Observation of Higgs Boson Decays to WW^* with Dilepton and Missing Transverse Momentum Events in the ATLAS Detector

by

Yesenia Hernández Jiménez

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

VNIVERSITAT DE VALÈNCIA

Universitat de València (IFIC-CSIC)
Departament de Física Atòmica, Molecular i Nuclear

Supervised by:
Dr. Juan A. Valls Ferrer
Dr. Bruce Mellado García

Valencia, 2015

El **Dr. Juan Antonio Valls Ferrer**, Científico Titular del Consejo Superior de Investigaciones Científicas, y el **Dr. Bruce Mellado García**, Profesor de University of Witwatersrand,

CERTIFICAN:

Que la presente memoria, **"*Observation of Higgs Boson Decays to* WW^* *with Dilepton and Missing Transverse Momentum Events in the ATLAS Detector*"** ha sido realizada bajo nuestra dirección en el Departamento de Física Atómica, Molecular y Nuclear de la Universitat de València por Yesenia Hernández Jiménez y constituye su tesis para optar al grado de doctor en Física por la Universitat de València.

Y para que así conste, en cumplimiento de la legislación vigente, firmamos el presente Certificado,

Dr. Juan A. Valls Ferrer

Dr. Bruce Mellado García

Declaration of Authorship

This dissertation is submitted to the Universitat de València for its evaluation and constitutes the work of my own PhD thesis. The results presented in this report are due to my own work, except where explicit reference is made to the work of others. These results have not been submitted for another academic qualification to this or any other university.

Signed: _____

Date: _____

Acknowledgements

First I would like to express my sincere gratitude to my supervisor Dr. Juan A. Valls Ferrer. He has been of great support starting from my first days at IFIC and CERN. I am very thankful for giving me such an amazing opportunity to participate in the Higgs discovery and support me to spend long stays at CERN. My sincere thanks also goes to Dr. Bruce Mellado. I have worked very closely with him over these years having sleepless nights to provide results on time before the deadlines. He has supported me throughout my research with his patience, knowledge and guidance while allowing me to develop this work in my own way. I have really enjoyed the intellectual discussions on physics we had and I am very grateful for all the things he has taught me.

I want to thank University of Valencia - IFIC and Departament de Física Atòmica, Molecular i Nuclear that support me academically, institutionally and economically. Special thanks to Prof. Antonio Ferrer and Prof. Emilio Higón for giving enormous advices about how to conduct research in large collaborations as ATLAS. I am also very thankful for all the time they have devoted in the thesis comments, as well as Carlos Solans, Irene Vichou, Luca Fiorini, Pierre Savard and Richard Teuscher.

I had lots of luck to be a part of the TileCal, EtMiss and Higgs HWW groups of ATLAS, and I am very thankful for the friendly and cheerful professional environment I worked within. I am very much indebted to all ATLAS colleagues that I have met, specially in my stays at CERN. I am deeply grateful to all of them for their continuous efforts on this analysis. They have made lots of meaningful suggestions and have given me ideas to improve my works. Special thanks to the colleagues from TileCal at the University of Valencia - IFIC group for helping me and for explaining such a lot of things about TileCal electronics and physics. My everyday life at IFIC and CERN would not be the same without them: Alberto, César, Fernando, Damián, Leo, Luis, Pablo... and the Wits team: Harshna, Matt, Rob... Thank you guys for such a lot of funny and happy moments inside and outside the office.

Last but not least I am deeply grateful for the warm support of my family, despite to not quite understand my research work at all. The greatest thanks go to my mum and my brother because without their unending love and caring this thesis would not have been possible. Through all the ups and downs of the past years I have been able to keep going thanks to Manolo. You always have been there for me

and with me even when we were separated by thousands of kilometers. Thanks for making me a better person. I love you.

This work would not be the same without the help from many people who gave me their encouragement and support in different ways. To all of them I dedicate this thesis.

Contents

Chapter 1

Standard Model Theory and the Higgs Boson

1.1 Introduction

The Standard Model (SM) constitutes one of the most successful achievements in modern physics. It provides a theoretical framework which is able to describe the known experimental facts in particle physics with high precision. All particles described in the SM have been observed and, up to now, no significant deviations from the theory have been found. The SM predicts a new scalar particle known as the Higgs boson. The Higgs boson breaks electroweak symmetry and provides mass to the fundamental particles. This is a key building block in the SM and the topic of this thesis. Before the Large Hadron Collider (LHC) started to collect data from proton-proton collisions, the Higgs boson was the only elementary particle in the SM that had not been observed.

The aim of this chapter is to describe the formulation of the SM and present the role of the Higgs boson. Section 1.2 presents an overview of the elementary particles. This is followed by the description of the quantum electrodynamics and quantum chromodynamics theories in Section 1.2.2 and 1.2.3, respectively. The unification of the electroweak is given in Section 1.2.4. In Section 1.3.1, the electroweak symmetry breaking mechanism and the Higgs boson in the SM are discussed. Then, the phenomenology of proton-proton collisions is described in Section 1.5. Finally, different production modes and the main decay channels of the Higgs boson in proton-proton collisions are presented in Section 1.6 and 1.7, respectively.

1.2 The Standard Model

The SM unifies three of the four fundamental forces of the Universe: strong, weak and electromagnetic interactions. Apart from gravity, the SM of particle physics is able to describe nature in terms of fundamental constituents and their interactions.

1.2.1 Elementary Particles in the SM

The SM involves two kinds of particles, those carrying charge and those which mediate interactions by coupling directly to charge. The physical nature of charge depends on the specific theory. Three such kind of charges appear in the SM, the so-called colour, weak isospin and weak hypercharge.

In the SM the mediators of the fundamental interactions are the spin-1 gauge bosons: one massless photon (γ), eight massless gluons (g) and three massive bosons ($W^{\pm}$ and Z). The photons mediate the electromagnetic interactions between electrically charged particles. The photon is a massless particle and it is described by the theory of Quantum Electrodynamics (QED). The $W^{\pm}$ and Z bosons are the elementary particles that mediate the weak interactions between fermions (quarks and leptons).

These are massive and while the Z particle is electrically neutral, the $W^{\pm}$ bosons carry an electric charge of ± 1 times the electron charge respectively. These three gauge bosons along with the photon are grouped together in the electroweak (EW) interaction. Finally, the gluons are massless particles that mediate the strong interactions between particles with colour charge, as the quarks. Because the gluons have colour charge, they can also interact among themselves. The gluons and their interactions are described by the Quantum Chromodynamics (QCD) theory. Table 1.1 summarises the three fundamental forces unified by the SM and their associated gauge bosons.

TABLE 1.1: Fundamental forces and their associated gauge bosons. J denotes the spin and P the parity of the particles [1].

Interaction	**Name**	**Symbol**	**Charge [e]**	$\boldsymbol{J^P}$	**Mass [GeV]**
Electromagnetic	Photon	γ	0	1^-	0
Weak	W bosons	$W^{\pm}$	± 1	1	80.385 ± 0.015
	Z boson	Z	0	1	91.1876 ± 0.0021
Strong	Gluon	g	0	1^-	0

The fermionic matter content in the SM is given by 12 elementary particles: the known leptons and quarks. There are six quarks: up (u), down (d), charm (c), strange (s), top (t), and bottom (b), and six leptons: electron (e), muon (μ), tau (τ), and their corresponding neutrino partners: electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ). The three neutrinos are neutral in terms of electric charge and they are assumed to be massless in the SM [1].

All fermions are particles of spin $\frac{1}{2}$ and each of them has its corresponding anti-particle with opposite quantum numbers. According to the spin-statistics theorem [4], fermions respect the Pauli exclusion principle. The difference between leptons and quarks relies on the coupling to the strong interaction. The physical nature of the leptons shows that they are colourless and they do not interact via the strong force. Oppositely, quarks carry colour and they do interact via strong processes. The quarks are grouped together forming the known hadrons: baryons with half-integer spin (three quarks), and mesons with integer spin (quark and anti-quark). This phenomenon is called colour confinement. Quarks also carry electric charge and weak isospin, hence they interact with another fermions both electromagnetically and via the weak interaction. Table 1.2 lists the three fermion generations.

TABLE 1.2: Fermion generations overview from Ref. [1]. Masses and electric charges in units of the electron charge are quoted. J denotes the spin and P the parity of the particles.

Fermion	**I generation**	**II generation**	**III generation**	**Charge** $[e]$	$\boldsymbol{J^P}$
Quarks	u	c	t	$+2/3$	$1/2^+$
	$2.3^{+0.7}_{-0.5}\,\text{MeV}$	$1.275 \pm 0.025\,\text{GeV}$	$173.5 \pm 0.6 \pm 0.8\,\text{GeV}$		
	d	s	b	$-1/3$	$1/2^+$
	$4.8^{+0.7}_{-0.3}\,\text{MeV}$	$95 \pm 5\,\text{MeV}$	$4.18 \pm 0.03\,\text{GeV}$		
Leptons	e^-	μ^-	τ^-	-1	$1/2^+$
	$0.511 \pm (0.11 \times 10^{-7})\,\text{MeV}$	$106 \pm (35 \times 10^{-7})\,\text{MeV}$	$1776.82 \pm 0.16\,\text{MeV}$		
	ν_e	ν_μ	ν_τ	0	$1/2^+$
	$< 2.05\,\text{eV}$ (95% CL)	$< 0.19\,\text{eV}$ (90% CL)	$< 18.2\,\text{eV}$ (95% CL)		

Lastly, the SM theory predicts a new particle that completes the fundamental particle spectrum: the Higgs boson. It appears trough the spontaneous electroweak symmetry breaking mechanism, which is a necessary ingredient in the SM for ensuring gauge invariance. It arises from the need to provide mass to the electroweak mediators, i.e., the $W^\pm$ and Z bosons, as will be discussed in the following sections.

[1] Note that several experiments with solar, atmospheric, reactor and accelerator neutrinos have provided compelling evidence for oscillations of neutrinos caused by nonzero neutrino masses [2, 3].

The theory constrains some aspects of the Higgs particle: it is a massive boson (although its mass is not predicted), with no spin, electric charge, or colour charge. In this sense, the position of the Higgs boson in the zoo of the fundamental particles in the SM is unique, as is its role. Figure 1.1 collects the fundamental particles in the SM.

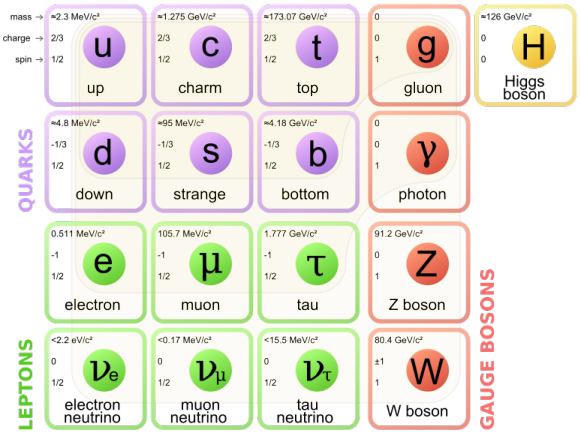

FIGURE 1.1: The SM of elementary particles with the three generations of fermions in the left-hand columns, gauge bosons in the fourth column and the Higgs boson on the right column.

1.2.2 Quantum Electrodynamics

Historically, the first of the gauge field theories was electrodynamics. Its modern version, Quantum Electrodynamics (QED), is the most thoroughly verified physical theory yet constructed. QED represents the best introduction to the SM, which both incorporates and extends it.

Let us consider the Lagrangian describing a free Dirac fermion with mass m and electric charge Qe,

$$\mathcal{L}_0 = i\overline{\psi}(x)\gamma^\mu\partial_\mu\psi(x) - m\overline{\psi}(x)\psi(x)\,, \tag{1.1}$$

where $\psi(x)$ is the fermion field and $\overline{\psi}(x) \equiv \psi^\dagger\gamma^0$ is its adjoint. Note that natural units ($\hbar = c = 1$) has been used in Eq. 1.1 and they will be assumed in the rest of the chapter. $\mathcal{L}_0$ is invariant under global $U(1)$ transformations,

$$\psi(x) \quad \xrightarrow{U(1)} \quad \psi'(x) \equiv \mathrm{e}^{iQ\theta}\psi(x)\,, \tag{1.2}$$

where $Q\theta$ is an arbitrary real constant hence,

$$\overline{\psi}(x) \to \mathrm{e}^{-i\theta}\overline{\psi}(x) \quad \text{and} \quad \partial_\mu\psi(x) \to \mathrm{e}^{iQ\theta}\partial_\mu\psi(x)\,. \tag{1.3}$$

The phase of $\psi(x)$ is then a pure convention-dependent quantity without any physical meaning. However, the free Lagrangian is no longer invariant if one allows the phase transformation to depend on the space-time coordinate, i.e., under local phase redefinitions $\theta = \theta(x)$.

Explicitly, the local gauge transformation is,

$$\psi(x) \quad \xrightarrow{U(1)} \quad \psi'(x) \equiv \mathrm{e}^{iQ\theta(x)}\psi(x)\,, \tag{1.4}$$

which clearly makes the Lagrangian non invariant due to the extra term from the derivative of $\theta(x)$,

$$\partial_\mu\psi(x) \quad \xrightarrow{U(1)} \quad \mathrm{e}^{iQ\theta(x)}(\partial_\mu + iQ\partial_\mu\theta(x))\psi(x)\,. \tag{1.5}$$

Thus, if a given phase convention has been adopted at a reference point, the same convention should be taken at all space-time points. This does not look reasonable. As an illustration, let us consider a mechanism which annihilates electric charge at one point in space and simultaneously create an equal amount of charge at another spatial point. The charge is conserved globally in this case, but it would be in conflict with special relativity. Electric charge has to be conserved locally. This property of local conservation is at the heart of the local gauge symmetries exhibited by the fundamental interactions.

The gauge principle is the requirement that the $U(1)$ phase invariance should hold locally. This is only possible if one adds an extra piece to the Lagrangian in Eq. 1.1, transforming in such a way as to cancel the offending term in Eq. 1.5. The needed modification is completely fixed by the transformation, hence one can introduce a new field A_μ which transforms as,

$$A_\mu(x) \quad \xrightarrow{U(1)} \quad A'_\mu(x) \equiv A_\mu(x) - \frac{1}{e}\partial_\mu\theta\,. \tag{1.6}$$

Defining the covariant derivative as,

$$\mathcal{D}_\mu\psi(x) \equiv [\partial_\mu + ieQA_\mu(x)]\psi(x)\,, \tag{1.7}$$

which has the required property of transforming like the field itself,

$$\mathcal{D}_\mu\psi(x) \quad \xrightarrow{U(1)} \quad \mathcal{D}'_\mu\psi(x) \equiv \mathrm{e}^{iQ\theta}\mathcal{D}_\mu\psi(x)\,. \tag{1.8}$$

The newly obtained Lagrangian,

$$\mathcal{L} \equiv i\overline{\psi}(x)\gamma^\mu\mathcal{D}_\mu\psi(x) - m\overline{\psi}(x)\psi(x) = \mathcal{L}_0 - eQA_\mu(x)\overline{\psi}(x)\gamma^\mu\psi(x)\,, \tag{1.9}$$

is now invariant under local $U(1)$ transformations. The price we had to pay was the introduction of the new vector field A_μ that couples to ψ through the last term in Eq. 1.9. Actually, it is just the familiar vertex of QED which contains the electromagnetic interaction between charged particles and its mediator, the photon.

If one wants A_μ to be a true propagating field, one needs to add a gauge-invariant kinetic term,

$$\mathcal{L}_{\text{kin}} = -\frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) + \frac{1}{2}m_A^2A^\mu A_\mu\,, \tag{1.10}$$

where $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$ is the usual electromagnetic field strength which is invariant under the local gauge transformation given in Eq. 1.5. However, the term $A^\mu A_\mu$ is not. In light of this, the gauge field must be massless: $m_A = 0$ and experimentally we have measured that $m_\gamma < 1 \times 10^{-18}\,\text{eV}$ [1].

The final QED Lagrangian can be written as,

$$\mathcal{L}_{\text{QED}} = i\overline{\psi}(x)\gamma^\mu\partial_\mu\psi(x) - m\overline{\psi}(x)\psi(x) - \frac{1}{4}F_{\mu\nu}F^{\nu\mu} - eQA_\mu(x)\overline{\psi}(x)\gamma^\mu\psi(x)\,, \tag{1.11}$$

which gives rise to the Maxwell equations and specifies the current produced by Dirac particles (J^μ),

$$\partial_\mu F^{\mu\nu} = J^\nu \equiv eQ\overline{\psi}\gamma^\mu\psi\,. \tag{1.12}$$

Thus, the requirement of local gauge invariance, applied to the free Dirac Lagrangian, generates all electrodynamics, so it leads to a very successful quantum field theory.

1.2.3 Quantum Chromodynamics

The large number of known mesonic and barionic states clearly signals the existence of a deeper level of elementary constituents of matter: quarks. Assuming that mesons are $M \equiv q\overline{q}$, while baryons have three quark constituents: $B \equiv qqq$; one

can classify the entire hadronic spectrum. However, in order to satisfy the Fermi-Dirac statistics one needs to assume the existence of a new quantum number, colour, such that each species of quark may have $N_{\rm C} = 3$ different colours: q^α, $\alpha = 1, 2, 3$ (red, green and blue). In order to avoid the existence of non-observed extra states with non-zero colour, one needs to further postulate that all asymptotic states are colourless, i.e., singlets under rotations in colour space. This assumption is known as the confinement hypothesis, because it implies the non-observability of free quarks: since quarks carry colour they are confined within colour-singlets bound states.

Let us denote q_f^α a quark field of colour α and flavour f. To simplify the equations, let us adopt a vector notation in colour space: $q_f^T \equiv (q_f^1\,,\, q_f^2\,,\, q_f^3)$. The free Lagrangian,

$$\mathcal{L}_0 = \sum_f \overline{q}_f \left(i\gamma^\mu \partial_\mu - m_f\right) q_f \tag{1.13}$$

is invariant under arbitrary global $SU(3)_{\rm C}$ transformations in colour space,

$$q_f^\alpha \to (q_f^\alpha)' = U_\beta^\alpha q_f^\beta, \quad UU^\dagger = U^\dagger U = 1\,, \quad \det U = 1\,. \tag{1.14}$$

The $SU(3)_{\rm C}$ matrices can be written in the form,

$$U = {\rm e}^{i\frac{\lambda^a}{2}\theta_a}\,, \tag{1.15}$$

where $\frac{\lambda^a}{2}$ $(a = 1, 2, ..., 8)$ denote the generators of the fundamental representation of the $SU(3)_{\rm C}$ algebra, and θ_a are arbitrary parameters. The matrices λ^a are traceless and satisfy the commutation relations,

$$\left[\frac{\lambda^a}{2}, \frac{\lambda^b}{2}\right] = i\, f^{abc}\, \frac{\lambda^c}{2}\,, \tag{1.16}$$

with f^{abc} the $SU(3)_{\rm C}$ structure constants, which are real and totally antisymmetric. As in the QED case, we can now require the Lagrangian to be also invariant under local $SU(3)_{\rm C}$ transformations, $\theta_a = \theta_a(x)$. To satisfy this requirement, we need to change the quark derivatives by covariant objects. Since we have now eight independent gauge parameters, eight different gauge bosons $G_a^\mu(x)$, the so-called gluons, are needed,

$$\mathcal{D}^\mu q_f \equiv \left[\,\partial^\mu + i g_s\, \frac{\lambda^a}{2}\, G_a^\mu(x)\right] q_f \equiv \left[\,\partial^\mu + i g_s\, G^\mu(x)\right] q_f\,. \tag{1.17}$$

Note that we have introduced the compact matrix notation,

$$\left[G^{\mu}(x)\right]_{\alpha\beta} \equiv \left(\frac{\lambda^a}{2}\right)_{\alpha\beta} G_a^{\mu}(x)\,. \tag{1.18}$$

We want $D^{\mu}q_f$ to transform in exactly the same way as the colour-vector q_f. This fixes the transformation properties of the gauge fields,

$$\mathcal{D}^{\mu} \to (\mathcal{D}^{\mu})' = U\mathcal{D}^{\mu}U^{\dagger}\,, \qquad G^{\mu} \to (G^{\mu})' = UG\mu U^{\dagger} + \frac{i}{g_s}(\partial^{\mu}U)U^{\dagger}\,. \tag{1.19}$$

Under an infinitesimal $SU(3)_{\mathrm{C}}$ transformation,

$$\begin{aligned} q_f^{\alpha} &\to \left(q_f^{\alpha}\right)' = q_f^{\alpha} + i\left(\frac{\lambda^a}{2}\right)_{\alpha\beta}\delta\theta_a\, q_f^{\beta}\,, \\ G_a^{\mu} &\to (G_a^{\mu})' = G_a^{\mu} - \frac{1}{g_s}\partial^{\mu}\left(\delta\theta_a\right) - f^{abc}\,\delta\theta_b\, G_c^{\mu}\,. \end{aligned} \tag{1.20}$$

The gauge transformation of the gluon field is more complicated than the one obtained in QED for the photon. The non-commutativity of the $SU(3)_{\mathrm{C}}$ matrices gives rise to an additional term involving the gluon fields themselves. For constant $\delta\theta_a$, the transformation rule for the gauge fields is expressed in terms of the structure constants f^{abc}. Thus, the gluon fields belong to the adjoint representation of the colour group. Note also that there is a unique $SU(3)_{\mathrm{C}}$ coupling g_s. In QED it was possible to assign arbitrary electromagnetic charges to the different fermions. Since the commutation relation given in Eq. 1.16 is non-linear, this freedom does not exist for $SU(3)_{\mathrm{C}}$.

To build a gauge-invariant kinetic term for the gluon fields, we introduce the corresponding field strengths,

$$\begin{aligned} G^{\mu\nu}(x) &\equiv -\frac{i}{g_s}\left[\mathcal{D}^{\mu},\, \mathcal{D}^{\nu}\right] = \partial^{\mu}G^{\nu} - \partial^{\nu}G^{\mu} + ig_s\,\left[G^{\mu},\, G^{\nu}\right] \equiv \frac{\lambda^a}{2}\,G_a^{\mu\nu}(x)\,, \\ G_a^{\mu\nu}(x) &= \partial^{\mu}G_a^{\nu} - \partial^{\nu}G_a^{\mu} - g_s\, f^{abc}\, G_b^{\mu}\, G_c^{\nu}\,. \end{aligned} \tag{1.21}$$

Taking the proper normalisation for the gluon kinetic term, we finally have the $SU(3)_{\mathrm{C}}$ invariant Lagrangian of QCD,

$$\mathcal{L}_{\mathrm{QCD}} = -\frac{1}{4}\,G_a^{\mu\nu}\,G^a_{\mu\nu} + \sum_f \overline{q}_f\,\left(i\gamma^{\mu}\mathcal{D}_{\mu} - m_f\right)\,q_f\,. \tag{1.22}$$

It is worthwhile to decompose the Lagrangian into its different pieces,

$$\begin{aligned}\mathcal{L}_{\rm QCD} = &-\frac{1}{4}\left(\partial^\mu G_a^\nu - \partial^\nu G_a^\mu\right)\left(\partial_\mu G_\nu^a - \partial_\nu G_\mu^a\right) + \sum_f \overline{q}_f^\alpha \left(i\gamma^\mu \mathcal{D}_\mu - m_f\right) q_f^\alpha \\ &- g_s\, G_a^\mu \sum_f \overline{q}_f^\alpha\, \gamma_\mu \left(\frac{\lambda^a}{2}\right)_{\alpha\beta} q_f^\beta \\ &+ \frac{g_s}{2} f^{abc}\left(\partial^\mu G_a^\nu - \partial^\nu G_a^\mu\right) G_\mu^b\, G_\nu^c - \frac{g_s^2}{4} f^{abc} f_{ade}\, G_b^\mu\, G_c^\nu\, G_\mu^d\, G_\nu^e\,.\end{aligned} \tag{1.23}$$

The first line in Eq. 1.23 contains the correct kinetic term for the different fields, which gives rise to the corresponding propagators. The colour interaction between quarks and gluons is given by the second line. It involves the $SU(3)_{\rm C}$ matrices λ^a. Finally, the $G_a^{\mu\nu}\, G_{\mu\nu}^a$ term generates the cubic and quartic gluon self-interactions shown in the last line. The strength of these interactions is given by the same coupling g_s which appears in the fermionic piece of the Lagrangian.

In spite of the rich physics contained in it, the Lagrangian in Eq. 1.23 looks very simple because of its colour symmetry properties. All interactions are given in terms of a single universal coupling g_s, which is called the strong coupling constant. The existence of self-interactions among the gauge fields is a new feature that was not present in QED. Hence, it seems then reasonable to expect that these gauge self-interactions could explain properties like asymptotic freedom (strong interactions become weaker at short distances) and confinement (the strong forces increase at large distances), which do not appear in QED.

1.2.4 Electroweak Unification

The electroweak (EW) theory unifies the weak and electromagnetic interactions. Historically, the basic structure was formulated by Sheldon Glashow (1961) [5] and the complete form was developed by Steven Weinberg (1967) [6] and Abdus Salam(1968) [7].

In the Glashow-Weinberg-Salam model left-handed leptons and quarks are assigned to weak isospin doublets, i.e., they transform as doublets under $SU(2)_L$, where the L subindex denotes the left-handed property of the $SU(2)$ symmetry group. Massive leptons and quarks can exist in right-handed states and these are assigned to weak isospin singlets. They are unaffected by $SU(2)_L$ transformations. The EW theory assumes zero mass for the neutrinos which are therefore uniquely assigned to left-handed doublets. In order to incorporate the electric charge and bring about the unification of the weak and electromagnetic interactions a new

gauge symmetry, $U(1)_Y$, was introduced. It is a $U(1)$ symmetry similar to QED based on the hypercharge Y.

An overview of all SM fermions and their electroweak quantum numbers: charge Q, weak isospin T, its third component T_3 and hypercharge Y, are given in Tab. 1.3. Those are related through the analogue of the well-known Gell-Mann - Nishijima relation,

$$Q = T_3 + \frac{Y}{2}\,. \tag{1.24}$$

The left-handed isospin doublets ψ_L and the right-handed isospin singlets ψ_R transform under the action of the $SU(2)_L \otimes U(1)_Y$ direct product group as follows,

$$\begin{aligned} &\psi_L \to \psi'_L = \mathrm{e}^{i\alpha^a(x)\,T_a + i\beta(x)Y}\,\psi_L\,, \quad a\,=\,1,\,2,\,3 \\ &\psi_R \to \psi'_R = \mathrm{e}^{i\beta(x)Y}\,\psi_R \end{aligned} \tag{1.25}$$

where $\alpha^a(x)$ and $\beta(x)$ are local phases and $T_a/2$ and Y are the generators of the $SU(2)_L$ and $U(1)_Y$ groups of gauge transformations, respectively. Since there are now four gauge parameters, $\alpha^a(x)$ and $\beta(x)$, there are as well four different gauge fields needed.

TABLE 1.3: Electroweak quantum numbers for the SM fermions. The charge is denoted by Q and is given in units of the electron charge. The weak isospin is represented by T, its third component by T_3 and the hypercharge by Y [1].

	Generation			**Quantum numbers**			
Fermions	1^{st}	2^{nd}	3^{rd}	T	T_3	Y	$Q[e]$
Leptons	$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L$	$1/2$ $1/2$	$1/2$ $-1/2$	-1 -1	0 -1
	e^-_R	μ^-_R	τ^-_R	0	0	-2	-1
Quarks	$\begin{pmatrix} u \\ d' \end{pmatrix}_L$	$\begin{pmatrix} c \\ s' \end{pmatrix}_L$	$\begin{pmatrix} t \\ b' \end{pmatrix}_L$	$1/2$ $1/2$	$1/2$ $-1/2$	$1/3$ $1/3$	$2/3$ $-1/3$
	u_R	c_R	t_R	0	0	$4/3$	$2/3$
	d_R	s_R	b_R	0	0	$-2/3$	$1/3$

We can write the free Lagrangian for the left-handed fermion doublet and a right-handed singlet as,

$$\mathcal{L} = i\overline{\psi}_L\gamma^\mu\partial_\mu\psi_L + i\overline{\psi}_R\gamma^\mu\partial_\mu\psi_R\,. \tag{1.26}$$

The Lagrangian in Eq. 1.26 can be made invariant by introducing the covariant derivative,

$$\mathcal{D}_\mu = \partial_\mu + i\frac{g}{2}W^a_\mu T_a + i\frac{g'}{2}B_\mu Y\,, \tag{1.27}$$

where W_μ^a ($a = 1, 2, 3$) and B_μ denote the gauge fields related to the $3 + 1$ degrees of freedom from the $SU(2)_L \otimes U(1)_Y$ symmetry group. The coupling constants g and g' determine the strength of the coupling to the $SU(2)_L$ and $U(1)_Y$ gauge fields, respectively. The corresponding field strength tensors of the gauge fields are given by,

$$\begin{aligned} W_a^{\mu\nu} &= \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g\epsilon_{abc} W_\mu^b W_\nu^c \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \,, \end{aligned} \tag{1.28}$$

where ϵ_{abc} denotes the totally antisymmetric tensor.

Finally, we introduce the kinetic terms for he gauge fields, $-\frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu}$ and $-\frac{1}{4} B_{\mu\nu} B^{\mu\nu}$, and we arrive at the gauge invariant EW Lagrangian,

$$\mathcal{L}_{\text{EW}} = \sum_j i\overline{\psi}_L^j \gamma^\mu \mathcal{D}_\mu \psi_L^j + \sum_k i\overline{\psi}_R^k \gamma^\mu \mathcal{D}_\mu \psi_R^k - \frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \,. \tag{1.29}$$

where the sum in i and k runs over all doublets and singlets listed in Tab. 1.3. In addition to the kinetic energy term, self-coupling of the $W_{\mu\nu}$ fields also appear in the theory. Note that mass terms such as $\frac{1}{2} m^2 B_\mu B^\mu$ are not gauge invariant and therefore can not be added to the Lagrangian.

The gauge fields $W_a^{\mu\nu}$ and B_μ do not carry the experimentally observed quantum numbers for the $W^\pm$ and Z bosons and the photon, hence they can not directly be identified with these elementary particles. Instead, a linear combination of these gauge fields leads to the physically observable states according to,

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp i W_\mu^2 \right) \,, \\ Z_\mu &= \cos\theta_W W_\mu^3 - \sin\theta_W B_\mu \,, \\ A_\mu &= \sin\theta_W W_\mu^3 + \cos\theta_W B_\mu \,, \end{aligned} \tag{1.30}$$

where $W_\mu^\pm$ and Z_μ denote the fields of the weak gauge bosons, A_μ the photon field and θ_W the weak mixing angle. The combinations above allow to relate the electric charge e and the electroweak couplings by,

$$e = g \sin(\theta_W) = g' \cos(\theta_W) \,. \tag{1.31}$$

The principle of local gauge invariance works beautifully for the QED and QCD interactions. Moreover, it allows to unify the weak and electromagnetic interactions while keeping the renormalisability of the theory [8, 9]. Nevertheless, its

application in the EW theory was stymied by the fact that the gauge fields have to be massless. Whereas the photon and the gluons are massless, the $W^{\pm}$ and Z bosons as well as the leptons, are indeed massive objects [1]. Hence, it is necessary to introduce a mechanism into the model to give mass to the gauge bosons. This mechanism is known as the Brout-Englert-Higgs (BEH) mechanism.

1.3 The BEH Mechanism in the SM

The EW Lagrangian given in Eq. 1.29 does not fully describe the reality. As was discussed in Section 1.2.4, gauge bosons have to be massless particles to keep the invariance of the theory. While it is fine for the photon in QED, the physical $W^{\pm}$ and Z bosons are quite heavy particles ($\sim 100\,\text{GeV}$, see Tab. 1.1). In this section, the concept of the spontaneous symmetry breaking (SSB) will be discussed as this is the base of the BEH mechanism.

1.3.1 Spontaneous Symmetry Breaking

In order to generate masses, we need to break the gauge symmetry in some way. However, we also need a fully symmetric Lagrangian to preserve renormalisability. This dilemma is solved by getting non-symmetric results from an invariant Lagrangian. Imagine a invariant Lagrangian which has a degenerate set of states with minimal energy. If one of those states is arbitrarily selected as the ground state of the system, it is said that there is spontaneous symmetry breaking (SSB). This is one of the most important concepts in quantum field theory (QFT). In a QFT, the ground state is the vacuum, thus the SSB mechanism will appear when there is a symmetric Lagrangian, but a non-symmetric vacuum.

To illustrate the SSB concept, let us consider the following Lagrangian,

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - V(\phi)\,, \qquad V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4\,. \tag{1.32}$$

This Lagrangian has reflexion symmetry: it is invariant under the $\phi \to -\phi$ operation. In order to have a ground state, the potential should be bounded from below as $\phi \to \infty$, i.e., the parameter λ has to be positive. Then, we can find the minimum of the potential by setting,

$$\frac{\partial V(\phi)}{\partial\phi} = \phi\left(\mu^2 + \lambda\phi^2\right) = 0\,, \tag{1.33}$$

which bring to the next two possibilities:

- Case 1: $\mu^2 \geq 0$. The potential has only the trivial minimum $\phi = 0$. Then the Lagrangian describes a spin-zero particle of mass μ and quartic coupling λ.
- Case 2: $\mu^2 < 0$. The minimum is obtained for those field configurations satisfying,

$$\phi = \pm\sqrt{\frac{-\mu^2}{\lambda}} \equiv \pm v \neq 0 \tag{1.34}$$

Figure 1.2 shows the potential $V(\phi)$ for the two cases discussed above. As it can be seen, the left plot has the minimum at $\phi = 0$ (represented by the green vertical line). However, in the second case (right plot), the $\phi = 0$ point is not a minimum. Instead, there are two minima at $\phi = \pm v$, which are, obviously, non-zero values. Since the field ϕ takes on the value v in the ground state, v is called the vacuum expectation value (VEV) of the field ϕ.

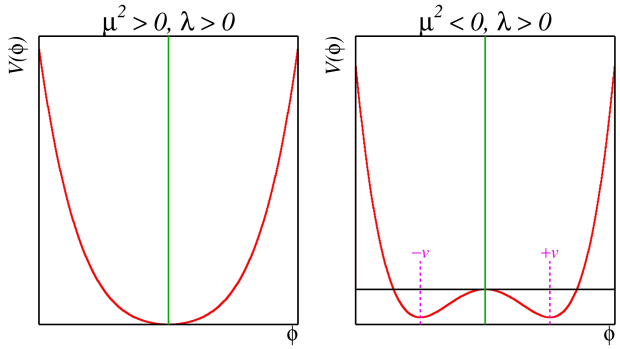

FIGURE 1.2: The potential $V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{2}\lambda\phi^4$ for $(\mu^2 > 0\,,\, \lambda > 0)$ and $(\mu^2 < 0\,,\, \lambda > 0)$ on the left and right, respectively.

To determine the particle spectrum, we must study the theory in the region of the minimum,

$$\phi(x) = v + \eta(x) \tag{1.35}$$

so we are expanding around $\eta = 0$. Substituting Eq. 1.35 into the Lagrangian $\mathcal{L}$ in Eq. 1.32 we obtain,

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\eta\,\partial^\mu\eta\right) - \left(\lambda v^2\eta^2 + \lambda v\eta^3 + \frac{1}{4}\lambda\eta^4\right) + \text{constant}\,. \tag{1.36}$$

This Lagrangian represents the description of a particle of mass $m_\eta^2 = 2\lambda v^2 = -2\mu^2$, and two interactions: a cubic one of strength λv and a quartic one of strength $\frac{\lambda}{4}$.

The two descriptions of the theory in terms of ϕ or η must be equivalent if the problem is solved exactly. The scalar particle described by the theory with $\mu^2 < 0$ is a real scalar because at the minimum of the potential there is a non-zero VEV.

The obtained Lagrangian in Eq. 1.36 seems not to preserve the reflection symmetry in the new field η because of the cubic term. Nevertheless, the original symmetry is still there but not in an obvious way. All we did was to add a constant shift to the field, so the physics described by both Lagrangians in Eq. 1.32 and in Eq. 1.36, have to be the same. The only piece that does not preserve the symmetry is the choice of a specific VEV value, i.e., the fact of selecting $+v$ instead of $-v$ in Eq. 1.35. Since for each possible ground state there corresponds to an equivalent physical theory, any one of them can serve as the ground state. However, the selection of one vacuum state means that it is not longer invariant under the action of the symmetry group. When it happens, it is said that the symmetry is spontaneously broken, which is an unfortunate description since the symmetry is not really broken, just expressed differently. The phenomenon described here is known as spontaneous symmetry breaking.

1.3.2 The Goldstone Theorem

Now, let us consider a complex scalar field $\phi \equiv \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$, described by the Lagrangian,

$$\mathcal{L} = \partial_\mu \phi^\dagger \, \partial^\mu \phi - V(\phi)\,, \quad V(\phi) = \mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \, \phi\right)^2 . \tag{1.37}$$

$\mathcal{L}$ is invariant under global phase transformations of the scalar field $\phi(x) \rightarrow \phi'(x) \equiv \mathrm{e}^{i\theta}\,\phi(x)$, where θ is a constant. Written in terms of ϕ_1 and ϕ_2 the Lagrangian becomes,

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi_1)^2 + \frac{1}{2}(\partial_\mu \phi_2)^2 - V(\phi_1, \phi_2)\,, \quad V(\phi_1, \phi_2) = \frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2) + \frac{1}{4}\lambda(\phi_1^2 + \phi_2^2)^2 . \tag{1.38}$$

As in Section 1.3.1, considering λ positive, there are two possibilities for the minimum condition of the potential. The first possibility, $\mu^2 > 0$, is just the usual situation with a single ground state.

The other case, $\mu^2 < 0$, with SSB, is more interesting. The potential has the minimum along a circle of radius,

$$\phi_1^2 + \phi_2^2 = \frac{-\mu^2}{\lambda} \equiv v^2 . \tag{1.39}$$

This potential is represented in Fig. 1.3.

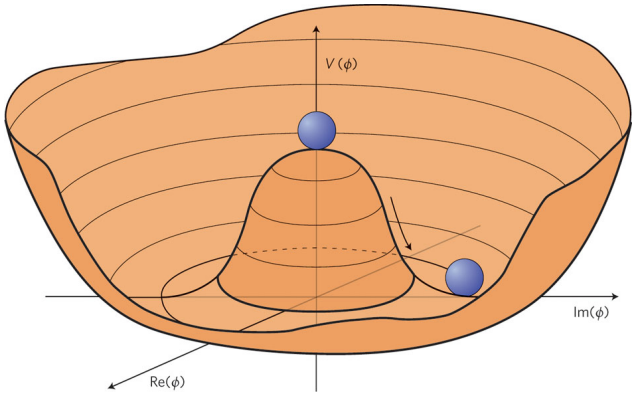

FIGURE 1.3: The potential for a complex scalar field ϕ. Note that the minima of $V(\phi)$, when $\mu^2 < 0$ and $\lambda > 0$, have infinite degenerate possible values along the circle of radius v (see text) on the real components of the field plane.

As before, to expand around v^2, we have to choose some point on the circle, which will break the symmetry for the solutions. Let us pick up the point $\phi_1 = v$, $\phi_2 = 0$, and write, with η and ϕ real,

$$\phi = \frac{1}{\sqrt{2}}(v + \eta(x) + i\rho(x))\,. \tag{1.40}$$

Substituting this in Eq. 1.38, we again find a Lagrangian that can be interpreted in terms of particles and their interactions,

$$\begin{aligned}\mathcal{L} =& \frac{1}{2}(\partial_\mu\rho)^2 + \frac{1}{2}(\partial_\mu\eta)^2 \\ &+ \mu^2\eta^2 - \frac{1}{2}(\mu^2 + \lambda v^2)\rho^2 \\ &- \lambda\nu(\eta\rho^2 + \eta^3) - \frac{\lambda}{2}\eta^2\rho^2 - \frac{1}{4}\eta^4 - \frac{\lambda}{4}\rho^4 \\ &+ \text{constant}\,.\end{aligned} \tag{1.41}$$

The terms in the first line are normal kinetic terms. The first term in the second line tells us that the η field corresponds to a particle of $m_\eta^2 = 2|\mu^2|$. Remarkably, the second term in ρ^2 vanished, since the value inside the parenthesis is null by definition, implying that the ρ field particle is massless. The fact that there are massless excitations associated with the SSB mechanism is a completely general result, known as the Goldstone theorem [10]: if a Lagrangian which is invariant under a continuous symmetry group G (in our case it was $U(1)$) is spontaneously

broken (due to the choice of a particular ground state), then there must exist as many massless spin-0 particles (Goldstone bosons) as broken generators. It is clear how the massless particle arises: the potential is a minimum along a circle. Along this circle the potential is flat, so there is no resistance to motion around the circle, which is the meaning of the massless excitation.

1.3.3 The BEH Mechanism

At first sight, the Goldstone theorem discussed in Section 1.3.2 has very little to do with the mass problem of the EW theory. In fact, it makes it worse in the sense that we need massive states and not massless ones. However, something very interesting happens when we consider the Lagrangian invariant under local gauge transformations. From Section 1.2.2, we know that the local gauge transformations require the introduction of a vector field A_μ and the Lagrangian should be written in terms of the covariant derivative,

$$\mathcal{L} = \left(\mathcal{D}_\mu \phi\right)^\dagger \left(\mathcal{D}_\mu \phi\right) - \mu^2 \phi^\dagger \phi - \lambda \left(\phi^\dagger \phi\right)^2 , \tag{1.42}$$

where we have not written the kinetic term $(-\frac{1}{4} F_{\mu\nu} F^{\mu\nu})$ since it does not enter in the analysis. We want to choose $\mu^2 < 0$ and write conveniently the field $\phi(x)$ profiting from the local gauge invariance,

$$\phi(x) = \frac{v + h(x)}{\sqrt{2}} , \tag{1.43}$$

with h being a real scalar field. Then the Lagrangian now takes the form,

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \left(\partial_\mu h\right) \left(\partial^\mu h\right) + \frac{1}{2} g^2 v^2 A_\mu A^\mu - \lambda v^2 h^2 - \lambda v h^3 \\ & - \frac{1}{4} h^4 + g^2 v h A^\mu A_\mu + \frac{1}{2} g^2 h^2 A_\mu A^\mu . \end{aligned} \tag{1.44}$$

One can see in the second term of Eq. 1.44 that the gauge boson A has now acquired mass $M_A = gv$. Note that it is non-zero only when the gauge symmetry is spontaneously broken, i.e., because of selecting a particular VEV.

The massless Goldstone boson of Section 1.3.2 has now become the longitudinal polarisation state of the gauge boson. This phenomenon is sometimes referred to as the gauge boson having "eaten" the Goldstone boson. The mechanism described is the so-called Brout-Englert-Higgs (BEH) mechanism which was developed in 1964

by three independent groups: by Robert Brout and François Englert [11]; by Peter Higgs [12]; and by Gerald Guralnik, Carl Hagen, and Thomas Kibble [13].

From the BEH mechanism, there arises a new single real boson h. From the third term in Eq. 1.44, we can see that it has a mass $M_h = \sqrt{2\lambda v^2}$. It is the so-called Higgs or BEH boson.[2] Note that the gauge boson mass is fixed if g^2 and v are known, but the mass of the Higgs boson is not predicted by the model since it depends on the strength parameter λ.

To apply the BEH mechanism to the EW theory, one has to extend the Higgs field in the group symmetry of $SU(2)$,

$$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} , \tag{1.45}$$

where ϕ^+ and ϕ^0 are each complex fields,

$$\phi^+ = \frac{1}{\sqrt{2}} (\phi_1 + i\phi_2) , \qquad \phi^0 = \frac{1}{\sqrt{2}} (\phi_3 + i\phi_4) . \tag{1.46}$$

The Lagrangian in the $SU(2)$ space has the same form,

$$\mathcal{L} = \partial_\mu \phi^\dagger \, \partial^\mu \phi - V(\phi) , \quad V(\phi) = \mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \, \phi \right)^2 . \tag{1.47}$$

As before we want to study the potential which is invariant under the local gauge transformations,

$$\phi(x) \rightarrow \phi'(x) = \mathrm{e}^{i \, \vec{\alpha}(x) \, \vec{\tau}/2} \, \phi(x) , \tag{1.48}$$

where τ_i are the Pauli matrices and α_i are parameters. Proceeding as before in Section 1.3.1 and 1.3.2, one can see that there is an infinite set of degenerate states with minimum energy, satisfying,

$$\phi^\dagger \phi = \frac{-\mu^2}{2\lambda} = \frac{v^2}{2} . \tag{1.49}$$

By choosing a direction in SU(2) space, the symmetry gets spontaneously broken. Considering one of the possible vacuum states, the appropriate choice is,

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} , \tag{1.50}$$

[2] In this thesis this boson will be referred to Higgs boson as that term has long since passed into common parlance, as was decided by the CERN Council in 2012 [14].

corresponding to $\phi_3 = v$ and $\phi_1 = \phi_2 = \phi_4 = 0$. Once more, we can study the particle spectrum by expanding around the ground state,

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} . \tag{1.51}$$

Again, we benefit from the local gauge invariance to make this simple choice. This amounts to three fields coming from the Goldstone theorem when we have chosen a particular vacuum state. Below we will see that these three massless bosons are just what are needed for the longitudinal parts of the $W^{\pm}$ and Z bosons. Then, adding the $U(1)_Y$ symmetry and writing the covariant derivative, the Lagrangian of Eq. 1.47 can be rewritten. The piece generating the gauge boson masses is,

$$(\mathcal{D}^{\mu}\phi)^{\dagger} \left(\mathcal{D}_{\mu}\phi\right) = \frac{v^2}{8} \left[g^2 \left((W_{\mu}^1)^2 + (W_{\mu}^2)^2 \right) + \left(g W_{\mu}^3 - g' B_{\mu} \right)^2 \right] . \tag{1.52}$$

Regarding the relations given in Eq. 1.30, the first term in Eq. 1.52 becomes,

$$\left(\frac{1}{2} v g_2 \right)^2 (W^+)_{\mu} (W^-)^{\mu} , \tag{1.53}$$

yielding the W mass,

$$m_W = \frac{gv}{2} . \tag{1.54}$$

Similarly, for the massive neutral gauge boson, Z, we obtain,

$$m_Z = \frac{v}{2} \sqrt{g^2 + g'^2} . \tag{1.55}$$

Finally, since no $A_{\mu}A^{\mu}$ appears, $M_A = 0$ which is expected since it is the photon term.

Thus, by SSB of the symmetry group $SU(2)_L \times U(1)_Y$, three Goldstone bosons have been absorbed by the $W^{\pm}$ and Z bosons to form their longitudinal components. Now these gauge bosons have acquired a proper mass term in the Lagrangian. Since the $U(1)_Y$ symmetry is still unbroken, the photon which is its generator, remains massless, as it should be. There is a new massive fundamental scalar boson predicted by the model: the Higgs boson. Its mass is given in terms of the signal strength, $M_H = v\sqrt{2\lambda}$.

Finally, the BEH mechanism provides a simple way to have massive leptons and quarks in the SM as well. Since we have introduced an additional scalar doublet into the model, we can add an interaction term for the leptons into the EW

Lagrangian,

$$\mathcal{L}_{int} = g_e \left(\overline{L} \phi e_R^- + \phi^\dagger \overline{e_R^-} L \right) \tag{1.56}$$

where the second term is the Hermitian conjugate of the first. Since $L = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}$ and $\phi = \begin{pmatrix} \phi^+ \phi^0 \end{pmatrix}$, the term $\overline{L}\phi = \overline{\nu_{eL}}\phi^+ + \overline{e_L^-}\phi^0$ is an $SU(2)$ invariant.

Following the previous analysis for the EW gauge bosons, we can calculate the mass spectrum by replacing,

$$\phi \to \begin{pmatrix} 0 \\ \frac{v+H}{\sqrt{2}} \end{pmatrix} , \tag{1.57}$$

where v is the Higgs VEV and H represents the Higgs boson. Substituting this into Eq. 1.56 gives,

$$\mathcal{L}_{int} = \frac{g_e v}{\sqrt{2}} \left(\overline{e_L^-} e_R^- + \overline{e_R^-} e_L^- \right) + \frac{g_e}{\sqrt{2}} \left(\overline{e_L^-} e_R^- + \overline{e_R^-} e_L^- \right) H \, , \tag{1.58}$$

where the first term has exactly the expected form for the mass of the electron, so

$$m_e = \frac{g_e v}{\sqrt{2}} \, . \tag{1.59}$$

Since g_e is arbitrary, the value of the electron mass is not predicted by the theory. The second term in Eq. 1.58 contains the electron-Higgs interaction, whose strength is proportional to the electron mass. The rest of the leptons and quarks masses can be generated in the same way.

1.4 The Higgs Boson Mass Constraints

The SM is the combination of the previously discussed theories of EW and strong interactions. Its Lagrangian has the symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$. After spontaneous symmetry breaking part of this symmetry group is reduced. $SU(2)_L \times U(1)_Y$ becomes $U(1)_{em}$ which is the usual group of classical QED.

The Higgs boson is predicted by the theory when the BEH mechanism is applied to the EW Lagrangian in order to get massive weak bosons. The Higgs boson mass is unpredicted by the SM theory as was discussed in Section 1.3.3. Nevertheless, constraints can be derived from internal consistency conditions. Upper bounds on the mass can be generated by assuming that the SM can be extended

up to a scale Λ before perturbation breaks down and new dynamical phenomena emerge.

There are several ways to get constraints on the Higgs mass despite the general lack of prediction from the theory, the first of which comes from one-loop EW radiative corrections. The Higgs boson contributes to radiative corrections on the top quark and W boson masses. Therefore, precision measurements of EW parameters, like the top quark and W boson masses or the weak mixing angle (θ_W) of the W and Z bosons can constrain the Higgs boson mass.

Combining the high precision measurements of these masses taken at the LEP [15] and Tevatron [16] colliders, leads to a $\Delta\chi^2$ fit of the Higgs boson mass. Figure 1.4 shows this fit, from the measurements mentioned above, as a function of the Higgs boson mass. The preferred value for the Higgs boson mass is at 87^{+35}_{-26} GeV at a 68% confidence level (CL), which corresponds to the minimum of the fitting curve. This is not a proof that the Higgs boson exists but it gives an idea for the mass range in which it is expected. The upper limit on the Higgs mass is quoted for $m_{\mathrm{H}} < 186\,\mathrm{GeV}$ at 95% CL and direct experimental searches by the LEP experiment placed a lower limit of $m_{\mathrm{H}} > 114.4\,\mathrm{GeV}$ at 95% CL [17].

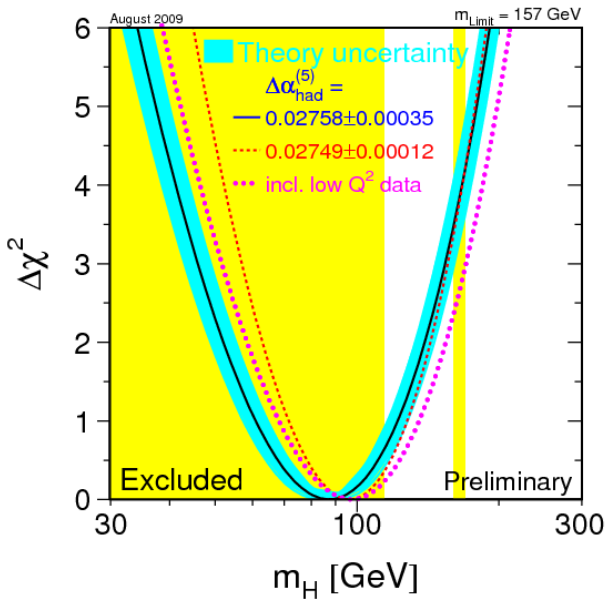

FIGURE 1.4: Limits on the Higgs mass within the SM from precision electroweak constraints, and direct Higgs searches by the LEP and Tevatron experiments.

1.5 Phenomenology of Proton-Proton Collisions

The calculation of production cross sections at proton colliders takes into account the fact that protons are composite objects. The process of interest takes place between the constituents of the protons but these processes are accompanied by interactions of the residual objects. The aim of this section is to give a brief summary of the most important aspects of proton-proton collisions. A detailed description can be found in Ref. [18].

1.5.1 General Aspects of Proton-Proton Collisions

Protons are not fundamental particles, instead they are composed of elementary constituents known as partons which are valence quarks, sea quarks and gluons. In proton-proton collisions, these basic constituents interact with each other. The dominant processes are strong interactions which are described by the theory of QCD. The amount of transferred momentum in the parton interaction allows to classify the processes as hard or soft. In the case of hard scattering, the cross section can be calculated using perturbation theory due to the small strong coupling parameter. Oppositely, for soft processes the strong coupling strength becomes larger and the scattering has to be calculated using non-perturbative QCD. The soft scattering constitutes the majority of the cases in proton-proton collisions. A hard scattering process is usually accompanied by soft interactions which occur with the partons not participating in the hard scatter process.

A simplified representation of an interaction of two partons a and b, which are constituents of the protons A and B, respectively is given in Fig. 1.5.

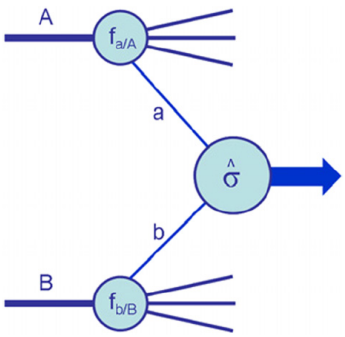

FIGURE 1.5: Representation of the structure of a generic hard-scattering process from Ref. [19].

The hard scatter process $ab \to X$ produces some final state X and hadronised proton remnants. The latter are referred to as the underlying event (UE). The cross section of the hadrons A and B can be expressed by,

$$\sigma_{AB} = \sum_{a,b} \int dx_a \, dx_b \, f_{a/A}(x_a, Q^2) f_{b/B}(x_b, Q^2) \hat{\sigma} \,, \tag{1.60}$$

where $\hat{\sigma}$ denotes the partonic cross section of the initial state partons a and b. The parton distribution function (pdf) is given by $f_{a/A}(x_a, Q^2)$ ($f_{b/B}(x_b, Q^2)$). This describes the probability to find a parton a (b) carrying the longitudinal momentum fraction x_a (x_b) at moment transfer Q^2 of the hadron A (B). The pdfs are not predicted by QCD perturbation theory but they can be measured in inelastic scattering experiments for a given Q^2 [20]. Figure 1.6 shows the pdfs of the proton for different transfer scales Q^2.

In the hard scattering many quarks and gluons are produced. The partons carry colour charge, so they can radiate through bremsstrahlung process. The quarks can radiate gluons which can radiate as well or create $q\overline{q}$ pairs, forming a parton shower (PS). Such PS can be also produced from the initial state partons which do not take part in the hard scatter process. When the emissions come from the incoming partons, they are known as initial state radiation (ISR). If the emission is related to the outgoing partons, then it is called final state radiation (FSR).

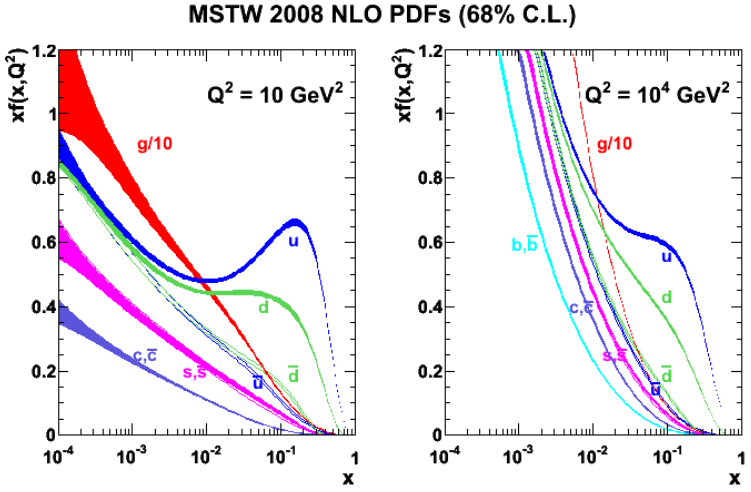

FIGURE 1.6: Parton distribution functions of the proton as determined for the MSTW08 PDF set for $Q^2 = 10\,\mathrm{GeV}^2$ and $Q^2 = 10^4\,\mathrm{GeV}^2$ on the left and right, respectively. The bands reflect the uncertainties at the 68% confidence level [20].

1.5.2 Luminosity

In colliders the luminosity (L) is the ratio of the number of events detected (N) of one type in a certain time (t) to the interaction cross section (σ),

$$L = \frac{1}{\sigma}\frac{dN}{dt}\,. \tag{1.61}$$

The luminosity has dimensions of events per time and per area. This is measured in units of $\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

Today's colliders all employ bunched beams. If two bunches containing n_1 and n_2 particles colliding head-on with frequency f_{coll}, a basic expression for the luminosity is,

$$L = f_{coll}\frac{n_1 n_2}{4\pi\sigma_x\sigma_y}\,, \tag{1.62}$$

where σ_x and σ_y characterise the transverse beam sizes in the horizontal and vertical directions. The expected number of events (N_{exp}) is the product of the cross section of interest (σ) and the time integral over the instantaneous luminosity ($\mathcal{L}$),

$$N_{exp} = \sigma \times \int L\,dt = \sigma \times \mathcal{L}, \tag{1.63}$$

where the quantity $\mathcal{L}$ is called integrated luminosity which is used to quantify the total amount of collisions in a period of time. Cross sections are usually measured in units of barns. [3]

1.5.3 Expected Cross Section at Hadron Colliders

The formalism described in the previous sections is used to obtain predictions for some SM cross sections at hadron colliders. Figure 1.7 presents an overview of different cross sections for some processes as a function of the center-of-mass energy for the Tevatron and the LHC colliders.

The total inelastic proton-proton cross section (σ_{tot}) is about 6 orders of magnitude higher than the cross section of W or Z bosons. The Higgs boson with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ production cross section is predicted to be about ten to eleven, depending on the production mechanism, orders of magnitude smaller than the total inelastic proton-proton cross section. Thus, very high luminosities are needed to produce sufficient rate of such processes.

[3] $1\,\mathrm{barn} = 10^{-28}\,\mathrm{m}^2 = 10^{-24}\,\mathrm{cm}^2$.

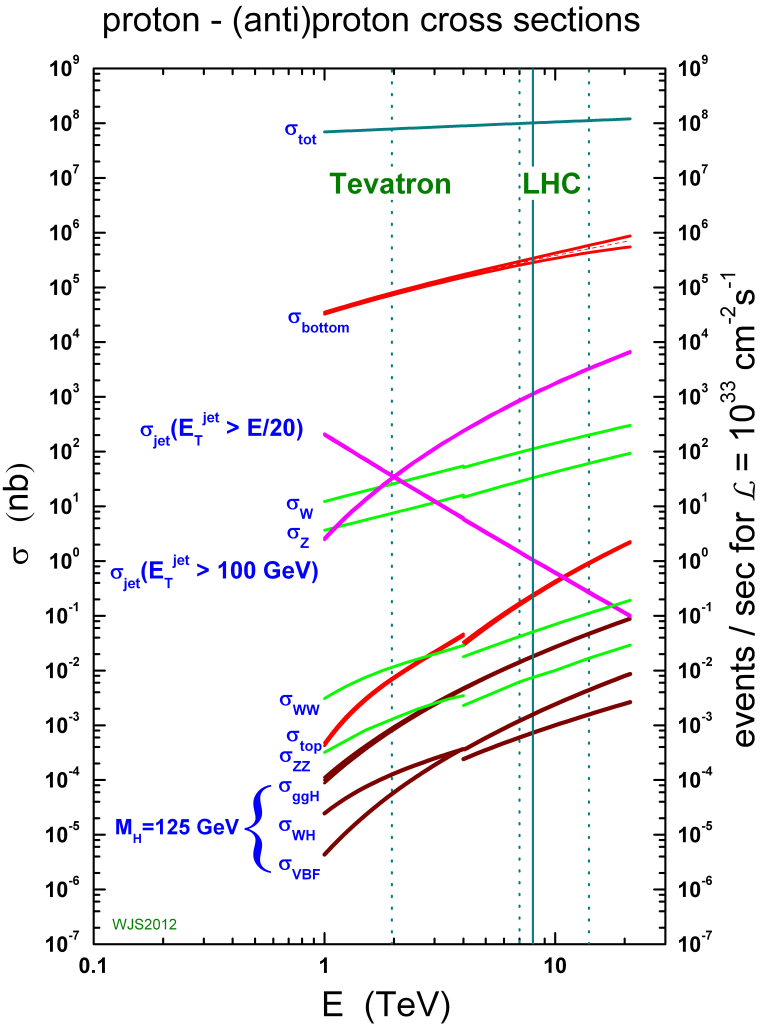

FIGURE 1.7: Cross sections and expected number of events for an integrated luminosity of $10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ of SM processes in proton-proton (LHC) and anti-proton-proton (Tevatron) collisions as a function of the center-of-mass energy [18].

1.6 Higgs Boson Production Modes

In the SM, the main production mechanisms for the Higgs boson at hadron colliders make use of the fact that the Higgs boson couples preferentially to heavy particles. It includes the massive W and Z vector bosons, the top quark and, to a lesser extent, the bottom quark. Other production processes, such as the associated production

with a single top, b-quarks and diffractive production play minor roles and are not considered here.

The four main production modes are thus:

- **Gluon-gluon fusion:** $gg \rightarrow H$
- **Vector boson fusion:** $qq \rightarrow V^*V^* \rightarrow qq + H$
- **Associated production with W/Z bosons:** $q\overline{q} \rightarrow V + H$
- **Associated production with top quark:** $gg,\, q\overline{q} \rightarrow t\bar{t}H$

The Feynman diagrams of these four main production mechanisms processes are shown in Fig. 1.8.

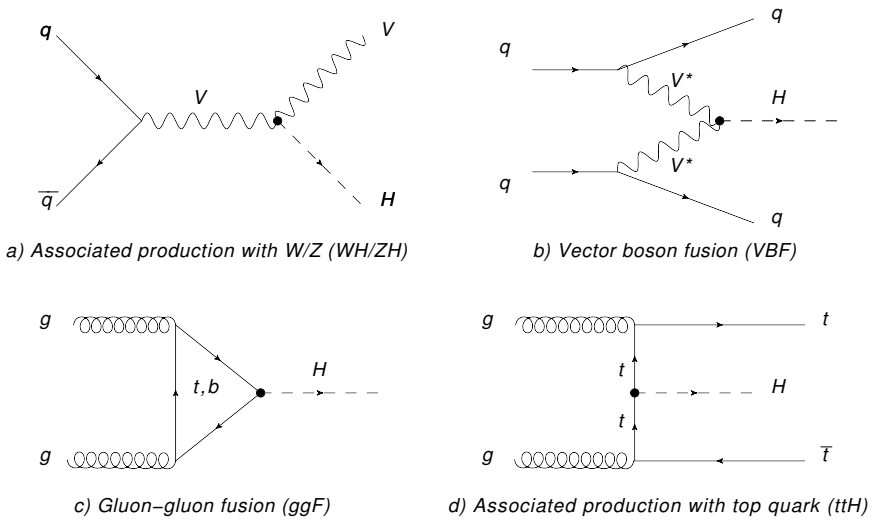

FIGURE 1.8: The dominant SM Higgs boson production modes in hadron colliders.

Figure 1.9 shows the cross sections for the dominant Higgs boson production as a function of the Higgs boson mass in proton-proton collisions at a center-of-mass energy of 8 TeV. The cross section decreases rapidly with increasing Higgs mass for all production modes.

Table 1.4 shows the cross section for the main production mechanisms for a Higgs boson mass of $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ in proton-proton collisions at a centre-of-mass of 8 TeV.

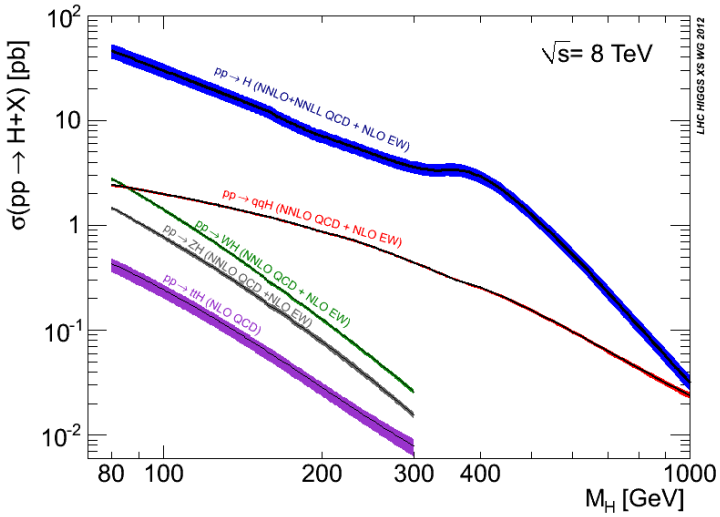

FIGURE 1.9: Higgs boson production cross sections in proton-proton collisions at a center-of-mass energy of 8 TeV.

TABLE 1.4: The dominant Higgs boson production mechanisms for proton-proton collisions and their cross section (σ) at a center-of-mass energy of 8 TeV for a Higgs boson mass with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. The corresponding errors are expressed in percent [21].

Production mode	Representation	σ (pb) ($m_{\mathrm{H}} = 125\,\mathrm{GeV}$)
ggF	$gg \rightarrow H$	$19.52^{+14.7\%}_{-14.7\%}$
VBF	$qq \rightarrow qqH$	$1.58^{+2.8\%}_{-3.0\%}$
WH	$qq \rightarrow WH$	$0.70^{+3.7\%}_{-4.1\%}$
ZH	$qq \rightarrow ZH$	$0.39^{+5.1\%}_{-5.0\%}$
ttH	$gg \rightarrow ttH$	$0.13^{+11.6\%}_{-17.0\%}$

1.6.1 Gluon-gluon Fusion Production

Gluon-gluon fusion (ggF) is the main production mechanism of Higgs bosons in high-energy proton-proton collisions throughout the entire mass range up to 1 TeV. The gluon coupling to the Higgs boson in the SM is mainly mediated by triangular loops of top quarks as shown in Figure 1.8-c). The loop can also be mediated by a bottom quark but the top quark is preferred because of its larger Yukawa coupling to the Higgs boson compared to the bottom quark.

The dynamics of the ggF mechanism is controlled by strong interactions. In QCD perturbation theory, its cross section is proportional to the square of the QCD

coupling constant. The ggF cross section has been calculated up to next-to-next-to-next-to-leading order (NNNLO) in QCD using the large-m_t limit [22, 23]. This approximation has been tested at next-to-leading order (NLO). This calculation has been compared to the exact Born cross section result, with the full dependence on the mass of the top quark, yielding differences only of a few percent [24]. The computation of the ggF cross section includes NLO electroweak corrections and QCD soft-gluon resummation up to next-to-next-to-leading logarithmic (NNLL) terms. The detailed description on the procedure is given in Ref. [25].

The Higgs boson can also be produced in association with one or more jets with high momenta. At leading order (LO) the processes are just $gg \to Hg$ and $gg \to Hgg$ which are mediated by triangles up to pentagon diagrams. Figure 1.10 presents some example Feynman diagrams for these processes.

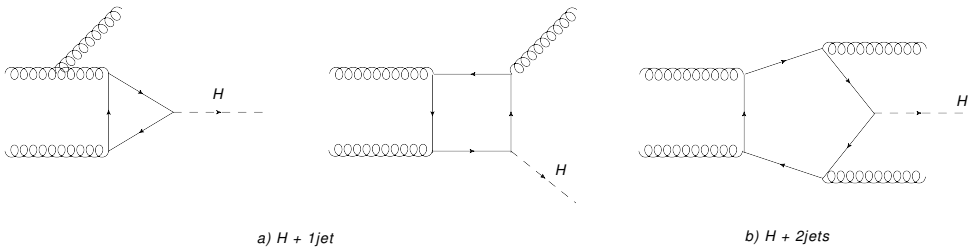

FIGURE 1.10: Feynman diagrams for the Higgs boson production through ggF in association with jets at proton-proton colliders.

The kinematics of Higgs signals in association with jets differ significantly from that of known SM backgrounds. These properties can be exploited to select corners of phase space where the expected signal-to-background ratios are larger than in a purely inclusive approach [26]. This is the basic strategy that follows the Higgs boson search in Chapter 5.

1.6.2 Vector Boson Production

The second dominant process arises from the vector boson fusion (VBF) mechanism, which is one order of magnitude lower than the ggF for a Higgs boson with mass $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, as shown in Fig. 1.9. This production mode is characterised by the presence of two forward jets arising from the two outgoing quarks. Figure 1.8 b) presents the Feynman diagram for the VBF production mechanism.

The production of a Higgs boson accompanied by two jets mainly receives contributions at hadron colliders from ggF, VBF and associated production modes. For the former, the Higgs boson couples to a weak boson that links two quarks lines,

which are dominated by t- and u-channel-like diagrams. For the latter, the two jets appear when the associated vector boson decays into a pair of quarks. In the genuine VBF channel, the hard jet pairs have a strong tendency to be forward-backward directed while the Higgs boson decay products are expected in the central region of the detector. This is in contrast to other jet-production mechanisms, as for example $t\bar{t}$ decay processes, offering a good background suppression. Since quarks from the incoming protons carry large momenta, the invariant mass of the two additional quarks in the VBF production mechanism is expected to be larger than for QCD processes where gluons are radiated off the incoming quarks. To measure the Higgs couplings in VBF, specific cuts should be applied in order to suppress events from Higgs accompanied by two jets via ggF, which becomes a new background to the VBF signal. In the ggF channel, as was discussed in Section 1.6.1, the Higgs boson can be radiated off a heavy-quark loop that couples to any parton of the incoming hadrons via gluons. Although the final states are similar, the kinematic distributions of the jets are different. Applying appropriate event selection criteria, it is possible to sufficiently suppress the ggF Higgs boson mechanism with respect to the VBF one. The ggF and VBF production modes are of particular interest for the analysis presented in Chapter 5.

1.6.3 Associated Production Modes

The next contributing production mechanism is the associated production with vector bosons represented by V, where $V = W, Z$ in Fig. 1.8 a). This is essentially a Drell-Yan process in which the $W^{\pm}$ or Z boson radiates a Higgs boson. Hence, these Higgs boson production processes are usually referred to as Higgs-strahlung. The cross section of the associated production mode with vector bosons is three (WH) and five (ZH) orders of magnitude lower than the ggF.

The last mechanism is the associated production with the top quark. This is the lowest contributing production process and its corresponding Feynman diagram is shown in Fig. 1.8 d).

1.7 Higgs Boson Decay Modes

In the SM, once the Higgs boson mass is fixed, the properties of the Higgs particle is uniquely determined. The Higgs couplings to gauge bosons are proportional to the square of the boson masses and the coupling to fermions are directly proportional

to the fermion masses,

$$g_{Hf\overline{f}} = \frac{m_f}{v}\,, \quad g_{HVV} = \frac{2m_V^2}{v}\,, \quad g_{HHVV} = \frac{2m_V^2}{v^2}$$
$$g_{HHH} = \frac{3m_H^2}{v}\,, \quad g_{HHHH} = \frac{3m_H^2}{v^2}\,. \tag{1.64}$$

The Higgs boson has the tendency to decay into the heaviest ones allowed kinematically. Thus, the dominant decay mechanisms involve the coupling to the Higgs boson to the weak bosons and/or the third generation of fermions.

Figure 1.11 shows the main Feynman diagrams for the Higgs boson decays into fermions and gauge bosons. Since photons and gluons are massless, they do not couple directly to the Higgs boson at tree level. Nevertheless, they can be generated via loops involving heavy virtual W bosons and heavy virtual quarks, as shown in the middle Feynman diagram in Fig. 1.11.

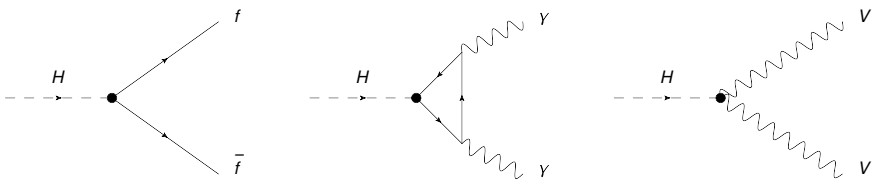

FIGURE 1.11: Feynman diagrams of the Higgs boson decay modes.

Figure 1.12 presents the branching ratio for different decay modes of the Higgs boson as a function of its mass in proton-proton collisions. Below the threshold for the production of a pair of W bosons, i.e. $m_{\mathrm{H}} < 2\,m_W$, the predominant decay channel is $H \to \overline{b}b$. The decays into a tau pair, a charm quark pair and into two gluons are one order of magnitude lower. All they together contribute less than $\sim 15\%$ on the total Higgs boson decay.

As shown in Fig. 1.12, above $m_{\mathrm{H}} \sim 130\,\mathrm{GeV}$ the dominant Higgs decay mode is into a pair of W bosons. Note that below the W^+W^- mass threshold the Higgs decay mode can be produced if one of the W bosons is virtual. The branching ratio for the $H \to ZZ$ decay reaches its maximum above the $m_{\mathrm{H}} \sim 200\,\mathrm{GeV}$ threshold. Similarly, the decay into a pair of top quarks (top and anti-top) starts being significant above a Higgs boson mass of 350 GeV. Table 1.5 presents an overview of the dominant Higgs boson decay modes for a Higgs boson mass $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ [21].

It is important to mention that besides the branching ratio, the properties and features of these decays play a central role. One clear example is the $b\overline{b}$ final

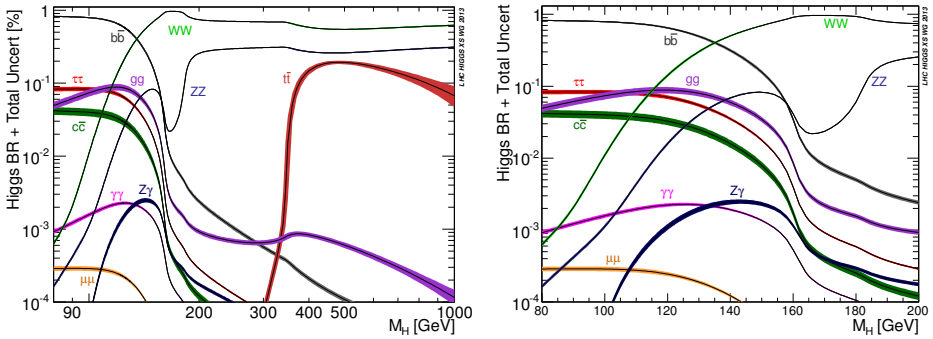

FIGURE 1.12: Branching ratios of the different Higgs boson decay modes as a function of the Higgs boson mass. The left plot extends the Higgs mass range up to 1 TeV. The right plot only shows the decay modes for the low-mass range, i.e., $m_{\mathrm{H}} < 200\,\mathrm{GeV}$.

TABLE 1.5: Branching ratios for the dominant Higgs boson decay modes for a Higgs boson mass $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ with the errors expressed in percent [21].

Decay mode	BR $\times 10^{-2}$ ($m_{\mathrm{H}} = 125\,\mathrm{GeV}$)
$H \to b\bar{b}$	$57.7^{+3.2\%}_{-3.3\%}$
$H \to WW^{(*)}$	$21.5^{+4.3\%}_{-4.2\%}$
$H \to gg$	$8.57^{+10.2\%}_{-10.0\%}$
$H \to \tau\bar{\tau}$	$6.32^{+5.7\%}_{-5.7\%}$
$H \to c\bar{c}$	$2.91^{+12.2\%}_{-12.2\%}$
$H \to ZZ^{(*)}$	$2.64^{+4.3\%}_{-4.2\%}$
$H \to \gamma\gamma$	$0.228^{+5.0\%}_{-4.9\%}$
$H \to \mu\bar{\mu}$	$0.0217^{+6.0\%}_{-5.8\%}$

state, for which we expect the highest branching ratio for a Higgs boson mass of $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, as shown in Tab. 1.5. In this case, the signal is inaccessible in the ggF mechanism due to the direct QCD production of b-quark pairs produced in proton-proton collisions. This search becomes more feasible in case of VBF or VH production modes since they provide additional characteristics that help to discriminate the SM background contamination. The counter-example is the $H \to \gamma\gamma$ decay which becomes one of the most powerful channels for the low mass search despite its very low branching ratio. The feasibility of this channel heavily relies on excellent photon resolution due to the expected small signal-to-background ratio for the inclusive analysis. In general, final states involving electrons and muons,

or photons are more beneficial for the search of the Higgs boson since they present a clearer signature in the environment of proton-proton collisions. In light of this, the $H \rightarrow ZZ^* \rightarrow 4\ell$ and the $H \rightarrow WW^{(*)} \rightarrow \ell\nu\ell\nu$ (with $\ell = e, \mu$) are expected to provide high sensitivity. The former is commonly referred to as the "Golden channel" as with four leptons in the final state the signal is easy to trigger on, thereby the signal-to-background ratio is increased. Moreover, this decay mode allows for full reconstruction of the Higgs boson mass. The latter benefits from high branching ratio for a wide mass range. The Higgs boson decaying into a W pair is the most significant channel at $m_{\mathrm{H}} \sim 160\,\mathrm{GeV}$ and it has considerable sensitivity at $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. The final state with two high p_{T} leptons is the search on which this thesis is focused. The analysis strategy is fully covered in Chapter 5.

Chapter 2

The ATLAS Experiment at the CERN Large Hadron Collider

2.1 Introduction

The European Organization for Nuclear Research (CERN) serves to provide particle accelerators and other infrastructure needed for high-energy physics research. It was founded in the northwest of Geneva in 1954 deriving its name from the acronym from the French "Conseil Européen pour la Recherche Nucléaire". Today CERN has 21 member states and over 600 institutes and universities around the world are involved contributing in different ways. Numerous experiments have been constructed at CERN following international collaborations such as the Large Hadron Collider (LHC).

In the LHC, the last and biggest element in the accelerator complex at CERN, particle beams are accelerated up to the record energy of 6.5 TeV per beam. The ring of the LHC stands astride the Franco-Swiss border, near Geneva. It has a circumference of ~ 27 km which is built at a mean depth of 100 m. Figure 2.1 shows an aerial view composition of the LHC tunnel and its perimeter on the surface.

The two beams are brought into collision in 4 points located in the center of the four main detectors situated in the LHC tunnel. The two biggest experiments at the LHC, ATLAS (A Toroidal LHC Apparatus) [27] and CMS (Compact Muon Solenoid) [28], use general-purpose detectors to investigate the largest

range of physics possible. Their main aim is to confirm or exclude the Higgs boson discussed in Chapter 1. There are two detectors with more specific roles, LHCb (LHC-beauty) [29] and ALICE (A Large Ion Collider Experiment) [30]. The latter focuses its research on quark-gluon plasma, simulating the conditions that existed shortly after the Big Bang. LHCb investigates the dominant amount of matter with respect to antimatter that is observed in the Universe nowadays. Figure 2.2 presents a diagram of the underground locations of the main four experiments at LHC: ATLAS, CMS, ALICE and LHCb.

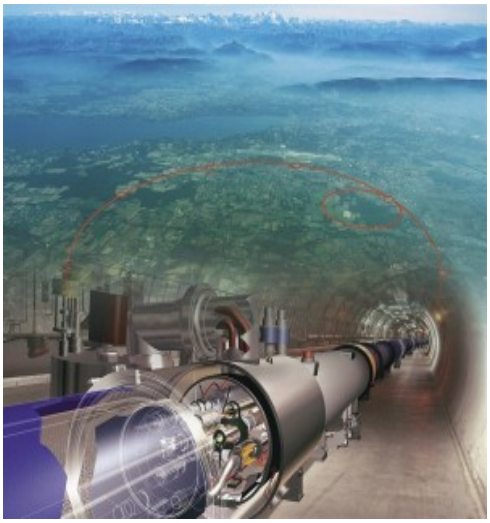

FIGURE 2.1: Aerial map of the LHC tunnel perimeter overlapped with the underground beamline view.

There are three smaller experiments at the LHC: TOTEM (Total Cross Section, Elastic Scattering and Diffraction Dissociation at the LHC), MoEDAL (Monopole and Exotics Detector at the LHC) and LHCf (LHC forward experiment). TOTEM uses detectors positioned on either side of the CMS interaction point, while LHCf is made up of two detectors which sit along the LHC beamline since it focus on particles brushing past each other rather than meeting head on when the beams collide. MoEDAL uses detectors deployed near LHCb to search for the theoretical magnetic monopole particles.

This chapter starts introducing the LHC complex in Section 2.2. The ATLAS detector, which delivered the data for this thesis, will be described in Section 2.3. There, an overview of the characteristics of the ATLAS components and their performance are described.

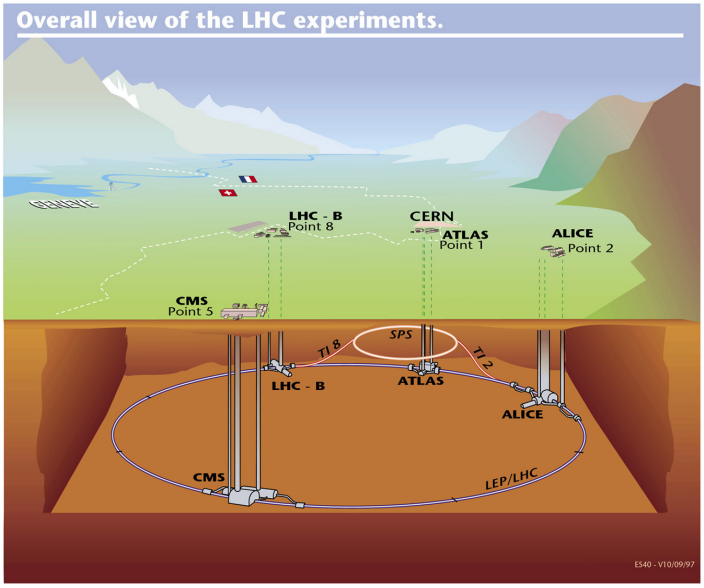

FIGURE 2.2: Diagram of the locations of the main experiments located at the LHC tunnel: ATLAS, CMS, ALICE and LHCb.

2.2 The Large Hadron Collider

The LHC physics goals require large collision energy since the cross-sections of processes of interest, so-called signal, raise faster with increasing collision energy with respect to the cross-sections of most background processes (see Fig. 1.7). In light of this, stable high intensity beams are needed in order to produce as many energetic collisions as possible.

The technological development to reach these goals have been achieved in the LHC. It has delivered both proton-proton and heavy ion (Pb^{82+}) collisions efficiently since 2008. This section provides the basis on the LHC accelerator complex and its design following the detailed description in Refs. [31–33].

2.2.1 LHC Accelerator Complex and Design

The accelerator complex at CERN is a succession of machines that accelerate particles to increasingly higher energies. Each machine boosts the energy of a beam of particles, before injecting the beam into the next machine in the sequence. Figure 2.3 illustrates the accelerators complex design.

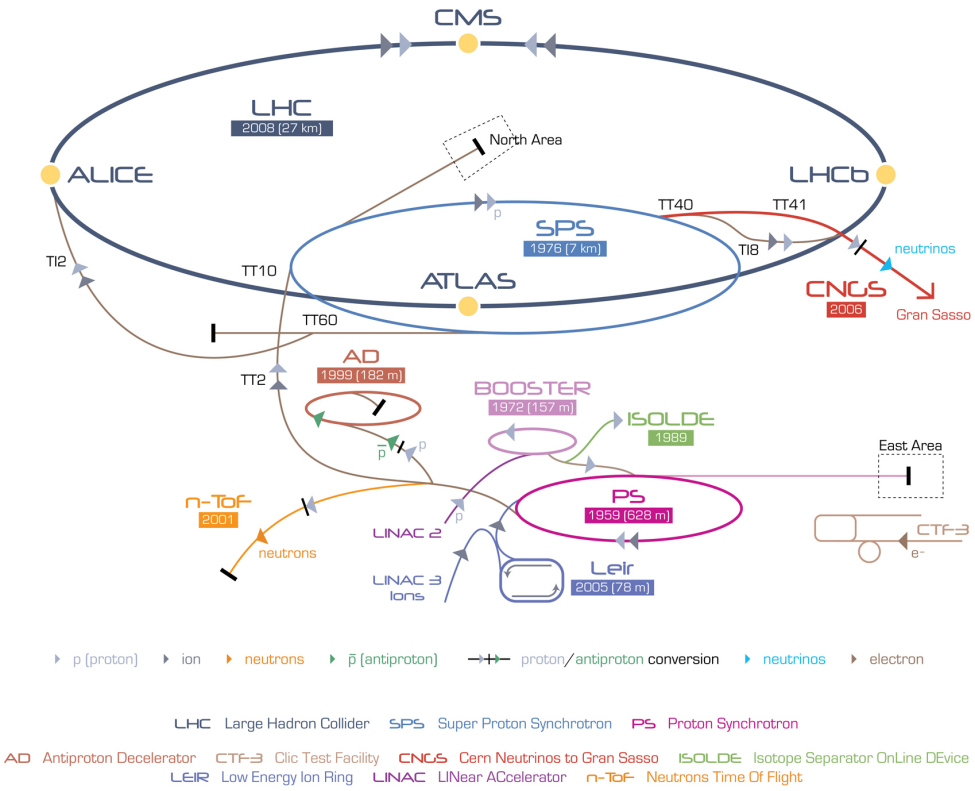

FIGURE 2.3: The CERN's accelerators complex from Ref. [31].

Protons are pre-accelerated in four increasingly large machines before being injected to the LHC ring. The beams injected in the LHC are actually a collection of proton bunches. Each bunch is about $20-30\,\mu$m in diameter, and a few centimetres long. The timing and control provided by the LHC is so precise that bunches only cross paths, producing collisions, within the four caverns of the LHC in which are located the detectors. The LHC challenge is to get as many bunches as possible circulating into the ring, each with maximum number of protons. To do that, the LHC complex starts with a simple bottle of hydrogen gas and an electric field to strip the electrons from the hydrogen atoms.

The first chain of the accelerator system is a linear accelerator, Linac 2. This delivers bunches of protons and gets them up to an energy of 50 MeV. From there, the protons are dumped into the first circular accelerator, called Proton Synchrotron Booster (PSB). This hardware dates from 1972, and it manages to get the protons up to 1.4 GeV in 1.2 seconds. It also starts squeezing the bunches down so that they have

a smaller cross-section. The beam is transferred to the Proton Synchrotron (PS), which was built in 1959. It has a 628 m circumference, and it takes 3.6 seconds to get two injections of bunches up to 25 GeV. In the last step of the LHC complex, the protons are sent to the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV. The protons are finally transferred to the two tubes of the LHC, the so-called beam pipes. As was introduced previously, protons are not the only particles accelerated in the LHC. Lead ions for the LHC are originate from a source of vaporised lead and enter Linac 3 before being collected and accelerated in the Low Energy Ion Ring (LEIR). They then follow the same route to maximum energy as the protons.

The LHC tunnel was originally constructed between 1984 and 1989 for the CERN Large Electron-Positron machine (LEP [15]). Inside the LHC, the two high-energy particle beams travel in separate beam pipes. The beam in one pipe circulates clockwise while the beam in the other pipe circulates anticlockwise. The beams circulate inside the LHC under a high vacuum condition. Moreover, electromagnetic devices are used to keep the particles in their orbits. The main components of the accelerator are superconducting dipole magnets, operating at a temperature of 1.9 K and designed for producing magnetic fields of 8.33 T. Superfluid helium is used to cool down the more than 1200 dipole magnets. Finally, quadrupole magnets focus the beam, and accelerating cavities, that are electromagnetic resonators, keep the bunches at a constant energy. It takes about 4 minutes to fill each beam pipe of the LHC, and 20 minutes for the protons to reach their maximum energy of 4 TeV. When that energy is reached, the proton beams collide in the center of the experiments. The details of the LHC beam parameters are quoted in Tab. 2.1.

Circumference	26.7 km
Radius	4.24 km
Number of magnets	9593
Number of dipole magnets	1232
Number of quadrupole magnets	392
Nominal magnetic field strength	8.33 T
Dipole operating temperature	1.9 K
Inelastic cross-section	60 mb
Total cross-section	100 mb
Revolution frequency	11.25 kHz
Bunch frequency	40.08 MHz

TABLE 2.1: Design parameters of the LHC operating at $\sqrt{s} = 14$ TeV.

2.2.2 LHC Operation 2010-2012

The first proton beams were successfully circulated in the main ring of the LHC on September 10^{th} 2008. Nine days later there was an unfortunate incident produced by a faulty electrical connection. This damaged over 50 superconducting magnets and led to a long technical intervention that delayed the research program by 14 months [34]. On November 20^{th} 2009, proton beams were successfully circulated in the LHC tunnel again, with the first recorded proton-proton collisions occurring three days later at a center-of-mass energy of 900 GeV and 2.36 TeV. On March 30^{th} 2010, the collisions took place between two 3.5 TeV beams, setting a world record for the highest-energy man-made particle collisions. During 2010 and 2011, the LHC produced $\sqrt{s} = 7\,\text{TeV}$ proton collisions increasing the number of bunches per beam from 200 to 1380. In 2012 the energy was increased to 4 TeV per beam, delivering $\sqrt{s} = 8\,\text{TeV}$ proton collisions.

The LHC program completed its first period of proton-proton collisions (Run I) in 2013. During the first long shutdown in 2013-2014 the experiments of the LHC have been updating its detectors in order to measure the future collisions at the center-of-mass energy of $13 - 14\,\text{TeV}$. The LHC re-started its activity by mid of 2015. Table 2.2 compares the parameters conditions of the LHC proton-proton collisions in each of the years of the Run I and the design quantities.

	2010	2011	2012	Design
Centre-of-mass energy	7 TeV	7 TeV	8 TeV	14 TeV
Peak luminosity [$\text{cm}^{-2}\text{s}^{-1}$]	2×10^{32}	3.65×10^{33}	7.73×10^{33}	10^{34}
Protons per bunch ($\times 10^{11}$)	$0.1 - 1.2$	$0.6 - 1.2$	1.48	1.15
Number of bunches	< 200	200-1380	1380	2808
Average collisions per bunch-crossing	≤ 3	9.1	20.7	22
Time between bunches [ns]	≥ 150	75/50	50	25
Delivered luminosity	$48.1\,\text{pb}^{-1}$	$5.46\,\text{fb}^{-1}$	$22.8\,\text{fb}^{-1}$	—

TABLE 2.2: Overview of machine parameters of the LHC operation during the Run I collision years compared to the design values.

One of the most important characteristics of the LHC data is the luminosity. Colliders' luminosity depends exclusively on beams parameters and can be calculated as follows [33],

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

where N_b is the number of particles per bunch, n_b is the number of bunches per beam, f_{rev} is the frequency of complete turns around the ring, γ_r is the relativistic

gamma factor for particles in the beam, ϵ_n is the beam emittance which is a measure of how the particles depart from the ideal trajectory, β^* is the beta function at the collision point giving the envelope for the particle motion, and F is the luminosity reduction factor due to the crossing angle at the interaction point (≤ 1). Figure 2.4 shows the peak instantaneous luminosity delivered to the ATLAS detector per day during the proton-proton collisions of 2010, 2011 and 2012.

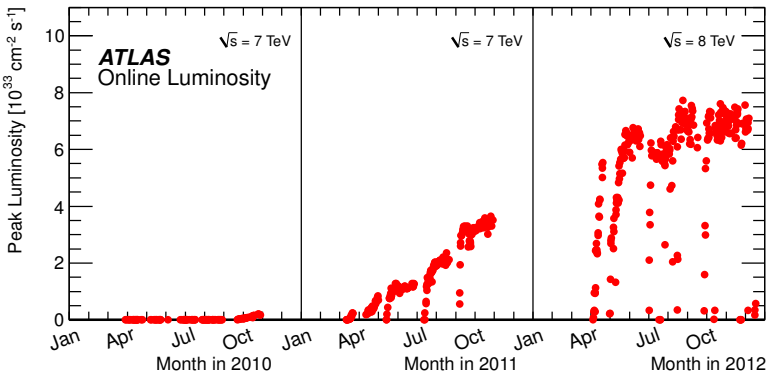

FIGURE 2.4: The peak instantaneous luminosity delivered to ATLAS per day versus time during the proton-proton collisions of 2010, 2011 and 2012 from Ref. [35].

The cumulative luminosities versus time delivered by the LHC, and recorded by the ATLAS detector are shown in Fig. 2.5. The cumulative luminosity is obtained integrating the instantaneous luminosity over time (see Eq. 1.63). A total integrated luminosity of $5.46\,\text{fb}^{-1}$ was delivered by the LHC at a collision energy of 7 TeV in the year 2011 of which $5.08\,\text{fb}^{-1}$ was recorded by ATLAS. In the year 2012 an integrated luminosity of $22.8\,\text{fb}^{-1}$ was delivered by the LHC at a collision energy of 8 TeV of which $21.3\,\text{fb}^{-1}$ was recorded by the ATLAS detector. These data samples are analysed in Chapter 5 for the search of the Higgs boson in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ final state. During 2010, there was recorded $48.1\,\text{pb}^{-1}$ which served for studies on the detector performance and efficiency.

From Eq. 2.1 is clear that the more protons per bunch, as well as the more bunches circulating at once into the LHC ring, the higher the LHC's instantaneous luminosity. Thus, having the maximum energy of collisions at the LHC gives the chance of exploring rare events. However, having high number of protons per bunch crossing and/or high number of bunches in the beams of the LHC enhance the probability of having multiple proton-proton interactions at the same event. This

effect is referred to as pile-up and it can be separated in the next two types depending on the origin of the additional proton-proton interactions,

- *in-time pile-up*: additional inelastic proton-proton interactions from the same bunch crossings. The higher number of protons per bunch the higher in-time pile-up effect.
- *out-of-time pile-up*: additional proton-proton interactions from nearby bunch crossings. The lower bunch spacing the higher out-of-time pile-up effect.

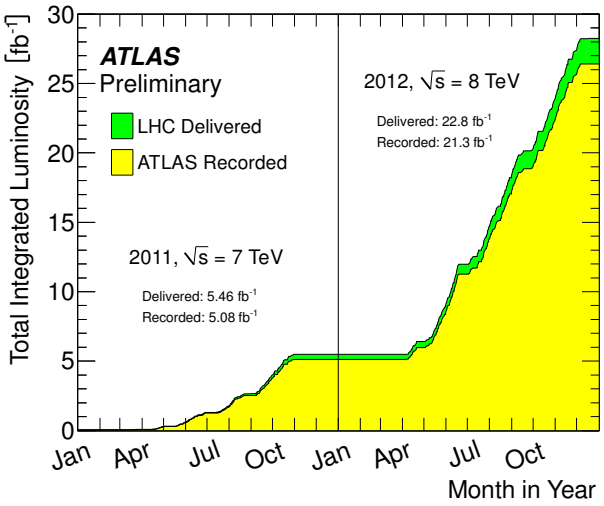

FIGURE 2.5: Total integrated luminosity delivered by the LHC (green), and recorded by ATLAS (yellow) at $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$ in 2011 and 2012, respectively [35].

The in-time pile-up has the largest impact on the physics analyses for the 2010-2012 running conditions. However, for the design 25 ns bunch spacing, it is expected that the out-of-time pile-up increases its contribution since the different bunches will be closer to their neighbours. In the following, the in-time pile-up will be referred simply as pile-up. The pile-up is directly related with the instantaneous luminosity and it is defined by [36],

$$\mu = \frac{L\,\sigma_{inel}}{n_b\,f_{rev}}\,, \tag{2.2}$$

where σ_{inel} is the inelastic cross-section which is taken to be 71.5 mb for 7 TeV collisions and 73.0 mb for 8 TeV collisions. The rest of variables are defined as in

Eq. 2.1. Figure 2.6 shows the luminosity-weighted distribution of the mean number of interactions per crossing for the 2011 and 2012 proton-proton collisions.

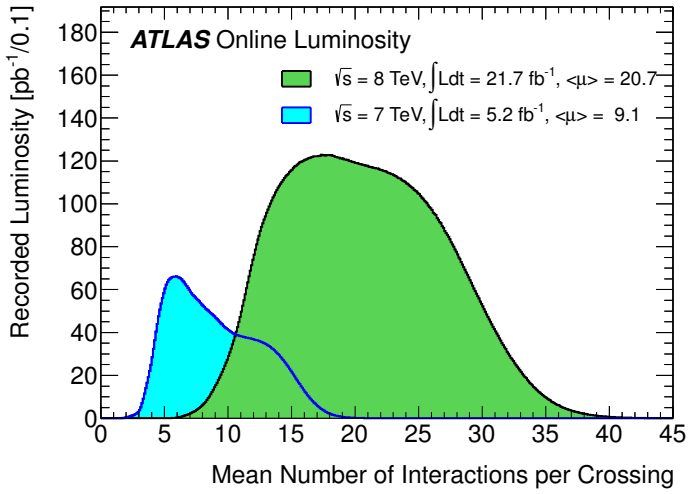

FIGURE 2.6: Luminosity weighted distributions of the mean number of interactions per bunch crossing for 2011 and 2012 from Ref. [35]. The integrated luminosities and the mean μ values are given for 2011 and 2012 running conditions.

The increase in pile-up through the Run I data taking has been treated properly by the physics analyses. The simulated processes are weighted by the μ distribution obtained from data in order to reproduce the same running conditions as the recorded collisions. However, the high pile-up environment suffered specially during 2012 has an important impact in the missing transverse momentum reconstruction. This quantity is typically represented by the symbol $\not\!\!E_T$ and this will be deeply analysed in Chapter 4. A big effort has been concentrated to recover the degraded performance of the $\not\!\!E_T$ with the presence of high pile-up as it is presented in Chapters 4 and 6.

2.3 The ATLAS Experiment

The general purpose of the ATLAS experiment is to investigate a large range of physics processes that might become detectable in the high energetic collisions of the LHC. The whole range of investigations using ATLAS include confirmations or improved measurements of the SM of particles, as well as studies of hypothetical

phenomena beyond the SM. One of the most important goals of ATLAS is to detect the Higgs boson particle. Since its discovery in July 2012, the efforts focus on the measurements of the Higgs boson properties. Up to now no deviations from the SM theory have been founded. This section summarises the design and characteristics of the ATLAS detector and its main components following the description in Refs. [37–40].

