompt backgrounds are very large at hadron colliders, the probability to select non-prompt leptons must be very small in order to not be dominated by the non-prompt background.

Figure 4.4 shows a representation of a non-prompt lepton produced in a jet via hadron decay. Non-prompt leptons are generally surrounded by the other particles composing the jet. If the non-prompt lepton is produced in a secondary decay, its impact parameter can be significantly different from zero. Prompt leptons have tracks and calorimeter energy deposits with distinctive properties that can hardly be mimicked by other particles. The rejection of non-prompt leptons is based on requirements on the isolation, the impact parameter and the quality of the detector signal associated with the leptons. The criteria used in the $t\bar{t}H$ and new physics analyses are described in the relative chapters. The estimate of the non-prompt background is also analysis specific. However, both analyses use data-driven methods for the estimate of this background, because non-prompt lepton production and reconstruction is not easy to simulate. The modeling of the parton showering and the hadronization processes must be accurate as well as the decay modes. The way in which a hadron can mimic the interaction of a lepton with the detector must also be accurate. The simulated statistics must be also very large since the selection probability of non-prompt leptons is small and the cross-section of non-prompt sources are very large. Data-driven methods are therefore preferred.

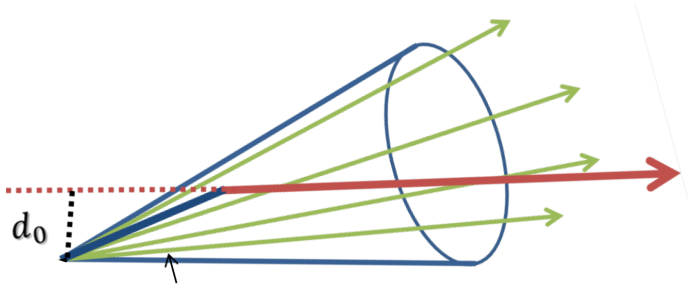


Figure 4.4: A visualisation of the production of a non-prompt lepton from a hadron decay in a jet. The blue line is the trajectory of the hadron that decays into a lepton (red arrow) and additional particles. The green arrows represent the hadrons forming the jet. The black dashed line shows the non-zero impact parameter of the lepton trajectory due to its production away from the primary vertex.

Chapter 5

New physics in same-sign leptons

The analysis of same-sign lepton pairs produced in proton-proton collisions constitutes an opportunity for discovery of new physics in a low background environment. The production of same-sign lepton pairs in the Standard Model is limited to rare processes, where multiple vector bosons and top quarks are typically produced. Additional same-sign lepton pairs can arise from non-prompt, charge misID and photon conversion backgrounds, as described in Chapter 4. These processes are also rare because of the ability of the ATLAS detector to discriminate them from prompt leptons, as described in Section 2.4.

Several theoretical scenarios, including Supersymmetry [44], universal extra dimensions [45], Left-right symmetric models [46–49], the little Higgs model [50], fourth generation quarks [52–54] and vector-like quarks [51], predict production of same-sign lepton pairs beyond the Standard Model. Non-Standard Model production of same-sign lepton pairs is also expected for models that provide an explanation for the neutrino mass generation, like see-saw models [63–70] and radiative models [71–74].

Given the variety of models potentially producing same-sign leptons, a wide range of final state topologies are possible. For instance, the same-sign lepton pair production may or may not be resonant, and additional particles, like other leptons, photons, jets and invisible particles, can be produced in the final state. It is also possible that Nature produces same-sign leptons in unknown ways. Because of this potential multitude of non-Standard Model same-sign production mechanisms of lepton pairs, a generic and inclusive approach to the analysis of same-sign lepton pairs is chosen.

The lepton kinematic measurement is limited only by the detector acceptance and minimal requirements on the activity in the event besides the same-sign lepton pair are imposed. Reconstructed leptons are selected with very stringent quality criteria to suppress the reducible backgrounds from non-prompt leptons, charge misID leptons and photon conversion.

The anomalous production of same-sign pairs is searched for in the invariant mass spectra of $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ reconstructed pairs. The form of the results obtained does not rely on any specific theory and allows the testing of several new physics models.

An additional study is performed on same-sign lepton pairs in search of di-lepton resonances. Two of the mechanisms of neutrino mass generation, the Type-II see-saw model [83, 152–155] and the classic Zee-Babu model [71–73], produce resonant same-sign lepton pairs from the decay of doubly charged particles. Given the good resolution of the energy of reconstructed leptons in ATLAS, the lepton pairs would form a narrow resonance in the same-sign lepton invariant mass distribution. The search for resonances in reconstructed same-sign lepton pairs has therefore tremendous potential to investigate these models.

This chapter describes the search for the production of same-sign lepton pairs due to physics beyond Standard Model. Examples of production modes of same-sign lepton pairs from new physics are given in Section 5.1. The data sample and simulations used in the analysis are described in Section 5.2, and the event selection requirements are summarized in Section 5.3. The estimation and validation of the Standard Model backgrounds are discussed in Section 5.4 and the associated uncertainties are evaluated in Section 5.5. The statistical analysis of the selected same-sign lepton pairs is presented in Section 5.6. Results for both the inclusive search for new physics and the search for doubly charged particles are also described there.

The analysis presented in this chapter is the result of collaboration between the author and six colleagues. The author has led the signal region selection optimization and the background determination and validation. He has also defined the fiducial selection, calculated the fiducial efficiency used in the inclusive search and estimated the efficiency of the detection of doubly charged particles with the ATLAS detector. The work has been published in JHEP [2].

5.1 New physics signatures

Chapter 1 has introduced two candidate extensions of the Standard Model that provide a mechanism for the neutrino mass generation. The see-saw mechanisms [63–70] require new fields: neutral fermion $SU(2)_L$ singlets (N_{R}) for Type-I, a new scalar $SU(2)_L$ triplet (ϕ) for Type-II and new fermion $SU(2)_L$ triplets (Σ) for Type-III. In radiative models [71–74], the loop diagram that generates the neutrino mass contains two scalar $SU(2)_L$ singlets ($k^{\pm\pm}$ and $h^{\pm}$) in the classic Zee-Babu model; diquarks and leptoquarks (S_{DQ} and S_{LQ}) in the Coloured Zee-Babu model.

The production of the particles predicted by these models at the LHC has been investigated [71–74, 83, 152–155], and both see-saw and Zee-Babu models lead to non-Standard Model production of same-sign lepton pairs. Diagrams of same-sign lepton pair productions via see-saw modes at the LHC are given in Figure 5.1. The heavy Majorana neutrino N_{R} of the Type-I model can be produced via a virtual $W^{\pm}$ and can decay into a $W^{\pm}$ and a lepton. Heavy Majorana neutrinos are also predicted by the Left-right symmetric model, where a heavy right-handed $W^{\pm}$ boson ($W_{\mathrm{R}}^{\pm}$) is also introduced. Therefore

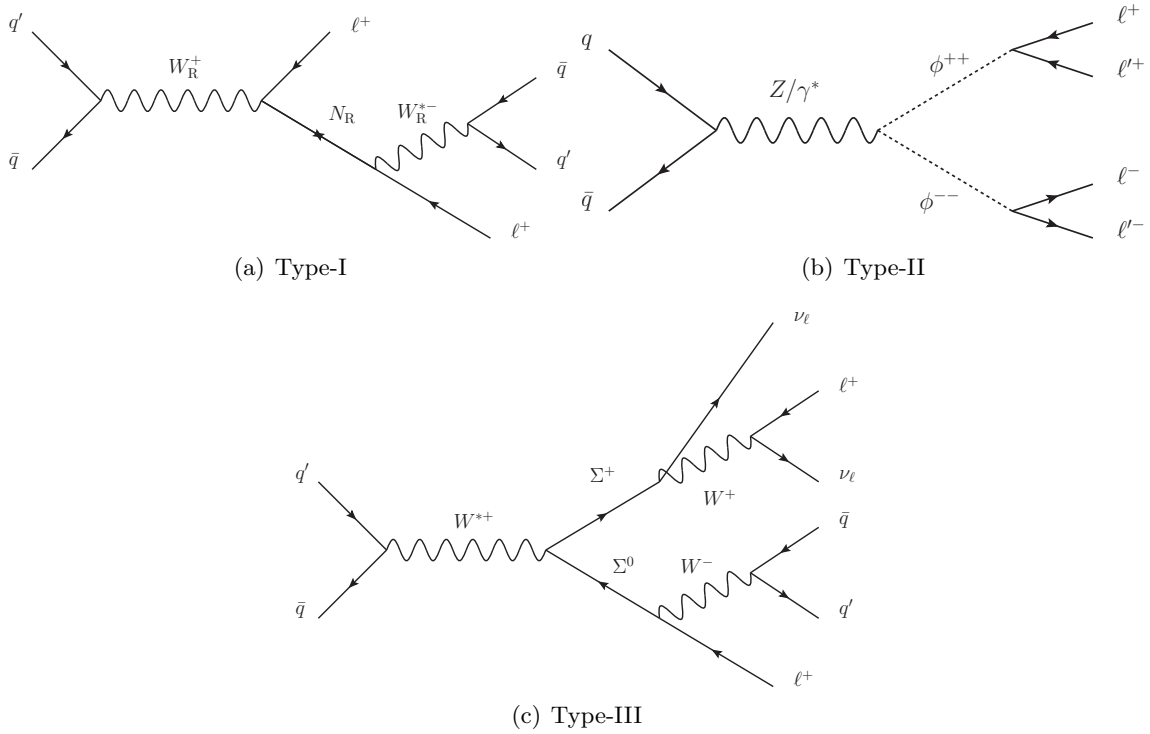


Figure 5.1: Tree level Feynman diagrams of same-sign lepton pairs productions in Type-I (a), Type-II (b) and Type-III (c) see-saw models.

in the Left-right symmetric model, the N_R can also be produced in the decay of a $W_R^{\pm}$. The production of a $W_R^{\pm}$ from quarks and its decay into a N_R and a lepton are shown in Figure 5.1(a). The N_R can decay either into a lepton and a $W^{\pm}$, or into a lepton and two quarks via a virtual $W_R^{\pm}$ exchange. The final state contains a pair of non-resonant same-sign leptons and additional jets or leptons depending on the $W^{\pm}$ decay mode.

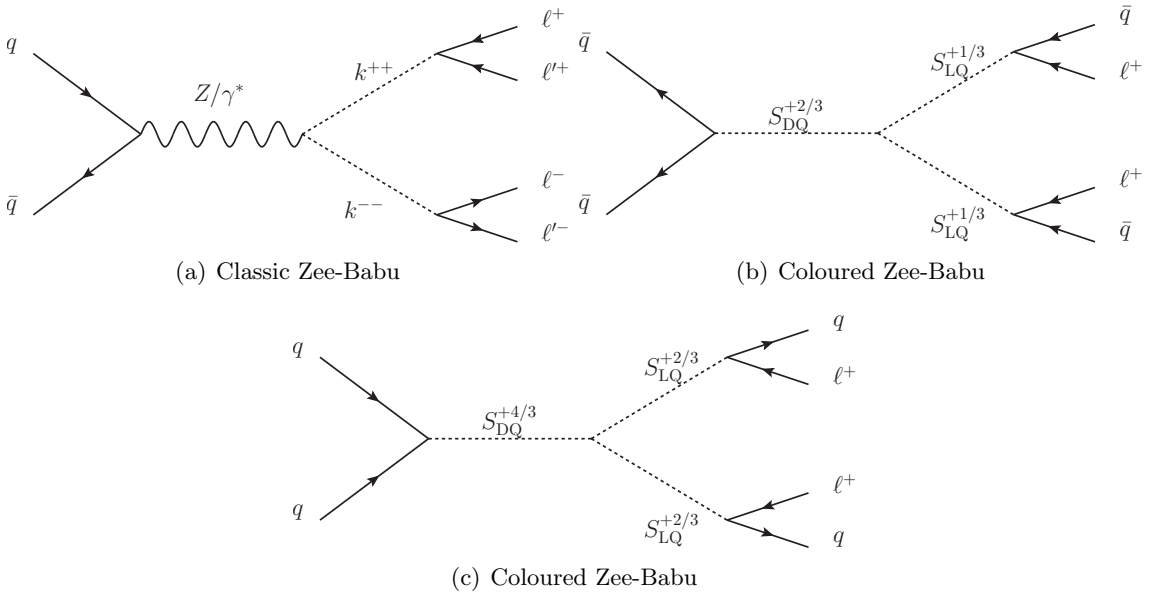
The scalar triplet ϕ in the Type-II model has a doubly charged component ($\phi^{\pm\pm}$). The decay of this particle produces same-sign lepton pairs or same-sign $W^{\pm}$ pairs with a ratio that depends on the vacuum expectation value of the triplet [156]. The latter decay mode would be observed in the same-sign lepton pairs as a non-resonant excess, and the former as a distinctive resonance. One of the main mechanisms of production of doubly charged particles at the LHC is the pair production via a Z/γ^* exchange shown in Figure 5.1(b). The final state of this process contains two pairs of resonant same-sign leptons, making it one of the cleanest signatures at the LHC. The Left-right symmetric model can generate Type-II neutrino masses via a scalar $SU(2)_L$ triplet and a scalar $SU(2)_R$ (the additional symmetry of the model) triplet. The doubly charged components of these fields are called $H_L^{\pm\pm}$ and $H_R^{\pm\pm}$, respectively. Their pair production at the LHC is analogous to the process shown in Figure 5.1(b). The Left-right symmetric model is used as a benchmark model for the production of doubly charged particles and a heavy Majorana neutrino in this analysis.

The neutral lepton Σ^0 and the charged lepton $\Sigma^{\pm}$, composing the lightest triplet of

the Type-III see-saw model, could be produced in association at the LHC via exchange of a virtual $W^{\pm}$. The decay of these particles produces a $W^{\pm}$ and either a neutrino or a lighter charged lepton. The final state of this process contains a same-sign lepton pair and jets as shown in Figure 5.1(c). The signature of this process is similar to that of the Type-I see-saw model but with additional E_{T}^{miss}.

In the classic Zee-Babu model, the neutrino mass generation occurs at two loop level and is mediated by a doubly charged particle, $k^{\pm\pm}$. The production of this particle at the LHC is via pair production similar to the production of $\phi^{\pm\pm}$ in the Type-II see-saw model (Figure 5.2(a)), and leads to the generation of two resonant same-sign lepton pairs.

Diquarks of the Coloured Zee-Babu model, S_{DQ}, with charges $\pm 2/3$ and $\pm 4/3$, are expected to be produced from quark pairs. The decay of S_{DQ} generates same-sign leptoquarks, S_{LQ}, that then decay into a charged lepton and a quark. The whole process leads to the production of a same-sign lepton pair and two quarks as shown in Figure 5.2(b,c).



(a) Classic Zee-Babu

(b) Coloured Zee-Babu

(c) Coloured Zee-Babu

Figure 5.2: Tree level Feynman diagrams of same-sign lepton pairs productions in classic Zee-Babu (a) and Coloured Zee-Babu models (b,c).

Additional modes of same-sign lepton pair production are expected from models of Supersymmetry [44], universal extra dimensions [45], vector-like quarks [51], fourth generation quarks [52–54] and others. These processes usually generate several particles in addition to the same-sign lepton pair. In these cases, a dedicated search taking into account the whole final state topology is typically more effective, however the fourth generation down-type quark b' pair production is considered anyway. The production diagrams of b' quarks and their decay to same-sign leptons is shown in Figure 5.3. The final state also contains a high and variable number of jets. This model is used as benchmark to test the validity of the analysis results for processes with busy final states.

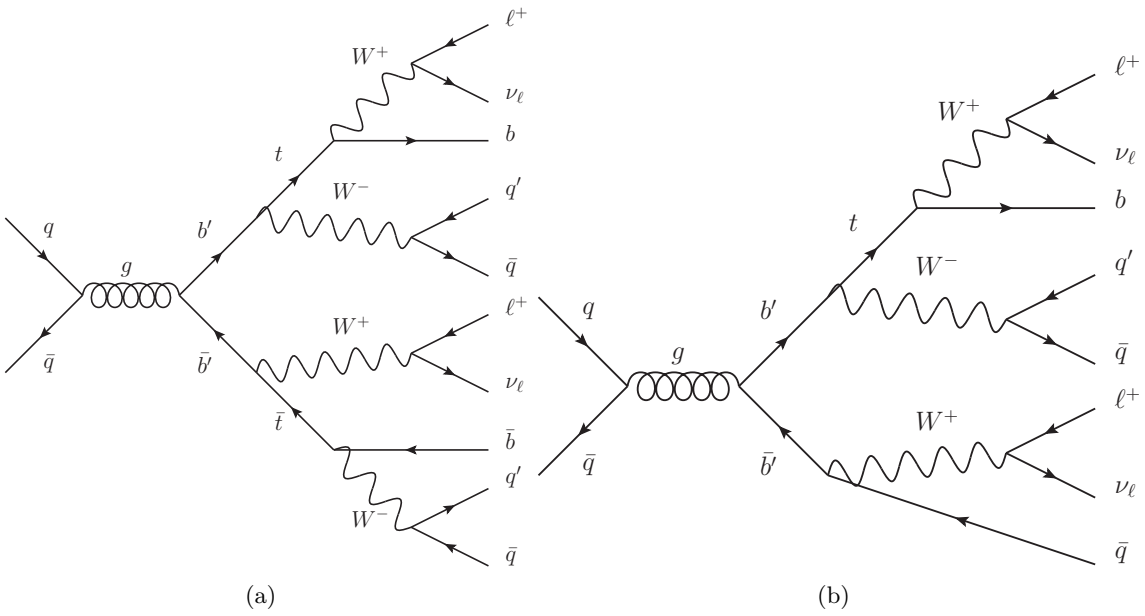


Figure 5.3: Feynman diagrams of fourth generation down-type quark pair production and decay into top quarks (a) or a top quark and a lighter quark (b).

In summary, the non-Standard Model production mechanisms of same-sign lepton pairs considered in this analysis are:

- pair production of $H_{\mathrm{L}}^{\pm\pm}$ via a virtual Z/γ^* exchange in the context of the Left-right symmetric model (two resonant same-sign pairs);
- production of a right-handed W boson (W_{R}) decaying to a charged lepton and a right-handed neutrino (N_{R}) in the context of the Left-right symmetric model. The right-handed neutrino further decays into a lepton and jets;
- production of a diquark (S_{DQ}) with charge $\pm 2/3$ or $\pm 4/3$ decaying into two same-sign leptoquarks (S_{LQ}): $pp \to S_{\mathrm{DQ}} \to S_{\mathrm{LQ}} S_{\mathrm{LQ}} \to \ell^{\pm}\ell^{\pm} qq$
- production of a fourth generation heavy $b'\bar{b}'$ pair with the b' quarks decaying into a W boson and either a top quark or an up-type quark.

These processes cover a variety of signatures, with multiple particles produced in addition to the same-sign lepton pair. This variety is needed for a model-independent interpretation of the results.

5.2 Data and simulation

The search for same-sign lepton pairs is performed on the proton-proton collisions collected by ATLAS at a centre-of-mass energy of $\sqrt{s} = 8$ TeV. Only events collected with fully operational detectors are considered. Events affected by significant noise, beam background

or cosmic ray background are also excluded. The integrated luminosity of the dataset is 20.3 fb^{-1}.

Simulated events are generated to model prompt and charge misID backgrounds that may rise from Standard Model processes. Non-Standard Model processes are also studied with simulations. The simulation of collision events is a multi-step process that uses dedicated programs to model specific aspects of the interaction. In addition to the process generation, the simulation takes into account the interaction of particles with the ATLAS detector and the resulting readout signals generated. Details of the simulation have been given in Section 2.3.

The simulated Standard Model processes are summarized in Table 5.1 for prompt backgrounds and Table 5.2 for charge misID and photon conversion backgrounds. For each process the generator and the cross-section used to scale the simulated yields are given. The complete set of simulations used in the generation of a process is specified in Table 2.1. For some processes, event filters are applied to select specific decay modes, to require a certain number of radiated partons or to apply kinematic constraints. The cross-sections provided take into account the efficiency of these event filters.

The simulated cross-sections are also corrected with K-factors to account for higher order corrections in perturbation theory. The K-factors are also given in Table 5.1 and Table 5.2.

The prompt backgrounds modeled with simulations are: diboson production ($W^{\pm}W^{\pm}$, $W^{\pm}Z$ and ZZ), $t\bar{t}W^{\pm}$, $t\bar{t}Z$ and multi-parton interaction (MPI) generating $W^{\pm}W^{\pm}$, $W^{\pm}Z$ and ZZ. In MPI, two hard scatterings occur in the same proton-proton collision, each producing either a $W^{\pm}$ or a Z boson. The simulation of the diboson production also includes the singly resonant production $W^{\pm}\gamma^{*}$ and $Z\gamma^{*}$. The $W^{\pm}W^{\pm}$ production proceeds via the t-category exchange of a gluon and results in two jets in the final state in addition to the two same-sign $W^{\pm}$ bosons. Triboson production processes ($W^{\pm}W^{+}W^{-}$, $ZW^{+}W^{-}$ and ZZZ) have also been simulated and are found to be negligible for this analysis.

The processes producing charge misID background that are estimated with simulations are Z/γ^{*}, $t\bar{t}$, $tW^{\pm}$ and $W^{+}W^{-}$ production. The Z/γ^{*} simulation includes also the production of radiated photons that contributes to the photon conversion background. $W^{\pm}\gamma$ production is also simulated.

In the estimate of the non-prompt background, simulated $W^{\pm}+\text{jets}$ samples are used to evaluate the fraction of prompt background in events with one electron. The simulations used to model this process are given in Table B.1.

The processes of new physics estimated with simulations are:

- pair production of $H_{\text{L}}^{\pm\pm}$ with $H_{\text{L}}^{\pm\pm}$ mass between 100 GeV and 1 TeV;
- production of a W_{R} boson decaying to a charged lepton and a right-handed neutrino (N_{R}). The mass ranges considered are 1 TeV $\leq m(W_{\text{R}}) \leq$ 2 TeV and 250 GeV $\leq m(N_{\text{R}}) \leq$ 1.5 TeV;

- production of a fourth generation heavy $b'\bar{b}'$ pair with mass between 400 GeV and 1 TeV;
- diquark (S_{DQ}) production and decay into two same-sign leptoquarks (S_{LQ}). The mass ranges considered are 2.5 TeV $\leq m(S_{\mathrm{DQ}}) \leq$ 3.5 TeV and 1 TeV $\leq m(S_{\mathrm{LQ}}) \leq$ 1.4 TeV.

The full list of simulated signal processes and their generators is given in Table B.2.

All the simulations are corrected for the pileup conditions as observed in data, and have data-driven corrections applied for particle detection efficiencies, energy calibration and energy resolution, as specified in Section 2.4.

Process	Generator	$\sigma \times \varepsilon_{filter}$ [pb]	K-factor
$W^{\pm}Z \to \ell\ell\ell\nu$	Sherpa	9.75	1.06
$ZZ \to \ell\ell\ell\ell$	Sherpa	8.74	1.11
$W^{\pm}W^{\pm} + 2p$	MadGraph+Pythia8	0.344	1.0
$t\bar{t}W^{\pm}$	MadGraph+Pythia6	0.104	1.17
$t\bar{t}W^{\pm} + p$	MadGraph+Pythia6	0.053	1.17
$t\bar{t}W^{\pm} + \geq 2p$	MadGraph+Pythia6	0.041	1.17
$t\bar{t}Z$	MadGraph+Pythia6	0.068	1.35
$t\bar{t}Z + p$	MadGraph+Pythia6	0.045	1.35
$t\bar{t}Z + \geq 2p$	MadGraph+Pythia6	0.04	1.35
MPI $W^{+}W^{-}$	Pythia8	0.012	1.0
MPI $W^{\pm}Z$	Pythia8	0.0075	1.0
MPI ZZ	Pythia8	0.012	1.0
$W^{\pm}W^{+}W^{-}$	MadGraph+Pythia6	0.005	1.0
$ZW^{+}W^{-}$	MadGraph+Pythia6	0.0016	1.0
ZZZ	MadGraph+Pythia6	0.0003	1.0

Table 5.1: Simulated processes used for the prompt background modeling. For each process, the generator configuration, the cross-section times the generator filter efficiency and the K-factor used to scale the simulated cross-section value to the latest theoretical calculation are given. The full simulation setup is specified in Table 2.1. $X + np$ with n integer and X a generic process indicates the number of partons (quarks or gluons) produced at matrix element level in addition to the process X.

Process	Generator	$\sigma \times \varepsilon_{filter}$ [pb]	K-factor
Charge misID background processes			
$Z/\gamma^* \to e^+e^-$ ($m_{\ell\ell} \in [10, 60]$ GeV)	Alpgen+Herwig	3477.9	1.195
$Z/\gamma^* \to e^+e^- + 1p$ ($m_{\ell\ell} \in [10, 60]$ GeV)	Alpgen+Herwig	108.72	1.195
$Z/\gamma^* \to e^+e^- + 2p$ ($m_{\ell\ell} \in [10, 60]$ GeV)	Alpgen+Herwig	52.83	1.195
$Z/\gamma^* \to e^+e^- + 3p$ ($m_{\ell\ell} \in [10, 60]$ GeV)	Alpgen+Herwig	11.31	1.195
$Z/\gamma^* \to e^+e^-$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	711.82	1.229
$Z/\gamma^* \to e^+e^- + 1p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	155.17	1.229
$Z/\gamma^* \to e^+e^- + 2p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	48.74	1.229
$Z/\gamma^* \to e^+e^- + 3p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	14.23	1.229
$Z/\gamma^* \to e^+e^- + 4p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	3.76	1.229
$Z/\gamma^* \to e^+e^- + \geq 5p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	1.09	1.229
$Z/\gamma^* \to \mu^+\mu^-$ ($m_{\ell\ell} \in [10, 60]$ GeV)	Alpgen+Herwig	3477.1	1.195
$Z/\gamma^* \to \mu^+\mu^- + 1p$ ($m_{\ell\ell} \in [10, 60]$ GeV)	Alpgen+Herwig	108.75	1.195
$Z/\gamma^* \to \mu^+\mu^- + 2p$ ($m_{\ell\ell} \in [10, 60]$ GeV)	Alpgen+Herwig	52.74	1.195
$Z/\gamma^* \to \mu^+\mu^- + 3p$ ($m_{\ell\ell} \in [10, 60]$ GeV)	Alpgen+Herwig	11.29	1.195
$Z/\gamma^* \to \mu^+\mu^-$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	712.11	1.229
$Z/\gamma^* \to \mu^+\mu^- + 1p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	154.77	1.229
$Z/\gamma^* \to \mu^+\mu^- + 2p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	48.91	1.229
$Z/\gamma^* \to \mu^+\mu^- + 3p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	14.22	1.229
$Z/\gamma^* \to \mu^+\mu^- + 4p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	3.78	1.229
$Z/\gamma^* \to \mu^+\mu^- + \geq 5p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	1.09	1.229
$Z/\gamma^* \to \tau^+\tau^-$ ($m_{\ell\ell} \in [10, 60]$ GeV)	Alpgen+Herwig	3478.0	1.195
$Z/\gamma^* \to \tau^+\tau^- + 1p$ ($m_{\ell\ell} \in [10, 60]$ GeV)	Alpgen+Herwig	108.67	1.195
$Z/\gamma^* \to \tau^+\tau^- + 2p$ ($m_{\ell\ell} \in [10, 60]$ GeV)	Alpgen+Herwig	52.73	1.195
$Z/\gamma^* \to \tau^+\tau^- + 3p$ ($m_{\ell\ell} \in [10, 60]$ GeV)	Alpgen+Herwig	11.33	1.195
$Z/\gamma^* \to \tau^+\tau^-$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	711.69	1.229
$Z/\gamma^* \to \tau^+\tau^- + 1p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	155.17	1.229
$Z/\gamma^* \to \tau^+\tau^- + 2p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	48.74	1.229
$Z/\gamma^* \to \tau^+\tau^- + 3p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	14.23	1.229
$Z/\gamma^* \to \tau^+\tau^- + 4p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	3.76	1.229
$Z/\gamma^* \to \tau^+\tau^- + \geq 5p$ ($m_{\ell\ell} \in [60, 2000]$ GeV)	Alpgen+Herwig	1.09	1.229
$t\bar{t}\,(\geq 1W^\pm \to \ell\nu_\ell)$	MC@NLO	112.9	1.2158
$tW^\pm$	MC@NLO	20.7	1.08
$W^+W^- \to \ell\nu\ell\nu$	Sherpa	5.5	1.07
$ZZ \to \ell\ell\nu\nu$	Sherpa	0.5	1.14
Photon conversion background processes			
$W^\pm\gamma$ ($W^\pm \to e^\pm\nu$)	Sherpa	162.88	1.0
$W^\pm\gamma$ ($W^\pm \to \mu^\pm\nu$)	Sherpa	162.88	1.0
$W^\pm\gamma$ ($W^\pm \to \tau^\pm\nu$)	Sherpa	162.96	1.0

Table 5.2: Simulated processes used for the charge misID and photon conversion background modeling. For each process, the generator configuration, the cross-section times the generator filter efficiency and the K-factor used to scale the simulated cross-section value to the latest theoretical calculation are given. The full simulation setup is specified in Table 2.1. $X + np$ with n integer and X a generic process indicates the number of partons (quarks or gluons) produced at matrix element level in addition to the process X.

5.3 Event selection

The aim of the selection is to collect a pure sample of events with prompt same-sign lepton pairs. The reducible backgrounds are suppressed with stringent lepton identification requirements, while to reduce Standard Model prompt backgrounds, events containing Z boson decays to leptons are rejected. The requirements on the same-sign lepton pair kinematic variables are minimal in order to perform an inclusive measurement of same-sign lepton pair production that is not model dependent.

5.3.1 Lepton selection

The reconstruction of leptons in ATLAS has been described in Section 2.4. A set of general requirements on the acceptance and the quality of reconstructed leptons has also been introduced in Section 2.4.1.2 for electrons and in Section 2.4.2.1 for muons. However, leptons used in this analysis are required to pass additional criteria.

Two levels of selection are defined for reconstructed leptons: *loose* and *tight*. The leptons selected by the *loose* requirements (*loose* leptons) are used for the data-driven estimate of the non-prompt background, as explained in Section 5.4.4. The *tight* requirements select a sub-sample of *loose* leptons of higher quality (*tight* leptons), and these leptons are used in the definition of the signal region selection. Table 5.3 contains the definitions of *loose* and *tight* leptons.

The lepton transverse momentum requirements, $p_{\mathrm{T}} > 20$ GeV and at least one with $p_{\mathrm{T}} > 25$ GeV, have been chosen to ensure high and flat trigger efficiency in all categories, and to avoid a large contamination of non-prompt backgrounds as they typically produce low p_{T} leptons. The requirements on the impact parameters are relative to the precision of z_0 and d_0. The measured d_0 has to be less than three times its uncertainty, while the requirement on z_0 depends on the θ of the lepton. These conditions suppress leptons produced by particles with a measurable lifetime like heavy flavour hadrons and leptons produced in interactions with the detector. The $d_0 < 0.2$ mm condition is applied to reconstructed muons to avoid the cosmic ray background. Muons with poorly measured charge are rejected by requiring the charge measurement in the IDet and the MS to be the same ($Q(IDet) = Q(MS)$). Electrons are selected if they satisfy the *medium++* identification criterion (Section 2.4.1.1); the conversion requirements from the *tight++* identification criterion are also applied to reject electrons from charge misID and photon conversion.

To avoid the case where a particle is reconstructed multiple times, when two reconstructed particles are found to be very close in ΔR, one of them is rejected. For instance, a muon can, in rare cases, release enough energy in Ecal and therefore be reconstructed both as a muon and as an electron. In such cases, the reconstructed electron is removed. Jets close to electrons are also removed.

Requirements on the *etcone* and *ptcone* isolations, and $\Delta R(\ell, jet)$, further reduce

the non-prompt and charge misID backgrounds. The isolation requirements have been optimized using simulation and opposite-sign di-lepton data. This study has been done by collaborators, hence it is just briefly described in the following. The efficiency of the isolation selection for prompt leptons is measured in $H_{\mathrm{L}}^{\pm\pm}$ pair production simulations; $c\bar{c}$ and $b\bar{b}$ simulated events are used to determine the non-prompt lepton rejection power. The isolation efficiency has also been measured in data with an opposite-sign lepton pair with invariant mass compatible with the Z mass, and in $Z/\gamma^* \to \ell^+\ell^-$ simulations. No significant differences have been observed between the efficiencies in data and simulation as a function of the lepton η and p_{T}, and the number of primary vertices. Thus, no corrections are applied to the efficiency of the isolation selection in simulated events.

Jets are used in the lepton selection to reject leptons of non-prompt origin. Jets are also used in the estimate of the background with non-prompt electrons, as described in Section 5.4.4.2. The baseline criteria for the jet selection are described in Section 2.4.6. In addition, all jets considered have $\Delta R\,(jet, e) > 0.2$, where e is any reconstructed electron, to avoid the use of jets reconstructed from prompt electron signals in the calorimeters. If a lepton is close to a high energy selected jet, determined by $\Delta R\,(\ell, jet) \leq 0.4$ for jets with $p_{\mathrm{T}}\,(jet) > 25$ GeV $+ p_{\mathrm{T}}\,(\ell) \times 0.05$, the lepton is removed because it is considered non-prompt. The efficiency of this selection requirement has been measured in data and simulation from a $Z \to \ell^+\ell^-$ enriched region. The two measurements are compatible and no correction is applied to the simulated events.

Alternative definitions of *loose* and *tight* leptons are used in the validation of the non-prompt background estimate. These selections modify the nominal selection such that more non-prompt leptons pass the *tight* and the *loose* selections. More details are given in Section 5.4.5.4.

Nominal Selection	Muon	Electron
loose	$p_{\mathrm{T}} > 20$ GeV	
	$p_{\mathrm{T}} > 25$ GeV for the highest p_{T} (leading) lepton of the pair	
	$\lvert z_0 \sin\theta\rvert < 1$ mm	
	$\lvert d_0\rvert/\sigma(d_0) < 3$ and $\lvert d_0\rvert < 0.2$ mm	$\lvert d_0\rvert/\sigma(d_0) < 3$
	$Q(IDet) = Q(MS)$	*medium++* ID
		not flagged as from photon conversion
		$\Delta R\,(e, \mu) > 0.2$
	$\Delta R\,(\ell, jet) > 0.4$ if $p_{\mathrm{T}}\,(jet) > 25$ GeV $+ p_{\mathrm{T}}\,(\ell) \times 0.05$	
tight	$ptcone30/p_{\mathrm{T}} < 0.07$	$ptcone30/p_{\mathrm{T}} < 0.1$
	$etcone30 < 3.5 + (p_{\mathrm{T}} - 20) \times 0.06$ GeV	$etcone20 < 3 + (p_{\mathrm{T}} - 20) \times 0.037$ GeV
		tight++ ID

Table 5.3: Definitions of the *loose* and *tight* selections applied to muons and electrons. The selection of jets used for the $\Delta R\,(\ell, jet)$ requirement is described in the text.

5.3.2 Trigger requirements

Data containing pairs of leptons are selected using the di-lepton triggers. These triggers were chosen after a careful analysis of the trigger options suitable for di-lepton event acquisition: the single lepton and di-lepton triggers described in Section 2.4.4. Simulated $W^{\pm}Z$ events where at least two leptons pass the lepton selection (Section 5.3.1) are used to study the trigger performances. Events with two electrons can be collected with the single electron trigger (SETrig) or the di-electron trigger (DETrig), while the single muon trigger (SMTrig) and the di-muon trigger (DMTrig) are used for $\mu^{\pm}\mu^{\pm}$ events. For mixed flavour $e^{\pm}\mu^{\pm}$ events, the single lepton triggers combination (SETrig or SMTrig) and the electron-muon trigger (EMTrig) are tested. Table 5.4 compares the efficiencies of the simulated multi-lepton events to pass the several trigger requirements in each lepton flavour category.

$e^{\pm}e^{\pm}$ category	SETrig	DETrig
Efficiency (%)	99.2 ± 0.5	97.6 ± 0.8
$e^{\pm}\mu^{\pm}$ category	SETrig or SMTrig	EMTrig
Efficiency (%)	97.9 ± 0.7	85.9 ± 1.8
$\mu^{\pm}\mu^{\pm}$ category	SMTrig	DMTrig
Efficiency (%)	93.2 ± 1.1	93.8 ± 1.0

Table 5.4: Efficiencies to trigger the acquisition of multi-lepton $W^{\pm}Z$ events for several trigger options and in the three lepton flavour categories considered in this analysis. The error reflects the statistical uncertainty of the simulated samples.

The di-lepton triggers and the single lepton triggers have comparable efficiency except in the $e^{\pm}\mu^{\pm}$ category. Here the single lepton trigger combination has 10% higher efficiency than the EMTrig. However, there are differences in selecting di-lepton pairs that have to be taken into account. Contrary to the di-lepton triggers, the single lepton triggers:

- do not apply any requirement to the other leptons in the event;
- impose isolation requirements on the trigger lepton to limit the firing rate to a level adequate for acquisition.

The former difference is beneficial to the signal acceptance; the latter is detrimental to the ability to precisely determine the non-prompt background.

The leptons selected for analysis, the *tight* leptons, are required to pass more stringent requirements than those imposed by any of these triggers. However, the leptons used to estimate the non-prompt background have looser isolation requirements compared to those for the single lepton trigger. To avoid biases, the *loose* lepton selection should require at least the same level of selection as the triggers. This means that to use single lepton triggers, the *loose* lepton must require some *ptcone* isolation. This condition undermines

the purpose of the *loose* lepton definition for selecting a large sample of non-prompt leptons. The sample of *loose* leptons needs to be unbiased by any isolation requirements.

Di-lepton pairs selected by the single lepton triggers are also found inconsistent with the method for the non-prompt background estimate for another reason: the triggers do not set any condition on the second lepton in the event. The definition of the non-prompt regions used in the non-prompt background estimate (Section 5.4.4.2 and Section 5.4.4.3) is such that any non-prompt lepton satisfy the trigger requirements. The non-prompt background estimate can be biased in events where the second lepton fails the trigger requirement of the non-prompt region. No bias is observed for muons, but a significant excess of predicted non-prompt background is found when the electron fails the trigger (see Figure B.1(b,d)). More details on this issue are given in Section B.2.

Given the potential issues affecting the background estimate in events selected by the single lepton triggers, only di-lepton triggers are used in this analysis. In the $e^{\pm}\mu^{\pm}$ category, the EMTrig is combined with a single electron trigger without isolation requirements (SENoIsoTrig) that works for electrons with $p_{\mathrm{T}} > 60$ GeV. This additional trigger is used to recover efficiency losses at high electron p_{T}. The EMTrig trigger requires the leakage in the hadronic calorimeter of the energy associated with the electron to be smaller than 1 GeV at L1. This requirement does not consider the absolute energy of the interacting electron. Energetic electrons can therefore produce leakage and not be accepted by the EMTrig. The SENoIsoTrig recover the inefficiency at high p_{T} and does not bias the non-prompt background estimate because no isolation requirement is applied. The DETrig does not show loss of efficiency for large electron p_{T} and therefore the SENoIsoTrig is not used in $e^{\pm}e^{\pm}$ events.

The chosen trigger configuration ensures high efficiency over the whole p_{T} range considered and allows for a reliable estimate of the non-prompt background in each category. Any pair of same-sign leptons selected in the analysis is required to activate the di-lepton trigger specific to the lepton flavour combination of the pair.

5.3.3 Selection of same-sign pairs

The events must contain at least two leptons (electrons or muons) satisfying the selection criteria in Section 5.3.1 and with the same electric charge. The pairs activate the event acquisition via the di-lepton triggers described in the previous section.

The selection does allow for additional leptons and other reconstructed particles like jets, photons and τ_{had} in the event. If more same-sign lepton pairs are reconstructed, any pair is considered for analysis. The lepton pairs are divided in three categories depending on their flavour: $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$, $\mu^{\pm}\mu^{\pm}$. These categories are studied separately.

The expected fraction of events with multiple same-sign lepton pairs in the same category for the considered backgrounds is as little as 0.1%. These events usually contain four leptons with null total charge ($\ell^{+}\ell^{+}\ell^{-}\ell^{-}$) forming two same-sign lepton pairs. In new

physics processes the generation of multiple same-sign pairs can be favoured. For instance, for the $\phi^{\pm\pm}$ pair production, both same-sign lepton pairs are reconstructed between 0.5% and 10% of the time depending mainly on the $\phi^{\pm\pm}$ mass, but also on the flavour category. Therefore the use of all reconstructed same-sign lepton pairs increases the sensitivity to some new physics models.

The same-sign lepton pairs considered have invariant mass $m(\ell^{\pm}, \ell^{\pm})$ larger than 15 GeV to avoid contamination from low-mass hadronic resonances, like the J/ψ and the Υ.

In the $e^{\pm}e^{\pm}$ category, the $m(e, e)$ region between 70 and 110 GeV is excluded from the signal region. The charge misID background generated by $Z/\gamma^* \to e^+e^-$ events in this region is very large and does not allow precision measurement of non-Standard Model physics. Rather, this region is used to calculate corrections for the charge misID background estimate and to validate it (Section 5.4.2.4 and Section 5.4.5.3).

The $W^{\pm}Z$ and ZZ productions are the main prompt backgrounds for this analysis. Such processes can be reduced by requiring that events with same-sign lepton pairs have no opposite-sign pairs of leptons with same flavour and with invariant mass within 10 GeV from the Z mass.

Some new physics models can produce Z bosons in events with same-sign leptons. The results presented are still valid for these models, but are less sensitive than dedicated analyses. For the processes of new physics described in Section 5.1, no Z boson is produced. In particular, in models with doubly charged particles, these particles cannot be produced in exclusive association with a Z boson in proton-proton collisions due to charge conservation. Even though these processes have no Z boson, some events could still be rejected due to accidental combinations of opposite-sign and same flavour leptons from different sources.

The constraint on the difference $|m(\ell^+, \ell^-) - m(Z)|$ has been tuned to optimize the sensitivity to $H_{\mathrm{L}}^{\pm\pm}$ pair production in the context of the Left-right symmetric model for several $H_{\mathrm{L}}^{\pm\pm}$ mass hypotheses. The significance is evaluated considering the total Standard Model background (prompt + non-prompt + charge misID + photon conversion) prediction and the signal expected for pair production of $H_{\mathrm{L}}^{\pm\pm}$ as a function of the excluded invariant mass interval around the Z mass. The significance is calculated as:

$$\sigma = \sqrt{2 \times ((s + B)\ln(1 + s/B) - s)} \tag{5.1}$$

where s is the expected signal yield for a given $H_{\mathrm{L}}^{\pm\pm}$ mass hypothesis, and $B = b + \Delta b^2$ is the total expected background plus its expected systematic error squared.

The two limit cases where Δb is either 0% or 30% of the total expected background yield have been considered. Uncertainties on predictions are expected to be below 10% for prompt backgrounds and around 20% for non-prompt backgrounds, hence 30% is chosen as the maximum systematic uncertainty on the expected background.

Figure 5.4 shows the expected significance for several $H_{\mathrm{L}}^{\pm\pm}$ masses as a function of the excluded invariant mass interval. The significance for null systematic uncertainty is shown for each flavour category. For $\mu^{\pm}\mu^{\pm}$, the category most sensitive to the Z mass requirement, the significance for a 30% systematic uncertainty is also shown. The significance is expressed in relation to the baseline significance calculated for no Z mass requirement.

The significance rises steeply with the increase of the excluded interval around the Z mass up to about $5-10$ GeV depending on the category and the $H_{\mathrm{L}}^{\pm\pm}$ mass. It then flattens out for high $H_{\mathrm{L}}^{\pm\pm}$ masses, while it decreases for low masses due to the higher probability of accidental di-lepton combinations compatible with the Z mass.

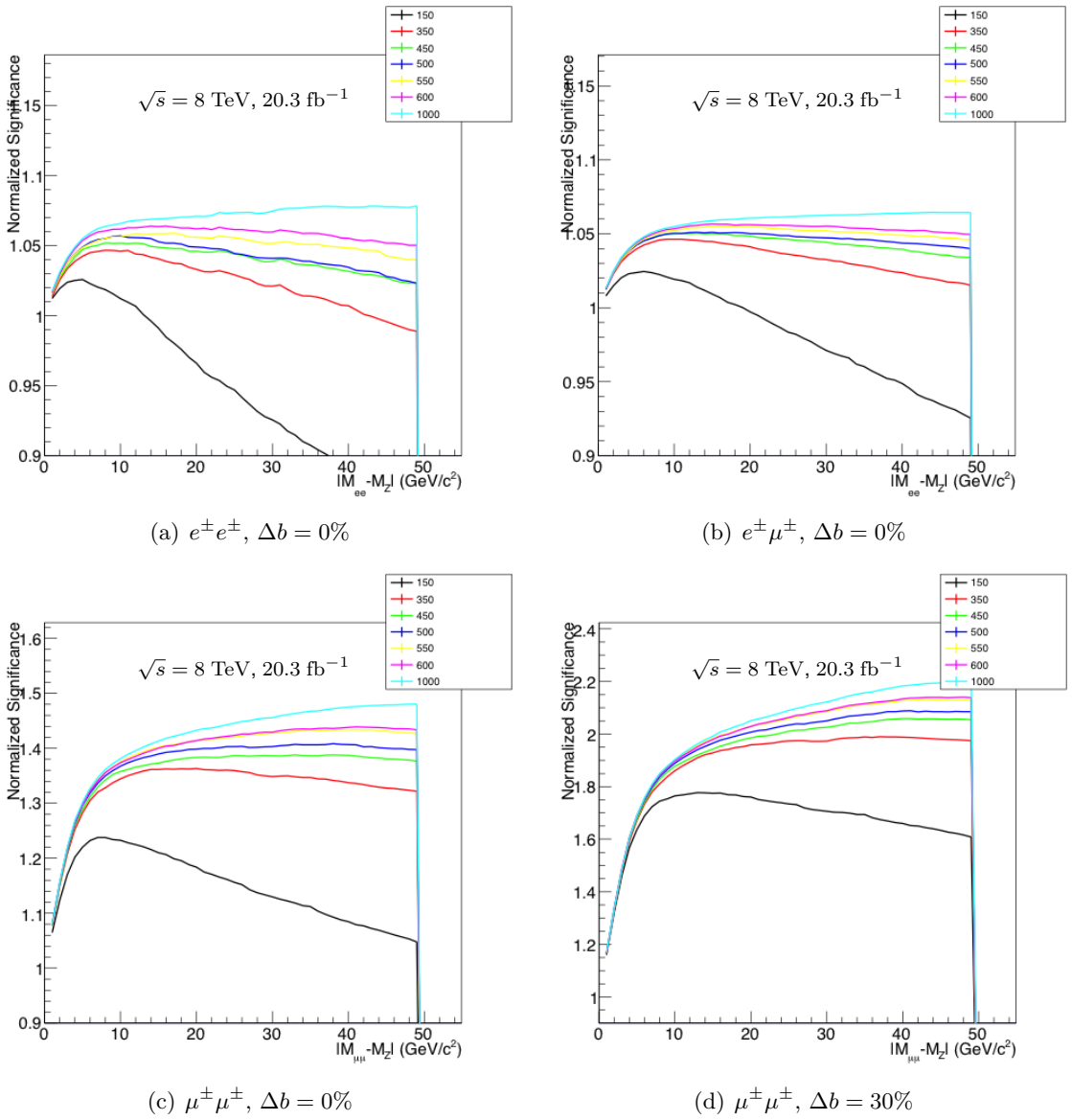
The biggest improvement in significance is achieved in the $\mu^{\pm}\mu^{\pm}$ category since $W^{\pm}Z$ and ZZ are the dominant backgrounds in that category. Other backgrounds dominate in the $e^{\pm}e^{\pm}$ and $e^{\pm}\mu^{\pm}$ categories. The significance in the $\mu^{\pm}\mu^{\pm}$ category for $|m(\ell^{+},\ell^{-}) - m(Z)| < 10$ GeV increases by $20-40$%, depending on the $H_{\mathrm{L}}^{\pm\pm}$ mass for $\Delta b = 0$%.

If systematic uncertainties are considered, the significance increases even more. Larger exclusion intervals are favoured. The significance, however, still flattens out or decreases above 10 GeV. For $\Delta b = 30$% and $|m(\ell^{+},\ell^{-}) - m(Z)| < 10$ GeV, the significance increases of $70-90$% in the $\mu^{\pm}\mu^{\pm}$ category, about 10% in the $e^{\pm}e^{\pm}$ category, and about 20% in the $e^{\pm}\mu^{\pm}$ category. The smallest improvement is for $H_{\mathrm{L}}^{\pm\pm}$ with a mass of 150 GeV. For this mass the significance decreases if the exclusion interval is extended beyond 10 GeV, even considering systematic uncertainties. Resonant Standard Model production of opposite-sign lepton pairs is indeed confined within 10 GeV of the Z mass and any further extension of the exclusion range would reduce just non-resonant processes. The production of $H_{\mathrm{L}}^{\pm\pm}$ pairs forms non-resonant opposite-sign lepton pairs as well. The lighter the $H_{\mathrm{L}}^{\pm\pm}$, the lower the invariant masses of the opposite-sign lepton combinations. For a mass of 150 GeV a consistent amount of combinations falls in the window around the Z mass, so the extension of the exclusion region reduces the significance. For higher masses the combinations have generally invariant masses sensibly higher than $m(Z)$, and larger exclusion ranges improve the significance.

Based on these optimization studies, events are excluded if the opposite-sign same flavour lepton pair has $|m(\ell^{+},\ell^{-}) - m(Z)| < 10$ GeV. Events that lie within the Z mass interval are still used to validate the prompt background estimate as explained in Section 5.4.5.

Many prompt background events also produce $W^{\pm}$. To reject such background upper limits on E_{T}^{miss} and M_{T} would be needed. However, several non-Standard Model processes can produce neutrinos or other invisible particles in the final state. Therefore, no selection is applied to suppress backgrounds with $W^{\pm}$ decaying to lepton and neutrinos.

The requirements described in this section define the events and the di-lepton pairs used in this analysis. For background validation, alternative regions are defined by changing some of the event and lepton selection requirements as described in Section 5.4.5.



(a) $e^{\pm}e^{\pm}$, $\Delta b = 0\%$

(b) $e^{\pm}\mu^{\pm}$, $\Delta b = 0\%$

(c) $\mu^{\pm}\mu^{\pm}$, $\Delta b = 0\%$

(d) $\mu^{\pm}\mu^{\pm}$, $\Delta b = 30\%$

Figure 5.4: Expected significance for $H_{\mathrm{L}}^{\pm\pm}$ pair production for several $H_{\mathrm{L}}^{\pm\pm}$ mass hypotheses as a function of the size of the excluded invariant mass interval around the Z mass. The significance is expressed in terms of the significance calculated for no Z mass requirement. The significance is computed assuming no systematic uncertainties for the $e^{\pm}e^{\pm}$ (a) and the $e^{\pm}\mu^{\pm}$ (b) categories. For the $\mu^{\pm}\mu^{\pm}$ category the significance for both no systematic uncertainty (c) and 30% uncertainty on the background prediction (d) is calculated.

5.4 Background estimate

The Standard Model processes that produce same-sign leptons in the final state, or that for experimental reasons can be reconstructed as same-sign leptons, have been described in Chapter 4. These backgrounds are grouped as prompt backgrounds, the charge misID background from opposite-sign lepton pair sources, backgrounds with photon conversions and backgrounds with non-prompt leptons.

This section describes the methods used to estimate these types of backgrounds.

5.4.1 Prompt backgrounds

The prompt background processes producing same-sign lepton pairs in decreasing order of importance for this analysis are: $W^{\pm}Z$, ZZ, $W^{\pm}W^{\pm}$, $t\bar{t}W^{\pm}$, multi parton diboson (MPI VV) and $t\bar{t}Z$ productions. At large di-lepton invariant masses (> 500 GeV), the $W^{\pm}W^{\pm}$ yields are expected to be as large as those from $W^{\pm}Z$. The Feynman diagrams for the production of $W^{\pm}Z$, ZZ and $W^{\pm}W^{\pm}$, with subsequent decay into final states with a positively charged same-sign lepton pair, are shown in Figure 5.5. The $t\bar{t}W^{\pm}$, $t\bar{t}Z$ and MPI VV productions constitute less than 5% of the expected prompt background. Other Standard Model processes with same-sign leptons give negligible contributions to the backgrounds in the signal regions.

Simulations are used to model the sources of prompt background. The list of simulated samples and their cross-sections is given in Table 5.1. Only reconstructed leptons generated by prompt leptons produced at generator level from the decay of a $W^{\pm}$ boson, a Z/γ^* boson or a τ lepton are considered in these simulations. Non-prompt leptons are not considered because they are accounted for in the data-driven non-prompt background estimate.

5.4.2 Charge misID backgrounds

The main processes producing opposite-sign lepton pairs that are reconstructed with the same charge due to charge mis-identification effects are, in decreasing order of importance, Z/γ^*, $t\bar{t}$, W^+W^- and $tW^{\pm}$. In the $e^{\pm}\mu^{\pm}$ category $t\bar{t}$ and W^+W^- become dominant at di-lepton invariant masses above 100 GeV. Other processes have been estimated to give negligible contributions.

The charge misID background is modeled with simulations with corrections applied to the charge misID rate. The list of simulated samples and their cross-sections is given in Table 5.2. Only reconstructed leptons generated by prompt leptons or in trident events are selected. Non-prompt leptons are ignored.

The measurement of the electron charge misID rate, its validation and its application to the estimate of background events with charge misID electrons, are described in the following. The charge misID rate for muons has been studied in Section 4.2.2 and found to be negligible.

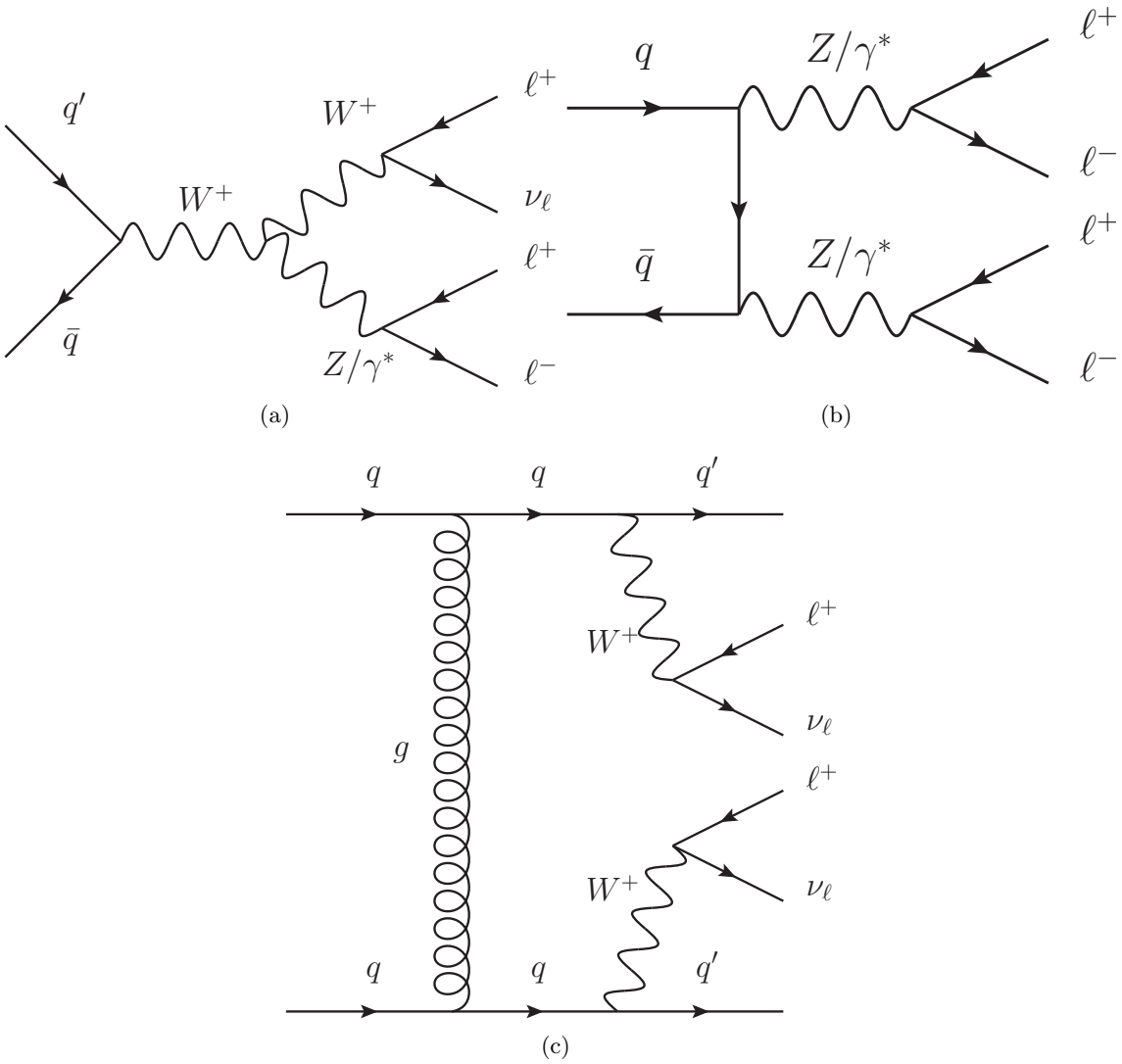


Figure 5.5: Tree level Feynman diagrams for the productions of $W^{\pm}Z$ (a), ZZ (b) and $W^{\pm}W^{\pm}$ (c). The decay to final states with a positively charged lepton pair is shown. Other tree level diagrams are possible.

5.4.2.1 Measurement of the charge misID rate

The origins of charge misID electrons are trident interactions, incorrect charge measurements and incorrect charge associations (see Section 4.2). Since the charge misID is due to the interaction of the electron with the detector, the charge misID probability of an electron is independent of the process that produced it, and only the electron kinematic variables matter. The sources of charge misID electrons are accounted for by simulations. However, due to the complexity of the charge misID processes, a study has been dedicated to measure and compare the probability of charge misID for electrons in data and simulation.

The events used for this measurement are di-electron events with $80 < m(e,e) < 100$ GeV. In the case of same-sign electron pairs, 98% of the events are expected to be produced by $Z/\gamma^* \to e^+e^-$, where the charge of one electron is mis-identified. For opposite-sign electrons, the interval of invariant mass $80 < m(e^+,e^-) < 100$ GeV is populated at 99% by Z/γ^* production. The lepton pairs, either opposite-sign or same-sign, with invariant mass within $80 < m(e,e) < 100$ GeV are used to calculate the charge misID probability for electrons. The electrons are required to pass the *tight* selection used in the analysis (Table 5.3). Note that the same-sign electrons used in the calculation of the charge misID rate are not used in the search for new physics. The same-sign electron pairs with invariant mass in the range $70 < m(e^\pm,e^\pm) < 110$ GeV are indeed removed from the signal region of the analysis.

The charge misID probability is measured by the minimization of a negative log likelihood function built with the observed numbers of same-sign and opposite-sign lepton pairs. This method is called the *likelihood method* [157].

Suppose the electron kinematic observables are classified in a finite number of states and ϵ_i is the charge misID probability for electrons in the *i-th* state. In this analysis, these states are defined by intervals in p_{T} and η in which the electron belongs. For a process producing a pair of prompt opposite-sign electrons in the *i-th* and *j-th* states, the possible final states, neglecting reconstruction inefficiencies, are:

1. e^+e^- without any charge misID electron, with a probability of $(1-\epsilon_i)\times(1-\epsilon_j)$,
2. e^+e^- with two charge misID electrons, with a probability of $\epsilon_i \times \epsilon_j$,
3. $e^\pm e^\pm$ where only one of the two electrons is charge misID, with a probability of $\epsilon_i \times (1-\epsilon_j) + (1-\epsilon_i)\times\epsilon_j$.

Since ϵ_i is very small for every state i, the probability of reconstructing same-sign leptons can be approximated to $\epsilon_i+\epsilon_j$ and events with two charge misID electrons can be neglected.

If N^{ij} opposite-sign leptons are produced, the expected number of reconstructed same-sign leptons reconstructed, λ^{ij}, is:

$$\lambda^{ij} = N^{ij}(\epsilon_i + \epsilon_j). \tag{5.2}$$

The probability of observing k same-sign electron events with λ expected is described by the Poisson distribution:

$$P(k,\lambda) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}, \tag{5.3}$$

Thus, the probability of observing N_{SS}^{ij} same-sign electron events when N^{ij} are produced is:

$$P(N_{SS}^{ij}, N^{ij}|\epsilon_i,\epsilon_j) = \frac{[N^{ij}(\epsilon_i+\epsilon_j)]^{N_{SS}^{ij}} \cdot e^{-N^{ij}(\epsilon_i+\epsilon_j)}}{N_{SS}^{ij}!}. \tag{5.4}$$

The likelihood L of the parameters ϵ_i, ϵ_j given N^{ij} and N_{SS}^{ij} is:

$$L(\epsilon_i,\epsilon_j|N_{SS}^{ij}, N^{ij}) = P(N_{SS}^{ij}, N^{ij}|\epsilon_i,\epsilon_j) \tag{5.5}$$

Taking into account all possible ij combinations, the likelihood for the charge misID rate is:

$$L(\vec{\epsilon}|\vec{N}_{ss},\vec{N}) = \prod_{i,j} \frac{[N^{ij}(\epsilon_i+\epsilon_j)]^{N_{ss}^{ij}} e^{-N^{ij}(\epsilon_i+\epsilon_j)}}{N_{ss}^{ij}!}, \tag{5.6}$$

where $\vec{\epsilon}$, $\vec{N}_{ss}$ and $\vec{N}$ are vectors containing all the ϵ_i, N_{SS}^{ij} and N^{ij}, respectively.

The likelihood function is used to obtain ϵ_i for each state i. The charge misID rates ϵ_i are obtained by minimizing the $-\ln L$ defined as:

$$-\ln L(\vec{\epsilon}|\vec{N}_{ss},\vec{N}) = \sum_{i,j} N^{ij}(\epsilon_i+\epsilon_j) - \ln[N^{ij}(\epsilon_i+\epsilon_j)]N_{ss}^{ij} \tag{5.7}$$

where the terms which do not depend on the rates ϵ_i, and hence do not influence the minimization, are omitted.

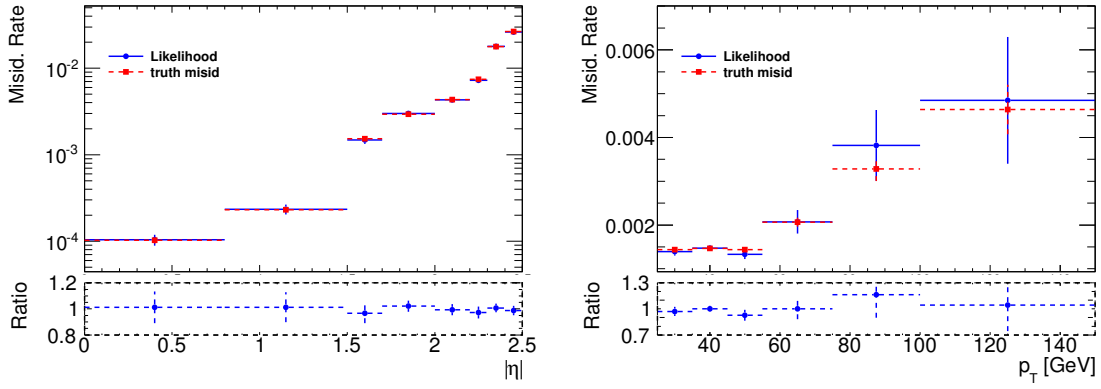
The *likelihood method* allows to use the full available electron sample, since it uses the information of both leptons in the pair. Therefore, the result is more statistically powerful compared to other methods used in the literature [157].

5.4.2.2 Validation of the charge misID rate measurement

The *likelihood method* is validated with Z/γ^* simulated events with $80 < m(e,e) < 100$ GeV. Only events with exactly two reconstructed electrons are considered. The rate ϵ_i is evaluated using either the *likelihood method* or the event generator information. In the latter case, ϵ_i is extracted by comparing the charge of the reconstructed electrons to that of the generator level electrons produced in the Z decay.

The charge misID rates are measured for intervals of p_{T} and $|\eta|$. The width of the p_{T} intervals is variable to ensure small enough statistical uncertainty in each one. The $|\eta|$ range is divided into intervals such that a similar amount of material is crossed by electrons in the same interval. The absolute value of η is used since intervals opposite in η result in the same rate. The rates as obtained from the *likelihood method* and the event

generator information are shown in Figure 5.6.



(a) (b)

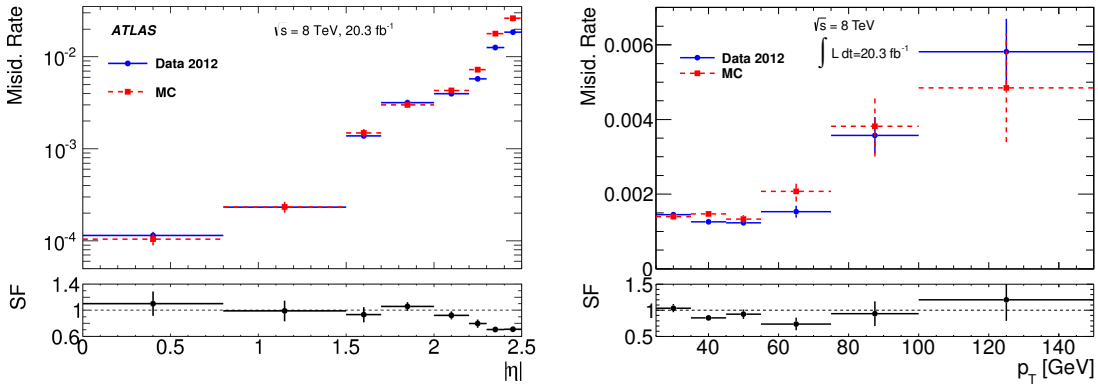
Figure 5.6: Charge misID rate as a function of $|\eta|$ (left) and p_{T} (right) as obtained using the *likelihood method* (solid blue) and using the event generator information (dashed red) in simulated $Z/\gamma^* \to e^+e^-$ events. The error bars correspond to the method errors in the first case and the statistical error in the second case. The upper panels show the rates and the lower ones the ratio of the rates measured by the *likelihood method* to the rates based on generator information.

The charge misID rate increases rapidly with $|\eta|$ by up to two orders of magnitude. The material crossed by electrons before reaching the calorimeter increases with $|\eta|$. This leads to a greater probability of trident interactions and multiple scattering, and consequently of charge misID. The charge misID rate increases to a minor degree with p_{T}. The accuracy of the charge measurement indeed worsens at higher p_{T} and trident events are more frequent due to the higher electron energy. The *likelihood method* results are compatible with the probabilities based on the generator information within the statistical uncertainty of the simulated sample for each p_{T} and $|\eta|$ interval. Thus, the *likelihood method* proves itself to be a solid method for the measurement of charge misID rates.

5.4.2.3 Charge misID rate in data and simulation

The charge misID rates obtained with the *likelihood method* in data and simulation are compared in Figure 5.7 in $|\eta|$ and p_{T} intervals. The rates as a function of p_{T} are measured up to 140 GeV due to the lack of electrons with higher p_{T} in Z/γ^* events. The rates are in good agreement along the whole p_{T} range with the largest difference being 25% for $55 < p_{\mathrm{T}} < 75$ GeV. A more significant difference is observed for $|\eta| > 2.2$. In that range the probability of charge misID is the highest and the simulation overestimates the charge misID rate by up to 30%. A pure simulation-based estimate of the charge misID background would therefore over-predict the background yields. The rates measured in data as a function of the electron η are used instead.

A cross-check of the measured rates is performed with opposite-sign di-electron pairs scaled according to the charge misID rates to obtain an estimate of charge misID same-

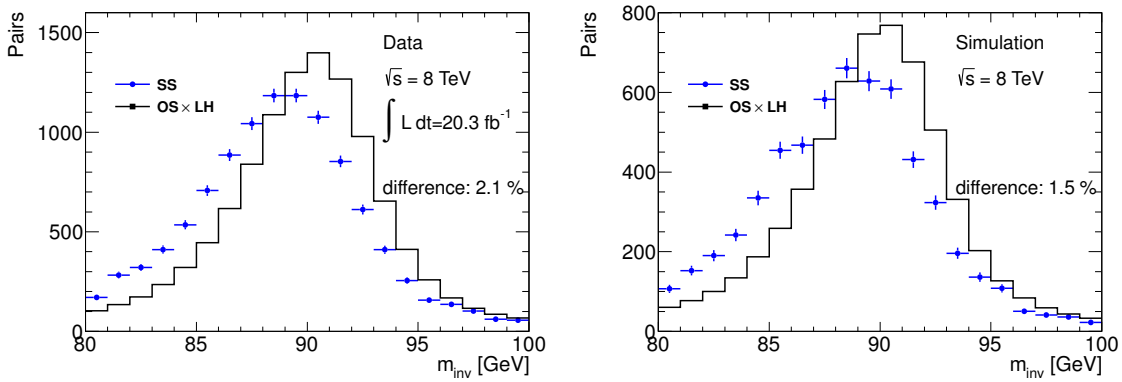


(a) (b)

Figure 5.7: Charge misID rate as a function of $|\eta|$ (left) and p_{T} (right) as measured in data and simulation using the *likelihood method*. The upper panels show the charge misID rates, while the lower ones show the scale factors (SF) which are the ratio of the rates measured in data to the rates measured in simulation. The error bars correspond to the uncertainty from the *likelihood method*.

sign pairs. The correction for opposite-sign pairs is $\frac{\epsilon_i+\epsilon_j-2\epsilon_i\epsilon_j}{1-\epsilon_i-\epsilon_j+2\epsilon_i\epsilon_j}$, where ϵ_i and ϵ_j are calculated according to the pseudorapidity of the leptons in the pair. The scaled opposite-sign di-electron pairs are compared to the observed same-sign di-electron pairs in the range $80 < m\,(e,e) < 100$ GeV. The comparison is performed in both data and simulation.

The invariant mass distributions of the pairs used in the test are shown in Figure 5.8. The distribution for scaled opposite-sign di-electron pairs is centred on the Z mass, while



(a) (b)

Figure 5.8: Same-sign lepton invariant mass distributions compared with those for opposite-sign leptons scaled by the charge misID rates in data (left) and simulations (right). The difference in yields is also indicated. The overall normalization of the simulated distributions is arbitrary.

the same-sign di-electron pairs have, on average, smaller invariant mass. This is due to the lower energy of electrons produced in the trident events compared to the electron at the origin of the interaction. For the same reason, events with same-sign pairs have a

lower probability of being reconstructed, and this leads to about a 2% difference in the background yield.

5.4.2.4 Estimation of the background with charge misID electrons

The prediction of the charge misID background is based on both data and simulation. Since the opposite-sign pairs corrected for charge misID do not take into account the subdivision of the energy that occurs in the trident processes, the use of simulated same-sign events is preferred for the modeling of the kinematic behaviour of the charge misID background. However, given the differences observed in charge misID rate between data and simulation at large pseudorapidity (see Figure 5.7(a)), the simulated yields are scaled by corrections that are a function of the $|\eta|$ of the charge misID electron. The charge misID electron is identified by means of the generator level information in the simulation.

The charge misID rate corrections SF_i are the ratios of the charge misID rates in data and in simulation as a function of $|\eta|$:

$$SF_i = \epsilon_i^{data} / \epsilon_i^{Simulation} \tag{5.8}$$

where i is the *i-th* interval in $|\eta|$. The sources of uncertainties on the estimate of these corrections considered are:

- the dependence on the process of production of the electrons;
- the trigger requirements;
- the selection of the $Z/\gamma^* \to e^+e^-$ candidates;
- the electron selection, and in particular the electron isolation requirement;
- the increase of the charge mis-reconstruction at high p_{T}.

To investigate the dependence of the charge misID rate on the additional activity in the event, the charge misID rate is measured in $t\bar{t}$ simulated events and compared to the rate for Z/γ^*. More jets are indeed expected in the final state of $t\bar{t}$ events. The rates are found to be compatible, however the statistical errors on the rates from $t\bar{t}$ are large.

The bias of the trigger requirement on the charge misID rate is investigated. Di-electron pairs in $Z/\gamma^* \to e^+e^-$ simulation are selected without any trigger requirements or by imposing the DETrig. The charge misID rates calculated in the two cases are compatible within uncertainties and therefore no systematic uncertainty is assigned.

Biases due to the $Z/\gamma^* \to e^+e^-$ event selection are also studied by changing the invariant mass requirement. Charge misID rates are calculated for $75 < m\,(e,e) < 105$ GeV and $85 < m\,(e,e) < 95$ GeV. The differences in rates with respect to the nominal ones are smaller than 8 % and are taken as systematic uncertainties.

The charge misID rates have been estimated for leptons that pass the *tight* selection. However, charge misID background also populates the regions containing *loose* electrons, which differ from the *tight* selection mainly in the isolation requirements (see Table 5.3). These regions are used for the non-prompt background estimate and must be cleaned of charge misID background contamination. The charge misID background estimate in regions with *loose* leptons is the same described in this section and the same charge misID corrections are applied to both *tight* and *loose* electrons. To check for potential biases due to the different lepton selections, an estimate of charge misID rates for electrons with loosened isolation (See Table 5.11) is made and the differences from the nominal rates are used as a systematic uncertainty. The largest difference found is 11%.

The systematic uncertainties on the charge misID corrections from the above sources as well as from the statistical uncertainties from the *likelihood method* are shown in Figure 5.9. The total error is also shown and varies between 6% and 20% depending on the pseudorapidity. For electrons with $p_{\mathrm{T}} > 100$ GeV, an additional uncertainty of 20% is assigned. This is to take into account that, at high p_{T}, the fraction of charge misID electrons due to wrong track measurement increases (see Figure 4.2). The uncertainty has been determined by using a distorted geometry in simulations to emulate the imperfect alignment of the detector that can affect the charge measurement.

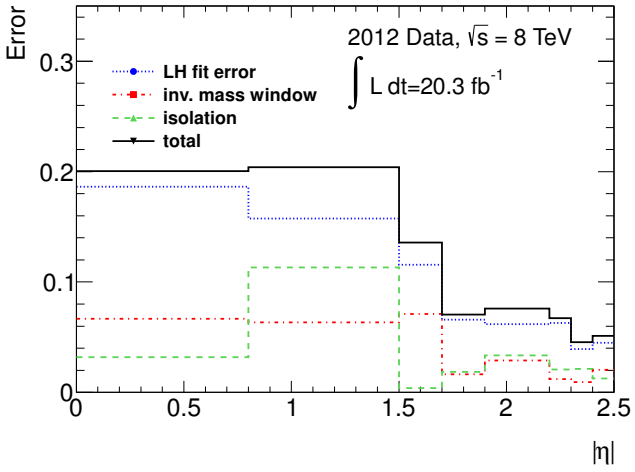


Figure 5.9: Relative uncertainties on the charge misID rate corrections as a function of $|\eta|$ from the *likelihood method* statistical error (blue), the interval of $m\,(e,e)$ used in the measurement (red), and the variation of isolation requirements (green). The total relative uncertainty (black) is also shown.

The charge misID background prediction is obtained by summing the expected same-sign lepton pairs from simulated $Z/\gamma^* + \text{jets}$, $t\bar{t}$, $W^{\pm}\,W^{\mp}$ and $tW^{\pm}$ production, scaled appropriately by the charge misID correction factors. The same-sign di-electron data are compared with the charge misID estimate and the residual Standard Model backgrounds in the $80 < m\,(e^{\pm},e^{\pm}) < 100$ GeV interval for validation in Section 5.4.5.3.

5.4.3 Photon conversion backgrounds

Processes with prompt leptons and photons are expected to give a small but non-negligible contribution to the background in the $e^{\pm}e^{\pm}$ and $e^{\pm}\mu^{\pm}$ categories. This is due to the possibility of photon conversion with one of the produced electrons having the same charge of the prompt lepton produced in the event. The processes with converted photons relevant for this analysis are $W^{\pm}\gamma$ and $Z\gamma$. These backgrounds are modeled with simulations. The list of simulated samples and their cross-sections is given in Table 5.2. The $Z\gamma$ production is included in the Z/γ^* simulation. Only reconstructed leptons generated by prompt leptons or photon conversion are considered.

Contrary to the case of processes with charge misID electrons, there is no trivial way to isolate a sample of events populated by photon conversions. Thus, the accuracy, which simulations have in emulating the photon conversion process, is hard to directly evaluate. The photon conversion is one of the two electromagnetic interactions with the detector characterizing a trident event. Therefore, photon conversions are expected to be simulated with a degree of accuracy similar to (or even better than) trident events. Since differences have been found in the charge misID rates measured in data and simulation, especially in the region where trident events occur more frequently, differences of the same order are expected for photon conversion rates. The predictions from simulation are therefore scaled by the same corrections used for charge misID processes as a function of the η of the electron from photon conversion (Equation 5.8). The uncertainties on charge misID corrections are also used for photon conversion corrections.

5.4.4 Non-prompt backgrounds

Backgrounds with non-prompt leptons contribute substantially to the same-sign di-lepton background. The main sources of non-prompt electrons are hadrons reconstructed as electrons and decays of hadrons containing b- or c-quarks. Non-prompt muons are mainly produced in heavy flavour hadron decays. Smaller contributions come from pion and kaon decays and hadronic showers piercing the calorimeters and activating the reconstruction of muons.

The main processes composing the non-prompt backgrounds in this analysis are QCD multi-jet and $W^{\pm}$ + jets productions with smaller contributions from Z + jets and $t\bar{t}$ productions. The simulation of non-prompt backgrounds is limited by the statistical uncertainty and the uncertainty on the modeling of non-prompt lepton reconstruction. The non-prompt background is therefore obtained from data.

5.4.4.1 The Fake Factor Method

The estimate of the non-prompt background is based on the Fake Factor Method. This data-driven method uses events enriched in non-prompt leptons in regions orthogonal to

the signal region to estimate the non-prompt background contributions in the signal region. The method accounts for backgrounds with either one or multiple non-prompt leptons. The Fake Factor Method is also used to estimate the non-prompt background in the validation regions of the background estimate, described in Section 5.4.5.

The non-prompt background estimate based on the Fake Factor Method proceeds through the following steps:

- definition of two lepton selections, *loose* and *tight*;
- calculation of fake factors in a non-prompt-dominated region;
- application of the fake factors to the same-sign lepton pairs with at least one *non-tight* lepton.

The method considers leptons that satisfy two levels of selection. The *tight* (ℓ) definition corresponds to the lepton selection with the highest quality standards and provides the best rejection of non-prompt leptons. It is the lepton selection used in the signal region of the analysis or, more generally, in the regions where the non-prompt background has to be estimated (regions of interest). In the region of interest both leptons in the selected pairs must pass the *tight* selection requirements. The *loose* ($\breve{\ell}$) selection has a loosened set of quality requirements allowing more non-prompt leptons to be selected. The definitions of *tight* and *loose* are given in Section 5.3.1 for the signal regions and in Section 5.4.5.4 for the validation regions of the non-prompt background estimate. The *non-tight* leptons ($\not{\ell}$) are defined as the $\breve{\ell}$ leptons failing the *tight* selection.

The method is based on the calculation of the fake factors f_ℓ^{FF} that represent the relative probability of a non-prompt lepton of flavour ℓ to be identified as *tight* rather than *non-tight*.

The f_ℓ^{FF} are estimated in regions dominated by non-prompt leptons, the non-prompt regions. These regions are defined in Section 5.4.4.2 and Section 5.4.4.3 for non-prompt electrons and non-prompt muons, respectively. The non-prompt leptons in this region must have origin and kinematic behaviour as similar as possible to those in the region of interest, but the two regions must be independent. The fake factors are calculated as:

$$f_\ell^{FF} = \frac{N_\ell}{N_{\not\ell}} \tag{5.9}$$

where N_ℓ and $N_{\not\ell}$ are respectively the number of *tight* and *non-tight* leptons selected in the non-prompt regions. The predicted contributions from other sources of background (prompt, charge misID, photon conversion) are subtracted from N_ℓ and $N_{\not\ell}$.

The fake factors f_ℓ^{FF} can depend on the lepton kinematic variables and the additional activity in the event. The composition of the non-prompt leptons is connected to the p_{T} and η of the non-prompt leptons. The efficiency of the *loose* and *tight* requirements can vary with η and p_{T} and the f_ℓ^{FF} is therefore calculated as a function of the η and p_{T} of the

non-prompt lepton. In this analysis, leptons are grouped in intervals of η and p_{T} whose widths are set according to the statistics available. For each interval in η and p_{T}, a fake factor is estimated. These factors are combined to obtain the fake factor as a function of both p_{T} and η:

$$f_{\ell}^{FF}\left(p_{\mathrm{T}},\eta\right)=\frac{f_{\ell}^{FF}\left(p_{\mathrm{T}}\right)f_{\ell}^{FF}\left(\eta\right)}{\left\langle f_{\ell}^{FF}\left(\eta\right)\right\rangle} \tag{5.10}$$

where $\left\langle f_{\ell}^{FF}\left(\eta\right)\right\rangle$ is the overall fake factor inclusive in lepton kinematic.

The estimate of the number of pairs with non-prompt leptons in the signal region is obtained from the number of pairs with at least one *non-tight* lepton corrected for the associated scale factor, as:

$$N_{\ell\ell}{}^{FF}=\sum^{N_{\ell\not\ell}}f_{\ell}^{FF}\left(p_{\mathrm{T}}\left(\ell\right),\left|\eta\left(\ell\right)\right|\right)-\sum^{N_{\not\ell\not\ell'}}f_{\ell}^{FF}\left(p_{\mathrm{T}}\left(\ell\right),\left|\eta\left(\ell\right)\right|\right)\times f_{\ell}^{FF}(p_{\mathrm{T}}(\ell'),|\eta(\ell')|) \tag{5.11}$$

where the first sum is over pairs with only one *non-tight* lepton and f_{ℓ}^{FF} is calculated with the *non-tight* lepton p_{T} and η; the second sum is over the pairs with two *non-tight* leptons and f_{ℓ}^{FF} is calculated for both leptons. The second term avoids double counting of pairs with two non-prompt leptons.

In case of leptons with different flavours, the pairs with one *tight* electron and one *non-tight* muon ($e\not\mu$) and the pairs with one *non-tight* electron and one *tight* muon ($\not e\mu$) must both be considered separately. Equation 5.11 takes the form:

$$\begin{aligned}N_{e\mu}^{FF}=\sum^{N_{e\not\mu}}f_{\mu}^{FF}\left(p_{\mathrm{T}}\left(\not\mu\right),\left|\eta\left(\not\mu\right)\right|\right)+\sum^{N_{\not e\mu}}f_{e}^{FF}\left(p_{\mathrm{T}}\left(\not e\right),\left|\eta\left(\not e\right)\right|\right)+\\ -\sum^{N_{\not e\not\mu}}f_{e}^{FF}\left(p_{\mathrm{T}}\left(\not e\right),\left|\eta\left(\not e\right)\right|\right)\times f_{\mu}^{FF}\left(p_{\mathrm{T}}\left(\not\mu\right),\left|\eta\left(\not\mu\right)\right|\right)\end{aligned} \tag{5.12}$$

Equation 5.11 assumes that $N_{\ell\not\ell}$, $N_{\not\ell\not\ell'}$ contain non-prompt lepton pairs only. The contributions of prompt, charge misID and photon conversion lepton pairs is therefore subtracted. Due to the lepton-dependent corrections applied, the subtraction is done only at the end. The prediction of the number of non-prompt di-lepton pairs becomes:

$$\begin{aligned}N_{\ell\ell}{}^{FF}\left(\mathrm{non-prompt}\right)=N_{\ell\ell}{}^{FF}\left(\mathrm{data}\right)-N_{\ell\ell}{}^{FF}\left(\mathrm{prompt}\right)-N_{\ell\ell}{}^{FF}\left(\mathrm{charge\ misID}\right)+\\ -N_{\ell\ell}{}^{FF}\left(\mathrm{photonconversion}\right)\end{aligned} \tag{5.13}$$

The Fake Factor Method allows the prediction of the kinematic distributions of non-prompt di-lepton pairs in the signal region by using the distributions for pairs with *non-tight* leptons appropriately corrected with fake factors.

5.4.4.2 Measurement of the electron fake factors

The measurement of the electron fake factors is performed on a "di-jet" sample which is enriched in events with either a jet/hadron misidentified as an electron or an electron from heavy flavour hadron decays. The di-jet events are reconstructed with exactly one electron, the non-prompt electron, and one jet pointing in the opposite azimuthal direction (recoiling jet). The angular distance in the transverse plane of the two, $\Delta\phi\,(e, jet)$, must be greater than 2.4. The electron satisfies the *loose* electron selection for the region of interest and the jet must have $p_{\mathrm{T}} > 30$ GeV in addition to the requirements described in Section 2.4.6.1.

The events are selected by a set of single electron triggers that do not require track isolation. The trigger identification requirements are also less stringent than those of the *loose* selection. The trigger condition therefore does not bias the definition of the lepton selection or consequently the f_{ℓ}^{FF} measurement. The triggers have different p_{T} thresholds and are used to calculate fake factors in several p_{T} ranges (see Table 5.5). Some triggers are pre-scaled, i.e. the events that fire these triggers are recorded only once every PS times due to the otherwise high rate of triggering. Pre-scaled triggers allow the selection of leptons with looser selection requirements at the price of reduced statistics.

PS can be as big as 1000 for this analysis. The lower the p_{T} threshold of the trigger, the higher the PS. The simulated prompt backgrounds have to be scaled by the inverse of the PS of the trigger used in order to be properly subtracted from data in these regions.

p_{T} range [GeV]	Trigger threshold at EF	PS
$20 < p_{\mathrm{T}} < 25$	15 GeV	975.97
$25 < p_{\mathrm{T}} < 60$	24 GeV	7.397
$p_{\mathrm{T}} > 60$	60 GeV	1

Table 5.5: p_{T} thresholds and pre-scale factors of the triggers used to collect electrons for the measurement of the electron fake factors in several p_{T} ranges.

To suppress contamination from prompt electron backgrounds, an event selection is applied. Events are rejected when:

- the transverse mass $M_{\mathrm{T}}\left(e, E_{\mathrm{T}}^{miss}\right) > 40$ GeV, to reduce $W^{\pm}$ and $t\bar{t}$ production;
- additional electrons are reconstructed and pass the *loose++* identification criteria (see Section 2.4.1.1) as expected in Z/γ^{*} and $t\bar{t}$ production;
- additional electrons, no matter their quality, are reconstructed and a pair of electrons in the event has invariant mass between 80 and 100 GeV. This reduces Z production.

The residual prompt backgrounds ($W^{\pm}+$jets, $t\bar{t}$, $Z/\gamma^{*}+$jets and diboson) are subtracted using simulation, before the selected electron is used to calculate f_{e}^{FF} as a function of p_{T} and η. The simulations used to model the $W^{\pm}+$jets background are listed in Table B.1.

The following sources of systematic uncertainty on the fake factor derivation are considered:

- prompt background subtraction via simulation;
- selection of the recoiling jet;
- non-prompt electron type;
- trigger biases.

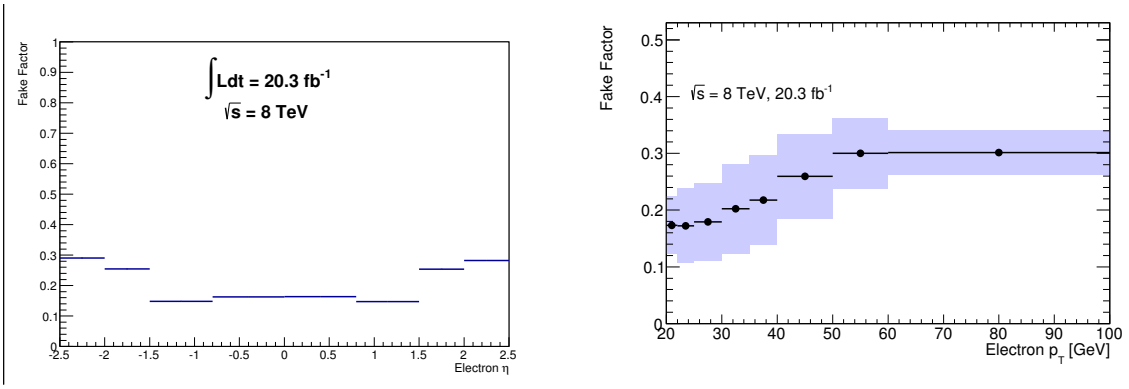
The uncertainty on the estimation of the prompt background to be subtracted in the di-jet region is estimated to be as big as 10% and accounts for theoretical uncertainties and experimental uncertainties (see Section 5.5). This systematic uncertainty affects mainly the high p_{T} fake factors, since the fraction of prompt lepton events increases at high p_{T}.

The jet selection can influence the electron fake factor value as a function of p_{T}. Indeed, the jet p_{T} and the electron p_{T} in the di-jet sample are correlated since the two have opposite directions in the transverse plane and there is only soft additional activity in the event. The recoiling jet p_{T} balances the p_{T} of the underlying jet faking the electron selection. Therefore, the electron p_{T} and the additional energy around the electron should balance the recoiling jet p_{T} too. For electrons with $p_{\mathrm{T}} > 20$ GeV, the recoiling jet has an average p_{T} of 30 GeV, hence the reason for the jet p_{T} threshold value. For a given electron p_{T}, the requirement on the jet p_{T} biases the activity around the electron or, in other words, biases its isolation. A higher recoiling jet p_{T}, means higher activity around the electron, and thus a smaller probability of passing the isolation requirements at the basis of the fake factor calculation. An uncertainty is assessed by varying the recoiling jet p_{T} requirement up to 50 GeV. This systematic uncertainty is dominant at low electron p_{T} and can be as large as 30%. It decreases significantly at higher electron p_{T}.

The fake factors can depend on the types of process that fake the electron reconstruction. If the composition of the non-prompt electrons used to estimate f_e^{FF} is different from that in the region of interest, the non-prompt prediction is biased. The fake factors have been calculated for electrons reconstructed from jets with light quark content and with heavy flavour content. The discrimination of the two types is done using a b-tagging algorithm, described in Section 2.4.7, applied to the jet closest to the electron candidate. The jet requirement $\Delta R\,(jet, e) > 0.2$ is not applied in this case, since the target is the jet producing the non-prompt electron. Fake factors are calculated separately for events with and without a b-tagged jet associated to the electron. The maximum difference with the nominal fake factors is 5% and is used as a systematic uncertainty.

To ensure that the fake factors are not biased by the trigger requirement, the fake factors obtained with the triggers in Table 5.5 are compared in overlapping p_{T} intervals. The fake factors obtained are compatible within the statistical uncertainties and no uncertainty is added.

The electron fake factors as a function of p_{T} and η are shown in Figure 5.10, and increase as a function of p_{T} and $|\eta|$. The total uncertainty varies between approximately 40% at $p_{\mathrm{T}} \approx 20$ GeV and 13% for $p_{\mathrm{T}} \approx 100$ GeV. Due to a lack of di-jet events with reconstructed electrons with $p_{\mathrm{T}} > 100$ GeV, the value of f_e^{FF} for $60 < p_{\mathrm{T}} < 100$ GeV electrons is also used above 100 GeV. However, the uncertainty is increased to 100%.



(a) (b)

Figure 5.10: Electron fake factors f_e^{FF} as a function of η (a) and p_{T} (b). The blue shaded area in the p_{T} plot corresponds to the statistical and systematic uncertainties added in quadrature. For the η plot only the statistical error, although too small to be visible, is included.

Electron fake factors have also been calculated for other electron selections and are used in the estimate of the non-prompt background in the validation regions shown in Section 5.4.5.4.

5.4.4.3 Measurement of the muon fake factors

The muon fake factor f_μ^{FF} is determined using a region that contains mainly muons from decays of heavy flavour hadrons, the main type of non-prompt muon expected in the signal regions. Given that the lifetimes of b- and c-hadrons are of the order of 10^{-12} s, the muons produced in their decays have a large impact parameter. Same-sign di-muon events with large d_0 are dominated by $b\bar{b}$ and $c\bar{c}$ production, and constitute an ideal sample to measure the muon fake factors. Contributions from other sources such as $t\bar{t}$ and $W^{\pm}$+jets are not relevant. The muons in these events fulfil the Fail-d_0 selection (see Table 5.13), where the muon must have a significant transverse impact parameter ($|d_0|/\sigma(d_0) > 3$ and $|d_0| < 10$ mm). The DMTrig is used to collect data in this region and the event selection follows the requirements in Section 5.3 for the $\mu^{\pm}\mu^{\pm}$ event category. The prompt muon contamination is almost negligible because both muons have large d_0 and the same charge. Nevertheless the prompt background expected from simulation is subtracted.