The ATLAS machine is the largest of the four detectors installed in the LHC tunnel. ATLAS is about 45 m long, more than 25 m high and has an overall weight of approximately 7000 tones. In the centre of the detector the two beams of protons or heavy ions circulating in the LHC collide at high energies. The particles produced in each collision emerge from the centre of the detector in all directions. The ATLAS detector has been designed to record the paths and energies of the particles produced from the LHC collisions. In light of this, different components are built to measure different types of particles. ATLAS is composed of the Inner Detector (ID), electromagnetic and hadronic calorimeters systems, and the Muon Spectrometer (MS). Figure 2.7 presents a view of the ATLAS detector and its main components.

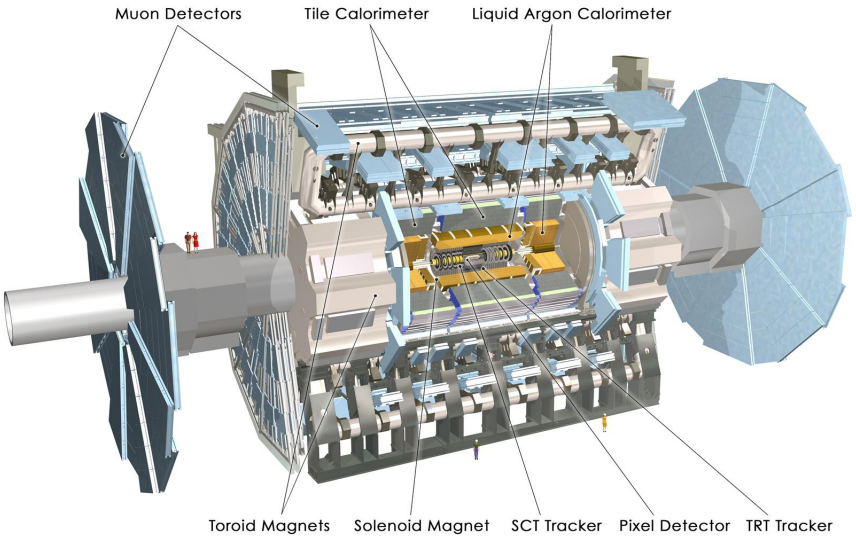

FIGURE 2.7: The ATLAS detector and its main components.

The ID is located in the innermost part of ATLAS. It is symmetrically built around the beam pipe of the LHC and it is designed to reconstruct tracks and decay vertices with high efficiency. It measures the trajectories of the charged particles

emerged from the collisions. The inner detector is embedded in a solenoidal magnet which generates a magnetic field of 2 T. The curvature of the trajectories which results from the the magnetic field bending power, is used to calculate the momentum of charged particles passing trough it. Electromagnetic and hadronic calorimeters surround the solenoid magnet. They are designed to measure the deposited energy and to reconstruct the direction of the different types of particles for which they are sensitive to. The last layer of ATLAS is formed by the muon spectrometer and a toroid magnet. The muon tracking system measures the paths of charged particles crossing the calorimeters. The trajectories are bent by the magnetic deflection provided by three superconducting air-core toroid magnets, which generate a field of 0.5 T.

The ATLAS detector is optimised to obtain high resolution measurements of different types of particles. Figure 2.8 presents an illustration of the signatures of different particles passing through the ATLAS detector from Ref. [41].

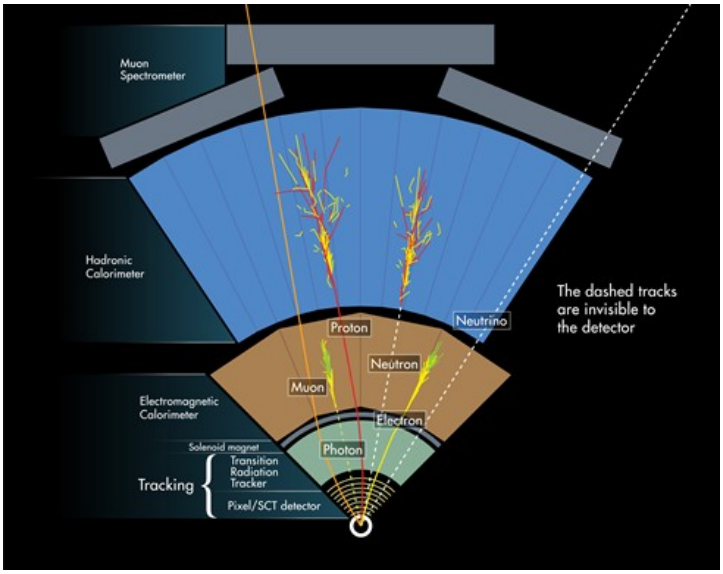

FIGURE 2.8: Computer generated image representing how the different components of the ATLAS detector measure different type of particles.

Basic requirements for the ATLAS design are the following,

- Efficient tracking at high luminosity for high-p_{T} lepton-momentum measurements, electron and photon identification, τ-lepton and heavy-flavour identification, and full event reconstruction capability at low luminosity.

- Very good electromagnetic calorimetry for electron and photon separation and measurement, complemented by a full-coverage hadronic calorimetry for accurate jet and $\not{E}_T$ measurements.
- High-precision muon measurements, with the capability of guaranteeing accurate measurements at the highest luminosity using the external muon spectrometer alone.
- Triggering and measurement of particles at low-p_{T}, providing high efficiencies for most physics processes of interest at LHC.
- Large acceptance in pseudo-rapidity (η) with almost full azimutal angle (ϕ) coverage. The ATLAS coordinate system is described below.

ATLAS uses a coordinate system with the origin at the point where the beams collide, the so-called interaction point (IP). The IP is located in the centre of the detector. The z-axis is situated along the beam line, with the side-A of the detector defined as that with the positive z. The x-y plane is perpendicular to the beam axis, with the positive x-axis pointing from the detector to the center of the LHC ring and the positive y-axis pointing upwards towards the surface. The azimuthal angle (ϕ) is measured around the beam axis, and the polar angle (θ) is the angle from the positive z-axis towards the y-axis. The rapidity is an important variable defined as,

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) , \tag{2.3}$$

where E is the energy of the particle and p_z is the particles momentum component in the z direction. More common usage has the pseudorapidity since it depends only on the polar angle of the particle's trajectory,

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

The pseudorapidity is actually the limit of the rapidity when setting the masses to zero. The value of $\eta = 0$ corresponds to $\theta = \frac{\pi}{2}$ and as the polar angle approaches zero, the pseudorapidity tends towards infinity. The advantage of this particular definition is that differences in pseudorapidity are invariant under boosts along the z axis. The distance ΔR between two objects in the pseudorapidity-azimuthal angle space is defined as,

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} , \tag{2.5}$$

where $\Delta\eta$ and $\Delta\phi$ are the differences between the object coordinates in η and ϕ, respectively.

2.3.1 The Inner Detector

The ID is the closest element of the ATLAS detector, being located directly around the beam pipe. It is built symmetrically with respect to the beams crossing point, covering a length of 7 m and a radius of 1.15 m. The ID is surrounded by the solenoid magnet and the LAr electromagnetic calorimeter. The magnetic field configuration of the ID is based on an inner thin superconducting solenoid surrounding the ID cylinder with a radius of 1.2 m and a length of 5.3 m. It provides an axial magnetic field of 2 T in the centre of the tracking volume.

The ID is responsible for tracking and vertex reconstruction. It provides excellent momentum and vertex measurements of charged particle tracks above $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ up to very high momentum. Using additional information from the calorimeter and muon systems, the ID also contributes to electron, photon, and muon identification, and supplies extra signatures for short-lived particle decay vertices. It is formed by three highly granular subsystems: the Pixel detector followed by the Semiconductor Tracker (SCT) layers, and the Transition Radiation Tracker (TRT) detector made of sensitive straws. The layout of the ID provides full tracking coverage over the range $|\eta| \leq 2.5$ and it is shown in Fig. 2.9. In the barrel region, the high-precision detector layers are distributed on concentric cylinders around the beam axis, while the end-cap detectors are mounted on disks perpendicular to the beam axis.

Each track of a charged particle can be fully identified by the combination of the parameters obtained from the three elements of the ID. These parameters are given at the point of closest approach to the nominal beam axis ($x = 0$ and $y = 0$) and they include: the impact parameters in $x - y$ and $r - z$ planes (d_0 and z_0); the azimutal and polar angles (ϕ and θ); and the charged curvature (q/p). The relative precision of the measurement is well complemented by the different components of the ID, so that no single measurement dominates the resolution and efficiency of the detector.

The Pixel detector is the innermost element of the ATLAS detector and it consists of three highly granulated cylindrical layers of pixel detectors. Each of the three pixel detector contains 1744 identical rectangular modules with a nominal size of $50 \times 400\,\mu\mathrm{m}^2$ and $250\,\mu$m thickness. The innermost layer of pixels is as close as 5 cm to the beam pipe and it is responsible for the determination of the interaction

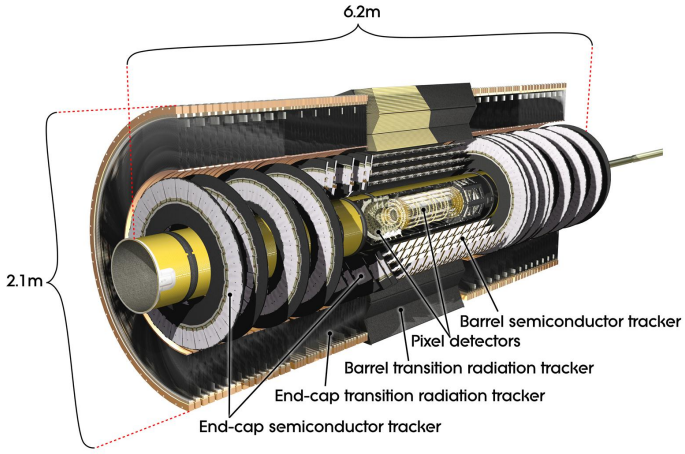

FIGURE 2.9: The layout of the ATLAS Inner Detector.

point and tagging of short-lived particles such as τ leptons and b-quarks. For this last reason, this layer is also known as B-layer and it suffers the highest radiation. Combining the measurements of the three layers of pixel detectors a precision of $10\,\mu$m in the transverse direction ($r-\phi$) and $115\,\mu$m in the longitudinal direction (z for the barrels and r for the end-caps) is achieved.

The SCT is the second innermost detector in ATLAS and it is located between the pixel and the TRT detectors. The basic principle of the semiconductor detectors is that the passage of ionizing radiation creates electron-hole pairs in the semiconductor which are collected by an electrical field. The barrel of the SCT contains four cylindrical layers of silicon micro-strip detectors while there are nine discs in the end-cap region. In the barrel region one set of strips in each layer is oriented parallel to the beam direction while another set of strips is running radially in the end-cap region. Each module of silicon detector is $6.36 \times 6.40\,\text{cm}^2$ and contains 780 readout trips. The SCT is designed to provide complementary measurements in the intermediate radial range, contributing to the measurement of momentum, impact parameter and vertex position. The achieved spatial resolution of the SCT is $16\,\mu$m in the $r-\phi$ plane and $580\,\mu$m in the longitudinal direction (z for the barrel and r for the end-caps). The SCT highly suffers from radiation damage so, it is necessary to operate the silicon sensors at low temperatures of approximately -5 to $-10\,^\circ$C to maintain adequate noise performance after radiation damage.

The outermost element of the ID is the TRT which is a gaseous straw detector composed of many layers of tubes of 4 mm diameter interleaved with transition

radiation material. The straws are filled with a mixture of 70% Xe, 27% CO_2, and 3% O_2. The main goals of the TRT are to enhance the tracking capability and to provide particle identification. In particular, the TRT distinguishes electrons from pions and other charged particles. Charged particles crossing a boundary between different dielectric media emit transition radiation with an intensity proportional to the Lorentz factor $\gamma = E/m$. These radiated photons have the energy of typically several keV, hence they are absorbed in the Xenon-based gas mixture of the straw tubes. Electrons have a lower mass compared to pions and thus emit a significant amount of transition radiation. This effect is used to achieved a better particle identification.

2.3.2 Calorimeters

After traversing the ID and the solenoid, particles produced at the LHC collisions enter in the ATLAS calorimetry system. The basis for the construction of the ATLAS calorimeters is to assemble absorber and detection mediums. When a particle interacts with the absorber material it produces a shower of secondary particles, of lower energies, which are detected in the active medium. The nature of the interaction is different for different types of particles: leptons and photons interact with matter via the electromagnetic interaction while hadrons may interact via the strong interaction. In light of this, two sets of calorimeters are used in ATLAS in order to provide good resolution of electromagnetic showers as well as a good containment of the wider hadronic showers for a large energy range. Thus, the ATLAS calorimetry system is composed of an electromagnetic calorimeter system (EM) which is based on lead and Liquid-Argon (LAr) with accordion geometry, and hadronic calorimeters based on a sampling technique with plastic scintillator embedded in an steel absorber. The central barrels are made out of plastic tiles, therefore the name of the hadronic central component: TileCal. The full calorimeter system covers the pseudorapidity range up to $|\eta| < 4.9$ and it contains the EM calorimeter covering the range $|\eta|<3.2$, a barrel hadronic calorimeter covering $|\eta|<1.7$, hadronic end-cap calorimeters covering $1.4<|\eta|<3.2$, and forward calorimeters covering $3.2<|\eta|<4.8$. At larger rapidities, where higher radiation resistance is needed, the radiation-hard technology is used for all the calorimeters: ElectroMagnetic End-Cap (EMEC), the Hadronic End-cap Calorimeter (HEC) and the Forward Calorimeter (FCal). A scheme with all the calorimeters for ATLAS can be seen in Fig. 2.10.

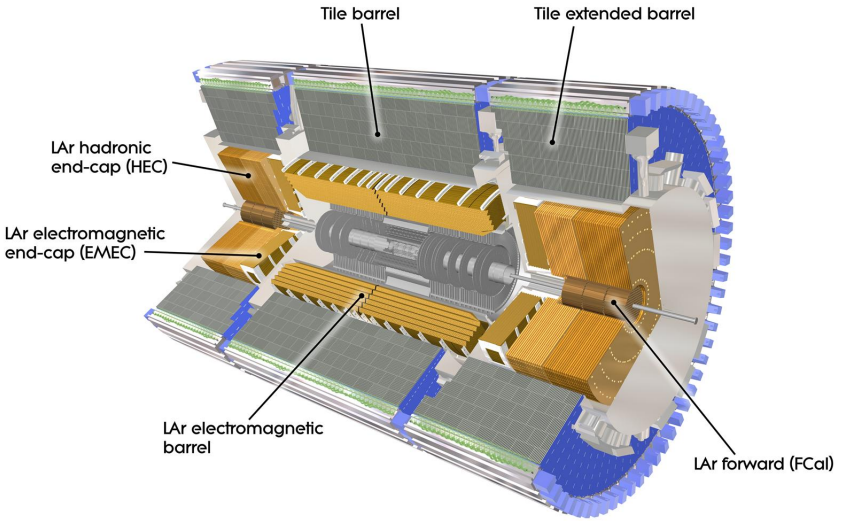

FIGURE 2.10: ATLAS calorimeters system.

2.3.2.1 Electromagnetic Calorimeter

The LAr EM calorimeter consists of two identical half-barrels, separated by a small gap of 4 mm at $z = 0$, covering the region $|\eta| < 1.47$ and two end-caps comprising two coaxial wheels covering the range $1.375 < |\eta| < 3.2$.

Each half-barrel is divided into 16 modules and it is made of 1024 accordion-shaped absorbers arranged with a complete ϕ symmetry around the beam axis. Between each pair of absorbers, there are two liquid argon gaps, separated by a readout electrode. In the region $|\eta| < 2.5$, each module is divided into three longitudinal layers with decreasing granularity, while in the $|\eta| > 2.5$ range the LAr EM calorimeter is segmented into two samplings as shown in Fig. 2.11. The first layer has the finest granularity and it allows for precise measurements of the electromagnetic shower shape. The cells in the second layer have granularities of 0.025×0.025 and 0.050×0.025 in $\Delta\eta \times \Delta\phi$ in the barrel and end-cap regions, respectively. Most of the energy of the electromagnetic showers originating from electrons and photons is deposited in the second layer which has a thickness of about 16 radiation lengths (X_0). The third layer owns a depth of about $2\,X_0$ and a granularity of 0.050×0.025 in $\Delta\eta \times \Delta\phi$.

The EMEC, the HEC and the FCal calorimeters are placed inside the end-cap cryostat. The EMEC uses the same technique as in the LAr EM barrel and it

covers the range 1.375<$|\eta|$<3.2. The HEC covers the region 1.5<$|\eta|$<3.2 and it uses copper plates as absorbers placed with parallel geometry in this case. The FCAL is extended in the 3.2<$|\eta|$<4.9 region and it provides larger electromagnetic coverage as well as hadronic shower measurements by using copper and tungsten as absorbers, respectively. The total thickness of the end-cap calorimeter system is above 26 X_0.

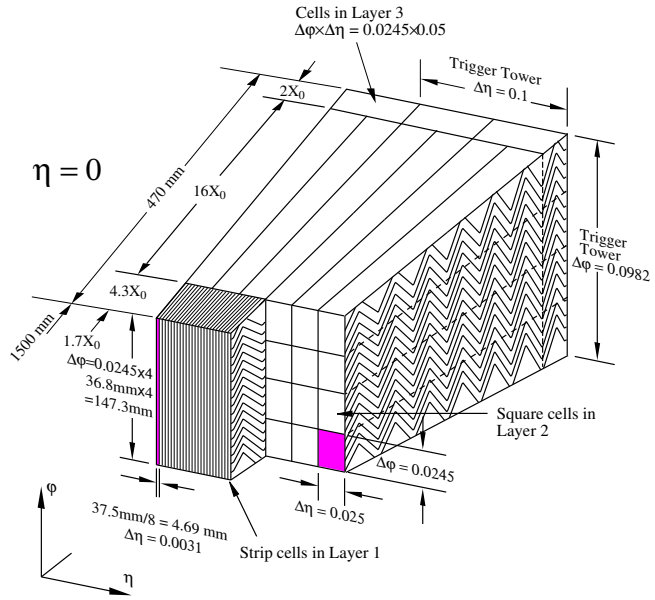

FIGURE 2.11: Diagram of a LAr EM calorimeter barrel module. It is shown the longitudinal segmentation, the cell size and the accordion structure.

2.3.2.2 Hadronic Calorimeter

The hadronic calorimetry system at the ATLAS detector is mainly designed to determine the energy and direction of the hadronic showers as well as to contribute to the measurement of the $\not\!\!E_T$ quantity [42]. Detailed studies for the different $\not\!\!E_T$ definitions used by ATLAS are described in Chapter 4. The ATLAS hadronic calorimetry surrounds the EM calorimeter and it has a high coverage extending its region up to $|\eta| < 5$. In order to maximise the efficiency of the detector different designs and techniques are used depending on different η ranges.

The TileCal calorimeter covers $|\eta|$<1.7 and it is divided into one central long barrel with a length of 5.6 m and two extended barrels with a length of 2.9 m each. There is a gap between the central and extended barrels of 0.6 m, which is needed for the ID detector and LAr calorimeter services. The inner radius of the TileCal

detector is approximately 2.2 m and the outer radius approximately 4.2 m. Each TileCal barrel contains 64 wedge-shaped modules, where the scintillator tiles are oriented radially and normal to the beam line. This achieves an almost full azimuthal angle coverage. The modules are made out of plastic scintillating tiles, which are embedded in a steel absorber structure as shown in Fig. 2.12.

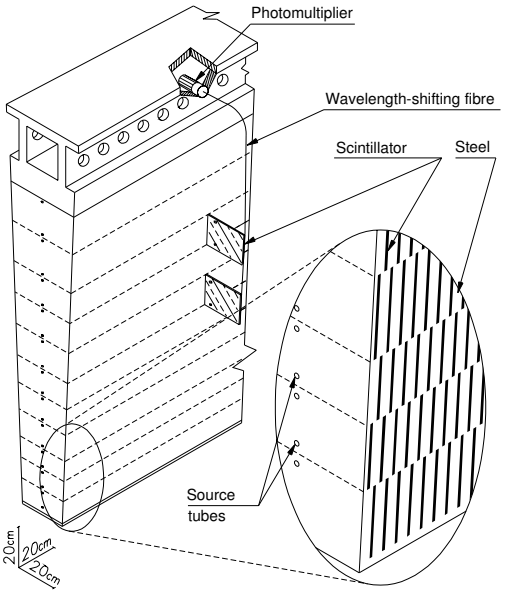

FIGURE 2.12: TileCal module components and structure.

When the particles cross the TileCal, light is produced in the scintillating material. The intensity of this light is proportional to the energy deposited by the particle in the calorimeter. The produced light is collected using wave-length shifting fibres and conducted to the PMTs that convert it to an electrical signal. The front-end and digitizing electronics are situated in the back-beam region of the modules on the so-called drawers. The motherboard is the basic element that holds together all the electronics in a drawer. The PMTs are read-out in groups of 12 by a motherboard, which sends out the digitalized data to the TileCal back-end electronics. The back-end electronics are installed in the counting rooms of the ATLAS cavern, in a low radiation environment. The Read-Out Driver (ROD) system is the central element of the back-end electronics. This uses the data from the front-end electronics as input to the online reconstruction algorithms in the first level of trigger as described in Section 2.3.5. High performance on the TileCal electronics and measurements are crucial for having a reliable reconstruction of the

hadronic showers. The TileCal energy and time offline reconstruction using the Optimal Filtering Algorithm are described in more detail in Chapter 3.

In the region $1.5 < |\eta| < 4.9$ the LAr techniques are used for the two end-caps ($1.5 < |\eta| < 3.2$) and the two high density forward calorimeters ($3.2 < |\eta| < 4.9$) as described in Section 2.3.2.1. The hadronic end-caps are made up of two equal diameter wheels. The first wheel is built out of 25 mm copper plates as absorber and the second wheel uses 50 mm copper plates. Compared to iron, copper has a shorter interaction length that allows to increase the size of the LAr gaps between plates, thereby reducing the electronic noise, the integration time and pile-up noise. In both wheels the absorber plates are separated by 8.5 mm gaps filled with liquid-argon and a structure of three electrodes that divide the gap into four drift spaces of ~ 1.8 mm.

The forward calorimetry should be efficient at forward jet tagging and $\not{E}_T$ reconstruction. The forward calorimeters are high density detectors in order to accommodate at least 9 interaction lengths of active material in rather short longitudinal space. Each forward calorimeter is divided into three longitudinal sections. In the first section the absorber is copper while in the second and third sections is tungsten. The calorimeter consists of a metal matrix (the absorber) filled with rods (electrodes). Liquid-argon is the active medium and fills the gaps between the matrix and the rods.

2.3.3 The Muon Spectrometer

The ATLAS muon spectrometer has been designed to make efficient use of the magnet bending power with a coverage of $|\eta|<3$. It provides projective towers in η and ϕ and is made out of practical chamber dimensions for production, transport and installation [43]. Figure 2.13 shows the position of the muon chambers.

The spectrometer is divided into three regions: barrel region ($|\eta| < 1.05$), transition region ($1.05 < |\eta| < 1.4$) and end-cap region ($|\eta| > 1.4$). Four different technologies have been used depending on spatial and timing resolution, resistance to radiation and engineering considerations: Monitored Drift Tube chambers (MDT), Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC).

The MDT chambers are composed of multilayers of high-pressure drift tubes. Each multilayer is mounted on each side of the support structure. The drift tubes are made of aluminium, 30 mm of diameter, with a central wire of W-Re. They work at 3 bar absolute pressure with a non-flammable mixture of $Ar - CO_2$.

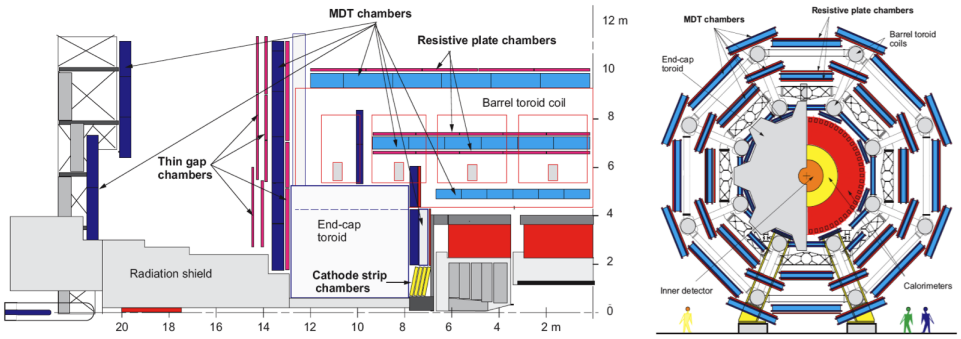

FIGURE 2.13: The ATLAS muon spectrometer in rz (left) and xy views (right).

The CSCs are multi-wire proportional chambers operated with a mixture of $Ar - CO_2 - CF_4$. The distance between anode wires (2.5 mm) equals the distance to the cathode. The cathode readout is segmented into strips (5.08 mm) orthogonal to the anode wires. The precision coordinate is obtained by measuring the induced avalanche in the segmented cathode, achieving space resolutions better than 60 μm.

The RPC is a gaseous parallel-plate detector with a typical space-time resolution of 1 cm $\times$ 1 ns with digital readout. It is composed by two parallel resistive plates made out of bakelite. The plates are separated by spacers that define the size of the gas gaps. The gas is a mixture of $C_2H_2F_4$. A uniform electric field of a few kV/mm produces the avalanche multiplication of ionization electrons. The signal is read out via capacitative coupling to metal strips placed at both sides of the detector and grounded.

The TGC is built with 50 μm wires separated by 2 mm. The wires are placed between two graphite cathodes at a distance of 1.6 mm. Behind the graphite cathodes, strips or pads are located to perform a capacitive readout in any desired geometry. Some advantages of these chambers are a fast signal, typical rise time 10 ns and low sensitivity to mechanical deformations.

In the barrel region the chambers are situated in three concentric cylinders (the so-called stations) around the beam axis at a radial distance of 4.5 m, 7 m and 10 m. MDT chambers are used for high precision measurements and RPC for triggering. The low-p_{T} muon trigger uses two double-layer RPCs located on each side of the middle station, while the high-p_{T} trigger uses one triple layer chamber located at the outer barrel muon station. In the transition and end-cap region most of the chambers are installed perpendicular to the beam axis as it is shown in Fig. 2.13. In the transition region ($1.05 < |\eta| < 1.4$) the muon track is measured

with three vertical stations, placed inside or near the barrel magnet. In the end-cap region ($|\eta| > 1.4$), the stations are located before and after the end-cap toroid magnets and a third one near the cavern wall. The trigger is provided by the TGC chambers while precision measurements are provided by the MDT chambers at small η and the CSC chambers at large rapidity.

2.3.4 Magnetic Field

The main purpose of the ATLAS magnetic field is to bend particles in order to perform momentum measurements. The ATLAS magnetic field is optimised to increase the identification power of the sub-detectors in a light and open structure which minimises scattering effects [44]. This consists of a central solenoid servicing the inner detector with an axial magnetic field of 2 T, surrounded by eight large scale air-core toroids generating a tangential magnetic field of approximately 0.5 T and 1 T for the muon spectrometer in the barrel and end-cap regions respectively (Fig. 2.14). The Nb-Ti superconductor in a copper matrix technology is used in this case. The magnet system weights 1300 t and is cooled by liquid He at 4.5 K.

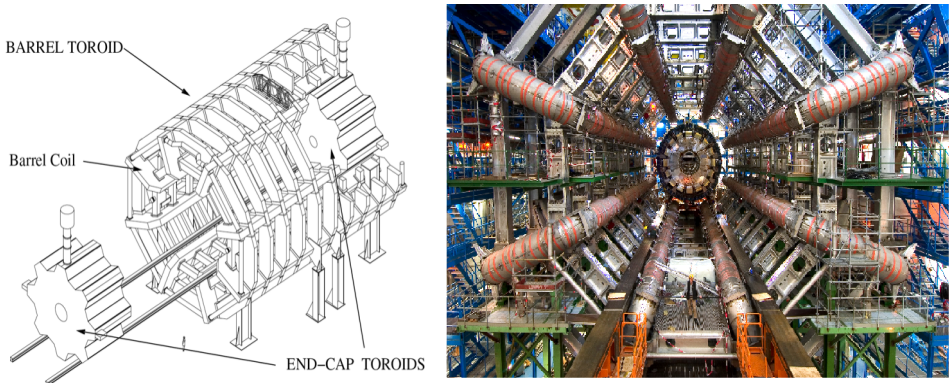

FIGURE 2.14: Scheme of the ATLAS superconducting air-core toroid magnet system (left) and picture of the central toroid (right).

2.3.5 Trigger System

The interactions in the ATLAS detectors create an enormous flow of data. To digest the data, ATLAS uses an advanced trigger system to tell the detector which events to record and which to ignore. Complex data-acquisition and computing systems are then used to analyse the collision events produced at the LHC [45]. The

major challenge for the online selection of interesting events is the high efficiency requirement to reduce the original event rates of 40 MHz down to 200 Hz. This selection has to be fast and efficient since the selected events are stored permanently and used by the physics analysis. The ATLAS trigger system is composed of three levels of event selection where each level refines the decision made at the previous level by applying additional selection criteria. The three distinct levels illustrated in Fig. 2.15 are briefly described in the following.

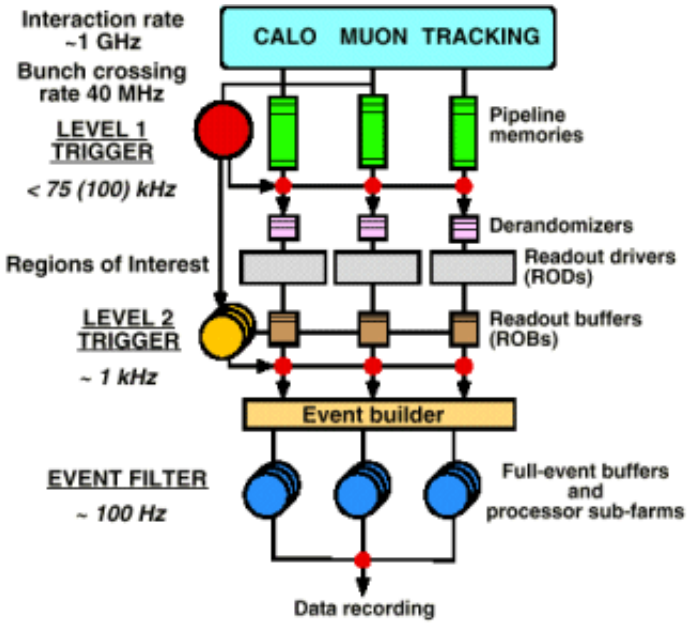

FIGURE 2.15: Schema of the ATLAS Trigger and Data Acquisition system from Ref. [45]. The Level 1 trigger system receives the data directly from the front-end electronics of the muon and calorimeter sub-systems. The data reconstructed by the ROD boards is transferred to the Level 2 for all the sub-systems. The Event Filter reduces by a factor 10 the data rate and its output is then recorded.

The Level 1 (L1) trigger stage is hardware-based and uses a limited amount of the total detector information to reach a decision whether to keep an event in less than 2.5 μs, reducing the event rate from up to 40 MHz to about 75−100 kHz. During this time the data from the sub-detectors are initially stored in pipeline memories. The L1 trigger uses reduced-granularity information from the calorimeter and muon systems and searches for high transverse momentum signatures originating from electrons, photons, jets and hadronically decaying τ leptons as well as large $\not{E}_T$ measurement. The possible trigger objects rely on the so-called Regions of Interest

(RoI's) which are defined by their coordinates in η and ϕ. The topoclustering algorithm is one of the algorithms used in ATLAS for merging cells with energies above a certain threshold to reconstruct an object. This clustering algorithm is described in detail in Chapter 3. This is evaluated through the performance of the energy reconstructed in TileCal.

Once the RoI's are defined, they are held in read-out buffers (ROBs) until they are processed by the Level 2 (L2) trigger. Then they can be either discarded or accepted, in which case they are transferred by the DAQ system to the storage system for the next level of triggering.The Level 2 (L2) trigger is software-based. The selection is largely based on the full- granularity information of all sub-detectors in the RoI's. A sequence of dedicated L2 algorithms is executed for each L1 RoI to compute event feature quantities to determine if the candidate object should be retained. The average processing time available for L2 algorithms is 40 ms and a reduced rate of approximately 3 kHz is achieved. The data accepted by the L2 trigger systems are further passed on to the Event Builder (EB), which performs a full reconstruction of the event.

The final online selection is performed by the Event Filter (EF) that typically uses the same algorithms as the offline reconstruction taking the full detector information into account. The event processing time is of about 4 s per event and it achieves the additional event rejection to reduce the output rate to about 200 Hz. The events selected by the EF are finally stored in the CERN computer centre for further offline processing and analysis.

Chapter 3

Description and Performance of the TileCal Noise

3.1 Introduction

The TileCal is the central component of the ATLAS hadronic calorimeter system, as described in Chapter 2. The features of the energy reconstruction in TileCal affect the performance of physics observables such as $\not{E}_T$ and jets. Specially, the noise treatment in TileCal has a direct impact on signal processing and thus, it causes effects on the performance of topological clusters. These clusters are the baseline for $\not{E}_T$ and jet algorithms, which rely on the energy deposits in the ATLAS calorimeter system. Hence, it is crucial to have a good response and reliable measurements in TileCal cells.

The aim of this chapter is to present the performance of the TileCal noise using randomly triggered events collected by the ATLAS detector during 2008 and 2009. Firstly, the energy reconstruction algorithm in TileCal cells is presented in Section 3.2. The Optimal Filtering algorithm is used for time and energy reconstruction in TileCal cells and described in detail. The reconstruction of the energy deposited in the TileCal cells is used as input for the ATLAS Topological Clustering algorithm. This algorithm merges together neighbouring cells as long as the signal in the cells is significant compared to noise as described in Section 3.3. The performance of different configurations is evaluated through shape-based topocluster quantities, so-called topocluster moments. The conclusions extracted from these results motivate the introduction of an improved noise description which accounts for

the non-Gaussian observed contribution. The two-Gaussian approach is presented in Section 3.4. Finally, Section 3.5 contains the results on topoclusters and $\not{E}_T$ performance comparing both, one- and two-Gaussian noise procedures.

3.1.1 Data and Monte Carlo Samples

Four data samples taken at the ATLAS detector during 2008 and 2009 are used for the noise studies presented in this chapter. Their characteristics are summarised in Tab. 3.1. The first three data samples are performed under comparable detector conditions. These are closed detector and both solenoid and toroid fields on. All of them are collected without collisions; instead a random trigger from cosmic interactions is used. The main differences with collision data are the timing, due to cosmic events being asynchronous with time in LHC machine, and the particle direction, since cosmic events are not coming from the interaction point. Therefore, randomly triggered events are ideal for energy reconstruction and noise studies since they allow to test detector performance without any extra contribution of energy deposits from particles generated at the collisions. Moreover, the events are recorded with absence of LHC beams except for the data sample listed at the bottom in Tab. 3.1. The TileCal cell energy in these randomly triggered events is reconstructed using the Optimal Filtering algorithm described in Section 3.2.1. The TileCal noise description methods are quoted in the sixth column and described in Section 3.4.

Run Number	Year	Month	Trigger	LHC conditions	Noise description	Number of events
91890	2008	October	Random cosmics	No beams	one-Gaussian	6000
121513	2009	July	Random cosmics	No beams	one-Gaussian	6060
137909	2009	November	Random cosmics	No beams	two-Gaussian	109927
140535	2009	November	Random cosmics	One beam	two-Gaussian	5540

TABLE 3.1: Data samples used in the TileCal noise performance studies presented in Chapter 3 and Appendix A. They were collected by the ATLAS detector with absence of collisions between October 2008 and November 2009. The number of events are quoted in the last column.

The results presented in this chapter are obtained with the Athena software releases from 14.5.0 to 15.3.0. Moreover, the official ATLAS package CaloRec-02-08-62 has been modified accordingly to obtain different configurations of topoclusters and to compare their performance. Finally, simulated samples are used to compare the

observed performance of the TileCal in real data in contrast with $W \to \ell\nu$ events generated with Pythia. This sample contains over 10^4 simulated events.

3.2 The Energy Reconstruction

The TileCal is divided into four partitions in η, two in the central long barrel (LBC and LBA) and two in the extended barrels (EBC and EBA). The nomenclature of the TileCal partitions refers to long (L) or extended (E) barrels (B) and it ends assessing the sign of the coordinate position along the beam axis (A and C for positive and negative η side respectively). The TileCal is also segmented in depth into three layers as shown in Fig. 3.1. The innermost layer contains type A cells. The layer in the middle contains BC cells in the long barrel and B cells in the extended barrels. The cells in the outermost TileCal layer are known as D cells. In addition to the standard cells, the Intermediate Tile Calorimeter (ITC) cells are located in the intersection between the crack region and the extended barrel. They cover the regions $0.8 < \eta < 1.0$ (labelled D4 and C10 in Fig. 3.1) and $1.0 < \eta < 1.6$ (E cells) [42, 46].

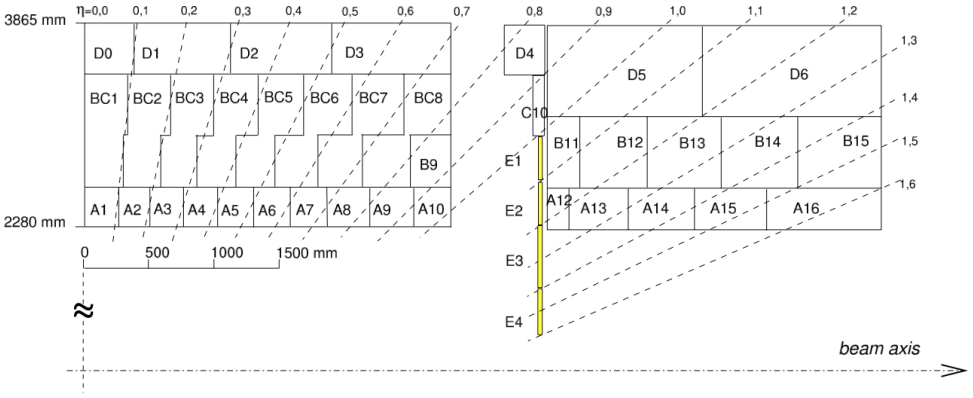

FIGURE 3.1: Segmentation in depth and η of the TileCal modules in LBA (left) and EBA (right) partitions.

In total, TileCal has 5182 cells, which corresponds to 9836 read-out channels. The light produced in the scintillating tiles is read out on two radial sides by wavelength shifting fibres which are bundled together in groups that form the TileCal cells. Since most of the cells are read out by two channels, the energy of the cells is defined as the sum of the energies obtained in each of the channels connected to the cell. The signal collected by the photomultipliers (PMTs) is digitised each 25 ns.

Seven samples are then used to reconstruct the amplitude, time and pedestal of the pulse using the Optimal Filtering (OF) algorithm.

3.2.1 The Optimal Filtering Algorithm

TileCal uses the OF algorithm to reconstruct the pedestal, amplitude and phase of the digitised signal. In this section a brief description of the method is given. Details of the OF algorithm can be found in Refs. [47–49].

The signal produced by the TileCal electronics can by expressed by the following equation,

$$S(t) = Ag(t - \tau) + p\,, \tag{3.1}$$

where $g(t)$ represents the pulse shape as a function of the time (t). A is the amplitude of the signal, τ refers to the relative phase and p the pedestal level. These magnitudes are illustrated in Fig. 3.2 and defined below.

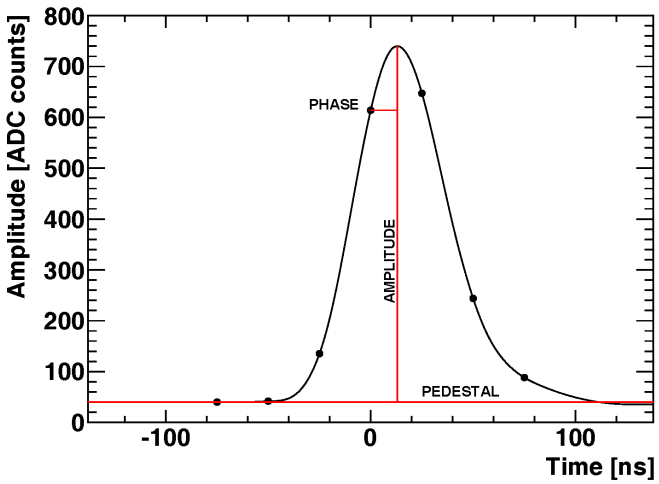

FIGURE 3.2: Pulse shape with the definition of amplitude, reconstructed phase and pedestal. The points represent the seven samples transmitted to the read-out detector electronics.

- The pedestal is the obtained measurement in absence of particles crossing the detector. This contains information related to the electronic noise contribution. In the presence of particles crossing the TileCal cells, this quantity represents the baseline of the expected signal pulse shape.

- The amplitude is the distance from the pedestal to the maximum of the reconstructed peak. The energy deposited by a particle passing through a cell is proportional to the signal amplitude. Several constants should be applied in order to obtain calibrated energy measurements.

- The phase is defined as the time between the peak of the pulse (τ) and the expected time of the pulse (τ_0), which is taken as the 4th sample by convention. This reference time is calculated with calibration systems for each channel, taking into account the time of flight of the particles from the interaction point and the length of the wavelength shifting optical fibres.

The seven samples that are transmitted to the back-end are the inputs to the OF method. The procedure to compute the amplitude, phase and pedestal magnitudes with the OF algorithm is a liner combination of the samples like the following,

$$A = \sum_{i=1}^{7} a_i \, S_i \, , \tag{3.2}$$

$$\tau = \frac{1}{A} \sum_{i=1}^{7} b_i \, S_i \, , \tag{3.3}$$

$$p = \frac{1}{A} \sum_{i=1}^{7} c_i \, S_i \, , \tag{3.4}$$

where S_i represents the i-th digital sample, A is the amplitude, τ is the phase of the pulse and, a_i, b_i and c_i are weights obtained from the signal pulse shape of the PMTs and the correlation of noise between digital samples.

The phase obtained by the OF is correlated with the reference phase (τ_0) used for the computation of the weights. If the weights are calculated for $\tau_0 = 0$, as in Fig. 3.2, then the phase provided by the OF corresponds to our definition of phase τ. However, if the weights have been obtained for any other phase, then the reconstructed phase by the OF is $\tau + \tau_0$. The phase reconstructed by the OF algorithm refers to the time between the expected phase of a pulse produced by a particle coming from the interaction point (which is the input for computing the weights) and the actual peak of the reconstructed pulse.

3.2.2 The Optimal Filtering with Iterations

The OF results rely on having a fixed and known offset between the signal peak and the collisions time for each TileCal read-out channel. However, signals caused

by cosmic rays are random and asynchronous with respect to the LHC collisions, hence the energy reconstruction will differ from the one caused by collision events. Note that this is the case for the non-collision data used in the studies presented here. Nevertheless, the OF algorithm can still be applied if it uses proper weights for each randomly triggered event according to the time position of the signal. With this purpose the OF algorithm explained above is extended to an iterative version, which is used when the expected time of the signal is not fixed.

The OF with iterations computes the phase in three iterations. The phase obtained in each iteration is used as the input reference time to select the proper weights for the next iteration. The first iteration starts from weights computed at time equal to zero. The following iterations select weights that are closer to the time of the pulse. The OF with iterations method is defined by the following equations,

$$A_k = \sum_{i=1}^{7} a_i\big|_{\tau_{k-1}} S_i \,, \tag{3.5}$$

$$\tau_k = \frac{1}{A_k} \sum_{i=1}^{7} b_i\big|_{\tau_{k-1}} S_i \,, \tag{3.6}$$

$$p_k = \sum_{i=1}^{7} c_i\big|_{\tau_{k-1}} S_i \,, \tag{3.7}$$

where S_i represents the i-th digital sample, k is the iteration index [0,2] with $k = 0$ corresponding to $\tau_0 = 0$. Finally, the set of weights (a_i, b_i, c_i) are different from each iteration k. They have been computed at each reference phase, τ_{k-1}. Those are stored in the ATLAS conditions database as described in Section 3.2.3.

The OF with iterations method is implemented in the ATLAS offline software and this is the default reconstruction algorithm of the signal measured by TileCal cells. The OF with iterations algorithm is used to obtain the results presented in this chapter. In the following and for sake of simplicity, the iterative method of the OF will be simply referred to as the OF algorithm.

3.2.3 Energy Calibration

The calibrated energy reconstructed in one channel (E_{ch}) is obtained by weighting the amplitude of the signal by several constants as follows,

$$E_{\text{ch}} = A \times C_{ADC \to pC} \times C_{pC \to GeV} \times C_{Cs} \times C_{\text{Laser}} \,, \tag{3.8}$$

where A is the signal amplitude in units of Analog to Digital Converter (ADC) counts, $C_{ADC \to pC}$ is the conversion factor of ADC counts to charge, $C_{pC \to GeV}$ is the conversion factor of charge to energy in GeV, C_{Cs} is the correction factor of non-uniformities after the gain performed by the Cesium–137 radioactive source calibration system [50, 51] and C_{Laser} refers to qualitative constants of the PMTs measured in between Cs scans [52]. The calibration constants shown in Eq. 3.8 are stored in the database for each TileCal channel, as well as the several sets of weights (a_i, b_i, c_i) required by the OF algorithm for different values of the expected time of the pulse. In this way, the OF algorithm can be evaluated offline using the last version of the calibration and weights parameters stored in this database.

3.3 The Topological Clustering Algorithm

Due to the nature of the strong interaction, the hadronic showers are formed by many calorimeter cells and are expanded in both lateral and longitudinal directions. Clustering algorithms are designed to group these cells and to sum up the total deposited energy within each cluster. The algorithm used in ATLAS with this purpose is the Topological Clustering algorithm. This algorithm starts with a seed cell and iteratively adds to it neighbouring cells. A requirement is applied to select cells with energy measurements incompatible with a noise fluctuation. With this requirement, a new cell-based structure known as a topocluster is built. Topoclusters are further used for object reconstruction algorithms in ATLAS together with other measurements provided by the rest of the detector systems. The evaluation of the topoclusters formed in TileCal is crucial from the point of view of jets and $\not{E}_T$ measurements. This section summarises the procedure of the Topological Clustering algorithm from Ref. [53].

The basic idea of the Topological Clustering algorithm is to group neighbouring cells that have enough signal compared with the expected noise. The energy significance threshold (s) is defined as the signal to noise ratio given by,

$$s = \frac{|E|}{\sigma}\,, \tag{3.9}$$

where $|E|$ is the absolute value of the cell's energy and σ is the expected noise value for such a cell. Note that using the absolute energy ensures symmetry in the noise spectrum. The Topological Clustering algorithm assumes that a normal distribution describes properly the noise amplitudes of all ATLAS cells. The energy significance threshold shown in Eq. 3.9 is measured in units of Gaussian sigmas. The

noise parameters σ are obtained from the root mean square (RMS) of the energy distribution for every cell in randomly triggered events. Those σ values are stored in the database in order to be used by the Topological Clustering algorithm.

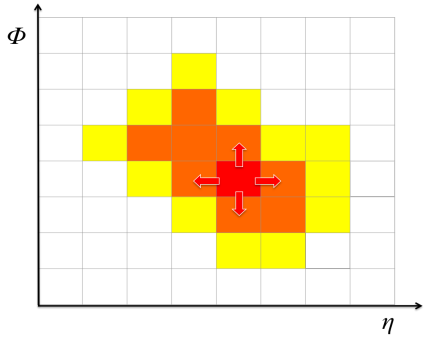

FIGURE 3.3: Representation of topocluster formation in the $\eta - \phi$ plane. First, the seed cell (red) is expanded in all directions. If the added cells satisfy the neighbouring threshold (orange), the topocluster includes them in a second iteration. The algorithm ends by adding cells with low significance (yellow).

A representation of the procedure to form topoclusters is presented in Fig. 3.3. The Topological Clustering algorithm starts by finding the seed cells. These should have an energy significance above a large threshold (s_s). Then, neighbouring cells are added to the cells tagged as seeds in the first step if the energy significance of the formers is above a low threshold (s_c). Moreover, cells around a seed cell can serve as additional seeds to further expand the cluster from them. It happens if they satisfy an energy significance above a medium threshold (s_n). Typically, the definition of neighbouring cells includes the surrounding cells within the same calorimeter layer. Optionally, the set of neighbouring cells can also include cells overlapping partially in η and ϕ in adjacent layers and/or adjacent calorimeter systems. Finally, the topoclusters may include bad cells if they satisfy any of the noise thresholds described above. However, the total energy of the topocluster does not account for the reconstructed energy in the cells labelled as bad cells. The energy of the topocluster is then calibrated and corrected for energy deposited outside the cluster, in dead material or bad cells as detailed in Refs. [53–55].

The default threshold values used by the Topological Clustering algorithm are $s_s = 4$, $s_n = 2$ and $s_c = 0$. This is also known as the (4,2,0) configuration. The selection of these values is optimal to find efficiently low energy clusters. The lowest threshold at the perimeter of the cluster ensures that the tails of the hadronic showers are not discarded. The large s_s and s_n values guarantee that the measured energy

is incompatible with a noise fluctuation. Assuming the electronic noise follows a normal distribution, the probability of tagging noise as a seed or neighbouring cell is below $6.3 \times 10^{-3}\%$ and 4.6% respectively [53].

3.3.1 Topocluster Moments

The total amount of energy contained in a topocluster object is obtained by summing up the energy of all cells contained in it. This is given by the following expression,

$$E_{\text{topo}} = \sum_{i}^{N} E_i \,, \tag{3.10}$$

where the index i runs over the N cells forming the topocluster. Note that negative contributions enter in the definition given in Eq. 3.10. Assuming a normal distribution for the electronic noise of the calorimeter, these noise contributions would cancel on average, hence $E_{\text{topo}} = 0$. Any deviation may indicate the presence of non-gaussian noise sources.

Apart from the total energy of a topocluster, it is also important to evaluate its shape. This section introduces two shape-related topocluster variables calculated from positive energy depositions. Typically, a cluster moment of a certain degree n in an observable m, defined for a cell constituent of the cluster, is given by the following expression,

$$< m^n > = \frac{1}{E_{\text{norm}}} \times \sum_{i|E_i>0} E_i \, m_i^n \,, \quad \text{where } E_{\text{norm}} = \sum_{i|E_i>0} E_i \,. \tag{3.11}$$

In Eq. 3.11 the index i, in both sums, runs over the cells with positive energy only, as mixing negative and positive weights could lead to unphysical behaviour. Typical observables for first and second moments are radial and longitudinal distances from the shower axis and the shower centre respectively. Once the shower axis $\vec{s}$ and the shower centre $\vec{c}$ are defined, two other quantities are calculated, the radial distance of the i-th cell from the shower axis,

$$r_i = |(\vec{x_i} - \vec{c}) \times \vec{s}| \,, \tag{3.12}$$

and the longitudinal distance of the i-th cell from the center along the shower axis,

$$\lambda_i = (\vec{x_i} - \vec{c}) \cdot \vec{s} \,. \tag{3.13}$$

The longitudinal and lateral extensions of a topocluster can be measured in terms of the second moments in λ ($<\lambda^2>$) and r ($<r^2>$) using Eq. 3.11 but with $n = 2$. Specifying topocluster dimensions in this way describes a spheroid with two axes of respective lengths $\sqrt{<\lambda^2>}$ and $\sqrt{<r^2>}$.

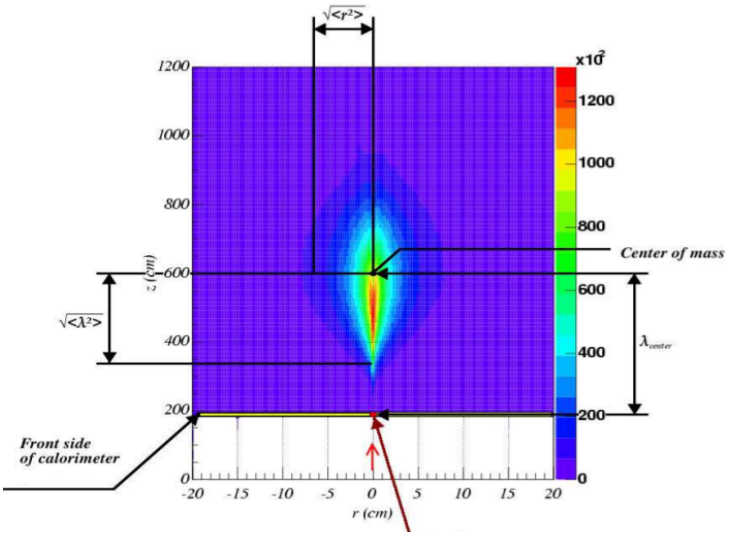

FIGURE 3.4: Schematic picture of a tau shower shape and its relevant variables, such as the RMS of the transverse extension in r ($\sqrt{<r^2>}$) and the RMS of the longitudinal extension in λ ($\sqrt{<\lambda^2>}$) from Ref. [56].

Figure 3.4 shows the schematic picture of a tau shower and its related longitudinal and transversal moments. In the following sections, the second moment in λ and the normalised second moment in r will be evaluated for different topocluster configurations. The normalised second moment in r is given by,

$$\frac{<r^2>}{<r^2_{\text{core}}>}, \tag{3.14}$$

where $<r^2_{\text{core}}>$ is computed by the two most energetic cells in the topocluster using a fixed value of $r = 40\,\text{mm}$ in Eq. 3.11, as described in Section A.2. For sake of simplicity, the normalised second moment in r will be referred to as $<r^2>$ in the rest of the chapter.

3.3.2 Performance of the Topocluster Moments

This section summarises a set of comparative performance studies on the Topological Clustering algorithm using different significance threshold values: s_s, s_n, and s_c, as

defined in Section 3.3. In total, there are 18 configurations of topoclusters analysed, which are classified into four groups. These groups are listed in Tab. 3.2. The complete set of distributions is presented in Appendix A.

$s_s = 4$	$s_s = 4.5$	$s_s = 5$	$s_n = 2$
(4, 1.5, 0)	(4.5, 1.5, 0)	(5, 1.5, 0)	(3, 2, 0)
(4, 2, 0)	(4.5, 2, 0)	(5, 2, 0)	(3.5, 2, 0)
(4, 2.5, 0)	(4.5, 2.5, 0)	(5, 2.5, 0)	(4, 2, 0)
(4, 3, 0)	(4.5, 3 ,0)	(5, 3, 0)	(4.5, 2, 0)
(4, 2, 0.5)			(5, 2, 0)
(4, 2, 1)			(5.5, 2, 0)
			(6, 2, 0)

TABLE 3.2: Topocluster configurations used for studies presented in Appendix A. The results in Chapter 3 are focussed on the performance of the first column.

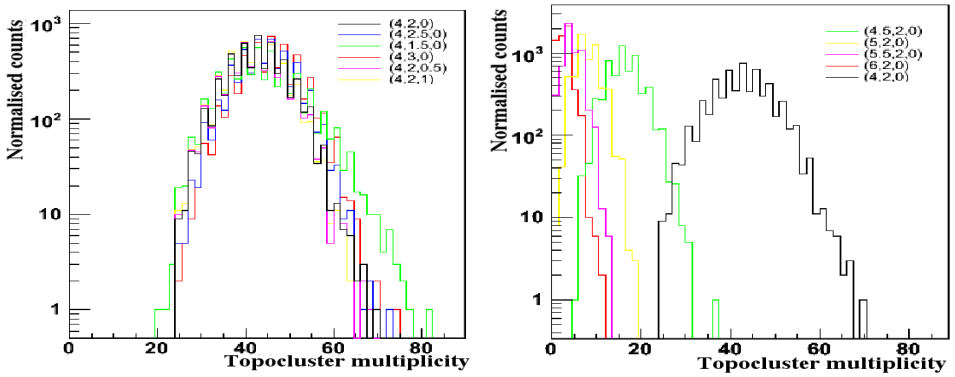

FIGURE 3.5: Topocluster multiplicity for different configurations (s_s,s_n,s_c) using randomly triggered events collected by the ATLAS detector during 2008. Distributions have been normalised to the number of topoclusters obtained with the nominal (4,2,0) Topological Clustering configuration.

The effect on the topocluster multiplicity due to selecting different threshold values as inputs for the Topological Clustering algorithm is shown in Fig. 3.5. The top plot maintains fixed significance for the seed cells $s_s = 4$ while the bottom distribution compares different configurations with $s_n = 2$ and $s_c = 0$. The s_s value has significantly higher impact on the topocluster multiplicity, as expected.

Moreover, it is observed that the currently used $s_n = 2$ and $s_c = 0$ thresholds provide the minimum number of formed topoclusters in each configuration group. In this light, the following studies evaluate several topocluster configurations fixing $s_n = 2$ and $s_c = 0$. The results will focus on the comparative performance using

different s_s values. Table 3.3 summarises the number of formed topoclusters from Fig. 3.5 and the mean energy per topocluster. The results show a tendency to higher positive mean energy values as s_s increases.

Figure 3.6 shows the second moment in r and λ spectrums obtained from randomly triggered events collected by ATLAS during 2008. These results point to the presence of two contributions of the noise which form two sets of topoclusters,

- small sized topoclusters: $\sqrt{< r^2 >} \in [0, 0.2]$ or $\sqrt{< \lambda^2 >} \in [0, 400]$ mm,
- large sized topoclusters: $\sqrt{< r^2 >} \in [0.2, 1.4]$ or $\sqrt{< \lambda^2 >} \in [400, 1200]$ mm.

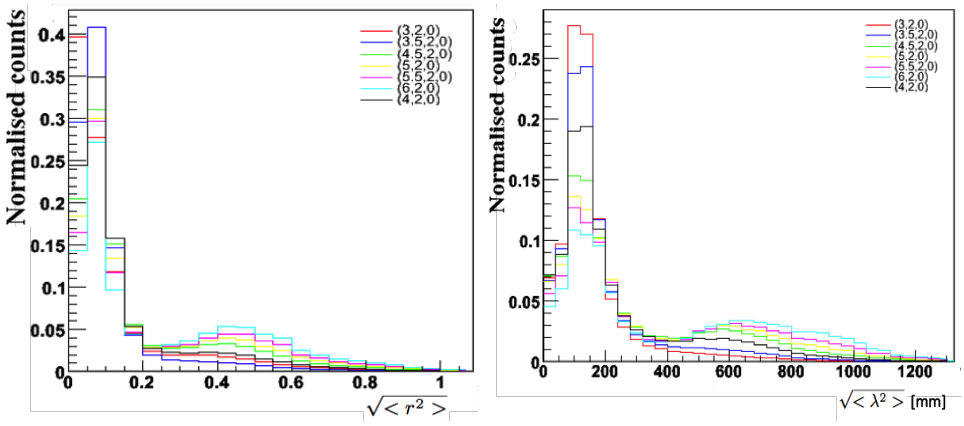

FIGURE 3.6: Squared root of the normalised second moment in r (left) and second moment in λ (right) for different values of the s_s significance threshold. The data correspond to randomly triggered events collected by the ATLAS detector during 2008. Distributions have been normalised to unity for comparison.

The same trends are obtained in the 2008 data as well as in the initial 2009 data collected by the ATLAS detector under the same conditions. The Topological Clustering configuration (4,2,0) is applied for both data years in the topoclusters moments shown in Fig. 3.7. The contribution from large topoclusters in 2009 data sample is $\sim$ 10% higher than in 2008 data (see Tab. 3.4). Hence, we refute the hypothesis that this is a spurious effect in the TileCal during 2008 since this behaviour is observed for both data periods. Additional investigations on this effect are collected in Appendix A. These results clearly show that the energy contribution from TileCal is dominant in large topoclusters.

Configuration	N_{topo}	E (MeV)	E_{T} (MeV)
(3, 2, 0)	>150	15.8	8.03
(3.5, 2, 0)	140.28	27.9	14.7
(4, 2, 0)	42.57	51.7	27.8
(4.5, 2, 0)	16.12	75.3	42.4
(5, 2, 0)	7.77	92.8	52.3
(5.5, 2, 0)	4.30	102	59.0
(6, 2, 0)	2.54	112	65.0

TABLE 3.3: Multiplicity, mean of the energy and transverse energy of topoclusters formed with different configurations: (s_s, 2, 0), where s_s ranges from 3 to 6 in steps of 0.5. The data correspond to randomly triggered events collected by the ATLAS detector during 2008.

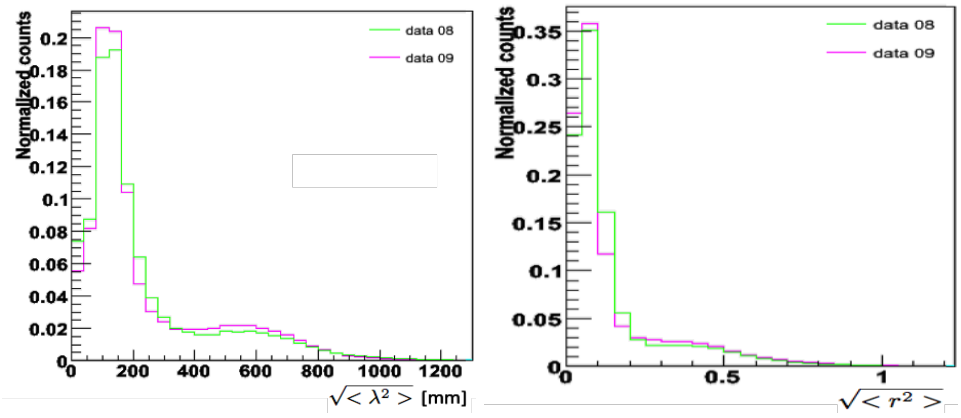

FIGURE 3.7: Second moment in λ (left) and normalised second moment in r (right) for reconstructed topoclusters with the Topological Clustering algorithm using the (4,2,0) configuration. Data collected by ATLAS in 2008 (green) and 2009 (pink) periods normalised to unity are compared.

Due to the stochastic nature of hadronic showers in randomly triggered events we mainly expect small topoclusters. However, Figs. 3.6 and Fig. 3.7 show a noticeable contribution of large topoclusters in all studied configurations and for several data periods. Large topoclusters can be affected by,

- cells which concentrate a large fraction of the total energy (so-called hot spots),
- source of coherent noise affecting an extended area in the detector (e.g. electronic cross-talk effects).

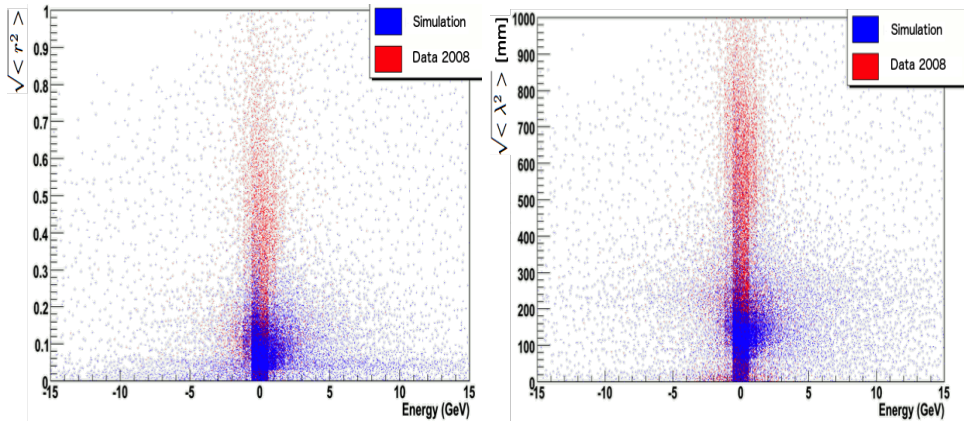

FIGURE 3.8: Normalised second moment in r (left) and second moment in λ (right) as a function of energy from reconstructed topoclusters in randomly triggered events in data (red) and simulation (blue).

The two dimensional correlation of the topocluster shape-related parameters as a function of the topocluster energy is shown in Fig. 3.8. These distributions compare randomly triggered data events collected by ATLAS during 2008 with $W \rightarrow \ell\nu$ simulated process.

The large topoclusters obtained using collected data are not observed in the MC sample. Hence, there is a noticeable effect in the Topological Clustering algorithm which is not taken into account in the simulation. The presence of large topoclusters observed in data for different s_s threshold values motivate to evaluate the modelling of the cell noise in TileCal. A more reliable description on the noise using a two-Gaussian model is described in Section 3.4 following Ref. [57]. The benefits on the topoclusters performance and $\not\!\!E_T$ measurement from the introduction of the two-Gaussian model are presented in Section 3.5.

3.4 Description of the TileCal Noise

TileCal noise constants are measured and kept up to date because the noise is the input to the algorithm reconstructing topoclusters, as described in Section 3.3. The Topological Clustering algorithm assumes that a normal distribution describes properly the noise contribution of the cells. However, the results on the performance of the Topological Clustering algorithm using topocluster moments shown in Section 3.3.2 clearly illustrate that TileCal noise is not properly described with the single Gaussian approach.

The measured noise contribution in a channel extrapolated from topocluster results can be considered as a linear combination of an intrinsic term (electronic noise) and a correlation term (coherent noise). The latter may be caused by cross-talk effects between channels situated in the same motherboard as described in Section 2.3.2.2. The relation of the topocluster features with the energy treatment of the TileCal noise motivates the evaluation of the noise description at the cell level. These investigations are based on the development of a new model to improve the noise description in the detector. This section describes the treatment of the dominant and intrinsic electronic noise in the TileCal, as well as the introduction of the two-Gaussian method [57].

3.4.1 The TileCal Electronic Noise

Electronic noise in TileCal is derived from standalone runs with absence of signal from the PMTs or injected calibration charge. These are called pedestal runs and are used to compute two sets of noise constants: Digital Noise computed for each channel and measured in ADC counts, and Cell Noise constants corresponding to the noise of each calorimeter cell and gain combination, measured in MeV. All these constants are stored in the database. Digital Noise constants are calculated before energy is reconstructed by the OF algorithm. Cell Noise constants are calculated after reconstruction. Problematic channels are masked in this process; for these the noise is read out but never used. These noise constants have a direct impact on the energy reconstruction in each channel and on a number of physics observables.

The Cell Noise is used as an input to the Topological Clustering algorithm. The σ values in Eq. 3.9 were obtained by fitting a Gaussian distribution function to the energy distribution of the events in several pedestal runs. If the energy distribution were Gaussian RMS/σ = 1, however, the results obtained show that this ratio is larger. Due to this behaviour, using the width of a normal distribution to define seed cells degrades the performance of the topoclustering algorithm in the low η region. Figure 3.9 illustrates the non-Gaussian nature of the TileCal cell noise. The plot shows the energy distribution of a typical TileCal cell for randomly triggered events collected in 2008. The OF algorithm is used to reconstruct the energy of the two channels forming the cell, being the energy of the cell the average. A free parameter fit to a Gaussian distribution is overlaid. Strong deviations from the Gaussian assumption are visible in the tails of the distribution, the relevant region for the Topological Clustering algorithm.

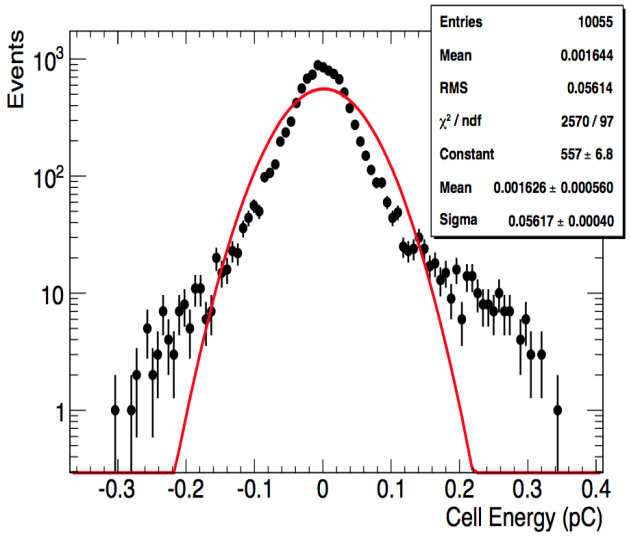

FIGURE 3.9: Reconstructed energy (in pC) of a typical cell for randomly triggered events collected by the ATLAS detector during 2008 from Ref. [57]. The distribution is fitted with a Gaussian function represented by the red line.

The shape of the reconstructed amplitude of a cell in randomly triggered events provides tails in the distribution as seen in Fig. 3.9. To improve the TileCal performance, a new approach for the noise description was developed.

3.4.2 Two-Gaussian Description of the TileCal Noise

In early tests during the ATLAS test beam the electronic noise was described by a single Gaussian. When the TileCal was installed in the ATLAS cavern and connected to the Low Voltage Power Supplies (LVPS) the noise increased significantly. In light of these features, a new noise modelling was needed to provide an accurate energy. A two-Gaussian function with three independent parameters was adopted in the spring of 2009. This will be referred to as two-Gaussian method in the following.

The general two-Gaussian probability density function (pdf) is given by the following equation,

$$f_{2g\,pdf} = \frac{1}{1+R}\left(\frac{1}{\sqrt{2\pi}\,\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} + \frac{R}{\sqrt{2\pi}\,\sigma_2}e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}\right). \tag{3.15}$$

Here, σ_1 and σ_2 are the sigmas of the two Gaussians and R is their relative normalisation. The mean values of the distributions μ_1 and μ_2 are constrained to

$\mu_1 = \mu_2 = 0$. This is a good approximation for TileCal cells and allows efficient storage in the database.

The result of a fit to the energy distribution of a typical TileCal cell with a two-Gaussian function is illustrated in Fig. 3.10. The parameters μ_1 and μ_2 are left as free parameters to better illustrate their typical values. They are represented by $G1\mu$ and $G2\mu$ respectively, which are small and can be approximated by 0. The parameters $G1\sigma$ and $G2\sigma$ corresponds to σ_1 and σ_2 in Eq. 3.15.

The comparison in Fig. 3.10 clearly shows that the two-Gaussian approach is more accurate in modelling the noise shape of the TileCal cells. As input to the Topological Clustering algorithm an equivalent sigma $\sigma_{eq}(E)$ is introduced. This is defined to give the same significance as the one σ region for a Gaussian pdf (e.g. $\int_{-\sigma_{eq}}^{\sigma_{eq}} f_{1g\,pdf} = 0.68$). The $\sigma_{eq}(E)$ is introduced to measure the E/σ_{eq} of the two-Gaussian pdf in units of σ of a normal distribution, with the purpose of comparing the performance of both pdfs' descriptions.

The equivalent significance for an energy deposition (E) and a two-Gaussian pdf can be expressed as,

$$\frac{E}{\sigma_{eq}(E)} = \sqrt{2}\,\mathrm{erf}^{-1}\left(\frac{\sigma_1 \mathrm{erf}\left(\frac{E}{\sqrt{2}\,\sigma_1}\right) + R\,\sigma_2\,\mathrm{erf}\left(\frac{E}{\sqrt{2}\,\sigma_2}\right)}{\sigma_1 + R\,\sigma_2}\right), \qquad (3.16)$$

where erf is the error function. [1]

One advantage of this definition of σ_{eq} is that there is a common unit for noise description for TileCal and LAr cells, so that the topological clustering algorithm is able to cluster cells in both calorimeters. These advantages are detailed in Ref. [57]. The input to the Topological Clustering algorithm is the σ_{eq} parameter in Eq. 3.16. Figure 3.11 illustrates the improvement in the estimation of the E/σ using randomly triggered events when the two-Gaussian template is used and the noise is expressed in terms of the σ_{eq} width.

3.5 Performance of the two-Gaussian Description

The benefit of using the two-Gaussian description in estimating the noise compatibility of energy deposits can be illustrated through the improvement of the performance of the topoclusters created in TileCal. The more accurate description of the TileCal noise using the two-Gaussian description has to be reflected in the reduction

[1] $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$

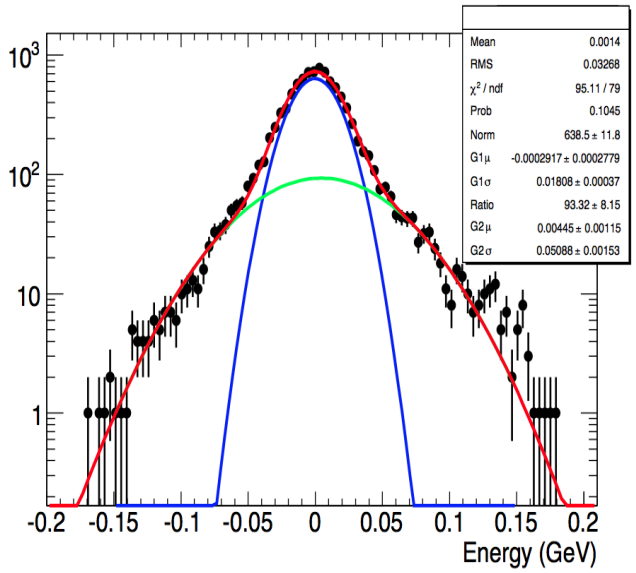

FIGURE 3.10: Reconstructed amplitude distribution of a typical cell for randomly triggered events collected by the ATLAS detector during 2008, from Ref. [57]. The two-Gaussian fit is shown in red. The blue and the green functions are its first and second Gaussian components respectively.

of the number of topoclusters and their size compared to the previous noise treatment using the single Gaussian approach. Furthermore, this improvement should be also observed by studying the TileCal contribution to the $\not{E}_T$ measurement. The aim of this section is to evaluate effects on the topoclusters formation when the two-Gaussian approach is applied instead of the one-Gaussian parametrisation.

Figure 3.12 shows the number of topoclusters formed using the nominal (4,2,0) Topological Clustering algorithm configuration for both one-Gaussian and two-Gaussian noise descriptions. These distributions are obtained for different runs collected by the ATLAS detector during 2008 and 2009 with randomly triggered events. The mean number of topoclusters formed with the two-Gaussian description for the TileCal noise is reduced by over a factor of two compared with the one-Gaussian model. Moreover, a reduction in the width of the topocluster multiplicity is observed. The two-Gaussian description of the TileCal noise limits the formation of topoclusters by using a larger noise constant σ_{eq}, as described in Section 3.4.1.

Figure 3.13 compares the second moment in r and λ of the topoclusters obtained for both TileCal noise descriptions on two runs used previously and a new

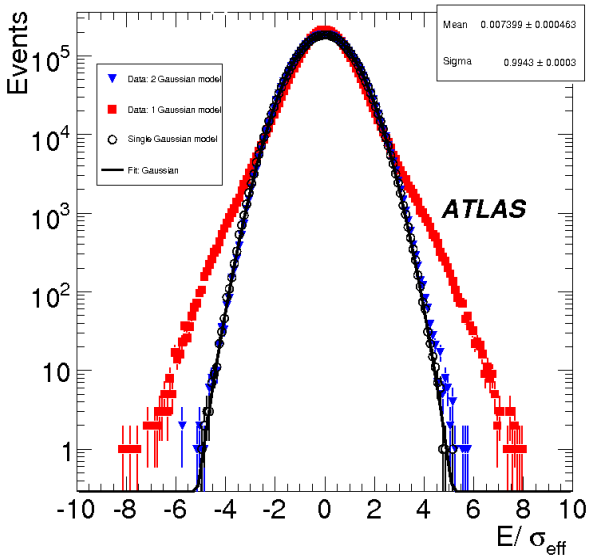

FIGURE 3.11: The energy deposited in TileCal cells divided by the noise constant stored in the database for different noise description models using randomly triggered events collected by the ATLAS detector during September 2008, from Ref. [57]. The black points correspond to the expected distribution for ideal one-Gaussian TileCal noise. The red and blue triangles are the measured E/σ_{eq} values from data using one- and two-Gaussian descriptions respectively. The distribution obtained with two-Gaussian approach is fitted with a Gaussian function represented by the black line.

sample with the two-Gaussian method applied. The existence of large topoclusters is highly reduced when the two-Gaussian approach is required in the computation of the TileCal noise constants. In order to evaluate this improvement, the ratio of the number of large over small topoclusters ($N_{\text{Large}}/N_{\text{Small}}$) comparing different TileCal noise descriptions is studied. The values are quoted in Table 3.4. The presence of large topoclusters using the two-Gaussian approach for the TileCal noise is reduced by almost a factor 10 with respect to the one-Gaussian treatment.

3.5.1 Noise Effects on Missing Transverse Momentum

The missing transverse momentum is a fundamental quantity which relies on the energy measured in the topoclusters (see Chapter 4). The E/σ thresholds used to construct topoclusters, as wells as the approach for describing the noise in TileCal,

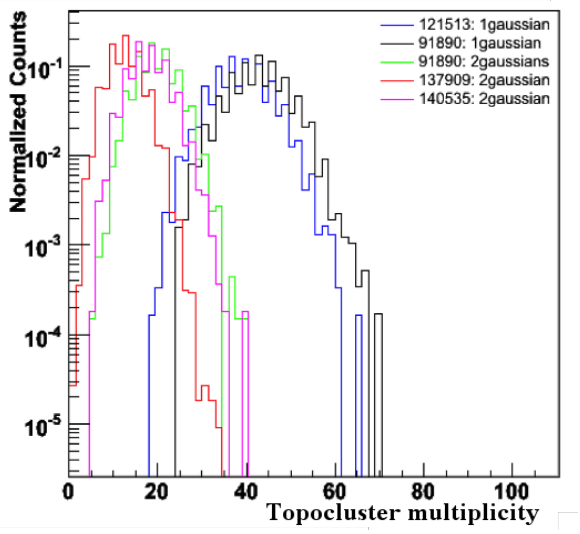

FIGURE 3.12: The topocluster multiplicity using the one-Gaussian (black and blue lines) and two-Gaussian (red, pink and green lines) descriptions for the TileCal noise using randomly triggered events collected by the ATLAS detector during 2008 and 2009. The legend shows the run number of each data sample. The distributions are normalised to unity.

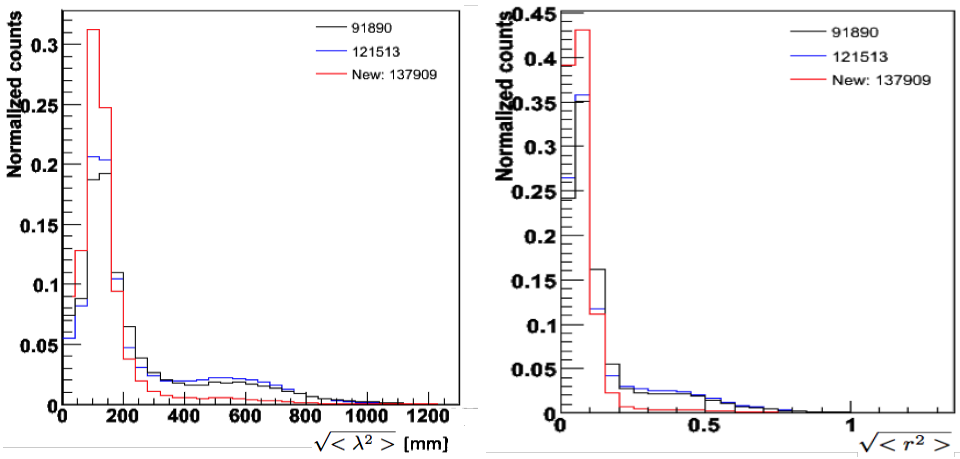

FIGURE 3.13: Second moment in λ (left) and normalised second moment in r (right) for different noise reconstructions using data recorded by the ATLAS detector during 2008 and 2009 years with a randomly trigger selection. The two-Gaussian noise description (red line) reduces the contribution of large topoclusters with respect to the one-Gaussian approach (blue and black lines). The number of counts are normalised to unity.

Run Number	Year	Data Type	Noise Treatment	$N_{\text{Large}}/N_{\text{Small}}$
91890	2008	random trigger	one-Gaussian	0.26
121513	2009	random trigger	one-Gaussian	0.34
137909	2009	random trigger	two-Gaussian	0.046

TABLE 3.4: Several randomly triggered ATLAS samples using one-Gaussian or two-Gaussian descriptions are listed. The contribution of large topoclusters is evaluated through the ratio values shown in the last column ($N_{\text{Large}}/N_{\text{Small}}$). The reduction in the formation of large topoclusters is almost an order of magnitude when applying the two-Gaussian method.

have a direct impact on the performance of this measurement. In order to investigate the effects of the TileCal noise, the negative vectorial sum of the transverse energy from all topoclusters in an event is evaluated under different scenarios. This magnitude will be referred to as MET in the following and is computed as,

$$\text{MET} = \sqrt{\left(-\sum_{i}^{N} E_{\text{T},x}(i)\right)^2 + \left(-\sum_{i}^{N} E_{\text{T},y}(i)\right)^2}, \tag{3.17}$$

where the index i runs over the total number of topoclusters (N) per event. $E_{\text{T},x}$ and $E_{\text{T},y}$ represent the longitudinal (x) and perpendicular (y) components of the E_{T} vector respectively.

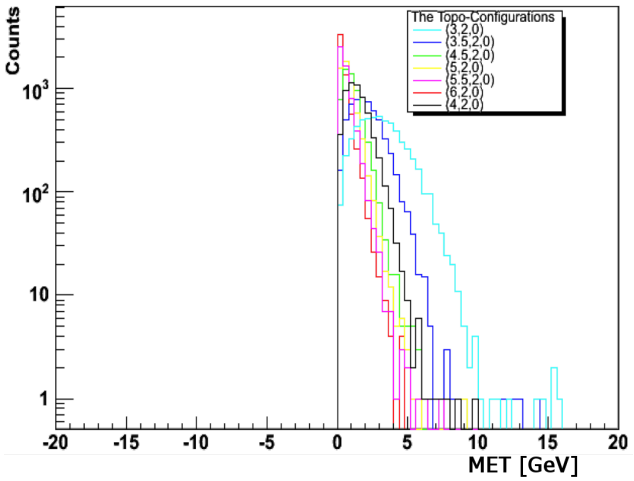

FIGURE 3.14: MET spectrum from randomly triggered events collected by the ATLAS detector during 2008 using different Topological Clustering algorithm configurations. The one-Gaussian approach is considered in the noise description.

The MET quantity is positive defined, as explicitly shown in Fig. 3.14. This distribution illustrates the dependence of the MET spectrum with respect to different Topological Clustering algorithm configurations. The width and tails observed in the MET distributions, computed as defined in Eq. 3.17, are directly correlated with the threshold value used in the definition of the seed cell in the Topological Clustering algorithm. In light of this, the MET measurement may depend on the selected TileCal noise description approach. To evaluate the impact of the two-Gaussian approach in the MET performance the nominal Topological Clustering configuration (4,2,0) is selected as a baseline.

Figure 3.15 compares the MET spectrums using both one-Gaussian and two-Gaussian methods to obtain the noise constants. In general, the datasets processed with the two-Gaussian approach present less tails than the cases using the one-Gaussian description. The differences are also noticeable at the level of 1 GeV considering the reconstructed peak for each method. There is an improvement in the resolution of the MET that can be evaluated by comparing the width of the spectrums in the different cases. Table 3.5 quotes the RMS values of the MET spectrums extracted from Fig. 3.15. Note that the run 91890 has been reprocessed using the two-Gaussian approach for direct comparison with the one-Gaussian. These results point to a reduction of 33-50% in the width of the MET distribution when the two-Gaussian approach is used with respect to the one-Gaussian.

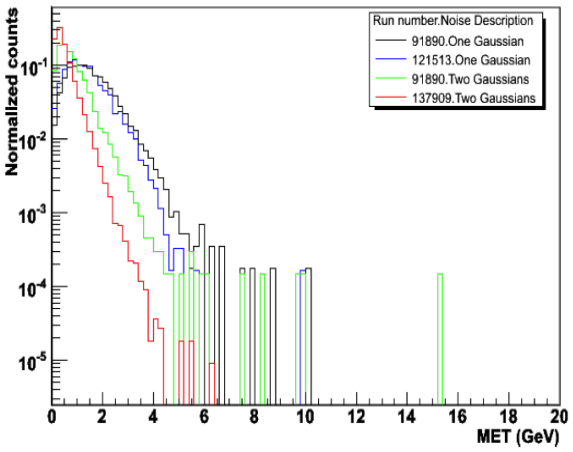

FIGURE 3.15: MET spectrum using both one-Gaussian and two-Gaussian models for the noise in TileCal. The data samples correspond to randomly triggered events collected by the ATLAS detector during 2008 and 2009. In the legend shows the run number and the noise description for each data sample. The distributions are normalised to unity for comparison.

Run	Year	Data Type	Noise Treatment	RMS [MeV]
91890	2008	randomly trigger	one-Gaussian	899
$91890^{(R)}$	2008	randomly trigger	two-Gaussians	632
121513	July 2009	randomly trigger	one-Gaussian	798
137909	Nov. 2009	randomly trigger	two-Gaussians	380
140535	Nov. 2009	randomly trigger + 1 beam	two-Gaussians	420

TABLE 3.5: RMS of the MET distributions obtained using randomly triggered events from non collision runs collected by the ATLAS detector during 2008 and 2009. The run $91890^{(R)}$ refers to the re-process procedure apply to this sample with the TileCal noise constants obtained from the two-Gaussian method.

In order to study the effect of the two-Gaussian approach for TileCal noise description in the tails of the MET distributions, the following investigations consider randomly triggered events satisfying MET $>$ 3 GeV and $|E_T| > 0.5$ GeV. These selected events will provide information related to the topocluster characteristics that form the tails observed in the MET distributions in Fig. 3.15. Figure 3.16 shows the second moment in λ for topoclusters satisfying the above energetic requirements when the one-Gaussian method is used. The results show that the high values of MET in randomly triggered events are mostly produced by large topoclusters ($\sqrt{<\lambda^2>} > 400$ mm). The contribution of the tails in the MET distribution is up to 15%. Hence any reduction of large topoclusters will reduce the MET tails.

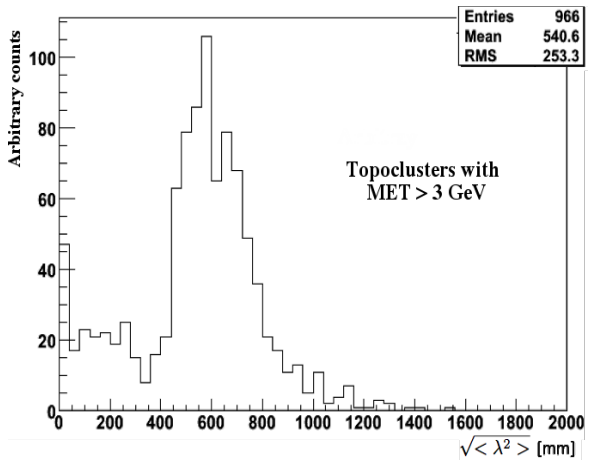

FIGURE 3.16: Second moment in λ using topoclusters satisfying the requirements in MET $>$ 3 GeV and $|E_T| > 0.5$ GeV. The randomly triggered events were collected by the ATLAS detector during 2009 using the one-Gaussian description for the TileCal noise constants (run 121513).

Figure 3.17 shows the noise occupancy in the $\eta - \phi$ plane of topoclusters with different energy thresholds. This distribution is expected to be homogeneous since the data sample corresponds to randomly triggered events collected by the ATLAS detector during 2009. However, there is a large contribution of low energetic topoclusters mainly localised in TileCal at $|\eta| < 1.6$ and $\phi \sim 2.5$ rad.

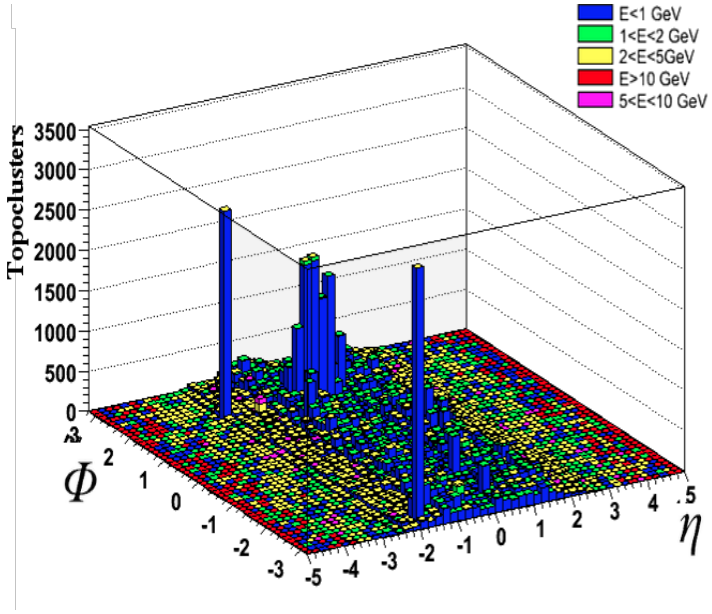

FIGURE 3.17: Noise occupancy in the $\eta - \phi$ plane for topoclusters with different energy thresholds. Low energetic topoclusters are mainly localised at $|\eta| < 1.6$ and $\phi \sim 2.5$ rad. The randomly triggered events were collected by the ATLAS detector during 2009 using the one-Gaussian description for the TileCal noise constants (run 121513).

Figure 3.18 shows the noise occupancy in the $x - y$ plane of large topoclusters located in $|\eta| < 1.6$ TileCal region and those topoclusters producing the observed MET tails. This plot shows a high occupancy of large topoclusters in a TileCal region around the point $(x, y) = (0.0, 3.0)$ m. Half of the events illustrated in Fig. 3.18 contain topoclusters located in this region. This affects the computation of the MET since the presence of large energetic topoclusters unbalances the measurement of the MET quantity thereby creating the observed long tails. The identification of this anomaly in the TileCal performance requires that action be taken for the affected cells. These actions go from trying to recover the problematic channels to label them as bad channels in the database.

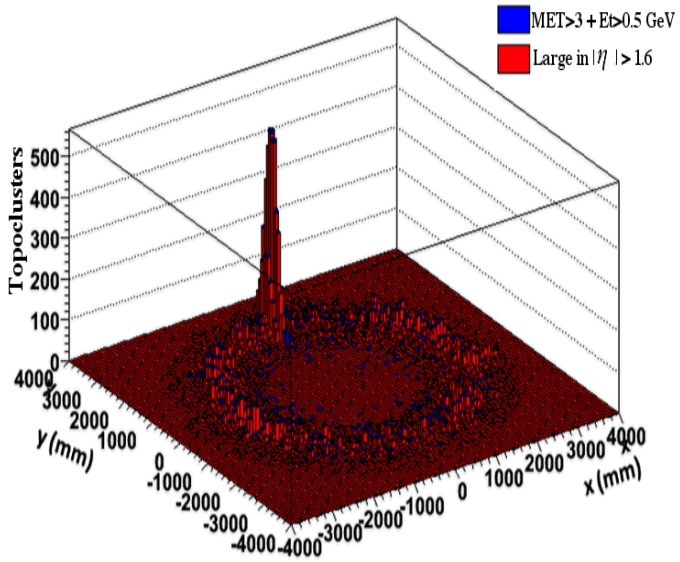

FIGURE 3.18: Noise occupancy in the $x-y$ plane of large topoclusters in TileCal region $|\eta| < 1.6$ (red) and those satisfying the requirements in MET $> 3\,\text{GeV}$ and $|E_T| > 0.5\,\text{GeV}$ (blue). Large topoclusters are mainly produced around the region $(x, y) = (0.0, 3.0)\,\text{m}$. The randomly triggered events were collected by the ATLAS detector during 2009 using the one-Gaussian description for the TileCal noise constants (run 121513).

Run	Noise description	Total events	Events with MET $> 3\,\text{GeV}$	Contribution of MET tails
121513	one-Gaussian	6060	966	16%
137909	two-Gaussians	109927	136	0.1%

TABLE 3.6: Contribution of the MET tails observed in Fig. 3.15 for randomly triggered events from non collision runs collected by the ATLAS detector during 2008 and 2009. The last column quotes the fraction of events with MET $> 3\,\text{GeV}$ over the total number of processed events for each collection of data.

The introduction of the two-Gaussian approach for the derivation of the TileCal noise constants allows to reduce the formation of large topoclusters by almost a factor 10, as shown in Tab. 3.4. Moreover, this procedure relies on an improvement on the MET resolution up to a factor 2, as quoted in Tab. 3.5. The benefit of identifying the region with large topoclusters forming the MET tails can be quantified by comparing the contribution of large topoclusters in the high MET region. Table 3.6 quotes the high reduction in the MET tails formation comparing

randomly triggered events for two data collections. The two-Gaussian approach and the treatment of the hot spots in TileCal leads to a reduction of the MET tails from 16% to 0.1%.

Finally, Fig. 3.19 illustrates the noise occupancy of the topoclusters satisfying the requirements on MET $>$ 3 GeV and $|E_T| > 0.5$ GeV after considering the two-Gaussian approach for the TileCal noise constants derivation and after the considerations taken over the highly populated area shown in Fig. 3.18. The occupancy is more homogeneous, as expected for randomly triggered events.

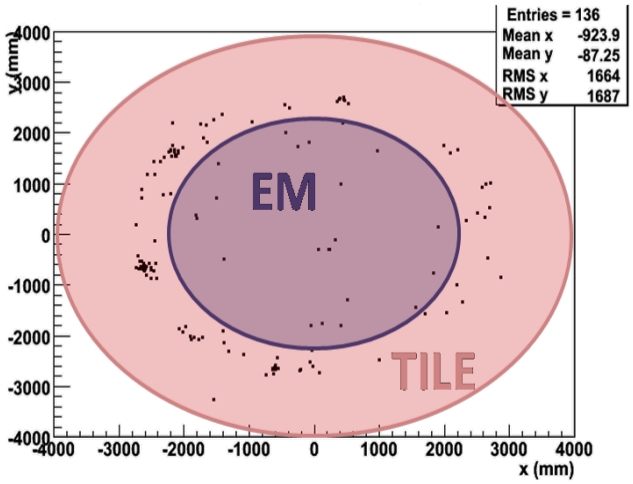

FIGURE 3.19: Noise occupancy in the $x - y$ plane for topoclusters satisfying MET $>$ 3 GeV and $|E_T| > 0.5$ GeV. The randomly triggered events were collected by the ATLAS detector during 2009 using the two-Gaussian description for the TileCal noise constants (run 137909). The region covered by the TileCal and the EM detectors is coloured in pink and purple respectively.

3.6 Conclusions

The reconstruction of the cell energy and time in the TileCal detector is provided by the OF algorithm. The determination of the cell noise is crucial for the object reconstruction algorithms since they rely on these noise values for discriminating the signal. The impact of the TileCal noise constants stored in the ATLAS conditions database has been evaluated through the objects formed by the Topological Clustering algorithm. The results from the topocluster performance allow to identify anomalies in the TileCal noise description using randomly triggered events. The

more accurate description of the TileCal noise constants made by the two-Gaussian method, as well as the treatment of localised hot spots result in a direct benefit the performance of the topoclusters and the MET measurement. The comparison of the topocluster moments using the two-Gaussian method is crucial to validate the two-Gaussian approach. The results show better MET resolution and a reduced population in the high MET region using randomly triggered events collected by the ATLAS detector during 2008 and 2009.

Other important contribution to the cell noise is the effect from simultaneous proton-proton interactions in the same bunch-crossing. A more reliable TileCal noise description with increasing pile-up conditions is described in Ref. [58]. Finally, a complete study on the performance of the Topological Clustering algorithm in the ATLAS calorimeters during LHC Run I can be found in Ref. [54].

Chapter 4

Missing Transverse Momentum in ATLAS

4.1 Introduction

The missing transverse momentum is a fundamental quantity to reconstruct physics processes produced at hadron colliders. Some particles, such as neutrinos or new stable weakly-interacting particles, traverse matter with a negligible probability of interaction. Hence, no direct evidence of them can be measured in a general purpose detector. However, the total momenta in the perpendicular plane to the beam axis has to be conserved, so any transverse momentum imbalance may signal the presence of such undetectable particles. In this light, the missing transverse momentum in ATLAS is defined as the negative vector sum of the transverse energy measured from all detected particles in an event. The symbol $\not{E}_T$ will be used in the following to represent the magnitude of this vector.

An optimised reconstruction and calibration of the $\not{E}_T$ is crucial in any search involving processes with low interacting particles in the final state. This is the case of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis described due to the presence of two neutrinos. Because of the complexity of measuring undetectable particles, this task represents one of the main challenges in collider experiments. Limited detector coverage and resolution, non-instrumented regions, as well as cosmic rays and beam-halo particles crossing the detector, can affect the $\not{E}_T$ reconstruction.

The most important sources affecting the $\not\!\!E_T$ measurement in ATLAS are,

- **Pile-up**. Additional proton-proton collisions superimposed to the hard scattering process may originate particles which deposit energy in the detector. These extra energetic contributions will enter in the $\not\!\!E_T$ computation, affecting significantly the genuine measurement. The high pile-up environment suffered during 2012 at the LHC motived different investigations to reduce this effect in the $\not\!\!E_T$ reconstruction.

- **Efficiency of ATLAS calorimeters**. The capability to discriminate noise from signal when reconstructing the energy at the cell level is directly propagated to the high-p_{T} objects tagged as leptons, photons, jets... As extensively reported in Chapter 3, the two-Gaussian approach for extracting the TileCal noise constants is validated through topoclusters' quantities. Topoclusters are used as inputs in the $\not\!\!E_T$ computation, hence the calorimeter noise description significantly affects the $\not\!\!E_T$ measurement, as discussed in Section 3.5.1.

This chapter provides a full overview on the $\not\!\!E_T$ performance in terms of data/MC comparisons, resolution, response and tails of several $\not\!\!E_T$ measurements computed under different approaches. Section 4.2 defines three different algorithms for reconstructing the $\not\!\!E_T$ magnitude in ATLAS. Characteristics and information of the data and simulated samples analysed in this chapter are detailed in Section 4.3. For all three $\not\!\!E_T$ definitions, the quality of data and simulation agreement is reported in Section 4.4. Section 4.5 presents the investigations developed to quantify the performance as well as comparisons between the different $\not\!\!E_T$ measurements in $Z \rightarrow \ell\ell$ and $H \rightarrow WW^{(*)} \rightarrow \ell\nu\ell\nu$ processes. Finally, the treatment for obtaining the systematic uncertainties of the $\not\!\!E_T$ measurement is described in Section 4.6.