Muons from the decay of heavy flavour hadrons also tend to be surrounded by other particles and therefore to be less isolated than prompt muons. For these muons, there is a correlation between the d_0 and the isolation requirements. As a consequence, the

fake factor, which has isolation in its definition, can be sensitive to the d_0 requirement. The fake factors for muons selected with the Nominal selection (Table 5.3) and the Fail-d_0 selection (Table 5.13) are measured in $b\bar{b}$ and $c\bar{c}$ simulations. The overall fake factor is 30% larger for the Nominal selection, as shown in Table 5.6, confirming the correlation. No significant p_{T} dependence is observed though. The muon fake factor is therefore measured in data with muons passing the Fail-d_0 selection and increased by 30% to account for the correlation between d_0 and isolation.

	Nominal isolation
$\lvert d_0\rvert/\sigma(d_0) < 3$, $\lvert d_0\rvert < 0.2$ mm	0.191 ± 0.005
$\lvert d_0\rvert/\sigma(d_0) > 3$, $\lvert d_0\rvert < 10$ mm	0.146 ± 0.05
Correction factor	1.30 ± 0.05

Table 5.6: Muon fake factors derived from $b\bar{b}$ and $c\bar{c}$ simulations for muons passing the Nominal and Fail-d_0 selection and the corresponding correction factor.

The sources of uncertainty considered for the muon fake factor estimate are:

- the statistical uncertainty on data that increases at high p_{T} for *non-tight* muons up to 16.6%;
- the uncertainty on the prompt background subtraction that is found to be negligible for any p_{T} and η;
- the uncertainty associated with the correction factor from small to large d_0. The correction factor is found to vary by 16% requiring exactly two or more than two muons in the event. The difference is included as a systematic uncertainty in addition to the statistical error on the correction factor;
- the uncertainty associated with the type of non-prompt muons results in a systematic uncertainty of 0.6% as described below.

The composition of non-prompt muons that fulfil the Nominal selection and the Fail-d_0 selection can be different, with the latter expected to have a larger fraction of muons from heavy flavour hadron decays. The fraction of non-prompt muons from light flavour quarks can be obtained by studying the difference between the reconstructed muon momentum in the Inner Detector, p^{IDet}, and the reconstructed momentum in the Muon Spectrometer corrected for the energy loss in crossing the calorimeters (about 3 GeV usually), p^{MS}. The fractional momentum loss Δp is defined as:

$$\Delta p = (p^{IDet} - p^{MS})/p^{IDet} \tag{5.14}$$

For prompt muons and muons from heavy flavour hadronic decays, the fractional momentum loss is expected to be null because the same particle is responsible for the track in

the Inner Detector and the Muon Spectrometer. For muons originating from decays of light flavour hadrons such as a pion or a kaon, or a hadronic shower leaking in the Muon Spectrometer, a large positive momentum loss is expected. Templates of Δp for heavy flavour-originating muons and light flavour muons have been obtained from $t\bar{t}$ simulation (see Figure 5.11). The fraction of the two types of non-prompt muons in data is extracted

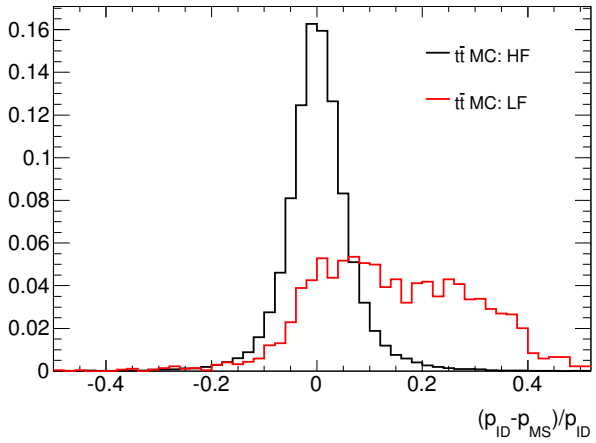


Figure 5.11: Fractional momentum loss for muons from heavy flavour (black) and light flavour (red) hadrons in simulated $t\bar{t}$ events. The overall normalization of the distributions is arbitrary.

with a template fit.

Table 5.7 shows the fractions obtained for *non-tight* muons passing the Nominal selection and *non-tight* muons selected with the Fail-d_0 requirements. The light flavour fraction

Muon Selection	Light flavour fraction
Nominal region	1.50 ± 0.03 %
Fail-d_0	0.61 ± 0.01 %

Table 5.7: Light flavour fraction for *non-tight* muons collected using the Nominal selection and the Fail-d_0 selection.

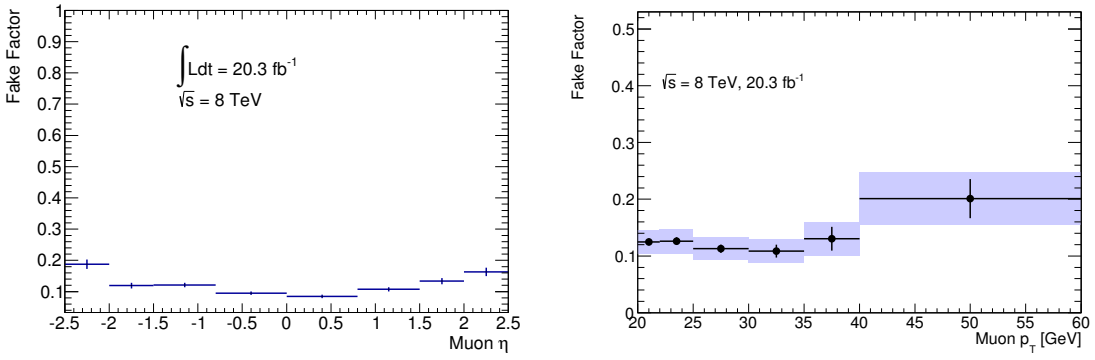
is very small in both cases, but is smaller for the Fail-d_0 selection, as anticipated. The fake factors are calculated for light flavour non-prompt muon events with Equation 5.15 exploiting the asymmetric behaviour of these muons. The fraction of heavy flavour muons indeed cancels out in the difference in the numerator and the denominator because of their symmetric Δp distribution.

$$f_{LF} = \frac{N_\ell\,(\Delta p > 0.1) - N_\ell\,(\Delta p < 0.1)}{N_{\not\ell}\,(\Delta p > 0.1) - N_{\not\ell}\,(\Delta p < 0.1)} \tag{5.15}$$

The light flavour fake factor obtained is 1.7 times higher than the nominal. Considering the light flavour fractions measured, a systematic uncertainty of 0.6% is estimated.

The total uncertainty on the muon fake factor f_μ^{FF} is 17% at low p_{T} and increases to 23% for $p_{\mathrm{T}} \approx 60$ GeV. A 100% uncertainty is used for $p_{\mathrm{T}} > 60$ GeV due to a lack of statistics.

The fake factors as a function of muon p_{T} and η after applying the correction factor from high to low impact parameter significance are shown in Figure 5.12. The fake factor for muons with $p_{\mathrm{T}} > 60$ GeV is equal to that for the highest p_{T} interval.



(a) (b)

Figure 5.12: Muon fake factors f_μ^{FF} measured as a function of η (a) and p_{T} (b). The blue shaded area corresponds to the statistical and systematic uncertainties added in quadrature. For η fake factors only the statistical error is shown.

The procedure explained in this section has also been applied to calculate the fake factors for the muons selected with the Weak isolation requirements (Table 5.11) that are used in the estimate of the non-prompt background in the validation regions shown in Section 5.4.5.4.

5.4.5 Background validation regions

The predictions for the various types of backgrounds are tested using validation regions orthogonal to the signal regions. Opposite-sign di-lepton pairs and same-sign di-lepton pairs in events with a $Z \to \ell^+\ell^-$ decay candidate are considered. These samples are used to validate the modeling of processes producing prompt leptons and the modeling of the prompt lepton reconstruction in simulations. The opposite-sign validation region also tests the production of the processes responsible for the charge misID background.

The charge misID rate corrections and the modeling of charge misID lepton kinematics are tested in $e^\pm e^\pm$ pairs with invariant mass close to the Z mass. Non-prompt background enhanced validation regions are instead created by altering the lepton impact parameter, the identification, the isolation or the p_{T} criteria. The aim of these regions is to validate the estimation of the Fake Factor Method method.

Table 5.8 contains a summary of the validation region definitions and their target background. The validation regions are divided by lepton pair flavour for a total of 19

Validation regions	Selection and target of the validation
Opposite-sign leptons	Two opposite-sign leptons, signal regions selection otherwise
	Corrections for simulated samples (normalization, efficiencies, lepton energy scale and resolution)
Prompt leptons	A same flavour opposite-sign lepton pair in the event with $\lvert m(\ell^+,\ell^-) - m(Z)\rvert < 10$ GeV, signal regions selection otherwise
	Predictions of prompt background with Z
Charge misID Z peak electron	Same-sign di-electron pair with $80 < m(e^\pm, e^\pm) < 100$ GeV, signal regions selection otherwise
	Charge misID rate corrections and charge misID modeling
Weak isolation electron	Electron selection as in Table 5.11
	Fake factors and non-prompt background predictions
Fail-ID electron	Electron selection as in Table 5.12
	Fake factors and non-prompt background predictions
Weak isolation muon	Muon selection as in Table 5.11
	Fake factors and non-prompt background predictions
Fail-d_0 muon	Muon selection as in Table 5.13
	Fake factors and non-prompt background predictions
Low-p_{T} muon	Muon selection as in Table 5.14
	Fake factors and non-prompt background predictions

Table 5.8: Description of the validation regions designed to test the several sources of background expected from Standard Model processes with same-sign lepton pairs.

regions. The fraction of expected same-sign lepton pairs of a certain type of background in each validation region is given in Table 5.9.

5.4.5.1 Opposite-sign lepton validation regions

A region with opposite-sign lepton pairs (the opposite-sign lepton validation region) is defined to validate the modeling of prompt leptons in simulation (selection efficiency, energy calibration and resolution) as well as the modeling of opposite-sign processes like Z/γ^* and $t\bar{t}$ that are used for the charge misID background estimate. The selection follows the criteria described in Section 5.3 with the exception of the total charge of the lepton pair which is required to be 0 (opposite-sign leptons). The opposite-sign lepton validation region is divided into lepton flavour categories and the estimated fraction of prompt background events is 98% or more in all categories.

The invariant mass distribution for e^+e^-, $e^\pm\mu^\mp$ and $\mu^+\mu^-$ pairs is shown in Figure 5.13. Z/γ^* production is dominant in the e^+e^- and $\mu^+\mu^-$ categories, accounting for 99% of the events, while $t\bar{t}$ production is the largest in the $e^\pm\mu^\mp$ category, with a fraction of 63%. The numbers of total predicted and observed lepton pairs are given in Table 5.15. The observed number of pairs is higher than predicted by 3%, 5% and 2% in the e^+e^-, $e^\pm\mu^\mp$ and $\mu^+\mu^-$ categories respectively, but this is within the theoretical systematic uncertainties on Z/γ^*

	Composition of electron pairs (%)			
Validation region	Prompt	Non-prompt	Charge misID	Photon conv.
Opposite-sign electrons	99.8	0.2	–	< 0.1
Prompt electrons	82.2	17.8	< 0.1	< 0.1
Charge misID Z peak electron	1.1	1.6	97.1	0.2
Weak isolation on both electrons	0.4	89.2	10.3	< 0.1
Weak isolation on leading electron	7.0	81.9	8.4	2.7
Weak isolation on sublead. electron	5.3	83.4	8.5	2.8
Fail-ID electron	1.5	57.0	37.1	4.4
	Composition of electron–muon pairs (%)			
Validation region	Prompt	Non-prompt	Charge misID	Photon conv.
Opposite-sign leptons	97.6	2.0	–	0.4
Prompt leptons	88.6	10.2	1.1	< 0.1
Weak isolation on electron	12.5	71.5	12.8	3.2
Weak isolation on muon	12.5	86.3	0.5	0.7
Fail-d_0 muon	15.1	79.1	4.5	1.3
Low-p_{T} muon	40.1	44.7	9.0	6.2
	Composition of muon pairs (%)			
Validation region	Prompt	Non-prompt		
Opposite-sign muons	99.6	0.4		
Prompt muons	96.5	3.5		
Weak isolation on both muons	1.0	99.0		
Weak isolation on leading muon	21.0	79.0		
Weak isolation on sublead. muon	15.9	84.1		
Fail-d_0 muon	19.0	81.0		

Table 5.9: Contributions of the background categories to the total Standard Model prediction in the validation regions. For the opposite-sign lepton regions, Z/γ^*, $t\bar{t}$, W^+W^- and $tW^{\pm}$ processes are included in the prompt background category.

and $t\bar{t}$ production. A good modeling of the background kinematic behaviour is observed in all ranges of invariant mass studied. This confirms that the lepton reconstruction is accurate within the expected degree of uncertainty at any energy.

5.4.5.2 Prompt lepton validation regions

A region with events containing lepton pairs from a Z decay (prompt lepton validation region) is defined to validate the modeling of leptons in simulation (selection efficiency, energy calibration and resolution) as well as the modeling of processes that produce prompt same-sign lepton pairs like $W^{\pm}Z$ and ZZ. The selection follows the criteria described in Section 5.3 with the exception that there must be at least one pair of same flavour opposite-sign leptons in the event (in addition to the same-sign lepton pair) and its invariant mass must be compatible with the Z mass ($|m(\ell^+,\ell^-) - m(Z)| < 10$ GeV).

The prompt lepton validation region is divided into lepton flavour categories. The composition of the background in this region is given in Table 5.9. The invariant mass distribution for $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ pairs is shown in Figure 5.14. Prompt backgrounds are dominant and account for the 82.2%, 88.6% and 96.5% of the expected predictions in the $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ categories, respectively. The numbers of total predicted and observed lepton pairs in these regions are collected in Table 5.15.

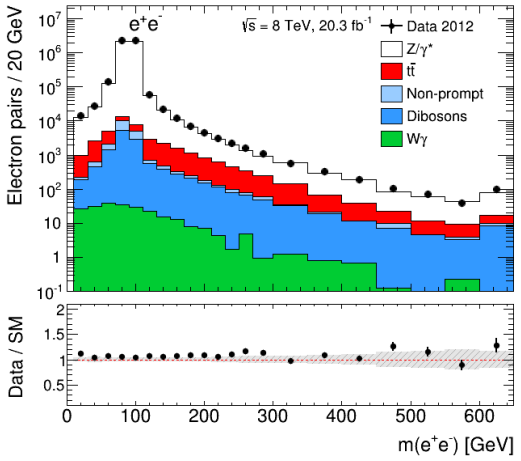
The statistical analysis of data in the signal regions is performed in several $m(\ell^{\pm},\ell^{\pm})$ intervals, as described later in Section 5.6.3. The background predictions in these validation regions are compared to data in the same $m(\ell^{\pm},\ell^{\pm})$ intervals used in the analysis. The ratio between data and expected pairs for each $m(\ell^{\pm},\ell^{\pm})$ interval is given Table 5.10. The predictions are compatible with the observed numbers of pairs for each category and invariant mass interval.

Category	Ratio between observed and expected pairs for $m(\ell^{\pm},\ell^{\pm})$						
	> 15 GeV	> 100 GeV	> 200 GeV	> 300 GeV	> 400 GeV	> 500 GeV	> 600 GeV
$e^{\pm}e^{\pm}$	0.97 ± 0.09	0.95 ± 0.10	0.96 ± 0.17	0.93 ± 0.27	1.3 ± 0.6	1.5 ± 1.0	1.3 ± 1.9
$e^{\pm}\mu^{\pm}$	1.05 ± 0.07	1.06 ± 0.08	1.25 ± 0.14	1.11 ± 0.21	1.5 ± 0.4	2.2 ± 0.9	2.5 ± 1.5
$\mu^{\pm}\mu^{\pm}$	1.10 ± 0.07	1.13 ± 0.09	1.03 ± 0.14	1.16 ± 0.26	0.8 ± 0.4	0.9 ± 0.7	1.3 ± 1.8

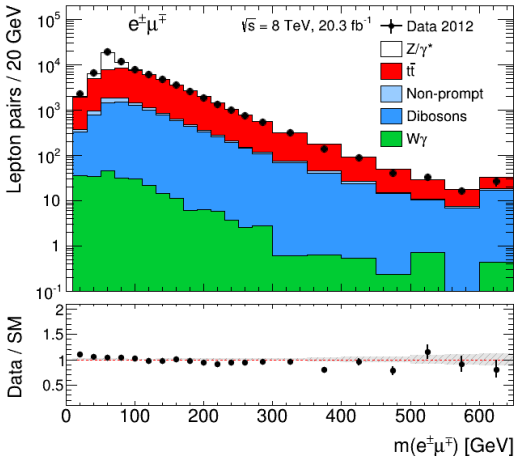
Table 5.10: Ratio between observed and expected same-sign pairs in the prompt lepton validation region for $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ categories and for various requirements on the di-lepton invariant mass. The uncertainties account for both statistical and systematic errors.

5.4.5.3 Charge misID Z peak electron validation region

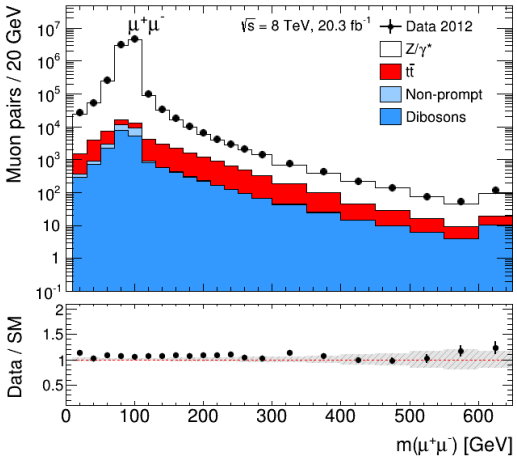
The di-electron pairs with $80 < m(e,e) < 100$ GeV have been used to calculated corrections to the charge misID rate. These corrections are verified by comparing same-sign di-electron pairs data with the charge misID background prediction in the same interval of invariant mass, $80 < m(e,e) < 100$ GeV. The selection otherwise follows the criteria



(a) e^+e^-

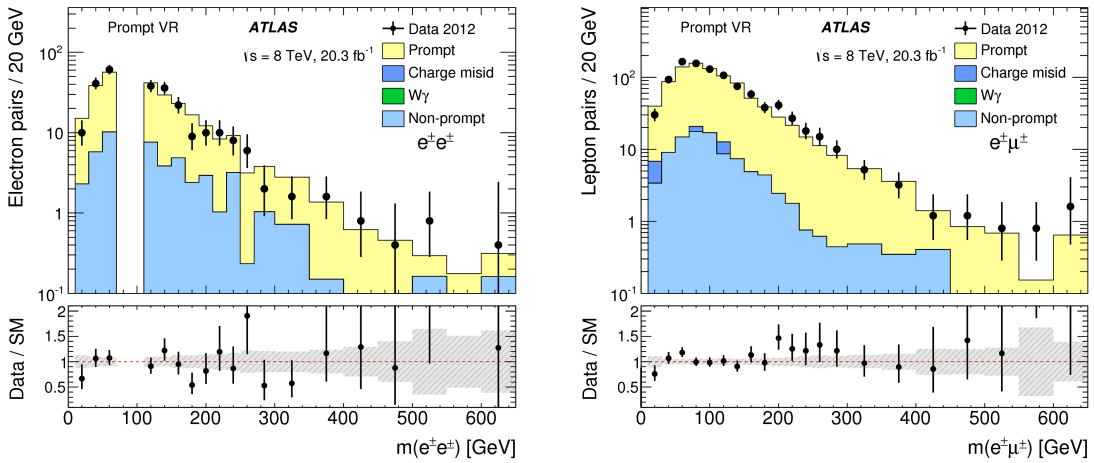


(b) $e^\pm\mu^\mp$



(c) $\mu^+\mu^-$

Figure 5.13: Invariant mass distributions for e^+e^- (a), $e^\pm\mu^\mp$ (b), and $\mu^+\mu^-$ (c) pairs in the opposite-sign leptons validation region. Events in data (black dots) are compared to the expected stacked contributions from Standard Model processes, as described in Section 5.4. Prompt background components, Z/γ^*, $t\bar{t}$ and VV, are shown separately. Overflow events are added to the last bin. The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the predictions. Note that the bin width is variable but the contents of all bins are normalized as if they are 20 GeV wide.



(a) $e^{\pm}e^{\pm}$ (b) $e^{\pm}\mu^{\pm}$

(c) $\mu^{\pm}\mu^{\pm}$

Figure 5.14: Invariant mass distributions of $e^{\pm}e^{\pm}$ (a), $e^{\pm}\mu^{\pm}$ (b) and $\mu^{\pm}\mu^{\pm}$ (c) pairs in the prompt lepton validation region. In the $e^{\pm}e^{\pm}$ category, the mass range between 70 GeV and 110 GeV is omitted because it is not included in the signal region. Events in data (black dots) are compared to the expected stacked contributions from Standard Model processes, as described in Section 5.4. Overflow events are added to the last bin. The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the predictions. Note that the bin width is variable but the contents of all bins are normalized as if they are 20 GeV wide.

described in Section 5.3. The composition of the background in this region is given in Table 5.9 and the invariant mass distribution of the selected electron pairs is shown in Figure 5.15. The charge misID background is dominant and accounts for 97.1% of the expected prediction. The numbers of total predicted and observed lepton pairs are given in Table 5.15 and are compatible within uncertainties.

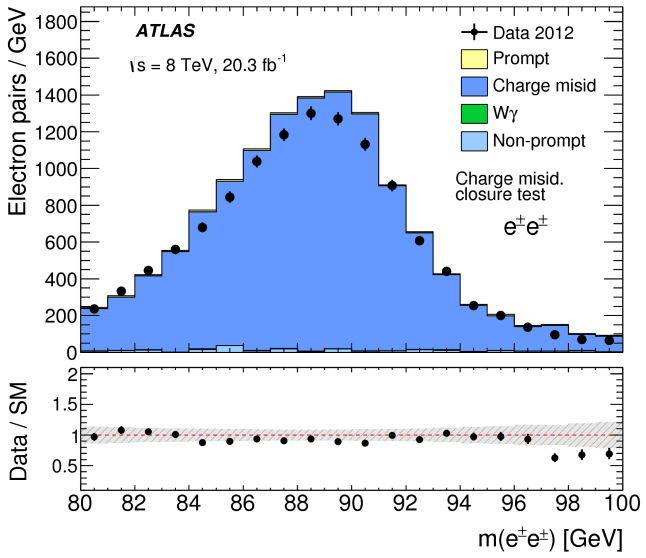


Figure 5.15: Invariant mass distributions of $e^{\pm}e^{\pm}$ pairs in the charge misID Z peak electron validation region. Events in data (black dots) are compared to the expected stacked contributions from Standard Model processes, as described in Section 5.4. The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the predictions.

5.4.5.4 Non-prompt leptons validation regions

Regions with enhanced non-prompt lepton backgrounds (non-prompt leptons validation regions) are defined by inverting some of the lepton selection requirements used to suppress non-prompt leptons, like impact parameter, identification, isolation and p_{T}. These regions are used to test the predictions of the Fake Factor Method and are divided into lepton flavour categories to test both electron and muon fake factors.

The event selection follows the criteria described in Section 5.3 but one or both leptons in the pair is selected with alternative requirements to those specified in Table 5.3. The alternative selections are:

- the Weak Isolation selection, defined in Table 5.11, collects leptons that fail the nominal isolation selection;
- the Fail-ID Electron selection (Table 5.12) has electrons with *medium++* identification at most;

- the Fail-d_0 Muon selection (Table 5.13) requires the muons to have large transverse impact parameter;
- the Low-p_{T} Muon selection (Table 5.14) limits the muon p_{T} to the range 18 to 20 GeV, where more non-prompt muons are expected.

Weak Isolation Selection	Muon	Electron
loose	same as *loose* muon in Table 5.3 $ptcone30/p_{\mathrm{T}} > 0.07$ or $etcone30 > 3.5 + (p_{\mathrm{T}} - 20) \times 0.06$ GeV	same as *loose* electron in Table 5.3 $ptcone30/p_{\mathrm{T}} > 0.1$ or $etcone20 > 3 + (p_{\mathrm{T}} - 20) \times 0.037$ GeV
tight	$ptcone30 < 4 + 0.07 \times p_{\mathrm{T}}$ GeV $etcone30 < 7.5 + (p_{\mathrm{T}} - 20) \times 0.06$ GeV	$ptcone30 < 4 + 0.1 \times p_{\mathrm{T}}$ GeV $etcone20 < 7 + (p_{\mathrm{T}} - 20) \times 0.037$ GeV *tight++* ID

Table 5.11: Definitions of the *tight* and *loose* Weak Isolation lepton selections used for the validation of the non-prompt background estimate. The differences from the nominal selections in Table 5.3 are highlighted.

	Fail-ID Electron Selection
loose	$p_{\mathrm{T}} > 20$ GeV $p_{\mathrm{T}} > 25$ GeV for the highest p_{T} (leading) lepton of the pair $\lvert z_0 \sin\theta \rvert < 1$ mm $\lvert d_0 \rvert / \sigma(d_0) < 3$ *loose++* ID fail *tight++* ID $\Delta R\,(e, \mu) > 0.2$ $\Delta R\,(e, jet) > 0.4$ if $p_{\mathrm{T}}\,(jet) > 25$ GeV $+ p_{\mathrm{T}}\,(e) \times 0.05$
tight	$ptcone30/p_{\mathrm{T}} < 0.1$ $etcone20 < 3 + (p_{\mathrm{T}} - 20) \times 0.037$ GeV *medium++* ID not flagged as from photon conversion

Table 5.12: Definitions of the *tight* and *loose* Fail-ID electron selections used for the validation of the non-prompt background estimate. The differences from the nominal selections in Table 5.3 are highlighted.

For each alternative lepton selection used in these regions, the fake factors are recalculated using the procedure explained in Section 5.4.4, but the associated systematic uncertainties are not re-evaluated. For any alternative lepton selection, the relative systematic uncertainty for a certain scale factor is assumed to be equal to the relative uncertainty of the nominal.

Four validation regions with $e^{\pm}e^{\pm}$ pairs are defined to test the predictions of background with non-prompt electrons:

Fail-d_0 Muon Selection	
loose	$p_{\mathrm{T}} > 20$ GeV
	$p_{\mathrm{T}} > 25$ GeV for the highest p_{T} (leading) lepton of the pair
	$\lvert z_0 \sin\theta \rvert < 1$ mm
	$\lvert d_0 \rvert / \sigma(d_0) > 3$ and $\lvert d_0 \rvert < 10$ mm
	$Q(IDet) = Q(MS)$
	$\Delta R\,(\mu, jet) > 0.4$ if $p_{\mathrm{T}}\,(jet) > 25$ GeV $+ p_{\mathrm{T}}\,(\mu) \times 0.05$
tight	$ptcone30/p_{\mathrm{T}} < 0.07$
	$etcone30 < 3.5 + (p_{\mathrm{T}} - 20) \times 0.06$ GeV

Table 5.13: Definitions of the *tight* and *loose* Fail-d_0 muon selections used for the validation of the non-prompt background estimate. The differences from the nominal selections in Table 5.3 are highlighted.

Low-p_{T} Muon Selection	
loose	18 GeV $< p_{\mathrm{T}} <$ 20 GeV
	$\lvert z_0 \sin\theta \rvert < 1$ mm
	$\lvert d_0 \rvert / \sigma(d_0) < 3$ and $\lvert d_0 \rvert < 0.2$ mm
	$Q(IDet) = Q(MS)$
	$\Delta R\,(\mu, jet) > 0.4$ if $p_{\mathrm{T}}\,(jet) > 25$ GeV $+ p_{\mathrm{T}}\,(\mu) \times 0.05$
tight	$ptcone30/p_{\mathrm{T}} < 0.07$
	$etcone30 < 3.5 + (p_{\mathrm{T}} - 20) \times 0.06$ GeV

Table 5.14: Definitions of the *tight* and *loose* Low-p_{T} muon selections used for the validation of the non-prompt background estimate. The differences from the nominal selections in Table 5.3 are highlighted in red.

Weak isolation on both electrons : the electrons fail the nominal electron isolation but pass a looser definition as described in Table 5.11.

Weak isolation on the leading electron : the electron with higher p_{T} fails the nominal electron isolation but passes a looser definition, described in Table 5.11. The electron with lower p_{T} passes the nominal selection, described in Table 5.3. Since one of the electrons passes the nominal selection, this validation region can test the nominal electron fake factors.

Weak isolation on the subleading electron : the electron with lower p_{T} fails the nominal electron isolation but passes a looser definition, described in Table 5.11. The electron with higher p_{T} passes the nominal selection, described in Table 5.3. Since one of the electrons passes the nominal selection, this validation region can test the nominal electron fake factors.

Fail-ID electron : at least one electron fails the nominal electron identification but passes a looser definition described in Table 5.12. One electron may pass the nominal selection, described in Table 5.3, and therefore this validation region can test the nominal electron fake factors.

The composition of the background in the $e^{\pm}e^{\pm}$ non-prompt validation region is given in Table 5.9. The non-prompt background is dominant in all validation regions and accounts for 89.2%, 81.9%, 83.4% and 57.0% of the expected predictions in the Weak isolation on both electrons, Weak isolation on the leading electron, Weak isolation on the subleading electron and Fail-ID electron regions, respectively. The invariant mass distributions of $e^{\pm}e^{\pm}$ pairs in these regions are shown in Figure 5.16 and the numbers of total predicted and observed electron pairs are collected in Table 5.15. The observed number of pairs is compatible with the prediction in each region, with the maximum difference in the Fail-ID electron region equal to 0.7 times the total uncertainty on the prediction.

Four validation regions with $\mu^{\pm}\mu^{\pm}$ pairs are defined to test the non-prompt muon predictions:

Weak isolation on both muons : the muons fail the nominal muon isolation but pass a looser definition as described in Table 5.11.

Weak isolation on the leading muon : the muon with higher p_{T} fails the nominal muon isolation but passes a looser definition as described in Table 5.11. The muon with lower p_{T} passes the nominal selection, described in Table 5.3. Since one of the muons passes the nominal selection, this validation region can test the nominal muon fake factors.

Weak isolation on the subleading muon : the muon with lower p_{T} fails the nominal muon isolation but passes a looser definition as described in Table 5.11. The muon

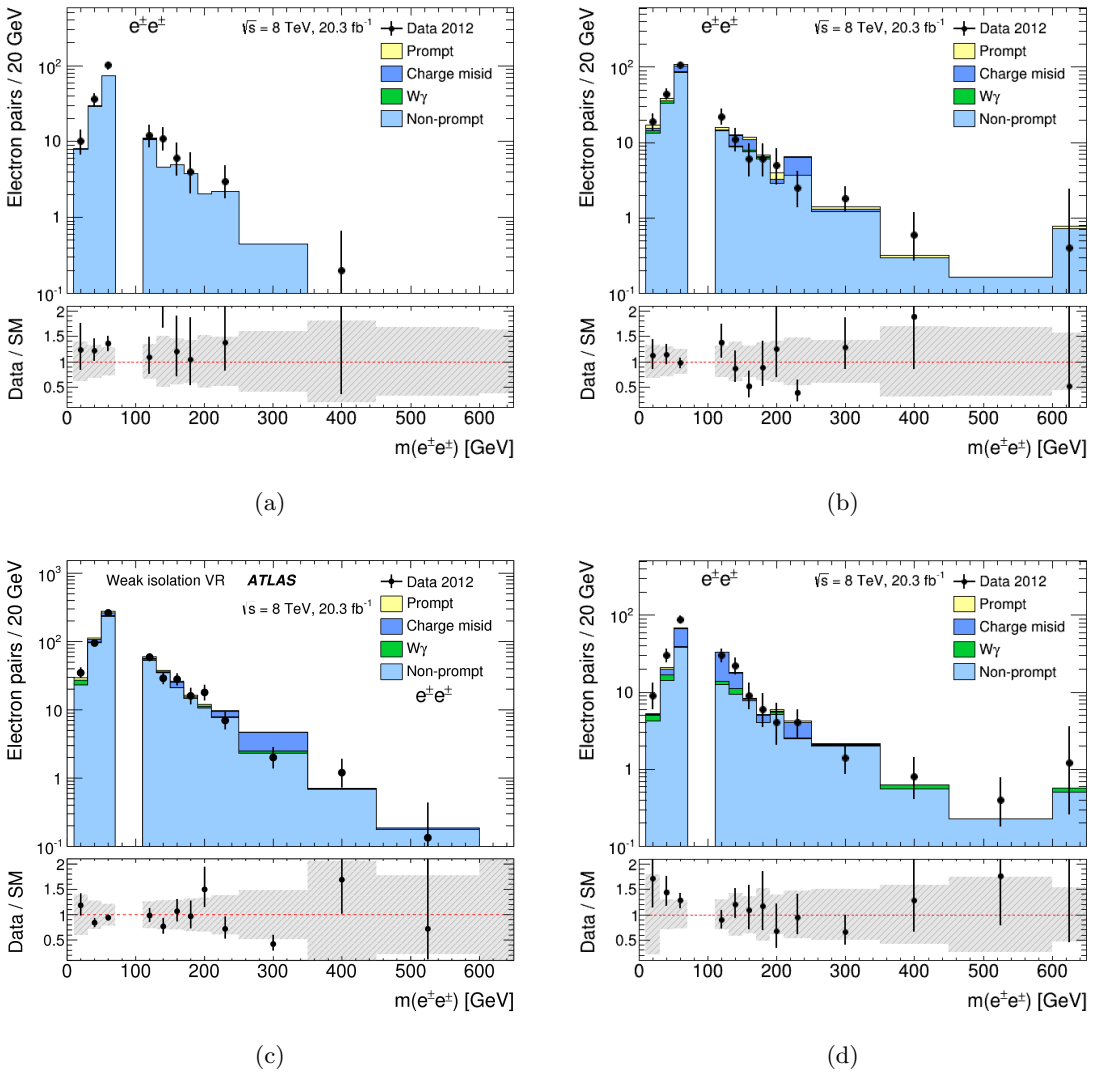


Figure 5.16: Invariant mass distributions of $e^{\pm}e^{\pm}$ pairs in the non-prompt electron validation regions. The Weak isolation selection is applied to both electrons in (a), to the leading electron in (b), to the subleading electron in (c), while the Fail-ID selection is applied to electrons in (d). The mass range between 70 GeV and 110 GeV is omitted because it is not included in the signal region. Events in data (black dots) are compared to the expected stacked contributions from Standard Model processes, as described in Section 5.4. Overflow events are added to the last bin. The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the predictions. Note that the bin width is variable but the contents of all bins are normalized as if they are 20 GeV wide.

with higher p_{T} passes the nominal selection, described in Table 5.3, and therefore this validation region can test the nominal muon fake factors.

Fail-d_0 muon : at least one muon fails the nominal muon d_0 impact parameter requirement but passes a looser definition as described in Table 5.13. One muon may pass the nominal selection, described in Table 5.3, and therefore this validation region can test the nominal muon fake factors.

The compositions of the background in these regions are given in Table 5.9. The non-prompt background is dominant in all validation regions and accounts for 99.0%, 79.0%, 84.1% and 81.0% of the expected predictions in the Weak isolation on both muons, Weak isolation on the leading muon, Weak isolation on the subleading muon and Fail-d_0 muon regions, respectively. The invariant mass distributions for $\mu^{\pm}\mu^{\pm}$ pairs in these regions are shown in Figure 5.17 and the numbers of total predicted and observed lepton pairs are collected in Table 5.15. The observed number of pairs is compatible with the prediction in each region, with the maximum difference in the Weak isolation on the subleading muon region equal to 0.5 times the total uncertainty on the prediction.

Finally, four validation regions with $e^{\pm}\mu^{\pm}$ pairs are defined to test the non-prompt muon and the non-prompt electron predictions:

Weak isolation on the electron : the electron fails the nominal electron isolation but passes a looser definition as described in Table 5.11. The muon passes the nominal selection described in Table 5.3. Since the muon passes the nominal selection, this validation region can test the nominal muon fake factors.

Weak isolation on the muon : the muon fails the nominal muon isolation but passes a looser definition as described in Table 5.11. The electron passes the nominal selection described in Table 5.3. Since the electron passes the nominal selection, this validation region can test the nominal electron fake factors.

Fail-d_0 muon : the muon fails the nominal muon d_0 impact parameter requirement but passes a looser definition as described in Table 5.13. The electron passes the nominal selection described in Table 5.3 and therefore this validation region can test the nominal electron fake factors.

Low-p_{T} muon : the muon fails the nominal $p_{\mathrm{T}} > 20$ GeV requirement and has $18 < p_{\mathrm{T}} < 20$ GeV as described in Table 5.14. The electron passes the nominal selection described in Table 5.3 and therefore this validation region can test the nominal electron fake factors.

The compositions of the background in these regions are given in Table 5.9. The non-prompt background is dominant in all validation regions and accounts for 71.5%, 86.3%, 79.1% and 44.7% of the expected predictions in the Weak isolation on the electron, Weak

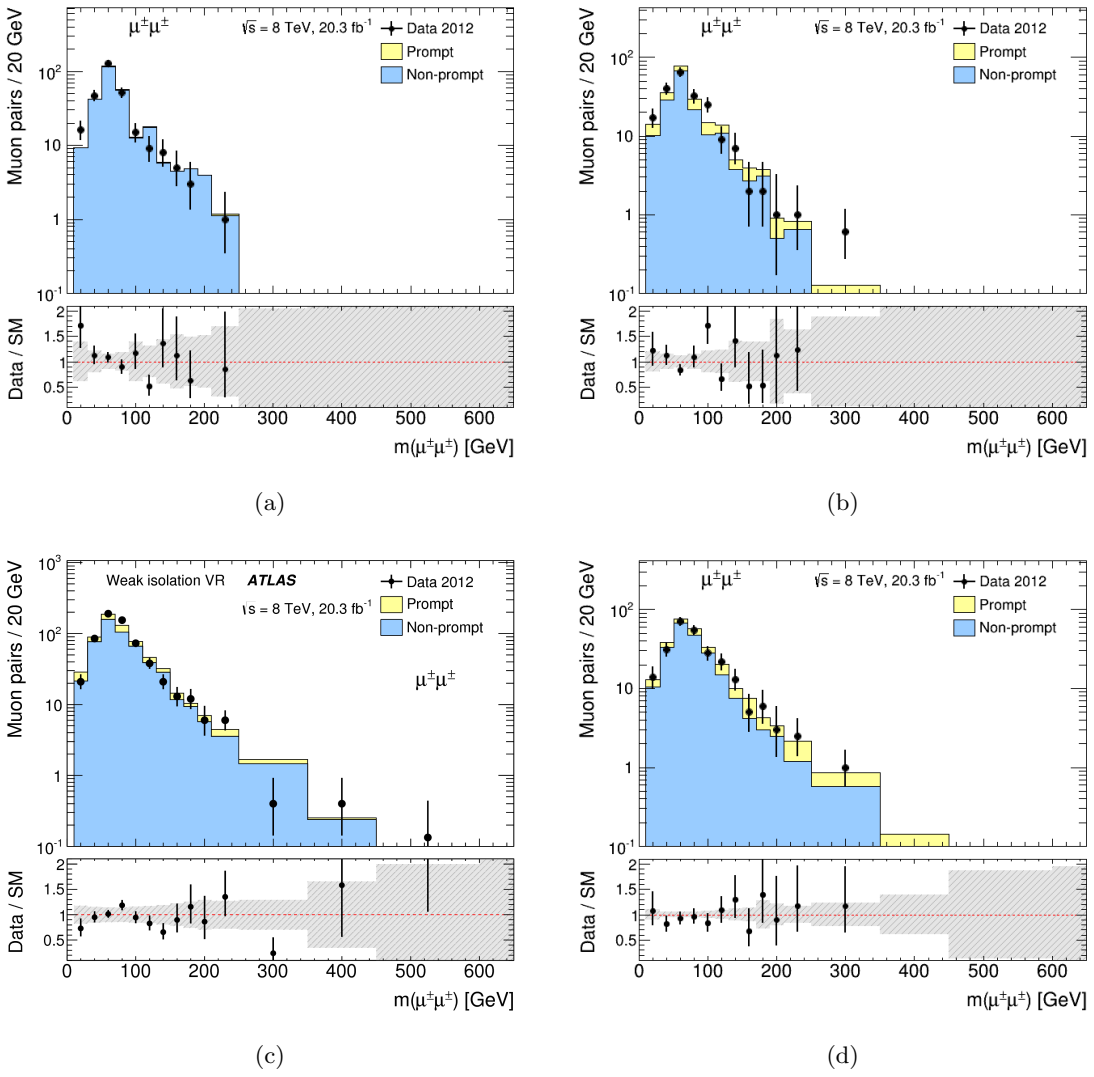


Figure 5.17: Invariant mass distributions of $\mu^{\pm}\mu^{\pm}$ pairs in the non-prompt muon validation regions. The Weak isolation selection is applied to both muons in (a), to the leading muon in (b), to the subleading muon in (c), while the Fail-d_0 selection is applied to one of the muons in (d). Events in data (black dots) are compared to the expected stacked contributions from Standard Model processes, as described in Section 5.4. Overflow events are added to the last bin. The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the predictions. Note that the bin width is variable but the contents of all bins are normalized as if they are 20 GeV wide.

isolation on the muon, Fail-d_0 muon and Low-p_{T} muon regions, respectively. The invariant mass distributions for $e^{\pm}\mu^{\pm}$ pairs in these regions are shown in Figure 5.18 and the numbers of total predicted and observed lepton pairs are collected in Table 5.15. The observed number of pairs is compatible with the prediction in each region, with the maximum difference in the Fail-d_0 muon region equal to 1.7 times the total uncertainty on the prediction.

5.4.5.5 Validation regions summary

The same-sign lepton pairs expected for Standard Model processes and observed in data for each validation region are collected in Table 5.15. The agreement between data and prediction is typically within the error. The largest difference is observed in the $\mu^{\pm}\mu^{\pm}$ Fail-d_0 muon validation region and is 1.7 times the uncertainty on the prediction. The estimate of all the Standard Model background components is therefore considered realistic.

	Number of electron pairs		
Validation region	Predictions	Data	Difference/Error
Opposite-sign electrons	4740000 ± 330000	4895830	-0.5
Prompt electrons	275 ± 23	268	$+0.3$
Charge misID Z peak electron	12700 ± 1300	11793	$+0.7$
Weak isolation on both electrons	280 ± 130	285	0.0
Weak isolation on leading electron	190 ± 60	224	-0.6
Weak isolation on subleading electron	620 ± 120	574	$+0.4$
Fail-ID electron	195 ± 32	217	-0.7
	Number of electron–muon pairs		
Validation region	Predictions	Data	Difference/Error
Opposite-sign leptons	70400 ± 4700	71771	-0.3
Prompt leptons	950 ± 60	1001	-0.8
Weak isolation on electron	750 ± 150	965	-1.4
Weak isolation on muon	790 ± 130	800	-0.1
Fail-d_0 muon	249 ± 19	216	$+1.7$
Low-p_{T} muon	211 ± 12	201	$+0.8$
	Number of muon pairs		
Validation region	Predictions	Data	Difference/Error
Opposite-sign muons	8144000 ± 10000	8216983	-0.7
Prompt muons	651 ± 43	714	-1.5
Weak isolation on both muons	280 ± 40	283	-0.1
Weak isolation on leading muon	199 ± 25	199	0.0
Weak isolation on subleading muon	697 ± 90	652	$+0.5$
Fail-d_0 muon	250 ± 31	255	-0.2

Table 5.15: Expected and observed numbers of lepton pairs in each validation region. The uncertainties on the predictions include statistical and systematic uncertainties. The column "Difference/Error" is calculated by dividing the difference between the prediction and data by the uncertainty of the prediction.

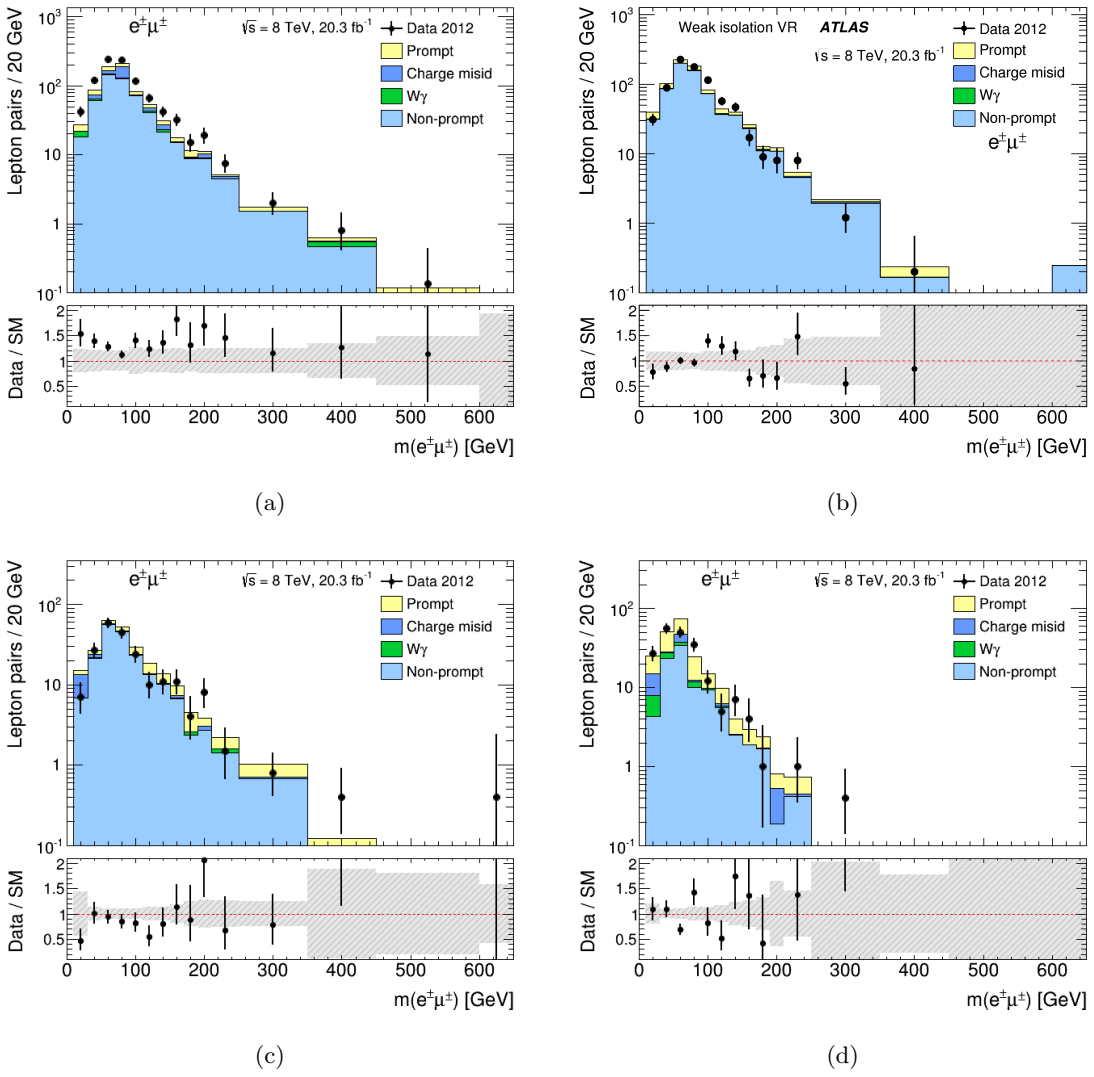


Figure 5.18: Invariant mass distributions of $e^{\pm}\mu^{\pm}$ pairs in the non-prompt lepton validation regions. The Weak isolation selection is applied to the electron in (a) and to the muon in (b). The Fail-d_0 selection is applied to the muon in (c) while the Low-p_{T} selection is applied to the muon in (d). Events in data (black dots) are compared to the expected stacked contributions from Standard Model processes, as described in Section 5.4. Overflow events are added to the last bin. The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the predictions. Note that the bin width is variable but the contents of all bins are normalized as if they are 20 GeV wide.

5.5 Systematic uncertainties

The systematic uncertainties affecting the predictions of Standard Model backgrounds and of new physics models can be classified as: theoretical, statistical, experimental or data-driven uncertainties.

Theoretical uncertainties. For each simulated process the uncertainties on the production cross-section and the modeling of the final state are considered. The renormalization and factorization scales in the fixed-order QCD calculations are varied by a factor of two and one half relative to their nominal values. The PDF and the value of the strong coupling constant α_s are also varied to determine uncertainties following dedicated procedures [158]. The modeling of specific processes in simulation is tested with different generators, parton shower and hadronization models. These uncertainties are summed together into an overall uncertainty on the production cross-section of the process. The uncertainties on the production cross-section for each simulated process are given in Table 5.16. Variations on the selection efficiency are not considered, since the sources of uncertainty considered affect mainly the hadronic activity in the event which is not particularly relevant for the selection applied in this analysis. For the $W^{\pm}W^{\pm}$ and the multi parton diboson production (MPI VV) the cross-section uncertainty is conservatively set to 50% and 100% respectively, since at the time of the analysis no experimental measurement of these processes were sensitive enough to constrain the cross-section value. The cross-section uncertainty for pair production of $H_{\mathrm{L}}^{\pm\pm}$ is 10-15% depending on the mass of the particle [152]. No theoretical uncertainty is evaluated for the other new physics signals considered because the analysis procedure used is not affected by the uncertainty on the cross-section of these processes.

Simulation statistical uncertainties. The number of events generated for processes modeled with simulation is at least one order of magnitude larger than what is expected in 20.3 fb^{-1} and is usually large enough to make the statistical uncertainty on predictions small compared to other uncertainties (less than 3% for $m(\ell^{\pm},\ell^{\pm}) > 15$ GeV). However, when small regions of the phase space are considered, like the $m(\ell,\ell) > 500$ GeV region, the statistical uncertainty can be dominant. The largest statistical uncertainty on the background prediction is 40% for $m(e^{\pm},e^{\pm}) > 600$ GeV.

Experimental uncertainties. These account for the finite accuracy in the simulation of the collisional conditions of the LHC as well as the event acquisition and particle detection. The integrated luminosity of proton-proton collisions at the ATLAS interaction point is measured with a 2.8% uncertainty [89]. The modeling of pileup conditions is described in Section 2.3. An uncertainty is applied based on the variation of the predictions due to the change of the mean value of the average number of interactions per bunch crossing within

its uncertainty. This uncertainty is smaller than 1%.

The uncertainties associated to the lepton reconstruction and the event trigger have been described in Section 2.4.5 and are propagated to the predictions in the signal regions. The largest uncertainties come from the electron identification efficiencies and from the efficiency for lepton pairs to trigger the acquisition of the event. For the simulation of some new physics samples, including $H_{\mathrm{L}}^{\pm\pm}$ pair production, the fast simulation [130] of the particle interaction with the detector has been used. The main difference compared to the standard simulation is the use of a parametrisation of the calorimeter response, which generally results in a reduced accuracy on electron and hadron reconstruction. A few $H_{\mathrm{L}}^{\pm\pm}$ pair production samples have been generated with both fast and full simulation. The average difference in number of same-sign lepton pairs in each flavour category is used as a systematic uncertainty on the predictions from fast simulated samples. An uncertainty of 2.9% (1.6%) is applied to fast simulation in the $e^{\pm}e^{\pm}$ ($e^{\pm}\mu^{\pm}$) category. The uncertainty for the $\mu^{\pm}\mu^{\pm}$ category is 0.6%.

Data-driven uncertainties. The uncertainty on the non-prompt background prediction with the Fake Factor Method has two components. Part of the uncertainty is due to the error on the fake factor estimate, and is described in Section 5.4.4.2 and Section 5.4.4.3 for electron and muon fake factors respectively. In addition, the uncertainties on the predictions of non-prompt lepton pairs with *non-tight* leptons, $N_{\ell\not{\ell}}$ and $N_{\not{\ell}\not{\ell}'}$, used in Equation 5.11, are accounted for. The uncertainties considered are the statistical error on data and the systematic uncertainties on the subtracted backgrounds.

For the charge misID and the photon conversion backgrounds, the uncertainties on the charge misID rate described in Section 5.4.2.4 are considered in addition to the other uncertainties applied to simulated samples.

Process	Production cross-section uncertainty	
$W^{\pm}Z$	7%	[159]
ZZ	5%	[159]
$t\bar{t}W^{\pm}$, $t\bar{t}Z$	22%	[160–162]
$W^{\pm}W^{\pm}$	50%	
MPI VV	100%	
Z/γ^{*}	7%	[163]
$W^{+}W^{-}$	7%	[159]
$t\bar{t}$, $tW^{\pm}$	5%	[164–170]
$W^{\pm}\gamma$	14%	

Table 5.16: Systematic uncertainties on the production cross-sections of the simulated backgrounds.

The impact of the sources of uncertainty considered on the predictions of $H_{\mathrm{L}}^{\pm\pm}$ pair production and of Standard Model background categories is shown in Table 5.17. When

sources of systematic uncertainties differently affect these processes a range of uncertainties is given. Note that the uncertainties are relative to the prediction for the processes specified and not for the total Standard Model background. Table 5.18, Table 5.19 and Table 5.20, instead, list the uncertainties on the total Standard Model predictions in the $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ categories, respectively. The uncertainties are given for several $m\,(\ell^{\pm},\ell^{\pm})$ ranges as these are used in the statistical analysis of the results.

The Fake Factor Method uncertainty is dominant in all categories. For large masses a significant uncertainty arises from the limited statistics in simulation and theoretical uncertainties. The former is particularly large in the $e^{\pm}e^{\pm}$ category because of the low numbers of simulated charge misID pairs in this region; the latter is mainly due to the larger fraction of $W^{\pm}W^{\pm}$ production expected as prompt backgrounds, with 50% uncertainty.

Systematic uncertainties on different physics processes from the same source are considered 100% correlated. For instance, the charge misID rate uncertainties for the Z/γ^{*}, $t\bar{t}$, $W^{+}W^{-}$ and $W^{\pm}\gamma$ are 100% correlated. The experimental uncertainties are correlated, while the theoretical and simulation statistical uncertainties are independent.

Source	Processes affected	Uncertainty		
		$e^{\pm}e^{\pm}$	$e^{\pm}\mu^{\pm}$	$\mu^{\pm}\mu^{\pm}$
Determination of non-prompt background	Non-prompt background	22%	24%	17%
Electron charge misidentification rate	Charge misID background	9%	1.2%	
Photon conversion rate	$W^{\pm}\gamma$	13%	11%	
Electron simulation	Signal and simulated background	2.7-2.9%	1.4-1.7%	
Muon simulation	Signal and simulated background		0.3-0.6%	0.7-1%
Trigger	Signal and simulated background	2.1-2.6%	2.1-2.6%	2.6-2.9%
Fast simulation	Signal	2.9%	1.6%	0.6%
Luminosity	Signal and simulated background	2.8%	2.8%	2.8%
Pileup	Signal and simulated background	0.1-0.4%	0.1-0.7%	0.1%
Simulation statistics	Signal and simulated background	1.6-6%	1-9%	1.3-3%
Theoretical cross-section	Simulated background	5-7%	5-6%	6%

Table 5.17: Percentage uncertainties on the numbers of predicted same-sign pairs for the type of process specified in the second column, for the mass range $m(\ell^{\pm}, \ell^{\pm}) > 15$ GeV. A range of uncertainties is given when the source of systematic uncertainty differently affects the signal, the prompt background, the charge misID background and the photon conversion background estimate.

Source	$m(e^{\pm}, e^{\pm})$			
	> 15 GeV	> 200 GeV	> 400 GeV	> 600 GeV
Determination of non-prompt background	6%	5%	12%	22%
Electron charge misID and Conversion rate	3.4%	3.2%	2.9%	4%
Electron simulation	1.6%	1.5%	1.6%	2.2%
Trigger	1.4%	1.6%	1.7%	2.3%
Luminosity	2.0%	1.9%	1.8%	2.4%
Pileup	0.1%	0.9%	1.2%	5%
Simulation statistics	2.8%	6%	18%	41%
Theoretical cross-section	5%	4%	4%	6%

Table 5.18: Percentage uncertainties on the expected Standard Model background in the $e^{\pm}e^{\pm}$ category for several $m(e^{\pm}, e^{\pm})$ ranges.

Source	$m(e^{\pm}, \mu^{\pm})$			
	> 15 GeV	> 200 GeV	> 400 GeV	> 600 GeV
Determination of non-prompt background	8%	7%	10%	13%
Electron charge misID and Conversion rate	1.0%	1.2%	0.9%	1.9%
Electron simulation	0.8%	0.9%	0.7%	0.9%
Muon simulation	0.2%	0.2%	0.3%	0.3%
Trigger	1.6%	1.8%	1.7%	1.9%
Pileup	0.1%	0.3%	<0.1%	1.2%
Luminosity	1.8%	2.0%	1.9%	2.2%
Simulation statistics	1.4%	2.2%	6%	13%
Theoretical cross-section	3.6%	5%	6%	6%

Table 5.19: Percentage uncertainties on the expected Standard Model background in the $e^{\pm}\mu^{\pm}$ category for several $m(e^{\pm}, \mu^{\pm})$ ranges.

Source	$m(\mu^{\pm},\mu^{\pm})$			
	>15 GeV	>200 GeV	>400 GeV	>600 GeV
Determination of non-prompt background	4%	7%	32%	1.1%
Muon simulation	0.5%	0.6%	0.7%	1.0%
Trigger	2.0%	2.5%	2.5%	2.9%
Luminosity	2.1%	2.5%	2.4%	2.8%
Pileup	<0.1%	0.3%	0.7%	0.9%
Simulation statistics	0.9%	3.3%	9%	25%
Theoretical cross-section	6%	9%	13%	21%

Table 5.20: Percentage uncertainties on the expected Standard Model background in the $\mu^{\pm}\mu^{\pm}$ category for several $m(\mu^{\pm},\mu^{\pm})$ ranges.

5.6 Results

The data sample selected with the requirements presented in Section 5.3 is now analysed. The background due to Standard Model processes that produces same-sign lepton pairs has been extensively studied in Section 5.4. The modeling of each background component, namely prompt, charge misID, non-prompt and photon conversion backgrounds, appears to be accurate as confirmed by the comparison with data in the validation regions defined in Section 5.4.5. The determination and validation of the background predictions is performed before analysing the data events accepted in the signal regions to avoid biases. Once the the background estimate is consolidated, data in the signal regions are analysed and compared to the estimated backgrounds. The analysis signal regions are studied in Section 5.6.1. Data and Standard Model predictions are compared to test the existence of beyond Standard Model processes producing same-sign leptons. Non-resonant anomalous production is studied in Section 5.6.3, while doubly charged resonances are searched for in Section 5.6.4.

5.6.1 Signal regions

Table 5.21 indicates the observed and expected numbers of same-sign lepton pairs in each category and for several $m(\ell^{\pm},\ell^{\pm})$ requirements. The expected pairs for each type of background are also given. The contributions to the background divided according to individual processes can be found in Appendix B (Table B.3, Table B.4 and Table B.5). Figure 5.19 shows the distributions of the invariant mass of same-sign lepton pairs in the

signal regions categorized by lepton flavour. For the $e^{\pm}e^{\pm}$ and $e^{\pm}\mu^{\pm}$ categories, pairs are found with masses up to around 1 TeV and are added to the last mass interval in the corresponding figures. The events with the most massive observed same-sign lepton pair in each category are displayed as recorded by ATLAS in Figure 5.20. The background composition varies widely with the flavour category.

In the $\mu^{\pm}\mu^{\pm}$ category, the prompt background is dominant and constitutes 74% of the total background. The $W^{\pm}Z$ production is the biggest of the prompt backgrounds, but for $m(\mu^{\pm},\mu^{\pm}) > 500$ GeV, ZZ and $W^{\pm}W^{\pm}$ production can be as large as the $W^{\pm}Z$ production.

In the $e^{\pm}e^{\pm}$ category, the dominant sources of same-sign di-electron pairs are charge misID processes accounting for 54% of the total background. Charge misID di-electron pairs are also predicted at very large $m(e^{\pm},e^{\pm})$, since the charge measurement is less accurate for high p_{T} electrons. However, the estimate of this background in this region is affected by a large uncertainty due to the limited simulated statistics. The larger number of events in the $e^{\pm}e^{\pm}$ categories compared to $\mu^{\pm}\mu^{\pm}$, is mainly due to charge misID processes, which do not produce $\mu^{\pm}\mu^{\pm}$ pairs, and secondly to the larger selection of non-prompt electrons compared to non-prompt muons. The prediction of prompt backgrounds in the $\mu^{\pm}\mu^{\pm}$ category instead exceeds that in the $e^{\pm}e^{\pm}$ category. The Standard Model processes considered have the same probability of producing same-sign muons or same-sign electrons. However, due to the higher efficiency of selecting muons compared to electrons, the prompt prediction in the $\mu^{\pm}\mu^{\pm}$ category is higher.

The $e^{\pm}\mu^{\pm}$ category contains large fractions of prompt and non-prompt backgrounds (40% and 35%, respectively) and smaller contributions from charge misID and photon conversion backgrounds. The fraction of prompt processes increases at large $m(e^{\pm},\mu^{\pm})$. The prompt and non-prompt background predictions in this category are larger than in the others. In diboson processes, which constitute a large fraction of prompt background, the probability of producing same-sign leptons with different flavour is twice the probability for same flavour leptons. In processes with only one prompt lepton, such as $W^{\pm}+$jets and $t\bar{t}$, the other reconstructed lepton can be either a non-prompt electron or a non-prompt muon in the $e^{\pm}\mu^{\pm}$ category, but it has to be exactly a non-prompt electron (muon) in the $e^{\pm}e^{\pm}$ ($\mu^{\pm}\mu^{\pm}$) category. This explains the larger backgrounds in this region and the larger number of observed lepton pairs.

The same-sign lepton pairs are also studied divided by total charge (+2 and −2), since Standard Model and new physics processes can be asymmetric in charge. The valence quarks in protons are mainly up type (charge +2/3), thus the final states with positive total charge are favoured. Prompt backgrounds with positive total charge, like $W^{+}Z$, are more likely than those with negative total charge, like $W^{-}Z$. This is true also for non-prompt backgrounds like $W^{\pm}+$ jets production. New physics models can also be asymmetric in the di-lepton charge sign. The production of right-handed W bosons is one of these cases. Table B.6 compares data with predictions separately for $\ell^{+}\ell^{+}$ and $\ell^{-}\ell^{-}$

pairs.

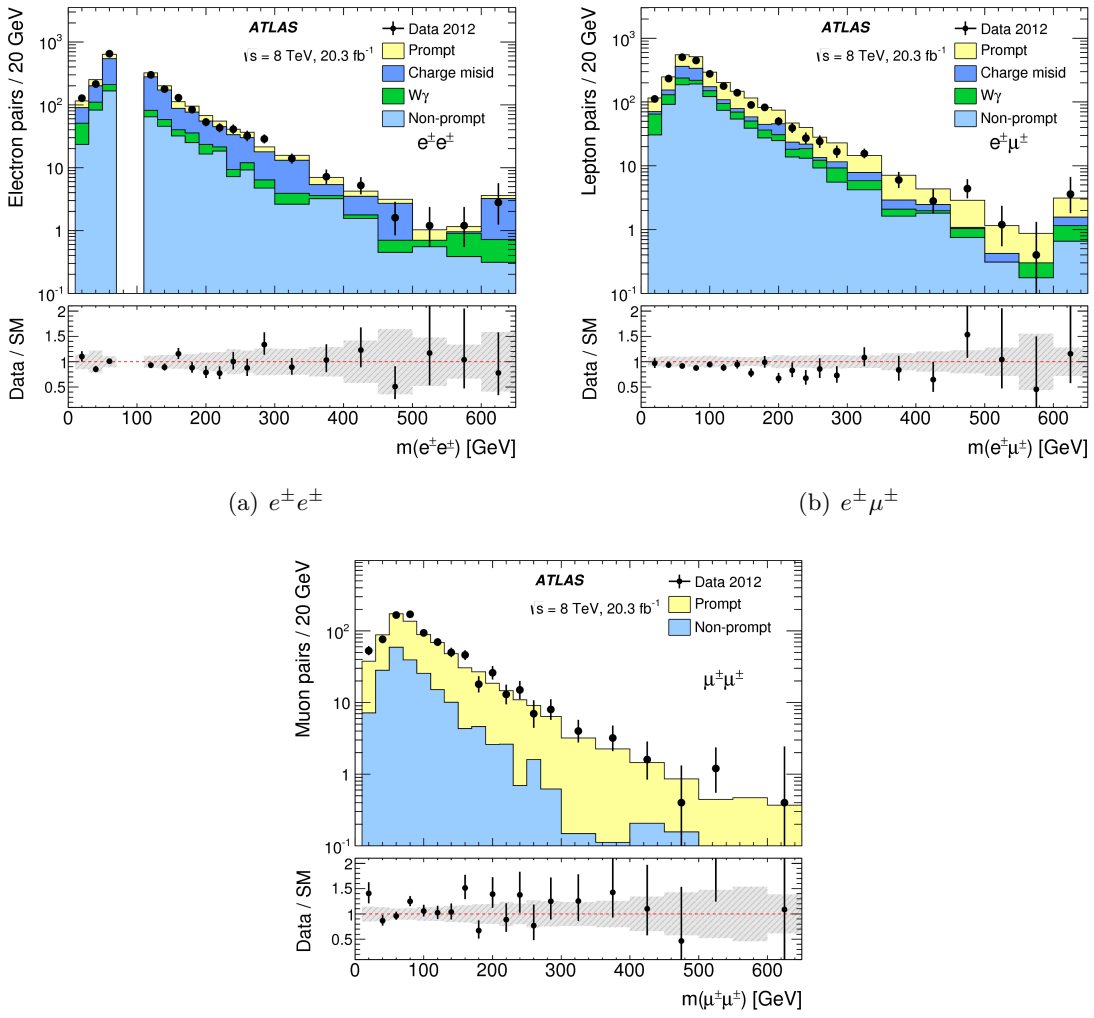
General good agreement is seen in all categories in terms of total yields and modeling of Standard Model processes within the given uncertainties. No significant indication of new physics is observed. The largest difference is found in the $e^{\pm}\mu^{\pm}$ category where the prediction exceeds the observation by 1.8 times the error on the prediction for $m\,(e^{\pm},\mu^{\pm}) > 200$ GeV.

	Number of electron pairs					
$m\,(e^{\pm},e^{\pm})$ [GeV]	Prompt	Non-prompt	Charge misID	Photon Conv.	Total SM	Data
> 15	347 ± 25	520 ± 120	1020 ± 150	180 ± 40	2060 ± 190	1976
> 100	174 ± 14	250 ± 50	550 ± 80	75 ± 16	1050 ± 100	987
> 200	51.5 ± 4.9	72 ± 13	150 ± 27	22 ± 5	296 ± 31	265
> 300	15.7 ± 1.9	23 ± 5	43 ± 12	8.0 ± 2.3	89 ± 14	83
> 400	5.3 ± 0.9	8.1 ± 2.4	16 ± 8	3.8 ± 1.3	33 ± 8	30
> 500	2.3 ± 0.5	3.1 ± 1.5	6 ± 5	2.7 ± 1.0	14 ± 5	13
> 600	0.91 ± 0.28	$0.8^{+1.0}_{-0.8}$	6 ± 5	1.0 ± 0.6	9 ± 5	7
	Number of electron–muon pairs					
$m\,(e^{\pm},\mu^{\pm})$ [GeV]	Prompt	Non-prompt	Charge misID	Photon Conv.	Total SM	Data
> 15	1030 ± 50	910 ± 220	370 ± 40	270 ± 50	2580 ± 240	2315
> 100	458 ± 26	340 ± 80	87 ± 11	104 ± 20	990 ± 90	859
> 200	130 ± 9	79 ± 17	29 ± 4	28 ± 6	265 ± 22	226
> 300	43 ± 5	24 ± 6	9.5 ± 1.9	8.1 ± 2.4	84 ± 8	85
> 400	16.0 ± 2.1	9.2 ± 3.0	2.5 ± 0.8	2.7 ± 1.1	31 ± 4	31
> 500	6.8 ± 1.1	2.8 ± 1.5	1.5 ± 0.4	1.6 ± 0.8	12.6 ± 2.1	13
> 600	3.5 ± 0.7	1.6 ± 1.0	0.9 ± 0.4	1.2 ± 0.7	7.4 ± 1.5	9
	Number of muon pairs					
$m\,(\mu^{\pm},\mu^{\pm})$ [GeV]	Prompt	Non-prompt	Total SM	Data		
> 15	580 ± 40	203 ± 34	780 ± 50	843		
> 100	245 ± 21	56 ± 11	301 ± 24	330		
> 200	67 ± 7	8.7 ± 2.3	76 ± 8	87		
> 300	20.7 ± 2.9	1.9 ± 1.0	22.6 ± 3.1	27		
> 400	7.7 ± 1.5	1.2 ± 0.9	9.0 ± 1.7	9		
> 500	2.9 ± 0.8	$0.32^{+0.41}_{-0.32}$	3.2 ± 0.9	4		
> 600	0.9 ± 0.4	$0.0^{+0.2}_{-0.0}$	0.9 ± 0.4	1		

Table 5.21: Expected and observed numbers of same-sign di-lepton pairs in the $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ categories for various invariant mass intervals. The errors shown are the overall uncertainties on the predictions.

5.6.2 Statistical analysis

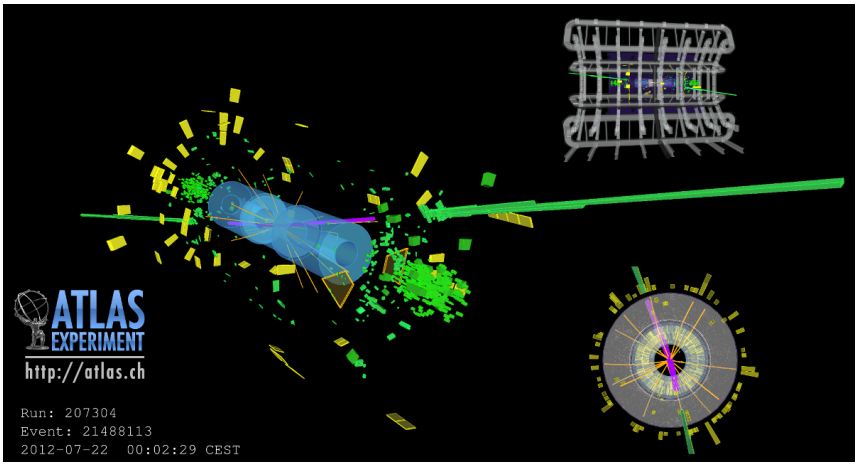
A statistical analysis of data is performed to quantify the compatibility of the observed data with the Standard Model background estimate. The predictions of Standard Model processes and the relative uncertainties have been evaluated for each signal region. Since several processes compose the background and each process is affected by a variety of



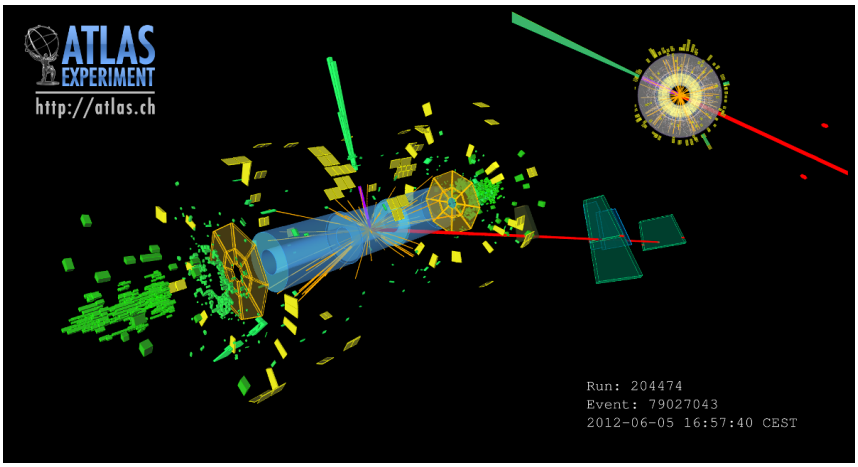
(a) $e^{\pm}e^{\pm}$ (b) $e^{\pm}\mu^{\pm}$

(c) $\mu^{\pm}\mu^{\pm}$

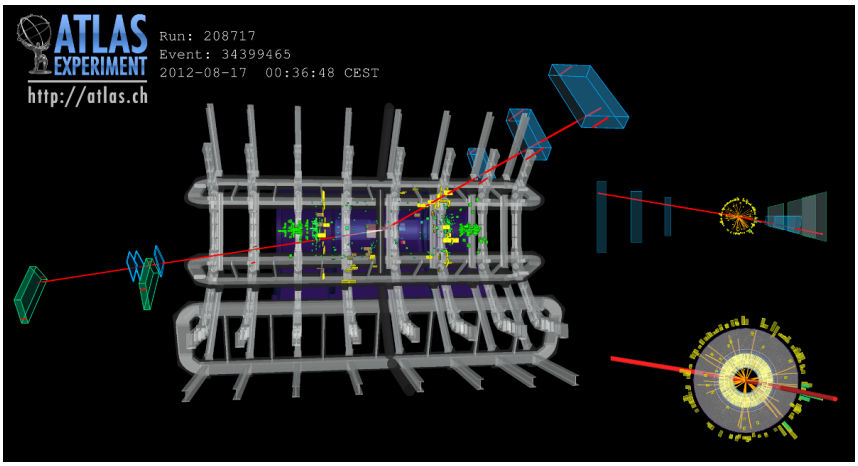
Figure 5.19: Invariant mass distributions of (a) $e^{\pm}e^{\pm}$ (b) $e^{\pm}\mu^{\pm}$ and (c) $\mu^{\pm}\mu^{\pm}$ pairs in the signal regions. The mass range between 70 GeV and 110 GeV is not included in the $e^{\pm}e^{\pm}$ category since this region is highly affected by charge misID background and is used for its estimate. Events in data (black dots) are compared to the expected stacked contributions from Standard Model processes, as described in Section 5.4. Overflow events are added to the last bin. The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the predictions. Note that the bin width is variable but the contents of all bins are normalized as if they are 20 GeV wide.



(a) $e^{\pm}e^{\pm}$



(b) $e^{\pm}\mu^{\pm}$



(c) $\mu^{\pm}\mu^{\pm}$

Figure 5.20: Displays of data events with same-sign leptons collected by ATLAS. The events have one reconstructed $e^{\pm}e^{\pm}$ (a), $e^{\pm}\mu^{\pm}$ (b) and $\mu^{\pm}\mu^{\pm}$ (c) pair with invariant masses of 964 GeV, 736 GeV and 628 GeV, respectively. The electrons are represented by purple tracks and muons by red tracks. The blue boxes show the modules of the muon spectrometer activated and the green and yellow boxes display the magnitudes of the energy deposits in the electromagnetic calorimeter and hadronic calorimeter, respectively. Additional charged particles reconstructed in the Inner Detector are shown in orange.

systematic uncertainties, a statistical model is built to allow the coherent use of all available information.

5.6.2.1 Statistical model

The statistical model is built with HistFactory [171], and can handle:

- multiple processes (signals and backgrounds);
- multiple regions of estimate (categories);
- unconstrained normalization parameters;
- variations in normalization due to systematic uncertainties;
- variations in shape of the process distributions due to systematic effects;
- variations due to the statistical uncertainties on the predicted processes;
- correlation of variations among multiple categories and processes.

HistFactory has also been used to define the statistical model of the analysis described in the next chapter. In the following, a generic description of the statistical model is given. The properties of the statistical model specific to this analysis are discussed at the end of the section. Since the analyses used in this thesis study the number of events or the number of di-lepton pairs in specific categories, and not the kinematic properties of the events in these regions, the handling of shape variations is omitted.

The statistical model is composed of the probability density functions associated with predictions in the considered categories. The core of the model is formed by Poisson probability functions that are combined with other probability density functions to reflect the *a priori* knowledge of the predicted events.

For each category c of the analysis and each process s, the number of predicted events $\nu_{cs}\left(\vec{\alpha}_{cs}\right)$ is expressed via a set of parameters $\vec{\alpha}_{cs}$. These parameters model the way in which the predictions change under variations of the quantities associated to them, such as luminosity, cross-sections, reconstruction efficiencies, charge misID efficiencies and fake factors. Some of the parameters in $\vec{\alpha}_{cs}$ can be free (Parameters of Interest or PoIs), while others are constrained by auxiliary measurements (Nuisance Parameters or θ). The models used in this thesis have only one PoI that is indicated with μ. The PoI is the quantity that is investigated in the analysis. For instance, μ could be the cross-section of the signal targeted by the analysis.

The Nuisance Parameters are used to account for systematic uncertainties on the predictions. When an uncertainty is correlated among processes or among categories, a single parameter is used to model the variation of the predictions in each category and for each process associated with that uncertainty. For instance, the luminosity uncertainty affects

each simulated process and each category in the same way, therefore one nuisance parameter is used for it; on the other hand, the statistical uncertainty on a simulated event for a specific process and a specific category is not correlated with the simulated statistical uncertainty in other processes and categories. The model therefore uses one Nuisance Parameter associated to the statistical uncertainty for each region and each process.

The model of the total expected production in a category c is obtained by summing up the process predictions ν_{cs}:

$$\nu_c\left(\vec{\alpha}_c\right) = \sum_{s\in\text{processes}} \nu_{cs}\left(\vec{\alpha}_{cs}\right) \tag{5.16}$$

where $\vec{\alpha}_c = \bigcup_{s\in\text{processes}} \vec{\alpha}_{cs}$. The probability of observing n_c events in the category c when $\nu_c\left(\vec{\alpha}_c\right)$ are expected follows the Poisson statistics:

$$\text{Pois}\left(n_c|\nu_c\left(\vec{\alpha}_c\right)\right) = \frac{\left(\nu_c\left(\vec{\alpha}_c\right)\right)^{n_c} e^{-\nu_c(\vec{\alpha}_c)}}{n_c!} \tag{5.17}$$

Since the predictions are affected by uncertainties, these have to be reflected in the model. The probability density function in Equation 5.17 is corrected with functions f_p of the Nuisance Parameters θ_p to include information on the uncertainties and values of these parameters obtained from auxiliary measurements. Each θ_p, where p is the index of the Nuisance Parameter, has an associated f_p, where the type of function depends on the type of uncertainty associated with the Nuisance Parameter. The function is different for experimental uncertainties, data-driven uncertainties, theoretical uncertainties and simulated statistical uncertainties.

For experimental systematic uncertainties and data-driven uncertainties the measured value a_p of the Nuisance Parameter θ_p is expected to be distributed as a Gaussian with width σ_p equal to the uncertainty on the parameter, as estimated from the auxiliary measurement:

$$G(a_p|\theta_p,\sigma_p) = \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp\left[-\frac{(a_p-\theta_p)^2}{2\sigma_p^2}\right] \tag{5.18}$$

The Nuisance Parameter can assume either positive or negative values.

The theoretical uncertainties affect the production cross-section of processes. Sometimes these uncertainties can be as large as 100%. If a Gaussian constraint is used, the Nuisance Parameter could assume a value such that the predicted cross-section, and consequently the predicted yields, would be negative. This has a non-physical meaning and therefore must be avoided. A Gaussian truncated at negative values can lead to numerical instabilities, hence a log-normal probability function is used in this case:

$$P_{LN}(n_p|\theta_p,\sigma_p) = \frac{1}{\sqrt{2\pi}\ln k_p}\frac{1}{n_p} \exp\left[-\frac{\ln(n_p/\theta_p)^2}{2(\ln k_p)^2}\right] \tag{5.19}$$

where $\ln n_p$ and $\ln k_p$ are proportional to the theoretical estimate and its uncertainty, here called θ_p and σ_p for consistency. This function avoids negative predictions and for small uncertainties assumes the form of the Gaussian pdf.

The events used to form predictions for a given process are scaled with several correction factors, like pileup, reconstruction efficiency and others. These are event-by-event corrections, thus the Poisson distribution cannot represent the probability distribution associated to the statistics of the predicted events. The Gamma distribution, which is an extension of the Poisson distribution to real numbers, is used instead:

$$\text{Gamma}(a_p|\theta_p) = \frac{\theta_p^{a_p} e^{-\theta_p}}{\Gamma(a_p+1)} \tag{5.20}$$

The probability of observing n_c events in category c is given by the product of the Poisson distribution and the constraint functions for the Nuisance Parameters:

$$p(n_c|\vec{\alpha}_c) = \text{Pois}\left(n_c|\nu_c\left(\vec{\alpha}_c\right)\right) \prod_{p\in\text{parameters}} f_p \tag{5.21}$$

f_p can be either $G(a_p|\theta_p,\sigma_p)$, $P_{LN}(n_p|\theta_p,k_p)$ or $\text{Gamma}(a_p|\theta_p)$ depending on the type of uncertainty associated with the parameter p.

If more event categories are considered, a combined model with a Poisson distribution for each category is used:

$$p(\vec{n}|\vec{\alpha}) = \prod_{c\in\text{categories}} \text{Pois}\left(n_c|\nu_c\left(\vec{\alpha}\right)\right) \prod_{p\in\text{parameters}} f_p \tag{5.22}$$

The vectors $\vec{\alpha}$ contain the Parameter of Interest (μ), the Nuisance Parameters (θ) from each category, $\vec{n} = [n_1, n_2, ..., n_K]$, are the numbers of events observed in each category, and K is the number of categories. The likelihood function for the set of parameters (μ,θ) is obtained from the probability density function with given observed events in each category $\vec{n}$:

$$L(\mu,\theta) = L(\vec{\alpha}|\vec{n}) = p(\vec{n}|\vec{\alpha}) \tag{5.23}$$

In this analysis, the likelihood $L(\mu,\theta)$ is built to investigate the number of di-lepton pairs in the signal regions. The signal regions, separated by lepton flavour, are not combined. A likelihood is used for each di-lepton invariant mass interval and each flavour category considered. The PoI of this analysis is the number of pairs due to non-Standard Model processes in the category considered. The uncertainties described in Section 5.5 are accounted for with θ. The experimental uncertainties are correlated among simulated processes; charge misID uncertainties are correlated among charge misID and photon conversion processes.

5.6.2.2 Test Statistic

The compatibility of the predictions for certain μ values with the observed data is quantified by the test statistic based on the profile likelihood ratio method [172]. The profile likelihood ratio $\lambda(\mu)$ is the ratio between the conditional maximum likelihood fit $L(\mu, \hat{\hat{\theta}}(\mu))$ and the unconditional maximum likelihood fit $L(\hat{\mu}, \hat{\theta})$. In the unconditional fit, θ and μ vary simultaneously and assume values that maximize the likelihood: $\hat{\mu}, \hat{\theta}$. In the conditional fit, μ is fixed and θ varies to the set of values $\hat{\hat{\theta}}(\mu)$ that maximize the likelihood for the given value of μ. The resulting profile likelihood ratio is a function that depends only on μ:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})} \tag{5.24}$$

In this thesis μ represents the magnitude of a signal (either a new physics process or $t\bar{t}H$ production), hence it has physical meaning only when positive. $\lambda(\mu)$ is modified to account for downward fluctuations of the observed events:

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})} & \hat{\mu} \geq 0 \\ \\ \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(0, \hat{\hat{\theta}}(0))} & \hat{\mu} < 0 \end{cases} \tag{5.25}$$

The calculation of the significance of an excess of data over the background uses the test statistic:

$$\tilde{q}_0 = \begin{cases} -2 \ln \tilde{\lambda}(0) & \hat{\mu} > 0 \\ \\ 0 & \hat{\mu} \leq 0 \end{cases} \tag{5.26}$$

The value of $\tilde{q}_0$ increases with the level of incompatibility between the data and the background-only hypothesis. Given the probability density function of $\tilde{q}_0$, $f(\tilde{q}_0)$, the probability of the test statistic being equal or greater than the observed value, $\tilde{q}_0^{\text{obs}}$, is called p_0 and is given by:

$$p_0 = \int_{\tilde{q}_0^{\text{obs}}}^{\infty} f(\tilde{q}_0) d\tilde{q}_0 \tag{5.27}$$

$f(\tilde{q}_0)$ is a known parametric function under certain hypotheses (asymptotic approximation) and can be determined via the "Asimov" data set [172]. The p_0 is also expressed in terms of Gaussian standard deviations σ. In particle physics the standard for the evidence of an excess is to have at least 3σ significance.

To evaluate the validity of the signal-plus-background null hypothesis for a given value

of μ, the test statistics $\tilde{q}_\mu$ is built according to:

$$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ \\ 0 & \hat{\mu} > \mu \end{cases} \tag{5.28}$$

The value of $\tilde{q}_\mu$ is higher when the model for a given μ is less compatible with data. Note that $\tilde{q}_0$ is zero if the data fluctuate downward, but $\tilde{q}_\mu$ is zero if the data fluctuate upward.

The probability that the $\tilde{q}_\mu$ test statistic for a given μ assumes a value equal to or higher than the observed value, $\tilde{q}_\mu^{\text{obs}}$, is:

$$p_\mu = \int_{\tilde{q}_\mu^{\text{obs}}}^{\infty} f(\tilde{q}_\mu) d\tilde{q}_\mu \tag{5.29}$$

The probability density function of $\tilde{q}_\mu$, $f(\tilde{q}_\mu)$, is also known under certain hypotheses [172].

If no significant excess is present in data, upper limits can be set on μ as described in the next section.

5.6.2.3 The CL_S method

The upper limit at 95% Confidence-Level (CL) is defined as the highest value of μ for which p_μ is equal to 5%. In other words, it is the maximum value of μ for which the signal-plus-background hypothesis is compatible with the observed events with a probability higher than 5%.

When data fluctuates downward below the expected background, this upper limit on μ can assume an arbitrarily small value beyond the sensitivity of the measurement. This problem is solved by the use of limits calculated with the CL_S method [173]. The value of the limit is given by

$$CL_S(\mu) = \frac{p_\mu}{1 - p_b} \tag{5.30}$$

The numerator is the probability associated to the signal-plus-background hypothesis for a given μ, while the denominator quantifies the probability of the background-only hypothesis, where p_b is given by:

$$p_b = 1 - \int_{\tilde{q}_{\mu=0}^{obs}}^{\infty} f(\tilde{q}_{\mu=0}) d\tilde{q}_{\mu=0} \tag{5.31}$$

The 95% CL_S upper limit is given by the value of μ for which $CL_S(\mu) = 0.05$. In case of downward fluctuations, the observed results would be compatible with neither the background-plus-signal nor the background-only hypotheses, but the ratio of the two probabilities would be seizable, and by product the upper limit. For upwards fluctuations the CL_S upper limit is similar to the CL upper limit.

5.6.3 Inclusive search for new physics

The observations in the signal regions are interpreted as 95% CL_S upper limits on the number of same-sign lepton pairs due to non-Standard Model processes. The pairs are divided according to lepton flavour and di-lepton invariant mass intervals: $m(\ell^{\pm},\ell^{\pm}) > m_{th}$ where $m_{th} \in [15, 100, 200, 300, 400, 500, 600]$ GeV. The limits for each of these categories are shown in Table 5.22 and are the statistical combinations of the limits for pairs with positive total charge and those with negative total charge. The observed and expected limits agree within 2σ.

	95% CL_S upper limit					
	$e^{\pm}e^{\pm}$		$e^{\pm}\mu^{\pm}$		$\mu^{\pm}\mu^{\pm}$	
Mass range	Expected	Observed	Expected	Observed	Expected	Observed
$m(\ell^{\pm},\ell^{\pm}) > 15$ GeV	$382^{+101(269)}_{-126(164)}$	312	$410^{+49(223)}_{-81(145)}$	294	$110^{+39(83)}_{-31(59)}$	151
$m(\ell^{\pm},\ell^{\pm}) > 100$ GeV	$185^{+61(136)}_{-55(66)}$	135	$152^{+55(91)}_{-25(69)}$	119	$55^{+20(46)}_{-21(30)}$	78
$m(\ell^{\pm},\ell^{\pm}) > 200$ GeV	$66^{+26(51)}_{-16(25)}$	52	$51^{+19(38)}_{-9(25)}$	34	$22^{+8(20)}_{-7(14)}$	27
$m(\ell^{\pm},\ell^{\pm}) > 300$ GeV	$32^{+13(29)}_{-4(16)}$	32	$25^{+10(21)}_{-7(12)}$	27	$12^{+5(10)}_{-1(5)}$	16
$m(\ell^{\pm},\ell^{\pm}) > 400$ GeV	$20^{+7(15)}_{-2(4)}$	20	$15^{+4(12)}_{-5(7)}$	16	$7.8^{+3.0(6.8)}_{-1.9(2.8)}$	8.5
$m(\ell^{\pm},\ell^{\pm}) > 500$ GeV	$12^{+4(9)}_{-3(3)}$	11	$10^{+3(10)}_{-3(5)}$	11	$5.0^{+1.8(4.6)}_{-1.1(1.8)}$	7.6
$m(\ell^{\pm},\ell^{\pm}) > 600$ GeV	$9.7^{+3.3(6.2)}_{-2.0(2.6)}$	10.0	$7.9^{+2.5(5.8)}_{-2.8(3.5)}$	9.3	$4.1^{+1.1(3.1)}_{-0.6(1.7)}$	4.5

Table 5.22: The 95% CL_S upper limit on the number of same-sign lepton pairs due to non-Standard Model physics in each signal region and for several invariant mass ranges. The 1σ (2σ) errors on the expected limit are also given.

These limits are independent of any theoretical assumption on new physics models. The format of these results is, however, not practical for testing new physics models. Each process would need to be simulated following the chain of events explained in Section 2.3, and the event selection would need to be applied to obtain the expected number of same-sign pairs for that model in the signal regions. The process of testing can be simplified if the limits are independent of experimental effects. This is achieved by the definition of a fiducial selection and ultimately the calculation of a fiducial cross-section. The fiducial selection is defined for generator level particles and resembles the selection applied to the reconstructed lepton pairs. The efficiency with which pairs that pass the fiducial selection are reconstructed and pass the analysis selection is called the fiducial efficiency. The fiducial efficiency tends to represent the efficiency with which the fiducial selected lepton pairs are reconstructed. The fiducial cross-section of a process is the production cross-section of the process multiplied by the efficiency of generating lepton pairs with the topology specified by the fiducial selection. In the next sections, the 95% CL_S upper limits on the fiducial cross-sections for new physics processes are calculated. The procedure uses

benchmark processes for the calculation of the fiducial efficiencies, but is such that the impact of these processes on the limits is minimal.

5.6.3.1 Fiducial selection and efficiency

The definition of the fiducial selection has a key role in the calculation of model independent and effective limits. Its aim is to select generator level di-lepton pairs that have characteristics suitable to be reconstructed and to pass the analysis selection. The region shaped by the fiducial selection is called the fiducial region. The more the fiducial selection is similar to the analysis selection applied to reconstructed leptons, the more the fiducial selection assumes the meaning of reconstruction efficiency of the di-lepton pair. As a consequence, the fiducial efficiency is less dependent from the characteristics of the production process and the limits that can be obtained from it are "largely" model independent.

The fiducial selection applied to generator level particles in this analysis is summarized in Table 5.23. The leptons fulfil the same kinematic requirements on p_{T} and η imposed

Selection	Electron requirement	Muon requirement
Leading p_{T} lepton	$p_{\mathrm{T}} > 25$ GeV	$p_{\mathrm{T}} > 25$ GeV
Subleading p_{T} lepton	$p_{\mathrm{T}} > 20$ GeV	$p_{\mathrm{T}} > 20$ GeV
Lepton η	$\lvert\eta\rvert < 1.37$ or $1.52 < \lvert\eta\rvert < 2.47$	$\lvert\eta\rvert < 2.5$
Isolation	$ptcone30/p_{\mathrm{T}} < 0.1$	$ptcone30/p_{\mathrm{T}} < 0.07$
Selection	Event selection	
Lepton pair	Same-sign lepton pair with $m(\ell^{\pm}, \ell^{\pm}) > 15$ GeV	
Electron pair	For $e^{\pm}e^{\pm}$ pairs $m(e^{\pm}, e^{\pm}) < 70$ GeV or $m(e^{\pm}, e^{\pm}) > 110$ GeV	
Z veto	No opposite-sign same flavour lepton pair with $\lvert m(\ell^{\pm}, \ell^{\mp}) - m(Z)\rvert < 10$ GeV	

Table 5.23: Definition of the fiducial selection applied to particles at generator level to select same-sign lepton pairs.

on the reconstructed leptons. Isolation requirements are also applied.

etcone and *ptcone* isolation requirements are calculated using stable particles (excluding neutrinos) and stable charged particles in place of reconstructed energy deposits and tracks, respectively. For the *ptcone*, the sum is over particles with $p_{\mathrm{T}} > 1$ (0.4) GeV and within $\Delta R = 0.3$ from the muon (electron) direction. The particles used to calculate the *etcone* have $p_{\mathrm{T}} > 1$ GeV and $\Delta R = 0.3$ (0.2) from the muon (electron) direction. The lepton itself is not included.

The *ptcone* distribution for generator level leptons and reconstructed leptons have been found to be compatible, and therefore a *ptcone* requirement equivalent to that of the analysis is included in the fiducial selection (the *ptcone* value cannot exceed 7% of the muon p_{T} and 10% of the electron p_{T}).

The *etcone* value calculated from generated particles is instead generally different from the *etcone* calculated from calorimeter signals. The resolution effects dominate the shape of the reconstructed *etcone* distribution, making it not reproducible at particle level. The

application of an *etcone* requirement on generator level particles would lead to an increase of the leakage and to no significant improvement in the fiducial efficiency. The leakage is defined as the fraction of di-lepton pairs that pass the analysis selection due to generated di-lepton pairs outside the fiducial region. The definition of the fiducial selection should be such that the leakage is minimal. If a consistent portion of lepton pairs, successfully selected in the analysis, lies outside the fiducial region, the exclusion power of the limits would be reduced. No *etcone* requirement is therefore applied at generator level.

The ΔR requirements on leptons with close-by jets have also been studied and are not applied at particle level for arguments similar to the *etcone* case.

The requirements on the event and di-lepton pair kinematic variables applied in the analysis selection are also considered in the fiducial selection. The pair of generated same-sign lepton pairs must satisfy $m(\ell^{\pm}, \ell^{\pm}) > 15$ GeV and not be in the range $70 < m(e^{\pm}, e^{\pm}) < 110$ GeV for the $e^{\pm}e^{\pm}$ category. Events with an opposite-sign same flavour lepton pair with $|m(\ell^{\pm}, \ell^{\mp}) - m(Z)| < 10$ GeV are excluded.

The fiducial efficiency, ϵ_{fid}, is the ratio of the number of selected same-sign lepton pairs to the number of generator level same-sign lepton pairs in the fiducial region. The fiducial efficiency ideally represents the average reconstruction efficiency for di-lepton pairs. However, because the reconstruction efficiency may be different for di-lepton pairs in different parts of the fiducial region, the fiducial efficiency can depend on the process considered. For instance, the efficiency of reconstruction for a lepton varies with its energy and direction, so processes with different lepton kinematic behaviour can have different fiducial efficiencies. Additional fiducial efficiency differences among processes can be due to the impossibility of reproducing the whole analysis selection applied on reconstructed pairs at generator level due to resolution effects. For example, since the *etcone* requirement is not part of the fiducial selection, processes with less isolated leptons could have a lower fiducial efficiency than processes with isolated particles.

Variations in fiducial efficiency are studied by considering several processes that form same-sign lepton pairs. The selected processes cover several regions of the same-sign lepton system kinematic phase space and have various levels of additional activity in the event. The model considered are described in Section 5.2. They are: pair production of $H_{\text{L}}^{\pm\pm}$, production of a W_{R} boson, production of $b'\bar{b}'$ pairs and diquark production. Several mass hypotheses are included.

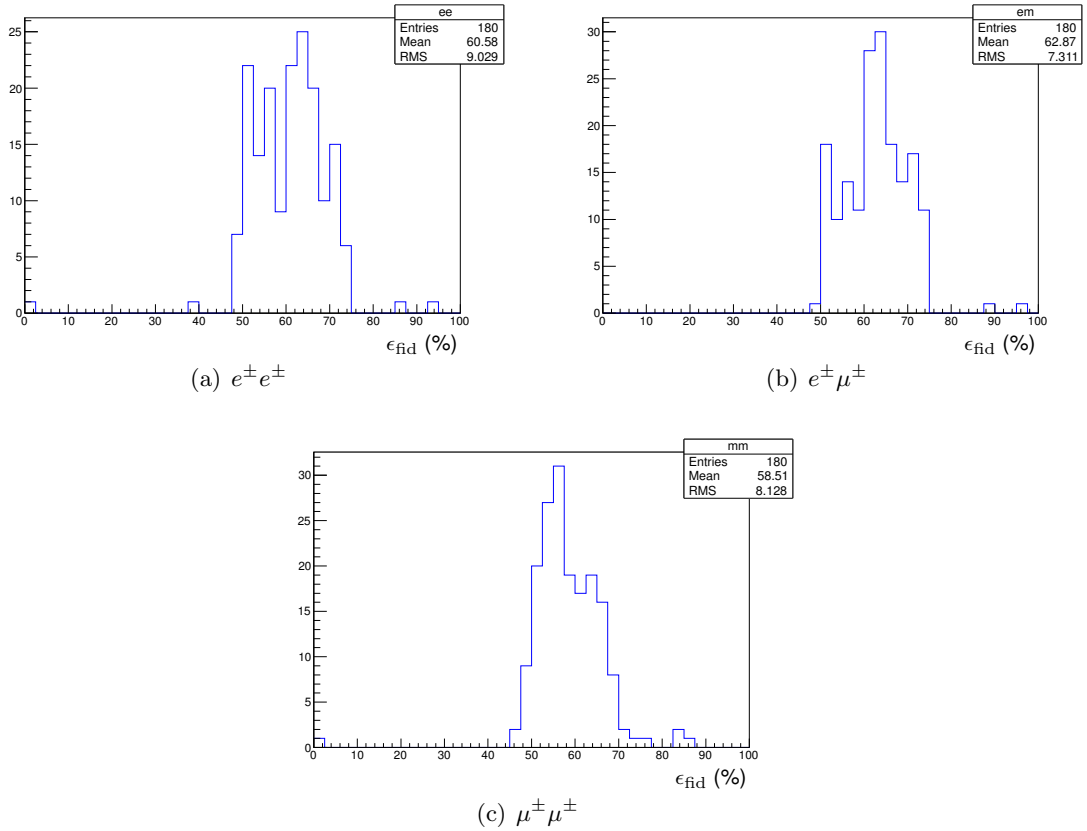
For each process, the fiducial selection efficiencies are calculated for several di-lepton invariant mass requirements: $m_{th} \in [15, 100, 200, 300]$ GeV. The distributions of ϵ_{fid} obtained from the considered processes in the three categories are shown in Figure 5.21. Outliers are due to production of $H_{\text{L}}^{\pm\pm}$ with mass smaller than or equal to the invariant mass requirement. These measurements have low-to-null lepton pairs leading to meaningless results. They are therefore excluded.

The fiducial efficiencies vary between 46% and 74% with similar values for the $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ categories. The lowest values of ϵ_{fid} are found in the case of $b'\bar{b}'$ pair

production, while W_{R} boson production and $H_{\mathrm{L}}^{\pm\pm}$ pair production have the highest efficiencies. The value of ϵ_{fid} increases with the di-lepton invariant mass requirement for each process and it is therefore expected a higher ϵ_{fid} for invariant mass thresholds of $[400, 500, 600]$ GeV. The fiducial efficiencies have also been derived separately for positively and negatively charged same-sign lepton pairs and no charge dependence has been found.

The leakage is limited to less than 10% for each process, category and mass value. The highest leakage is found for production of a W_{R} boson and the lowest for diquark production.

The most inclusive and model independent limits can be obtained by using the lowest measured fiducial efficiency in each category: 48.3% for the $e^{\pm}e^{\pm}$ category, 49.7% for the $e^{\pm}\mu^{\pm}$ category, and 46.0% for the $\mu^{\pm}\mu^{\pm}$ category.



(a) $e^{\pm}e^{\pm}$ (b) $e^{\pm}\mu^{\pm}$

(c) $\mu^{\pm}\mu^{\pm}$

Figure 5.21: The distributions of the fiducial efficiencies calculated for the processes considered in this analysis divided according to flavour category: $e^{\pm}e^{\pm}$ (a), $e^{\pm}\mu^{\pm}$ (b) and $\mu^{\pm}\mu^{\pm}$ (c).

5.6.3.2 Fiducial cross-section limits

The determination of limits on the number of produced non-Standard Model lepton pairs, N_{95}, and on the fiducial efficiency, ϵ_{fid}, allow the calculation of 95% CL_S upper limits

on the cross-section for new physics processes producing same-sign leptons in the fiducial region:

$$\sigma_{95}^{fid}(\ell^{\pm}\ell^{\pm}) = \frac{N_{95}}{\epsilon_{\text{fid}}\mathcal{L}} \tag{5.32}$$

where $\mathcal{L}$ is the integrated luminosity of the collected data. The observed and expected 95% CL_S upper limits on the fiducial cross-section are shown in Figure 5.22 and in Table 5.24 separately for each signal region category. Table 5.24 also shows the limits separately for positively and negatively charged pairs. Limits are calculated for pairs with lower invariant mass requirement $m_{th} \in [15, 100, 200, 300, 400, 500, 600]$ GeV. The observed limits range between 32 fb and 0.48 fb combining positively and negatively charged pairs. Smaller limits are found for larger m_{th}. The $\mu^{\pm}\mu^{\pm}$ category limits are stronger due to the lower Standard Model background. For all final states the observed limits are generally within 2σ of the expected limits.

The choice of the lower fiducial efficiency in each category from a set of heterogeneous processes leads to conservative but general limits on σ_{95}^{fid}. The limits can therefore be used to test other models producing same-sign lepton pairs. To investigate a process, the final state at generator level must be simulated. The simulation of the detection process is not needed. The selection described in Table 5.23 and one of the lower invariant mass requirements can be applied to the lepton pairs in the simulated events to obtain the fraction of generated same-sign pairs that pass this selection (A). The cross-section of a process can then be compared with the observed 95% upper limit σ_{95} given by:

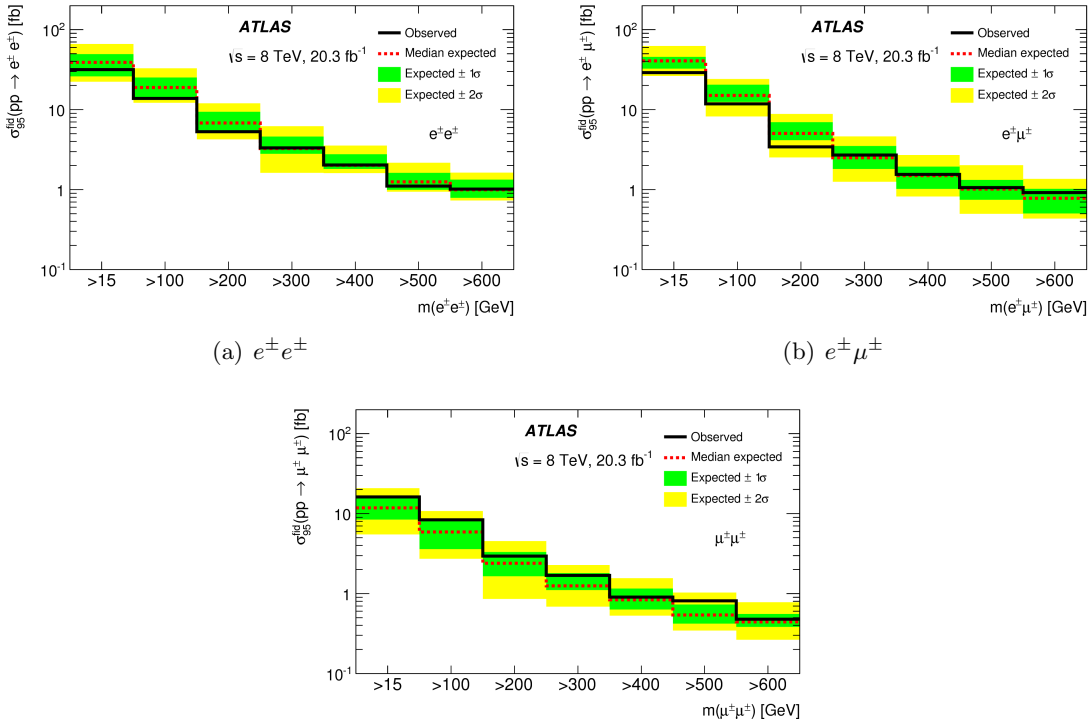
$$\sigma_{95} = \frac{\sigma_{95}^{fid}}{R_{\text{pair}}A} \tag{5.33}$$

where R_{pair} is the average number of same-sign lepton pairs produced per event. The lowest σ_{95} limit can be obtained by choosing the best invariant mass requirements, which depends on the behaviour of the process kinematic. Processes with lepton charge asymmetry may be able to be tested more effectively by using the positively or negatively charged pair limits.

This analysis has targeted generic anomalous production of same-sign lepton pairs by imposing minimal kinematic requirements and by using data to set fiducial cross-section limits. This procedure allows to test the prediction of any new physics process that produces same-sign lepton pairs. In some cases, however, the limits may not be as sensitive as process-specific studies. For instance, processes producing a Z boson, or with intense hadronic activity, have low probability of being selected (small A), hence the resulting limits are weak. For processes where same-sign leptons are the main signature, the limits set in this analysis can be very powerful. The next section is dedicated to the setting of limits on the production of doubly charged particles.

	95% CL_S upper limit [fb]					
	$e^{\pm}e^{\pm}$		$e^{\pm}\mu^{\pm}$		$\mu^{\pm}\mu^{\pm}$	
Mass range	Expected	Observed	Expected	Observed	Expected	Observed
> 15 GeV	39^{+10}_{-13}	32	41^{+5}_{-8}	29	12^{+4}_{-3}	16
> 100 GeV	19^{+6}_{-6}	14	$15.1^{+5.5}_{-2.6}$	11.8	$5.9^{+2.2}_{-2.3}$	8.4
> 200 GeV	$6.8^{+2.6}_{-1.7}$	5.3	$5.0^{+1.9}_{-0.9}$	3.4	$2.4^{+0.9}_{-0.8}$	2.9
> 300 GeV	$3.3^{+1.3}_{-0.4}$	3.3	$2.5^{+1.0}_{-0.7}$	2.7	$1.25^{+0.55}_{-0.15}$	1.69
> 400 GeV	$2.02^{+0.74}_{-0.21}$	2.03	$1.5^{+0.4}_{-0.5}$	1.6	$0.83^{+0.32}_{-0.20}$	0.91
> 500 GeV	$1.25^{+0.36}_{-0.26}$	1.10	$1.02^{+0.30}_{-0.27}$	1.06	$0.54^{+0.19}_{-0.12}$	0.82
> 600 GeV	$0.99^{+0.34}_{-0.20}$	1.02	$0.78^{+0.24}_{-0.28}$	0.92	$0.44^{+0.11}_{-0.06}$	0.48
	$e^{+}e^{+}$		$e^{+}\mu^{+}$		$\mu^{+}\mu^{+}$	
Mass range	Expected	Observed	Expected	Observed	Expected	Observed
> 15 GeV	27^{+11}_{-6}	28	25^{+10}_{-4}	23	$9.5^{+3.3}_{-3.1}$	14
> 100 GeV	$14.3^{+5.4}_{-2.8}$	13.5	$11^{+4}_{-2.1}$	9	$5.0^{+1.6}_{-1.3}$	6.3
> 200 GeV	$5.4^{+2.0}_{-1.4}$	4.6	$3.6^{+1.3}_{-0.7}$	3.6	$2.2^{+0.8}_{-0.5}$	3.6
> 300 GeV	$2.5^{+0.9}_{-0.6}$	2.0	$1.9^{+0.8}_{-0.5}$	2.6	$1.11^{+0.46}_{-0.29}$	1.42
> 400 GeV	$1.59^{+0.47}_{-0.34}$	1.64	$1.10^{+0.46}_{-0.23}$	1.39	$0.74^{+0.27}_{-0.17}$	0.74
> 500 GeV	$1.44^{+0.34}_{-0.36}$	1.55	$0.79^{+0.21}_{-0.22}$	0.89	$0.42^{+0.24}_{-0.10}$	0.38
> 600 GeV	$1.27^{+0.37}_{-0.26}$	1.10	$0.65^{+0.14}_{-0.16}$	0.77	$0.37^{+0.09}_{-0.05}$	0.32
	$e^{-}e^{-}$		$e^{-}\mu^{-}$		$\mu^{-}\mu^{-}$	
Mass range	Expected	Observed	Expected	Observed	Expected	Observed
> 15 GeV	23^{+8}_{-5}	19	$19.0^{+8.0}_{-2.8}$	16.0	$6.8^{+2.7}_{-1.5}$	8.3
> 100 GeV	$10.8^{+4.4}_{-2.4}$	9.0	$8.2^{+2.2}_{-2.1}$	5.6	$3.5^{+1.4}_{-0.9}$	5.1
> 200 GeV	$3.9^{+1.4}_{-1.2}$	3.5	$2.8^{+1.2}_{-0.9}$	1.5	$1.41^{+0.54}_{-0.33}$	1.29
> 300 GeV	$2.1^{+0.7}_{-0.5}$	2.6	$1.6^{+0.6}_{-0.4}$	1.3	$0.79^{+0.30}_{-0.16}$	1.0
> 400 GeV	$1.56^{+0.41}_{-0.31}$	1.35	$0.91^{+0.34}_{-0.26}$	0.77	$0.52^{+0.20}_{-0.13}$	0.59
> 500 GeV	$0.69^{+0.27}_{-0.17}$	0.64	$0.62^{+0.12}_{-0.12}$	0.65	$0.355^{+0.139}_{-0.013}$	0.683
> 600 GeV	$0.58^{+0.21}_{-0.08}$	0.61	$0.49^{+0.16}_{-0.10}$	0.59	$0.332^{+0.014}_{-0.011}$	0.454

Table 5.24: The 95% CL_S upper limits on the fiducial cross-section of new physics in same-sign lepton pairs divided according to lepton flavour ($e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$) and lower bound on the lepton pair mass. The expected limits and their 1σ uncertainties are given together with the observed limits derived from data. Limits are given separately for positively and negatively charged pairs and for their combination.



(a) $e^{\pm}e^{\pm}$ (b) $e^{\pm}\mu^{\pm}$

(c) $\mu^{\pm}\mu^{\pm}$

Figure 5.22: Observed and expected 95% CL_S upper limits on the fiducial cross-section of new physics in same-sign lepton pairs divided according to flavour $e^{\pm}e^{\pm}$ (a), $e^{\pm}\mu^{\pm}$ (b) and $\mu^{\pm}\mu^{\pm}$ (c) and as a function of the lower bound on the lepton pair mass. The green and yellow bands show the $\pm 1\sigma$ and $\pm 2\sigma$ errors on the expected limits.

5.6.4 Doubly charged particle search

The cleanest way to search for doubly charged particles at the LHC is to look for same-sign lepton pair resonances. As already discussed at the beginning of this chapter, doubly charged particles are expected in the Type-II see-saw and classic Zee-Babu models of neutrino mass generation. The Left-right symmetric model, which allows Type-II see-saw neutrino mass production, proposes two doubly charged particles: $H_{\mathrm{L}}^{\pm\pm}$ and $H_{\mathrm{R}}^{\pm\pm}$. According to this model, the dominant production mode of these particles at the LHC is pair production via a virtual Z/γ^* exchange, as shown in Figure 5.1(b). This analysis studies the $H_{\mathrm{L}}^{\pm\pm}$ pair production in the Left-right symmetric model and sets limits on its cross-section as a function of the doubly charged particle mass. Simulations of this process are used to determine the expected reconstructed width of the doubly charged resonances and their reconstruction efficiencies as a function of their mass, as described in more detail later. The limits can apply to other models that allow doubly charged particle pair production as long as the width of the doubly charged particle is negligible compared to the experimental resolution, and the production dynamic is the same as that of the $H_{\mathrm{L}}^{\pm\pm}$. This is the case for the doubly charged particle $k^{\pm\pm}$, of the classic Zee-Babu model, whose production mode is identical to that of the $H_{\mathrm{L}}^{\pm\pm}$. The $H_{\mathrm{R}}^{\pm\pm}$ boson has a smaller coupling to the Z boson than $H_{\mathrm{L}}^{\pm\pm}$, hence a smaller cross-section. The final state kinematic behaviour of the pair production mode is the same though. The decay widths of both $k^{\pm\pm}$ and $H_{\mathrm{R}}^{\pm\pm}$ are negligible, therefore the limits obtained on the $H_{\mathrm{L}}^{\pm\pm}$ pair production are also valid for these particles.

The doubly charged particles can decay either into pairs of same-sign fermions or pairs of same-sign $W^{\pm}$ bosons. For the $H_{\mathrm{L}}^{\pm\pm}$, the decay mode depends on the vacuum expectation value of the triplet of which the doubly charged field is part [174]. For small vacuum expectation values (< 0.1 MeV), a more natural condition for Type-II see-saw models, the decay to fermions is favoured and this is the case considered here. There are no indirect constraints on any individual coupling governing these decays, but there are constraints on the product of couplings. For example, the product of the $e^{\pm}e^{\pm}$ and $e^{\pm}\mu^{\pm}$ decay widths is constrained by the upper limit on the $\mu^+ \to e^+e^+e^-$ branching ratio to be less than 1.0×10^{-12} [146].

The final state of $H_{\mathrm{L}}^{\pm\pm}$ pair production is characterized by two resonant same-sign lepton pairs. The sample of same-sign lepton pairs collected in this analysis is scanned in ranges of invariant mass compatible with the resolution of the reconstructed $H_{\mathrm{L}}^{\pm\pm}$ mass. Since no significant excesses are observed, 95% CL_S upper limits are calculated on the number of same-sign pairs due to new physics in each interval and these are translated into limits on the production cross-section of $H_{\mathrm{L}}^{\pm\pm}$ $H_{\mathrm{R}}^{\pm\pm}$ and $k^{\pm\pm}$ pairs as a function of their mass. Lower limits on the doubly charged particle masses are also set considering the theoretical production cross-sections of these particles.

5.6.4.1 Determination of the mass intervals for the search

To study the production of doubly charged resonances, the same-sign lepton pairs collected in data and the Standard Model expectation are compared in several windows of invariant mass and divided according to lepton flavour. The mass intervals have to be large enough to contain most of the $H_{\mathrm{L}}^{\pm\pm}$ events but also avoid unnecessary Standard Model background. The interval is centred on the considered $H_{\mathrm{L}}^{\pm\pm}$ mass and the half-width of the interval is parametrized as a function of the $H_{\mathrm{L}}^{\pm\pm}$ mass:

$$a \times m(H_{\mathrm{L}}^{\pm\pm})/\mathrm{GeV} + b \times m(H_{\mathrm{L}}^{\pm\pm})^2/\mathrm{GeV}^2 \tag{5.34}$$

where the coefficients a and b depend on the flavour of the lepton pair. The width of the reconstructed resonance increases with the mass, however the increase depends on the flavour of the lepton pair. Muons have better energy resolution than electrons at low energies but worse at higher energies. The intrinsic width of the $H_{\mathrm{L}}^{\pm\pm}$ resonance is narrow for the mass range considered, and the width of the reconstructed mass peak is consequently dominated by the detector resolution. The invariant mass distribution of reconstructed $\mu^{\pm}\mu^{\pm}$ pairs is therefore broader than that for $e^{\pm}e^{\pm}$ pairs at larger masses. For 300 GeV $H_{\mathrm{L}}^{\pm\pm}$, the resolution in $e^{\pm}e^{\pm}$ pairs is already better than that in the other flavour categories, as shown in Figure 5.23. The coefficients a and b have been tuned to have

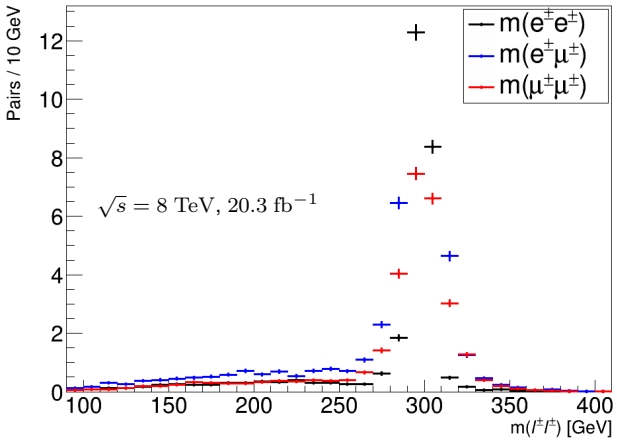


Figure 5.23: The invariant mass distributions of reconstructed same-sign lepton pairs originating from the decay of a $H_{\mathrm{L}}^{\pm\pm}$ with mass of 300 GeV. The decays to $e^{\pm}e^{\pm}$ (black), $e^{\pm}\mu^{\pm}$ (blue) and $\mu^{\pm}\mu^{\pm}$ (red) are shown separately.

about 80% of the reconstructed same-sign pairs from $H_{\mathrm{L}}^{\pm\pm}$ pair production falling inside the mass interval for any $H_{\mathrm{L}}^{\pm\pm}$ mass condition. These numbers have been calculated using simulated $H_{\mathrm{L}}^{\pm\pm}$ with masses between 50 and 1000 GeV (see Table B.2). The coefficients that determine the width of the invariant mass interval as a function of the $H_{\mathrm{L}}^{\pm\pm}$ mass are given in Table 5.25.

Category	$e^{\pm}e^{\pm}$	$e^{\pm}\mu^{\pm}$	$\mu^{\pm}\mu^{\pm}$
a	0.04	0.06	0.06
b	$0.2 \cdot 10^{-4}$	$0.7 \cdot 10^{-4}$	$1.5 \cdot 10^{-4}$

Table 5.25: The values of the coefficients used in Equation 5.34 to calculate the half-width of the invariant mass interval investigated in the analysis.

The same-sign lepton pair invariant mass distribution for each flavour category is scanned using a sliding window whose half-width changes with mass as explained above. Data and Standard Model expectations are compared but no significant differences are found. The 95% CL_S upper limits on the number of same-sign lepton pairs due to new physics are calculated as a function of the $H_{\mathrm{L}}^{\pm\pm}$ mass. The limits on the production cross-section of $H_{\mathrm{L}}^{\pm\pm}$ $H_{\mathrm{R}}^{\pm\pm}$ and $k^{\pm\pm}$ pairs as a function of their mass are then obtained, as described in Section 5.6.4.3. The acceptance times efficiency of reconstructing a $H_{\mathrm{L}}^{\pm\pm}$ decay into leptons is calculated to extract the cross-section limits.

5.6.4.2 Reconstruction acceptance and efficiency

A key element for setting limits on the $H_{\mathrm{L}}^{\pm\pm}$ production cross-section is the calculation of the acceptance and reconstruction efficiency of the pairs of same-sign leptons. The same-sign lepton pair has to pass the analysis selection and its invariant mass must be within the mass window around the $H_{\mathrm{L}}^{\pm\pm}$ mass defined above. In other words, the percentage of $H_{\mathrm{L}}^{\pm\pm}$ particles produced by pair production that are reconstructed, selected and that fall inside the mass window must be obtained as a function of the $H_{\mathrm{L}}^{\pm\pm}$ mass. The product of the acceptance and the reconstruction efficiency ($A \times \epsilon$) is calculated for each simulated $H_{\mathrm{L}}^{\pm\pm}$ mass point and for each decay mode. The value of $A \times \epsilon$ between simulated mass points is obtained from a fit with an empirical function. The function has the form shown in Equation 5.35. It has four free parameters (p_0, p_1, p_2 and p_3) and a fifth (p_4) determined by requiring continuity at $m(H_{\mathrm{L}}^{\pm\pm}) = 450$ GeV for the $e^{\pm}e^{\pm}$ category and at $m(H_{\mathrm{L}}^{\pm\pm}) = 300$ GeV for the $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ categories.

$$A \times \epsilon = \begin{cases} p_0 \cdot (1 - e^{-(m-p_1)/p_2}), & \text{if } m < 300 \text{ GeV } (450 \text{ GeV for } e^{\pm}e^{\pm}) \\ p_3 + p_4 \cdot m, & \text{if } m \geq 300 \text{ GeV } (450 \text{ GeV for } e^{\pm}e^{\pm}) \end{cases} \tag{5.35}$$

The fit to the $A \times \epsilon$ as a function of the $H_{\mathrm{L}}^{\pm\pm}$ mass is show in Figure 5.24 for each $H_{\mathrm{L}}^{\pm\pm}$ decay mode. The values of the parameters obtained from the fit are summarized in Table 5.26. The efficiency is smaller at lower masses due to the lower probability for low energy leptons to pass the analysis selection. For $H_{\mathrm{L}}^{\pm\pm}$ masses above 300 GeV the $A \times \epsilon$ assumes stable values between 45% and 50% in all categories.

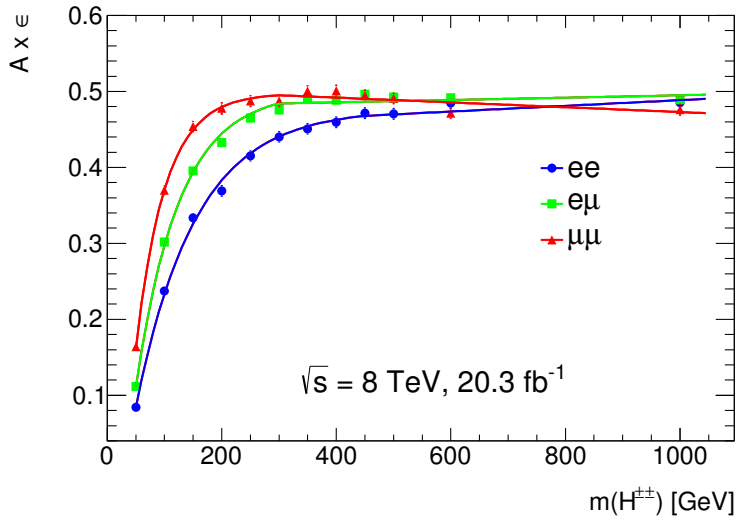


Figure 5.24: The product of the acceptance and the reconstruction efficiency ($A \times \epsilon$) of the same-sign lepton pair from the $H_{\mathrm{L}}^{\pm\pm}$ decay computed as a function of the $H_{\mathrm{L}}^{\pm\pm}$ mass and for the $e^{\pm}e^{\pm}$ (blue), $e^{\pm}\mu^{\pm}$ (green), and $\mu^{\pm}\mu^{\pm}$ (red) decay modes. The fitted $A \times \epsilon$ curves as a function of the $H_{\mathrm{L}}^{\pm\pm}$ mass are also shown.

Parameter	Value		
	$e^{\pm}e^{\pm}$	$e^{\pm}\mu^{\pm}$	$\mu^{\pm}\mu^{\pm}$
p_0	4.76×10^{-1}	4.99×10^{-1}	4.98×10^{-1}
p_1	$2.94 \times 10^{+1}$	$3.06 \times 10^{+1}$	$2.95 \times 10^{+1}$
p_2	$1.05 \times 10^{+2}$	$7.64 \times 10^{+1}$	$5.14 \times 10^{+1}$
p_3	4.51×10^{-1}	4.79×10^{-1}	5.05×10^{-1}

Table 5.26: The values obtained from the fit of the parametric function in Equation 5.35 to the $A \times \epsilon$ dependence on the $H_{\mathrm{L}}^{\pm\pm}$ mass for each decay mode.

5.6.4.3 Limits on doubly charged particle production

The 95% CL_S upper limit on the cross-section for pair production of $H_{\mathrm{L}}^{\pm\pm}$ $H_{\mathrm{R}}^{\pm\pm}$ and $k^{\pm\pm}$ multiplied by the branching ratio (BR) in each of the three flavour categories is extracted as a function of the particle mass. These limits are obtained from the 95% CL_S upper limits on the number of same-sign lepton pairs due to new physics and the value of $A \times \epsilon$ calculated above.

The expected number of $H_{\mathrm{L}}^{\pm\pm}$ pairs produced at the LHC is given by:

$$N^{HH} = \sigma^{HH} \cdot \mathcal{L} \tag{5.36}$$

where σ^{HH} is the pair production cross-section and $\mathcal{L}$ is the integrated luminosity. The expected number of $H_{\mathrm{L}}^{\pm\pm}$ decays in a given lepton flavour mode with branching ratio BR at the LHC is therefore:

$$N^{H} = 2 \cdot N^{HH} \cdot \mathrm{BR} \tag{5.37}$$

Given the likelihood of the $H_{\mathrm{L}}^{\pm\pm}$ decay to be reconstructed ($A \times \epsilon$) and the 95% CL_S upper limit on the number of reconstructed $H_{\mathrm{L}}^{\pm\pm}$ decays (N_{95}^{H}), a 95% CL_S upper limit on σ^{HH} times branching ratio can be set:

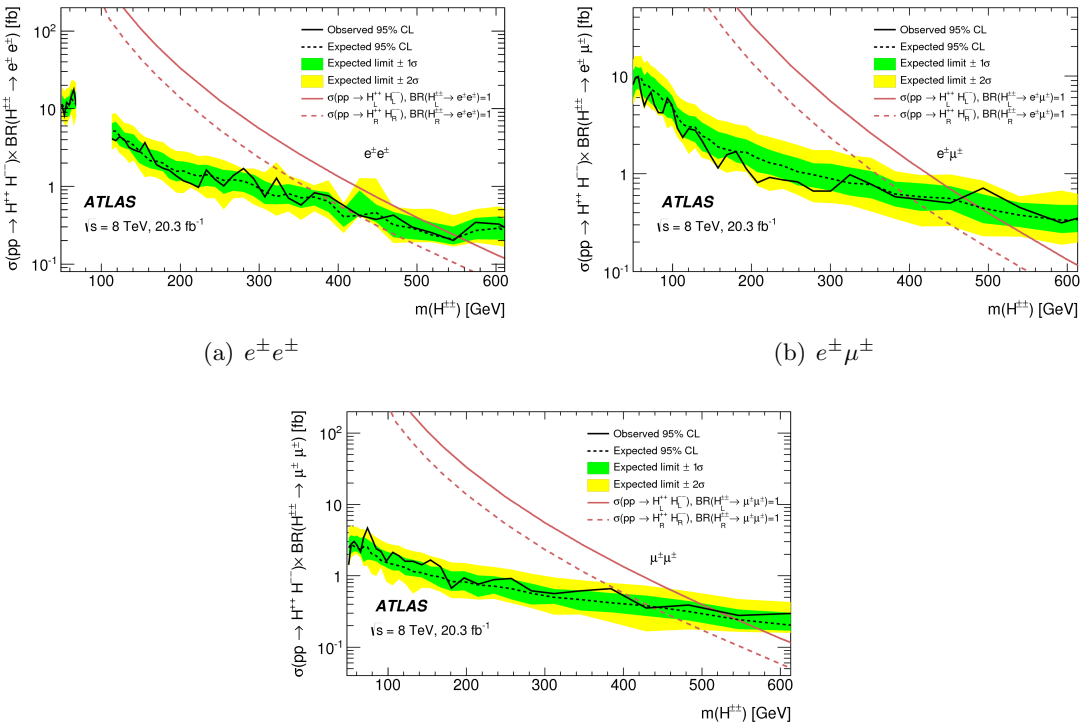
$$\sigma_{95}^{HH} \times \mathrm{BR} = \frac{N_{95}^{H}}{2 \cdot A \times \epsilon \cdot \mathcal{L}} \tag{5.38}$$

The limits on the pair production cross-section of doubly charged particles times branching ratio are shown in Figure 5.25 as a function of the particle mass. The cross-section values expected for $H_{\mathrm{L}}^{\pm\pm}$ and $H_{\mathrm{R}}^{\pm\pm}$ pair production in the Left-right symmetric model are also shown.

The limits are of the order of femtobarns for $m(H_{\mathrm{L}}^{\pm\pm}) < 200$ GeV and around 0.3 fb for $m(H_{\mathrm{L}}^{\pm\pm}) > 500$ GeV. The limits in the $\mu^{\pm}\mu^{\pm}$ category are stronger in the low mass region than in the other categories thanks to the smaller Standard Model background. At high masses, the limits are similar among flavour categories. This is due to the larger mass window used for $\mu^{\pm}\mu^{\pm}$ resonances which increases the background contamination. The scatter between adjacent mass intervals in the observed limits is due to fluctuations in the yields derived from limited statistics. In general, good agreement is seen between observed and expected limits with maximum deviations of 2σ.

5.6.4.4 Mass limits

Considering the theoretical predictions for the pair productions of $H_{\mathrm{L}}^{\pm\pm}$ and $H_{\mathrm{R}}^{\pm\pm}$, a lower limit on the mass of these particles can be set. The 95% CL_S lower mass limits for $H_{\mathrm{L}}^{\pm\pm}$ and $H_{\mathrm{R}}^{\pm\pm}$ obtained in each flavour category, assuming BR = 1 in that category, are given in Table 5.27. In addition to the uncertainties on the cross-section limits, the 1σ errors on the expected limits take into account the theoretical uncertainties on the cross-section



(a) $e^{\pm}e^{\pm}$

(b) $e^{\pm}\mu^{\pm}$

(c) $\mu^{\pm}\mu^{\pm}$

Figure 5.25: Observed and expected 95% CL_S upper limits on the cross-section of $H_{\mathrm{L}}^{\pm\pm}$ pair production times BR as a function of the $H_{\mathrm{L}}^{\pm\pm}$ mass for the $e^{\pm}e^{\pm}$ (a), $e^{\pm}\mu^{\pm}$ (b) and $\mu^{\pm}\mu^{\pm}$ (c) decay modes. The green and yellow bands show the $\pm1\sigma$ and $\pm2\sigma$ errors on the expected limits. Also shown are the expected cross-sections as a function of mass for $H_{\mathrm{L}}^{\pm\pm}$ (solid red) and $H_{\mathrm{R}}^{\pm\pm}$ (dashed red) pair production in the Left-right symmetric model. The mass range between 70 GeV and 110 GeV is not included in the $e^{\pm}e^{\pm}$ category since this region is highly affected by charge misID background and has been used for its estimate.

	95% CL_S lower limit [GeV]					
	$e^{\pm}e^{\pm}$		$e^{\pm}\mu^{\pm}$		$\mu^{\pm}\mu^{\pm}$	
Signal	Expected	Observed	Expected	Observed	Expected	Observed
$H_{\mathrm{L}}^{\pm\pm}$	553 ± 30	551	487 ± 41	468	543 ± 40	516
$H_{\mathrm{R}}^{\pm\pm}$	425 ± 30	374	396 ± 34	402	435 ± 33	438

Table 5.27: Observed and expected 95% CL_S lower limits on the mass of $H_{\mathrm{L}}^{\pm\pm}$ and $H_{\mathrm{R}}^{\pm\pm}$ bosons, assuming a 100% branching ratio to $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ or $\mu^{\pm}\mu^{\pm}$ pairs. The 1σ variations are also shown for the expected limits.

for $H_{\mathrm{L}}^{\pm\pm}$ and $H_{\mathrm{R}}^{\pm\pm}$ production (about 10%). The uncertainties are symmetrized to reduce the effect of statistical fluctuations.

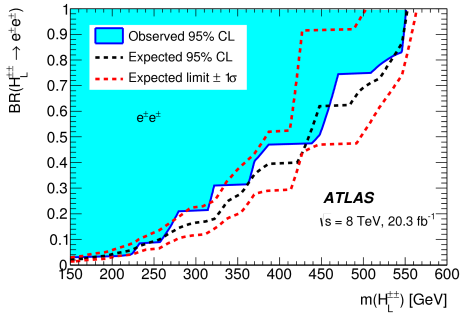
The best limits are obtained for $H_{\mathrm{L}}^{\pm\pm}$ in the $e^{\pm}e^{\pm}$ category with a limit of 553 GeV and for $H_{\mathrm{R}}^{\pm\pm}$ in the $\mu^{\pm}\mu^{\pm}$ category with a limit around 435 GeV. The limits are from 10 to 20% worse in the $e^{\pm}\mu^{\pm}$ category due to the larger background at high invariant masses from $W^{\pm}Z$ production. The signal yields are similar in all three categories, but the $W^{\pm}Z$ background is approximately twice as large in the $e^{\pm}\mu^{\pm}$ category than in the other categories.

Lower mass limits are also calculated as a function of the decay branching ratio in each category. These 95% CL_S limits are displayed in Figure 5.26 for both $H_{\mathrm{L}}^{\pm\pm}$ and $H_{\mathrm{R}}^{\pm\pm}$. The mass limits on the $H_{\mathrm{L}}^{\pm\pm}$ mass are stronger due to the higher expected production cross-section. The limits obtained for the $H_{\mathrm{L}}^{\pm\pm}$ are also valid for the $k^{\pm\pm}$ predicted in the classic Zee-Babu model as the cross-sections and decay kinematics are identical.

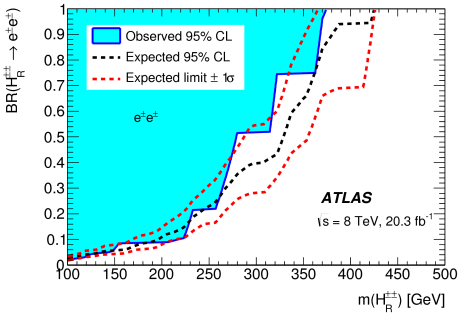
5.6.5 Summary

This analysis has investigated the production of same-sign lepton pairs due to physics beyond the Standard Model. No significant excess over the predicted Standard Model background is observed. An inclusive search has been developed to set upper limits on the fiducial cross-section of new physics processes producing same-sign lepton pairs. These limits, shown in Figure 5.22, can be used to test new physics models after calculating the efficiency with which same-sign lepton pairs produced in these processes pass the fiducial selection given in Table 5.23. A comparison of these limits with other measurements is not trivial due to the analysis-dependent definition of fiducial efficiencies. For processes with significant additional activity beside the same-sign lepton pair, dedicated analyses may lead to better limits.

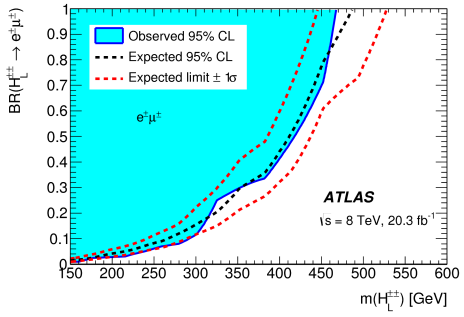
The search for $H_{\mathrm{L}}^{\pm\pm}$, $k^{\pm\pm}$ and $H_{\mathrm{R}}^{\pm\pm}$ pair production has not found significant excesses attributable to a doubly charged resonance. The limits obtained (Table 5.27 and Figure 5.26) improve on the results of previous analyses by ATLAS [175], CMS [176] and CDF [177]. Compared to the results based on the 7 TeV data collected by ATLAS [175], the limits on the doubly charged particle masses have been increased by 30-40%. ATLAS has recently also published results with 13 TeV data [178] obtaining limits compatible with those obtained in this thesis (570 GeV for $H_{\mathrm{L}}^{\pm\pm}$ and 420 GeV for $H_{\mathrm{R}}^{\pm\pm}$).



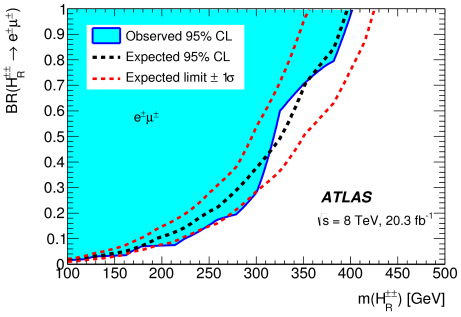
(a) $e^{\pm}e^{\pm}$



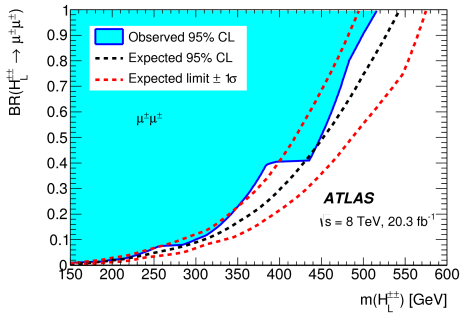
(b) $e^{\pm}e^{\pm}$



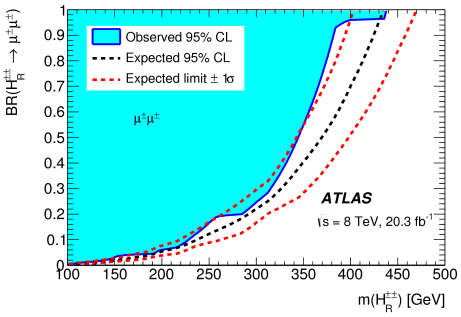
(c) $e^{\pm}\mu^{\pm}$



(d) $e^{\pm}\mu^{\pm}$



(e) $\mu^{\pm}\mu^{\pm}$



(f) $\mu^{\pm}\mu^{\pm}$

Figure 5.26: Observed and expected 95% CL_S lower limits on the $H_{\mathrm{L}}^{\pm\pm}$ (left colum) and $H_{\mathrm{R}}^{\pm\pm}$ (right column) masses as a function of the branching ratio to $e^{\pm}e^{\pm}$ (top) $e^{\pm}\mu^{\pm}$ (centre) and $\mu^{\pm}\mu^{\pm}$ (bottom) pairs. The cyan area is excluded by observation and the expected limit is shown as a black dashed line. The band delimited by dashed red lines is the $\pm 1\sigma$ error on the expected limits.

Chapter 6

Study of $t\bar{t}H$ production

This chapter investigates the production of the Higgs boson in association with a top anti-top quark pair at the LHC. The Higgs boson and its properties have been introduced in Section 1.2. The ATLAS and CMS experiments have discovered this particle in 2012 and measured its mass to be $m_H = 125.09 \pm 0.21$ (stat.) ± 0.11 (syst.) GeV [26]. Despite evidences of the Higgs boson in numerous production and decay modes, the associated production of a Higgs boson with a top anti-top quark pair ($t\bar{t}H$) remains unobserved. The $t\bar{t}H$ production has indeed the smallest cross-section among the Standard Model Higgs boson production modes that can be investigated at the LHC.

The study presented here is part of the first search for the $t\bar{t}H$ production mode with the data collected by the ATLAS experiment at $\sqrt{s} = 8$ TeV. The targeted final state comprises one same-sign lepton pair ($e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ or $\mu^{\pm}\mu^{\pm}$) and jets. There are no additional reconstructed leptons or τ_{had}. This final state, labelled as $2\ell 0\tau_{\text{had}}$, is expected to be the most sensitive to $t\bar{t}H$ production. The $t\bar{t}H$ production has also been searched in final states with three leptons (3ℓ), four leptons (4ℓ), one same-sign lepton pair and one τ_{had} ($2\ell 1\tau_{\text{had}}$), and one lepton and two τ_{had} ($1\ell 2\tau_{\text{had}}$). These studies have been combined and published [3].

Table 6.1 shows the contribution of the Higgs boson decay modes to the investigated final states (categories). These analyses target mainly the decay modes to WW^*, $\tau^+\tau^-$ and ZZ^*. In the $2\ell 0\tau_{\text{had}}$ category, about 80% of the signal events are produced via $H \to WW^{(*)}$ decays and 15% via $H \to \tau^+\tau^-$ decays. The $t\bar{t}H$ productions with the Higgs boson decaying to $b\bar{b}$ and $\gamma\gamma$ are studied in independent analyses based on 8 TeV data [31,32].

This chapter focuses on the study of the $2\ell 0\tau_{\text{had}}$ analysis, where the author had major involvement. In this category, the author has conducted the validation of the background estimate and the evaluation of the systematic uncertainties. He has also studied alternative approaches to the estimate of the non-prompt backgrounds, while the non-prompt and charge misID background estimates used for publication and the event selection have been developed by collaborators. The contributes by the author to the other categories are

	Higgs boson decay mode			
Category	WW^*	$\tau\tau$	ZZ^*	Other
$2\ell 0\tau_{\text{had}}$	80%	15%	3%	2%
3ℓ	74%	15%	7%	4%
$2\ell 1\tau_{\text{had}}$	35%	62%	2%	1%
4ℓ	69%	14%	14%	3%
$1\ell 2\tau_{\text{had}}$	4%	93%	0%	3%

Table 6.1: Fraction of $t\bar{t}H$ events expected for different Higgs boson decay modes in each category. $H \to \mu^+\mu^-$ and $H \to b\bar{b}$ are the main decays contributing to the "other" column.

summarized in Appendix C.

The goal of this study is to measure the $t\bar{t}H$ signal strength (μ), which is defined as the ratio of the measured $t\bar{t}H$ cross-section to the one expected in Standard Model. The results are compared with other measurements of $t\bar{t}H$ production and are used to determine the Top Yukawa coupling at the end of the chapter.

6.1 $t\bar{t}H$ production and decay topology

The production of the Higgs boson with a $t\bar{t}$ pair at the LHC is initiated by gluon-gluon interaction at leading order in the perturbation theory. The top-quarks subsequently decay to a b-quark and a $W^{\pm}$ boson ($t \to W^+b$, $\bar{t} \to W^-\bar{b}$), while the Higgs boson can decay in several modes as illustrated in Figure 1.1(b). Leptons with same charge are produced in the $t\bar{t}H$ decay when, for instance, the Higgs boson decays to WW^*, and in addition one $W^{\pm}$ from the Higgs boson and one $W^{\pm}$ from a t-quark with same charge decay into leptons as shown in Figure 6.1(a). In addition to leptons, the final state presents four jets (sometimes more from additional radiation) and at least two of those are initiated by a b-quark. When the Higgs boson decays to $\tau^+\tau^-$ or ZZ^*, the final state can have same-sign leptons, but an additional lepton or τ_{had} is also produced, as shown in Figure 6.1(b,c). These processes enters in the $2\ell 0\tau_{\text{had}}$ category when the additional lepton or τ_{had} is not detected. The collision data are therefore scanned in search for events with same-sign lepton pairs and multiple jets.

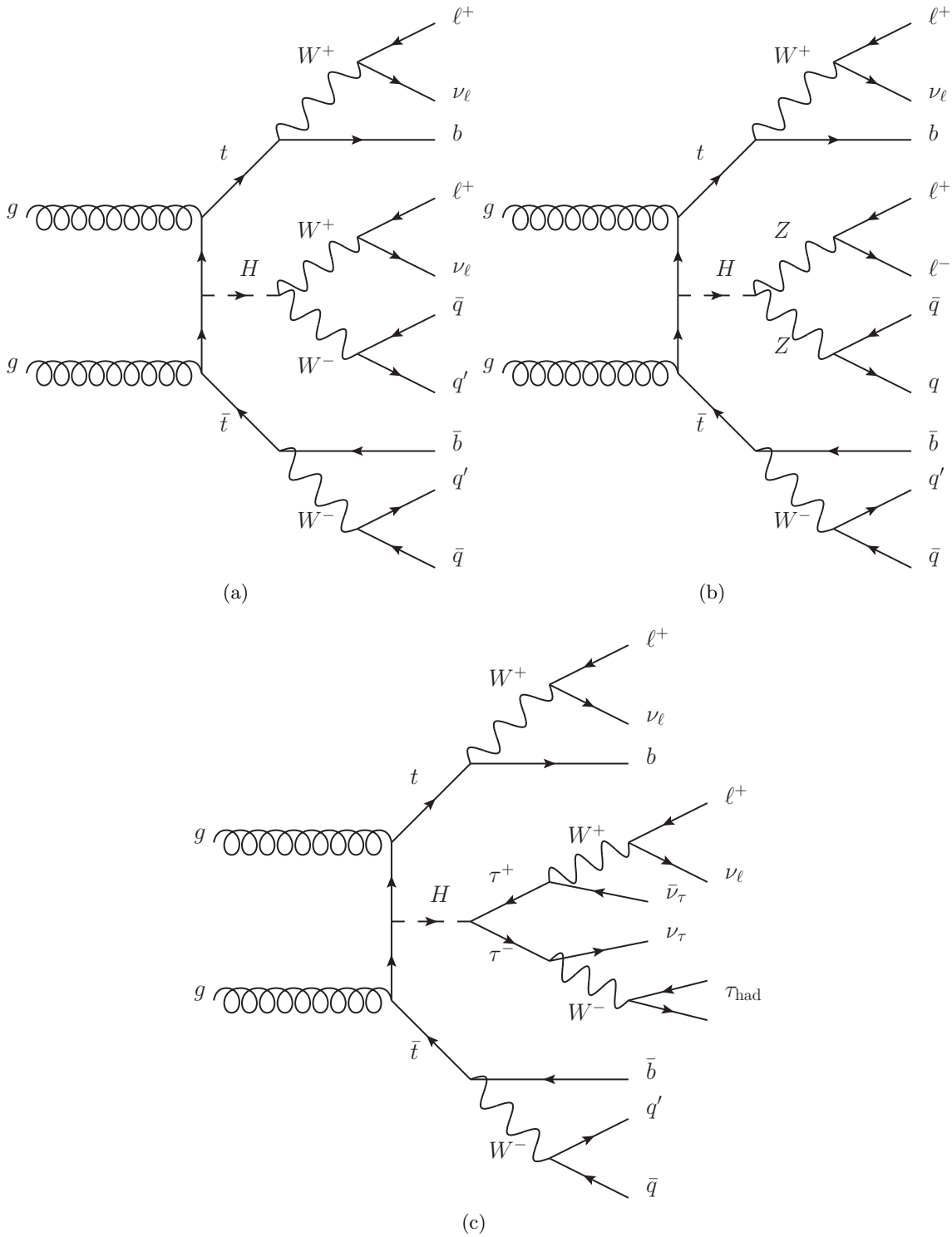


Figure 6.1: The Feynman diagram of tree level $t\bar{t}H$ production and decay to final states with a same-sign lepton pair. The Higgs boson decays to WW^*, $\tau^+\tau^-$ and ZZ^* in (a), (b), (c) respectively.

6.2 Data and simulation

The analysis of the $t\bar{t}H$ production is performed on the proton-proton collisions collected by ATLAS at a centre-of-mass energy of $\sqrt{s} = 8$ TeV. Only events with full operational detectors are considered. Events affected by significant noise, beam background and cosmic ray background are also excluded. The integrated luminosity of the dataset is 20.3 fb^{-1}.

Simulation are used in the estimate of $t\bar{t}H$ production and Standard Model backgrounds. The simulation of collisional events is a multi-step process that uses dedicated programs to model specific aspects of the interaction. In addition to the process generation, the simulation takes into account the interaction of particles with the ATLAS detector and the resulting readout signals generated. Details on the simulation have been given in Section 2.3.

The simulations used in the estimate of the signal and background processes are listed in the Table 6.2. For each process the generator, the cross-section used to scale the simulated yields and the integrated luminosity corresponding to the statistics generated in the simulation are given. The complete set of simulations used in the generation of a process is specified in Table 2.1. The simulations of processes are sometimes divided by final state and filters can be applied at matrix element to select a specific decay mode, to simulate additional partons or to apply kinematic requirements. This is done in order to simulate the processes of interest with enough statistics. The cross-sections provided take into account the efficiency of these event filters and higher order corrections to the production cross-section in the perturbation theory.

The $t\bar{t}H$ production is computed at NLO in QCD and includes all the Higgs boson decay modes with branching ratios calculated at NNLO in QCD [15, 179–182].

The prompt background and the photon conversion background are estimated with simulations in this analysis. The relevant processes producing prompt same-sign leptons are: $t\bar{t}W^{\pm}$, $t\bar{t}Z$, tZ, $W^{\pm}Z$, ZZ, $t\bar{t}W^{+}W^{-}$, $tHqb$ and $tHW^{\pm}$. The $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ simulations consider up to two additional partons and up to one additional parton at matrix element, respectively. For diboson processes, the SHERPA $q\bar{q}$ and qg originated samples include diagrams with additional partons at the matrix element. The processes producing Z bosons, like $t\bar{t}Z$, tZ, $W^{\pm}Z$, ZZ productions also include off-shell Z and virtual photon components. The production of the Higgs boson with a single t-quark are considered a background in this analysis. The $tHqb$ and $tHW^{\pm}$ productions are simulated but the s-channel production of tH is neglected due to the much smaller cross-section in Standard Model.

The main process contributing to photon conversion background is the $W^{\pm}\gamma$ production with additional hadronic activity. This process is simulated with up to five extra partons at matrix element. The $t\bar{t}$, $tW^{\pm}$, $Z+$jets and $W^{\pm}+$jets productions constitute the main non-prompt and charge misID backgrounds in this analysis. The contributions due to these processes in the signal regions are extracted from data, however simulations of

these processes have been used during the analysis as specified in Section C.1. The simulations considered are summarized in Table C.1. The $Z+$jets and $W^{\pm}+$jets contributions are produced with up to five additional partons at matrix element.

All the simulations are corrected to match the pileup conditions observed in data, and the real detector performances such as particle detection efficiencies, energy calibration and energy resolution, as specified in Section 2.4.

Table 6.2: Simulated processes used for signal and background modeling. For each process, the generator, the cross-section (including the generator filter efficiency and the K-factor) and the integrated luminosity equivalent to the statistics generated is given. The full simulation setup is specified in Table 2.1. $X + np$ with $n = 1, 2, 3, 4, 5$ and X a generic process indicates the number of partons (quarks or gluons) produced at matrix element level in addition to the process.

Process	Generator	$\sigma \times \varepsilon_{filter}$ $\times K$-factor [fb]	Equivalent $\mathcal{L}$ [fb^{-1}]
Signal processes			
$t\bar{t}H\left(t\bar{t} \to b\bar{b}qqqq\right)$	PowHel+Pythia8	59.09	2146.5
$t\bar{t}H\left(t\bar{t} \to b\bar{b}qq\ell\nu_\ell\right)$	PowHel+Pythia8	56.63	2238.9
$t\bar{t}H\left(t\bar{t} \to b\bar{b}\ell\nu_\ell\ell\nu_\ell\right)$	PowHel+Pythia8	13.58	9332.0
Background processes			
$t\bar{t}W^{\pm}$	MadGraph+Pythia6	122.84	2633.2
$t\bar{t}W^{\pm} + p$	MadGraph+Pythia6	62.98	6347.7
$t\bar{t}W^{\pm} + \geq 2p$	MadGraph+Pythia6	48.95	7450.8
$t\bar{t}Z\,(\geq 1\ell)$	MadGraph+Pythia6	93.23	5362.0
$t\bar{t}Z\,(Z/\gamma^* \to \ell^+\ell^-) + \geq 1p$	MadGraph+Pythia6	12.55	7968.0
$t\bar{t}Z\,(no\ Z/\gamma^* \to \ell^+\ell^-) + \geq 1p$	MadGraph+Pythia6	48.85	8188.2
tZ inclusive ($tW^{\pm}$ channel, $Z/\gamma^* \to \ell^+\ell^-$)	MadGraph+Pythia6	4.13	24212.8
tZ inclusive ($s+t$ channel, $Z/\gamma^* \to \ell^+\ell^-$)	MadGraph+Pythia6	3.12	32051.3
$tHqb\,(H \to W^+W^-)$	MadGraph+Pythia8	3.65	273595.6
$tHW^{\pm}$	MG5+Herwig	1.43	34925.5
$t\bar{t}t\bar{t}$	MadGraph+Pythia8	0.69	291441.1
$t\bar{t}W^+W^-$	MadGraph+Pythia8	1.90	104890.0
$gg \to H \to ZZ \to 4\ell$	Powheg+Pythia8	5.57	35913.1
$W^{\pm}W^{\pm} + 2p$	MadGraph+Pythia8	344.42	551.7
$W^{\pm}\gamma^* \to \ell\nu ee(m_{\ell\ell} < 7$ GeV)	MadGraph+Pythia6	11256.00	35.5
$W^{\pm}\gamma^* \to \ell\nu\mu\mu(m_{\ell\ell} < 7$ GeV)	MadGraph+Pythia6	2769.18	108.2
$W^{\pm}\gamma^* \to \ell\nu\tau\tau(m_{\ell\ell} < 7$ GeV)	MadGraph+Pythia6	295.81	101.4
$W^{\pm}Z \to \ell\ell\ell\nu$	Sherpa	10231.79	263.8
$W^{\pm}Z \to \ell\ell\ell\nu + 2p$(6EWcoupling)	Sherpa	12.56	1592.5
$ZZ \to \ell\ell\ell\ell + 2p$(6EWcoupling)	Sherpa	0.74	679628.6
$ZZ \to 4e(m_{\ell\ell} > 4$ GeV)	Powheg+Pythia8	66.71	16488.7

Table 6.2: Simulated processes used for signal and background modeling. For each process, the generator, the cross-section (including the generator filter efficiency and the K-factor) and the integrated luminosity equivalent to the statistics generated is given. The full simulation setup is specified in Table 2.1. $X + np$ with $n = 1, 2, 3, 4, 5$ and X a generic process indicates the number of partons (quarks or gluons) produced at matrix element level in addition to the process.

Process	Generator	$\sigma \times \varepsilon_{filter}$ $\times K$-factor [fb]	Equivalent $\mathcal{L}$ [fb^{-1}]
$ZZ \to 2e2\mu(m_{\ell\ell} > 4$ GeV)	Powheg+Pythia8	141.29	11319.4
$ZZ \to 2e2\tau(m_{\ell\ell} > 4$ GeV)	Powheg+Pythia8	99.54	11048.7
$ZZ \to 4\mu(m_{\ell\ell} > 4$ GeV)	Powheg+Pythia8	67.06	16398.8
$ZZ \to 2\mu2\tau(m_{\ell\ell} > 4$ GeV)	Powheg+Pythia8	100.30	10960.8
$ZZ \to 4\tau(m_{\ell\ell} > 4$ GeV)	Powheg+Pythia8	7.79	38491.4
$gg \to ZZ \to 4e$	gg2ZZ+Herwig	0.67	134328.4
$gg \to ZZ \to 4\mu$	gg2ZZ+Herwig	0.67	134328.4
$gg \to ZZ \to 2e2\mu$	gg2ZZ+Herwig	1.35	66666.7
$Z\gamma^* \to 4e(m_{\ell\ell1} > 4,\ m_{\ell\ell2} < 4$ GeV)$+p$	Sherpa	2828.60	282.8
$Z\gamma^* \to 2\mu2e(m_{\ell\ell1} > 4,\ m_{\ell\ell2} < 4$ GeV)$+p$	Sherpa	2345.00	255.4
$Z\gamma^* \to 2\tau2e(m_{\ell\ell1} > 4,\ m_{\ell\ell2} < 4$ GeV)$+p$	Sherpa	1593.30	31.9
$Z\gamma^* \to 2e2\mu(m_{\ell\ell1} > 4,\ m_{\ell\ell2} < 4$ GeV)$+p$	Sherpa	486.10	308.6
$Z\gamma^* \to 4\mu(m_{\ell\ell1} > 4,\ m_{\ell\ell2} < 4$ GeV)$+p$	Sherpa	508.40	294.8
$Z\gamma^* \to 2\tau2\mu(m_{\ell\ell1} > 4,\ m_{\ell\ell2} < 4$ GeV)$+p$	Sherpa	422.90	354.7
$Z\gamma^* \to 2e2\tau(m_{\ell\ell1} > 4,\ m_{\ell\ell2} < 4$ GeV)$+p$	Sherpa	4.00	5000.0
$Z\gamma^* \to 2\mu2\tau(m_{\ell\ell1} > 4,\ m_{\ell\ell2} < 4$ GeV)$+p$	Sherpa	4.00	5000.0
$Z\gamma^* \to 4\tau(m_{\ell\ell1} > 4,\ m_{\ell\ell2} < 4$ GeV)$+p$	Sherpa	4.10	4878.0
$W^{\pm}\gamma + 4p$	Alpgen+Herwig	2440.76	149.5
$W^{\pm}\gamma + 5p$	Alpgen+Herwig	536.04	111.9
$W^{\pm}\gamma\ (\gamma \to e^+e^-)$	Alpgen+Herwig	82935.65	172.4
$W^{\pm}\gamma + p\ (\gamma \to e^+e^-)$	Alpgen+Herwig	30712.27	174.1
$W^{\pm}\gamma + 2p\ (\gamma \to e^+e^-)$	Alpgen+Herwig	13396.59	216.4
$W^{\pm}\gamma + 3p\ (\gamma \to e^+e^-)$	Alpgen+Herwig	5156.53	166.7

6.3 Selection of $2\ell0\tau_{\text{had}}$ events

The final state of the $t\bar{t}H$ production can be rather crowded as shown in Figure 6.1. The decay modes producing same-sign leptons also generate jets, b-jets and non-vanishing E_{T}^{miss}. Compared to other Standard Model processes producing same-sign leptons, the $t\bar{t}H$ have a generally larger jet and b-jet multiplicity, while large E_{T}^{miss} is present also in several

backgrounds. The selection of $t\bar{t}H$ events is therefore based on optimized requirements to target prompt leptons from $t\bar{t}H$ processes and on the counting of jets and b-jets in the event. The other analysis categories, 3ℓ, 4ℓ, $2\ell 1\tau_{\mathrm{had}}$ and $1\ell 2\tau_{\mathrm{had}}$, also adopt the same strategy.

Requirements aiming at reconstructing the sequence of decays initiated by the a $t\bar{t}H$ production have not been used in this analysis due to the combinatorial background arising from the several sources the leptons and jets can be originated by. In addition, the limited statistics expected in the signal regions of this analysis makes the comparison of continuous variables not more effective than a yields comparison. The expected number $t\bar{t}H$ events produced in the $\sqrt{s} = 8$ TeV collisions at the ATLAS interaction point is about 2500 and less than 2% of these events have decayed in a final state with same-sign leptons.

6.3.1 Lepton and jet selection

The techniques used to reconstruct and identify leptons and jets as well as the definition of a preliminary particle selection have been described in Section 2.4. The leptons and jets are required to satisfy additional conditions as explained in the following.

The selection of leptons and τ_{had} are listed in Table 6.3, while the conditions for selecting jets and b-jets are in Table 6.4. The leptons have requirements on the longitudinal and the transverse impact parameter of the lepton track in the Inner Detector (z_0 and d_0 respectively) to suppress leptons produced in the decay of long lifetime particles like heavy flavour hadrons and in interactions of other particles with the detector. Isolation requirements reduce the number of reconstructed leptons originated in a jet. The electron *very tight Likelihood* identification criterion is imposed to reduce the probability of hadrons to be reconstructed as electrons. B-jets are identified using a multivariate technique (MV1 b-tagging) as explained in Section 2.4.7. Several properties of the jet and of the tracks associated with it are considered to assess b-jet likeliness of the jet. The analysis also looks for τ_{had} in order to distinguish between the $2\ell 0\tau_{\mathrm{had}}$ and the $2\ell 1\tau_{\mathrm{had}}$ categories. The τ_{had} selection is described in Section 2.4.8, but the minimum p_{T} requirement is raised to 25 GeV in this analysis.

To avoid that multiple particles are reconstructed from the interaction of a single particle with the detector, reconstructed particles very close in direction are removed according to the rules in Table 6.5. The distance ΔR between any two particles is used in this process. When an electron is superimposed to a muon, the electron is removed. A muon can indeed deposit energy in the calorimeter and be reconstructed as an electron, while the electron is absorbed by the calorimeter and cannot generate a signal in the MS. The jet reconstruction algorithm considers the electron energy deposit in the electromagnetic calorimeter as candidate energy cluster for the jet reconstruction, therefore jets overlapping with electrons are removed. Hadrons forming jets can hardly reach the MS but muons can be produced in their decays. When a muon is superimposed to a jet, the muon has a

high chance to be a non-prompt one and is removed. The ΔR requirement is optimized to maximize the acceptance for prompt muons at a fixed rejection factor for non-prompt muon candidates.

The selection described above establishes the pool of particles and jets that are used to define the event categories used in this analysis. This common selection ensures that the dataset analysed in each category are independent and can be combined in the statistical analysis. Further requirements on particles are imposed to optimize the sensitivity of each category. The next section describes the selection of the $2\ell 0\tau_{\text{had}}$ category, while the other categories are described in Appendix C.

Muon	Electron	τ_{had}
$p_{\text{T}} > 10$ GeV	$p_{\text{T}} > 10$ GeV	$p_{\text{T}} > 25$ GeV
$\lvert z_0 \sin\theta \rvert < 1$ mm	$\lvert z_0 \sin\theta \rvert < 1$ mm	
$\lvert d_0 \rvert / \sigma(d_0) < 3$	$\lvert d_0 \rvert / \sigma(d_0) < 4$	
	very tight Likelihood ID	
$ptcone20/p_{\text{T}} < 0.1$	$ptcone20/p_{\text{T}} < 0.05$	
$etcone20/p_{\text{T}} < 0.1$	$etcone20/p_{\text{T}} < 0.05$	

Table 6.3: Baseline selections for light leptons and τ_{had}. A pre-selection is applied to the reconstructed particles as explained in Section 2.4.

Jet selection
$p_{\text{T}} > 25$ GeV
$\lvert\eta\rvert < 2.5$
B-jet selection
Pass jet selection and the MV1 b-tagging criterion

Table 6.4: Selection applied on reconstructed jets and b-jets. A pre-selection is applied to the reconstructed jets as explained in Section 2.4.6.

Kept Particle	Removed Particle	ΔR requirement
Muon	Electron	< 0.1
Electron (leading p_{T})	Electron (subleading p_{T})	< 0.1
Electron	Jet	< 0.3
Jet	Muon	$< 0.04 + \frac{10 \text{ GeV}}{p_{\text{T}}(\mu)}$
Electron	τ_{had}	< 0.2
Muon	τ_{had}	< 0.2
τ_{had}	Jet	< 0.3

Table 6.5: The sequence of criteria used to remove particles and jets that overlap in the $\eta \times \phi$ phase space.

6.3.2 $2\ell 0\tau_{\mathrm{had}}$ event category

The $2\ell 0\tau_{\mathrm{had}}$ category is defined by imposing conditions on the number of the reconstructed particles:

- number of reconstructed leptons, $N(\ell)$, equal to two
- number of reconstructed τ_{had}, $N(\tau_{\mathrm{had}})$, equal to zero

Other criteria are applied to define the signal region and other regions used for background estimate and validation. The selection of signal region events is summarized in Table 6.6. The leptons must have the same electric charge and $p_{\mathrm{T}} > 20$ GeV to reduce the background from non-prompt leptons. The isolation requirements for muons are tightened to further reduce the non-prompt background. Electrons with pseudorapidity $|\eta| > 1.37$ are rejected because of the high rate of charge misID electrons at large $|\eta|$.

The number of reconstructed jets $N(\text{jet})$ must be at least four including at least one identified as b-jet to reduce prompt backgrounds and low jet multiplicity non-prompt backgrounds. The selected events are separated by lepton flavour ($e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$) and number of jets (exactly four jets and at least five jets) into six categories with different fractions of signal and background. The higher jet multiplicity categories have larger signal fraction but smaller statistics. The division of the events in several categories lead to a higher overall sensitivity when the results from these categories are combined.

Events are acquired with the single lepton triggers described in Section 2.4.4. In the $e^{\pm}e^{\pm}$ category one of the selected electrons must have activated the SETrig. Similarly, one of the selected muons in the $\mu^{\pm}\mu^{\pm}$ category leptons has activated the SMTrig. Either the electron or the muon can activate the respective single lepton trigger in the $e^{\pm}\mu^{\pm}$ category. The triggered lepton must have $p_{\mathrm{T}} > 25$ GeV to be in the p_{T} range of high trigger acceptance.

As for the other analyses presented in this thesis, data must have been acquired during optimal running conditions, and events containing signals identified as detector effects, beam background and cosmic ray showers are rejected. In all events, the reconstructed interaction vertex with the highest $\sum p_{\mathrm{T}}^2(\text{track})$ must have at least three associated tracks with $p_{\mathrm{T}} > 0.4$ GeV.

The event selection for the other categories contributing to the measurement of the $t\bar{t}H$ cross-section are described in Section C.3.

Leptons	Jets	B-jets	τ_{had}
$N(\ell) = 2$ $\Sigma Q_\ell = \pm 2$ $\lvert\eta(\text{electron energy cluster})\rvert < 1.37$ $p_{\text{T}} > 20$ GeV ≥ 1 lepton with $p_{\text{T}} > 25$ GeV and trigger compatibility $ptcone20/p_{\text{T}} < 0.05$ $etcone20/p_{\text{T}} < 0.05$	$N(\text{jet}) \geq 4$	$N(b\text{-jet}) \geq 1$	$N(\tau_{\text{had}}) = 0$

Table 6.6: Event selections for the $2\ell 0\tau_{\text{had}}$ category.

6.4 Background estimate

The background that populates the $2\ell 0\tau_{\text{had}}$ signal regions is composed by prompt background, charge misID background from Standard Model processes with opposite-sign lepton pairs, processes with photon conversions and with non-prompt leptons. The origin of each background is described in Chapter 4. The processes contributing to these categories however differ from those considered for the new physics search described in Chapter 5. This is due mainly to the requirement of high jet multiplicity applied in this analysis that has not been considered before. The contributions from prompt backgrounds are determined with simulations. The contributions from charge misID and non-prompt backgrounds are determined with data-driven methods described in the following.

6.4.1 Prompt backgrounds

The Standard Model process with topology most similar to the $t\bar{t}H$ in the $2\ell 0\tau_{\text{had}}$ final state is the associated production of a top anti-top pair with a vector boson, either $W^\pm$ or Z/γ^*, called $t\bar{t}W^\pm$ and $t\bar{t}Z$, respectively. $t\bar{t}W^\pm$ is the biggest prompt background in the $2\ell 0\tau_{\text{had}}$ category and $t\bar{t}Z$ the second biggest. The leading order Feynman diagrams of $t\bar{t}W^\pm$, $t\bar{t}Z$ productions with the consequent decays in final states with positively charged same-sign leptons are shown in Figure 6.2.

Diboson production processes ($W^\pm Z$, ZZ and $W^\pm W^\pm$) can also enter the signal region selection if additional partons are produced from initial or final state radiation. The contributions with a virtual photon in place of a Z are also considered. These processes will be referred to as VV and their contributions in the $2\ell 0\tau_{\text{had}}$ category come in order of magnitude after $t\bar{t}W^\pm$ and $t\bar{t}Z$.

A smaller contribution comes from the single top production in association with a Z/γ^* (tZ). The irreducible $t\bar{t}W^+W^-$ background is also considered although small compared to other prompt backgrounds. Other rare processes are considered but their expected contribution to the signal regions yields is almost negligible. These processes include the Higgs boson production via gluon-gluon fusion and decay to $H \to ZZ^{(*)}$, the Higgs boson

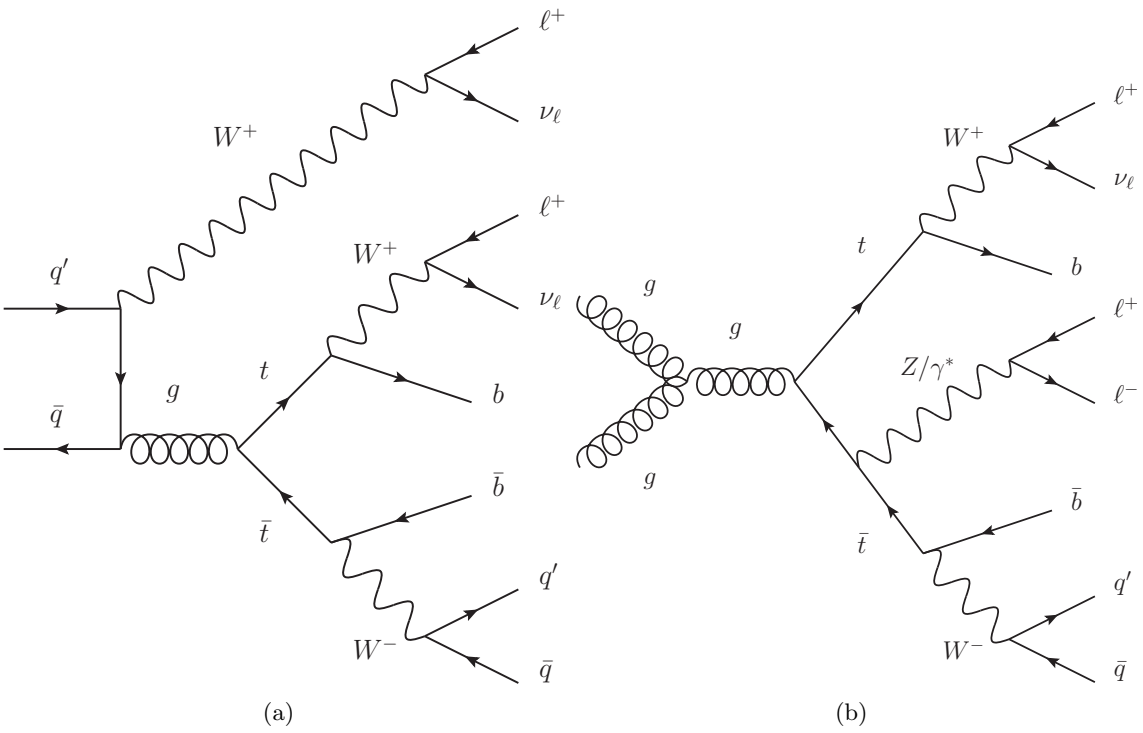


Figure 6.2: The Feynman diagram of tree level for $t\bar{t}W^{\pm}$ (a) and $t\bar{t}Z$ (b) production. The decay to final states with a positively charged lepton pair is shown.

production in association with a single top quark ($tHqb$, $tHW^{\pm}$) and $t\bar{t}t\bar{t}$ production. In the validation regions and signal regions shown in Section 6.5 and Section 6.7.1, respectively, the small contributions from tZ, $t\bar{t}W^+W^-$, $H \to ZZ^{(*)}$, $tHqb$, $tHW^{\pm}$ and $t\bar{t}t\bar{t}$ are plotted together and labeled as "Rare" to simplify the exposition.

The backgrounds with prompt leptons are modeled with simulations as described in Section 6.2. The simulated events with reconstructed non-prompt or charge misID leptons are discarded since already considered in the respective background estimate methods. All the simulations are corrected for differences between data and simulation in pileup conditions, particle detection efficiencies, energy calibration and energy resolution by means of data-driven corrections. These corrections are specified in Section 2.4.

6.4.2 Charge misID background

The events that constitute the charge misID background in this analysis produce high jet and b-jets multiplicities in addition to two prompt opposite-sign leptons. The $t\bar{t}$ production is the dominant source of this background followed by Z/γ^* events. Smaller backgrounds are due to $tW^{\pm}$ and W^+W^- production. Processes like $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ can also form opposite-sign leptons and can contribute to the charge misID background in a small amount.

Since the charge mis-identification is due to detection conditions, the ways in which

charge misID occurs for electrons and muons are different (see Section 4.2). The electron charge misID background is significant and its estimate is described in the following sections. The muon charge misID background is instead negligible as determined in in Section 4.2.2.

6.4.2.1 Estimate of the background with charge misID electrons

The strategy used to estimate the electron charge misID background in this analysis is similar to what used in the new physics search. First, the charge misID rate is estimated using the *likelihood method* described in Section 5.4.2.1. The rate obtained is not identical to that used in the new physics search due to differences in the electron selection requirements. The rate is then applied to data events passing the event selection but having two opposite-sign leptons in the final state to obtain the estimate of the charge misID background in the signal regions.

6.4.2.2 Measurement of the electron charge misID rate

Since the probability of an electron to be reconstructed with wrong charge does not depend on the process of production but rather on the kinematic of the electron, the charge misID rate can be measured in principle from any process. A sample of Z events is used to measure the charge misID rate since it can be selected with higher purity and higher statistics than other charge misID processes, like $t\bar{t}$. Events with two electrons selected according to the selection in Table 6.3 and either mutual charge are considered. The single electron trigger is used to acquire these events and the electron that activates the trigger must have $p_{\mathrm{T}} > 25$ GeV. The electron pairs must be compatible with a Z decay by restricting the $m\,(e, e)$ to near the Z mass value. For opposite-sign leptons, $m\,(e^+, e^-)$ must be between 78.5 and 102.3 GeV; for same-sign leptons, $76.5 < m\,(e^\pm, e^\pm) < 101.3$ GeV. The intervals of invariant mass have been determined by fitting the distribution of the Z peak with a Gaussian function and considering an interval as large as eight standard deviations centred at the mean value of the fit. For same-sign leptons the interval is larger and shifted to lower values due to the energy lost by trident event electrons.

The charge misID rate is then estimated with the *likelihood method* described in Section 5.4.2.1, as a function of $|\eta|$ and p_{T} of the electron. The electrons from $t\bar{t}$ events have generally p_{T} higher than those from Z decays. The statistics of Z events with high p_{T} (> 130 GeV) electron is not enough to make accurate measurements of the charge misID rates in that p_{T} range. Hence, the *likelihood method* is used to measure charge misID rates $\epsilon(|\eta|, p_{\mathrm{T}})$ in data up to 130 GeV of electron p_{T} and an extrapolation to higher p_{T} is performed based on $t\bar{t}$ simulations. The charge misID rate estimated for electrons with $p_{\mathrm{T}} \in [90, 130]$ GeV is multiplied by correction factors $\alpha_{t\bar{t}}$ as a function of $|\eta|$ obtained from

$t\bar{t}$ simulated events. The charge misID rates for high p_{T} electrons are defined as

$$\epsilon(|\eta|, p_{\mathrm{T}} > 130 \text{ GeV}) = \epsilon(|\eta|, p_{\mathrm{T}} \in [90, 130] \text{ GeV}) \times \alpha_{t\bar{t}}(|\eta|) \tag{6.1}$$

The correction $\alpha_{t\bar{t}}(|\eta|)$ is defined as the ratio of the charge misID rate for electron with $p_{\mathrm{T}} > 130$ GeV and that for electron with $p_{\mathrm{T}} \in [90, 130]$ GeV measured in the $t\bar{t}$ simulation:

$$\alpha_{t\bar{t}}(|\eta|) = \frac{\epsilon(|\eta|, p_{\mathrm{T}} > 130 \text{ GeV})_{t\bar{t}}}{\epsilon(|\eta|, p_{\mathrm{T}} \in [90, 130] \text{ GeV})_{t\bar{t}}} \tag{6.2}$$

The rates $\epsilon(|\eta|, p_{\mathrm{T}})_{t\bar{t}}$ are measured using the generator information of the simulation, i.e. by counting the fraction of reconstructed electrons with charge opposite to the electron generated in the $W^{\pm}$ decay. The extrapolation is valid under the hypothesis that the $t\bar{t}$ simulation predicts accurately the fraction of charge misID electrons with $p_{\mathrm{T}} > 130$ GeV compared to those with $p_{\mathrm{T}} \in [90, 130]$ GeV.

The estimated charge misID rates are shown in Figure 6.3. The error bars include statistical and systematic uncertainties.

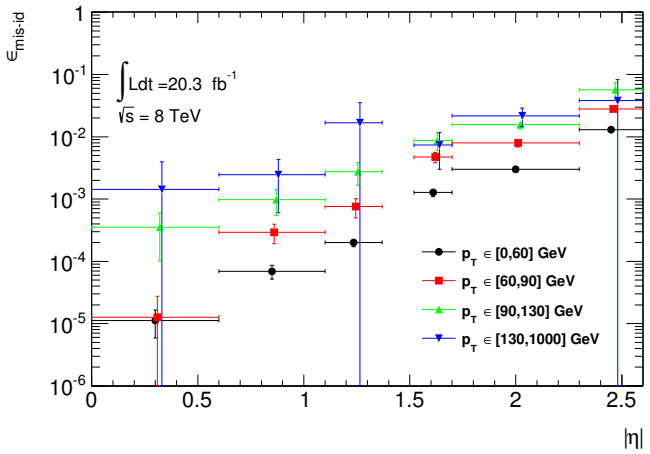


Figure 6.3: Electron charge misID rates measured with the *likelihood method* on Z data (black points, red squares and green triangles) as a function of electron $|\eta|$ and extrapolated for $p_{\mathrm{T}} > 130$ GeV (blue triangles). Error bars show the total uncertainties.

The uncertainties considered in the measurement of the charge misID rates are:

- The statistical uncertainty from the *likelihood method*. It is the dominant uncertainty for $p_{\mathrm{T}} < 130$ GeV.
- The statistical uncertainty on the extrapolation factor $\alpha_{t\bar{t}}(|\eta|)$ that affects only the $p_{\mathrm{T}} > 130$ GeV region and is dominant there.
- The variation of charge misID rates due to the variation of the Z mass interval of measurement by 1 σ. This uncertainty is bigger at low $|\eta|$ where the statistics collected is lower.

- The difference between the charge misID rates measured with the *likelihood method* on simulated $Z \to e^+e^-$ events and those evaluated using the generator information similarly to what done in Section 5.4.2.2.
- The difference between charge misID rates for positrons and electrons. This was investigated with a dedicated study that found some differences at large $|\eta|$ and negligible otherwise.

The magnitude of each uncertainty relative to the rate in the several $(|\eta|, p_{\mathrm{T}})$ intervals considered is shown in Figure 6.4. Since the sources of uncertainties are independent, the total uncertainty on the rates is calculated as the quadratic sum of the uncertainties.

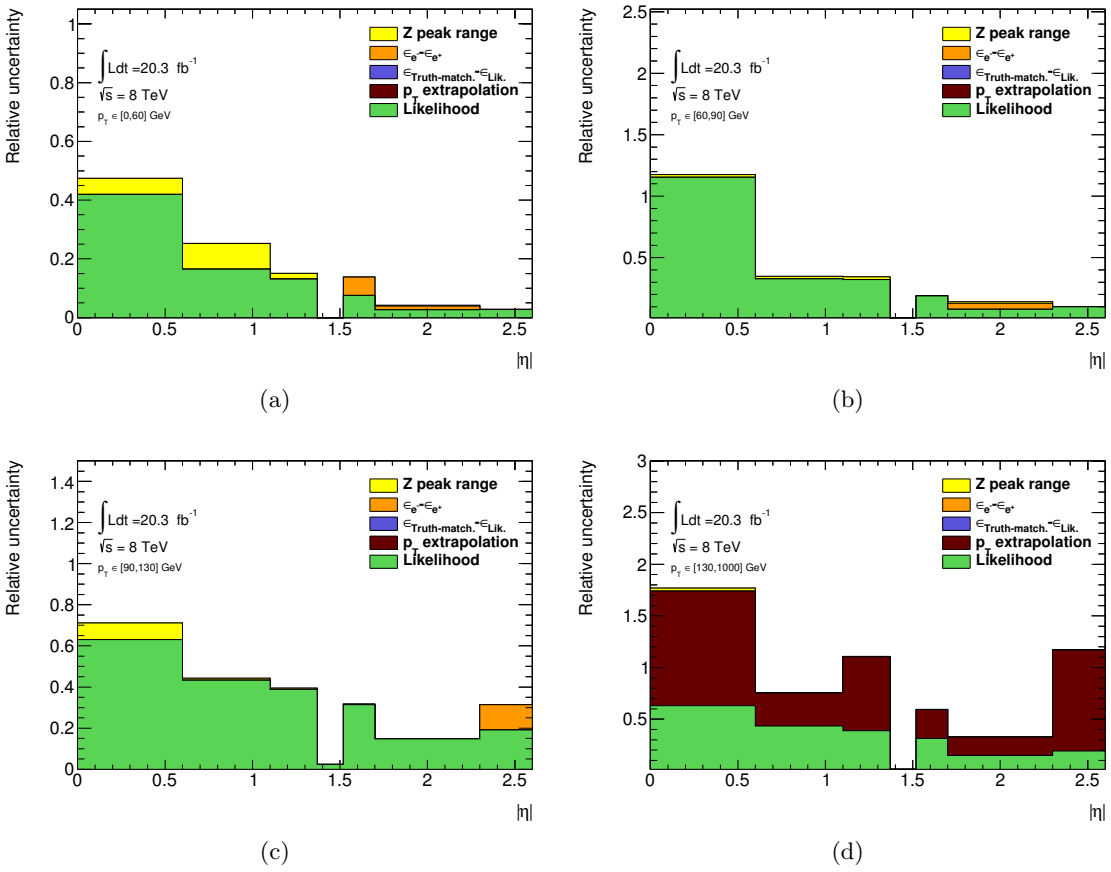


Figure 6.4: Relative uncertainty on the charge misID rate for intervals of $|\eta|$ and $p_{\mathrm{T}} \in [0, 60]$ GeV (a), $p_{\mathrm{T}} \in [60, 90]$ GeV (b), $p_{\mathrm{T}} \in [90, 130]$ GeV (c), $p_{\mathrm{T}} > 130$ GeV (d).

6.4.2.3 Determination of the charge misID background in the signal regions

Known the electron charge misID rates, it is possible to estimate the charge misID background in the analysis signal regions by applying these rates to events with opposite-sign leptons. The data events must be selected with the same selection criteria of the signal regions except for the relative charge sign requirement of the leptons. These samples are

very pure of prompt opposite-sign leptons and the non-prompt background is negligible. Events with an $e^{\pm}\mu^{\mp}$ pair are used to estimate the background in the $e^{\pm}\mu^{\pm}$ category; analogously events with a $e^{+}e^{-}$ pair the background in the $e^{\pm}e^{\pm}$ category. If N^i $e^{\pm}\mu^{\mp}$ events are produced with the electron in the $i-th$ interval of $|\eta|$ and p_{T}, then $N^i_{SS} = \epsilon_i \times N^i$ are expected to be reconstructed as same-sign leptons and $N^i_{OS} = (1-\epsilon_i) \times N^i$ as opposite-sign leptons. The estimated number of same-sign lepton events in the $e^{\pm}\mu^{\pm}$ category due to charge misID background is therefore:

$$N_{SS} = \sum_i \frac{\epsilon_i}{1-\epsilon_i} \times N^i_{OS} \tag{6.3}$$

where the sum is over all the possible intervals $(|\eta|, p_{\mathrm{T}})$. In the $e^{\pm}e^{\pm}$ category, the charge misID background estimate is complicated by the fact that both electrons can be reconstructed with opposite charge. The estimate number of charge misID events in the $e^{\pm}e^{\pm}$ category is:

$$N_{SS} = \sum_i \sum_{j \geq i} \frac{\epsilon_i + \epsilon_j - 2\epsilon_i\epsilon_j}{1-\epsilon_i-\epsilon_j+2\epsilon_i\epsilon_j} \times N^{ij}_{OS} \tag{6.4}$$

where both sums are over the $(|\eta|, p_{\mathrm{T}})$ intervals. In addition to the total event yields, one can obtain the distributions of the charge misID background from opposite-sign data distribution scaled with charge misID rate corrections. The shapes of the predicted pseudorapidity distribution of the leading p_{T} electron for the charge misID background and $t\bar{t}H$ signal in the $e^{\pm}e^{\pm}$ category are shown in Figure 6.5. The charge misID background has a clear excess at $|\eta| > 1.5$ due to the increase of charge misID rate with $|\eta|$, while in $t\bar{t}H$ events smaller values of $|\eta|$ are favoured. For this reason, events with an electron with $|\eta| > 1.5$ are excluded from the signal regions. Since the region $1.37 < |\eta_{cl}| < 1.52$ was already excluded due to the transition between calorimeter units, the electron considered in the signal region results in being limited to $|\eta_{cl}| < 1.37$.