4.2 Missing Transverse Momentum Reconstruction

This section presents three different algorithms developed in ATLAS for reconstructing the $\not\!\!E_T$ magnitude. These are differentiated by which part of the detector provides the energetic measurements that enter in each calculation. There are two different approaches to obtain the $\not\!\!E_T$ measurement.

Calorimeter-based $\not\!\!E_T$ definitions make use of the energy reconstructed by the calorimeters. The first definition is described in Section 4.2.1 and will referred as $E_{\mathrm{T}}^{\mathrm{miss}}$ in the following. Due to the increase in the average of interactions per bunch crossing during 2012 data taking conditions, a correction was developed to minimise

the impact from pile-up in the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement. This pile-up suppression technique relies on track information and is applied to several $E_{\mathrm{T}}^{\mathrm{miss}}$ terms as described in Section 4.2.2. This correction defines the pile-up suppressed calorimeter-based measurement, known as $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$.

The $\not{E}_T$ computation described in Section 4.2.3 uses as input the transverse momentum of the tracks produced by charged-particles traversing the ID system. This measurement efficiently suppresses pile-up contributions since only tracks associated to the hard scattering process are selected. The symbol $E_{\mathrm{T}}^{\mathrm{miss,track}}$ will be used in the following for referring to this track-based measurement.

4.2.1 Calorimeter-Based $\not{E}_T$: $E_{\mathrm{T}}^{\mathrm{miss}}$

The $E_{\mathrm{T}}^{\mathrm{miss}}$ magnitude is reconstructed from energetic deposits in the calorimeters and muons reconstructed in the MS [59, 60]. These energetic deposits are associated with a reconstructed and identified high-p_{T} parent object in the following order: electrons (e), photons (γ), taus (τ), high-p_{T} jets (jets), and muons (μ). Remaining energetic contributions, not associated to any such objects, are also considered in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation through the so-called soft-term. This term includes energetic measurements from topoclusters reconstructed by the Topological Clustering algorithm with (4, 2, 0) configuration, as described in Section 3.3. The selection criteria applied to each high-p_{T} object identification is detailed in Appendix B. All these contributions are accounted for in separated non-overlapped terms to compute the $x-$ and $y-$components of the $E_{\mathrm{T}}^{\mathrm{miss}}$ as follows,

$$E_{x(y)}^{\mathrm{miss}} = E_{x(y)}^{\mathrm{miss},\,e} + E_{x(y)}^{\mathrm{miss},\,\gamma} + E_{x(y)}^{\mathrm{miss},\,\tau} + E_{x(y)}^{\mathrm{miss,\,jets}} + E_{x(y)}^{\mathrm{miss,\,SoftTerm}} + E_{x(y)}^{\mathrm{miss},\,\mu}\,, \quad (4.1)$$

where each term is calculated from the negative sum of the reconstructed energy of the objects, projected onto the x and y directions.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ magnitude and its azimuthal coordinate (ϕ^{miss}) are then defined by the following expressions,

$$E_{\mathrm{T}}^{\mathrm{miss}} = \sqrt{\left(E_x^{\mathrm{miss}}\right)^2 + \left(E_y^{\mathrm{miss}}\right)^2}\,, \quad (4.2)$$

$$\phi^{\mathrm{miss}} = \arctan\left(E_y^{\mathrm{miss}},\, E_x^{\mathrm{miss}}\right)\,. \quad (4.3)$$

4.2.2 $E_{\mathrm{T}}^{\mathrm{miss}}$ with Pile-up Suppression: $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$

Some of the terms in Eq. 4.1 have a proper pile-up suppression technique applied already which is included at the reconstruction level for such objects. However, the soft- and jet-term specially suffer from pile-up. Tracks provide an excellent method to mitigate this effect. This approach relies on vertices information for subtracting pile-up contributions in the soft- and jet-term [61, 62].

Tracks can be associated with the hard scattering process through the main reconstructed vertex, so-called primary vertex (PV). In ATLAS, the PV is defined by the maximum sum of the transverse momenta of tracks ($p_{\mathrm{T}}^{\mathrm{track}}$) emerging from the vertex,

$$\sum_{\mathrm{tracks}}^{\mathrm{PV}} \left(p_{\mathrm{T}}^{\mathrm{track}}\right)^2 = \max . \tag{4.4}$$

In general for $\not{E}_T$ reconstruction, energetic contributions not associated with the PV can be safely interpreted as originating from one of the additional pile-up interactions, so they can be completely excluded. Pile-up suppression in the $E_{\mathrm{T}}^{\mathrm{miss}}$ definition is then achieved by scaling the soft-term with the pile-up contribution in each event, and by correcting the jet-term in order to reject jets produced by extra interactions. The soft-term vertex-fraction (STVF) is a weight factor computed as the fraction of scalar sum of $p_{\mathrm{T}}^{\mathrm{track}}$ for tracks associated with the PV relative to the total scalar sum of $p_{\mathrm{T}}^{\mathrm{track}}$ including pile-up interactions,

$$\mathrm{STVF} = \frac{\sum_{\mathrm{SoftTerm}}^{\mathrm{PV}} p_{\mathrm{T}}^{\mathrm{track}}}{\sum_{\mathrm{SoftTerm}}^{\mathrm{All}} p_{\mathrm{T}}^{\mathrm{track}}} , \tag{4.5}$$

with STVF being $0 \leq \mathrm{STVF} \leq 1$. The sums in Eq. 4.5 are taken over the tracks unmatched to high-p_{T} physics objects in the other $E_{\mathrm{T}}^{\mathrm{miss}}$ terms and tracks in the numerator have to satisfy association criterion with the PV. This track selection criteria used to compute the STVF factor is detailed in Appendix B, Section B.2. This pile-up correction is then applied by scaling the soft-term in Eq. 4.1 by the STVF factor,

$$E_{\mathrm{T,\,STVF}}^{\mathrm{miss,\,SoftTerm}} = \mathrm{STVF} \cdot E_{\mathrm{T}}^{\mathrm{miss,\,SoftTerm}} . \tag{4.6}$$

The second pile-up correction concerns jets used to compute the jet-term. This correction relies on identification of jets originated in the hard scattering process, through the use of tracking and vertexing information. By combining tracks

and their primary vertices with calorimeter jets it is possible to define a discriminant, known as jet vertex fraction (JVF) which measures the probability that a jet was originated from the PV. The JVF is very similar to the STVF correction shown in Eq. 4.5, but now applied to tracks associated with a jet,

$$\mathrm{JVF} = \frac{\sum\limits_{\mathrm{jet}}^{\mathrm{PV}} p_{\mathrm{T}}^{\mathrm{track}}}{\sum\limits_{\mathrm{jet}}^{\mathrm{All}} p_{\mathrm{T}}^{\mathrm{track}}} . \tag{4.7}$$

The JVF quantity is assigned jet by jet although it is limited and can not always be computed. The JVF only can be calculated for jets well within the ID acceptance ($|\eta_{\mathrm{jet}}| < 2.4$), and with associated tracks to the vertices. The pile-up suppression of the jet-term requires jets with $20 < p_{\mathrm{T}}^{\mathrm{jet}} < 50\,\mathrm{GeV}$ in the ID volume to satisfy the JVF> 0 criteria. If the condition on the JVF is not satisfied, their signals are completely excluded from the jet-term. This filter efficiently selects soft p_{T}-jets coming from the PV, while still provides rejection of jets produced by pile-up vertices (JVF $= 0$). Jets with larger p_{T}, or those for which the JVF can not be computed, will always be inputs to the jet-term. The new jet-term considering the JVF filter described above is denoted as $E_{\mathrm{T,\,JVF}}^{\mathrm{miss,\,jets}}$.

The components of the calorimeter-based $\not\!\!E_T$ with pile-up suppression components ($E_{x(y)}^{\mathrm{miss,\,STVF}}$) have the same form as in Eq. 4.1,

$$E_{x(y)}^{\mathrm{miss,\,STVF}} = E_{x(y)}^{\mathrm{miss,\,e}} + E_{x(y)}^{\mathrm{miss},\,\gamma} + E_{x(y)}^{\mathrm{miss},\,\tau} + \mathbf{E}_{\mathbf{x(y),\,JVF}}^{\mathbf{miss,\,jets}} + \mathbf{E}_{\mathbf{x(y),\,STVF}}^{\mathbf{miss,\,SoftTerm}} + E_{x(y)}^{\mathrm{miss},\,\mu} , \tag{4.8}$$

except that now the two highlighted contributions refer to the pile-up suppressed terms discussed above. In this calculation with pile-up mitigated terms not only the magnitude is corrected,

$$E_{\mathrm{T}}^{\mathrm{miss,\,STVF}} = \sqrt{\left(E_x^{\mathrm{miss,\,STVF}}\right)^2 + \left(E_y^{\mathrm{miss,\,STVF}}\right)^2} , \tag{4.9}$$

but also its direction,

$$\phi^{\mathrm{miss,\ STVF}} = \arctan\left(E_y^{\mathrm{miss,\,STVF}},\, E_x^{\mathrm{miss,\,STVF}}\right) . \tag{4.10}$$

4.2.3 Track-Based $\not{E}_T$: $E_{\mathrm{T}}^{\mathrm{miss,track}}$

This section describes a method to estimate the $\not{E}_T$ differently to the calorimeter-based $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ approaches presented in Section 4.2.1 and 4.2.2, respectively. This new algorithm only relies on reconstructed tracks in the ID to compute the transverse momentum imbalance in the event [63, 64]. The main advantage of using tracks relies on the selection of the PV, which provides a powerful discrimination against pile-up contributions. However, this approach is also limited due to the characteristics of the ATLAS tracker system compared with the calorimeters. Only charged-particle information can be used in the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ reconstruction, since neutral particles are not measured in the ID. Moreover, the geometrical coverage of the ID ($|\eta| < 2.4$) is smaller than the calorimeters ($|\eta| < 4.9$).

The track-based $\not{E}_T$ variable is computed as the vector sum of the reconstructed momentum of the tracks measured with the ID. The nominal x- and y-components of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ are calculated as in the following expression,

$$E_{x(y)}^{\mathrm{miss,\,track}} = -\sum_{\mathrm{tracks}}^{\mathrm{PV}} p_{x(y)}^{\mathrm{track}}\,, \tag{4.11}$$

where the tracks refer to those satisfying the full selection criteria described in Appendix B, and having an association with the PV (see Section B.3 for details). Finally, the track-based magnitude and its direction are obtained as follows,

$$E_{\mathrm{T}}^{\mathrm{miss,track}} = \sqrt{\left(E_x^{\mathrm{miss,\,track}}\right)^2 + \left(E_y^{\mathrm{miss,\,track}}\right)^2}\,. \tag{4.12}$$

$$\phi^{\mathrm{miss,\,track}} = \arctan\left(E_y^{\mathrm{miss,\,track}},\, E_x^{\mathrm{miss,\,track}}\right)\,. \tag{4.13}$$

The $E_{\mathrm{T}}^{\mathrm{miss,track}}$ can be considered as complementary to the previous calorimeter-based quantities. The $E_{\mathrm{T}}^{\mathrm{miss}}$ will provide an overestimated measurement in events with high pile-up since all interactions are grouped together in the calculation. Conversely, the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ can be mis-measured in some event topologies because the presence of neutral particles or particles within $|\eta| > 2.4$. This complementarity benefits the isolation of a process of interest with genuine $\not{E}_T$ from other contributions without genuine $\not{E}_T$, specially under high pile-up conditions. This is the strategy of the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis described in Chapter 5. The $H\to WW^{(*)}\to \ell\nu\ell\nu$ criterion use combined requirements on $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ to further suppress Z decay events, which form the dominant background of the search.

4.2.4 $\not\!\!E_T$ Related Variables

This section defines two quantities related to the $\not\!\!E_T$. First, the scalar sum of the total transverse energy measured by the ATLAS calorimetric system is presented. This will be used to evaluate the performance of the different $\not\!\!E_T$ approaches. The other $\not\!\!E_T$-related quantity is very useful to suppress $Z \to ee$ and $Z \to \mu\mu$ processes. This variable relies on the proximity of the $\not\!\!E_T$ with high-p_{T} particles in the final state.

4.2.4.1 Total Transverse Energy in the Calorimeters

The total transverse energy in the calorimeters ($\sum E_{\mathrm{T}}$) is an important quantity to parametrise the $\not\!\!E_T$ performance. This quantity is defined as the scalar sum of the transverse energy of the contributions entering in the $\not\!\!E_T$ computation. According to the $E_{\mathrm{T}}^{\mathrm{miss}}$, the $\sum E_{\mathrm{T}}$ is given by the following expression,

$$\sum E_{\mathrm{T}} = \sum E_{\mathrm{T}}^{e} + \sum E_{\mathrm{T}}^{\gamma} + \sum E_{\mathrm{T}}^{\tau} + \sum E_{\mathrm{T}}^{\mathrm{jet}} + \sum E_{\mathrm{T}}^{\mathrm{SoftTerm}} . \qquad (4.14)$$

Note that muons do not appear in Eq. 4.14 because their momenta are measured from MS and ID reconstructed tracks. The resolution on calorimeter objects is much different from ones obtained with tracks, which motivates this omission. The $\sum E_{\mathrm{T}}$ measurement is sensitive to both, the number of energetic depositions included in the calculation and their reconstructed E_{T}, so this quantity describes the hardness of the event.

4.2.4.2 $\not\!\!E_T$ Relative to the Closest Lepton or Jet

Events with mis-measured energies of leptons or jets will reconstruct the $\not\!\!E_T$ poorly. In such cases, the $\not\!\!E_T$ will most likely point in the direction of the mis-measured object. Aiming to reduce the rate of processes that arise from these mis-measurements, a correction is introduced on the $\not\!\!E_T$ magnitude. The new quantity is built projecting the nominal $\not\!\!E_T$ in the direction of the closest jet or lepton ($\not\!\!E_{T,Rel}$). Accordingly to the $E_{\mathrm{T}}^{\mathrm{miss}}$ definition, the relative $E_{\mathrm{T}}^{\mathrm{miss}}$ magnitude ($E_{\mathrm{T,Rel}}^{\mathrm{miss}}$) is computed as follows,

$$E_{\mathrm{T,Rel}}^{\mathrm{miss}} = \begin{cases} E_{\mathrm{T}}^{\mathrm{miss}}, \text{ if } \Delta\phi \geqslant \pi/2 \\ \\ E_{\mathrm{T}}^{\mathrm{miss}} \times \sin\Delta\phi, \text{ if } \Delta\phi < \pi/2 \end{cases} ,$$

where $\Delta\phi$ refers to the minimum angle between the $E_{\mathrm{T}}^{\mathrm{miss}}$ direction and the nearest lepton or jet (obj): $\Delta\phi = \min(|\phi^{\mathrm{miss}} - \phi^{\mathrm{obj}}|)$. In the same way, this projection can be applied to the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ definitions. Those will be represented by the symbols $E_{\mathrm{T,Rel}}^{\mathrm{miss,STVF}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$, respectively.

Figure 4.1 shows the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ spectrums obtained from data and simulation in $Z \to \mu\mu$ events. This comparison illustrates the benefit of using $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ instead of $E_{\mathrm{T}}^{\mathrm{miss}}$ to further suppress processes without genuine $\not\!\!E_T$. The number of events located at $E_{\mathrm{T}}^{\mathrm{miss}} < 20\,\mathrm{GeV}$ is almost an order of magnitude higher in the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ distribution and the tails are significantly smaller. The mean and RMS from the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ distribution are reduced by $\sim 30\%$ and $\sim 20\%$ with respect to the $E_{\mathrm{T}}^{\mathrm{miss}}$ values. These behaviours are also observed in $Z \to ee$ events, as well as from $E_{\mathrm{T,Rel}}^{\mathrm{miss,STVF}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ comparisons.

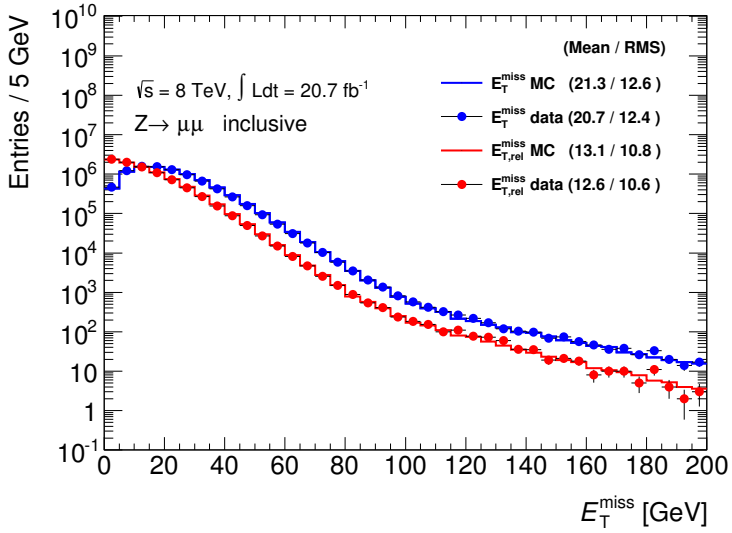

FIGURE 4.1: $E_{\mathrm{T}}^{\mathrm{miss}}$, in blue, and $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$, in red, spectrums for 8 TeV data (dots) and simulated MC samples (lines). The events are selected to be compatible within the $Z \to \mu\mu$ process. The legend shows the mean and RMS values.

In the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis strategy described in Chapter 5, lower bounds on $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ magnitudes are applied in final states with same flavoured leptons. The combination of both definitions highly suppresses Z/DY contributions in benefit of the Higgs boson signal significance. The description of the $\not\!\!E_T$ selection criteria used in this analysis and in particular, the latest $\not\!\!E_T$ optimisation results are presented in Section 5.6.3 and Chapter 6 respectively.

4.3 Data Samples and Event Selection

In this chapter, the performance of the $\not\!\!E_T$ is evaluated using ATLAS data collected during 2012 using proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$ in the LHC. Details on LHC operation are given in Chapter 2, Section 2.2. The most important feature in the 2012 LHC running period is the enhance of pile-up interactions with respect to 2011. The mean number of inelastic collisions per bunch crossing for 2012 is up to 35 with an average of 20 over the full data set, as shown in Fig. 2.6.

Simulation samples are treated to generate the pile-up environment in data properly. Pile-up modelling for all MC samples comes from PYTHIA8. These extra collisions are overlaid to the physics process that is being simulated before reconstructing objects in the event as tracks, photons, leptons, jets, etc. In MC events, the genuine $\not\!\!E_T$ is calculated from all generated non-interacting particles in the simulated process. It will be referred to as true missing transverse momentum in the following and the symbol $E_{\text{T}}^{\text{miss, True}}$ will be used to denote this quantity. Several MC generators and parton shower models are used for the following results. MC samples are generated over a large range of jet activity in order to quantify its effect on the $\not\!\!E_T$ measurements. Details on the different MC generators are described in Section 5.3.

The criterion for selecting each event topology first consider a set of quality requirements at the level of reconstructed objects. The selection of leptons and jets, as well as the ATLAS reconstruction algorithms, are described in Section 5.6. Candidate events should satisfy a selection that guarantees compatibility with the process of interest. $Z \to e^{\pm}e^{\mp}$ and $Z \to \mu^{\pm}\mu^{\mp}$ decays are interesting for comparing observed data with simulation and, for evaluating the $\not\!\!E_T$ performance in events without genuine $\not\!\!E_T$ measurement. For sake of simplicity, these events will be grouped and referred to as $Z \to \ell\ell$ in the following. $Z \to \ell\ell$ events are required to have two well-reconstructed high-p_{T} leptons with opposite charge and same flavour. Moreover, the invariant mass of the dilepton system has to be consistent within the Z mass peak, $66 < m_{\ell\ell} < 116\,\text{GeV}$. Processes with genuine $\not\!\!E_T$ measurement are also studied since they allow to compare the different $\not\!\!E_T$ measurements with the expected $E_{\text{T}}^{\text{miss, True}}$ quantity. For the performance investigations in Section 4.5.4, the $H \to WW^{(*)} \to \ell\nu\ell\nu$ process is used. The details on the generators of the Higgs boson samples, with $m_{\text{H}} = 125\,\text{GeV}$, are quoted in Tab. 5.1.

4.4 Data and Simulation Comparisons

This section collects some basic $\not{E}_T$-related quantities for the three $\not{E}_T$ reconstructions defined in Section 4.2. These results compare ATLAS collected data with MC expectations in $Z \to \ell\ell$ events. The advantages of using this process are the clean signature of the two leptons from the Z decay and the large amount of Z bosons produced at LHC. Most of the $\not{E}_T$ measured in such events results from imperfections in computing the $\not{E}_T$ magnitude, as object selection efficiency, or in detector response. Hence, this event topology allows to evaluate detector noise and pile-up effects in the different $\not{E}_T$ reconstructions.

TABLE 4.1: Expected and observed event yields with 8 TeV ATLAS data (Observed) and simulation from MC (Total Bkg.). Different columns show the expected event yields for each SM process satisfying the $Z \to \ell\ell$ selection criteria. The non-WW column includes WZ, ZZ and $W\gamma$ processes. Groups separated by a horizontal line consider different jet activity in the final state. The quoted uncertainties are only due to sample statistics.

Selection	WW	non WW	$t\bar{t}$	Single Top	Z+jets	W+jets	Total Bkg.	Observed	$\frac{Data}{MC}$
Inclusive	2558 ± 12	4970 ± 15	13036 ± 41	1240 ± 5	14157500 ± 7800	18960 ± 180	14213900 ± 7800	14399977	1.01
0 jets	1671 ± 10	2480 ± 8	158 ± 4	96.1 ± 1.4	10875400 ± 6800	13730 ± 150	10894100 ± 6800	11037562	1.01
1 jet	659 ± 6	1554 ± 9	1802 ± 15	577.4 ± 3.4	2431000 ± 3300	3700 ± 80	2443100 ± 3300	2483850	1.02
$\geq$ 2 jets	228.1 ± 4.3	936 ± 9	11076 ± 36	566.5 ± 3.5	851100 ± 1200	1530 ± 40	876700 ± 1200	878565	1.00

Table 4.1 shows the event yields from simulation and 2012 ATLAS data after applying the criteria for selecting $Z \to \ell\ell$ events. The total number of events in data and simulation are compatible within the statistical errors. The MC decomposition of the different processes shows that 99.6% of the events are expected to be produced by a $Z \to \ell\ell$ decay while other SM processes that may have the same reconstructed final state only contribute by ∼0.4%. The purity of a selected phase space is considered as the relative event yields obtained with the sample of interest over the total number of selected events, considering other SM processes as well. In this case, the Z purity in events with no jets nor exactly 1 jet exceeds the 99%. The purity of the Z sample in presence of at least two jets is ∼ 97%. This reduction on the Z purity is due to the relatively higher contribution of $t\bar{t}$ events. Other SM processes satisfying Z selection, including $t\bar{t}$, WW and WZ diboson events, involve genuine $\not{E}_T$ measurement, so they will mostly contribute to the tails of the $\not{E}_T$ distribution.

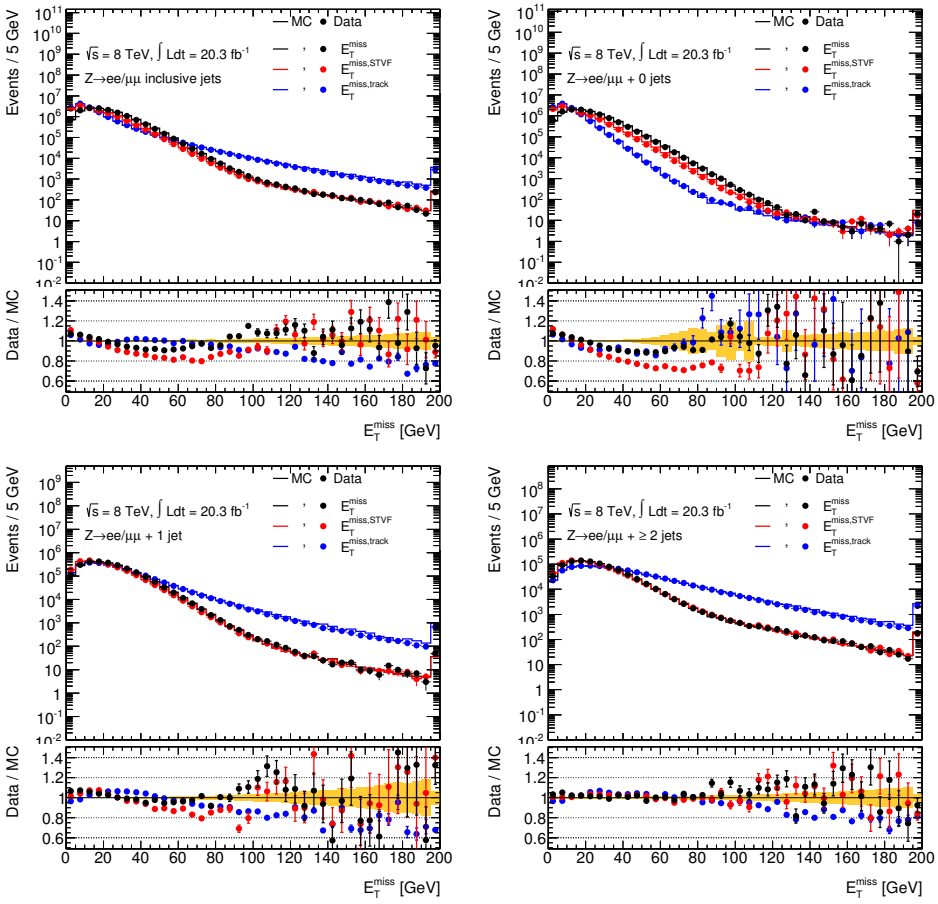

FIGURE 4.2: Different $\not\!\!E_T$ reconstruction in $Z \to \ell\ell$ candidates for 8 TeV ATLAS data and MC simulation: $E_{\mathrm{T}}^{\mathrm{miss}}$ (black), $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ (red) and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ (blue). Distributions are shown for different jet multiplicities: inclusive (top-left), 0-jet (top-right), 1-jet (bottom-left) and at least two jets (bottom-right).

Figure 4.2 shows the different $\not\!\!E_T$ spectrums obtained from 2012 ATLAS data and MC simulation for $Z \to \ell\ell$ candidate events with different number of jets in the final state. In general, the agreement between data and the MC is reasonable. However, for some regions of the distributions the discrepancy between data and MC can be up to 30%. There are two MC features responsible for the observed differences,

- soft jet activity is difficult to model, and
- concrete contributions from each pile-up interaction can not be predicted, as they may be originated by many possible and different final states.

The highest data/MC difference is observed for the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ measurement in events without jets. This is due to the bad modelling of the simulated soft-term which accounts for low-p_{T} jets by definition. Moreover, the computation of the STVF factor in MC results in an overcorrection of the STVF soft-term as this relies on tracks from pile-up vertices. In events with jets, the contribution of the jet-term becomes significant and reduces the effects from the STVF soft-term. In the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction the discrepancies do not exceed 10%. As in the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ case, the soft-term is not well described by the simulation. However, the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction shows better agreement than the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ since the STVF factor is not applied for the former. In summary, the limitations of the simulated soft-term affects both calorimeter-based definitions. The inaccurate pile-up modelling in MC enhances the disagreement between data and simulation through the application of the STVF factor.

Conversely, differences between data and MC in the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ measurement do not originate from pile-up effects since only tracks from the PV vertex are used. In this case, the discrepancies are up to $\sim 20\%$ and are mainly located in the tails of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ spectrum. The source of this disagreement is related to the simulation of the ID activity in events with high number of jets in the final state.

TABLE 4.2: Mean and RMS values expressed in GeV from the different $\not{E}_T$ distributions collected in Fig. 4.2. ATLAS data events collected during 2012 at $\sqrt{s} = 8\,\mathrm{TeV}$ being consistent with a $Z \to \ell\ell$ decay are used considering different jet multiplicities.

$Z \to \ell\ell$	$E_{\mathrm{T}}^{\mathrm{miss}}$		$E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$		$E_{\mathrm{T}}^{\mathrm{miss,track}}$	
	Mean	RMS	Mean	RMS	Mean	RMS
0 jets	18.6	11.9	13.4	10.3	9.5	6.8
1 jet	22.3	12.5	18.1	12.4	24.8	18.7
≥ 2 jets	23.1	12.0	22.7	12.2	38.8	29.1

Table 4.2 quotes the mean and RMS values for data distributions from Fig. 4.2. For Z events without jets the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ performs better than the calorimeter-based quantities, as it provides the smallest mean and RMS values. The $E_{\mathrm{T}}^{\mathrm{miss,track}}$ measurement efficiently rejects the contribution from pile-up interactions. These extra interactions may deposit energy in the calorimeters, hence they will enter in the soft-term of the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ measurements. The latter suppresses pile-up contribution through the STVF correction, which decreases the mean and RMS values by $\sim 30\%$ with respect to the $E_{\mathrm{T}}^{\mathrm{miss}}$ in events without jets. However, the

$E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ presents the highest data/MC discrepancy due to the complexity of simulating pile-up contributions, as discussed above.

In general, the presence of jets degrades all three $\not\!\!E_T$ definitions. High jet activity originates wider spectrums and longer tails as shown in Fig. 4.2. The $E_{\mathrm{T}}^{\mathrm{miss,track}}$ measurement is specially sensitive to jet activity, being substantially more affected than calorimeter-based $\not\!\!E_T$ definitions as observed in Tab. 4.2. This is due to jets that scape detection in the ID, as those produced by neutral particles or outside the ID coverage range. The omission of neutral and forward jets in the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ computation results in an increase of $\sim 50\%$ in the mean and RMS values with respect to calorimeter-based definitions in events with at least two jets.

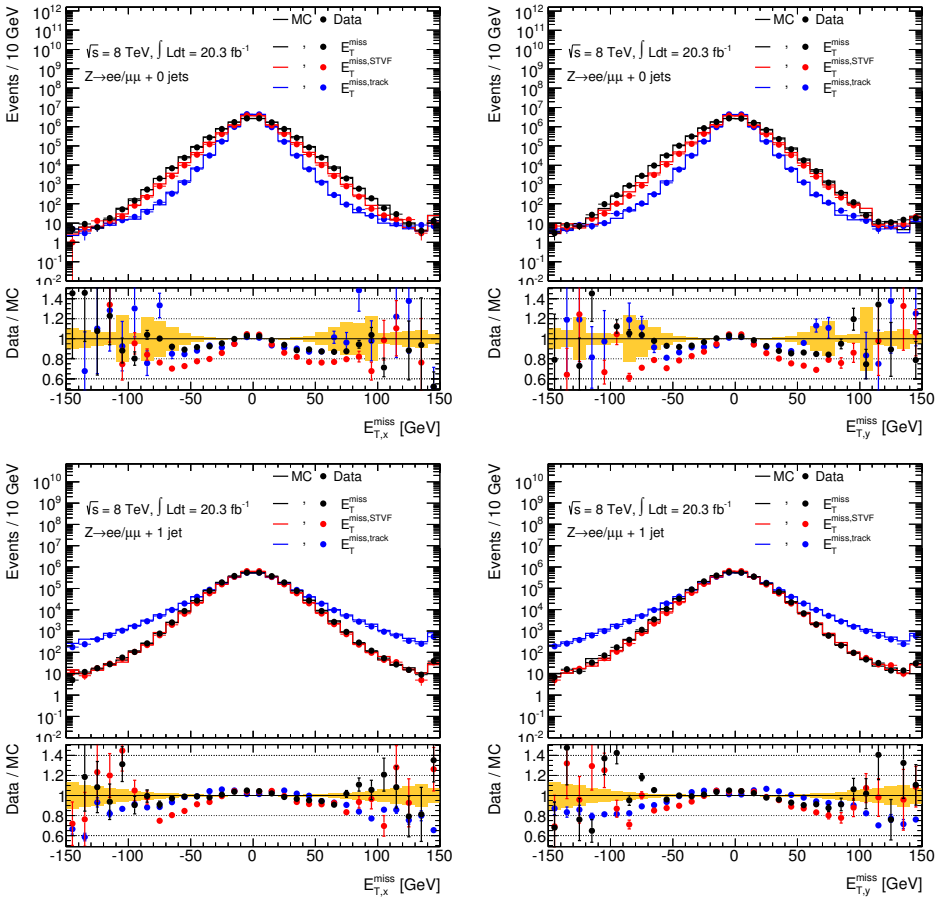

FIGURE 4.3: x and y components of the different $\not\!\!E_T$ reconstructions in $Z \to \ell\ell$ candidates for 8 TeV ATLAS data and simulation. Events are for $ee + \mu\mu$ final states with 0 and 1 jet on top and bottom, respectively.

Figure 4.3 compares the x- and y-components for each $\not\!\!E_T$ definition in $Z \to \ell\ell$ events in data and simulation. The directional components of the $\not\!\!E_T$ are expected to be centered at zero and described by a Gaussian in events without genuine $\not\!\!E_T$ measurement. However, pile-up interactions can affect the width of these distributions by introducing fluctuations that make them no longer Gaussian. In addition, the limitation of some $\not\!\!E_T$ reconstructions to efficiently account for all significant energy depositions, also worsens the Gaussian shape of the x and y-components. The $E_{\mathrm{T}}^{\mathrm{miss,track}}$ directional components show the narrowest distributions in events without jets. However, when considering events with jet activity, the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ measurement changes to a much wider distribution and higher populated tails than the calorimeter-based reconstructions.

TABLE 4.3: Mean and RMS values expressed in GeV of the x-component for each $\not\!\!E_T$ reconstruction. ATLAS data events collected during 2012 at $\sqrt{s} = 8\,\mathrm{TeV}$ being consistent with a $Z \to \mu\mu$ decay are used considering events with 0-jets and exactly 1-jet.

$Z \to \mu\mu$	$E_{\mathrm{T}}^{\mathrm{miss}}$		$E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$		$E_{\mathrm{T}}^{\mathrm{miss,track}}$	
	Mean	RMS	Mean	RMS	Mean	RMS
0 jets	-0.103	16.5	0.00752	12.43	-0.0771	8.92
1 jet	-0.0515	18.08	0.0367	16.43	-0.168	21.45

In order to numerically establish the degradation of the performance with the presence of jets, Tab. 4.3 quotes the mean and RMS values of the x-components for each $\not\!\!E_T$ reconstruction. Due to the distortion of the distributions shown in Fig. 4.3, the quality of the Gaussian fitted width is not a good estimator of the $\not\!\!E_T$ reconstruction. Using the RMS instead is more appropriated to comparatively quantify the performance of the different $\not\!\!E_T$ definitions. The RMS gives a quantitative description of the tails of the $\not\!\!E_T$ distributions and will be used in the following. In general, all three definition are centered at zero as represented by the mean values in Tab. 4.3. Hence, there is not a privileged direction which points to a bias in the $\not\!\!E_T$ reconstruction. The $E_{\mathrm{T}}^{\mathrm{miss,track}}$ measurement achieves the smallest RMS value in Z events without jets. This is almost a factor 2 lower than the RMS obtained for the directional components in the $E_{\mathrm{T}}^{\mathrm{miss}}$ case. The $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ slightly corrects the contribution of pile-up in the tails of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution. This is due to the suppression of extra interactions achieved through the application of the STVF factor. The comparison of these two calorimeter-based quantities show that the STVF correction reduces the RMS of the $E_{\mathrm{T}}^{\mathrm{miss}}$ x-component by $\sim 25\%$. The presence of a high-p_{T} jet degrades all three $\not\!\!E_T$ measurements, being the $E_{\mathrm{T}}^{\mathrm{miss,track}}$

the most affected quantity. Results show a factor 3 difference between the RMS of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ in events with a jet with respect to events without jets. $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ are more robust under the presence of jets in the event. The jet effects in calorimeter-based reconstructions enhance the RMS by $\sim 10\%$ with respect to the case without jets.

Besides the poor performance of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ measurement in events with jets, the computation of the $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ recovers the quality of the reconstruction up to the same level of the other calorimeter-based magnitudes. Figure 4.4 shows the comparisons between the different $\not\!\!E_{T,Rel}$ spectrums in data and MC simulation.

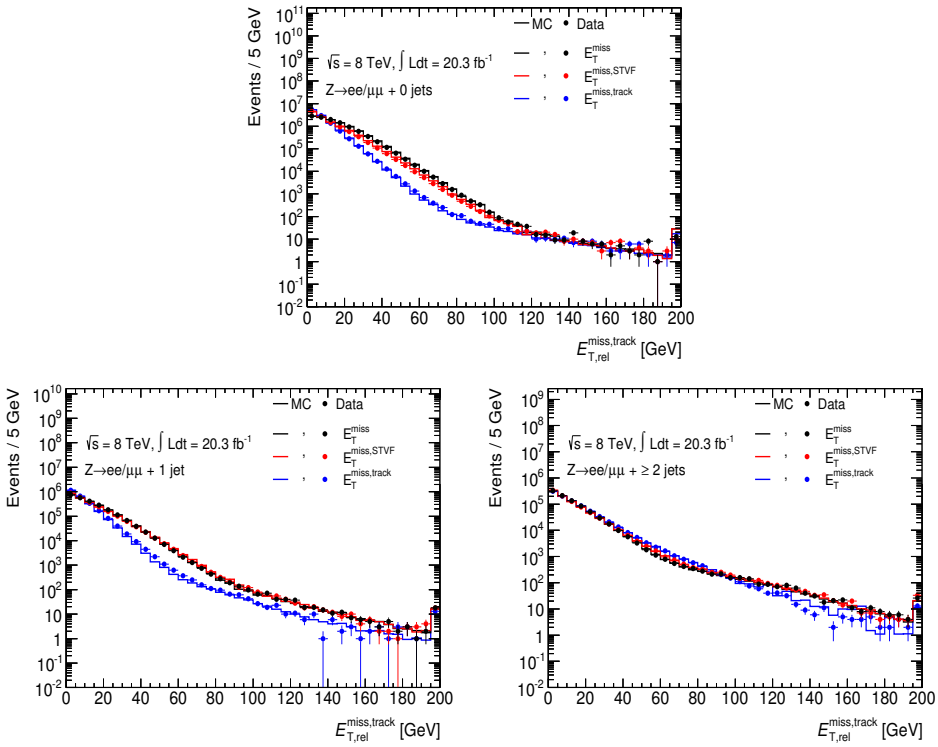

FIGURE 4.4: $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ comparisons for 2012 data and MC simulation in $Z \rightarrow \ell\ell$ events. The events are divided by different number of jets in the final state.

In general, the $\not\!\!E_{T,Rel}$ improves the nominal $\not\!\!E_T$ measurement in $Z \rightarrow \ell\ell$ events. The bulk of the $\not\!\!E_{T,Rel}$ distributions are closer to zero and there is a significant reduction of the tails compared with the nominal $\not\!\!E_T$ distributions from Fig. 4.2. The $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ is the most performant reconstruction in events without jets, as expected from the nominal $E_{\mathrm{T}}^{\mathrm{miss,track}}$ results. In addition, the $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ distribution improves to the level of the calorimeter-based reconstructions in events with jet activity as well. As observed in the $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ spectrums, the width becomes narrower and the tails are similar than the ones obtained with $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,STVF}}$ measurements. The reason of this behaviour is that the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ reconstruction tends to be in the direction of the mis-measured jet from the ID system. Hence, the projection of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$, through the computation of the $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$, is highly favoured. In light of the results, the $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ is used, in combination with the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$, to further suppress $Z \rightarrow \ell\ell$ events in the $H \rightarrow WW^{(*)} \rightarrow \ell\nu\ell\nu$ search.

4.5 $\not\!\!E_T$ Performance

Up to this section, the different approaches for reconstructing the $\not\!\!E_T$ in ATLAS have been defined. In addition, the results from basic data and MC comparisons discussed in Section 4.4 show characteristics and features for each $\not\!\!E_T$ measurement depending on pile-up environment and jet activity in the final state. This section summarises the main results on the performance of the different $\not\!\!E_T$ measurements in terms of pile-up dependence, resolution, scale, linearity and direction.

4.5.1 $\not\!\!E_T$ Dependence with Pile-Up

To evaluate the correlation of the $\not\!\!E_T$ with respect to pile-up, the mean and RMS for each of the different $\not\!\!E_T$ measurements are obtained in bins of the average number of interactions per bunch-crossing. The results from $Z \rightarrow \ell\ell$ events with different jet multiplicities in data and simulation are shown in Fig. 4.5. The most important feature from these distributions is that the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ reconstruction is much more robust against pile-up than the calorimeter-based measurements. This is expected since only tracks associated to the PV are included in the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ computation. The stability of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ reconstruction regardless of high pile-up conditions motivates many analysis searches for using this measurement to isolate the signal. Conversely, extra energetic depositions coming from pile-up interactions are entering in $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ computations, so these definitions have a strong correlation with pile-up as observed in high dependence Fig. 4.5. When the jet activity increases

the performance of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ worsens dramatically due to the ID limitations. Besides the results still show a very stable $E_{\mathrm{T}}^{\mathrm{miss,track}}$ measurement, flat and almost independent of the pile-up conditions, the mean and RMS values are significantly larger than the ones obtained with the calorimeter-based measurements. These results agree with the features observed in Section 4.4 and confirm the weakness of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ definition for considering high-p_{T} neutral particles.

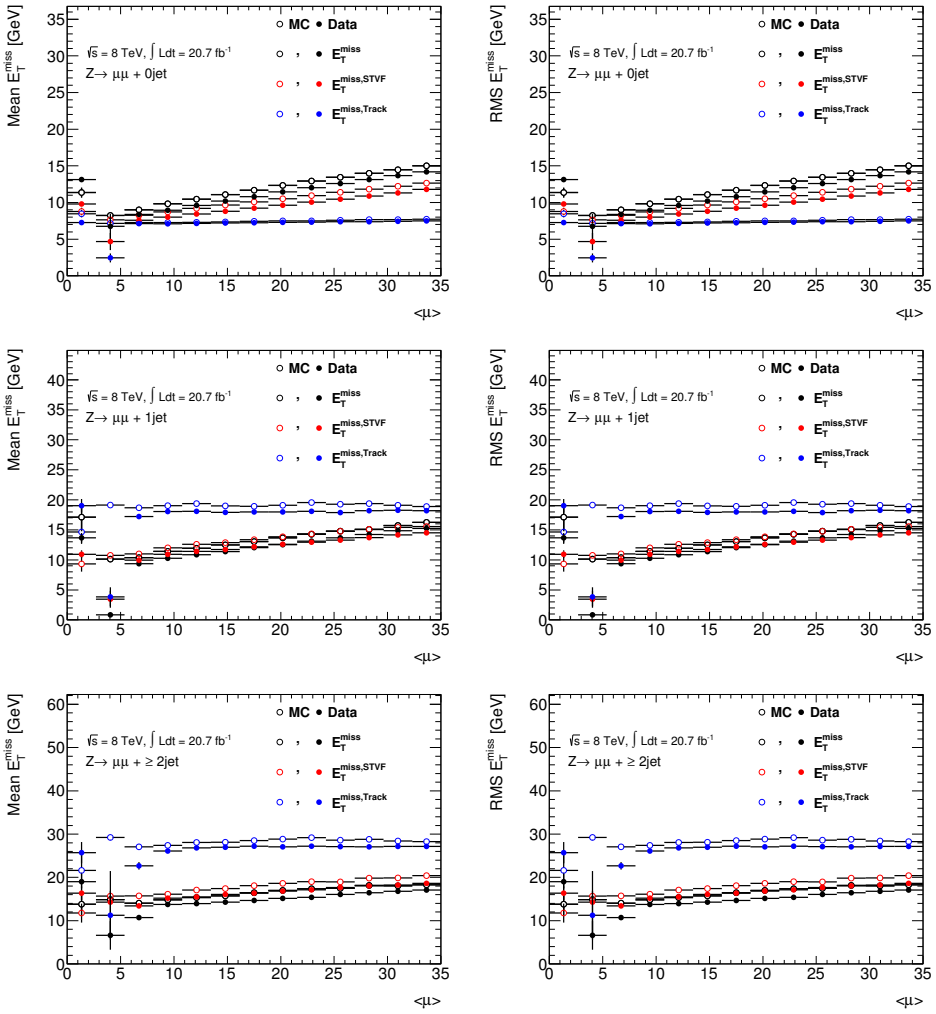

FIGURE 4.5: Mean (left) and RMS (right) in GeV of the different $\not{E}_T$ reconstructions against the average of interactions per bunch-crossing. $Z \rightarrow \mu\mu$ events in data (closed circles) and simulation (open circles) with 0, 1 and at least 2 jets are shown on top, middle and bottom, respectively.

$E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ measurements show a high correlation with pile-up, as expected. The observed trends are significantly higher than in the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ case but the slope remains constant independently of the number of reconstructed jets in the event. This effect points to a relatively lower contribution of the soft-term and manifests the difference between calorimeter-based and track-based approaches when the jet-term becomes significant. The pile-up suppression technique applied through the STVF factor for computing the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ measurement, improves the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution in events without jets. In events with jets, both calorimeter-based definitions provide similar correlation with pile-up.

Pile-up effects on $\not\!\!E_T$ can also be evaluated through the $\sum E_{\mathrm{T}}$ measurement. Extra energetic depositions in the calorimeters are inputs to the $\sum E_{\mathrm{T}}$, as defined in Section 4.2.4.1, hence the $\not\!\!E_T$ stability with pile-up can be investigated in bins of $\sum E_{\mathrm{T}}$ as well. Figure 4.6 presents the mean of the different $\not\!\!E_T$ definitions as a function of $\sum E_{\mathrm{T}}$ for data and MC in $Z \to \mu\mu$ and $Z \to ee$ events separately.

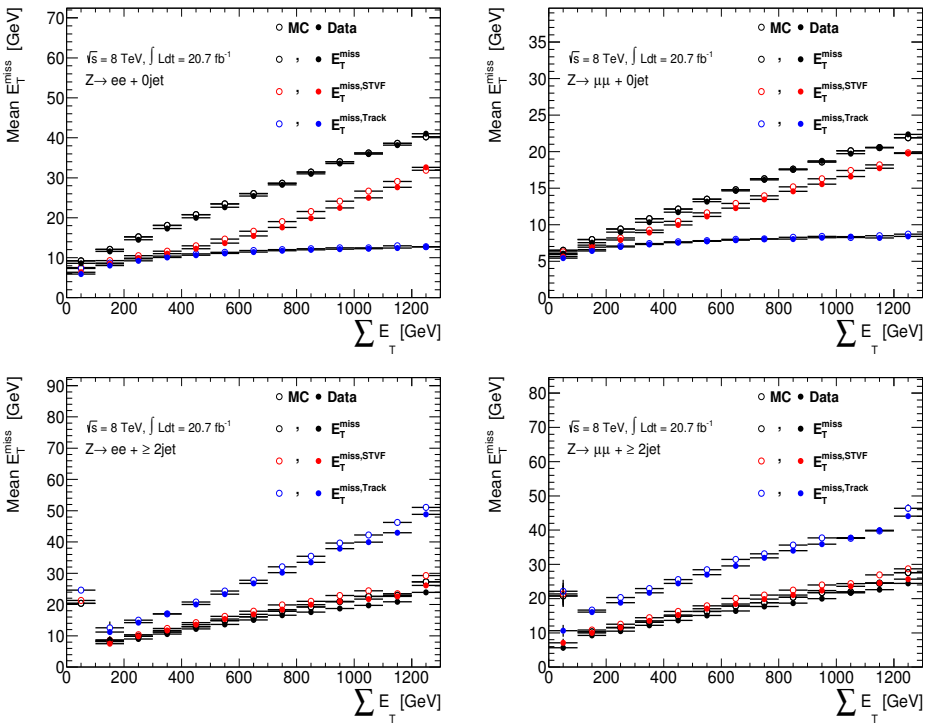

FIGURE 4.6: Mean of the different $\not\!\!E_T$ quantities in bins of $\sum E_{\mathrm{T}}$ in $Z \to ee$ and $Z \to \mu\mu$ events on the left and right, respectively. 2012 ATLAS data (full circle) and simulation from MC (open circle) are represented. Top distributions have applied a jet veto and bottom plots are obtained selecting events with at least two jets.

Data and MC show reasonable agreement for $\sum E_{\mathrm{T}} > 100\,\mathrm{GeV}$ in all three $\not\!\!E_T$ measurements and for both ee and $\mu\mu$ final states. In $Z \to ee$ events, the trends are higher than in the $\mu\mu$ final state for all $\not\!\!E_T$ measurements. This is due primarily to the Bremsstrahlung of electrons, as well as to the better resolution of muons in the ATLAS detector. Results in events without jets show that the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ reconstruction is very stable with respect to the $\sum E_{\mathrm{T}}$ measurement, while the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ present higher dependence. The application of the STVF factor to the soft-term in $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ improves the performance of the nominal $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement. However, in the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ case the difference between data and MC increases with respect to $E_{\mathrm{T}}^{\mathrm{miss}}$. This is due to the poor modelling of the STVF factor, as discussed in Section 4.4. When considering high jet activity, the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ looses its performance and becomes the most correlated measurement with the $\sum E_{\mathrm{T}}$ of the event. High $\sum E_{\mathrm{T}}$ events tend to have more high-p_{T} neutral particles which are not accounted for in $E_{\mathrm{T}}^{\mathrm{miss,track}}$ computation. Finally, both calorimeter-based definitions provide similar dependence with $\sum E_{\mathrm{T}}$ in events with jets. This points to a lower relative contribution of the soft-term, and so of the STVF correction effects, due to the enhanced jet-term in such final states.

4.5.2 $\not\!\!E_T$ Resolution

The $\not\!\!E_T$ resolution is an important indicator of the $\not\!\!E_T$ performance. This is typically presented as the RMS from the distributions of the directional components of each $\not\!\!E_T$ measurement in bins of $< \mu >$ and $\sum E_{\mathrm{T}}$. The RMS width of the x- and y-components are added in quadrature to obtain the $\not\!\!E_T$ resolution.

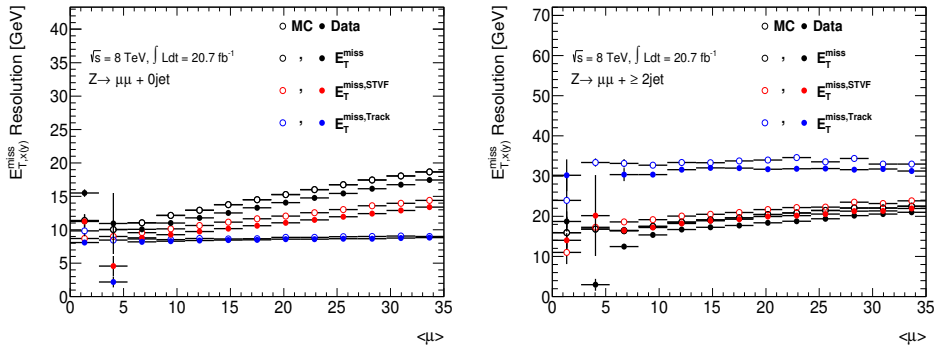

FIGURE 4.7: $\not\!\!E_T$ resolution from the different $\not\!\!E_T$ directional components in bins of the average of interactions per bunch-crossing. 2012 ATLAS data (full circle) and simulation from MC (open circle) are represented. $Z \to \mu\mu$ simulated events with 0 and at least 2 jets are shown on left and right, top, respectively.

Figure. 4.7 and Fig. 4.8 show the $\not\!\!E_T$ resolution in terms of the directional components and the RMS of the $\not\!\!E_T$ distributions in bins of $< \mu >$ and $\sum E_{\mathrm{T}}$, respectively. 2012 ATLAS data and simulated events consistent with $Z \to \ell\ell$ process are used. These results agree with the investigations presented in Section 4.5.1. Same trends and features are found for all three $\not\!\!E_T$ measurements. The use of track information greatly improves the resolution over the calorimeter-based measurements in events without jets since the ID system provides a powerful discriminator against pile-up. For events with jets, the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ performances are superior to the $E_{\mathrm{T}}^{\mathrm{miss,track}}$. The $E_{\mathrm{T}}^{\mathrm{miss,track}}$ resolution worsens because of missing high-p_{T} neutrals, specially coming from jets.

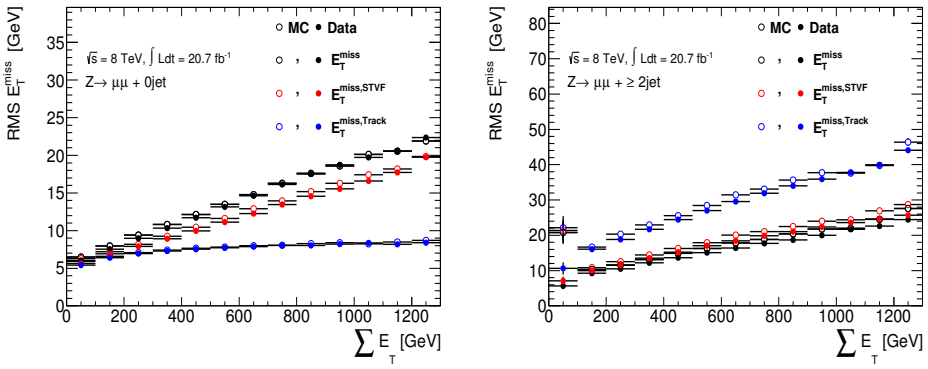

FIGURE 4.8: $\not\!\!E_T$ resolution from the different $\not\!\!E_T$ directional components in bins of the average of interactions per bunch-crossing. 2012 ATLAS data (full circle) and simulation from MC (open circle) are represented. $Z \to \mu\mu$ simulated events with 0 and at least 2 jets are shown on left and right, respectively.

Events with genuine $\not\!\!E_T$ provide an important topology for validating the different $\not\!\!E_T$ measurements. The expected presence of neutrinos in the final state means that the reconstructed $\not\!\!E_T$ is not just a measurement of the fluctuations around zero, as in the $Z \to \ell\ell$ process. In topologies with genuine $\not\!\!E_T$ the resolution can only be studied in MC simulation events since the information of the $E_{\mathrm{T}}^{\mathrm{miss,\,True}}$ is needed. In light of this, the resolution of the two $\not\!\!E_T$ components for each definition is estimated from the RMS width of

$$E_{\mathrm{T},\,i}^{\mathrm{miss}} \; - \; E_{\mathrm{T},\,i}^{\mathrm{miss,\;True}} \,, \tag{4.15}$$

where i represents the directional $\not\!\!E_T$ components.

The resolution of the different $\not\!\!E_T$ measurements using $H \to WW^{(*)} \to \ell\nu\ell\nu$ simulated events, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, are shown in Fig. 4.9. These results combine

all lepton flavour final states and the two most contributive production mechanisms: ggF and VBF in order to increase the statistics. Same trends and conclusions as in the results obtained with $Z \to \ell\ell$ events presented above are observed now in simulated events with genuine $\not\!\!E_T$ measurement.

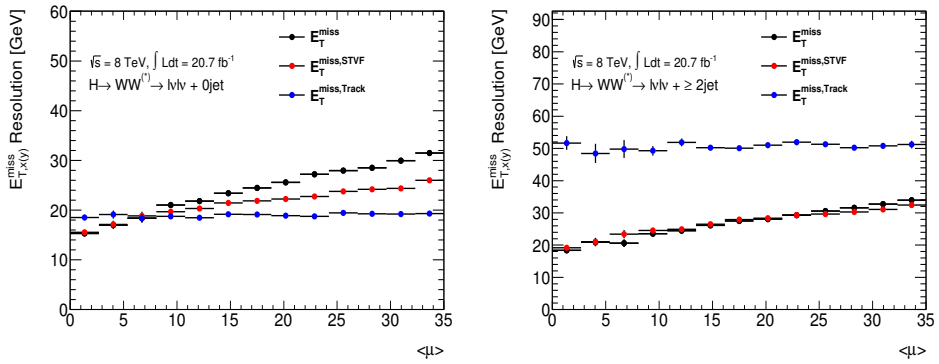

FIGURE 4.9: $\not\!\!E_T$ resolution from Eq. 4.15 for the different $\not\!\!E_T$ measurements in bins of the average of interactions per bunch-crossing. $H \to WW^{(*)} \to \ell\nu\ell\nu$ simulated events, with $\ell = e, \mu$ and $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, are shown combining all lepton flavour final states and the two most contributive production mechanisms: ggF and VBF. Events with 0 and at least 2 jets are represented on the left and right, respectively.

4.5.3 $\not\!\!E_T$ Scale

The $\not\!\!E_T$ scale provides a good global indicator of the $\not\!\!E_T$ resolution performance. Investigations on the $\not\!\!E_T$ scale allow to check the control over fluctuations in energy and momentum measurements due to pile-up interactions and the choice of objects used in the $\not\!\!E_T$ computations. The $\not\!\!E_T$ response relies on the scale of the $\not\!\!E_T$ balance using better measured and calibrated quantities like the transverse momentum of the Z boson (p_{T}^Z). A bias in the scale of the $\not\!\!E_T$ reconstruction implies a systematic under or overestimated of some portion of the event transverse momentum, e.g. due to detector calibration.

In this section, the $\not\!\!E_T$ scale is obtained using the $Z \to \ell\ell$ topology. In such kind of events, the transverse momentum of the Z boson defines an axis in the transverse plane of the ATLAS detector. This is used to check that the $\not\!\!E_T$ along this axis balances the p_{T}^Z. Hence, in events from $Z \to \ell\ell$ decays one can define an axis in the transverse plane such that the component of $\not\!\!E_T$ along this axis is sensitive to detector resolution and biases [59].

The normalised direction of the axis of the Z boson (A_Z) is defined by the reconstructed momenta of the leptons,

$$A_Z = \frac{p_{\mathrm{T}}^{\ell^+} + p_{\mathrm{T}}^{\ell^-}}{|\, p_{\mathrm{T}}^{\ell^+} + p_{\mathrm{T}}^{\ell^-} \,|}, \tag{4.16}$$

where p_{T}^{ℓ} are the vector transverse momenta of the lepton and anti-lepton. The direction of A_Z thus reconstructs the transverse direction of motion of the Z boson. The mean value of the projection of $\not\!\!E_T$ onto the longitudinal axis ($< E_{\mathrm{T}}^{\mathrm{miss}} \cdot A_Z >$) measures the $\not\!\!E_T$ scale, as this axis is sensitive to the balance between the leptons and the hadronic recoil.

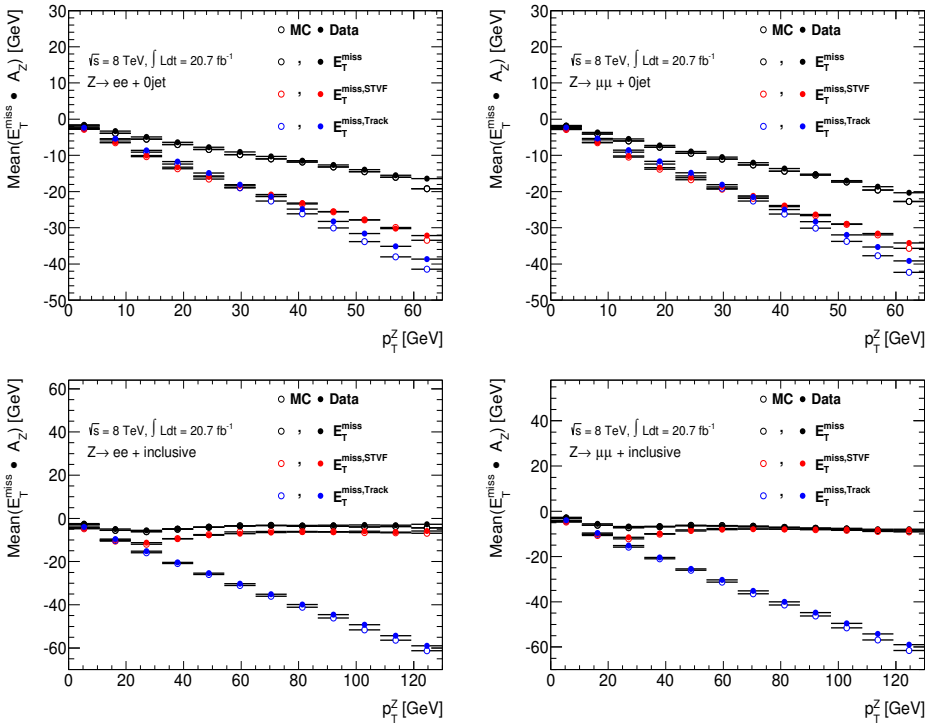

FIGURE 4.10: Mean of $< E_T \cdot p_{\mathrm{T}}^Z >$ for the different $\not\!\!E_T$ measurements in bins of p_{T}^Z for 8 TeV ATLAS data (closed markers) and MC simulation (open markers). The results are separated for $Z \to ee$ and $Z \to \mu\mu$ final states on the the left and right, respectively. Distributions on top have a jet veto applied and on the bottom corresponds to the inclusive jet case.

Figure 4.10 shows the $\not\!\!E_T$ scale in bins of p_{T}^Z in $Z \to \ell\ell$ events with different number of jets in the final state. The agreement between data and simulation is very good for low p_{T}^Z range, which benefits of higher statistics. Moreover, ee and $\mu\mu$

final states produce compatible results. The $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement provides the lowest scale deviation in all cases, while presence of neutral particles clearly degrades the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ scale measurements. The correction introduced by the STVF factor in the soft-term modifies the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ direction with respect to the nominal $E_{\mathrm{T}}^{\mathrm{miss}}$. Since the STVF factor ignores contributions from neutral particles, the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ direction is biased and so its scale. This effect can be observed in events without jets in the final state. In events with jets, the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ scale improves to the $E_{\mathrm{T}}^{\mathrm{miss}}$ level since the STVF soft-term is less contributive. The ID limited detection is expected to specially affect the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ scale measurement. The $E_{\mathrm{T}}^{\mathrm{miss,track}}$ scale shows the highest deviation from zero for all jet cases, just performing slightly better than the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ up to $p_{\mathrm{T}}^{Z} = 30\,\mathrm{GeV}$ in events without jets.

4.5.4 $\not\!\!E_T$ Linearity

Another useful magnitude in events with genuine $\not\!\!E_T$ is the linearity. This measures the consistency between the magnitude of the reconstructed and the simulated $\not\!\!E_T$ measurement. Note that the latter is invaluable for data samples with hard neutrinos, so the linearity will be only computed for simulated events. The linearity of the $\not\!\!E_T$ is defined as the mean value of the following ratio,

$$\text{Linearity} = \frac{E_{\mathrm{T}}^{\mathrm{miss}} - E_{\mathrm{T}}^{\mathrm{miss,\,True}}}{E_{\mathrm{T}}^{\mathrm{miss,\,True}}}\,. \tag{4.17}$$

The linearity magnitude is expected to be zero if the $\not\!\!E_T$ is reconstructed at the correct scale.

In Fig. 4.11, the linearity for $H \to WW^{(*)} \to \ell\nu\ell\nu$ simulated events is shown in bins of the $E_{\mathrm{T}}^{\mathrm{miss,\,True}}$. Any threshold on the reconstructed $\not\!\!E_T$ definitions is applied for this plot, however, a bias for low $E_{\mathrm{T}}^{\mathrm{miss,\,True}}$ values is observed. The relative difference with respect to $E_{\mathrm{T}}^{\mathrm{miss,\,True}}$ is positive when this quantity is small. This effect extends up to 50 GeV and it is related to the finite resolution of the / E_T measurement as well as to the fact that the reconstructed $\not\!\!E_T$ is positive by definition.

The results in events with jets show that when $E_{\mathrm{T}}^{\mathrm{miss,\,True}} > 50\,\mathrm{GeV}$, the linearity is better than 5% for $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ reconstructions. The $E_{\mathrm{T}}^{\mathrm{miss,track}}$ presents a stronger negative linearity because of the lack of neutral particles. This shows a much better behaviour in agreement with the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ linearity for events without high-p_{T} jets, as shown in the upper distribution.

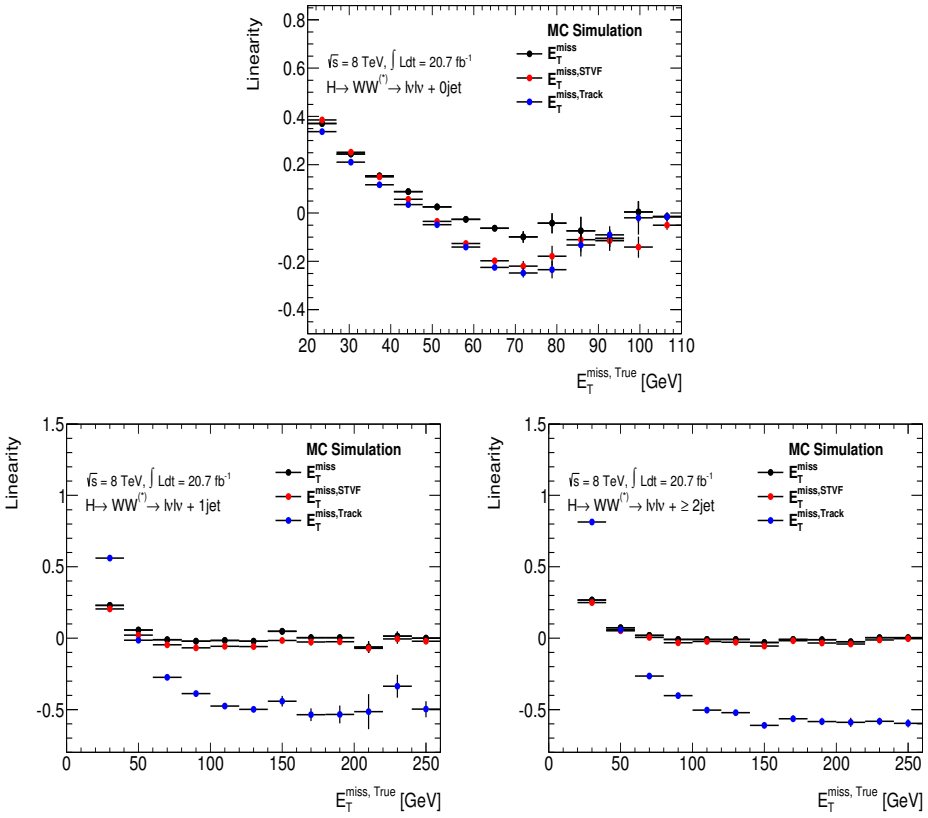

FIGURE 4.11: Linearity using the different $\not\!\!E_T$ measurements in bins of $E_{\mathrm{T}}^{\mathrm{miss,\,True}}$ for $H\to WW^{(*)}\to \ell\nu\ell\nu$ MC simulated events with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. Results are separated for 0 jets on top, 1 jet on bottom-left and at least 2 jets on bottom-right.

In events without jets, all $\not\!\!E_T$ definitions produce negative linearity values when $E_{\mathrm{T}}^{\mathrm{miss,\,True}} > 50\,\mathrm{GeV}$. The negative trend improves from $E_{\mathrm{T}}^{\mathrm{miss,\,True}} > 70\,\mathrm{GeV}$. This points to a defect on the calibration of jets in the limit of the selection threshold. Around 3% of the events with $50 < E_{\mathrm{T}}^{\mathrm{miss,\,True}} < 70\,\mathrm{GeV}$ have a truth jet with $p_{\mathrm{T}}^{\mathrm{True}} > 20\,\mathrm{GeV}$ which was not considered in the $\not\!\!E_T$ reconstructions due to the jet selection quality requirements. The p_{T} of this missing jet is underestimated in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation and more so in the other definitions. As the $E_{\mathrm{T}}^{\mathrm{miss,\,True}}$ increases in the denominator of the linearity given in Eq. 4.17, the values pull closer to zero.

4.5.5 $\not\!\!E_T$ Direction

The direction of the $\not\!\!E_T$ is an extremely important magnitude as well. The performance of the $\not\!\!E_T$ direction is evaluated using physics processes with genuine $\not\!\!E_T$ from simulation. This allows to check the resolution of the $\not\!\!E_T$ direction as a function of the $E_{\mathrm{T}}^{\mathrm{miss,\ True}}$. The difference in the $\not\!\!E_T$ direction from the generator value is computed by taking the difference in the azimuthal angle of the reconstructed $\not\!\!E_T$ direction ($\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, E_{\mathrm{T}}^{\mathrm{miss,\ True}})$). The RMS of this distribution is taken as the resolution of the direction of each $\not\!\!E_T$ definition.

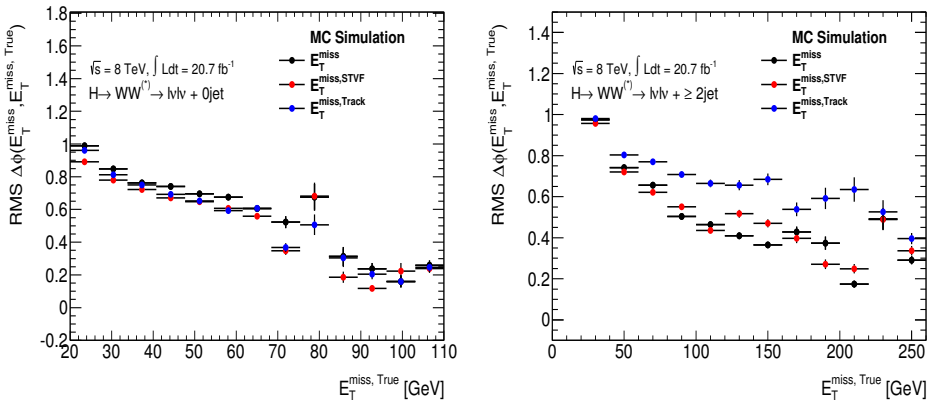

FIGURE 4.12: Resolution on $\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, E_{\mathrm{T}}^{\mathrm{miss,\ True}})$ for $H\to WW^{(*)}\to \ell\nu\ell\nu$ simulated events without jets (left) and with at least two jets (right).

Figure 4.12 shows the resolution of the $\not\!\!E_T$ direction in bins of $E_{\mathrm{T}}^{\mathrm{miss,\ True}}$ using $H\to WW^{(*)}\to \ell\nu\ell\nu$ simulated events from MC with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. In general, as the $E_{\mathrm{T}}^{\mathrm{miss,\ True}}$ value increases the angular resolution improves for all three $\not\!\!E_T$ reconstructions. Events without jets present very similar angular resolution for all $\not\!\!E_T$ measurements. The resolution of the $\not\!\!E_T$ direction reaches a plateau when $E_{\mathrm{T}}^{\mathrm{miss,\ True}} > 80\,\mathrm{GeV}$. In events with jet activity, the presence of jets plays an important role and makes the distinction between the calorimeter-based and track-based $\not\!\!E_T$ approaches. In this case, the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ presents a much worse directional resolution than the calorimeter-based measurements. Missing neutral particles severally deteriorates the angular resolution of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$.

4.6 Systematic Uncertainty on $\not\!\!E_T$

The $\not\!\!E_T$ measurements, as defined in Section 4.2, are defined by the sum of several terms associated to different reconstructed objects. The uncertainty of each individual term is evaluated given the knowledge of the reconstructed objects that are then used to build it. The uncertainties provided for the electrons, muons, jets, taus, and photons are propagated into their respective term. The overall systematic uncertainty on the $\not\!\!E_T$ measurement is calculated by combining the uncertainties on each object. This section therefore focuses on the derivation of systematic uncertainties for the different $\not\!\!E_T$ soft-terms.

Another important $\not\!\!E_T$ uncertainty source comes from soft energetic contributions, i.e., those entering in the computation of the $\not\!\!E_T$ measurements but not associated to any high-p_{T} object. Note that these extra contributions will explicitly enter in the soft-term of the calorimeter-based definitions as shown in Eq. 4.1 and Eq. 4.8. For the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ case, the corresponding soft-term can be interpreted as the contribution of those tracks considered in the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ computation, which are not associated to any well reconstructed object in the event. The $E_{\mathrm{T}}^{\mathrm{miss,track}}$ soft-term is not specified in Eq. 4.11 but this can easily defined as the magnitude of the vector difference between the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ and the transverse momentum from all high-p_{T} particles,

$$E_{x(y)}^{\text{track, SoftTerm}} = E_{x(y)}^{\mathrm{miss,track}} - \sum_{e} p_{x(y)}^{e} - \sum_{\mu} p_{x(y)}^{\mu} - \sum_{\gamma} p_{x(y)}^{\gamma} - \sum_{\tau} p_{x(y)}^{\tau} - \sum_{\text{jet}} p_{x(y)}^{\text{jet}} \,. \tag{4.18}$$

In the following, these extra contributions not associated to the high-p_{T} objects will be referred to as soft-term for all $\not\!\!E_T$ approaches, taking into account the distinction between the calorimeter-based and track-based origins.

Two methods for evaluating systematic uncertainties on scale and resolution of the different $\not\!\!E_T$ soft-terms have been developed in ATLAS and are extensively documented in Ref. [59]. The first method considers the agreement between data and simulation in $Z \to \mu\mu$ events without jets. This topology is used to estimate the systematics on the remaining contributions of the $\not\!\!E_T$ measurement after subtracting the high-p_{T} and well reconstructed objects of the event. In light of this, $Z \to \mu\mu$ without jets events are optimal for extracting scale systematic uncertainty of the $\not\!\!E_T$ soft-terms as only muons are expected to contribute to the $\not\!\!E_T$ measurement. Any extra contribution will affect the scale and resolution of the $\not\!\!E_T$ soft-term. The projection of the different $\not\!\!E_T$ soft-terms onto the transverse direction of the Z boson momentum allows to test any bias in the scale. The systematic uncertainty

on the $\not\!\!E_T$ scale is estimated by comparing data and MC agreement of the projected soft-terms in bins of $\sum E_{\mathrm{T}}$. Similarly to the results obtained in Fig. 4.10, the size of the average deviation between data and the simulation is taken as the scale systematic uncertainty. The size of the average deviation from unity is about 8% for the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement. This value is typically taken as a flat uncertainty on the absolute scale. The systematic uncertainty on the resolution is estimated by evaluating the discrepancies between data and MC in the x- and y-components of each $\not\!\!E_T$ measurement. The systematic uncertainties on the resolution are obtained in bins of $\sum E_{\mathrm{T}}$, similarly to the results discussed in Fig. 4.7. The uncertainty on the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft-term resolution is $\sim$3%.

The second approach for evaluating the systematic uncertainties on the $\not\!\!E_T$ exploits the balance between the soft-term and the total transverse momentum of all particles that may have been originated by a certain process. The latter is known as $p_{\mathrm{T}}^{\mathrm{hard}}$ and it is defined as the vector sum of the transverse momenta of the hard objects in the event. The x- and y-components of the $p_{\mathrm{T}}^{\mathrm{hard}}$ are computed as,

$$p_{x(y)}^{\mathrm{hard}} = \sum_{e} p_{x(y)}^{e} + \sum_{\mu} p_{x(y)}^{\mu} + \sum_{\gamma} p_{x(y)}^{\gamma} + \sum_{\tau} p_{x(y)}^{\tau} + \sum_{\mathrm{jet}} p_{x(y)}^{\mathrm{jet}} + \sum_{\nu} p_{x(y)}^{\nu}\,, \quad (4.19)$$

The $p_{\mathrm{T}}^{\mathrm{hard}}$ direction defines an axis in the transverse plane of the ATLAS detector in which the soft-term is expected to balance the $p_{\mathrm{T}}^{\mathrm{hard}}$ magnitude. $p_{\mathrm{T}}^{\mathrm{hard}}$ can be regarded as the true value of the $\not\!\!E_T$ soft-term and allows to evaluate $\not\!\!E_T$ scale and resolution systematic uncertainties in events with jets. Equation 4.19 only can be computed in MC samples since the $\sum \vec{p}^{\,\nu}_{\mathrm{T}}$ contribution is unknown in data. However, in $Z \to \ell\ell$ processes the component from non-interacting particles can be safety neglected. In this case, the $p_{\mathrm{T}}^{\mathrm{hard}}$ magnitude can be investigated in both data and simulated events. To evaluate the mean and resolution of the $\not\!\!E_T$ soft-term, this is decomposed along the $p_{\mathrm{T}}^{\mathrm{hard}}$ direction and along the orthogonal direction, referred to as longitudinal and perpendicular directions, respectively. The longitudinal component is sensitive to scale and resolution differences between the data and simulation since the soft-term should balance the $p_{\mathrm{T}}^{\mathrm{hard}}$. The perpendicular component is only sensitive to differences in resolution since the mean is very close to 0 in data and MC because the hadronic recoil only affects the parallel component. The discrepancies between data and simulation are considered as the systematic uncertainties for the scale and resolution. The average uncertainties are about 3% for both longitudinal and perpendicular directions. All $\not\!\!E_T$ reconstruction approaches can be applied with small systematic uncertainties on the full measurement.

4.7 Conclusions

The increasing number of pile-up interactions during 2012 data-taking has required several investigations into the performance of the $\not\!\!E_T$ reconstruction on ATLAS. The features of several approaches based on different object information to build the unbalanced transverse momentum in ATLAS events have been explored in this chapter. The pure calorimeter-based $E_{\mathrm{T}}^{\mathrm{miss}}$ definition highly depends on the number of pile-up interactions which degrades its performance in term of resolution and tails. Two pile-up suppressed $\not\!\!E_T$ reconstructions are developed in order to deal with pile-up contributions: $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track}}$. However, these measurements come with new features. In the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ reconstruction, the MC modelling of the STVF factor produces an under-calibrated soft-term which results in discrepancies between data and simulation of about 20% in events without jets. For events with jets, the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ perform similarly because the dominant component is the jet-term. For such event topologies, the contribution of the soft-term is relatively lower and so are its effects. Lastly, an alternative measurement based on tracks coming from the hardest reconstructed vertex is also studied. The $E_{\mathrm{T}}^{\mathrm{miss,track}}$ measurement is very stable with respect to pile-up interactions, but limited ID coverage and missing high-p_{T} neutral particles lead to large degradation in the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ linearity and scale, specially in events with high jet activity.

Chapter 5

$H \to WW^{(*)} \to \ell\nu\ell\nu$ Analysis

5.1 Introduction

The SM of particle physics has been tested by many experiments over the last four decades and it has been shown to successfully describe high energy phenomena. However, the mechanism that breaks electroweak symmetry in the SM was not confirmed until 2012. The BEH mechanism of electroweak symmetry breaking was published in 1964 [11–13]. This provides masses to the SM particles and predicts the existence of a new elementary particle known as the Higgs boson, as addressed in Chapter 1.

Early in the 21st century, direct searches at LEP and Tevatron excluded at 95% confidence level (CL) a Higgs boson with a mass below 114.4 GeV [17] and in the regions $147 < m_{\mathrm{H}} < 180\,\mathrm{GeV}$ and $100 < m_{\mathrm{H}} < 103\,\mathrm{GeV}$ [65], respectively. In 2011, various search results using proton-proton collision data at the LHC with a center-of-mass energy of 7 TeV and corresponding to an integrated luminosity of $4.7\,\mathrm{fb}^{-1}$ were reported by the ATLAS experiment. No significant excess of events over the expected background was observed and the results allowed to exclude at 95% CL the mass ranges $112.9 - 115.5\,\mathrm{GeV}$, $131 - 238\,\mathrm{GeV}$, and $251 - 466\,\mathrm{GeV}$ [66]. By the summer of 2012 the ATLAS detector had collected the first $5.8 fb^{-1}$ of integrated luminosity at $\sqrt{s} = 8\,\mathrm{TeV}$.

On July 4th, 2012, the ATLAS and CMS Collaborations reported the observation of a new particle compatible with the SM Higgs boson, with a mass of approximately $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. These results were achieved by combining results from several final states of the Higgs decay. The CMS Collaboration combined five decay modes observing an excess with a local significance of 5.0 σ [67]. The ATLAS Collaboration combined results from $H \to \gamma\gamma$, $H \to ZZ^{(*)} \to 4\ell$ and $H \to WW^{(*)} \to \ell\nu\ell\nu$ analyses obtaining a local significance of the observed excess corresponding to $5.9\,\sigma$ by adding the first $\sim 5.8\,\mathrm{fb}^{-1}$ at 8 TeV to the previous 7 TeV data [68]. In 2013, results from several Higgs boson decay searches were upgraded using the total data samples recorded by the ATLAS detector in the complete Run I data taking period, corresponding to an integrated luminosity of about $25\,\mathrm{fb}^{-1}$. The results confirmed the existence of the Higgs boson particle and stated the evidence of the Higgs boson production through VBF. Results on Higgs boson coupling to fermions show that all measurements are consistent with expectations for the SM Higgs boson [69].

This chapter presents the results published in Ref. [69] in the search of the Higgs boson, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, into a pair of W bosons using 7 TeV and 8 TeV data collected by the ATLAS detector. This decay mode was one of the three used in the Higgs boson discovery. The analysis considers contributions from the $H \to WW^{(*)} \to \ell\nu\ell\nu$ (with $\ell = e, \mu$) decay mode where the Higgs boson is produced through ggF and VBF production mechanisms. The $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis has been performed using the complete data sample collected with the ATLAS detector at the LHC during 2012 and 2011 at $\sqrt{s} = 8\,\mathrm{TeV}$ and $\sqrt{s} = 7\,\mathrm{TeV}$, respectively and corresponding to a total integrated luminosity of $20.7\,\mathrm{fb}^{-1}$ and $4.7\,\mathrm{fb}^{-1}$ [70]. Section 5.2 presents an overview of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis. Section 5.3 and Section 5.4 describe Monte Carlo and data samples, respectively. Section 5.6 summarises selection of reconstructed objects in the final state, as leptons and jets. Section 5.7 describes the analysis strategy, introducing background sources and key variables for enhancing the sensitivity of the search. Section 5.9 details the estimation of the backgrounds yields. Section 5.10 presents the systematic uncertainties. The focus of the chapter is on the analysis of the 8 TeV data, but the analysis of the 7 TeV data is also briefly described in Section 5.8. Finally, Section 5.12 presents the evidence for the Higgs-like boson results and Section 5.13 states the conclusions.

5.2 Analysis Overview

The $H \to WW^{(*)} \to \ell\nu\ell\nu$ decay mode is particularly sensitive in the mass range $120 < m_{\mathrm{H}} < 200\,\mathrm{GeV}$. The Higgs boson decay branching ratio to WW falls off with decreasing m_{H} below $m_{\mathrm{H}} = 2M_{\mathrm{W}}$ but it is still just over 20% at $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, as can be seen on the left plot in Fig. 5.1. The Higgs boson production total cross section increases up to 30% at $\sqrt{s} = 8\,\mathrm{TeV}$ compared with $\sqrt{s} = 7\,\mathrm{TeV}$ [25] as is shown, on the right, in Fig. 5.1. The benefit of running at higher center-of-mass energy comes at the price of having multiple proton-proton collisions per bunch crossing, the so-called pile-up, as was discussed in Chapter 4. This affects several aspects of the analysis; as for example, the higher Drell-Yan background contamination due to an increased rate of fake $\not\!\!E_T$ measurements.

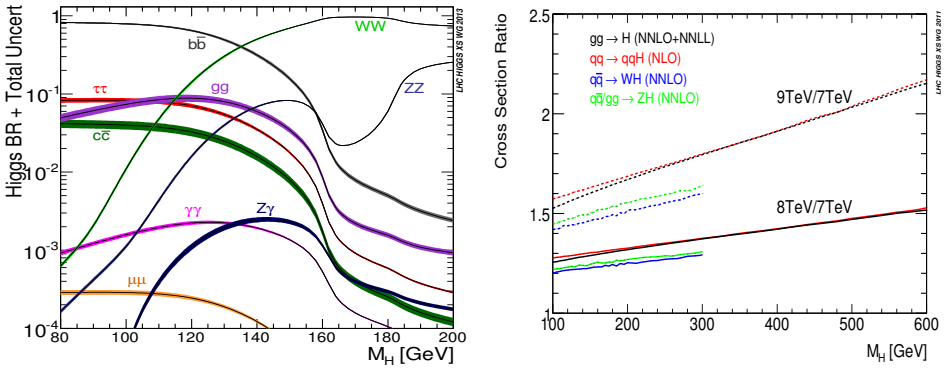

FIGURE 5.1: Higgs boson decay branching ratios (left) and production cross section ratios between $\sqrt{s} = 7, 8\,\mathrm{TeV}$ in solid lines; $\sqrt{s} = 7, 9\,\mathrm{TeV}$ dashed lines (right).

There are four possible final states in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ (with $\ell = e, \mu$) search: $e\nu e\nu$, $\mu\nu\mu\nu$, $e\nu\mu\nu$, and $\mu\nu e\nu$. Note that the first symbol denotes the charged lepton with the higher p_{T}. The analysis is performed in the two possible combinations depending on the flavour of the final leptons. This results in the same-flavour, $e\nu e\nu$ and $\mu\nu\mu\nu$; and different-flavour, $e\nu\mu\nu$ and $\mu\nu e\nu$, categories. This strategy allows to optimise the selection of events with a favourable signal-to-background ratio due to the different background processes that dominate in each case. The notation $e\mu{+}\mu e$ and $ee{+}\mu\mu$ will be used in the following to denote the different-flavour and same-flavour final states, respectively.

The search is further divided depending on whether the events contain zero, one or at least two jets [26]. These categories are referred to as $H{+}0j$, $H{+}1j$, and $H{+}2j$ analyses, respectively. The $H{+}0j$ and $H{+}1j$ categories are suitable to select

Higgs boson events produced via the dominant ggF production mode. The H+0j category relies on Higgs events produced through the ggF with no jets in the final state. However, an important fraction of the candidate events will be accompanied by a jet produced from radiation of a quark or gluon off the initial state gluons. These events are recovered in the H+1j analysis. Higgs events produced through the VBF production mechanism are characterised by at least two high-p_{T} and widely separated jets coming from the quarks emitting the boson. These are selected by requiring at least two jets in the final state in the H+2j analysis. The H+0j category with $e\mu$+μe final states provides most of the sensitivity of the Higgs boson decaying into a pair of W boson search. Nevertheless, the analysis expands the scope of the production mode and the decay channels combining in a global fit a transverse mass variable to discriminate the presence of the Higgs boson signal.

The analysis procedure is determined by blinding the kinematic region where a signal might be expected. Data events passing the selection, designed to isolate a signal from the Higgs boson with a mass between $110 - 140\,\mathrm{GeV}$ are excluded during analysis developments. Optimisation investigations and control of the quality of processes predictions are performed without looking at the Higgs signal phase space. Once the agreement between observed data and MC simulation is judged satisfactory, the signal region is analysed. The blinding criteria provide robustness and confidence to the analysis selection as they prevent against possible biases based on the desire to exclude or see an excess.

5.3 Monte Carlo Samples

The signal contributions considered in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis include ggF and VBF production mechanisms as well as associated production of the Higgs boson with a W or Z boson. The $t\bar{t}H$ production mechanism is neglected due to its smaller cross section. For the decay of the Higgs boson, only the fully leptonic modes are considered, $H\to WW^{(*)}\to \ell\nu\ell\nu$ (with $\ell = e,\,\mu$), which include the contributions from leptonic τ decays. The branching fraction for the decay as a function of m_{H} is calculated using PROPHECY4F [71, 72] with HDECAY used to calculate the total width [73].

The MC generators used to model signal and background processes are listed in Tab. 5.1 and Tab. 5.2, respectively. The ggF signal cross section is computed at next-to-next-to-leading order (NNLO) in QCD [23, 74–78] for the ggF production mechanism using the MSTW2008 PDF set [20]. Next-to-leading order (NLO)

electroweak (EW) corrections are also applied [79, 80], as well as QCD soft-gluon resummations up to next-to-next-to-leading log (NNLL) [81]. These calculations are detailed in Refs. [82–84], and they assume factorisation between the QCD and EW corrections. The VBF signal cross section is computed with approximate NNLO QCD corrections [85] and full NLO QCD and EW corrections [86–88]. The cross sections of the associated production mechanism are calculated up to NNLO QCD corrections [89, 90] and NLO EW corrections [91].

TABLE 5.1: MC generators used to model the signal and the corresponding product of the cross section (σ) and branching fraction ($\mathcal{B}$) at $\sqrt{s} = 8$ TeV.

Signal	MC generator	$\sigma \cdot \mathcal{B}$ (pb)
ggF	POWHEG [92]+PYTHIA8 [93]	0.44
VBF	POWHEG+PYTHIA8	0.035
VH	PYTHIA8	0.13

TABLE 5.2: MC generators used to model SM background processes and the corresponding product of the cross section (σ) and branching fraction ($\mathcal{B}$) at $\sqrt{s} = 8$ TeV.

Process	MC generator	$\sigma \cdot \mathcal{B}$ (pb)
$q\bar{q}, gq \to WW$	POWHEG+PYTHIA6 [94]	5.7
$q\bar{q}$, $gq \to WW$+2j	Sherpa [95] with no $\mathcal{O}(\alpha_s)$ terms	0.039
$gg \to WW$	GG2WW 3.1.2 [96, 97]+HERWIG [98]	0.16
$t\bar{t}$	MC@NLO [99]+HERWIG	240
Single top: tW, tb	MC@NLO+HERWIG	28
Single top: tqb	AcerMC [100]+PYTHIA6	88
Z/γ^*, inclusive	ALPGEN+HERWIG	16000
$Z^{(*)}Z^{(*)} \to 4\ell$	POWHEG+PYTHIA8	0.73
$WZ/W\gamma^*, m_{Z/\gamma^*}$>7	POWHEG+PYTHIA8	0.83
$W\gamma^*, m_{\gamma^*} \leq 7$	MadGraph [101–103]+PYTHIA6	11
$W\gamma$	ALPGEN+HERWIG	370

Separate programs are used to generate the hard scattering process and to model hadronisation, PS and UE, as described in Section 1.5.1. PYTHIA8 [93] or PYTHIA6 [94] are used in the latter three steps for signal and some background processes. When HERWIG [98] is used for hadronisation and PS, UE is modelled with JIMMY [104]. W+jets, Z/γ^*+jets and $W\gamma$ processes are generated using *ALPGEN*+*HERWIG* with MLM matching scheme [105]. In addition, Sherpa [95] is used for both hard-scattering process and PS modelling of the VBF production mechanism. Cross sections for the $W\gamma$ and $W\gamma^*$ processes are normalised to the MCFM [106] NLO predictions. These normalisation factors (K-factors) are calculated to be 1.15 for $W\gamma$ and 2.01 for $W\gamma^*$.

The pdf (see Section 1.5.1) set from CT10 [107] is used for POWHEG and MC@NLO samples; CTEQ6L1 [108] is used for ALPGEN, MadGraph, PYTHIA6 and PYTHIA8 samples. Acceptances and efficiencies are obtained from a full simulation [109] of the ATLAS detector using GEANT4 [110]. In two exceptional cases ($q\bar{q}$, $gq \to WW$ and single top processes) a fast simulation is used to increase MC statistics. The simulation incorporates a model of the event pile-up conditions in the collected data considering the effects described in Section 2.2.2. These are multiple proton-proton collisions in the same bunch crossing (in-time pile-up) and nearby bunch crossings (out-of-time pile-up).

5.4 Data Samples

The main goal during the data-taking process is to receive and record, under good detector conditions, the maximum amount of integrated luminosity. Stable accelerator operation and optimal conditions for the full detector system are required in order to accept an event by the ATLAS analyses. Any change in the LHC conditions, calibrations and tests are labelled by the run number. Run numbers are grouped into periods if they share common data-taking conditions. Data periods are cataloged by a letter and a digit, like period Ax, where x represents an integer. The letter is incremented when differences in data-taking conditions or detector operation are expected to be significant, as after technical stop periods. The digit in the period is incremented for small changes. The ATLAS data samples at 8 TeV divided by periods and with their corresponding luminosity, are listed in Tab. 5.3.

TABLE 5.3: Data-taking periods and corresponding integrated luminosity for the full 20.7 fb^{-1} dataset collected by ATLAS detector during 2012 at 8 TeV.

Data period	Luminosity (fb^{-1})
A	0.84
B1-B3	0.30
B4-B8	2.10
B8-B11	1.19
B12-B14	1.64
C1-C9	1.54
D1-D8	3.37
E1-E5	2.70
G-L	7.00

The beams provided to the LHC main ring are typically circulating for several hours. As the data are recorded, people in the control room monitor detector hardware and software operations quality, and take actions if problems are spotted. Soon thereafter sub-detector and physics groups further check the quality of the data. The granularity of these decisions is around 1-2 minutes, known as 1 luminosity block (lumiblock). Flags signalling the quality of the data are associated with each lumiblock feed into a report, the so-called Good Runs List. This is the baseline for any ATLAS analysis.

5.5 Trigger and Pre-selection

The $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis selects events with exactly two oppositely charged leptons. These events are triggered and forced to satisfy quality requirements on the reconstructed vertices. This set of preliminary requirements is known as pre-selection and they are applied in order to discriminate potentially badly reconstructed events. This section describes the trigger criteria and the pre-selection requirements used in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis.

5.5.1 Trigger

The analysis relies on single lepton triggers to select events with high-p_{T} leptons. The increasing number of pile-up events during 2012 data acquisition results in an enhanced event rate. To address this rising event rate the trigger threshold for primary single lepton triggers in 2012 was stricter than for the 2011 data-taking period. The 2012 data samples used in this search are triggered requiring at least one electron or one muon candidate with $p_{\mathrm{T}} > 24$ GeV. A combination of two different lepton identification criteria is chosen for electron and muon triggers. This allows to recover efficiency at high p_{T}.

For the electron triggers case, there is a hadronic leakage cut which consists of a veto on hadronic energy of more than or equal to 1 GeV deposited in the hadronic layers of the calorimeter, including EM clusters. The inefficiency of this selection for large energy deposit justifies the combination of electron triggers. Similarly, for the muon triggers case, a combination of two isolation criteria is used during the event selection.

Furthermore, the lepton triggers have a relative track-based isolation requirement,

$$\frac{\sum\limits_{\text{tracks}} p_{\text{T}}(\text{tracks})}{p_{\text{T}}(\ell)} < A\,, \tag{5.1}$$

where $\sum\limits_{\text{tracks}} p_{\text{T}}(\text{tracks})$ is the scalar sum of the p_{T} of the tracks measured in the ID and having a $p_{\text{T}} > 1\,\text{GeV}$ within a cone of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.2$ around the lepton (ℓ). $p_{\text{T}}(\ell)$ is the transverse momentum of the lepton candidate and A is the threshold value corresponding to 0.1 and 0.12 for the electron and the muon trigger, respectively.

Because of the detector geometry, the acceptance of the lepton triggers is limited to $|\eta| < 2.4$ and additional non-isolated triggers are used. The lepton trigger efficiencies (ϵ) have been determined using the Z tag-and-probe method and are measured as a function of p_{T}, η, and data-taking period using $Z \to \ell\ell$ (with $\ell = \mu, e$) candidate events [111]. The efficiencies are approximately 90% for electrons and 90% (70%) for muons in the end-cap (barrel). In order to account for trigger performance mis-modelling in simulated samples, scale factors relying on these efficiencies in both data and MC samples are computed as follows,

$$\text{scale factor} = \frac{1 - \prod\limits_{n=1}^{N} (1 - \epsilon_{\text{Data,n}})}{1 - \prod\limits_{n=1}^{N} (1 - \epsilon_{\text{MC,n}})}\,, \tag{5.2}$$

where N is the number of objects satisfying analysis specific selection criteria. $\epsilon_{\text{Data,n}}$ and $\epsilon_{\text{MC,n}}$ are the trigger efficiencies for the objects determined with data and MC expectation, respectively.

The additional $4.6\,\text{fb}^{-1}$ of data collected at $\sqrt{s} = 7\,\text{TeV}$ in 2011 use inclusive single-lepton triggers. The required p_{T} thresholds are 18 GeV for muons and between 20 and 22 GeV for electrons.

5.5.2 Pre-selection

This section describes the event pre-selection, as well as the main background processes and the associated rejection criteria. The background composition depends on the final state lepton flavour combination. In light of this, the requirements based on the dilepton invariant mass ($m_{\ell\ell}$) are optimised for $e\mu$+μe and ee+$\mu\mu$ final states separately.

As introduced in Section 5.4, data quality criteria are applied to data events in order to suppress non-collision backgrounds such as cosmic-ray muons, beam-related backgrounds, and noise in the calorimeters. Only data events tagged in the Good Runs List are used in the analysis. Moreover, general event selection requires a vertex with at least three associated charged-particle tracks with $p_{\mathrm{T}} > 400\,\mathrm{MeV}$ and consistent with the beam spot position.

Events with exactly two high-p_{T} leptons with opposite charge are selected by the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis. The leading lepton of the pair, corresponding to the lepton with higher transverse momentum, is required to have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$. The other lepton, known as the sub-leading lepton, is required to have $p_{\mathrm{T}} > 15\,\mathrm{GeV}$. The event is accepted by the analysis if at least one of the leptons fires one of the triggers discussed in Section 5.5.1. Then, a lower bound on the dilepton invariant mass is applied in order to suppress contributions from γ^* and Υ decays. The threshold values are 10 and 12 GeV for $e\mu{+}\mu e$ and $ee{+}\mu\mu$ final states, respectively. For the $ee{+}\mu\mu$ case, an additional criteria on the $m_{\ell\ell}$ consistent with the Z boson mass peak ($|\, m_{\ell\ell} - m_Z| < 15\,\mathrm{GeV}$) is required to reduces $Z \to \ell\ell$ contribution. The dilepton invariant mass distributions for events satisfying the pre-selection requirements described above, except the Z veto, are shown in Fig. 5.2.