There are few differences between this charge misID background estimate and that used in the new physics search (Section 5.4.2.1). The estimate used here is totally data-driven with charge misID rates estimated in data (except for electrons with $p_{\mathrm{T}} > 130$ GeV) and applied to opposite-sign data. In Chapter 5 instead, corrections derived from rates measured in data and simulation were applied to simulated events. The difference in strategy is justified by the type of analysis. This analysis compares the number of observed and expected events with a rare topology: large jet and b-jet multiplicities in addition to same-sign leptons. Only a tiny fraction of Z/γ^* and $t\bar{t}$ events is accepted and in order to produce accurate predictions, large simulated statistics would be needed. Further, the modeling of simulations in this rare topology has never been tested before and this would be reflected in large systematic uncertainties. The full data-driven approach is therefore preferred, but such method does not compensate for differences in kinematics between the

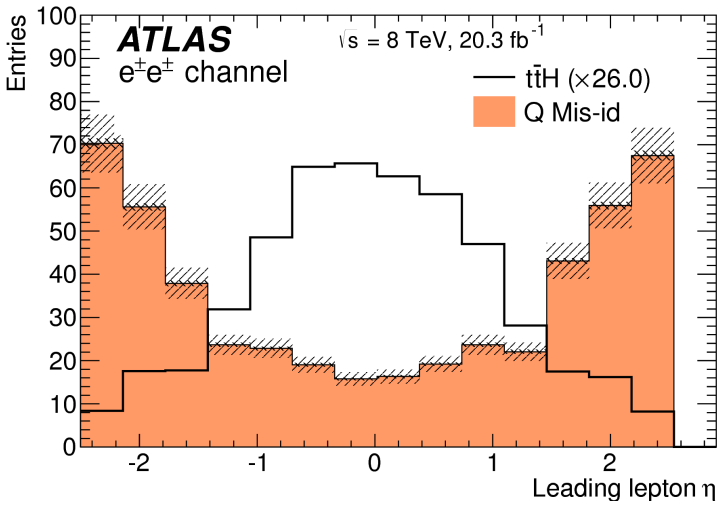


Figure 6.5: Pseudorapidity distribution of the higher p_{T} lepton in the $2\ell 0\tau_{\mathrm{had}}$ $e^{\pm}e^{\pm}$ categories for charge misID background and $t\bar{t}H$ signal expectation. The hatched region shows the uncertainty on the background prediction. The signal is magnified 26 times.

prompt leptons and the charge misID leptons. The latter ones, if created in a trident event, would have energy and direction different from the electron initiating the interaction. This has not a large impact on the $t\bar{t}H$ analysis because the analysis considers predictions in number of events in the signal region and, does not make use of the lepton kinematics[1]. In the search for new physics where the di-lepton invariant mass is scanned for excesses, the correct modeling of the kinematic is required. The simulation with corrected charge misID rates is hence used in that case.

6.4.3 Photon conversion backgrounds

Processes radiating a photon and having additional prompt leptons and jets can be a background for the $t\bar{t}H$ analysis when the photon converts in a $e^{+}e^{-}$ pair due to interaction with detector material. These events are accounted for in the simulation of the prompt backgrounds. In addition, $W^{\pm}\gamma$ events are simulated and their contribution is added to the contributions from the other VV processes. These simulations are corrected for detection differences between data and simulation as specified in Section 2.4.

6.4.4 Non-prompt backgrounds

Processes that produce one prompt lepton and large jet and b-jet multiplicities in the final state constitute the non-prompt background for this analysis. In these processes, a non-prompt lepton is produced in the jet activity and forms with a prompt lepton the

[1]The analysis imposes p_{T} and $|\eta|$ restrictions but does not used the lepton kinematic distribution to measure the signal. The correction of the charge misID kinematic is therefore not necessary.

same-sign pair. For this analysis, the $t\bar{t}$ production is the largest non-prompt background. Smaller sources are *single-top*, $tW^{\pm}$, $W^{\pm}$ + jets, Z + jets and $W^{+}W^{-}$.

A $t\bar{t}$ simulation is studied to investigate the type of non-prompt leptons that pass the analysis selection. The $t\bar{t}$ must have two reconstructed same-sign leptons, one prompt and one non-prompt, and one reconstructed b-jet. The lepton type is determined accessing the generator information of the simulation. The 95% of non-prompt electrons and the 99% of non-prompt muons are produced in the decay of hadrons containing b-quarks. In other words, the non-prompt background is nearly produced by a single source which makes simpler and more reliable estimate. The remaining non-prompt leptons are produced by other hadrons or are hadrons reconstructed as leptons.

Since the jet multiplicity is a key element of the non-prompt background estimate, the fraction of non-prompt leptons from hadron containing b-quarks is also studied as a function of the jet multiplicity in the event. This fraction is steady for muons and decreases at higher multiplicity for electrons due to the larger number of light-flavour particles produced in jets that can be reconstructed as electrons. The variation of the b-quarks originated non-prompt fraction does not exceed 10%.

6.4.4.1 The θ Method

The non-prompt background is estimated in this analysis with a simplified version of the Fake Factor Method introduced in Section 5.4.4.1. This method is called the θ Method and uses correction factors θ_{ℓ} and the number of non-prompt events estimated in orthogonal non-prompt background dominated regions to predict the non-prompt background yields in the signal regions. The estimate of the non-prompt background in the signal regions where the leptons satisfy a selection called here *tight* (ℓ) consists of:

- Definition of a *non-tight* ($\not{\ell}$) lepton selection to target non-prompt leptons.
- Definition of a non-prompt-dominated region orthogonal to the signal regions.
- Evaluation of the θ_{ℓ} factor, the fraction of events with *tight* leptons to events with *non-tight* leptons in the non-prompt-dominated region.
- Application of the θ_{ℓ} factor to events with a *non-tight* lepton, but satisfying the signal region selection otherwise.

The *tight* selection for electrons and muons correspond to that used in the $2\ell 0\tau_{\text{had}}$ signal region, defined in Table 6.3 with additional requirements from Table 6.6. The *non-tight* selection is designed to target non-prompt leptons that, in this case, are the soft and not isolated leptons typical of the decay of heavy flavour hadrons.

The definitions of *non-tight* muon differs from the *tight* one for the p_{T} requirement: the *non-tight* muon has $6 < p_{\text{T}} < 10$ GeV. Further a *non-tight* muon directionally overlapping with a jet according to the requirement in Table 6.5 is not removed. This is due to the

p_{T} dependent ΔR requirement ($\Delta R\,(jet, \mu) < 0.04 + \frac{10\ \mathrm{GeV}}{p_{\mathrm{T}}(\mu)}$) that would be otherwise too large.

For electrons, the *non-tight* selection differs from the *tight* one for the identification and isolation criteria: *non-tight* electrons fails the *very tight Likelihood* criterion, have $ptcone20/p_{\mathrm{T}} > 0.05$ and have $etcone20/p_{\mathrm{T}} > 0.05$.

The θ_ℓ factor is the ratio of the number of ℓ non-prompt leptons to the number of $\not{\ell}$ non-prompt leptons calculated in the non-prompt dominated region. θ_ℓ has the same meaning of the fake factor f_ℓ^{FF}, but is not calculated as a function of the lepton kinematic. In this analysis, the θ_ℓ factors are estimated in a region with two same-sign and same flavour leptons, either $e^\pm e^\pm$ or $\mu^\pm\mu^\pm$, at least one b-jet and low jet multiplicity: $1 \leq N\,(\mathrm{jet}) \leq 3$ in $e^\pm e^\pm$ events and $2 \leq N\,(\mathrm{jet}) \leq 3$ in $\mu^\pm\mu^\pm$ events. In the $\mu^\pm\mu^\pm$ case, the restriction to two or three jets enhances the fraction of $t\bar{t}$ events in the non-prompt background. The one jet events are indeed enriched in other processes like $W^\pm$ + jets. On the other side, $e^\pm e^\pm$ events with $|m\,(e^\pm, e^\pm) - m\,(Z)\,| < 10$ GeV are removed to reduce the fraction of charge misID background in this region. Only events with two *tight* leptons ($N_{\ell\ell}$) or with one *tight* lepton and one *non-tight* ($N_{\ell\not{\ell}}$) are considered.

The θ_ℓ factors is calculated as the ratio of $N_{\ell\ell}$ to $N_{\ell\not{\ell}}$:

$$\theta_\ell = \frac{N_{\ell\ell}}{N_{\ell\not{\ell}}} \tag{6.5}$$

Any contribution from prompt, charge misID and photon conversion background to the yields in these regions are subtracted from data before θ_ℓ is calculated. The θ_ℓ factors for electrons, θ_e, and for muons, θ_μ, are obtained from $e^\pm e^\pm$ and $\mu^\pm\mu^\pm$ events, respectively, and are given in Table 6.7. Note the larger value of θ_μ does not mean that the fraction

Factor	θ_e	θ_μ
Value	0.016 ± 0.006	0.25 ± 0.05

Table 6.7: The measured values of the θ_ℓ factors.

of background with non-prompt muons in the signal regions is larger, but just that the fraction of non-prompt muons selected as *tight* in relation to those selected as *non-tight* is higher.

Assumed that the θ_ℓ factors do not depend on the multiplicity of jets in the event, the estimate of non-prompt events with two *tight* leptons and n jets is obtained by multiplying the number of events with n jets and one *tight* +*non-tight* lepton pair by θ_ℓ. The non-prompt background estimates produced by the θ Method, $N^\theta_{\ell\ell'}$, in each category $\ell, \ell' = e, \mu$ are:

$$N^\theta_{ee}\,(n\ \mathrm{jets}) = N_{e\not{e}}\,(n\ \mathrm{jets}) \times \theta_e \tag{6.6}$$

$$N^{\theta}_{\mu\mu}\,(n \text{ jets}) = N_{\mu\not{\ell}}\,(n \text{ jets}) \times \theta_{\mu} \tag{6.7}$$

$$N^{\theta}_{e\mu}\,(n \text{ jets}) = N_{e\not{\ell}}\,(n \text{ jets}) \times \theta_{\mu} + N_{\mu\not{\ell}}\,(n \text{ jets}) \times \theta_{e} \tag{6.8}$$

where $N_{\ell\not{\ell}}$ is the number of data events with one *tight* +*non-tight* lepton pair after subtraction of prompt, charge misID and photon conversion backgrounds. The charge misID background in events with a *non-tight* electron is not estimated with the *likelihood method*. The charge misID contributions from $t\bar{t}$, Z/γ^{*} + jets, $tW^{\pm}$ and $W^{+}W^{-}$ productions are estimated with pure simulations. The simulations considered are given in Table C.1.

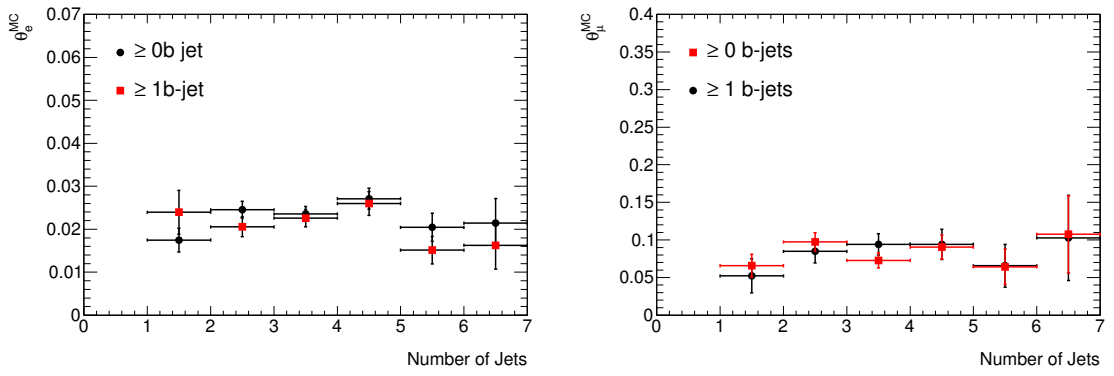
The non-prompt background predictions for the $2\ell 0\tau_{\text{had}}$ signal regions are given by Equation 6.6, Equation 6.7 and Equation 6.8 for number of jets $n = 4$ and $n \geq 5$. Since the $e^{\pm}\mu^{\pm}$ events are not used in the determination of θ_e and θ_μ, Equation 6.8 with $1 < n \leq 3$ can be used as validation region of the non-prompt background estimate. $N^{\theta}_{e\mu}$ is compared with data after background subtraction in this region with 2 or 3 jets and 1 b-jet and an $e^{\pm}\mu^{\pm}$ pair:

$$\frac{N^{\theta}_{e\mu}}{N_{e\mu}(\text{Data}) - N_{e\mu}(\text{prompt} + \text{charge misID} + \text{photonconversion})} = 1.43 \pm 0.46 \text{ (stat.)} \tag{6.9}$$

The θ Method is based on the assumption that the θ_ℓ factors for electrons and muons are stable with respect to the jet multiplicity. This has been tested using $t\bar{t}$ simulations. The selected events contain one non-prompt and one prompt lepton determined via generator information with same charge and can contain one reconstructed b-jet. The value of θ_e and θ_μ calculated from these events as a function of the jet multiplicity is shown in Figure 6.6. The θ_ℓ seem compatible within the statistical error for any jet multiplicity. Lower θ_e values are found for $N\,(\text{jet}) \geq 5$, but these are measured with less accuracy. The θ_μ value for $N\,(\text{jet}) = 1$ is smaller too, but events with one jet are excluded from the estimate.

As a test of the θ Method, the prediction of $t\bar{t}$ non-prompt background from simulation $N^{MC}_{\ell\ell}$ is compared to $N^{MC}_{\ell\not{\ell}} \times \theta^{MC}_{\ell}$, the estimate from the θ Method applied to the simulated events with the θ^{MC}_{ℓ} factors measured in the same simulation (closure test). The test is repeated for different numbers of b-jets and variations of the lepton selection and of the generator used in the simulation. The largest difference between $N^{MC}_{\ell\ell}$ and $N^{MC}_{\ell\not{\ell}} \times \theta^{MC}_{\ell}$ among any of these cases is taken as systematic uncertainty of the method ($\Delta\theta^{\text{syst}}_{\mu}$(closure) and $\Delta\theta^{\text{syst}}_{e}$(closure)). Additional systematic uncertainties are considered.

The uncertainties on the background (prompt, charge misID and photon conversion) subtracted from data when estimating θ_ℓ and $N_{\ell\not{\ell}}$, are accounted for and propagated to the non-prompt estimate. The correlations between the uncertainties on the subtraction and those on the prompt, charge misID and photon conversion estimates are taken into



(a) (b)

Figure 6.6: Stability of θ_e and θ_μ in $t\bar{t}$ simulated events as a function of the number of jets and requiring any number of b-jets (black) or at least one b-jet (red).

account. Since charge misID rate is not calculated for ℓ electrons, for $N_{\ell\ell}$ events, the subtracted charge misID background is estimated with simulations and an uncertainty on this estimate of 100% is applied.

The influence of sources of non-prompt background other than $t\bar{t}$ production to the prediction of non-prompt background is evaluated. Processes like $W^{\pm}$+jets and Z/γ^*+jets are expected to be suppressed at high jet multiplicity. However, the fraction of $W^{\pm}+$jets and Z/γ^*+jets in the low jet multiplicity used to estimate the θ_ℓ factors is larger. If for these processes θ_ℓ is different from that measured in $t\bar{t}$ events, a bias can be introduced in the non-prompt background estimate. The fraction of single boson production in the studied sample is varied by changing the b-jet p_{T} requirement and introducing a requirement on the E_{T}^{miss}. The largest variation in the value of θ_ℓ obtained from these tests is taken as systematic uncertainty due to non-$t\bar{t}$ processes on the non-prompt background estimate ($\Delta\theta_\mu^{\mathrm{syst}}$(non-$t\bar{t}$) and $\Delta\theta_e^{\mathrm{syst}}$(non-$t\bar{t}$)).

In addition to the uncertainties listed above, the statistical uncertainties on the θ_ℓ ($\Delta\theta_e^{\mathrm{stat}}$ and $\Delta\theta_\mu^{\mathrm{stat}}$) and on the yields in the $\ell\ell$ regions ($\Delta N_{\ell\ell}^{\mathrm{stat}}$) are also accounted. The total uncertainties on the non-prompt predictions are given in Table 6.8 for each signal region category. The larger uncertainties are due to the statistical error and the differences in the closure tests with $t\bar{t}$ simulations.

The θ Method is different from the Fake Factor Method presented in Section 5.4.4.1. One difference is that θ_ℓ is an overall factor and does not account for lepton kinematic dependences. In other words, if the same *tight* and *non-tight* definitions are used for the θ Method and the Fake Factor Method, θ_ℓ would be the average fake factor:

$$\theta = \langle f\,(p_{\mathrm{T}}, \eta)\rangle \tag{6.10}$$

Uncertainties (%)		Channels					
		4 jets			≥ 5 jets		
		$e^{\pm}e^{\pm}$	$\mu^{\pm}\mu^{\pm}$	$e^{\pm}\mu^{\pm}$	$e^{\pm}e^{\pm}$	$\mu^{\pm}\mu^{\pm}$	$e^{\pm}\mu^{\pm}$
Statistical	$\Delta\theta_e^{\text{stat}}$	40	–	9.7	40	–	14
	$\Delta\theta_\mu^{\text{stat}}$	–	22	16	–	22	14
	$\Delta N_{\ell\ell}^{\text{stat}}$	6.8	20	13	8.4	30	16
Systematic	$\Delta\theta_e^{\text{syst}}$(closure)	22	–	6.8	27	–	8.9
	$\Delta\theta_\mu^{\text{syst}}$(closure)	–	25	19	–	31	21
	$\Delta\theta_\mu^{\text{syst}}$(non-$t\bar{t}$)	–	14	11	–	14	9.2
	$\Delta\theta_e^{\text{syst}}$(non-$t\bar{t}$)	19	–	4.7	19	–	6.5
	charge misID ($\ell\ell$)	2.2	–	0.7	2.4	–	0.8
Total		**50**	**41**	**32**	**52**	**50**	**37**
Correlated Systematic	charge misID ($\ell\ell$)	24	–	6.8	24	–	7.9
	$t\bar{t}W^{\pm}$ cross-section	4.0	4.1	5.1	4.0	4.1	4.5
	$t\bar{t}W^{\pm}$ PDFs[2]	2.7	2.7	1.8	2.8	2.7	3.2
	$t\bar{t}Z$	0.5	0.5	0.5	0.5	0.5	0.6

Table 6.8: Summary of the uncertainties (in %) on the non-prompt background estimate in the $e^{\pm}e^{\pm}$, $\mu^{\pm}\mu^{\pm}$ and $e^{\pm}\mu^{\pm}$ categories. The "correlated" systematic uncertainties are anti-correlated to the systematic on other background processes on the signal region.

Unless the fake factors are not affected by the lepton kinematic, the θ Method would not predict accurately the kinematic properties of the non-prompt background. This does not create a problem in this analysis because the kinematic properties of the signal region events are not investigated. The results are obtained from the comparison of data yields with the prediction. Since the available statistics for this analysis is limited, the fake factors would have been dominated by the statistical error, preventing the observation of any significant p_{T} or η dependence.

Another difference with the Fake Factor Method is that in the θ Method also assumes that the non-prompt background due to events with two non-prompt leptons is considered negligible compared to the one with only one non-prompt lepton. This is a reasonable assumption since the main non-prompt background is $t\bar{t}$ with one prompt lepton.

In the next years, ATLAS will collect more statistics and more accurate background estimates will be feasible. The uncertainties related to the non-prompt background estimate have the largest impact on this measurement as will be shown later. One goal for future analyses is to improve the non-prompt background estimate. Alternative methods of estimate have been studied by the author and described in the following section.

6.4.5 Non-prompt background estimates for future analyses

In the second half of 2015, LHC has started to collide protons at a centre-of-mass energy of 13 TeV. The $t\bar{t}$ production cross-section at this energy is more than three times larger the one at $\sqrt{s} = 8$ TeV. Furthermore, the LHC has the potential to produce data in the order of

tens of inverse femtobarns in 2016 and 2017. As a result, the number of $t\bar{t}$ events produced could be several times larger than in 2012. Assuming detection conditions similar to those at $\sqrt{s} = 8$ TeV[3], the non-prompt background due to $t\bar{t}$ would also increase by a similar fraction. In terms of non-prompt background estimate, each region used will contain more non-prompt background events leading to a reduction of the statistical uncertainty on the non-prompt estimate. More complex techniques could therefore be employed including those resolving the kinematics of the signal events against the non-prompt background and offering more discrimination power. This would allow to investigate alternative analysis strategies. For instance, a template fit of kinematic distributions in the signal regions could be used in place of the yields comparison to obtain higher sensitivity.

This section investigates the use of the Fake Factor Method, introduced in Section 5.4.4.1, and another technique called the Matrix Method [183] as candidate methods to estimate non-prompt background in future $t\bar{t}H$ analyses. The Matrix Method is introduced in the next section. Both methods require the definition of two levels of lepton selection and a non-prompt background enriched region where correction factors are calculated. The lepton selections used in this study are defined in Section 6.4.5.2, while several non-prompt background enriched regions are investigated in Section 6.4.5.3. The Fake Factor Method and Matrix Method predictions are then compared using simulated $t\bar{t}$ events and $\sqrt{s} = 8$ TeV data in Section 6.4.5.5 and Section 6.4.5.6, respectively.

6.4.5.1 The Matrix Method

The Matrix Method estimate of non-prompt background is based on the calculation of the efficiencies for non-prompt and prompt leptons to pass certain selection criteria. These efficiencies are used to scale data in non-prompt enriched regions and obtain the non-prompt background prediction in the signal regions. This method accounts for backgrounds with either one or multiple non-prompt leptons. The Matrix Method is based on the definition of:

- a *loose* and a *tight* lepton selection (as in the Fake Factor Method);
- a non-prompt lepton dominated region orthogonal to the signal regions (similarly to the Fake Factor Method) used to calculate the efficiency for non-prompt leptons selected with the *loose* requirements to pass the *tight* selection;
- a prompt lepton dominated region orthogonal to the signal regions used to calculate the efficiency for prompt leptons selected with the *loose* requirements to pass the *tight* selection;
- a data sample with *loose* leptons scaled with factors built from prompt and non-prompt efficiencies to obtain the non-prompt background prediction.

[3]Differences in the running conditions, e.g. pileup, and in the reconstruction techniques could alter the efficiency to select non-prompt leptons.

Equivalently to the Fake Factor Method, reconstructed leptons have to pass a preliminary *loose* ($\breve{\ell}$) selection. A mix of prompt and non-prompt leptons is selected. Implications due to the presence of leptons from other background sources are discussed later. The *loose* leptons are divided in *tight* (ℓ) and *non-tight* ($\not{\ell}$) leptons depending on whether the lepton satisfies the *tight* selection or not, respectively. The *tight* selection matches the leptons selection used in the signal regions and aims at reducing the fraction of non-prompt leptons. Typically this is obtained by imposing isolation and identification requirements. As a consequence, non-prompt leptons are predominantly classified as *non-tight* leptons.

The efficiencies for non-prompt and prompt leptons to pass the *tight* selection, f_ℓ and r_ℓ respectively, are calculated in ad-hoc regions populated by events with non-prompt or prompt leptons:

$$f_\ell = \frac{N_\ell\,(\text{non} - \text{prompt lepton})}{N_{\breve{\ell}}\,(\text{non} - \text{prompt lepton})} \tag{6.11}$$

$$r_\ell = \frac{N_\ell\,(\text{prompt lepton})}{N_{\breve{\ell}}\,(\text{prompt lepton})} \tag{6.12}$$

where N_ℓ and $N_{\breve{\ell}}$ are respectively the number of leptons of the type specified to pass the *tight* and *loose* selection, respectively. The efficiency for prompt leptons to pass the *tight* selection must be much higher than that for non-prompt leptons to obtain an accurate estimate of the non-prompt background. The efficiencies for prompt and non-prompt leptons are calculated in regions orthogonal to the signal regions. Since the characteristics of the event can affect the probability of leptons to pass the *loose* and *tight* selections, the definition of these orthogonal regions should be as close as possible to the signal regions. For non-prompt enriched regions, the composition of the selected non-prompt lepton sample (e.g. the fractions of non-prompt leptons from heavy flavour hadrons and mis-identified particles) must be similar to the one in the signal region to not introduce biases. Candidates for non-prompt enriched regions are studied in Section 6.4.5.3. The f_ℓ and r_ℓ are calculated in data, but potential contaminations from other types of leptons (charge misID, photon conversion electrons and non-prompt leptons for r_ℓ or prompt leptons for f_ℓ) are subtracted from N_ℓ and $N_{\breve{\ell}}$.

The f_ℓ and r_ℓ can be calculated as a function of a set of parameters, for instance the lepton $|\eta|$ and p_{T}. This allows to better model the kinematics of the non-prompt background instead of just predicting yields. The efficiency of the *tight* selection is indeed expected to not be constant because the event kinematic influences the selection requirements. The variables that are expected to have the larger impact on f_ℓ and r_ℓ are the lepton $|\eta|$ and p_{T} because of differences in the detection performances and selection efficiency of leptons depending on their direction and momentum. Once the prompt and non-prompt efficiencies are obtained, it is possible to relate the number of events with prompt and non-prompt leptons selected with the *loose* requirement, N_r and N_f, to the number of expected events with *tight* and *non-tight* selected leptons, N_ℓ and $N_{\not{\ell}}$. For events with one lepton, this

relation takes the form of:

$$\begin{pmatrix} N_\ell \\ N_{\not\ell} \end{pmatrix} = \begin{pmatrix} r_\ell & f_\ell \\ 1-r_\ell & 1-f_\ell \end{pmatrix} \begin{pmatrix} N_r \\ N_f \end{pmatrix} = M\left(r_\ell, f_\ell\right) \begin{pmatrix} N_r \\ N_f \end{pmatrix} \tag{6.13}$$

The matrix $M\left(r_\ell, f_\ell\right)$, which this method takes its name from, can be inverted (if $r_\ell \neq f_\ell$) to find N_f and consequently the number of events with one non-prompt lepton passing the *tight* selection, $N_{f\to\ell} = f_\ell N_f$.

In the case of events with two leptons, the estimate of the non-prompt background is more complex. The events can have lepton pairs formed by any combination of prompt and non-prompt leptons. For convenience, events are divided in N_{rr}, N_{rf}, N_{fr} and N_{ff}, where the first subscript refers to the type of the leading p_T lepton and the second to the type of the subleading one. The subscripts r indicates a prompt lepton, while f a non-prompt one. The prompt and non-prompt efficiencies are used to obtain the relations that connect N_{rr}, N_{rf}, N_{fr} and N_{ff} to the numbers of events with any combination of *tight* (ℓ) and *non-tight* ($\not\ell$) reconstructed leptons ($N_{\ell\ell}$, $N_{\ell\not\ell}$, $N_{\not\ell\ell}$ and $N_{\not\ell\not\ell}$):

$$\begin{pmatrix} N_{\ell\ell} \\ N_{\ell\not\ell} \\ N_{\not\ell\ell} \\ N_{\not\ell\not\ell} \end{pmatrix} = M\left(r_1, r_2, f_1, f_2\right) \begin{pmatrix} N_{rr} \\ N_{rf} \\ N_{fr} \\ N_{ff} \end{pmatrix} \tag{6.14}$$

$$M\left(r_1, r_2, f_1, f_2\right) = \begin{pmatrix} r_1 r_2 & r_1 f_2 & f_1 r_2 & f_1 f_2 \\ r_1(1-r_2) & r_1(1-f_2) & f_1(1-r_2) & f_1(1-f_2) \\ (1-r_1) r_2 & (1-r_1) f_2 & (1-f_1) r_2 & (1-f_1) f_2 \\ (1-r_1)(1-r_2) & (1-r_1)(1-f_2) & (1-f_1)(1-r_2) & (1-f_1)(1-f_2) \end{pmatrix}$$

where the subscripts 1 and 2 refer respectively to the leading and subleading p_T lepton in the pair. The efficiencies r_1 (f_1) can be different from r_2 (f_2) because they are parametrised according to the p_T, $|\eta|$ and flavour of the lepton. The matrix $M\left(r_1, r_2, f_1, f_2\right)$ of Equation 6.14 can be inverted if $r_1 \neq f_1$ and $r_2 \neq f_2$ in order to obtain the number of events with *loose* lepton pairs with non-prompt leptons: N_{rf}, N_{fr} and N_{ff}. The conditions $r_1 > f_1$ and $r_2 > f_2$ are typically satisfied because the *tight* selection is defined such that is more efficiently passed by prompt leptons than non-prompt ones.

The number of events with non-prompt leptons that pass the *tight* selection, ${N_{f\to\ell\ell}}^{MM}$, is obtained multiplying N_{rf}, N_{fr} and N_{ff} for the efficiencies of the leptons that constitute the pair to pass the *tight* selection:

$${N_{f\to\ell\ell}}^{MM} = r_1 f_2 N_{rf} + f_1 r_2 N_{fr} + f_1 f_2 N_{ff} \tag{6.15}$$

Equation 6.15 can also be rewritten in terms of $N_{\ell\ell}$, $N_{\ell\not\ell}$, $N_{\not\ell\ell}$ and $N_{\not\ell\not\ell}$ as:

$$\begin{aligned} {N_{f\to\ell\ell}}^{MM} &= AN_{\ell\ell} + BN_{\ell\not\ell} + CN_{\not\ell\ell} + DN_{\not\ell\not\ell} \\ A &= 1 - r_1 r_2 (1-f_1)(1-f_2)\alpha \\ B &= r_1 r_2 (1-f_1) f_2 \alpha \\ C &= r_1 r_2 f_1 (1-f_2)\alpha \\ D &= -r_1 r_2 f_1 f_2 \alpha \\ \alpha &= \tfrac{1}{(r_1-f_1)(r_2-f_2)} \end{aligned} \tag{6.16}$$

Note that $A, D \leq 0$ and $B, C \geq 0$ for $r_1 > f_1$ and $r_2 > f_2$. The closer r_i is to f_i, the larger $|A|$, B, C and $|D|$ become and, consequently, the more sensitive ${N_{f\to\ell\ell}}^{MM}$ is to statistical fluctuations on $N_{\ell\ell}$, $N_{\ell\not\ell}$, $N_{\not\ell\ell}$ and $N_{\not\ell\not\ell}$. Therefore, the Matrix Method works best when $r_i \gg f_i$.

To account for the dependence of f_ℓ and r_ℓ on the lepton p_{T}, $|\eta|$ and flavour, Equation 6.16 becomes:

$$\begin{aligned} {N_{f\to\ell\ell}}^{MM} = &\sum^{N_{\ell\ell}} A\left(\ell_1, p_{\mathrm{T}1}, |\eta_1|, \ell_2, p_{\mathrm{T}2}, |\eta_2|\right) + \sum^{N_{\ell\not\ell}} B\left(\ell_1, p_{\mathrm{T}1}, |\eta_1|, \ell_2, p_{\mathrm{T}2}, |\eta_2|\right) + \\ &+ \sum^{N_{\not\ell\ell}} C\left(\ell_1, p_{\mathrm{T}1}, |\eta_1|, \ell_2, p_{\mathrm{T}2}, |\eta_2|\right) + \sum^{N_{\not\ell\not\ell}} D\left(\ell_1, p_{\mathrm{T}1}, |\eta_1|, \ell_2, p_{\mathrm{T}2}, |\eta_2|\right) \end{aligned} \tag{6.17}$$

where the coefficients A, B, C, D are now calculated with the efficiencies $f_i = f_i\left(\ell_i, p_{\mathrm{T}i}, |\eta_i|\right)$ and $r_i = r_i\left(\ell_i, p_{\mathrm{T}i}, |\eta_i|\right)$ that are functions of the lepton properties (transverse momentum, pseudorapidity and flavour). The ${N_{f\to\ell\ell}}^{MM}$, as computed in Equation 6.17, is the Matrix Method estimate of the non-prompt background in events with two leptons. In the limit $r_\ell \to 1$, ${N_{f\to\ell\ell}}^{MM}$ is equal to ${N_{\ell\ell}}^{FF}$ (Equation 5.11) used in the Fake Factor Method[4].

The main difference between the Matrix Method and Fake Factor Method lies in the way in which the backgrounds other than the non-prompt one are accounted for. In the Fake Factor Method, background events other than non-prompt ones are subtracted from $N_{\ell\not\ell}$, $N_{\not\ell\ell}$ and $N_{\not\ell\not\ell}$ (Equation 5.13), while in the Matrix Method the prompt background is excluded by definition. In this case, the events are scaled with different factors, which account for prompt background contaminations.

The description of the Matrix Method given above considers only events with prompt and non-prompt leptons. Special considerations must be made for the $e^\pm e^\pm$ and $e^\pm \mu^\pm$ categories where events with electrons from charge mis-identification or photon conversion constitute a significant background. In this analysis, the photon conversion background is almost negligible, but the charge misID background can be comparable to the non-prompt and prompt backgrounds in magnitude. The Matrix Method provides the correct

[4]Note that the non-prompt efficiency is connected to the fake factor by the relation: $f_\ell^{FF} = \frac{f_\ell}{1-f_\ell}$.

estimate of non-prompt background if the efficiency for charge misID and photon conversion electrons to pass the *tight* selection is equal to that for prompt electrons. If not, the Matrix Method would still give the correct estimate for the other backgrounds if r_ℓ is not interpreted any more as an efficiency for prompt leptons but rather as an average efficiency for leptons from any process but the non-prompt ones. In that case, N_{rf}, N_{fr} and N_{rr} would represent the events with at least one lepton that is not a non-prompt lepton. The efficiency r_ℓ should be measured in a region with a composition of prompt, charge misID and photon conversion electron events as close as possible to the one expected in the signal regions. Alternatively, efficiencies should be calculated in regions enhanced with either prompt or charge misID or photon conversion electrons and then combined according to their expected fractions in the signal regions. These complications could be avoided by choosing an appropriate *loose* electron definition, such that r_ℓ for prompt and charge misID electrons are similar. In fact, r_ℓ and f_ℓ depend on the requirements chosen to differentiate between *loose* and *tight* leptons. The *tight* lepton definition is typically optimized to reject background in the signal regions, while the *loose* selection can be adjusted as best suits for the non-prompt estimate. The definition of *loose* electron can be calibrated in order to have small efficiency for non-prompt electrons but large and similar efficiencies for prompt, charge misID and photon conversion electrons. Considerations on the *loose* selection definitions are given below.

As explained in Section 4.2, there are several types of charge misID electrons. Their reconstructed signals can present differences from those of prompt leptons, leading to differences in selection efficiencies. If the charge misID electron is due to a wrong reconstructed track curvature, the electron has similar properties to a prompt one, but a less accurate cluster energy over track momentum ratio (E_{cl}/p), which is used in the electron identification. If a mismatch between track and cluster occurs, the impact parameter and the ratio E_{cl}/p of the reconstructed electron could be different compared to prompt electrons. An electron produced in the interaction with the detector, i.e. in a trident event or in a photon conversion, has generally an impact parameter larger than a prompt lepton. Isolation and E_{cl}/p can also be biased by the additionally produced electrons.

If requirements on the E_{cl}/p, impact parameter and isolation are used to discriminate *tight* and *non-tight* electrons, there could be a difference in the selection efficiency for prompt, charge misID and photon conversion electrons. However, these properties are effective to discriminate prompt and non-prompt leptons: the leptons produced in the decays of an heavy flavour hadrons have a non-vanishing impact parameter; electron produced in jets are less isolated than prompt electrons; hadrons reconstructed as electrons would have $E_{cl}/p \neq 1$ due to the difference in the type of interactions with the calorimeters. The use of requirements on these variables, especially on the isolation, to define the *tight* electron selection are therefore important to satisfy the conditions of the Matrix Method ($r_\ell \gg f_\ell$). In this study, the impact parameter requirement is used as part of the *loose* electron selection, therefore it has no affect on the efficiencies for charge misID and photon

conversion electrons. However, both E_{cl}/p and isolation requirements are included in the *tight* electron definition. These requirements may reduce more the selection efficiency for charge misID and photon conversion electrons than for prompt electrons. The magnitude of this difference should be investigated. A solution to reduce potential biases due to the isolation requirements could be the use of a condition on the electromagnetic energy deposits surrounding the electron within the *loose* selection. Additional isolation conditions would then be used in the *tight* selection. The charge misID and photon conversion electron with additional electrons would be removed by the *loose* selection and the additional selection would have a small impact on their selection efficiency. On the other hand, a large fraction of non-prompt electrons would be able to pass the *loose* selection but not the *tight* selection. Since the collisional conditions in the $\sqrt{s} = 13$ TeV data involve a reassessment of the lepton selection definitions, the study of differences in selection efficiency for charge misID, photon conversion and prompt electrons and the optimization of the *loose* selection are postponed to the analysis of new data.

6.4.5.2 Preliminary information for non-prompt background studies

To use the Fake Factor Method and the Matrix Method [183], the leptons are required to pass a preliminary selection, the *loose* selection, and are then tested against the *tight* selection. The *tight* selection corresponds to the selection passed by the leptons in the region where the non-prompt background is estimated (usually the signal regions of the analysis). The lepton selections used for this study are given in Table 6.9. The *loose* selection does not require isolation and uses a looser identification for electrons and a relaxed d_0 condition for muons to ensure a high statistics of non-prompt leptons in the regions used for the non-prompt background estimate. The *tight* selection for this study is not identical to that of the published $t\bar{t}H$ analysis. As a consequence, the signal regions may also be somewhat different. The differences in the lepton selection are highlighted in red and are not significant for the aim of the study. Furthermore, the lepton selection will undergo a re-optimization in the analysis of $\sqrt{s} = 13$ TeV data.

The jet selection, the τ_{had} selection and the removal criteria for overlapping particles follow the indications of the main analysis (Table 6.4 and Table 6.5). The definition of the $2\ell 0\tau_{\text{had}}$ signal region is identical to the main analysis (Table 6.6).

The prompt and photon conversion backgrounds are estimated with simulations as in the main analysis, while the charge misID background is obtained from simulations with corrected charge misID rates following the method used in the new physics search (Section 5.4.2.1). The sources of charge misID background considered are $Z/\gamma^* + \text{jets}$, $t\bar{t}$, *single-top*, $tW^{\pm}$ and W^+W^-.

Selection	Muon	Electron
loose	$p_{\mathrm{T}} > 10$ GeV	$p_{\mathrm{T}} > 10$ GeV
	$\lvert\eta\rvert < 2.5$	$\lvert\eta_{cl}\rvert < 1.37$ or $1.52 < \lvert\eta_{cl}\rvert < 2.47$
	$\lvert z_0 \sin\theta\rvert < 1$ mm	$\lvert z_0 \sin\theta\rvert < 1$ mm
	$\lvert d_0\rvert/\sigma(d_0) < 10$	$\lvert d_0\rvert/\sigma(d_0) < 3$
	$Q(IDet) = Q(MS)$	*medium++* ID
		not flagged as from photon conversion
tight	$ptcone20/p_{\mathrm{T}} < 0.05$	$ptcone20/p_{\mathrm{T}} < 0.05$
	$etcone20/p_{\mathrm{T}} < 0.05$	$etcone20/p_{\mathrm{T}} < 0.05$
	$\lvert d_0\rvert/\sigma(d_0) < 3$	*tight++* ID

Table 6.9: Definitions of the *loose* and *tight* selections applied to muons and electrons used for the non-prompt background estimate. The *tight* requirements are applied in addition to the *loose* ones. The differences from the lepton selection of the published analysis are highlighted in red.

6.4.5.3 Non-prompt enriched region for fake factor and efficiency measurements

The fake factors, f_ℓ^{FF}, and the non-prompt efficiencies, f_ℓ, are estimated in a region dominated by non-prompt background and with large statistics. The same "non-prompt region" can used for both the Fake Factor Method and the Matrix Method. These methods assume that the selection efficiencies calculated for non-prompt leptons in the non-prompt region are equal to the selection efficiencies for non-prompt leptons in the signal regions. This is expected if the composition of the non-prompt leptons and the event activity in the non-prompt region are similar to those in the signal regions. The event activity can for instance affect the lepton isolation and as a consequence the selection efficiency.

The $t\bar{t}$ events are the main source of non-prompt background in the signal regions and produce non-prompt leptons predominantly via the decays of b-quark flavoured hadrons. Non-prompt muons are produced almost exclusively in this way, while non-prompt electrons can also come from light flavour hadrons. A region dominated by $t\bar{t}$ non-prompt background is therefore a natural candidate to estimate f_ℓ^{FF} and f_ℓ. Non-prompt regions dominated by other processes and other types of non-prompt leptons would introduce biases on the measurement due to the different non-prompt lepton composition and event activities. However, the resulting systematic uncertainties may be offset by reduced statistical errors or by smaller uncertainties on other systematic sources (like background contamination).

Three regions enriched in non-prompt leptons are investigated. The measurement of f_ℓ^{FF} for electrons is relatively more complex than for muons because the non-prompt electrons have multiple origins and the non-prompt regions can be significantly contaminated by charge misID and photon conversion background. Therefore this section focuses mainly on the study of non-prompt regions for electrons. A discussion about the non-prompt region used for muons is given at the end of this section.

The electron non-prompt regions studied are:

$W^{\pm}+\text{jets}$ regions: a non-prompt electron is reconstructed from a jet in event with a candidate $W^{\pm} \to \mu^{\pm}\nu$ decay. Events with opposite-sign and same-sign leptons are considered separately with dedicated selections to reduce the prompt background:

- one *tight* muon with $p_{\mathrm{T}} > 25$ GeV that has activated the SMTrig;
- $E_{\mathrm{T}}^{miss} > 25$ GeV and $M_{\mathrm{T}}\left(\mu, E_{\mathrm{T}}^{miss}\right) > 40$ GeV to select $W^{\pm} \to \mu^{\pm}\nu$ events;
- one *loose* electron;
- if opposite-sign leptons, then $N\left(\text{jet}\right) = 0$;
- if same-sign leptons, then $30 < m\left(e^{\pm}, \mu^{\pm}\right) < 150$ GeV.

The last two requirements are applied to suppress the dominant prompt background: $t\bar{t}$ production in events with opposite-sign leptons and diboson production in $e^{\pm}\mu^{\pm}$ events. Events having electrons with large impact parameters are also studied to increase the non-prompt background fraction. In that case, the electrons selected have $|d_0|/\sigma(d_0) > 3$ instead of $|d_0|/\sigma(d_0) < 3$ and pass the other *loose* requirements.

$Z+\text{jets}$ regions: a non-prompt electron is reconstructed from a jet in event with a candidate $Z \to \mu^{+}\mu^{-}$ decay. Two selections (M_{T} and $\Delta\phi$) are studied to reduce the prompt background in these regions:

- two opposite-sign *tight* muons, one of which has $p_{\mathrm{T}} > 25$ GeV and has activated the SMTrig;
- $|m\left(\mu^{+}, \mu^{-}\right) - m_Z| < 15$ GeV to select $Z \to \mu^{+}\mu^{-}$ events;
- one *loose* electron;
- for the M_{T} selection:
  - $M_{\mathrm{T}}\left(e, E_{\mathrm{T}}^{miss}\right) < 40$ GeV;
  - $E_{\mathrm{T}}^{miss} + M_{\mathrm{T}}\left(e, E_{\mathrm{T}}^{miss}\right) < 60$ GeV.
- for the $\Delta\phi$ selection:
  - $\Delta\phi\left(e, E_{\mathrm{T}}^{miss}\right) < 0.5$;
  - $E_{\mathrm{T}}^{miss} < 30$ GeV.

The last two requirements are applied to suppress in a complementary way the $W^{\pm}Z$ prompt background.

$t\bar{t}$ region: a non-prompt electron is reconstructed from a jet in event with an additional lepton and b-jets. The selection targets $t\bar{t}$ events where one $W^{\pm}$ decays in hadrons and the other in leptons. The selection requires:

- one *tight* lepton with $p_{\mathrm{T}} > 25$ GeV that has activated the SLTrig;

- if the *tight* lepton is an electron, then $|\eta| < 1.37$ to reduce charge misID contaminations;
- $N\,(b\text{-jet}) \geq 1$;
- $0 < N\,(\text{jet}) < 4$ to avoid overlaps with the signal regions;
- one *loose* electron;
- same-sign leptons to avoid the large prompt opposite-sign background;
- if two electrons then $|m\,(e^{\pm}, e^{\pm}) - m_Z| > 10$ GeV to avoid $Z \to e^+e^-$ events with charge misID electrons.

One of the challenges with the non-prompt region definition is the suppression of background from either prompt, photon conversion or charge misID sources. These backgrounds can be significant especially in events where the *loose* electron satisfies also the *tight* selection. Before the fake factors or efficiencies are calculated from the non-prompt region, the expected background of events with prompt, photon conversion and charge misID electrons must be subtracted from data. The number of observed events in the considered regions and the expected background fractions are given in Table 6.10. The largest background contamination and the lowest statistics is found in events with *tight* electrons, as expected.

Region	*tight* electron		*non-tight* electron	
	Data	Background fraction (%)	Data	Background fraction (%)
$W^{\pm}$ + jets (opposite-sign lep.)	7818	86%	9327	27%
$W^{\pm}$ + jets (opposite-sign lep., $\lvert d_0\rvert/\sigma(d_0) > 3$)	1022	72%	2710	22%
$W^{\pm}$ + jets (same-sign lep.)	1303	65%	6997	15%
$W^{\pm}$ + jets (same-sign lep., $\lvert d_0\rvert/\sigma(d_0) > 3$)	421	42%	3725	11%
Z + jets ($\Delta\phi$ sel.)	128	46%	517	25%
Z + jets (M_{T} sel.)	338	52%	1468	27%
$t\bar{t}$ (same-sign lep.)	355	47%	2435	11%

Table 6.10: Observed event yields in the electron non-prompt regions divided for *tight* and *non-tight* electrons, together with the expected background contamination from prompt, charge misID and photon conversion processes.

The largest background is found in the $W^{\pm}$ + jets regions with opposite-sign leptons. The distributions of the electron p_{T} and $|\eta|$ in these regions are shown in Appendix C (Figure C.1 and Figure C.2). The $N\,(\text{jet}) = 0$ requirement suppresses $t\bar{t}$ events with prompt leptons, but the W^+W^- and $Z \to \tau^+\tau^-$ processes are still dominant ($> 70\%$ of the yields) and make these regions not suitable for f_ℓ^{FF} (and f_ℓ) measurements.

The $W^{\pm}$ + jets regions with same-sign leptons have less statistics but a larger fraction of non-prompt events. This is due to the smaller number of prompt events with same-sign leptons opposed to the similar number of non-prompt events with same-sign and with

opposite-sign leptons. These regions can contain a significant charge misID background though.

The distributions of the p_{T} and $|\eta|$ for *tight* electrons and for *non-tight* electrons in the $W^{\pm}+$jets region with same-sign leptons are shown in Figure 6.7[5]. The *non-tight* electron region has less than 15% of background, while in events with *tight* electrons the expected backgrounds accounts for the 65% of data. The background contribution is larger at high p_{T} and uniform in $|\eta|$. The majority of the background is due to $W^{\pm}Z$ production.

Events where the *loose* electrons have $|d_0|/\sigma(d_0) > 3$ instead of $|d_0|/\sigma(d_0) < 3$ have also been considered. The inversion of the impact parameter requirement reduces the prompt background fraction, but increases the fraction of photon conversion background and charge misID background from trident events. The statistics is smaller than the other $W^{\pm}+$ jets regions, but the overall background fraction is smaller too. The main background is due to $t\bar{t}$ events with charge misID electrons and $W^{\pm}\gamma$. The distributions of the p_{T} and $|\eta|$ of *tight* electrons and *non-tight* electrons are shown in Figure 6.8. As observed previously, charge misID electrons and electrons from photon conversion have higher production rate at larger $|\eta|$. In the pseudorapidity range to $|\eta| < 1.37$, which is the requirement imposed in the $2\ell 0\tau_{\mathrm{had}}$ category, the background contamination for *tight* electrons reduces to 25%.

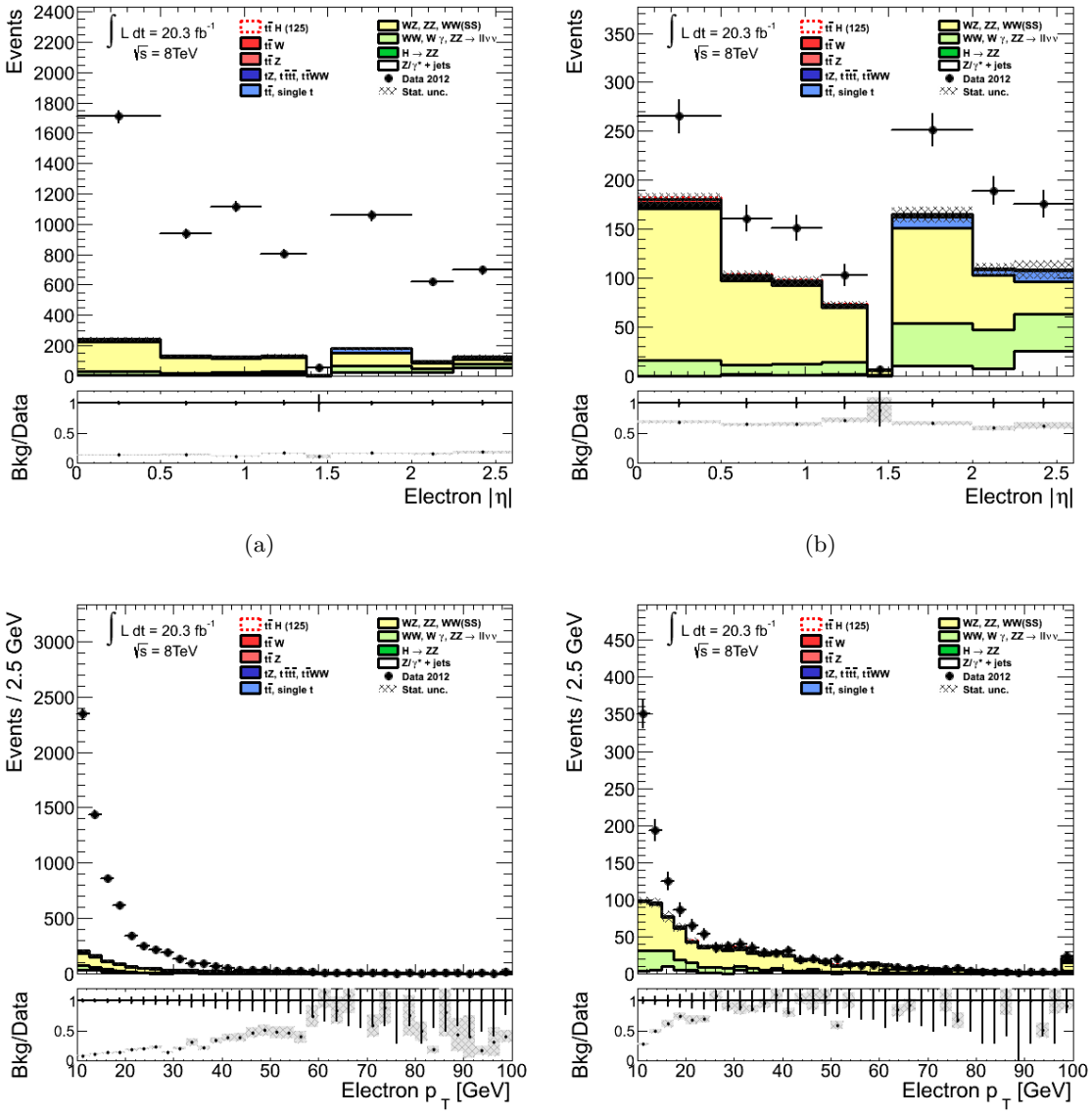
The $Z+$ jets non-prompt region selected with the $\Delta\phi$ requirement has very little statistics and therefore is not suitable for measuring fake factors as a function of the lepton kinematics. The distributions of the p_{T} and $|\eta|$ for *tight* electrons and *non-tight* electrons in this region are shown in Appendix C (Figure C.3), but have not been used to calculate fake factors.

The $Z+$jets region that uses M_{T} requirements has larger statistics. The p_{T} and $|\eta|$ distributions for *tight* and *non-tight* electrons in this region are shown in Figure 6.9. The main background in this region is the $W^{\pm}Z$ production. The total background in events with *tight* electrons is about 50%.

The $t\bar{t}$ non-prompt region is one of the purest in non-prompt electrons. The electron p_{T} and $|\eta|$ distributions in this region are shown in Figure 6.10. Events with either $e^{\pm}e^{\pm}$ or $e^{\pm}\mu^{\pm}$ leptons are considered. In the first case, if both electrons pass the *tight* selection, have $|\eta| < 1.37$ and have activated the trigger, the electron with lower p_{T} is used in the determination of the fake factors. The $t\bar{t}$ simulation shows indeed that the lowest p_{T} lepton in the pair is more likely to be the non-prompt one. The region with *tight* electrons is affected, especially for large $|\eta|$, by a significant contribution of charge misID background from $t\bar{t}$ and $Z+$jets productions. The purity is still one of the best among the regions investigated though.

The fake factors have been measured in both $W^{\pm}+$jets regions with same-sign leptons,

[5]Note that the charge misID and photon conversion backgrounds in the non-prompt regions with same-sign leptons are not shown as a whole but rather as individual Standard Model processes ($t\bar{t}$, *single-top*, Z/γ^*, W^+W^- and $W^{\pm}\gamma$).



(a) (b)

(c) (d)

Figure 6.7: The $|\eta|$ (top) and p_{T} (bottom) distributions of electrons passing (right) and failing (left) the *tight* selection in the $W^{\pm}$ + jets same-sign lepton non-prompt region. Events in data (black dots) are compared to the expected stacked background contributions, non-prompt background excluded. Overflow events are added to the last bin. The lower panel shows the fraction of expected background yields. The error bars on the data points and the dashed band on the simulated background show the statistical uncertainties.

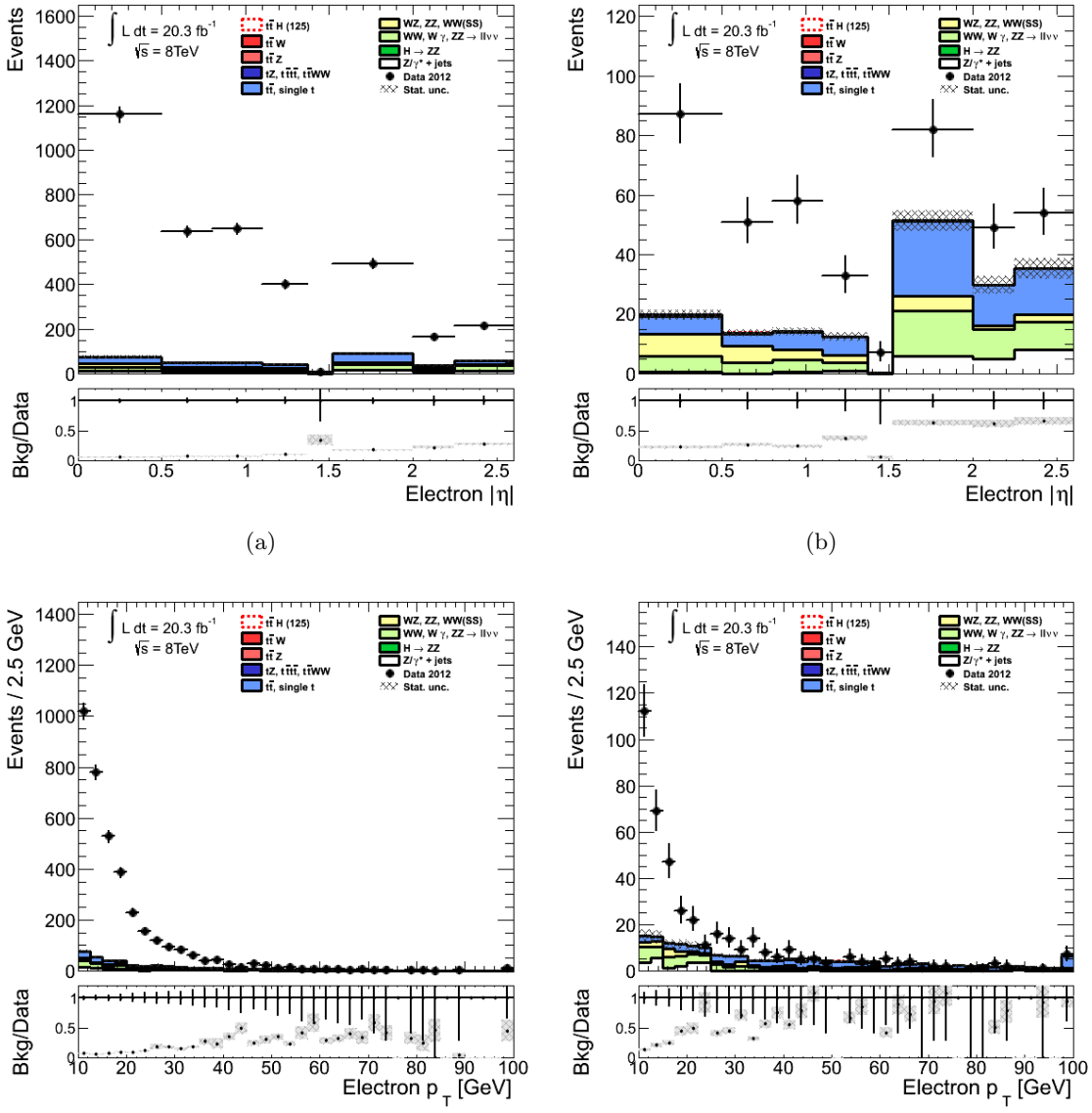
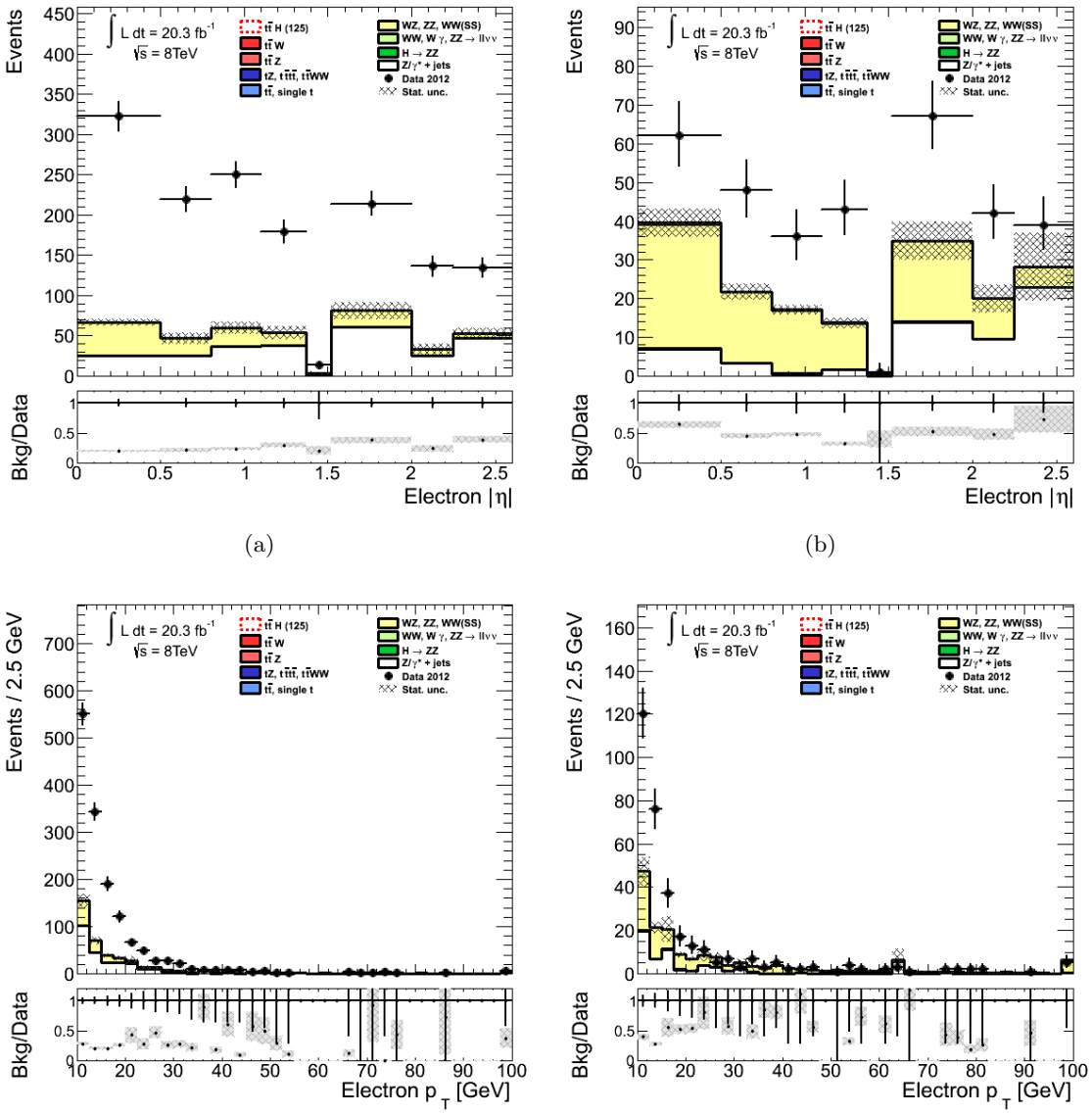


Figure 6.8: The $|\eta|$ (top) and p_{T} (bottom) distributions of electrons with $|d_0|/\sigma(d_0) > 3$ passing (right) and failing (left) the *tight* selection in the $W^{\pm}$+jets same-sign lepton non-prompt region. Events in data (black dots) are compared to the expected stacked background contributions, non-prompt background excluded. Overflow events are added to the last bin. The lower panel shows the fraction of expected background yields. The error bars on the data points and the dashed band on the simulated background show the statistical uncertainties.



(a) (b)

(c) (d)

Figure 6.9: The $|\eta|$ (top) and p_{T} (bottom) distributions of electrons passing (right) and failing (left) the *tight* selection in the $Z+$jets M_{T} selection non-prompt region. Events in data (black dots) are compared to the expected stacked contributions from prompt Standard Model processes. Overflow events are added to the last bin. The lower panel shows the fraction of expected prompt background yields. The error bars on the data points and the dashed band on the simulated background show the statistical uncertainties.

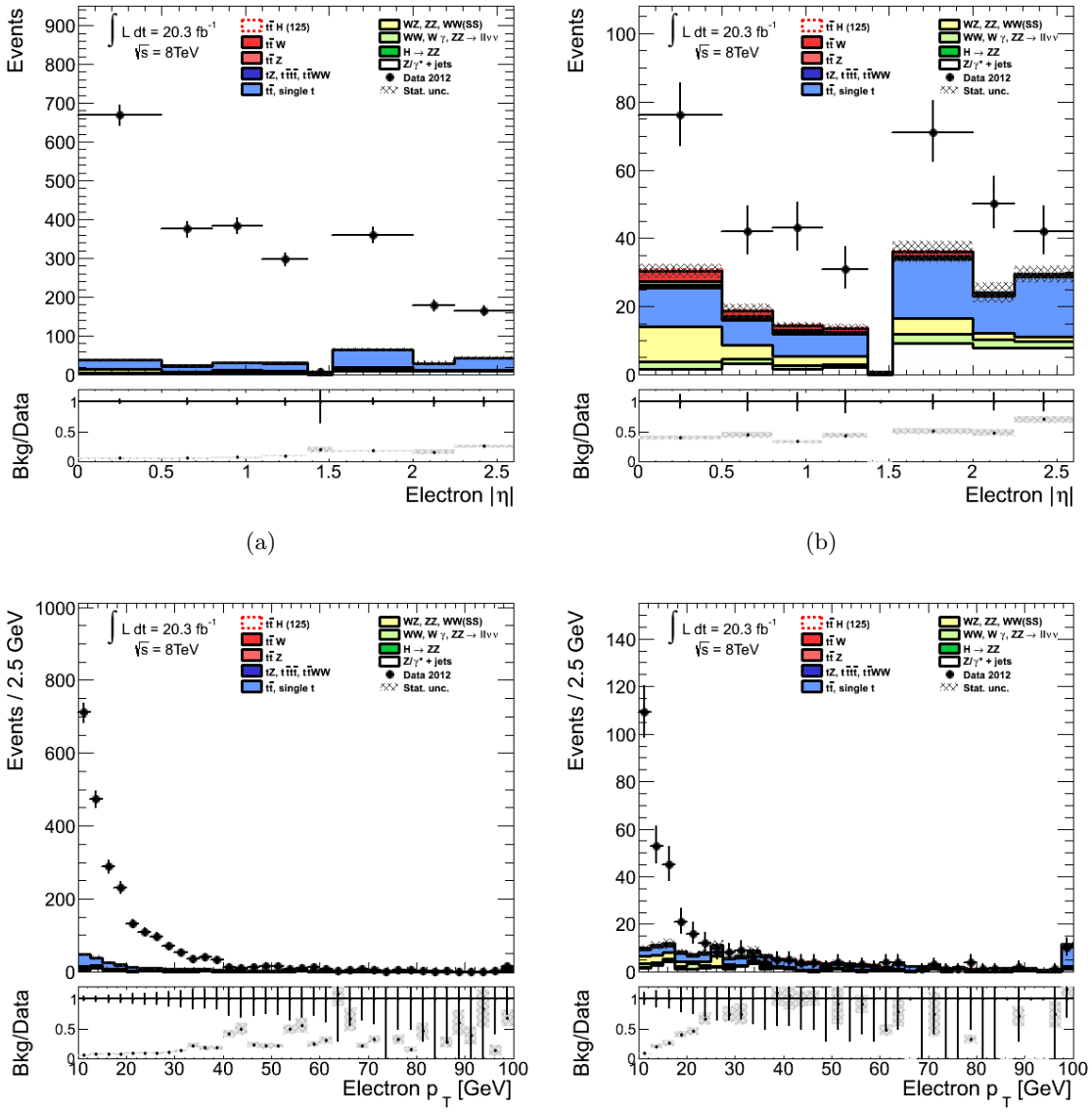
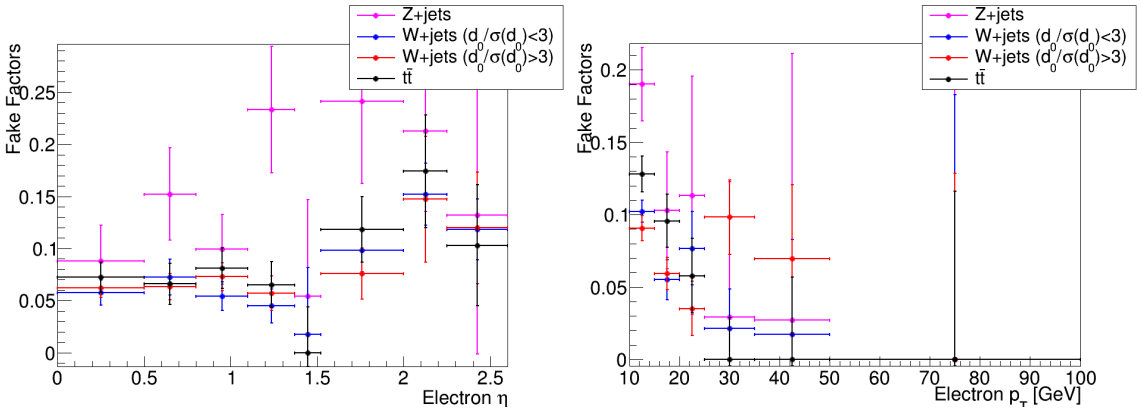


Figure 6.10: The $|\eta|$ (top) and p_{T} (bottom) distributions of electrons passing (right) and failing (left) the *tight* selection in the $t\bar{t}$ non-prompt region. Events in data (black dots) are compared to the expected stacked background contributions, non-prompt background excluded. Overflow events are added to the last bin. The lower panel shows the fraction of expected background yields. The error bars on the data points and the dashed band on the simulated background show the statistical uncertainties.



(a) (b)

Figure 6.11: Electron fake factors as a function of $|\eta|$ (left) and p_{T} (right) obtained from the Z+jets M_{T} selection region (magenta), from the $W^{\pm}$ + jets same-sign lepton with electron $|d_0|/\sigma(d_0) < 3$ (blue) and electron $|d_0|/\sigma(d_0) > 3$ (red) regions and from the $t\bar{t}$ region (black). The error bars account for the statistical uncertainties on data and the subtracted background.

in the Z+jets region with M_{T} requirements and in the $t\bar{t}$ region. The values obtained from each region as a function of the electron p_{T} and $|\eta|$ are shown in Figure 6.11. The error on each fake factor accounts for the statistical uncertainty on data and on the subtracted background. All measurements are statistically compatible. Apart the fake factors from the Z+jets region, that have the largest uncertainties, the fake factors have overall similar values and similar statistical uncertainties.

The fake factors are larger at low p_{T} and for $|\eta| > 1.52$. The p_{T} dependences of the fake factors appear slightly different for the $t\bar{t}$ and the $W^{\pm}$+jets regions. The fake factors calculated in the $t\bar{t}$ region seem to be higher at low p_{T} and seem to decrease faster with p_{T}. The origin of this difference could be a bias in the background subtraction or a dependence on the targeted process and in particular on the composition of the non-prompt electron sample collected. These regions are indeed populated by different background processes and the selected non-prompt electrons are produced more often from heavy flavour hadron decays in the $t\bar{t}$ region than in the $W^{\pm}$ + jets regions. The effect of these sources of systematic uncertainties on the fake factors should be further investigated. Given the change in experimental conditions in the 13 TeV data, the study of the systematic effects affecting the fake factors calculation is postponed to the new data.

The fake factors used in the remaining of this study are those calculated from the $t\bar{t}$ non-prompt region because the non-prompt electrons in this region are produced mainly by $t\bar{t}$ production as expected to occur in the signal regions. Systematic uncertainties related to the composition of non-prompt electron sample are probably smaller than in any other non-prompt region, as well as the background subtraction uncertainties because of the smallest background contamination[6]. In future analyses, a proper evaluation of the

[6]This last point should be further investigated because the charge misID background, dominant in

non-prompt compositions and of the systematic uncertainties could allow a combination of the fake factors from different regions. This could improve the precision of the estimate if the statistical error is still the dominant uncertainty on the fake factors.

The $t\bar{t}$ non-prompt region is additionally used to calculate the efficiencies f_ℓ for the Matrix Method. The f_ℓ are indeed obtained from the f_ℓ^{FF} via: $f_\ell^{FF} = \frac{f_\ell}{1-f_\ell}$.

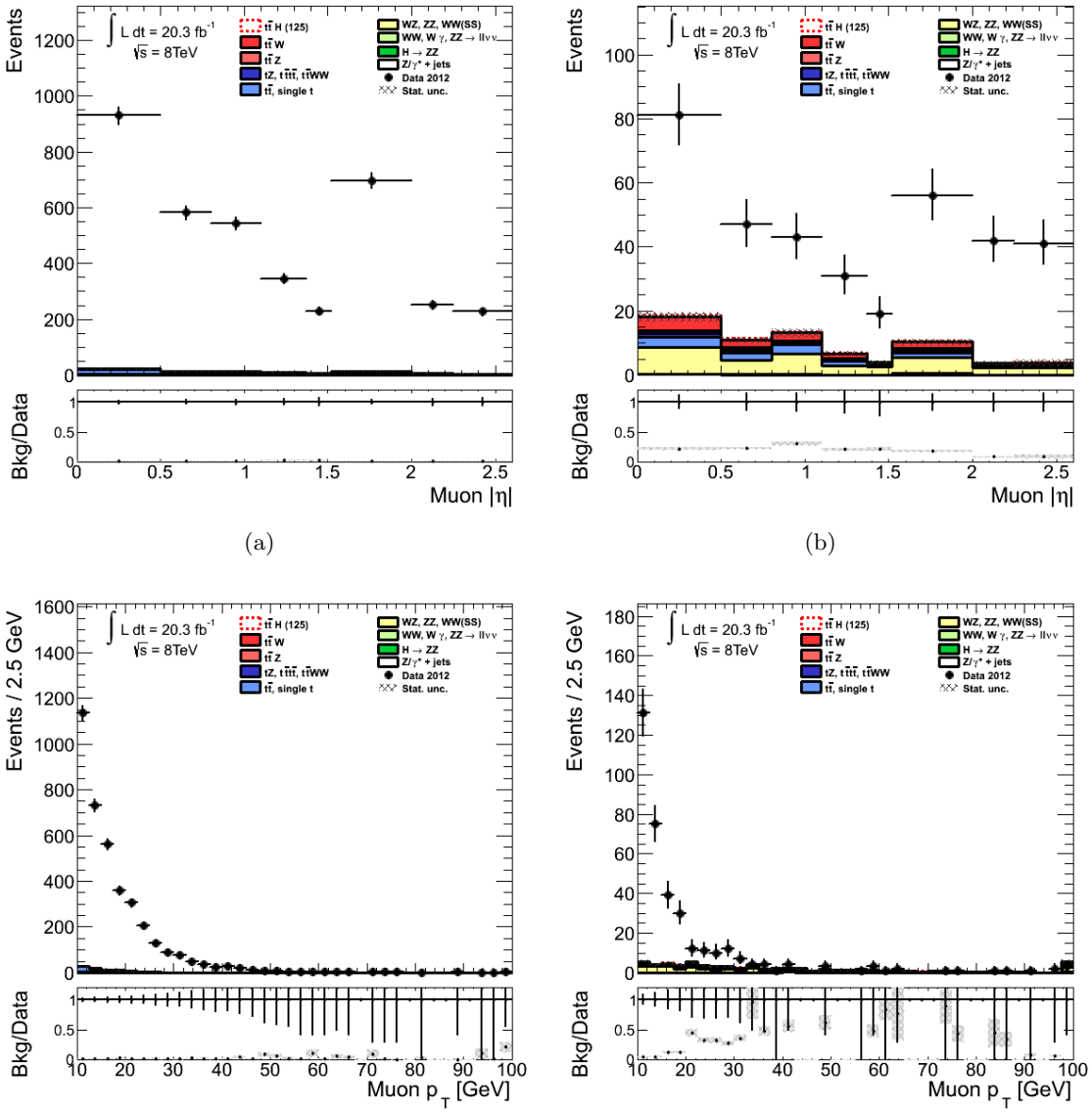
A $t\bar{t}$ enriched region with non-prompt muons is used to measure muon fake factors and efficiencies. The definition of the muon $t\bar{t}$ non-prompt region is:

- one *tight* lepton with $p_{\mathrm{T}} > 25$ GeV that has activated the SLTrig;
- if the *tight* lepton is an electron, then $|\eta| < 1.37$ to reduce charge misID contaminations;
- $N\,(b\text{-jet}) \geq 1$;
- $0 < N\,(\text{jet}) < 4$;
- one *loose* muon;
- same-sign leptons to avoid prompt opposite-sign lepton backgrounds;

This region differs from the electron one for the selection of *loose* muons instead of *loose* electrons. When two muons pass the *tight* selection, have $p_{\mathrm{T}} > 25$ GeV and activate the SMTrig, there is an ambiguity in the choice of the non-prompt muon in the pair. The muon with lower p_{T} is considered non-prompt, as supported by simulation studies, and is used in the determination of the fake factors. The distributions of the non-prompt muon p_{T} and $|\eta|$ in the $t\bar{t}$ non-prompt region are shown in Figure 6.12. This region is considerably purer than the non-prompt electron one because there is very little charge misID background contamination. The total expected background is less than 20% and is due mainly to prompt same-sign lepton processes. Given the low background contamination and the similarity in non-prompt muon composition with the signal regions, the $t\bar{t}$ non-prompt region is that used to estimate muon fake factors and efficiencies. A combination of several non-prompt regions may be considered in future to improve the statistics.

The fake factors calculated from the $t\bar{t}$ non-prompt regions are shown for electrons and muons as a function of p_{T} and $|\eta|$ in Figure 6.13. Despite the different selections, the fake factors for muons and electrons have similar magnitudes and p_{T} dependence. The non-prompt leptons are reconstructed about ten times more frequently as *non-tight* leptons than as *tight* ones. The fake factors seem to decrease with p_{T}. The electron fake factors have a larger error mainly due to the larger background contamination in the electron non-prompt region. Non-prompt efficiencies, f_ℓ, are obtained from f_ℓ^{FF} via: $f_\ell = \frac{f_\ell^{FF}}{1+f_\ell^{FF}}$.

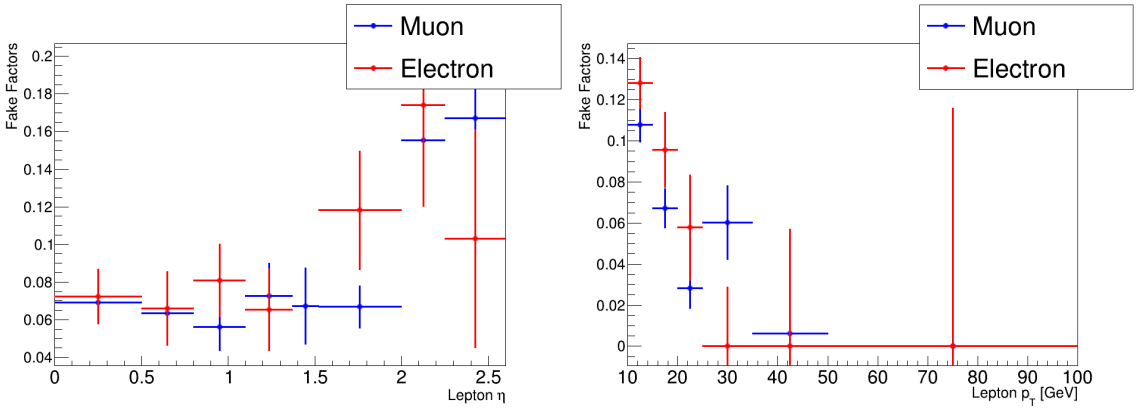
the $t\bar{t}$ region, may have a larger uncertainty than the prompt background, dominant in the $W^\pm$ + jets regions. Therefore the uncertainty on the background subtraction could also even result bigger in the $t\bar{t}$ non-prompt region.



(a) (b)

(c) (d)

Figure 6.12: The $|\eta|$ (top) and p_{T} (bottom) distributions of muons passing (right) and failing (left) the *tight* selection in the $t\bar{t}$ non-prompt region. Events in data (black dots) are compared to the expected stacked background contributions, non-prompt background excluded. Overflow events are added to the last bin. The lower panel shows the fraction of expected background yields. The error bars on the data points and the dashed band on the simulated background show the statistical uncertainties.



(a) (b)

Figure 6.13: Fake factors as a function of the lepton $|\eta|$ (left) and p_{T} (right) for muons (blue) and electrons (red) as obtained from the $t\bar{t}$ non-prompt regions. The error bars account for the statistical uncertainties on data and the subtracted background.

For this study, only statistical uncertainties on data and predictions have been considered. However, a preliminary study of the biases due to the use of SLTrig on the fake factors has been performed. This is reported in Section C.2.3 and shows a potential bias in the muon fake factors. A full evaluation of systematic uncertainties should be performed in future analyses, similarly to what has been done for the fake factors estimated in Section 5.4.4.2 and Section 5.4.4.3.

6.4.5.4 Prompt background enriched regions for the evaluation of r_ℓ

The Matrix Method requires the measurement of *tight* selection efficiencies for prompt leptons, r_ℓ. The measurement is performed in a region highly pure in prompt leptons. Opposite-sign di-lepton events are produced in large quantity by processes with prompt leptons. Therefore the r_ℓ are estimated in a region with opposite-sign leptons and additional jets and b-jets to be close to the signal region selection, the prompt region. The region is defined by:

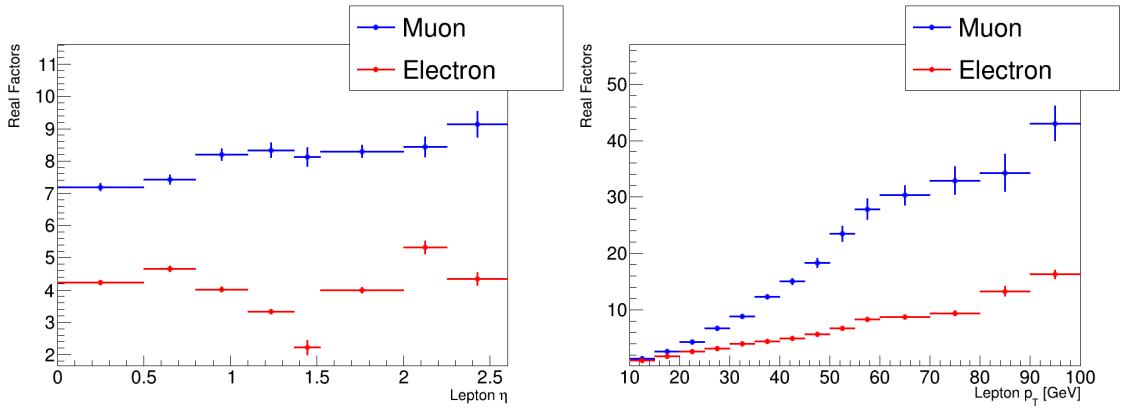
- one *tight* lepton with $p_{\mathrm{T}} > 25$ GeV that has activated the SLTrig;
- $N\,(b\text{-jet}) \geq 1$;
- $0 < N\,(\text{jet}) < 4$;
- one *loose* lepton;
- opposite-sign leptons.

If the leptons have same flavour and both are *tight*, have $p_{\mathrm{T}} > 25$ GeV and are compatible with the SLTrig, the lepton used to calculate the efficiencies is chosen randomly since both are expected to be prompt leptons. The selection is identical to that defining the

$t\bar{t}$ non-prompt regions except for the requirement of opposite-sign leptons. The electron efficiencies are obtained from events with e^+e^- and $e^\pm\mu^\mp$ pairs, while the muon efficiencies from events with $\mu^+\mu^-$ and $e^\pm\mu^\mp$ pairs. The b-jet and jet requirements are introduced to select events with an activity similar to the signal regions. The requirement $0 < N\,(\text{jet}) < 4$ is set to match the condition on the hadronic activity of the non-prompt region. Since the isolation of a prompt lepton and so the prompt efficiencies can depend on the additional hadronic activity in the event, the selection should resemble the signal region one. Uncertainties on prompt efficiencies due to the prompt region selection should be evaluated in future analyses. Currently, the statistics in the $N\,(\text{jet}) \geq 4$ region is too small to measure r_ℓ, but when more data will be collected, the use of this requirement may be convenient in terms of overall statistical and systematic uncertainties.

The main process in the prompt region is the $t\bar{t}$ production with both $W^\pm$ decaying to leptons. In events with same flavour leptons, the $Z/\gamma^* + \text{jets}$ production is also important. Events with *non-tight* leptons present a small contamination of non-prompt background. The impact of the background contamination on the prompt efficiencies and the Matrix Method predictions is studied in Section C.2.2. The bias on the non-prompt predictions due to this contamination is smaller than 4% in the $e^\pm e^\pm$ category, 7% in the $e^\pm\mu^\pm$ category and 10% in the $\mu^\pm\mu^\pm$ category. This effect is small compared to the other uncertainties of the method in this study, but it might be significant in future analyses.

For convenience and consistence, real factors (r_ℓ^{FF}) are defined as the ratio of the number of prompt leptons passing the *tight* selection to the number of prompt leptons failing it. The real factors calculated from the prompt region are shown for electrons and muons as a function of p_T and $|\eta|$ in Figure 6.13. The prompt efficiencies r_ℓ are obtained from r_ℓ^{FF} via: $r_\ell = \frac{r_\ell^{FF}}{1+r_\ell^{FF}}$. On average prompt electrons have four times more chances to



(a) (b)

Figure 6.14: Real factors as a function of the lepton $|\eta|$ (left) and p_T (right) for muons (blue) and electrons (red) as obtained from the prompt regions. The error bars account for the statistical uncertainty.

be reconstructed as *tight* electron than *non-tight*. Eight out of nine times *loose* prompt

muons are reconstructed as *tight* muons. Real factors increase with p_{T} for both leptons, since the isolation requirements allow for more activity around the higher p_{T} leptons.

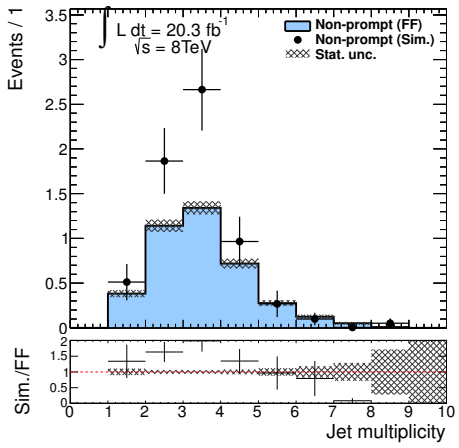
For this study no systematic uncertainties are considered for r_{ℓ}^{FF}, however a preliminary study of the biases due to the use of SLTrig on the real factors has been performed. This is reported in Section C.2.3 and shows biases in the measurement of electron real factors. A full evaluation of systematic uncertainties should be performed in future analyses similarly to what has been done for fake factors in Section 5.4.4.2 and Section 5.4.4.3.

6.4.5.5 Validation studies with $t\bar{t}$ simulations

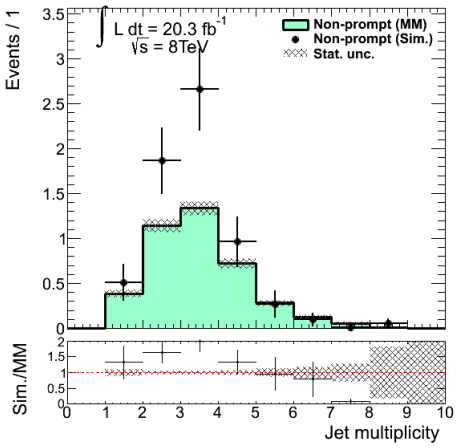
The predictive power of the Matrix Method and the Fake Factor Method is tested on simulated $t\bar{t}$ events. In this case, the inputs needed by the methods are obtained exclusively from simulation: the $t\bar{t}$ sample is used to calculate the non-prompt and prompt efficiencies (and the fake factors); then $t\bar{t}$ events are scaled according to the Fake Factor Method and Matrix Method procedures. In the $t\bar{t}$ non-prompt regions, defined in Section 6.4.5.3, the simulated events with charge misID electrons are excluded, so that only non-prompt background events are used to calculate non-prompt efficiencies and fake factors. In the prompt regions instead, the events with non-prompt leptons are not considered in the determination of the prompt efficiencies. The $t\bar{t}$ events that pass the $2\ell 0\tau_{\mathrm{had}}$ category selection and have two *loose* same-sign leptons of which at least one is non-prompt, are then scaled with the factors obtained from efficiencies following the prescriptions of the methods.

The accuracy of the two methods is tested comparing their non-prompt background predictions with the estimate obtained directly from simulation. If the base assumptions of the Matrix Method and the Fake Factor Method are valid, their predictions should be compatible with those from pure simulation. Differences may arise especially in the modeling of the event kinematic, if the parametrisation of fake factors and efficiencies is not sufficiently refined. For instance, these quantities may depend on other properties beside the lepton p_{T} and $|\eta|$ or may significantly change in value within the intervals which the p_{T} and $|\eta|$ ranges are divided into. The complexity of the parametrisation is limited by the statistics available. If significant differences are observed, these should be considered as systematic uncertainties associated to the methods.

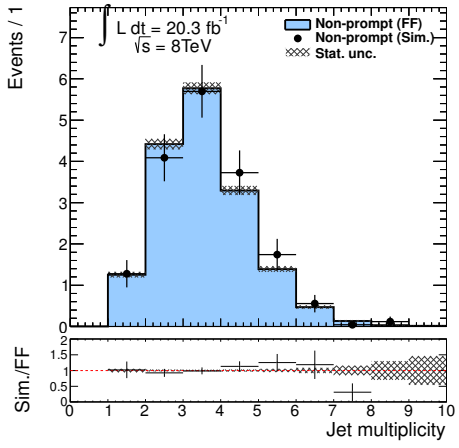
The distributions of the jet multiplicity in the $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ categories estimated directly from simulation, from the Matrix Method and from the Fake Factor Method applied to simulated events are shown in Figure 6.15. The predictions of the Fake Factor Method and the associated uncertainties are within 1% of those from the Matrix Method. The reason is that in this test the prompt efficiencies are very close to the unity and have uncertainties negligible compared to those of non-prompt efficiencies. Thus, the scaling applied by the Matrix Method is very similar to the one of the Fake Factor Method. Furthermore, this Fake Factor Method estimate does not require background



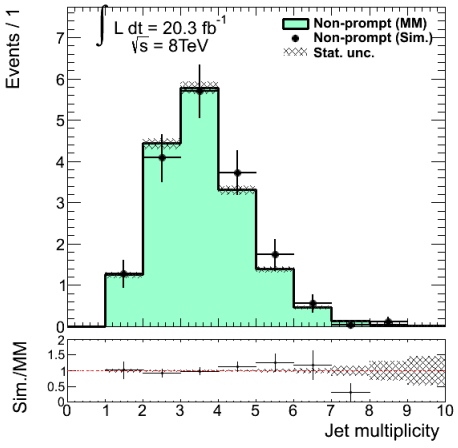
(a) $e^{\pm}e^{\pm}$



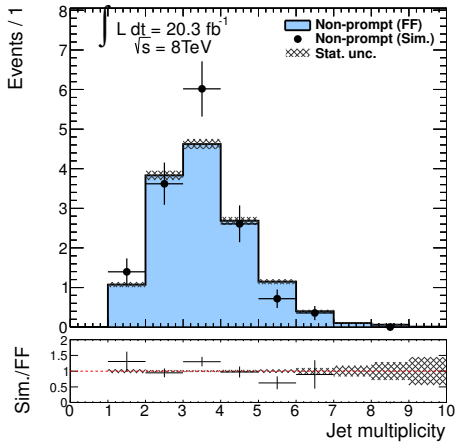
(b) $e^{\pm}e^{\pm}$



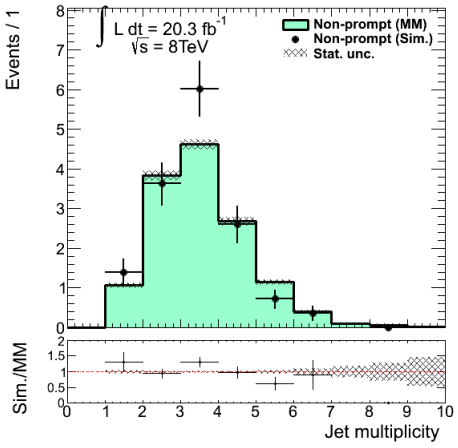
(c) $e^{\pm}\mu^{\pm}$



(d) $e^{\pm}\mu^{\pm}$



(e) $\mu^{\pm}\mu^{\pm}$



(f) $\mu^{\pm}\mu^{\pm}$

Figure 6.15: Distributions of the jet multiplicity for $e^{\pm}e^{\pm}$ (top), $e^{\pm}\mu^{\pm}$ (centre) and $\mu^{\pm}\mu^{\pm}$ (bottom) events in the $t\bar{t}$ simulation. Simulated $t\bar{t}$ non-prompt events (black dots) are compared to the predictions from the Fake Factor Method applied to the $t\bar{t}$ simulated events (light blue) and from the Matrix Method applied to the same events (light green). Overflow events are added to the last bin. The lower panel shows the ratio between the predictions from simulation and from the data-driven methods. The error bars and the dashed band show the statistical uncertainties on simulation and on the data-driven predictions, respectively. Uncertainties on the fake factors and the rates are not included.

subtraction because all the events considered contain non-prompt leptons. The predictions of the two methods are therefore equivalent as shown in more details in Section C.2.4. In the rest of this section only the Matrix Method predictions are discussed, but the same statements apply to the Fake Factor Method predictions.

In Figure 6.15, the requirement of $N\,(\text{jet}) \geq 4$, that defines the $2\ell 0\tau_{\text{had}}$ category, has been dropped to show the agreement in both the signal regions and the low $N\,(\text{jet})$ regions ($N\,(\text{jet}) \leq 3$). The yields predicted from simulation and by the Matrix Method in the signal regions and the low $N\,(\text{jet})$ regions are shown in Table 6.11. The uncertainty on the simulated yields is the statistical error on the simulated events that have two *tight* leptons. The statistical error for the Matrix Method is smaller due to the additional statistics with *non-tight* leptons used in the prediction. For the Matrix Method, the error due to the statistical uncertainty on non-prompt efficiencies (f_ℓ) is also given. In the low $N\,(\text{jet})$ regions, the statistical error is partly included in the f_ℓ uncertainty because the scaled events are also those used to calculate the non-prompt efficiencies. In this case, the predictions can be compared without considering the statistical errors. The uncertainty due to the statistics in the prompt regions, used to measure the prompt efficiencies, is less than 0.1 and therefore is omitted.

The uncertainty on the f_ℓ measurement has the largest impact on the Matrix Method predictions. This uncertainty is obtained by re-evaluating the predictions with all the efficiencies scaled either up or down of their uncertainty simultaneously. The uncertainties on the efficiencies in different $|\eta|$ and p_{T} ranges are therefore assumed correlated, even if the measurements in different ranges are independent from each other. The statistical uncertainty quoted for f_ℓ is therefore conservative. A more accurate estimate of the uncertainty could be obtained by generating pseudo-experiments, where each efficiency is varied independently within its error. The standard deviation of the predictions from the pseudo-experiments would give a better estimate of the uncertainty.

The low $N\,(\text{jet})$ regions are also called the closure regions because contain mostly the same events used to compute the non-prompt efficiencies. If the procedures of estimate are valid, they should provide predictions compatible with the pure simulation in this region.

In the signal regions, the jet multiplicity requirement is higher than in the regions where efficiencies are calculated. Indeed, efficiencies are deliberately calculated in a sample independent and with low signal contamination. In order for Matrix Method to work in the signal regions, the measured efficiencies should not depend significantly on the jet multiplicity. The comparison of the Matrix Method predictions and the pure simulation in the signal regions tests this assumption.

Table 6.11 shows that the Matrix Method and the pure simulation produce compatible estimates. Indeed most of them agree within the statistical error on simulations. The largest difference is found in the $e^{\pm}e^{\pm}$ closure region, where the Matrix Method estimate is 40% lower than the simulation. The power of the Matrix Method to model the non-prompt background kinematic is also tested. This method indeed should be able to

Region	Simulation	Matrix Method
$e^{\pm}e^{\pm}$ (N (jet) ≤ 3)	5.0 ± 0.6(stat.)	2.9 ± 0.1(stat.)$^{+6.0}_{-1.5}(f_\ell)$
$e^{\pm}e^{\pm}$ (N (jet) ≥ 4)	1.4 ± 0.3(stat.)	1.2 ± 0.1(stat.)$^{+2.8}_{-0.6}(f_\ell)$
$e^{\pm}\mu^{\pm}$ (N (jet) ≤ 3)	11.1 ± 0.9(stat.)	11.5 ± 0.2(stat.)$^{+12.7}_{-4.9}(f_\ell)$
$e^{\pm}\mu^{\pm}$ (N (jet) ≥ 4)	6.2 ± 0.7(stat.)	5.4 ± 0.1(stat.)$^{+6.6}_{-2.3}(f_\ell)$
$\mu^{\pm}\mu^{\pm}$ (N (jet) ≤ 3)	11.0 ± 0.9(stat.)	9.5 ± 0.2(stat.)$^{+5.9}_{-3.7}(f_\ell)$
$\mu^{\pm}\mu^{\pm}$ (N (jet) ≥ 4)	3.7 ± 0.6(stat.)	4.4 ± 0.1(stat.)$^{+3.5}_{-1.7}(f_\ell)$

Table 6.11: The predicted $t\bar{t}$ non-prompt background yields from simulation and from the Matrix Method in the $2\ell 0\tau_{\text{had}}$ signal regions and the closure regions. For Matrix Method in addition to the statistical error, the statistical uncertainty on f_ℓ is accounted. The uncertainty on r_ℓ is negligible for each category.

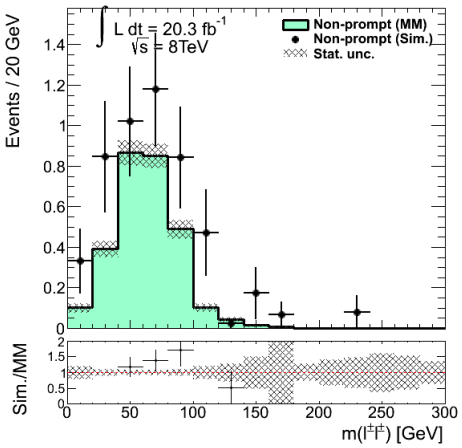
reproduce the properties of the non-prompt background if the measured efficiencies are properly parametrised. The invariant mass distributions in signal and closure regions as obtained from simulation and from the Matrix Method are shown in Figure 6.16. In the regions with larger statistics ($e^{\pm}\mu^{\pm}$ with N (jet) ≤ 3, $e^{\pm}\mu^{\pm}$ with N (jet) ≥ 4 and $\mu^{\pm}\mu^{\pm}$ with N (jet) ≤ 3), the Matrix Method is able to reproduce accurately the shape of the simulation distribution. In the other regions, the simulation distributions are subject to large fluctuations, but there is still agreement within uncertainties. The Fake Factor Method predicts shapes and associated uncertainties very similar to the Matrix Method (Figure C.8 and Figure C.9). The methods are equivalently valid to predict the non-prompt background properties.

The tests performed on the Matrix Method and Fake Factor Method with $t\bar{t}$ simulated events show that the methods deliver solid non-prompt background estimates. The yields predicted by the methods in the closure regions are compatible with those from simulation validating the procedures and the parametrisations of the efficiencies. The good agreement in the signal regions shows that the efficiencies calculated from the N (jet) ≤ 3 region are applicable to N (jet) ≥ 4 events without introducing additional uncertainties[7]. The methods are therefore able to predict the yields and the kinematic properties of the non-prompt background events accepted in the $2\ell 0\tau_{\text{had}}$ signal regions. The limit with both methods seems to be the statistics available in the non-prompt region used to calculate the fake factors and non-prompt efficiencies.

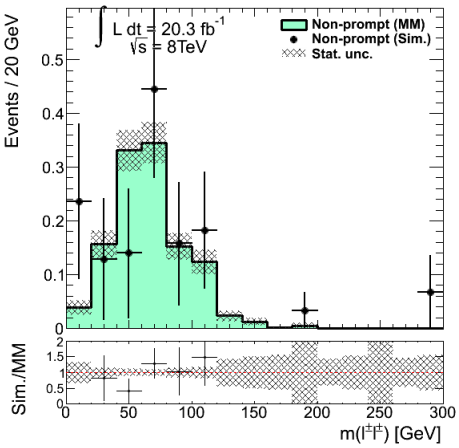
6.4.5.6 Validation with data

After having tested the internal consistency of the method with simulation, the predictions of non-prompt backgrounds in the $2\ell 0\tau_{\text{had}}$ signal regions and in the closure regions as obtained with the Matrix Method and Fake Factor Method are validated in data. In this case, the methods are completely based on data: the factors and efficiencies are mea-

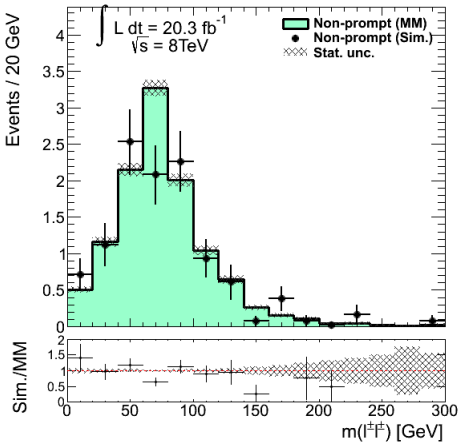
[7] At least in this condition where the predictions are dominated by the non-prompt efficiency statistical error.



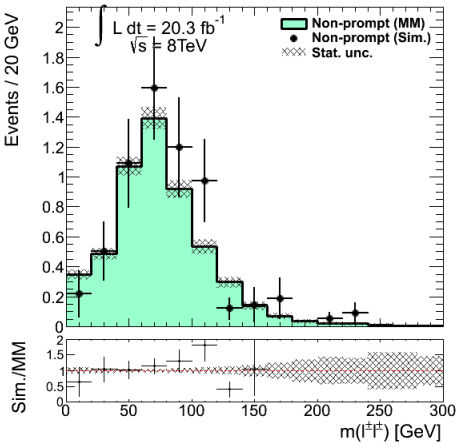
(a) $e^{\pm}e^{\pm}$, closure region



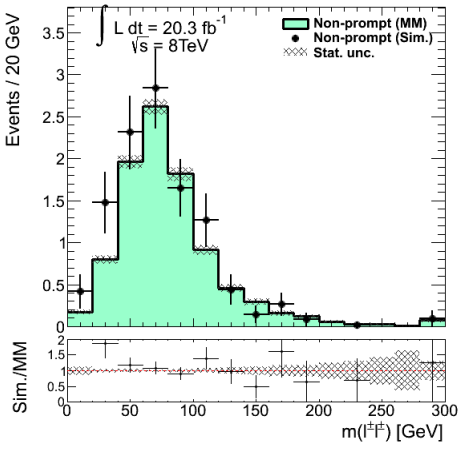
(b) $e^{\pm}e^{\pm}$, signal region



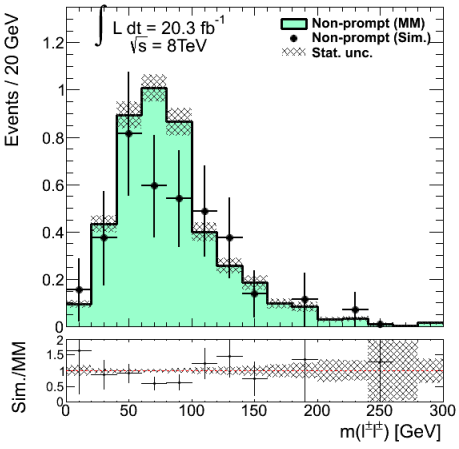
(c) $e^{\pm}\mu^{\pm}$, closure region



(d) $e^{\pm}\mu^{\pm}$, signal region



(e) $\mu^{\pm}\mu^{\pm}$, closure region



(f) $\mu^{\pm}\mu^{\pm}$, signal region

Figure 6.16: Distributions of the di-lepton invariant mass for $e^{\pm}e^{\pm}$ (top), $e^{\pm}\mu^{\pm}$ (centre) and $\mu^{\pm}\mu^{\pm}$ (bottom) events in $t\bar{t}$ simulation. The distributions are shown separately for the closure regions (left) and for the signal regions (right). Simulated $t\bar{t}$ non-prompt events (black dots) are compared to the predictions from the Matrix Method applied to the $t\bar{t}$ simulated events (light green). Overflow events are added to the last bin. The lower panel shows the ratio between the predictions from simulation and from the Matrix Method. The error bars and the dashed band show the statistical uncertainties on simulation and on the Matrix Method predictions, respectively. Uncertainties on the rates are not included.

sured in data as shown in Section 6.4.5.3 and are used to scale data events with same-sign *loose* leptons. The Matrix Method and the Fake Factor Method predicted yields in the closure regions and in the signal regions are summarized in Table 6.12. The jet multiplicity distributions of the non-prompt background in the $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ categories obtained from the two methods are compared in Figure 6.17(right). The methods pro-

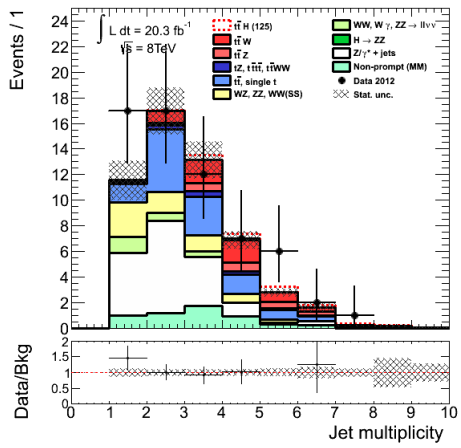
Region	Fake Factor Method	Matrix Method
$e^{\pm}e^{\pm}$ (N (jet) ≤ 3)	$3.6 \pm 0.5(\text{stat.})^{+13.0}_{-0.5}(f_{\ell}^{FF})$	$3.9 \pm 0.5(\text{stat.})^{+11.6}_{-2.1}(f_{\ell})$
$e^{\pm}e^{\pm}$ (N (jet) ≥ 4)	$1.4 \pm 0.3(\text{stat.})^{+4.4}_{-0.5}(f_{\ell}^{FF})$	$1.5 \pm 0.3(\text{stat.})^{+4.1}_{-0.8}(f_{\ell})$
$e^{\pm}\mu^{\pm}$ (N (jet) ≤ 3)	$17 \pm 1(\text{stat.})^{+20}_{-7}(f_{\ell}^{FF})$	$18 \pm 1(\text{stat.})^{+21}_{-8}(f_{\ell})$
$e^{\pm}\mu^{\pm}$ (N (jet) ≥ 4)	$6.2 \pm 0.6(\text{stat.})^{+8.0}_{-3.2}(f_{\ell}^{FF})$	$6.3 \pm 0.6(\text{stat.})^{+8.4}_{-2.7}(f_{\ell})$
$\mu^{\pm}\mu^{\pm}$ (N (jet) ≤ 3)	$16 \pm 1(\text{stat.})^{+10}_{-7}(f_{\ell}^{FF})$	$17 \pm 1(\text{stat.})^{+11}_{-6}(f_{\ell})$
$\mu^{\pm}\mu^{\pm}$ (N (jet) ≥ 4)	$6.0 \pm 0.5(\text{stat.})^{+4.5}_{-2.7}(f_{\ell}^{FF})$	$6.1 \pm 0.5(\text{stat.})^{+4.8}_{-2.4}(f_{\ell})$

Table 6.12: The predicted non-prompt background yields from Fake Factor Method and from the Matrix Method in the $2\ell 0\tau_{\text{had}}$ signal regions and the closure regions. For Matrix Method (Fake Factor Method) in addition to the statistical error, the uncertainty on f_{ℓ} (f_{ℓ}^{FF}) is accounted. The uncertainty on r_{ℓ} is negligible for each category.

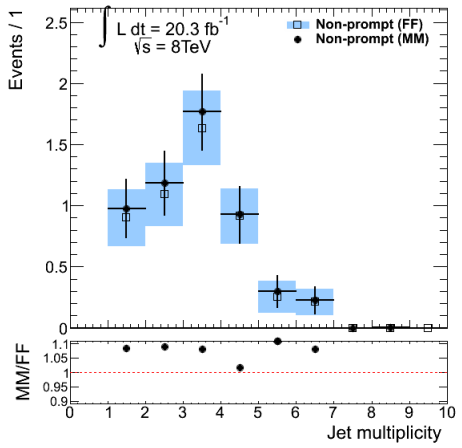
duce very similar estimates with comparable uncertainties. The largest uncertainty on the methods is by far due to the precision which the f_{ℓ} and f_{ℓ}^{FF} are calculated with. The uncertainties on the r_{ℓ} have a negligible effect on the Matrix Method predictions and are omitted in Table 6.12. In the $\mu^{\pm}\mu^{\pm}$ category, the differences between the methods are as little as 2%, while in the $e^{\pm}e^{\pm}$ category the Matrix Method predictions can exceed the Fake Factor Method ones by 10%. The origin of the differences lies in the way the prompt, charge misID and photon conversion backgrounds are treated in the two methods as explained in Section 6.4.5.1. The larger difference in the categories with electrons has to be attributed to the additional background produced by charge misID processes.

The predictions of the methods are tested against data. Figure 6.17(left) shows the jet multiplicity distributions observed in data and the total background predictions, where the non-prompt background is obtained with the Matrix Method. Predictions are compatible with data for any jet multiplicity and lepton flavour combination. The largest difference is found in $\mu^{\pm}\mu^{\pm}$ events with four jets. This difference is well within the uncertainty associated to the non-prompt efficiencies. The effect of these uncertainties on the total background predictions are shown in Figure C.11. Equivalent predictions are obtained with the Fake Factor Method since the differences in the predicted non-prompt background are not significant, as shown in Figure 6.17(right).

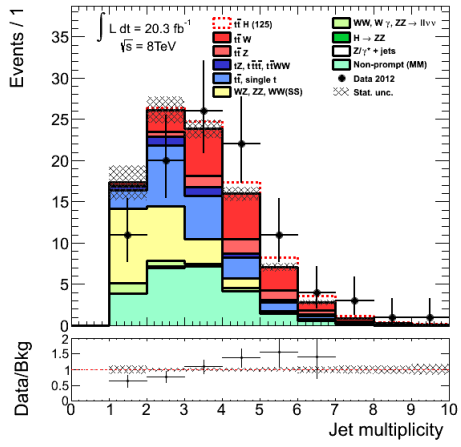
The modeling of the event kinematic is tested in the closure regions. The di-lepton invariant mass distributions in data and predictions for $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ events are shown in Figure 6.18. Agreement is found in any region, however the statistics and the large fraction of backgrounds from other sources do not allow to perform sensitive tests, especially in the $e^{\pm}e^{\pm}$ category. In the $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ categories that are less



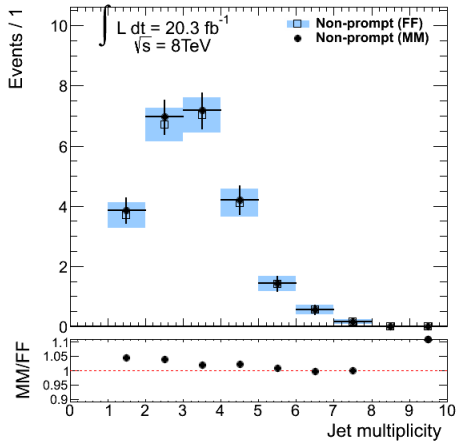
(a) $e^{\pm}e^{\pm}$



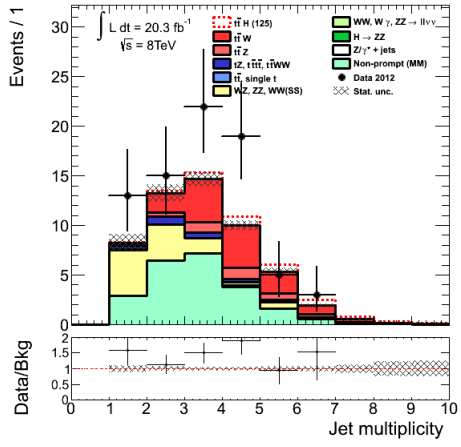
(b) $e^{\pm}e^{\pm}$



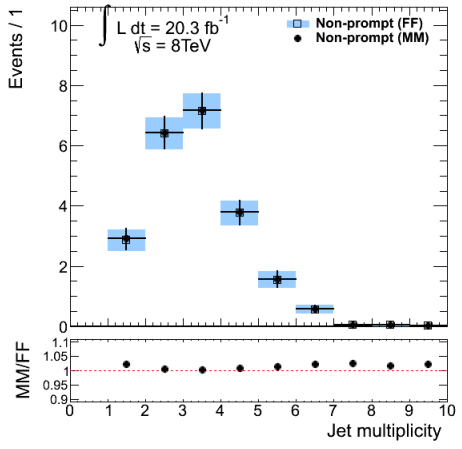
(c) $e^{\pm}\mu^{\pm}$



(d) $e^{\pm}\mu^{\pm}$

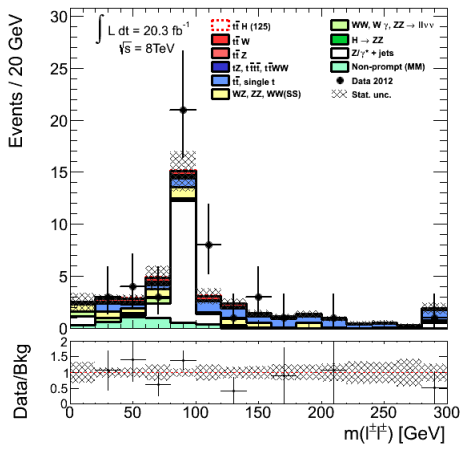


(e) $\mu^{\pm}\mu^{\pm}$

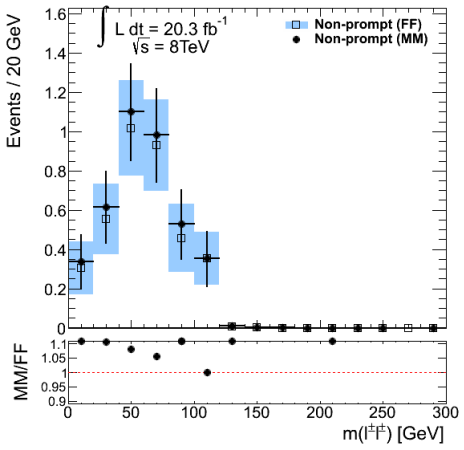


(f) $\mu^{\pm}\mu^{\pm}$

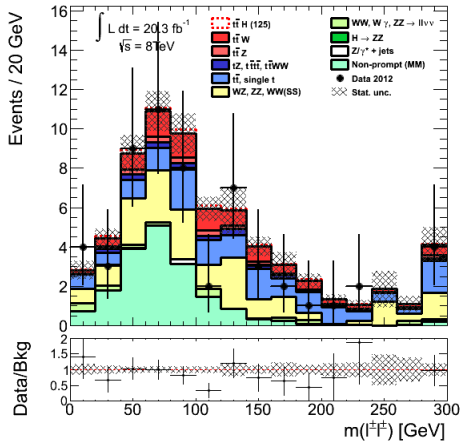
Figure 6.17: Distributions of the jet multiplicity for $e^{\pm}e^{\pm}$ (top), $e^{\pm}\mu^{\pm}$ (centre) and $\mu^{\pm}\mu^{\pm}$ (bottom) events. (left): Events in data (black dots) are compared to the expected stacked contributions from Standard Model processes with their ratio shown in the lower panel. The error bars and the dashed band show the statistical uncertainties on the data and the predictions, respectively. (right):Matrix Method predictions (black dots) are compared to Fake Factor Method predictions (open squares) with their ratio shown in the lower panel. The error bars and the light blue band show the statistical uncertainties on the Matrix Method and the Fake Factor Method predictions, respectively. In all the plots, overflow events are added to the last bin.



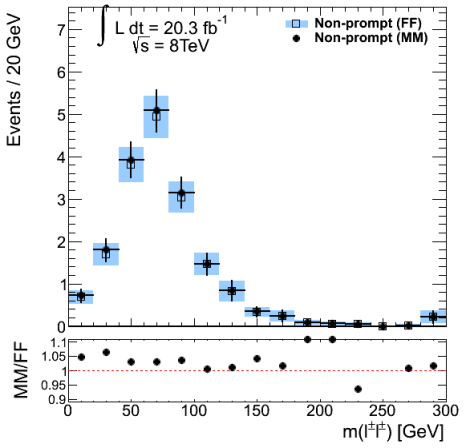
(a) $e^{\pm}e^{\pm}$



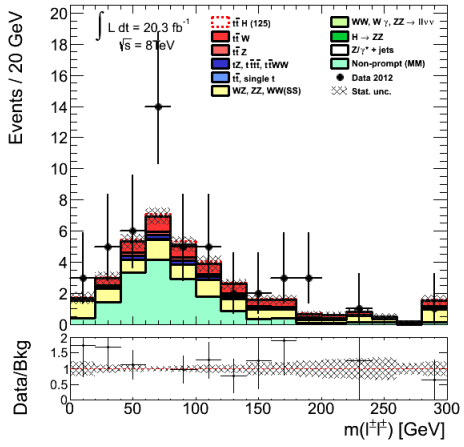
(b) $e^{\pm}e^{\pm}$



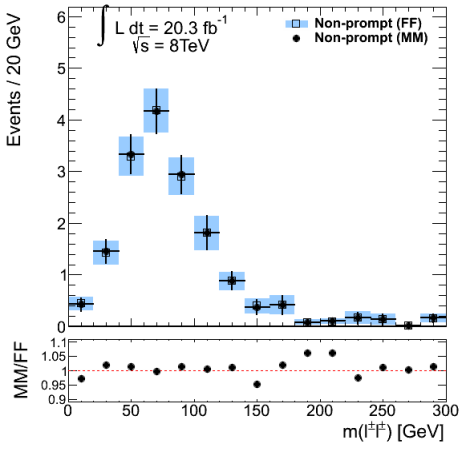
(c) $e^{\pm}\mu^{\pm}$



(d) $e^{\pm}\mu^{\pm}$



(e) $\mu^{\pm}\mu^{\pm}$



(f) $\mu^{\pm}\mu^{\pm}$

Figure 6.18: Distributions of the di-lepton invariant mass for $e^{\pm}e^{\pm}$ (top), $e^{\pm}\mu^{\pm}$ (centre) and $\mu^{\pm}\mu^{\pm}$ (bottom) events in the closure region (N (jet) ≤ 3). (left): Events in data (black dots) are compared to the expected stacked contributions from Standard Model processes with their ratio shown in the lower panel. The error bars and the dashed band show the statistical uncertainties on the data and the predictions, respectively. (right):Matrix Method predictions (black dots) are compared to Fake Factor Method predictions (open squares) with their ratio shown in the lower panel. The error bars and the light blue band show the statistical uncertainties on the Matrix Method and the Fake Factor Method predictions, respectively. In all the plots, overflow events are added to the last bin.

contaminated by other backgrounds, the methods seem to model well the invariant mass distributions. The comparison of di-lepton invariant mass distributions for signal regions events shows no significant differences as the statistics collected in these regions is even smaller (see Figure C.10).

The limited statistics do not allow to determine whether the Matrix Method or the Fake Factor Method give a better modeling of the non-prompt background. However, it would be important to understand if any of the two methods can improve the estimates obtained with the θ Method. The θ Method is a simplified version of the Fake Factor Method, where only one overall fake factor, θ_ℓ, is measured and events with two *non-tight* leptons are neglected. Assuming the latter condition valid, the Fake Factor Method and the θ Method should in principle produce similar estimates in terms of yields predictions, but there are exceptions. Suppose that the fake factors depend heavily on a certain variable, the p_{T} for instance. If the p_{T} distribution of non-prompt leptons is significantly different in the non-prompt regions and in the signal regions (or where the non-prompt background is estimated), the θ Method prediction presents a bias because the average *tight*/*non-tight* ratio in the signal regions is different from the value of θ_ℓ. By using fake factors parametrised in p_{T} instead, their average would adjust to the *tight*/*non-tight* ratio expected in the signal regions. In this study, the predictions obtained by the two methods in the signal regions are compatible but the limited statistics do not allow to investigate if a bias in the θ Method estimate exists. What can be investigated instead is the ability of the methods to model the non-prompt background kinematic.

As already mentioned, the θ Method itself is not designed to predict the properties of the estimated background. However, the yields predicted by the θ Method could be used to scale non-prompt background distributions obtained either from simulation or from data with *non-tight* leptons. The former solution requires the study of the systematic uncertainties associated with the modeling of non-prompt leptons in simulation. This is typically challenging, which is the reason of the search for alternative solutions based on data. The use of data containing *non-tight* leptons to obtain shape distributions assumes that there is no difference in kinematics between the events with *tight* and *non-tight* non-prompt leptons. This is also partly assumed in the Matrix Method and the Fake Factor Method but the methods parametrise their corrections factors as a function of those variables which are more affected by the change in the lepton selection condition from *non-tight* to *tight*. In this analysis, the fake factors and efficiencies have been parametrised as a function of the p_{T} and $|\eta|$ of the non-prompt lepton. The probability of a non-prompt lepton to pass *tight* selection depends indeed by its kinematic as the non-constant fake factors in Figure 6.13 show. Therefore, samples with *tight* and *non-tight* leptons present at least a difference in the kinematic of the non-prompt leptons that pass the *tight* and *non-tight* selections.

The clearest difference for what concerns kinematic modeling between the Fake Factor Method and the θ Method, which uses the shapes of data with *non-tight* leptons, should be observed in the p_{T} and $|\eta|$ distributions of the non-prompt leptons. Since the

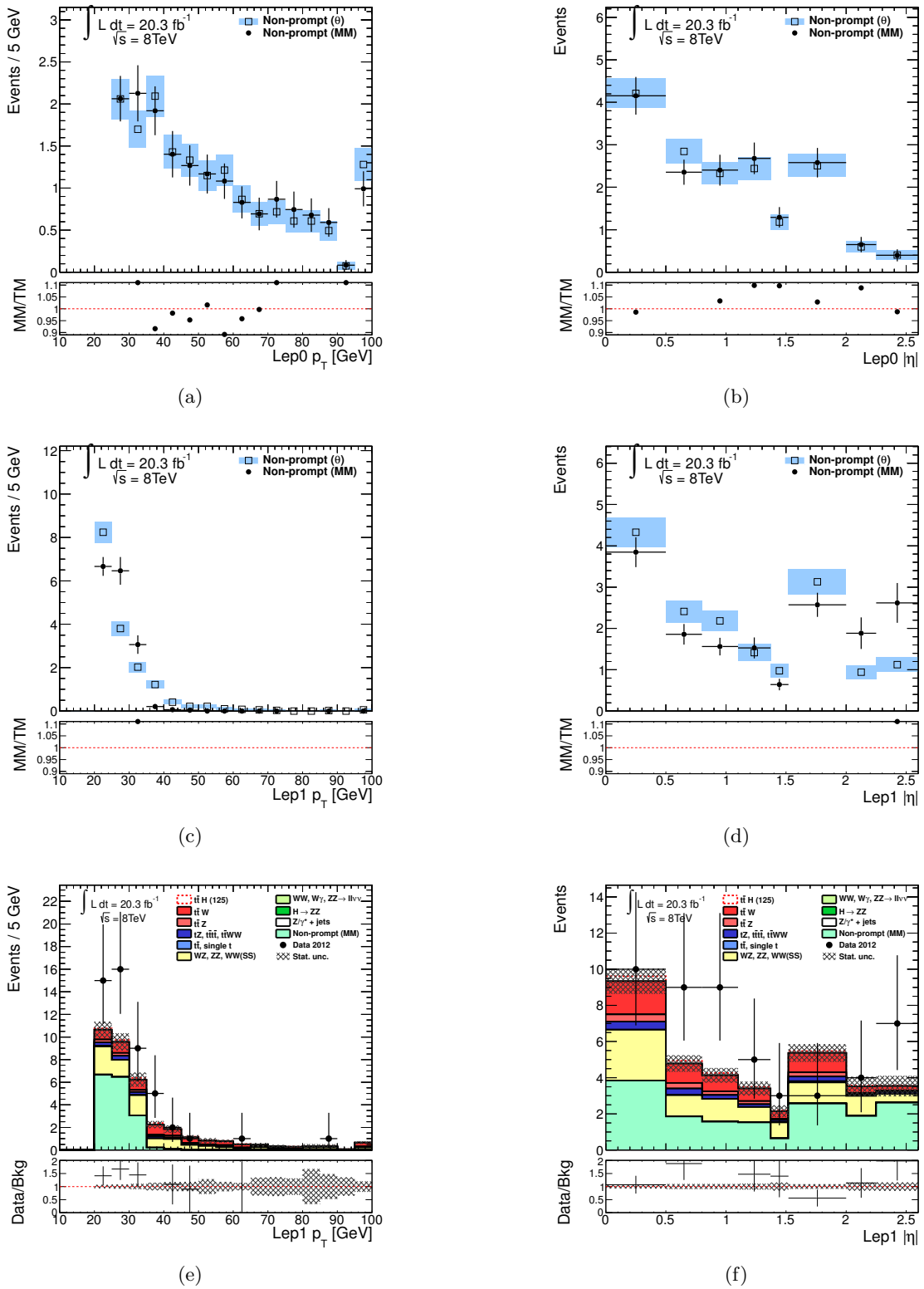
muon fake factors have been estimated with a precision superior to the electron ones, and the $\mu^{\pm}\mu^{\pm}$ category presents the lowest background beside the non-prompt one, the non-prompt lepton kinematic modeling of the θ Method and the Matrix Method are studied in the $\mu^{\pm}\mu^{\pm}$ closure region. The non-prompt lepton is typically the one with lower p_{T} in the pair, as emerged from simulation studies. Thus, a difference in modeling of the θ Method and the Fake Factor Method is expected mainly in the p_{T} and $|\eta|$ distributions of the lowest p_{T} lepton. Comparisons of the p_{T} and $|\eta|$ distributions of the leading (Lep0) and subleading (Lep1) p_{T} leptons in $\mu^{\pm}\mu^{\pm}$ events obtained with the two methods are shown in Figure 6.19. The predictions of the leading p_{T} lepton distributions are similar but some remarkable differences can be observed in the subleading p_{T} lepton distributions. The largest differences are observed for $p_{\mathrm{T}} < 35$ GeV and $|\eta| > 2$, where the Matrix Method estimate exceeds the θ Method one. The comparison of the total Standard Model prediction with data, Figure 6.19(bottom), seems to indicate that the distributions obtained with the Matrix Method are more in agreement with data in these regions[8]. The comparison of the predictions in the other lepton flavour final states do not lead to conclusive observations due to the larger uncertainties and background contaminations. The signal regions have not enough statistics to test the modeling of the Matrix Method and the θ Method. The use of the Matrix Method and Fake Factor Method seems to lead to some advantages over the θ Method, in particular if variables sensitive to the non-prompt lepton kinematic are used in future analysis.

6.4.5.7 Summary of the non-prompt background studies

The predictive power of the Matrix Method and Fake Factor Method for the estimate of non-prompt backgrounds in the $t\bar{t}H$ analysis has been investigated with data and $t\bar{t}$ simulation. The performed tests show no biases with the predictions of the two methods. However, the study is limited by the statistics and by the large uncertainty associated with the f_{ℓ}^{FF} and f_{ℓ} measurements attributable as well to the low statistics in the non-prompt regions.

With these limitations, the Matrix Method and the Fake Factor Method do not provide predictions significantly superior to those of the θ Method which are used in the published analysis. Some difference are observed in the modeling of the variables associated to the non-prompt lepton kinematic, like the subleading p_{T} lepton distributions. Indeed, the f_{ℓ}^{FF} and the f_{ℓ} show a dependence on these variables, which is not accounted for by the θ Method. The modeling with the θ Method, with the current statistics, is still compatible with those by the other methods, but with more data it would eventually exhibit significant biases in the determination of the non-prompt background kinematic. The Matrix Method and Fake Factor Method should instead be able to correct the kinematic

[8]There are observations which would favourite the θ Method prediction, like the $0.5 < |\eta| < 1$ region, but the differences in predictions of the two methods here are significantly smaller.



(a) (b) (c) (d) (e) (f)

Figure 6.19: The p_{T} (left) and $|\eta|$ (right) distributions of the leading p_{T} lepton (top) and the subleading p_{T} lepton (centre) expected from the Matrix Method (black dots) and the θ Method (open squares) in the $\mu^{\pm}\mu^{\pm}$ closure region ($N\,(\mathrm{jet}) \leq 3$). The error bars and the light blue band show the statistical uncertainty on the Matrix Method and the θ Method predictions, respectively. For the subleading p_{T} lepton the observed events and the total Standard Model expected ones are also shown (bottom). The error bars and the dashed band show the statistical uncertainty on the data and the predictions, respectively.

biases thanks to the additional information intrinsic to the parametrisation of efficiencies and fake factors. These methods are therefore preferable in future analyses that will make use of the kinematic properties of the signal regions.

In light of these studies, the Matrix Method and Fake Factor Method predictions are equivalent. If smaller f_ℓ^{FF} and f_ℓ uncertainties will be achieved, the predictions and the associated uncertainties obtained with the Matrix Method and Fake Factor Method may become significantly diverse. The two methods indeed deal differently with the sources of leptons other than the non-prompt processes. The r_ℓ of the Matrix Method are measured in the prompt region from data, thus the statistics in this region and the contamination from non-prompt background are sources of uncertainty. Another uncertainty comes from the fact that the r_ℓ should also account for the selection efficiencies for charge misID and photon conversion electrons in the $e^\pm\mu^\pm$ and $e^\pm e^\pm$ regions, as discussed in Section 6.4.5.1. This may be the largest uncertainty on the electron r_ℓ, because the efficiencies of these types of electrons may be significantly different from the prompt ones and their determination requires the definition of regions enriched of charge misID and photon conversion electrons selected with *tight* and *non-tight* requirements. On the other side, the Fake Factor Method requires an accurate estimate of the prompt, charge misID and photon conversion backgrounds in events with *non-tight* leptons and otherwise identical to the signal regions. These backgrounds must indeed be subtracted from data as described in Section 5.4.4.1. The background predictions may have a significant error, especially large statistical fluctuations, because of the very selective requirements that define these regions. Theoretical uncertainties on the simulated processes may also be large in these regions. Future analyses will establish which method have the lowest associated systematic uncertainties, thus which one is preferable.

The main problem to solve at the moment in order to obtain a better non-prompt background prediction seems to be the statistics especially in the non-prompt regions where the non-prompt efficiencies and fake factors are calculated. A reduction of this uncertainty is achievable in future since the dataset collected at 13 TeV will contain a higher non-prompt background statistics and multiple non-prompt regions could also be combined to obtain more precise efficiency measurements.

6.5 Validation regions

Validation regions are used to establish the quality of the background predictions. Each validation region is designed to be enriched of a specific background so that its modeling can be validated via comparison with data. The requirements used for the particle selection in the validation regions is identical to the signal regions (see Section 6.3.1). The validation regions however are orthogonal to the signal regions and differ for particle multiplicity and lepton kinematic.

The background in these regions is estimated according to Section 6.4, with few ex-

ceptions as specified below. The $t\bar{t}W^{\pm}$ region is the only validation region with significant non-prompt background contamination, and this is estimated with the θ Method. For the other validation regions, the non-prompt background prediction is obtained from simulations. The considered non-prompt background processes are described in Section C.1, and the process contributing the most to the non-prompt background is $t\bar{t}$ production. In the validation regions with opposite-sign leptons, the processes listed in Table C.1 are also used to estimate the prompt background. The uncertainties on background predictions include experimental, theoretical, statistical and data-driven uncertainties as described in Section 6.6.

The main backgrounds for the $2\ell 0\tau_{\text{had}}$ category are non-prompt, charge misID, $t\bar{t}W^{\pm}$, $t\bar{t}Z$ and $W^{\pm}Z$ processes. A validation test of the non-prompt prediction with the θ Method has been described in Section 6.4.4.1 (see Equation 6.9). The θ factors and the charge misID rates have been further tested with simulations. The author had no involvement in these studies that are described elsewhere [184].

The definitions of the validation regions is designed to target the dominant background processes of the $2\ell 0\tau_{\text{had}}$ category are defined in Table 6.13. Additional validation regions are used to test processes dominant in other signal region categories. These regions are described in Section C.4.

Validation region	Selection and target of the validation
$t\bar{t}$ opposite-sign (e^+e^-,$e^{\pm}\mu^{\mp}$,$\mu^+\mu^-$)	two opposite-sign leptons, leading lepton $p_{\text{T}} > 25$ GeV, $N\,(\text{jet}) \geq 2$ and $N\,(b\text{-jet}) \geq 1$, $\lvert m_{\ell\ell}^{OS\ SF} - m_Z\rvert > 10$ GeV Check b-tag corrections for simulated samples, check composition of the data used to estimate charge misID background
$t\bar{t}W^{\pm}$	Same as the $2\ell 0\tau_{\text{had}}$ category selection but $N\,(b\text{-jet}) \geq 2$ and $N\,(\text{jet}) \leq 3$ Verify $t\bar{t}W^{\pm}$ normalization, modeling
$t\bar{t}Z$	Same as the 3ℓ category selection (see Table C.2), but $\lvert m_{\ell\ell}^{OS\ SF} - m_Z\rvert < 10$ GeV Verify $t\bar{t}Z$ normalization, modeling
$W^{\pm}Z$	Same as the 3ℓ category selection (see Table C.2), but $\lvert m_{\ell\ell}^{OS\ SF} - m_Z\rvert < 10$ GeV, $N\,(b\text{-jet}) = 0$ and any number of jets Verify $W^{\pm}Z$ normalization and $N\,(\text{jet})$ spectrum
$W^{\pm}\gamma^*$	Same as the 3ℓ category selection (see Table C.2), but $N\,(b\text{-jet}) = 0$ and any number of jets Verify $W^{\pm}\gamma^*$ modeling
$W^{\pm}Z + b$-jet	Same as the 3ℓ category selection (see Table C.2), but $\lvert m_{\ell\ell}^{OS\ SF} - m_Z\rvert < 10$ GeV, $N\,(b\text{-jet}) \geq 1$ and any number of jets Verify $W^{\pm}Z$ modeling when heavy flavour hadrons are detected

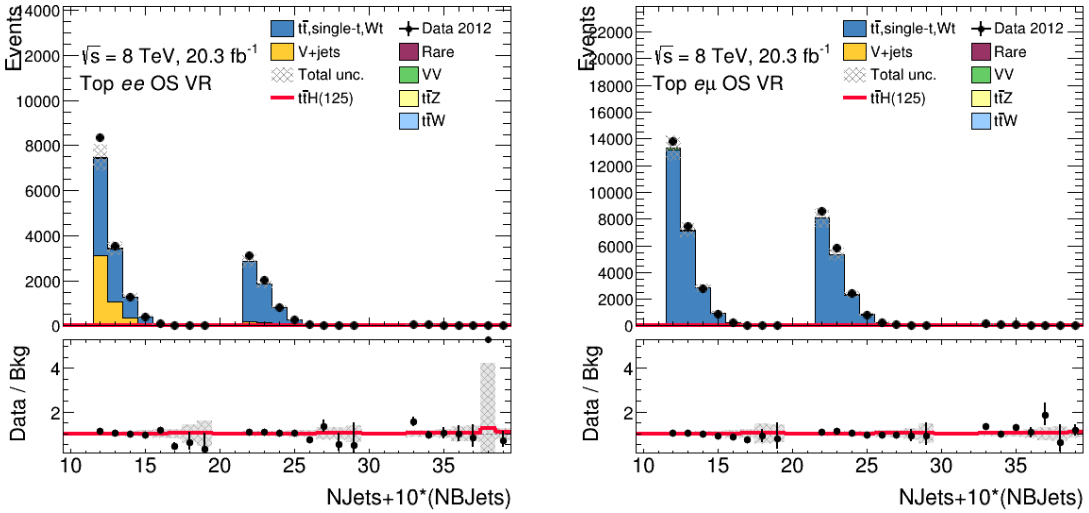
Table 6.13: Definition of the validation regions designed to check the prediction of the dominant backgrounds affecting the $2\ell 0\tau_{\text{had}}$ category.

$t\bar{t}$ opposite-sign validation region. A region with opposite-sign leptons, at least two jets and at least one b-jet is defined to test the modeling of prompt leptons, jets and b-jets reconstruction in simulations. Events with same flavour lepton pairs with $|m_{\ell\ell}^{OS\ SF} - m_Z| < 10$ GeV are vetoed to exclude Z + jets events The region is dominated by $t\bar{t}$ events where both the $W^{\pm}$ bosons from the t and $\bar{t}$ decay into leptons. Figure 6.20 shows the jet and b-jet multiplicity for these events. To conveniently represent the two quantities, the combination $N\,(\text{jet}) + 10 \cdot N\,(b\text{-jet})$ is shown. Data and predictions are in agreement for any number of jets and for any lepton flavour combination. This indicates a good modeling of the reconstruction and selection efficiencies of leptons and jets in simulation. Furthermore, the events of this region with $N\,(\text{jet}) \geq 4$ are used in the charge misID background estimate[9]. The almost total $t\bar{t}$ nature of the predictions and the good agreement with data, indicate that processes with non-prompt leptons, like $W^{\pm}$ + jets and QCD multi-jet productions, are a negligible background. The charge misID prediction in the same-sign regions is therefore not influenced significantly by background contaminations.

$t\bar{t}W^{\pm}$ validation region. In the decay of a $t\bar{t}W^{\pm}$ event in same-sign leptons two b-quarks and additional quarks are produced (see Figure 6.2(a)). Therefore the $t\bar{t}W^{\pm}$ validation region is defined by $N\,(b\text{-jet}) \geq 2$ and $N\,(\text{jet}) \leq 3$ in addition to two same-sign leptons. The number of jets is limited to three to avoid overlaps with the signal regions. The combined jet and b-jet multiplicity and the electron multiplicity of the $t\bar{t}W^{\pm}$ validation regions are shown in Figure 6.21. The electron multiplicity is chosen to conveniently display the lepton flavour composition of the events in this region ($e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$). The non-prompt, charge misID and $t\bar{t}W^{\pm}$ processes have similar magnitude and together form the 90% of the background. This region therefore validates also the predictions for non-prompt and charge misID backgrounds. The $e^{\pm}e^{\pm}$ region is particularly sensitive to the charge misID background, while the $\mu^{\pm}\mu^{\pm}$ one to non-prompt and $t\bar{t}W^{\pm}$ processes. Data and predictions agree within uncertainties in any flavour category and for any jet multiplicity.

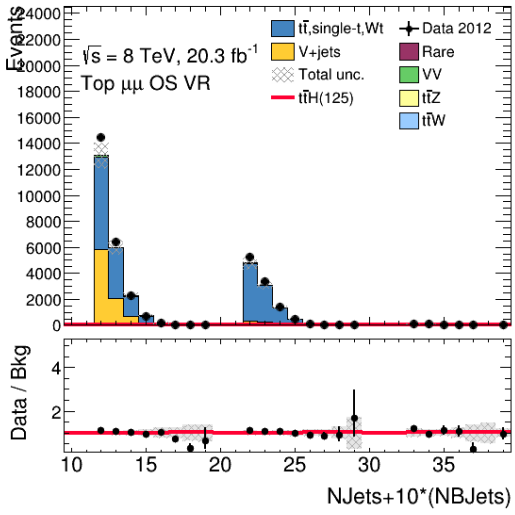
$t\bar{t}Z$ validation region. $t\bar{t}Z$ events are selected by requiring three leptons and the presence of an opposite-sign same flavour (OS SF) lepton pair compatible with the Z decay ($|m_{\ell\ell}^{OS\ SF} - m_Z| < 10$ GeV). The $t\bar{t}Z$ decay produces also two b-quarks and additional quarks (see Figure 6.2(b)). The selected events have $N\,(\text{jet}) \geq 4$ and $N\,(b\text{-jet}) \geq 1$. Events with $N\,(b\text{-jet}) \geq 2$ and $N\,(\text{jet}) = 3$ are also accepted. The jet and b-jet multiplicity and the lepton flavour composition of the $t\bar{t}Z$ validation regions are shown in Figure 6.22. The largest discrepancy between data and predictions is in events with three muons and is 1.6 times the error. Good agreement is observed otherwise.

[9]Events with $|m_{\ell\ell}^{OS\ SF} - m_Z| < 10$ GeV are used as well.



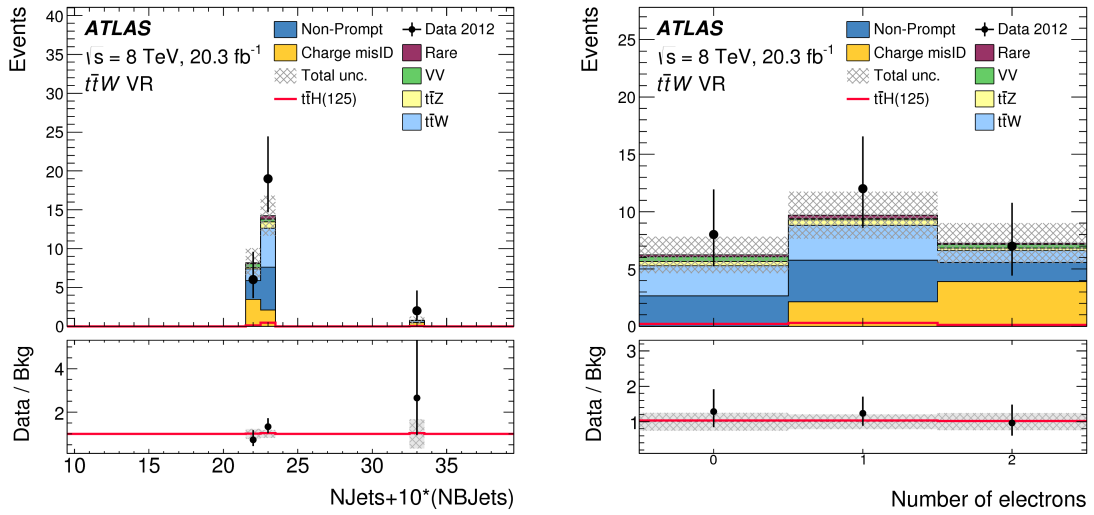
(a) e^+e^-

(b) $e^{\pm}\mu^{\mp}$



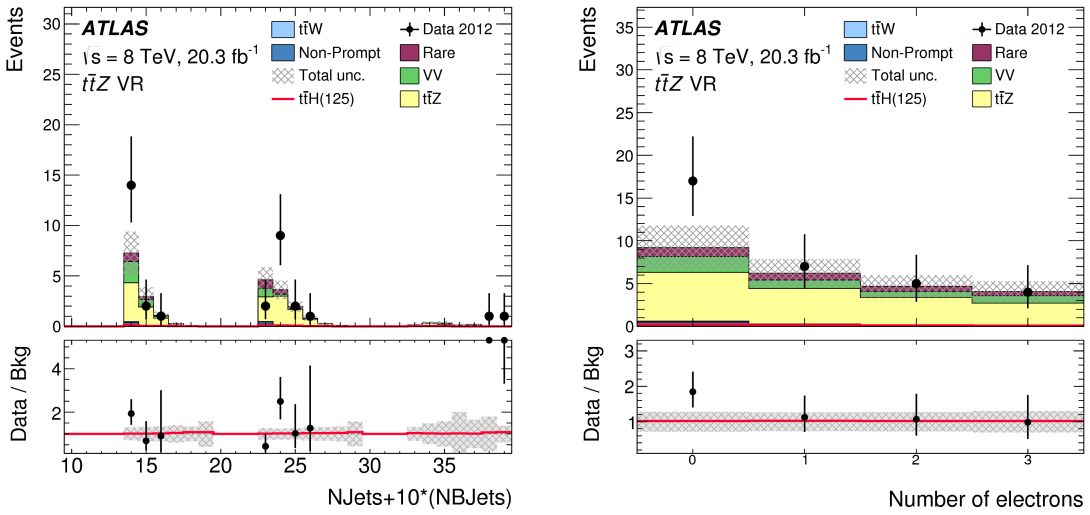
(c) $\mu^+\mu^-$

Figure 6.20: Distributions of the combined jet and b-jet multiplicities ($N\,(\text{jet}) + 10 \cdot N\,(b\text{-jet})$) in the $t\bar{t}$ opposite-sign validation region for e^+e^- (a), $e^{\pm}\mu^{\mp}$ (b) and $\mu^+\mu^-$ (c) events. In case of $N\,(\text{jet}) > 9$ events, these are accounted in the $9 + 10 \cdot N\,(b\text{-jet})$ interval to avoid mixing with events with an additional b-jet. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.



(a) (b)

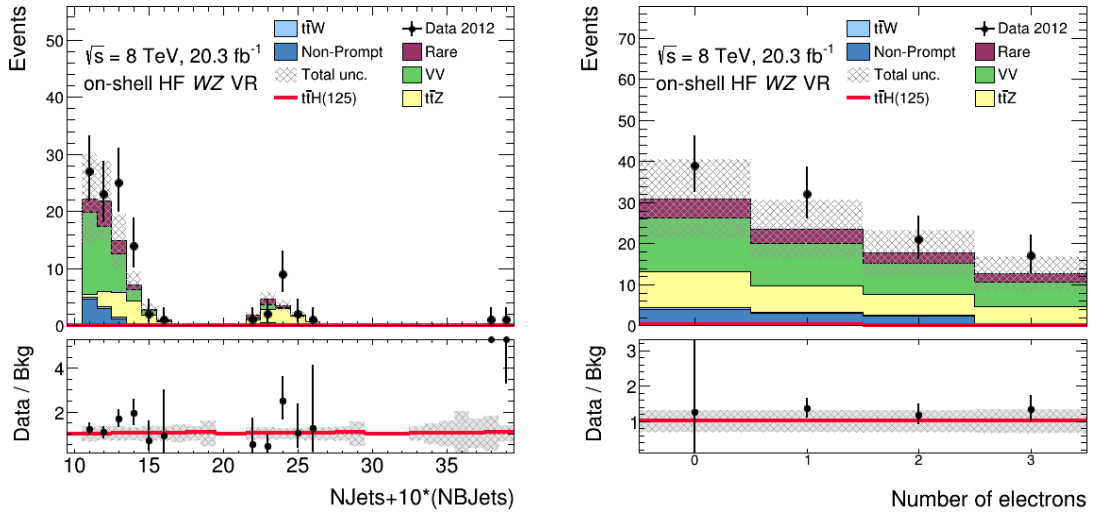
Figure 6.21: Distributions of the combined jet and b-jet multiplicities (left) and electron multiplicity (right) in the $t\bar{t}W^{\pm}$ validation region. In case of $N\,(\text{jet}) > 9$ events, these are accounted in the $9 + 10 \cdot N\,(b\text{-jet})$ interval of the $N\,(\text{jet}) + 10 \cdot N\,(b\text{-jet})$ plot to avoid mixing with events with an additional b-jet. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.



(a) (b)

Figure 6.22: Distributions of the combined jet and b-jet multiplicities (left) and electron multiplicity (right) in the $t\bar{t}Z$ validation region. In case of $N\,(\text{jet}) > 9$ events, these are accounted in the $9 + 10 \cdot N\,(b\text{-jet})$ interval of the $N\,(\text{jet}) + 10 \cdot N\,(b\text{-jet})$ plot to avoid mixing with events with an additional b-jet. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.

$W^{\pm}Z$+b-jet **validation region.** $W^{\pm}Z$ events are targeted requiring three leptons and an opposite-sign same flavour (OS SF) lepton pair with $|m_{\ell\ell}^{OS\ SF} - m_Z| < 10$ GeV. Contrary to the $t\bar{t}Z$ production, $W^{\pm}Z$ events are generally produced with less partons. The modeling of $W^{\pm}Z$ events in events with b-jets is tested by requiring $N\,(b\text{-jet}) \geq 1$ and any number of jets. Figure 6.23 shows the jet and b-jet multiplicity and the lepton flavour composition of this region. Data and predictions are in good agreement. The large jet multiplicity region is however only marginally populated by $W^{\pm}Z$ production[10].



(a) (b)

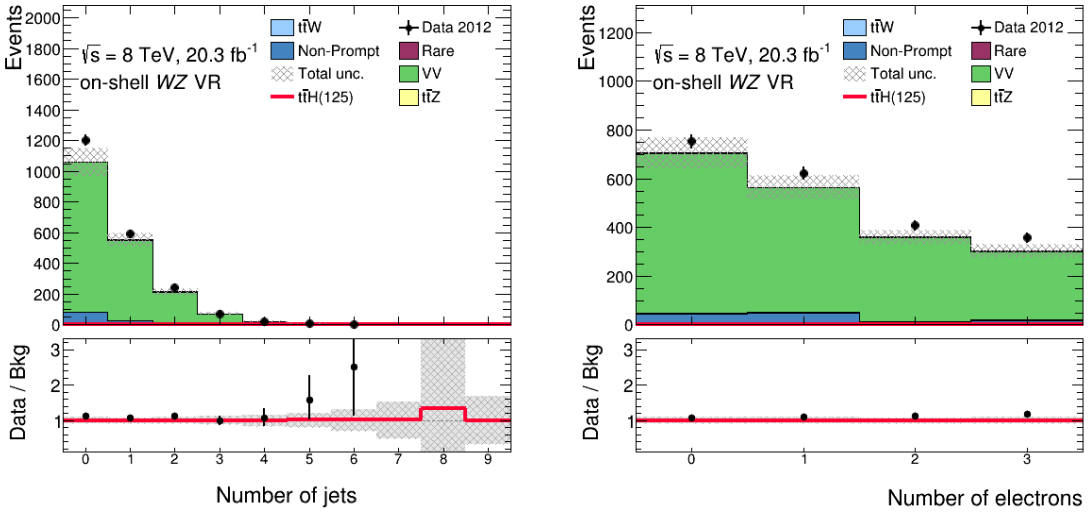
Figure 6.23: Distributions of the combined jet and b-jet multiplicities (left) and electron multiplicity (right) in the $W^{\pm}Z + b$-jet validation region. In case of $N\,(\text{jet}) > 9$ events, these are accounted in the $9 + 10 \cdot N\,(b\text{-jet})$ interval of the $N\,(\text{jet}) + 10 \cdot N\,(b\text{-jet})$ plot to avoid mixing with events with an additional b-jet. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.

$W^{\pm}Z$ **and** $W^{\pm}\ \gamma^*$ **validation regions.** The modeling of $W^{\pm}Z$ events with high multiplicity of reconstructed jets is tested in events with $N\,(b\text{-jet}) = 0$. The events must contain three leptons and an OS SF lepton pair with $|m_{\ell\ell}^{OS\ SF} - m_Z| < 10$ GeV. This region offers a way to test the $W^{\pm}Z$ production complementary to the $W^{\pm}Z + b$-jet validation region. The jet multiplicity and the lepton flavour composition of the selected events are shown in Figure 6.24. The $W^{\pm}Z$ production is the main process also at large jet multiplicity. There is good agreement between data and predictions.

The modeling of off-shell production ($W^{\pm}\ \gamma^*$) is also tested. In this case, the requirement of an opposite-sign same flavour (OS SF) lepton pair is inverted ($|m_{\ell\ell}^{OS\ SF} - m_Z| \geq$

[10]Note that by definition the hight jet multiplicity region is the $t\bar{t}Z$ validation region.

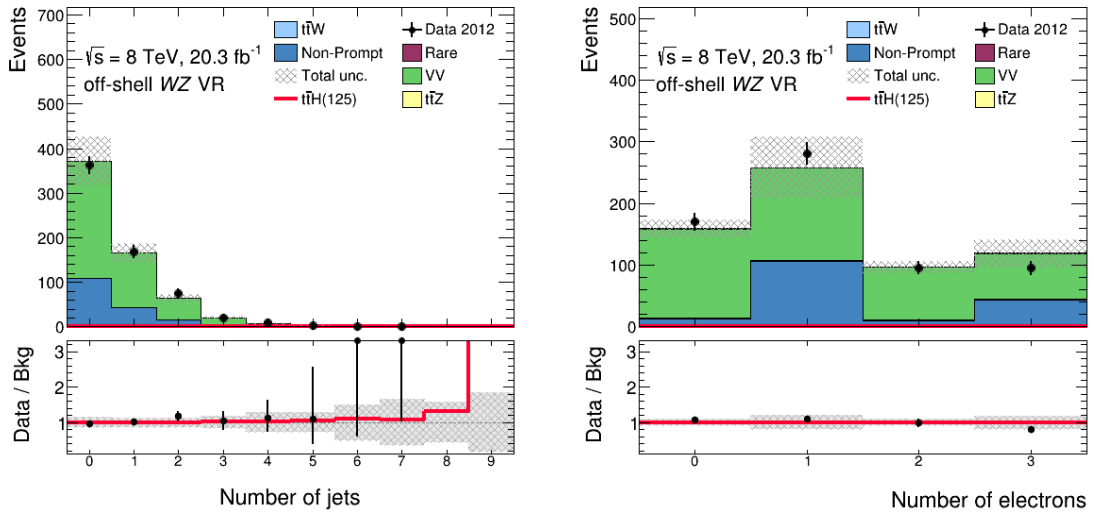
10 GeV). The jet multiplicity and the lepton flavour composition of these events are shown in Figure 6.25. Good agreement is found for any jet multiplicity and lepton flavour category.



(a) (b)

Figure 6.24: Distributions of the jet multiplicity (left) and electron multiplicity (right) in the $W^{\pm}Z$ validation region. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.

The data yields and the predictions in the validation regions described in this section are summarized in Table 6.14, Table 6.15 and Table 6.16. The contributions of the most relevant processes to the predictions are also given. Although the predictions tend to be smaller than the observed yields, the differences are everywhere compatible with the uncertainties on the predictions. The contamination of signal events in the validation regions is negligible. The highest contamination is expected in the $t\bar{t}W^{\pm}$ validation region and is less than 3%.



(a) (b)

Figure 6.25: Distributions of the jet multiplicity (left) and electron multiplicity (right) in the $W^{\pm}\gamma^{*}$ validation region. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.

	$t\bar{t}$ opposite-sign
V + jets	14800 ± 1600
$t\bar{t}$, *single-top*, $tW^{\pm}$	77800 ± 5300
$t\bar{t}W^{\pm}$	122 ± 30
$t\bar{t}Z$	99 ± 31
VV	483 ± 44
Rare	25.5 ± 3.6
$t\bar{t}H$	65 ± 12
Tot. Background	93400 ± 6600
Data	99016
Data/Background	1.060 ± 0.075

Table 6.14: Predicted and observed yields on the $t\bar{t}$ opposite-sign lepton validation region. The yields from the $e^{+}e^{-}$, $e^{\pm}\mu^{\mp}$ and $\mu^{+}\mu^{-}$ final states are combined. The full systematic uncertainties on the predictions are shown. Statistical data uncertainty is not included in the ratios.

	$t\bar{t}W^{\pm}$
Charge misID	6.0 ± 1.9
Non-prompt	8.0 ± 2.2
$t\bar{t}W^{\pm}$	6.7 ± 1.7
$t\bar{t}Z$	1.06 ± 0.36
VV	0.78 ± 0.71
Rare	0.58 ± 0.12
$t\bar{t}H$	0.62 ± 0.13
Tot. Background	23.1 ± 4.1
Data	27
Data/Background	1.17 ± 0.30

Table 6.15: Predicted and observed yields on the $t\bar{t}W^{\pm}$ validation regions. The yields from the $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ final states are combined. The full systematic uncertainties on the predictions are shown. Statistical data uncertainty is not included in the ratios.

	$W^{\pm}Z$	$W^{\pm}\gamma^{*}$	$W^{\pm}Z + b$-jet	$t\bar{t}Z$
Non-prompt	127 ± 56	172 ± 76	9.7 ± 4.3	0.71 ± 0.31
$t\bar{t}W^{\pm}$	0.301 ± 0.093	2.40 ± 0.62	1.39 ± 0.35	0.39 ± 0.10
$t\bar{t}Z$	6.4 ± 2.1	1.84 ± 0.59	24.0 ± 7.5	15.8 ± 5.0
VV	1780 ± 140	442 ± 36	46 ± 23	5.6 ± 3.2
$Rare$	10.4 ± 2.4	3.6 ± 1.3	12.4 ± 1.4	2.94 ± 0.39
$t\bar{t}H$	0.201 ± 0.041	0.95 ± 0.19	0.81 ± 0.15	0.53 ± 0.10
Tot. Background	1920 ± 160	622 ± 85	93 ± 32	25.4 ± 8.3
Data	2140	642	109	33
Data/Background	1.113 ± 0.093	1.03 ± 0.15	1.17 ± 0.41	1.30 ± 0.48

Table 6.16: Predicted and observed yields on in the three lepton validation regions. The full systematic uncertainties on the predictions are shown. Statistical data uncertainty is not included in the ratios.

6.6 Systematic uncertainties

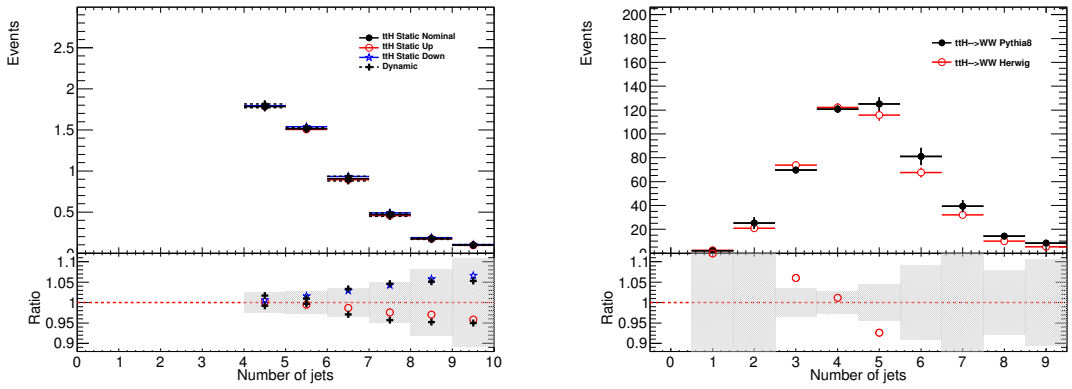
The uncertainties on the prediction of Standard Model backgrounds and the $t\bar{t}H$ signal are divided in four groups: theoretical, statistical, experimental and data-driven uncertainties.

Theoretical uncertainties. For each simulated process the uncertainties on the production cross-section and the modeling of the final state are evaluated by varying the parameters of the event generation process. These are the renormalization and factorization scales, the strong coupling constant, the parton density functions (PDF) tunes, the parton shower and the underlying event models. For the predictions of the signal and the main prompt backgrounds, $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ production, the effects of the variations are accounted for separately in the determination of the results. For secondary prompt backgrounds, an overall uncertainty that combines the uncertainties from different sources is instead considered.

The uncertainties on both the production cross-section and the category acceptances are evaluated. The category acceptance is the probability for events of the considered process to pass the signal region selection. Since the analysis signal regions have requirements on the number of jets and b-jets, the category acceptance is sensitive to parton shower modeling, PDF and QCD scales. Note that in the search for new physics described in Chapter 5, where jets are not used to define the signal regions, only the uncertainty on the total cross-section are evaluated. The category acceptance uncertainty for a given source is calculated separately in each category of the $t\bar{t}H$ analysis. The acceptance and cross-section uncertainties due to a given source are correlated among categories.

In the nominal $t\bar{t}H$ simulation the renormalization and factorization scales are both set to $\mu_0 = m(t) + m(H)/2$ (static scale). The $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ simulations use as renormalization and factorization scales $\mu_0 = m(t)$ and $\mu_0 = m(t) + m(Z)/2$, respectively. To assess the uncertainty due to the choice of the renormalization and factorization scales,

the total cross-section and the category acceptance are calculated for scales $2\mu_0$ and $\mu_0/2$. The difference with the nominal scale is taken as systematic uncertainty. For $t\bar{t}H$, a scale that depends on the kinematic of the event (dynamic scale) is also considered: $\mu = (M_{\mathrm{T}}(t)M_{\mathrm{T}}(\bar{t})M_{\mathrm{T}}(H))^{1/3}$ where $M_{\mathrm{T}}(p) = \sqrt{m^2(p) + p_{\mathrm{T}}^2(p)}$, thus the scale depend on the p_{T} of the generated particles. The comparison of static and dynamic scales allows to understand how the $t\bar{t}H$ system kinematic is sensitive to extra QCD radiation. The effects of the scale choice on the $t\bar{t}H$ predictions in the $2\ell 0\tau_{\mathrm{had}}$ category are shown in Figure 6.26(a). The difference between dynamic and static scale is as big as the difference between the nominal static scale and its variations. Since both differences account for the same source of uncertainty, the difference between static and dynamic scale (made symmetric) is taken as scale uncertainty. For $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ productions, the difference between the nominal static scale and its variations is the uncertainty considered.



(a) renormalization and factorization scale

(b) parton showering and underlying event

Figure 6.26: Impact of the theoretical uncertainties on the $t\bar{t}H$ predictions in the $2\ell 0\tau_{\mathrm{had}}$ category as a function of jet multiplicity. (left): Effect of the renormalization and factorization scale choice. The distributions obtained with the static scale, its variations by a factor of 2 and the dynamic scale are shown. (right): Effect of the parton shower and underlying event modeling on the jet multiplicity distribution. The PYTHIA8 and HERWIG modelings are shown. The absolute normalization of the histograms is arbitrary. The lower panel shows the ratio between the prediction of nominal and the alternative simulations. The gray bands in the lower panels represent the statistical uncertainties on the predictions of the nominal simulation.

The uncertainty of the $t\bar{t}H$ event selection due to the PDF modeling and the choice of the strong coupling constant α_{s} is estimated comparing the predictions from three PDF sets: CT10 [96], MSTW2008NLO [185] and NNPDF2.1 [186]. For each set, the predictions are evaluated for several α_{s} values. The width of the envelope is taken as systematic uncertainty following a dedicated procedure [160]. The uncertainties on $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ productions due to the PDFs and the strong coupling are obtained analogously.

The parton shower and the underlying event in the $t\bar{t}H$, $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ simulations are modeled with PYTHIA. An alternative modeling for the parton showering and the underlying event is offered by HERWIG++. The comparison of the two simulations is used

to estimate the systematic uncertainties on the modeling of these processes. Figure 6.26(b) shows the jet multiplicity for $t\bar{t}H$ events in the $2\ell 0\tau_{\text{had}}$ category (the N (jet) requirement is omitted) produced with PYTHIA8 and HERWIG++. The predictions can be as far as 20% for some jet multiplicities. The systematic uncertainty on the parton shower and underlying event modeling is calculated as the difference in category acceptance predicted by the two simulations. An uncertainty has been estimated for each signal region category. The same procedure is also used to calculate the uncertainties on the underlying event and parton shower modeling for $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ productions.

For the $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ simulations, uncertainties on the matrix element-to-parton shower matching scheme (matching) and on the LO modeling of jet multiplicity (LO modeling) are also evaluated. The uncertainty on the matching is obtained comparing the predictions of the baseline simulation for this analysis (MADGRAPH) with an alternative simulation (SHERPA). The difference in acceptance between the two simulations in each category is taken as uncertainty on the matching. In the $t\bar{t}H$ simulation, the matching is mediated by POWHEG-BOX and the associated uncertainty is found to be small [114]. The $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ productions are simulated at LO in QCD, although their cross-sections are calculated at NLO. In order to determine the accuracy with which the LO calculation models the jet multiplicity (so the category acceptance), these processes have been also simulated with SHERPA and SHERPA-ONELOOP. The former simulation performs a LO calculation, while the latter considers also the next order in perturbation theory. The difference in category acceptance obtained from the two simulations is taken as uncertainty on the LO modeling. The $t\bar{t}H$ production is simulated at NLO in QCD, so no LO modeling uncertainty has been considered for it.

The diboson production cross-sections are known with few percent accuracy as shown in Table 5.16. Because of the jet and b-jet requirements in the signal regions, the uncertainties on the predictions of diboson processes can be significantly higher. The selection favourites events where b-quarks or c-quarks are produced in addition to VV, mainly $W^{\pm}Z + b$-jet production. Since the constraints from analyses [187] on the production of this process are not very strong, a conservative 50% uncertainty is assigned to $W^{\pm}Z$ production and a 25% one for the other diboson productions.

The theoretical uncertainties considered for each simulated process are summarized in Table 6.17.

Simulation statistical uncertainties. The processes modeled with simulations are usually produced with a statistics that is tens of times larger than the expected statistics in data (see Table 6.2). The resulting uncertainty on the predictions is typically small. Few exceptions in this analysis are the processes that produce preferentially low jet and b-jet multiplicities and that do not contain filters to increase the statistics of events produced with extra partons. The statistical uncertainty is particularly relevant for VV simulations, and can be as big as 50% in some signal regions.

Type	Description	Uncertainty
	$t\bar{t}H$	
QCD Scale	Cross Section	$^{+3.8\%}_{-9.3\%}$
	Category Acceptance	0.0-2.6%
PDF+α_S	Cross Section	± 8.1%
	Category Acceptance	0.3-1.0%
Parton Shower	Category Acceptance	0.2-1.4%
	$t\bar{t}W^{\pm}$	
QCD Scale	Cross Section	±12%
	Category Acceptance	1.5-1.6%
PDF+α_S	Cross Section	± 9%
	Category Acceptance	1.1-4.8%
Matching	Category Acceptance	0.2-10.2%
LO Modeling	Category Acceptance	0.1-15.6%
Parton Shower	Category Acceptance	2.4-13.0%
	$t\bar{t}Z$	
QCD Scale	Cross Section	±12%
	Category Acceptance	1.3-2.9%
PDF+α_S	Cross Section	± 9%
	Category Acceptance	0.9-2.7%
Matching	Category Acceptance	0.5-16.0%
LO Modeling	Category Acceptance	3.5-10.5%
Parton Shower	Category Acceptance	2.4-13.0%
	VV	
Overall	$W^{\pm}Z$	± 50%
	ZZ, W^+W^-	± 25%
	$W^{\pm}\gamma$, $W^{\pm}\gamma^*$ and $Z\gamma^*$	± 25%

Table 6.17: Summary of theoretical uncertainties considered in the $t\bar{t}H$ analysis. The category acceptance uncertainties are estimated for each signal regions and the table report the minimum and maximum uncertainty found among these categories.

Experimental uncertainties. The level of accuracy to which the LHC collisional conditions and the ATLAS particle detection are known and are reproduced in simulations is reflected in uncertainties on the predicted yields. There are multiple sources of experimental uncertainties. The integrated luminosity of the data sample analysed is measured with a 2.8% uncertainty [89]. The uncertainty on the average number of interactions per bunch crossing ($\langle\mu\rangle$) is about 3%. An uncertainties due to pileup conditions is obtained by varying $\langle\mu\rangle$ of as much as its uncertainty in simulations. The variation in the predicted yields is less than 0.5% in the $2\ell 0\tau_{\text{had}}$ category and smaller than 2% in any other category.

Uncertainties on the lepton, τ_{had}, jet and b-jet modeling are also considered. These uncertainties are described in Section 2.4 and include selection efficiencies, energy and momentum calibration and energy and momentum resolution and event acquisition efficiencies. The sources of uncertainties for lepton, jet, b-jet and τ_{had} are given in Table 2.2, Table 2.3, Table 2.4 and Table 2.5, respectively.

The three most important sources of experimental uncertainties affecting the predictions of $t\bar{t}H$, $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ yields in each $2\ell 0\tau_{\text{had}}$ category are listed in Table 6.18, Table 6.19 and Table 6.20. The luminosity uncertainty is not considered here since it is a constant 2.8% in any category and any process, but it is obviously accounted in the determination of the results. The uncertainties related to the lepton isolation and the jet

Category	1^{st} Unc. Down-Up(%)	2^{nd} Unc. Down-Up(%)	3^{rd} Unc. Down-Up(%)
$e^{\pm}e^{\pm}$ + 4jets	El. Isolation	El. ID	b-tag eff. b-quark 5
	-4.0 4.1	-1.6 1.6	-1.2 1.2
$e^{\pm}e^{\pm}+\geq$ 5jets	El. Isolation	JES Flav. Comp.	JES Method Modeling 1
	-4.1 4.2	-3.8 2.4	-3.8 2.3
$e^{\pm}\mu^{\pm}$ + 4jets	El. Isolation	Muon Isolation	JES Flav. Comp.
	-2.0 2.0	-2.0 2.0	1.3 -2.4
$e^{\pm}\mu^{\pm}+\geq$ 5jets	JES Flav. Comp.	JES Method Modeling 1	El. Isolation
	-2.9 3.2	-2.5 3.1	-2.0 2.00
$\mu^{\pm}\mu^{\pm}$ + 4jets	Muon Isolation	JES Flav. Comp.	JES Flav. Resp.
	-3.9 4.0	3.9 -2.2	3.3 -1.0
$\mu^{\pm}\mu^{\pm}+\geq$ 5jets	Muon Isolation	JES Flav. Comp.	JES Method Modeling 1
	-3.9 4.0	-3.4 2.7	-3.1 3.1

Table 6.18: The three most relevant experimental sources of uncertainty on the $t\bar{t}H$ expected yields in the $2\ell 0\tau_{\text{had}}$ categories.

energy calibration are dominant in all the categories and for each simulated process. The main uncertainties for other signal region categories are shown in Table C.6, Table C.7 and Table C.8.

An overall experimental systematic uncertainty is calculated by summing in quadrature the uncertainties from independent sources. The overall uncertainties are shown in Table 6.21 for the $2\ell 0\tau_{\text{had}}$ categories and in Table C.9 for the other categories. The ex-

Category	1^{st} Unc. Down-Up(%)	2^{nd} Unc. Down-Up(%)	3^{rd} Unc. Down-Up(%)
$e^{\pm}e^{\pm}$ + 4jets	El. Isolation -4.3 4.4	JES Flav. Comp. -2.5 -1.5	b-tag eff. b-quark 5 -1.8 1.8
$e^{\pm}e^{\pm}+ \geq$ 5jets	El. Isolation -4.4 4.5	JES Flav. Comp. -3.3 4.8	JES Method Modeling 1 -4.0 4.6
$e^{\pm}\mu^{\pm}$ + 4jets	El. Isolation -2.2 2.2	Muon Isolation -2.0 2.0	JES Method Modeling 1 0.2 2.7
$e^{\pm}\mu^{\pm}+ \geq$ 5jets	JES Flav. Comp. -5.2 3.5	JES Method Modeling 1 -4.9 3.9	JES pileup ρ -2.9 2.7
$\mu^{\pm}\mu^{\pm}$ + 4jets	Muon Isolation -4.0 4.1	b-tag eff. b-quark 5 -1.3 1.3	JES Flav. Comp. 0.4 1.51
$\mu^{\pm}\mu^{\pm}+ \geq$ 5jets	Muon Isolation -4.1 4.2	JES Flav. Comp. -3.7 2.2	JES Method Modeling 1 -3.5 2.4

Table 6.19: The three most relevant experimental sources of uncertainty on the $t\bar{t}W^{\pm}$ expected yields in the $2\ell 0\tau_{\text{had}}$ categories.

Category	1^{st} Unc. Down-Up(%)	2^{nd} Unc. Down-Up(%)	3^{rd} Unc. Down-Up(%)
$e^{\pm}e^{\pm}$ + 4jets	El. Isolation -4.2 4.3	JES pileup NPV -2.6 -2.3	JES Method Modeling 1 -1.7 -3.4
$e^{\pm}e^{\pm}+ \geq$ 5jets	El. Isolation -4.5 4.6	JES Flav. Comp. -1.5 5.3	JES Method Modeling 1 -1.5 4.7
$e^{\pm}\mu^{\pm}$ + 4jets	JER — 3.6	El. Isolation -2.2 2.2	Muon Isolation 2.0 -2.0
$e^{\pm}\mu^{\pm}+ \geq$ 5jets	JES Method Modeling 1 -4.3 2.4	JES Flav. Comp. -3.0 2.0	El. Isolation -2.3 2.3
$\mu^{\pm}\mu^{\pm}$ + 4jets	Muon Isolation -4.0 4.1	JES Flav. Comp. 2.9 3.9	JER — 3.3
$\mu^{\pm}\mu^{\pm}+ \geq$ 5jets	JES Flav. Comp. -5.5 4.2	Muon Isolation -4.1 4.2	JES Method Modeling 1 -4.5 3.3

Table 6.20: The three most relevant experimental sources of uncertainty on the $t\bar{t}Z$ expected yields in the $2\ell 0\tau_{\text{had}}$ categories.

pected uncertainties from experimental sources on the simulated yields vary between 3.5% and 12% depending on the category and the process considered. In the determination of the results, the experimental sources of uncertainty are accounted separately to improve the statistical power.

Total Systematic Uncertainty	$e^{\pm}e^{\pm}$ + 4jets Down	Up	$e^{\pm}e^{\pm}$ + $\geq$ 5jets Down	Up	$e^{\pm}\mu^{\pm}$ + 4jets Down	Up	$e^{\pm}\mu^{\pm}$ + $\geq$ 5jets Down	Up
$t\bar{t}H$	-4.7	5.8	-8.2	6.1	-5.1	3.5	-5.5	6.4
$t\bar{t}W^{\pm}$	-7.2	5.5	-8.7	11.3	-3.6	6.2	-9.7	8.0
$t\bar{t}Z$	-9.7	5.1	-5.9	11.0	-4.1	6.2	-8.4	5.0
Total Systematic Uncertainty	$\mu^{\pm}\mu^{\pm}$ + 4jets Down	Up	$\mu^{\pm}\mu^{\pm}$ + $\geq$ 5jets Down	Up				
$t\bar{t}H$	-5.2	7.5	-7.3	6.8				
$t\bar{t}W^{\pm}$	-4.5	5.2	-8.6	6.9				
$t\bar{t}Z$	-5.2	8.7	-9.7	8.2				

Table 6.21: Sum in quadrature of all the experimental systematic uncertainties on the number of $t\bar{t}H$, $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ event predictions divided per signal region. The luminosity uncertainty is not considered in the sum.

Data-driven uncertainties. The non-prompt and the charge misID backgrounds are estimated in this analysis using data-driven methods. For the θ Method, used to evaluate the non-prompt background, the uncertainties on the measured θ_{ℓ} and the predictions of events with a *non-tight* lepton $N_{\ell\not\ell}$ are accounted. The statistical error and the systematic uncertainties have been estimated in Section 6.4.4.1.

The uncertainty on the charge misID predictions is due to the uncertainties on the charge misID rates described in Section 6.4.2.2 and to the statistics of opposite-sign lepton data used in the estimate (Equation 6.3 and Equation 6.3).

6.7 Results

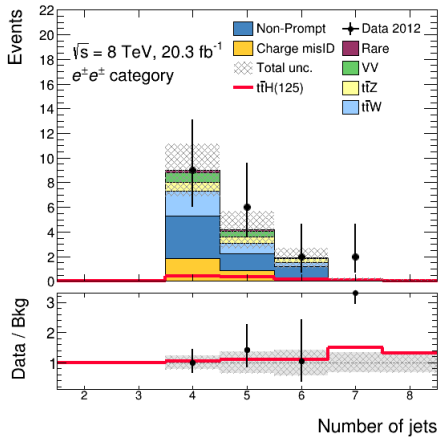
The data events observed in the signal region categories are analysed in this section. The expected contributions by Standard Model processes to the event yields in these categories have been estimated as explained in Section 6.4. The uncertainties affecting the signal and background predictions have been obtained in Section 6.6. The signal and background determinations have been performed before analysing the data events accepted in the signal regions to avoid biases. Data in the signal regions are here studied in search for evidence of the $t\bar{t}H$ production. The analysis signal regions are displayed in Section 6.7.1 and the measurement of the $t\bar{t}H$ cross-section is described in Section 6.7.2. The results are also interpreted as constraints on the Top Yukawa coupling for several scenarios.

6.7.1 Signal regions

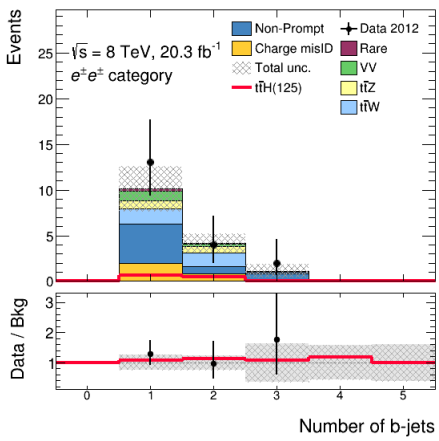
The number of data events observed in the $2\ell 0\tau_{\text{had}}$ signal regions is given in Table 6.22, together with the expected signal and background events. The largest number of signal events is expected in the $e^{\pm}\mu^{\pm}$ category. The $t\bar{t}H$ events have preferentially five or more jets while the background prevails in four jets. Thus the division of the signal regions in jet multiplicity increase the sensitivity of the $t\bar{t}H$ cross-section measurement. In most of the categories, the largest contribution to the background is due to non-prompt lepton events, followed by $t\bar{t}W^{\pm}$ events. The $t\bar{t}W^{\pm}$ production is instead the dominant background in the $\mu^{\pm}\mu^{\pm}+ \geq 5$jets category.

Overall, predictions are in agreement with data. The largest uncertainty on the background predictions comes from the non-prompt background estimate. For this reason, investigations on alternative methods for non-prompt background estimates, like those described in Section 6.4.5, are important for the sensitivity of future analyses. The largest differences between data and predictions are observed in the $e^{\pm}e^{\pm}+ \geq 5$jets, $e^{\pm}\mu^{\pm}+ \geq 5$jets and $\mu^{\pm}\mu^{\pm} + 4$jets. The jet and b-jet multiplicities for observed and expected events in the $2\ell 0\tau_{\text{had}}$ categories are shown in Figure 6.27. The largest excess of data over the prediction is observed in $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ events with two b-jets. Additional distributions of $2\ell 0\tau_{\text{had}}$ event properties are given in Section C.6. In that case, the shape of the non-prompt background predictions are obtained from simulation.

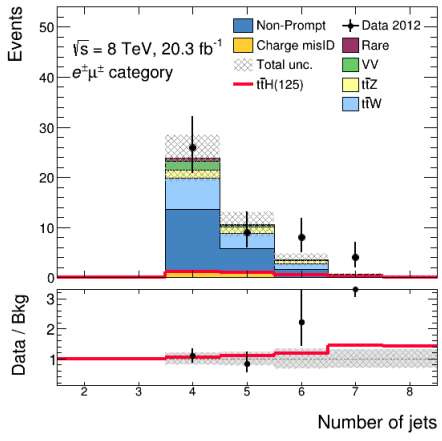
The event yields for the other analysis categories (3ℓ, 4ℓ, $2\ell 1\tau_{\text{had}}$ and $1\ell 2\tau_{\text{had}}$) are summarized in Table C.10. The combined jet and b-jet multiplicities and the lepton flavour composition of the events in these categories are shown in Section C.7. Among all the $t\bar{t}H$ analysis categories, the $2\ell 0\tau_{\text{had}}$ and the 3ℓ categories present the largest observed excess and have also the largest expected number of signal events. The findings in each category are used in the next section to measure the $t\bar{t}H$ production cross-section and are combined to provide a more solid measurement.



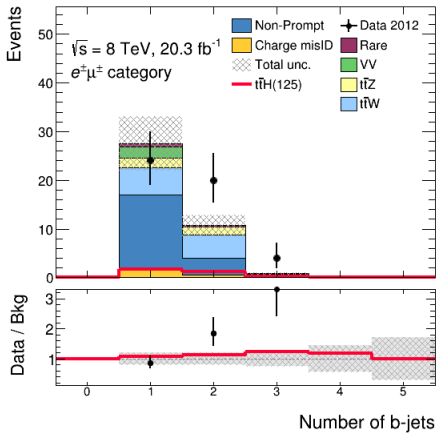
(a) $e^{\pm}e^{\pm}$



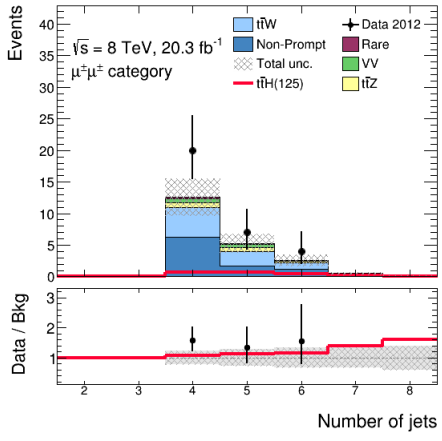
(b) $e^{\pm}e^{\pm}$



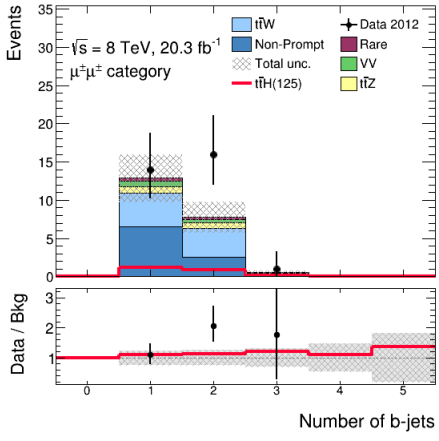
(c) $e^{\pm}\mu^{\pm}$



(d) $e^{\pm}\mu^{\pm}$



(e) $\mu^{\pm}\mu^{\pm}$



(f) $\mu^{\pm}\mu^{\pm}$

Figure 6.27: Distributions of the jet multiplicity (left) and the b-jet multiplicity (right) in the $2\ell 0\tau_{\text{had}}$ signal region for $e^{\pm}e^{\pm}$ (top), $e^{\pm}\mu^{\pm}$ (centre) and $\mu^{\pm}\mu^{\pm}$ (bottom) events. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.

Category	Charge misID	Non-prompt	$t\bar{t}W^{\pm}$	$t\bar{t}Z$	VV	Total bkg.	$t\bar{t}H$	Observed
$e^{\pm}e^{\pm}$ + 4jets	1.8 ± 0.7	3.4 ± 1.7	2.0 ± 0.4	0.75 ± 0.20	0.74 ± 0.42	9.1 ± 2.1	0.44 ± 0.06	9
$e^{\pm}\mu^{\pm}$ + 4jets	1.4 ± 0.6	12 ± 4	6.2 ± 1.0	1.5 ± 0.3	1.9 ± 1.0	24 ± 5	1.16 ± 0.14	26
$\mu^{\pm}\mu^{\pm}$ + 4jets	–	6.3 ± 2.6	4.7 ± 0.9	0.80 ± 0.22	0.53 ± 0.30	12.7 ± 2.9	0.74 ± 0.10	20
$e^{\pm}e^{\pm}+ \geq$ 5jets	1.1 ± 0.5	2.3 ± 1.2	1.4 ± 0.4	0.98 ± 0.26	0.47 ± 0.29	6.5 ± 1.8	0.73 ± 0.14	10
$e^{\pm}\mu^{\pm}+ \geq$ 5jets	0.85 ± 0.35	6.7 ± 2.4	4.8 ± 1.2	2.1 ± 0.5	0.38 ± 0.30	15 ± 3	2.13 ± 0.41	22
$\mu^{\pm}\mu^{\pm}+ \geq$ 5jets	–	2.9 ± 1.4	3.8 ± 0.9	0.95 ± 0.25	0.69 ± 0.39	8.6 ± 2.2	1.41 ± 0.28	11

Table 6.22: Expected and observed event yields in each flavour category and jet multiplicity of the $2\ell 0\tau_{\text{had}}$ signal region. The overall uncertainties on the predictions are also given. Rare processes (tZ, $t\bar{t}W^{+}W^{-}$, $H \to ZZ^{(*)}$, $tHqb$, $tHW^{\pm}$ and $t\bar{t}t\bar{t}$) are included in the total expected background estimate.

6.7.2 Measurement of the $t\bar{t}H$ production cross-section

The aim of the analysis is to search for a evidence of $t\bar{t}H$ production. The search is performed by comparing the data observed in the signal regions with the expected Standard Model processes. To do so, a statistical model of the type described in Section 5.6.2 is used. The model combines the information from each category into a likelihood functions where the uncertainties are treated with Nuisance Parameters and the $t\bar{t}H$ signal strength is an unconstrained parameter. The signal strength μ is defined as the ratio of the measured $t\bar{t}H$ production cross-section to the cross-section predicted by Standard Model. $\mu = 1$ is the expected signal strength from Standard Model, while $\mu = 0$ represents the background only hypothesis. In this model the decay modes and the branching ratio of the Higgs boson are fixed to the Standard Model values. Correlations among the predictions in any category are taken into account: experimental uncertainties are correlated among simulated processes and categories; data-driven uncertainties are correlated among categories; theoretical uncertainties are correlated among categories but not among processes; statistical errors are not correlated in simulated and observed events.

The statistical model is used to measure the $t\bar{t}H$ signal strength, to assign probabilities to μ hypotheses and to set CL_S upper limits on μ. The signal strength is determined with a maximum likelihood fit to the observed data yields. The fit returns the best fit value of the signal strength and of the Nuisance Parameters. The best value for μ in each of the $2\ell 0\tau_{\text{had}}$ categories and for their combination are shown in Figure 6.28. For each final state flavour, the $N\,(\text{jet}) = 4$ and the $N\,(\text{jet}) \geq 5$ constitute two separate categories that are combined in the fit. This improves the sensitivity compared to the single category with $N\,(\text{jet}) \geq 4$. The best value of μ from the combination of the $2\ell 0\tau_{\text{had}}$ categories is $2.8^{+2.1}_{-1.9}$. The largest source of error is the statistical uncertainty, which is about 1.5 in the combination. The $e^{\pm}\mu^{\pm}$ category is the most sensitive because of the largest signal fraction and statistics.

The best fit values for μ in the other signal regions categories and the best value for the combined fit of all the categories signal regions are shown in Figure 6.29. The μ values are all compatible with the Standard Model hypothesis. The biggest discrepancy between the predicted μ and the Standard Model value is in the $1\ell 2\tau_{\text{had}}$ category and it is 1.1 times the error. This category is the less sensitive to the $t\bar{t}H$ production and the negative value of μ is due to the predicted background being larger than the observed events. In the 4ℓ category, the total expected yields would assume a non-physical negative value for $\mu < -0.17$, therefore that range is excluded. The $2\ell 0\tau_{\text{had}}$ and the 3ℓ categories have the best sensitivity. They both measure a signal strength larger than the Standard Model one but compatible within uncertainties. The value of μ from the combined fit is $2.1^{+1.4}_{-1.2}$.