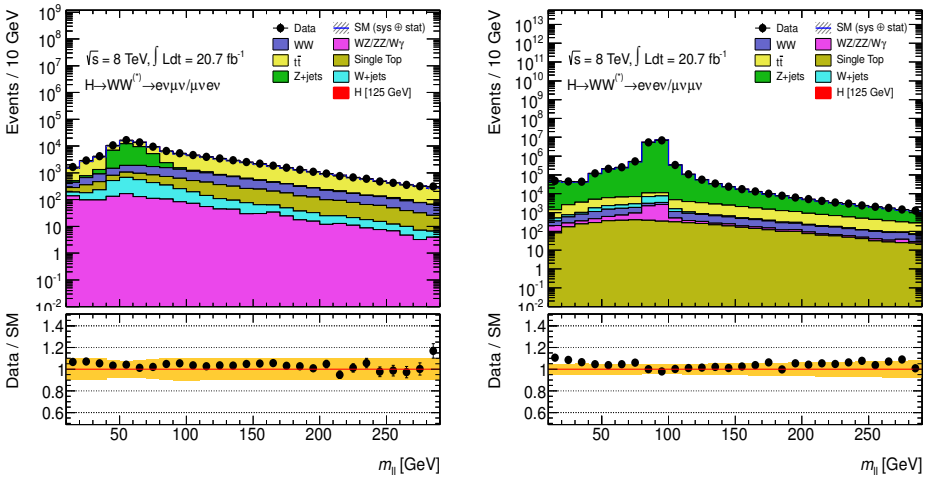

FIGURE 5.2: Dilepton invariant mass distribution in 8 TeV data for $e\mu{+}\mu e$ and $ee{+}\mu\mu$ final states on the left and right, respectively. The expected composition of the SM backgrounds from MC is also shown. The events satisfy all pre-selection requirements, except the Z veto for the $ee{+}\mu\mu$ final states. The lower part of each plot shows the ratio between the data and the background expectation, with the yellow band indicating the total statistical and systematic uncertainty.

There are five major background processes that can satisfy the requirements described above: multi-jets, $W+$ jets, Drell-Yan, top quark and diboson pair decays. In the following, top background refers to the combined $t\bar{t}$ and single top processes (tW, tb, and tqb), unless stated otherwise. The Drell-Yan processes (γ^*, Υ, Z) are noted as Z+jets in tables and legends. In the following these processes will be referred to as DY in the text. The non-WW diboson backgrounds (WZ, ZZ, $W\gamma$, $W\gamma^*$) are noted as VV in tables and $WZ/ZZ/W\gamma$ in legends. Multi-jets and W bosons produced in association with hadronic jets can be a large source of background if a jet is misidentified as a lepton. For sake of simplicity, both processes will be combined and referred to as $W+$ jets.

Table 5.4 collects the expected event yields for the Higgs boson signal, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, the SM background from MC simulation, and the observed 8 TeV data events at each pre-selection stage for $e\mu{+}\mu e$ final states. Similarly, Tab. 5.5 presents the event yields for $ee{+}\mu\mu$ final states.

	Signal $[m_{\mathrm{H}} = 125]$	Total bkg.	Observed	WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets
Trigger	631.5 ± 3.3	116270 ± 160	122800	10721 ± 26	4120 ± 50	53840 ± 80	5545 ± 31	37010 ± 120	5024 ± 24
Lepton p_{T}	529.1 ± 3.0	103980 ± 150	107737	10153 ± 25	3200 ± 40	51020 ± 80	5240 ± 30	29340 ± 110	5024 ± 24
Opp. charge	508.2 ± 2.7	99130 ± 140	103713	10128 ± 25	1577 ± 29	50900 ± 80	5198 ± 30	28800 ± 100	2504 ± 20
$m_{\ell\ell} > 10$	504.2 ± 2.7	99020 ± 140	103579	10118 ± 25	1546 ± 28	50860 ± 80	5193 ± 30	28810 ± 100	2497 ± 20

TABLE 5.4: Pre-selection event yields in 8 TeV data for $e\mu{+}\mu e$ final states. The observed and expected yields for the Higgs signal and SM background are shown. The composition of the total background is given on the right. Energies, masses, and momenta are in units of GeV. All uncertainties are statistical.

	Signal $[m_{\mathrm{H}} = 125]$	Total bkg.	Observed	WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets
Trigger	663.2 ± 3.5	14605000 ± 8000	14707614	10924 ± 26	9030 ± 40	53990 ± 80	5514 ± 31	14513000 ± 8000	12550 ± 270
Lepton p_{T}	560.4 ± 3.2	14321000 ± 8000	14389435	10360 ± 25	8035 ± 34	51180 ± 80	5207 ± 30	14233000 ± 8000	12550 ± 270
Opp. charge	536.2 ± 3.0	14298000 ± 8000	14369989	10339 ± 25	6662 ± 26	51070 ± 80	5170 ± 30	14214000 ± 8000	11130 ± 270
$m_{\ell\ell} > 12$	523.0 ± 2.9	14284000 ± 8000	14320704	10303 ± 25	6587 ± 25	50920 ± 80	5150 ± 30	14201000 ± 8000	10230 ± 270
Z veto	495.4 ± 2.7	1311500 ± 2000	1372772	9230 ± 26	2004 ± 22	41790 ± 70	4248 ± 28	1251000 ± 1900	3640 ± 120

TABLE 5.5: Pre-selection event yields in 8 TeV data for $ee{+}\mu\mu$ final states. The observed and expected yields for the Higgs signal and SM background are shown. The composition of the total background is given on the right. Energies, masses, and momenta are in units of GeV. All uncertainties are statistical.

5.6 Object Selection

Proton-proton collision events producing $H \to WW^{(*)} \to \ell\nu\ell\nu$ candidates are selected based on pre-selection requirements, object selections and, final kinematic cuts depending on the flavour composition of the final leptons and the number of jets (N_{jets}) accompanying the Higgs boson candidate event. The pre-selection requirements are addressed in Section 5.5. This section focuses on defining the objects used in the analysis: leptons, jets and missing transverse momentum.

5.6.1 Leptons

Electron candidates are selected by applying a set of tight identification criteria using a combination of tracking and calorimetric information. Muon candidates are identified by matching tracks reconstructed in the ID and in the MS. Requirements on the number of hits in all three components of the ID (pixels, SCT, and TRT) provide background rejection, particularly against pion and kaon decays-in-flight.

5.6.1.1 Electron Selection

In 2012, electron candidates are reconstructed by the Gaussian Sum Fitter (GSF) algorithm, where candidate electron or photon conversion tracks with $p_{\text{T}} > 1\,\text{GeV}$ are refitted based on a non-linear generalisation of the Kalman Fitter after loose matching between tracks and EM clusters [112]. This algorithm helps, in particular, to account for bremsstrahlung losses.

The nominal analysis relies on electron candidates within the geometrical acceptance ($|\,\eta\,| < 2.47$), excluding the transition region between the barrel and the end-cap regions of the calorimeter ($1.37 < |\,\eta\,| < 1.52$), and with transverse energy greater than 15 GeV. Here the candidate's energy is taken from the calorimeter cluster with a small correction, i.e., rescaling with respect to transverse energy for data in order to account for layer calibrations, gain corrections and intermodule widening effects. A smearing with respect to calorimeter energy resolution is applied to the MC samples as well. However, its η and ϕ are taken from the track information. Figure 5.3 shows the η distributions for electron candidates satisfying pre-selection requirements.

Electrons are also required to satisfy reconstruction quality requirements: candidates are excluded if they lie in regions of the LAr calorimeter encountering transient issues during the run, such as dead regions of the detector, when the event

was recorded. Then the electron candidates are selected using an identification quality which consists of calorimeter shower shape, track quality, track-cluster matching and transition radiation energy criteria. This is re-optimised for the 2012 data in order to cope with the increased level of pile-up and trigger rates.

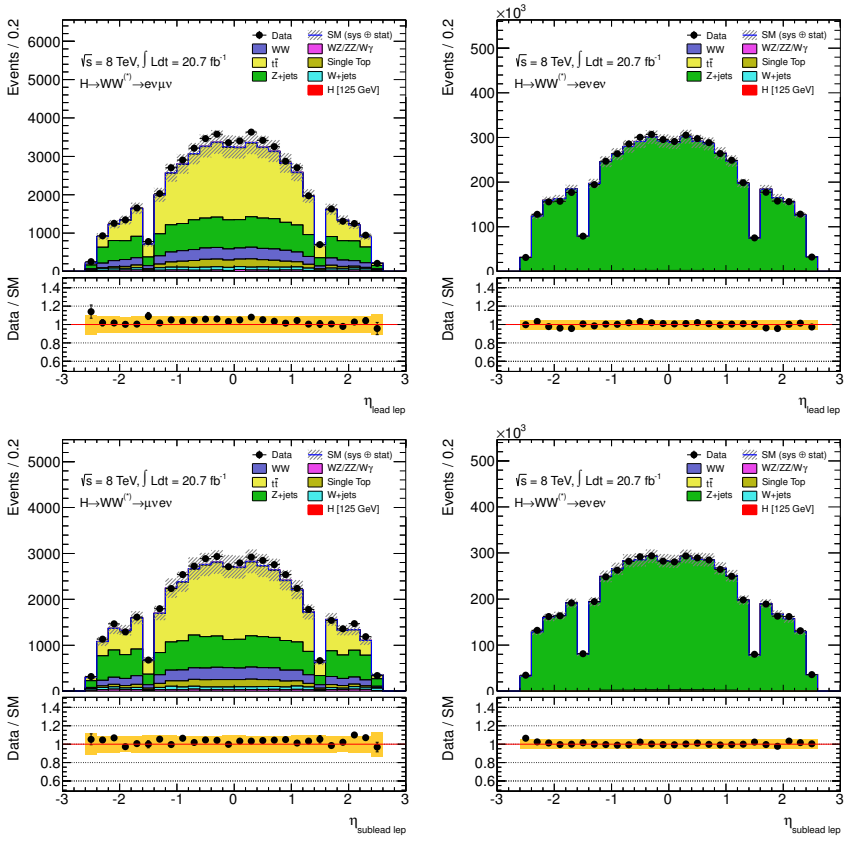

FIGURE 5.3: The leading (top row) and sub-leading (bottom row) electron η distributions for the $e\mu$+μe and ee final states on the left and right, respectively. The distributions are obtained after the application of the lepton p_{T} thresholds and the low $m_{\ell\ell}$ requirement. The lower part of each plot shows the ratio between the data and the background expectation from MC, with the yellow band indicating the total statistic and systematic uncertainty in the normalization.

In addition to the set of lepton standard requirements described above, the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis further suppresses SM backgrounds, aiming, in particular, to reduce the W+jets contamination. This is done by applying not only distance requirements with respect to the PV (see Section 4.2.2), but also isolation cuts where tracks and the calorimeter cells energy are matched. The former, known as the transverse impact parameter (d_0), requires that tracks point back to the PV

with a d_0 significance less than 3. The d_0 significance is defined as the ratio of the d_0 to its measured uncertainty as follows,

$$\frac{d_0\, w.r.t\, PV}{sig(d_0\, w.r.t\, PV)}\,. \tag{5.3}$$

Also the difference in z position with respect to the PV times $\sin\theta$, where θ represents the azimuthal angle, is required to be less than 0.4 mm.

The Topological Clustering algorithm addressed in Chapter 3 is used for calorimeter isolation requirements. The topological isolation energy is based on the sum from topoclusters with positive energy, inside a cone of radius ΔR. Moreover, the pile-up noise is also corrected. This correction is based on the event-by-event ambient energy density to make it more robust against pile-up [111]. Given the large background from $W+$ jets production, the isolation criteria have been tightened in 2012 with respect to the 2011 data. Track isolation uses the scalar sum of the transverse momenta of all tracks within a 0.3 cone around the electron normalized to the electron p_{T} (pTcone30/$p_{\mathrm{T}} \equiv \sum_{\mathrm{tracks}}(p_{\mathrm{T}}^{\mathrm{tracks}})/p_{\mathrm{T}}$). Optimisation studies on the track isolation lead to p_{T} dependent cuts. The calorimeter isolation similarly uses the scalar sum of cell energy deposits projected onto the transverse plane for all cells within a 0.3 cone around the electron. Then a p_{T} independent relative requirement is applied (Etcone30/$p_{\mathrm{T}} \equiv \sum_{\mathrm{cell}}(E_T^{\mathrm{cell}})/p_{\mathrm{T}}$). The threshold values required for electron candidates are summarised in Tab. 5.6.

TABLE 5.6: Isolation and impact parameter requirements for electron candidates in $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis.

Cut	Threshold
Etcone30/p_{T}	< 0.16
pTcone30/p_{T}	$< 0.12\,(0.16)$ if $15 \leq p_{\mathrm{T}} < 25$ GeV ($p_{\mathrm{T}} \geq 25$ GeV)
d_0 significance	< 3.0
$z_0 \sin\theta$	0.4 mm

The small differences between observed data and MC prediction arising from identification, impact parameter, isolation and the calorimeter pile-up correction are accounted for as a function of the reconstructed p_{T} for each electron candidate. These scale factors are measured using the Z tag-and-probe technique and they are consistent with unity to within 1-2%. These factors are applied as "weights" in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis. Figure 5.4 shows the p_{T} distribution for the selected electron candidates after the pre-selection criteria are applied.

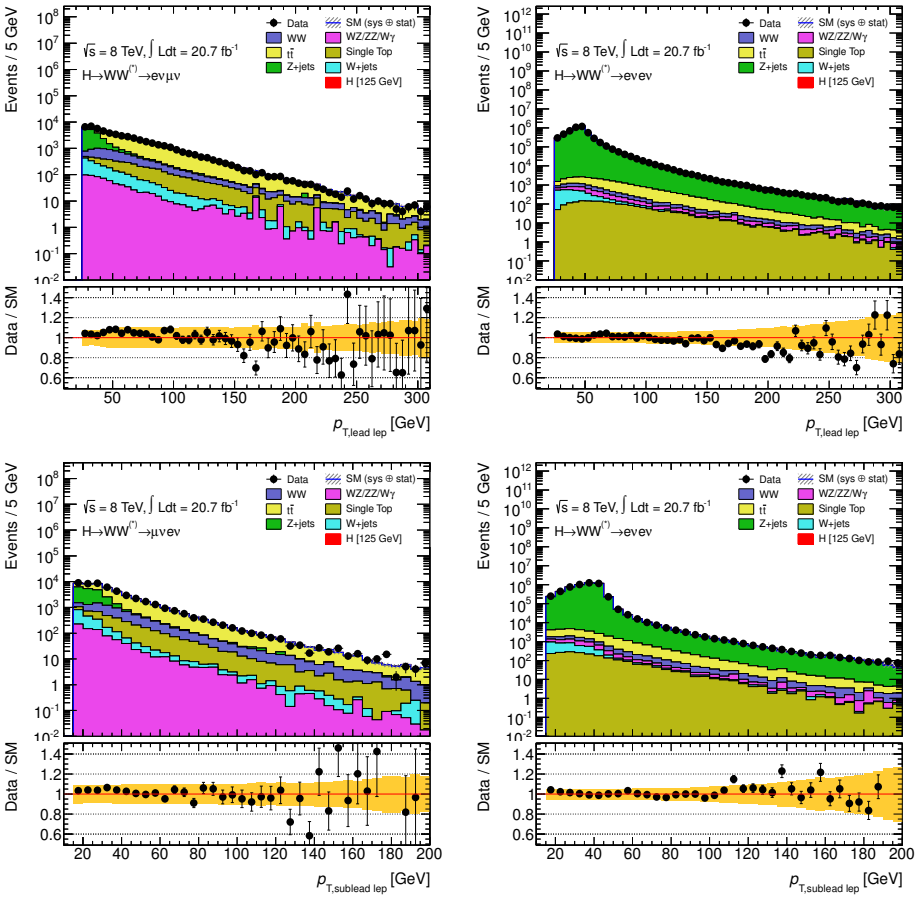

FIGURE 5.4: The leading (top row) and sub-leading (bottom row) electron p_{T} distributions for the $e\mu$+μe and ee final states on the left and right, respectively. The distributions are obtained after the application of the lepton p_{T} thresholds and the low $m_{\ell\ell}$ requirement. The lower part of each plot shows the ratio between the data and the background expectation from MC, with the yellow band indicating the total statistic and systematic uncertainty in the normalization.

Finally, the selected electrons have to satisfy an object isolation requirement, known as overlap removal. An electron is vetoed if it is within a ΔR cone of 0.1 with a well identified muon, as described in the Section 5.6.1.2. Additionally, if two electrons are matched within a ΔR cone of 0.1, the lower-p_{T} electron will be discarded.

5.6.1.2 Muon Selection

Muon candidates are reconstructed using an algorithm that combines the information coming from the ID and the MS. The muon trajectory is measured by both systems separately at first and then combined using a statistical combination approach [113].

To reduce mis-identification and to improve on the muon momentum resolution, the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis applies quality requirements on the muon track reconstruction. Muon tracks are required to have at least two hits in the pixel detector (one of them in the first layer) and six or more hits in the SCT. Tracks are vetoed if they have more than two holes in the SCT and pixel detectors, as well as tracks with an excessive amount of outlier hits in the TRT. In addition, kinematic cuts of $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| < 2.4$ are applied.

Isolation cuts are required to the muons in order to reduce backgrounds from $W+$ jets and multi-jets production. To deal with high pile-up, the calorimeter isolation quantity is corrected for its dependence on the number of reconstructed vertices. A linear correction was applied to the muon calorimeter isolation in order to maintain a flat efficiency with respect to pile-up. The isolation correction is applied to observed data and simulation in the same manner. Then the isolation cuts, PtCone30/p_{T} and EtConeCor30/p_{T}, are optimised in different p_{T} bins. A scan over the significance, including the effect of the large systematic uncertainty of the $W+$ jets, is used to find the optimal threshold value,

$$\frac{N_{\mathrm{sig}}}{\sqrt{N_{\mathrm{bkg}} + \sigma^2_{W+\,\mathrm{jets}}}}\,. \tag{5.4}$$

To account for larger impact parameter uncertainties at high η, the analysis also uses $z_0 \sin\theta$ and d_0 significance thresholds. A summary of all the selection criteria for muon candidates is summarised in Tab. 5.7.

Table 5.7: Isolation and impact parameter requirements for muon candidates in $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis.

Cut	Threshold
EtConeCor30/p_{T}	$< 0.014 p_{\mathrm{T}} - 0.15$ and < 0.20
PtCone30/p_{T}	$< 0.01 p_{\mathrm{T}}$ and < 0.15
d_0 significance	< 3.0
$z_0 \sin\theta$	$< 1.0\,\mathrm{mm}$

Finally, to quantify the differences between MC prediction and observed data, all the selection requirements are computed using a Z tag-and-probe sample. The derived scale factor is consistent with 1 within a percent systematic error. The muon efficiency after all selection criteria described above is around 70% for a Higgs boson with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. Figure 5.5 and Fig. 5.6 show the p_{T} and η distributions of the selected muon candidates after the pre-selection requirements, respectively.

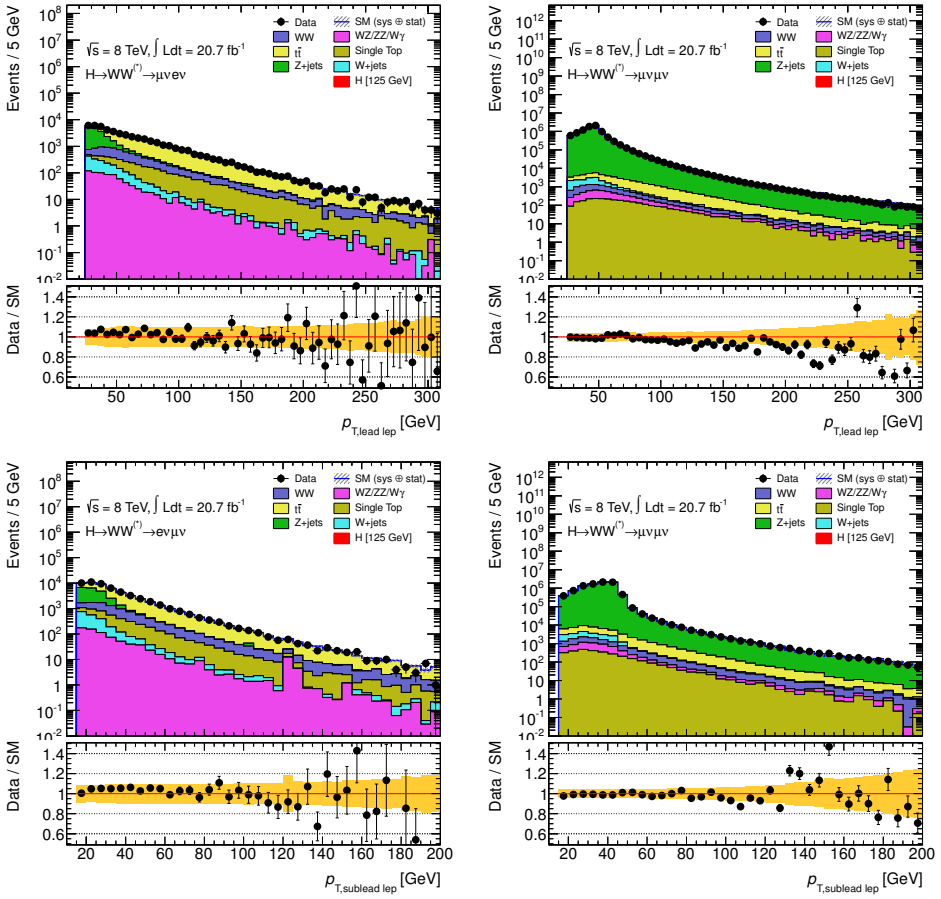

FIGURE 5.5: The leading (top row) and sub-leading (bottom row) muon p_{T} distributions for the $e\mu$+μe and $\mu\mu$ final states on the left and right, respectively. The distributions are obtained after the application of the lepton p_{T} thresholds and the low $m_{\ell\ell}$ requirement. The lower part of each plot shows the ratio between the data and the background expectation from MC, with the yellow band indicating the total statistic and systematic uncertainty in the normalization.

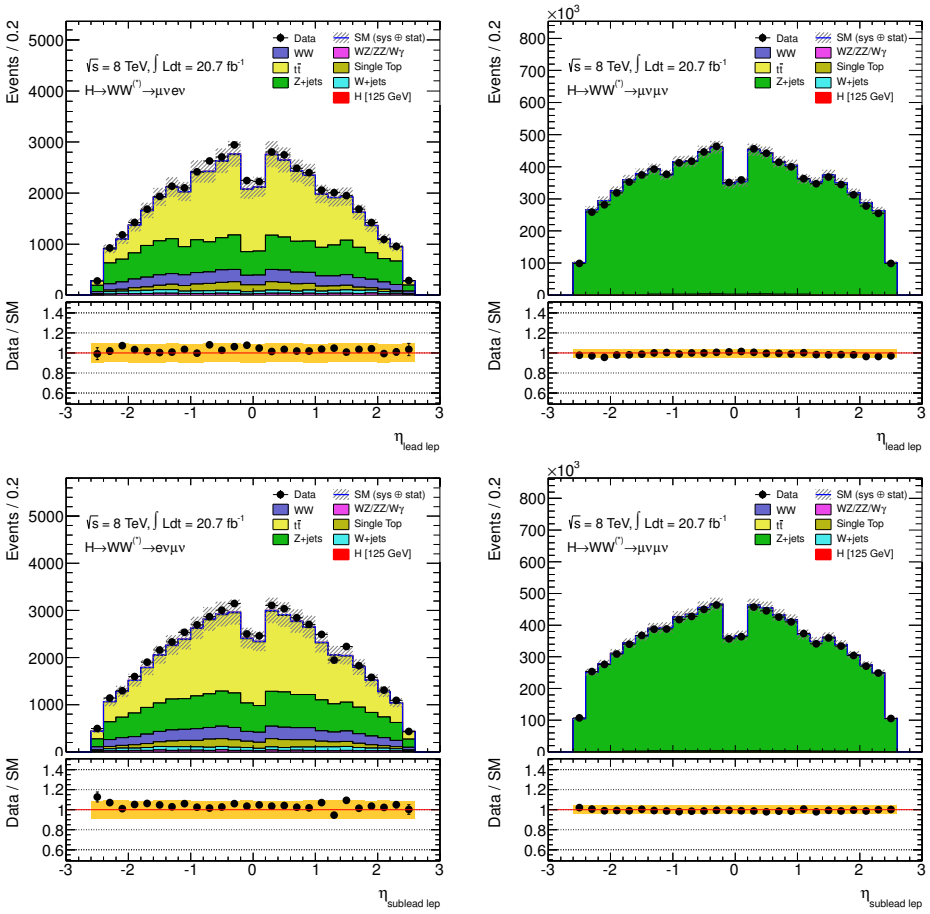

FIGURE 5.6: The leading (top row) and sub-leading (bottom row) muon η distributions for the $e\mu$+μe and $\mu\mu$ final states on the left and right, respectively. The distributions are obtained after the application of the lepton p_{T} thresholds and the low $m_{\ell\ell}$ requirement. The lower part of each plot shows the ratio between the data and the background expectation from MC, with the yellow band indicating the total statistic and systematic uncertainty in the normalization.

5.6.2 Jets

One of the features of high-energy proton-proton collisions at the LHC is the production of highly collimated sprays of energetic hadrons originating from quarks and gluons in the primary collisions. The large centre-of-mass energy at the LHC enables the production of Lorentz-boosted heavy particles as well. When massive particles decay hadronically, the products tend to be collimated in a localized area of the detector. In such situations, the resulting hadrons are clustered into a single object known as a jet.

In the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis the jets are reconstructed using the anti-k_t algorithm with distance parameter $R = 0.4$ [114]. The anti-k_t algorithm acts by iteratively merging the nearest object constructed from topoclusters deposited in the calorimeters.

Pile-up interactions contribute with spurious energy to the input of the jet algorithms. This undesirable effect should be under control and reduced as much as possible. With this purpose, two data-derived corrections have been developed to mitigate the jet energy dependence on pile-up. The first relies on the product of the event p_{T} density and the jet area [115]. The second correction depends on the number of reconstructed vertices and the mean number of expected interactions, which primarily affects jets in the forward region. Note that this second correction also addresses the effects of out-of-time pile-up. After these two corrections, energy- and η-dependent MC calibrations (typically 40%) are applied to all jets. Finally, a residual correction (few percent) from *in situ* measurements is applied to refine the jet calibration [116].

ATLAS analyses require jets satisfying general quality criteria. This primary selection rejects jets that do not originate from the initial hard scatter. The main backgrounds to jets coming from collisions events are: beam-gas events, where one proton of the beam collides with the residual gas within the beam pipe; beam-halo events, caused by interactions in the tertiary collimators in the beam-line far away from the detector; cosmic ray muons overlapping in-time with collision events; and electronic noise in the calorimeters. The jet quality cuts efficiently reject jets coming from the background processes described above, while keeping a high efficiency, over 99.8%, for jets produced in proton-proton collisions.

In the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis, two selection criteria have been optimised to further suppress pile-up dependence: the p_{T} threshold, used for jet counting, and the JVF (see Section 4.2.2), to reduce the contribution of pile-up jets in the central region. The selection requires $|\,\mathrm{JVF}\,| > 0.5$ for jets with $p_{\mathrm{T}} < 50\,\mathrm{GeV}$. Since JVF relies on tracking information, this can only be computed for jets within $|\,\eta\,| < 2.4$. This motivates the study of the p_{T} threshold separately for central and forward jets, as the latter set does not have any pile-up suppression. Optimisation investigations result in requiring $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ for jets within $|\,\eta\,| < 2.4$, and $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ if the jet is located in $2.4 < |\,\eta\,| < 4.5$. Finally, to be consistent with the statement of well isolated objects, the analysis vetoes jets within a cone of $\Delta R = 0.3$ with lepton candidates, as they are defined in Section 5.6.1. Figure 5.7 shows the p_{T} and η distributions of the higher p_{T} jet candidate from 8 TeV events, containing at least one jet in the final state, after applying the pre-selection requirements.

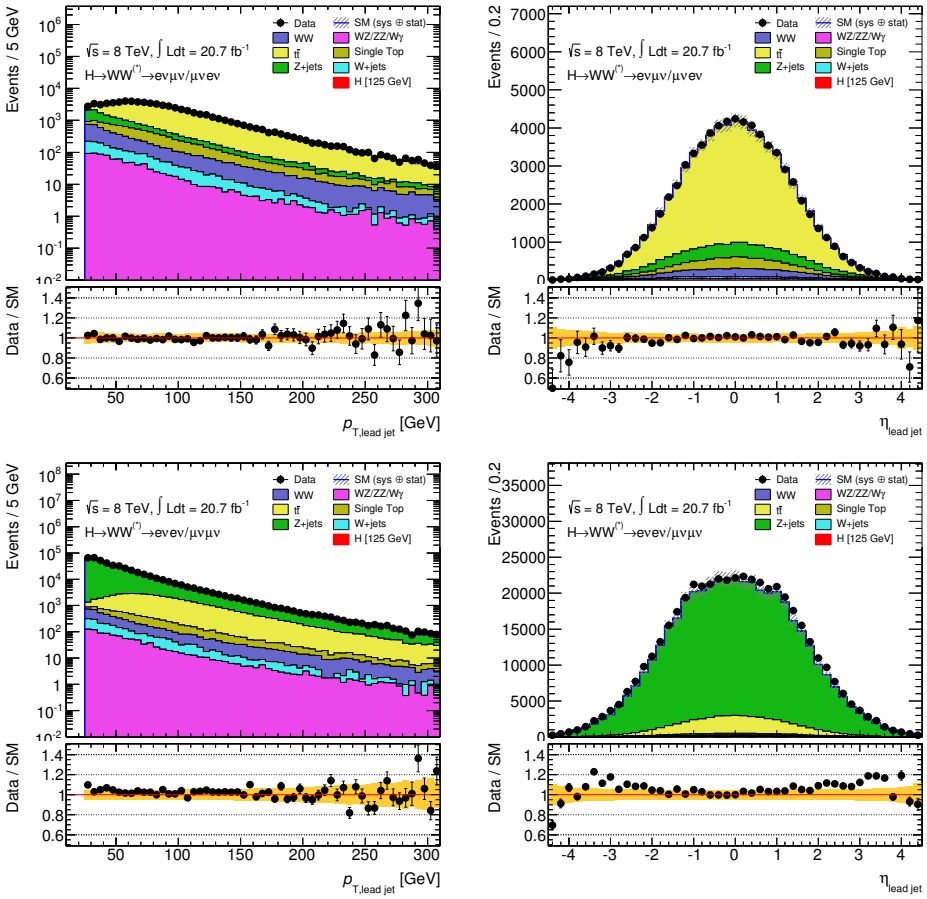

FIGURE 5.7: The p_{T} (left) and η (right) distributions of the leading jet for all events from 8 TeV data satisfying the pre-selection criteria. The $e\mu$+μe final states are shown on top and ee+$\mu\mu$ on the bottom. The lower part of each plot shows the ratio between the observed data and the background expectation from simulation, with the yellow band indicating the total statistic and systematic uncertainty in the normalization.

b-tagging jets

The number of jets identified by a b-tagging algorithm ($N_{b\text{-jet}}$) is used for the H+1j and H+2j analyses. In these two categories top quark events, both $t\bar{t}$ and electroweak production of single top quarks, form a substantial background. If $t\bar{t}$ and Wt processes decay leptonically, the final state can not be distinguished from the Higgs boson signal as far as leptons are concerned. Therefore, the most obvious way to suppress these backgrounds is to exploit the fact that nearly all top quark decays lead to b-quark jets (b-jet). In this light, the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis vetoes

events in which any jet is tagged as a b-jet. The b-tagging algorithm combines the result of a baseline tagging algorithm together with impact parameter and secondary vertex information in a neural network architecture [117, 118]. The b-tagging performance is calibrated for several working point efficiencies for 2012 running conditions, as detailed in Ref. [119, 120]. The $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis uses the 85% efficiency operating point based on the results obtained from optimisation studies. It is found that the dependence of the statistical signal significance and the operating point is negligible. Since these significance studies are performed without considering the effect of systematics, it was decided to minimise the remaining top quark background by selecting the option with the highest efficiency.

5.6.3 Missing Transverse Momentum and Hadronic Recoil

The $\not\!\!E_T$ treatment is one of the most important components of the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis. The signature of the Higgs boson in this decay mode involves the presence of two neutrinos coming from the W bosons in the final state. In light of this, the $\not\!\!E_T$ measurement becomes a key variable in order to suppress processes with non-genuine $\not\!\!E_T$ such as Drell-Yan and multi-jets.

Nevertheless, the high pile-up environment, specially during the 2012 data-acquisition period, generated a challenging situation in terms of $\not\!\!E_T$ measurement and resolution, as extensively investigated in Chapter 4. This section summarises the different $\not\!\!E_T$ reconstruction methods used in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis.

5.6.3.1 $\not\!\!E_T$ Selection

The $\not\!\!E_T$ spectrum presents significant differences depending on the background processes that dominate in each lepton analysis category. Figure 5.8 shows the $E_{\mathrm{T,rel}}^{\mathrm{miss}}$ distributions after the pre-selection requirements for $e\mu{+}\mu e$ and $ee{+}\mu\mu$ final states separately. Moreover, the background composition in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis varies with the number of jets associated with the selected event. For this reason, the $\not\!\!E_T$ measurement also is dependent on the number of jets in the final state. Based on these features, the $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement is optimised in each of the jet categories for $e\mu{+}\mu e$ and $ee{+}\mu\mu$ final states separately after the pre-selection requirements.

In events with $N_{\mathrm{jet}} \leq 1$, multi-jets and Drell-Yan events are suppressed using $E_{\mathrm{T,rel}}^{\mathrm{miss}}$ quantities. In the $e\mu{+}\mu e$ lepton category, Drell-Yan decaying into a pair of τ leptons process, is the main background contribution. The $E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 25\,\mathrm{GeV}$ threshold provides the highest signal-to-background ratio using the 7 TeV data samples.

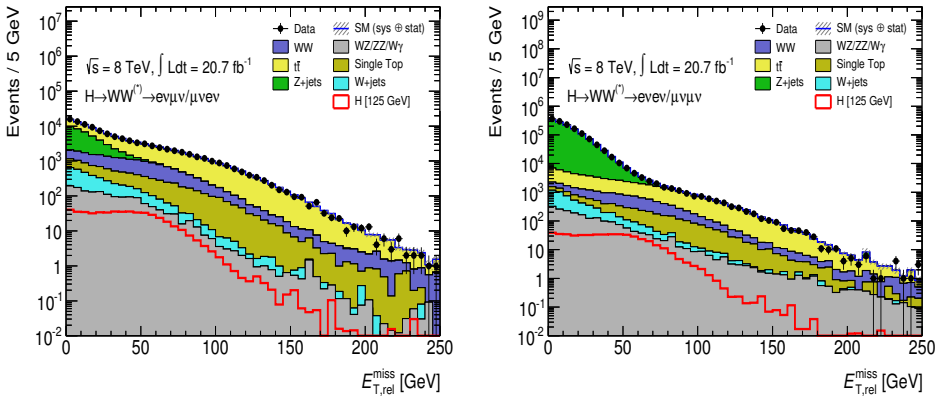

FIGURE 5.8: The $E_{\mathrm{T,rel}}^{\mathrm{miss}}$ distributions for different (same) flavour events after pre-selection cuts on the left (right) from 8 TeV data. The background expectation from the MC simulation is also shown. The Higgs signal is overlaid as a red curve.

In this case, the $\not\!\!E_T$ measurement still tends to be small in these events since the neutrinos from the τ decays are usually back-to-back. The other backgrounds represent genuine $\not\!\!E_T$ measurement, as well as the Higgs signal, so the investigations show that a higher threshold value results in a lower signal-to-background ratio. For the features mentioned above, it is decided to keep this threshold value for the 8 TeV data as well. Figure 5.9 shows the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ distribution after the pre-selection requirements for $e\mu$+μe events with up to 1 jet.

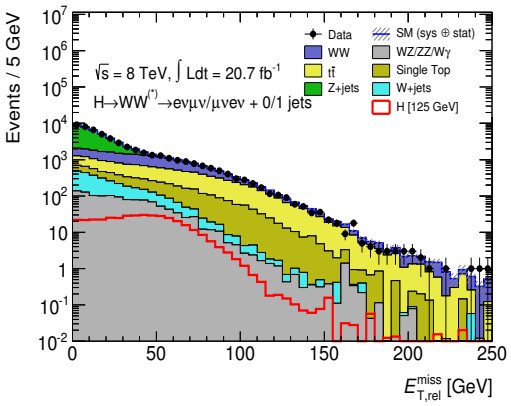

FIGURE 5.9: The $E_{\mathrm{T,rel}}^{\mathrm{miss}}$ distribution for $e\mu$+μe final states from 8 TeV data for events passing the pre-selection criterion and with up to 1 jet. The background expectation from the MC simulation is also shown. The Higgs signal is overlaid as a red curve.

For ee+$\mu\mu$ final states, a higher threshold is used given the large amount of Drell-Yan background, as can be observed in the left distribution of Fig. 5.10. Due to the fake $\not\!\!E_T$ measurement in events with high pile-up, especially for the 8 TeV data, a combined requirement using two $\not\!\!E_T$ reconstruction methods is applied to further reduce this background contamination. The optimisation results, based on the first $\sim 6\,\text{fb}^{-1}$ of data at 8 TeV, represents the best choice of a combination of cuts on $E_{\text{T,Rel}}^{\text{miss}}$ and $E_{\text{T,Rel}}^{\text{miss,track}}$. The optimal operating point, in terms of signal acceptance and background rejection, is found at 45 GeV. The benefit of combining two $\not\!\!E_T$ measurements to suppress Drell-Yan events is shown in the right distribution of Fig. 5.10, which shows the $E_{\text{T,Rel}}^{\text{miss,track}}$ spectrum after the $E_{\text{T,rel}}^{\text{miss}} > 45\,\text{GeV}$ requirement.

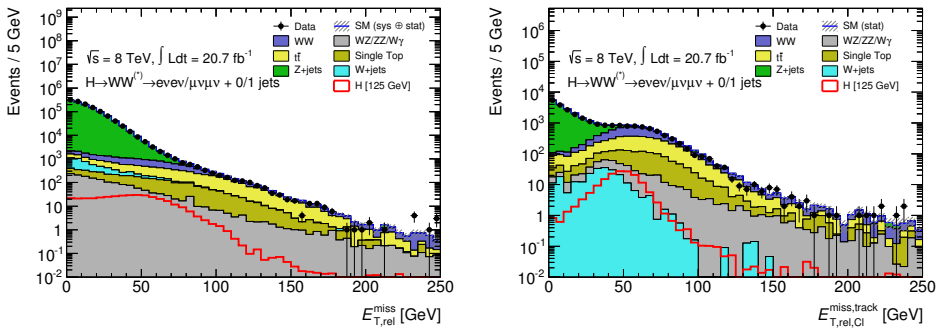

FIGURE 5.10: Left: the $E_{\text{T,rel}}^{\text{miss}}$ distribution from 8 TeV data for ee+$\mu\mu$ events with $N_{\text{jets}} \leq 1$ and passing the pre-selection cuts. Right: $E_{\text{T,Rel}}^{\text{miss,track}}$ after the pre-selection and $E_{\text{T,rel}}^{\text{miss}} > 45\,\text{GeV}$ requirement.

The additional threshold on $E_{\text{T,Rel}}^{\text{miss,track}}$ is applied near the end of the event selection. The reason for this is that the CRs (see Section 5.9) can benefit from higher statistics if the $E_{\text{T,Rel}}^{\text{miss,track}}$ requirement is not applied after the pre-selection.

The projected $\not\!\!E_{T,Rel}$ measurement is biased in events with high jet activity. A high number of jets in the final state increases the probability of randomly computing the projection in the direction of any nearby jet. In light of this, the H+2j category does not uses projected $\not\!\!E_{T,Rel}$ quantities because of the large number of reconstructed jets in the final state. The $e\mu$+μe final states require $E_{\text{T}}^{\text{miss}} > 20\,\text{GeV}$. The $E_{\text{T}}^{\text{miss}}$ distribution for $e\mu$+μe final states in the H+2j category is shown in Fig. 5.11. For ee+$\mu\mu$ final states, the $\not\!\!E_T$ selection combines $E_{\text{T}}^{\text{miss}} > 45\,\text{GeV}$ and $E_{\text{T}}^{\text{miss, STVF}} > 35\,\text{GeV}$ requirements. The combination of the high level of hadronic activity from the extra jets, in addition to the high pile-up conditions, leads to the $E_{\text{T,STVF}}^{\text{miss}}$ reconstruction performing better than $E_{\text{T}}^{\text{miss,track}}$ for these events [121]. The

right distribution in Fig. 5.12 shows the $E_{\mathrm{T,STVF}}^{\mathrm{miss}}$ spectrum for $ee{+}\mu\mu$ final states in the $H{+}2j$ category after requiring $E_{\mathrm{T}}^{\mathrm{miss}} > 45\,\mathrm{GeV}$. This plot shows how the additional threshold on $E_{\mathrm{T,STVF}}^{\mathrm{miss}} > 35\,\mathrm{GeV}$ helps in reducing the Drell-Yan events that survive after the cut on $E_{\mathrm{T}}^{\mathrm{miss}}$ with a low impact on the signal acceptance.

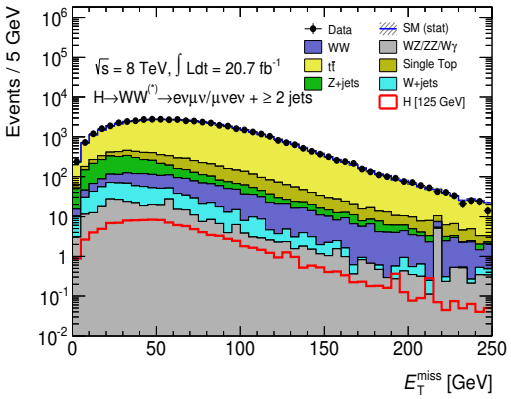

FIGURE 5.11: The $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution for $e\mu{+}\mu e$ final states after the pre-selection cuts for $H{+}2j$ category. 8 TeV data and MC simulation are shown. The Higgs boson signal, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, is overlaid as a red curve.

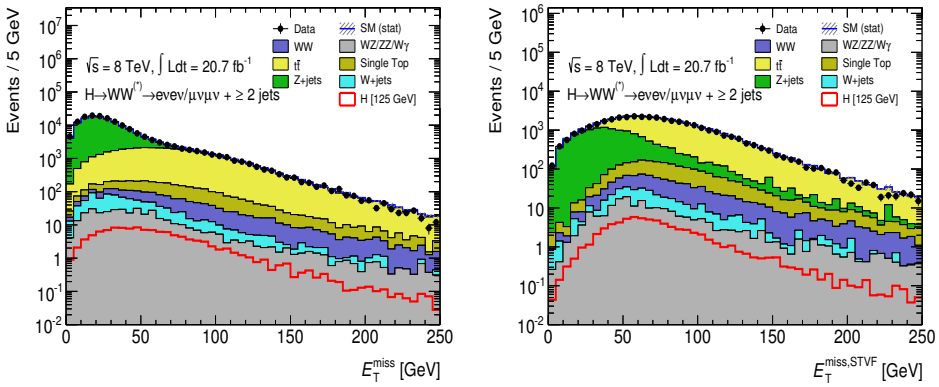

FIGURE 5.12: Left: $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution for $ee{+}\mu\mu$ final states in the $H{+}2j$ category after applying the pre-selection requirements. Right: $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ distribution is shown for event candidates satisfying the $E_{\mathrm{T}}^{\mathrm{miss}} > 45\,\mathrm{GeV}$ criteria. 8 TeV data and MC simulation for SM processes are shown. The Higgs boson signal, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, is overlaid as a red curve in both plots.

Table 5.8 summarises the different $\not\!\!E_T$ reconstructions and their corresponding threshold values in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis for each lepton final state and jet category.

TABLE 5.8: Threshold values on different $\not\!\!E_T$ and $\not\!\!E_{T,Rel}$ reconstruction techniques applied in each of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis categories. The symbol "-" indicates no $\not\!\!E_T$ selection in the corresponding reconstruction and category.

Reconstruction	H+0j and H+1j Analyses		H+2j Analysis	
	ee+$\mu\mu$	$e\mu$+μe	ee+$\mu\mu$	$e\mu$+μe
$E_{\mathrm{T}}^{\mathrm{miss}}$	$E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 45$ GeV	$E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 25$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 45$ GeV	$E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV
$E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$	-	-	$E_{\mathrm{T}}^{\mathrm{miss,\,STVF}} > 35$ GeV	-
$E_{\mathrm{T}}^{\mathrm{miss,track}}$	$E_{\mathrm{T}}^{\mathrm{miss,track}} > 45$ GeV	-	-	-

5.6.3.2 Soft Hadronic Recoil in Drell-Yan Events

For the H+0j and H+1j categories with ee+$\mu\mu$ final states an additional quantity is used to further suppress Drell-Yan contamination. In what follows, this variable will be referred to as f_{recoil}. The f_{recoil} quantity is a measurement of the soft hadronic recoil opposite to the system of the leptons and any accompanying jet for $N_{\mathrm{jets}} \leq 1$. This quantity is defined for the H+0j analysis as the ratio of the recoil momentum and the transverse momentum of the dilepton system ($p_{\mathrm{T}}^{\ell\ell}$),

$$\frac{\left| \sum_{\text{jets with } p_{\mathrm{T}}>10 \text{ GeV}} |\mathrm{JVF}| \times \vec{p}_{\mathrm{T}}^{\,\mathrm{jet}} \right|}{p_{\mathrm{T}}^{\ell\ell}}. \tag{5.5}$$

In the H+1j analysis, $p_{\mathrm{T}}^{\ell\ell j}$ is used in place of $p_{\mathrm{T}}^{\ell\ell}$, where it is the modulus of the vectorial sum defined by $\vec{p}_{\mathrm{T}}^{\,\ell\ell j} = \vec{p}_{\mathrm{T}}^{\,\ell\ell} + \vec{p}_{\mathrm{T}}^{\,\mathrm{jet}}$. The numerator in Eq. 5.5 represents the recoil momentum. This is defined as the vectorial sum p_{T} of the low-p_{T} jets in the ϕ quadrant opposite to $\vec{p}_{\mathrm{T}}^{\,\ell\ell}$ and $\vec{p}_{\mathrm{T}}^{\,\ell\ell j}$ for H+0j and H+1j analyses, respectively. Low-p_{T} jets are defined as those with $p_{\mathrm{T}} > 10$ GeV and each of them is weighted by its JVF value.

The f_{recoil} distribution of Drell-Yan events is distinct from that of non Drell-Yan processes, because of the different event topology. The dilepton system in Drell-Yan events is balanced by soft hadronic activity, resulting in large values of f_{recoil}, whereas the dilepton system in WW, top, Higgs signal, and similar processes is balanced by a combination of recoiling neutrinos and soft hadronic activity, which results in small values of f_{recoil}. Figure 5.13 shows the f_{recoil} distributions for Drell-Yan, non Drell-Yan and signal processes in simulated events with no jets in the final state.

The thresholds used in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis are $f_{\text{recoil}} < 0.05$ for the H+0j bin and 0.2 for the H+1j mode; this results in a Z/DY background rejection of $\sim$ 80% and $\sim$ 50%, respectively. Note that the requirement on f_{recoil} is evaluated almost at the end of the event selection (see Section 5.7), following the strategy used for applying the $E_{\text{T,Rel}}^{\text{miss,track}} > 45\,\text{GeV}$ requirement.

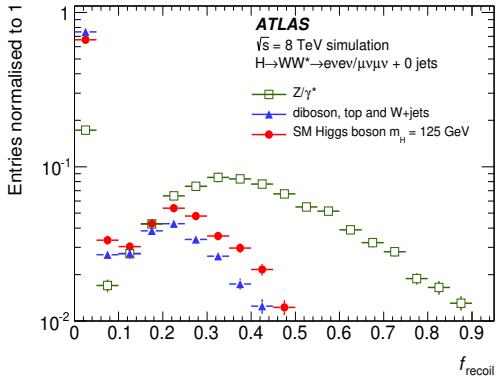

FIGURE 5.13: The f_{recoil} distributions for ee+$\mu\mu$ with no jets final state after the selection requirement on $m_{\ell\ell} < 50\,\text{GeV}$. Simulated Drell-Yan, non Drell-Yan and signal processes are shown. The total number of events is normalised to unity for easier comparison.

5.7 Event Selection

Basic pre-selection criteria used in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis have been stablished in Section 5.5, whereas object reconstruction requirements are addressed in Section 5.6. In the current section, kinematic event selection is presented for each jet category in which this analysis is divided.

5.7.1 Analyses Categorised in N_{jets}

The $H\to WW^{(*)}\to \ell\nu\ell\nu$ events are categorised by the number of jets in the final state. This leads to the H+0j, H+1j and, H+2j categories, following Ref. [26] introduced in Section 5.2. There are two main reasons for this jet categorisation as a strategy. These are differences in the Higgs boson production mechanism and different SM background rates and composition.

For the former, Tab. 5.9 quotes the fraction of expected Higgs boson signal, with $m_{\text{H}} = 125\,\text{GeV}$, produced by ggF and VBF for different number of jets in

the final state. As observed, the Higgs boson production mechanism is related to the number of final jets. In $N_{\text{jets}} \leq 1$ events, the signal mainly originates from the ggF production mechanism, whereas in the $N_{\text{jets}} \geq 2$ case the Higgs boson is almost entirely produced by the VBF process.

TABLE 5.9: Expected fractions of the ggF and VBF production modes for a Higgs boson with $m_{\text{H}} = 125\,\text{GeV}$ in each jet category after the $H \to WW^{(*)} \to \ell\nu\ell\nu$ pre-selection and $\not{E}_T$ requirements for all lepton final states combined.

N_{jets}		Production Mechanism	
		ggF	VBF
≤ 1	$= 0$	53.5%	2%
	$= 1$	25%	11%
≥ 2		21.5%	87%

In addition, the latter motivation refers to the different SM background rate and composition depending on the number of jets in the event. After the $\not{E}_T$ selection described in Section 5.6.3.1, the main backgrounds are top quark and Drell-Yan for $e\mu$+μe and ee+$\mu\mu$ final states, respectively. However, the contribution of these processes is very different for each N_{jets} case, as shown in Fig. 5.14.

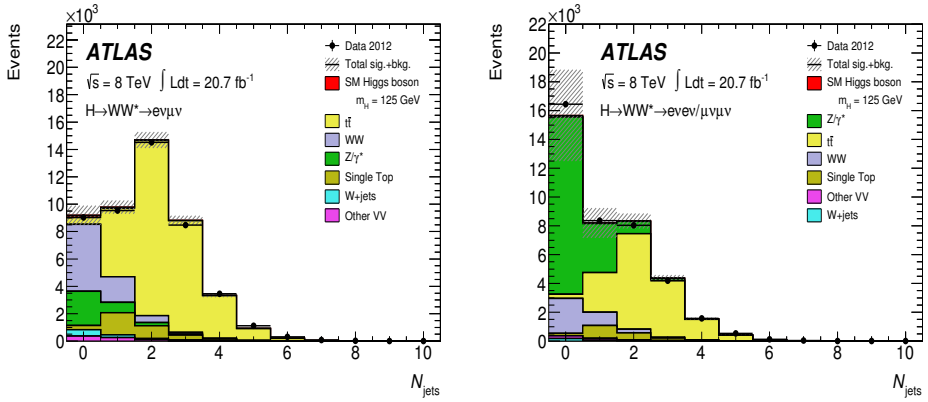

FIGURE 5.14: N_{jets} distributions for $e\mu$+μe (left) and ee+$\mu\mu$ (right) from 8 TeV data after pre-selection and $\not{E}_T$ requirements: $E_{\text{T,rel}}^{\text{miss}} > 25\,\text{GeV}$ and $> 45\,\text{GeV}$, respectively. The SM background expectation from the MC simulation is also shown. The hatched area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

In light of this, to optimise the $H \to WW^{(*)} \to \ell\nu\ell\nu$ selection specific requirements are considered in each jet category. These are evaluated depending on the

kinematic differences between the Higgs boson signal and the significant backgrounds in each case.

5.7.2 H+0j Specific Selection

The H+0j category selects events without reconstructed jets satisfying the selection defined in Section 5.6.2. The jet veto removes most of the top quark backgrounds that survive after the $E_{\mathrm{T,rel}}^{\mathrm{miss}}$ requirement. As observed in Fig. 5.14, events without jets are dominated by WW and Drell-Yan backgrounds for the $e\mu$+μe and ee+$\mu\mu$ final states, respectively. First, a cut on the opening angle between the dilepton system and the $E_{\mathrm{T}}^{\mathrm{miss}}$ direction ($\Delta\phi_{ll,E_{\mathrm{T}}^{\mathrm{miss}}}$) is required to be greater than $\pi/2$. This helps to remove pathological events in which the $E_{\mathrm{T}}^{\mathrm{miss}}$ is pointing in the direction of the lepton pair. Figure 5.15 shows that these are mostly Drell-Yan events. Moreover, the $\Delta\phi_{ll,E_{\mathrm{T}}^{\mathrm{miss}}}$ requirement has a very low cost in signal acceptance.

In the next selection stage, the transverse momentum of the dilepton system is required to be $p_{\mathrm{T}}^{\ell\ell} > 30\,\mathrm{GeV}$. This criterion mainly rejects Drell-Yan contribution since it populates the low-$p_{\mathrm{T}}^{\ell\ell}$ region, as observed in Fig. 5.16 for $e\mu$+μe and ee+$\mu\mu$ final states, separately.

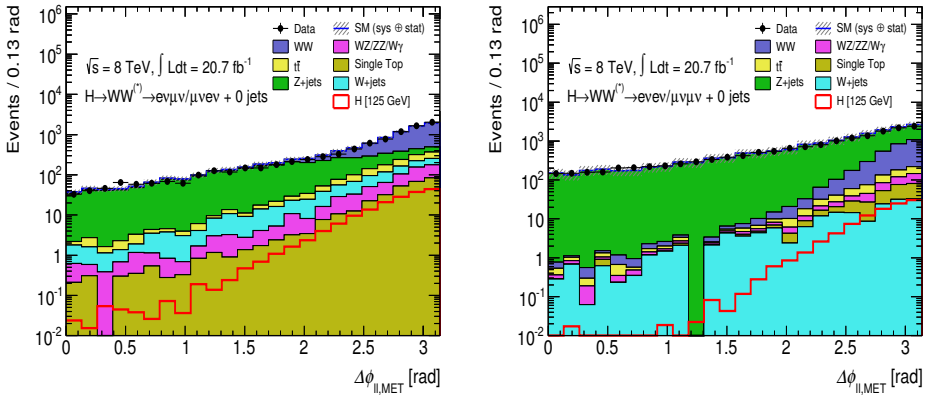

FIGURE 5.15: $\Delta\phi_{ll,E_{\mathrm{T}}^{\mathrm{miss}}}$ distributions for H+0j events in 8 TeV data. The plots are shown for $e\mu$+μe, on the left, and ee+$\mu\mu$, on the right, after the jet veto. The background expectation from the MC simulation is also shown. The hatched area represents the total uncertainty on the sum of the signal and background yields from statistical, experimental, and theoretical sources. The expected Higgs signal, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, is represented as a red curve.

After applying the $p_{\mathrm{T}}^{\ell\ell}$ cut, the Drell-Yan contamination is reduced by a factor ~ 15 and the WW background represents $\sim 70\%$ of the total background in $e\mu$+μe final states. For ee+$\mu\mu$, Z/DY processes are reduced to the level of the expected

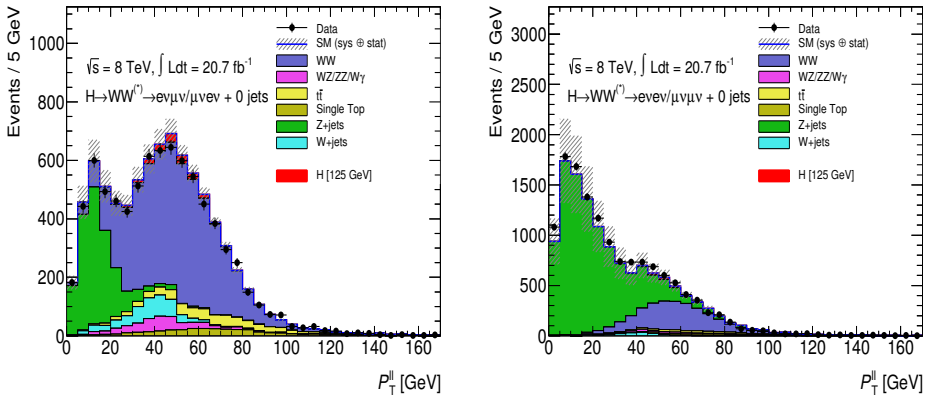

FIGURE 5.16: $p_{\mathrm{T}}^{\ell\ell}$ distributions for H+0j events in 8 TeV data. The plots are shown for $e\mu$+μe on the left and ee+$\mu\mu$ on the right after the requirement on $\Delta\phi_{ll,E_{\mathrm{T}}^{\mathrm{miss}}}$. The hatched area represents the total uncertainty on the sum of the signal and background yields from statistical, experimental, and theoretical sources. The expected Higgs signal ($m_{\mathrm{H}} = 125$ GeV) is stacked on top of the SM backgrounds and it is represented by the red area.

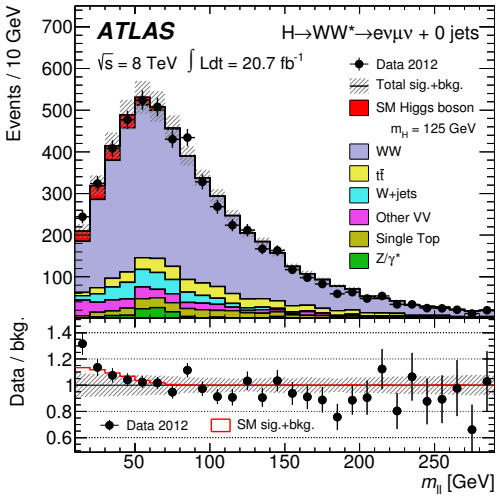

FIGURE 5.17: The $m_{\ell\ell}$ distribution for $e\mu$+μe events with $N_{\mathrm{jets}} = 0$ in 8 TeV data. The events with $m_{\ell\ell} < 50$ GeV correspond to the signal region except that the $\Delta\phi_{\ell\ell} < 1.8$ requirement is not applied here. The Higgs signal is stacked on top of the SM background. The hatched area represents the total uncertainty on the sum of the signal and background yields from statistical, experimental, and theoretical sources. The lower part of the left plot shows the ratio of the data to the predicted SM background. For comparison, the expected ratio of the signal plus background to the background alone is also shown as a red curve.

WW background. Then, a set of topological selections takes advantage of the configuration of the two leptons. The leptons emerge in the same direction due to the spin correlations of $H \to WW^{(*)}$ decay and the $V-A$ structure of the W decay. The invariant mass of the dilepton system ($m_{\ell\ell}$) is required to be small ($< 50\,\text{GeV}$). The $m_{\ell\ell}$ distribution is shown in Fig. 5.17 for the $H{+}0j$ category with $e\mu{+}\mu e$ final states. It is observed that the Higgs expected signal, with a mass of 125 GeV, is located in the low-$m_{\ell\ell}$ region, whereas approximately half of the WW contribution forms a long tail in the high-$m_{\ell\ell}$ excluded region.

Lastly, the azimuthal separation between the two leptons ($\Delta\phi_{\ell\ell}$) is required to be $\Delta\phi_{\ell\ell} < 1.8$ radians. The $\Delta\phi_{\ell\ell}$ distributions are shown in Fig. 5.18 for $e\mu{+}\mu e$ and $ee{+}\mu\mu$, separately. Moreover, in $ee{+}\mu\mu$ final states, additional cuts on $f_{\text{recoil}} < 0.05$ and $E_{\text{T,Rel}}^{\text{miss,track}} > 45\,\text{GeV}$ are applied to further reduce the Drell-Yan background as was discussed in Section 5.6.3. Table 5.10 and Tab. 5.11 show the expected number of signal ($m_{\text{H}} = 125\,\text{GeV}$) and background events after applying each cut in the $H{+}0j$ analysis, for an integrated luminosity of $20.7\,\text{fb}^{-1}$, for the $e\mu{+}\mu e$ and $ee{+}\mu\mu$ final states, respectively.

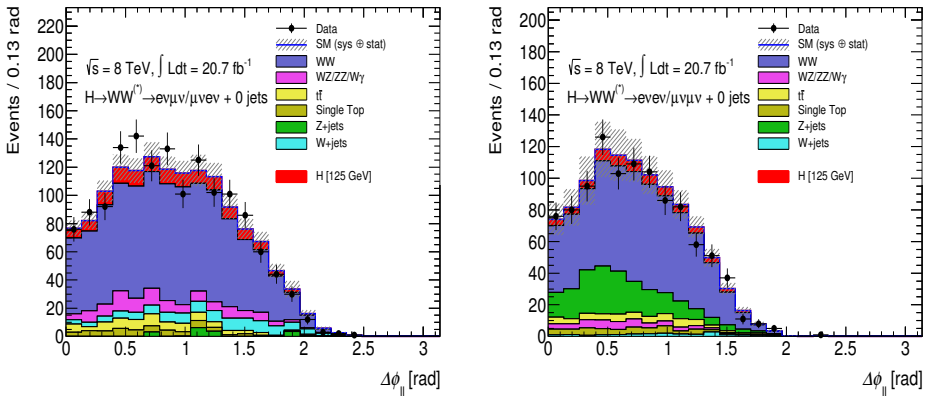

FIGURE 5.18: $\Delta\phi_{\ell\ell}$ distributions for $H{+}0j$ events in 8 TeV data. The plots are shown for $e\mu{+}\mu e$ on the left and $ee{+}\mu\mu$ on the right after the requirement on $m_{\ell\ell}$. Note that for the $ee{+}\mu\mu$ final states, the $E_{\text{T}}^{\text{miss,track}} > 45\,\text{GeV}$ cut is applied at this stage. The hatched area represents the total uncertainty on the sum of the signal and background yields from statistical, experimental, and theoretical sources. The expected Higgs signal ($m_{\text{H}} = 125\,\text{GeV}$) is stacked on top of the SM backgrounds and it is represented by the red area.

At the end of the $H{+}0j$ selection, nearly the entire background is composed of WW events. In $e\mu{+}\mu e$ case the second and third most contributing processes are non-WW diboson and $W+$jets, respectively. In the $ee{+}\mu\mu$, Drell-Yan events represent the second main contamination, whereas $W+$jets is the smallest.

TABLE 5.10: Event yields for the H+0j category in 8 TeV data for $e\mu$+μe final states. The observed and expected yields for the Higgs signal and SM background are shown. The composition of the total background is given on the right. Energies, masses, and momenta are in units of GeV. All uncertainties are statistical.

	Observed	Total bkg.	Signal [$m_H = 125$]	WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets
$E_{T,Rel}^{miss}$>25	46526	46660 ± 80	327.6 ± 2.2	7432 ± 23	778 ± 21	30700 ± 60	3360 ± 25	3356 ± 40	840 ± 10
$N_{jets} = 0$	9024	9000 ± 40	171.7 ± 1.6	4901 ± 19	368 ± 12	506 ± 8	310 ± 7	2440 ± 32	470 ± 7
$\Delta\phi_{ll,E_T^{miss}} > \frac{\pi}{2}$	8100	8120 ± 40	170.4 ± 1.6	4839 ± 19	356 ± 12	491 ± 8	305 ± 8	1687 ± 29	437 ± 6
$p_T^{\ell\ell} > 30$	5497	5485 ± 27	156.0 ± 1.5	4048 ± 17	288 ± 12	450 ± 8	280 ± 8	100 ± 14	319 ± 5
$m_{\ell\ell} < 50$	1453	1308 ± 14	124.0 ± 1.3	964 ± 8	110 ± 6	68.5 ± 3.2	45.5 ± 2.8	18 ± 7	101.5 ± 2.4
$\Delta\phi_{\ell\ell} < 1.8$	1399	1244 ± 13	119.2 ± 1.3	925 ± 8	107 ± 6	67.2 ± 3.1	44.4 ± 2.8	13 ± 7	87.5 ± 2.3

TABLE 5.11: Event yields for the H+0j category in 8 TeV data for ee+$\mu\mu$ final states. The observed and expected yields for the Higgs signal and SM background are shown. The composition of the total background is given on the right. Energies, masses, and momenta are in units of GeV. All uncertainties are statistical.

	Observed	Total bkg.	Signal [$m_H = 125$]	WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets
$E_{T,Rel}^{miss}$>45	39330	38430 ± 190	189.2 ± 1.7	3691 ± 16	404 ± 11	15540 ± 50	1776 ± 18	16710 ± 190	306 ± 14
$N_{jets} = 0$	16446	15550 ± 160	103.7 ± 1.2	2436 ± 13	191 ± 5	281 ± 6	175 ± 6	12300 ± 160	172 ± 11
$\Delta\phi_{ll,E_T^{miss}} > \frac{\pi}{2}$	13697	12970 ± 140	103.5 ± 1.2	2431 ± 13	188 ± 5	277 ± 6	174 ± 6	9740 ± 140	161 ± 10
$p_T^{\ell\ell} > 45$	5670	5650 ± 70	99.3 ± 1.2	2300 ± 13	172 ± 5	264 ± 6	167 ± 5	2610 ± 70	133.9 ± 3.6
$m_{\ell\ell} < 50$	2314	2393 ± 22	84.0 ± 1.1	759 ± 7	64.1 ± 2.8	53.4 ± 2.9	42.2 ± 2.7	1412 ± 20	62.1 ± 2.5
$E_{T,Rel}^{miss,track}$>45	1032	993 ± 10	62.9 ± 0.9	646 ± 7	41.5 ± 2.0	46.7 ± 2.7	38.8 ± 2.5	200 ± 5	19.4 ± 1.5
$\Delta\phi_{\ell\ell} < 1.8$	1026	983 ± 10	62.5 ± 0.9	644 ± 7	41.5 ± 2.0	46.0 ± 2.7	38.8 ± 2.5	195 ± 5	18.4 ± 1.5
$f_{recoil} < 0.05$	671	647 ± 7	41.8 ± 0.8	515 ± 6	29.8 ± 1.6	19.4 ± 1.8	21.8 ± 1.9	48.6 ± 2.6	12.0 ± 1.3

5.7.3 H+1j Specific Selection

The H+1j category requires events with exactly one jet. These events are mainly dominated by top quark events, as it can be seen in Fig. 5.14. The top quark contamination represents about 70% of the total background in the H+1j analysis

for $e\mu$+μe final states. In the ee+$\mu\mu$ case, the contribution from top decay processes is comparable to the expected Drell-Yan yield. To suppress the $t\bar{t}$ contamination, jets tagged as b-jets are vetoed ($N_{b\text{-jet}} = 0$). This b-jet veto achieves a reduction in the $t\bar{t}$ yield by a factor ~ 4 in both $e\mu$+μe and ee+$\mu\mu$ final states. In order to reject $Z \to \tau\tau$ events in $e\mu$+μe final states, a requirement on the $\tau\tau$ invariant mass ($m_{\tau\tau}$) is required; this is similar to the $Z \to \ell\ell$ veto applied in the pre-selection for ee+$\mu\mu$ final states. The $m_{\tau\tau}$ is reconstructed using the collinear approximation [122] which assumes that the neutrinos are collinear with the leptons in the τ decay and, that they are the only source of $\not{E}_T$. If the mass of the $\tau\tau$ system is consistent with the Z boson mass; i.e., satisfying $|m_{\tau\tau} - m_Z| < 25\,\text{GeV}$, the event is rejected.

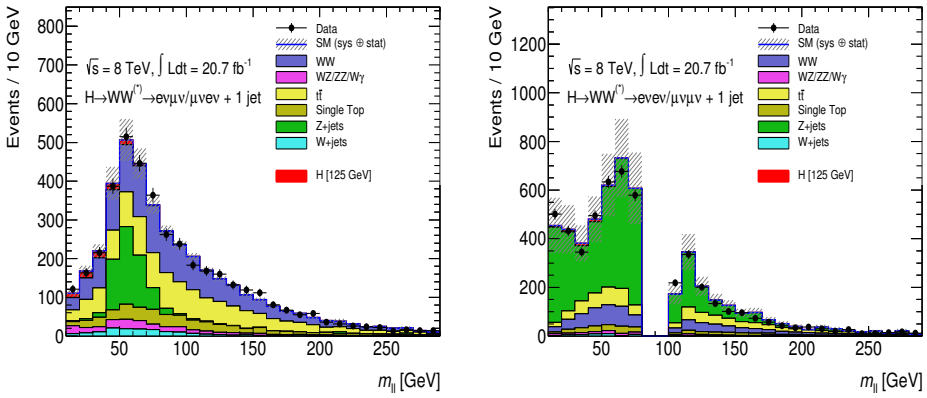

FIGURE 5.19: $m_{\ell\ell}$ distributions for H+1j events in 8 TeV data. The plots are shown for $e\mu$+μe, on the left, and ee+$\mu\mu$, on the right, after the b-jet and $Z \to \tau\tau$ vetoes. The hatched area represents the total uncertainty on the sum of the signal and background yields from statistical, experimental, and theoretical sources. The expected Higgs boson signal, $m_{\text{H}} = 125\,\text{GeV}$, is stacked on top of the SM backgrounds and it is represented by the red area.

Lastly, H+1j final candidates must pass the requirements on $m_{\ell\ell} < 50\,\text{GeV}$ and $\Delta\phi_{\ell\ell} < 1.8$, as they are described for the H+0j category in Section 5.7.2. Figure 5.19 shows the $m_{\ell\ell}$ distributions after the b-jet and $Z \to \tau\tau$ vetoes. The $\Delta\phi_{\ell\ell}$ distributions are represented after the upper bound on $m_{\ell\ell}$. Note that to further reduce the Z/DY contribution in ee+$\mu\mu$ final states, extra requirements are applied on $E_{\text{T,Rel}}^{\text{miss,track}} > 45\,\text{GeV}$ and $f_{\text{recoil}} < 0.2$.

Table 5.12 and 5.13 show the observed 8 TeV data events and the expected yields for the Higgs boson signal, with $m_{\text{H}} = 125\,\text{GeV}$, and SM background applying the corresponding H+1j requirements for ee+$\mu\mu$ and $e\mu$+μe final states, respectively.

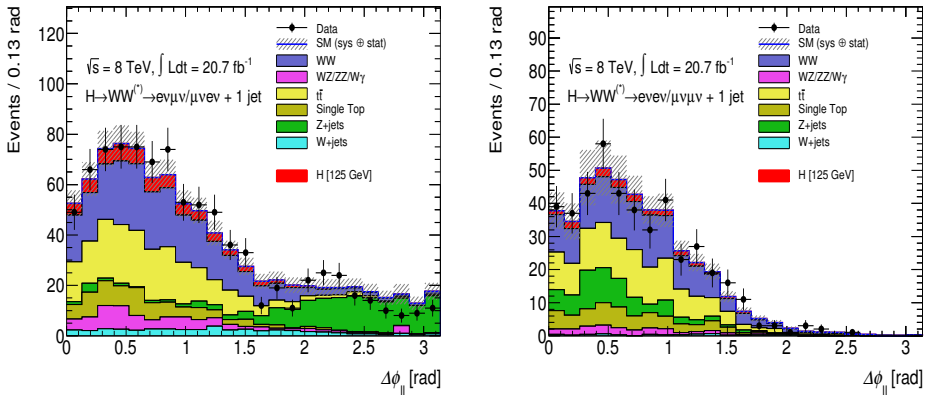

FIGURE 5.20: $\Delta\phi_{\ell\ell}$ distributions for H+1j events in 8 TeV data. The plots are shown for $e\mu$+μe on the left and ee+$\mu\mu$ on the right after the requirement on $m_{\ell\ell}$. Note that for the ee+$\mu\mu$ final states, the $E_{\mathrm{T}}^{\mathrm{miss,track}} > 45$ GeV cut is applied at this stage. The hatched area represents the total uncertainty on the sum of the signal and background yields from statistical, experimental, and theoretical sources. The expected Higgs signal, $m_{\mathrm{H}} = 125$ GeV, is stacked on top of the SM backgrounds and it is represented by the red area.

TABLE 5.12: Event yields for the H+1j category in 8 TeV data for ee+$\mu\mu$ final states. The observed and expected yields for the Higgs signal and SM background are shown. The composition of the total background is given on the right. Energies, masses, and momenta are in units of GeV. All uncertainties are statistical.

	Observed	Total bkg.	Signal ($m_{\mathrm{H}} = 125$)	WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets
$E_{\mathrm{T,Rel}}^{\mathrm{miss}}$>45	39330	38430 ± 190	189.2 ± 1.7	3691 ± 16	404 ± 11	15540 ± 50	1776 ± 18	16710 ± 190	306 ± 14
$N_{\mathrm{jets}} = 1$	8354	8110 ± 80	54.3 ± 1.0	820 ± 7	137 ± 8	2744 ± 19	885 ± 13	3470 ± 80	60 ± 9
$N_{b\text{-jet}} = 0$	5192	4800 ± 70	47.7 ± 0.9	723 ± 7	120 ± 8	719 ± 10	256 ± 7	2930 ± 70	44 ± 8
$m_{\ell\ell} < 50$	1773	1537 ± 16	37.6 ± 0.7	194.5 ± 3.5	34.8 ± 1.7	166 ± 5	64.9 ± 3.4	1057 ± 14	20.1 ± 1.5
$E_{\mathrm{T,Rel}}^{\mathrm{miss,track}} > 45$	440	418 ± 7	21.4 ± 0.6	148.1 ± 3.0	20.6 ± 1.3	128 ± 5	51.9 ± 3.1	64 ± 4	5.1 ± 0.8
$\Delta\phi_{\ell\ell} < 1.8$	430	407 ± 7	20.4 ± 0.5	143.2 ± 3.0	19.9 ± 1.2	126 ± 4	51.0 ± 3.1	63 ± 4	4.5 ± 0.7
$f_{\mathrm{recoil}} < 0.2$	346	315 ± 6	16.2 ± 0.5	128.4 ± 2.8	17.4 ± 1.2	97 ± 4	44.3 ± 2.9	25.1 ± 2.1	3.1 ± 0.6

5.7.4 H+2j Specific Selection

The H+2j category requires at least two jets in the final state. After applying the $N_{\mathrm{jets}} \geq 2$ requirement, nearly the entire total background comes from top quark and

TABLE 5.13: Event yields for the H+1j category in 8 TeV data for $e\mu$+μe final states. The observed and expected yields for the Higgs signal and SM background are shown. The composition of the total background is given on the right. Energies, masses, and momenta are in units of GeV. All uncertainties are statistical.

	Observed	Total bkg.	Signal [$m_H = 125$]	WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets
$E_{\mathrm{T,Rel}}^{\mathrm{miss}}$>25	46526	46660 ± 80	327.6 ± 2.2	7432 ± 23	778 ± 21	30700 ± 60	3360 ± 25	3356 ± 40	840 ± 10
$N_{\mathrm{jets}} = 1$	9527	9460 ± 40	96.7 ± 1.2	1656 ± 10	265 ± 12	4981 ± 25	1604 ± 17	757 ± 15	195 ± 5
$N_{b\text{-jet}} = 0$	4320	4239 ± 25	84.8 ± 1.1	1460 ± 10	224 ± 9	1272 ± 13	457 ± 9	667 ± 14	160 ± 4
$Z \to \tau\tau$ veto	4138	4024 ± 25	84.4 ± 1.1	1417 ± 9	217 ± 9	1220 ± 13	436 ± 9	580 ± 14	155 ± 4
$m_{\ell\ell} < 50$	886	829 ± 11	63.4 ± 0.9	269 ± 4	69 ± 5	216 ± 6	79 ± 4	149 ± 5	45.6 ± 2.3
$\Delta\phi_{\ell\ell} < 1.8$	728	650 ± 9	58.8 ± 0.9	247 ± 4	60.5 ± 3.5	204 ± 6	76 ± 4	27.8 ± 3.3	34.5 ± 2.0

Z/DY processes in $e\mu$+μe and ee+$\mu\mu$ final states, respectively. In order to deal with such background contaminations, the b-veto is applied in addition to the Z/DY $\to \tau\tau$ veto, as is done in the H+1j category with $e\mu$+μe final states. Further $t\bar{t}$ rejection is obtained by requiring a small total transverse momentum ($p_{\mathrm{T}}^{\mathrm{tot}} < 45\,\mathrm{GeV}$). The $p_{\mathrm{T}}^{\mathrm{tot}}$ is the magnitude of the following vectorial sum,

$$\vec{p_{\mathrm{T}}}^{\,\mathrm{tot}} = \vec{p}_{\mathrm{T}}^{\,\ell\ell} + \vec{p}_{\mathrm{T}}^{\,jj} + \vec{E}_{\mathrm{T}}^{\mathrm{miss}}\,, \tag{5.6}$$

where $\vec{p}_{\mathrm{T}}^{\,jj}$ is the vectorial sum of the transverse momenta of the two highest-p_{T} jets. These will be referred to as tagged jets in the following.

After the above selection, a set of VBF-specific topological variables are exploited in order to enhance the Higgs boson signal significance. The VBF process is characterised by the kinematics of the pair of tagged jets and the activity in the rapidity gap between them. For the former, the invariant mass of the tagged jets (m_{jj}) and their rapidity gap separation (Δy_{jj}) are used. The rapidity gap of the two tagged jets is defined as a positive quantity: $\Delta y_{jj} \equiv |y_{j_1} - y_{j_2}|$. The events are required to satisfy $\Delta y_{jj} > 2.8$ and $m_{jj} > 500\,\mathrm{GeV}$. Figure 5.21 and 5.22 show the Δy_{jj} and m_{jj} distributions, respectively.

The next two selections rely on the activity within the rapidity gap between the tagged jets. There is a relatively low level of hadronic activity expected for the VBF signal since weak bosons do not exchange colour. The number of extra jets in this gap quantifies such activity, so requiring the absence of jets in this region

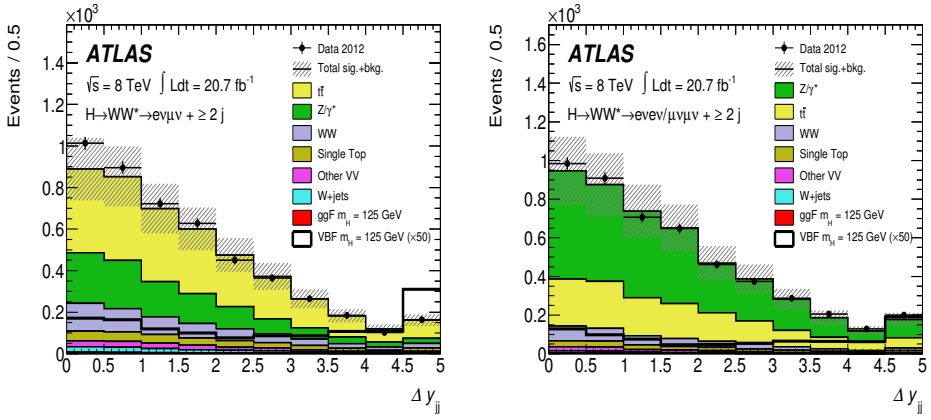

FIGURE 5.21: The Δy_{jj} distributions for the H+2j analysis category in 8 TeV data, after the b-jet veto and $p_{\mathrm{T}}^{\mathrm{tot}} < 45$ GeV requirements with $e\mu$+μe and ee+$\mu\mu$ final states. The expected SM background contributions are shown. The VBF signal magnified by a factor of 50 is represented by the black curve. The hatched area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

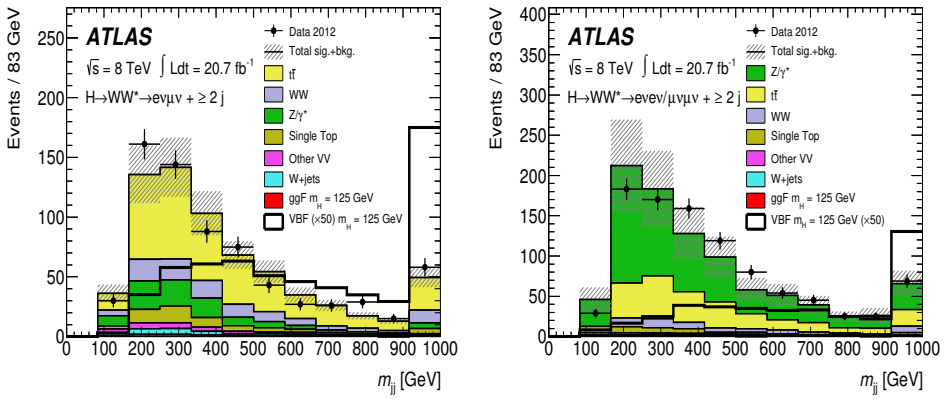

FIGURE 5.22: The m_{jj} distributions for the H+2j analysis category in 8 TeV data, after the $|\Delta y_{jj}| > 2.8$ requirement with $e\mu$+μe and ee+$\mu\mu$ final states. The expected SM background contributions are shown. The VBF signal magnified by a factor of 50 is represented by the black curve. The hatched area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

suppresses processes where the jets are produced via QCD radiation. In addition, the Higgs boson decay products tend to be in this central region because of the high boost transferred to them. In this light, the VBF topological requirements include both an outside lepton and a central jet activity requirements. The first is known as the central jet veto (CJV). This rejects events containing any jet with $p_{\mathrm{T}} > 20$ GeV

inside the rapidity gap defined by the tagged jets. This results in a reduction of events produced via ggF process.

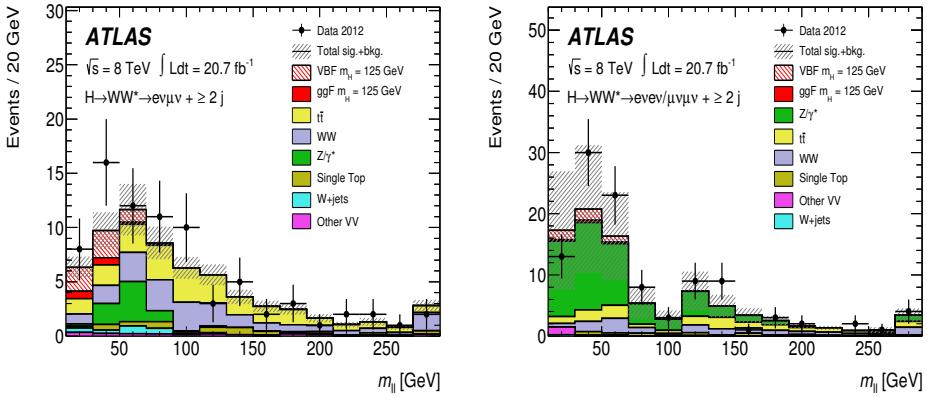

FIGURE 5.23: The $m_{\ell\ell}$ distributions for the H+2j analysis category in 8 TeV data, after the OLV requirement (which accepts only events with both leptons between the two tagging jets) with $e\mu$+μe and ee+$\mu\mu$ final states. The expected SM background contributions are shown. The hatched area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

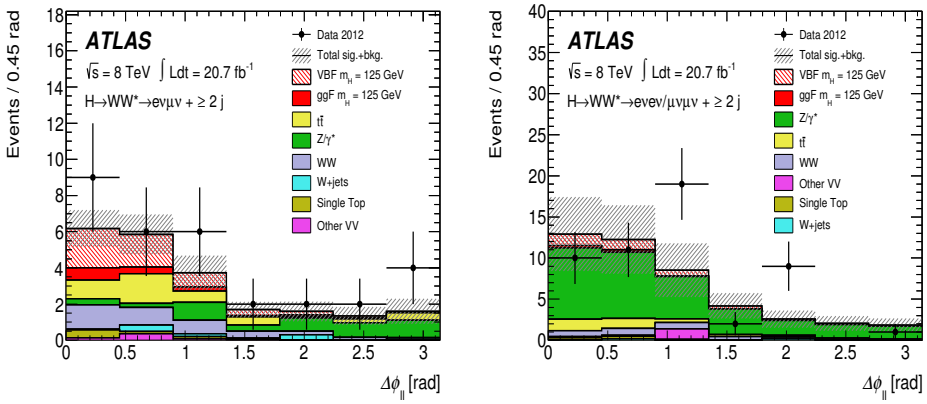

FIGURE 5.24: The $\Delta\phi_{\ell\ell}$ distributions for H+2j analysis category in 8 TeV data, after the $m_{\ell\ell} < 60$ GeV selection with $e\mu$+μe and ee+$\mu\mu$ final states. The expected SM background contributions are shown. The hatched area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

After the CJV criterion the contribution of $t\bar{t}$ is not dominant and is reduced to the level of WW in $e\mu$+μe final states. The second requirement is the so-called outside lepton veto (OLV). This requires the lepton to be within the rapidity gap

between the tagged jets. These two jet activity vetoes achieve a factor ∼3 reduction of the total background, whereas the VBF Higgs signal is kept at the same level. Finally, requirements on $m_{\ell\ell} < 60\,\text{GeV}$ and $\Delta\phi_{\ell\ell} < 1.8$ are applied for the H+0j and H+1j categories. The $m_{\ell\ell}$ and $\Delta\phi_{\ell\ell}$ distributions are shown in Fig. 5.23 and Fig. 5.24, respectively.

Table 5.14 and Tab. 5.15 show the expected number of VBF signal ($m_{\text{H}} = 125\,\text{GeV}$) and background events after applying each cut in the H+2j analysis, using a total integrated luminosity of $20.7\,\text{fb}^{-1}$, for the $e\mu$+μe and ee+$\mu\mu$ final states, respectively.

TABLE 5.14: Event yields for the H+2j category in 8 TeV data with $e\mu$+μe final states. The observed and expected yields for the Higgs signal and SM background are shown. In this table, the ggF Higgs boson signal is treated as background and so, included in "Total bkg." column. The column tagged as VBF includes the Higgs boson contribution from VH production mechanism, which is negligible after the VBF-related criteria. The composition of the total background is given on the right. Energies, masses, and momenta are in units of GeV. All uncertainties are statistical.

	Observed	Total bkg.	Signal $m_{\text{H}} = 125\,\text{GeV}$		WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets
			VBF	ggF						
$E_{\text{T}}^{\text{miss}}$>20	84324	83810 ± 130	77.5 ± 1.4	376.5 ± 2.2	8801 ± 23	1292 ± 26	49910 ± 80	5051 ± 31	16570 ± 80	1766 ± 16
$N_{\text{jets}} \geq 2$	48723	47740 ± 80	42.6 ± 0.9	66.9 ± 0.9	939 ± 7	299 ± 15	41850 ± 70	2368 ± 21	1811 ± 25	437 ± 8
$N_{b\text{-jet}} = 0$	5852	5697 ± 32	30.6 ± 0.7	49.1 ± 0.8	685 ± 6	202 ± 7	2932 ± 21	351 ± 8	1310 ± 18	171 ± 5
$\mathbf{p}_{\text{T}}^{\text{tot}} < 45$	4790	4615 ± 29	26.7 ± 0.7	40.8 ± 0.7	591 ± 6	155 ± 7	2319 ± 18	287 ± 8	1168 ± 17	126 ± 4
$Z \to \tau\tau$ veto	4007	3846 ± 26	24.5 ± 0.7	38.0 ± 0.7	544 ± 6	141 ± 7	2148 ± 18	264 ± 7	673 ± 14	108 ± 4
$\lvert\Delta y_{jj}\rvert > 2.8$	696	678 ± 10	11.9 ± 0.23	9.50 ± 0.34	100.0 ± 2.3	24.8 ± 3.4	377 ± 7	55.1 ± 3.1	95 ± 5	19 ± 2
$m_{jj} > 500$	198	170 ± 4	7.54 ± 0.12	2.93 ± 0.19	33.7 ± 1.2	5.6 ± 0.6	93.4 ± 3.0	11.4 ± 1.2	18.9 ± 2.5	4.4 ± 0.7
CJV	92	77.6 ± 2.4	6.30 ± 0.11	1.74 ± 0.15	25.5 ± 1.0	2.8 ± 0.4	30.2 ± 1.5	5.3 ± 0.8	9.3 ± 1.2	3.1 ± 0.6
OLV	78	58.5 ± 2.1	6.07 ± 0.10	1.57 ± 0.14	18.7 ± 0.8	2.05 ± 0.32	22.5 ± 1.3	4.3 ± 0.7	7.3 ± 1.2	2.4 ± 0.5
$m_{\ell\ell} < 60$	31	16.4 ± 1.3	5.49 ± 0.10	1.48 ± 0.14	3.8 ± 0.4	0.66 ± 0.21	4.48 ± 0.69	0.70 ± 0.31	4.4 ± 0.8	1.0 ± 0.4
$\Delta\phi_{\ell\ell} < 1.8$	23	12.3 ± 1.0	5.11 ± 0.09	1.34 ± 0.13	3.5 ± 0.4	0.63 ± 0.21	3.7 ± 0.7	0.70 ± 0.31	1.9 ± 0.5	0.56 ± 0.30

5.7.5 Selection Summary and Signal Discriminant Variables

All the event selection requirements in each of the jet categories H+0j, H+1j and H+2j have been presented in Section 5.7.2, 5.7.3 and 5.7.4, respectively. A

TABLE 5.15: Event yields for the H+2j category in 8 TeV data with ee+$\mu\mu$ final states. The observed and expected yields for the Higgs signal and SM background are shown. In this table, the ggF Higgs boson signal is treated as background and so, included in "Total bkg." column. The column tagged as VBF includes the Higgs boson contribution from VH production mechanism, which is negligible after the VBF-related criteria. The composition of the total background is given on the right. Energies, masses, and momenta are in units of GeV. All uncertainties are statistical.

	Observed	Total bkg.	Signal [$m_{\mathrm{H}}=125$]		WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets
			VBF	ggF						
$E_{\mathrm{T}}^{\mathrm{miss}}$>45 $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$>35	58690	56930 ± 210	45.5 ± 1	198.8 ± 2	3924 ± 20	604 ± 10	29300 ± 60	2863 ± 20	19620 ± 190	463 ± 16
$N_{\mathrm{jets}} \geq 2$	32877	32170 ± 110	26.4 ± 0.7	39.7 ± 0.7	537 ± 6	186 ± 9	24540 ± 60	1388 ± 16	5420 ± 90	191 ± 7
$N_{b\text{-jet}} = 0$	65388	6370 ± 80	18.9 ± 0.6	29.6 ± 0.6	394 ± 5	129 ± 7	1747 ± 16	203 ± 6	3810 ± 80	58 ± 4
$p_{\mathrm{T}}^{tot} < 45$	4903	4830 ± 70	16.7 ± 0.5	24.4 ± 0.6	336 ± 4	93 ± 5	1375 ± 14	171 ± 6	2790 ± 70	42.9 ± 2.9
$\mid \Delta y_{jj} \mid > 2.8$	958	926 ± 33	8.06 ± 0.20	6.23 ± 0.28	61.1 ± 1.7	12.7 ± 1.3	253 ± 6	35.3 ± 2.4	552 ± 33	6.2 ± 1.1
$m_{jj} > 500$	298	245 ± 6	5.55 ± 0.10	2.10 ± 0.16	23.5 ± 1.0	4.1 ± 1.1	62.4 ± 2.5	9.3 ± 1.1	139 ± 5	1.4 ± 0.6
CJV	147	119 ± 4	4.65 ± 0.10	1.10 ± 0.12	16.6 ± 0.8	2.8 ± 1.1	19.3 ± 1.3	4.1 ± 0.7	72.6 ± 3.4	0.68 ± 0.38
OLV	108	82.7 ± 3.3	4.45 ± 0.09	0.93 ± 0.11	12.5 ± 0.6	2.3 ± 1.1	14.3 ± 1.1	3.1 ± 0.6	49.0 ± 2.7	0.30 ± 0.30
$m_{\ell\ell} < 60$	52	38.7 ± 2.5	4.03 ± 0.09	0.81 ± 0.10	3.23 ± 0.34	1.7 ± 1.1	3.8 ± 0.6	0.80 ± 0.30	28.3 ± 2.1	0.14 ± 0.24
$\Delta\phi_{\ell\ell} < 1.8$	42	33.0 ± 2.4	3.70 ± 0.09	0.72 ± 0.09	2.82 ± 0.32	1.7 ± 1.1	3.3 ± 0.5	0.74 ± 0.30	23.6 ± 2.0	0.06 ± 0.21

summary of the complete $H\to WW^{(*)}\to \ell\nu\ell\nu$ selection strategy can be found in Tab. 5.16. The final signal discriminants $m_{\ell\ell}$ and m_{T}, which appear at the bottom of the table, are introduced at the end to further separate the Higgs boson candidates from the remaining SM background processes. The sample of events at the stage corresponding to the selection on these discriminants variables is discussed in this section.

The $m_{\ell\ell}$ spectrums for the H+0j and H+1j categories are shown in Fig. 5.17 and 5.19, respectively. The signal-to-background ratio (S/B) varies in these distributions for different bins. In light of this, the sample is further subdivided into two regions at $m_{\ell\ell} = 30\,\mathrm{GeV}$ for $N_{\mathrm{jets}} \leq 1$ with $e\mu$+μe final states in order to improve the signal to background ratio.

A transverse mass variable (m_{T}) is used in this analysis to test for the presence of the Higgs boson, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, decaying into a pair of W bosons. The m_{T} quantity is defined as,

TABLE 5.16: Summary of the $H\to WW^{(*)}\to \ell\nu\ell\nu$ event selection criteria divided in each of the jet analysis categories. The criteria specific to $e\mu$+μe and ee+$\mu\mu$ are noted as such; otherwise, they apply to both. Pre-selection applies to all N_{jets} modes. The rapidity gap is the y range spanned by the two leading jets. The $m_{\ell\ell}$ split is at 30 GeV. Energies, masses, and momenta are in units of GeV.

	$H\to WW^{(*)}\to \ell\nu\ell\nu$ Selection		
Pre-selection	Two isolated leptons ($e\,\mu$) with opposite charge Leptons with $p_{\text{T}}^{\text{lead}}>25$ and $p_{\text{T}}^{\text{sublead}}>15$ $e\mu$+μe: $m_{\ell\ell}>10$ ee+$\mu\mu$: $m_{\ell\ell}>12$ and $\mid m_{\ell\ell}-m_Z\mid>15$		
Category	H+0j	H+1j	H+2j
Missing transverse momentum and hadronic recoil	$e\mu$+μe: $E_{\text{T,rel}}^{\text{miss}}>25$ ee+$\mu\mu$: $E_{\text{T,rel}}^{\text{miss}}>45$ ee+$\mu\mu$: $E_{\text{T,Rel}}^{\text{miss,track}}>45$ ee+$\mu\mu$: $f_{\text{recoil}}<0.05$	$e\mu$+μe: $E_{\text{T,rel}}^{\text{miss}}>25$ ee+$\mu\mu$: $E_{\text{T,rel}}^{\text{miss}}>45$ ee+$\mu\mu$: $E_{\text{T,Rel}}^{\text{miss,track}}>45$ ee+$\mu\mu$: $f_{\text{recoil}}<0.2$	$e\mu$+μe: $E_{\text{T}}^{\text{miss}}>20$ ee+$\mu\mu$: $E_{\text{T}}^{\text{miss}}>45$ ee+$\mu\mu$: $E_{\text{T,STVF}}^{\text{miss}}>35$ -
General selection	- $\mid\Delta\phi_{\ell\ell,MET}\mid>\pi/2$ $p_{\text{T}}^{\ell\ell}>30$	$N_{b\text{-jet}}=0$ - $e\mu$+μe: $Z\to\tau\tau$ veto	$N_{b\text{-jet}}=0$ $p_{\text{T}}^{tot}<45$ $e\mu$+μe: $Z\to\tau\tau$ veto
VBF topology	- - - -	- - - -	$m_{jj}>500$ $\mid\Delta y_{jj}\mid>2.8$ CJV OLV
$H\to WW^{(*)}\to \ell\nu\ell\nu$ topology	$m_{\ell\ell}<50$ $\Delta\phi_{\ell\ell}<1.8$ $e\mu$+μe: split $m_{\ell\ell}$ Fit m_{T}	$m_{\ell\ell}<50$ $\Delta\phi_{\ell\ell}<1.8$ $e\mu$+μe: split $m_{\ell\ell}$ Fit m_{T}	$m_{\ell\ell}<60$ $\Delta\phi_{\ell\ell}<1.8$ - Fit m_{T}

$$m_{\text{T}}=\sqrt{(E_{\text{T}}^{\ell\ell}+E_{\text{T}}^{\text{miss}})^2-\mid\vec{p}_{\text{T}}^{\,\ell\ell}+\mathbf{E}_{\text{T}}^{\text{miss}}\mid^2}\,, \tag{5.7}$$

where $E_{\text{T}}^{\ell\ell}=\sqrt{\mid\vec{p}_{\text{T}}^{\,\ell\ell}\mid^2+m_{\ell\ell}^2}$ and $|\vec{p}_{\text{T}}^{\,\ell\ell}|=p_{\text{T}}^{\ell\ell}$.

Figure 5.25 shows the m_{T} distributions with the expected signal and the composition of the SM backgrounds after the full analysis selection criteria. Each category and lepton final states are presented separately. As it can be seen, the background composition and rate change for different m_{T} windows. The highest S/B is in a region of m_{T} around m_{H}: $0.75\,m_{\text{H}}<m_{\text{T}}<m_{\text{H}}$ for $N_{\text{jets}}\leq 1$ and $m_{\text{T}}<1.2\,m_{\text{H}}$ for H+2j. To illustrate the relative size of the signal, the expected S/B in the above-mentioned m_{T} range is 0.14, 0.16, and 0.26 for H+0j, H+1j and, H+2j, respectively, combining all lepton final states.

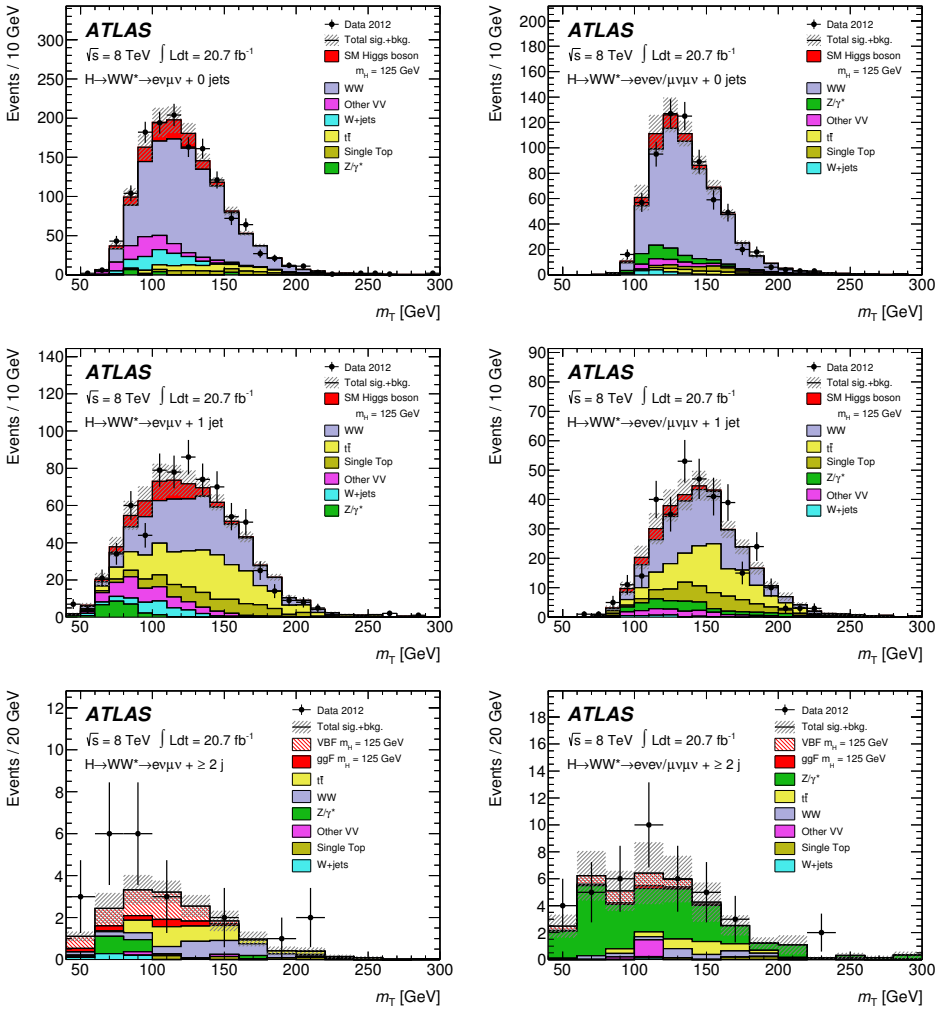

FIGURE 5.25: m_{T} distributions for final candidate events in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis using the 8 TeV data. The plots are shown for the $e\mu$+μe (left) and ee+$\mu\mu$ (right) final states in $N_{\mathrm{jets}} = 0$ (top), $N_{\mathrm{jets}} = 1$ (middle), and $N_{\mathrm{jets}} \geq 2$ final states (bottom). The distributions are shown prior to splitting the samples into two $m_{\ell\ell}$ regions for the $e\mu$+μe final state in the $N_{\mathrm{jets}} = 0$ and $= 1$ cases. The signal is stacked on top of the background. For the $N_{\mathrm{jets}} \geq 2$ analysis, the signal is plotted separately for the ggF and VBF production processes. The hatched area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

Note that the final discriminant depends explicitly on $E_{\mathrm{T}}^{\mathrm{miss}}$, so it is crucial that the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement is well understood. The choice of the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction entering in the m_{T} computation is based on previous studies using the first $5.8\,\mathrm{fb}^{-1}$ data at 8 TeV [123]. Since similar separation power is found in all

three m_{T} reconstruction methods, it is decided to use the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement in the m_{T} computation since $E_{\mathrm{T}}^{\mathrm{miss}}$ is common to all analysis categories (see Tab. 5.8). The performance of different m_{T} reconstruction methods is further investigated in Chapter 6.

Finally, the statistical treatment benefits from the S/B differences between different bins in the $m_{\ell\ell}$ and m_{T} distributions and exploits this feature to better discriminate the signal. The likelihood fit technique is described in detail in Section 5.11.

5.8 Re-analysis of 7 TeV Data

The analysis of the $4.6\,\mathrm{fb}^{-1}$ collected at $\sqrt{s} = 7\,\mathrm{TeV}$ in 2011 was published in Ref. [124]. The purpose of reanalysing this dataset is to apply those improvements that the analysis has acquired throughout 2012, as well as to make the selections and MC samples compatible with the 8 TeV analysis. This simplifies and facilitates the combination of the 2011 and 2012 results. The object and event selections closely follow the requirements described in Section 5.7 used in the analysis of the 8 TeV data. However, due to both technical differences in the data and MC samples as well as different properties in the data-taking conditions, some criteria have been slightly modified to adequate the selection. The differences in the selection and the results obtained re-analysing the data taken during the 2011 period are given in this section.

At the object and pre-selection level, the requirements used in the 2011 reanalysis are almost the same as the ones described in Section 5.6 and 5.5, respectively. Nevertheless, there are two exceptions: the identification algorithm for electrons and the JVF threshold used for jet selection. In the reanalysis of the 7 TeV data the algorithm for the electron identification is non-GSF. Due to the lower pile-up conditions during the 7 TeV period, the JVF selection is required to be above 0.75.

In terms of the $E_{\mathrm{T}}^{\mathrm{miss}}$ criteria, the relevant variables and thresholds for the $e\mu$+μe final state are the same in both 2011 and 2012 analyses. However, in the ee+$\mu\mu$ final state the criteria are loosened for the H+0j and H+1j cases since the level of Drell-Yan contamination is lower in 2011 than in 2012. For the H+0j category with ee+$\mu\mu$ final states, the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ threshold is 35 GeV and the second requirement on $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ is not applied. Instead, the $p_{\mathrm{T}}^{\ell\ell}$ cut is slightly tightened to 40 GeV. In the H+1j category the only requirement on $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ is also lowered to

40 GeV and a requirement on the transverse momentum of the system defined by the two leptons and the jet is added: $p_{\mathrm{T}}^{\ell\ell j} > 35\,\mathrm{GeV}$. The f_{recoil} cut is also used in $ee{+}\mu\mu$ final states following the 2012 criteria: $f_{\mathrm{recoil}} < 0.2$ and < 0.5 are applied to the $H{+}0j$ and $H{+}1j$ categories, respectively. The selection of the $H{+}2j$ analysis for the 7 TeV period completely follows the criteria used in 2012, as is described in Section 5.7.4. The event yields for the 2011 data are collected in Appendix C.

In summary, the 2011 selections that differ from the 2012 analysis are listed below,

- The GSF algorithm is not used for the electron identification.
- The jet vertex fraction for the reconstructed jets is changed: $|JVF| > 0.75$.
- For the $ee{+}\mu\mu$ final states:
  - $E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 35(40)\,\mathrm{GeV}$ for the $H{+}0j$ ($H{+}1j$) analysis and the additional cut on $E_{\mathrm{T,Rel}}^{\mathrm{miss,trk}}$ is not applied.
  - $f_{\mathrm{recoil}} < 0.2$ and < 0.5 for the $H{+}0j$ and $H{+}1j$ analysis, respectively.
  - $p_{\mathrm{T}}^{\ell\ell} > 40\,\mathrm{GeV}$ for the $H{+}0j$ and $p_{\mathrm{T}}^{\ell\ell j} > 35\,\mathrm{GeV}$ for the $H{+}1j$.

The largest improvement with respect to the published analysis in Ref. [124] is a 40% reduction of the $W+$jets contamination. This is accomplished by tightening the lepton isolation requirement following the 2012 analysis. Other improvements between the published results and the re-analysis of the 7 TeV data are achieved by changing some MC generators, specially the WW background. All the background estimation techniques mimic the 2012 analysis, as is described in Section 5.9.

Figure 5.26 shows the expected and observed m_{T} distribution from events satisfying all selection criteria for $H{+}0j$ and $H{+}1j$ categories. These distributions are equivalent to Fig. 5.25 for 8 TeV. The corresponding plots for $H{+}2j$ are omitted due to limited statistics at the end of the event selection, as Tab. C.4 shows. The expected Higgs boson signal, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, is stacked on top of the background spectrum.

5.9 Background Estimation

The search for the Higgs boson in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ decay mode involves a high variety of SM background contamination. This feature forces the analysis to develop validation techniques for the background predictions. The MC evaluation relies on comparisons between the expected yields and the observed rate events in several

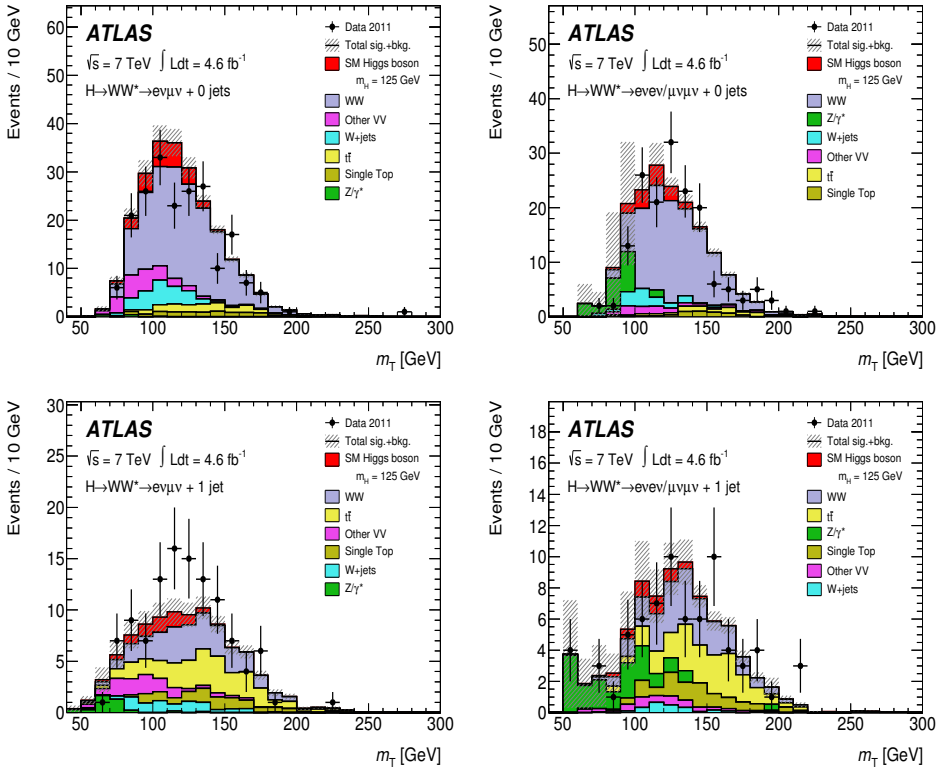

FIGURE 5.26: Distribution of transverse mass for the signal regions of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis of the 7 TeV data with $N_{\text{jets}} = 0$, top, and $N_{\text{jets}} = 1$, bottom. The plots are shown for the $e\mu$+μe, on the left, and ee+$\mu\mu$, on the right. The WW, top-quark, and ZDY backgrounds predicted by MC simulation are scaled using the normalisation from the corresponding CRs described in the text, and the W+jet prediction is from the data-driven estimate. The signal is stacked on top of the background. The hatched area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

enriched samples. Control Regions (CR) and validation regions (VR) are designed to evaluate the MC predictions. Those regions are defined by the compromise of the following statements,

- Be as pure as possible for the relevant background process.
- Preserve as much statistics as possible in order to avoid fluctuations.
- Select Higgs boson signal-depleted regions.

In general, the requirements applied for the control and validation regions are similar to those used in the event selection, but with some criteria reversed or

modified. In this section, the estimation procedures for each of the backgrounds in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis are presented. Moreover, data-driven methods for background shapes and normalization are described, and their performance is discussed in detail. Table 5.17 summarises the treatment of the backgrounds. An overview of the CRs and VRs used for each background source is presented below.

TABLE 5.17: Summary on background treatment in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis. The estimation procedures for various background processes are given in four categories: normalised using a CR; data-derived estimate (Data); normalised using the MC; and normalised using the MC, but validated in a CR (MC + VR). The "($e\mu$+μe)" terms denote that for the ee+$\mu\mu$ final states, in the same N_{jets} mode, the $e\mu$+μe region is used instead, for reasons of purity and/or statistics. The "(merged)" terms indicate that the fully combined $e\mu$+μe + ee+$\mu\mu$ CR is used for all lepton final states.

	WW	Top	$Z/\text{DY}\to\tau\tau$	$Z/\gamma^*\to\ell\ell$	W+jets	VV
H+0j						
$e\mu$+μe	CR	CR	CR	MC	Data	MC + VR
ee+$\mu\mu$	CR ($e\mu$+μe)	CR ($e\mu$+μe)	CR ($e\mu$+μe)	Data	Data	MC + VR
H+1j						
$e\mu$+μe	CR	CR	CR	MC	Data	MC + VR
ee+$\mu\mu$	CR ($e\mu$+μe)	CR ($e\mu$+μe)	CR ($e\mu$+μe)	Data	Data	MC + VR
H+2j						
$e\mu$+μe	MC	CR (merged)	CR	MC	Data	MC
ee+$\mu\mu$	MC	CR (merged)	CR ($e\mu$+μe)	Data	Data	MC

5.9.1 Background Overview

As shown in previous sections, there is a large number of SM processes whose final decay states may be reconstructed as two high-p_{T} leptons with a high $\not\!\!E_T$ measurement. At the LHC, the main background contributions to the search for the Higgs boson in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis are WW and top quark production. These processes contain two isolated high-p_{T} leptons, from the W bosons, in the final state. In such processes, the measurement of the $\not\!\!E_T$ is real due to the presence of high-p_{T} neutrinos as well. The MC WW background in both the H+0j and H+1j analyses and the MC top background in the H+1j and H+2j categories are normalised to the data yields in the corresponding CRs, after subtracting contributions from processes other than the targeted one. The resulting estimated event yield for that process is extrapolated from the CR to the signal region: the value from the normalization is used as a correction factor to the MC. In general, the $e\mu$+μe CRs, with higher statistics and significantly higher purity, are used as CRs for the corresponding ee+$\mu\mu$ final states. The only exception is the top background

in the $H+2j$ analysis, in which all lepton final states are combined since the Drell-Yan contribution is negligible in this CR. Details on the normalisation of the top and WW backgrounds are addressed in Sections 5.9.6 and 5.9.7, respectively.

For the $W+$jets background case, one of the jet decay products can be mis-reconstructed as the second high-p_{T} lepton. Only a small fraction of those events satisfy the full event selection, besides the large cross section of the $W+$jets process. However, the m_{T} shape for the $W+$jets and multi-jet backgrounds is very similar to the Higgs boson, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, and the final fit can not constrain this contamination in the signal region. Hence, it is crucial to have under control the $W+$jets background. This type of process is difficult to model reliably with simulation. In light of this, this is estimated from data for all jet multiplicities, as described in Section 5.9.2.

Backgrounds from diboson processes other than WW also have signal-like kinematics. These backgrounds add up to 10% of the total predicted background in the $H+0j$ and $H+1j$ categories and are of the same magnitude as the signal. The shapes and normalisation of the non-WW diboson processes are estimated using simulation, as described in Section 5.9.3. These processes produce same-charge and opposite-charge leptons pairs, as well as the reconstructed objects from the $W+$jets decay products. The same-charge events which satisfy the full event selection form the VR for the above-mentioned backgrounds. The MC predictions, together with the $W+$jets data-derived estimation, are evaluated using the observed events in this VR. Note that the VR is not used to normalise the backgrounds, as the CR does, but to assess MC generators in a Higgs boson free region.

Finally, the Drell-Yan background is another important contamination source, especially for the $ee+\mu\mu$ final states, which has been shown to increase significantly in 2012 data. For this reason, using the Z peak as CR is subject to large uncertainties from mis-modelling of the $E_{\mathrm{T}}^{\mathrm{miss}}$ tails as a function of $m_{\ell\ell}$. Therefore, data-derived methods are used to estimate the DY background for $ee+\mu\mu$, as described in Section 5.9.4. In the case of $e\mu+\mu e$, where the majority of the background comes from $Z/\mathrm{DY} \to \tau\tau$, a CR is used to normalise the MC prediction to the observed data events. The details are described in Section 5.9.5.

The correlations introduced among the backgrounds by the presence of other processes in the CRs are fully incorporated in the statistical fit procedure used to test the background-only hypothesis and to extract the signal strength. This will be presented later in Section 5.12. In the following, each background estimate is described after those which are relevant for the extrapolation. In light of this, the largest background, WW, is described last.

5.9.2 W+jets Background Estimation

Events in which W bosons are produced in association with jets may become one of the background contaminations in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ search when the jet is reconstructed as a lepton. The rate at which jets are mis-identified as leptons may not be accurately described in the simulation because these events are due to rare fragmentation processes or complex interactions within the detector. In light of this, the $W+$ jets background is estimated entirely from data by defining a phase space enriched with $W+$ jets events.

The $W+$ jets background contribution is estimated using a CR in which one of the two leptons fully satisfies the lepton definition described in Section 5.6, and the other lepton only passes loosened requirements. This selection enhances the rate of jets mis-identified as leptons. The non-$W+$ jets contributions to the $W+$ jets CR are subtracted using MC. The purity of $W+$ jets events in this CR is about 90% in the electron channel and 80% in the muon channel.

The $W+$ jets background in the signal region is obtained from the control sample by applying an extrapolation factor, referred to as the fake factor (f_ℓ) in the following. The fake factor relates the $W+$ jets estimation in the CR to the $W+$ jets contamination in the signal region using di-jet data events. The fake factor is defined as the ratio of the number of objects satisfying the full lepton identification (N_{id}) to those satisfying the loosened criteria; so-called anti-id selection ($N_{anti-id}$),

$$f_\ell \equiv \frac{N_{\mathrm{id}}}{N_{anti-id}} \text{ with } \ell = e \text{ or } \mu\,. \tag{5.8}$$

The fake factor in Eq. 5.8 has been calculated as a function of the p_{T} and η of the anti-identified electron or muon, separately. The $W+$ jets yield in the signal region is then calculated by scaling the number of events in the $W+$ jets CR ($N_{\mathrm{id}+anti-id}$) by the corresponding fake factor as follows,

$$N_{W+\,\mathrm{jets}} = f_\ell \times N_{id+anti-id} \tag{5.9}$$

The $W+$ jets yield estimation includes a prediction of multi-jets background, in which both leptons are due to the mis-identified jets. The multi-jets contribution to the signal region has been explicitly calculated and it is found to be less than 5% of the total $W+$ jets background, justifying the decision to not subtract out the QCD component from the $W+$ jets background estimation.

The fake factor uncertainty is the main uncertainty source on the $W+$ jets background estimation. The fake factor uncertainty is divided into the following

sources:

- The difference in fake factor in the di-jet sample and $W+$jets CR. Differences in jet kinematics and heavy flavour fraction may cause the fake factor to be different in these two samples. A systematic uncertainty, evaluated from MC, is included to account for the sample dependence.
- The effect of pile-up on the fake factor due to changing data acquisition conditions. This is estimated by studying the fake factor variation as a function of the average number of interactions per bunch crossing. The level of systematic uncertainty is at most 10% depending on p_{T} and η.
- The uncertainty associated to the real lepton contamination from W/Z events in the di-jet CR. This is estimated by varying the requirements to subtract this contamination in the di-jet sample. In general, this systematic uncertainty is below 1%.

The total fake factor uncertainty is driven by the differences in jet composition between di-jet and $W+$jets samples and it represents about 45% (40%) for mis-identified electrons (muons). This systematic uncertainty is treated as uncorrelated between electrons and muons. This reduces the effective uncertainty on the total $W+$jets background, which yields approximately 30% across different N_{jets} categories.

Same-Charge Validation Region

In order to assess the prediction on the $W+$jets background, an independent set of events is used. They form the same-charge VR whise event selection follows the requirements described in Section 5.7, except that two leptons with the same charge are selected. The processes producing the majority of same-charged dilepton events, namely, $W+$jets, $W\gamma$ $W\gamma^{(*)}$, WZ and ZZ, are all backgrounds to $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis.

The comparison of the expected and observed rate and kinematics of the same-charged events is used to validate these background predictions. For illustration, the m_{T} distribution of same-charge $H{+}0j$ events passing the selection after $p_{\mathrm{T}}^{\ell\ell}$ requirement is shown in Fig. 5.27. The total uncertainty on the background prediction shown in these figures includes the systematic uncertainties on the $W+$jets background and the other non-WW diboson backgrounds. The predicted and observed number of events, as well as the modelling of the kinematic distributions,

present a satisfactory agreement overall within the uncertainties in the same-charge VR.

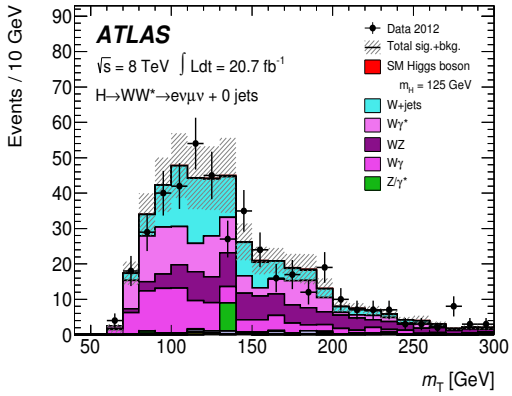

FIGURE 5.27: Distribution of m_{T} in the $N_{\mathrm{jets}} = 0$ same-charge validation region of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis for the 8 TeV data, after the $p_{\mathrm{T}}^{\ell\ell}$ selection. The W+jets prediction is from the data-driven estimate. The hatched area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources. The contributions from WW, ZZ, $t\bar{t}$, and single top are negligible and omitted from the legend.

5.9.3 Non-WW Diboson Backgrounds

The diboson backgrounds, other than the WW signature, consist of $W\gamma$, $W\gamma^*$, WZ, and ZZ processes. The first three processes generate three leptons in the final state. Nevertheless, a high-p_{T} lepton can be emitted very close to a low-p_{T} lepton that is undetected or does not pass the lepton counting kinematic requirements. In such cases, the reconstructed final state is mis-identified as formed by two leptons. The dominant of all these backgrounds is the $W\gamma$ process.

The $W\gamma$ background arises from the photon converting into an electron-positron pair, while the W decay provides the other lepton and the $\not{E}_T$ signature. The simulation of the $W\gamma$ background is checked in a modified same-charge VR in which the electron selection requirements that reject photon conversions are reversed. In this region, a high $W\gamma$ purity of approximately 80% is obtained. Figure 5.28 shows the transverse mass distribution in the $W\gamma$ validation region for events with zero and one jet on the left and right, respectively.

The $W\gamma^*$ background originates from the associated production of a W boson with a virtual photon, where the photon internally converts to a pair of charged

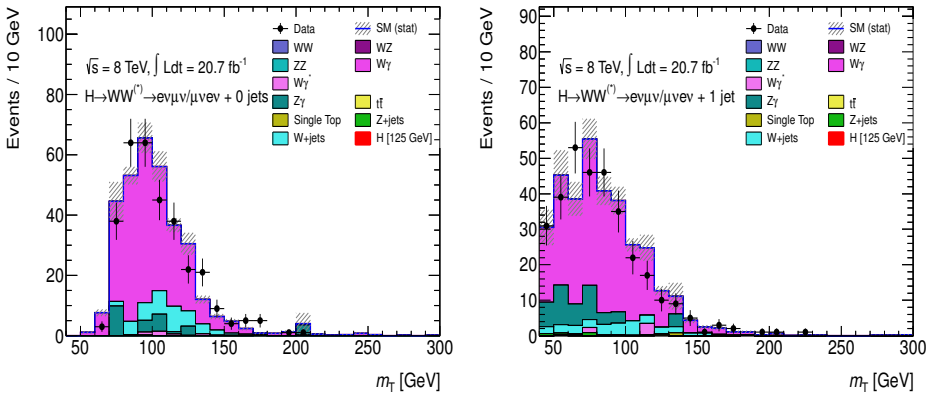

FIGURE 5.28: m_{T} distributions for the same-charge $W\gamma$ validation region in the $H{+}0j$, on the left, and $H{+}1j$, on the right, selection. The $e\mu$ and μe channels are combined. The hatched area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

leptons. Because the conversion occurs before detector material and the cross-section is very low, it is challenging to isolate this background directly in data.

The most pure phase space for the $W\gamma^*$ and WZ processes, in comparison with other backgrounds, is obtained for same-sign leptons at the selection stages of $p_{\mathrm{T}}^{\ell\ell}$ and b-jet veto for the $H{+}0j$ and $H{+}1j$ categories, respectively. After these requirements, the $W\gamma$ contribution is less prominent and the $W\gamma^*$ and WZ purity is about 50%. Several kinematic variables are scrutinised to evaluate the simulation of these backgrounds. Data and MC agree well within the total uncertainties. For illustration, the transverse momentum of the leading lepton distributions in the same-charge VR with $H{+}0j$ and $H{+}1j$ are shown in Fig. 5.29. All the non-WW backgrounds are combined.

The remaining non-WW diboson background originates from the ZZ process. This background contributes with a small fraction of the total expected SM background at the end of the event selection. This is entirely predicted using MC simulation.

Finally, the difference between the data and MC predictions in the VR is taken as a systematic uncertainty. The uncertainty on the total non-WW diboson background in the signal region is 16% and 22% for $H{+}0j$ and $H{+}1j$, respectively.

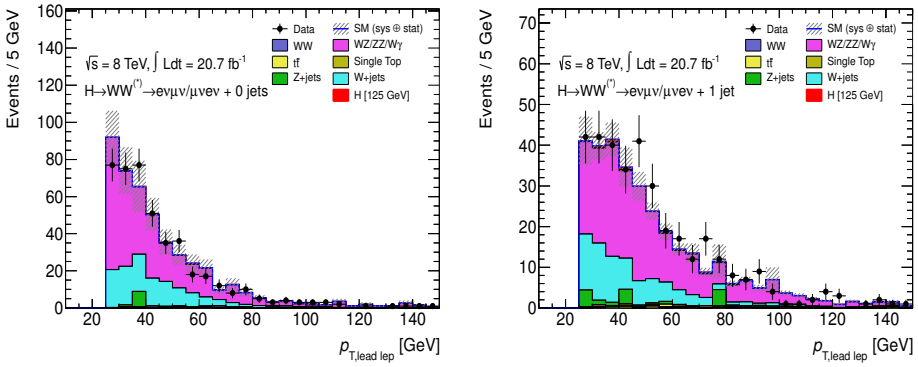

FIGURE 5.29: Distribution of the leading lepton p_{T} for H+0j events, on the left, and H+1j events, on the right, in the same-charge validation region of the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis for the 8 TeV data, after the $p_{\mathrm{T}}^{\ell\ell}$ and the b-jet veto selection, respectively. The W+ jets prediction is from the data-driven estimate. The $e\mu$ and μe channels are combined. The hatched area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

5.9.4 $Z/\gamma^*\to\ell\ell$ Background and the Pacman Method

This section describes the treatment of the Z/DY background decaying to ee or $\mu\mu$. In the $H\to WW^{(*)}\to \ell\nu\ell\nu$ ee+$\mu\mu$ final states, the $p_{\mathrm{T}}^{\ell\ell}$ boost requirement suppresses this background. The majority of the Z/γ^* events are produced nearly at rest, so the boost of the dilepton system must be balanced by a hadronic recoil system. Moreover, these events are produced without genuine $\not\!\!E_T$ measurement. However, due to the enhanced Z/γ^* contamination in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ search, because of the $\not\!\!E_T$ contributions from pile-up interactions, the f_{recoil} variable is introduced in the H+0j and H+1j categories, as discussed in Section 5.6.3.2. The f_{recoil} distribution presents a clear shape distinction between DY and all other processes with genuine $\not\!\!E_T$ measurement, including the Higgs boson signal. The f_{recoil} definition relies on low-p_{T} jets, which are difficult to model reliably with simulation. Hence, it is essential to validate the shape and efficiency of this variable using data. The data-driven technique that estimates the remaining Z/γ^* contribution in ee+$\mu\mu$ final states after the f_{recoil} requirement, is known as the Pacman method [123, 125].

The Pacman method consists of a template fit of the f_{recoil} data distribution considering ee+$\mu\mu$ final states in the signal region (SR), and two f_{recoil} templates extracted from data in two CRs and fitted to data in the SR. The normalisations of the two templates are free parameters of the fit. For the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis, a simplified version of the Pacman method is introduced, as described

below. The full f_{recoil} shape is replaced by two bins: events passing or failling the f_{recoil} requirement. Instead of extracting templates, efficiencies of a cut on f_{recoil} are measured using the same CRs. Moreover, instead of fitting the SR, a extraction of a single analytical solution of the normalisation becomes possible.

The number of observed events before and after the f_{recoil} requirement in the SR ($N_{\text{SR}}^{\text{obs}}$ and $n_{\text{SR}}^{\text{obs}}$, respectively) can be defined in terms of the number of DY events (N^{DY}), and non-DY ($N^{\text{non-DY}}$) as follows,

$$\mathrm{N}_{\text{SR}}^{\text{obs}} = \mathrm{N}^{\text{DY}} + \mathrm{N}^{\text{non-DY}} \tag{5.10}$$

$$\mathrm{n}_{\text{SR}}^{\text{obs}} = \epsilon^{\text{DY}}\mathrm{N}^{\text{DY}} + \epsilon^{\text{non-DY}}\mathrm{N}^{\text{non-DY}}\,, \tag{5.11}$$

where ϵ^{DY} represents the Z/γ^* efficiency of the f_{recoil} requirement, and $\epsilon^{\text{non-DY}}$ is the efficiency for backgrounds other than Z/γ^*. By solving the equations above, the number of Z/γ^* events in the SR passing the f_{recoil} requirement is,

$$n^{\text{DY}} = \epsilon^{\text{DY}}N^{\text{DY}} = \epsilon^{\text{DY}} \times \frac{n_{\text{SR}}^{\text{obs.}} - \epsilon^{\text{non-DY}}N_{\text{SR}}^{\text{obs.}}}{\epsilon^{\text{DY}} - \epsilon^{\text{non-DY}}}\,. \tag{5.12}$$

Equation 5.12 presents all the needed ingredients in order to calculate the Z/γ^* contribution in the SR. However, it is not possible to measure the f_{recoil} efficiencies explicitly from observed events. How the Pacman method estimates these efficiencies is presented below.

The non-DY efficiency is calculated directly from the number of $e\mu$+μe events above and below the f_{recoil} threshold in the signal region,

$$\epsilon^{\text{non-DY}} = \frac{n_{e\mu+\mu e}^{\text{non-DY}}}{N_{e\mu+\mu e}^{\text{non-DY}}}\,. \tag{5.13}$$

The Z/γ^* efficiency is defined as the fraction of events in data passing the f_{recoil} requirement in the Z CR. This CR is defined by applying the pre-selection requirements, but reversing the Z veto, $|m_{\ell\ell} - m_Z| < 15\,\text{GeV}$. In this phase space, the contamination of backgrounds other than Z/γ^* is not negligible and it has to be subtracted. For this purpose, the efficiency for non-Z/DY backgrounds is also calculated in the $e\mu$+μe Z CR as follows,

$$\epsilon_{Z\,\text{CR}}^{\text{non-DY}} = \frac{n_{e\mu+\mu e}^{\text{non-DY}}}{N_{e\mu+\mu e}^{\text{non-DY}}}\,. \tag{5.14}$$

Then, the Z/DY efficiency in the SR with ee+$\mu\mu$ final states ($\epsilon_{\text{est.}}^{\text{DY}}$) is estimated using the Z CR,

$$\epsilon_{\text{est.}}^{\text{DY}} = \frac{n_{\text{est.}}^{\text{DY}}}{N_{\text{est.}}^{\text{DY}}} = \frac{n_{Z\,\text{CR}}^{\text{obs.}} - \epsilon_{Z\,\text{CR}}^{\text{non-DY}} N^{\text{non-DY}}}{N_{Z\,\text{CR}}^{\text{obs.}} - N^{\text{non-DY}}}, \tag{5.15}$$

where all quantities are referring to the ee+$\mu\mu$ final states in the Z CR. The non-DY contributions ($N^{\text{non-DY}}$) in the Z CR are obtained from MC simulation. Lastly, introducing Eq. 5.15 and 5.14 into Eq. 5.12, the estimation on the Z/γ^* background, in ee+$\mu\mu$ events, can be obtained using the number of events after and before the f_{recoil} requirement in several data samples.

The Pacman method makes two assumptions when estimating the f_{recoil} cut efficiencies. First, $\epsilon^{\text{non-DY}}$ and $\epsilon_{Z\,\text{CR}}^{\text{non-DY}}$ are measured using $e\mu$+μe final states and applied to ee+$\mu\mu$ events in the same phase space. Second, the ϵ^{DY} is estimated from the Z CR and then used in the SR with ee+$\mu\mu$ Higgs boson candidate events. These two assumptions will be treated as uncertainties of the Pacman method, as described below.

Systematic uncertainties are assessed for each of the assumptions of the Pacman method. The difference in the non-DY f_{recoil} efficiency between $e\mu$+μe and ee+$\mu\mu$ final states is checked in simulation and validated in data by using the low-p_{T} objects in the recoil calculation, but it is computed in the region perpendicular to the $\vec{p}_{\text{T}}^{\,\ell\ell}$ and $\vec{p}_{\text{T}}^{\,\ell\ell j}$ direction for H+0j and H+1j, respectively. The difference in the Z/γ^* f_{recoil} efficiency between the Z CR and the SR with the ee+$\mu\mu$ final states is checked in simulation and validated in data in the $E_{\text{T,rel}}^{\text{miss}} < 30\,\text{GeV}$ region. The differences summed in quadrature are taken as a systematic uncertainty on the extrapolated efficiency. The largest contribution is the 27% assigned for the Z/γ^* extrapolation from the Z CR to the low-$m_{\ell\ell}$ region. The total uncertainty on this background is 60% and 80% in the H+0j and H+1j categories, respectively.

The ABCD Method for H+2j Category

The H+2j category does not use the f_{recoil} quantity to reject Z/γ^* background due to the difficulty of measuring it in events with a high number of jets and the lower statistics. Instead, this evaluates the Z/γ^* contribution in ee+$\mu\mu$ final states using a data-derived technique, called the ABCD method. The ABCD method estimates the Z/γ^* background in the $E_{\text{T}}^{\text{miss}}$-$m_{\ell\ell}$ plane. The selected events are required to satisfy the pre-selection criteria in addition to the following cuts: $N_{\text{jets}} \geq 2$, $N_{b\text{-jet}} = 0$, $p_{\text{T}}^{\text{tot}} < 45\,\text{GeV}$, and $m_{jj} > 500\,\text{GeV}$. For statistical reasons,

the full VBF-specific selection can not be applied and only the m_{jj} requirement is incorporated into the method. Moreover, there is also a correlation with m_{jj} and the Z/γ^* background estimation which motivates the inclusion. The ABCD method uses four regions, as shown in Fig. 5.30. The x-axis separates the regions in low-$m_{\ell\ell}$ ($m_{\ell\ell} < 60\,\text{GeV}$), and around the Z peak ($|m_{\ell\ell} - m_Z| < 15\,\text{GeV}$). The y-axis is divided into two $E_{\text{T}}^{\text{miss}}$ regions. The former mimics the $\not{E}_T$ selection used in the H+2j category: $E_{\text{T}}^{\text{miss}} > 45\,\text{GeV}$ and $E_{\text{T}}^{\text{miss, STVF}} > 35\,\text{GeV}$. The latter refers to the low $E_{\text{T}}^{\text{miss}}$ region, which inverts the above $\not{E}_T$ requirements.

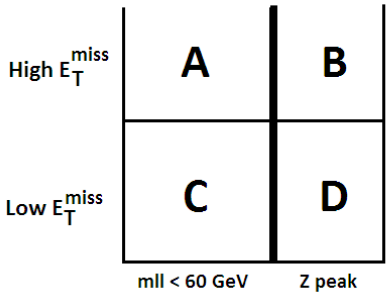

FIGURE 5.30: Schematic representation of the ABCD regions defined in the $E_{\text{T}}^{\text{miss}}$-$m_{\ell\ell}$ plane.

The ABCD method uses a data-derived estimate of region A (A^{est}) using those of B, C and D; where contributions from non-Z sources are subtracted from the last regions.

$$A^{\text{est}} = C \times \frac{B}{D}\,. \tag{5.16}$$

This method assumes that the extrapolation from region B to A behaves in a similar way as the extrapolation from D to C. The above equation can be calculated for data and simulation, resulting in the following normalisation factor (NF),

$$\text{NF} = \frac{A^{\text{est}}_{\text{data}}}{A^{\text{est}}_{\text{MC}}} = \frac{B_{\text{data}}\frac{C_{\text{data}}}{D_{\text{data}}}}{B_{\text{MC}}\frac{C_{\text{MC}}}{D_{\text{MC}}}}\,. \tag{5.17}$$

This equation can be written in terms of a full MC-based term, denoted as f_{corr},

$$\text{NF} \;=\; f_{\text{corr}}\frac{B_{\text{data}}\frac{C_{\text{data}}}{D_{\text{data}}}}{A_{\text{MC}}}\,, \tag{5.18}$$

where,

$$f_{\text{corr}} = \frac{A_{\text{MC}}/B_{\text{MC}}}{C_{\text{MC}}/D_{\text{MC}}}\,. \tag{5.19}$$

It is found that this correction is $3 \pm 10\,(\text{stat.})\%$ which is applied to the simulation.

The ABCD estimate in the Z CR is applied to the low $m_{\ell\ell}$ region. In order to propagate this estimate further in the SR, the normalisation factor extracted from the ratio of simulation and data is computed at the $m_{\ell\ell}$ requirement stage level. The resulting normalisation factor is 0.81 ± 0.06 (stat.).

Lastly, a VBF CR is defined to compute the data-MC cut efficiency ratio of the VBF-specific analysis requirements. As mentioned above, the Z CR does not contain the full VBF-selection in order to increase the statistics of this sample. Since the VBF-specific requirements do not involve lepton kinematics, a higher statistics sample of $Z \to ee$ and $Z \to \mu\mu$ is used. The systematic uncertainty is evaluated by re-deriving the normalisation factor in a region with low-$E_{\text{T}}^{\text{miss}}$ values, enriched with $Z \to \ell\ell$ events, and without $m_{\ell\ell}$ requirement. The total uncertainty on the background results in 15%, which is dominated by experimental sources.

5.9.5 $Z/\text{DY} \to \tau\tau$ Control Region

The $Z/\text{DY} \to \tau\tau$ background makes a noticeable contribution to the $e\mu$+μe final states, particularly at the early stages of the event selection. A $Z/\text{DY} \to \tau\tau$ CR is defined by applying the $H\to WW^{(*)}\to \ell\nu\ell\nu$ event criteria up to the jet selection, including the selection of $E_{\text{T,Rel}}^{\text{miss}} > 25\,\text{GeV}$, and in addition, requiring $m_{\ell\ell} < 80\,\text{GeV}$ and $\Delta\phi_{\ell\ell} > 2.8$. The last requirement is what really separates the Z CR from the SR.

In the H+$0j$ category the sample is 94% pure and the contamination from other background sources is estimated using simulation, except for the W+jets background, which is estimated from data, as is done for the SR. The normalisation factors for the $Z/\text{DY} \to \tau\tau$ background to be applied in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis are derived from the ratio of the background subtracted event yields in data, divided by the expected $Z/\text{DY} \to \tau\tau$ event yield from MC. Since no significant difference is observed between the $e\mu$ and μe final states, the combined $e\mu$+μe final states are used to derive the normalisation factors, which are then applied to both final states. The resulting normalisation factor is $0.90 \pm 0.03\,(\text{stat.}) \pm 0.11\,(\text{syst.})$, which is applied to the $Z/\text{DY} \to \tau\tau$ MC prediction in the SR, as well as to the $Z \to ee$ and $Z \to \mu\mu$ processes, which represent less than 5% of the total Z/γ^*, in the $e\mu$+μe final states.

The $Z/\mathrm{DY} \to \tau\tau$ CR in the $H{+}1j$ analysis is built on an inclusive jet sample and it follows the above prescription, regarding the $m_{\ell\ell}$ and $\Delta\phi_{\ell\ell}$ requirements. The normalisation factors obtained in the $H{+}1j$ category are consistent with those derived when not making any jet selection, but with much larger statistical uncertainties. The decision is, therefore, made to use the normalisation factors from the inclusive jet multiplicity phase space for the $H{+}1j$ category in order to avoid statistical fluctuations. The resulting CR purity is 74% and the normalisation factor is $0.91 \pm 0.03\,(\mathrm{stat.}) \pm 0.09\,(\mathrm{syst.})$.

Finally, the $H{+}2j$ category follows the above prescription, but requiring at least two jets, $N_{b\text{-jet}} = 0$ and, $\mathbf{p}_{\mathrm{T}}^{\mathrm{tot}} < 45\,\mathrm{GeV}$ to define a CR with 67% purity. The purity decreases as the jets are included in the $Z/\mathrm{DY} \to \tau\tau$ CR due to the enhanced top background contamination, which is normalised by its correction factor and subtracted using simulation. The resulting data-MC normalisation factor in this CR is $0.93{\pm}0.11\,(\mathrm{stat.})$. This sample is used to correct the modelling of $E_{\mathrm{T}}^{\mathrm{miss}}$ while a VBF CR is defined to compute the data-MC cut efficiency ratio of the VBF related cuts. The total relative uncertainty on this background is 32%.

5.9.6 Top Control Region

Top background, which includes $t\bar{t}$, tW, tb and tqb with $t \to bW \to b\ell\nu$ processes, is a dominant process that produces high-p_{T} leptons, genuine $\not{E}_T$ measurement, and b-quark jets. The background prediction from simulation is normalised using a CR, which is defined separately for the different jet multiplicities.

The top background for the $H{+}0j$ category is estimated using two CRs. One consists of $e\mu{+}\mu e$ final states passing the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ requirement but without any requirements on the number of jets. The sample is dominated by top quark events, as shown in Fig. 5.14. This is used to set the normalisation of the background, after accounting for the contribution of non-top processes, which are estimated from simulation, except for the $W+\mathrm{jets}$ contribution, which is estimated from data. The second CR is a subset of the first, which has $N_{b\text{-jet}} \geq 1$, and is used to correct the efficiency for the jet veto requirement in top events ($\epsilon_{0-\mathrm{jet}}$). The correction uses the fraction of events in the b-tagged CR to have no jets reconstructed in addition to the one that is tagged ($P_1^{\mathrm{b-tag}}$), and makes the approximation $\epsilon_{0-\mathrm{jet}} \approx (P_1^{\mathrm{b-tag}})^2$, where the power of two is motivated by the presence of two b-jets in $t\bar{t}$. The approximation is not exact because of kinematic correlations between the two b-tagged jets, the presence of QCD radiation and single top, and the effects of b-tagging. Nevertheless, the ratio $\epsilon_{0-\mathrm{jet}}/(P_1^{\mathrm{b-tag}})^2$ is sufficiently stable with respect to experimental uncertainties

and the details of the top sample kinematics allow the estimation of the true $\epsilon_{0-\text{jet}}$ from the one measured in simulation [126]. Because b-tagging selects a very pure sample of top quark events, $P_1^{\text{b-tag}}$ can be estimated in both data and simulation, and $f_{0-\text{jet}}^{\text{MC}}$ is corrected by multiplying it by the ratio $(P_1^{\text{b-tag,data}})^2/(P_1^{\text{b-tag,MC}})^2$. The MC normalisation factor of $1.07 \pm 0.03\,(\text{stat.})$ is found and the total uncertainty on this background is 13%.

In the $H{+}1j$ analysis, the top background represents the main contamination source at the first stages of the event selection in $e\mu{+}\mu e$ final states. In this case, this contribution is the highest one after the full event selection, representing more than 33% of the total expected background. The top prediction is normalised to the data in a CR defined by reversing the b-jet veto and removing the requirements on $\Delta\phi_{\ell\ell}$ and $m_{\ell\ell}$. The resulting samples are primarily top quark events, and the small contribution from other sources is accounted for using simulation and the data-derived $W+$jets estimate. The predicted and observed transverse mass distribution of events in these samples is shown on the left plot in Fig. 5.31.

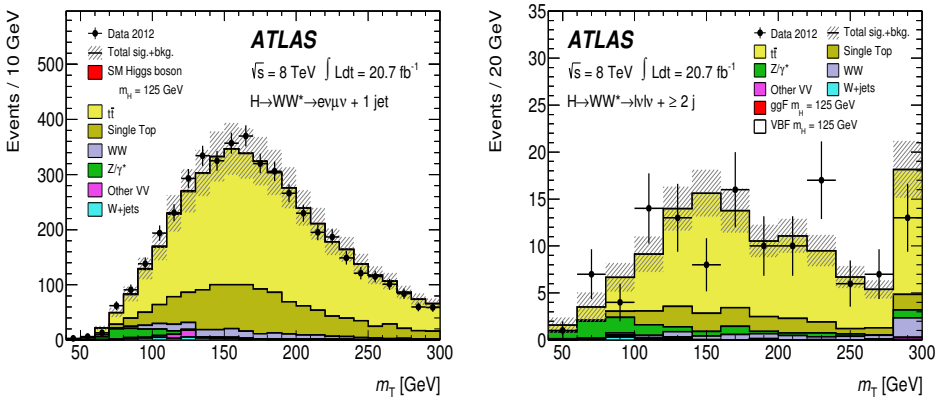

FIGURE 5.31: Distributions of m_{T} in the $N_{\text{jets}} = 1$, on the left, and $N_{\text{jets}} \geq 2$, on the right, top-quark background control regions for the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis of the 8 TeV data. The MC expectation is normalised to the data. The right-most bin in the $N_{\text{jets}} \geq 2$ distribution contains events that would lie beyond the right edge of the figure. The hatched area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

A good agreement is observed between data and MC for the number of events in the CR. The lepton kinematics modelling of the simulation is also satisfactory. The resulting normalisation factor for the $H{+}1j$ analysis is $1.04 \pm 0.02\,(\text{stat.})$ and the total uncertainty on the estimated top quark background in the signal region is 28%.

In the H+2j category the simulated top background is normalised to the data using a control sample defined in a similar way as in the H+1j analysis described above. Hence, the definition of the top background CR selects events with $N_{b\text{-jet}} = 1$. The justification of the $N_{b\text{-jet}} = 1$ requirement is based on the analysis of the truth jet composition of the $t\bar{t}$ MC sample. Those studies show that the sample is quickly dominated for $N_{b\text{-jet, truth}} = 1$ events as the VBF-specific selection is applied. This implies that the top background estimation is sensitive to the modelling of additional jets. Due to the poor statistics after this requirement, the normalisation factor computed at this level is used to weight the top contribution from simulation to the dilepton requirements in the SR.

Due to the low statistics of this CR with ee+$\mu\mu$ final states, all lepton flavours are combined. In the top background CR, the requirements are imposed sequentially in the same order as done in the SR. The transverse mass distribution is shown for the top background CR after the OLV cut in the right plot in Fig. 5.31. Finally, the resulting normalisation factor is $0.59 \pm 0.07\,(\text{stat.})$ which reflects the limitation of the non-VBF simulation in the corner of phase space with $m_{jj} > 500\,\text{GeV}$ and $|\,\Delta y_{jj}\,| > 2.8$. The total uncertainty for this background, including both statistical and systematic effects, is 39%.

5.9.7 WW Background Estimation

The WW background is the dominant background in H+0j category, comparable to the top background in the H+1j, and still significant in the H+2j category. The predictions in the H+0j and H+1j categories are normalised using CRs. The spirit of the method to normalise the WW background relies on events with a dilepton invariant mass different from the SR. In the H+2j case, the WW prediction is obtained entirely from simulation because of the difficulty in isolating a kinematic region that is sufficiently free of top background while still retaining high statistics.

The WW CR for the H+0j category is defined using the same selection as the SR except that the $\Delta\phi_{\ell\ell}$ requirement is dropped and the $m_{\ell\ell}$ bound is modified: $50 \leq m_{\ell\ell} < 100\,\text{GeV}$. The selection of the $m_{\ell\ell}$ range is based on the reduction of the theoretical uncertainties extrapolated from the CR to the SR. In the H+0j category, the WW CR is ∼70% pure using this selection. A factor is used to extrapolate the WW contribution into the SR. It is evaluated by subtracting from the data events the predictions from MC backgrounds other than the WW and the data-derived estimation for the W+jets case. The normalisation factor is found to be $1.16 \pm 0.04\,(\text{stat.})$. This factor is used to scale the MC WW contributions in the SR.

Note that because of the large Z/γ^* contamination in the WW CR in ee+$\mu\mu$ final states, only the $e\mu$+μe events are used to extract the WW normalisation factor which is then applied to all lepton flavour cases. The H+0j analysis also uses a VR which is defined as the CR but with the difference on the $m_{\ell\ell}$ requirement: $m_{\ell\ell} > 100\,\text{GeV}$. This is used to double-evaluate the modelling of the simulation but not to scale the WW background from MC.

In the H+1j category, the definition of the WW CR follows the same selection as the corresponding SR as well. The difference with respect to the H+0j WW CR is that the upper bound on $m_{\ell\ell}$ is replaced with a lower bound of 80 GeV: $m_{\ell\ell} > 80\,\text{GeV}$. In the H+1j category, events from WW contribute about 40% of the total number of events in the WW CR due to the large contamination of top quark processes. The same procedure to obtain the scale factor is used in this case. The resulting WW normalisation factor is $1.03 \pm 0.06\,(\text{stat.})$.

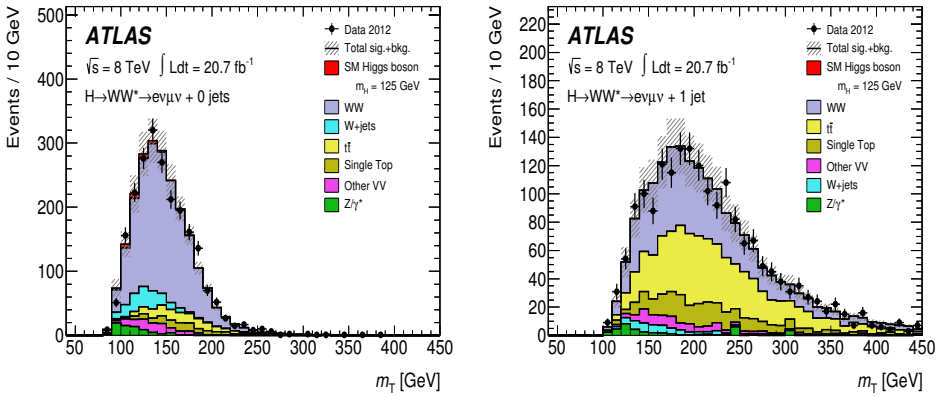

FIGURE 5.32: Distributions of m_{T} in the $N_{\text{jets}} = 0$, on the left, and $N_{\text{jets}} = 1$, on the right, WW CR in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis for the 8 TeV data. The MC expectation is normalised to the data. The top-quark and $Z/\text{DY} \to \tau\tau$ backgrounds are scaled using the normalisation factors derived from the corresponding CRs described in the text. The hatched area represents the uncertainty on the signal and background yields from statistical, experimental, and theoretical sources.

Note that the normalisation factors derived from the top and $Z/\text{DY} \to \tau\tau$ CRs are applied to their corresponding contributions when these backgrounds are subtracted to the data. The observed and predicted m_{T} distributions for events in the WW CRs are shown in Fig. 5.32 for the H+0j and H+1j categories separately. The total uncertainty on the predicted WW background in the SR is 7.4%, 37%, and 37% for the H+0j, H+1j, and H+2j categories, respectively.

5.9.8 Summary of Background Estimates

The background estimation methods, and comparisons of data to MC agreement in several control samples used in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis have been presented in this section. Table 5.18 shows the expected and observed number of events in the different CRs discussed above, which are used to normalise the MC prediction using the observed yields. These include the WW in the $H{+}0j$ and $H{+}1j$ categories, $Z/\mathrm{DY}\to\tau\tau$ in all jet multiplicities cases and top background in $H{+}1j$ and $H{+}2j$.

The normalisation factors do not directly affect the final results of the analysis since all CRs are implemented in a systematic way in the full likelihood. They are a first guess at the appropriate background normalisation and are solely used to give a more accurate representation of the final background expectation in plots and event yield numbers.

TABLE 5.18: CR yields for 8 TeV data in different $H\to WW^{(*)}\to \ell\nu\ell\nu$ categories. The observed (Obs.) and expected event yields for the Higgs boson, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, signal (Higgs Signal) and the total SM background (Total Bkg.) processes are given. The composition of N_{bkg} is given on the right. For $H{+}2j$, the quoted Higgs signal corresponds to the VBF+VH production mechanisms combined, whereas the ggF process is added to the Total Bkg. column. In general, no normalisation factors are applied with the following exception: the top and $Z/\mathrm{DY}\to\tau\tau$ normalisation factors are applied for the corresponding estimates in the WW CRs. All uncertainties are statistical.

Control Region	Obs.	Total bkg.	Higgs Signal	WW	VV	$t\bar{t}$	Single Top	Z/DY	$W+\mathrm{jets}$
WW									
$H{+}0j$	2224	1970±17	31±0.7	1383±9.3	100±6.8	152±4.4	107±4.3	68±10	160±3.6
$H{+}1j$	1897	1893±17	1.9±0.3	752±6.8	88±5.5	717±9.5	243±6.7	37±7.5	56±2.5
$Z\to\tau\tau$									
$H{+}0j$	1935	2251±31	2.5±0.2	61.0±1.9	8.5±1.1	4.5±0.8	2.7±0.6	2113±31	61±3.8
$H{+}1j$	2884	3226±34	7.5±0.3	117±2.7	22±3.1	570±8.4	50±3	2379±32	88±4.3
$H{+}2j$	212	224±7	0.6±0.1	13±1	4±1	44±3	5±1	148±6	9±1
Top									
$H{+}1j$	4926	4781±26	12±0.5	184±3.7	43±9.5	3399±20	1049±13	72±3.1	35±2.2
$H{+}2j$	126	201±5	1.6±0.1	6.4±0.4	1.0±0.3	157±4	26±2	9±1	0.3±0.4

The distributions in the CRs and VRs show satisfactory agreement between the data and the MC given the systematic uncertainties on the latter, which are dominated by the overall theoretical uncertainties on the various background contributions. These uncertainties do not propagate to the SR because they are replaced by the statistical uncertainties on the data. The extrapolation uncertainties are discussed in more detail in the next section.

5.10 Systematic Uncertainties

The systematic uncertainties on the expected yields and cross sections can be divided into two categories: experimental and theoretical. The former takes into account uncertainties derived from the detector measurement such as the jet energy scale or the b-jet tagging efficiency. The latter includes uncertainties such as the estimation of the effect of higher-order terms through variations of the QCD scale inputs to MC calculations. Some of these uncertainties are correlated between the signal and background predictions, so the impact of each uncertainty is calculated by varying the parameter in question and coherently recalculating the signal and background event yields. For the largest backgrounds normalised using CRs (WW for $N_{\text{jets}} \leq 1$ and top in $H{+}1j$ and $H{+}2j$ categories), the theoretical and experimental uncertainties on the extrapolation are described below and the total uncertainties on these backgrounds, as quoted in Section 5.9, are summarised at the end of this section.

5.10.1 Theoretical Systematic Uncertainties on the Signal

Theoretical uncertainties on the signal production cross sections include uncertainties on the QCD renormalisation and factorisation scales, the PDF model used to evaluate the cross section and acceptance, and the underlying event and parton shower model used in the signal model [21, 24]. To evaluate the uncertainties from the QCD factorisation and renormalisation scales, the scales are independently varied up and down by a factor of two. For the ggF signal contribution in the $H{+}0j$ and $H{+}1j$ analyses, the QCD scale uncertainties on the inclusive cross sections for events with $N_{\text{jets}} \geq 0$, ≥ 1, and ≥ 2 are assumed to be independent [127].

Those uncertainties are approximately 8%, 20%, and 70%, respectively, and are calculated using the inclusive ggF process from the HHNLO program [128, 129]. They are converted into uncertainties on the cross sections in exclusive jet multiplicity final states according to the prescription in Refs. [21, 127, 130]. The uncertainties on the inclusive cross sections are shared across the exclusive jet multiplicity categories, and in practice introduce anti-correlations among the predicted signal yield for the different final states representing the migration of events among different jet multiplicities. The sums in quadrature of those uncertainties are 17% and 37% for $H{+}0j$ and $H{+}1j$, respectively.

Scale uncertainties on the ggF process as it appears in $H{+}2j$ are evaluated using the same procedure. In this case, two inclusive ggF processes are considered: events with at least two jets and passing VBF-specific jet requirements but ignoring

the CJV, and events with at least three jets, at least one of which would cause the event to fail the CJV. MCFM [131] is used to evaluate the cross sections under varied renormalisation and factorisation scales. A relative scale uncertainty of 43% is assigned on the cross section for ggF events passing the VBF selection results.

The total scale uncertainty on the signal combines the ggF and VBF contributions. For the VBF signal, the QCD scale uncertainty on the inclusive cross section is estimated to be less than a percent and therefore is negligible. The large scale uncertainties on the ggF mode are correspondingly diluted in the uncertainty on the total signal yield, particularly for higher jet multiplicities. The corresponding uncertainties on the total signal yield are 17%, 30%, and 7% for the H+0j, H+1j and H+2j analyses, respectively. The total QCD scale uncertainty on the signal includes an additional contribution of about 4%, corresponding to the QCD scale uncertainty on the acceptance alone, which is correlated among the jet multiplicities.

The PDF uncertainties on the signal cross section and acceptance are evaluated following Refs. [20, 107, 132, 133], using the envelopes of error sets as well as different PDF sets, applied separately to quark-quark, quark-gluon, and gluon-gluon initiated processes. The dependence on the used value of $\alpha_s(M_Z)$ is also included. The relative PDF uncertainty is 8% for the ggF and $t\bar{t}H$ processes and $3-4$% for the quark-initiated VBF and VH processes. These uncertainties are estimated by following the prescription in Ref. [132] and by using the PDF sets of CT10 [107], MSTW [20], and NNPDF [133]. The PDF uncertainties are assumed to be completely correlated among processes with identical initial states, regardless of whether or not they are signal or background processes.

Uncertainties on the POWHEG+PYTHIA8 modelling of signal processes, particularly the sensitivity to the underlying event and parton shower model, are estimated by comparison to MC@NLO+HERWIG. The resulting uncertainties are 3% for the H+0j signal and 10% for the H+1j signal, anti-correlated between the jet multiplicity bins. For the H+2j analysis, the uncertainty on the effect of UE modelling is evaluated through comparison of POWHEG+PYTHIA6 samples generated with and without the UE, and is 9% for the ggF process and 3% for the VBF process.

5.10.2 Dominant Systematic Uncertainties on Background Processes

When carrying out the limit fit the total background shape uncertainty is dominated by the uncertainty on the normalization of each background contributing to it.

Hence, the total background shape uncertainty is mainly affected by changes in the relative background contributions.

For backgrounds such as WW and top, that are evaluated through extrapolation from a signal-depleted CR, theoretical uncertainties are reduced compared to those on the absolute MC normalisation. The extrapolation to the signal region must still be derived from simulation, so some theoretical uncertainties remain. The parameters are defined generally as the ratio of the number of events passing the signal region selection to the number passing the CR selection as evaluated in simulation, $\alpha = N_{\mathrm{SR}}/N_{\mathrm{CR}}$. These are discussed in more detail below. For small backgrounds, such as $W\gamma^{(*)}$ and WZ, the background acceptance is completely evaluated from simulation and calculated cross sections are used for their normalisation. As a result, the associated theoretical uncertainties are larger than those for backgrounds using CRs.

For WW, the parameters α_{WW}^{0j} and α_{WW}^{1j} denote the extrapolation parameters for $H{+}0j$ and $H{+}1j$, respectively. The uncertainties on these parameters are evaluated according to the prescription of Ref. [21]. Four main sources of uncertainty on the normalisation have been considered: QCD renormalisation and factorisation scales, dependence on PDF, dependence on the MC generator and dependence on the UE and PS model. Scale uncertainties have been computed using the MCFM generator by varying the renormalisation and factorisation scales by a factor of 2. PDF uncertainties are calculated as for the signal, using the same generator as used for the central value of the α parameters.

The signal extraction procedure relies on the precise knowledge of the modelling by simulation. These uncertainties are evaluated by comparing the α extrapolation parameters from different generators: *Powheg*+*Pythia8* and MCFM. However, *MC@NLO* is not included in this comparison because the calculation excludes singly-resonant processes and does not treat spin correlations at the matrix element level. The UE and PS uncertainties are evaluated by comparing the predictions of Powheg interfaced with Pythia8, Pythia6, and Herwig. The α parameters are found to be positively correlated between $H{+}0j$ and $H{+}1j$, as well as for all of subdivisions of the signal region by lepton flavour and $m_{\ell\ell}$. The total quoted uncertainties are about 2% and 4–6% for the $H{+}0j$ and $H{+}1j$ categories, respectively. These values are summarised in Table 5.19. The modelling and scale uncertainties have been checked using *aMC@NLO* [134, 135], which gives the same results within the statistical uncertainties of the comparison.

Because the m_{T} distribution is used in the analysis to estimate the signal yield, an additional theoretical uncertainty is evaluated on the shape of this distribution

TABLE 5.19: Uncertainties on the extrapolation parameters α for the WW background in the H+0j and H+1j categories. Uncertainties due to the QCD scale, PDF, parton shower (PS), underlying event (UE), and modelling of the NLO qq, $gq \to WW$ processes are given. Each source, represented by a column, is assumed to be uncorrelated, but for a given source the uncertainties are assumed to be fully correlated among all signal regions with H+0j and H+1j. A relative sign between two entries in a column indicates anti-correlation between those signal regions for that source of uncertainty.

Final State	Range (GeV)	QCD scale (%)	PS, UE (%)	PDF (%)	Modelling (%)
H+0j					
$e\mu$+μe	$10 < m_{\ell\ell} < 30$	0.9	0.2	1.5	−1.2
$e\mu$+μe	$30 \leq m_{\ell\ell} < 50$	0.9	0.8	1.1	−1.4
ee+$\mu\mu$	$12 < m_{\ell\ell} < 50$	1.0	0.3	1.1	1.7
H+1j					
$e\mu$+μe	$10 < m_{\ell\ell} < 30$	1.6	0.5	2.0	−5.1
$e\mu$+μe	$30 \leq m_{\ell\ell} < 50$	1.5	0.5	1.8	−5.0
ee+$\mu\mu$	$12 < m_{\ell\ell} < 50$	1.4	0.6	1.7	−3.1

for the dominant WW background. It is computed by comparing the m_{T} shape predicted by the MCFM, SHERPA, POWHEG+PYTHIA6, and MC@NLO+HERWIG generators, as well as a comparison among showering algorithms. The resulting maximal variations in the normalised m_{T} distributions are about 20% and are concentrated in the tails of the distribution. The envelope of the distributions from the comparison, which is dominated by the differences between MCFM and MC@NLO+HERWIG, is taken as a relative shape uncertainty on the POWHEG m_{T} distribution.

The dominant uncertainties on the top background for the H+0j category are the theoretical uncertainties on the component derived from MC simulation. These total to 10% and include the effects of QCD scale, initial- and final-state radiation, generator/PS model, the relative normalisation of $t\bar{t}$ and single top, and the interference between single top and $t\bar{t}$, which is neglected when using separate $t\bar{t}$ and single top Monte Carlo samples. The top background for H+1j and H+2j categories evaluated by extrapolation from a signal-depleted CR, as is the case for WW, but the associated uncertainty is dominated by experimental uncertainties, to be described in the next section. For H+1j, the uncertainty of 8% on α is evaluated by comparison of simulated $t\bar{t}$ and single top events with different QCD tunes for initial- and final-state radiation. For H+2j, the uncertainty of 15% on α is evaluated by comparing the modelling of various generators after the VBF-related selection.

The WW yield in H+2j is predicted from simulation. Two types of contributions are considered: QCD and electroweak processes. For the former, a total

uncertainty of 42% is dominated by QCD scale and PDF variations. For the latter, which are non-negligible in $H+2j$ category, a total uncertainty of 11% is obtained by considering the QCD scale, the interference between QCD and Higgs boson processes, and the difference between the SHERPA and MadGraph generators.

5.10.3 Uncertainties on Backgrounds Normalised to Control Regions

For the backgrounds normalised using CRs (WW for the $H+0j$ and $H+1j$ categories and top in the $H+1j$ and $H+2j$ analyses), the sources of uncertainty can be grouped into three categories,

- the statistical uncertainty,
- two uncertainties related to the simulation-based extraction from the CR to the signal region: theoretical and experimental,
- the uncertainty on the other contribution processes in the CR, which are subtracted from the data yield to get the estimated number of events from the targeted background in the corresponding CR.

These sources, and the resulting total uncertainty, are summarised in Table 5.20. The uncertainties on α are defined and described above. The statistical uncertainty is derived from the number of events in the corresponding CR, which can be found in Table 5.18. The uncertainties from the normalisation of other processes in the CR, as represented here, are necessarily approximate because of the correlations among the backgrounds. However, these correlations are fully represented in the statistical procedure to extract the results, as discussed in Section 5.11.2.

5.10.4 Experimental Uncertainties

The experimental uncertainties affect both the expected signal and background yields, and are mainly associated with reconstruction efficiency, energy scale and resolution of the different objects (leptons, jets, and $E_{\mathrm{T}}^{\mathrm{miss}}$) in the event. The most significant contributions are from the jet energy scale and resolution, the b-tagging efficiency, and the 30% uncertainty on the fake factor used to calculate the $W+$ jets background. There is also an experimental uncertainty associated to the integrated luminosity which is 3.6% for the 2012 data. It is derived, following the methodology of Ref. [136], from a preliminary calibration of the luminosity scale derived from beam-separation scans of April 2012.

TABLE 5.20: Total relative uncertainties on backgrounds normalised using CRs. The statistical component (Stat.) is from the CR yields; the theoretical uncertainties (Theory) are from the α extrapolation parameter; the experimental (Exp.) uncertainties are given. The approximate uncertainties on the normalisation of other processes in the CR (Crosstalk) are given. The WW and top in $H{+}1j$ are anti-correlated due to the b-jet selection, so that the uncertainties partially cancel.

Estimate	Stat. (%)	Theory (%)	Exp. (%)	Crosstalk (%)	Total (%)
WW					
$H{+}0j$	2.9	1.6	4.4	5.5	7.4
$H{+}1j$	6	5	4	9	37
Top					
$H{+}1j$	2	8	22	16	29
$H{+}2j$	10	15	29	19	39

The jet energy scale is determined from a combination of test beam, simulation, and *in situ* measurements [137]. Its uncertainty is split into several independent components: η intercalibration of jets from the central to the forward region, high-p_{T} jets, MC non-closure, topologies with close-by jets, different quark/gluon composition and response, the b-jet energy scale, impact from in-time and out-of-time event pile-up, and *in situ* jet energy corrections. The latter is further divided into several different categories depending on the physical source of the uncertainty. The jet energy scale uncertainty, for jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\,\eta\,| < 4.5$, is 1–5% depending on p_{T} and η. The jet energy resolution varies from 5% to 20% as a function of the jet p_{T} and η. The relative uncertainty on the resolution, as determined from *in situ* measurements, ranges from 2% to 40%, with the largest value of the resolution and relative uncertainty occurring at the p_{T} threshold of the jet selection. The reconstruction, identification, and trigger efficiencies for electrons and muons, as well as their momentum scales and resolutions, are estimated using $Z \to \ell\ell$, $J/\psi \to \ell\ell$, and $W \to \ell\nu$ decays ($\ell = e, \mu$). The resulting uncertainties are all smaller than 1%. The exception to this is the uncertainty on the electron selection efficiency, which varies between 2% and 5% as a function of p_{T} and η.

The efficiency of the b-tagging algorithm is calibrated using samples containing muons reconstructed in the vicinity of jets [118]. The resulting uncertainty on the b-jet tagging efficiency varies between 5% and 12% as a function of jet p_{T}.

The changes in jet energy and lepton energy/momentum due to systematic variations are propagated to $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T,STVF}}^{\mathrm{miss}}$; the changes in the high-p_{T} object

energy/momentum and in the $E_{\mathrm{T}}^{\mathrm{miss}}$ quantities are therefore, fully correlated. Additional contributions to the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T,STVF}}^{\mathrm{miss}}$ uncertainty arise from jets with $p_{\mathrm{T}} < 20\,\mathrm{GeV}$ as well as from low-energy calorimeter deposits not associated with reconstructed physics objects [59]; their effect on the total signal and background yields is about 3%.

Lepton momentum scale uncertainties are also propagated to the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ calculation. In addition, uncertainties are assigned to the scale and resolution of the remaining $E_{\mathrm{T}}^{\mathrm{miss,track}}$ component not associated with charged leptons. These uncertainties are calculated by comparing the properties of $E_{\mathrm{T}}^{\mathrm{miss,track}}$ in Z events in real and simulated data, as a function of the sum of the hard p_{T} objects in the event.

In the fit to the m_{T} distribution to extract the signal yield, the predicted m_{T} shape from simulation is used for all of the backgrounds except $W+$jets. For $W+$jets, the shape is taken from the same data which is used to normalise the background estimate, with the same fake factor applied. For the other backgrounds, the impact of experimental uncertainties on the m_{T} shapes for the individual backgrounds and signal are evaluated, and no statistically significant dependence is observed for the majority of the experimental uncertainties. Those experimental uncertainties, which do produce statistically significant variations of the shape, have no appreciable effect on the final results because the uncertainty on the m_{T} shape of the total background is dominated by the uncertainties on the normalisations of the individual contaminations.

5.11 Statistical Model and Signal Extraction

In this section the statistical treatment for the analysis of the Higgs boson decaying into a pair of W bosons is presented. Section 5.11.1 summarises the results published in Ref. [138], in which the general formalism of a search as a statistical test is outlined. The statistical significance of an observed signal can be quantified by means of a p-value or its equivalent Gaussian significance, as discussed below. Section 5.11.2 presents the statistical procedure for fitting the m_{T} spectrum used in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis. Details on the methods specific to the treatment of each background process, as well as the combination across jet bins and across years are discussed.

5.11.1 Formalism of a Search as a Statistical Test

For purposes of discovering a new signal process, the null hypothesis (H_0) can be defined as describing only known processes, the so-called background-only hypothesis. This is to be tested against the alternative hypothesis (H_1), which includes both background as well as the signal. When setting limits, the model with signal plus background is tested against the background-only hypothesis.

To summarise the outcome of such a search one quantifies the level of agreement of the observed data with a given hypothesis H by computing a p-value. The p-value is the probability, under the assumption of H, of finding data of equal or higher incompatibility with the predictions of H. Moreover, one can regard the hypothesis as excluded if its p-value is observed below a specified threshold. It is possible to convert the p-value into an equivalent significance, Z, defined such that a Gaussian distributed variable found Z standard deviations above its mean has an upper-tail probability equal to p. That is,

$$Z = \phi^{-1}(1-p)\,, \tag{5.20}$$

where ϕ^{-1} is the quantile (inverse of the cumulative distribution) of the standard Gaussian. For example, a significance of $Z > 5$ corresponds to $p < 2.9 \times 10^{-7}$.

The sensitivity of an analysis can be reported by the median, so-called expected, significance that one would obtain with a given measurement under the assumption of the hypothesis. In light of this, the sensitivity to discovery of a given signal process H_1 could be characterised by the median value, under the assumption of H_1, of the value of Z obtained from a test of H_0.

Consider an experiment where for each selected event one measures the values of certain kinematic variables. Suppose for each event in the signal sample one measures a variable x and uses these values to construct an N-bin histogram with $n = (n_1, ..., n_N)$. The expectation value of n_i can be written $E[n_i] = \mu s_i + b_i$, where the mean number of entries in the ith bin from the signal and background are,

$$s_i = s_{tot} \int_{\text{bin } i} f_s(x|\vec{\theta}_s)\,\mathrm{d}x\,, \qquad b_i = b_{tot} \int_{\text{bin } i} f_b(x|\vec{\theta}_b)\,\mathrm{d}x\,. \tag{5.21}$$

Here the parameter μ determines the strength of the signal process, with $\mu = 0$ corresponding to the background-only hypothesis and $\mu = 1$ being the nominal signal hypothesis. The functions $f_s(x|\vec{\theta}_s)$ and $f_b(x|\vec{\theta}_b)$ are the probability density functions (pdfs) of the variable x for signal and background events, and $\vec{\theta}_s$ and $\vec{\theta}_b$ represent parameters that characterize the shapes of pdfs. The quantities s_{tot} and

b_{tot} are the total mean numbers of signal and background events. Here we will use $\vec{\theta} = (\vec{\theta}_s, \vec{\theta}_b, b_{tot})$ to denote all of the nuisance parameters.

In addition to the measured histogram $\vec{n}$ often subsidiary measurements are made in order to constrain the nuisance parameters. These then give a set of values $\vec{m} = (m_1, ..., m_M)$ for the number of entries in each of the M bins. The expectation value of m_i can be written $E[m_i] = u_i(\vec{\theta})$, where the u_i are calculable quantities depending on the parameters $\vec{\theta}$.

In this case, the likelihood function is the product of the Poisson probabilities for all bins,

$$\mathcal{L}(\mu, \vec{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} \mathrm{e}^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} \mathrm{e}^{-u_k} \,. \tag{5.22}$$

To test a hypothesis value of μ we consider the profile likelihood ratio [139],

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}})}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}})} \,, \tag{5.23}$$

where $\hat{\hat{\vec{\theta}}}$, in the numerator, denotes the value of $\vec{\theta}$ that maximizes $\mathcal{L}$ for the specified μ, i.e., it is the conditional maximum-likelihood (ML) estimator of $\vec{\theta}$; and thus it is a function of μ. The $\hat{\vec{\theta}}$, in the denominator, is the maximized (unconditional) likelihood function, i.e., $\hat{\mu}$ and $\hat{\vec{\theta}}$ are the ML estimators.

From the definition of $\lambda(\mu)$ in Eq. 5.23, it can be seen that $0 \leq \lambda \leq 1$, with λ near 1 implying better agreement between the data and the hypothesized value of μ. Equivalently, it is convenient to use the statistic,

$$t_\mu = -2 \ln \lambda(\mu) \,, \tag{5.24}$$

as the basis of a statistical test. Higher values of the t_μ thus correspond to increasing incompatibility between the data and μ. To quantify the level of agreement, between the observed data and the hypothesis, the p-value is computed,

$$p_\mu = \int_{t_{\mu,\mathrm{obs}}}^{\infty} f(t_\mu|\mu) \, \mathrm{d}t_\mu \,, \tag{5.25}$$

where $t_{\mu,\mathrm{obs}}$ is the value of the statistic t_μ observed from the data and $f(t_\mu|\mu)$ denotes the pdf of t_μ under the assumption of the signal strength μ.

The statistic test for discovery: For discovery purposes it is convenient to

test the $\mu = 0$ hypothesis. Rejecting $\mu = 0$ amounts to discovering a new signal. In such a case it is possible to define the test such that the data are only regarded as discrepant with the hypothesis $\mu = 0$ if an excess of events is observed, hence $\hat{\mu} > 0$. The test is defined as,

$$q_0 = \begin{cases} -2\ln\lambda(0) & \text{if } \hat{\mu} \geq 0\,, \\ 0 & \text{if } \hat{\mu} < 0\,, \end{cases} \tag{5.26}$$

where $\lambda(0)$ is the profile likelihood ratio for $\mu = 0$ as defined in Eq. 5.23. The p-value calculated for this test statistic will then take the form,

$$p_0 = \int_{t_{0,\text{obs}}}^{\infty} f(q_0|0)\,\mathrm{d}q_0\,. \tag{5.27}$$

This is, the primary role of the p-value is to quantify the probability that the background-only model gives a statistical fluctuation as big as the one seen or bigger.

The statistic test for upper limits: For purposes of establishing an upper limit on the strength parameter μ, it can be defined,

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \text{if } \hat{\mu} \leq \mu\,, \\ 0 & \text{if } \hat{\mu} > \mu\,, \end{cases} \tag{5.28}$$

where $\lambda(\mu)$ is the profile likelihood ratio from Eq. 5.23. The reason for setting $q_\mu = 0$ for $\hat{\mu} > \mu$ is that when setting an upper limit, one would not regard data with $\hat{\mu} > \mu$ as representing less compatibility with μ than the data obtained, and therefore this is not taken as part of the rejection region of the test. From the definition of the test statistic one sees that higher values of q_μ represent greater incompatibility between the data and the hypothesis value of μ.

For the case in which the model considers $\mu \geq 0$, if the data provides $\hat{\mu} < 0$, the best level of agreement between the data and any physical value of μ occurs for $\mu = 0$. In this case the profile likelihood ratio can be defined as below,

$$\tilde{\lambda}(\mu) = \begin{cases} \dfrac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\hat{\vec{\theta}}})} & \text{if } \hat{\mu} \geq 0\,, \\ \dfrac{\mathcal{L}(\mu, \hat{\hat{\vec{\theta}}}(\mu))}{\mathcal{L}(0, \hat{\hat{\vec{\theta}}}(0))} & \text{if } \hat{\mu} < 0\,. \end{cases} \tag{5.29}$$

Where $\hat{\hat{\vec{\theta}}}(0)$ and $\hat{\hat{\vec{\theta}}}(\mu)$ refer to the conditional ML estimators of $\vec{\theta}$ given a strength parameter of 0 and μ, respectively.

Then, the variable $\tilde{\lambda}(\mu)$ can be used instead of $\lambda(\mu)$ in Eq. 5.23 to obtain the corresponding test statistic, which is denoted by $\tilde{q}_\mu$,

$$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \text{if } \hat{\mu} \leq \mu \\ 0 & \text{if } \hat{\mu} > \mu \end{cases} = \begin{cases} -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(0,\hat{\hat{\theta}}(0))} & \text{if } \hat{\mu} < 0\,, \\ -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}(\hat{\mu},\hat{\theta})} & \text{if } 0 \leq \hat{\mu} \leq \mu\,, \\ 0 & \text{if } \hat{\mu} > \mu\,. \end{cases} \tag{5.30}$$

This test statistic is known as the alternative test statistic and it is used in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis for setting upper limits on the parameter of interest μ.

Frequentist method CL_s: It can be that the effect of a given hypothesis μ is very small relative to the background-only prediction. This means that the pdf for both will be almost the same and the probability to reject μ if it is true will be close to the probability to reject μ if $\mu = 0$. In this case one could exclude hypotheses to which one has essentially no sensitivity. This effect is known as spurious exclusion. The problem of spurious exclusion has been known for more than 30 years [140]. In the 1990s this was re-examined for the LEP Higgs search leading to the modified frequentist method, known as $\mathrm{CL_s}$ [141]. This method is used to compute 95% confidence intervals on the signal strength parameter μ. The $\mathrm{CL_s}$ solution is to base the test not on the usual p-value, $\mathrm{CL_{s+b}}$, but rather to divide this by $\mathrm{CL_b}$, which is approximately one minus the p-value of the background-only hypothesis,

$$\mathrm{CL_s} = \frac{\mathrm{CL_{s+b}}}{\mathrm{CL_b}} = \frac{p_\mu}{1-p_\mathrm{b}}\,, \tag{5.31}$$

where the p_μ and p_b are the p-values derived from the pdf distributions $f(\tilde{q}_\mu|\ \mu,\hat{\theta}_\mu)$ and $f(\tilde{q}_\mu|\ 0,\hat{\theta}_0)$, respectively:

$$p_\mu = \int_{\tilde{q}_{\mu,\mathrm{obs}}}^{\infty} f(\tilde{q}_\mu|\ \mu)d\tilde{q}_\mu\,, \qquad p_b = \int_{-\infty}^{\tilde{q}_{\mu,\mathrm{obs}}} f(\tilde{q}_\mu|\ 0)d\tilde{q}_\mu\,. \tag{5.32}$$

The 95% CL upper limit on μ is the solution to $\mathrm{CL_s} = 0.05$.

5.11.2 Statistical Procedure in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ Analysis

The statistical treatment in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis follows the procedure covered in Ref. [138] which is summarised in Section 5.11.1.

The $H\to WW^{(*)}\to \ell\nu\ell\nu$ search uses the likelihood function ($\mathcal{L}$) computed as the product of Poisson functions for each phase space defined in the analysis. The product is done over all lepton final states: ee, $\mu\mu$, $e\mu$ and μe in each of the three jet bin categories: $N_{\text{jets}} = 0$, $N_{\text{jets}} = 1$, and $N_{\text{jets}} \geq 2$. In the Poisson term for the signal region μ scales the expected signal yield, with $\mu = 0$ corresponding to no signal and $\mu = 1$ corresponding to the Higgs boson hypothesis with $m_{\text{H}} = 125\,\text{GeV}$. The signal strength μ is found by maximising $\mathcal{L}$ that is defined using the m_{T} distribution for events passing the event selection described in Section 5.7. The final cut on the transverse mass is left out as this spectrum will be used in the fitting procedure. As mentioned in Section 5.7.5, the samples for the $e\mu$+μe final state in $N_{\text{jets}} \leq 1$ are split at $m_{\ell\ell} = 30\,\text{GeV}$, treating them as separate signal regions.

The MC is not distributed homogeneously across the transverse mass range, as Fig. 5.25 shows. This degrades the performance and leaves many bins with low MC statistics. For this reason, the m_{T} distribution in the signal region is mapped separately in each lepton final state such that the sum of the backgrounds is uniformly distributed across the mass range. The number of bins used for the remapping is driven by the available MC statistics at the end of the event selection. The $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis uses five, three and four bins in the H+0j, H+1j and H+2j categories, respectively. The exception is the H+2j category in 2011 data for which no binning is carried out, since the statistics are very low after the full event selection is applied, as seen in Tab. C.4.

The background treatment uses extrapolation factors to describe how the fitted background rates translate from the CR to the signal region. Rather than using these normalisation factors values, an equivalent parametrisation is used,

$$\mathcal{L}(\mu, \mu_b) = P(N|\mu s + \mu_b b_{\text{SR}}^{\text{exp}}) \times P(M|\mu_b b_{\text{CR}}^{\text{exp}}) \tag{5.33}$$

where b_{SR}^{exp} and b_{CR}^{exp} are expected background yields in the signal and CR determined by simulation, μ is the signal strength parameter, and $\mu_b b_{CR}^{exp}$ is the background strength parameter.

The strength parameters are allowed to move freely to best fit the data. The expected signal and background yields in the Poisson distribution are allowed to

vary within the allowed range of the relevant systematic uncertainties. Such an uncertainty is parametrised by the corresponding nuisance parameter θ that can be constrained by an unit Gaussian,

$$G(\tilde{\theta}|\theta, 1) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\frac{(\tilde{\theta}-\theta)^2}{2}} , \tag{5.34}$$

or a Poisson distribution,

$$\mathcal{P}(\tilde{\theta}|\theta\alpha) = \frac{(\theta\alpha)^{\tilde{\theta}} \mathrm{e}^{-\theta\alpha}}{\tilde{\theta}!} , \tag{5.35}$$

where α is a constant taken as the nominal value of $\tilde{\theta}$. When adding the nuisance parameter, θ, one can consider a constraint $N(\tilde{\theta}|\theta)$ representing an auxiliary measurement $\tilde{\theta}$ related to the nuisance parameter θ. In practice, there are as many nuisance parameters as uncertainties and it will be referred to as the vector of the nuisance parameters, $\vec{\theta}$.

Expanding the simple likelihood given in Eq. 5.33 to the one used in this analysis, a product over lepton flavours and jet multiplicities is done. Because the m_{T} distribution is binned, a product over the m_{T} bins is also present. The strength parameters μ_b, that were introduced in Eq. 5.33, are applied to the WW background in the H+0j and H+1j analyses, as well as the top background in the H+1j and H+2j analyses. The other backgrounds are added to the Poisson expectations. The full likelihood can be written as,

$$\mathcal{L}(\mu, \vec{\theta}) = \prod_{k=ee,e\mu,\mu e,\mu\mu} \prod_{j=0}^{N_{\mathrm{jets}}} \prod_{i=1}^{N_{\mathrm{bins}}} \mathcal{P}(N_{ijk}|\mu s_{ijk} + \sum_{m}^{N_{\mathrm{bkg}}} b_{ijkm}) \times \prod_{i=1}^{N_\theta} N(\tilde{\theta}|\theta) . \tag{5.36}$$

The signal and background expectations are functions of the nuisance parameters θ. These functions are parametrised such that the response of s and b to each θ is factorised from the nominal value of the expected rate. That is, $s = s_0 \times \prod \nu(\theta)$ and similarly, $b = b_0 \times \prod \nu(\theta)$; where the form of $\nu(\theta)$ depends on the systematic source. There are four general cases,

- Case 1. Systematics that do not change the m_{T} shape, flat systematics, take the form $\nu_{\mathrm{flat}}(\theta) = \kappa^\theta$, where κ is determined by measuring ν_{flat} at $\theta = \pm 1$. In this case, the constraint term on θ that is present in the likelihood is a unit gaussian.

- Case 2. If the systematic affects the shape, the shape variation is first separated into a flat component and a pure shape component, such that varying pure shape component of s or b has no effect on the expected rate. The flat component is treated as described above. The pure shape component uses vertical linear interpolation to estimate the variation, and so it is distributed as a truncated gaussian. Explicitly, $\nu_{\text{shape}}(\theta) = 1 + \epsilon\theta$, where ϵ is again determined by measuring ν_{shape} at $\theta = \pm 1$ and the constraint is a unit gaussian. The truncation is imposed such that $\nu_{\text{shape}}(\theta < \frac{-1}{\epsilon}) = 0$. Note that systematic sources can have both a normalisation, case 1, and shape component. In this case, the same θ is shared between both functions $\nu_{\text{flat}}(\theta)$ and $\nu_{\text{shape}}(\theta)$.

- Case 3. This case takes into account the treatment of purely statistical uncertainties. It refers to uncertainties from MC statistics or data-driven methods. This constrain represents an auxiliary measured number of events $\tilde{\theta}$ with an expected number $\theta\alpha$. It is the Poisson probability as given in Eq. 5.35.

- Case 4. The final case is where it is used a CR to constrain the normalisation of a background. The expected number of events is $\lambda = \mu s + \theta b_{\text{target}} + \sum_i^{N_{\text{bkg}-1}} b_i$, where b_{target} is the background targeted by the CR. This properly takes into account the contamination due to both the signal and other backgrounds. In the full likelihood there are four nuisance parameters representing the strengths of the two major backgrounds: WW in $H{+}0j$ and $H{+}1j$ analyses, and top in $H{+}1j$ and $H{+}2j$ analyses. Moreover, the strength parameters multiply the expected background anywhere that background is present. In this way, the contaminations among the various CRs are treated properly. Although there are only four unique background strength parameters, a separate Poisson constraint is present for each leptonic final state and jet multiplicity which means 12 constraints of this form.

Because each θ represents a different systematic source, one θ can affect multiple signal and background rates in a correlated way. For all samples, shape uncertainties due to b-tagging, electron, trigger, and isolation efficiency, are included. For the $W+$jets, the shape uncertainty on the fake rate is also included. Finally, the $E_{\text{T}}^{\text{miss}}$ and pileup shape uncertainties on the ggF signal are added.

The use of a fit, instead of a selection of events in a range of m_{T}, increases the sensitivity of the analysis but also generates additional systematic uncertainties on the modelling of the shape of the m_{T} distribution for the backgrounds. In light of

this, an uncertainty on the modelling of the m_{T} shape is determined by comparing several generators and showering simulations, as presented in Section 5.10.

The statistical workspaces for the 2011 and 2012 data taking periods are made separately. Hence, to produce the combined results for the $7 + 8$ TeV analyses, it is important to take into account the correlated effect of systematic uncertainties across years. In general, all nuisance parameters are treated as 100% correlated except for those which are statistical in origin or have a different source in the two datasets as the statistical component of the jet energy scale calibration, b-tagging efficiency, luminosity uncertainty or the uncertainty on the soft term in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation.

In Section 5.12, the results are reported with the signal significance and the corresponding p_0 value, the 95% CL exclusion curves, the signal strength parameter μ, and a two-dimensional plot of μ vs. m_{H}.

5.12 Results

This section presents the results of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis using the 7 TeV and 8 TeV datasets collected by the ATLAS detector. First, the comparison of observed and expected signal and background yields is given in Section 5.12.1, following the event selection for 2012 and the re-analysis of the 2011 data shown in Section 5.7 and Section 5.8, respectively. The statistical interpretations of the 7 TeV and the 8 TeV analysis are given in Section 5.12.2 and Section 5.12.3, respectively. The interpretation of the combined 7 TeV and 8 TeV results is discussed in Section 5.12.4. Then the statistical results for the ggF and VBF production modes are presented in Section 5.12.5. Finally, the results in Section 5.12.6 present the measurement of the Higgs boson production cross section.

5.12.1 Expected Signal and Background Event Yields

In Section 5.7 the observed and expected signal and background yields in the signal regions of the 8 TeV modes are given in Tabs. 5.10, 5.11, 5.13, 5.12, 5.14, and 5.15 for each of the final states in which the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis is divided. Figure 5.25 shows the distributions of the transverse mass after the selection using the 2012 dataset in each of the $H{+}0j$, $H{+}1j$ and $H{+}2j$ analyses, for the $e\mu{+}\mu e$ and $ee{+}\mu\mu$ final states separately using 8 TeV data. In general, an excess of events relative to the total background is seen.

Similar tables for the re-analysis of the 7 TeV data are presented in Section 5.8. Tables C.2, C.3, and C.4 of Appendix C show the observed and expected event yields using the 2011 dataset. The corresponding m_{T} distributions for the 7 TeV data are shown in Fig. 5.26 for $H{+}0j$ and $H{+}1j$ categories combining all final lepton flavours.

The summary for the ggF production mechanism of the Higgs boson, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis is presented in Fig. 5.33. It shows the transverse mass distribution after the full selection for $N_{\mathrm{jets}} \leq 1$ combining both 2011 and 2012 data periods. It can be seen that the region with $m_{\mathrm{T}} > 150\,\mathrm{GeV}$ is depleted of signal contributions. The level of agreement of the data with the expectation in the high m_{T} region, which is different from those used to normalise the backgrounds, illustrates the quality of the background estimates. The lower insert of the plot shows that the observed spectrum after the total estimated background is subtracted from the data is similar to that expected from the Higgs boson signal, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. Equivalently, for the VBF production mode of the Higgs boson, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, Fig. 5.34 shows the transverse mass distribution after the full selection for the $H{+}2j$ category considering $e\mu{+}\mu e$ final states and combining the 2011 and 2012 datasets.

The summary of the observed and expected yields for the signal and background processes after the full event selection is given in Tab. 5.21 for the 8 TeV data and in Tab. 5.22 for the 7 TeV data. To reflect better the sensitivity of the analysis additional thresholds on m_{T} have been applied, as specified in Section 5.7.5: $0.75\,m_{\mathrm{H}} < m_{\mathrm{T}} < m_{\mathrm{H}}$ for $N_{\mathrm{jets}} \leq 1$ and $m_{\mathrm{T}} < 1.2\,m_{\mathrm{H}}$ for $N_{\mathrm{jets}} \geq 2$. Nevertheless, note that the full transverse mass range is used for the fit procedure, as detailed in Section 5.11.2. The yield results are shown for the $e\mu{+}\mu e$ and $ee{+}\mu\mu$ final states combined. The VBF process contributes 2%, 12%, and 81% of the Higgs boson ($m_{\mathrm{H}} = 125\,\mathrm{GeV}$) expected yield in the signal region of the $H{+}0j$, $H{+}1j$, and $H{+}2j$ analyses, respectively. The uncertainties in the tables include the systematic uncertainties discussed in Section 5.10 and correspond to those entering the statistical procedure. The total uncertainty on the background is calculated accounting for the correlations among the individual backgrounds and includes all contributions. The total number of observed events in the m_{T} window defined above is 218 in the 7 TeV and 1195 in the 8 TeV data, to be compared with the total expected SM background 213 ± 13 and 1040 ± 60, respectively.

The main sources of systematic uncertainties are summarised in Tab. 5.23. As for Tabs. 5.21 and 5.22, the values are for events in the m_{T} range. Moreover, the constraints from CRs are included. The uncertainties are shown by source rather than by their impact on the signal or a particular background. The leading sources,

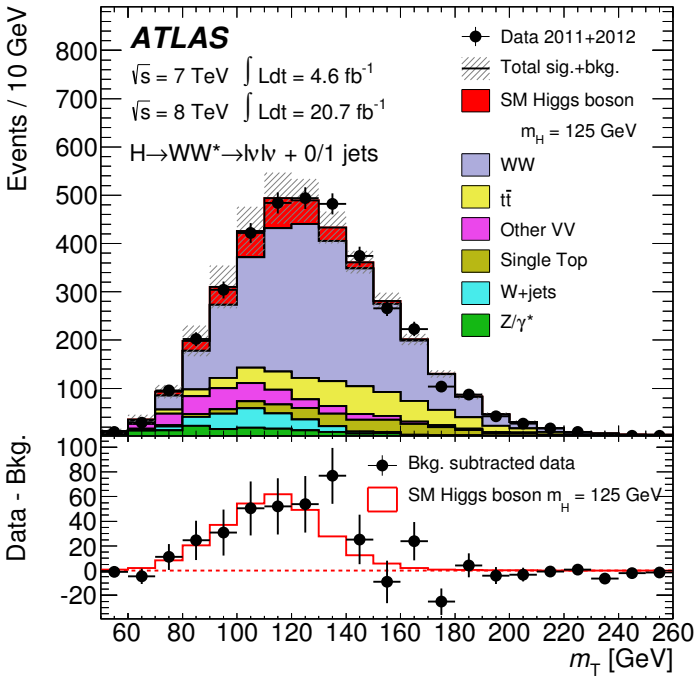

FIGURE 5.33: The transverse mass distributions for events passing the full selection of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis summed over all lepton flavours for final states with $N_{\text{jets}} \leq 1$. The signal is stacked on top of the background in red. The hatched area represents the total uncertainty on the sum of the signal and background yields from statistical, experimental, and theoretical sources. In the bottom part of the plot, the residuals of the data with respect to the estimated background are shown, compared to the expected m_{T} distribution of a Higgs boson with $m_{\text{H}} = 125\,\text{GeV}$.

i.e., those resulting in at least 4% uncertainty on the total signal or background yield in at least one N_{jets} category, are reported. Larger uncertainties from the QCD renormalisation and factorisation scales affect the predicted distribution of the ggF signal among the exclusive jet bins and can produce migration between categories. Their impact on the signal yield is summarised in Table 5.23, in addition to other non-negligible contributions: parton shower and underlying event modelling, as well as acceptance uncertainties due to QCD scale variations. The largest impact on the signal expectation comes from the knowledge of the jet energy scale and resolution, which is up to 6% in the H+2j category.

For the WW background in the $N_{\text{jets}} \geq 1$ final states, the theoretical uncertainties on the transfer factors include the impact of missing higher-order QCD corrections, PDF variations, and MC modelling choices, as described in Section 5.10.

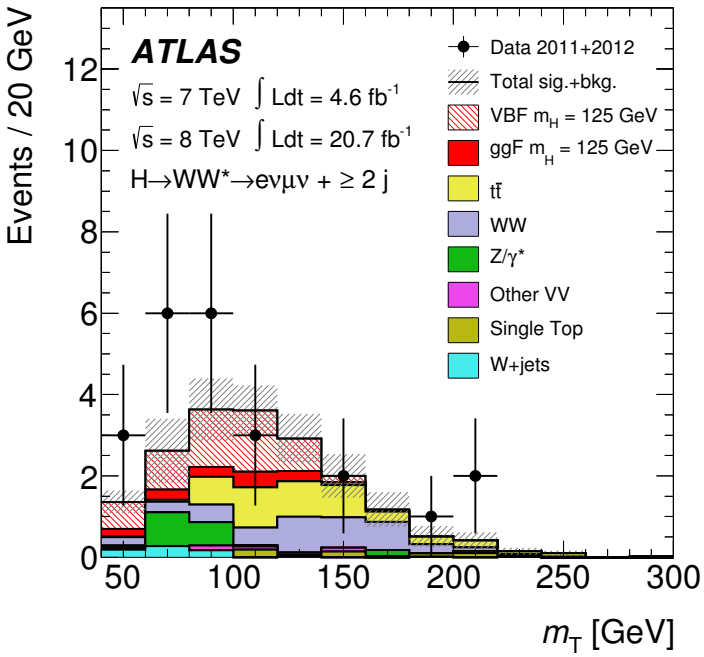

FIGURE 5.34: The transverse mass distributions for events passing the full selection of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis for $e\mu$+μe final states with $N_{\text{jets}} \geq 2$. The signal of the Higgs boson with $m_{\text{H}} = 125\,\text{GeV}$ is stacked on top of the background. It is shown separately for the ggF and VBF production processes. The hatched area represents the total uncertainty on the sum of the signal and background yields from statistical, experimental, and theoretical sources.

They amount to $\pm 1\%$ and $\pm 2\%$ relative to the total predicted background in the $N_{\text{jets}} = 0$ and $N_{\text{jets}} = 1$ final states, respectively. For the WW yield in the $N_{\text{jets}} \geq 2$ category, which is obtained from simulation, the systematic uncertainty relative to the total expected background is 4%. The leading uncertainties on the top-quark background are experimental. The b-tagging efficiency is the most important of these, and it appears in Tab. 5.23 primarily through its effect on this background.

The W+ jets transfer factor uncertainty is dominated by differences in the jet composition between dijet and W+ jets samples as covered in Section 5.9.2. The uncertainties on the muon and electron transfer factors are treated as correlated among the N_{jets} categories but uncorrelated with each other. The impact on the total background uncertainty is at most $\pm 2.5\%$. The main uncertainty on the Drell-Yan contribution in the H+0j and H+1j categories comes from the use of the f_{recoil} (Section 5.6.3.2) efficiency for the estimation of the Drell-Yan contamination in the signal region for the ee+$\mu\mu$ final states, as described in Section 5.9.4.

TABLE 5.21: The number of events observed in the 8 TeV data and expected from a Higgs signal ($m_{\mathrm{H}} = 125\,\mathrm{GeV}$) and backgrounds inside the transverse mass regions $0.75\,m_{\mathrm{H}} < m_{\mathrm{T}} < m_{\mathrm{H}}$ for $N_{\mathrm{jets}} \leq 1$ and $m_{\mathrm{T}} < 1.2\,m_{\mathrm{H}}$ for $N_{\mathrm{jets}} \geq 2$. All lepton flavours are combined. The total background is shown at the top of the table and its main components are shown at the bottom. Note that for $N_{\mathrm{jets}} \geq 2$ the ggF contribution of the Higgs boson signal is added to the total background expectation. The quoted uncertainties include the statistical and systematic contributions and account for anticorrelations between the background predictions.