In the $2\ell 0\tau_{\text{had}}$ channel, the systematic uncertainties and the statistical uncertainty contribute similarly to the total uncertainty on μ, while in the other categories the statistical uncertainty is dominant due to the lower observed and expected yields. The dominant

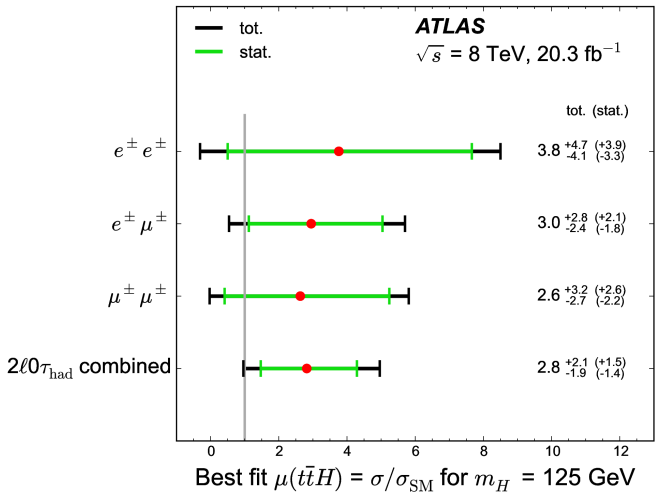


Figure 6.28: Best fit values of the signal strength parameter $\mu = \sigma_{t\bar{t}H,\mathrm{obs}}/\sigma_{t\bar{t}H,\mathrm{SM}}$ for the $2\ell 0\tau_{\mathrm{had}}$ categories, where the $= 4$ jet and ≥ 5 jet categories are combined for each of the $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$, and $\mu^{\pm}\mu^{\pm}$ lepton final states.

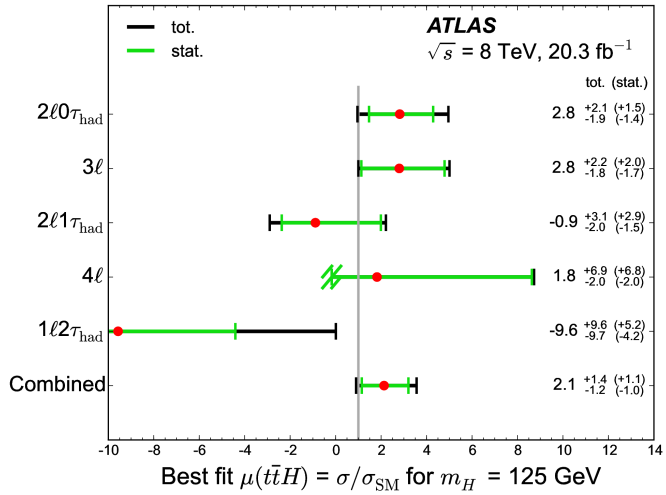


Figure 6.29: Best-fit values of the signal strength parameter $\mu = \sigma_{t\bar{t}H,\mathrm{obs}}/\sigma_{t\bar{t}H,\mathrm{SM}}$. For the 4ℓ Z-depleted category, $\mu < -0.17$ results in a negative expected total yield and so the lower uncertainty is truncated at this point.

systematic uncertainties in the combined fit are listed in Table 6.23. The largest uncertainties are due to the θ_ℓ used in the $2\ell 0\tau_{\text{had}}$ non-prompt background estimate and the theoretical uncertainties.

Source	$\Delta\mu$	
θ_μ in the $2\ell 0\tau_{\text{had}}$ category	$+0.38$	-0.35
$t\bar{t}W^\pm$ acceptance	$+0.26$	-0.21
$t\bar{t}H$ inclusive cross section	$+0.28$	-0.15
Jet energy calibration	$+0.24$	-0.18
θ_e in the $2\ell 0\tau_{\text{had}}$ category	$+0.26$	-0.16
$t\bar{t}H$ acceptance	$+0.22$	-0.15
$t\bar{t}Z$ inclusive cross section	$+0.19$	-0.17
$t\bar{t}W^\pm$ inclusive cross section	$+0.18$	-0.15
Muon isolation efficiency	$+0.19$	-0.14
Luminosity	$+0.18$	-0.14

Table 6.23: Contributions to the signal strenth uncertainty from the main sources of systematic uncertainty.

The maximum likelihood fit not only determines the optimal μ value but also the optimal value for the Nuisance Parameters. Sometimes the data in the signal regions can constrain significantly the error on the Nuisance parameters, especially if the accessory measurement of the parameter has large uncertainty. The post-fit values of the Nuisance parameters are all compatible with their pre-fit values within the errors. The combined fit constrains significantly the NPs related to the θ factors of the non-prompt background estimate in the $2\ell 0\tau_{\text{had}}$ categories and to the non-prompt background normalization in the $1\ell 2\tau_{\text{had}}$ category[11]. The fitted values for these Nuisance Parameters are also those that differ more from the pre-fit value. The difference is -1.0 times the error for the NP associated to the non-prompt background in the $1\ell 2\tau_{\text{had}}$ category and $+0.4$ times the error for the NP associated to the θ_ℓ in the $2\ell 0\tau_{\text{had}}$ category.

The compatibility of the signal-plus-background ($\mu = 1$) and the background-only ($\mu = 0$) hypotheses with the observed data are quantified by the test statistics based on the profile likelihood ratio method (see Section 5.6.2.2). The background only hypothesis has an observed p-value of 0.037 and an expected one of 0.18, corresponding to a significance expressed in terms of Gaussian standard deviations of 1.8σ and 0.9σ respectively. The observations are compatible with the $\mu = 1$ hypothesis with a p-value of 0.18 corresponding to 0.9σ. These results reflect the excess of data compared to the predictions which promotes a $\mu > 1$.

The 95% CL_S upper limits on the signal strength are also determined using the CL_S method [172, 173] introduced in Section 5.6.3. The observed and expected upper limits in the $2\ell 0\tau_{\text{had}}$ category and in the other categories are listed in Table 6.24 and shown

[11] In the $1\ell 2\tau_{\text{had}}$ category, the non-prompt background estimate is not obtained with the θ Method.

in Figure 6.30. The expected limits are calculated for background only and background plus Standard Model signal hypotheses. The stronger limits are set in the $2\ell 0\tau_{\text{had}}$ and 3ℓ channels. The observed combined limit is $\mu < 4.7$ and the expected for background only is $\mu < 2.4$.

	Observed	Expected Limit					
Channel	Limit	-2σ	-1σ	Median ($\mu = 0$)	$+1\sigma$	$+2\sigma$	Median ($\mu = 1$)
$2\ell 0\tau_{\text{had}}$	6.7	2.1	2.8	3.9	5.7	8.4	5.0
3ℓ	6.8	2.0	2.7	3.8	5.7	8.5	5.1
$2\ell 1\tau_{\text{had}}$	7.5	4.5	6.1	8.4	13	21	10
4ℓ	18	8.0	11	15	23	39	17
$1\ell 2\tau_{\text{had}}$	13	10	13	18	26	40	19
Combined	4.7	1.3	1.8	2.4	3.6	5.3	3.7

Table 6.24: Observed and expected 95% CL_S upper limits on μ for various categories and their combination. For the background only hypothesis, the median, the 1σ and the 2σ error bands are indicated. The median for the Standard Model signal hypothesis is presented in the last column.

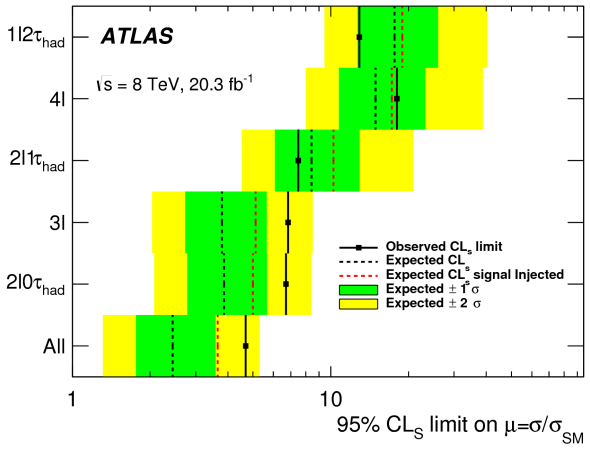
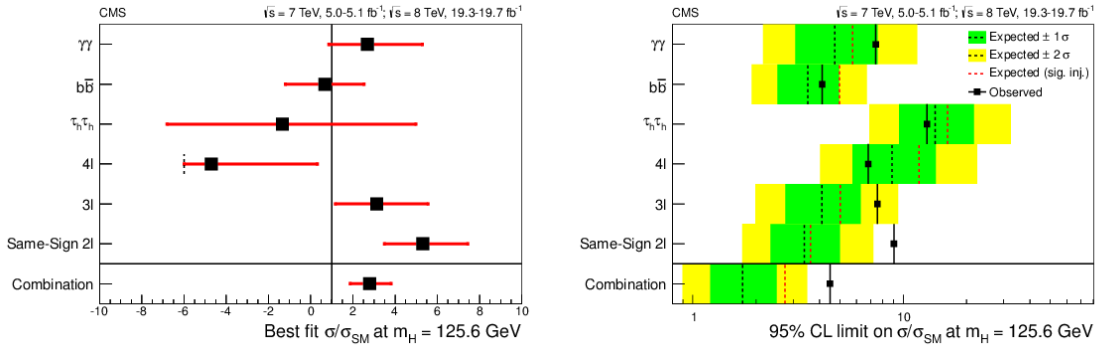


Figure 6.30: Observed and expected 95% CL_S upper limits on μ for various categories and their combination. The black dashed line and the red dashed lines show the expected limits for background only hypothesis and Standard Model signal plus background, respectively. The black line indicates the observed limit.

Excluding the analysis presented here, the most accurate measurements of $t\bar{t}H$ production performed so far are by the CMS collaboration on 8 TeV data [30], and very recently by ATLAS in 13 TeV data [33]. The measured values by CMS of the signal strength and the 95% CL_S upper limits are shown in Figure 6.31. The μ value from the $2\ell 0\tau_{\text{had}}$ category fit is $5.3^{+2.1}_{-1.8}$ and the observed 95% CL_S upper limits is 9.0. The sensitivity of this analysis and the CMS one to $t\bar{t}H$ production is very similar and both observe a signal strength greater than 1 in the $2\ell 0\tau_{\text{had}}$ category. The CMS analysis however measures a larger μ with an observed limit that exceeds the expected limit for background only hypothesis by

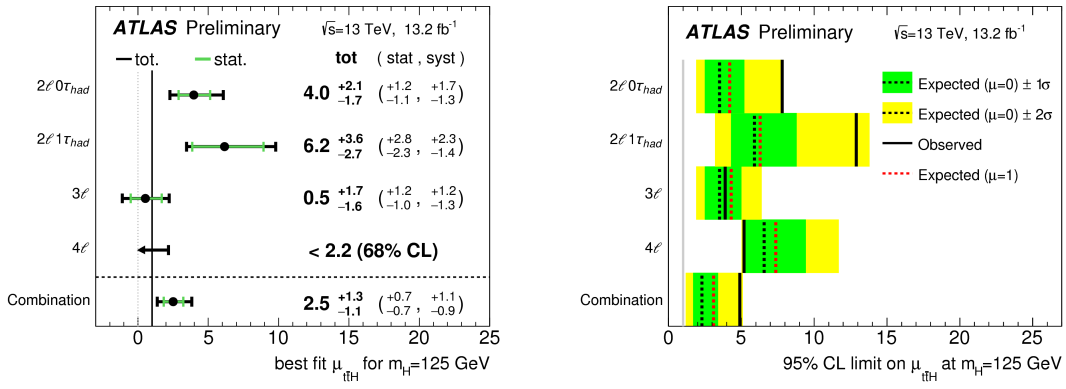


(a) (b)

Figure 6.31: The signal strenght measured by the CMS collaboration for several categories (left), and the respective 95% CL_S upper limits (right).

more than 2σ. The excess is driven by the $\mu^{\pm}\mu^{\pm}$ category where $\mu = 8.5^{+3.3}_{-2.7}$. The signal strength measured by CMS in the $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ categories are, respectively, $2.7^{+4.6}_{-4.1}$ and $1.8^{+2.5}_{-2.3}$, in agreement with the results obtained by ATLAS.

A preliminary measurement of the $t\bar{t}H$ signal strength on a 13.2 fb^{-1} sample of 13 TeV proton-proton collisions has been performed by ATLAS with the same analysis strategy described here. The measured values of the signal strength and the 95% CL_S upper limits are



(a) (b)

Figure 6.32: The signal strenght measured by the ATLAS collaboration at $\sqrt{s} = 13$ TeV for several categories (left), and the respective 95% CL_S upper limits (right).

shown in Figure 6.32. In the $2\ell 0\tau_{\text{had}}$ category, the observed μ is $4.0^{+1.2}_{-1.1}$ (stat.)$^{+1.7}_{-1.3}$ (syst.) and the combination with the other categories gives a $\mu = 2.5 \pm 0.7$ (stat.)$^{+1.1}_{-0.9}$ (syst.). The results obtained in the new data are compatible with what found in this thesis and have similar precision. It is to note that the observed μ is again larger than 1 despite the use of an independent data sample. The improvement in precision from the collection of a larger dataset at 13 TeV and the use of a more precise determination of the backgrounds are

expected to clarify if there is an anomalous event production in these regions or if what has been observed is just a statistical fluctuation.

The ATLAS and CMS collaborations have investigated the mechanisms of production and the decay modes of the Higgs boson with numerous analyses of $\sqrt{s} = 7$ and 8 TeV data, including the search of $t\bar{t}H$ production in the final states with leptons. The results of these analyses have been combined to achieve the most accurate measurements of the production and decay modes of the Higgs boson and lastly of the coupling of the Higgs boson to Standard Model particles. ATLAS and CMS have published independent results [188, 189] and have performed a combination of their measurements [34]. The results of this combination are obtained in the context of the κ-framework [14] and are described in Section C.8. The measurements of the $t\bar{t}H$ signal strength (μ), of the relative cross-section of $t\bar{t}H$ and gluon-gluon fusion productions ($\sigma_{t\bar{t}H}/\sigma_{ggF}$) and of the Top Yukawa coupling modifier (κ_t) have been relevantly influenced by the results obtained in this thesis. A signal strength for $t\bar{t}H$ production of $2.3^{+0.7}_{-0.6}$ has been measured from the combination. The observed relative cross-section $\sigma_{t\bar{t}H}/\sigma_{ggF}$ value is $0.022^{+0.007}_{-0.006}$, which is more than 2σ higher than the expected value of 0.0067 ± 0.0010. The ATLAS only combination is 1σ higher than the expected value and has a smaller uncertainty than the CMS one.

The observations can also be reinterpreted in terms of coupling strength of the Higgs boson to Standard Model particles, but assumptions on the Higgs boson boson interaction vertices have to be done. In the Standard Model scenario, different production processes and decay modes are sensitive to specific coupling constants as shown in Table 1.2. The $t\bar{t}H$ production is sensitive to the Top Yukawa coupling, but the gluon-gluon fusion production and the $H \to \gamma\gamma$ decay are even more sensitive in this condition. If non-Standard Model particles are introduced, the $t\bar{t}H$ production would still probe directly the Top Yukawa coupling, but the gluon-gluon fusion production and the $H \to \gamma\gamma$ decay would have their coupling modified by the interactions of the non-Standard Model particles with the Higgs boson field.

The Top Yukawa coupling strength is measured in units of the expected values for a Standard Model Higgs boson with 125 GeV mass (coupling modifiers) under three hypotheses: no interaction vertices of the Higgs boson with beyond Standard Model (BSM) particles; the Higgs boson cannot decay to BSM particles ($BR(BSM) = 0$); the coupling modifiers for vector bosons κ_W and κ_Z must be at most 1 ($\kappa_V \leq 1$). The measured κ_t under these assumption are given in Table 6.25. The large κ_t observed in the BSM scenarios is a consequence of the $t\bar{t}H$ signal strength measurement, while in the Standard Model case the κ_t is heavily constrained by the gluon-gluon fusion and $H \to \gamma\gamma$ measurements. The choice of the condition for the BSM scenario, either $BR(BSM) = 0$ or $\kappa_V \leq 1$, has negligible impact on the κ_t measurement.

In a typical new physics scenario, one would observe a κ_t value compatible with one, and modifiers associated to the gluon-gluon fusion production and $H \to \gamma\gamma$ decay diagrams (κ_g and κ_γ) significantly different. The current findings instead indicates a $\kappa_t > 1$ and

Hypothesis	Combination Measured κ_t	Combination Expected κ_t uncertainty	ATLAS Measured κ_t	CMS Measured κ_t
$BR(BSM) = 0$	$1.43^{+0.23}_{-0.22}$	$^{+0.27}_{-0.32}$	$1.31^{+0.35}_{-0.33}$	$1.45^{+0.42}_{-0.38}$
$\kappa_V \leq 1$	$1.40^{+0.24}_{-0.21}$	$^{+0.26}_{-0.39}$	$1.32^{+0.31}_{-0.33}$	$1.51^{+0.33}_{-0.32}$
no BSM particles	$0.87^{+0.15}_{-0.15}$	$^{+0.15}_{-0.18}$	$0.98^{+0.21}_{-0.20}$	$0.77^{+0.20}_{-0.18}$

Table 6.25: The measurement of the modifier κ_t of the Top Yukawa coupling for several parametrisations using ATLAS and CMS data. The first two parametrisations allow for non-Standard Model particles in the loop couplings but have different hypotheses: in the former, the Higgs boson cannot decay to BSM particles, in the latter, BSM particles can contribute to the Higgs boson width but $\kappa_V \leq 1$ with $V = W, Z$.

nearly unit value for κ_g and κ_γ. If this scenario is confirmed, the Top Yukawa coupling for the discovered particle at 125 GeV is not that expected for a Standard Model Higgs boson and therefore the particle itself is not the Higgs boson. In addition, there should be BSM particles which contribute to the gluon-gluon fusion production and to the $H \to \gamma\gamma$ decay to compensate for the non-one κ_t value. This condition seems highly unlikely but cannot be excluded a priori. An anomaly in $t\bar{t}H$ production measurement and so in the value of κ_t could be also the consequence of unknown processes that match the $t\bar{t}H$ production final state. For instance, BSM particles may be produced in association with top quark pairs, and result in a large multitude of jets and leptons (photons and b-jets for analyses considering other $t\bar{t}H$ decay modes). These processes would constitute a background for this analysis which has not been taken into account resulting in an inflation of the measured $t\bar{t}H$ signal strength.

At this stage, the observed κ_t deviation is not significant enough to constitute an evidence of new physics but it is something that should be further investigated with new data.

Conclusions

This thesis has explored the physics of same-sign lepton pair production at the LHC analysing the data collected by the ATLAS experiment. The targets of this study have been the $t\bar{t}H$ and non Standard Model productions which lead to final states with same-sign lepton pairs. The calibration of the τ_{had} energy in data via the study of $Z \to \tau^+\tau^-$ events has also been described. Although this study does not constitute the core of this thesis, it has important outcomes as improves the τ_{had} energy modeling in analyses studying processes with τ leptons. One of these analyses is the search for $t\bar{t}H$ production, where one or more τ_{had} are required is some categories ($2\ell 1\tau_{\text{had}}$ and $1\ell 2\tau_{\text{had}}$) and are prohibited in others ($2\ell 0\tau_{\text{had}}$). If there is a difference in the τ_{had} energy calibration for real and simulated τ_{had}, the use of a τ_{had} p_{T} requirement in the category definitions leads to differences in the way real and simulated events would fill these categories. In the $t\bar{t}H$ analysis, where the signal strength is extracted from comparisons of expected and observed events in several categories, an accurate τ_{had} energy calibration results fundamental. The calibration study has found a correction to the energy calibration in data of 0.8% $\pm$ 1.3% (stat.) $\pm$ 0.6% (syst.) for 1-track τ_{had} and 1.1% $\pm$ 1.4% (stat.) $\pm$ 0.7% (syst.) for 3-tracks τ_{had}.

In the main part of the thesis, the production of same-sign lepton pairs at the LHC has been studied. The Standard Model contribution to final states of the $t\bar{t}H$ and new physics searches has been characterised. The prompt, charge misID, photon conversion and non-prompt backgrounds have been estimated with dedicated techniques, which have been validated with data. The non-prompt background estimate carries one of the largest uncertainties in both analyses. For this background, alternative methods of estimate have been explored. Provided enough statistics, the Matrix Method and Fake Factor Method could improve the modeling of the non-prompt background compared to the θ Method. However, the study has not been able to determine whether the Matrix Method or the Fake Factor Method would perform better in future analyses.

The search for new physics in same-sign lepton pairs has not observed significant production over the Standard Model prediction. Data has been used to set 95% CL_S upper limits on the fiducial cross-section of new physics processes producing same-sign lepton pairs. Limits are placed as a function of the di-lepton invariant mass within a fiducial region which is defined according to the event selection criteria. The observed limits vary between 0.48 fb and 32 fb depending on the di-lepton invariant mass and

lepton flavour combination. These limits have general validity and can be used to test new physics models upon calculation of the fiducial region acceptance. 95% CL_S limits on the production cross-section of doubly charged bosons in the context of the Left-right symmetric model and on their mass are also obtained. The observed lower limits on the $H_{\mathrm{L}}^{\pm\pm}$ mass vary between 468 and 551 GeV depending on the lepton flavour category and between 374 and 438 GeV for the $H_{\mathrm{R}}^{\pm\pm}$ mass. These limits improve by 30-40% previous results.

The analysis of $t\bar{t}H$ production in the $2\ell 0\tau_{\mathrm{had}}$ category has led to the measurement of a $t\bar{t}H$ signal strength $\mu = 2.8^{+2.1}_{-1.9}$. The combination of the results from this and other categories gives a value for the signal strength of $2.1^{+1.4}_{-1.2}$. The observed data are compatible with the signal-plus-background (background-only) hypothesis with a p-value of 0.18 (0.037). The observed excess is therefore not statistically significant, even though other measurements by ATLAS and CMS confirm a signal strength greater than one. These observations may constitute hints of new physics, such as an anomalous Top Yukawa coupling or non-Standard Model processes contributing to the yields in the $2\ell 0\tau_{\mathrm{had}}$ signal region, which seems to have largest excess. They could also be due to statistical fluctuations.

The collision of protons at a centre-of-mass energy of 13 TeV has started at the LHC in 2016 and may conceal interesting physics. The observation of $t\bar{t}H$ production is one of the main goals of the ATLAS collaboration, but the study of final states with same-sign leptons may also reveal new physics connected with the mechanisms of neutrino mass generation or anomalies in the coupling of the top quark with the Higgs boson.

Bibliography

[1] ATLAS Collaboration, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at* $\sqrt{s} = 8$ *TeV*, Eur. Phys. J. C **75** (2015) 303, arXiv:1412.7086 [hep-ph].

[2] ATLAS Collaboration, *Search for anomalous production of prompt same-sign lepton pairs and pair-produced doubly charged Higgs bosons with* $\sqrt{s} = 8$ *TeV pp collisions using the ATLAS detector*, JHEP **03** (2015) 041, arXiv:1412.0237 [hep-ex].

[3] ATLAS Collaboration, *Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector*, Phys. Lett. B **749** (2015) 519, arXiv:1506.05988 [hep-ex].

[4] S. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.

[5] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[6] A. Salam, *Weak and electromagnetic interactions in Elementary particle theory: relativistic groups and analyticity.* N. Svartholm, ed., p. 367. Almqvist & Wiksell, Stockholm, 1968. Proceedings of the eighth Nobel symposium.

[7] J. Goldstone, *Field Theories with Superconductor Solutions*, Nuovo Cim. **19** (1961) 154.

[8] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13** (1964) 321.

[9] P. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Rev. Lett. **12** (1964) 132.

[10] P. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) 508.

[11] G. Guralnik et al., *Global conservation laws and massless particles*, Phys. Rev. Lett. **13** (1964) 585.

[12] Y. Nambu and G. Jona-Lasinio, *Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I*, Phys. Rev. **122** (1961) 345. http://link.aps.org/doi/10.1103/PhysRev.122.345.

[13] M. Gell-Mann and M. Lévy, *The axial vector current in beta decay*, Nuovo Cim. **16** no. 4, (1960) 705. http://dx.doi.org/10.1007/BF02859738.

[14] LHC Higgs Cross Section Working Group Collaboration, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, Tech. Rep. CERN-2013-004, CERN, Geneva, 2013. arXiv:1307.1347 [hep-ph].

[15] LHC Higgs Cross Section Working Group Collaboration, *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, Tech. Rep. CERN-2011-002, CERN, Geneva, 2011. arXiv:1101.0593 [hep-ph].

[16] Tevatron Electroweak Working Group, CDF, DELPHI, SLD Electroweak and Heavy Flavour Groups, ALEPH, LEP Electroweak Working Group, SLD, OPAL, D0, L3 Collaboration, *Precision Electroweak Measurements and Constraints on the Standard Model*, Tech. Rep. CERN-PH-EP-2010-095, 2010. arXiv:1012.2367 [hep-ex].

[17] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.164. http://cds.cern.ch/record/1129806.

[18] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[19] P. Teixeira-Dias, *Higgs boson searches at LEP*, J. Phys. Conf. Ser. **110** (2008) 042030, arXiv:0804.4146 [hep-ex].

[20] CDF and D0 Collaboration, *Updated Combination of CDF and D0 Searches for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data*, Tech. Rep. FERMILAB-CONF-12-318-E, Fermilab, Chicago, 2012. arXiv:1207.0449 [hep-ex].

[21] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[22] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv:1207.7214 [hep-ex].

[23] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv:1207.7235 [hep-ex].

[24] ATLAS Collaboration, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, Phys. Lett. B **726** (2013) 120, arXiv:1307.1432 [hep-ex].

[25] CMS Collaboration, *Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV*, Phys. Rev. D **92** (2015) 012004, arXiv:1411.3441 [hep-ex].

[26] ATLAS and CMS Collaboration, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, arXiv:1503.07589 [hep-ex].

[27] ATLAS Collaboration, *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector*, JHEP **1504** (2015) 117, arXiv:1501.04943 [hep-ex].

[28] ATLAS Collaboration, *Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector*, JHEP **01** (2015) 069, arXiv:1409.6212 [hep-ex].

[29] CMS Collaboration, *Evidence for the direct decay of the 125 GeV Higgs boson to fermions*, Nature Phys. **10** (2014) 557, arXiv:1401.6527 [hep-ex].

[30] CMS Collaboration, *Search for the associated production of the Higgs boson with a top-quark pair*, JHEP **1409** (2014) 087, (erratum *ibid* **1410** (2014), 106), arXiv:1408.1682 [hep-ex].

[31] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Eur. Phys. J. C **75** (2015) 349, arXiv:1503.05066 [hep-ex].

[32] ATLAS Collaboration, *Search for $H \to \gamma\gamma$ produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector*, Phys. Lett. B **740** (2015) 222, arXiv:1409.3122 [hep-ex].

[33] ATLAS Collaboration, *Search for the Associated Production of a Higgs Boson and a Top Quark Pair in Multilepton Final States with the ATLAS Detector*, Tech. Rep. ATLAS-CONF-2016-058, CERN, Geneva, Aug, 2016. http://cds.cern.ch/record/2206153.

[34] ATLAS and CMS Collaboration, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV*, JHEP **08** (2016) 045, arXiv:1606.02266 [hep-ex].

[35] O. Eberhardt et al., *Impact of a Higgs boson at a mass of 126 GeV on the standard model with three and four fermion generations*, Phys. Rev. Lett. **109** (2012) 241802, arXiv:1209.1101 [hep-ph].

[36] S. P. Martin, *A Supersymmetry primer*, Adv. Ser. Direct. High Energy Phys. **18** (1997) 1, arXiv:hep-ph/9709356 [hep-ph].

[37] J. A. Aguilar-Saavedra et al., *Handbook of vectorlike quarks: Mixing and single production*, Phys. Rev. D **88** no. 9, (2013) 094010, arXiv:1306.0572 [hep-ph].

[38] M. Carena et al., *Light Stops, Light Staus and the 125 GeV Higgs*, JHEP **08** (2013) 087, arXiv:1303.4414 [hep-ph].

[39] N. Arkani-Hamed et al., *2:1 for Naturalness at the LHC?*, JHEP **01** (2013) 149, arXiv:1207.4482 [hep-ph].

[40] D. Carmi et al., *Interpreting LHC Higgs Results from Natural New Physics Perspective*, JHEP **07** (2012) 136, arXiv:1202.3144 [hep-ph].

[41] N. Arkani-Hamed et al., *The Hierarchy problem and new dimensions at a millimeter*, Phys. Lett. B **429** (1998) 263, arXiv:hep-ph/9803315 [hep-ph].

[42] A. D. Sakharov, *Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32. [Usp. Fiz. Nauk161,61(1991)].

[43] L. Canetti et al., *Matter and Antimatter in the Universe*, New J. Phys. **14** (2012) 095012, arXiv:1204.4186 [hep-ph].

[44] R. Barnett et al., *Discovering supersymmetry with like sign dileptons*, Phys. Lett. B **315** (1993) 349, arXiv:hep-ph/9306204 [hep-ph].

[45] ATLAS Collaboration, *Search for microscopic black holes in a like-sign dimuon final state using large track multiplicity with the ATLAS detector*, Phys. Rev. D **88** no. 7, (2013) 072001, arXiv:1308.4075 [hep-ex].

[46] J. Pati and A. Salam, *Lepton number as the fourth "color"*, Phys. Rev. D **10** (1974) 275, (erratum *ibid* **11** (1975), 703).

[47] R. Mohapatra and J. Pati, *Left-right gauge symmetry and an "isoconjugate" model of CP violation*, Phys. Rev. D **11** (1975) 566.

[48] G. Senjanovic and R. Mohapatra, *Exact left-right symmetry and spontaneous violation of parity*, Phys. Rev. D **12** (1975) 1502.

[49] T. Rizzo, *Doubly charged Higgs bosons and lepton number violating processes*, Phys. Rev. D **25** (1982) 1355, (addendum *ibid* **27** (1983), 657).

[50] A. Hektor et al., *Testing neutrino masses in little Higgs models via discovery of doubly charged Higgs at LHC*, Nucl. Phys. B **787** (2007) 198, arXiv:0705.1495 [hep-ph].

[51] A. Belyaev et al., *Phenomenology of littlest Higgs model with T^- parity: including effects of T^- odd fermions*, Phys. Rev. D **74** (2006) 115020, arXiv:hep-ph/0609179 [hep-ph].

[52] S. Bar-Shalom et al., *Two Higgs doublets, a 4th generation and a 125 GeV Higgs: a review*, Adv. High Energy Phys. **2013** (2013) 672972, arXiv:1208.3195 [hep-ph].

[53] B. Holdom et al., *Four Statements about the Fourth Generation*, PMC Phys. A **3** (2009) 4, arXiv:0904.4698 [hep-ph].

[54] A. Alok et al., *Constraints on the Four-Generation Quark Mixing Matrix from a Fit to Flavor-Physics Data*, Phys. Rev. D **83** (2011) 073008, arXiv:1011.2634 [hep-ph].

[55] S. Dimopoulos and G. L. Landsberg, *Black holes at the LHC*, Phys. Rev. Lett. **87** (2001) 161602, arXiv:hep-ph/0106295 [hep-ph].

[56] R. N. Mohapatra and J. C. Pati, *A Natural Left-Right Symmetry*, Phys. Rev. D **11** (1975) 2558.

[57] G. Senjanovic, *Spontaneous Breakdown of Parity in a Class of Gauge Theories*, Nucl. Phys. B **153** (1979) 334.

[58] N. G. Deshpande et al., *Left-right symmetric electroweak models with triplet Higgs*, Phys. Rev. D **44** (1991) 837.

[59] M. Schmaltz and D. Tucker-Smith, *Little Higgs review*, Ann. Rev. Nucl. Part. Sci. **55** (2005) 229, arXiv:hep-ph/0502182 [hep-ph].

[60] A. Djouadi and A. Lenz, *Sealing the fate of a fourth generation of fermions*, Phys. Lett. B **715** (2012) 310, arXiv:1204.1252 [hep-ph].

[61] Super-Kamiokande Collaboration, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562, arXiv:hep-ex/9807003 [hep-ex].

[62] SNO Collaboration, *Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87** (2001) 071301, arXiv:nucl-ex/0106015 [nucl-ex].

[63] A. Davidson and W. Kameshwar, *Family mass hierarchy from universal seesaw mechanism*, Phys. Rev. Lett. **60** (1988) 1813.

[64] J. Schechter and J. Valle, *Neutrino masses in SU(2) ⊗ U(1) theories*, Phys. Rev. D **22** (1980) 2227.

[65] M. Magg and C. Wetterich, *Neutrino mass problem and gauge hierarchy*, Phys. Lett. B **94** (1980) 61.

[66] T. Cheng and L. Li, *Neutrino masses, mixings, and oscillations in SU(2)ÃŮU(1) models of electroweak interactions*, Phys. Rev. D **22** (1980) 2860.

[67] G. Gelmini and M. Roncadelli, *Left-handed neutrino mass scale and spontaneously broken lepton number*, Phys. Lett. B **99** (1981) 411.

[68] G. Lazarides et al., *Proton lifetime and fermion masses in an SO(10) model*, Nucl. Phys. B **181** (1981) 287.

[69] M. Ernest, *Pathways to naturally small neutrino masses*, Phys. Rev. Lett. **81** (1998) 1171, arXiv:hep-ph/9805219 [hep-ph].

[70] R. Foot et al., *See-saw neutrino masses induced by a triplet of leptons*, Zeitschrift für Physik C Particles and Fields **44** no. 3, (1989) 441.

[71] A. Zee, *Quantum Numbers of Majorana Neutrino Masses*, Nucl. Phys. B **264** (1986) 99.

[72] K. Babu, *Model of 'Calculable' Majorana Neutrino Masses*, Phys. Lett. B **203** (1988) 132.

[73] M. Nebot, and others, *Prospects for the Zee-Babu model at the CERN LHC and low energy experiments*, Phys. Rev. D **77** (2008) 093013, arXiv:0711.0483 [hep-ph].

[74] M. Kohda, H. Sugiyama, and K. Tsumura, *Lepton number violation at the LHC with leptoquark and diquark*, Phys. Lett. B **718** (2013) 1436, arXiv:1210.5622 [hep-ph].

[75] R. Battye and A. Moss, *Evidence for Massive Neutrinos from Cosmic Microwave Background and Lensing Observations*, Phys. Rev. Lett. **112** no. 5, (2014) 051303, arXiv:1308.5870 [astro-ph.CO].

[76] A. Goobar et al., *A new bound on the neutrino mass from the sdss baryon acoustic peak*, JCAP **0606** (2006) 019, arXiv:astro-ph/0602155 [astro-ph].

[77] KamLAND-Zen Collaboration, *Search for Majorana Neutrinos near the Inverted Mass Hierarchy Region with KamLAND-Zen*, Phys. Rev. Lett. **117** no. 8, (2016) 082503, arXiv:1605.02889 [hep-ex].

[78] H. Georgi and S. L. Glashow, *Unity of All Elementary-Particle Forces*, Phys. Rev. Lett. **32** (1974) 438. http://link.aps.org/doi/10.1103/PhysRevLett.32.438.

[79] Z. Xing, *Tev Neutrino Physics at the Large Hadron Collider*, Int. J. Mod. Phys. A **24** (2009) 3286, arXiv:0901.0209 [hep-ph].

[80] T. Han and B. Zhang, *Signatures for Majorana Neutrinos at Hadron Colliders*, Phys. Rev. Lett. **97** (2006) 171804. http://link.aps.org/doi/10.1103/PhysRevLett.97.171804.

[81] F. del Aguila et al., *Heavy neutrino signals at large hadron colliders*, JHEP **10** (2007) 047, arXiv:hep-ph/0703261 [hep-ph].

[82] R. Franceschini, T. Hambye, and A. Strumia, *Type-III see-saw at LHC*, Phys. Rev. **D78** (2008) 033002, arXiv:0805.1613 [hep-ph].

[83] P. Perez et al., *Testing a Neutrino Mass Generation Mechanism at the LHC*, Phys. Rev. D **78** (2008) 071301, arXiv:0803.3450 [hep-ph].

[84] K. S. Babu and C. N. Leung, *Classification of effective neutrino mass operators*, Nucl. Phys. B **619** (2001) 667, arXiv:hep-ph/0106054 [hep-ph].

[85] P. W. Angel et al., *Origin of neutrino masses at the LHC: $\Delta L = 2$ effective operators and their ultraviolet completions*, Phys. Rev. D **87** no. 7, (2013) 073007, arXiv:1212.6111 [hep-ph].

[86] The LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** no. 08, (2008) S08005.

[87] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[88] J. Caron, *The LHC injection complex. L'ensemble d'injection du LHC*, Tech. Rep. LHC-PHO-1993-008, CERN, Geneva, May, 1993. https://cds.cern.ch/record/841568.

[89] ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **73** (2013) 2518, arXiv:1302.4393 [hep-ex].

[90] ATLAS Collaboration, *Minimum Bias and Underlying Event Measurements with ATLAS*, MPI@LHC 2011 no. ATL-PHYS-PROC-2012-033, (2012) 11, arXiv:1202.2090 [hep-ex].

[91] ATLAS Collaboration, *Muons in the Calorimeters: Energy Loss Corrections and Muon Tagging*, Tech. Rep. ATL-PHYS-PUB-2009-009, CERN, Geneva, Mar, 2009. http://cds.cern.ch/record/1169055.

[92] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data*, Eur. Phys. J. C **74** (2014) 3130, arXiv:1407.3935 [hep-ex].

[93] ATLAS Collaboration, *Performance of the ATLAS muon trigger in pp collisions at* $\sqrt{s} = 8$ *TeV*, Eur. Phys. J. C **75** (2015) 120, arXiv:1408.3179 [hep-ex].

[94] J. Pumplin et al., *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*, JHEP **0207** (2012) 012, arXiv:hep-ph/0201195 [hep-ph].

[95] P. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, Phys. Rev. D **78** (2008) 013004, arXiv:0802.0007 [hep-ph].

[96] H. Lai et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, arXiv:1007.2241 [hep-ph].

[97] A. Sherstnev and R. Thorne, *Parton Distributions for LO Generators*, Eur. Phys. J. C **55** (2008) 553, arXiv:0711.2473 [hep-ph].

[98] M. Mangano et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **0307** (2003) 001, arXiv:hep-ph/0206293 [hep-ph].

[99] T. Binoth et al., *Gluon-induced QCD corrections to* $pp \to ZZ \to l\bar{l}l'\bar{l}'$, arXiv:0807.0024 [hep-ph].

[100] T. Binoth et al., *Gluon-induced W-boson pair production at the LHC*, JHEP **12** (2006) 046, arXiv:hep-ph/0611170 [hep-ph].

[101] S. Frixione, and B. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP **06** (2002) 029, arXiv:hep-ph/0204244 [hep-ph].

[102] G. Corcella et al., *HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)*, JHEP **0101** (2001) 010, arXiv:hep-ph/0011363 [hep-ph].

[103] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **1407** (2014) 079, arXiv:1405.0301 [hep-ph].

[104] J. Alwall et al., *MadGraph 5 : Going Beyond*, JHEP **1106** (2011) 128, arXiv:1106.0522 [hep-ex].

[105] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **0411** (2004) 040, arXiv:hep-ph/0409146 [hep-ph].

[106] S. Frixione et al., *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **0711** (2007) 070, arXiv:0709.2092 [hep-ph].

[107] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **1006** (2010) 43, arXiv:1002.2581 [hep-ph].

[108] S. Frixione et al., *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **0709** (2009) 126, arXiv:0707.3088 [hep-ph].

[109] T. Melia et al., *W^+W^-, WZ and ZZ production in the POWHEG BOX*, JHEP **1111** (2011) 078, arXiv:1107.5051 [hep-ph].

[110] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. C **71** (2011) 1547, arXiv:1009.2450 [hep-ph].

[111] S. Alioli et al., *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP **0909** (2009) 111, (erratum *ibid* **1002** (2010), 011), arXiv:0907.4076 [hep-ph].

[112] B. Kersevan and E. Richter-Was, *The Monte Carlo Event Generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1*, arXiv:hep-ph/0405247 [hep-ph].

[113] G. Bevilacqua et al., *HELAC-NLO*, Comput. Phys. Commun. **184** (2013) 986, arXiv:1110.1499 [hep-ph].

[114] M. Garzelli et al., *Standard Model Higgs boson production in association with a top anti-top pair at NLO with parton showering*, Europhys. Lett. **96** (2011) 11001, arXiv:1108.0387 [hep-ph].

[115] T. Sjöstrand et al., *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820 [hep-ph"].

[116] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **0902** (2009) 007, arXiv:0811.4622 [hep-ph].

[117] J. Stanisław et al., *TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons*, Comput. Phys. Commun. **64** (1991) 275.

[118] P. Golonka and Z. Wąs, *PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays*, Eur. Phys. J. C **45** (2006) 97, arXiv:hep-ph/0506026 [hep-ph].

[119] T. Sjöstrand et al., *High-energy-physics event generation with Pythia 6.1*, Comput. Phys. Commun. **135** (2001) 238, arXiv:hep-ph/0010017 [hep-ph].

[120] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639, arXiv:0803.0883 [hep-ph].

[121] J. Butterworth et al., *Multiparton interactions in photoproduction at HERA*, Z. Phys. C **72** (1996) 637, arXiv:9601371 [hep-ph].

[122] P. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. D **82** (2010) 074018, arXiv:1005.3457 [hep-ph].

[123] ATLAS Collaboration, *New ATLAS event generator tunes to 2010 data*, Tech. Rep. ATL-PHYS-PUB-2011-008, CERN, Geneva, Apr, 2011. http://cds.cern.ch/record/1345343.

[124] ATLAS Collaboration, *ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11*, Tech. Rep. ATL-PHYS-PUB-2011-009, CERN, Geneva, Jul, 2011. http://cds.cern.ch/record/1363300.

[125] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, Tech. Rep. ATL-PHYS-PUB-2012-003, CERN, Geneva, Aug, 2012. https://cds.cern.ch/record/1474107.

[126] S. Gieseke, and others, *Colour reconnections in Herwig++*, Eur. Phys. J. C **72** (2012) 2225, arXiv:1206.0041 [hep-ph].

[127] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv:1005.4568 [physics.ins-det].

[128] S. Agostinelli et al., *Geant4: a simulation toolkit*, NIM **A 506** (2003) 250.

[129] G. Folger and J. Wellisch, *String parton models in GEANT4*, eConf **C0303241** (2003) MOMT007, arXiv:nucl-th/0306007 [nucl-th].

[130] ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATL-PHYS-PUB-2010-013. http://cds.cern.ch/record/1300517.

[131] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data*, Eur. Phys. J. C **74** (2014) 2941, arXiv:1404.2240 [hep-ex].

[132] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data*, Tech. Rep. ATLAS-CONF-2014-032, CERN, Geneva, Jun, 2014. http://cds.cern.ch/record/1706245.

[133] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, Tech. Rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr, 2008. https://cds.cern.ch/record/1099735.

[134] M. Wielers et al., *Single Electron Trigger Performance Plots*, Tech. Rep. ATL-COM-DAQ-2012-146, CERN, Geneva, Jun, 2012. https://cds.cern.ch/record/1456795.

[135] ATLAS Collaboration, *Performance of the ATLAS Electron and Photon Trigger in p-p Collisions at $\sqrt{s} = 7$ TeV in 2011*, Tech. Rep. ATLAS-CONF-2012-048, CERN, Geneva, May, 2012. https://cds.cern.ch/record/1450089.

[136] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, arXiv:1603.02934 [hep-ex].

[137] C. Cojocaru et al., *Hadronic calibration of the ATLAS liquid argon end-cap calorimeter in the pseudorapidity region $1.6 < |\eta| < 1.8$ in beam tests*, NIM **A 531** (2004) 481, arXiv:physics/0407009 [physics].

[138] T. Barillari et al., *Local Hadronic Calibration*, Tech. Rep. ATL-LARG-PUB-2009-001-2, CERN, Geneva, Jun, 2008. https://cds.cern.ch/record/1112035.

[139] M. Cacciari et al., *The anti-k_t jet clustering algorithm*, JHEP **0804** (2008) 063, arXiv:0802.1189 [hep-ph].

[140] M. Cacciari and G. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. B **641** (2006) 57, arXiv:hep-ph/0512210 [hep-ph].

[141] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. C **75** (2015) 17, arXiv:1406.0076 [hep-ex].

[142] ATLAS Collaboration, *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS*, Tech. Rep. ATLAS-CONF-2015-037, CERN, Geneva, Aug, 2015. https://cds.cern.ch/record/2044941.

[143] ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, Tech. Rep. ATLAS-CONF-2014-046, CERN, Geneva, Jul, 2014. https://cds.cern.ch/record/1741020.

[144] ATLAS Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*, Tech. Rep. ATLAS-CONF-2011-102, CERN, Geneva, Jul, 2011. https://cds.cern.ch/record/1369219.

[145] ATLAS Collaboration, *Calibration of ATLAS b-tagging algorithms in dense jet environments*, Tech. Rep. ATLAS-CONF-2016-001, CERN, Geneva, Feb, 2016. https://cds.cern.ch/record/2127958.

[146] Particle Data Group Collaboration, *Review of Particle Physics*, Chin. Phys. C **38** (2014) 090001.

[147] L. Breiman et al., *Classification and regression trees*. Chapman and Hall/CRC, 1984.

[148] Y. Freund and R. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, Journal of Computer and System Sciences **55** no. 1, (1997) 119. http://www.sciencedirect.com/science/article/pii/S002200009791504X.

[149] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at 8 TeV*, Tech. Rep. ATLAS-CONF-2013-082, CERN, Geneva, Aug, 2013. https://cds.cern.ch/record/1570993.

[150] ATLAS Collaboration, *Determination of the tau energy scale and the associated systematic uncertainty in proton-proton collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector at the LHC in 2012*, Tech. Rep. ATLAS-CONF-2013-044, CERN, Geneva, Apr, 2013. http://cds.cern.ch/record/1544036.

[151] ATLAS Collaboration, *Modelling* $Z \to \tau\tau$ *processes in ATLAS with* τ*-embedded* $Z \to \mu\mu$ *data*, JINST **10** (2015) P09018, arXiv:1506.05623 [hep-ex].

[152] M. Mühlleitner and M. Spira, *A Note on Doubly-Charged Higgs Pair Production at Hadron Colliders*, Phys. Rev. D **68** (2003) 117701, hep-ph/0305288.

[153] A. Akeroyd and M. Aoki, *Single and Pair Production of Doubly Charged Higgs Bosons at Hadron Colliders*, Phys. Rev. D **72** (2005) 035011, hep-ph/0506176.

[154] W. Chao et al., *Correlative signatures of heavy Majorana neutrinos and doubly-charged Higgs bosons at the Large Hadron Collider*, Phys. Lett. B **666** (2008) 451, arXiv:0804.1265 [hep-ph].

[155] F. Aguila and J. Aguilar-Saavedra, *Distinguishing seesaw models at LHC with multi-lepton signals*, Nucl.Phys. B **813** (2009) 22, arXiv:0808.2468 [hep-ph].

[156] C. Chiang et al., *Search for doubly charged Higgs bosons using the same-sign diboson mode at the LHC*, Phys. Rev. D **85** (2012) 095023, arXiv:1202.2014 [hep-ph].

[157] ATLAS Collaboration, *Search for anomalous production of prompt like-sign lepton pairs at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **12** (2012) 007, arXiv:1210.4538 [hep-ex].

[158] M. Botje et al., *The PDF4LHC Working Group Interim Recommendations*, arXiv:1101.0538 [hep-ph].

[159] J. Campbell et al., *Vector boson pair production at the LHC*, JHEP **1107** (2011) 018, arXiv:1105.0020 [hep-ph].

[160] J. Campbell and K. Ellis, *$t\bar{t}W^{+-}$ production and decay at NLO*, JHEP **07** (2012) 052, arXiv:1204.5678 [hep-ph].

[161] A. Kardos et al., *Top quark pair production in association with a Z-boson at NLO accuracy*, Phys. Rev. D **85** (2012) 054015, arXiv:1111.0610 [hep-ph].

[162] M. Garzelli et al., *$t\bar{t}W^{\pm}$ and $t\bar{t}Z$ Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects*, JHEP **1211** (2012) 056, arXiv:1208.2665 [hep-ph].

[163] S. Catani et al., *Vector boson production at hadron colliders: A fully exclusive QCD calculation at NNLO*, Phys. Rev. Lett. **103** (2009) 082001, arXiv:0903.2120 [hep-ph].

[164] M. Cacciari et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. B **710** (2012) 612, arXiv:1111.5869 [hep-ph].

[165] P. Bärnreuther et al., *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** (2012) 132001, arXiv:1204.5201 [hep-ph].

[166] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **12** (2012) 054, arXiv:1207.0236 [hep-ph].

[167] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP **01** (2013) 080, arXiv:1210.6832 [hep-ph].

[168] M. Czakon et al., *The total top quark pair production cross-section at hadron colliders through $O(\alpha_s^4)$*, Phys. Rev. Lett. **110** (2013) 252004, arXiv:1303.6254 [hep-ph].

[169] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. D **81** (2010) 054028, arXiv:1001.5034 [hep-ph].

[170] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W or H*, Phys. Rev. D **82** (2010) 054018, arXiv:1005.4451 [hep-ph].

[171] ROOT Collaboration, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, Tech. Rep. CERN-OPEN-2012-016, New York U., New York, Jan, 2012. https://cds.cern.ch/record/1456844.

[172] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, (erratum *ibid* **73** (2013), 2501), arXiv:1007.1727 [physics.data-an].

[173] A. Read, *Presentation of search results: The CL(s) technique*, J. Phys. G **28** (2002) 2693.

[174] S. Kanemura et al., *First constraint on the mass of doubly-charged Higgs bosons in the same-sign diboson decay scenario at the LHC*, Phys. Lett. B **726** (2013) 316, arXiv:1305.2383 [hep-ph].

[175] ATLAS Collaboration, *Search for doubly-charged Higgs bosons in like-sign dilepton final states at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. C **72** (2012) 2244, arXiv:1210.5070 [hep-ex].

[176] CMS Collaboration, *A search for a doubly-charged Higgs boson in pp collisions at $\sqrt{s} = 7$* TeV, Eur. Phys. J. C **72** (2012) 2189, arXiv:1207.2666 [hep-ex].

[177] CDF Collaboration, *Inclusive Search for New Physics with Like-Sign Dilepton Events in ppbar Collisions at $\sqrt{s} = 1.96$* TeV, Phys. Rev. Lett. **98** (2007) 221803, hep-ex/0702051.

[178] ATLAS Collaboration, *Search for doubly-charged Higgs bosons in same-charge electron pair final states using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2016-051, CERN, Geneva, Aug, 2016. http://cds.cern.ch/record/2206133.

[179] A. Djouadi et al., *HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension*, Comput. Phys. Commun. **108** (1998) 56, arXiv:hep-ph/9704448 [hep-ph].

[180] A. Bredenstein et al., *Precise predictions for the Higgs-boson decay $H \to WW/ZZ \to 4$ leptons*, Phys. Rev. D **74** (2006) 013004, arXiv:hep-ph/0604011 [hep-ph].

[181] S. Actis et al., *NNLO Computational Techniques: The Cases H —> gamma gamma and H —> g g*, Nucl. Phys. B **811** (2009) 182, arXiv:0809.3667 [hep-ph].

[182] A. Denner et al., *Standard Model Higgs-Boson Branching Ratios with Uncertainties*, Eur. Phys. J. C **71** (2011) 1753, arXiv:1107.5909 [hep-ph].

[183] 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*, Tech. Rep. ATLAS-CONF-2014-058, CERN, Geneva, Oct, 2014. https://cds.cern.ch/record/1951336.

[184] C. Lester, *A search for the Higgs boson produced in association with top quarks in multilepton final states at ATLAS*. PhD thesis, Pennsylvania U., Feb, 2015. https://cds.cern.ch/record/2037201.

[185] A. Martin et al., *Parton distributions for the LHC*, Eur. Phys. J. C **63** (2009) 189, arXiv:0901.0002 [hep-ph].

[186] R. Ball et al., *Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology*, Nucl. Phys. B **849** (2011) 296, arXiv:1101.1300 [hep-ph].

[187] ATLAS Collaboration, *Measurements of $W^{\pm}Z$ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector and limits on anomalous gauge boson self-couplings*, Phys. Rev. D **93** no. 9, (2016) 092004, arXiv:1603.02151 [hep-ex].

[188] ATLAS Collaboration, *Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = =$ 7 and 8 TeV in the ATLAS experiment*, Eur. Phys. J. C **76** no. 1, (2016) 6, arXiv:1507.04548 [hep-ex].

[189] CMS Collaboration, *Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV*, Eur. Phys. J. C **75** no. 5, (2015) 212, arXiv:1412.8662 [hep-ex].

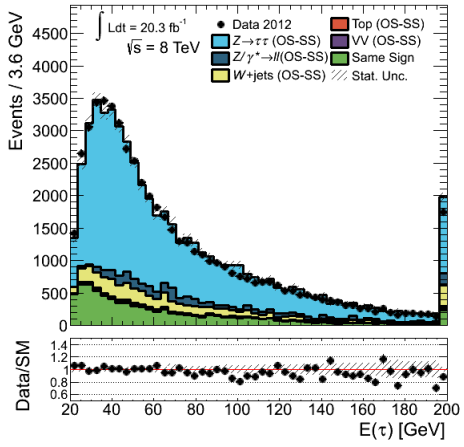
Appendix A

Additional information about the τ_{had} energy calibration

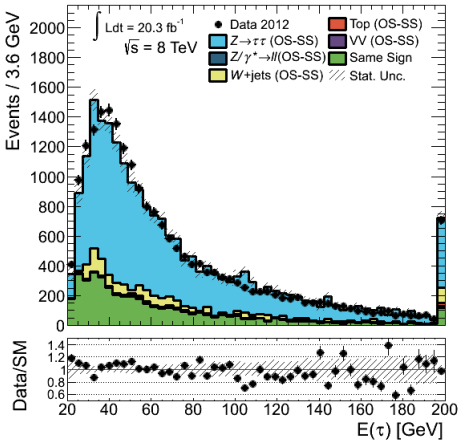
This appendix contains additional information about the analysis described in Chapter 3.

A.1 Lepton and τ_{had} kinematic in the signal region

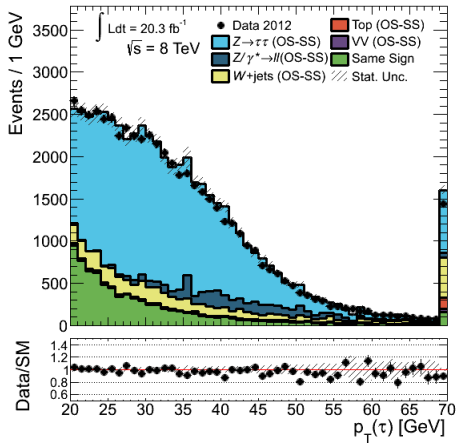
This section collects the distributions of the τ_{had} and the muon kinematic in the $Z \to \tau^+\tau^-$ events used for the TES measurement. Figure A.1 shows the kinematic distributions of 1-track and 3-tracks τ_{had}, while Figure A.2 shows the muon kinematic and E_{T}^{miss} for the same events.



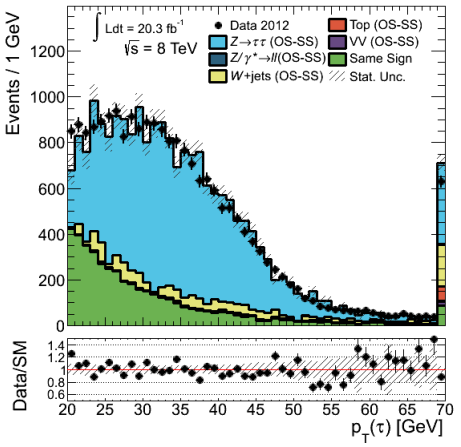
(a) 1-track τ_{had}



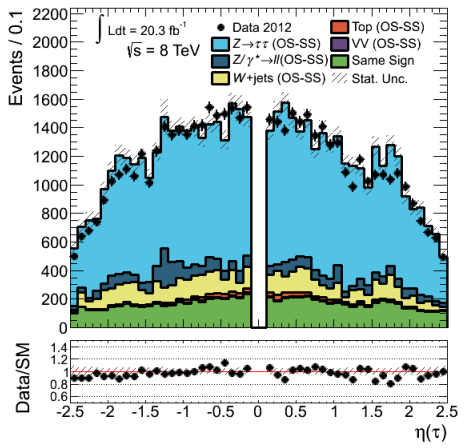
(b) 3-tracks τ_{had}



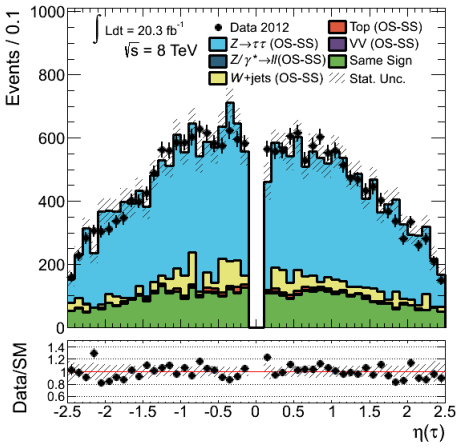
(c) 1-track τ_{had}



(d) 3-tracks τ_{had}

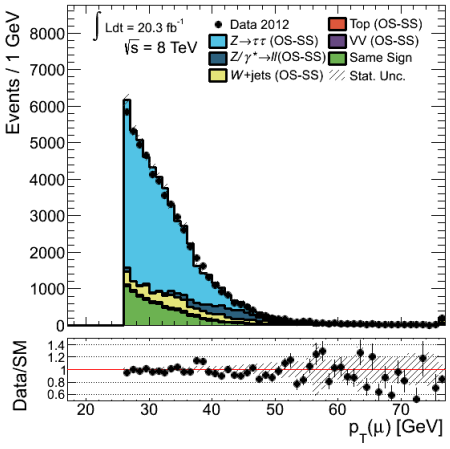


(e) 1-track τ_{had}

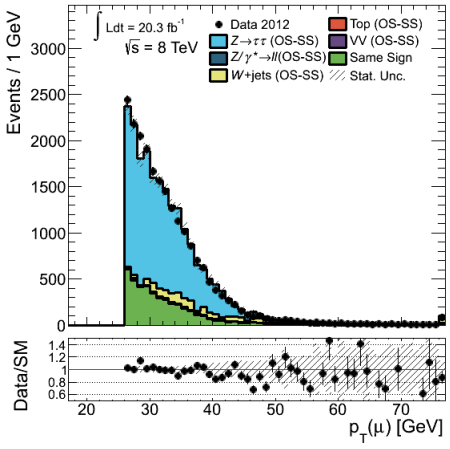


(f) 3-tracks τ_{had}

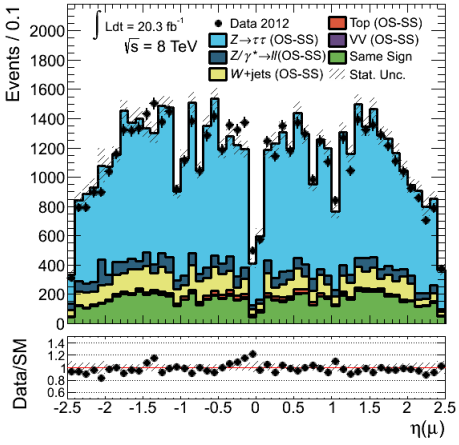
Figure A.1: Energy (top), transverse momentum (centre) and pseudorapidity (bottom) distributions of τ_{had} with one (left) or three (right) associated tracks in $Z \to \tau^+\tau^-$ events selected according to Section 3.2. Events in data (black dots) are compared to the expected stacked contributions from $Z \to \tau^+\tau^-$ (light blue) and background processes determined as described in Section 3.3. Only statistical uncertainties are shown. Underflow and overflow events are added to the first and last bin, respectively. The lower panel shows the ratio of observed events in data to the sum of expected signal and background events.



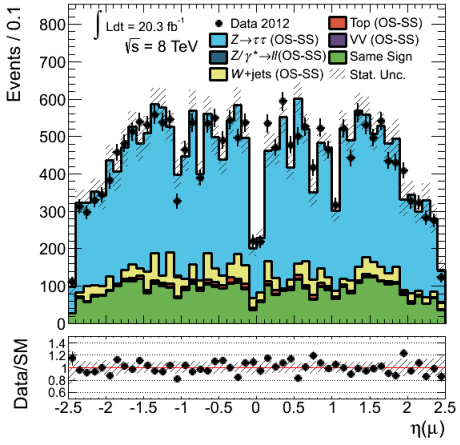
(a) 1-track τ_{had}



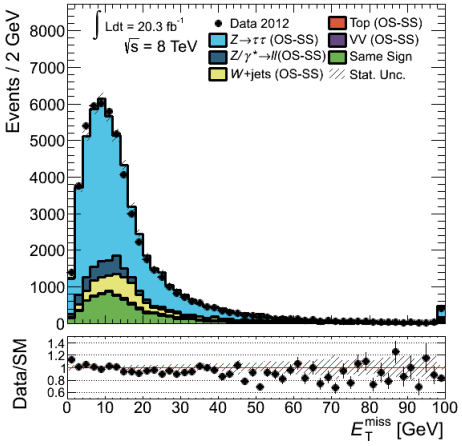
(b) 3-tracks τ_{had}



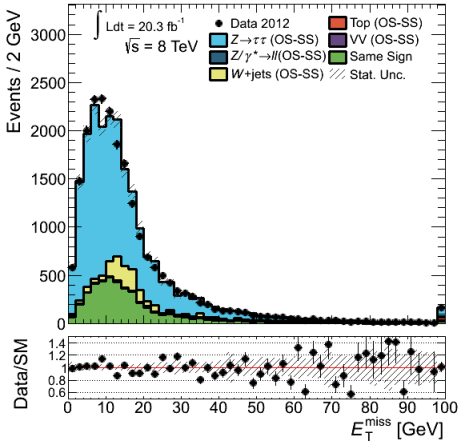
(c) 1-track τ_{had}



(d) 3-tracks τ_{had}



(e) 1-track τ_{had}



(f) 3-tracks τ_{had}

Figure A.2: Muon transverse (top), muon pseudorapidity (centre) and E_{T}^{miss} (bottom) distributions in $Z \to \tau^+\tau^-$ events selected according to Section 3.2 and τ_{had} with one (left) or three (right) associated tracks. Events in data (black dots) are compared to the expected stacked contributions from $Z \to \tau^+\tau^-$ (light blue) and background processes determined as described in Section 3.3. Only statistical uncertainties are shown. Underflow and overflow events are added to the first and last bin, respectively. The lower panel shows the ratio of observed events in data to the sum of expected signal and background events.

A.2 Estimators of TES variations

The effect of a TES variation in simulated background events only on the M_{vis} estimators described in Section 3.4 is shown in Figure A.3.

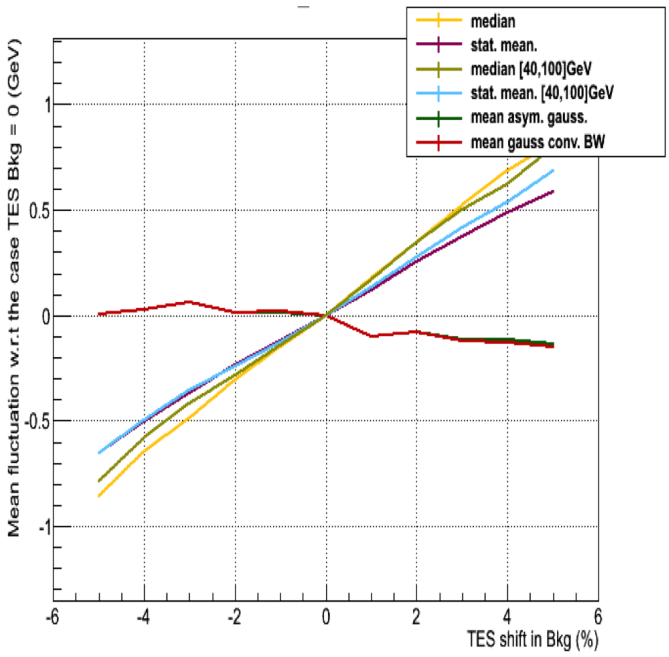


Figure A.3: Difference in the value of the estimator measured in data due to the variation of the TES in background events. The estimator considered are the mean, the median and the parametric functions Bifurcated Gaussian (dark green) and Gaussian $\otimes$ Breit-Wigner (red). The results for mean and median are also obtained for 40 GeV $< M_{vis} <$ 100 GeV.

The difference in the value of the Composed Bifurcated Gaussian and Landau fit parameters in data and simulation as a function of α is shown in Figure A.4. each parameter is highly non-linear for variations of α including the maximum position. The Composed Bifurcated Gaussian and Landau is therefore not used in the estimate of TES differences.

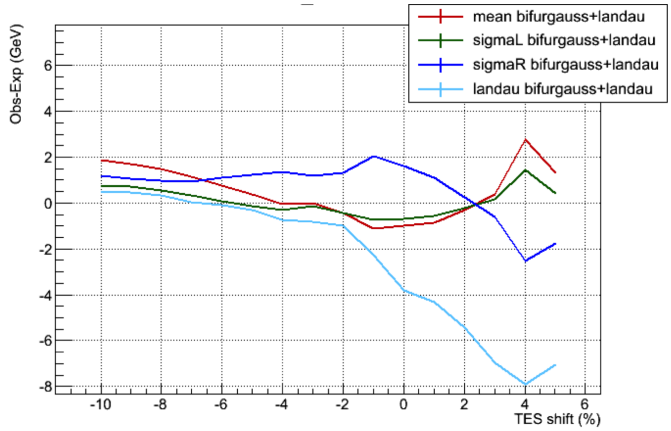


Figure A.4: The difference in the value of the Composed Bifurcated Gaussian and Landau fit parameters in data and simulation as a function of α. The peak position parameter is given in red.

A.3 Other TES corrections

Table A.1 shows the correction to apply to the TES in data in order to obtain matching between the maximum position of the M_{vis} in data and simulation estimated with a Polynomial (5th order) fit. τ_{had} in the *barrel* region and in the *end-cap* regions are also studied separately.

τ_{had} selection	TES correction
$\lvert\eta\rvert < 0.8$	0.4% ± 1.5% (stat.) ± 0.8% (syst.)
$0.8 < \lvert\eta\rvert < 2.5$	0.7% ± 1.2% (stat.) ± 0.6% (syst.)

Table A.1: Corrections for the TES of τ_{had} reconstructed in data in the *barrel* region and in the *end-cap* regions.

Appendix B

Additional information about new physics searches

This chapter contains additional information regarding the new physics search described in Chapter 5:

Section B.1 : list of simulated samples used to model new physics and to evaluate the prompt background in the electron fake factor regions;

Section B.2 : analysis of the single lepton trigger bias on the non-prompt background estimate;

Section B.3 : additional distributions from the signal regions and expected yields divided by individual production processes and pair total charge;

Section B.4 : CL_S limits divided by pair total charge.

B.1 Simulated samples for new physics and for $W^{\pm}$ + jets background estimate

Simulated $W^{\pm}$+jets samples have been used to model the prompt background in the region of estimate of the electron fake factors as described in Section 5.4.4.2. The background from $W^{\pm}$ + jets, as well as from other processes with at least a prompt lepton, must be subtracted in order to have an accurate measurement of the electron fake factors. The simulations used to model the $W^{\pm}$ + jets production are listed in Table B.1.

Several models of new physics producing same-sign lepton pairs have been considered in this analysis as described in Section 5.1. These models are studied with simulations and are listed in Table B.2. For each process the masses of the non-Standard Model particles involved are varied to study how the detection probability of the same-sign lepton pair changes with these parameters.

Process	Generator	$\sigma \times \varepsilon_{filter}$ [pb]	K-factor
$W^{\pm} \to e^{\pm}\nu$	Alpgen+Herwig	8037.1	1.176
$W^{\pm} \to e^{\pm}\nu + 1p$	Alpgen+Herwig	1579.2	1.176
$W^{\pm} \to e^{\pm}\nu + 2p$	Alpgen+Herwig	472.2	1.176
$W^{\pm} \to e^{\pm}\nu + 3p$	Alpgen+Herwig	133.9	1.176
$W^{\pm} \to e^{\pm}\nu + 4p$	Alpgen+Herwig	35.6	1.176
$W^{\pm} \to e^{\pm}\nu + \geq 5p$	Alpgen+Herwig	10.6	1.176
$W^{\pm} \to \tau^{\pm}\nu$	Alpgen+Herwig	8035.8	1.176
$W^{\pm} \to \tau^{\pm}\nu + 1p$	Alpgen+Herwig	1579.8	1.176
$W^{\pm} \to \tau^{\pm}\nu + 2p$	Alpgen+Herwig	477.6	1.176
$W^{\pm} \to \tau^{\pm}\nu + 3p$	Alpgen+Herwig	133.8	1.176
$W^{\pm} \to \tau^{\pm}\nu + 4p$	Alpgen+Herwig	35.6	1.176
$W^{\pm} \to \tau^{\pm}\nu + \geq 5p$	Alpgen+Herwig	10.5	1.176

Table B.1: The $W^{\pm}+$jets simulations used to subtract prompt background contaminations from the electron fake factor region. For each process, the generator configuration, the cross-section times the generator filter efficiency and the K-factor used to scale the simulated cross-section value to the latest theoretical calculation are given. The full simulation setup is specified in Table 2.1. $X + np$ with n integer and X a generic process indicates the number of partons (quarks or gluons) produced at matrix element level in addition to the process X.

Process	Generator
$H_{\mathrm{L}}^{++}H_{\mathrm{L}}^{--} \to \ell^+\ell^+\ell^-\ell^-$ ($m(H_{\mathrm{L}}^{\pm\pm}) = 50$ GeV)	Pythia8
$H_{\mathrm{L}}^{++}H_{\mathrm{L}}^{--} \to \ell^+\ell^+\ell^-\ell^-$ ($m(H_{\mathrm{L}}^{\pm\pm}) = 100$ GeV)	Pythia8
$H_{\mathrm{L}}^{++}H_{\mathrm{L}}^{--} \to \ell^+\ell^+\ell^-\ell^-$ ($m(H_{\mathrm{L}}^{\pm\pm}) = 150$ GeV)	Pythia8
$H_{\mathrm{L}}^{++}H_{\mathrm{L}}^{--} \to \ell^+\ell^+\ell^-\ell^-$ ($m(H_{\mathrm{L}}^{\pm\pm}) = 200$ GeV)	Pythia8
$H_{\mathrm{L}}^{++}H_{\mathrm{L}}^{--} \to \ell^+\ell^+\ell^-\ell^-$ ($m(H_{\mathrm{L}}^{\pm\pm}) = 250$ GeV)	Pythia8
$H_{\mathrm{L}}^{++}H_{\mathrm{L}}^{--} \to \ell^+\ell^+\ell^-\ell^-$ ($m(H_{\mathrm{L}}^{\pm\pm}) = 300$ GeV)	Pythia8
$H_{\mathrm{L}}^{++}H_{\mathrm{L}}^{--} \to \ell^+\ell^+\ell^-\ell^-$ ($m(H_{\mathrm{L}}^{\pm\pm}) = 350$ GeV)	Pythia8
$H_{\mathrm{L}}^{++}H_{\mathrm{L}}^{--} \to \ell^+\ell^+\ell^-\ell^-$ ($m(H_{\mathrm{L}}^{\pm\pm}) = 400$ GeV)	Pythia8
$H_{\mathrm{L}}^{++}H_{\mathrm{L}}^{--} \to \ell^+\ell^+\ell^-\ell^-$ ($m(H_{\mathrm{L}}^{\pm\pm}) = 450$ GeV)	Pythia8
$H_{\mathrm{L}}^{++}H_{\mathrm{L}}^{--} \to \ell^+\ell^+\ell^-\ell^-$ ($m(H_{\mathrm{L}}^{\pm\pm}) = 500$ GeV)	Pythia8
$H_{\mathrm{L}}^{++}H_{\mathrm{L}}^{--} \to \ell^+\ell^+\ell^-\ell^-$ ($m(H_{\mathrm{L}}^{\pm\pm}) = 600$ GeV)	Pythia8
$H_{\mathrm{L}}^{++}H_{\mathrm{L}}^{--} \to \ell^+\ell^+\ell^-\ell^-$ ($m(H_{\mathrm{L}}^{\pm\pm}) = 1000$ GeV)	Pythia8
$W_{\mathrm{R}}^{\pm} \to N_{\mathrm{R}}\ell^{\pm}$ ($m(W_{\mathrm{R}}^{\pm}) = 1000$ GeV, $m(N_{\mathrm{R}}) = 250$ GeV)	Pythia8
$W_{\mathrm{R}}^{\pm} \to N_{\mathrm{R}}\ell^{\pm}$ ($m(W_{\mathrm{R}}^{\pm}) = 1000$ GeV, $m(N_{\mathrm{R}}) = 500$ GeV)	Pythia8
$W_{\mathrm{R}}^{\pm} \to N_{\mathrm{R}}\ell^{\pm}$ ($m(W_{\mathrm{R}}^{\pm}) = 1000$ GeV, $m(N_{\mathrm{R}}) = 750$ GeV)	Pythia8
$W_{\mathrm{R}}^{\pm} \to N_{\mathrm{R}}\ell^{\pm}$ ($m(W_{\mathrm{R}}^{\pm}) = 1400$ GeV, $m(N_{\mathrm{R}}) = 350$ GeV)	Pythia8
$W_{\mathrm{R}}^{\pm} \to N_{\mathrm{R}}\ell^{\pm}$ ($m(W_{\mathrm{R}}^{\pm}) = 1400$ GeV, $m(N_{\mathrm{R}}) = 700$ GeV)	Pythia8
$W_{\mathrm{R}}^{\pm} \to N_{\mathrm{R}}\ell^{\pm}$ ($m(W_{\mathrm{R}}^{\pm}) = 1400$ GeV, $m(N_{\mathrm{R}}) = 1050$ GeV)	Pythia8
$W_{\mathrm{R}}^{\pm} \to N_{\mathrm{R}}\ell^{\pm}$ ($m(W_{\mathrm{R}}^{\pm}) = 2000$ GeV, $m(N_{\mathrm{R}}) = 500$ GeV)	Pythia8
$W_{\mathrm{R}}^{\pm} \to N_{\mathrm{R}}\ell^{\pm}$ ($m(W_{\mathrm{R}}^{\pm}) = 2000$ GeV, $m(N_{\mathrm{R}}) = 1000$ GeV)	Pythia8
$W_{\mathrm{R}}^{\pm} \to N_{\mathrm{R}}\ell^{\pm}$ ($m(W_{\mathrm{R}}^{\pm}) = 2000$ GeV, $m(N_{\mathrm{R}}) = 1500$ GeV)	Pythia8
$b' \to qW^{\pm}$ ($m(b') = 400$ GeV)	Pythia8
$b' \to qW^{\pm}$ ($m(b') = 500$ GeV)	Pythia8
$b' \to qW^{\pm}$ ($m(b') = 600$ GeV)	Pythia8
$b' \to qW^{\pm}$ ($m(b') = 650$ GeV)	Pythia8
$b' \to qW^{\pm}$ ($m(b') = 700$ GeV)	Pythia8
$b' \to qW^{\pm}$ ($m(b') = 800$ GeV)	Pythia8
$b' \to qW^{\pm}$ ($m(b') = 900$ GeV)	Pythia8
$b' \to qW^{\pm}$ ($m(b') = 950$ GeV)	Pythia8
$b' \to qW^{\pm}$ ($m(b') = 1000$ GeV)	Pythia8
$b' \to tW^{\pm}$ ($m(b') = 450$ GeV)	Pythia8
$b' \to tW^{\pm}$ ($m(b') = 500$ GeV)	Pythia8
$b' \to tW^{\pm}$ ($m(b') = 550$ GeV)	Pythia8
$b' \to tW^{\pm}$ ($m(b') = 600$ GeV)	Pythia8
$b' \to tW^{\pm}$ ($m(b') = 650$ GeV)	Pythia8
$b' \to tW^{\pm}$ ($m(b') = 700$ GeV)	Pythia8
$b' \to tW^{\pm}$ ($m(b') = 800$ GeV)	Pythia8
$b' \to tW^{\pm}$ ($m(b') = 900$ GeV)	Pythia8
$b' \to tW^{\pm}$ ($m(b') = 1000$ GeV)	Pythia8
$S_{\mathrm{DQ}} \to S_{\mathrm{LQ}}S_{\mathrm{LQ}}$ ($m(S_{\mathrm{DQ}}) = 2500$ GeV, $m(S_{\mathrm{LQ}}) = 1000$ GeV)	MadGraph+Pythia8
$S_{\mathrm{DQ}} \to S_{\mathrm{LQ}}S_{\mathrm{LQ}}$ ($m(S_{\mathrm{DQ}}) = 2500$ GeV, $m(S_{\mathrm{LQ}}) = 1200$ GeV)	MadGraph+Pythia8
$S_{\mathrm{DQ}} \to S_{\mathrm{LQ}}S_{\mathrm{LQ}}$ ($m(S_{\mathrm{DQ}}) = 3000$ GeV, $m(S_{\mathrm{LQ}}) = 1000$ GeV)	MadGraph+Pythia8
$S_{\mathrm{DQ}} \to S_{\mathrm{LQ}}S_{\mathrm{LQ}}$ ($m(S_{\mathrm{DQ}}) = 3000$ GeV, $m(S_{\mathrm{LQ}}) = 1200$ GeV)	MadGraph+Pythia8
$S_{\mathrm{DQ}} \to S_{\mathrm{LQ}}S_{\mathrm{LQ}}$ ($m(S_{\mathrm{DQ}}) = 3500$ GeV, $m(S_{\mathrm{LQ}}) = 1000$ GeV)	MadGraph+Pythia8
$S_{\mathrm{DQ}} \to S_{\mathrm{LQ}}S_{\mathrm{LQ}}$ ($m(S_{\mathrm{DQ}}) = 3500$ GeV, $m(S_{\mathrm{LQ}}) = 1200$ GeV)	MadGraph+Pythia8
$S_{\mathrm{DQ}} \to S_{\mathrm{LQ}}S_{\mathrm{LQ}}$ ($m(S_{\mathrm{DQ}}) = 3500$ GeV, $m(S_{\mathrm{LQ}}) = 1400$ GeV)	MadGraph+Pythia8

Table B.2: Simulation used to model beyond Standard Model processes that produce same-sign lepton pairs. The full simulation setup is specified in Table 2.1. The simulation cross-section is not used in the analysis and therefore is not shown.

B.2 Biases of the SLTrig in the $e^{\pm}\mu^{\pm}$ category

The electrons used to estimate the fake factors in Section 5.4.4.2 are responsible for the event acquisition via the activation of one of the single electron triggers in Table 5.5. These triggers check for energy leakages in the hadronic calorimeter of the cluster associated to the electron at L1 for electrons with $p_{\mathrm{T}} < 60$ GeV. This condition affects the lepton selection and consequently the electron fake factors estimate. To be consistent, the electrons used in the analysis should satisfy the same leakage requirement. The use of di-lepton triggers enforces the application of this requirement on any electron considered making the lepton pair selection consistent with the non-prompt background estimate.

If single lepton triggers are used instead, one of the leptons in the pair does not have to satisfy the trigger requirements, hence its selection may be incompatible with the non-prompt background estimate. No significant issues have been observed in the usage of pairs with not triggered muons. However, a bias is observed when not triggered electrons are considered. This is the case of events with $e^{\pm}\mu^{\pm}$ pairs where only the single muon trigger has been activated.

Figure B.1 shows the invariant mass distribution of $e^{\pm}\mu^{\pm}$ pairs where the electron passes the nominal selection (see Table 5.3) and the muon passes either the Fail-d_0 selection (see Table 5.13) or the Weak isolation selection (see Table 5.11). The events where the single electron trigger is activated or only the muon trigger fires are shown separately. The background predictions are in agreement with data when the electron trigger is activated, but an excess of predicted non-prompt background is observed when only the muon trigger is activated. To avoid biases in the non-prompt background estimate, the EMTrig is preferred to the SLTrig in the $e^{\pm}\mu^{\pm}$ category.

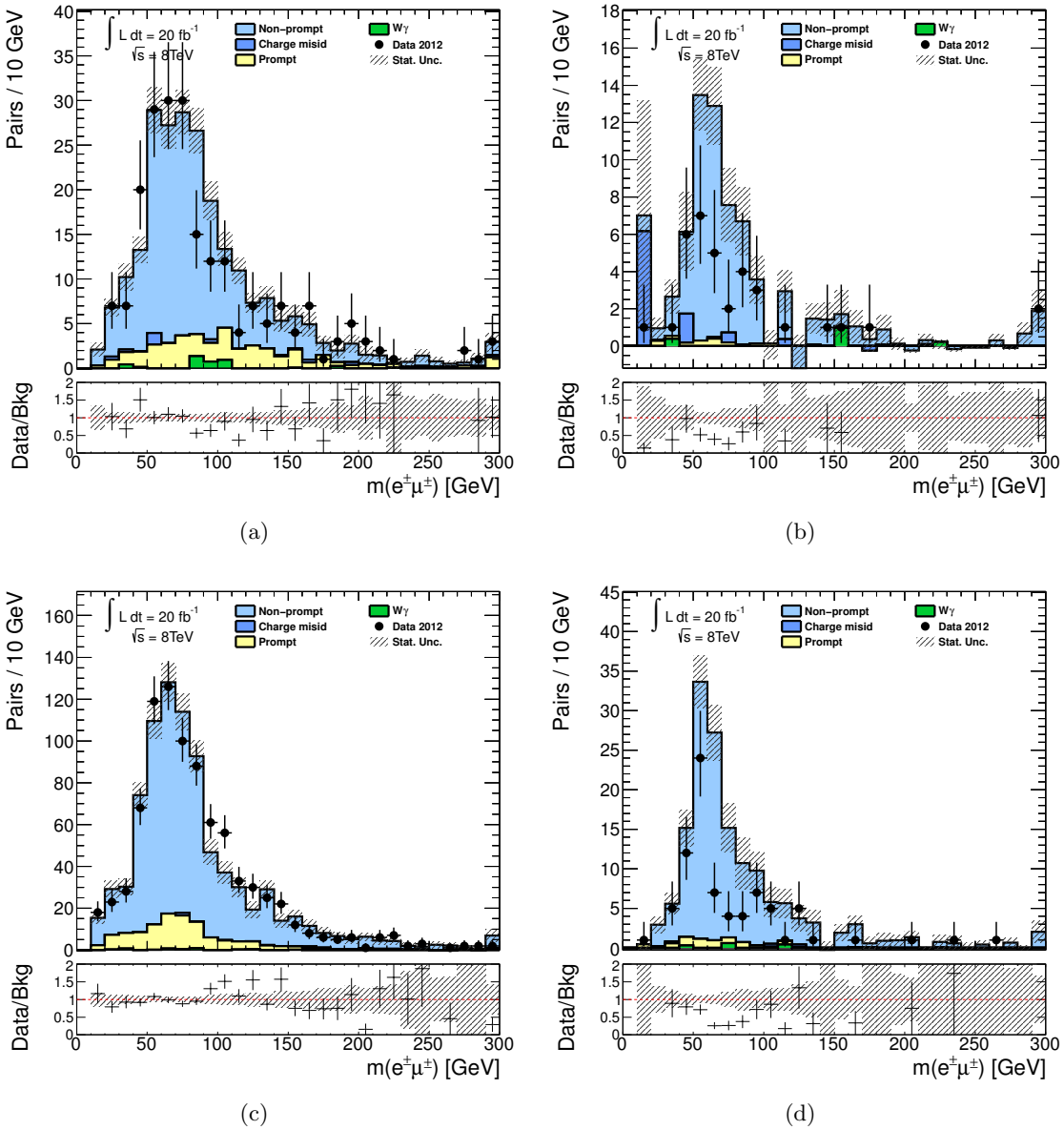


Figure B.1: Invariant mass distributions of $e^{\pm}\mu^{\pm}$ pairs with muon passing the Fail-d_0 selection (top) or the Weak Isolation selection (bottom). The pairs are divided on the base of the activation of the SETrig: active (left) and inactive (right). Events in data (black dots) are compared to the expected stacked contributions from Standard Model processes, as described in Section 5.4. Overflow events are added to the last bin. The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the predictions.

B.3 Additional signal region tables and distributions

The expected number of $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ in the signal regions divided by single process are shown in Table B.3, Table B.4 and Table B.5. The expected and observed number of pairs in the three lepton flavour categories divided by total pair charge are instead shown in Table B.6. The distributions of the p_{T} and η of the leading p_{T} lepton are shown in Figure B.2, while the distributions of the subleading p_{T} lepton are shown in Figure B.3.

Sample	Number of electron pairs with $m(e^{\pm}, e^{\pm})$						
	> 15 GeV	> 100 GeV	> 200 GeV	> 300 GeV	> 400 GeV	> 500 GeV	> 600 GeV
Non-prompt	518.57 ± 120.17	247.49 ± 49.5	71.67 ± 13.15	22.66 ± 4.8	8.13 ± 2.42	3.12 ± 1.49	0.78 ± 1.01
$W^{\pm}Z$	234.36 ± 22.24	132.79 ± 12.76	37.12 ± 3.9	10.95 ± 1.43	3.23 ± 0.61	1.5 ± 0.4	0.5 ± 0.22
ZZ	86.05 ± 7.21	22.42 ± 2.11	6.75 ± 0.84	1.78 ± 0.37	0.61 ± 0.2	0.34 ± 0.16	0.21 ± 0.12
$W^{\pm}W^{\pm}$	14.99 ± 7.59	12.1 ± 6.14	5.55 ± 2.84	2.35 ± 1.22	1.22 ± 0.66	0.4 ± 0.24	0.16 ± 0.11
MPI VV	4.04 ± 4.06	1.6 ± 1.61	0.38 ± 0.39	0.06 ± 0.07	0.02 ± 0.02	0 ± 0	0 ± 0
$t\bar{t}W^{\pm}$	5.33 ± 1.23	3.83 ± 0.89	1.32 ± 0.32	0.44 ± 0.11	0.14 ± 0.04	0.08 ± 0.03	0.03 ± 0.01
$t\bar{t}Z$	1.73 ± 0.41	1.2 ± 0.29	0.4 ± 0.1	0.11 ± 0.04	0.03 ± 0.01	0.02 ± 0.01	0.01 ± 0.01
Prompt Total	346.51 ± 24.95	173.94 ± 14.44	51.52 ± 4.93	15.7 ± 1.92	5.25 ± 0.92	2.34 ± 0.49	0.91 ± 0.28
$W^{\pm}\gamma$	175.25 ± 36.28	74.89 ± 15.62	22.42 ± 5.15	8.04 ± 2.26	3.84 ± 1.31	2.69 ± 1.05	1.02 ± 0.57
Z/γ^{*}	968.61 ± 145.63	513.53 ± 77.7	130.91 ± 26.99	36.1 ± 12.17	12.8 ± 7.89	4.79 ± 4.86	4.79 ± 4.86
$t\bar{t}$	36.92 ± 6.01	30.1 ± 4.99	14.55 ± 2.8	5.05 ± 1.32	2.15 ± 0.78	1.05 ± 0.58	1.18 ± 0.56
$W^{+}W^{-}$	13.01 ± 2.34	10.74 ± 1.96	4.85 ± 0.97	1.86 ± 0.45	0.68 ± 0.22	0.43 ± 0.16	0.28 ± 0.13
Charge misID Total	1018.54 ± 145.78	554.37 ± 77.89	150.31 ± 27.16	43.01 ± 12.25	15.62 ± 7.93	6.27 ± 4.89	6.25 ± 4.89
Total Background	2058.86 ± 193.92	1050.69 ± 94.67	295.92 ± 30.99	89.41 ± 13.49	32.83 ± 8.44	14.41 ± 5.25	8.96 ± 5.04

Table B.3: Expected numbers of $e^{\pm}e^{\pm}$ pairs for various di-electron invariant mass, $m(e^{\pm}, e^{\pm})$, intervals. The errors shown are the overall uncertainties on the predictions.

Sample	Number of electron-muon pairs with $m(e^{\pm}, \mu^{\pm})$						
	> 15 GeV	> 100 GeV	> 200 GeV	> 300 GeV	> 400 GeV	> 500 GeV	> 600 GeV
Non-prompt	910 ± 220	342 ± 77	79 ± 17	23.7 ± 5.8	9.2 ± 3.0	2.8 ± 1.5	1.6 ± 1.0
$W^{\pm}Z$	710 ± 38	359 ± 20	98.5 ± 6.1	30.8 ± 2.5	10.9 ± 1.2	4.51 ± 0.75	2.32 ± 0.53
ZZ	235 ± 13	49.3 ± 3.2	12.6 ± 1.2	3.31 ± 0.53	1.33 ± 0.33	0.57 ± 0.22	0.39 ± 0.18
$W^{\pm}W^{\pm}$	46 ± 23	32 ± 16	13.2 ± 6.7	7.2 ± 3.7	3.1 ± 1.6	1.57 ± 0.84	0.74 ± 0.42
MPI VV	13 ± 13	4.6 ± 4.6	1.1 ± 1.1	$0.24^{+0.25}_{-0.24}$	$0.040^{+0.050}_{-0.040}$	$0.0021^{+0.0029}_{-0.0021}$	0.0 ± 0.0
$t\bar{t}W^{\pm}$	16.6 ± 8.3	10.2 ± 5.1	3.2 ± 1.6	1.15 ± 0.58	0.41 ± 0.21	0.127 ± 0.070	0.066 ± 0.038
$t\bar{t}Z$	6.4 ± 3.2	3.8 ± 1.9	1.37 ± 0.69	0.35 ± 0.18	0.143 ± 0.076	0.062 ± 0.036	0.020 ± 0.013
Prompt Total	1027 ± 49	458 ± 26	129.9 ± 9.3	43.1 ± 4.5	16.0 ± 2.1	6.8 ± 1.1	3.53 ± 0.70
Z/γ^{*}	259 ± 37	12.3 ± 7.8	0.000046 ± 0.000046	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
$t\bar{t}$	86 ± 11	52.8 ± 7.2	20.9 ± 3.7	6.0 ± 1.7	1.09 ± 0.62	0.34 ± 0.34	0.34 ± 0.34
$W^{+}W^{-}$	19.7 ± 2.3	14.2 ± 1.7	5.97 ± 0.89	2.70 ± 0.51	1.47 ± 0.34	1.14 ± 0.31	0.55 ± 0.21
$tW^{\pm}$	8.5 ± 2.1	7.9 ± 2.1	1.63 ± 0.82	0.80 ± 0.64	0.000 ± 0.014	0.0 ± 0.0	0.0 ± 0.0
Charge misID Total	372 ± 38	87 ± 11	28.5 ± 3.9	9.5 ± 1.9	2.54 ± 0.82	1.48 ± 0.41	0.89 ± 0.41
$W^{\pm}\gamma$	271 ± 50	104 ± 20	27.7 ± 6.1	8.1 ± 2.4	2.7 ± 1.1	1.55 ± 0.77	1.24 ± 0.69
Total Background	2580 ± 240	991 ± 88	265 ± 22	84.4 ± 8.3	30.5 ± 4.0	12.6 ± 2.1	7.4 ± 1.5

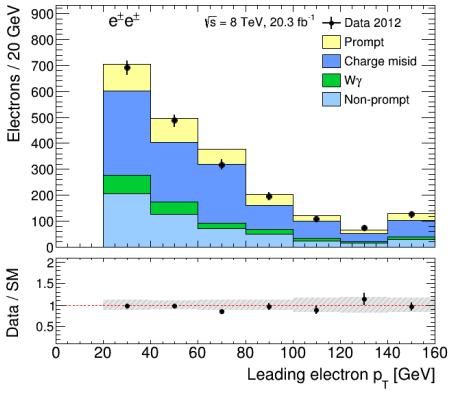
Table B.4: Expected numbers of $e^{\pm}\mu^{\pm}$ pairs for various electron-muon invariant mass, $m(e^{\pm}, \mu^{\pm})$, intervals. The errors shown are the overall uncertainties on the predictions.