N_{jets}	N_{obs}	N_{bkg}	N_{sig}
$= 0$	831	739 ± 39	100 ± 21
$= 1$	309	261 ± 28	41 ± 14
≥ 2	55	36 ± 4	10.9 ± 1.4

N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	$N_{Z/\mathrm{DY}}$	$N_{W+\,\mathrm{jets}}$
551 ± 41	58 ± 8	23 ± 3	16 ± 2	30 ± 10	61 ± 21
108 ± 40	27 ± 6	68 ± 18	27 ± 10	12 ± 6	20 ± 5
4.1 ± 1.5	1.9 ± 0.4	4.6 ± 1.7	0.8 ± 0.4	22 ± 3	0.7 ± 0.2

TABLE 5.22: The number of events observed in the 7 TeV data and expected from a Higgs signal ($m_{\mathrm{H}} = 125\,\mathrm{GeV}$) and backgrounds inside the transverse mass regions $0.75\,m_{\mathrm{H}} < m_{\mathrm{T}} < m_{\mathrm{H}}$ for $N_{\mathrm{jets}} \leq 1$ and $m_{\mathrm{T}} < 1.2\,m_{\mathrm{H}}$ for $N_{\mathrm{jets}} \geq 2$. All lepton flavours are combined. The total background is shown at the top of the table and its main components are shown at the bottom. Note that for $N_{\mathrm{jets}} \geq 2$ the ggF contribution of the Higgs boson signal is added to the total background expectation. The quoted uncertainties include the statistical and systematic contributions and account for anticorrelations between the background predictions.

N_{jets}	N_{obs}	N_{bkg}	N_{sig}
$= 0$	154	161 ± 11	25 ± 5
$= 1$	62	47 ± 6	7 ± 2
≥ 2	2	4.6 ± 0.7	1.2 ± 0.2

N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	$N_{Z/\mathrm{DY}}$	$N_{W+\,\mathrm{jets}}$
113 ± 10	12 ± 2	5 ± 1	4 ± 1	6 ± 3	21 ± 5
16 ± 6	5 ± 1	10 ± 3	6 ± 2	5 ± 2	5 ± 1
0.7 ± 0.2	-	0.7 ± 0.4	0.1 ± 0.1	2.4 ± 0.4	0.3 ± 0.2

5.12.2 7 TeV Results

Using the data collected in 2011 the expected significance for the Higgs boson at $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ is $1.8\,\sigma$, which corresponds to $p_0 = 0.04$. The observed significance using the 2011 data is $0\,\sigma$, which translates into $p_0 = 0.5$. This result is compatible

TABLE 5.23: The leading systematic uncertainties on the expected event yields for the 8 TeV analysis. All numbers are summed over lepton flavours. The first four rows are calculated for inclusive N_{jets} modes and redistributed to exclusive ones, as shown in Section 5.10. The QCD scale uncertainties on the inclusive ggF cross sections are anti-correlated between the exclusive N_{jets} modes. Sources contributing less than 4% to any column, and individual entries below 1%, are indicated with '-'.

	Signal processes (%)			Background processes (%)		
Source	H+0j	H+1j	H+2j	H+0j	H+1j	H+2j
Theoretical uncertainties						
QCD scale for ggF signal for $N_{\text{jets}} \geq 0$	13	-	-	-	-	-
QCD scale for ggF signal for $N_{\text{jets}} \geq 1$	10	27	-	-	-	-
QCD scale for ggF signal for $N_{\text{jets}} \geq 2$	-	15	4	-	-	-
QCD scale for ggF signal for $N_{\text{jets}} \geq 3$	-	-	4	-	-	-
Parton shower and UE model (signal only)	3	10	5	-	-	-
PDF model	8	7	3	1	1	1
$H \to WW$ branching ratio	4	4	4	-	-	-
QCD scale (acceptance)	4	4	3	-	-	-
WW normalisation	-	-	-	1	2	4
Experimental uncertainties						
Jet energy scale and resolution	5	2	6	2	3	7
b-tagging efficiency	-	-	-	-	7	2
f_{recoil} efficiency	1	1	-	4	2	-

with 1.8 σ with a Higgs boson signal at $m_{\text{H}} = 125\,\text{GeV}$. The highest value of $0.8\,\sigma$ ($p_0 = 0.22$) occurs at $m_{\text{H}} = 158\,\text{GeV}$.

The best-fit value of the signal strength at $m_{\text{H}} = 125\,\text{GeV}$ is $\mu = 0.0 \pm 0.6$. This result is consistent with the previous published analysis [68] using the 7 TeV data, $\mu = 0.5 \pm 0.7$ at $m_{\text{H}} = 125\,\text{GeV}$.

5.12.3 8 TeV Results

Using the data collected during the 2012 year the expected significance for the signal with $m_{\text{H}} = 125\,\text{GeV}$ is $3.5\,\sigma$ corresponding to $p_0 = 2 \times 10^{-4}$. The observed significance using the 2012 dataset is $4.3\,\sigma$, which corresponds to $p_0 = 1 \times 10^{-5}$. The highest value of $4.5\,\sigma$ ($p_0 = 4 \times 10^{-6}$) occurs at $m_{\text{H}} = 135\,\text{GeV}$. The best-fit signal strength μ at $m_{\text{H}} = 125\,\text{GeV}$ is,

$$\begin{aligned}\mu_{\text{obs, 8 TeV}} &= 1.26 \pm 0.24\,(\text{stat.}) \pm 0.21\,(\text{th. syst.}) \pm 0.14\,(\text{exp. syst.}) \pm 0.06\,(\text{lumi.}) \\ &= 1.26 \pm 0.35\ .\end{aligned}$$

The expected best-fit μ at $m_{\text{H}} = 125\,\text{GeV}$ is given below,

$$\begin{aligned}\mu_{\text{exp}} &= 1 \pm 0.23\,(\text{stat.}) \pm 0.23\,(\text{syst.}) \\ &= 1 \pm 0.33\,.\end{aligned}$$

The expected best-fit μ value in Ref. [142], obtained using $13\,\text{fb}^{-1}$ at $\sqrt{s} = 8\,\text{TeV}$, is $1 \pm 0.32\,(\text{stat.}) \pm 0.42\,(\text{syst.})$. For both analyses, the systematic uncertainty includes a small but non-negligible contribution from the statistical uncertainty on the analysis inputs derived from simulation. The expected precision is improved relative to Ref. [142] primarily because of the reduced extrapolation uncertainties from the WW CR in $H{+}0j$ and the increase in integrated luminosity. The statistical and systematic uncertainties are comparable in the present analysis.

5.12.4 Combined 7 TeV and 8 TeV Results

In this section, the expected and observed results are given for the combination of the 7 TeV and 8 TeV data periods for all jet multiplicities and all final lepton flavours. The expected significance of the Higgs signal with $m_{\text{H}} = 125\,\text{GeV}$ is $3.7\,\sigma$, which corresponds to $p_0 = 1 \times 10^{-4}$. The observed significance at $m_{\text{H}} = 125\,\text{GeV}$ is $3.8\,\sigma$ and it corresponds to $p_0 = 8 \times 10^{-5}$. The highest value of $4.1\,\sigma$ ($p_0 = 2 \times 10^{-5}$) occurs at $m_{\text{H}} = 140\,\text{GeV}$. Figure 5.35 shows the p_0 curves for the expected and the observed data. The Higgs boson signal expectation is for $m_{\text{H}} = 125.5\,\text{GeV}$, which corresponds to the best fit mass from the combined $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ results [143]. The shape and normalisation of the expected and observed p_0 curves are in agreement within the $\pm 1\,\sigma$ uncertainty band on the expected p_0 values across the mass range.

To measure the Higgs boson production strength, the parameter μ is determined from a fit to the data using the profile likelihood ratio for a fixed mass hypothesis corresponding to $m_{\text{H}} = 125\,\text{GeV}$. The excess of events corresponds to an observed strength parameter compatible with the Higgs boson prediction with $m_{\text{H}} = 125\,\text{GeV}$,

$$\begin{aligned}\mu_{\text{obs}} &= 1.01 \pm 0.21\,(\text{stat.}) \pm 0.19\,(\text{th. syst.}) \pm 0.12\,(\text{exp. syst.}) \pm 0.04\,(\text{lumi.}) \\ &= 1.01 \pm 0.31.\end{aligned}$$

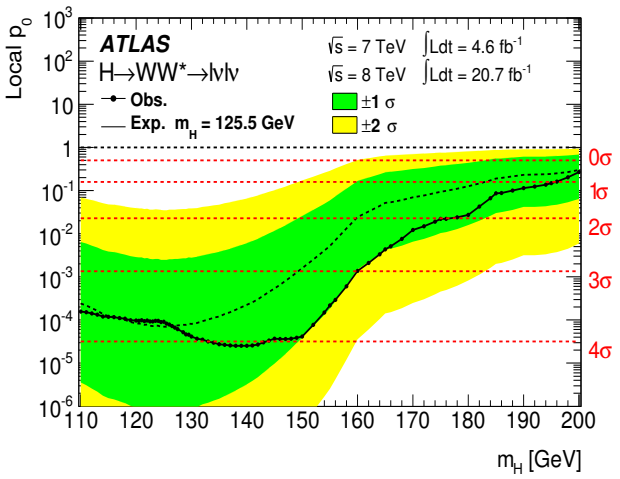

FIGURE 5.35: The expected (dashed) and observed (solid) local p_0 values as a function of m_{H} for the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis of the combined 7 TeV and 8 TeV data. The green band indicates $\pm 1\,\sigma$ uncertainty on the expected p_0 curve, and the yellow band represents $\pm 2\,\sigma$ uncertainty.

Table 5.24 lists the sources of the uncertainties on μ. The dominant systematic uncertainty is the theoretical uncertainty on the WW background normalisation. Another important contribution is the experimental systematic uncertainty, which is dominated by contributions from the b-tagging efficiency and the jet energy scale and resolution, as Tab. 5.23 shows. A significant contribution comes from the normalisation of the signal yield including the uncertainty on the cross section and the branching ratio.

TABLE 5.24: Leading uncertainties on the signal strength μ for the combined 7 and 8 TeV analysis.

Category	Source	Uncertainty, up (%)	Uncertainty, down (%)
Statistical	Observed data	+21	−21
Theoretical	Signal yield ($\sigma \cdot \mathcal{B}$)	+12	−9
Theoretical	WW normalisation	+12	−12
Experimental	Objects and DY estimation	+9	−8
Theoretical	Signal acceptance	+9	−7
Experimental	MC statistics	+7	−7
Experimental	W+jets fake factor	+5	−5
Theoretical	Backgrounds, excluding WW	+5	−4
Luminosity	Integrated luminosity	+4	−4
Total		+32	−29

Figure 5.36 shows the profile likelihood from the 2011+2012 analyses with two parameters of interest, μ and m_{H}. The scan of the likelihood ratio is shown in the μ-m_H plane. The $H \to WW^{(*)} \to \ell\nu\ell\nu$ result is compared to that of $H \to ZZ^{(*)} \to 4\ell$ [144] and $H \to \gamma\gamma$ [145]. The best-fit signal strength values for the three categories are consistent at 68% CL.

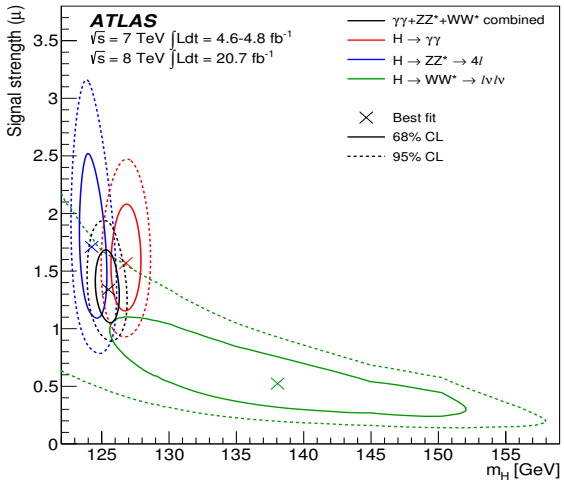

FIGURE 5.36: Two-dimensional likelihood contours in the (μ, m_{H}) plane for the $H \to \gamma\gamma$, $H \to ZZ \to 4\ell$ and $H \to WW^{(*)} \to \ell\nu\ell\nu$ analyses. The markers indicate the best-fit estimates in each case. The 68% (solid) and 95% (dashed) CL are shown. Mass scale systematic uncertainties are treated as uncorrelated between the three decay modes for the individual contours.

5.12.5 Separation of VBF and ggF Production Mechanisms

The measurements of the signal strength described in Section 5.12.4 do not provide direct information of the relative contributions of the different production mechanisms of the Higgs boson. However, in addition to the total signal strength, the signal strength of different production processes can be determined exploiting the sensitivity offered by the use of different jet multiplicities in this analysis.

The $7+8$ TeV combined data are fitted separating the VBF production mode of the Higgs boson, with $m_{\mathrm{H}} = 125$ GeV, from the ggF production mechanism. Statistical tests of a VBF signal are performed on the 7 TeV and 8 TeV data by considering the ggF signal as part of the background. The test defines μ_{VBF}, the signal strength parameter associated with the VBF process, as the parameter of interest. The ggF signal strength μ_{ggF} is profiled, and is constrained mainly by the $N_{\mathrm{jets}} \leq 1$ signal regions.

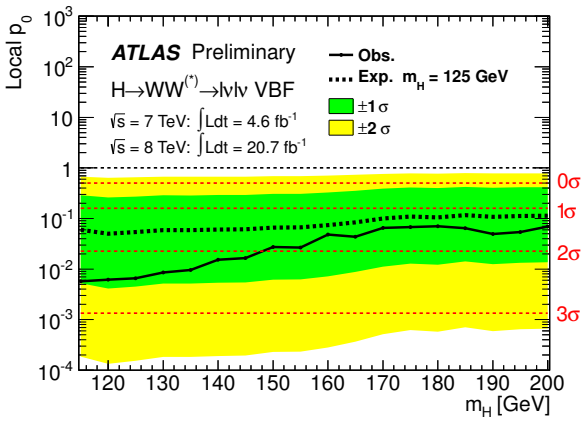

FIGURE 5.37: The expected (dashed) and observed (solid) local p_0 values as a function of the mass of the Higgs boson produced via VBF in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis combining 7 TeV and 8 TeV data. The green band indicates $\pm 1\,\sigma$ uncertainty on the expected p_0 values, and the yellow band represents $\pm 2\,\sigma$ uncertainties.

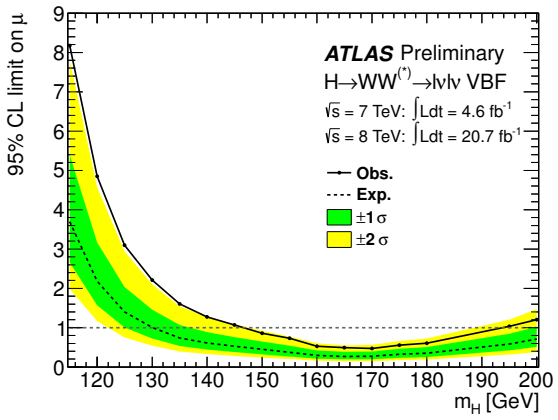

FIGURE 5.38: VBF production mode results for 95% CL upper limit using combined 7 TeV and 8 TeV data. The expected 95% CL upper limit is computed in the absence of the Higgs signal produced via VBF. The green band indicates $\pm 1\,\sigma$ uncertainty on the expected p_0 values, and the yellow band represents $\pm 2\,\sigma$ uncertainties.

Figure 5.37 compares the observed p_0 curve with the expected distribution in the presence of the Higgs boson with, $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, produced via the VBF process. The observed and expected results are compatible within the $\pm 1\,\sigma$ uncertainty band on the expected p_0 curve.

The 95% CL exclusion on the VBF strength parameter is shown in Fig. 5.38. The expected VBF signal significance at $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ is $1.6\,\sigma$, which corresponds to $p_0 = 0.05$. The observed significance from the 2011+2012 data result is $2.5\,\sigma$, which corresponds to $p_0 = 0.007$. The highest observed significance corresponds to $p_0 = 0.006$ and occurs at mass $m_{\mathrm{H}} = 115\,\mathrm{GeV}$. The observed strength parameter measured for the VBF production mode is consistent with the Higgs boson prediction with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$,

$$\begin{aligned} \mu_{\mathrm{obs,\ VBF}} &= 1.66 \pm 0.67\,(\mathrm{stat.}) \pm 0.42\,(\mathrm{syst.}) \\ &= 1.66 \pm 0.79. \end{aligned}$$

Similarly, μ_{ggF} has been measured using the $7+8\,\mathrm{TeV}$ combined data by considering the VBF signal as part of the background. In this test, μ_{VBF} is constrained mainly by the $H+2j$ signal region. The best-fit signal strength at $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ is,

$$\begin{aligned} \mu_{\mathrm{obs,\ ggF}} &= 0.82 \pm 0.24\,(\mathrm{stat.}) \pm 0.28\,(\mathrm{syst.}) \\ &= 0.82 \pm 0.36. \end{aligned}$$

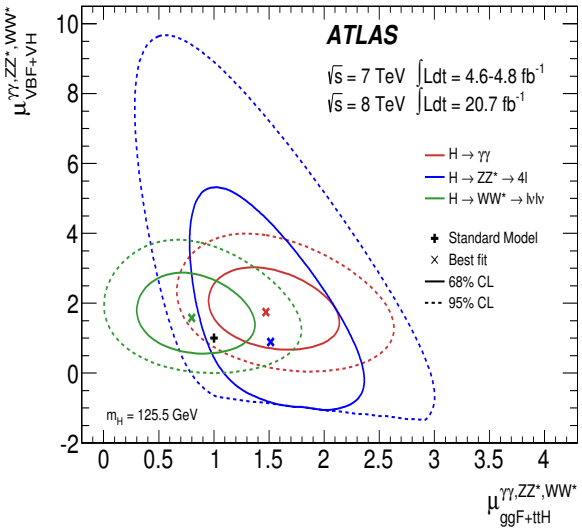

FIGURE 5.39: Likelihood contours in the ($\mu_{\mathrm{ggF+ttH}}$, $\mu_{\mathrm{VBF+VH}}$) plane for the $H \to \gamma\gamma$, $H \to ZZ^{(*)} \to 4\ell$, and $H \to WW^{(*)} \to \ell\nu\ell\nu$ final states for a Higgs boson with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. The best-fit to the data (x) and the 68% (solid) and 95% (dashed) CL curves are indicated, as well as the SM expectation (+).

The data are fitted separating vector-boson-mediated processes, VBF and VH, from gluon-mediated processes, ggF and ttH. Two signal strength parameters are introduced $\mu_{\text{ggF+ttH}} = \mu_{ggF} = \mu_{ttH}$ and $\mu_{\text{VBF+VH}} = \mu_{VBF} = \mu_{VH}$, which scale the SM-predicted rate to that observed. The results are shown in Fig. 5.39 for the different considered final states, $H \to \gamma\gamma$, $H \to ZZ^{(*)} \to 4\ell$, and $H \to WW^{(*)} \to \ell\nu\ell\nu$. The 95% CL contours of the measurements are consistent with the SM expectation.

A model-independent test can be done by measuring the ratio $\frac{\mu_{VBF+VH}}{\mu_{ggF+ttH}}$. The results of the fit to the data with the likelihood $\Lambda(\mu_{VBF+VH}/\mu_{ggF+ttH})$ are shown in Fig. 5.40.

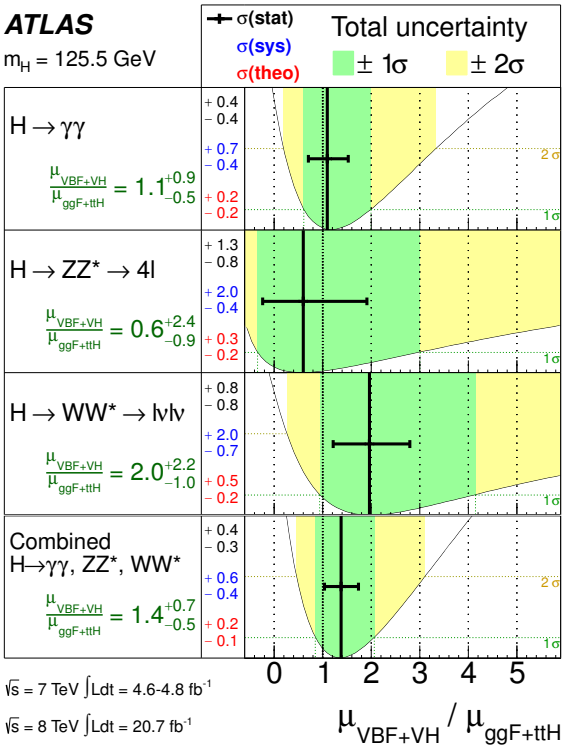

FIGURE 5.40: Measurements of the $\mu_{VBF+VH}/\mu_{ggF+ttH}$ ratios for the individual diboson final states and their combination, for a Higgs boson with $m_{\text{H}} = 125\,\text{GeV}$. The best-fit values are represented by the solid vertical lines, with the total $\pm 1\,\sigma$ and $\pm 2\,\sigma$ uncertainties indicated by the dark- and light-shaded band, respectively. The statistical uncertainties are indicated by the superimposed horizontal error bars. The numbers in the second column specify the contributions of the statistical uncertainty (top), the total (experimental and theoretical) systematic uncertainty (middle), and the theoretical uncertainty (bottom) on the signal cross section alone. The distributions of the likelihood ratios from which the total uncertainties are extracted are overlaid.

The $H\to WW^{(*)}\to \ell\nu\ell\nu$ result has a larger best-fit ratio,

$$\frac{\mu_{VBF+VH}}{\mu_{ggF+ttH}} = 2.0^{+2.2}_{-1.0}\,, \tag{5.37}$$

compared with the other channels. However this result is consistent with the $H\to\gamma\gamma$ and $H\to ZZ^{(*)}\to 4\ell$ analyses at 68% CL.

5.12.6 Total Cross Section Measurement

In the statistics formalism the signal strength μ, is defined as a factor scaling the MC yield to the observed data. In case we have K number categories, different for the final state flavour combinations and jet binning, the likelihood can be written as,

$$\mathcal{L}(\mu,\vec{\theta}) = \prod_{i=1}^{K} \mathcal{L}_i(\mu N_i^{\mathrm{MC}},\vec{\theta})\,, \tag{5.38}$$

where μ is the signal strength and $\vec{\theta}$ represents a set of nuisance parameters. The quantities N_i^{MC} can be written as,

$$N_i^{MC} = A_i(\theta)\sigma L\,, \tag{5.39}$$

where σ is the sum of the cross sections of the processes implemented in the MC simulation, L is the luminosity of the data sample and A_i is the signal acceptance for the category i as a function of the nuisance parameters that affect the systematic on the signal acceptance. Substituting Eq. 5.39 in Eq. 5.38 the likelihood takes the form,

$$\mathcal{L}(\mu,\vec{\theta}) = \prod_{i=1}^{K} \mathcal{L}_i(A_i(\theta)\mu\sigma L,\vec{\theta}) = \mathcal{L}(\sigma_{\mathrm{obs}},\ \vec{\theta})\,, \tag{5.40}$$

where we have defined $\sigma_{\mathrm{obs}} = \mu\sigma$. Equation 5.40 makes explicit that fitting μ is the same as fitting the total SM cross section σ.

The observed value of the product of the inclusive cross section $\sigma(pp\to H)$ and branching ratio $\mathcal{B}(H\to WW)$ for a Higgs boson signal at $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, is

$$\begin{aligned}(\sigma\cdot\mathcal{B})_{\mathrm{obs,\ 8\,TeV}} &= 6.0 \pm 1.1\,(\mathrm{stat.}) \pm 0.8\,(\mathrm{th.\ syst.}) \pm 0.7\,(\mathrm{exp.\ syst.}) \pm 0.3\,(\mathrm{lumi.})\,\mathrm{pb}\\ &= 6.0 \pm 1.6\,\mathrm{pb}.\end{aligned}$$

The corresponding expected value is

$$\begin{aligned}(\sigma\cdot\mathcal{B})_{\text{exp, 8 TeV}} &= 4.8\pm0.6\,(\text{cross section})\pm0.2\,(\text{branching ratio})\,\text{pb}\\ &= 4.8\pm0.7\,\text{pb}\,.\end{aligned}$$

The expected value is slightly smaller than the observed value, but they are consistent within the uncertainties.

The predicted value of $\sigma\cdot\mathcal{B}$ has been computed as the sum of the values reported in Ref. [21] (updated in Ref. [146]) for the production modes (ggF, VBF, and VH) used to evaluate the signal acceptance. The associated theoretical uncertainties are added linearly following the prescription in Ref. [146].

A cross-check of the results has been done using the yields in the m_{T} ranges described in Section 5.7.5 in lieu of the distribution. Table 5.21 gives the corresponding event yields and the uncertainties. The expected significance for the signal at $m_{\text{H}}=125\,\text{GeV}$ is lower at $3.3\,\sigma$ ($p_0=4\times10^{-4}$). The corresponding observed significance is $4.0\,\sigma$ ($p_0=3\times10^{-5}$).

The compatibility between the 7 TeV and 8 TeV results was tested using as parameter of interest in the statistic test the ratio of the best-fit μ values for each year, $\frac{\mu_{2012}}{\mu_{2011}}$. The statistical test comparing the 7 TeV and 8 TeV results shows that they are compatible within 1.8 σ.

5.13 Conclusions

The analysis of the $H\to WW^{(*)}\to\ell\nu\ell\nu$ process in the mass range of 115–200 GeV is presented in this chapter using the complete data sample of 2012 and 2011. The samples correspond to $20.7\,\text{fb}^{-1}$ at $\sqrt{s}=8\,\text{TeV}$ and $4.6\,\text{fb}^{-1}$ at $\sqrt{s}=7\,\text{TeV}$ collected with the ATLAS detector at the LHC.

An excess of events over the expected background is observed for $m_H\lesssim150\,\text{GeV}$. The signal significance at $m_{\text{H}}=125\,\text{GeV}$ is 3.8 standard deviations ($p_0=8\times10^{-5}$); the best fit signal strength at that mass is $\mu=1.01\pm0.31$. The observed VBF signal significance at $m_{\text{H}}=125\,\text{GeV}$ is 2.5 σ, with an observed strength parameter of $\mu=1.66\pm0.79$.

The measured value of the product of the cross section and the $WW^{(*)}$ branching ratio for a Higgs boson signal, with $m_{\text{H}}=125\,\text{GeV}$, at 8 TeV is $6.0\pm1.6\,\text{pb}$ while the expected value is $4.8\pm0.7\,\text{pb}$. The results are consistent with the predictions for the SM Higgs boson decaying to a pair of W bosons.

Chapter 6

Missing Transverse Momentum Optimisation

6.1 Introduction

After the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis described in Chapter 5 several investigations were focussed on optimising the selection of this search. The optimisation procedure relies on the complete 8 TeV ATLAS data to improve the background estimates and extend the phase space of the signal regions. The final $H \to WW^{(*)} \to \ell\nu\ell\nu$ results considering the new optimised selection are presented in Chapter 7.

One of the most significant improvements comes from optimising the $\not{E}_T$ criteria and the related m_{T} discriminant used in the likelihood fit. These studies exploit the differences between the expected Higgs boson signal, with $m_H = 125\,\mathrm{GeV}$, and the remaining SM backgrounds after the complete $H \to WW^{(*)} \to \ell\nu\ell\nu$ selection. There are several aspects that benefit the $\not{E}_T$ optimisation comparing with previous iterations. Mainly, the $\not{E}_T$ optimisation gains in statistics and more reliable running conditions. For the former, the enhance in statistics is originated by increasing the Higgs signal acceptance by lowering lepton p_{T} thresholds and including softer triggered events. This allows to evaluate the $\not{E}_T$ in each $H \to WW^{(*)} \to \ell\nu\ell\nu$ category with high enough statistics. Earlier studies were performed at the beginning of the event selection combining different final states in order to avoid statistical fluctuations, specially during 2011. For the latter, having the complete 2012 dataset allows to perform the investigations simulating the same running conditions as in collected data. For optimisation studies collected data are not considered due to the blinded

criteria, hence the procedure is only based on MC simulation. The MC mimics data pile-up conditions through the pile-up reweighting technique [147]. The full 2012 pile-up environment is considered in the $\not\!\!E_T$ optimisation for the first time. Previously, only initial data-taking periods were considered while more proton-proton collisions were being produced at the same time.

In addition, the $\not\!\!E_T$ optimisation mainly benefits from a new developed $\not\!\!E_T$ reconstruction. This new definition relies on the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ approach but including calorimeter depositions to account for particles escaping from ID detection. As extensively reported in Chapter 4, such particles are not entering in the original $E_{\mathrm{T}}^{\mathrm{miss,track}}$ computation, originating a deficient measurement and a poor reconstruction in some event topologies. However, the correction introduced by the new definition achieves the best resolution and stability against pile-up in events with jets. Moreover, this provides the highest separation power between the Higgs boson signal and the SM background when it is used to compute m_{T}.

The aim of this chapter is to present the improvements on the $\not\!\!E_T$ reconstruction methods and the results obtained optimising the criteria used in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ search. The definition of the improved version of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ reconstruction is presented in Section 6.2. An overview on the improvements performed in the calorimeter-based $\not\!\!E_T$ definitions is given in Section 6.3. Section 6.4 presents the performance results obtained comparing all four different $\not\!\!E_T$ reconstruction methods, including the new improved definition. Optimisation studies in each of the categories in which the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis is divided are described in Section 6.5. The results on the optimal $\not\!\!E_T$ usage and threshold values are presented, as well as the evaluation of other $\not\!\!E_T$-dependent variables. Finally, the performance, resolution and optimisation of the m_{T} variable is covered in Section 6.6.

6.2 Track-based Missing Transverse Momentum

It has been well studied in Chapter 4 that the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ reconstruction is very robust against pile-up interactions and provides the best resolution in events without jets. In light of this, a lower bound on the $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ quantity is used in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis to further suppress Z+jets contribution in ee+$\mu\mu$ with $N_{\mathrm{jets}} \leq 1$ final states (see Section 5.6.3). However, the performance of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ degrades quickly with the number of jets in the final state. As a consequence, the

$E_{\mathrm{T}}^{\mathrm{miss,track}}$ reconstruction is not used in the $H+2j$ analysis category nor in $e\mu+\mu e$ final states. The main two reasons ruining the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ measurement are,

- the spatial coverage of the ID ($|\eta| < 2.4$) limits the information on forward objects,
- the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ jet-term is mis-measured due to the presence of neutral particles which scape to the ID detection.

In order to improve the performance of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ reconstruction, a calorimeter-based correction is applied for recovering lepton' energies in the original $E_{\mathrm{T}}^{\mathrm{miss,track}}$ computation. This correction also allows to consider lepton radiation energy losses, as they are properly measured by the calorimeters. The improvements applied to the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ reconstruction are described in Section 6.2.1. The natural extension of the above correction applied to well reconstructed jets results in a new $\not\!\!E_T$ measurement defined in Section 6.2.2.

6.2.1 Improvements on $E_{\mathrm{T}}^{\mathrm{miss,track}}$ Reconstruction

The nominal $E_{\mathrm{T}}^{\mathrm{miss,track}}$ relies on tracks measured by the ID and associated to the PV, as defined in Section 4.2.3. In order to properly take into account lepton energy losses, such as electron radiative losses from bremsstrahlung, it is necessary to correct lepton components in the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ computation, as these are not measured by the ID. The aim of this correction is to add leptons' energy measured by the calorimeters, so it can replace the track p_{T} if the track matches one of the selected lepton. In general, the correction considers leptons defined by the $H\to WW^{(*)}\to \ell\nu\ell\nu$ strategy described in Chapter 7 (analysis leptons). In addition, leptons satisfying the set of requirements established by the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction (see Tab. B.1) are also included. The latter will be referred to as non-analysis leptons. Note that adding non-analysis leptons can introduce pile-up dependence. Thus, it is crucial to require a specific criterion to minimise the selection of non-analysis leptons originated from pile-up vertices. This is achieved by applying an impact parameter condition (maximum allowed distance of the lepton vertex with respect to the PV): $|z_0 \sin\theta| < 1.0\,\mathrm{mm}$, to non-analysis leptons as well (see Section 5.6.1).

Once the selected leptons are added to the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ calculation, the already included tracks associated to them have to be excluded. In other words, the p_{T} of the tracks are replaced by the fully reconstructed energy of the associated leptons. In light of this, the correction applied to the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ reconstruction benefits from higher energetic resolution since the fully reconstructed energy from depositions in

the calorimeters is used now instead of the measured p_{T} by the ATLAS tracker system. The track-lepton association is performed within a ΔR cone of 0.05.

Making the track association to the leptons that pass the quality criteria described above, the next two situations have to be considered.

- It is possible to add leptons in the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ calculation which do not have any associated tracks. Note that this can be expected if the tracks produced by the lepton do not satisfy the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ selection. This effect has been measured in simulated $Z \rightarrow ee$ events where $\sim 1\%$ of the events are found having one electron which is not matching to any selected track [148]. Thus, in addition to a precise energy measurement from the calorimeters, this leptons' correction provides a more accurate $E_{\mathrm{T}}^{\mathrm{miss,track}}$ measurement. The presence of these unmatched leptons motivates the extension of consistency with the PV criterion to the non-analysis leptons as well in order to minimise pile-up contributions.

- There are tracks which are not associated to any selected lepton after the track-lepton matching is done. Besides the ATLAS tracker does an excellent job at reconstructing tracks, there is a percentage of around $4 \times 10^{-3}\%$ of events in $Z \rightarrow \mu\mu$ simulated process having some tracks that are catastrophically measured with very high energies which are not associated to any selected lepton in the event. These tracks will be referred in the following as mis-reconstructed tracks. The mis-reconstructed tracks are studied using truth information from simulation finding to be mostly low-p_{T} pions and kaons [148]. They are found to be very isolated, with a large relative error on their transverse momentum ($\frac{\Delta p_T}{p_T}$), and tend to be pointing opposite to the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ direction. These characteristics allow to make a selection criteria in order to remove the mis-reconstructed tracks from the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ computation.

Finally, the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ computation including lepton's corrections can be expressed as follows,

$$E_{\mathrm{T}}^{\mathrm{miss,track}} = -\sum_{\text{i tracks}}^{\mathrm{PV}} \vec{p_{\mathrm{T}}}^{\,i} + \sum_{\ell \text{ leptons}} \left(\vec{p_{\mathrm{T}}}^{\,\ell,\mathrm{trk}} - \vec{p_{\mathrm{T}}}^{\,\ell,\mathrm{calo}} \right), \tag{6.1}$$

where the subindex i runs over all tracks associated to the PV of the event, $\vec{p_{\mathrm{T}}}^{\,\ell,\mathrm{trk}}$ refers to all tracks (trk) associated to the lepton (ℓ), and $\vec{p_{\mathrm{T}}}^{\,\ell,\mathrm{calo}}$ is the fully reconstructed p_{T} of the lepton measured by the calorimeters. The first term in Eq. 6.1

is the nominal $E_{\mathrm{T}}^{\mathrm{miss,track}}$ definition given in Eq. 4.11 and the second term contains the calorimeter correction for leptons.

6.2.2 Jet-Corrected Track-Based $\not{E}_T$ Definition: $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$

Until now, we have only considered two kinds of objects in the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ calculation: tracks and leptons. In other words, we only apply corrections assuming events without neutral particles in the final state. The extension of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ to events with jets, in which the latter are explicitly corrected in the calculation, brings us to the new jet-corrected and track-based reconstruction which will be referred in the following as $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$. The aim of the explicit jet implementation in the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ shares the same purpose as for the leptons case, i.e, including fully jet reconstructed energy in the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ computation. In addition, neutral particles which are not measured in the tracker system but they deposit their energy in the detector are also included. To use the full energy of the jets, the sum of momenta of the tracks in jets are replaced by their calorimeter energy in a similar way as the correction is applied for the leptons' case.

The jets used in the correction are analysis jets. They have $p_{\mathrm{T}} > 25(30)\,\mathrm{GeV}$ for $|\eta| < (>)2.4$, and they are required to have $|JVF| > 0.5$ for $p_{\mathrm{T}} < 50\,\mathrm{GeV}$. Tracks are considered to be associated with a jet if they are within a ΔR cone of 0.4 of the jet. The $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ is defined by the following expression,

$$E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} = -\sum_{\text{i trks}} \vec{p_{\mathrm{T}}}^{\,i} + \sum_{\ell \text{ leptons}} \left(\vec{p_{\mathrm{T}}}^{\,\ell,\mathrm{trk}} - \vec{p_{\mathrm{T}}}^{\,\ell,\mathrm{calo}}\right) + \sum_{\text{j jets}} \left(\vec{p_{\mathrm{T}}}^{\,j,\mathrm{trk}} - \vec{p_{\mathrm{T}}}^{\,j,\mathrm{calo}}\right), \tag{6.2}$$

where the first two terms are defined as in Eq. 6.1. Thus, in events without jets, $E_{\mathrm{T}}^{\mathrm{miss,track}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ are identical. $\vec{p_{\mathrm{T}}}^{\,j,\mathrm{trk}}$ refers to all tracks associated to the jet (j) and $\vec{p_{\mathrm{T}}}^{\,j,\mathrm{calo}}$ is the jet area corrected transverse momentum of the jet.

Lastly, an overlap removal criterion between leptons and jets is applied and the case of mis-reconstructed tracks is extended to the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ as well. Additional track replacement is applied for unmatched high p_{T} tracks considering a softer jet selection. The replacement of mis-reconstructed tracks matching low p_{T} jets by the energy of the latter provides an extra improvement on the resolution of 20%. For illustration, Fig. 6.1 presents the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ distributions in $Z \to \ell\ell$ events with 1 jet in the final state for 2012 data and MC. Comparing both plots, it is clear how the correction of the jet achieves a significative reduction of the tail shown by the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ spectrum. Moreover, a much better

data/MC agreement is observed for in the high $\not\!\!E_T$ region ($> 60\,\text{GeV}$). Similar improvements are also observed for the directional $\not\!\!E_T$ components in events with at least 2 jets in the $Z \to \ell\ell$ events. These are shown in Fig. 6.2. Narrower width shapes are achieved after applying the correction on the jets. Comparisons with the calorimeter-based $\not\!\!E_T$ reconstruction methods are collected in Appendix D.

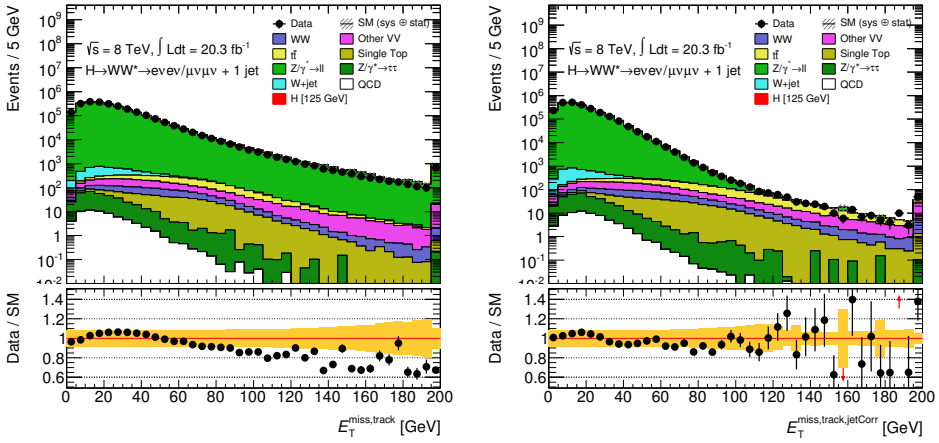

FIGURE 6.1: $E_{\mathrm{T}}^{\mathrm{miss,track}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ distributions in the $Z \to \ell\ell$ enriched region with exactly one jet with 8 TeV ATLAS data and simulation on the left right, respectively. ee+$\mu\mu$ final states are combined.

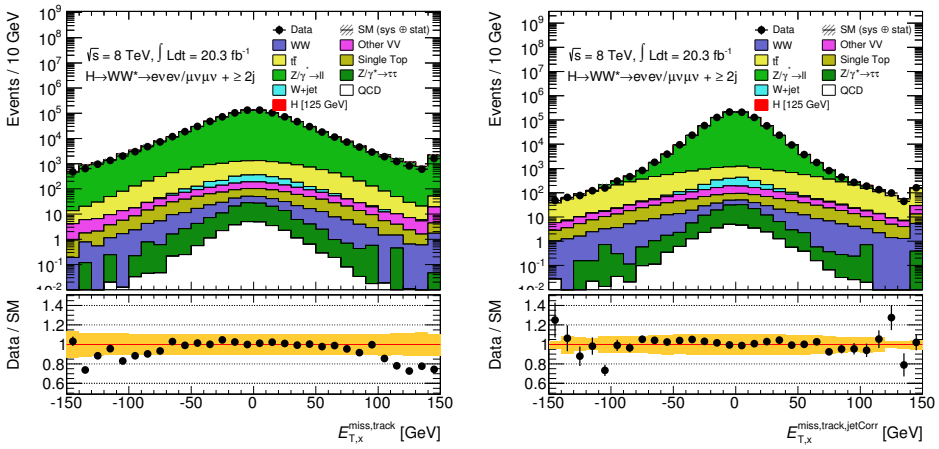

FIGURE 6.2: x-component of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ (left) and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ (right) reconstruction methods in the $Z \to \ell\ell$ enriched region with at least two jets with 8 TeV ATLAS data and simulation. ee+$\mu\mu$ final states are combined.

The two track-based $\not\!\!E_T$ definitions discussed above are not so track-based as they used to be. Instead, both of them incorporate measurements of the energy deposited by leptons in the calorimeters. However, the base of its construction relies on the selected tracks coming from the PV which still allows to reduce pile-up contributions very efficiently.

6.3 Improvements on $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$

The calorimeter-based $\not\!\!E_T$ variables are defined in Section 4.2.1 and 4.2.2. The aim of this section is to summarise the main improvement developed for them. The improvement affects to both calorimeter-based $\not\!\!E_T$ definitions, i.e., $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$.

The correction treats the energy loss of muons which in some cases could be entering twice in the calorimeter-based $\not\!\!E_T$ computations. When a muon deposits sufficient energy to form a topocluster, the topocluster is included either in a jet or in the CellOut term (soft term). However, the energy of the topocluster is also included into the p_T of the muon. In order to correct the double-counting, the measured energy loss is subtracted from the muon momentum when a set of specific identification criteria are satisfied. The effect of the correction has been studied using Higgs boson, with $m_H = 125\,\mathrm{GeV}$, decaying into a pair of W boson simulated samples. The results show that the reconstructed $\not\!\!E_T$ after the correction is closer to the truth $\not\!\!E_T$ and moreover, a better resolution is observed [148]. In light of the results, the correction on the muon energy loss has been considered for the $\not\!\!E_T$ optimisation in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis.

6.4 Performance of the $\not\!\!E_T$ Reconstruction

It is very important to check the performance of the improved $\not\!\!E_T$ reconstruction methods after all modifications adopted for each of them have been implemented. In addition, the new $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis object selection and trigger requirements have changed, as described in Chapter 7. These modifications have a direct impact in the $\not\!\!E_T$ performance, since the reconstructions depend on the analysis object selection criteria. In this section, a summary of comparative results using all four $\not\!\!E_T$ reconstruction methods are presented. These studies are indicative and they prove the superior performance of the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ reconstruction over the other $\not\!\!E_T$ measurements.

Data and simulation comparisons using the complete ATLAS 2012 dataset, with a total integrated luminosity of $20.3\,\text{fb}^{-1}$, for each of the $\not\!\!E_T$ reconstruction methods are collected in Appendix D.The first set of results are evaluated using the $Z \to \ell\ell$ selection, as this is defined in Section 4.3. The second collection contain comparisons in WW, $t\bar{t}$ and $Z \to \tau\tau$ enriched regions, which are defined in Sections 5.9.7, 5.9.6, and 5.9.5, respectively. The results are separated in different jet multiplicities and lepton flavours in the final state. Overall, there is good agreement between data and MC in all distributions. Besides some $\not\!\!E_T$ reconstruction methods present mis-modelling in $Z \to \ell\ell$ events, a satisfactory data/MC agreement is obtained for the $E_{\text{T}}^{\text{miss,track,jetCorr}}$ reconstruction.

The $\not\!\!E_T$ performance in events with genuine measurement is evaluated using the $H \to WW^{(*)} \to \ell\nu\ell\nu$ simulated process, with $m_H = 125\,\text{GeV}$. The ggF and VBF production mechanisms of the Higgs boson are evaluated separately by dividing the events by the number of jets in the final state. Figure 6.3 contains the relative difference on the directional $\not\!\!E_T$ components with respect to the expected measurement for all four $\not\!\!E_T$ reconstruction methods. The $E_{\text{T}}^{\text{miss,track}}$ and $E_{\text{T}}^{\text{miss,track,jetCorr}}$ are identical in events without jets and their distributions overlap. In processes with jets, the large improvement in resolution and correction of the bias when considering the jet correction in the $E_{\text{T}}^{\text{miss,track}}$ reconstruction is observed. The mean and RMS values from the x-component results shown in Fig. 6.3 are collected in Tab. 6.1. In addition, the number of events in the tails of each $\not\!\!E_T$ distribution is quoted as well.

$H \to WW^{(*)} \to \ell\nu\ell\nu$	**0-jet (ggF)**			**1-jet (ggF)**			**2-jet (VBF)**		
	Mean	RMS	Integral of tail	Mean	RMS	Integral of tail	Mean	RMS	Integral of tail
$E_{\text{T}}^{\text{miss}}$	0.011	1.4	39	0.05	1.5	28	0.03	1.5	14
$E_{\text{T}}^{\text{miss, STVF}}$	-0.025	1.3	34	0.021	1.5	27	0.07	1.5	15
$E_{\text{T}}^{\text{miss,track}}$	-0.022	1.2	28	0.13	1.8	40	0.41	2.0	26
$E_{\text{T}}^{\text{miss,track,jetCorr}}$	-0.022	1.2	28	0.020	1.4	23	0.05	1.4	13

TABLE 6.1: Summary on the performance resolution of the different $\not\!\!E_T$ reconstruction methods for a $H \to WW^{(*)} \to \ell\nu\ell\nu$ Higgs boson signal. The mean, RMS and integral of the tail values from the relative difference with respect to the truth value of Higgs boson signal simulated samples, with $m_H = 125\,\text{GeV}$, in 0, 1, and ≥ 2 jet multiplicities are given in each column. Mean and RMS values are given in GeV.

The tail contribution is defined as the integral of events with $|\not\!\!E_T - E_{\text{T}}^{\text{miss, True}}| > 80\,\text{GeV}$. In events without jets, the integral of the tail is lowest for the track-based $\not\!\!E_T$ reconstruction methods, as we expected from results in Chapter 4. In events with

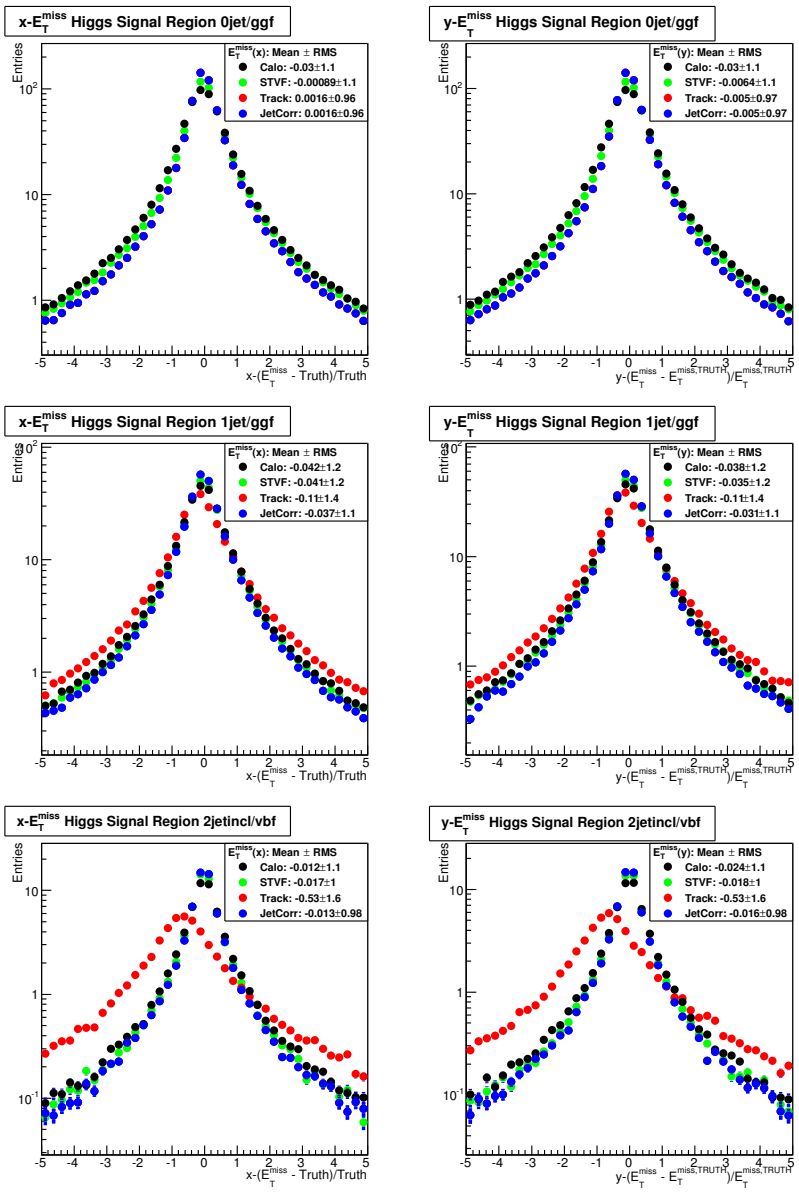

FIGURE 6.3: Relative difference on the x- and y-components of the four $\not{E}_T$ reconstruction methods with respect to the generated values by the MC simulation (Truth) are given on the left and right, respectively. Higgs signal simulation samples, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, are divided in the ggF (top and middle) and VBF (bottom) production mechanisms. The mean and RMS values from each distribution are quoted in the legend. The $E_{\mathrm{T}}^{\mathrm{miss,track}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ definitions are identical in events without jets.

jets, the changes introduced in the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ highly improve the performance compared with the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ results. Figure 6.4 shows the mean and RMS from the relative difference distributions in Fig. 6.3 as a function of $< \mu >$.

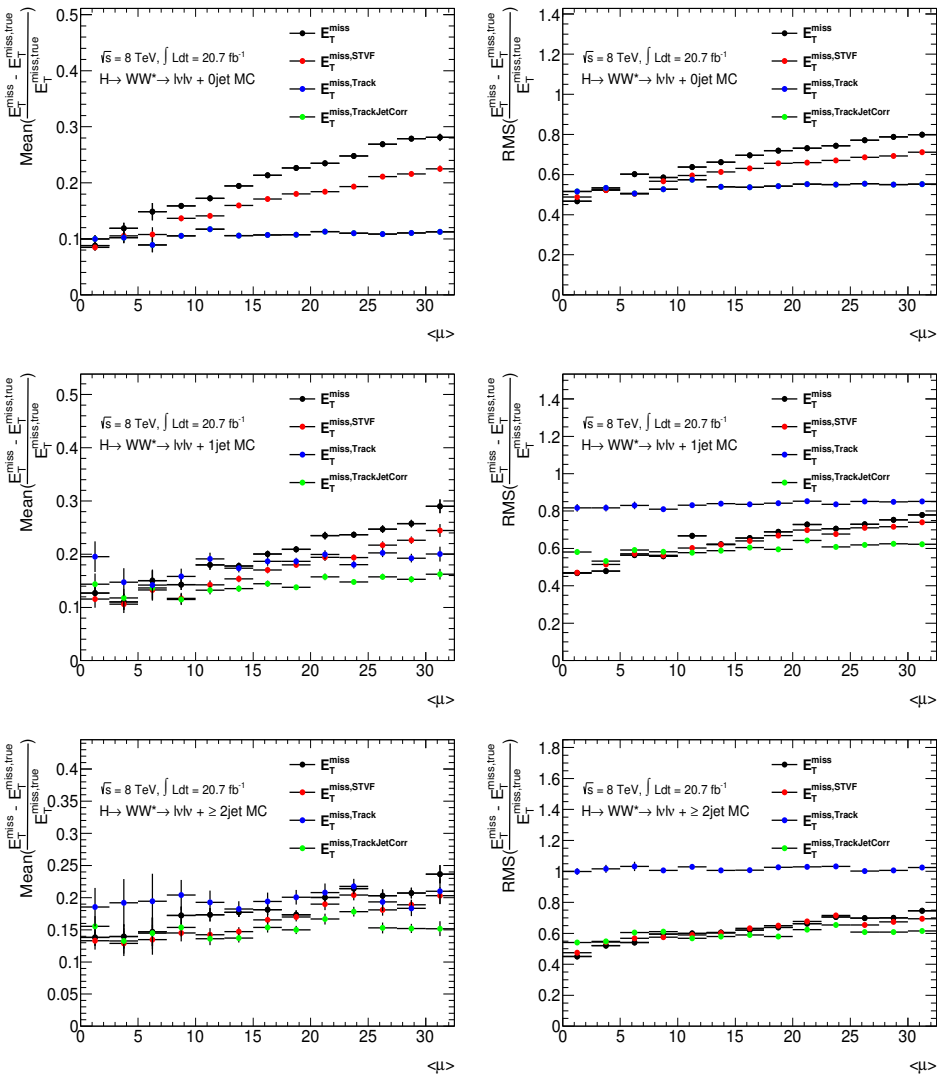

FIGURE 6.4: Resolution as a function of $< \mu >$ for each of the $\not{E}_T$ reconstruction methods divided in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ categories for the expected Higgs boson signal with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ compared with the generated value by simulation. The relative differences of the mean and RMS with respect to the true value from the MC are given on the left and right, respectively. The lepton flavours were summed. The $E_{\mathrm{T}}^{\mathrm{miss,track}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ are identical in events without jets.

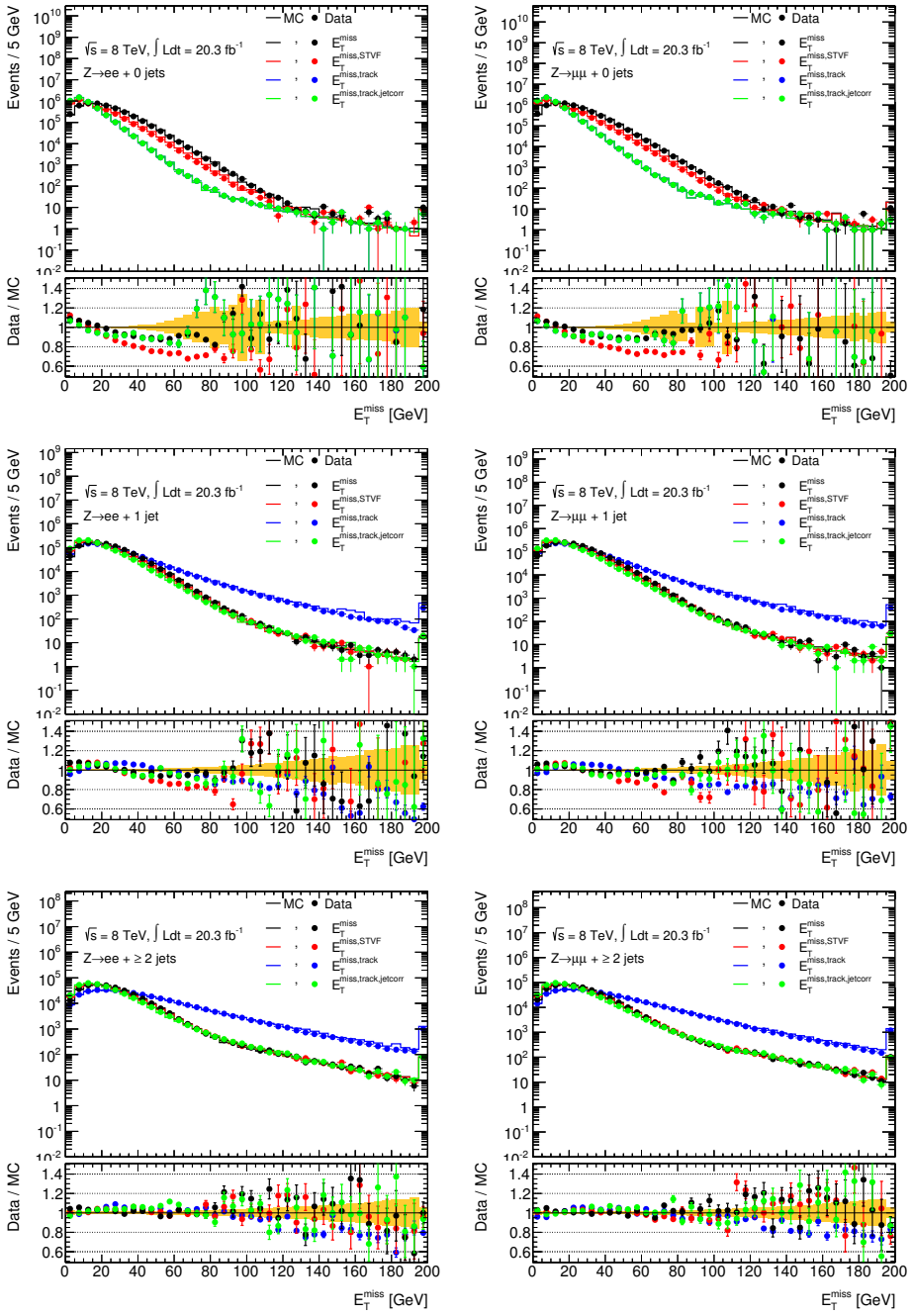

FIGURE 6.5: $\not{E}_T$ distributions in the $Z \to \ell\ell$ enriched regions with different number of jets in the final state in 8 TeV ATLAS data and MC simulation. The left side is $Z \to ee$ and the right side is $Z \to \mu\mu$. The values in the legend show the mean and RMS for each $\not{E}_T$ reconstruction and for easier comparison they are also shown in Tab. 6.2 for the 1 jet case. $E_{\mathrm{T}}^{\mathrm{miss,track}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ are identical in events without jets.

The performance improvements over the full 8 TeV datasets are also studied in events without genuine $\not{E}_T$ measurement. The $\not{E}_T$ distributions obtained in $Z \to \ell\ell$ events with different jet multiplicities are shown in Fig. 6.5. For easier comparison, the mean and RMS values for all four $\not{E}_T$ measurements in the $Z \to \mu\mu + 1\,j$ simulated process are collected in Tab. 6.2. The $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ has the lowest values in this process without expected $\not{E}_T$ measurement, even with presence of jet activity.

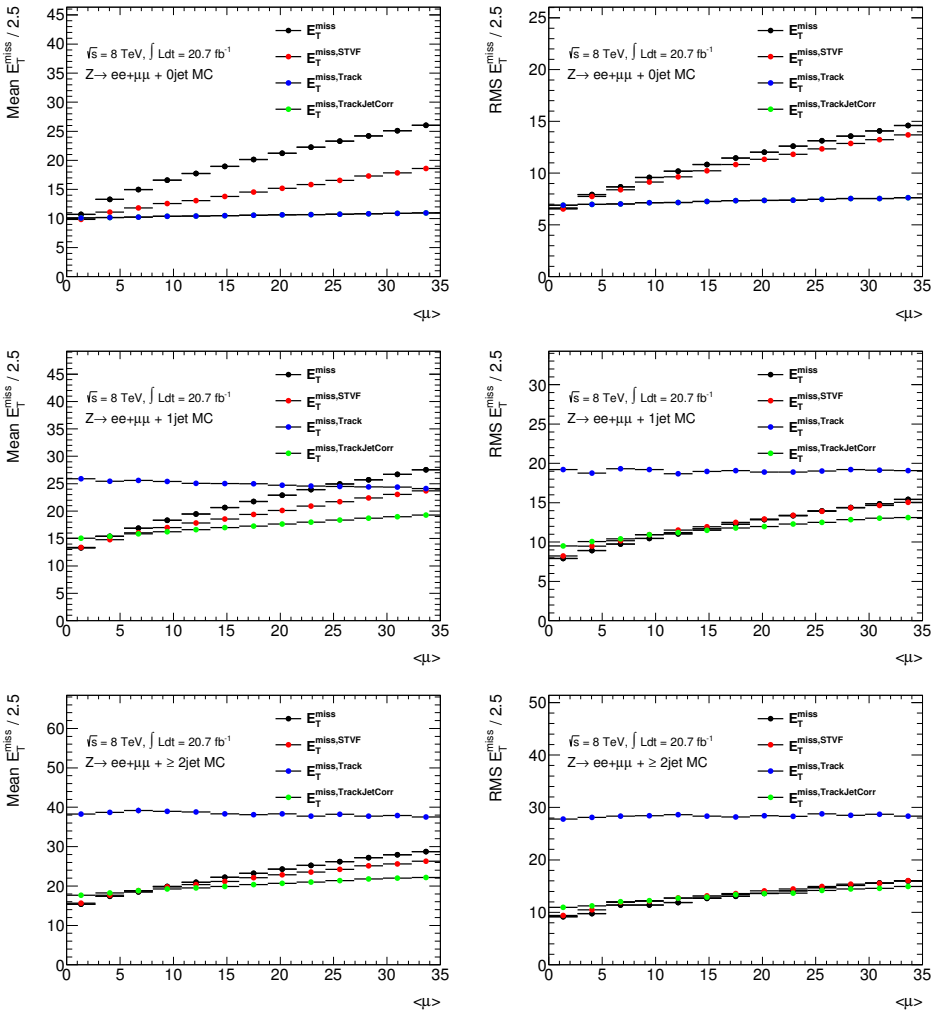

FIGURE 6.6: The mean (left) and RMS (right) values for each of the $\not{E}_T$ varieties using simulation for same flavour events in the Z control region for different jet multiplicities: 0 jet (top), 1 jet (middle) and ≥ 2 (bottom). $E_{\mathrm{T}}^{\mathrm{miss,track}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ are identical in events without jets.

	Mean	RMS	Integral of tail
$E_{\mathrm{T}}^{\mathrm{miss}}$	23	13.1	1543
$E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$	20	13.2	2437
$E_{\mathrm{T}}^{\mathrm{miss,track}}$	25	19.0	20274
$E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$	18	12.3	1163

TABLE 6.2: Summary on the performance resolution of the different $\not\!\!E_T$ reconstruction methods for the $H\to WW^{(*)}\to \ell\nu\ell\nu$ Higgs boson signal. The mean, RMS and integral of the tail values from the relative difference with respect to the generated value from the simulation. Mean and RMS values are given in GeV.

Figure 6.6 shows the dependence of the mean and the resolution for each of the $\not\!\!E_T$ varieties from Fig. 6.5 with respect to the mean number of interactions per bunch-crossing. The $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ is very stable with respect to pile-up, showing the lower trends than in calorimeter-based $\not\!\!E_T$ reconstruction methods while still maintaining the smallest mean and RMS values for the high $< \mu >$ range. All results presented up to here confirm that the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ has the best resolution and the smallest tail in different event topologies for processes with expected $\not\!\!E_T$ measurement, as well as in $Z \to \ell\ell$ events. The $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ reconstruction recovers the RMS value to the level of the calorimeter-based measurements, provides the lowest contribution of the tails, and corrects the positive bias introduced in the mean by mis-measured jets in the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ computation. The superior pile-up robustness of the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ and its lower mean and resolution in all jet multiplicities motivates the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis to evaluate the Higgs boson signal significance switching to the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ reconstruction in the selection strategy.

6.5 $\not\!\!E_T$ Optimisation in $H\to WW^{(*)}\to \ell\nu\ell\nu$ Search

From the performance studies presented above, the four different $\not\!\!E_T$ definitions,

- calorimeter-based $\not\!\!E_T$: $E_{\mathrm{T}}^{\mathrm{miss}}$;
- calorimeter-based $\not\!\!E_T$ with pile-up suppression: $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$;
- track-based $\not\!\!E_T$: $E_{\mathrm{T}}^{\mathrm{miss,track}}$;
- track-based $\not\!\!E_T$ with the correction for jets: $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$;

and their corresponding projected $\not\!\!E_{T,Rel}$ quantities (see Section 4.2.4.2) are evaluated in order to increase the Higgs boson signal significance with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis. This section focuses on the $\not\!\!E_T$ optimisation results obtained in each $H\to WW^{(*)}\to \ell\nu\ell\nu$ category. The optimisation procedure uses the full statistical fit to extract the expected significance of signal in different scenarios, in which different $\not\!\!E_T$ reconstruction methods and several lower bound values are studied. In general, and unless stated otherwise, $e\mu$+μe and ee+$\mu\mu$ final states are treated separately. To consider each $H\to WW^{(*)}\to \ell\nu\ell\nu$ signal region, the results are also divided by the number of jets: H+0j, H+1j and H+2j. In total, the $\not\!\!E_T$ optimisation is developed for six categories. Furthermore, when optimising the criteria on the $\not\!\!E_T$ measurement, all other analysis selections are applied following Chapter 5; with the exception of the inclusion of low-p_{T} leptons ($p_{\mathrm{T}} > 22,\,10\,\mathrm{GeV}$) and dilepton triggered events, as described in Chapter 7.

6.5.1 Different Flavour Channels: $e\mu$+μe

In the $e\mu$+μe category, the main remaining SM processes after the full selection present neutrinos in their final states, thus they produce genuine $\not\!\!E_T$ measurement as well as the Higgs boson signal does when decays into two W bosons. In light of this, it is expected to achieve poor background rejection power using a lower bound requirement on the $\not\!\!E_T$ measurement.

6.5.1.1 H+0j Analysis

Figure 6.7 presents the evolution on the background composition of the $\not\!\!E_T$ distribution at different stages of the event selection for $e\mu$+μe candidates in the H+0j category. Note that these distributions are obtained without any $\not\!\!E_T$ requirement and the data samples are not included in order not to bias the results. These plots show how the specific analysis requirements applied to select the final candidates sculpt the shape of the $\not\!\!E_T$ distribution. The background composition is specially changed after $p_{\mathrm{T}}^{\ell\ell}$ and $\Delta\phi_{\ell\ell}$ requirements because of the high rejection of the $Z/\mathrm{DY}\to\tau\tau$ background. These dilepton system variables are very correlated with the $\not\!\!E_T$ as shown in Fig. 6.8. Hence, it is expected that they change the background composition and, for extension, the shape of the $\not\!\!E_T$ for $e\mu$+μe final candidates.

Due to the requirements applied in the H+0j analysis with $e\mu$+μe final candidates, the background rejection power than can be achieved using a $\not\!\!E_T$ threshold is limited by the poor population of the low $\not\!\!E_T$ region. Fig. 6.7 illustrates this behaviour using the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ reconstruction but same effect is observed for the

other $\not\!\!E_T$ definitions. Appendix E collects the shape from MC simulation through the analysis selection stages for all four $\not\!\!E_T$ reconstruction methods.

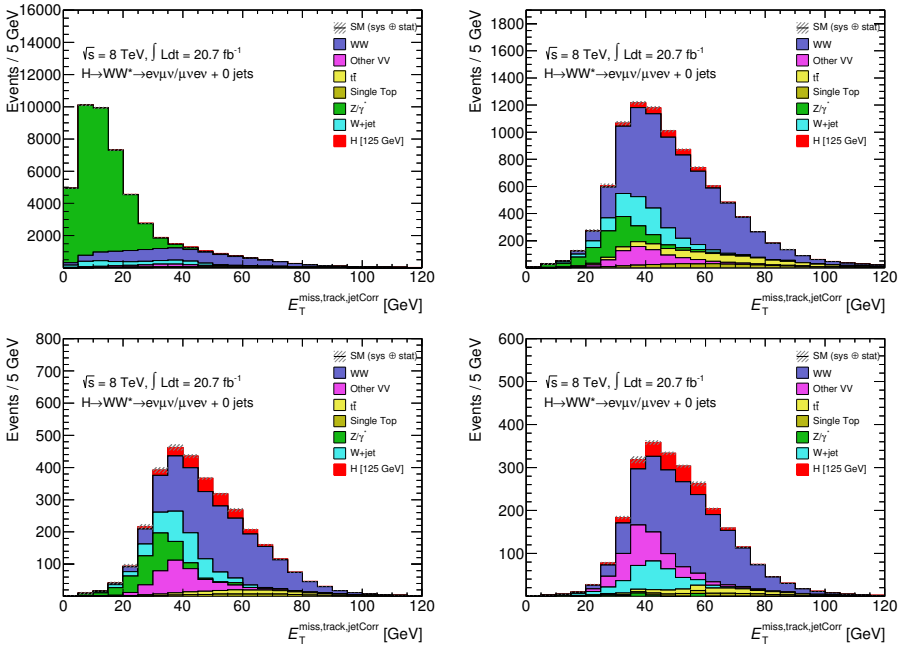

FIGURE 6.7: Evolution of $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ distribution through the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis selection of $e\mu$+μe final states in the H+0j category. The plots follow the selection in Section 5.7.2: jet veto (top left), $p_{\mathrm{T}}^{\ell\ell} > 30GeV$ (top right), $m_{\ell\ell} < 55\,\mathrm{GeV}$ (bottom left) and $\Delta\phi_{\ell\ell} < 1.8$ (bottom right).

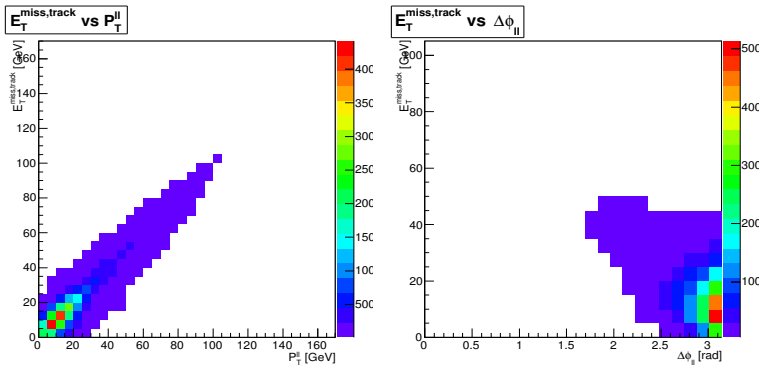

FIGURE 6.8: $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ as a function of $p_{\mathrm{T}}^{\ell\ell}$ (left) and $\Delta\phi_{\ell\ell}$ (right) correlation for the total expected background in $e\mu$+μe final states after jet veto but without $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold applied.

TABLE 6.3: Event yields from MC simulation for different $\not\!\!E_T$ scenarios for $e\mu$+μe final candidates in the H+0j category. The Higgs signal is considered for a mass of $m_H = 125\,\text{GeV}$ (first column) and the different background contributions are summed up in the second column. The last column represents the significance computed as in Eq. 6.3. Bottom table shows the total background separated by processes. The errors are only statistical.

	Signal [125 GeV]	Total Bkg.	$S/\sqrt{B}$
No $E_{\text{T}}^{\text{miss}}$ Cut	195.8 ± 1.6	2095 ± 13	4.28 ± 0.04
$E_{\text{T,Rel}}^{\text{miss}} > 25\,\text{GeV}$	175.1 ± 1.5	1860 ± 10	4.06 ± 0.04

	WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets	Multi-jets
No $E_{\text{T}}^{\text{miss}}$ Cut	1247.76	341.28	120.77	69.68	26.20	284.57	4.79
	± 5.14	± 6.37	± 1.30	± 0.71	± 8.75	± 4.73	± 0.29
$E_{\text{T,Rel}}^{\text{miss}} > 25\,\text{GeV}$	1153.73	278.66	109.08	64.38	9.03	243.11	2.06
	± 4.95	± 5.77	± 1.23	± 0.68	± 4.52	± 4.23	± 0.24

Table 6.3 shows the event yields for $e\mu$+μe final candidates in the H+0j analysis. This compares the $\not\!\!E_T$ selection used in Chapter 5 ($E_{\text{T,Rel}}^{\text{miss}} > 25$ GeV) with the scenario without $\not\!\!E_T$ requirement. From the quoted event yields it is clear that the total background rejection using $E_{\text{T,Rel}}^{\text{miss}} > 25$ GeV is of the level of the discarded Higgs boson signal with $m_{\text{H}} = 125\,\text{GeV}$: ∼11%. In light of this, there is a reduction of 5% in the signal significance computed by the figure of merit,

$$\frac{S}{\sqrt{B}}\,, \tag{6.3}$$

where S and B are the MC event yields for the Higgs boson and the total SM backgrounds respectively.

The shape differences between $\not\!\!E_T$ and its relative quantity have been studied as well. Figure 6.9 compares the $E_{\text{T}}^{\text{miss,track,jetCorr}}$ and $E_{\text{T,Rel}}^{\text{miss,track,jetCorr}}$ distributions for the Higgs signal shape and the total background for $e\mu$+μe final candidates in the H+0j analysis. The differences in shapes between $\not\!\!E_T$ and $\not\!\!E_{T,Rel}$ are small, since the projected quantity only populates $\sim 3-4\%$ more the low region of the spectrum. Moreover, it is found that for the same required threshold, up to 40 GeV, the Higgs signal acceptance is about 0.5% lower than the achieved background rejection when the $\not\!\!E_{T,Rel}$ is used instead of the non-projected variable.

Furthermore, the $E_{\text{T}}^{\text{miss,track,jetCorr}}$ presents the lowest Higgs signal contribution in the low $\not\!\!E_T$ range. The results are presented in Fig. 6.10 which shows the comparison of all the $\not\!\!E_T$ reconstruction methods using the simulated Higgs boson signal sample. Hence, applying the lower bound requirement on the $E_{\text{T}}^{\text{miss,track,jetCorr}}$ will enhance the signal efficiency in the H+0j category.

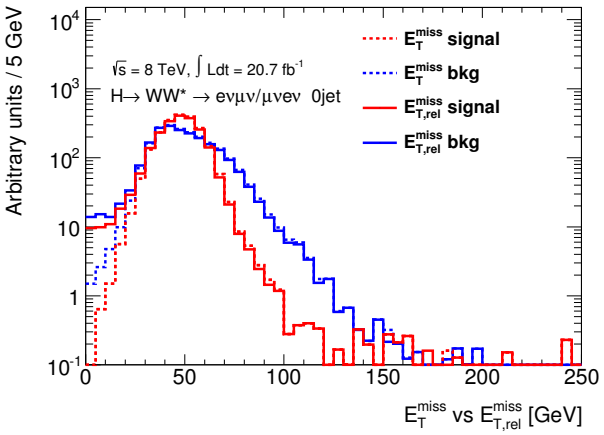

FIGURE 6.9: $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,track,jetCorr}}$ distributions for the expected Higgs boson signal (red), with $m_{\mathrm{H}} = 125[GeV]$, and the total SM background (blue) for $e\mu + \mu e$ final candidates in the $H{+}0j$ category.

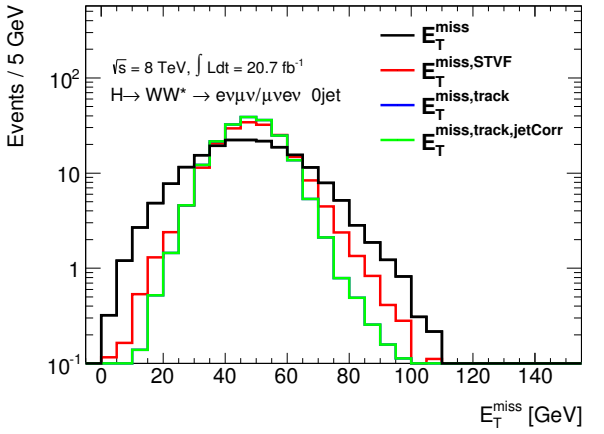

FIGURE 6.10: $\not\!\!E_T$ shape comparison for the Higgs signal sample at $m_H = 125\,\mathrm{GeV}$ from simulation. The events require leptons with different flavour and they satisfy the full $H{+}0j$ selection. The $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ has the lowest signal contribution in the low range: $\not\!\!E_T < 35\,\mathrm{GeV}$.

In order to find out which $\not\!\!E_T$ reconstruction and threshold is optimal, the result from each $\not\!\!E_T$ definition is obtained through the final statistical likelihood fit, which ensures to include all effects in both, signal and control regions. The fit is done using the m_{T} based on $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ as it improves the expected significance by 8% compared to result obtained with m_{T} based on $E_{\mathrm{T}}^{\mathrm{miss}}$ as discussed in Section 6.6.

TABLE 6.4: Expected Higgs boson, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, signal significances in $H\to WW^{(*)}\to\ell\nu\ell\nu$ analysis evaluating different $\not\!\!E_T$ reconstruction methods and lower bound values in steps of 5 GeV for $e\mu + \mu e$ final candidates in the H+0j category. The first row indicates the scenario without any $\not\!\!E_T$ selection applied.

Bound (GeV)	Nominal $\not\!\!E_T$		Projected $\not\!\!E_{T,Rel}$	
	$E_{\mathrm{T}}^{\mathrm{miss}}$	$E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$	$E_{\mathrm{T,Rel}}^{\mathrm{miss}}$	$E_{\mathrm{T,Rel}}^{\mathrm{miss,track,jetCorr}}$
0	3.238			
5	3.239	3.240	3.243	3.238
10	3.254	3.243	3.249	3.241
15	3.248	3.248	3.235	3.240
20	3.243	3.252	3.229	3.236
25	3.213	3.249	3.197	3.239

The scan is performed in steps of 5 GeV for the different $\not\!\!E_T$ definitions and its correspondent relative quantities. A summary of the fit results is given in Tab. 6.4, where the requirements on $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ are quoted. There is only about 2% difference in the expected significance between the different $\not\!\!E_T$ scenarios. This is consistent with the low background rejection that can be achieved given the low population in the low $\not\!\!E_T$ region and with Table 6.3, where the gain without $\not\!\!E_T$ requirement is discussed. Moreover, the requirement on $\not\!\!E_{T,Rel}$ quantities produces lower significance than using their non-projected $\not\!\!E_T$ partners. This is also in agreement with above discussion based on Fig. 6.9.

In light of the results, the requirement on $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} > 20$ GeV is chosen as the final cut for the $e\mu$+μe events in the H+0j analysis, when considering the following facts,

- it has better performance compared with respect to the other $\not\!\!E_T$ quantities. $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ has the best resolution and robustness against pile-up;
- it has the highest signal acceptance with respect to other $\not\!\!E_T$ varieties nor their projected $\not\!\!E_{T,Rel}$ quantities;
- the 20 GeV threshold is nearly the highest significance and improves it by 1.7% with respect to the previous requirement used up to now: $E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 25\,\mathrm{GeV}$;
- consistency in the definition of the $\not\!\!E_T$ requirement and the m_{T} used in the fit;
- consistency with respect to $e\mu$+μe final candidates in the H+1j analysis (see Section 6.5.1.2).

TABLE 6.5: Event yields from MC simulation for different $\not{E}_T$ scenarios with $e\mu$+μe final candidates in the H+0j category. The Higgs signal is considered for a mass of $m_H = 125\,\text{GeV}$ (first column) and the different background contributions are summed up in the second column. The last column represents the significance computed as in Eq. 6.3. Bottom table shows the total background separated by processes. The thresholds are given in units of GeV. The errors are only statistical.

	Signal [125 GeV]	Total Bkg.	$S/\sqrt{B}$
$E_{\text{T,Rel}}^{\text{miss}} > 25$	175.1 ± 1.5	1860 ± 10	4.06 ± 0.04
$E_{\text{T}}^{\text{miss,track,jetCorr}} > 20$	195.1 ± 1.6	2077 ± 13	4.28 ± 0.04

	WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets	Multi-jets
$E_{\text{T,Rel}}^{\text{miss}} > 25$	1153.73	278.66	109.08	64.38	9.03	243.11	2.06
	± 4.95	± 5.77	± 1.23	± 0.68	± 4.52	± 4.23	± 0.24
$E_{\text{T}}^{\text{miss,track,jetCorr}} > 20$	1245.57	338.85	119.42	69.25	22.83	277.14	3.75
	± 5.14	± 6.36	± 1.29	± 0.71	± 8.69	± 4.63	± 0.27

Table 6.5 shows the event yields at the end of the event selection for $e\mu$+μe final candidate events in the H+0j category. It compares the reference requirement $E_{\text{T,Rel}}^{\text{miss}} > 25\,\text{GeV}$ from Chapter 5 with the optimised threshold: $E_{\text{T}}^{\text{miss,track,jetCorr}} > 20$ GeV. There is a gain in the Higgs boson signal acceptance of about 11% using the new threshold. The total background yield with the new choice increases similarly. The highest background difference is observed in the Z+jets, which increase its rate by a factor ~ 3. However, the Z/DY is maintained at a manageable level since its contribution only represents about 1% of the total background contamination.

6.5.1.2 H+1j Analysis

This section evaluates the $\not{E}_T$ requirement for $e\mu$+μe final candidates in the H+1j analysis, similarly to the studies performed for H+0j in Section 6.5.1.1. The shape of the final candidates for the total background and signal is compared for each of the $\not{E}_T$ varieties and their corresponding $\not{E}_{T,Rel}$ quantity. Note that in this category, the projection is performed onto the nearest lepton or jet. After the full H+1j selection criteria, the $\not{E}_T$ spectrum is sculpted because of the correlation between the $\not{E}_T$ magnitude and dilepton topological requirements applied to this search, as well as it occurs in the H+0j analysis.

Table 6.6 compares the event yields for two $\not{E}_T$ scenarios. As before, the reference $\not{E}_T$ threshold from Chapter 5 ($E_{\text{T,Rel}}^{\text{miss}} > 25\,\text{GeV}$) is compared against the case without $E_{\text{T}}^{\text{miss}}$ requirement.

TABLE 6.6: Expected event yields from MC simulation for different $\not\!\!E_T$ scenarios with $e\mu + \mu e$ final candidates in the $H{+}1j$ category. The Higgs signal is considered for a mass of $m_H = 125\,\text{GeV}$ (first column) and the different background contributions are summed up in the second column. The last column represents the significance computed as in Eq. 6.3. Bottom table shows the total background separated by processes. The thresholds are given in units of GeV. The errors are only statistical.

	Signal [125 GeV]	Total Bkg.	$S/\sqrt{B}$
No $\not\!\!E_T$ cut	108.08 ± 0.55	1717.8 ± 10.6	2.60 ± 0.04
$E_{\text{T,Rel}}^{\text{miss}} > 25\,\text{GeV}$	68.01 ± 0.45	865.24 ± 5.67	2.31 ± 0.04

	WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets	Multi-jets
No $\not\!\!E_T$ cut	464.17	204.62	293.32	109.81	239.95	162.50	155.25
	± 2.91	± 4.67	± 2.00	± 1.04	± 5.91	± 6.74	± 1.57
$E_{\text{T,Rel}}^{\text{miss}} > 25$	343.25	98.53	222.11	84.12	24.84	69.21	14.20
	± 2.50	± 3.23	± 1.74	± 0.93	± 1.85	± 2.91	± 0.52

The expected Higgs boson signal yield increases by $\sim$ 40% when the $\not\!\!E_T$ requirement is dropped in this category. The total background is also enhanced by a factor 2, being the most increased processes W+jets, Z/DY, VV and multi-jets. These two facts result on the observed differences in the signal significance values. The scenario without $E_{\text{T}}^{\text{miss}}$ selection is not possible due to the need to reject the total background yield. Nevertheless, the improvement shown in Tab. 6.6 is related with the gain of the Higgs boson signal acceptance in the $H{+}1j$ analysis.

Compared to the $H{+}0j$ case, the W+jets background represents one of the main background sources in the $H{+}1j$ analysis. The main reason for the increase of the total background comparing with the results in Chapter 5 is the softer trigger requirements and the lower p_{T} thresholds applied to select the lepton candidates, specially in the W+jets process. The shape of the W+jets background is very signal-like in the m_{T} distribution which is used in the fit. Moreover, the systematic errors associated to this background are one of the highest. For these reasons, the optimisation studies are focus on the reduction specially of this background as well as keeping as much signal as possible.

The $\not\!\!E_T$ distributions for $H{+}1j$ $e\mu{+}\mu e$ final candidates extracted from simulation are presented in Fig. 6.11. As it can be observed, in the $E_{\text{T}}^{\text{miss,track,jetCorr}}$ distribution the backgrounds tend to be placed in the lower region of the spectrum compared with the other $\not\!\!E_T$ quantities. Hence, this variable would be able to reject more background than the rest of quantities. For easier illustration, the shape of the total background for each of the $\not\!\!E_T$ varieties are compared in Fig. 6.12.

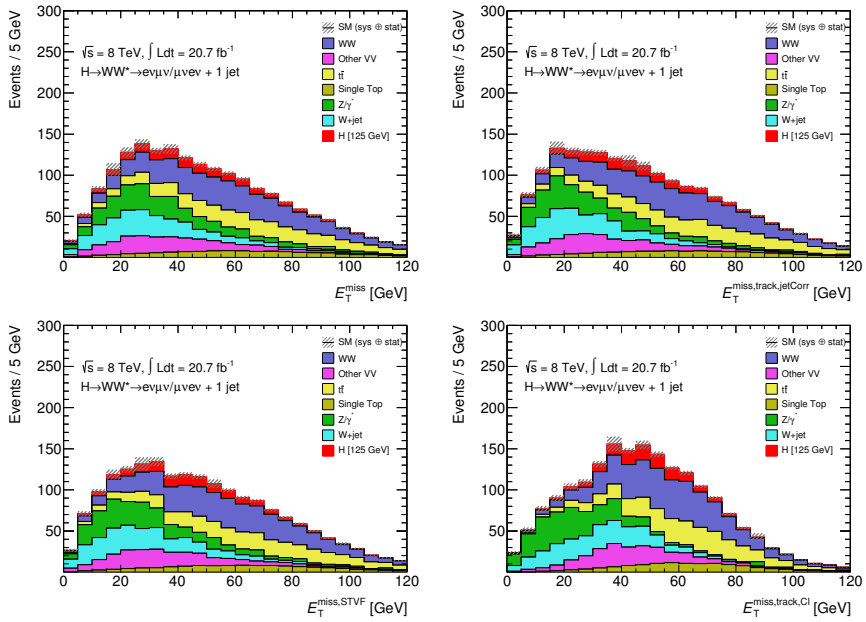

FIGURE 6.11: $\not{E}_T$ distributions from simulation after all event selection applied for $e\mu + \mu e$ final candidates in the $H{+}1j$ category. The different $\not{E}_T$ varieties are $E_{\mathrm{T}}^{\mathrm{miss}}$ in the top left, $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ in the top right, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ in the bottom left, and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ in the bottom right.

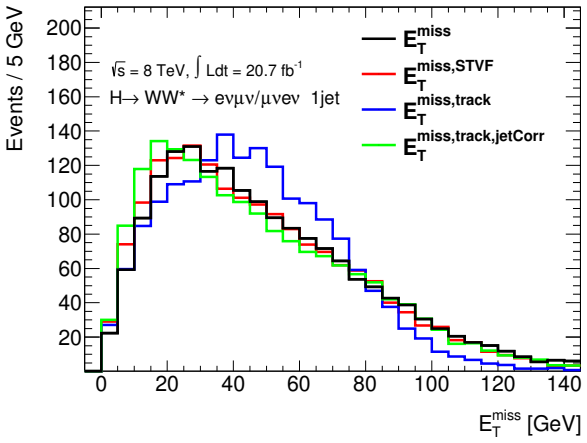

FIGURE 6.12: $\not{E}_T$ distributions different $\not{E}_T$ reconstruction methods for the expected SM background $e\mu{+}\mu e$ final candidates in the $H{+}1j$ category. The $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ (green line) spectrum has the highest background contribution in the low range: $\not{E}_T < 35\,\mathrm{GeV}$.

Besides a requirement on the $\not\!\!E_{T,Rel}$ quantity rejects more background events than using $\not\!\!E_T$, it also reduces the Higgs signal yield in the same level. Thus, a lower bound requirement on $\not\!\!E_T$ will preserve more signal events than using its projected quantity, as concluded from the studies on the $H+0j$ analysis discussed in Section 6.5.1.1. Figure 6.13 shows the differences between $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,track,jetCorr}}$ shapes for signal and background for the final candidates in the $H+1j$ analysis. About a 20% of the signal is included at the final candidates switching from the projected $\not\!\!E_{T,Rel}$ quantity to the nominal $\not\!\!E_T$ definition. In addition, the rejection power for background is softened at the level of the acceptance of the signal. In light of this, the improvement on the signal significance for this case is expected to be around 10%, as quoted in Tab. 6.6.

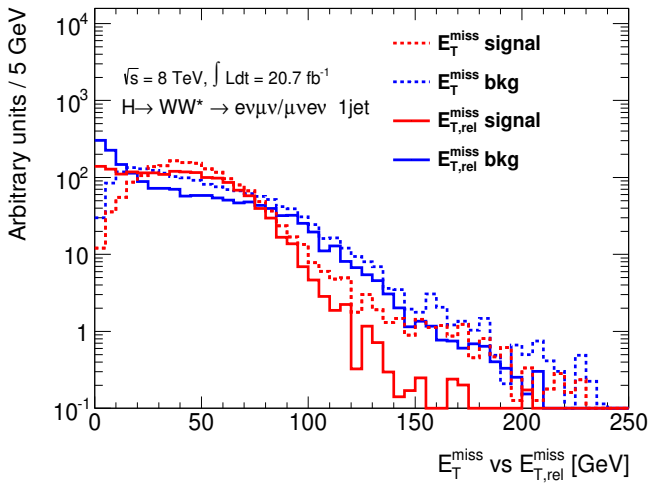

FIGURE 6.13: $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,track,jetCorr}}$ distributions for the expected Higgs signal with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ on the left and for all background on the right in $e\mu + \mu e$ final candidates for the $H+1j$ category. Bottom distributions show the relative difference of events per bin between $\not\!\!E_T$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ rejection.

The same optimisation procedure described for the $H+0j$ analysis is performed for the $H+1j$ case. The results summarised in Tab. 6.7 show the preference for lower cuts using the non-projected $\not\!\!E_T$ values. These results are consistent with the expectation using the shape and event yields comparisons for different $\not\!\!E_T$ scenarios. From the expected signal significance numbers, the lower bound threshold achieving the highest value is $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} > 10\,\mathrm{GeV}$. Besides this requirement enhances the Higgs signal yield by 55%, the rate on the $W+$jets and multi-jet processes increase by a factor 3.

TABLE 6.7: Expected Higgs boson, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, signal significances in $H\to WW^{(*)}\to\ell\nu\ell\nu$ analysis evaluating different $\not\!\!E_T$ reconstruction methods and lower bound values in steps of 5 GeV for $e\mu + \mu e$ final candidates in the $H{+}1j$ category. The first row indicates the scenario without any lower bound on $\not\!\!E_T$ applied.

Bound (GeV)	Nominal $\not\!\!E_T$		Projected $\not\!\!E_{T,Rel}$	
	$E_{\mathrm{T}}^{\mathrm{miss}}$	$E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$	$E_{\mathrm{T,Rel}}^{\mathrm{miss}}$	$E_{\mathrm{T,Rel}}^{\mathrm{miss,track,jetCorr}}$
0	3.238			
5	2.765	2.765	2.765	2.765
10	2.753	2.730	2.767	2.770
15	2.737	2.633	2.788	2.642
20	2.755	2.502	2.751	2.525
25	2.650	2.394	2.633	2.367
30	2.567	2.298	2.458	2.191

Aiming to reduce this background source, a new variable is introduced in the $H{+}1j$ analysis: the maximum of the transverse mass of the $\not\!\!E_T$ and one of the lepton candidates: $\max m_T(W)$. It is defined in Eq. 6.4.

$$\max m_T(W) = \sqrt{(p_T^{\ell_i} + \not\!\!E_T)^2 \;-\; |\vec{p}_T^{\,\ell_i} + \vec{\not\!\!E}_T|^2} \quad \text{where } \ell_i = 1, 2\,, \tag{6.4}$$

where the $p_T^{\ell_i}$ is the lepton candidate transverse momentum and $\not\!\!E_T$ is the neutrino transverse momentum of one of the W bosons. The $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ quantity is used to compute the $\max m_T(W)$ measurement because its higher performance and stability against high pile-up environment. The $\max m_T(W)$ quantity has a high separation power between Higgs boson signal, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, and processes like Z/DY and multi-jets as shown in Fig. 6.14.

The scan over different $\not\!\!E_T$ thresholds is performed again with the introduction of the $\max m_T(W) > 50\,\mathrm{GeV}$ requirement. The results show that the optimal threshold for the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ quantity is still 10 GeV but the difference with 20 GeV is below 1%. In addition, the introduction of the $\max m_T(W)$ benefits the $Z/\mathrm{DY}\to\tau\tau + 1\mathrm{jet}$ phase space, since it allows to increase the purity of this control region. The influence of the optimised $\not\!\!E_T$ selection in different control regions is discussed in Section 6.5.2. In light of the results, a combined requirement on $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} > 20\,\mathrm{GeV}$ and $\max m_T(W) > 50\,\mathrm{GeV}$ is selected for $e\mu{+}\mu e$ final candidates in the $H{+}1j$ category.

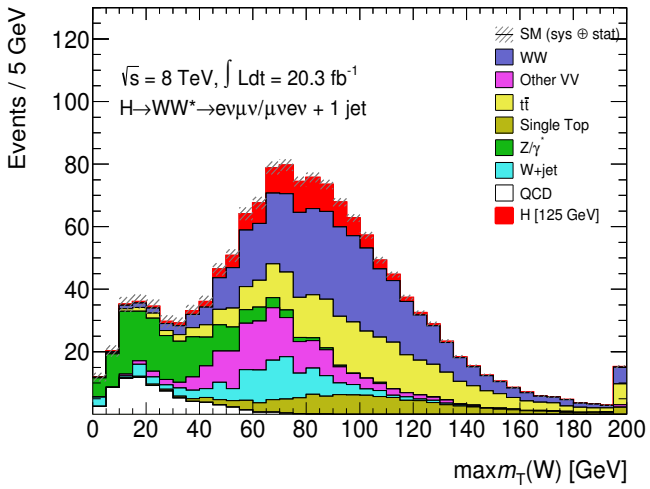

FIGURE 6.14: $\max m_T(W)$ distribution from MC simulation for $e\mu$+μe final candidates in the H+1j category. The distribution shows the high separation power in the low range of the spectrum between several background processes (Z/DY, W+jets and multi-jets) and the Higgs boson signal with $m_H = 125\,\text{GeV}$.

TABLE 6.8: Expected event yields from MC simulation for different $\not{E}_T$ scenarios with $e\mu + \mu e$ final candidates in the H+1j category. The Higgs signal is for $m_H = 125\,\text{GeV}$ (first column) and the different background contributions are summed up in the second column. The last column represents the significance from Eq. 6.3. Bottom table shows the total background separated by processes.
The thresholds are given in units of GeV. The errors are only statistical.

	Signal [125 GeV]	Total Bkg.	$S/\sqrt{B}$
$E_{\text{T,Rel}}^{\text{miss}} > 25$	68.01 ± 0.45	865.24 ± 5.67	2.31 ± 0.04
$E_{\text{T}}^{\text{miss,track,jetCorr}} > 20$ + $\max m_{\text{T}}(W) > 50$	87.18 ± 0.48	997.87 ± 6.10	2.76 ± 0.04

	WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets	Multi-jets
$E_{\text{T,Rel}}^{\text{miss}} > 25$	343.25	98.53	222.11	84.12	24.84	69.21	14.20
	± 2.50	± 3.23	± 1.74	± 0.93	± 1.85	± 2.91	± 0.52
$E_{\text{T}}^{\text{miss,track,jetCorr}} > 20$	399.52	123.01	254.12	96.59	26.19	88.11	6.06
+ $\max m_{\text{T}}(W) > 50$	± 2.70	± 3.62	± 1.86	± 0.96	± 2.06	± 2.89	± 0.39

Table 6.8 compares the expected event yields from MC for $e\mu$+μe final candidates in the H+1j category for the reference and new $\not{E}_T$ scenarios. The new selection achieves an improvement in the signal acceptance of ∼30% while still keeping all the background processes at the same order of magnitude; except for multi-jets process which is reduced by half due to the introduced $\max m_T(W)$ requirement.

6.5.1.3 H+2j Analysis

In the H+2j analysis at the end of the event selection, the event yields are very poor in statistics. In order to obtain reliable optimisation results, some of the thresholds are softened with respect to the reference selection described in Section 5.7.4. The following optimisation studies use a lower bound on the $m_{jj} > 250\,\text{GeV}$ and a softer $|\,\Delta y_{jj}\,| < 2$ requirement. The comparison of the event yields from simulation are presented in Tab. 6.9 for two different $E_{\text{T}}^{\text{miss}}$ choices: $E_{\text{T}}^{\text{miss}} > 20\,\text{GeV}$ and no $\not\!\!E_T$ requirement. The former is taken as reference from Chapter 5. From the quoted numbers, it can be observed that about 11% of the Higgs boson signal in the VBF production mode is concentrated in the $E_{\text{T}}^{\text{miss}} < 20\,\text{GeV}$ region, while for the total SM background the achieved rejection using this cut is $\sim 7\%$. The third and fourth columns of the table quote the signal significance defined in Eq. 6.5 and Eq. 6.6, respectively.

$$\text{Poisson sig} = \sqrt{2\left((S\,+\,B)\,\ln\left(1\,+\,\frac{S}{B}\right)\,-S\right)}\,, \tag{6.5}$$

$$\text{Syst sig} = \frac{S}{\sqrt{B+\sigma^2}}\,, \tag{6.6}$$

where S and B are the expected event yields for the Higgs boson signal and the total background contamination, respectively. σ refers to the estimated systematic uncertainty for each background source.[1] The significance signal values show an improvement of $\sim 10\%$ when no $\not\!\!E_T$ requirement is applied.

TABLE 6.9: Expected event yields from MC simulation for different $\not\!\!E_T$ scenarios with $e\mu$+μe final candidates in the H+2j category. The Higgs boson, with $m_{\text{H}} = 125\,\text{GeV}$, signal yield in the first column corresponds to the VBF and VH production mechanisms. The remaining background contributions and the ggF signal events are summed up in the second column. Last three columns quote the expected significance defined in Eq. 6.5, Eq. 6.6 and Eq. 6.3. The errors are only statistical.

	vbf+vh [125 GeV]	Total Bkg. (+ggf)	Poisson signif	Syst signif	S/B
No $E_{\text{T}}^{\text{miss}}$ Cut	4.84 ± 0.09	13.72 ± 1.11	1.24 ± 0.10	0.83 ± 0.02	0.35 ± 0.03
$E_{\text{T}}^{\text{miss}} > 20\,\text{GeV}$	4.37 ± 0.09	12.82 ± 1.08	1.16 ± 0.10	0.78 ± 0.02	0.34 ± 0.03

[1] The systematic uncertainties on the background yields for this purpose are extracted from the $H\to WW^{(*)}\to\ell\nu\ell\nu$ analysis described in Chapter 5.