Sample	Number of muon pairs with $m(\mu^{\pm},\mu^{\pm})$						
	>15 GeV	>100 GeV	>200 GeV	>300 GeV	>400 GeV	>500 GeV	>600 GeV
Non-prompt	203 ± 34	56 ± 11	8.7 ± 2.3	1.9 ± 1.0	1.22 ± 0.88	0.32 ± 0.41	$0.0^{+0.2}_{-0.0}$
$W^{\pm}Z$	412 ± 32	193 ± 15	49.8 ± 4.5	14.1 ± 1.6	4.72 ± 0.81	1.19 ± 0.37	0.11 ± 0.11
ZZ	113.3 ± 7.2	18.7 ± 1.6	4.66 ± 0.62	1.59 ± 0.35	0.60 ± 0.21	0.44 ± 0.18	0.25 ± 0.15
$W^{\pm}W^{\pm}$	37 ± 18	26 ± 13	10.7 ± 5.4	4.5 ± 2.3	2.3 ± 1.2	1.22 ± 0.67	0.55 ± 0.32
$t\bar{t}W^{\pm}$	6.1 ± 1.4	3.09 ± 0.71	0.91 ± 0.22	0.221 ± 0.066	0.069 ± 0.030	0.013 ± 0.013	0.013 ± 0.013
$t\bar{t}Z$	1.58 ± 0.36	0.88 ± 0.21	0.163 ± 0.048	0.060 ± 0.024	0.023 ± 0.012	$0.0^{+0.0}_{-0.0}$	$0.0^{+0.0}_{-0.0}$
MPI VV	9.5 ± 9.5	3.6 ± 3.6	0.94 ± 0.95	0.25 ± 0.26	0.0027 ± 0.0037	0.0025 ± 0.0036	0.0025 ± 0.0036
Prompt Total	579 ± 41	245 ± 21	67.2 ± 7.3	20.7 ± 2.9	7.7 ± 1.5	2.87 ± 0.79	0.92 ± 0.37
Total Background	782 ± 54	301 ± 24	75.9 ± 7.6	22.6 ± 3.1	9.0 ± 1.7	3.19 ± 0.89	$0.92^{+0.43}_{-0.37}$

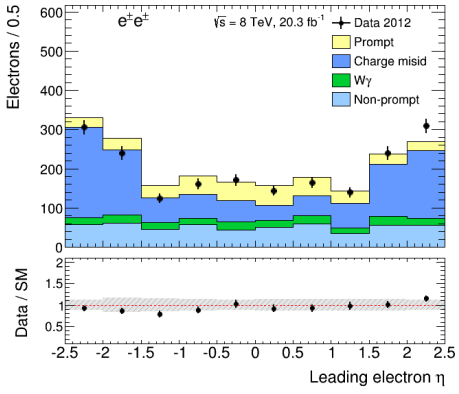
Table B.5: Expected numbers of $\mu^{\pm}\mu^{\pm}$ pairs for various di-muon invariant mass, $m(\mu^{\pm},\mu^{pm})$, intervals. The errors shown are the overall uncertainties on the predictions.

$m(\ell^+,\ell^+)$	e^+e^+ pairs		$e^+\mu^+$ pairs		$\mu^+\mu^+$ pairs	
[GeV]	Expected	Data	Expected	Data	Expected	Data
>15	1120 ± 100	1124	1440 ± 130	1327	454 ± 32	502
>100	610 ± 60	593	570 ± 50	523	184 ± 16	198
>200	187 ± 22	167	146 ± 13	143	48 ± 6	62
>300	61 ± 11	48	50 ± 5	56	15.3 ± 2.2	18
>400	19 ± 6	18	18.4 ± 2.6	21	6.2 ± 1.2	6
>500	9 ± 5	9	7.8 ± 1.4	8	2.6 ± 0.8	1
>600	7 ± 5	5	4.8 ± 1.1	6	0.8 ± 0.4	0
$m(\ell^-,\ell^-)$	e^-e^- pairs		$e^-\mu^-$ pairs		$\mu^-\mu^-$ pairs	
[GeV]	Expected	Data	Expected	Data	Expected	Data
>15	940 ± 100	852	1140 ± 110	988	328 ± 23	341
>100	440 ± 50	394	417 ± 40	336	117 ± 9	132
>200	109 ± 16	98	119 ± 11	83	27.6 ± 2.8	25
>300	29 ± 7	35	35 ± 4	29	7.3 ± 1.2	9
>400	14 ± 5	12	12.1 ± 2.3	10	2.7 ± 0.7	3
>500	5.0 ± 1.3	4	4.9 ± 1.5	5	$0.64^{+0.33}_{-0.26}$	3
>600	2.7 ± 0.9	2	2.5 ± 1.0	3	$0.09^{+0.23}_{-0.09}$	1

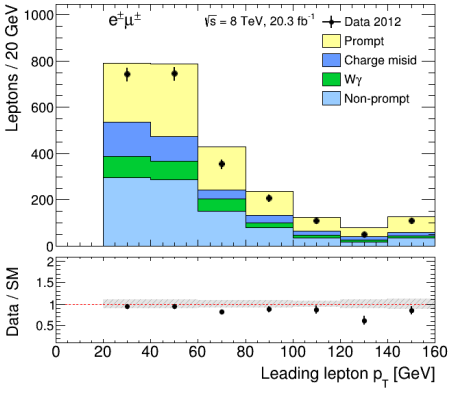
Table B.6: Expected and observed numbers of positively or negatively charged lepton pairs for various invariant mass, $m(\ell^{\pm},\ell^{\pm})$, intervals. The errors shown are the overall uncertainties on the predictions.



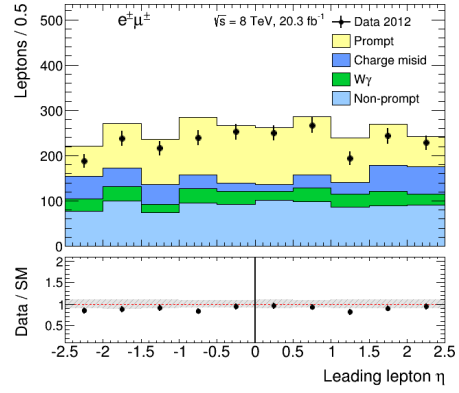
(a) $e^{\pm}e^{\pm}$



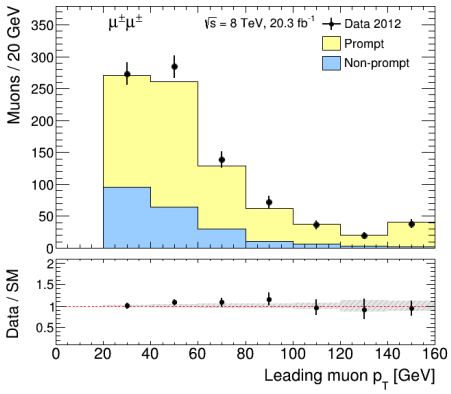
(b) $e^{\pm}e^{\pm}$



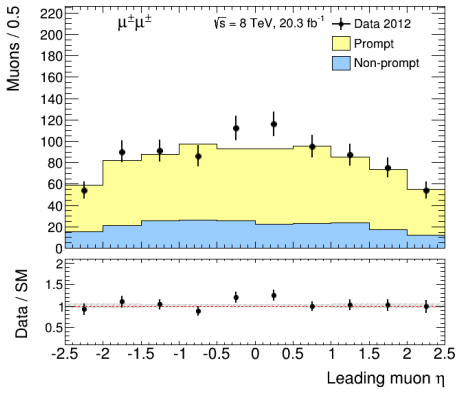
(c) $e^{\pm}\mu^{\pm}$



(d) $e^{\pm}\mu^{\pm}$

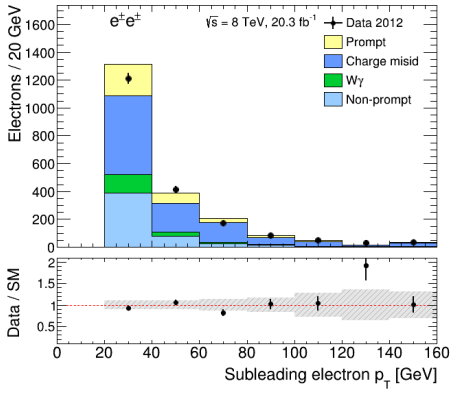


(e) $\mu^{\pm}\mu^{\pm}$

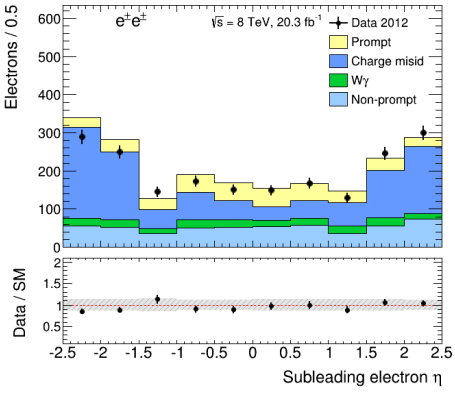


(f) $\mu^{\pm}\mu^{\pm}$

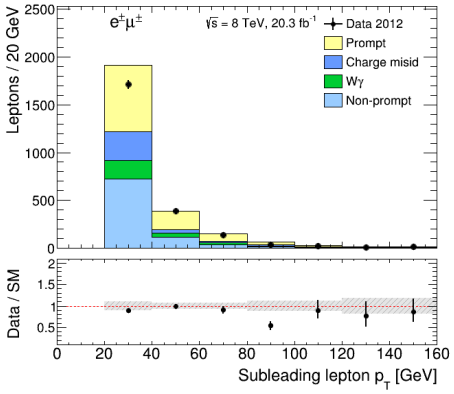
Figure B.2: Leading lepton p_{T} (left) and η (right) distributions in $e^{\pm}e^{\pm}$ (top) $e^{\pm}\mu^{\pm}$ (centre) and $\mu^{\pm}\mu^{\pm}$ (bottom) pairs. Events in data (black dots) are compared to the expected stacked contributions from Standard Model processes, as described in Section 5.4. Overflow events are added to the last bin. The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the predictions.



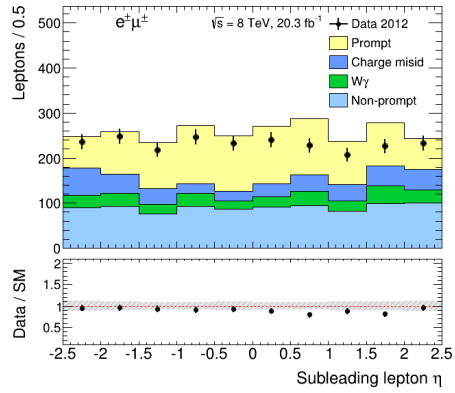
(a) $e^{\pm}e^{\pm}$



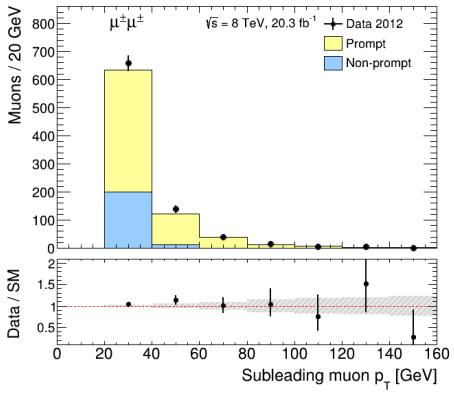
(b) $e^{\pm}e^{\pm}$



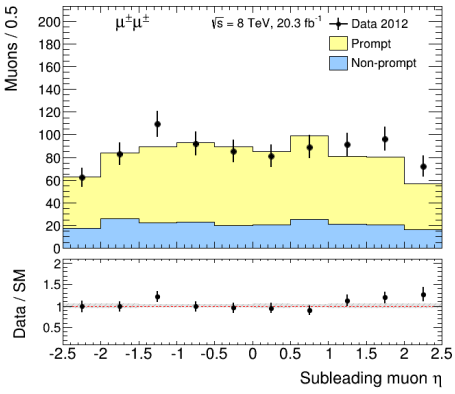
(c) $e^{\pm}\mu^{\pm}$



(d) $e^{\pm}\mu^{\pm}$



(e) $\mu^{\pm}\mu^{\pm}$



(f) $\mu^{\pm}\mu^{\pm}$

Figure B.3: Subleading lepton p_{T} (left) and η (right) distributions in $e^{\pm}e^{\pm}$ (top) $e^{\pm}\mu^{\pm}$ (centre) and $\mu^{\pm}\mu^{\pm}$ (bottom) pairs. Events in data (black dots) are compared to the expected stacked contributions from Standard Model processes, as described in Section 5.4. Overflow events are added to the last bin. The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the predictions.

B.4 Limits divided by pair charge sign

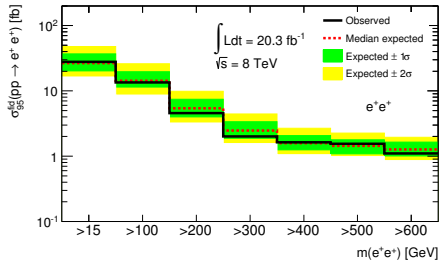
The 95% CL_S upper limits on the number of $e^{\pm}e^{\pm}$, $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ pairs due to non-Standard Model processes are given divided by pair total charge in Table B.7 and Table B.8. The 95% upper limits on the fiducial cross-sections divided by pair total charge, flavour category and invariant mass requirement are shown in Figure B.4.

	95% CL_S upper limit					
	e^+e^+		$e^+\mu^+$		$\mu^+\mu^+$	
Mass range	expected	observed	expected	observed	expected	observed
$M > 15$ GeV	$260.78^{+100.84(207.83)}_{-62.89(100.49)}$	272.66	$254.27^{+100.68(183.94)}_{-41.91(123.70)}$	235.64	$88.36^{+31.15(70.02)}_{-28.41(49.22)}$	134.74
$M > 100$ GeV	$140.02^{+52.89(115.03)}_{-27.63(55.40)}$	131.82	$108.69^{+35.25(83.19)}_{-21.47(58.80)}$	92.60	$47.05^{+15.11(33.73)}_{-11.80(24.64)}$	58.90
$M > 200$ GeV	$53.05^{+19.90(43.39)}_{-13.87(21.49)}$	44.57	$36.59^{+12.91(29.93)}_{-6.65(11.28)}$	36.70	$20.75^{+7.14(15.96)}_{-4.49(10.04)}$	33.54
$M > 300$ GeV	$24.14^{+8.78(19.16)}_{-5.68(8.84)}$	19.72	$19.35^{+7.66(17.71)}_{-5.20(7.30)}$	26.12	$10.37^{+4.26(9.71)}_{-2.75(3.69)}$	13.28
$M > 400$ GeV	$15.56^{+4.63(10.75)}_{-3.29(5.22)}$	16.02	$11.09^{+4.65(10.54)}_{-2.37(4.22)}$	14.04	$6.86^{+2.48(4.16)}_{-1.58(3.13)}$	6.95
$M > 500$ GeV	$14.13^{+3.34(8.07)}_{-3.49(4.48)}$	15.15	$7.94^{+2.15(5.97)}_{-2.20(2.74)}$	8.94	$3.95^{+2.26(6.27)}_{-0.91(1.05)}$	3.57
$M > 600$ GeV	$12.43^{+3.67(6.72)}_{-2.58(4.02)}$	10.77	$6.58^{+1.40(9.42)}_{-1.60(2.81)}$	7.78	$3.45^{+0.86(2.78)}_{-0.48(0.84)}$	3.02

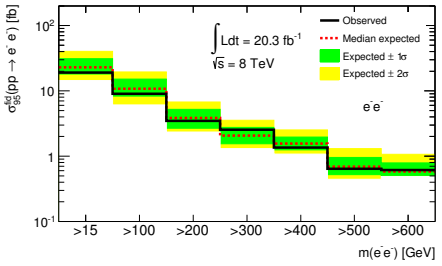
Table B.7: The 95% CL_S upper limit on the number of positively charged lepton pairs due to non-Standard Model physics in each signal region and for several invariant mass requirements. The 1σ (2σ) bands on the expected limit are also given.

	95% CL_S upper limit					
	e^-e^-		$e^-\mu^-$		$\mu^-\mu^-$	
Mass range	expected	observed	expected	observed	expected	observed
$M > 15$ GeV	$224.35^{+80.11(170.05)}_{-46.89(83.34)}$	186.74	$191.82^{+80.19(150.12)}_{-27.72(91.92)}$	161.46	$63.33^{+24.96(52.43)}_{-14.31(24.16)}$	77.50
$M > 100$ GeV	$105.39^{+42.59(85.42)}_{-23.62(45.49)}$	88.27	$82.81^{+21.64(51.43)}_{-21.12(27.08)}$	56.74	$32.47^{+12.71(26.54)}_{-8.44(15.90)}$	47.27
$M > 200$ GeV	$37.85^{+13.21(28.60)}_{-11.52(14.94)}$	34.06	$28.38^{+11.88(24.05)}_{-9.19(12.33)}$	15.36	$13.18^{+5.05(10.78)}_{-3.05(5.78)}$	12.06
$M > 300$ GeV	$20.30^{+6.70(14.37)}_{-5.13(7.52)}$	24.95	$15.68^{+5.57(13.42)}_{-3.94(6.53)}$	12.81	$7.34^{+2.83(6.60)}_{-1.49(2.23)}$	9.11
$M > 400$ GeV	$15.26^{+4.02(9.44)}_{-3.01(4.68)}$	13.24	$9.17^{+3.40(5.57)}_{-2.60(3.60)}$	7.79	$4.82^{+1.90(4.54)}_{-1.21(1.48)}$	5.48
$M > 500$ GeV	$6.79^{+2.61(6.03)}_{-1.67(2.45)}$	6.26	$6.25^{+1.16(2.83)}_{-1.25(2.37)}$	6.52	$3.32^{+1.30(3.02)}_{-0.12(0.27)}$	6.37
$M > 600$ GeV	$5.69^{+2.03(4.76)}_{-0.79(0.93)}$	5.99	$4.91^{+1.64(5.41)}_{-1.00(1.18)}$	5.91	$3.10^{+0.13(1.50)}_{-0.11(0.16)}$	4.24

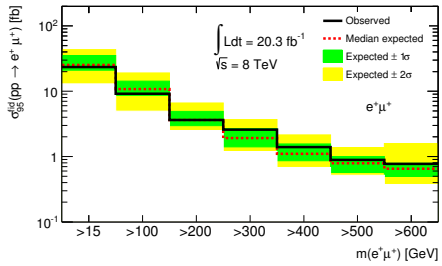
Table B.8: The 95% CL_S upper limit on the number of negatively charged lepton pairs due to non-Standard Model physics in each signal region and for several invariant mass requirements. The 1σ (2σ) bands on the expected limit are also given.



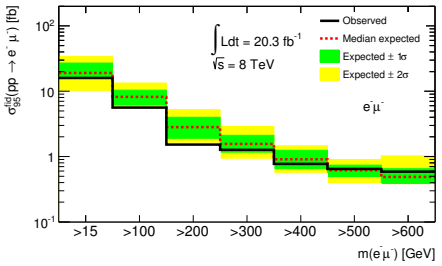
(a) e^+e^+



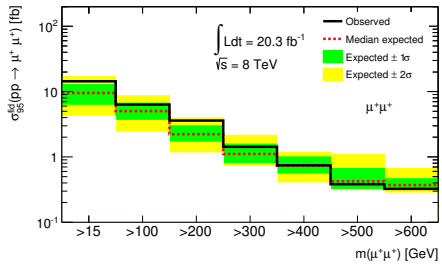
(b) e^-e^-



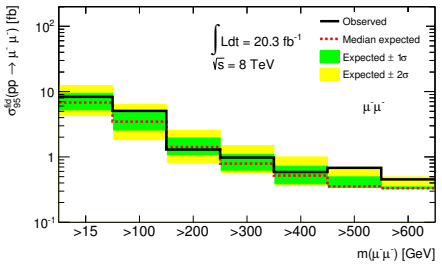
(c) $e^+\mu^+$



(d) $e^-\mu^-$



(e) $\mu^+\mu^+$



(f) $\mu^-\mu^-$

Figure B.4: Observed and expected 95% CL_S upper limits on the fiducial cross-section of new physics in e^+e^+ (a), e^-e^- (b), $e^+\mu^+$ (c), $e^-\mu^-$ (d), $\mu^+\mu^+$ (e) and $\mu^-\mu^-$ (f) pairs and as a function of the lower bound on the lepton pair mass. The green and yellow bands show the $\pm1\sigma$ and $\pm2\sigma$ bands on the expected limits.

Appendix C

Additional studies on $t\bar{t}H$ production

This chapter contains additional information regarding the $t\bar{t}H$ analysis described in Chapter 6:

Section C.1 : additional simulated samples used in validation studies;

Section C.2 : further studies on non-prompt background estimate;

Section C.3 : definition of the other signal regions used in the $t\bar{t}H$ cross-section combined measurement;

Section C.4 : additional control distributions and other validation regions;

Section C.5 : systematic uncertainties on the predictions in the additional signal regions used in the $t\bar{t}H$ cross-section combined measurement;

Section C.6 : additional distributions for events in the $2\ell 0\tau_{\text{had}}$ category;

Section C.7 : distributions for events in the other signal region categories.

C.1 Additional simulated samples for background studies and validation regions

Simulated $t\bar{t}$, *single-top*, $tW^{\pm}$, W^+W^-, $W^{\pm}$+jets and Z/γ^*+jets samples have been used to model the shapes of non-prompt backgrounds in the validation regions and the signal regions. However, in the regions with same-sign leptons, the non-prompt background yields for each jet multiplicity are obtained from data as described in Section 6.4.4. In the validation regions with opposite-sign leptons, these simulations are also used to determine the prompt background. In the studies of non-prompt background estimate in Section 6.4.5.3, these samples are again used to evaluate prompt backgrounds in the opposite-sign lepton

regions but also charge misID backgrounds in same-sign lepton regions. Finally, the $t\bar{t}$ simulations are used in the closure tests in Section 6.4.5.5 too.

The simulation used to generate $t\bar{t}$, *single-top*, $tW^{\pm}$, $W^{+}W^{-}$, $W^{\pm}$+jets and Z/γ^{*}+jets processes are listed in Table C.1.

Table C.1: Simulated processes used for charge misID and non-prompt background modeling. For each process, the generator, the cross-section and the luminosity equivalent to the statistics generated is given. The full simulation setup is specified in Table 2.1. $X + np$ with $n = 1, 2, 3, 4, 5$ and X a generic process indicates the number of partons (quarks or gluons) produced at matrix element level in addition to the process.

Process	Generator	Cross-section [fb]	Equivalent $\mathcal{L}$ [fb^{-1}]
$t\bar{t}(\geq 1W^{\pm} \to \ell\nu_{\ell})$	Powheg+Pythia6	137320.39	363.91
single-top (s channel, $W^{\pm} \to \ell\nu_{\ell}$)	Powheg+Pythia6	1817.64	660.1
single-top (t channel, $t \to bW^{+} \to \ell^{+}\nu_{\ell}$)	Powheg+Pythia6	18396	271.8
single-top (t channel, $\bar{t} \to \bar{b}W^{-} \to \ell^{-}\bar{\nu}_{\ell}$)	Powheg+Pythia6	9971.82	501.4
$tW^{\pm}$	Powheg+Pythia6	2337.67	424.8
$W^{+}W^{-} \to \ell\nu\ell\nu$	Sherpa	5679.00	475.4
$W^{\pm} \to e^{\pm}\nu$	Alpgen+Pythia6	9573431	0.36
$W^{\pm} \to e^{\pm}\nu + 1p$	Alpgen+Pythia6	1879129	4.39
$W^{\pm} \to e^{\pm}\nu + 2p$	Alpgen+Pythia6	569426.9	6.63
$W^{\pm} \to e^{\pm}\nu + 3p$	Alpgen+Pythia6	159138.7	6.28
$W^{\pm} \to e^{\pm}\nu + 4p$	Alpgen+Pythia6	42215.25	5.92
$W^{\pm} \to e^{\pm}\nu + \geq 5p$	Alpgen+Pythia6	12680.64	5.52
$W^{\pm} \to \mu^{\pm}\nu$	Alpgen+Pythia6	9575097	0.36
$W^{\pm} \to \mu^{\pm}\nu + 1p$	Alpgen+Pythia6	1881628	4.40
$W^{\pm} \to \mu^{\pm}\nu + 2p$	Alpgen+Pythia6	568117.9	6.64
$W^{\pm} \to \mu^{\pm}\nu + 3p$	Alpgen+Pythia6	159352.9	6.30
$W^{\pm} \to \mu^{\pm}\nu + 4p$	Alpgen+Pythia6	42818.58	5.84
$W^{\pm} \to \mu^{\pm}\nu + \geq 5p$	Alpgen+Pythia6	12386.71	1.58
$W^{\pm} \to \tau^{\pm}\nu$	Alpgen+Pythia6	9554153	0.37
$W^{\pm} \to \tau^{\pm}\nu + 1p$	Alpgen+Pythia6	1878653	1.25
$W^{\pm} \to \tau^{\pm}\nu + 2p$	Alpgen+Pythia6	568582	6.64
$W^{\pm} \to \tau^{\pm}\nu + 3p$	Alpgen+Pythia6	159471.9	6.30
$W^{\pm} \to \tau^{\pm}\nu + 4p$	Alpgen+Pythia6	41957.02	5.96
$W^{\pm} \to \tau^{\pm}\nu + \geq 5p$	Alpgen+Pythia6	12668.74	5.12
$W^{\pm} + b\bar{b}$	Alpgen+Pythia6	66261.58	7.16
$W^{\pm} + b\bar{b} + 1p$	Alpgen+Pythia6	53839.17	6.68
$W^{\pm} + b\bar{b} + 2p$	Alpgen+Pythia6	27662.74	6.3
$W^{\pm} + b\bar{b} + 3p$	Alpgen+Pythia6	13261.36	3.77
$W^{\pm} + c\bar{c}$	Alpgen+Pythia6	178726.10	7.13

Table C.1: Simulated processes used for charge misID and non-prompt background modeling. For each process, the generator, the cross-section and the luminosity equivalent to the statistics generated is given. The full simulation setup is specified in Table 2.1. $X + np$ with $n = 1, 2, 3, 4, 5$ and X a generic process indicates the number of partons (quarks or gluons) produced at matrix element level in addition to the process.

Process	Generator	Cross-section [fb]	Equivalent $\mathcal{L}$ [fb^{-1}]
$W^{\pm} + c\bar{c} + 1p$	Alpgen+Pythia6	157889.20	6.65
$W^{\pm} + c\bar{c} + 2p$	Alpgen+Pythia6	85450.33	6.12
$W^{\pm} + c\bar{c} + 3p$	Alpgen+Pythia6	36014.16	4.72
$W^{\pm} + c$	Alpgen+Pythia6	961389.1	6.64
$W^{\pm} + c + 1p$	Alpgen+Pythia6	318455.9	6.5
$W^{\pm} + c + 2p$	Alpgen+Pythia6	83089.37	6.2
$W^{\pm} + c + 3p$	Alpgen+Pythia6	24450.93	4.5
$W^{\pm} + c + 4p$	Alpgen+Pythia6	5125.21	3.9
$Z/\gamma^* \to e^+e^-$ ($m_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	875477.10	7.3
$Z/\gamma^* \to e^+e^- + 1p$ ($m_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	190859.10	7.0
$Z/\gamma^* \to e^+e^- + 2p$ ($m_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	59956.35	6.8
$Z/\gamma^* \to e^+e^- + 3p$ ($m_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	17496.75	6.3
$Z/\gamma^* \to e^+e^- + 4p$ ($m_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	4624.18	6.5
$Z/\gamma^* \to e^+e^- + \geq 5p$ ($m_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	1393.71	222.0
$Z/\gamma^* \to \mu^+\mu^-$ ($m_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	875895.30	7.5
$Z/\gamma^* \to \mu^+\mu^- + 1p$ ($m_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	190367.10	6.7
$Z/\gamma^* \to \mu^+\mu^- + 2p$ ($m_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	60161.76	6.7
$Z/\gamma^* \to \mu^+\mu^- + 3p$ ($m_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	17497.98	6.3
$Z/\gamma^* \to \mu^+\mu^- + 4p$ ($m_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	4654.07	6.4
$Z/\gamma^* \to \mu^+\mu^- + \geq 5p$ ($m_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	1395.07	220.3
$Z/\gamma^* \to \tau^+\tau^-$ ($m_{\ell\ell} \in [60, 2000]$GeV)[1]	Alpgen+Pythia6	34707.74	103.2
$Z/\gamma^* \to \tau^+\tau^- + 1p$ ($m_{\ell\ell} \in [60, 2000]$GeV)[1]	Alpgen+Pythia6	9009.69	270.5
$Z/\gamma^* \to \tau^+\tau^- + 2p$ ($m_{\ell\ell} \in [60, 2000]$GeV)[1]	Alpgen+Pythia6	3175.34	156.3
$Z/\gamma^* \to \tau^+\tau^- + 3p$ ($m_{\ell\ell} \in [60, 2000]$GeV)[1]	Alpgen+Pythia6	1028.86	173.4
$Z/\gamma^* \to \tau^+\tau^- + 4p$ ($m_{\ell\ell} \in [60, 2000]$GeV)[1]	Alpgen+Pythia6	317.18	106.1
$Z/\gamma^* \to \tau^+\tau^- + \geq 5p$ ($m_{\ell\ell} \in [60, 2000]$GeV)	Alpgen+Pythia6	1398.02	221.6
$Z/\gamma^* \to e^+e^-$ ($m_{\ell\ell} \in [10, 60]$GeV)[1]	Alpgen+Pythia6	43274.26	159.2
$Z/\gamma^* \to e^+e^- + 1p$ ($m_{\ell\ell} \in [10, 60]$GeV)[1]	Alpgen+Pythia6	26370.87	170.6
$Z/\gamma^* \to e^+e^- + 2p$ ($m_{\ell\ell} \in [10, 60]$GeV)[1]	Alpgen+Pythia6	8702.67	151.4
$Z/\gamma^* \to e^+e^- + 3p$ ($m_{\ell\ell} \in [10, 60]$GeV)[1]	Alpgen+Pythia6	2795.55	156.5
$Z/\gamma^* \to e^+e^- + 4p$ ($m_{\ell\ell} \in [10, 60]$GeV)[1]	Alpgen+Pythia6	777.16	140.7
$Z/\gamma^* \to e^+e^- + \geq 5p$ ($m_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Pythia6	824.28	89.1
$Z/\gamma^* \to \mu^+\mu^-$ ($m_{\ell\ell} \in [10, 60]$GeV)[1]	Alpgen+Pythia6	44947.85	259.8

[1] These samples have a filter to select events with two reconstructed leptons.

Table C.1: Simulated processes used for charge misID and non-prompt background modeling. For each process, the generator, the cross-section and the luminosity equivalent to the statistics generated is given. The full simulation setup is specified in Table 2.1. $X + np$ with $n = 1, 2, 3, 4, 5$ and X a generic process indicates the number of partons (quarks or gluons) produced at matrix element level in addition to the process.

Process	Generator	Cross-section [fb]	Equivalent $\mathcal{L}$ [fb^{-1}]
$Z/\gamma^* \to \mu^+\mu^- + 1p$ ($m_{\ell\ell} \in [10, 60]$GeV)[1]	Alpgen+Pythia6	27298.35	271.0
$Z/\gamma^* \to \mu^+\mu^- + 2p$ ($m_{\ell\ell} \in [10, 60]$GeV)[1]	Alpgen+Pythia6	8957.82	243.2
$Z/\gamma^* \to \mu^+\mu^- + 3p$ ($m_{\ell\ell} \in [10, 60]$GeV)[1]	Alpgen+Pythia6	2875.39	254.4
$Z/\gamma^* \to \mu^+\mu^- + 4p$ ($m_{\ell\ell} \in [10, 60]$GeV)[1]	Alpgen+Pythia6	794.01	137.7
$Z/\gamma^* \to \mu^+\mu^- + \geq 5p$ ($m_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Pythia6	825.54	96.3
$Z/\gamma^* \to \tau^+\tau^-$ ($m_{\ell\ell} \in [10, 60]$GeV)[1]	Alpgen+Pythia6	100.77	287.5
$Z/\gamma^* \to \tau^+\tau^- + 1p$ ($m_{\ell\ell} \in [10, 60]$GeV)[1]	Alpgen+Pythia6	176.04	170.4
$Z/\gamma^* \to \tau^+\tau^- + 2p$ ($m_{\ell\ell} \in [10, 60]$GeV)[1]	Alpgen+Pythia6	109.41	263.7
$Z/\gamma^* \to \tau^+\tau^- + 3p$ ($m_{\ell\ell} \in [10, 60]$GeV)[1]	Alpgen+Pythia6	52.13	396.0
$Z/\gamma^* \to \tau^+\tau^- + 4p$ ($m_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Pythia6	3084.48	70.0
$Z/\gamma^* \to \tau^+\tau^- + \geq 5p$ ($m_{\ell\ell} \in [10, 60]$GeV)	Alpgen+Pythia6	824.55	141.6
$Z/\gamma^* \to e^+e^- + b\bar{b}$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	10304.57	14.6
$Z/\gamma^* \to e^+e^- + b\bar{b} + 1p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	4001.07	20.0
$Z/\gamma^* \to e^+e^- + b\bar{b} + 2p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	1463.95	30.7
$Z/\gamma^* \to e^+e^- + b\bar{b} + 3p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	618.42	8.1
$Z/\gamma^* \to \mu^+\mu^- + b\bar{b}$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	10300.27	14.5
$Z/\gamma^* \to \mu^+\mu^- + b\bar{b} + 1p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	4002.42	20.0
$Z/\gamma^* \to \mu^+\mu^- + b\bar{b} + 2p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	1452.63	31.0
$Z/\gamma^* \to \mu^+\mu^- + b\bar{b} + 3p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	623.23	8.0
$Z/\gamma^* \to \tau^+\tau^- + b\bar{b}$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	10302.11	14.5
$Z/\gamma^* \to \tau^+\tau^- + b\bar{b} + 1p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	3988.52	20.1
$Z/\gamma^* \to \tau^+\tau^- + b\bar{b} + 2p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	1468.37	30.6
$Z/\gamma^* \to \tau^+\tau^- + b\bar{b} + 3p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	612.43	8.2
$Z/\gamma^* \to e^+e^- + c\bar{c}$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	19254.42	31.4
$Z/\gamma^* \to e^+e^- + c\bar{c} + 1p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	8480.36	30.7
$Z/\gamma^* \to e^+e^- + c\bar{c} + 2p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	3592.09	30.6
$Z/\gamma^* \to e^+e^- + c\bar{c} + 3p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	1403.55	28.5
$Z/\gamma^* \to \mu^+\mu^- + c\bar{c}$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	19248.27	31.2
$Z/\gamma^* \to \mu^+\mu^- + c\bar{c} + 1p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	8478.39	31.3
$Z/\gamma^* \to \mu^+\mu^- + c\bar{c} + 2p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	3588.65	32.0
$Z/\gamma^* \to \mu^+\mu^- + c\bar{c} + 3p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	1399.37	28.6
$Z/\gamma^* \to \tau^+\tau^- + c\bar{c}$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	19251.96	31.2
$Z/\gamma^* \to \tau^+\tau^- + c\bar{c} + 1p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	8484.42	31.2
$Z/\gamma^* \to \tau^+\tau^- + c\bar{c} + 2p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	3579.30	26.2

Table C.1: Simulated processes used for charge misID and non-prompt background modeling. For each process, the generator, the cross-section and the luminosity equivalent to the statistics generated is given. The full simulation setup is specified in Table 2.1. $X + np$ with $n = 1, 2, 3, 4, 5$ and X a generic process indicates the number of partons (quarks or gluons) produced at matrix element level in addition to the process.

Process	Generator	Cross-section [fb]	Equivalent $\mathcal{L}$ [fb^{-1}]
$Z/\gamma^* \to \tau^+\tau^- + c\bar{c} + 3p$ ($m_{\ell\ell} \in [40, 2000]$GeV)	Alpgen+Pythia6	1394.82	28.7

C.2 Further studies on the non-prompt background estimate

This section extends the study of the non-prompt background estimate described in Section 6.4.5.

C.2.1 More non-prompt regions

Section 6.4.5.3 compares candidates non-prompt regions for the estimate of fake factors and efficiencies. Among the studied regions, there are two $W^{\pm}$+jets regions with opposite-sign leptons and a Z + jets region with $\Delta\phi$ requirements. The $|\eta|$ and p_{T} distributions of electrons in the $W^{\pm}$ + jets non-prompt regions with opposite-sign leptons are shown in Figure C.1 for non-prompt electron with $|d_0|/\sigma(d_0) < 3$ and in Figure C.2 for those with $|d_0|/\sigma(d_0) > 3$. The figures display separately the distributions for *tight* and *non-tight* electrons. As stated in Section 6.4.5.3, the contamination of prompt background in these regions is dominant, thus these regions are not appropriate for the calculation of fake factors and efficiencies.

The electron p_{T} and $|\eta|$ distributions in events belonging to the Z+jets non-prompt region and selected with the $\Delta\phi$ requirements are shown in Figure C.3 separately for *tight* and *non-tight* electrons. The statistics in this region is very poor and there is a quite significant diboson contamination, thus the region is not appropriate for the calculation of fake factors and efficiencies.

C.2.2 Prompt regions

The prompt regions defined in Section 6.4.5.4 have been used to calculate r_ℓ. The p_{T} and $|\eta|$ distributions of the electrons and muons in the prompt regions are shown in Figure C.4

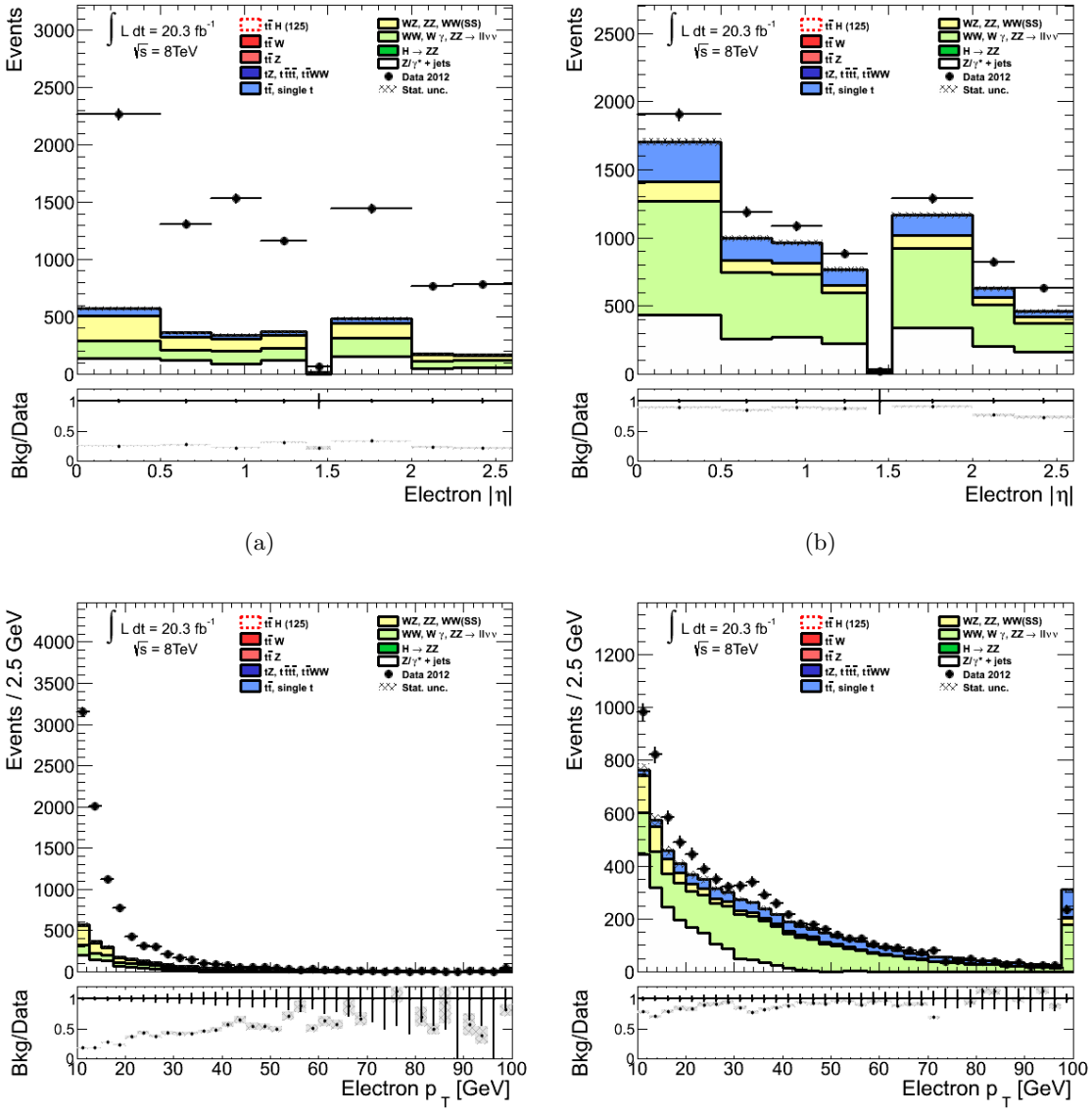


Figure C.1: The $|\eta|$ (top) and p_{T} (bottom) distributions of electrons passing (right) and failing (left) the *tight* selection in the $W^{\pm}$ + jets opposite-sign lepton non-prompt region. Events in data (black dots) are compared to the expected stacked contributions from prompt Standard Model processes. Overflow events are added to the last bin. The lower panel shows the fraction of expected prompt background yields. The error bars on the data points and the dashed band on the simulated background show the statistical uncertainties.

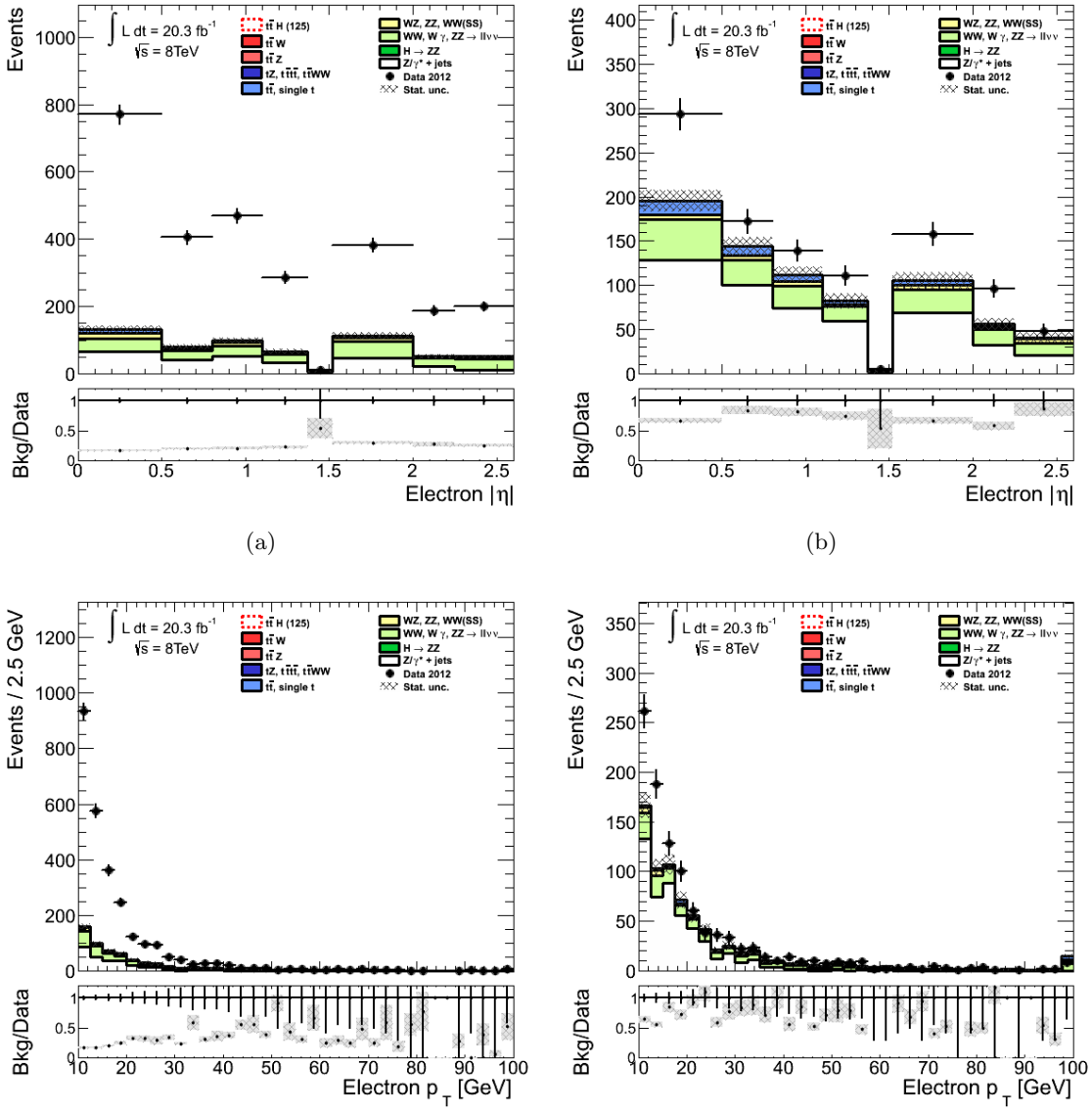
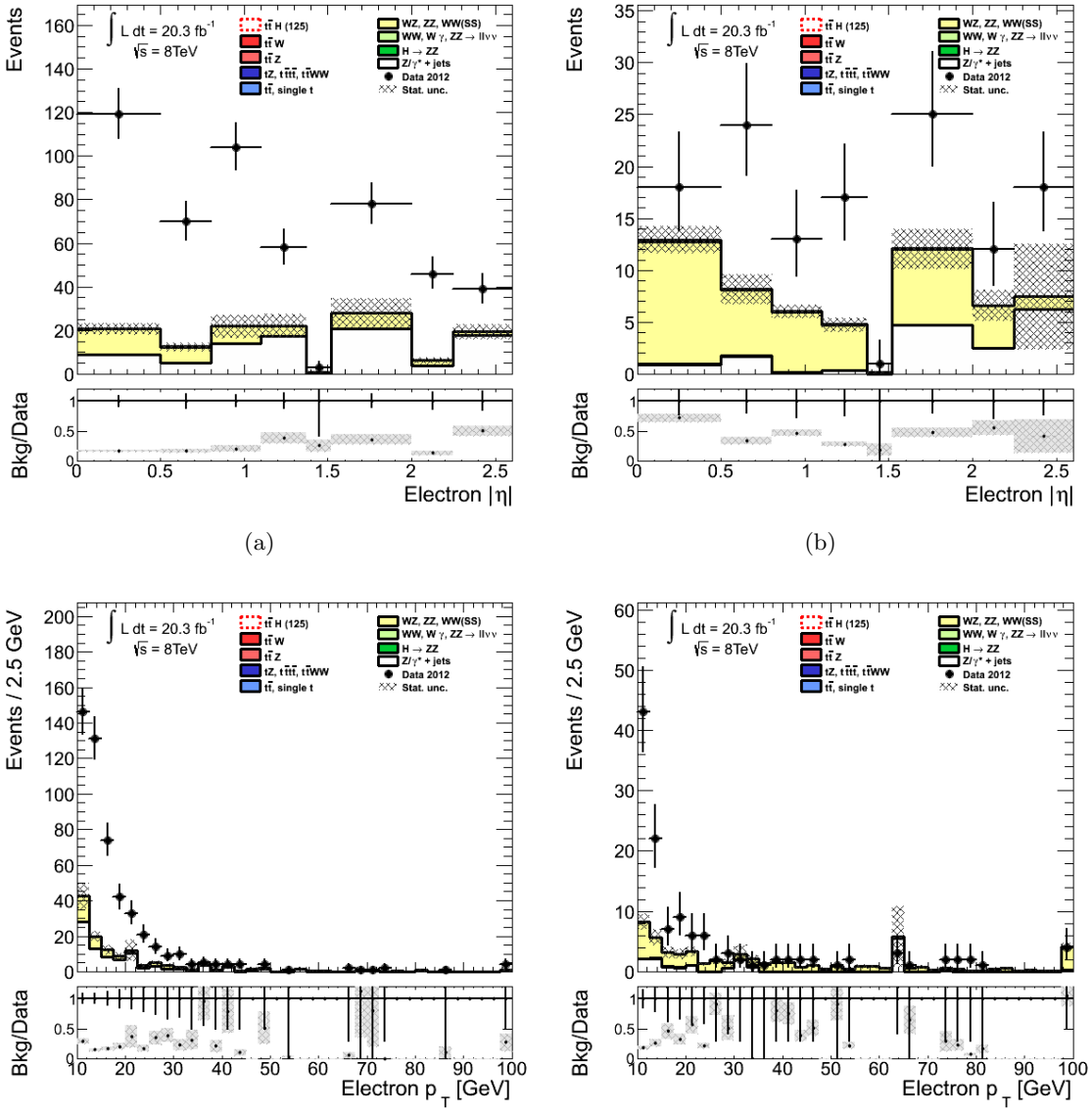


Figure C.2: The $|\eta|$ (top) and p_{T} (bottom) distributions of electrons with $|d_0|/\sigma(d_0) > 3$ passing (right) and failing (left) the *tight* selection in the $W^{\pm}$+jets opposite-sign lepton non-prompt region. Events in data (black dots) are compared to the expected stacked contributions from prompt Standard Model processes. Overflow events are added to the last bin. The lower panel shows the fraction of expected prompt background yields. The error bars on the data points and the dashed band on the simulated background show the statistical uncertainties.



(a) (b)

(c) (d)

Figure C.3: The $|\eta|$ (top) and p_{T} (bottom) distributions of electrons passing (right) and failing (left) the *tight* selection in the $Z+$jets $\Delta\phi$ selection non-prompt region. Events in data (black dots) are compared to the expected stacked contributions from prompt Standard Model processes. Overflow events are added to the last bin. The lower panel shows the fraction of expected prompt background yields. The error bars on the data points and the dashed band on the simulated background show the statistical uncertainties.

and Figure C.5, respectively. In events with *tight* electrons and with *tight* muons, there is a deficit of expected events in the Z mass peak. This has been observed also in the control regions of the main analysis. Because the effect is isolated to the Z mass region, the $Z+$jets simulations are underestimating the fraction of events passing the b-jet selection. A non-prompt background contamination, that is not accounted for in the predictions for this region, would not indeed be resonantly distributed around the Z mass. Therefore, whenever $Z+$jets simulations have been used in the analysis to model the background in events with b-jets, the predictions have been scaled to correct the mis-modeling. In the $t\bar{t}$ non-prompt region, where the $Z+$jets simulation is used in the estimate of the charge misID background, the scaling is also applied.

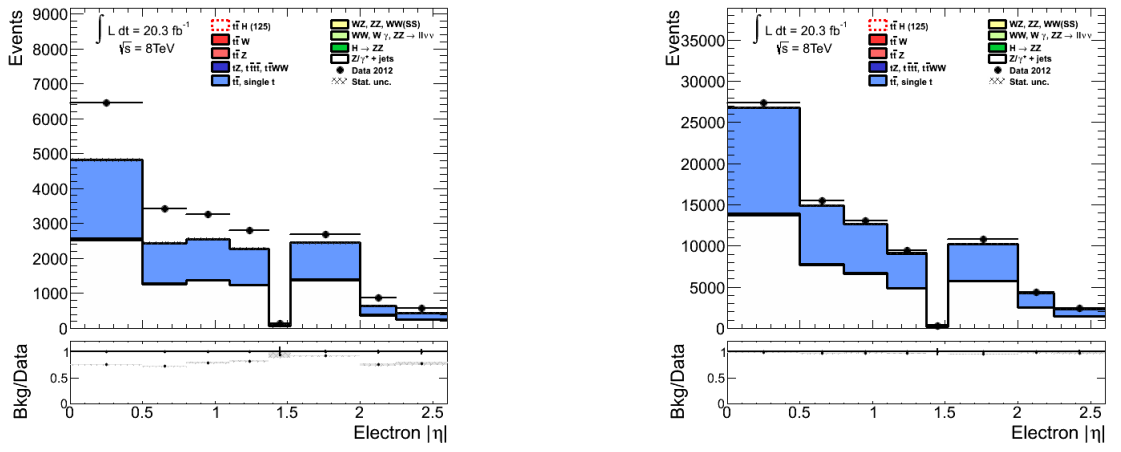
In the regions with *non-tight* leptons, the predictions are generally lower than the observations, in particular at lower lepton p_{T}. Since the predictions account only for prompt processes, the gap between data and predictions is due to non-prompt background unless the prompt production estimate is wrong. This has been verified by adding to the predictions the simulated processes in Table C.1 and allowing non-prompt leptons to be selected[2]. It is found that the gap is due almost totally to $t\bar{t}$ non-prompt background, and very marginally to $W^{\pm}+$jets non-prompt background.

The presence of a non-prompt background in the prompt regions affects the measurement of the r_{ℓ}. To evaluate the effect of the non-prompt background contamination and the $Z+$jets mis-modeling on the Matrix Method background estimate, the r_{ℓ} have been calculated on the prompt background predictions obtained from simulations instead that on data. The Matrix Method estimates of the non-prompt background obtained with r_{ℓ} from simulated prompt leptons differ by 4% in the $e^{\pm}e^{\pm}$ category, by 7% in the $e^{\pm}\mu^{\pm}$ category and by 10% in the $\mu^{\pm}\mu^{\pm}$ one from the estimates obtained with r_{ℓ} calculated in data. These uncertainties are negligible compared to those related to the f_{ℓ}, but in future studies they may be significant and may have to be considered.

C.2.3 Trigger biases

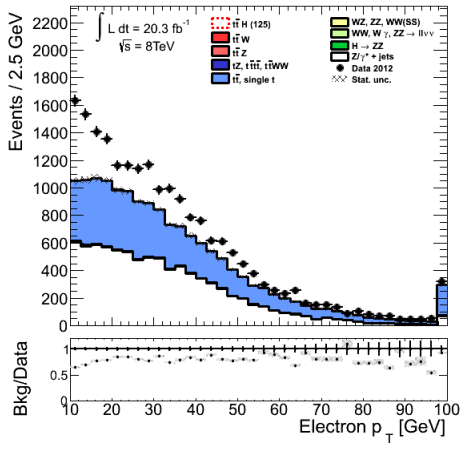
This section investigates if the condition of trigger activation by one of the leptons can alter the lepton probability to be selected as a *tight* lepton and, as a consequence, if can alter the non-prompt background estimate. The fake factors and the real factors have been calculated separately for leptons that activate the SLTrig and for leptons that fail the activation. Figure C.6 and Figure C.7 compare ,respectively, the fake factors and the real factors as a function of p_{T} obtained from leptons activating or failing the activation of the single lepton triggers (SETrig condition for electrons and SMTrig condition for muons). Electron and muon factors are studied separately. Note that the leptons with $p_{\mathrm{T}} < 25$ GeV are below the trigger threshold, thus no factors are calculated for lepton activating the

[2]The non-prompt leptons in simulations are usually excluded using the generator information since the non-prompt background is already accounted for with another method.

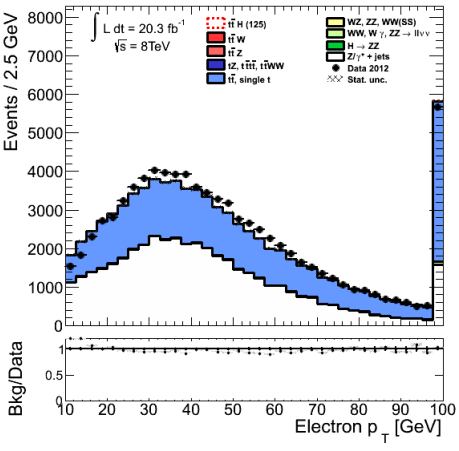


(a)

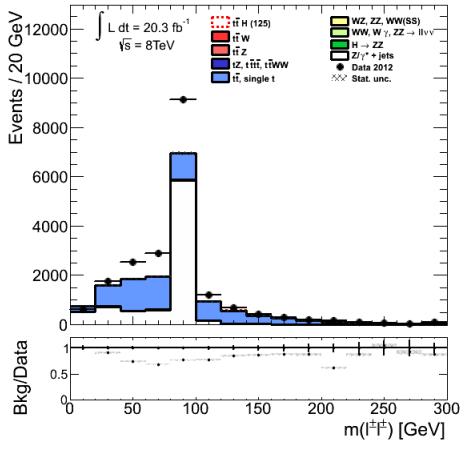
(b)



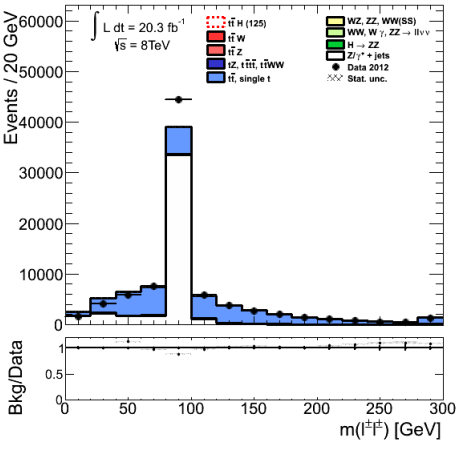
(c)



(d)

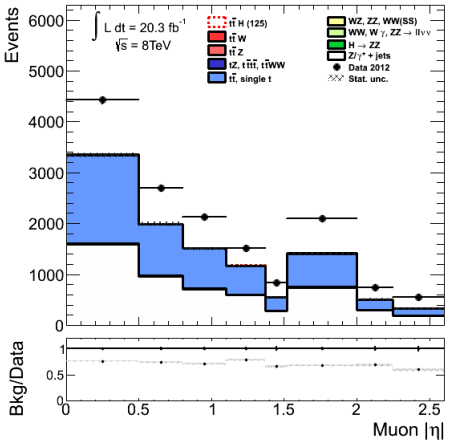


(e)

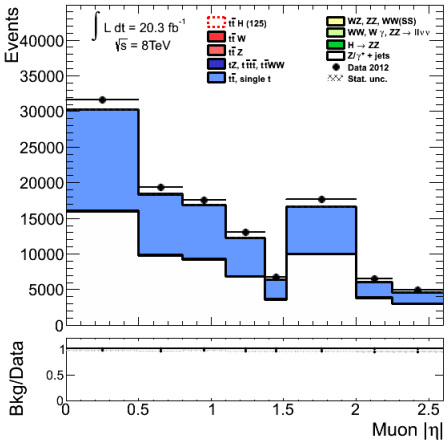


(f)

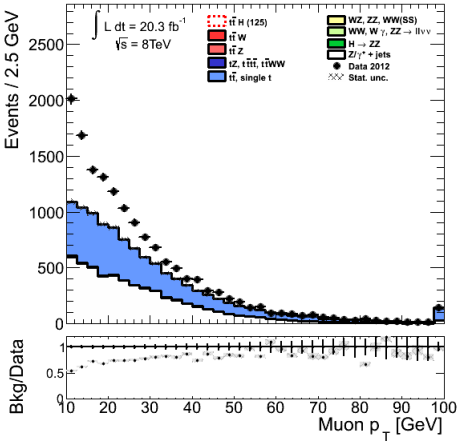
Figure C.4: Distributions of the electron $|\eta|$ (top), the electron p_{T} (centre) and $m\,(e^{\pm},\ell^{\mp})$ (bottom) in events with a *loose* electron passing (right) and failing (left) the *tight* selection in the prompt region. The plot shows the distribution of the invariant mass of the two opposite-sign leptons in the event. Events in data (black dots) are compared to the expected stacked contributions from prompt Standard Model processes. Overflow events are added to the last bin. The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points and the dashed band on the simulated background show the statistical uncertainties.



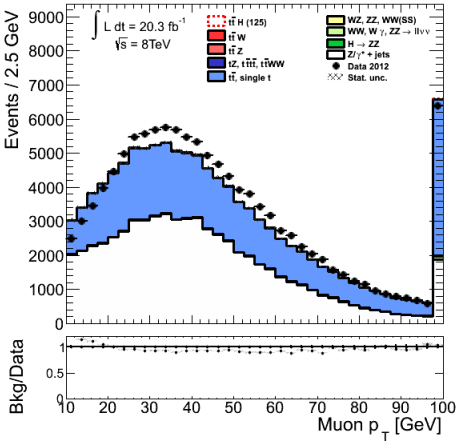
(a)



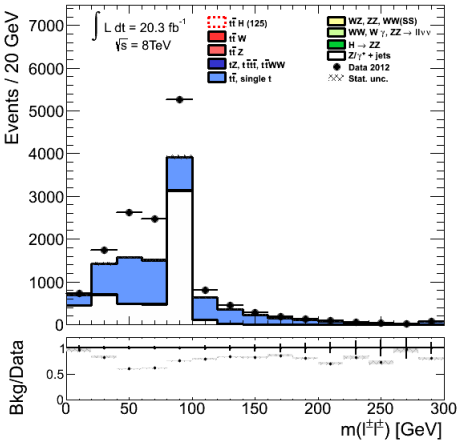
(b)



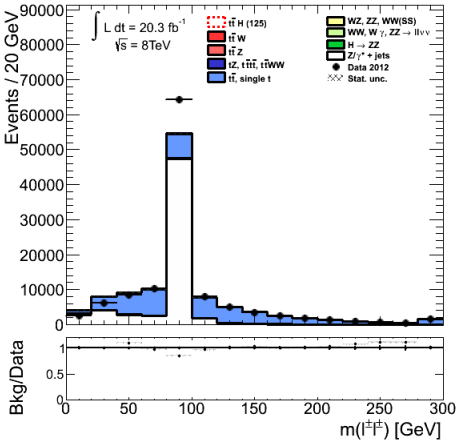
(c)



(d)

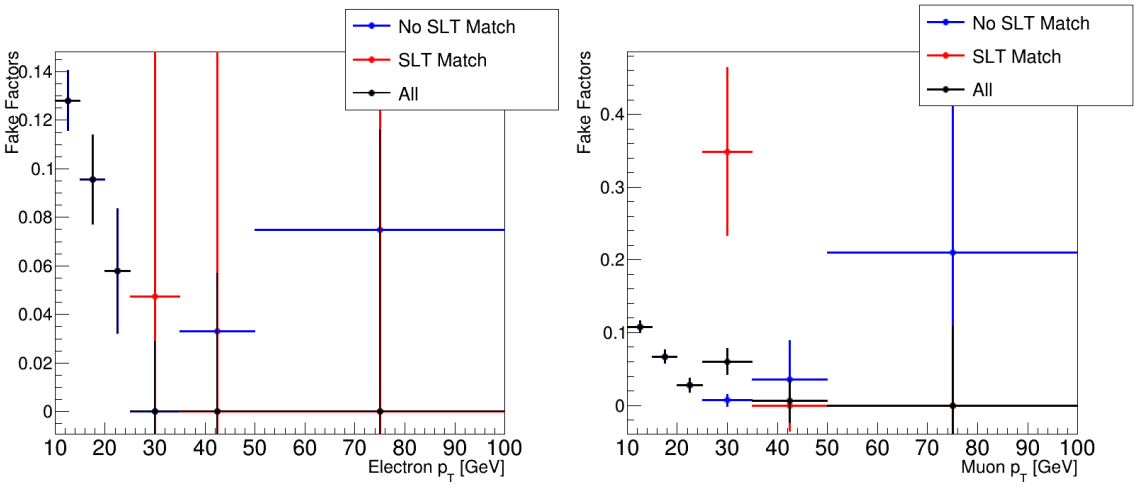


(e)



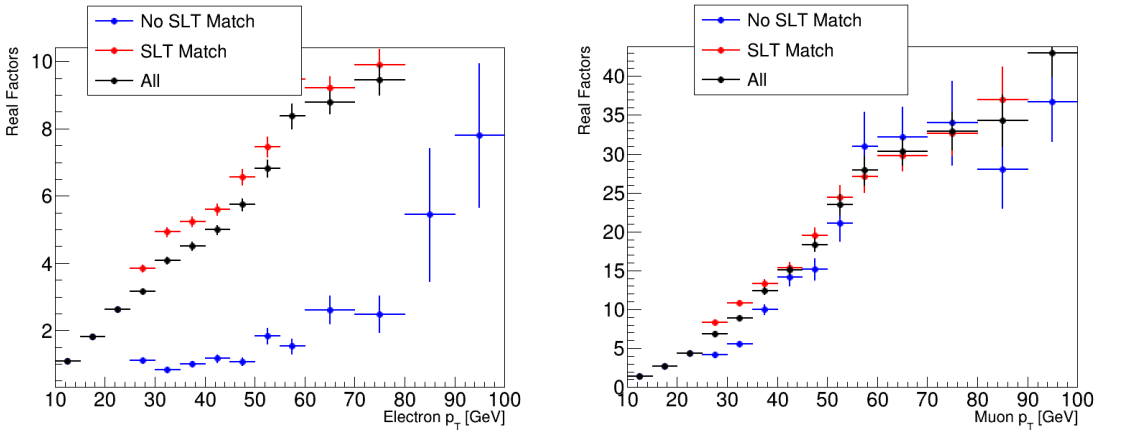
(f)

Figure C.5: Distributions of the muon $|\eta|$ (top), the muon p_{T} (centre) and $m\,(\mu^{\pm}, \ell^{\mp})$ (bottom) in events with a *loose* muon passing (right) and failing (left) the *tight* selection in the prompt region. The plot shows the distribution of the invariant mass of the two opposite-sign leptons in the event. Events in data (black dots) are compared to the expected stacked contributions from prompt Standard Model processes. Overflow events are added to the last bin. The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points and the dashed band on the simulated background show the statistical uncertainties.



(a) (b)

Figure C.6: Fake factors f_ℓ^{FF} as a function of the lepton p_T calculated on electrons (left) and muons (right) that activate the trigger (red), fail its activation (blue) and either activate it or not (black). The error bars account for the statistical uncertainty.



(a) (b)

Figure C.7: Real factors r_ℓ^{FF} as a function of the lepton p_T calculated on electrons (left) and muons (right) that activate the trigger (red), fail its activation (blue) and either activate it or not (black). The error bars account for the statistical uncertainty.

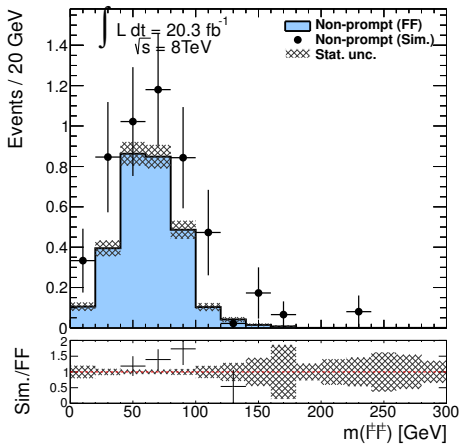
triggers. There are significant differences in the electron real factors and muon fake factors between $25 < p_{\mathrm{T}} < 35$ GeV. The statistics is too poor to conclude anything about muon fake factors for $p_{\mathrm{T}} > 35$ GeV and about electron fake factors. The muon real factors seem to not be particularly affected by the trigger condition. The differences observed can be justified by the isolation requirements that the leptons have to pass to activate the SLTrig. This condition alters the probability for a *loose* lepton to pass the *tight* selection, thus the measurements of f_{ℓ}^{FF} and r_{ℓ}^{FF} too. To avoid these biases in future analyses, any lepton selected with the *loose* requirement must satisfy also the requirements imposed by the trigger used to collect data at $\sqrt{s} = 13$ TeV. This condition would have been satisfied in this study collecting data with the DLTrig, whose conditions are looser than the requirements of the *loose* selection.

C.2.4 More on validation with $t\bar{t}$ simulations

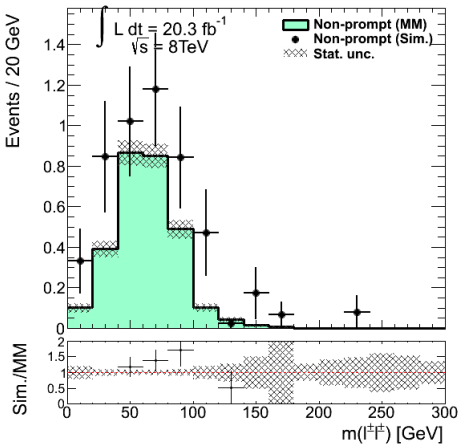
The Matrix Method has been applied to simulated $t\bar{t}$ non-prompt background and compared with the pure simulation prediction in Section 6.4.5.5. Here the predictions from the Fake Factor Method applied to $t\bar{t}$ simulations are also shown. The Matrix Method and the pure simulation predictions are also shown for comparison. Figure C.8 and Figure C.9 show respectively the distributions of the di-lepton invariant mass for events with $N\,(\mathrm{jet}) \leq 3$ and the di-lepton invariant mass for events with $N\,(\mathrm{jet}) \geq 4$ obtained from the simulation, the Matrix Method and the Fake Factor Method. The Matrix Method and Fake Factor Method predictions are almost identical and are both compatible with the pure simulation estimate.

C.2.5 More on validation with data

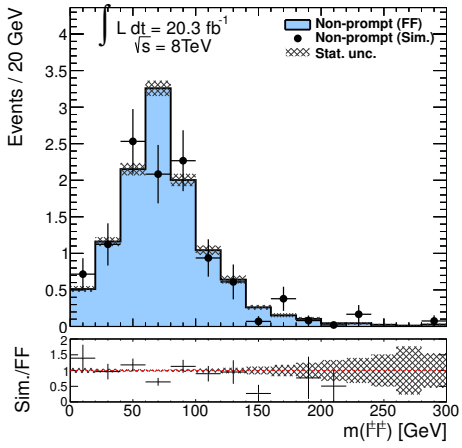
The predictions of the Matrix Method and the Fake Factor Method in the closure regions ($N\,(\mathrm{jet}) \leq 3$) for each same-sign lepton pair flavour combination are compared in Section 6.4.5.6. The predictions of the non-prompt background in the signal regions ($N\,(\mathrm{jet}) \geq 4$) are instead compared in Figure C.10(right). The observed data and the total Standard Model predictions are shown in Figure C.10(left). The effect of the systematic uncertainties associated with the fake factor and efficiency measurements on the jet multiplicity distribution of the total expected background is shown in Figure C.11.



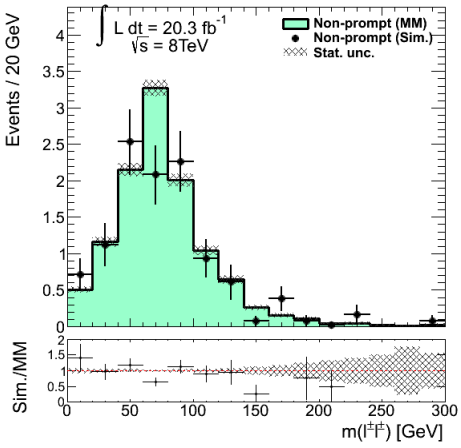
(a) $e^{\pm}e^{\pm}$



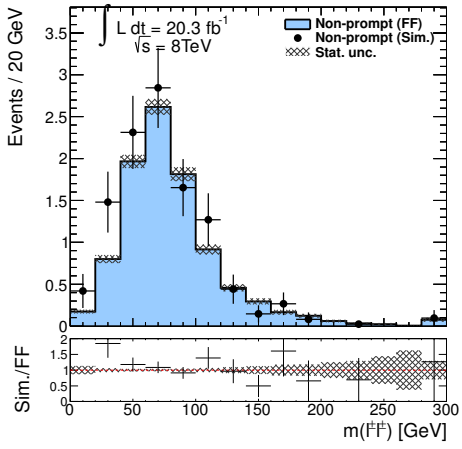
(b) $e^{\pm}e^{\pm}$



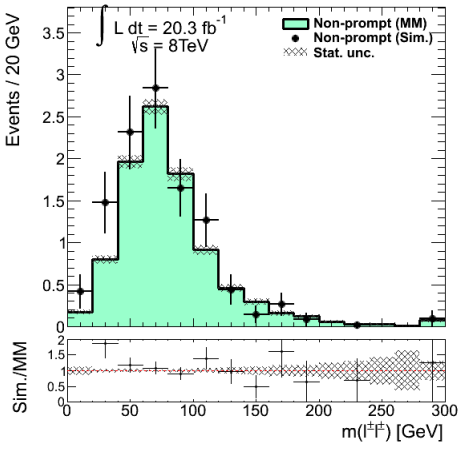
(c) $e^{\pm}\mu^{\pm}$



(d) $e^{\pm}\mu^{\pm}$

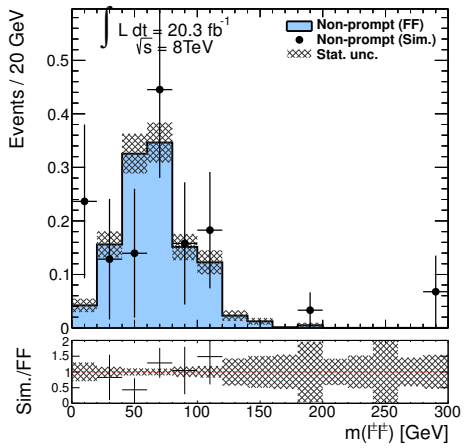


(e) $\mu^{\pm}\mu^{\pm}$

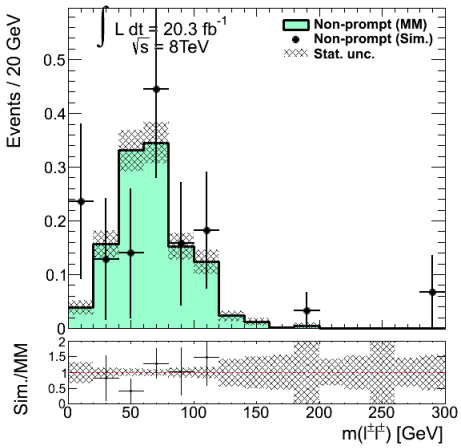


(f) $\mu^{\pm}\mu^{\pm}$

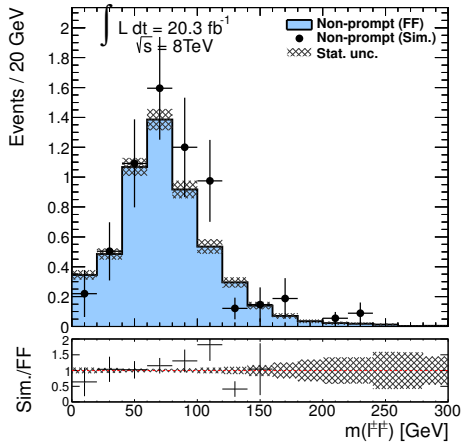
Figure C.8: Distributions of the di-lepton invariant mass for $e^{\pm}e^{\pm}$ (top), $e^{\pm}\mu^{\pm}$ (centre) and $\mu^{\pm}\mu^{\pm}$ (bottom) events in $N\,(\text{jet}) \leq 3$ region (closure region) in $t\bar{t}$ simulation. Simulated $t\bar{t}$ non-prompt events (black dots) are compared to the predictions from the Fake Factor Method applied to the $t\bar{t}$ simulated events (light blue) and from the Matrix Method applied to the same events (light green). Overflow events are added to the last bin. The lower panel shows the ratio between the predictions from simulation and from the data-driven methods. The error bars and the dashed band show the statistical uncertainties on simulation and on the data-driven predictions, respectively. Uncertainties on the fake factors and the rates are not included.



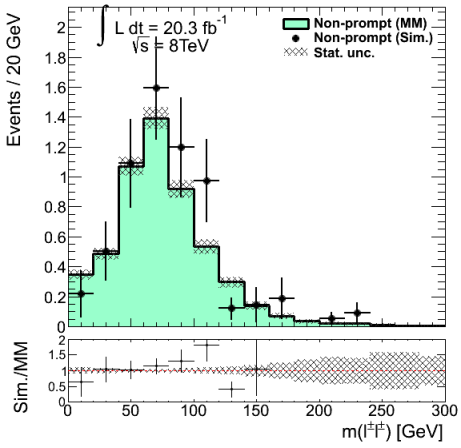
(a) $e^{\pm}e^{\pm}$



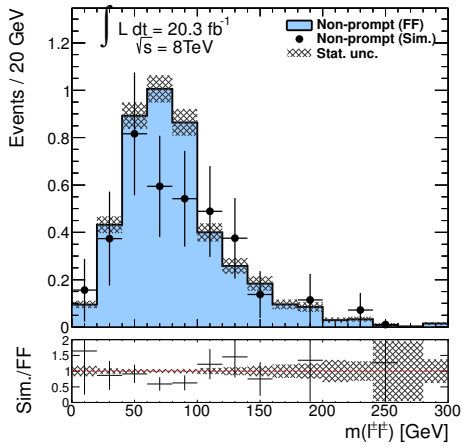
(b) $e^{\pm}e^{\pm}$



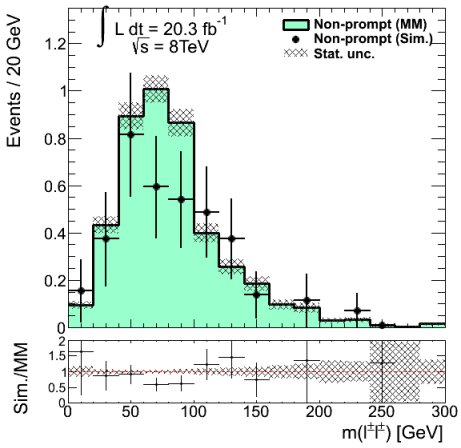
(c) $e^{\pm}\mu^{\pm}$



(d) $e^{\pm}\mu^{\pm}$

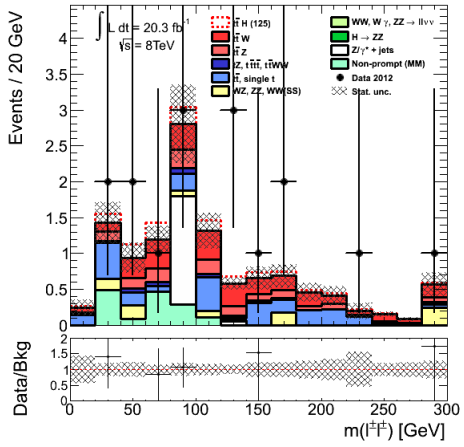


(e) $\mu^{\pm}\mu^{\pm}$

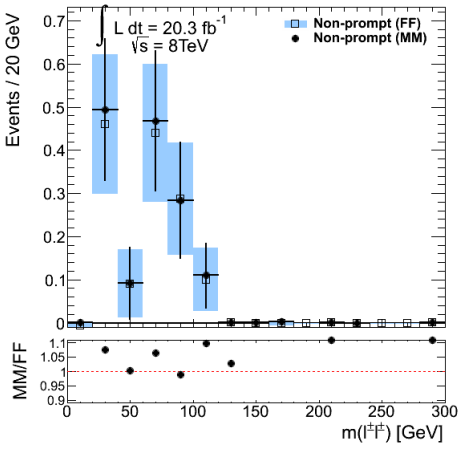


(f) $\mu^{\pm}\mu^{\pm}$

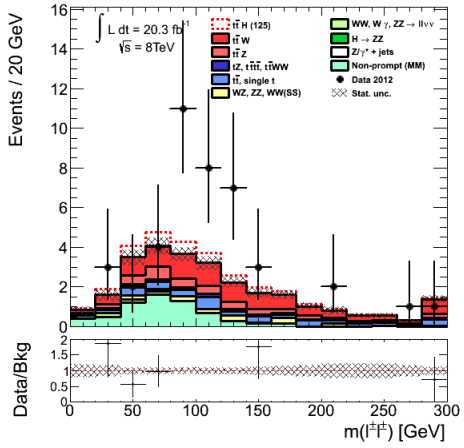
Figure C.9: Distributions of the di-lepton invariant mass for $e^{\pm}e^{\pm}$ (top), $e^{\pm}\mu^{\pm}$ (centre) and $\mu^{\pm}\mu^{\pm}$ (bottom) events in $N\,(\text{jet}) \geq 4$ region (signal regions) in $t\bar{t}$ simulation. Simulated $t\bar{t}$ non-prompt events (black dots) are compared to the predictions from the Fake Factor Method applied to the $t\bar{t}$ simulated events (light blue) and from the Matrix Method applied to the same events (light green). Overflow events are added to the last bin. The lower panel shows the ratio between the predictions from simulation and from the data-driven methods. The error bars and the dashed band show the statistical uncertainties on simulation and on the data-driven predictions, respectively. Uncertainties on the fake factors and the rates are not included.



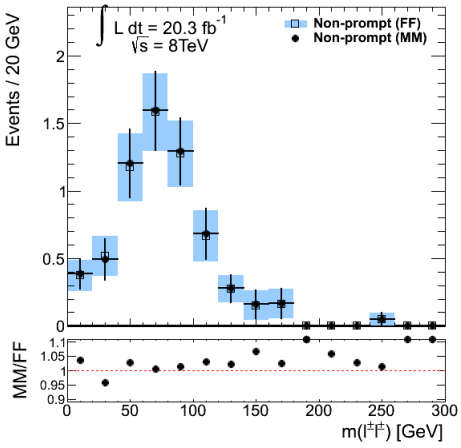
(a) $e^{\pm}e^{\pm}$



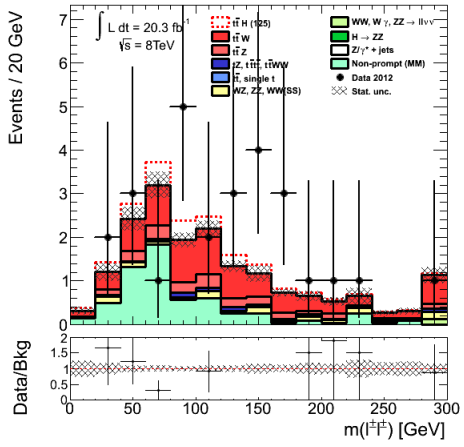
(b) $e^{\pm}e^{\pm}$



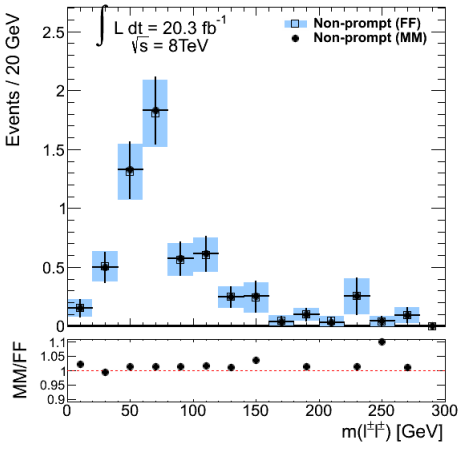
(c) $e^{\pm}\mu^{\pm}$



(d) $e^{\pm}\mu^{\pm}$

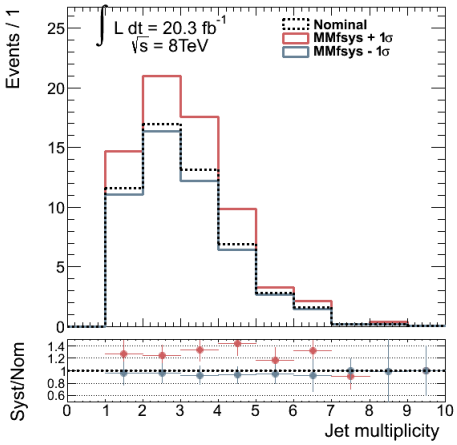


(e) $\mu^{\pm}\mu^{\pm}$

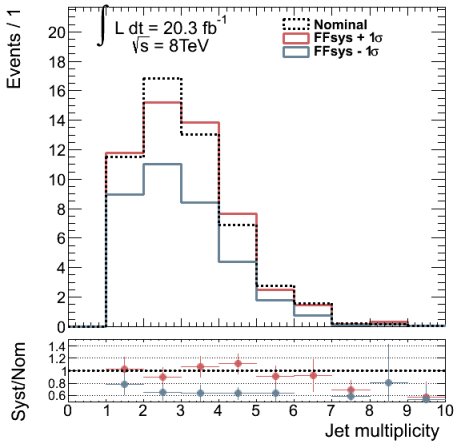


(f) $\mu^{\pm}\mu^{\pm}$

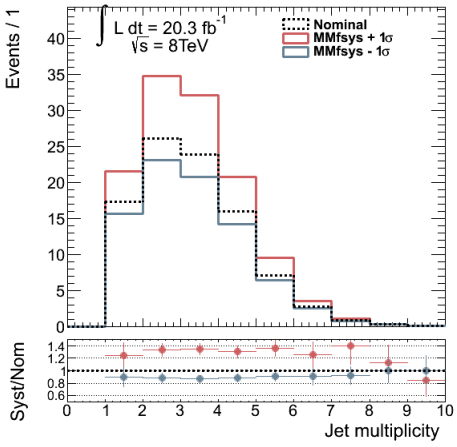
Figure C.10: Distributions of the di-lepton invariant mass for $e^{\pm}e^{\pm}$ (top), $e^{\pm}\mu^{\pm}$ (centre) and $\mu^{\pm}\mu^{\pm}$ (bottom) events in the signal regions (N (jet) ≥ 4). (left): Events in data (black dots) are compared to the expected stacked contributions from Standard Model processes with their ratio shown in the lower panel. The error bars and the dashed band show the statistical uncertainties on the data and the predictions, respectively. (right):Matrix Method predictions (black dots) are compared to Fake Factor Method predictions (open squares) with their ratio shown in the lower panel. The error bars and the light blue band show the statistical uncertainties on the Matrix Method and the Fake Factor Method predictions, respectively. In all the plots, overflow events are added to the last bin.



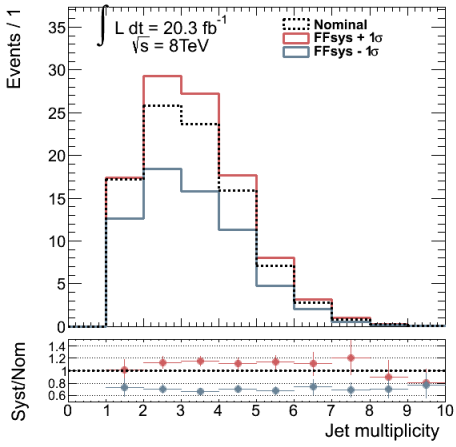
(a) $e^{\pm}e^{\pm}$



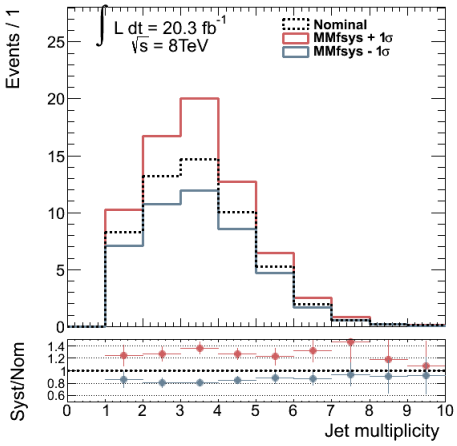
(b) $e^{\pm}e^{\pm}$



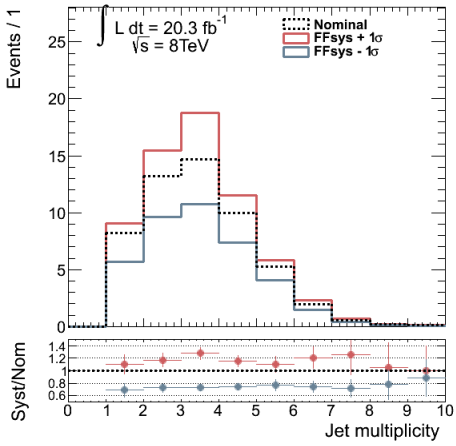
(c) $e^{\pm}\mu^{\pm}$



(d) $e^{\pm}\mu^{\pm}$



(e) $\mu^{\pm}\mu^{\pm}$



(f) $\mu^{\pm}\mu^{\pm}$

Figure C.11: Distributions of the total expected background events in $e^{\pm}e^{\pm}$ (top), $e^{\pm}\mu^{\pm}$ (centre) and $\mu^{\pm}\mu^{\pm}$ (bottom) events. The non-prompt background is predicted with the Matrix Method (Fake Factor Method) on the left (right) plots and the variation due to the increase and decrease of the f_{ℓ} (f_{ℓ}^{FF}) of their uncertainty is shown by the red and gray lines.

C.3 Definition of the other $t\bar{t}H$ signal regions

In addition to events with two same-sign leptons and no τ_{had}, other events have been used in the measurement of the $t\bar{t}H$ cross-section. The categorisation of these events has been described in the introduction of Chapter 6. This section describes the additional selection applied to each category to reduce the backgrounds. The description is taken from the published paper [3]. The summary of the selection applied to each category is given in Table C.2.

C.3.1 3ℓ category

Selected events are required to include exactly three light leptons with total charge equal to ± 1. Candidate events arising from non-prompt leptons overwhelmingly originate as opposite-sign prompt lepton events with one additional non-prompt lepton. As a result, the non-prompt lepton is generally one of the same-sign leptons. To reduce these backgrounds, a higher momentum threshold $p_{\text{T}} > 20$ GeV is applied to the same-sign leptons. No requirements are imposed on the number of τ_{had} candidates. In order to suppress the $t\bar{t}$ and $t\bar{t}V$ backgrounds, selected events are required to include either at least four jets of which at least one must be b-tagged, or exactly three jets of which at least two are b-tagged. To suppress the $t\bar{t}Z$ background, events that contain an opposite-sign same flavour lepton pair with the di-lepton invariant mass within 10 GeV of the Z mass are vetoed. Events containing an opposite-sign lepton pair with invariant mass below 12 GeV are also removed to suppress background from hadronic resonances that decay to light leptons.

C.3.2 $2\ell 1\tau_{\text{had}}$ category

Selected events are required to include exactly two same-sign leptons, one with $p_{\text{T}} > 25$ GeV and the other with $p_{\text{T}} > 15$ GeV, and in addition one τ_{had}. The reconstructed charge of the τ_{had} candidate has to be opposite to that of the leptons. In order to reduce $t\bar{t}$ and $t\bar{t}V$ backgrounds, events must include at least four jets. In order to suppress diboson and single boson backgrounds, at least one jet must be b-tagged. To suppress the $Z+$jets background, events with di-electron invariant mass within 10 GeV of the Z mass are vetoed.

C.3.3 4ℓ categories

Selected events are required to include exactly four light leptons with total charge equal to zero and leading (subleading) $p_{\text{T}} > 25$ (15) GeV. No requirements are applied on the number of τ_{had} candidates. In order to suppress the $t\bar{t}$ and $t\bar{t}V$ backgrounds, the selected events are required to include at least two jets of which at least one must be b-tagged. To suppress the $t\bar{t}Z$ background, events that contain an opposite-sign same

flavour lepton pair with di-lepton invariant mass within 10 GeV of the Z mass are vetoed. In order to suppress background contributions from resonances that decay to light leptons, all opposite-sign same flavour lepton pairs are required to have a di-lepton invariant mass greater than 10 GeV. The invariant mass of the four leptons is required to be between 100 and 500 GeV, which gives high acceptance for $t\bar{t}H$, where $H \to WW^{(*)} \to \ell\nu\ell\nu$, but rejects $Z \to 4\ell$ and high-mass $t\bar{t}Z$ events. Selected events are separated by the presence or absence of an opposite-sign same flavour lepton pair into two categories, referred to respectively as the Z-enriched and Z-depleted categories. In both cases the Z mass veto is applied, but background events in the Z-enriched category can arise from off-shell Z and $\gamma^* \to \ell^+\ell^-$ processes while in the Z-depleted category these backgrounds are absent.

C.3.4 $1\ell 2\tau_{\text{had}}$ category

Selected events are required to include exactly one light lepton with $p_{\text{T}} > 25$ GeV and exactly two τ_{had} candidates. The τ_{had} candidates must have opposite charge. In order to suppress the $t\bar{t}$ and $t\bar{t}V$ backgrounds, events must include at least three jets. In order to suppress diboson and single-boson backgrounds, at least one of the jets must be b-tagged. This final state is primarily sensitive to $H \to \tau^+\tau^-$ decays, allowing use of the invariant mass of the visible decay products of the $\tau_{\text{had}}\tau_{\text{had}}$ system (M_{vis}) as a signal discriminant. Signal events are required to satisfy $60 < M_{vis} < 120$GeV.