In Appendix E, Fig. E.6 shows the expected $\not\!\!E_T$ distributions for VBF Higgs boson signal and SM backgrounds considering $e\mu$+μe final candidates in the H+2j analysis. Besides the poor statistics at the end of the event selection, the shape comparison still provides information about the $\not\!\!E_T$ role in the analysis. All $\not\!\!E_T$ reconstruction methods perform similarly, except the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ which clearly mismeasures the energy of the jets, as extensively reported in Chapter 4. The common feature is that the VBF signal contribution is relatively higher than the total background in the low region of the distributions. In light of this, any lower bound on $\not\!\!E_T$ will reject more signal than background candidates. This is consistent with the improvement on the expected signal significance observed in Tab. 6.9 when the $\not\!\!E_T$ requirement is dropped.

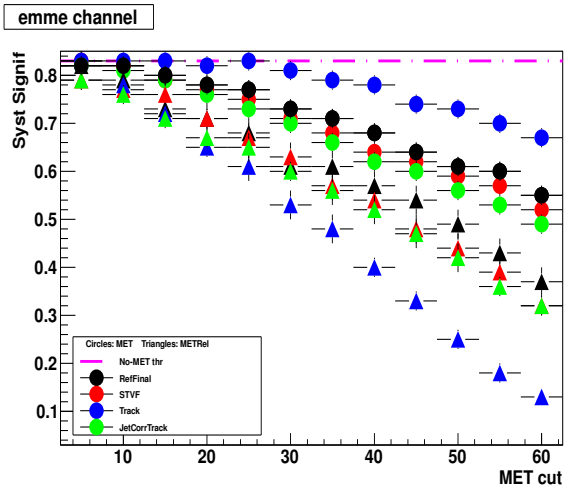

FIGURE 6.15: Figure of merit using systematic errors (see Eq. 6.6) for different $\not\!\!E_T$ reconstruction methods as a function of lower bound values for the $e\mu + \mu e$ in H+2j final state. The circles represent the non-projected $\not\!\!E_T$ quantities and the triangles show the relative $\not\!\!E_T$. The pink line is the corresponding significance for the no $\not\!\!E_T$ threshold scenario.

Figure 6.15 presents the systematic significance values obtained with different $\not\!\!E_T$ and $\not\!\!E_{T,Rel}$ thresholds in steps of 5 GeV. The VBF signal significance decreases as the lower bound on $\not\!\!E_T$ increases. For all scanned thresholds, the cut on the nominal $\not\!\!E_T$ reconstruction provides higher significance than using the projected $\not\!\!E_{T,Rel}$ quantity, as expected. The scenario without $\not\!\!E_T$ threshold provides the highest expected Higgs boson signal significance in the VBF search. As it happens for $e\mu$+μe final candidates when the ggF production mode is considered, all SM backgrounds at the end of the H+2j event selection have genuine $\not\!\!E_T$ measurement. Thus, a lower

bound on this quantity achieves very poor background rejection. In addition, the VBF Higgs boson signal contribution in the low $\not{E}_T$ region is relatively higher than for background processes because the jets that characterise the VBF production mechanism balance the $\not{E}_T$ of the event.

In order to validate the $\not{E}_T$ optimisation results, Tab. 6.10 collects the expected VBF Higgs boson signal significance results obtained through the statistical likelihood fit. The results from the quoted numbers are consistent with the conclusions extracted above. In light of this, the optimised selection does not apply any $\not{E}_T$ lower bound requirement for $e\mu$+μe final candidates in the H+$2j$ analysis, as this provides the highest expected Higgs signal significance with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ VBF search.

TABLE 6.10: Expected Higgs boson produced via VBF signal significance using different $\not{E}_T$ scenarios for $e\mu$+μe final candidates in the H+$2j$ category.

Threshold (GeV)	$E_{\mathrm{T}}^{\mathrm{miss}}$	$E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$
0	1.368	
5	1.368	1.364
10	1.350	1.353
15	1.351	1.336
20	1.324	1.302

6.5.2 New $\not{E}_T$ Selection: Changes in Control Regions

The optimisation results presented above point to lower $\not{E}_T$ thresholds in the $e\mu$+μe final candidates. Besides this increases the Higgs boson signal acceptance and maintains the background contribution similar to the reference level from Chapter 5, other aspects of the analysis have to be considered. Softening the $\not{E}_T$ requirement may affect the background composition of the different control regions. These regions serve to validate the optimised selection and provide information about the quality of the simulated processes through the statistical likelihood fit. In light of this, further investigations are developed in order to quantify the impact of the optimised $\not{E}_T$ thresholds in these control samples. This section presents the effects on background composition and the adopted changes for recovering the purity of the main control regions defined in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis.

6.5.2.1 $WW + 0j$ Control Region

Figure 6.16 shows the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ distribution for the WW CR in $e\mu$+μe final states without jets using the optimal $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} > 20\,\mathrm{GeV}$ threshold. This well explains the effect of the new $\not\!\!E_T$ selection. More Z+jets events survived due to the lower $\not\!\!E_T$ requirement. The composition of the SM process in the WW CR is compared for the reference and the optimal $\not\!\!E_T$ selection in Tab. 6.11.

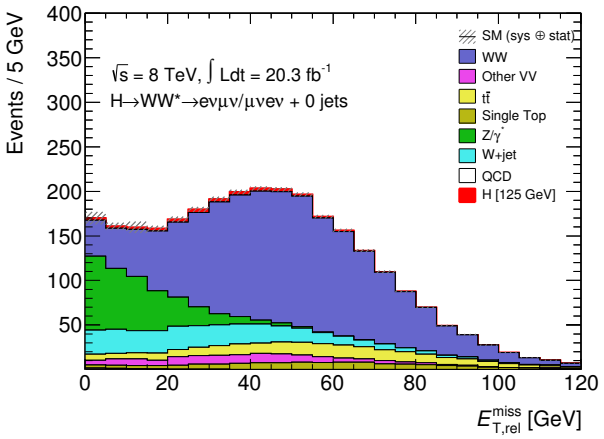

FIGURE 6.16: $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ distribution from MC simulation in the WW CR for $e\mu$+μe final states without jets. This shows the increase of Z+jets background in the low $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ region when the optimal $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} > 20\,\mathrm{GeV}$ threshold is applied.

TABLE 6.11: Expected event yields from MC simulation for WW CR using different $\not\!\!E_T$ scenarios with $e\mu + \mu e$ final states without jets. WW events are quoted in the first column while other SM processes are considered as background in the second column. The third column provides the WW purity in each phase space. Bottom table shows the total background separated by processes. The Higgs signal is considered for a mass of $m_H = 125\,\mathrm{GeV}$. The thresholds are given in units of GeV. The errors are only statistical.

	WW	Total Bkg.	Purity
$E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 25$	1454 ± 5	2064 ± 8	70%
$E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} > 20$	1743 ± 6	2869 ± 14	61%

	Signal [125 GeV]	$Z \to \tau\tau$ +γ/jets	$Z \to \ell\ell$ +γ/jets	W+jets + Multi-jets	Other Bkgs.
$E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 25$	22.6 ± 0.4	49.7 ± 2.9	3.7 ± 0.8	179.6 ± 3.1	377.5 ± 3.4
$E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} > 20$	35.4 ± 0.5	319 ± 7	24 ± 8	314 ± 4	467.1 ± 3.9

The quoted event yields in the table show that the Z+jets process increases by a factor ~ 6 using the new $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} > 20\,\mathrm{GeV}$ requirement. This is due to the lower $Z + jets$ rejection of the nominal $\not{E}_T$ reconstruction with respect to the projected $\not{E}_{T,Rel}$. This results in Z+jets events being the most significant contamination in the WW enriched sample. The enhanced Drell-Yan contribution in the WW CR worsens by 10% the purity of this region, as quoted in the fourth column. In order to keep the purity of WW CR at the same level than using the selection described in Chapter 5, a new upper bound in the angular distribution of the dilepton system is introduced. Figure 6.17 presents the $\Delta\phi_{\ell\ell}$ distributions from simulation in the WW for both $\not{E}_T$ scenarios.

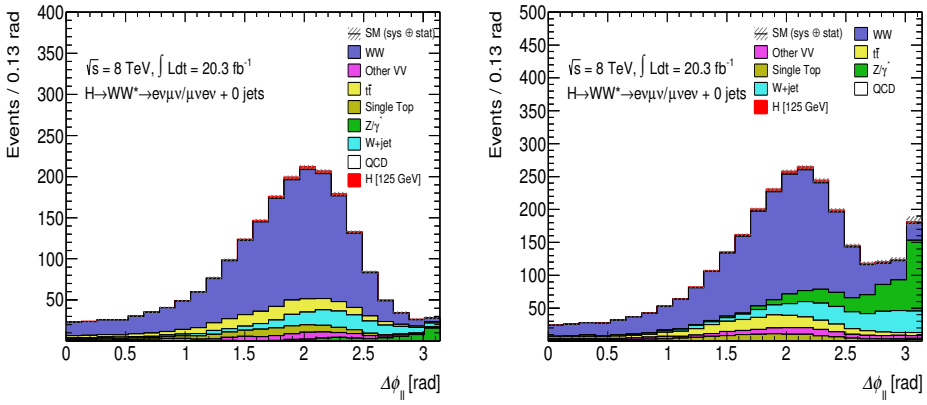

FIGURE 6.17: $\Delta\phi_{\ell\ell}$ distributions in the WW CR for $e\mu + \mu e$ final states without jets. On the left, the WW candidate events using the reference $E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 25\,\mathrm{GeV}$ requirement from Chapter 5 are shown. Right distribution shows the Z+jets contribution satisfying the new $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} > 20\,\mathrm{GeV}$ requirement, located at the high $\Delta\phi_{\ell\ell}$ region.

It is clear from the right distribution that an upper bound in this variable will reduce significantly the Z+jets contamination. The $\Delta\phi_{\ell\ell} < 2.6$ requirement is selected since it provides similar background composition than the obtained with the previous $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ threshold. The comparison of the event yields in the WW CR introducing the $\Delta\phi_{\ell\ell} < 2.6$ requirement is presented in Tab. 6.12. The purity of the WW CR increases up to 69% when the new $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} > 20\,\mathrm{GeV}$ threshold is applied in combination to $\Delta\phi_{\ell\ell} < 2.6$.

With the same purpose, the threshold on $\Delta\phi_{\ell\ell}$ is also introduced for the 0 jet top estimate to further reduce Z+jets events satisfying the new requirement on $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$.

TABLE 6.12: Expected event yields from MC simulation for WW CR using different $\not{E}_T$ scenarios with $e\mu + \mu e$ final states without jets. WW events are quoted in the first column while other SM processes are considered as background in the second column. The third column provides the WW purity in each phase space. Bottom table shows the total background separated by processes. The Higgs signal is considered for a mass of $m_H = 125\,\text{GeV}$. The thresholds are given in units of GeV. The errors are only statistical.

	WW	Total Bkg.	Purity
$E_{\text{T,Rel}}^{\text{miss}} > 25$	1454 ± 5	2064 ± 8	70
$E_{\text{T}}^{\text{miss,track,jetCorr}} > 20$ $+ \Delta\phi_{\ell\ell} < 2.6$	1603 ± 5	2316 ± 9	69

	Signal [125 GeV]	$Z \to \tau\tau$ $+\gamma$/jets	$Z \to \ell\ell$ $+\gamma$/jets	W+jets + Multi-jets	Other Bkgs.
$E_{\text{T,Rel}}^{\text{miss}} > 25$	22.6 ± 0.4	49.7 ± 2.9	3.7 ± 0.8	179.6 ± 3.1	377.5 ± 3.4
$E_{\text{T}}^{\text{miss,track,jetCorr}} > 20$ $+ \Delta\phi_{\ell\ell} < 2.6$	28.4 ± 0.4	106 ± 4	8.7 ± 1.3	184.4 ± 3.1	414 ± 4

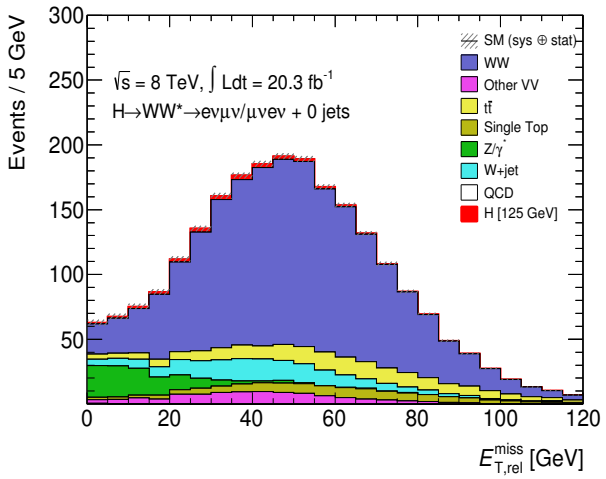

FIGURE 6.18: $\not{E}_{T,Rel}$ distribution in the WW control region for $e\mu + \mu e$ final states in the $H{+}0j$ category after adding the new selection: $\Delta\phi_{\ell\ell} < 2.6$. There is a significative reduction of the Drell-Yan background comparing with Fig 6.17.

Finally, Fig. 6.18 presents the $E_{\text{T,Rel}}^{\text{miss}}$ distribution in the new WW CR after applying $\Delta\phi_{\ell\ell} < 2.6$ requirement. This demonstrates the equivalence of the new WW CR since the remanent Z+jets contribution is reduced to the level of WW events in the $E_{\text{T,Rel}}^{\text{miss}} < 25\,\text{GeV}$ region.

6.5.2.2 WW+1j and Z/DY $\to \tau\tau$+1j Control Regions

Similar studies are also done in the WW+1j CR in order to evaluate the impact on the new $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} > 20\,\mathrm{GeV}$ threshold. The following investigations rely on the results reported in Section 6.5.1.2. The $\max m_T(W)$ is an interesting variable to study since the Z+jets background that survives to the requirement on $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$, is mostly located at the low $\max m_T(W)$ range, as shown in Fig. 6.14. In light of this, a high Z+jets background rejection is achieved by introducing an upper bound on this variable, as quoted in Tab. 6.8.

TABLE 6.13: Expected event yields from MC simulation for WW+1j CR using different $\not{E}_T$ scenarios with $e\mu + \mu e$ final states. WW events are quoted in the first column while other SM processes are considered as background in the second column. The third column provides the WW purity in each phase space. Bottom table shows the total background separated by processes. The Higgs boson signal, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, is considered as background. The thresholds are given in units of GeV. The errors are only statistical.

	WW	Total Bkg.	Purity
$E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 25$	740 ± 4	1646 ± 6	45%
$E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} > 20$	1445 ± 5	3539 ± 16	41%
$+\max m_{\mathrm{T}}(W) > 50$	1098 ± 5	2531 ± 12	43%

	Signal [125 GeV]	$Z \to \tau\tau$ +γ/jets	$Z \to \ell\ell$ +γ/jets	W+jets + Multi-jets	Other Bkgs.
$E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 25$	2.1 ± 0.3	17.1 ± 1.5	2.2 ± 0.6	78 ± 2	808 ± 4
$E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} > 20$	9.7 ± 0.5	196 ± 5	42 ± 12	347 ± 5	1507 ± 6
$+\max m_{\mathrm{T}}(W) > 50$	4.3 ± 0.4	77.3 ± 3.3	17 ± 9	165 ± 3	1174 ± 5

For the WW+1j CR, Tab. 6.13 presents the change in the event yields for two $\not{E}_T$ scenarios. Comparing with respect to the previous $E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 25\,\mathrm{GeV}$ requirement, the new event selection significantly increases the Z+jets and W+jets contamination in the WW + 1j CR, as expected. The consequence is that the WW+1j region decreases its purity from 45% to 41%. However, the introduction of the $\max m_{\mathrm{T}}(W) > 50\,\mathrm{GeV}$ threshold efficiently rejects $Z/\mathrm{DY} \to \tau\tau$ events. This recovers the purity to 43% in the WW+1j CR, as shown by the expected event yields in the last row of Tab. 6.13.

Figure 6.19 presents the $\max m_T(W)$ distribution from MC simulation for $e\mu$+μe final states in the WW+1j CR. Similarly to the the H+1j phase space, the low region of this spectrum is mostly populated by W+jets, multi-jets, and Z/γ^* backgrounds. Thus, the upper threshold $\max m_T(W) > 50\,\mathrm{GeV}$ is added to the event selection in the WW+1j CR in order to reject such contributions.

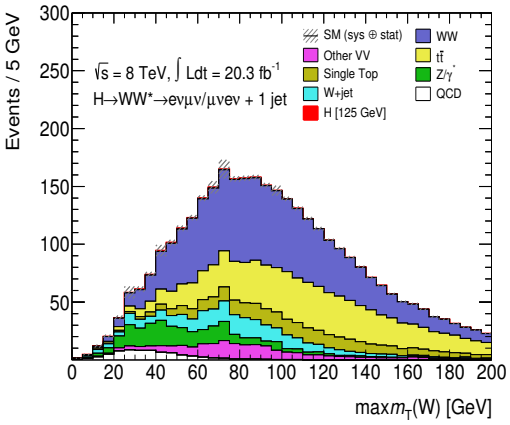

FIGURE 6.19: Expected $\max m_{\mathrm{T}}(W)$ distribution for $e\mu{+}\mu e$ final states in the $WW{+}1j$ CR from MC simulation.

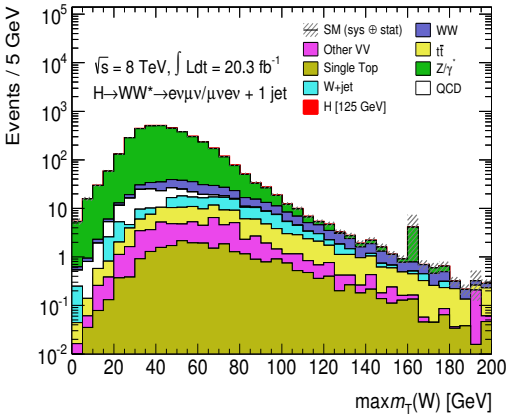

FIGURE 6.20: $max\, m_{\mathrm{T}}(W)$ distribution for $e\mu{+}\mu e$ final states in the $Z/\mathrm{DY} \to \tau\tau{+}1j$ CR from MC simulation. The $max\, m_{\mathrm{T}}(W)$ distribution is done after the selection on the $m_{\tau\tau}$ mass window to define the $Z/\mathrm{DY} \to \tau\tau{+}1j$ CR. The background composition motivates the introduction of the new upper bound requirement on $max\, m_{\mathrm{T}}(W)$.

The $Z/\mathrm{DY} \to \tau\tau{+}1j$ CR also benefits from the introduction of the $\max m_{\mathrm{T}}(W)$ quantity. The inversion of the $\max m_{\mathrm{T}}(W)$ threshold required in $H{+}1j$ and $WW{+}1j$ samples, results in an efficient selection of $Z/\mathrm{DY} \to \tau\tau$ candidates in this case. Figure 6.20 shows the $\max m_{\mathrm{T}}(W)$ distribution for $e\mu{+}\mu e$ final states being consistent with $Z/\mathrm{DY} \to \tau\tau{+}1j$ CR. The high discrimination power between the Z/γ^* and the

other backgrounds motivates the introduction of $\max m_{\mathrm{T}}(W) < 50\,\mathrm{GeV}$ requirement for improving the selection of $Z/\mathrm{DY} \to \tau\tau{+}1j$ candidate events.

Finally, the good resolution on the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ is extended to other variables relying on the $\not{E}_T$ measurement. This is the case of the $m_{\tau\tau}$ variable described in Section 5.7.3. The $m_{\tau\tau}$ spectrum is used to select events compatible with the $Z/\mathrm{DY} \to \tau\tau$ process by defining a window around the Z boson mass. The $m_{\tau\tau}$ computation relies on the $\not{E}_T$ measurement and different reconstructions have been evaluated in terms of resolution of the $Z/\mathrm{DY} \to \tau\tau$ peak. Figure 6.21 presents the $m_{\tau\tau}$ distribution based on the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ measurement in the top+$1j$ CR. The narrower Z/γ^* peak using the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ in the $m_{\tau\tau}$ computation, instead of other $\not{E}_T$ reconstruction methods, allows to veto higher number of Z/γ^* candidate events. In the $Z/\mathrm{DY} \to \tau\tau{+}1j$ CR, the benefit of using the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ reconstruction in the $m_{\tau\tau}$ computation achieves an improvement of $\sim 10\%$ on the purity of this sample.

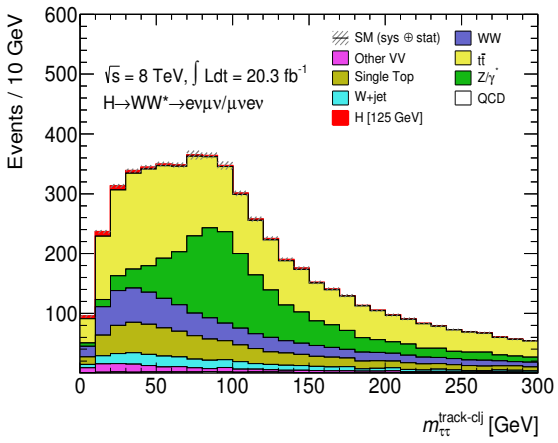

FIGURE 6.21: $m_{\tau\tau}$ distribution using the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ measurement for $e\mu{+}\mu e$ final states in the top+$1j$ CR from MC simulation.

6.5.2.3 Top+$2j$ and $Z/\mathrm{DY} \to \tau\tau{+}2j$ CRs

The $H{+}2j$ signal phase space and the $Z/\mathrm{DY} \to \tau\tau{+}2j$ CR make use of the $m_{\tau\tau}$ distribution to veto and select $Z/\mathrm{DY} \to \tau\tau$ candidates, respectively. The higher performance of the $m_{\tau\tau}$ quantity based on the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ reconstruction, motivates to replace this variable in the correspondent $H{+}2j$ regions as well.

In general, dropping the $\not{E}_T$ requirement in the H+2j analysis has low impact on the CRs event yields compared with the previous $E_{\mathrm{T}}^{\mathrm{miss}} > 20\,\mathrm{GeV}$ selection. Table 6.14 quotes the event yields from simulation in the top+2j and Z/DY $\rightarrow$ $\tau\tau$+2j CRs for the two $\not{E}_T$ scenarios discussed in Section 6.5.1.3. The last column shows the data/MC agreement in both cases. The high compatibility observed in the H+2j related regions with and without $\not{E}_T$ requirement, suggests to keep the selection referred to these enriched regions as in Chapter 5.

TABLE 6.14: Event yields from simulation and 2012 data (Obs.) for top+2j (top) and Z/DY $\rightarrow$ $\tau\tau$+2j (bottom) CRs using different $\not{E}_T$ scenarios. The ggF Higgs signal is considered for a mass of $m_H = 125\,\mathrm{GeV}$ and it is included in the total background column. The VBF Higgs signal contribution is negligible and thus excluded. Last column quotes the ratio between observed and expected events. The $\not{E}_T$ threshold is given in units of GeV. The errors are only statistical.

	WW	VV	$t\bar{t}$	Single Top	Z/γ^* +jets	W+jets + Multi-jets	Total Bkg. (+ggf)	Obs.	$\frac{\text{Data}}{\text{MC}}$
	Top+2j Control Region								
$E_{\mathrm{T}}^{\mathrm{miss}} > 20$	141 ± 3	41 ± 4	20020 ± 50	1368 ± 5	286 ± 8	277 ± 6	22140 ± 50	21536	0.97 ± 0.01
No $\not{E}_T$ cut	151 ± 3	46 ± 4	21300 ± 50	1468 ± 5	365 ± 9	328 ± 6	23670 ± 50	23028	0.97 ± 0.01
	Z/DY $\rightarrow$ $\tau\tau$+2j Control Region								
$E_{\mathrm{T}}^{\mathrm{miss}} > 20$	17 ± 1	5.3 ± 0.6	82 ± 4	8.3 ± 0.5	230 ± 6	26.7 ± 1.7	365 ± 8	330	0.90 ± 0.05
No $\not{E}_T$ cut	18 ± 1	6.3 ± 0.7	88 ± 4	8.6 ± 0.5	357 ± 8	37 ± 2	519 ± 10	474	0.91 ± 0.05

6.5.3 Same Flavour Final States: ee+$\mu\mu$

The $\not{E}_T$ requirements play an important role in the $H\rightarrow WW^{(*)}\rightarrow \ell\nu\ell\nu$ analysis when the ee+$\mu\mu$ final states are considered. The Z/γ^* process constitutes the main background source in this category, thus a lower bound on the $\not{E}_T$ spectrum is crucial to suppress this background. However, the large pile-up environment, specially suffered during 2012 data-taking, degrades the resolution of the $\not{E}_T$ measurement and causes Z/γ^* events to be reconstructed with significant $\not{E}_T$. This behaviour enhances Z/γ^* contamination in the $H\rightarrow WW^{(*)}\rightarrow \ell\nu\ell\nu$ search evolving ee+$\mu\mu$ final states. The performance of the different $\not{E}_T$ reconstruction methods in $Z \rightarrow \ell\ell$ events is addressed in Section 6.4. Data/MC comparisons for all four different $\not{E}_T$ reconstruction methods are collected in Appendix D .

The first investigations on $\not\!\!E_T$ optimisation for ee+$\mu\mu$ final states used as figure of merit the Higgs boson signal significance with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ as defined in Eq. 6.6. The systematic uncertainties for each of the processes at the end of the event selection are taken from Chapter 5. These uncertainties values are summarised bellow for the main background contributions,

- W+jets prediction is taken from data, as described in section 5.9.2. The systematic uncertainty of 30% is considered for both categories, H+0j and H+1j.
- WW, top and $Z \to \tau\tau$ are taken from MC and normalized to data in their respective $e\mu$+μe CRs. Systematic uncertainties of 7.4% (37%) for WW, 13% (30%) for top, and 14% (40%) for $Z \to \tau\tau$ in the H+0j (H+1j) category.
- $Z \to ee/\mu\mu$ estimated from data using the Pacman method with 60%(80%) systematic uncertainty, for H+0j (H+1j) category.
- Other diboson processes taken directly from simulation, with 16% (22%) systematic uncertainty, for H+0j (H+1j) category.

6.5.3.1 H+0j and H+1j Analyses

Aiming to reduce the huge Z/DY contamination, the projected $\not\!\!E_{T,Rel}$ magnitude is used for ee+$\mu\mu$ final states in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ search. Figure 6.22 compares the shape of the nominal $\not\!\!E_T$ measurement and its projected $\not\!\!E_{T,Rel}$ version for ee+$\mu\mu$ events satisfying the full $H \to WW^{(*)} \to \ell\nu\ell\nu$ selection, except the $\not\!\!E_T$ requirements. Using a lower bound on the projected $\not\!\!E_{T,Rel}$ distribution achieves about a factor 2 higher Z/γ^* suppression than with the nominal $\not\!\!E_T$ measurement. Table 6.15 quotes the mean values obtained from distributions in Fig. 6.22. For both set of comparisons, the projected $\not\!\!E_{T,Rel}$ quantity tends to lower mean values. This behaviour motivates the $\not\!\!E_{T,Rel}$ usage in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ search with ee+$\mu\mu$ final states when considering the H+0j and H+1j categories.

	$E_{\mathrm{T}}^{\mathrm{miss}}$	$E_{\mathrm{T,Rel}}^{\mathrm{miss}}$	$E_{\mathrm{T}}^{\mathrm{miss,track}}$	$E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$
H+0j	27.1	23.9	30.2	26.9
H+1j	22.4	12.5	24.0	8.43

TABLE 6.15: Mean values in GeV obtained for $\not\!\!E_T$ reconstruction methods in Fig. 6.22. The quoted values correspond to ee+$\mu\mu$ final candidates in the H+0j and H+1j categories in the top and bottom row, respectively.

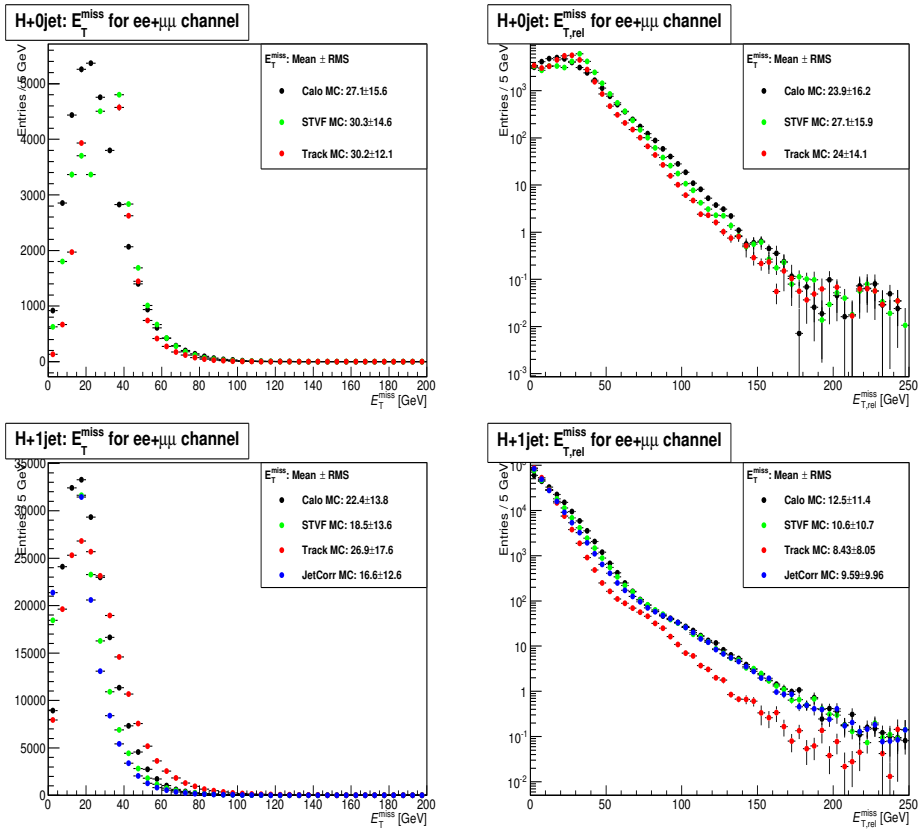

FIGURE 6.22: $\not\!\!E_T$ (left) and $\not\!\!E_{T,Rel}$ (right) shape comparisons from simulation for $ee+\mu\mu$ final candidates in the $H+0j$ and $H+1j$ categories on top and bottom, respectively.

As shown by right distributions in Fig. 6.22, the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ spectrum presents the highest contribution up to 45 GeV for both $H+0j$ and $H+1j$ analyses. In earlier studies, it was found that a lower bound on $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ provides the highest expected signal significance [111] compared with other $\not\!\!E_{T,Rel}$ magnitudes. In light of this, the optimisation procedure evaluates the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ in combination with requirements on the other $E_{\mathrm{T,Rel}}^{\mathrm{miss,STVF}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ measurements to further reject Z/γ^* background. Figure 6.23 shows the two-dimensional (2-D) scan of the lower $\not\!\!E_{T,Rel}$ bounds using two different combinations: $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ on the left, and $E_{\mathrm{T,Rel}}^{\mathrm{miss,STVF}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ on the right. The distributions on the top row are obtained computing the signal significance as described above using the event yields from simulation for $H+0j$ final candidates. The distributions on the bottom row are obtained for the $H+1j$ analysis. Each point in the 2-D grid shows the significance obtained by

using cut values at that point. The optimal threshold values for each combination are circled for the H+0j analysis. In this case, there is a $\sim$ 10% improvement in the expected significance for $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ combined thresholds. These optimisation investigations suggest that applying combined requirements on $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ leads to higher signal significance results than using $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,STVF}}$ thresholds, for both H+0j and H+1j.

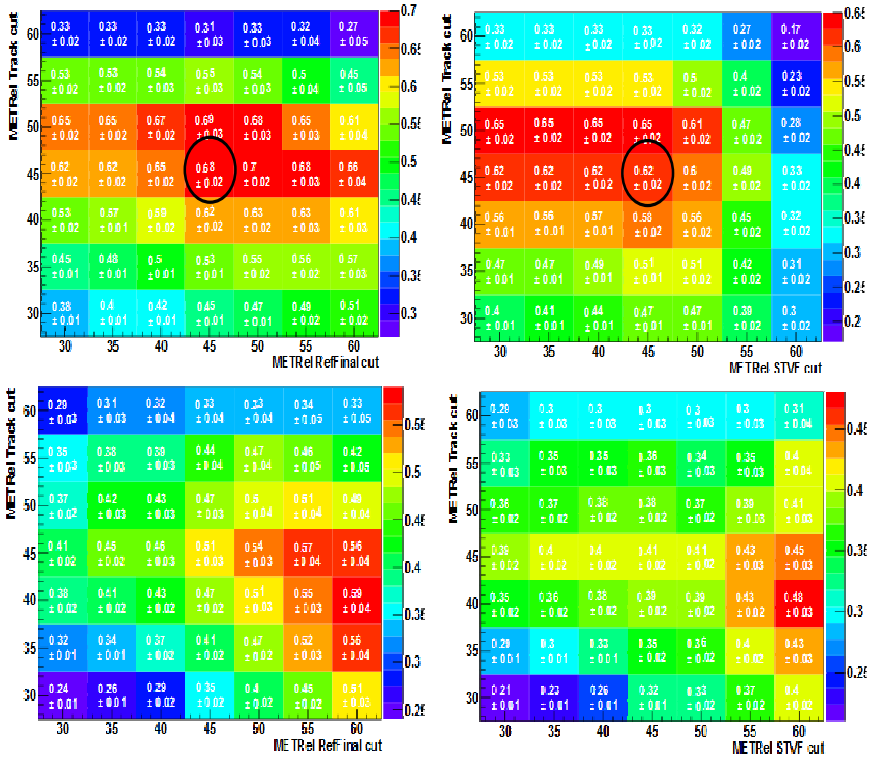

FIGURE 6.23: 2D signal significance scans for the optimization of lower bounds on missing transverse energy. $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ versus $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ is shown on the left and $E_{\mathrm{T,Rel}}^{\mathrm{miss,STVF}}$ versus $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ is on the right. The top distributions are for the 0-jet bin and the bottom ones are for 1-jet bin. The circle in the top plots highlights the optimal point.

Using the new $E_{\mathrm{T,Rel}}^{\mathrm{miss,track,jetCorr}}$ definition was also considered for the H+1j analysis as well. Note that for H+0j analysis this is identical to $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$. Once the projected quantities are computed, the comparison of both track-based definitions shows that $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ provides higher rejection and better shape separation of Z/DY background from the signal. This is illustrated in Fig. 6.24, for the remaining Z/DY contribution after the full H+1j selection is applied.

The nominal $E_{\mathrm{T}}^{\mathrm{miss,track}}$ has poor performance because of mis-measured neutral particles, as extensively reported in Section 6.4. This reconstruction presents higher width and tails than the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$. However, the results using the corresponding projected distributions show that the best resolution in Z/DY events is obtained for the $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ magnitude. This behaviour is observed for both ee and $\mu\mu$ final states.

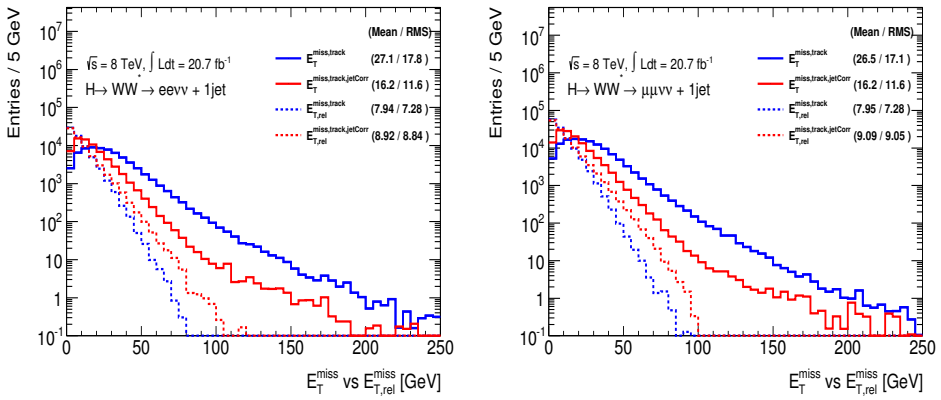

FIGURE 6.24: Comparisons of the shapes of $E_{\mathrm{T}}^{\mathrm{miss,track}}$ (blue) and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ (red) and their corresponding relative quantities (dashed lines) for Z/DY simulated samples, ee on the left and $\mu\mu$ final states on the right. Distributions are obtained after the full selection criteria for the H+1j analysis is applied.

The distributions in Fig. 6.25 show the azimuthal separation between $E_{\mathrm{T}}^{\mathrm{miss,track}}$ and the jet on the left and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ and the jet on the right, for events in the Z CR. It is clear that $E_{\mathrm{T}}^{\mathrm{miss,track}}$ tends to point in the direction of the jet, whereas $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ points mostly away from the jet, causing $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ to be projected more than $E_{\mathrm{T,Rel}}^{\mathrm{miss,track,jetCorr}}$. Hence, the combination achieving the highest Higgs signal significance in $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis with ee+$\mu\mu$ final states is $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$.

The $H\to WW^{(*)}\to \ell\nu\ell\nu$ strategy in ee+$\mu\mu$ final states, takes advantage of different and complementary methods to suppress Z/DY background to manageable level, while maintaining a reasonable Higgs boson signal acceptance. An additional measurement of the soft hadronic momentum recoiling the dilepton system is used to further reduce this background. The details on the f_{recoil} subject are addressed in Chapter 5, Section 5.6.3.2. The optimal selections to be applied on $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ and $E_{\mathrm{T,rel}}^{\mathrm{miss}}$ were found by scanning not only these variables but also f_{recoil} and $p_{\mathrm{T}}^{\ell\ell}$, since they are largely correlated and targeted at suppressing Z/DY events.

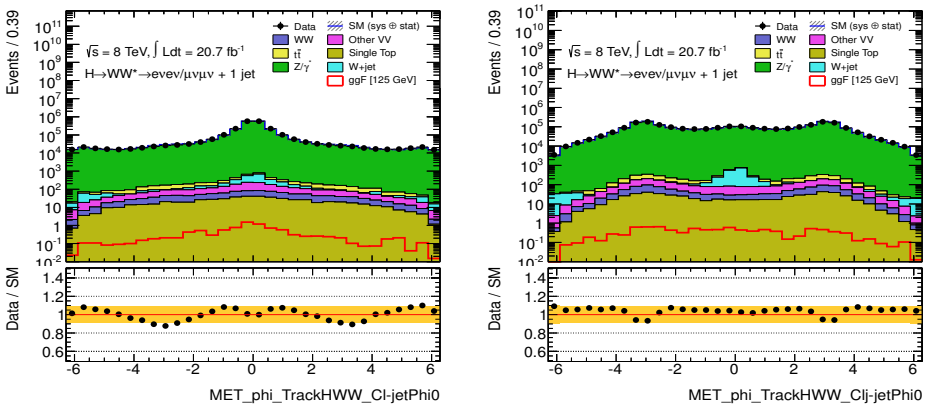

FIGURE 6.25: Azimuthal angle between $E_{\mathrm{T}}^{\mathrm{miss,track}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ with respect to the jet in $Z \rightarrow \ell\ell$ events with exactly one jet in the final state in 8 TeV data and MC simulation.

It is found that the optimal f_{recoil} and $p_{\mathrm{T}}^{\ell\ell}$ selections led to highest significances when used in combination with the following thresholds,

- H+0j: $E_{\mathrm{T,rel}}^{\mathrm{miss}} > 40$ GeV and $p_{\mathrm{T,rel}}^{\mathrm{miss,trk}} > 40$ GeV,
- H+1j: $E_{\mathrm{T,rel}}^{\mathrm{miss}} > 40$ GeV and $p_{\mathrm{T,rel}}^{\mathrm{miss,trk}} > 35$ GeV.

These become the new requirements for selection of ee+$\mu\mu$ candidates in the $H \rightarrow WW^{(*)} \rightarrow \ell\nu\ell\nu$ analysis up to one jet.

6.5.3.2 H+2j Analysis

As in the $N_{\mathrm{jets}} \leq 1$ jet categories, the ee+$\mu\mu$ final candidates in the H+2j analysis are dominated by Z/DY background. Thus, a lower bound in the $\not\!\!E_T$ spectrum is crucial to suppress this process. Note that the projected $\not\!\!E_{T,Rel}$ computation is biased in the H+2j case due to the high jet activity. This means the projected $\not\!\!E_{T,Rel}$ magnitude is computed almost randomly, also in events with genuine measurement, as the Higgs boson decay into a pair of W bosons.

Figure 6.26 shows the different $\not\!\!E_T$ reconstruction methods from MC simulated ee+$\mu\mu$ final candidates in the H+2j analysis. Besides the low available statistics, $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ spectrums concentrate higher number of background events at the low region. Hence, the lower mean values for the total background hypothesis are provided by these reconstruction methods, as shown by the quoted numbers in the legend. The $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ recovers $E_{\mathrm{T}}^{\mathrm{miss,track}}$ resolution and its performance is better than the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ measurement, used so far in this category.

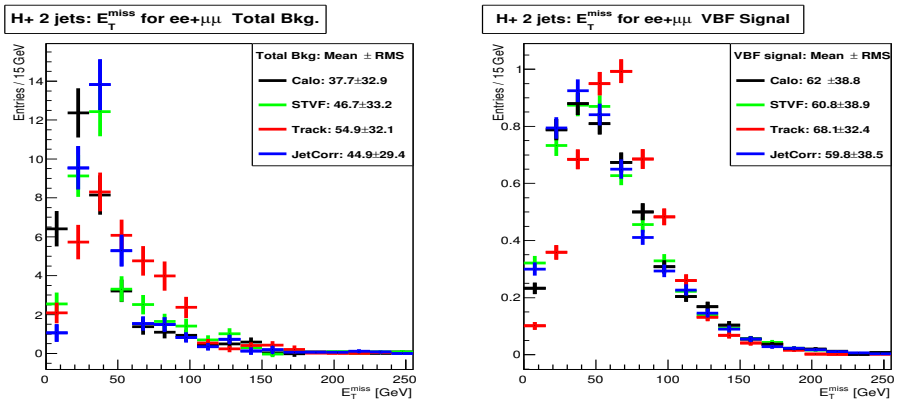

FIGURE 6.26: $\not\!\!E_T$ distributions for $ee{+}\mu\mu$ final candidates in the $H{+}2j$ category. Left distribution shows the total SM background and right plot shows the VBF production mechanism of the Higgs boson with $m_H = 125\,\text{GeV}$. In the legends the mean and RMS values in GeV are quoted for each of the $\not\!\!E_T$ reconstruction methods.

In light of these shape comparison results, it is expected to further reject Z/DY contamination applying a combined lower bound on $E_{\text{T}}^{\text{miss}}$ and $E_{\text{T}}^{\text{miss,track,jetCorr}}$ quantities.

Figure 6.27 shows the expected significance computed from Eq. 6.5 and 6.6 on the left and right, respectively; for each $\not\!\!E_T$ (circles) and its corresponding projected $\not\!\!E_{T,Rel}$ (triangles) magnitudes. In general, the $\not\!\!E_T$ requirements above 35 GeV provide higher significance values than the same threshold on the projected $\not\!\!E_{T,Rel}$ magnitudes. This is expected since the high jet activity in these events can overcorrect the projected $\not\!\!E_{T,Rel}$ quantities, as mentioned above. The highest significances are obtained for $E_{\text{T}}^{\text{miss}}$, since this concentrates more background events at the low region of the spectrum. Moreover, the second highest significance values are achieved after requiring a lower bound on $E_{\text{T}}^{\text{miss,track,jetCorr}}$. It improves the expected VBF Higgs boson signal significance by $\sim 9\%$ with respect the previous $E_{\text{T}}^{\text{miss, STVF}}$ requirement used up to now.

Finally, the optimal $\not\!\!E_T$ selection values in the $ee{+}\mu\mu$ $H{+}2j$ category, are found performing a similar 2-D scan procedure as for the $H{+}0j$ and $H{+}1j$ analyses. The optimal $\not\!\!E_T$ requirement values for $ee{+}\mu\mu$ final candidates in the $H{+}2j$ category are $E_{\text{T}}^{\text{miss}} > 55$ GeV and $E_{\text{T}}^{\text{miss,track,jetCorr}} > 50$ GeV.

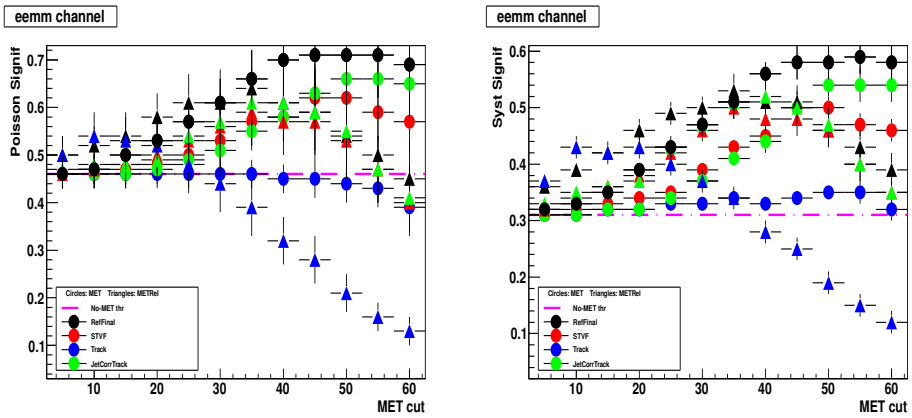

FIGURE 6.27: Higgs boson signal expected significance as a function of different $\not{E}_T$ threshold values and measurements for $ee{+}\mu\mu$ final candidates in the $H{+}2j$ analysis. Left distribution corresponds to the Poisson significance and right is the figure of merit which includes the systematic uncertainties for each background.

6.6 m_{T} Performance and Optimisation

As discussed in Chapter 5, the $\not{E}_T$ measurement is not only used to suppress background contributions, like Z/DY in $ee{+}\mu\mu$ final states, but also for computing other magnitudes as $m_{\tau\tau}$, $\max m_{\mathrm{T}}(W)$... One of the most relevant variable is the m_{T}, defined in Eq. 5.7. In this light, there are also four different m_{T} measurements depending on the $\not{E}_T$ reconstruction method entering in the m_{T} computation. The differences on the $\not{E}_T$ reconstruction methods are expanded to the m_{T} measurement as well. This dependence is crucial for optimisation purposes since the separation of the Higgs boson signal from the SM background is done through the m_{T} shape in the statistical likelihood fit. In the $H{\to}WW^{(*)}{\to}\ell\nu\ell\nu$ analysis, a fit in data to the m_{T} is used to extract the signal strength and to perform the optimisation results discussed in Section 6.5. So far, $E_{\mathrm{T}}^{\mathrm{miss}}$ has been used in the calculation of m_{T}. However, it has been shown that other measurements of $\not{E}_T$ have better resolution and show more stability against pile-up and may, therefore, provide better separation between signal and background. This section contains investigations on the different m_{T} measurement in terms of performance and expected signal significance.

Figure 6.28 shows that the separation power between the signal and diboson $W{+}$jets backgrounds is higher for the m_{T} based on $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ than for the other m_{T} reconstructions. The difference in shape and composition of the total background for each m_{T} reconstruction is clearly visible in the distributions.

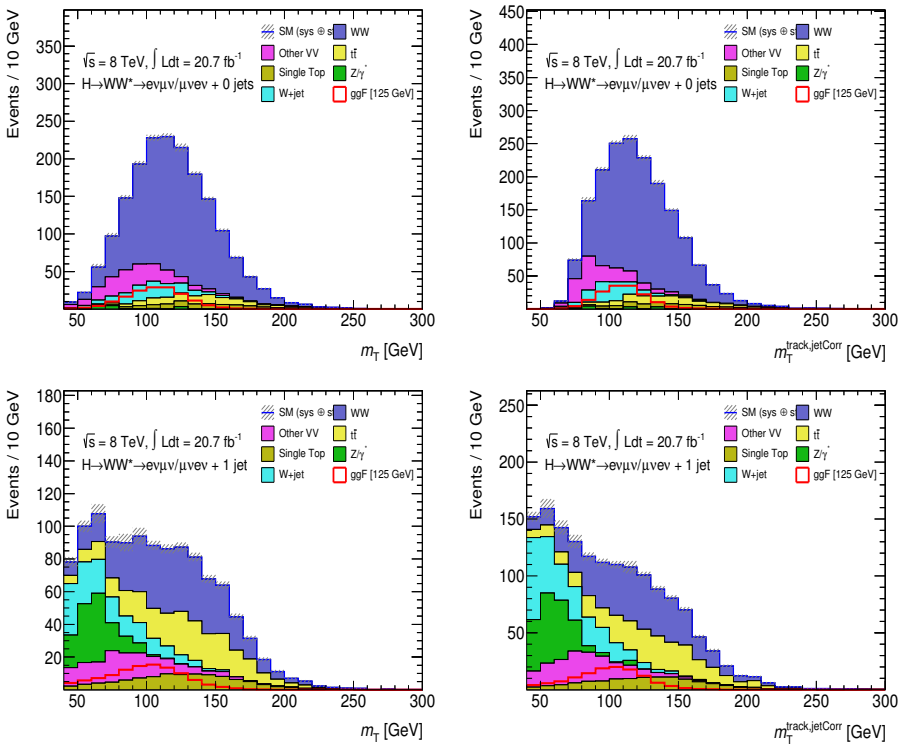

FIGURE 6.28: Expected m_{T} distributions based on $E_{\mathrm{T}}^{\mathrm{miss}}$ on the left, and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ on the right for $e\mu$+μe final candidates in the H+0j (top) and H+1j (bottom) categories.

The improvement in resolution achieved by the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ reconstruction in the m_{T} computation is also evaluated for MC simulated Higgs boson signal with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. Figure 6.29 compares the different m_{T} measurements using Higgs boson signal MC samples. The legend quotes the mean and RMS values for each m_{T} reconstruction.

A summary of the expected significance obtained when fitting different m_{T} definitions for the $e\mu$+μe final state in H+0j and H+1j analysis are presented in Tab. 6.16. For easier comparison, the relative differences are calculated with respect to the fit obtained with the $E_{\mathrm{T}}^{\mathrm{miss}}$-based m_{T}, as used so far. The best result is obtained by fitting $m_{\mathrm{T}}^{\mathrm{track,jetCorr}}$, which achieves an improvement of $\sim 13\%$. The gain when using $m_{\mathrm{T}}^{\mathrm{track,jetCorr}}$ comes from the higher separation power between the Higgs boson signal and backgrounds like $W\gamma$ and $W+$jets, which are not negligible contaminations at the end of the event selection in these categories. The statistical fit using $m_{\mathrm{T}}^{\mathrm{track,jetCorr}}$ benefits of this behaviour to extract the signal strength.

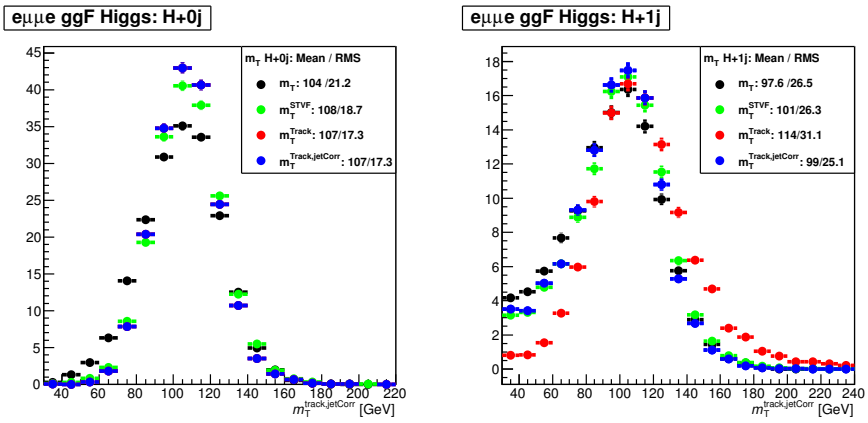

FIGURE 6.29: m_{T} distributions from MC simulation of a Higgs boson with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ produced through the ggF production mode for $e\mu{+}\mu e$ final candidates in the $H{+}0j$ and $H{+}1j$ on the left and right, respectively.

Z^{125}_{exp}	$e\mu{+}\mu e$ 0j	diff.	$e\mu{+}\mu e$ 1j	diff.
m_{T} fit	2.226	+0.0%	1.715	+0.0%
$m_{\mathrm{T}}^{\mathrm{STVF}}$ fit	2.472	+11.1%	1.807	+5.3%
$m_{\mathrm{T}}^{\mathrm{track}}$ fit	2.509	+12.7%	1.578	−8.0%
$m_{\mathrm{T}}^{\mathrm{track,jetCorr}}$ fit	2.509	+12.7%	1.954	+13.9%

TABLE 6.16: Expected Higgs boson with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ signal significance for different m_{T} reconstructions for $e\mu{+}\mu e$ final candidates in $H{+}0j$ and $H{+}1j$ categories. All relative differences ("diff" columns) are calculated with respect to the fit using m_{T}.

The good performance and the higher separation power between signal and background is also studied in $ee{+}\mu\mu$ final candidates. In Fig. 6.30, for the $H{+}0j$, and Fig. 6.31, for the $H{+}1j$, the shapes of the individual background source are compared directly for the different m_{T} reconstructions. In general, m_{T} computed with $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ shows the best resolution for the different background contaminations also in the $ee{+}\mu\mu$ final candidates. This behaviour is seen for the signal as well, as shown in Fig. 6.32, where the different shapes for a sample with a mass of 125 GeV are compared also to the true m_{T} value obtained from the simulation.

The expected uncertainty on the signal strength, as well as the expected significance, are determined for each of the different m_{T} flavour fits. The obtained results are presented in Tabs. 6.17 and 6.18. For comparison purposes, the relative differences (show as percentage) are calculated with respect to the fit in the $E_{\mathrm{T}}^{\mathrm{miss}}$-based m_{T} used until now. The best result is obtained by fitting m_{T} calculated with $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$.

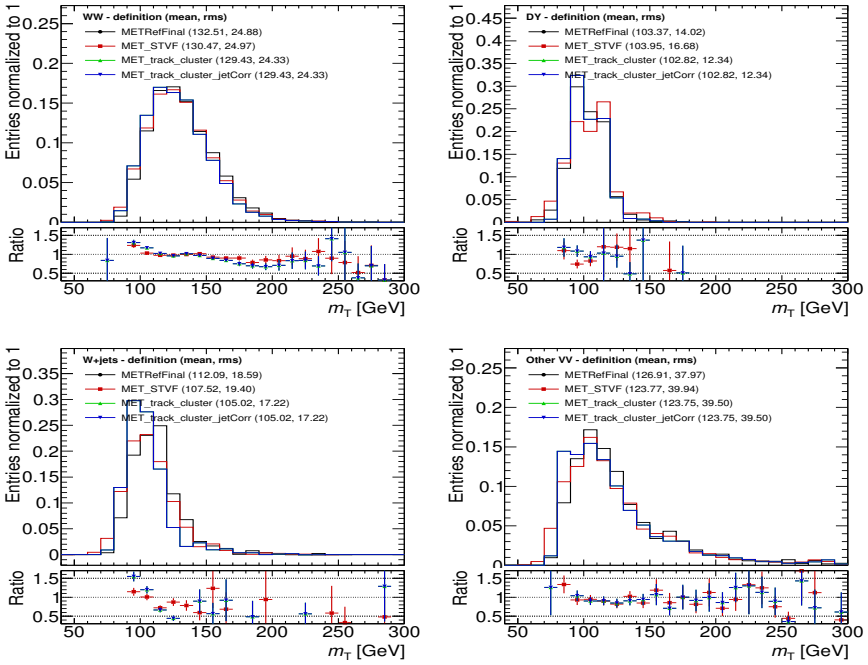

FIGURE 6.30: m_{T} shape comparisons from simulation for ee+$\mu\mu$ final candidates in H+0j category for different background sources - WW on top left, Z/DY on top right, W+jets on bottom left and other dibosons on the bottom right. In the m_{T} calculation, different $\not{E}_T$ reconstructions are used: $E_{\mathrm{T}}^{\mathrm{miss}}$ in black, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ in red and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ in green/blue.

TABLE 6.17: Expected uncertainty on the Higgs boson signal strength when fitting different m_{T} reconstructions. All relative differences ("diff" columns) are calculated with respect to the fit using m_{T}. Last row ("fit + double bins") doubles the number of m_{T} bins with respect to the previous row, i.e., 10/6 bins instead of 5/3 in the 0/1-jet case, respectively.

	H+0j		H+1j		H+0/1j	
$\hat{\mu}_{\mathrm{exp}}^{125}$	Strength	diff.	Strength	diff.	Strength	diff.
m_{T}	$1^{+1.026}_{-1.024}$	+0.0% +0.0%	$1^{+1.551}_{-1.388}$	+0.0% +0.0%	$1^{+0.828}_{-0.81}$	+0.0% +0.0%
$m_{\mathrm{T}}^{\mathrm{STVF}}$	$1^{+0.985}_{-0.969}$	+3.7% +4.9%	$1^{+1.534}_{-1.453}$	+1.0% −2.7%	$1^{+0.81}_{-0.799}$	+1.9% +1.2%
$m_{\mathrm{T}}^{\mathrm{track}}$	$1^{+0.971}_{-0.946}$	+4.9% +7.0%	$1^{+1.686}_{-1.593}$	−8.2% −11.6%	$1^{+0.817}_{-0.799}$	+1.1% +1.2%
$m_{\mathrm{T}}^{\mathrm{track,jetCorr}}$	$1^{+0.972}_{-0.947}$	+4.9% +6.9%	$1^{+1.544}_{-1.476}$	+0.4% −5%	$1^{+0.801}_{-0.787}$	+2.9% +2.5%
$m_{\mathrm{T}}^{\mathrm{track\text{-}jetCorr}}$ + double bins	$1^{+0.949}_{-0.916}$	+6.9% +9.7%	$1^{+1.416}_{-1.251}$	+7.3% +7.7%	$1^{+0.762}_{-0.735}$	+7.1% +8.2%

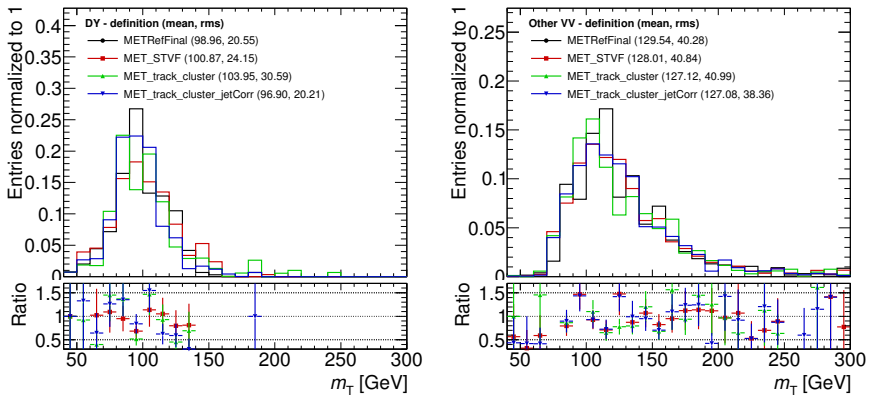

FIGURE 6.31: Transverse mass shapes for $ee{+}\mu\mu$ final candidates in the $H{+}1j$ category for different background sources - Z/DY on the left and other dibosons on the right - using different $\not{E}_T$ definitions for the m_{T} calculation: $E_{\mathrm{T}}^{\mathrm{miss}}$ in black, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ in red, $E_{\mathrm{T}}^{\mathrm{miss,track}}$ in green and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ in blue.

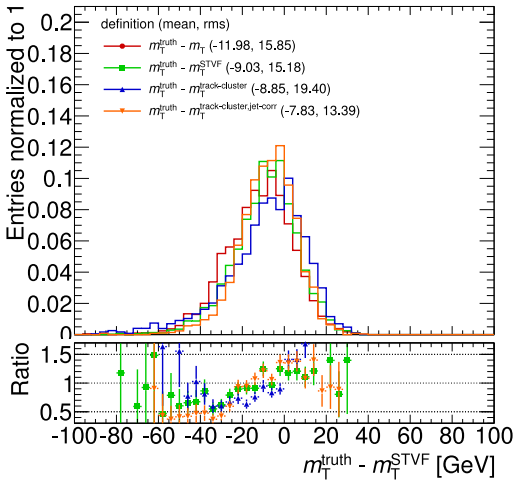

FIGURE 6.32: Difference between each of the m_{T} calculations and the generated m_{T} value from the simulation of the Higgs boson signal, with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, for $ee{+}\mu\mu$ final candidates in the $H{+}1j$ category. $m_{\mathrm{T}}^{\mathrm{track,jetCorr}}$ is represented in orange and it shows the smallest difference with respect to the true value and the best resolution.

The results varying the number of bins with $ee{+}\mu\mu$ final candidates are presented in the last rows in Tabs. 6.17 and 6.18. Doubling the number of bins in the fit of $m_{\mathrm{T}}^{\mathrm{track,jetCorr}}$ in the fit treatment achieves an extra improvement of $\sim 5\%$ in the expected significance results also in the same flavour channels.

TABLE 6.18: Expected Higgs boson with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ signal significance for different m_{T} measurements in $ee{+}\mu\mu$ final candidates for $H{+}0j$ and $H{+}1j$ categories. All relative differences ("diff" columns) are calculated with respect to the fit using m_{T}. Last row ("fit + double bins") doubles the number of m_{T} bins with respect to the previous row, i.e., 10/6 bins instead of 5/3 in the 0/1-jet case, respectively.

	$H{+}0j$		$H{+}1j$		$H{+}0/1j$	
Z^{125}_{exp}	Signif.	diff.	Signif.	diff.	Signif.	diff.
m_{T}	0.979	+0.0%	0.737	+0.0%	1.233	+0.0%
$m_{\mathrm{T}}^{\mathrm{STVF}}$	1.034	+6.1%	0.713	−4.0%	1.247	+1.3%
$m_{\mathrm{T}}^{\mathrm{track}}$	1.057	+8.6%	0.657	−13.2%	1.245	+1.1%
$m_{\mathrm{T}}^{\mathrm{track,jetCorr}}$	1.056	+8.5%	0.704	−5.5%	1.263	+2.8%
$m_{\mathrm{T}}^{\mathrm{track\text{-}jetCorr}}$ + double bins	1.09	+12.4%	0.812	+11.7%	1.355	+11.4%

TABLE 6.19: Expected significance ("Signif" columns) obtained when fitting different m_{T} measurements for $ee+\mu\mu$ and $e\mu+\mu e$ final states in the $H{+}2j$ analysis on the right and left, respectively. All relative differences ("diff" columns) are calculated with respect to the m_{T} result.

Z^{125}_{exp}	$ee{+}\mu\mu$		$e\mu{+}\mu e$	
	Signif.	diff.	Signif.	diff.
m_{T}	0.630	+0.0%	1.249	+0.0%
$m_{\mathrm{T}}^{\mathrm{track,jetCorr}}$	0.693	+10.0%	1.298	+3.9%

Same results and conclusions are obtained in the $H{+}2j$ analysis. The fit using the $m_{\mathrm{T}}^{\mathrm{track,jetCorr}}$ improves the expected significance by $\sim 4\%$ and $\sim 10\%$ for the $e\mu{+}\mu e$ and $ee{+}\mu\mu$ final states respectively. The results are quoted in Tab. 6.19.

Finally, different number of m_{T} bins are also studied, to take advantage of improved signal resolutions. Table 6.20 compares the expected significance when doubling the number of bins used in the fit procedure, i.e., 10 bins for the $H{+}0j$ analysis instead of the previous 5 bins, and 6 bins in the $H{+}1j$ analysis instead of the 3 bins used so far. The improvement in the $e\mu{+}\mu e$ channels is specially significative in the $H{+}1j$ analysis, as expected from Fig. 6.28, achieving a gain in the expected significance up to $\sim 6\%$. The final choice, based on the results presented in this section, is to use the $m_{\mathrm{T}}^{\mathrm{track,jetCorr}}$ for all $H{\to}WW^{(*)}{\to}\ell\nu\ell\nu$ categories. Moreover, the $H{+}0j$ and $H{+}1j$ analysis with $e\mu{+}\mu e$ highly benefit from doubling the number of bins in the fit procedure.

TABLE 6.20: Expected significance obtained when fitting m_{T} with different number of bins. All relative differences ("diff" columns) are calculated with respect to the fit using 5 and 3 bins for the H+0j and H+1j analysis, respectively.

N_{bins}	H+0j			H+1j		
	5 bins	10 bins	diff.	3 bins	6 bins	diff.
$e\mu$+μe	2.434	2.465	+1.3%	1.673	1.772	+5.9%
$e\mu$	2.057	2.077	+1.0%	1.405	1.468	+4.5%
μe	1.391	1.415	+1.7%	1.196	1.246	+4.2%

6.7 Conclusions

This chapter presents several $\not{E}_T$ investigations focused on improving the sensitivity of the $H\to WW^{(*)}\to \ell\nu\ell\nu$ search. The final optimised results highly benefit from the development of a new $\not{E}_T$ reconstruction. The new $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ measurement is based on the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ approach but replacing tracks by the corresponding calorimetric measurements and adding neutral particles, which are excluded in the original $E_{\mathrm{T}}^{\mathrm{miss,track}}$ reconstruction. The results show that the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ is able to recover the resolution in events with jets, while still maintains a good stability against pile-up, and smaller tails in $Z\to\ell\ell$ process. Additional investigations using event topologies with genuine $\not{E}_T$, also point to a more reliable measurement when using the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ reconstruction.

The strategy for optimising the $\not{E}_T$ criteria in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis relies on simulated final candidate events, computing the signal significance through the statistical likelihood fit. Given the background composition and contribution depends on the final state, the $\not{E}_T$ optimisation is developed in each $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis category. For $e\mu$+μe final candidates the different $\not{E}_T$ measurements perform very similarly at the end of the event selection. The low region of the spectrums is almost no populated since the main backgrounds, as well as the Higgs boson signal, are expected to have genuine $\not{E}_T$. In addition, the analysis requirements on $m_{\ell\ell}$ and $p_{\mathrm{T}}^{\ell\ell}$, which are correlated with the $\not{E}_T$ measurement, sculpt the $\not{E}_T$ shapes at the end of the selection. Hence, similar expected signal significance values are obtained by using any of the $\not{E}_T$ reconstruction methods. However, the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ is preferred because of its better performance and resolution. A conservative threshold of 20 GeV is used for H+0j and H+1j categories. The new $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ requirement increases the expected significance by 7% in the ggF-enriched analysis with respect to the previous selection.

Since the Higgs boson production via VBF is typically characterised by two emerging quarks, the $\not\!\!E_T$ measurement for the VBF Higgs signal is expected to be smaller than in the ggF production mode. The results confirm that the low region of the $\not\!\!E_T$ spectrums is mainly populated by signal events. In this light, the VBF strategy does not apply any threshold on the $\not\!\!E_T$ measurement in $e\mu$+μe final states. For the VBF-enriched search, the overall improvement due to the optimised $\not\!\!E_T$ selection is observed up to 14% in the expected significance results.

In the $H\rightarrow WW^{(*)}\rightarrow \ell\nu\ell\nu$ search, ee+$\mu\mu$ final states are affected by a huge Z/γ^* contribution, so combinations of several $\not\!\!E_T$ reconstruction methods have to be applied to further suppress this background. When the final state contain up to one jet, the requirement is done using the projected $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$ measurements. Investigations on the direction of the new $E_{\mathrm{T,Rel}}^{\mathrm{miss,track,jetCorr}}$ conclude that the rejection power of the original $E_{\mathrm{T}}^{\mathrm{miss,track}}$ is still higher. This is due to the latter reconstruction tends to point to the mis-measured jet, hence the projected $\not\!\!E_{T,Rel}$ magnitude benefits from this feature. However, for the VBF-enriched analysis with ee+$\mu\mu$ final candidates the $\not\!\!E_{T,Rel}$ magnitude may be biased because of the probability to randomly project the $\not\!\!E_T$ into the direction of any reconstructed jet. This points back to the usage of the nominal $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ measurement, complemented with a lower bound on $E_{\mathrm{T}}^{\mathrm{miss}}$.

The better $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ performance is exploited to also benefit other $\not\!\!E_T$-dependent variables, as the m_{T}. Results show a better resolution of the m_{T} measurement obtained by using the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ in the computation. This leads to higher separation between the Higgs signal and the remaining backgrounds, specially for multi-jets and non-WW diboson processes. The introduction of the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ in the m_{T} computation enhances the expected significance by 9% in the most sensitive category.

Chapter 7

Observation of Higgs Boson Decays to WW^* with ATLAS

7.1 Introduction

Chapter 5 describes the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis strategy and the results obtained using the whole data collected by the ATLAS detector during Run-I period. However, optimisation results on $\not{E}_T$ thresholds and m_{T} comparisons from Chapter 6 show that the sensitivity of the search can be enhanced. In light of this, a complete analysis optimisation procedure has been investigated using the total integrated luminosity at 8 TeV collected at ATLAS. The main developments rely on the introduction of more performant variables, as the jet-corrected track-based $\not{E}_T$; improvements on techniques for estimating backgrounds, which allow to reduce systematic uncertainties associated to them; and extensions of the signal phase space to increase the Higgs boson signal acceptance, as the introduction of a new H+2j ggF-enriched category. The whole strategy has undergone through optimisation investigations which use the likelihood fit to ensure modifications benefit the expected significance and no tensions between nuisance parameters are added. All the changes motivated by those studies are propagated to the 7 TeV data sample as well.

This chapter is structured as follows. Section 7.2 summarises the changes adopted in the different categories in which the search is divided. Section 7.3 presents the performance of the analysis using the new selection. The final candidates for the ggF-enriched region are shown in Section 7.3.1. The H+2j VBF-enriched analysis makes use of a new multivariate approach which replaces the cut-based analysis used

up to now. The basis of this procedure and its results are presented in Section 7.3.2. The last part of the chapter contains the results obtained through the likelihood fit in Section 7.4. The observation of the Higgs boson in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ channel and the evidence of the VBF production mechanism are found in Sections 7.4.1 and 7.4.2, respectively. The results on the measured signal strength for the inclusive search, as well as for the ggF and VBF production modes, are given in Section 7.4.3. Following that, Section 7.4.4 uses the previous signal strength in each production mode to test the compatibility of the fermionic and bosonic couplings of the Higgs boson with the SM prediction. Finally, the Higgs cross section measurements are given in Section 7.4.5.

7.2 New Developments on $H\to WW^{(*)}\to \ell\nu\ell\nu$ Search

The main improvements on the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis with respect to the strategy presented in Chapter 5 are summarised below.

7.2.1 Common Changes

The changes, which are common to all analysis categories for 8 TeV data, are briefly described in the following.

Dilepton triggers
The re-analysis of the 8 TeV data uses events selected with triggers that required either a single lepton or two leptons (dilepton). The single lepton triggers had more restrictive lepton identification requirements and higher p_{T} thresholds than the dilepton triggers. The benefit of requiring two leptons with opposite charge at the trigger level allows to lower the lepton p_{T} thresholds while still maintaining the trigger efficiency.

Electron identification
Improvement on electron identification which is based on a new likelihood technique [149]. This improves background rejection and allows to lower the $p_{\mathrm{T}}^{\mathrm{sub}}$ requirement.

Lepton p_{T} thresholds
The p_{T} threshold of the leading lepton is lowered from 25 to 22 GeV and the requirement on the subleading lepton is changed from 15 GeV to 10 GeV. Lowering the requirements on the p_{T} thresholds increases the signal acceptance. The gain in the signal is illustrated in the first two bins of p_{T} distributions shown in Fig. 7.1.

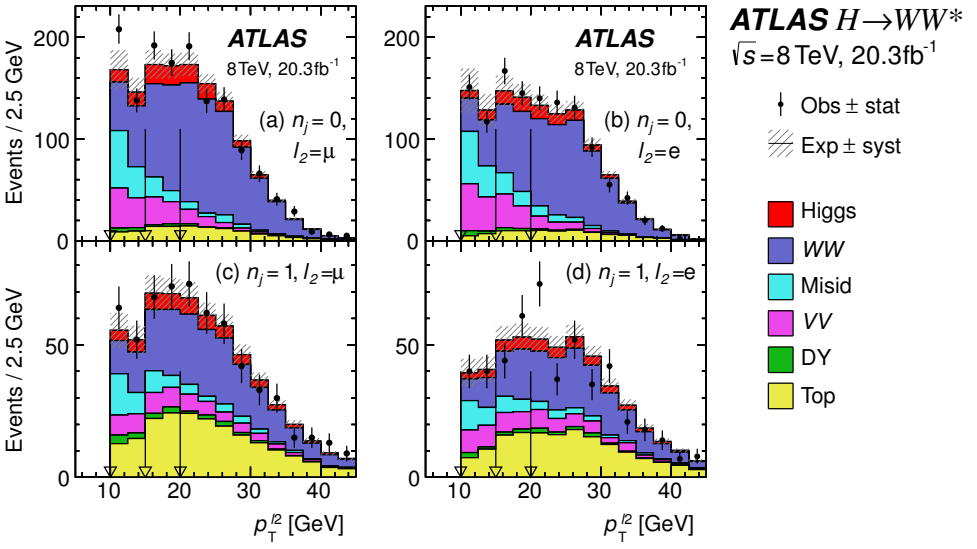

FIGURE 7.1: Subleading lepton p_{T} distributions for $e\mu$+μe final candidates in H+0j and H+1j categories in 8 TeV ATLAS data and MC simulation. The distributions are shown for two categories of events based on the flavour of the subleading lepton (ℓ_2). The plots are made after requiring all selections up to the m_{T} requirement, as shown in Tab. 7.3. The observed data points (Obs, •) with their statistical uncertainty (stat) are compared with the histograms representing the cumulative expected contributions (Exp, —), for which the systematic uncertainty (syst) is represented by the shaded band. This band accounts for experimental and theoretical uncertainties. The legend order follows the histogram stacking order of the plots. The arrows mark the bin boundaries for defining the signal regions used in the likelihood fit.

Switch to jet-corrected track-based $\not\!\!E_T$

The more performant jet-corrected track-based $\not\!\!E_T$ leads to a significant improvement when using this quantity for the $\not\!\!E_T$ thresholds. Moreover, its better resolution is used for defining $\not\!\!E_T$-dependent quantities as m_{T} and $m_{\tau\tau}$. Figure 7.2 compares the resolution of the $\not\!\!E_T$ and m_{T} quantities from simulated ggF signal events. Complete results on the jet-corrected track-based $\not\!\!E_T$ performance are detailed in Section 6.4. For sake of simplicity, the jet-corrected track-based $\not\!\!E_T$ quantity described in Chapter 6 will be referred to as $p_{\mathrm{T}}^{\mathrm{miss}}$ in the following.

Lower jet p_{T} threshold for b-veto

To further reduce contamination from top quark decays, the veto on b-tagged jets is extended to remove events with any b-tagged jet with $p_{\mathrm{T}} > 20$ GeV. Previously, only events with a b-tagged jet over the standard analysis jet counting thresholds

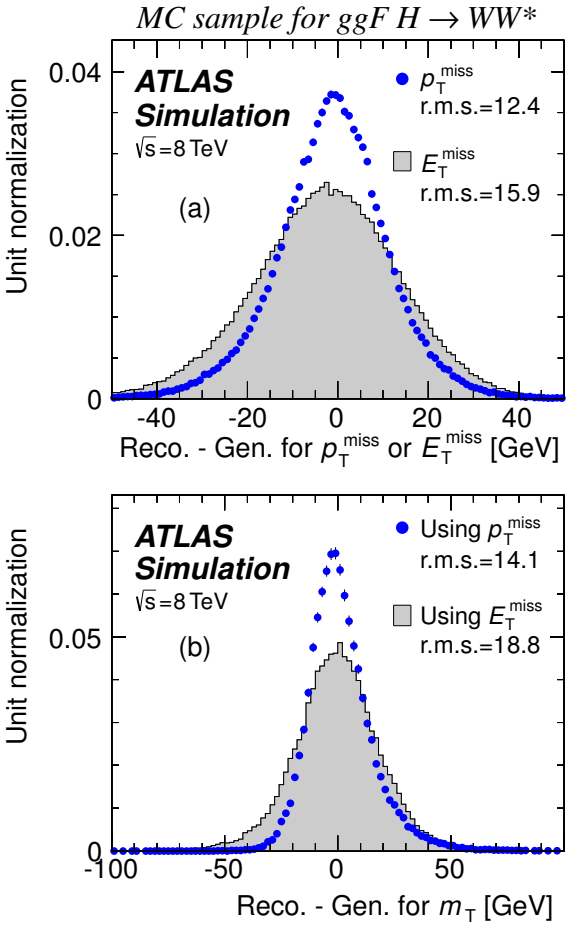

FIGURE 7.2: Resolutions of (a) $\not{E}_T$ and (b) m_{T} for the ggF signal MC in the $H+0j$ category. The comparisons are made between the calorimeter-based ($E_{\mathrm{T}}^{\mathrm{miss}}$) and the jet-corrected track-based ($p_{\mathrm{T}}^{\mathrm{miss}}$) measurements. The resolution is measured as the difference of the reconstructed (Reco.) and generated (Gen.) quantities. The RMS (r.m.s.) of the distributions are given in the legends in units of GeV.

were used. This was, $p_{\mathrm{T}} > 25(30)$ GeV for $|\eta| < (>)2.4$ (see Section 5.6.2).

Multi-jet estimation and Z/DY fake factor

Previously, the multi-jet contribution was relatively small and included in the $W+$jets data-driven method, as described in Section 5.9.2. However, lowering the p_{T} thresholds of lepton candidates enhances the contribution of multi-jet background. A

direct estimate of this contribution is calculated and included in the analysis separately. In addition, the fake factor used in the $W+$jets estimate is now derived from the Z/DY sample instead of the dijet sample used so far. The Z/DY sample composition and its p_{T} spectrum are closer to the $W+$jets control sample. Hence, the main advantages using the Z/DY fake factor are the more accurate estimation of the $W+$jets background, as well as, the reduction of the sample dependence systematic uncertainty.

Top theory
Previously, a non-uniform set of uncertainties from statistically-limited MC comparisons was used to determine the top quark background theory systematics. This procedure has been changed to take into account variations of the four independent sources: scale variations, PDF, PS and MC modelling uncertainties.

The changes for the 7 TeV data analysis mimic the new common selection changes to benefit from the improvements obtained using the 8 TeV optimisation results. Furthermore, it keeps consistency between the two data years and simplifies the final 7 + 8 TeV combination .

7.2.2 Changes in the ggF-enriched Analysis

This section summarises specific changes adopted in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis when the Higgs boson production mode is ggF.

$\not{E}_T$ requirements
The decrease on signal acceptance and the low background rejection motivated the change of the previous $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ requirement in the $e\mu$+μe final states. The more robust and performant $p_{\mathrm{T}}^{\mathrm{miss}}$ is now used to define the selection in this case. Complete optimisation studies of the $\not{E}_T$ in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis are presented in Section 6.5. Figure 7.3 shows the $p_{\mathrm{T}}^{\mathrm{miss}}$ spectrum at the preselection level with $N_{\mathrm{jets}} = 0$ and $N_{\mathrm{jets}} = 1$ separately. Vertical arrows in these distributions indicate the threshold of the $\not{E}_T$ selection requirement in $e\mu$+μe and ee+$\mu\mu$ final candidates.

Introduction of max$(m_{\mathrm{T}}(W))$
Processes with at least one real W boson typically have a large value of $\max(m_{\mathrm{T}}(W))$, as defined in Eq. 6.4, for at least one of the two leptons. In light of this, to further reduce multi-jet and Z/DY $\to \tau\tau$ contributions in the H+1j analysis with $e\mu$+μe final states it is possible to introduce a new lower bound in this quantity: $\max(m_{\mathrm{T}}(W)) > 50$ GeV. Moreover, this requirement is inverted in the CRs in order

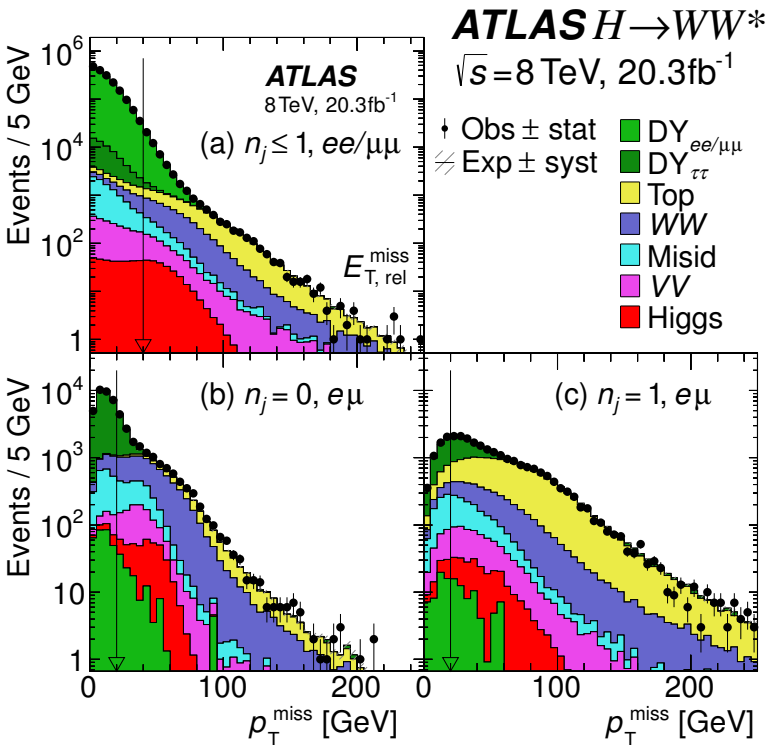

FIGURE 7.3: Top distribution corresponds to the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ distribution for the $ee{+}\mu\mu$ final states with $N_{\mathrm{jets}} \leq 1$. Bottom distributions represent the jet-corrected track-based $\not\!\!E_T$ quantity after applying the preselection criteria for different flavour final states. Bottom left plot corresponds to events without jets and bottom right has required $N_{\mathrm{jets}} = 1$. The arrows mark the thresholds of the selection requirements. See Fig. 7.1 for plotting details.

to increase the purity of the corresponding process in the background phase space. Details on the $\max(m_{\mathrm{T}}(W))$ magnitude are given in Section 6.5.1.2.

Double bining in m_{T} fit

The number of bins in which the m_{T} distribution is divided for the likelihood fit has been doubled: from 5 to 10 and from 3 to 6 in the $H{+}0j$ and $H{+}1j$ analysis, respectively. The double bining increases the expected significance by 3% and 5% in the $e\mu{+}\mu e$ and $ee{+}\mu\mu$ final states, respectively. Studies using the double bining in the m_{T} distribution are presented in Section 6.6.

Signal region boundaries

The upper bound of the ggF signal regions with $N_{\mathrm{jets}} \leq 1$ is moved from $m_{\ell\ell} < 50\,\mathrm{GeV}$

to $m_{\ell\ell} < 55\,\text{GeV}$. This change increases the signal acceptance and leads to a 1.5% improvement in the expected significance for the different flavour final state.

3D fit

The signal regions for the $H{+}0j$ and $H{+}1j$ analyses in the $e\mu{+}\mu e$ final states are subdivided in $m_{\ell\ell}$ and m_{T}, as well as in $p_{\text{T}}^{\text{sub}}$ now. The signal region is further divided using the $p_{\text{T}}^{\text{sub}}$ spectrum into three bins, at 15 GeV and 20 GeV boundaries, as shown in Fig 7.1. These additional regions facilitate the inclusion of data with lower $p_{\text{T}}^{\text{sub}}$ and increases the sensitivity, improving the expected significance by 7.6%. This improvement is due to the decrease in statistical uncertainties, which results from the addition of data with $10 < p_{\text{T}}^{\text{sub}} < 15\,\text{GeV}$, as well as exploiting different signal over background ratios in the different regions.

$H{+}2j$ ggF-enriched category

A new ggF-enriched category is added to the analysis. This additional signal region contains $N_{\text{jets}} \geq 2$ but is orthogonal to the VBF-enriched phase space. The non-overlap is guaranteed by requiring the events to fail at least one of the VBF-specific selections. In the $H{+}2j$ ggF-enriched category only the $e\mu{+}\mu e$ final state is analysed due to the relatively low expected significance achieved by the $ee{+}\mu\mu$ sample. This region has a large acceptance of ggF+2j events resulting in a signal purity of 3.3%, so it adds to the sensitivity of the search in spite of the large $t\bar{t}$ and Wt contaminations. The improvement in the expected significance by including the $H{+}2j$ ggF-enriched category is ∼3.8%.

Same-sign control region

In the $H{+}0j$ and $H{+}1j$ with different flavour final states, the $W\gamma^{(*)}$, Z/γ^*, WZ, and ZZ backgrounds are currently normalised using a control region defined with the exact selection of the signal regions, except without any subdivision in $m_{\ell\ell}$ or $p_{\text{T,sub}}$ and with the leptons required to have same sign. These backgrounds were previously estimated from statistically-limited simulated samples. The benefit of lowering the p_{T} thresholds for the leptons and softening the $\not{E}_T$ requirements enhance the statistics of this region. The same-sign control region improves the significance in the different flavour final states by about 4.5% due to the reduction on luminosity and theory uncertainties.

$N_{\text{jets}} = 1$ top estimate

The top background estimate in the $H{+}1j$ analysis is extrapolated from a b-tagged control region to the signal region, which has a veto on b-tagged jets. The efficiencies of the b-jet tagging and veto requirements are now extracted from an auxiliary

control region with exactly two jets, one or two of which must be tagged. Previously, it was taken from simulation. The new top estimate procedure reduces the impact on b-tagging efficiency and improves the expected significance by ∼1.2% in the different flavour final states.

$Z/\mathrm{DY} \to \tau\tau$ control region

The $Z/\mathrm{DY} \to \tau\tau$ background in the $H{+}1j$ analysis is now estimated using a control region with $N_{\mathrm{jets}} = 1$, $m_{\ell\ell} < 80\,\mathrm{GeV}$, and $m_{\tau\tau} > (m_Z - 25)\,\mathrm{GeV}$ in addition to the signal region selection up to $Z/\mathrm{DY} \to \tau\tau$ veto. The collinear mass approximation is computed using the $p_{\mathrm{T}}^{\mathrm{miss}}$ because of its better resolution in the peak around the Z mass compared with the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement. Figure 7.4 compares the $m_{\tau\tau}$ spectrum computed using these $\not{E}_T$ definitions. The $N_{\mathrm{jets}} = 0$ and $N_{\mathrm{jets}} = 1$ CRs are now both included in the profile likelihood fit as control regions for obtaining the corresponding normalisation factors.

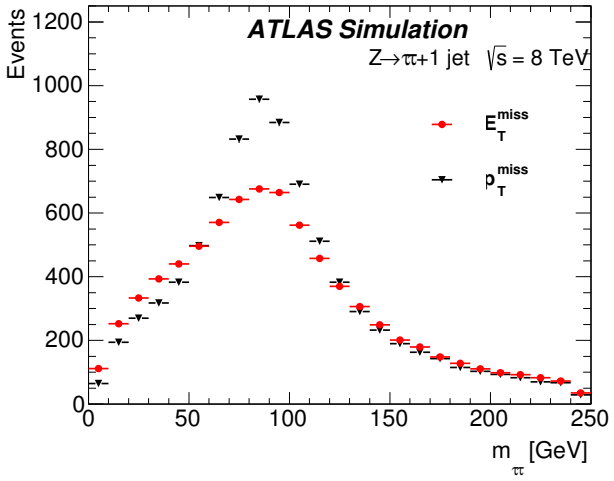

FIGURE 7.4: Collinear approximation of the $\tau\tau$ invariant mass with the default $Z/\mathrm{DY} \to \tau\tau$ MC simulation is shown, using the calorimeter-based ($E_{\mathrm{T}}^{\mathrm{miss}}$, red circles) and the $p_{\mathrm{T}}^{\mathrm{miss}}$ ($p_{\mathrm{T}}^{\mathrm{miss}}$, black triangles). The comparison is shown in the $N_{\mathrm{jets}} = 1$ signal region after the b-veto requirement. A much stronger peak around m_Z is observed if $m_{\tau\tau}$ is computed using the $p_{\mathrm{T}}^{\mathrm{miss}}$ definition. Hence, the requirement on $m_{\tau\tau} < 66\,\mathrm{GeV}$ applied for selecting Higgs candidate events is highly favoured by using $p_{\mathrm{T}}^{\mathrm{miss}}$ magnitude.

p_{T}^Z reweighting

The generated p_{T} of Z bosons in the MC simulation is reweighted using the ratio of $Z \to \mu\mu$ spectrums in data and MC.

Sherpa Z/γ^*
To improve the modelling of $e\mu$+μe final states produced with an electron converted from a photon, a Sherpa Z/γ^* prediction is included in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis.

Many of the above changes when applied together lead to a bigger improvement to the analysis than the sum of the individual changes. For example, when adding data in the $10 < p_{\mathrm{T}}^{\mathrm{sub}} < 15\,\mathrm{GeV}$ range and including the same-sign control region the final improvements in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis is enhanced. This is due to the statistical uncertainties are also lower in the same-sign control region with the additional low-p_{T} events.

7.2.3 Changes in the VBF-enriched Analysis

On top of the improvements described in Section 7.2.1, the following specific changes were imposed in the H+2j VBF-enriched analysis.

$\not{E}_T$ threshold
Extensive studies on cut optimisation were performed, as presented in Section 6.5.1.3 and 6.5.3.2. The most significant change is the decision of dropping the $\not{E}_T$ threshold in the $e\mu$+μe final states. This results in 6% gain in expected significance.

Improvement on MC generators
The baseline generators for top and WW backgrounds have been changed. For simulation of the top background the generator used currently is POWHEG, instead of the previous $MC@NLO$ choice. POWHEG $t\bar{t}$ MC models dijet kinematics better in the VBF phase space, which has been extensively studied in the top control region. For the WW expectation the POWHEG samples are replied by SHERPA which produces up to three jets from ME. Hence, the new generator is more suitable for WW+2jets prediction in the VBF-enriched signal region.

m_{jj} split
For the selection-based analysis, the signal regions is further split into two m_{jj} regions, $600 < m_{jj} < 1000\,\mathrm{GeV}$ and $m_{jj} > 1000\,\mathrm{GeV}$. The split improves the significance by 5% due to the VBF signal distribution presents longer m_{jj} tail than the backgrounds. This results in better signal over background ratio in the high m_{jj} region, which is exploited by the statistical fit procedure.

Boosted Decision Tree method

The main change in the strategy of the $H+2j$ VBF-enriched analysis is the introduction of a boosted decision tree (BDT) method as baseline approach instead of the previous selection-based procedure. The $H+2j$ VBF-enriched sample is analysed using a BDT multivariate method [150–152], that considers VBF Higgs boson production as signal and the rest of the processes as background, including ggF Higgs boson production. The BDT method relies on the correlation among different discriminating variables. Using BDT method for the VBF analysis brings up to 30% improvement with respect to the optimised cut-based strategy, which is also performed for cross-check purposes. More details about this method and the variables used as input to the BDT are given in Section 7.3.2.

7.2.4 Summary

After all the improvements described above, the strategy of the $H\to WW^{(*)}\to \ell\nu\ell\nu$ search results in five categories defined by combinations of lepton-flavour samples and number of jets in the ggF-enriched analysis. For the $H+2j$ VBF-enriched analysis, there are two categories differentiated by the flavour of the leptons candidates.

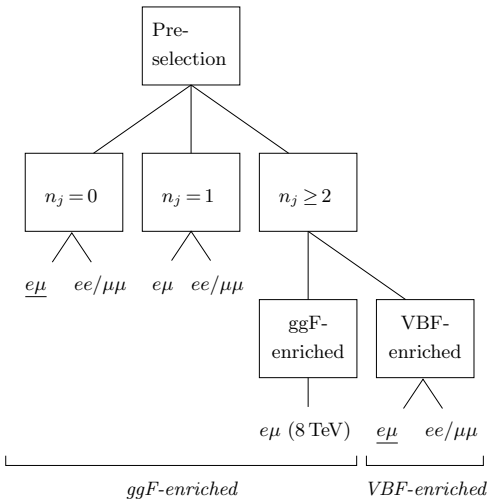

FIGURE 7.5: Analysis divisions in categories based on jet multiplicity (n_j) and lepton flavour final states ($e\mu$ and $ee/\mu\mu$). Note that the different flavoured leptons categories are denoted as $e\mu$, which refers to the combined $e\mu/\mu e$ final states. The most sensitive signal region for ggF production is $H+0j$ category with different flavoured leptons in the final states, while for VBF production this corresponds to the $H+2j$ category with $e\mu/\mu e$ final states. These two samples are underlined.

The schematic shown in Fig. 7.5 summarises the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis categories. The improvement in expected significance driven by the main changes adopted in the ggF-enriched analysis with respect to the strategy described in Chapter 5 are summarised in Tab. 7.1. The expected significance in the most sensitive final state highly increases due to two $p_{\mathrm{T}}^{\mathrm{miss}}$-related improvements. The introduction of the $p_{\mathrm{T}}^{\mathrm{miss}}$ for computing the m_{T} enhances the expected significance by 9% due to the better resolution of this definition compared with the previously used $E_{\mathrm{T}}^{\mathrm{miss}}$. Moreover, lowering the $\not\!\!E_T$ selection in $e\mu$+μe final states relies on 7% improvement in the expected significance when combining $N_{\mathrm{jets}} = 0$ and $N_{\mathrm{jets}} = 1$ jet categories. After all changes applied and the inclusion of the H+2j ggF-enriched region, the expected significance increases up to ∼56%.

TABLE 7.1: Expected significance (Z_0 exp) evolution throughout the main improvements in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ search for different flavour final states in the ggF-enriched phase space. The last two columns quote the Z_0 relative difference with respect to the previous change (Δ / Prev) and with respect to the reference value (Δ / Ref) extracted from the results presented in Chapter 5. The two highest improvements (highlighted in bold) are related with $p_{\mathrm{T}}^{\mathrm{miss}}$ benefits. All energy and transverse momentum quantities are given in GeV.

Change in $e\mu$+μe for ggF analysis	Z_0 exp (σ)	Δ / Prev (%)	Δ / Ref (%)
Reference (from Chapter 5)	2.80	-	-
Change in MC samples	2.84	1.3	1.3
Dilepton Triggers and $p_{\mathrm{T}}^{\mathrm{lead}} > 22$	3.04	4.0	8.6
Improved W+jets	3.15	3.6	12.5
m_{T} based on $p_{\mathrm{T}}^{\mathrm{miss}}$	3.43	**9.0**	22.5
Double m_{T} bins	3.62	5.5	29.3
$p_{\mathrm{T}}^{\mathrm{sub}} > 10$	3.73	3.0	33.2
Same-sign CR	3.87	3.7	38.2
$p_{\mathrm{T}}^{\mathrm{miss}} > 20$	4.23	**7.0**	51.1
Adding ggF+2j category	4.36	3.0	55.7

Table 7.2 quotes the improvements in the expected significance for the modifications adopted in the H+2j VBF-enriched analysis. Note that the second row, quoted as common changes, includes all the modifications described in Section 7.2.1 (dilepton triggers, lower lepton p_{T} thresholds...). The sum of the improvements related to the $\not\!\!E_T$ optimisation studies presented in Chapter 6 enhance the expected significance by ∼14%. The BDT technique, rather than the selection-based

approach, improves the sensitivity of the expected VBF results by 70% relative to the previous analysis described in Chapter 5.

TABLE 7.2: Expected significance evolution throughout the main improvements in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis for different flavour final states in the VBF-enriched phase space. Expected significance (Z_0 exp) evolution throughout the main improvements in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis for different flavour final states in the VBF-enriched phase space. The last two columns quote the Z_0 relative difference with respect to the previous change (Δ / Prev) and with respect to the reference value (Δ / Ref) extracted from the results presented in Chapter 5. All energy and transverse momentum quantities are given in GeV.

Change in $e\mu$+μe for VBF analysis	Z_0 exp (σ)	Δ / Prev (%)	Δ / Ref (%)
Reference (from Chapter 5)	1.31	-	-
Common changes	1.56	19	19
m_{T} based on $p_{\mathrm{T}}^{\mathrm{miss}}$	1.61	3.2	22.9
$\not{E}_T$ cut dropped	1.66	3.1	26.7
Other $p_{\mathrm{T}}^{\mathrm{miss}}$-based variables	1.78	7.2	40.5
m_{jj} split	1.93	8.4	47.3
BDT technique as baseline	2.23	15.5	70.2

7.3 New Selection Performance

This section summarises the new $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis selection after including the changes presented in Section 7.2. The final event yields in each analysis category for the 8 TeV data are shown in Section 7.3.1 and 7.3.2 for the ggF- and VBF-enriched regions, respectively. The BDT method used in the Higgs search for VBF production mode is also detailed in Section 7.3.2.

7.3.1 Gluon-gluon Fusion Enriched Region

After the improvements listed in Section 7.2, the Higgs signal acceptance is increased by 75% and 50% in the H+0j and H+1j categories, respectively, compared with the previous analysis. The higher signal acceptance is mainly achieved by switching from $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ to $p_{\mathrm{T}}^{\mathrm{miss}}$ in $e\mu$+μe final states, and lowering the $p_{\mathrm{T}}^{\mathrm{sub}}$ threshold to 10 GeV. Moreover, the inclusion of the dilepton triggers, in addition to single lepton triggers, allows reduction of the $p_{\mathrm{T}}^{\mathrm{lead}}$ threshold to 22 GeV. Finally, the signal kinematic region in the ggF-enriched search is extended from 50 to 55 GeV, which also enhances

the signal acceptance. The total signal acceptance, including all categories and production modes, at 8 TeV and for a Higgs boson mass of $m_{\mathrm{H}} = 125.36\,\mathrm{GeV}$ increases by roughly a factor two, from 5.3% to 10.2%. The selection requirements for the ggF-enriched analysis are summarised in Tab. 7.3.