Category	Leptons		Jets	B-jets	τ_{had}
Common	$N(\ell) \geq 2$	≥ 1 trigger compatible lepton			
3ℓ	$N(\ell) = 3$	$p_{\text{T}}(SS\ \ell) > 20$ GeV $\Sigma Q_\ell = \pm 1$ $\lvert m_{\ell\ell}^{OS\ SF} - m_Z \rvert > 10$ GeV	$N(\text{jet}) \geq 4$ & $N(b\text{-jet}) \geq 1$ or $N(\text{jet}) = 3$ & $N(b\text{-jet}) \geq 2$		-
$2\ell 1\tau_{\text{had}}$	$N(\ell) = 2$	$\Sigma Q_\ell = \pm 2$ $p_{\text{T}} > 15$ GeV ≥ 1 lepton with $p_{\text{T}} > 25$ GeV $\lvert m(e^\pm, e^\pm) - m_Z \rvert > 10$ GeV	$N(\text{jet}) \geq 4$	$N(b\text{-jet}) \geq 1$	$N(\tau_{\text{had}}) = 1$ $Q_\tau = -Q_\ell$
$1\ell 2\tau_{\text{had}}$	$N(\ell) = 1$	$p_{\text{T}} > 25$ GeV	$N(\text{jet}) \geq 3$	$N(b\text{-jet}) \geq 1$	$N(\tau_{\text{had}}) = 2$ $\Sigma Q_\tau = 0$ $60 < M_{vis} < 120$ GeV
4ℓ	$N(\ell) = 4$	$\Sigma Q_\ell = 0$ $p_{\text{T}}^{lead} > 25$ GeV $p_{\text{T}}^{sub} > 15$ GeV $m_{\ell\ell}^{OS\ SF} \geq 10$ GeV $100 < m_{4\ell} < 500$ GeV $\lvert m_{\ell\ell}^{OS\ SF} - m_Z \rvert > 10$ GeV	$N(\text{jet}) \geq 2$	$N(b\text{-jet}) \geq 1$	-
		Z-depleted: 0 $OS\ SF$ lepton pairs			
		Z-enriched: ≥ 1 $OS\ SF$ lepton pairs			

Table C.2: Summary of event selections for the signal region categories other than the $2\ell 0\tau_{\text{had}}$. "$OS\ SF$" stands for an opposite sign same flavour lepton pair. $m_{\ell\ell}^{OS\ SF}$ is the invariant mass of two leptons with opposite charge sign and of the same flavour, $m_{4\ell}$ is the invariant mass of the four lepton system and M_{vis} is the two τ_{had} system visible mass.

C.4 Additional control distributions and validation regions

This section contains additional distributions from the validation regions described in Section 6.5. The validation regions targeting the background processes in the other analysis signal regions (non-$2\ell 0\tau_{\text{had}}$) are also shown here. For some validation regions, the H_{T} distribution is shown. H_{T} is defined as the scalar sum of the p_{T} of the leptons and jets in the event. Table C.3 defines all the validation regions of the analysis and specifies the Figures containing the distributions associated to the regions. The yields of the validation regions not presented in Section 6.5 are given in Table C.5 and Table C.4. Predictions are generally compatible with the observed events within uncertainties. The largest discrepancies are observed in the inclusive opposite-sign validation region especially in the E_{T}^{miss} distribution. It has to be noted however that the non-prompt background is not estimated in this region. The discrepancies are compatible with a non-prompt background contamination from $W^{\pm}$ + jets and QCD multi-jet productions. In any case, this discrepancy does not affect the signal region predictions.

Validation region	Selection and target of the validation
Inclusive opposite-sign (e^+e^-,$e^\pm\mu^\mp$,$\mu^+\mu^-$)	two opposite-sign leptons, leading lepton $p_{\mathrm{T}} > 25$ GeV, $m(\ell^+,\ell^-) > 40$ GeV Verify the corrections applied to leptons in simulated samples. Figure C.12 and Figure C.13
$t\bar{t}$ opposite-sign (e^+e^-,$e^\pm\mu^\mp$,$\mu^+\mu^-$)	Same as inclusive opposite-sign, but $N(\text{jet}) \geq 2$ and $N(b\text{-jet}) \geq 1$, $\lvert m_{\ell\ell}^{OS\ SF} - m_Z\rvert > 10$ GeV Check b-tag corrections for simulated samples, check composition of the data used to estimate charge misID background. Figure C.14
$t\bar{t}W^\pm$	Same as the $2\ell 0\tau_{\mathrm{had}}$ category selection but $N(b\text{-jet}) \geq 2$ and $N(\text{jet}) \leq 3$ Verify $t\bar{t}W^\pm$ normalization, modeling. Figure C.15
$t\bar{t}Z$	Same as the 3ℓ category selection (see Table C.2), but $\lvert m_{\ell\ell}^{OS\ SF} - m_Z\rvert < 10$ GeV Verify $t\bar{t}Z$ normalization, modeling. Figure C.19
$W^\pm Z$	Same as the 3ℓ category selection (see Table C.2), but $\lvert m_{\ell\ell}^{OS\ SF} - m_Z\rvert < 10$ GeV, $N(b\text{-jet}) = 0$ and any number of jets. Figure C.16 Verify $W^\pm Z$ normalization and $N(\text{jet})$ spectrum
$W^\pm\gamma^*$	Same as the 3ℓ category selection (see Table C.2), but $N(b\text{-jet}) = 0$ and any number of jets Verify $W^\pm\gamma^*$ modeling. Figure C.17
$W^\pm Z + b$-jet	Same as the 3ℓ category selection (see Table C.2), but $\lvert m_{\ell\ell}^{OS\ SF} - m_Z\rvert < 10$ GeV, $N(b\text{-jet}) \geq 1$ and any number of jets Verify $W^\pm Z$ modeling when heavy flavour hadrons are detected. Figure C.18
ZZ	Same as the 4ℓ category selection (see Table C.2), but require two pairs of opposite-sign same flavour leptons have $\lvert m_{\ell\ell}^{OS\ SF} - m_Z\rvert < 10$ GeV and any number of jets and bjets. Verify ZZ normalization and $N(\text{jet})$ spectra. Figure C.20 and Figure C.21
Inclusive 4ℓ	Same as the ZZ validation region but at most one opposite-sign same flavour lepton pair with $\lvert m_{\ell\ell}^{OS\ SF} - m_Z\rvert < 10$ GeV Verify non-resonant 4ℓ production modeling. Figure C.22 and Figure C.23

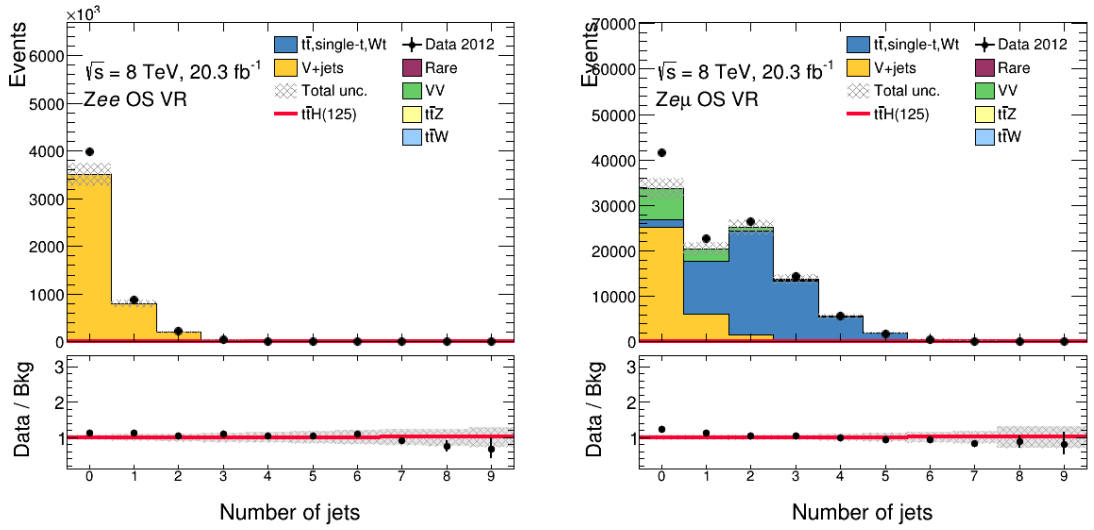
Table C.3: Definition of the validation regions designed to check the prediction of the backgrounds affecting the $t\bar{t}H$ analysis. The Figures showing the validation regions are also specified.

	Inclusive opposite-sign
V + jets	13390000 ± 850000
$t\bar{t}$, *single-top*, $tW^{\pm}$	115100 ± 7500
$t\bar{t}W^{\pm}$	150 ± 36
$t\bar{t}Z$	198 ± 62
VV	46800 ± 3000
Rare	153 ± 23
$t\bar{t}H$	70 ± 13
Tot. Background	13560000 ± 860000
Data	15047800
Data/Background	1.110 ± 0.070

Table C.4: Predicted and observed yields on the two lepton opposite-sign validation region inclusive in jet multiplicity. The yields from the e^+e^-, $e^{\pm}\mu^{\mp}$ and $\mu^+\mu^-$ final states are combined. The full systematic uncertainties on the predictions are shown. Statistical data uncertainty is not included in the ratios.

	ZZ	Inclusive 4ℓ
Non-prompt	0.0 ± 0.0	1.16 ± 0.51
$t\bar{t}W^{\pm}$	0.0 ± 0.0	0.0133 ± 0.0081
$t\bar{t}Z$	0.229 ± 0.077	4.6 ± 1.4
VV	117.4 ± 9.7	84.9 ± 7.5
Rare	0.0278 ± 0.0060	5.1 ± 2.3
$t\bar{t}H$	0.0035 ± 0.0012	0.461 ± 0.088
Tot. Background	117.7 ± 9.8	96 ± 10
Data	124	126
Data/Background	1.05 ± 0.13	1.32 ± 0.19

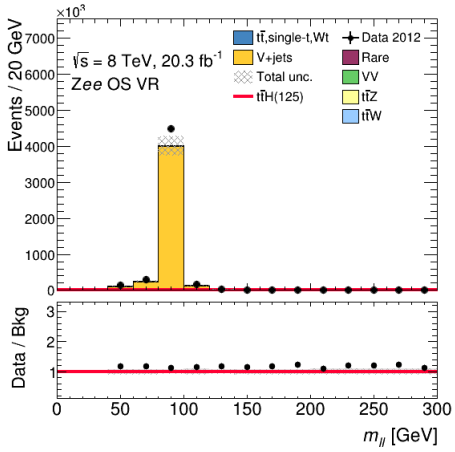
Table C.5: Predicted and observed yields on in the four lepton validation regions. The full systematic uncertainties on the predictions are shown. Statistical data uncertainty is not included in the ratios.



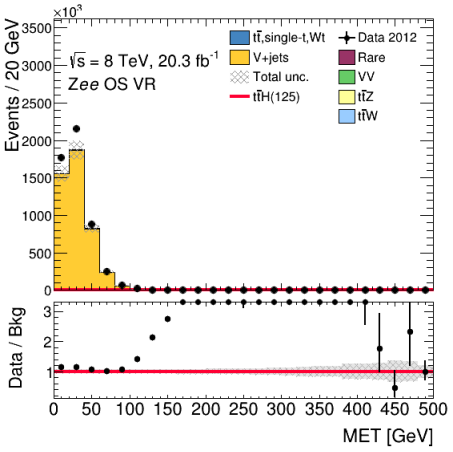
(a) e^+e^- (b) $e^\pm\mu^\mp$

(c) $\mu^+\mu^-$

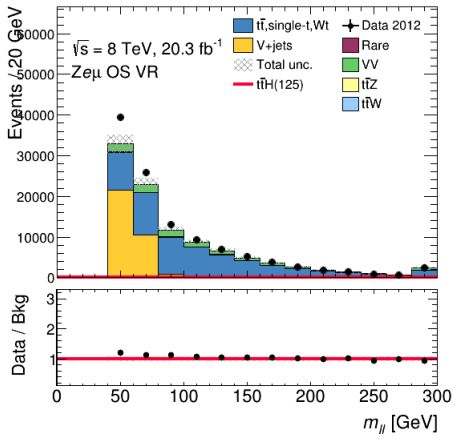
Figure C.12: Distributions of the jet multiplicity in the inclusive opposite-sign validation region for e^+e^- (a), $e^\pm\mu^\mp$ (b) and $\mu^+\mu^-$ (c) events. Events in data (black dots) are compared to the expected stacked contributions from prompt background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.



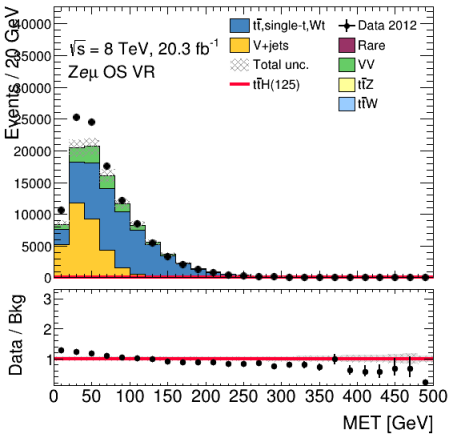
(a) e^+e^-



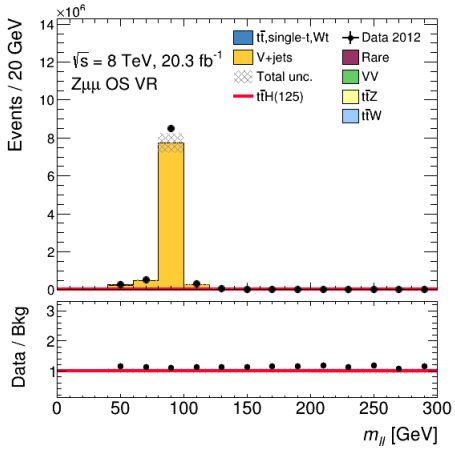
(b) e^+e^-



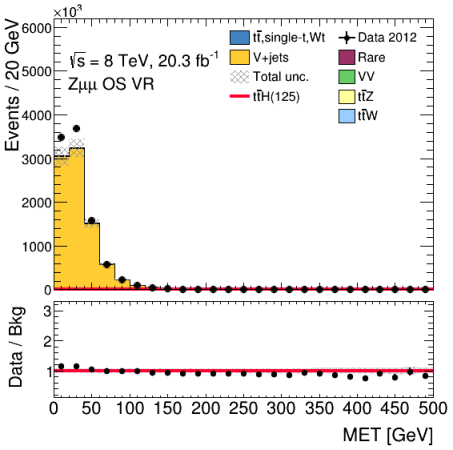
(c) $e^{\pm}\mu^{\mp}$



(d) $e^{\pm}\mu^{\mp}$

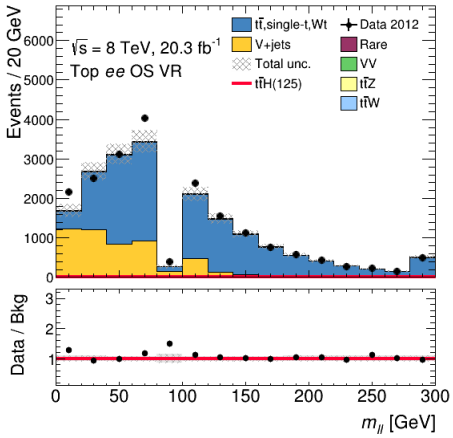


(e) $\mu^+\mu^-$

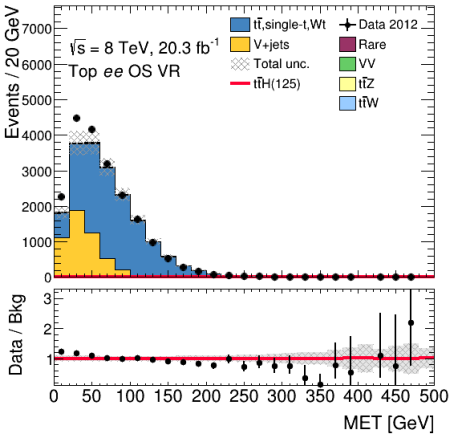


(f) $\mu^+\mu^-$

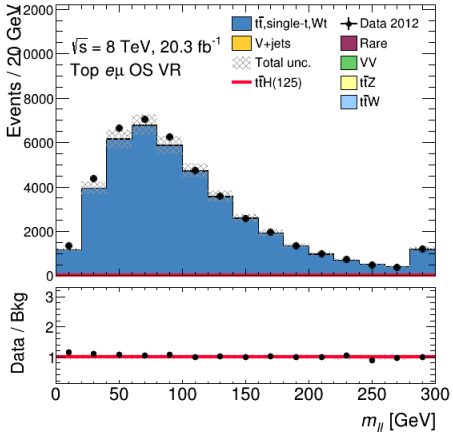
Figure C.13: Distributions of the di-lepton invariant mass (left) and E_{T}^{miss} (right) in the inclusive opposite-sign validation region for e^+e^- (top), $e^{\pm}\mu^{\mp}$ (centre) and $\mu^+\mu^-$ (bottom) events. Events in data (black dots) are compared to the expected stacked contributions from prompt background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.



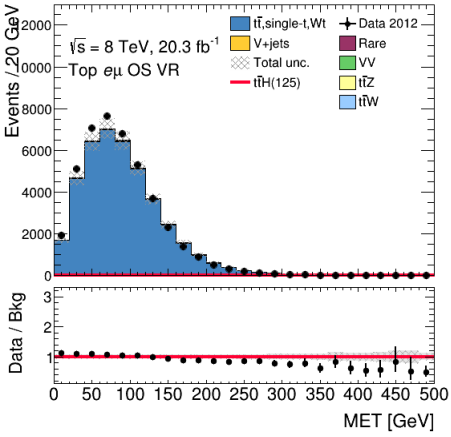
(a) e^+e^-



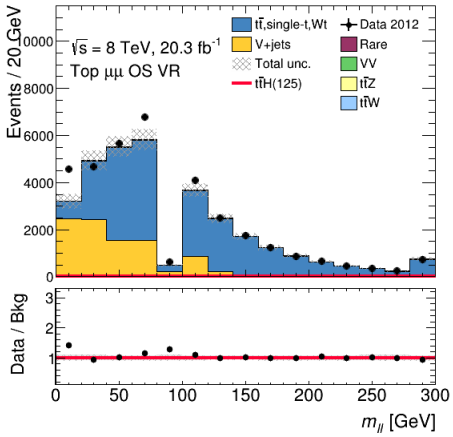
(b) e^+e^-



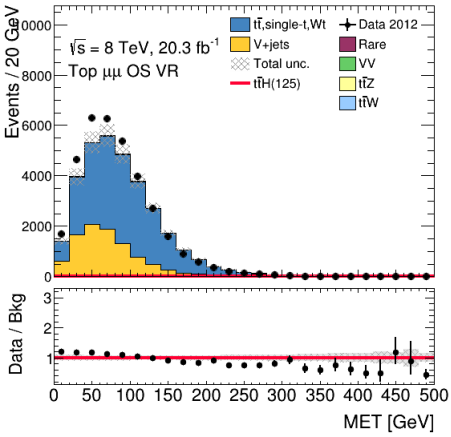
(c) $e^{\pm}\mu^{\mp}$



(d) $e^{\pm}\mu^{\mp}$

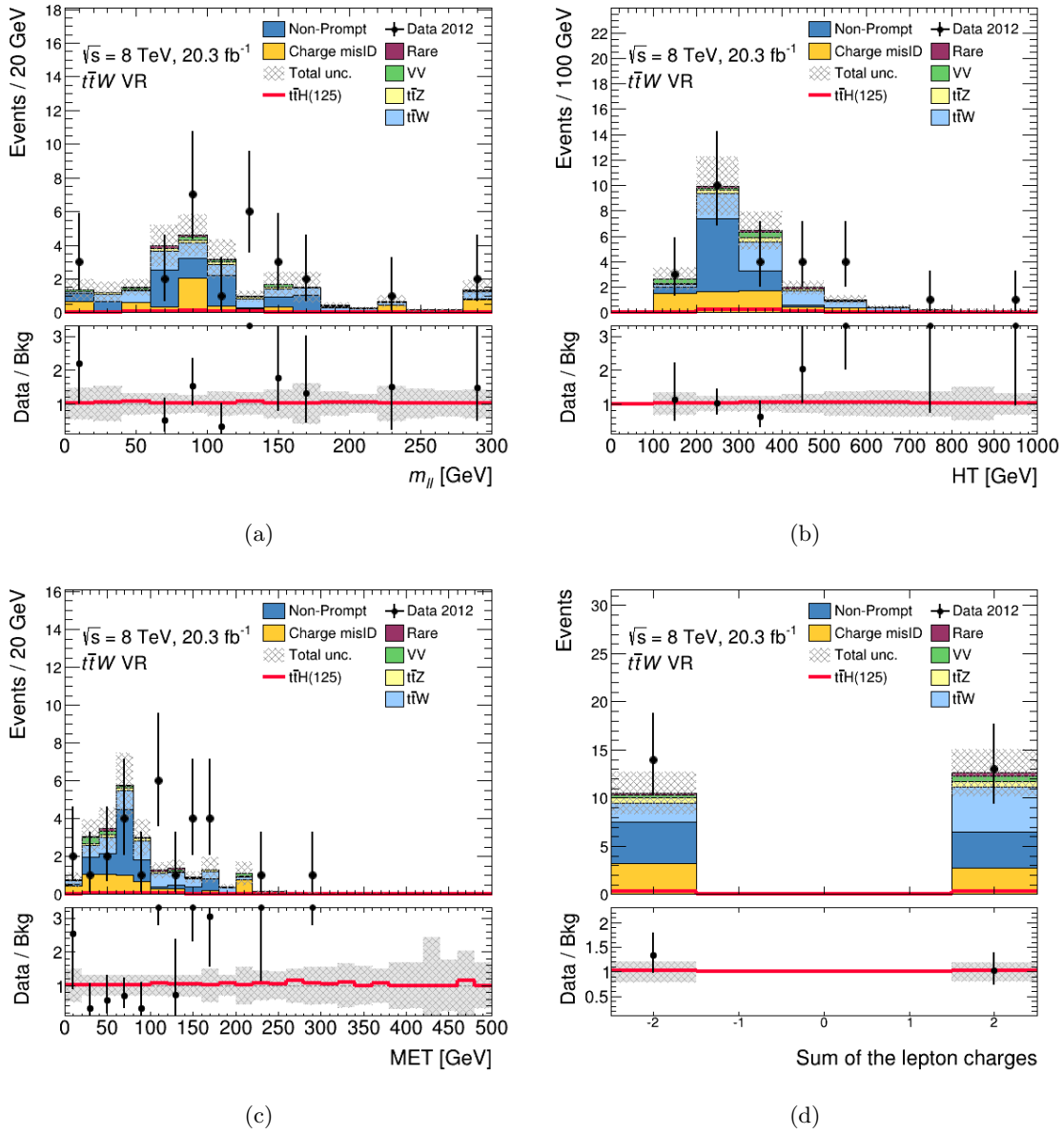


(e) $\mu^+\mu^-$



(f) $\mu^+\mu^-$

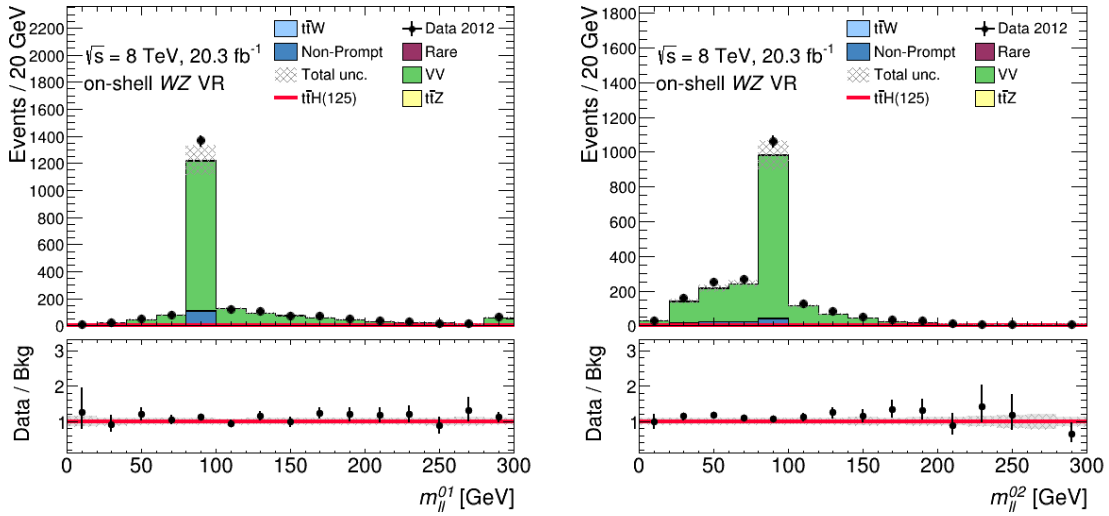
Figure C.14: Distributions of the di-lepton invariant mass (left) and E_{T}^{miss} (right) in the $t\bar{t}$ opposite-sign validation region for e^+e^- (top), $e^{\pm}\mu^{\mp}$ (centre) and $\mu^+\mu^-$ (bottom) events. Events in data (black dots) are compared to the expected stacked contributions from prompt background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.



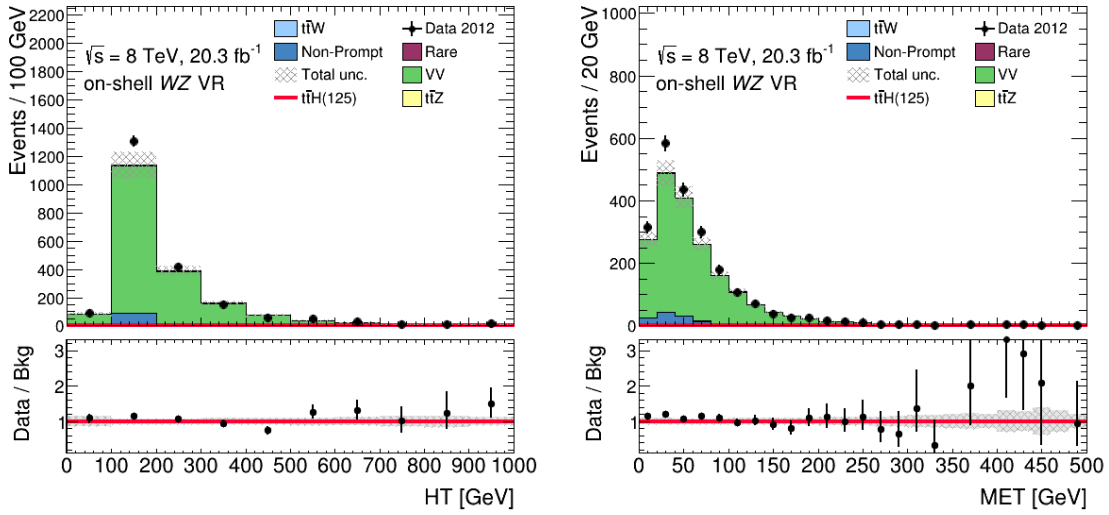
(a) (b)

(c) (d)

Figure C.15: Distributions of the di-lepton invariant mass (a), the H_{T} (b), the E_{T}^{miss} (c) and sum of the lepton charges (d) in the $t\bar{t}W^{\pm}$ validation region. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.



(a) (b)



(c) (d)

Figure C.16: Distributions of the two di-lepton opposite-sign invariant masses (a-b), the H_{T} (c) and the E_{T}^{miss} (d) in the $W^{\pm}Z$ validation region. The label 0 indicates the lepton with opposite charge with respect to the other leptons (1 and 2). The lepton 1 is the closest in ΔR to the lepton 0. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.

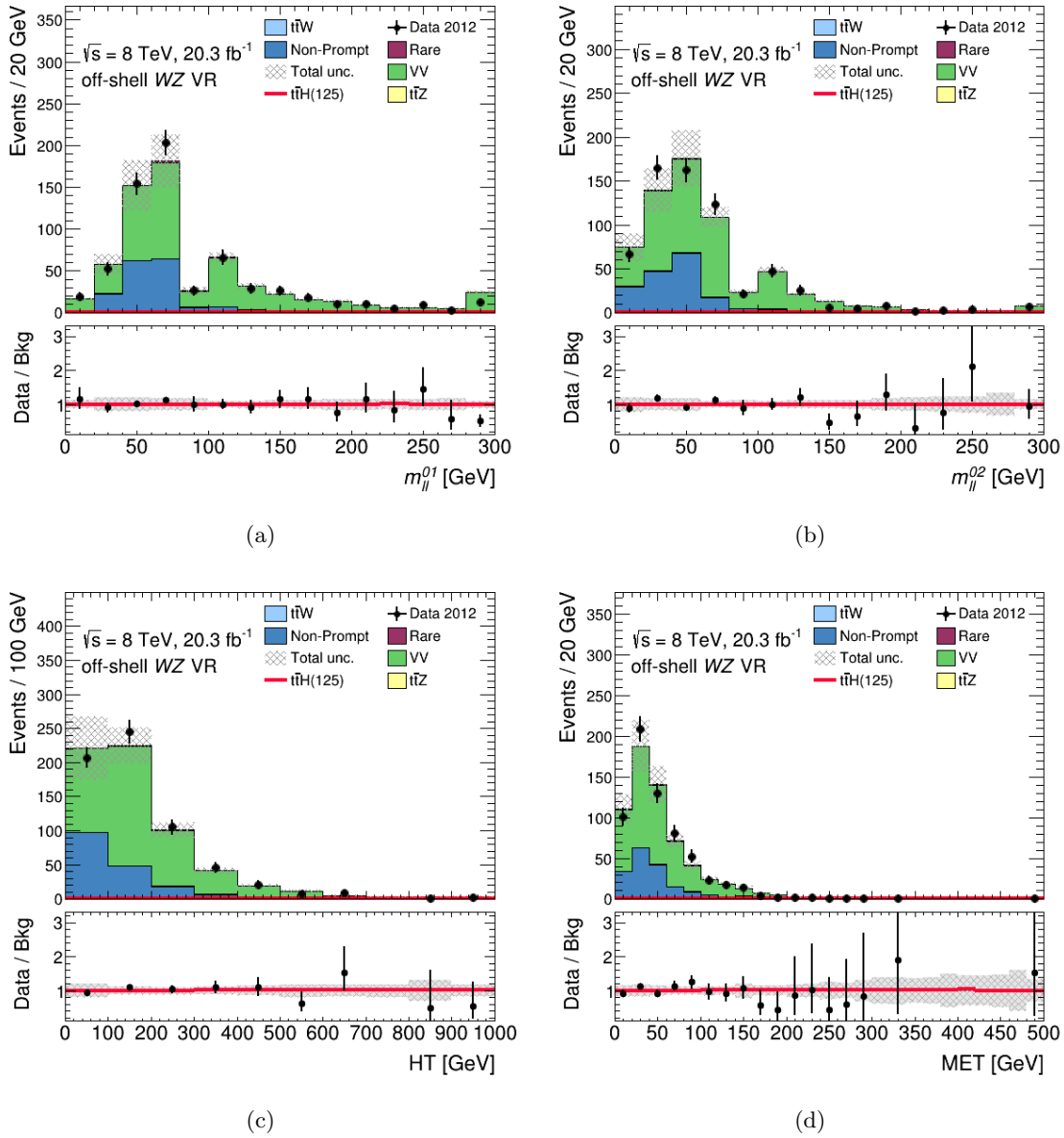
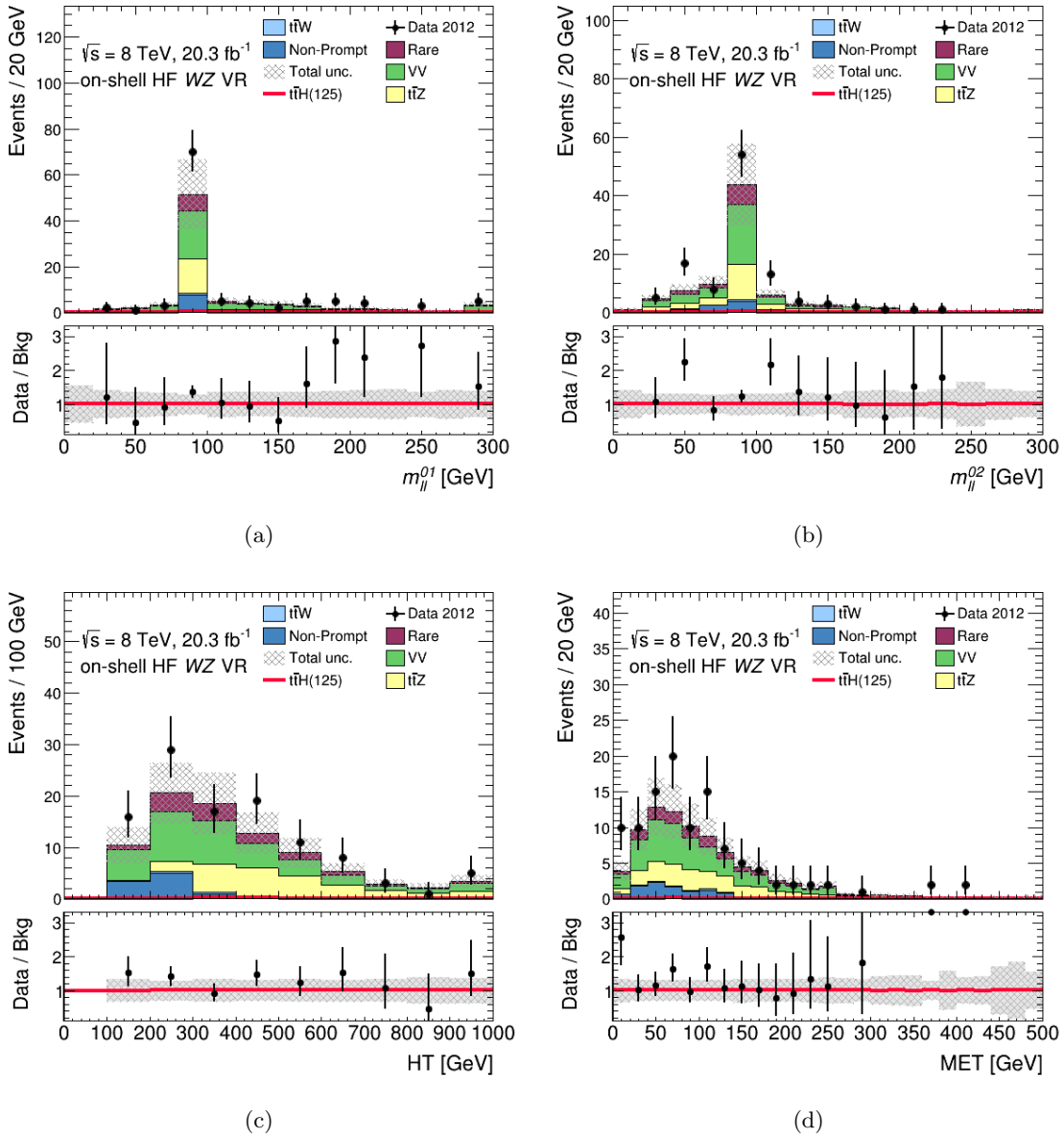


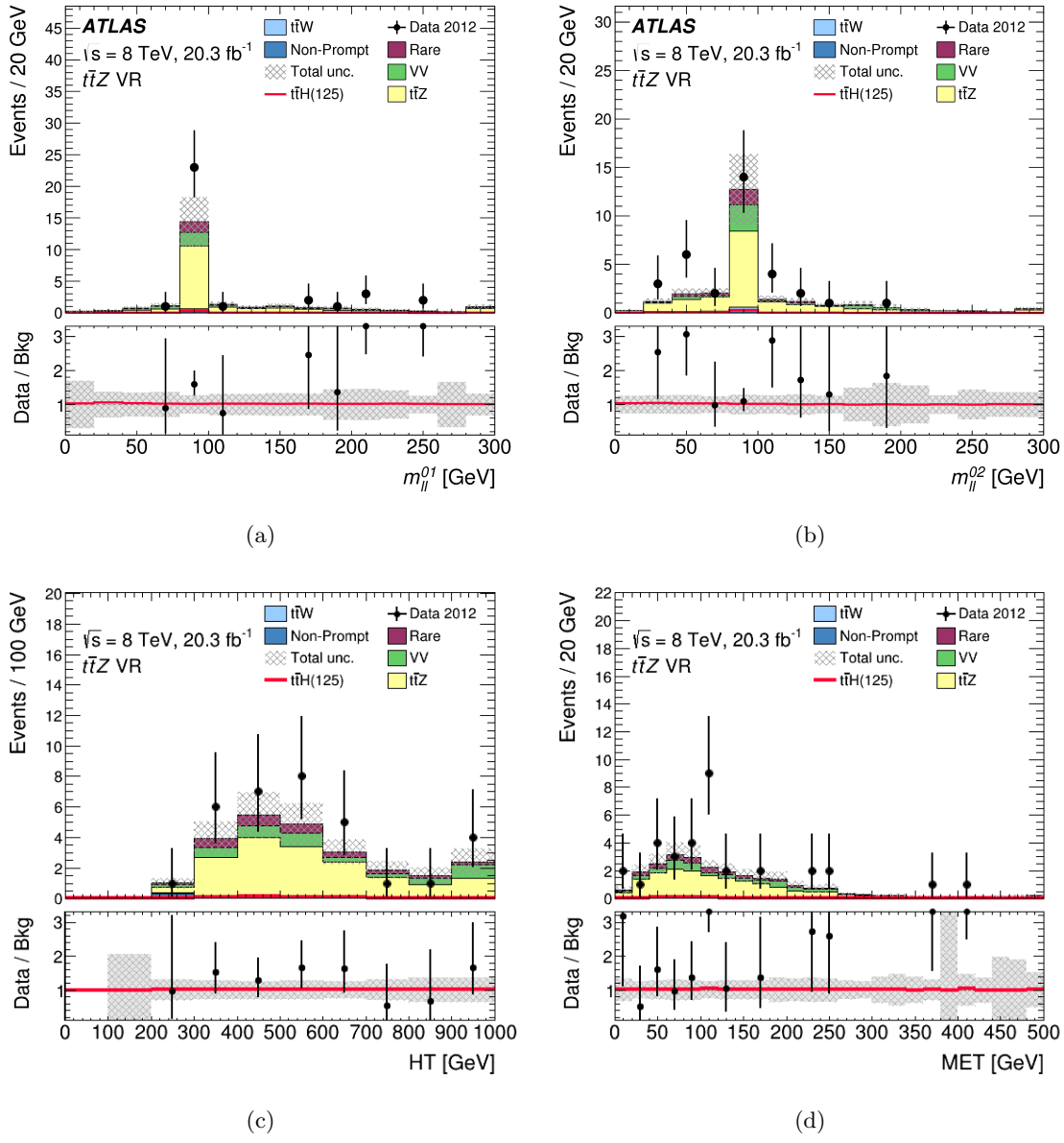
Figure C.17: Distributions of the two di-lepton opposite-sign invariant masses (a-b), the H_{T} (c) and the E_{T}^{miss} (d) in the $W^{\pm}\gamma^{*}$ validation region. The label 0 indicates the lepton with opposite charge with respect to the other leptons (1 and 2). The lepton 1 is the closest in ΔR to the lepton 0. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.



(a) (b)

(c) (d)

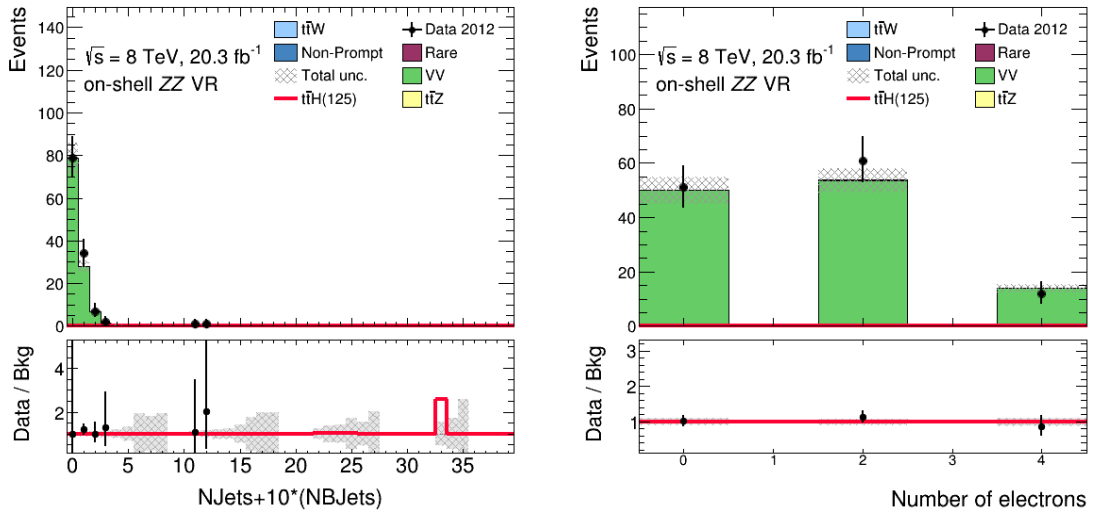
Figure C.18: Distributions of the two di-lepton opposite-sign invariant masses (a-b), the H_{T} (c) and the E_{T}^{miss} (d) in the $W^{\pm}Z + b$-jet validation region. The label 0 indicates the lepton with opposite charge with respect to the other leptons (1 and 2). The lepton 1 is the closest in ΔR to the lepton 0. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.



(a) (b)

(c) (d)

Figure C.19: Distributions of the two di-lepton opposite-sign invariant masses (a-b), the H_{T} (c) and the E_{T}^{miss} (d) in the $t\bar{t}Z$ validation region. The label 0 indicates the lepton with opposite charge with respect to the other leptons (1 and 2). The lepton 1 is the closest in ΔR to the lepton 0. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.



(a) (b)

Figure C.20: Distributions of the combined jet and b-jet multiplicities (left) and electron multiplicity (right) in the ZZ validation region. In case of $N\,(\text{jet}) > 9$ events, these are accounted in the $9 + 10 \cdot N\,(b\text{-jet})$ interval of the $N\,(\text{jet}) + 10 \cdot N\,(b\text{-jet})$ plot to avoid mixing with events with an additional b-jet. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.

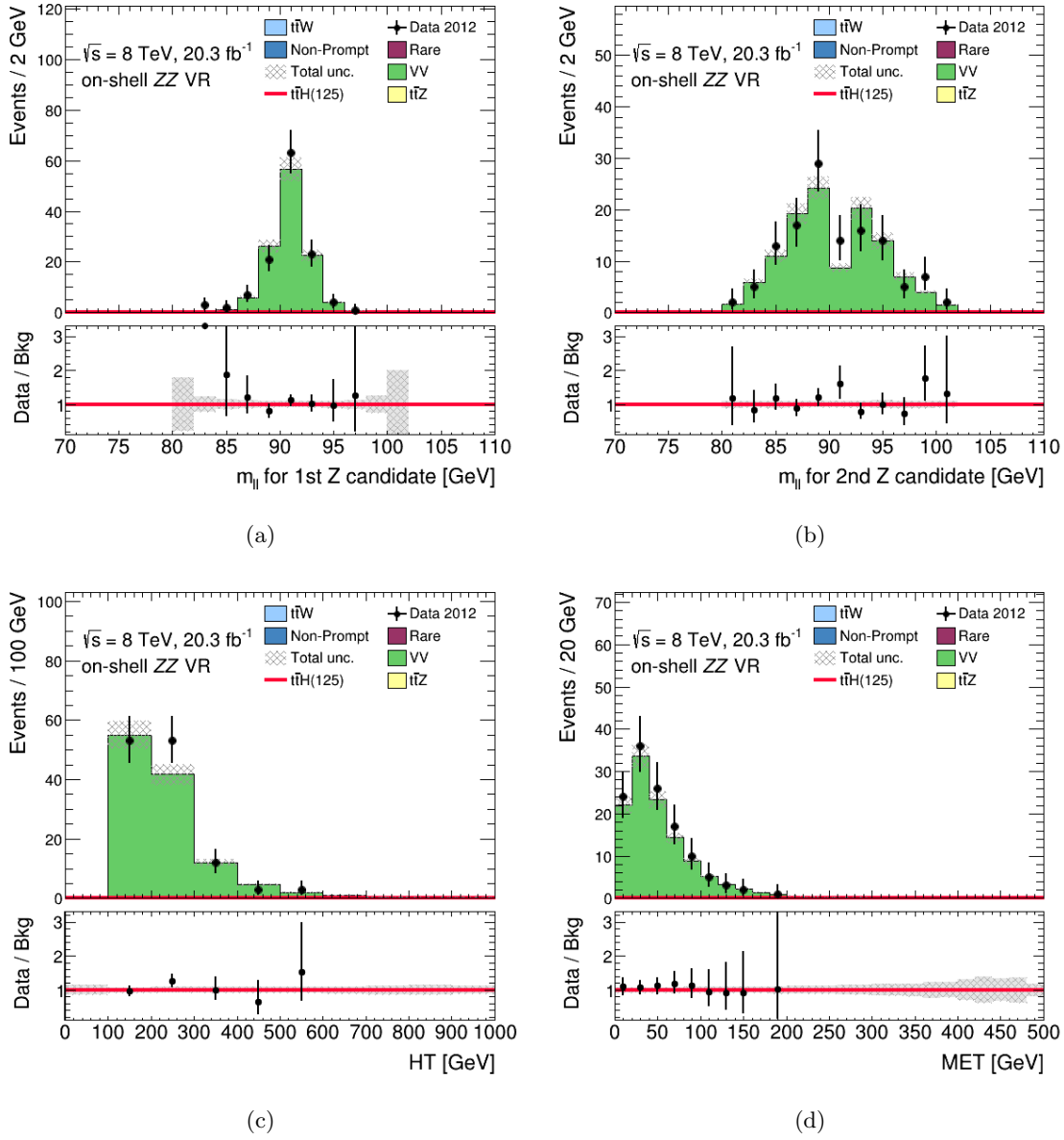
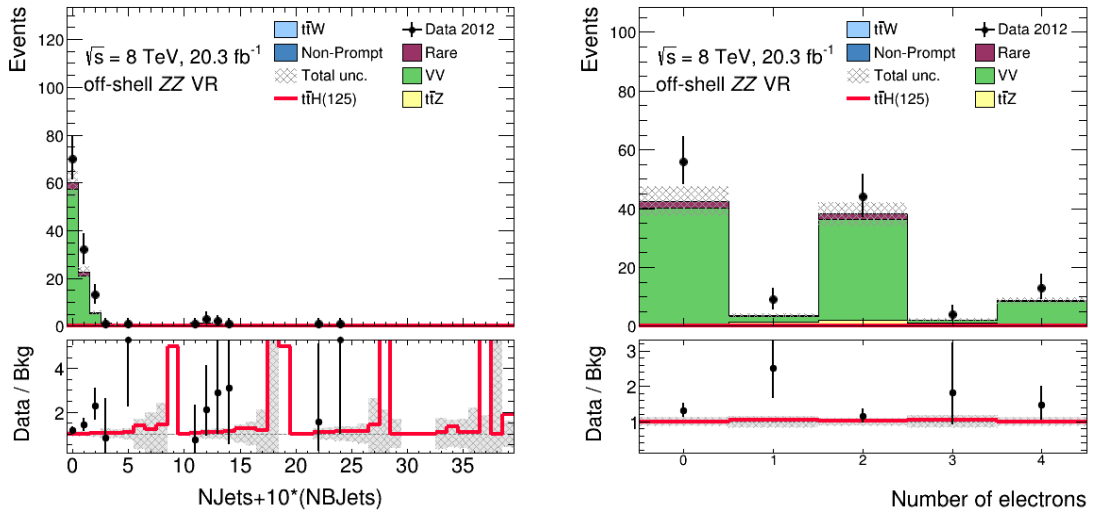
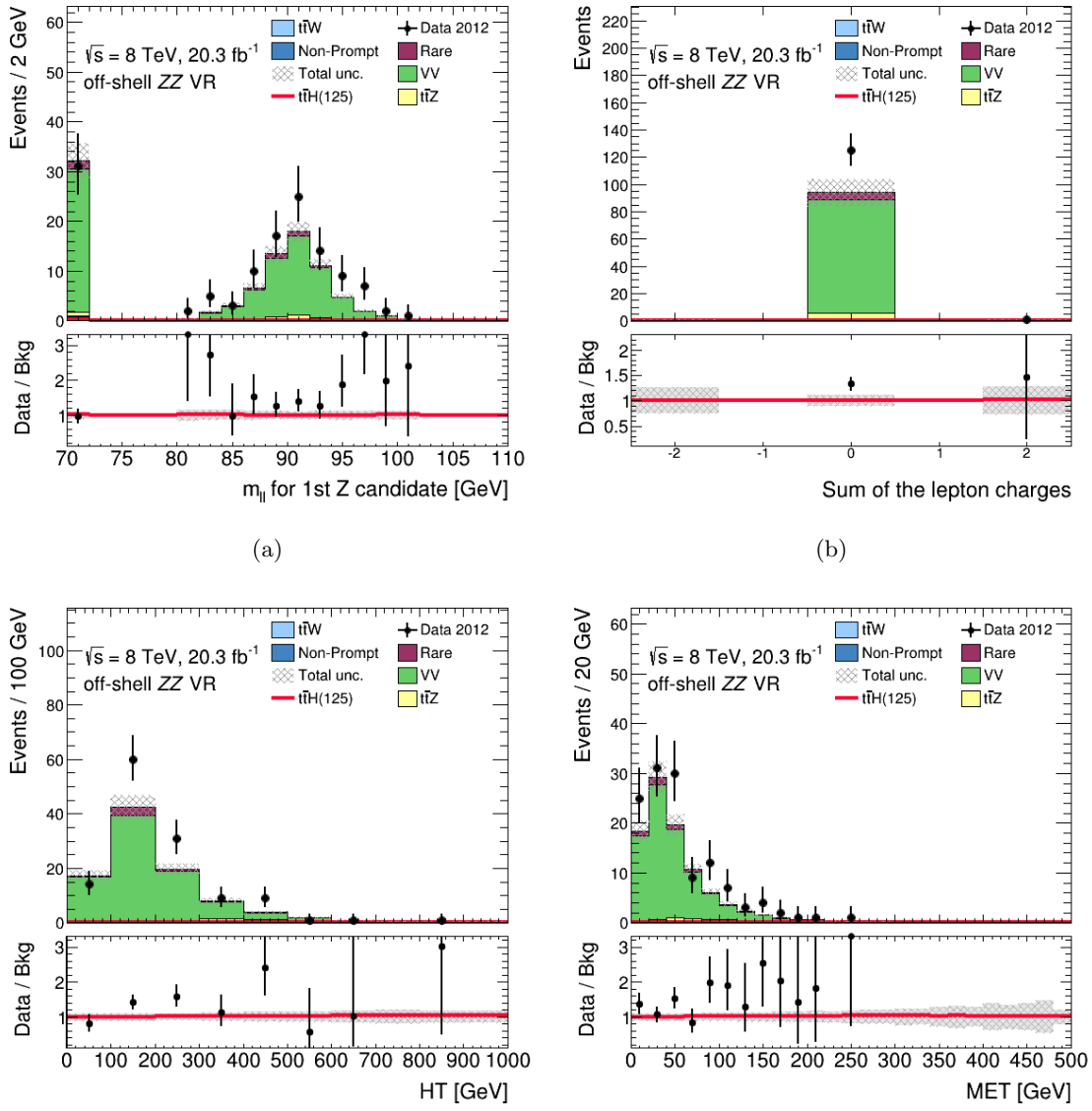


Figure C.21: Distributions of the two Z candidate lepton decay pairs invariant masses (a-b), the H_{T} (c) and the E_{T}^{miss} (d) in the ZZ validation region. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.



(a) (b)

Figure C.22: Distributions of the combined jet and b-jet multiplicities (left) and electron multiplicity (right) in the 4ℓ validation region. In case of $N\,(\text{jet}) > 9$ events, these are accounted in the $9 + 10 \cdot N\,(b\text{-jet})$ interval of the $N\,(\text{jet}) + 10 \cdot N\,(b\text{-jet})$ plot to avoid mixing with events with an additional b-jet. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.



(a) (b)

(c) (d)

Figure C.23: Distributions of the Z candidate lepton decay pair invariant mass (a), the sum of the lepton charges (b), the H_{T} (c) and the E_{T}^{miss} (d) in the 4ℓ validation region. If no Z candidate lepton decay pair is found the invariant mass is set arbitrarily to 70 GeV. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.

C.5 Systematic uncertainties for the other analysis categories

The largest experimental sources of systematic on the $t\bar{t}H$, $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ predictions in the signal regions described in Section C.3 are shown in Table C.6, Table C.7 and Table C.8.

Category	1^{st} Unc. Down-Up(%)	2^{nd} Unc. Down-Up(%)	3^{rd} Unc. Down-Up(%)
$1\ell 2\tau_{\text{had}}$	Tau ID	Tau Calibration	Trigger
	-6.6 6.8	-3.6 4.1	-1.3 1.3
$2\ell 1\tau_{\text{had}}$	Tau Calibration	Tau ID	JES Flav. Comp.
	-3.2 3.2	-2.7 2.7	-2.3 0.9
3ℓ	Muon Isolation	El. Isolation	JES Method Modeling 1
	-3.4 3.5	-2.6 2.7	-1.9 1.5
4ℓ	Muon Isolation	El. Isolation	El. ID
	-4.7 4.9	-3.5 3.6	-1.8 1.8

Table C.6: The three most relevant experimental sources of uncertainty on the $t\bar{t}H$ expected yields in the analysis categories.

Category	1^{st} Unc. Down-Up(%)	2^{nd} Unc. Down-Up(%)	3^{rd} Unc. Down-Up(%)
$1\ell 2\tau_{\text{had}}$	Tau Calibration	Tau ID	JES Flav. Comp.
	-5.0 8.6	-4.1 4.2	-4.0 0.1
$2\ell 1\tau_{\text{had}}$	Tau Calibration	Muon Isolation	JES pileup NPV
	-6.4 2.0	-2.3 2.3	0.2 3.2
3ℓ	Muon Isolation	b-tag eff. b-quark 5	El. Isolation
	-3.4 -3.5	-2.9 3.0	-2.9 2.9

Table C.7: The three most relevant experimental sources of uncertainty on the $t\bar{t}W^{\pm}$ expected yields in the analysis categories.

Category	1^{st} Unc. Down-Up(%)	2^{nd} Unc. Down-Up(%)	3^{rd} Unc. Down-Up(%)
$1\ell 2\tau_{\text{had}}$	Tau ID	Tau Calibration	JES Method Modeling 1
	-6.5 6.7	-3.5 4.9	-1.4 1.7
$2\ell 1\tau_{\text{had}}$	Muon Isolation	JES Flav. Comp.	JES Method Modeling 1
	-2.3 2.3	-3.0 2.0	-2.7 1.9
3ℓ	Muon Isolation	El. Isolation	JES Method Modeling 1
	-3.6 3.7	-2.7 2.8	-1.8 2.5
4ℓ	JER	Muon Isolation	El. Isolation
	— -5.2	-4.8 5.0	-3.8 3.9

Table C.8: The three most relevant experimental sources of uncertainty on the $t\bar{t}Z$ expected yields in the analysis categories.

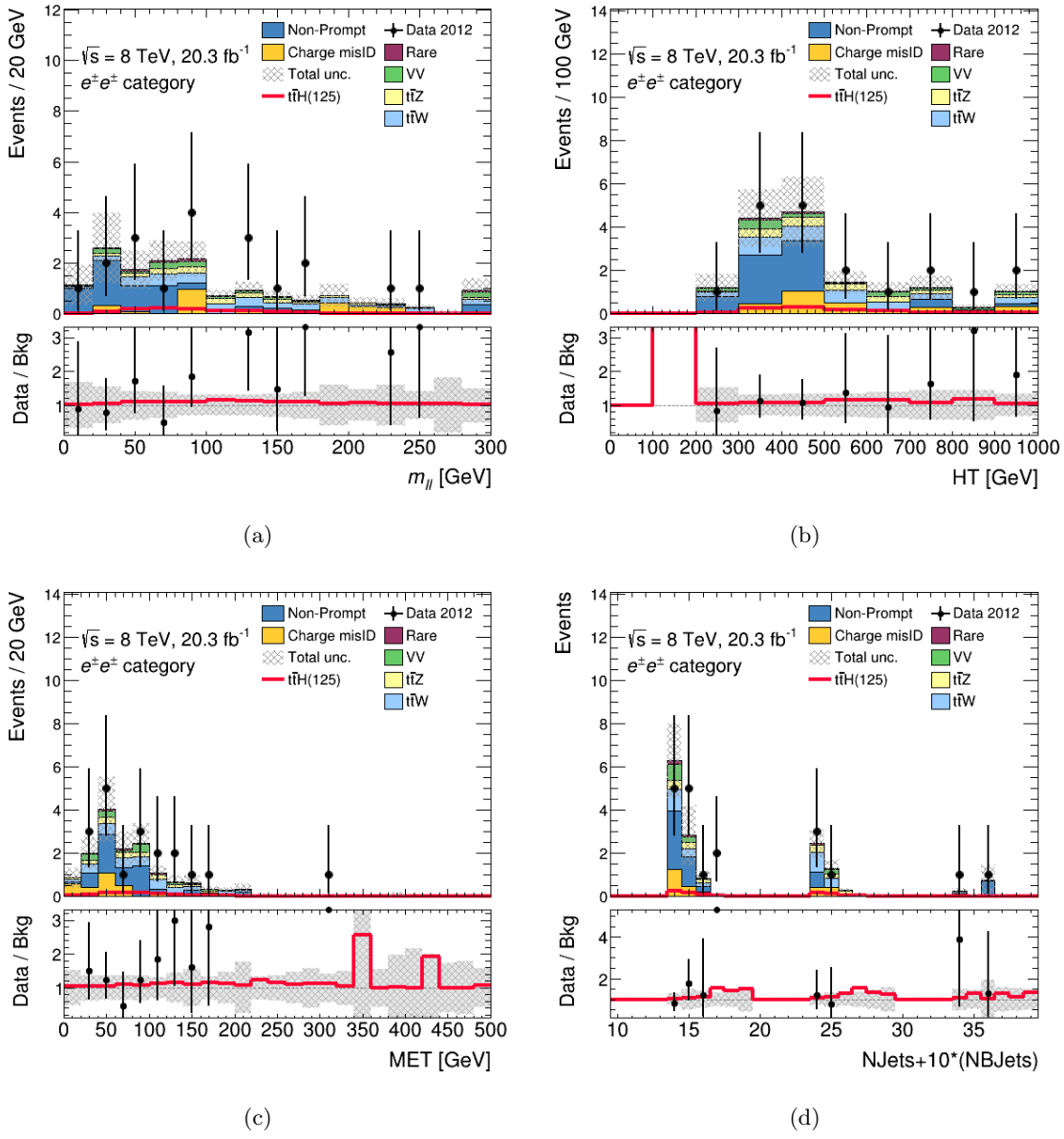
Total Systematic Uncertainty	$1\ell 2\tau_{\text{had}}$ Down-Up		$2\ell 1\tau_{\text{had}}$ Down-Up		3ℓ Down-Up		4ℓ Down-Up	
ttH	-7.9	8.4	-6.4	5.9	-5.8	5.6	-6.5	6.5
ttW	-9.3	10.0	-7.5	7.8	6.4	8.2	—	—
ttZ	-8.1	9.2	-7.5	7.2	-6.1	6.7	-9.6	6.9

Table C.9: Sum in quadrature of all the experimental systematic uncertainties on the number of $t\bar{t}H$, $t\bar{t}W^{\pm}$ and $t\bar{t}Z$ event predictions divided per signal region. The luminosity uncertainty is not considered in the sum.

C.6 Additional $2\ell 0\tau_{\text{had}}$ distributions

This section shows some of the properties of the events selected in the $2\ell 0\tau_{\text{had}}$ category. The di-lepton invariant mass, the E_{T}^{miss}, the H_{T}[3] and the $N\,(\text{jet}) + 10 \cdot N\,(b\text{-jet})$ distributions are displayed in Figure C.24, Figure C.25, Figure C.26 for the $e^{\pm}e^{\pm}$, the $e^{\pm}\mu^{\pm}$ and the $\mu^{\pm}\mu^{\pm}$ categories, respectively. The shape of the non-prompt background prediction is obtained from simulation. The predictions are in good agreement with observations.

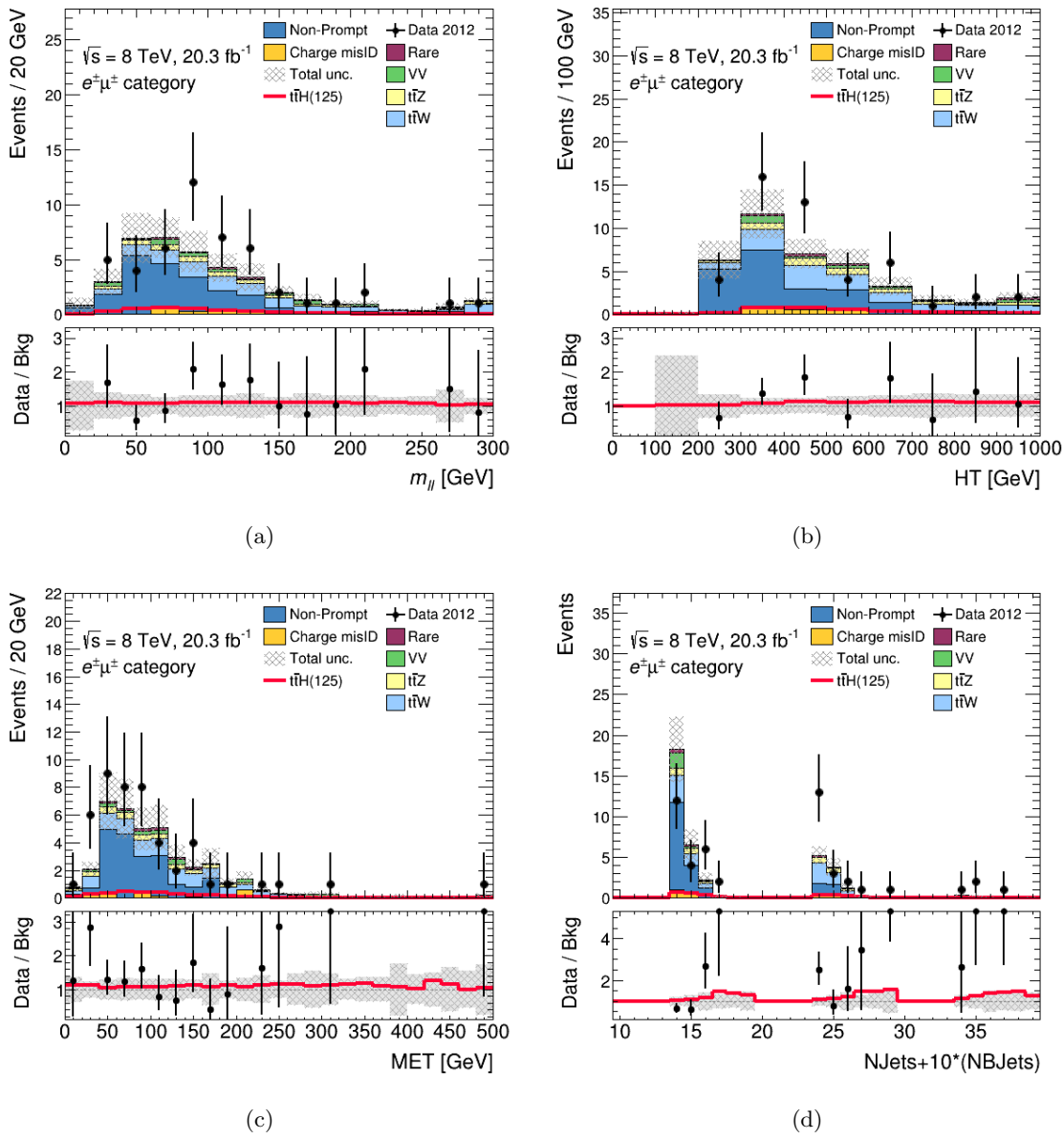
[3] H_{T} is defined as the scalar sum of the p_{T} of the leptons and jets in the event.



(a) (b)

(c) (d)

Figure C.24: Distributions of the di-lepton invariant mass (a), the H_{T} (b), the E_{T}^{miss} (c) and the combined jet and b-jet multiplicities (d) in the $e^{\pm}e^{\pm}$ $2\ell 0\tau_{\mathrm{had}}$ signal regions. In case of N (jet) > 9 events, these are accounted in the $9 + 10 \cdot N$ (b-jet) interval of the N (jet) $+ 10 \cdot N$ (b-jet) plot to avoid mixing with events with an additional b-jet. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.



(a) (b)

(c) (d)

Figure C.25: Distributions of the di-lepton invariant mass (a), the H_{T} (b), the E_{T}^{miss} (c) and the combined jet and b-jet multiplicities (d) in the $e^{\pm}\mu^{\pm}$ $2\ell 0\tau_{\mathrm{had}}$ signal regions. In case of $N\,(\mathrm{jet}) > 9$ events, these are accounted in the $9 + 10 \cdot N\,(b\text{-jet})$ interval of the $N\,(\mathrm{jet}) + 10 \cdot N\,(b\text{-jet})$ plot to avoid mixing with events with an additional b-jet. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.

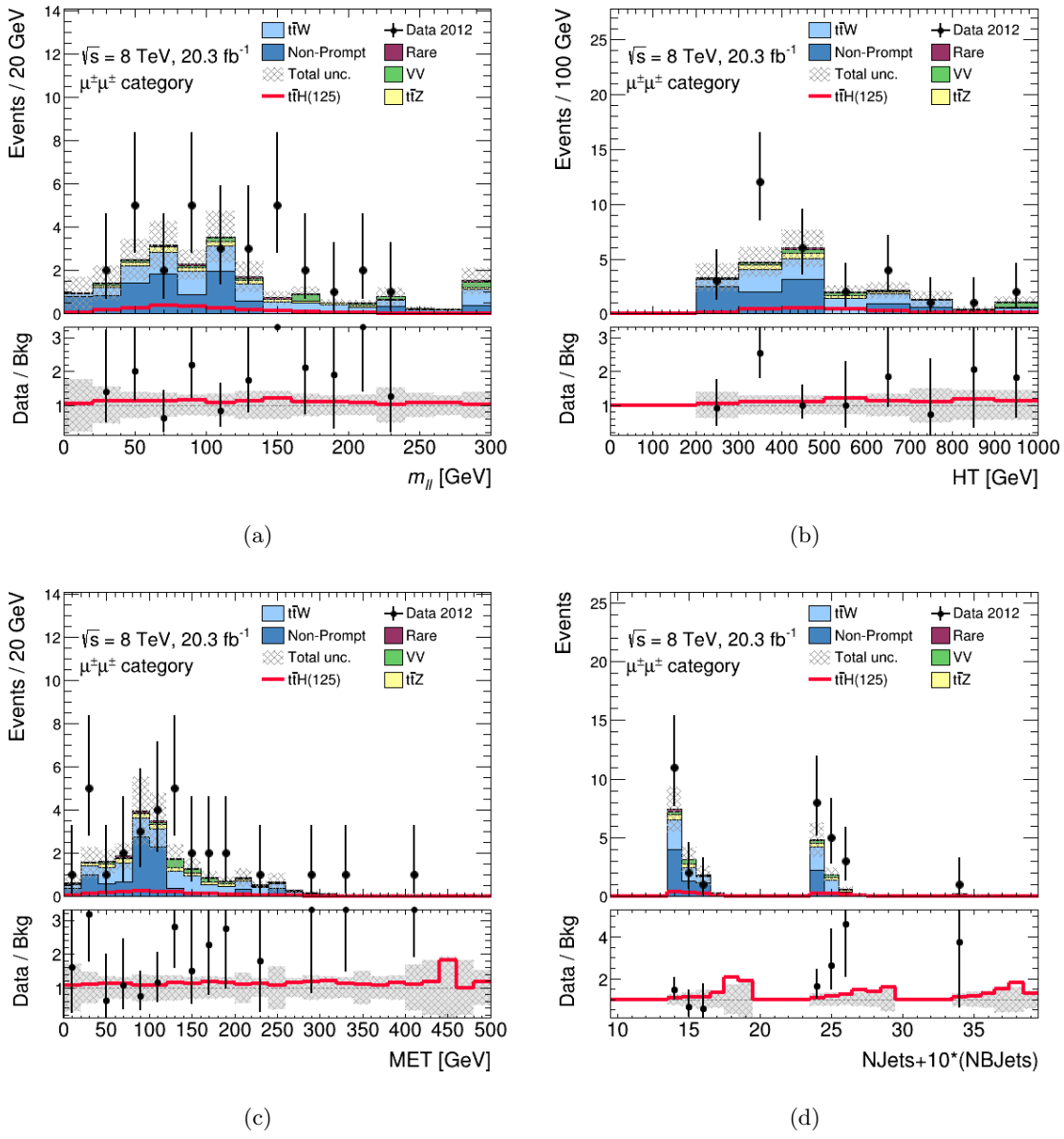


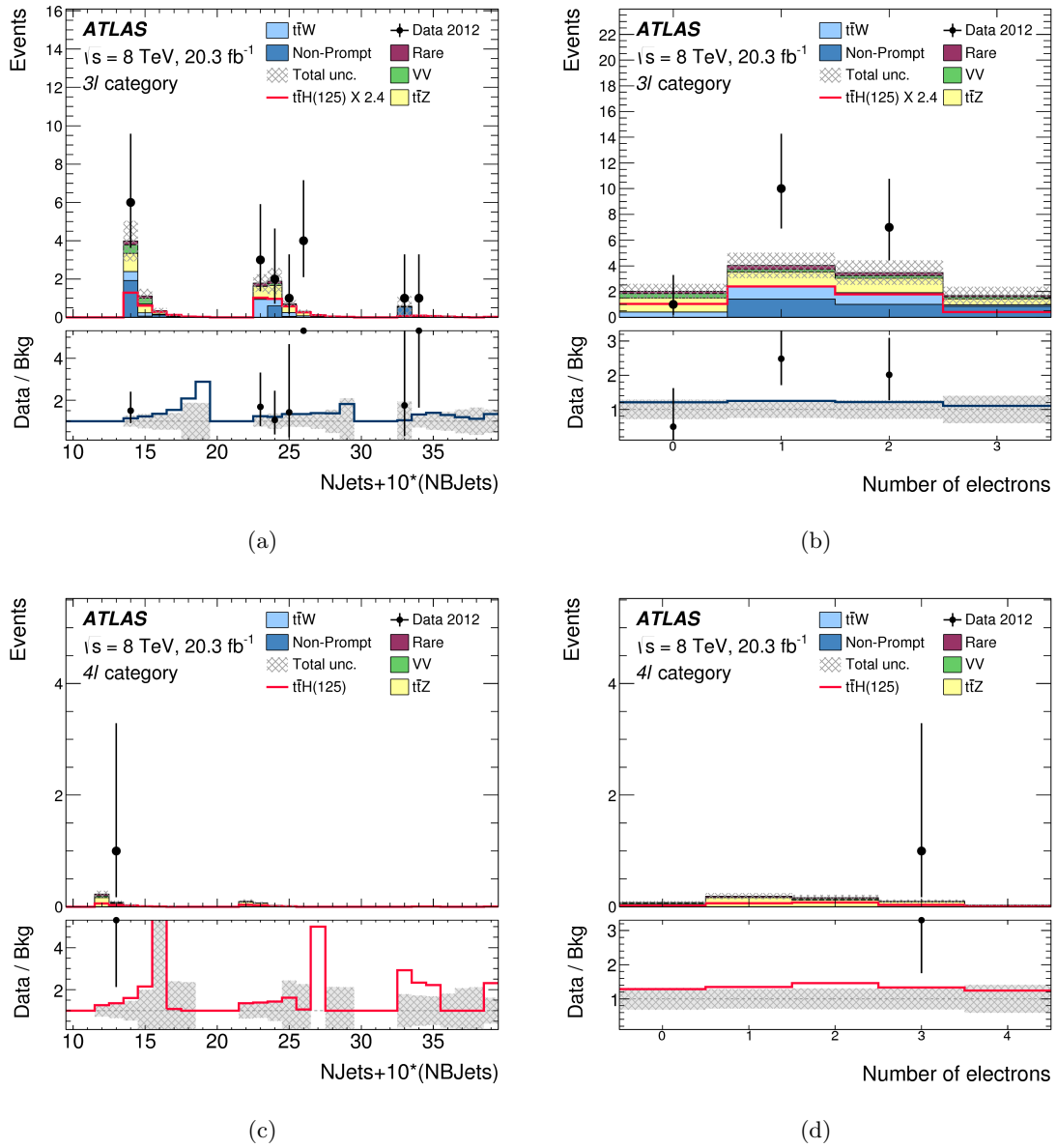
Figure C.26: Distributions of the di-lepton invariant mass (a), the H_{T} (b), the E_{T}^{miss} (c) and the combined jet and b-jet multiplicities (d) in the $\mu^{\pm}\mu^{\pm}$ $2\ell 0\tau_{\mathrm{had}}$ signal regions. In case of N (jet) > 9 events, these are accounted in the $9 + 10 \cdot N$ (b-jet) interval of the N (jet) $+ 10 \cdot N$ (b-jet) plot to avoid mixing with events with an additional b-jet. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.

C.7 Other signal regions

This section shows the signal regions with three leptons, four leptons or τ_{had} in the final state. The observed and expected events in each category are listed in Table C.10. The distributions of the combined jet and b-jet multiplicities and the lepton flavour composition of the events in the 3ℓ, 4ℓ, $2\ell 1\tau_{\text{had}}$ and $1\ell 2\tau_{\text{had}}$ categories are shown in Figure C.27 and Figure C.28. The observed events exceed the prediction in the 3ℓ category, where the number of signal events expected is higher. Data events are below expectations in the $1\ell 2\tau_{\text{had}}$ category dominated by non-prompt background from $t\bar{t}$ production and where the fraction od signal is the smallest. The observation in these regions are combined with those in the $2\ell 0\tau_{\text{had}}$ category in Section 6.7.2 to measure the $t\bar{t}H$ production cross-section.

Category	Non-prompt	$t\bar{t}W^{\pm}$	$t\bar{t}Z$	VV	Total bkg.	$t\bar{t}H$	Observed
3ℓ	3.2 ± 0.7	2.3 ± 0.7	3.9 ± 0.8	0.86 ± 0.55	11.4 ± 2.3	2.34 ± 0.35	18
4ℓ Z-enr.	$\lesssim 10^{-3}$	$\lesssim 3 \times 10^{-3}$	0.43 ± 0.12	0.05 ± 0.02	0.55 ± 0.15	0.17 ± 0.02	1
4ℓ Z-dep.	$\lesssim 10^{-4}$	$\lesssim 10^{-3}$	0.002 ± 0.002	$\lesssim 2 \times 10^{-5}$	0.007 ± 0.005	0.025 ± 0.003	0
$2\ell 1\tau_{\text{had}}$	$0.4\ ^{+0.6}_{-0.4}$	0.38 ± 0.12	0.37 ± 0.08	0.12 ± 0.11	1.4 ± 0.6	0.47 ± 0.08	1
$1\ell 2\tau_{\text{had}}$	15 ± 5	0.17 ± 0.06	0.37 ± 0.09	0.41 ± 0.42	16 ± 5	0.68 ± 0.13	10

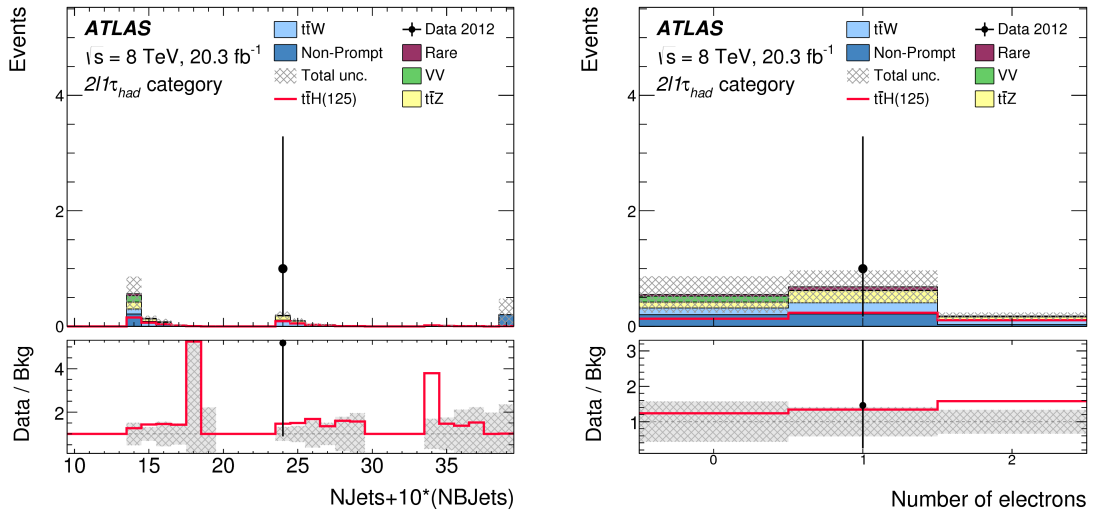
Table C.10: Expected and observed event yields in the analysis signal regions with three leptons, four leptons or τ_{had}. The overall uncertainties on the predictions are also given. Rare processes (tZ, $t\bar{t}W^{+}W^{-}$, $H \to ZZ^{(*)}$, $tHqb$, $tHW^{\pm}$ and $t\bar{t}t\bar{t}$) are included in the total expected background estimate.



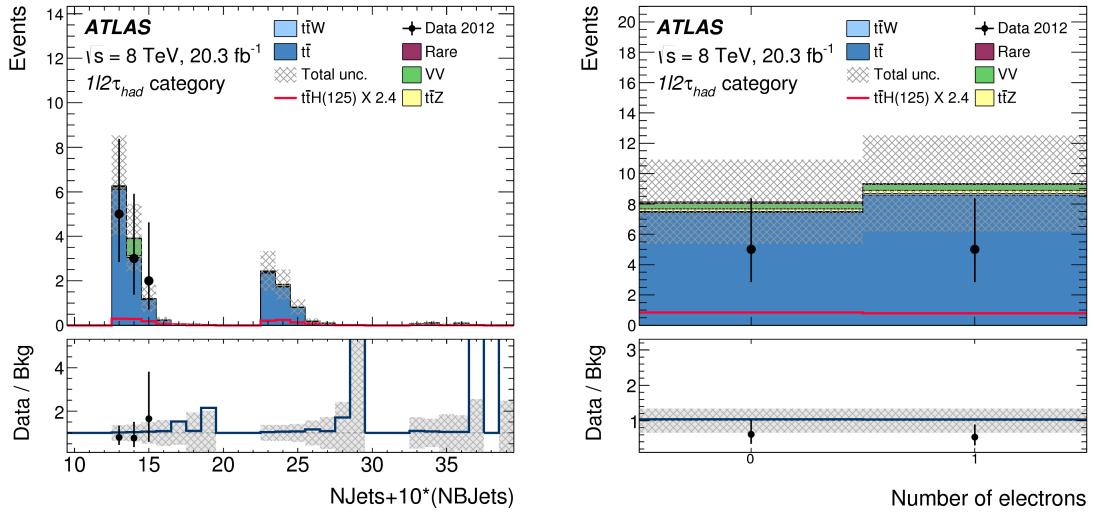
(a) (b)

(c) (d)

Figure C.27: Distributions of the combined jet and b-jet multiplicities (left) and the electron multiplicity (right) in the 3ℓ (top), 4ℓ (bottom) signal regions. In case of N (jet) > 9 events, these are accounted in the $9+10\cdot N$ (b-jet) interval of the N (jet) $+10\cdot N$ (b-jet) plot to avoid mixing with events with an additional b-jet. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.



(a) (b)



(c) (d)

Figure C.28: Distributions of the combined jet and b-jet multiplicities (left) and the electron multiplicity (right) in the $2\ell 1\tau_{\text{had}}$ (top) and $1\ell 2\tau_{\text{had}}$ (bottom) signal regions. In case of N (jet) > 9 events, these are accounted in the $9 + 10 \cdot N$ (b-jet) interval of the N (jet) $+ 10 \cdot N$ (b-jet) plot to avoid mixing with events with an additional b-jet. Events in data (black dots) are compared to the expected stacked contributions from Standard Model background processes. The expected signal contributions are overlaid to the backgrounds (red line). The lower panel shows the ratio of observed events in data to expected Standard Model events. The error bars on the data points show the statistical uncertainty and the dashed band shows the total uncertainties of the background predictions.

C.8 Measurement of the Top Yukawa coupling

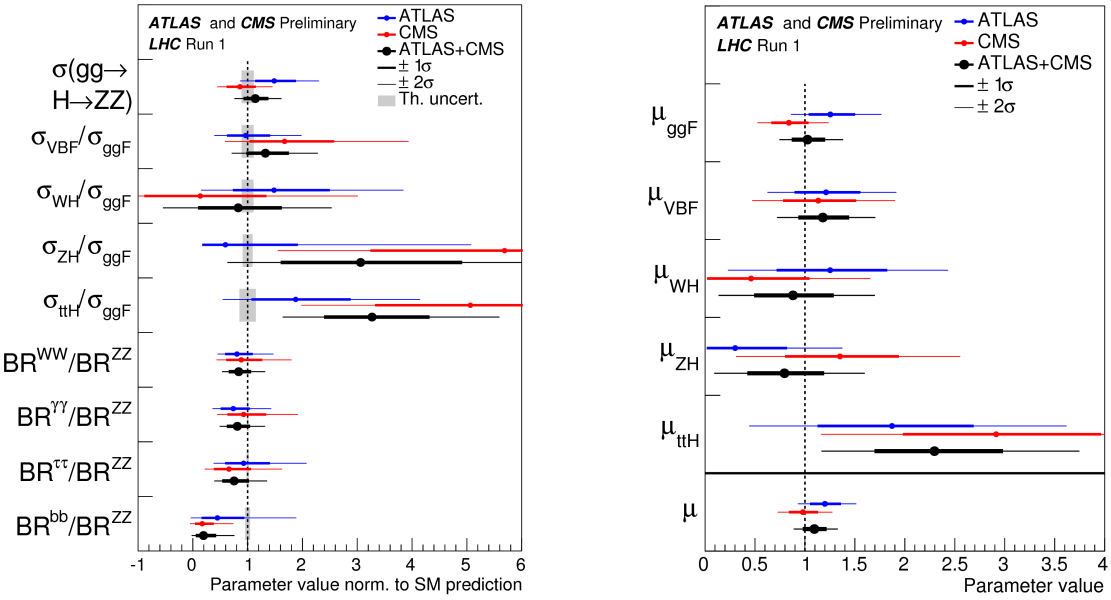
After the discovery of the Higgs boson, the ATLAS and CMS collaborations have investigated the mechanisms of production and the decay modes of the Higgs boson with numerous analyses of $\sqrt{s} = 7$ and 8 TeV data, including the search of $t\bar{t}H$ production in the final states with leptons. The results of these analyses have been combined to achieve the most accurate measurements of the production and decay modes of the Higgs boson and lastly of the coupling of the Higgs boson to Standard Model particles. ATLAS and CMS have published independent results [188, 189] and have performed a combination of their measurements [34]. The results of the combination are obtained in the context of the κ-framework [14] and are briefly described here with particular stress for those results related to the $t\bar{t}H$ analysis findings.

The assumptions made in the study are that the properties of the signal investigated are Standard Model Higgs boson-like. More precisely the particle is a scalar and even under CP, has the tensor structure of the Standard Model interactions and has width enough small to allow the factorisation of the production and decay mode. The production rate and decay branching ratios are not constrained but the kinematic of each production and decay mode is assumed to be the same of a Standard Model Higgs boson with 125 GeV mass. For small deviations of production and decay rates from Standard Model, the hypotheses made have negligible effects on the results [34].

The studied production modes are: gluon-gluon fusion (ggF), vector boson fusion (VBF) and associated production with vector bosons (VH) or a pair of top quarks ($t\bar{t}H$). The decay channels considered are: $H \to ZZ^{(*)}$, $H \to W^+W^-$, $H \to \gamma\gamma$, $H \to \tau^+\tau^-$, $H \to b\bar{b}$.

Several parametrisations have been used in the statistical model. The most generic parametrisation uses ratios of cross-sections and of branching ratios. The production cross-section times branching ratio of the process $gg \to H \to ZZ^{(*)}$, that has the smallest systematic uncertainties and one of the smallest overall uncertainties, is one of the PoIs. The other parameters are the ratio of the production cross-section for a mode to the ggF production cross-section and the ratio of the branching ratio of a decay mode to the branching ratio of $H \to ZZ^{(*)}$. Many theoretical and experimental systematic uncertainties cancel out in these ratios. There is no assumption on the total decay width. The measured parameters normalized to the value expected by Standard Model are shown in Figure C.29(a). The $t\bar{t}H$ analysis results affect mainly the measurement of $\sigma_{t\bar{t}H}/\sigma_{ggF}$ whose observed value is $0.022^{+0.007}_{-0.006}$. This is more than 2σ higher than the expected value of 0.0067 ± 0.0010. The ATLAS on ly combination is 1σ higher than the expected value and has a smaller uncertainty than those by CMS.

Production cross-sections and branching ratios cannot be measured simultaneously, since the observed rate depends by their products. Assumed that the Higgs boson branching ratios are those expected in Standard Model, the signal strength of the Higgs boson



(a) (b)

Figure C.29: (left): The measured values of the $gg \to H \to ZZ$ cross-section, of the production cross-sections relative to gluon-gluon fusion (ggF) and of the branching ratios relative to $H \to ZZ^{(*)}$ from the combination of ATLAS and CMS Higgs boson analyses. The values are normalized to the Standard Model predictions. (right): The measured signal strengths for the Higgs boson production modes from the combination of ATLAS and CMS Higgs boson analyses. Results from the ATLAS measurement and CMS measurement are shown in blue and red respectively. The 1σ (thick lines) and 2σ (thin lines) error intervals are indicated.

production mechanisms can be measured. The signal strength for the Higgs boson production modes are shown in Figure C.29(b). The signal strength for $t\bar{t}H$ production obtained from the combination is $2.3^{+0.7}_{-0.6}$; the value measured by ATLAS is $1.9^{+0.8}_{-0.7}$. The observed (expected) significance of the $t\bar{t}H$ production is 2.7 (1.6) for ATLAS and 3.6 (1.3) for CMS. On the other hand, the production modes can be fixed to their Standard Model values to measure the branching ratios. The $t\bar{t}H$ analysis is however less relevant for these measurements.

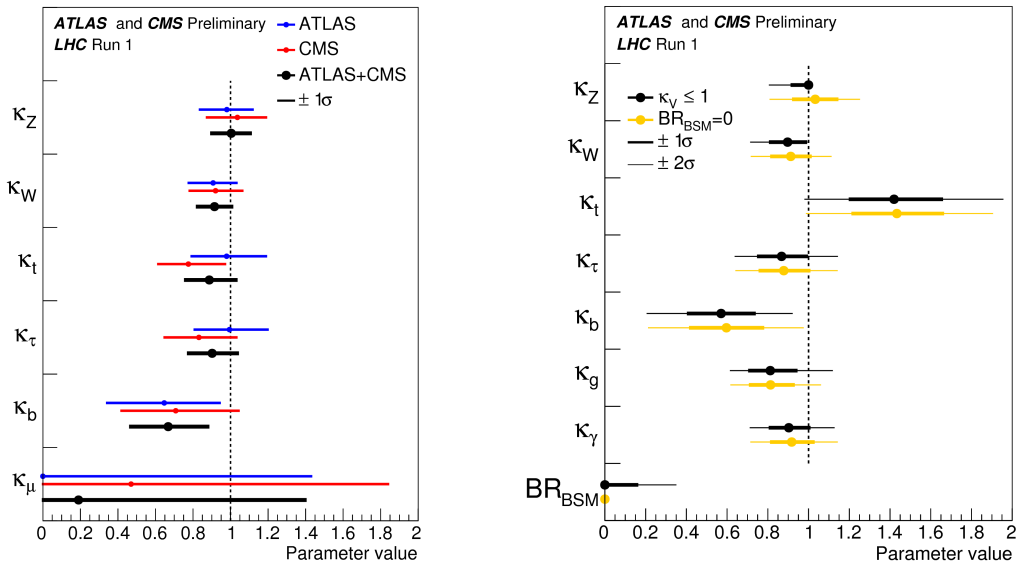
Finally, the results can be reinterpreted in terms of coupling strength of the Higgs boson to Standard Model particles. Different production processes and decay modes are sensitive to different coupling constants. The coupling values are measured in units of the expected values for a Standard Model Higgs boson with 125 GeV mass (coupling modifiers κ). The ggF production and the $H \to \gamma\gamma$ decay width diagrams contain loops and are studied through either the effective coupling modifiers (κ_g and κ_γ) or the modifiers of the Standard Model particles in the loops (κ_b, κ_W and κ_t).

The production cross-section and the decay widths parametrisation are expressed as a function of coupling modifiers in Table 1.2. The associated production of the Higgs boson and top quark pairs is the only process that depends exclusively from the Top Yukawa coupling modifier (κ_t). The $t\bar{t}H$ cross-section is proportional to κ_t^2. The Top Yukawa coupling is otherwise contributing to processes with loops like ggF and $H \to \gamma\gamma$, but they also depend on couplings to other Standard Model particles and potentially to beyond Standard Model (BSM) particles.

In the assumption of no interaction vertexes of the Higgs boson with BSM particles, i.e. there are no BSM particles in the loops and there is no Higgs boson decay mode to BSM particles, the parametrisation of the cross-sections and decay widths is based on the Yukawa coupling modifiers and the boson couplings modifiers: κ_t, κ_b, κ_τ, κ_μ, κ_W, κ_Z. The measured coupling modifiers are shown in Figure C.30(a) and the value of κ_t is given in Table 6.25.

To allow BSM particles in loops, the effective coupling modifiers κ_g and κ_γ are used in addition to the other parameters. These parameters would reflect contributions from new heavy particles with non-null electric charge or colour charge. Since BSM particles could contribute to the Higgs boson decays width and consequently to the branching ratios of each detectable decay mode, a hypothesis on the width is required. Two assumption has been made: the Higgs boson cannot decay to BSM particles ($BR(BSM) = 0$) or the coupling modifiers κ_W and κ_Z must be at most 1 ($\kappa_V \leq 1$). The measured coupling modifiers for the two conditions are shown in Figure C.30(b) and the value of κ_t is given in Table 6.25.

There is a difference between the measured Top Yukawa coupling modifier depending on whether BSM particles are allowed in the loops. In the Standard Model particle only case, the modifier is compatible with 1 within the error and is determined from the measurement of ggF production, $H \to \gamma\gamma$ decay and $t\bar{t}H$ production. When BSM



(a) (b)

Figure C.30: The measured couplings of the Higgs boson to Standard Model particles normalized to the Standard Model value under three hypotheses. (left): No BSM particles affecting the loops that couple the Higgs boson to gluons and to γ and no decay of the Higgs boson to BSM particles. Results from the ATLAS measurement and CMS measurement are shown in blue and red respectively. (right): BSM particles allowed in the loops but H cannot decay to BSM particles (yellow) or $\kappa_V \leq 1$ with $V = W^{\pm}, Z$ (black). The 1σ (thick lines) and 2σ (thin lines) error intervals are indicated.

particles are allowed, κ_t is determined from $t\bar{t}H$ production data only and its value is larger than the Standard Model one by two times the error. In the ATLAS combination, κ_t exceed the expected value by as much as the error. The choice of condition for the BSM scenario, either $BR(BSM) = 0$ or $\kappa_V \leq 1$, has negligible impact on the κ_t measurement.

In a typical new physics scenario, one would observe a κ_t value compatible with one, and modifiers associated to the gluon-gluon fusion production and $H \to \gamma\gamma$ decay diagrams (κ_g and κ_γ) significantly different. The current findings instead indicates a $\kappa_t > 1$ and nearly unit value for κ_g and κ_γ. If this scenario is confirmed, the Top Yukawa coupling for the discovered particle at 125 GeV is not that expected for a Standard Model Higgs boson and therefore the particle itself is not the Higgs boson. In addition, there should be BSM particles which contribute to the gluon-gluon fusion production and to the $H \to \gamma\gamma$ decay to compensate for the non-one κ_t value. This condition seems highly unlikely but cannot be excluded a priori. An anomaly in $t\bar{t}H$ production measurement and so in the value of κ_t could be also the consequence of unknown processes that match the $t\bar{t}H$ production final state. For instance, BSM particles may be produced in association with top quark pairs, and result in a large multitude of jets and leptons (photons and b-jets for analyses considering other $t\bar{t}H$ decay modes). These processes would constitute a background for this analysis which has not been taken into account resulting in an inflation of the measured $t\bar{t}H$ signal strength.

At this stage, the observed κ_t deviation is not significant enough to constitute an evidence of new physics but it is something that should be further investigated with new data.