TABLE 7.3: Event selection summary for the ggF-enriched search in the $H\to WW^{(*)}\to \ell\nu\ell\nu$ analysis for each jet category. Selection requirements specific to $e\mu$+μe and ee+$\mu\mu$ are noted as such; otherwise, they apply to both. A dash (-) indicates no selection. All energy-related values are in GeV.

Category	ggF-enriched		
	$N_{\mathrm{jets}} = 0$	$N_{\mathrm{jets}} = 1$	$e\mu$+μe: $N_{\mathrm{jets}} \geq 2$
Pre-selection	Two isolated leptons ($\ell = e, \mu$) Leptons with opposite charge $p_{\mathrm{T}}^{\mathrm{lead}} > 22$ and $p_{\mathrm{T}}^{\mathrm{sub}} > 10$ $e\mu$+μe: $m_{\ell\ell} > 10$ ee+$\mu\mu$: $m_{\ell\ell} > 12$ and $\lvert m_{\ell\ell} - m_Z \rvert > 15$		
$e\mu$+μe: $\not{E}_T$	$p_{\mathrm{T}}^{\mathrm{miss}} > 20$	$p_{\mathrm{T}}^{\mathrm{miss}} > 20$	$p_{\mathrm{T}}^{\mathrm{miss}} > 20$
ee+$\mu\mu$: $\not{E}_T$ and hadronic recoil	$E_{\mathrm{T,rel}}^{\mathrm{miss}} > 40$ $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}} > 40$ $f_{\mathrm{recoil}} < 0.1$	$E_{\mathrm{T,rel}}^{\mathrm{miss}} > 40$ $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}} > 35$ $f_{\mathrm{recoil}} < 0.1$	- - -
General selection	- $\lvert \Delta\phi_{\ell\ell,MET} \rvert > \pi/2$ $p_{\mathrm{T}}^{\ell\ell} > 30$	$N_{b\text{-jet}} = 0$ $e\mu$+μe: $m_{\mathrm{T}}^{\ell} > 50$ $e\mu$+μe: $Z/\mathrm{DY} \to \tau\tau$ veto	$N_{b\text{-jet}} = 0$ VBF orthogonality $Z/\mathrm{DY} \to \tau\tau$ veto
$H\to WW^{(*)}\to \ell\nu\ell\nu$ topology	$m_{\ell\ell} < 55$ $\Delta\phi_{\ell\ell} < 1.8$ $e\mu$+μe: split $m_{\ell\ell}$ and $p_{\mathrm{T}}^{\mathrm{sub}}$ Fit m_{T}	$m_{\ell\ell} < 55$ $\Delta\phi_{\ell\ell} < 1.8$ $e\mu$+μe: split $m_{\ell\ell}$ and $p_{\mathrm{T}}^{\mathrm{sub}}$ Fit m_{T}	$m_{\ell\ell} < 55$ $\Delta\phi_{\ell\ell} < 1.8$ – Fit m_{T}

Table 7.4 collects the final event yields for the Higgs boson signal and background processes in the ggF-enriched region for 8 TeV data. Figure 7.6 presents the m_{T} distribution for the H+0j, H+1j and H+2j ggF-enriched categories using the 8 TeV data.

7.3.2 Vector Boson Fusion Enriched Region

The baseline strategy for the VBF search relies on the BDT technique instead of the previous selection-based approach used so far. A decision tree is a collection of cuts designed to classify events as signal-like or background-like. A given signal event is correctly identified if it is placed in a signal-dominated leaf, and vice-versa for background events. After the initial tree is built, another tree is grown to better separate the signal and background events that were misidentified by the first tree. This proceeds iteratively until there is a collection of a specified number of trees, in

TABLE 7.4: ggF-enriched event yields in each category for the 8 TeV data. The N_{sig} columns show the expected signal yields from the ggF and VBF production modes. For each group separated by a horizontal line, the first line gives the combined values for the different final states. The quoted uncertainties include the theoretical and experimental systematic sources and those due to sample statistics. Values less than 0.1 (0.01) events are written as 0.0 (-).

Selection	Summary				Composition of N_{bkg}						
	N_{obs}	N_{bkg}	N_{sig}		N_{WW}	N_{top}		N_{misid}		N_{VV}	N_{DY}
			N_{ggF}	N_{VBF}		$N_{t\bar{t}}$	N_t	N_{Wj}	N_{jj}		
$N_{\text{jets}} = 0$	3750	3430 ±90	300 ±50	8 ±4	2250 ±95	112 ±9	195 ±15	360 ±60	16 ±5	420 ±40	78 ±21
$e\mu$	1430	1280 ±40	129 ±20	3.0 ±2.1	830 ±34	41 ±3	73 ±6	149 ±29	10.1 ±3.6	167 ±21	14 ±2.4
μe	1212	1106 ±35	97 ±15	2.5 ±0.6	686 ±29	33 ±3	57 ±5	128 ±31	3.8 ±1.5	184 ±23	14 ±2.4
ee+$\mu\mu$	1108	1040 ±40	77 ±15	2.4 ±1.7	740 ±40	39 ±3	65 ±5	82 ±16	2 ±0.5	68 ±7	50 ±21
$N_{\text{jets}} = 1$	1596	1470 ±40	102 ±26	17 ±5	630 ±50	150 ±10	385 ±20	108 ±20	8.2 ±3.0	143 ±20	51 ±13
$e\mu$	621	569 ±19	45 ±11	7.4 ±2	241 ±20	58 ±4	147 ±7	51 ±11	5.7 ±2	53 ±10	13.8 ±3.3
μe	508	475 ±18	35 ±9	6.1 ±1.4	202 ±17	45 ±3	119 ±6	37 ±9	2.3 ±0.9	60 ±10	9.3 ±2.5
ee+$\mu\mu$	467	427 ±21	22 ±6	3.6 ±1.8	184 ±15	46 ±4	119 ±10	19 ±4	0.2 ±0.1	31 ±4	28 ±12
$N_{\text{jets}} \geq 2$, $e\mu$+μe ggF	1017	960 ±40	37 ±11	13 ±1.4	138 ±28	56 ±5	480 ±40	54 ±25	62 ±22	56 ±18	117 ±21

a process known as boosting. A weighted average is taken from all these trees to form a BDT output discriminant with values ranging between -1 and $+1$.

In the VBF analysis, the BDT is trained using both $e\mu$+μe and ee+$\mu\mu$ categories. Training with all lepton flavour channels combined improves the performance because it makes use of the full statistical power of the training samples. The BDT is trained without events with sub-leading lepton p_{T} in the lowest range $10 < p_{\text{T}}^{\text{sub}} < 15\,\text{GeV}$, but applied to all events with $p_{\text{T}}^{\text{sub}} > 10\,\text{GeV}$ when fitting to the BDT output. Including low-p_{T} events in training brought solely marginal gain in expected sensitivity (∼0.8%, statistical uncertainty only), but would have required re-computing VBF-specific theory uncertainties. Furthermore, the same BDT trained with 8 TeV MC are used in the 7 TeV data re-analysis. In the VBF analysis the baseline approach uses as input 8 variables to the BDT technique. To determine the optimal number of variables, a grid scan is performed by using the ratio $S/\sqrt{B}$ as figure of merit and employing an $N-1$ minimal loss variable pruning

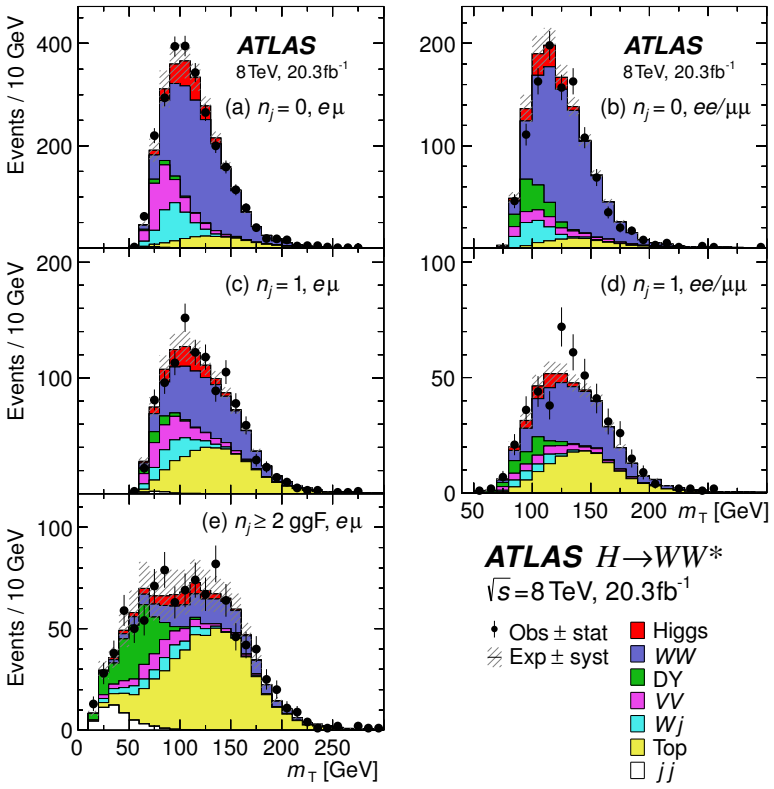

FIGURE 7.6: Distributions of the transverse mass, m_{T}, for the $N_{\mathrm{jets}} \leq 1$ and $N_{\mathrm{jets}} \geq 2$ ggF-enriched categories in the 8 TeV data analysis. The plots are made after requiring all selections up to m_{T} (see Tab. 7.3). See Fig. 7.1 for plotting details.

procedure. This starts with a BDT trained with a maximal set of potentially useful discriminating variables. Then, one variable at a time is removed to evaluate the performance of the $(N-1)$-variable BDT based on the above figure of merit. The best performing $(N-1)$-variable BDT is selected as the new benchmark and the last two steps are iterated until the performance drops significantly.

Before the BDT training, a few requirements are applied. These are common requirements for the BDT, as well as for the VBF selection-based analysis. The latter is also performed as a cross-check of the BDT results. The set of common cuts include pre-selection thresholds, same as for the ggF-enriched categories; the optimised $\not{E}_T$ selection which only concerns the $ee{+}\mu\mu$ category, following the results presented in Chapter 6, and a general selection vetoing events with b-tagged jets or with a reconstructed $m_{\tau\tau}$ consistent with the Z mass peak. Up to here,

the requirements mimic the same structure of the analysis for the VBF-enriched category described in Chapter 5 but including optimised definitions and threshold values. After this selection, a set of VBF topological requirements are also applied to passing events before they enter to the BDT. The VBF process is characterised by the kinematics of the pair of tagged jets and the activity in the rapidity gap between them. With respect to the former, the invariant mass of the tagging jets and their rapidity gap separation are inputs to the BDT. For the cross-check analysis, $|\,\Delta y_{jj}\,| > 3.6$ and $m_{jj} > 600\,\text{GeV}$ are required. For the latter, the population in the central region defined by the two tagged jets is exploited. There is a relatively low level of hadronic activity expected for VBF signal since weak bosons do not exchange colour. The number of extra jets in this gap quantifies such activity, so requiring the absence of jets in this region suppresses processes where the jets are produced via QCD radiation. On the other hand, the Higgs boson decay products tend to be in this central region because of the high boost transferred to them. In this light, the VBF topological requirements include an outside lepton and a central jet vetoes. Those are now expressed in terms of jet (and lepton) centrality, defined as follows,

$$C_{j3} = \left| \eta_{j3} - \frac{\sum \eta_{jj}}{2} \right| \Big/ \frac{\Delta\eta_{jj}}{2}\,, \tag{7.1}$$

where η_{j3} is the pseudorapidity of an extra well-reconstructed jet, $\sum \eta_{jj} = \eta_{j1} + \eta_{j2}$ and $\Delta\eta_{jj} = \left|\eta_{j1} - \eta_{j2}\right|$. The value of C_{j3} increases from zero (when η_{j3} is centered between the tagged jets), to unity (when η_{j3} is aligned in η with either of the tagged jets), and it is greater than unity when $|\eta_{j3}| > |\eta_{j1}$ or $|\eta_{j3}| > |\eta_{j2}$. In light of this, the centrality of any extra jet in the event is required to be $C_{j3} > 1$. The centrality of a given lepton with respect to the tagged jets (C_ℓ) is defined similarly to Eq. 7.1. A requirement of $C_\ell < 1$ is applied to each lepton in both BDT and cross-check analyses and the sum of lepton centralities, defined as,

$$\sum C_\ell = C_{\ell 1} + C_{\ell 2}\,, \tag{7.2}$$

is used as input to the BDT.

Aiming to further reduce top quark pair production, which is frequently accompanied by QCD radiation, the analysis uses the summed transverse momentum vector ($\vec{p_{\mathrm{T}}}^{\,\mathrm{sum}}$) of the final-state objects. This is defined as follows,

$$\vec{p_{\mathrm{T}}}^{\,\mathrm{sum}} = \overrightarrow{p_{\mathrm{T}}}^{\,\ell\ell} + \overrightarrow{p_{\mathrm{T}}}^{\,\mathrm{miss}} + \sum \overrightarrow{p_{\mathrm{T}}}^{\,j}\,, \tag{7.3}$$

where $\sum \vec{p_{\mathrm{T}}}^{j}$ is the sum of the transverse momenta of all jets in the event.

The $\vec{p_{\mathrm{T}}}^{\,\mathrm{sum}}$ magnitude is used as input to the BDT and is required to be $p_{\mathrm{T}}^{\mathrm{sum}} < 15\,\mathrm{GeV}$ in the cross-check analysis. Finally, the sum of the four combinations of lepton-jet invariant mass,

$$\sum m_{\ell j} \;=\; m_{\ell 1,j1} \;+\; m_{\ell 1,j2} \;+\; m_{\ell 2,j1} \;+\; m_{\ell 2,j2}\,, \tag{7.4}$$

is also used as an input to the BDT. Since in the VBF topology tagged jets are more forward whereas the leptons tend to be more central, the $\sum m_{\ell j}$ distributions for the VBF signal and SM background process show differences in shape. The other BDT input variables are those related to the $H\to WW^{(*)}\to \ell\nu\ell\nu$ decay topology: $m_{\ell\ell}$, $\Delta\phi_{\ell\ell}$ and m_{T}, which are also used in the ggF-enriched analysis. The cross-check analysis requires $\Delta\phi_{\ell\ell} < 1.8(2.8)$, for $p_{\mathrm{T}}^{\mathrm{sub}} > (<) 15\,\mathrm{GeV}$, and $m_{\ell\ell} < 50\,\mathrm{GeV}$.

Table 7.5 shows the selection requirements for both analysis, as well as the input variables to the BDT. The BDT output is known as BDT score (O_{BDT}) and it is used in the statistical fit procedure. The binning of the O_{BDT} distribution has been optimised to maximise the expected significance while keeping reasonable MC statistics in each bin, as it is the case for the ggF-analysis. The chosen configuration is four bins with boundaries at $[-0.48,\ 0.3,\ 0.78]$ over the range of $[-1,\ 1]$ in the BDT score. The bin with the lowest BDT score contains the majority of background ($> 90\%$) and thus substantially smaller signal-to-background ratio. In light of this, this bin is vetoed for the BDT and cross-check analyses, as shown in Tab. 7.5.

Table 7.6 collects the final event yields for the Higgs boson signal and background processes in the VBF-enriched region for 8 TeV data using BDT technique. The total signal contribution to the final candidates when only considering the VBF production mode is $\sim$16%. Distributions in Fig. 7.7 show the O_{BDT} outputs in the $e\mu$+μe and ee+$\mu\mu$ final states. In terms of VBF signal production, the third BDT bin provides the highest purity, with a signal-to-background ratio of approximately two. The m_{T} magnitude is an input to the BDT and the corresponding distributions after the BDT classification are also shown combining all three BDT bins.

7.4 Final Results in $H\to WW^{(*)}\to \ell\nu\ell\nu$ Search

In this section, the observation of the inclusive Higgs boson signal and evidence for the VBF production mode are stablished first. Following that, the excess in data is characterised using the SM Higgs boson as the signal hypothesis. Results include the inclusive signal strength as well as those for the individual ggF and VBF modes.

TABLE 7.5: Event selection summary for the VBF-enriched search using $N_{\text{jets}} \geq 2$. Selection requirements specific to $e\mu$+μe and ee+$\mu\mu$ are noted as such; otherwise, they apply to both. The variables used as inputs to the BDT are noted as such. A dash (-) indicates no selection. All energy-related values are given in GeV.

Category	VBF-enriched: $N_{\text{jets}} \geq 2$	
	Cut-based analysis	BDT analysis
Pre-selection	Two isolated leptons ($\ell = e, \mu$) Leptons with opposite charge Leptons with $p_{\text{T}}^{\text{lead}} > 22$ and $p_{\text{T}}^{\text{sub}} > 10$ $e\mu$+μe: $m_{\ell\ell} > 10$ ee+$\mu\mu$: $m_{\ell\ell} > 12, \mid m_{\ell\ell} - m_Z \mid > 15$	
$\not{E}_T$ requirement	$e\mu$+μe: No requirement ee+$\mu\mu$: $E_{\text{T}}^{\text{miss}} > 45$ and $p_{\text{T}}^{\text{miss}} > 45$	
General selection	$N_{b\text{-jet}} = 0$ $e\mu$+μe : $Z/\text{DY} \to \tau\tau$ veto	
VBF topology	$C_{\ell 1} < 1$ and $C_{\ell 2} < 1$ $C_{j3} > 1$ for $j3$ with $p_{\text{T}}^{j3} > 20$ $O_{\text{BDT}} \geq -0.48$	
	$\mathbf{p}_{\text{T}}^{\text{sum}} < 15$	Input 1 to BDT
	$\sum C_\ell$ not used	Input 2 to BDT
	$\mid \Delta y_{jj} \mid > 3.6$	Input 3 to BDT
	$m_{jj} > 600$	Input 4 to BDT
$H \to WW^{(*)} \to \ell\nu\ell\nu$ topology	$m_{\ell\ell} < 50$	Input 5 to BDT
	$\Delta\phi_{\ell\ell} < 1.8$ ($p_{\text{T}}^{\text{sub}} > 15$) $\Delta\phi_{\ell\ell} < 2.8$ ($p_{\text{T}}^{\text{sub}} < 15$)	Input 6 to BDT
	m_{T} split	Input 7 to BDT
BDT-specific	$\sum m_{\ell j}$ not used	Input 8 to BDT

Figure 7.8 presents the transverse mass distribution at the end of the event selection combining the 2011 and 2012 data samples in all flavour final states with $N_{\text{jets}} \leq 1$. A clear excess of ~ 500 data events over the expected background is observed. The excess agrees well within the expectation of the signal from a SM Higgs boson with $m_{\text{H}} = 125\,\text{GeV}$. The profile likelihood fit is used to search for a signal and to characterise the production rate in the ggF and VBF production modes. In the following, all results are quoted for a Higgs boson mass corresponding to the central value of the ATLAS measurement in the $ZZ \to 4\ell$ and $\gamma\gamma$ decay modes, $m_{\text{H}} = 125.36 \pm 0.41\,\text{GeV}$ [153].

TABLE 7.6: VBF-enriched post-fit yields in each category for the 8 TeV data. The N_{sig} columns show the expected signal yields from the ggF and VBF production modes, with values scaled to the observed combined signal strength (see Sec. 7.4). The first line gives the combined values for the different BDT bins and final lepton states. The quoted uncertainties include the theoretical and experimental systematic sources and those due to sample statistics. Values less than 0.01 events are written as −.

Selection	Summary				Composition of N_{bkg}						
	N_{obs}	N_{bkg}	N_{sig}		N_{WW}	N_{top}		N_{misid}		N_{VV}	N_{DY}
			N_{ggF}	N_{VBF}		$N_{t\bar{t}}$	N_t	N_{Wj}	N_{jj}		
$N_{\text{jets}} \geq 2$, VBF	130	99 ±9	7.7 ±2.6	21 ±3	11 ±3.5	5.5 ±0.7	29 ±5	4.7 ±1.4	2.8 ±1.0	4.4 ±0.9	38 ±7
$e\mu$+μe bin 1	37	36 ±4	3.3 ±1.2	4.9 ±0.5	5.0 ±1.5	3.0 ±0.6	15.6 ±2.6	3.2 ±1.0	2.3 ±0.8	2.3 ±0.7	3.6 ±1.5
$e\mu$+μe bin 2	14	6.5 ±1.3	1.4 ±0.5	4.9 ±0.5	1.7 ±0.7	0.3 ±0.4	2.0 ±1.0	0.4 ±0.1	0.3 ±0.1	0.7 ±0.2	0.6 ±0.2
$e\mu$+μe bin 3	6	1.2 ±0.3	0.4 ±0.3	3.8 ±0.7	0.3 ±0.1	0.1 ±0.0	0.3 ±0.1	− −	− −	0.1 ±0.0	0.2 ±0.1
ee+$\mu\mu$ bin 1	53	46 ±6	1.7 ±0.6	2.6 ±0.3	3.1 ±1.0	1.7 ±0.3	10.1 ±1.6	0.9 ±0.2	0.2 ±0.1	1.0 ±0.3	28 ±5
ee+$\mu\mu$ bin 2	14	8.4 ±1.8	0.7 ±0.3	3.0 ±0.4	0.9 ±0.3	0.3 ±0.2	1.2 ±0.5	0.2 ±0.1	− −	0.3 ±0.1	5.2 ±1.7
ee+$\mu\mu$ bin 3	6	1.1 ±0.4	0.2 ±0.2	2.1 ±0.4	0.1 ±0.1	0.1 ±0.0	0.2 ±0.1	− −	− −	− −	0.5 ±0.3

7.4.1 Observation of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ Decay Mode

The statistic test q_μ, as defined in Section 5.11, is used to quantify the significance of the excess of events observed in Fig. 7.8. The observed and expected p_0 are shown as a function of m_{H} in Fig. 7.9. The observed curve presents a broad minimum centered around $m_{\text{H}} = 130\,\text{GeV}$, in contrast with the higher p_0-values observed for the lower and higher values of m_{H}. The shapes of the observed and expected curves are in good agreement.

The probability p_0 can equivalently be expressed in terms of the number of standard deviations Z_0. The observed local significance for a SM Higgs boson signal with $m_{\text{H}} = 125.36\,\text{GeV}$ is 6.1 σ. The expected significance for a SM Higgs boson at the same mass is 5.8 σ. This result establishes the discovery for the SM Higgs boson signal in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ channel alone.

In order to assess the compatibility with the SM expectation for a Higgs boson of mass m_{H}, the observed best-fit $\hat{\mu}$ values as a function of m_{H} are shown in Fig. 7.10. The observed $\hat{\mu}$ is close to zero for $m_{\text{H}} > 160\,\text{GeV}$ and crosses unity around $m_{\text{H}} = 125\,\text{GeV}$.

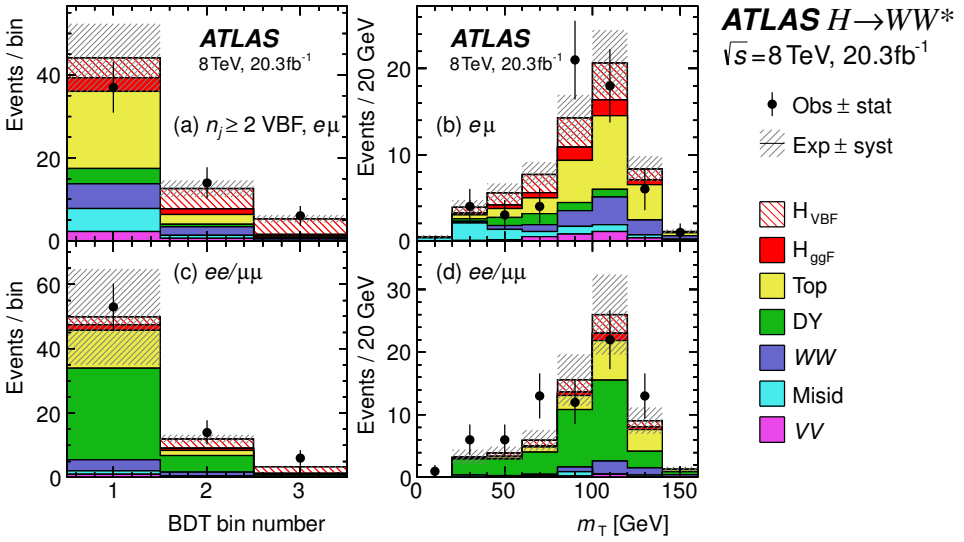

FIGURE 7.7: Post-fit BDT and transverse mass distributions in the $N_{\text{jets}} \geq 2$ VBF-enriched category in the 8 TeV analysis: (a) BDT output in $e\mu$+μe, (b) mT in $e\mu$+μe, (c) BDT output in ee+$\mu\mu$, and m_{T} in ee+$\mu\mu$. For (b) and (d), the three BDT bins are combined. See Fig. 7.1 for plotting details.

Figure 7.11 shows the two-dimensional likelihood contours of (m_{H}, μ). The value ($m_{\text{H}} = 125.36\,\text{GeV}$, $\mu = 1$) lies well within the 68% C.L. contour. It evidences that the observed signal is compatible with those in the high-resolution channels.

7.4.2 Evidence for VBF Production

The H+$2j$ VBF-enriched signal region was optimised for its specific sensitivity to the VBF production process, as described in Sec. 7.3.2. Nevertheless, as can be seen in Table 7.6, the ggF contribution to the VBF-enriched signal region is not negligible, approximately 30%, so it has to be profiled by the global fit together with the extraction of the significance of the signal strength of the VBF production process.

The global likelihood can be evaluated as a function of the ratio $\mu_{\text{VBF}}/\mu_{\text{ggF}}$, with both signal strengths varied independently. The significance derived from testing the ratio $\mu_{\text{VBF}}/\mu_{\text{ggF}} = 0$ is equivalent to the significance of testing $\mu_{\text{VBF}} = 0$, though testing the ratio is advantageous since the branching fraction cancels in $\mu_{\text{VBF}}/\mu_{\text{ggF}}$, while it is implicit in μ_{VBF}. The result is illustrated in Fig. 7.12, which

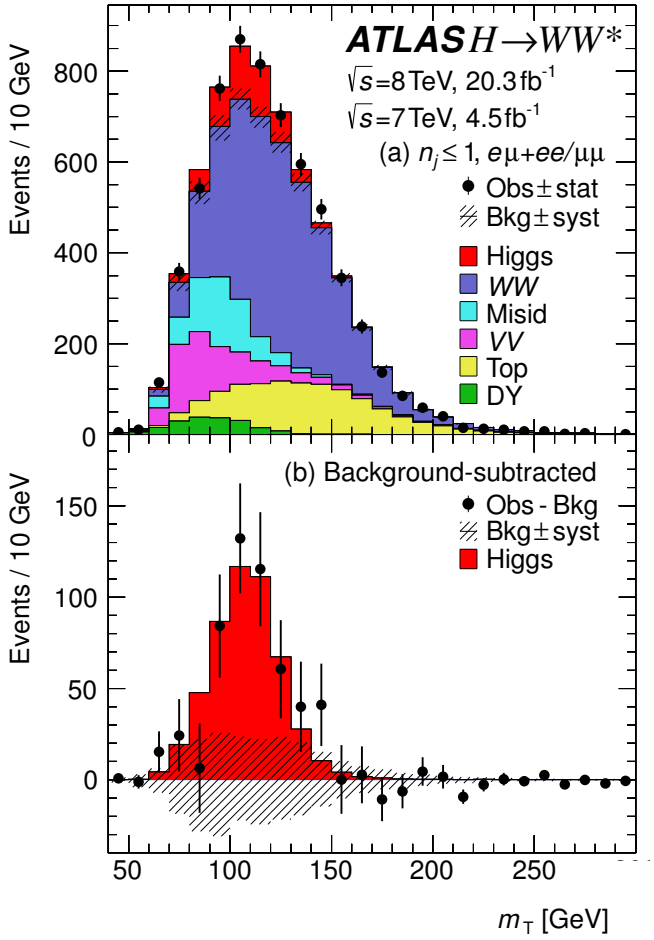

FIGURE 7.8: Post-fit combined transverse mass distribution for $N_{\text{jets}} \leq 1$ for all lepton-flavour samples in 7 and 8 TeV data analyses. The plot in (b) shows the residuals of the data with respect to the estimated background compared to the expected distribution for a SM Higgs boson with $m_{\text{H}} = 125\,\text{GeV}$. The error bars on the data are statistical, $\sqrt{\text{N}_{\text{obs}}}$. The uncertainty on the background (shown as the shaded band around 0) is up to about 25 events per m_{T} bin and partially correlated between bins. Background processes are scaled by post-fit normalisation factors and the signal processes by the observed signal strength μ from the likelihood fit to all regions. Their normalisations also include effects from the pulls of the nuisance parameters.

has a best-fit value for the ratio of,

$$\frac{\mu_{\text{VBF}}}{\mu_{\text{ggF}}} = 1.26\,^{+0.61}_{-0.45}\,(\text{stat.})\,^{+0.50}_{-0.26}\,(\text{syst.}) = 1.26\,^{+0.79}_{-0.53}\,. \tag{7.5}$$

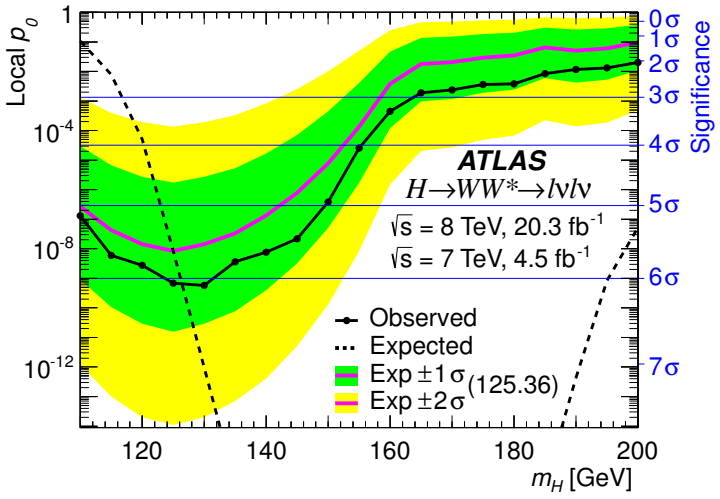

FIGURE 7.9: Local p_0 as a function of m_{H}. The observed values are shown as a solid line with points where p_0 is evaluated. The dashed line shows the expected values given the presence of a signal at each x-axis value. The expected values for $m_{\mathrm{H}} = 125.36\,\mathrm{GeV}$ are given as a solid line without points. The inner (outer) band shaded darker (lighter) represents the one (two) standard deviation uncertainty.

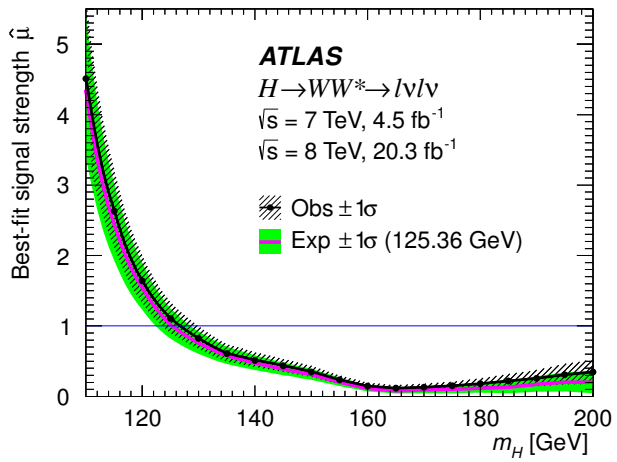

FIGURE 7.10: Best-fit signal strength $\hat{\mu}$ as a function of m_{H}. The observed values are shown as a solid line with points where $\hat{\mu}$ is evaluated. The expected values for $m_{\mathrm{H}} = 125.36\,\mathrm{GeV}$ are shown as a solid line without points. The dashed and shaded (solid) bands represent the 1 σ uncertainties for the observed (expected) values.

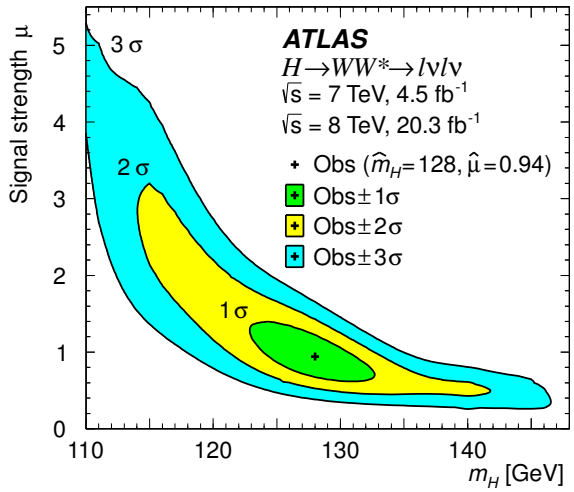

FIGURE 7.11: Observed signal strength μ as a function of m_{H} as evaluated by the likelihood fit. The shaded areas represent the one, two, and three σ contours with respect to the best fit values $\hat{m}_{\mathrm{H}}$ and $\hat{\mu}$.

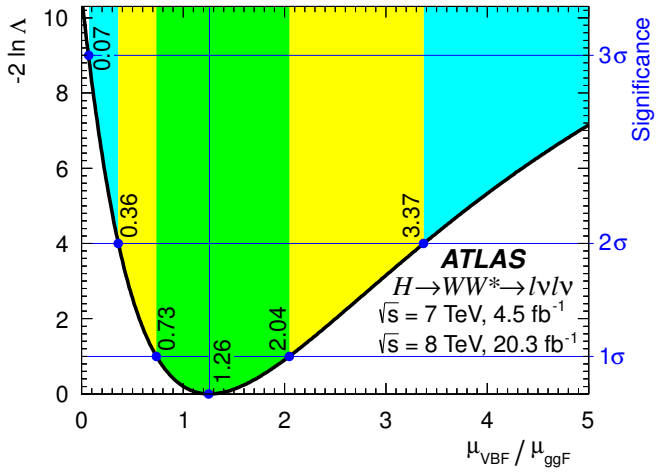

FIGURE 7.12: Likelihood scan as a function of $\mu_{\mathrm{VBF}}/\mu_{\mathrm{ggF}}$ for $m_{\mathrm{H}} = 125.36\,\mathrm{GeV}$. The value of the likelihood at $\mu_{\mathrm{VBF}}/\mu_{\mathrm{ggF}} = 0$ gives the significance of the VBF signal at 3.2 σ. The inner (middle) [outer] band shaded darker (lighter) [darker] represents the one (two) [three] standard deviation uncertainty around the central value represented by the vertical line.

The value of the likelihood at $\mu_{\mathrm{VBF}}/\mu_{\mathrm{ggF}} = 0$ can be interpreted as the observed significance of the VBF production process for $m_{\mathrm{H}} = 125.36\,\mathrm{GeV}$, and corresponds to 3.2 σ. The expected significance for the quoted mass is 2.7 σ. This result

establishes the evidence for the VBF production mode in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ final state.

This result was verified with the cross-check analysis. The expected and observed significances at $m_{\mathrm{H}} = 125.36\,\mathrm{GeV}$ are 2.1 σ and 3.0 σ, respectively. The probability that the difference in Z_0 values, obtained with the cross-check analysis, is larger than the one observed is 79%, reflecting good agreement.

7.4.3 Signal Strength μ

The parameter μ is used to characterise the inclusive Higgs boson signal strength as well as subsets of the signal regions or individual production modes. First, the ggF and VBF processes can be distinguished by using the normalisation parameter μ_{ggF} for the signal predicted for the ggF signal process, and μ_{VBF} for the signal predicted for the VBF signal process. This can be done for a fit to any set of the signal regions in the various categories. In addition, to check that the measured value is consistent among categories, different subsets of the signal regions can be fit. For example, the H+0j and H+1j categories can be compared, or the $e\mu$+μe and ee+$\mu\mu$ categories. To derive these results, only the signal regions are separated; the control region definitions do not change. In particular, the control regions defined using only $e\mu$+μe events are used, even when only ee+$\mu\mu$ signal regions are considered.

7.4.3.1 Combined Signal Strength

The combined Higgs signal strength μ, including 7 and 8 TeV data and all signal region categories, is:

$$\begin{aligned}
\mu &= 1.09 \;{}^{+0.16}_{-0.15}\,(\mathrm{stat.}) \;{}^{+0.08}_{-0.07}\left(\begin{smallmatrix}\mathrm{expt.}\\ \mathrm{syst.}\end{smallmatrix}\right) \;{}^{+0.15}_{-0.12}\left(\begin{smallmatrix}\mathrm{theo.}\\ \mathrm{syst.}\end{smallmatrix}\right) \;\pm 0.03\left(\begin{smallmatrix}\mathrm{lumi.}\\ \mathrm{syst.}\end{smallmatrix}\right) \\
&= 1.09 \;{}^{+0.16}_{-0.15}\,(\mathrm{stat.}) \;{}^{+0.17}_{-0.14}\,(\mathrm{syst.}) \\
&= 1.09 \;{}^{+0.23}_{-0.21}.
\end{aligned} \tag{7.6}$$

The expected value of μ is $1\,{}^{+0.16}_{-0.15}\,(\mathrm{stat.})\,{}^{+0.17}_{-0.13}\,(\mathrm{syst.})$.

The uncertainties are divided according to their source. The statistical uncertainty accounts for the number of observed events in the signal regions and profiled control regions. The statistical uncertainties from MC simulated samples, from non-profiled control regions, and from the extrapolation factors used in the W+ jets background estimate are all included in the experimental uncertainties here and for all results in this section. The theoretical uncertainty includes uncertainties on

the signal acceptance and cross section as well as theoretical uncertainties on the background extrapolation factors and normalisations.

7.4.3.2 Signal Strength in Each Category

In order to check the compatibility with the SM predictions of the ggF and VBF production processes, μ_{ggF} and μ_{VBF} can be simultaneously determined through a fit to all categories because of the different sensitivity to these processes in the various categories. In this fit, the VH contribution is included although there is no dedicated category for it, and the SM value for the ratio $\sigma_{\mathrm{VBF}}/\sigma_{\mathrm{VH}}$ is assumed. The VH production process contributes a small number of events, amounting to about 1% of the expected signal from the VBF process. It is included in the predicted signal yield, and where relevant, is grouped with the VBF signal. Technically, the signal strength $\mu_{\mathrm{VBF+VH}}$ is measured, but for sake of simplicity, the notation μ_{VBF} is used.

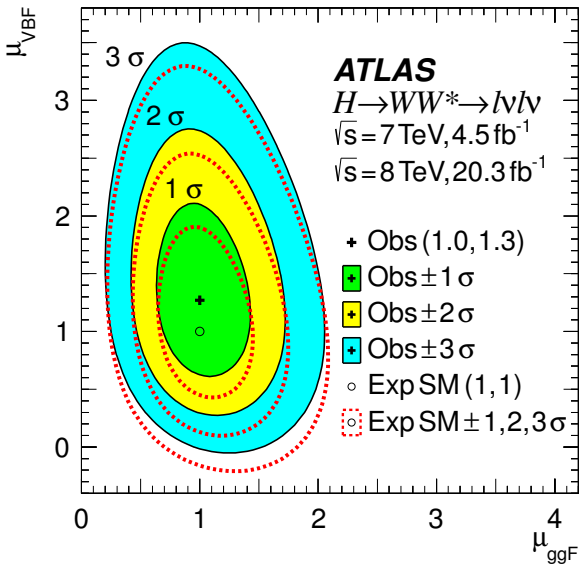

FIGURE 7.13: Likelihood scan as a function of μ_{ggF} and μ_{VBF}. The best-fit observed (expected SM) value is represented by the cross symbol (open circle) and its one, two, and three standard deviation contours are shown by solid lines surrounding the filled areas (dotted lines). The x- and y-axis scales are the same to visually highlight the relative sensitivity.

The signal strength results are shown in Table 7.7 for $m_{\mathrm{H}} = 126.36\,\mathrm{GeV}$. The table includes inclusive results as well as results for individual categories and

production modes. The expected and observed significance for each category and production mode is also shown.

TABLE 7.7: Signal significance Z_0 and signal strength μ. The expected (Exp) and observed (Obs) values are given; μ_{exp} is unity by assumption. For each group separated by a horizontal line, the first row gives the combined result. The total errors are quoted including the statistical and systematic sources and they are explicitly given for each observed Z_0 values. The first three set of values report the results for the inclusive search and when considering ggF and VBF production modes separately. All the quoted results are given assuming $m_{\text{H}} = 125.36\,\text{GeV}$.

Sample	Signal significance		Observed uncertainty				Observed
	Exp.	Obs.	Stat. err.		Syst. err.		central value
	Z_0	Z_0	$+$	$-$	$+$	$-$	μ_{obs}
All N_{jets}, signal	$5.76^{+0.23}_{-0.20}$	$6.06^{+0.23}_{-0.21}$	0.16	0.15	0.17	0.14	$1.09^{+0.23}_{-0.21}$
ggF as signal	$4.34^{+0.30}_{-0.24}$	$4.28^{+0.29}_{-0.26}$	0.19	0.19	0.22	0.18	$1.02^{+0.29}_{-0.26}$
VBF as signal	$2.67^{+0.50}_{-0.43}$	$3.24^{+0.53}_{-0.45}$	0.44	0.40	0.30	0.21	$1.27^{+0.53}_{-0.45}$
H+0j	$3.70^{+0.35}_{-0.30}$	$4.08^{+0.37}_{-0.32}$	0.22	0.22	0.30	0.23	$1.15^{+0.37}_{-0.32}$
$e\mu$	$2.89^{+0.41}_{-0.36}$	$3.07^{+0.43}_{-0.38}$	0.30	0.29	0.32	0.24	$1.08^{+0.43}_{-0.38}$
μe	$2.36^{+0.49}_{-0.44}$	$3.12^{+0.54}_{-0.48}$	0.38	0.37	0.39	0.30	$1.40^{+0.54}_{-0.48}$
ee+$\mu\mu$	$1.43^{+0.74}_{-0.70}$	$0.71^{+0.68}_{-0.66}$	0.45	0.44	0.51	0.50	$0.47^{+0.68}_{-0.66}$
H+1j	$2.60^{+0.51}_{-0.41}$	$2.49^{+0.50}_{-0.41}$	0.33	0.32	0.38	0.26	$0.96^{+0.50}_{-0.41}$
$e\mu$+μe	$2.56^{+0.51}_{-0.42}$	$2.83^{+0.56}_{-0.45}$	0.35	0.35	0.43	0.29	$1.16^{+0.56}_{-0.45}$
ee+$\mu\mu$	$1.02^{+1.12}_{-0.98}$	$0.21^{+1.02}_{-0.97}$	0.80	0.76	0.63	0.61	$0.19^{+1.02}_{-0.97}$
H+2j, ggF, $e\mu$	$1.21^{+0.96}_{-0.83}$	$1.44^{+0.91}_{-0.84}$	0.70	0.68	0.70	0.49	$1.20^{+0.99}_{-0.84}$
H+2j, VBF	$3.38^{+0.42}_{-0.36}$	$3.84^{+0.45}_{-0.38}$	0.36	0.33	0.27	0.19	$1.20^{+0.45}_{-0.38}$
$e\mu$+μe	$3.01^{+0.48}_{-0.40}$	$3.02^{+0.47}_{-0.39}$	0.40	0.35	0.24	0.16	$0.98^{+0.47}_{-0.39}$
ee+$\mu\mu$	$1.58^{+0.84}_{-0.67}$	$2.96^{+0.97}_{-0.78}$	0.83	0.71	0.51	0.33	$1.98^{+0.97}_{-0.78}$

The μ values are consistent with each other and with unity within the assigned uncertainties. In addition to provide a consistency check, these results illustrate the sensitivity of the different categories. For the overall signal strength, the contribution from the H+2j VBF-enriched category is second after the H+0j ggF category, and the H+2j ggF-enriched contribution is comparable to those in the $N_{\text{jets}} \leq 1$ ee+$\mu\mu$ categories. In all of these results, the signal acceptance for all production modes is evaluated assuming a SM Higgs boson.

7.4.4 Higgs Couplings to Fermions and Vector Bosons

The values of μ_{ggF} and μ_{VBF} can be used to test the compatibility of the fermionic and bosonic couplings of the Higgs boson with the SM prediction using a framework motivated by the leading-order interactions [25]. The parametrisation uses the scale factors applied to all fermionic couplings (κ_F), and applied to all bosonic couplings (κ_V). These parameters are unity for the SM.

In particular, the ggF production cross section is proportional to κ_F^2 through the top-quark or bottom-quark loops at the production vertex, and the VBF production cross section is proportional to κ_V^2. The branching fraction $\mathcal{B}_{H\to WW^*}$ is proportional to κ_V^2 and inversely proportional to a linear combination of κ_F^2 and κ_V^2. This model assumes that there are no non-SM decay modes, so the denominator corresponds to the total decay width in terms of the fermionic and bosonic decay amplitudes. The formulae, following Ref. [25], are

$$\begin{aligned} \mu_{\mathrm{ggF}} &\propto \frac{\kappa_F^2 \cdot \kappa_V^2}{(\mathcal{B}_{H\to f\bar{f}} + \mathcal{B}_{H\to gg})\,\kappa_F^2 + (\mathcal{B}_{H\to VV})\,\kappa_V^2} \\ \mu_{\mathrm{VBF}} &\propto \frac{\kappa_V^4}{(\mathcal{B}_{H\to f\bar{f}} + \mathcal{B}_{H\to gg})\,\kappa_F^2 + (\mathcal{B}_{H\to VV})\,\kappa_V^2}. \end{aligned} \tag{7.7}$$

The small contribution from $\mathcal{B}_{H\to gg}$ depends on both κ_F and κ_V and is not explicitly shown. Because $(\mathcal{B}_{H\to f\bar{f}} + \mathcal{B}_{H\to gg}) \approx 0.75$, κ_F^2 is the dominant component of the denominator for $\kappa_F^2 \leq 3\,\kappa_V^2$. As a result, the κ_F^2 dependence for the ggF process approximately cancels, but the rate remains sensitive to κ_V. Similarly, the VBF rate scales approximately with κ_V^4/κ_F^2 and the VBF channel provides more sensitivity to κ_F than the ggF channel does in this model. Because Eq. 7.7 contains only κ_F^2 and κ_V^2, this channel is not sensitive to the sign of κ_F or κ_V.

The likelihood scan as a function of κ_V and κ_F is shown in Fig. 7.14. Both the observed and expected contours are shown, and are in good agreement. The relatively low discrimination among high values of κ_F in the plot is due to the functional behaviour of the total ggF yield. The product $\sigma_{\mathrm{ggF}} \cdot \mathcal{B}$ does not depend on κ_F in the limit where $\kappa_F \gg \kappa_V$, so the sensitivity at high κ_F values is driven by the value of μ_{VBF}. The VBF process rapidly vanishes in the limit where $\kappa_F \gg \kappa_V$ due to the increase of the Higgs boson total width and the consequent reduction of the branching fraction to WW bosons. Therefore, within this framework, excluding $\mu_{\mathrm{VBF}} = 0$ excludes $\kappa_F \gg \kappa_V$.

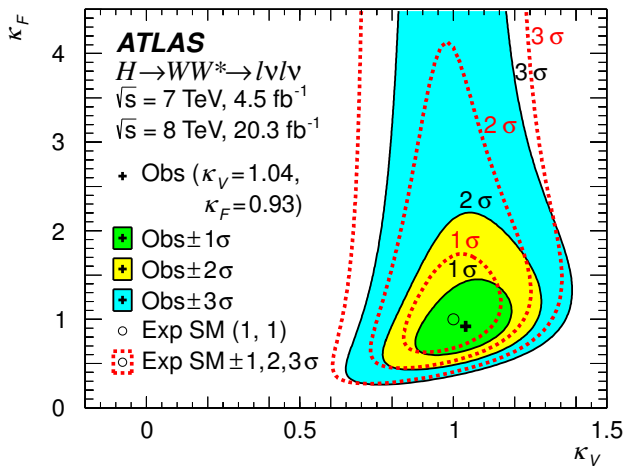

FIGURE 7.14: Likelihood scan as a function of κ_V and κ_F. The best-fit observed (expected SM) value is represented by the cross symbol (open circle) and its one, two, and three standard deviation contours are shown by solid lines surrounding the filled areas (dotted lines). NB. The y-axis spans a wider range than the x-axis.

The best fit values are,

$$\begin{array}{lllll} \kappa_F & = 0.93 & {}^{+0.24}_{-0.18} & {}^{+0.21}_{-0.14} & = 0.93 \; {}^{+0.32}_{-0.23} \\ \kappa_V & = 1.04 & {}^{+0.07}_{-0.08} & {}^{+0.07}_{-0.08} & = 1.04 \; \pm 0.11\,, \\ & & \text{(stat.)} & \text{(syst.)} & \end{array} \tag{7.8}$$

and their correlation is $\rho = 0.47$. The correlation is derived from the covariance matrix constructed from the second-order mixed partial derivatives of the likelihood, evaluated at the best-fit values of κ_F and κ_V.

7.4.5 Higgs Production Cross Section

The measured signal strength can be used to evaluate the product $\sigma \cdot \mathcal{B}_{H\to WW^*}$ for Higgs boson production at $m_{\mathrm{H}} = 125.36\,\mathrm{GeV}$, as well as for the individual ggF and VBF production modes. The central value is simply the product of μ and the predicted cross section used to define it. The uncertainties are similarly scaled, except for the theoretical uncertainties related to the total production yield, which do not apply to this measurement. These are the QCD scale and PDF uncertainties on the total cross sections, and the uncertainty on the branching fraction for $H \to WW^*$.

In practice, the corresponding nuisance parameters are fixed to their nominal values in the fit, effectively removing these uncertainties from consideration. Inclusive cross section measurements are performed for ggF and VBF production. The cross section is also measured for ggF production in defined fiducial volumes. This approach minimises the impact of theoretical uncertainties.

7.4.5.1 Inclusive cross sections

Inclusive cross sections are evaluated at both 7 and 8 TeV for the ggF production process and at 8 TeV for the VBF production process. The 7 TeV VBF cross section is not measured because of the large statistical uncertainty. The signal strengths used for ggF and VBF are determined through a simultaneous fit to all categories, as described in Sec. 7.4.3. The small VH contribution, corresponding to 0.9%, is neglected, and its expected fractional yield is added linearly to the total error. The 7 TeV signal strength ($\mu_{\text{ggF}}^{7\,\text{TeV}}$) and 8 TeV signal strengths ($\mu_{\text{ggF}}^{8\,\text{TeV}}$ and $\mu_{\text{VBF}}^{8\,\text{TeV}}$) are,

$$\begin{array}{lllll} \mu_{\text{ggF}}^{7\,\text{TeV}} & = 0.57 & {}^{+0.52}_{-0.51} & {}^{+0.36}_{-0.34} & {}^{+0.14}_{-0.004} \\ \mu_{\text{ggF}}^{8\,\text{TeV}} & = 1.09 & \pm 0.20 & {}^{+0.19}_{-0.17} & {}^{+0.14}_{-0.09} \\ \mu_{\text{VBF}}^{8\,\text{TeV}} & = 1.45 & {}^{+0.48}_{-0.44} & {}^{+0.38}_{-0.24} & {}^{+0.11}_{-0.06} \\ & & (\text{stat.}) & (\text{syst.}) & (\text{sig.}) \end{array} \tag{7.9}$$

where (sig.) indicates the systematic uncertainties on the total signal yield for the measured process, which does not affect the cross section measurement. The effect of uncertainties on the signal yield for other production modes is included in the systematic uncertainties. In terms of the measured signal strength, the inclusive cross section is defined as

$$\begin{aligned} \left(\sigma \cdot \mathcal{B}_{H\to WW^*}\right)_{\text{obs}} &= \frac{(\text{N}_{\text{sig}})_{\text{obs}}}{\mathcal{A}\cdot\mathcal{C}\cdot\mathcal{B}_{WW\to\ell\nu\ell\nu}} \cdot \frac{1}{\int L\text{dt}} \\ &= \hat{\mu}\cdot(\sigma\cdot\mathcal{B}_{H\to WW^*})_{\text{exp}}\,. \end{aligned} \tag{7.10}$$

In Eq. 7.10, $\mathcal{A}$ is the kinematic and geometric acceptance, and $\mathcal{C}$ is the ratio of the number of measured events to the number of events produced in the fiducial phase space of the detector. The product $\mathcal{A}\times\mathcal{C}$ is the total acceptance for reconstructed events.

The cross sections are measured using the last line of the equation, and the results are,

$$\begin{aligned}
\sigma^{7\,\mathrm{TeV}}_{\mathrm{ggF}} \cdot \mathcal{B}_{H\to WW^*} &= 2.0 \quad \pm 1.7 \quad {}^{+1.2}_{-1.1} \quad = 2.0 \quad {}^{+2.1}_{-2.0} \quad \mathrm{pb} \\
\sigma^{8\,\mathrm{TeV}}_{\mathrm{ggF}} \cdot \mathcal{B}_{H\to WW^*} &= 4.6 \quad \pm 0.9 \quad {}^{+0.8}_{-0.7} \quad = 4.6 \quad {}^{+1.2}_{-1.1} \quad \mathrm{pb} \\
\sigma^{8\,\mathrm{TeV}}_{\mathrm{VBF}} \cdot \mathcal{B}_{H\to WW^*} &= 0.51 \quad {}^{+0.17}_{-0.15} \quad {}^{+0.13}_{-0.08} \quad = 0.51 \quad {}^{+0.22}_{-0.17} \quad \mathrm{pb}\,. \\
&\qquad\qquad (\mathrm{stat.}) \quad (\mathrm{syst.})
\end{aligned} \tag{7.11}$$

The predicted cross section values are $3.3 \pm 0.4\,\mathrm{pb}$, $4.2 \pm 0.5\,\mathrm{pb}$, and $0.35 \pm 0.02\,\mathrm{pb}$, respectively.

7.4.5.2 Fiducial cross sections

Fiducial cross section measurements enable comparisons to theoretical predictions with minimal assumptions about the kinematics of the signal and possible associated jets in the event. The cross sections described here are for events produced within a fiducial volume closely corresponding to a ggF signal region.

The measured fiducial cross section is defined as follows,

$$\begin{aligned}
\sigma_{\mathrm{fid}} &= \frac{(\mathrm{N_{sig}})_{\mathrm{obs}}}{\mathcal{C}} \cdot \frac{1}{\int L\,\mathrm{dt}} \\
&= \hat{\mu} \cdot (\sigma \cdot \mathcal{B}_{H\to WW^*\to e\nu\mu\nu})_{\mathrm{exp}} \cdot \mathcal{A},
\end{aligned} \tag{7.12}$$

with the multiplicative factor $\mathcal{A}$ being the sole difference with respect to the inclusive cross section calculation. The measured fiducial cross section is not affected by the theoretical uncertainties on the total signal yield nor by the theoretical uncertainties on the signal acceptance. The total uncertainty is reduced compared to the value for the inclusive cross section because the measured signal yield is not extrapolated to the total phase space.

The correction factors for $H{+}0j$ and $H{+}1j$ events, $\mathcal{C}^{\mathrm{ggF}}_{0j}$ and $\mathcal{C}^{\mathrm{ggF}}_{1j}$, are evaluated using the standard signal MC sample. According to the simulation, the fraction of measured signal events within the fiducial volume is 85% for $N_{\mathrm{jets}} = 0$ and 63% for $N_{\mathrm{jets}} = 1$.

The values of the correction factors are,

$$\begin{aligned}
\mathcal{C}^{\mathrm{ggF}}_{0j} &= 0.507 \pm 0.027 \\
\mathcal{C}^{\mathrm{ggF}}_{1j} &= 0.506 \pm 0.022.
\end{aligned} \tag{7.13}$$

The experimental systematic uncertainty is approximately 5%. Remaining theoretical uncertainties on the $\mathcal{C}^{\mathrm{ggF}}$ values are found to be approximately 2% and are neglected. The acceptance of the fiducial volume is,

$$\begin{aligned} \mathcal{A}_{0j}^{\mathrm{ggF}} &= 0.206 \pm 0.030 \\ \mathcal{A}_{1j}^{\mathrm{ggF}} &= 0.075 \pm 0.017\,. \end{aligned} \tag{7.14}$$

The uncertainties on the acceptance are purely theoretical in origin and the largest contributions are from the effect of the QCD scale on the jet multiplicity requirements.

The cross section values are computed by fitting the μ values in the $N_{\mathrm{jets}} \leq 1$ categories. The VBF contribution is subtracted assuming the expected yield from the SM instead of using the simultaneous fit to the VBF signal regions as is done for the inclusive cross sections. The non-negligible ggF yield in the VBF categories would require an assumption on the ggF acceptance for different jet multiplicities, whereas the fiducial cross section measurement is intended to avoid this type of assumption. The effect of the theoretical uncertainties on the VBF signal yield is included in the systematic uncertainties on the cross sections. The obtained signal strengths are,

$$\begin{aligned} \mu_{0j,e\mu}^{\mathrm{ggF}} &= 1.39 \quad \pm 0.27 \quad {}^{+0.21}_{-0.19} \quad {}^{+0.27}_{-0.17} \\ \mu_{1j,e\mu}^{\mathrm{ggF}} &= 1.14 \quad {}^{+0.42}_{-0.41} \quad {}^{+0.27}_{-0.26} \quad {}^{+0.42}_{-0.17} \\ &\qquad\quad (\mathrm{stat.}) \quad (\mathrm{syst.}) \quad (\mathrm{sig.}) \end{aligned} \tag{7.15}$$

where (sig.) indicates the systematic uncertainties on the signal yield and acceptance, which do not apply to the fiducial cross section measurements. The corresponding cross sections, evaluated at $m_{\mathrm{H}} = 125.36\,\mathrm{GeV}$ and using the 8 TeV data, are,

$$\begin{aligned} \sigma_{\mathrm{fid},0j}^{\mathrm{ggF}} &= 27.6 \quad {}^{+5.4}_{-5.3} \quad {}^{+4.1}_{-3.9} &&= 27.6 \quad {}^{+6.8}_{-6.6} \ \mathrm{fb} \\ \sigma_{\mathrm{fid},1j}^{\mathrm{ggF}} &= 8.3 \quad {}^{+3.1}_{-3.0} \quad {}^{+3.1}_{-3.0} &&= 8.3 \quad {}^{+3.7}_{-3.5} \ \mathrm{fb}\,. \\ &\qquad (\mathrm{stat.}) \quad (\mathrm{syst.}) \end{aligned} \tag{7.16}$$

The predicted values are $19.9 \pm 3.3\,\mathrm{fb}$ and $7.3 \pm 1.8\,\mathrm{fb}$, respectively.

7.5 Conclusions

The results summarised in this chapter report the observation of Higgs boson decays to WW^* based on an excess over background of 6.1 σ in the dilepton final state, for which the SM expectation is 5.8 σ. Evidence of the VBF production mode is

obtained with a significance of 3.2 σ. The results are obtained from a data sample corresponding to an integrated luminosity of 25 fb^{-1} from $\sqrt{s} = 7$ and 8 TeV proton-proton collisions recorded by the ATLAS detector at the LHC. For a Higgs boson mass of 125.36 GeV, the ratio of the measured value to the expected value of the total production cross section times branching ratio is $1.09\,^{+0.16}_{-0.15}\,(\text{stat.})\,^{+0.17}_{-0.14}(\text{syst.})$. The corresponding signal strengths for the ggF and VBF production mechanisms are $1.02 \pm 0.19\,(\text{stat.})\,^{+0.22}_{-0.18}(\text{syst.})$ and $1.27\,^{+0.44}_{-0.40}\,(\text{stat.})\,^{+0.21}_{-0.30}(\text{syst.})$, respectively.

Chapter 8

Conclusions

This thesis presents a wide range of studies that go from energy reconstruction at the cell level in TileCal performance reconstruction of the $\not\!\!E_T$ measurement in ATLAS, and finishing with the discovery of the Higgs boson decaying into a pair of W bosons.

The reconstruction of the energy, as well as the time, in the TileCal cells is provided by the OF algorithm. These measurements are the inputs for object reconstruction algorithms which base their logic on the signal over noise ratio. In this light, the determination of the cell noise is crucial for an accurate event reconstruction in ATLAS. The impact of the TileCal noise constants has been evaluated through the performance of topoclusters since they clearly allow to identify anomalies in the TileCal noise. The results show the evidence of a coherent source of noise which is not properly described by the One-Gaussian approach used so far. This effect produces larger and wider structures for topoclusters in the first ATLAS data collected during 2008 and 2009 compared with the expectations from MC. The results motivated a new description of the TileCal noise constants using a Two-Gaussian method instead. The improvement using the Two-Gaussian description reduces the number of large topoclusters by a factor $\sim$10 in randomly triggered ATLAS events. The better description of the noise reduces the discrepancies between data and simulation as well. During these investigations, there were also found extremely energetic areas in TileCal. These hot spots are mainly originated by cells, probably affected by electronic damage, which poorly reconstruct the energy of the genuine signal. Bad cells producing hot spots were identified and properly treated at the detector operation level in order to ensure the quality of TileCal reconstruction. In addition to the improvements on topoclusters, the Two-Gaussian method also leads to a better missing transverse momentum measurement. The RMS of the

spectrum in data is reduced by a factor 2 and the contribution on the tails highly decreases from 16% to 0.1%. All these results confirm and validate the Two-Gaussian description of the TileCal noise which has been used for the whole Run I ATLAS data reconstruction.

The $\not\!\!E_T$ measurement relies on the momentum conservation law in the transverse plane to the beam axis. For a specific process, this measures the unbalanced transverse momentum from all the particles in the final state, so it is sensitive to the presence of undetectable particles, such as neutrinos. The results on the performance for the different $\not\!\!E_T$ reconstructions developed in ATLAS are crucial as this measurement plays an important role in many analysis searches. The energetic measurements from the particles produced in the LHC collisions are taken from the ATLAS calorimetric system in the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement. The $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement depends on the number of pile-up interactions since their final products may also deposit energy in the calorimeters. These extra energetic contributions are included in the $E_{\mathrm{T}}^{\mathrm{miss}}$ computation degrading the genuine measurement. The increasing pile-up environment at the LHC during 2012 motivated investigations on new approaches for improving the $\not\!\!E_T$ reconstruction in ATLAS. Two pile-up suppressed alternatives based on track information and vertex association were developed: $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track}}$, respectively. The former follows the calorimeter-based approach of the $E_{\mathrm{T}}^{\mathrm{miss}}$ but scaling down the soft term and rejecting pile-up jets. The latter relies on the energy measured from well reconstructed tracks in the inner detector which are associated to the primary vertex of the event. The performance of the several approaches based on the object information to build the unbalanced transverse momentum have been evaluated in terms of resolution, scale and linearity. Results show that besides the better stability against pile-up of the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ reconstructions, these approaches come with new features. For the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$, the poor modelling of tracks coming from pile-up interactions produces an under-calibrated soft term in MC. This results in discrepancies between data and simulation, specially in events without jets. In events with jets, the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ perform very similar because the dominant component is the jet term. The $E_{\mathrm{T}}^{\mathrm{miss,track}}$ measurement is very robust against extra interactions since only tracks associated to the vertex of the hardest process are included. However, limited ID coverage and missing high-p_{T} neutral particles lead to large degradation in the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ linearity and scale, specially in event topologies with high jet activity. Due to the variety on $\not\!\!E_T$ reconstructions and their behaviours depending on the event topology, the optimal measurement may be different based on the characteristics of the physics process to study.

The second part of the thesis describes the strategy of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis and reports the results using the complete ATLAS Run I data. This corresponds to about $25\,\text{fb}^{-1}$ at $\sqrt{s} = 7$ and $8\,\text{TeV}$ collected with the ATLAS detector at the LHC. The Higgs boson decaying into a pair of W bosons benefits from a larger BR compared with other final states for a wide range of the Higgs boson mass. This makes the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis one of the most important channels for the Higgs boson search. However, this analysis suffers from high background contamination, which difficulties the distinction between the Higgs boson signal and other processes that may have the same reconstructed final state. In addition, the analysis is not sensitive to the Higgs boson mass due to the presence of the two neutrinos coming from the W bosons. These two facts define the strategy of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis. The selection criteria should find an optimal compromise to be hard enough for rejecting as many background contributions as possible and, at the same time, soft enough for still keeping the Higgs boson signal. The analysis selects events with exactly two high-p_{T} well reconstructed leptons (electrons or muons) oppositely charged and with $\not{E}_T$ measurement originated by the final neutrinos. In order to deal with different background contributions the events are divided by the number of jets as well as by the flavour of the two leptons. This separation allows to adequate the selection since the background composition is different in each category. In general, final states with same flavour leptons are mostly populated by Z/γ^* background while events with different flavoured leptons are mainly originated by top quark processes. For the former, the analysis applies a combined requirement using several $\not{E}_T$ reconstructions in order to further suppress Z/γ^* contributions, for which non genuine $\not{E}_T$ measurement is expected. The latter vetoes jets which are considered as produced by a b quark from reconstruction algorithms. In addition, the division on the jet multiplicity also allows to distinguish Higgs candidates as originated by gluon-gluon fusion (with zero or up to one jet) or vector boson fusion (at least two jets) production mechanisms. This distinction leads to better separate Higgs signal from the remaining backgrounds in each case by exploiting the differences in dilepton kinematics and, when relevant, in jet based magnitudes. After all selection is applied, the transverse mass of the dilepton system and the $\not{E}_T$ of the Higgs candidate events is used as final discriminants in a statistical test. Given the importance of simulating all background processes correctly, the analysis builds different control regions to check the agreement between data and MC. The differences in the control regions are also inputs to the statistical procedure. This ensures that the likelihood fit includes them properly as associated uncertainties in the final results. The first results of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis showed an excess of events

over the expected background observed for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ with a signal significance of 3.8 σ, for which the expectation is 3.7 σ. The best fit signal strength at that mass is $\mu = 1.01 \pm 0.31$. The expected VBF signal significance at $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ is 1.6 σ and the observation results in 2.5 σ. The first $H \to WW^{(*)} \to \ell\nu\ell\nu$ results are consistent with the measurements from the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ searches. All ATLAS measurements from Higgs decaying into boson pair searches are combined allowing to observe an excess over the expectation with a local significance of 5 σ.

After these first results several studies were focussing on optimising the selection of the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis in order to enhance the sensitivity of the search. The final optimised results mainly benefit from the development of a new $\not\!\!E_T$ reconstruction, represented by the symbols $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ or $p_{\mathrm{T}}^{\mathrm{miss}}$. This new reconstruction is based on the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ approach but replacing tracks by the calorimetric measurements of the objects associated to them and adding jets which are missing in the original $E_{\mathrm{T}}^{\mathrm{miss,track}}$ computation. Although this may create a higher dependence with pile-up, the new approach still profits from pile-up rejection from the original track-based selection and with a much more accurate measurement in topologies with neutral particles in the final state. The results show that the $p_{\mathrm{T}}^{\mathrm{miss}}$ is able to recover the resolution in events with jets while still maintains a good stability with pile-up and smaller tails in $Z \to \ell\ell$ process. Additional investigations using event topologies with genuine $\not\!\!E_T$ also point to a more reliable measurements of the expected $\not\!\!E_T$ when using the $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. The strategy for optimising the $\not\!\!E_T$ criteria in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis is based on simulated final candidate events evaluated through the statistical likelihood fit. Given the composition and contribution of the different backgrounds depend on the final state, the $\not\!\!E_T$ optimisation is evaluated in each analysis category. The different $\not\!\!E_T$ measurements perform very similarly at the end of the event selection for $e\mu$+μe final states. The low region of the spectrums are almost not populated since the main backgrounds, as well as the Higgs boson signal, are expected to have genuine $\not\!\!E_T$. In addition, the analysis requirements on $m_{\ell\ell}$ and $p_{\mathrm{T}}^{\ell\ell}$, which are correlated with the $\not\!\!E_T$ measurement, sculpt the $\not\!\!E_T$ shapes at the end of the selection. Hence, there are almost no differences in the expected significance values using any of the $\not\!\!E_T$ reconstructions. However, the $p_{\mathrm{T}}^{\mathrm{miss}}$ is preferred because of its better performance and resolution. A conservative threshold of 20 GeV is used to deal with possible mis-measurements from multi-jets background in H+0j and H+1j analyses. Since the Higgs boson produced via VBF is typically characterised by two emerging quarks, the $\not\!\!E_T$ measurement for the Higgs signal is expected to be smaller than in the ggF production mode. In this light, the VBF strategy does not apply any threshold on the $\not\!\!E_T$ measurement since

the low region of the spectrum is mainly populated by signal events. Final states with same flavoured leptons are affected by a huge Z/γ^* contribution, so combining several $\not\!\!E_T$ reconstructions achieves further Z/γ^* rejection. In this case when the final state contains up to one jet, the requirement is done using the projected $\not\!\!E_{T,Rel}$ magnitude for the $E_{\mathrm{T}}^{\mathrm{miss}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ measurements. Investigations on the direction of the new $p_{\mathrm{T}}^{\mathrm{miss}}$ conclude that the rejection power of the original $E_{\mathrm{T}}^{\mathrm{miss,track}}$ is still higher. This is due to the fact that the latter tends to point to the mismeasured jets, hence the $\not\!\!E_{T,Rel}$ computation using $E_{\mathrm{T}}^{\mathrm{miss,track}}$ benefits to highly reject the Z/γ^* contribution. $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ computed with the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ still provides the best significance. For the VBF-enriched analysis, however, the $\not\!\!E_{T,Rel}$ magnitude may be biased because of the probability to randomly project the nominal measurement to any reconstructed jet. This points back to the usage of the $p_{\mathrm{T}}^{\mathrm{miss}}$ measurement, complemented with a purely calorimeter-based $E_{\mathrm{T}}^{\mathrm{miss}}$ threshold. The better $p_{\mathrm{T}}^{\mathrm{miss}}$ performance can be exploited to also benefit other $\not\!\!E_T$-dependent quantities used in the $H\to WW^{(*)}\to\ell\nu\ell\nu$, as the m_{T}. Results show a better resolution of the m_{T} measurement obtained by using the $p_{\mathrm{T}}^{\mathrm{miss}}$ in the computation. The usage of the $p_{\mathrm{T}}^{\mathrm{miss}}$ in the m_{T} leads to a better separation between the Higgs signal and the remaining backgrounds, specially for multi-jet and non-WW diboson processes. These optimal thresholds using the $p_{\mathrm{T}}^{\mathrm{miss}}$ measurement increase the expected significance by 7% in the ggF-enriched analysis of the $H\to WW^{(*)}\to\ell\nu\ell\nu$ search. The introduction of the $p_{\mathrm{T}}^{\mathrm{miss}}$ for the transverse mass computation enhances the expected significance by 9%. For the VBF-enriched search, the overall improvement due to the optimised $\not\!\!E_T$ selection is observed up to 14% in the expected significance results.

Finally, the optimisation of the $H\to WW^{(*)}\to\ell\nu\ell\nu$ analysis has been developed using the complete 8 TeV data sample. The main improvements rely on the introduction of more performant variables as the $p_{\mathrm{T}}^{\mathrm{miss}}$, new techniques for background estimation, and extensions of the Higgs signal phase space to enhance the sensitivity of the search. After the whole optimisation, the expected significance of the ggF production mode increases from 2.8 σ to 4.36 σ just in $e\mu$+μe final states. For the VBF production mode of the Higgs boson the overall gain is up to 70% due to the BDT technique applied now for the H+$2j$ category. The last $H\to WW^{(*)}\to\ell\nu\ell\nu$ results using Run I ATLAS data are reported at $m_{\mathrm{H}} = 125.36\,\mathrm{GeV}$. There is an excess over background of 6.1 σ observed for the $H\to WW^{(*)}\to\ell\nu\ell\nu$ analysis for which the SM expectation is 5.8 σ. Evidence of the VBF production mode is also obtained with a significance of 3.2 σ. The ratio of the measured value to the expected value of the total production cross section times branching ratio is $1.09\,^{+0.16}_{-0.15}\,(\mathrm{stat.})\,^{+0.17}_{-0.14}(\mathrm{syst.})$. The corresponding signal strengths for the ggF and VBF production mechanisms

are $1.02 \pm 0.19\,(\text{stat.})\,^{+0.22}_{-0.18}(\text{syst.})$ and $1.27\,^{+0.44}_{-0.40}\,(\text{stat.})\,^{+0.21}_{-0.30}(\text{syst.})$, respectively. The measurements are consistent with SM Higgs boson expectations and state the observation of the Higgs boson decaying to WW^* in ATLAS.

Capítulo 9

Resumen

9.1 El Experimento ATLAS en el LHC del CERN

9.1.1 El Gran Colisionador de Hadrones del CERN

El CERN (siglas en francés de *Conseil Européen pour la Recherche Nucléaire*) tiene como objetivo proporcionar aceleradores de partículas y otras infraestructuras necesarias para la investigación de la Física de Altas Energías. Desde su fundación en 1954, el CERN ha albergado varios experimentos como LEP [15], con los que se han obtenido resultados de alto impacto en la comunidad científica. Por ello, el CERN se ha consolidado como el centro pionero y de referencia en la Física de Partículas a nivel mundial. Hoy en día, el CERN está formado por 21 estados miembros y más de 600 institutos y universidades de todo el mundo participan en sus proyectos de investigación de distintas formas.

Actualmente, el CERN está sumergido en el programa de investigación del Gran Colisionador de Hadrones, LHC [32] (siglas en inglés de *Large Hadron Collider*). El LHC es un acelerador de partículas de 27 km de circunferencia que está construido en un túnel a 100 m de profundidad entre las fronteras de Francia y Suiza, cerca de la ciudad de Ginebra. En el LHC circulan dos haces de partículas en sentidos opuestos, los cuales se aceleran hasta la energía de 4 TeV. Cuando los haces alcanzan dicha energía, se hacen colisionar en el centro de los cuatro experimentos situados en el túnel del LHC: ATLAS [27], CMS [28], LHCb [29] y ALICE [30]. La Fig. 9.1 muestra la localización del CERN en la superficie y el perímetro del túnel del LHC en el subsuelo, así como la ubicación de los cuatro experimentos situados en el anillo del acelerador. La función de los detectores situados en el túnel del LHC es la de

medir las partículas producidas en las colisiones con el fin de excluir o confirmar distintas hipótesis teóricas.

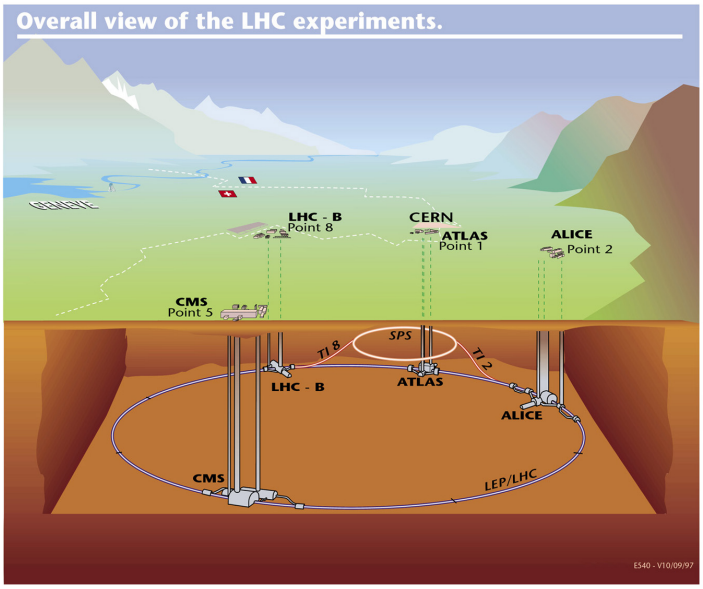

FIGURA 9.1: Diagrama de las ubicaciones de los principales experimentos situados en el túnel del LHC: ATLAS, CMS, ALICE y LHCb.

9.1.2 El Experimento ATLAS

El objetivo general del experimento ATLAS (siglas en inglés de *A Toroidal LHC Apparatus*) es investigar el amplio rango de procesos físicos que teóricamente podrían producirse en las colisiones a altas energías que tienen lugar en el LHC. Ello incluye confirmaciones o medidas mejoradas del Modelo Estándar, SM (siglas en inglés de *Standar Model*), de partículas elementales, así como descubrir nuevos e inesperados fenómenos al extender el rango energético de observación. Uno de los logros más importantes del experimento ATLAS fue el descubrimiento del bosón de Higgs [68], internacionalmente anunciado en Julio de 2012.

ATLAS es el detector más grande situado en el túnel del LHC. Tiene 45 m de largo, más de 25 m de alto y pesa unas 7000 t. En el centro del detector ocurren las colisiones de los haces del LHC, las cuales producen partículas que emergen del punto de interacción en todas direcciones. El detector ATLAS debe medir la trayectoria y energía de dichas partículas teniendo en cuenta que distintas partículas interaccionan con la materia de forma diferente. Por este motivo, el detector de ATLAS

está construido usando una serie de componentes especializados en la detección de un tipo concreto de partículas. El detector ATLAS está formado por el Detector Interno (ID, siglas en inglés de *Inner Detector*), de los calorímetros electromagnético y hadrónico (LAr y TileCal, respectivamente) y del Espectrómetro de Muones (MS, siglas en inglés de *Muon Spectrometer*). En la Fig. 9.2 se muestra una representación del detector de ATLAS y de sus principales componentes.

El ID está situado en la parte más interna del detector y más cercana a los haces del LHC. Este detector está diseñado para reconstruir los vértices de desintegración, así como la trayectoria y momento transverso de las partículas cargadas que emergen de ellos. El sistema de calorímetros envuelve el ID y en él se deposita la energía de las partículas que lo atraviesan para las cuales estos detectores están construidos. La última capa de ATLAS está ocupada por el MS que se encarga de medir las partículas cargadas que logran atravesar los calorímetros, como los muones.

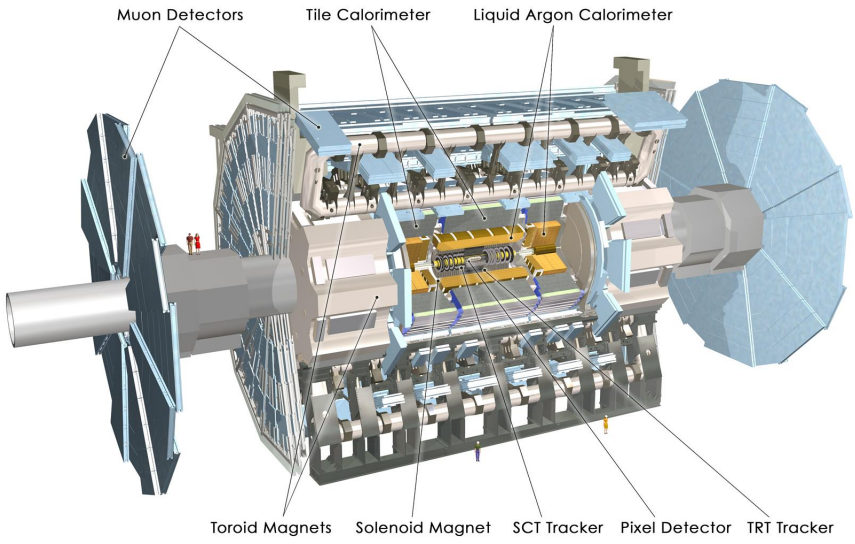

FIGURA 9.2: El detector de ATLAS y sus principales componentes.

9.1.3 El Calorímetro Hadrónico TileCal del Detector ATLAS.

Un esquema con todos los calorímetros del detector de ATLAS se muestra en la Fig. 9.3. La parte principal del sistema hadrónico de ATLAS es el detector TileCal [42]. El TileCal está diseñado para determinar la energía y dirección de las

cascadas hadrónicas, así como contribuir a la medida de la Energía Transversa Perdida (MET, siglas en inglés de *Missing Transverse Energy*).

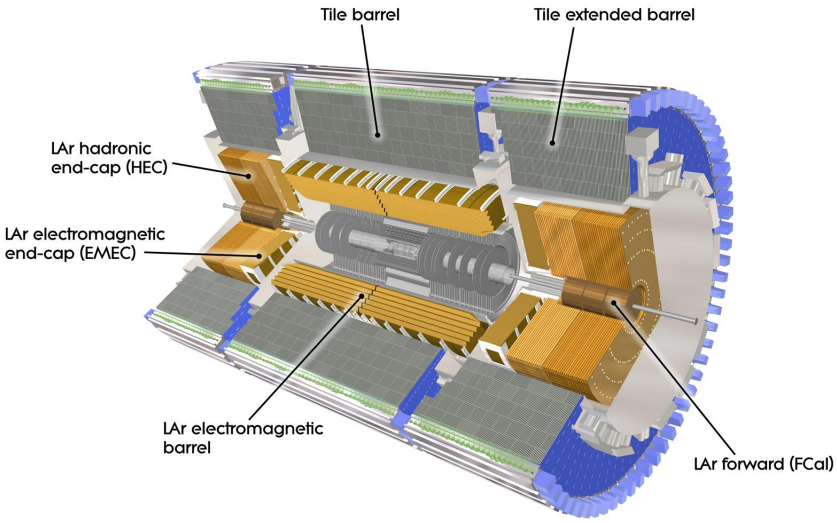

FIGURA 9.3: Calorímetros electromagnético y hadrónico del detector ATLAS.

El TileCal está dividido en cuatro particiones, dos en el gran barril central y otras dos en los barriles extendidos. Los barriles de TileCal están construidos por celdas de material centelleador, usado como material activo, e hierro, usado como material absorbente. Cada uno de los barriles de TileCal está formado por 64 módulos en los cuales el material centelleador está situado radialmente al eje que definen los haces de protones. En la Fig. 9.4 se muestra la estructura y componentes de un modulo del TileCal. La partículas que atraviesan el TileCal interaccionan con el material activo produciendo luz en el material centelleador. Esta luz es recogida por fibras ópticas y conducida hacia los fotomultplicadores los cuales amplifican y digitalizan la señal cada 25 ns, obteniendo siete muestras. Estas muestras son usadas para reconstruir la amplitud, tiempo y pedestal del pulso usando el algoritmo de Filtrado Óptimo (OF, siglas del inglés *Optimal Filtering* [48]).

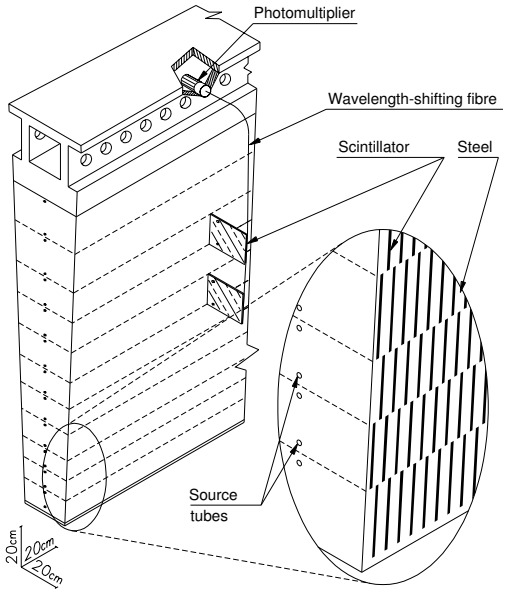

FIGURA 9.4: Estructura y componentes de uno de los módulos que forman el calorímetro hadrónico TileCal.

9.2 Descripción del Ruido en el Calorímetro TileCal

9.2.1 El Método Optimal Filtering y el Algoritmo de Topoclusters

Una buena medida de la energía depositada por las partículas en el TileCal es esencial para la reconstrucción de los procesos producidos en las colisiones del LHC. Los pulsos producidos por la electrónica del TileCal pueden ser caracterizados por tres parámetros, la amplitud (A), la fase (τ) y el pedestal (p). El procedimiento del algoritmo OF para reconstruir dichas magnitudes viene dado por las siguientes ecuaciones:

$$A = \sum_{i=1}^{7} a_i \, S_i \, , \tag{9.1}$$

$$\tau = \frac{1}{A} \sum_{i=1}^{7} b_i \, S_i \, , \tag{9.2}$$

$$p = \frac{1}{A} \sum_{i=1}^{7} c_i \, S_i \, , \tag{9.3}$$

donde S_i representa la muestra digital número i; y (a_i, b_i, c_i) son los pesos obtenidos para para una fase de referencia τ_0.

Las medidas de la energía proporcionadas por las celdas del TileCal son usadas por el algoritmo de Agrupación Topológica (en inglés *Topological Clustering*) [53]. Este algoritmo está diseñado para agrupar celdas que tienen una señal suficientemente elevada como para no ser producida por una fluctuación del ruido electrónico:

$$s = \frac{|E|}{\sigma} \tag{9.4}$$

donde t es el umbral definido para considerar la celda en el algoritmo de Agrupación Topológica, E es la energía medida por las celdas del TileCal y σ es el ruido electrónico asociado a dicha celda. Celdas que cumplen $s > s_s$ son consideradas semillas. A las celdas semillas se les añaden sus celdas colindantes siempre y cuando $s > s_n$. Éstas son consideradas celdas vecinas. Por último, todas las celdas colindantes a las celdas vecinas también son añadidas si $s > s_c$. Con las distintas condiciones en s para la definición de celdas semillas y celdas vecinas se consigue crear unas nuevas estructuras conocidas como topoclusters. Figura 9.5 muestra la multiplicidad de los topoclusters formados con distintos valores de s_s, s_n y s_c usando eventos tomados por el detector ATLAS durante 2008.

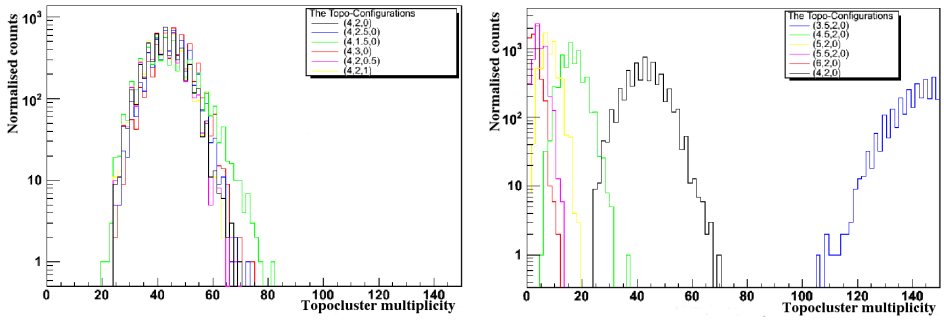

FIGURA 9.5: Multiplicidad de topoclusters para diferentes configuraciones de (s_s,s_n,s_c) obtenidas usando eventos tomados por el detector ATLAS durante 2008. Las distribuciones se han normalizado al número de topoclusters obtenidos con la configuración nominal (4,2,0).

Los topoclusters son las estructuras a partir de las cuales los algoritmos de reconstrucción de ATLAS, junto con información adicional de otras partes del detector, se reconstruyen las partículas producidas en las colisiones del LHC: muones y jets; así como la medida de la $\not{E}_T$. Por lo tanto, el estudio de evaluación de la reconstrucción de energía en TileCal, así como la descripción del ruido electrónico de

este detector son fundamentales para reconstruir correctamente los procesos físicos que se miden en ATLAS. Las investigaciones sobre diferentes parámetros espaciales de los topoclusters permiten evaluar la reconstrucción de la energía y la descripción del ruido en el TileCal. Típicamente estos parámetros están relacionados con la distancia radial (r) y longitudinal (λ) con respecto al eje y el centro de la cascada hadrónica, respectivamente.

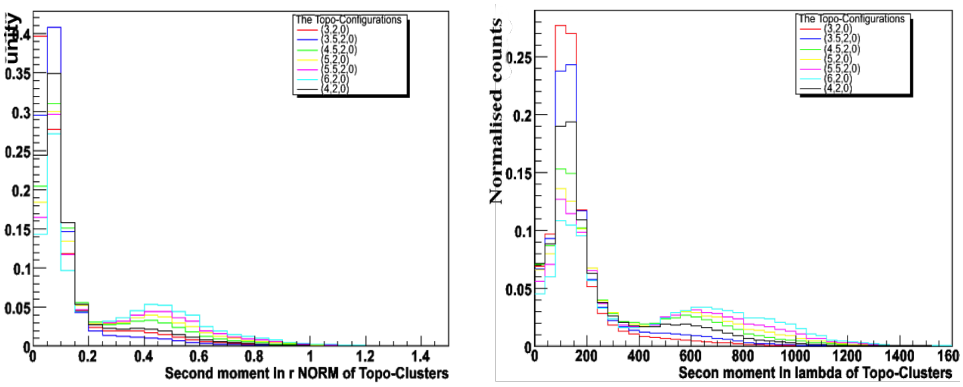

FIGURA 9.6: Momentos en r (izquierda) y λ (derecha) de topoclusters obtenidos usando distintas configuraciones de (s_s, s_n, s_c). Los datos corresponden a eventos tomados por el detector ATLAS durante 2008. Las distribuciones se han normalizado a la unidad para facilitar su comparación.

Figura 9.6 muestra los momentos en r y λ de los topoclusters obtenidos con distintos valores de s_s, s_n y s_c usando eventos recogidos por ATLAS durante 2008. Claramente existen dos contribuciones de ruido debido a los picos observados en las distribuciones de momentos r y λ de los topoclusters estudiados. Los resultados además muestran que la contribución que forma los topoclusters mayores aumenta con s_s. Los topoclusters obtenidos usando datos de ATLAS son comparados con la simulación del proceso $W \rightarrow \ell\nu$ en la Fig. 9.7. Las distribuciones muestran los momentos en r y λ de topoclusters formados con la configuración nominal de ATLAS (4,2,0) en función de la energía de los topoclusters (definida como la suma de energías de todas las celdas que forman el topocluster).

Las distribuciones señalan que topoclusters con mayores momentos en r y λ sólo se observan en datos y se concentran en regímenes de energía pequeños. Todos los resultados evidencian la existencia de ruido coherente en el TileCal y motivaron una mejora en la descripción del ruido del detector durante 2009.

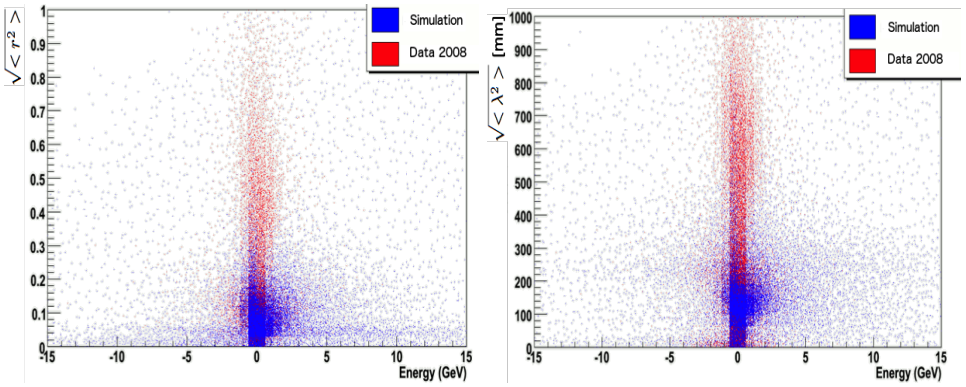

FIGURA 9.7: Segundos momentos en r (derecha) y λ (izquierda) como función de la energía de los topoclusters formados con la configuración (4,2,0) en datos tomados en ATLAS (rojo) y en la simulación del proceso $W \rightarrow \ell\nu$ en MC (azul).

9.2.2 Método de la Doble Gaussiana para la Descripción del Ruido en TileCal

Típicamente, el ruido electrónico de un detector es aleatorio, por lo que puede ser descrito por una distribución Gaussiana. Sin embargo, la distribución de la energía reconstruida por las celdas de TileCal presenta colas que no son descritas por una función Gaussiana [57]. Por lo tanto, la medida del ruido en una celda puede ser considerada como una combinación lineal de un término intrínseco de ruido aleatorio y un término que incluye efectos de correlación entre celdas, lo cual crea la componente de ruido coherente. Figura 9.8 ilustra el método de la doble Gaussiana desarrollado para la descripción del ruido de las celdas de TileCal.

Figura 9.9 compara el momento en λ de los topoclusters reconstruidos en TileCal usando una (en azul y negro) y dos (en rojo) Gaussianas para la descripción del ruido. Los eventos usados son aleatorios y fueron tomados por el detector ATLAS durante 2008 y 2009. La contribución de topoclusters es reducida un factor 10 debido a la mejora de la descripción del ruido obtenida con el método de la Doble Gaussiana. Los resultados también muestran una mejora en la comparación de los datos con la simulación de MC. Adicionalmente, estudios de la energía transversa perdida usando topoclusters (MET) producidos con ambas descripciones de ruido electrónico muestran hasta un factor 2 de reducción en el RMS de dichas distribuciones usando la descripción de la Doble Gaussiana. Por último, la contribución de las colas de los espectros de MET es fuertemente suprimida pasando de 16 % a 0.1 %.

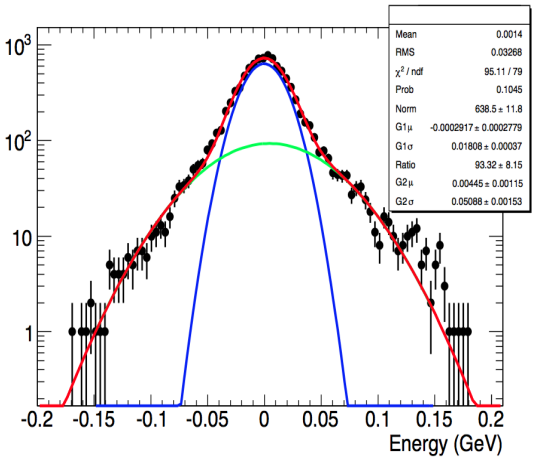

FIGURA 9.8: Reconstrucción de la energía de una celda típica de ATLAS usando eventos aleatorios tomados por ATLAS durante 2008 [57]. El método de la Doble Gaussiana es mostrado en rojo, En azul y verde las funciones por separado de la primera y segunda componentes del ruido, respectivamente.

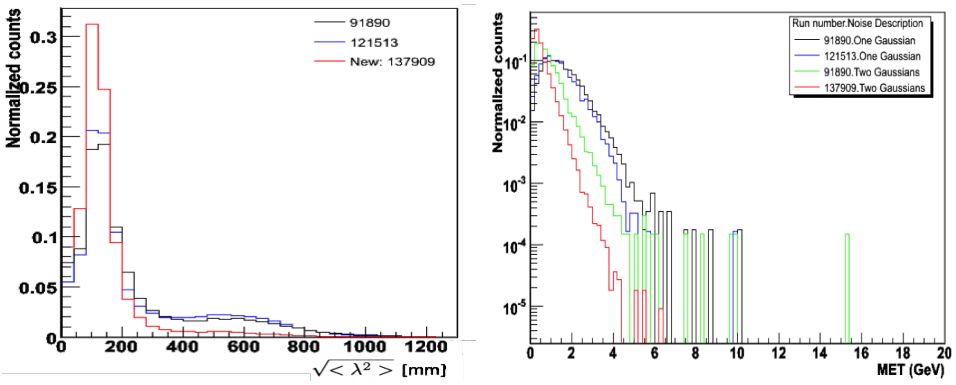

FIGURA 9.9: En la izquierda, momento en λ para distintas descripciones del ruido de TileCal usando datos tomados en ATLAS durante 2008 y 2009. En la derecha, espectros de la energía transversa perdida (MET) a partir de topoclusters obtenidos con distintos métodos para describir el ruido en TileCal. La descripción de la Doble Gaussiana (en rojo) reduce la contribución de topoclusters con $\lambda > 400$ con respecto a la descripción usando una única Gaussiana (azul y negro). Además la Doble Gaussiana (roja y verde) reduce la anchura y las colas en la distribución de MET comparado con la descripción del ruido dado por una única Gaussiana (azul y negro) Las distribuciones han sido normalizadas a la unidad para facilitar su comparación.

9.3 La Energía Transversa Perdida en ATLAS

Los métodos de reconstrucción de la energía transversa perdida ($\not\!\!E_T$, en inglés *Missing Transverse Energy*) se basan en la ley de conservación de momento en el plano transversal al eje de los haces del LHC. Esta magnitud mide el desequilibrio en el momento transverso a partir de todas las partículas obtenidas en el estado final de un determinado proceso. Debido a esto, la medida de la $\not\!\!E_T$ es sensible a la presencia de partículas que atraviesan el detector ATLAS sin depositar energía en él, como los neutrinos. Las investigaciones sobre la medida de la $\not\!\!E_T$ en ATLAS son esenciales dado que esta medida juega un papel importante en muchos análisis.

9.3.1 Energía Transversa Perdida: Definiciones

En ATLAS se han desarrollado dos métodos para calcular la $\not\!\!E_T$ los cuales se diferencian en la parte del detector que proporciona la información de las partículas producidas en un evento. Las conocidas como medidas de $\not\!\!E_T$ calorimétricas mayormente se basan en la reconstrucción de energía dada por los calorímetros electromagnético y hadrónico, y el espectrómetro de muones de ATLAS. Así pues, las deposiciones de energía son asociadas a los objetos en el siguiente orden: electrones (e), fotones (γ), taus (τ), jets y muones (μ). Las deposiciones de energía que no están asociadas a ningún objeto, también son consideradas en el cálculo de la $\not\!\!E_T$ a través del término $E_{\mathrm{T}}^{\mathrm{miss,\ SoftTerm}}$. Todas estas contribuciones son introducidas en el cálculo de las componentes de $E_{\mathrm{T}}^{\mathrm{miss}}$ en las direcciones x e y:

$$E_{x(y)}^{\mathrm{miss}} = E_{x(y)}^{\mathrm{miss},\,e} + E_{x(y)}^{\mathrm{miss},\,\gamma} + E_{x(y)}^{\mathrm{miss},\,\tau} + E_{x(y)}^{\mathrm{miss,\,jets}} + E_{x(y)}^{\mathrm{miss,\,SoftTerm}} + E_{x(y)}^{\mathrm{miss},\,\mu}\,, \quad (9.5)$$

donde cada término está calculado como la suma negativa de las energía de los objetos proyectadas en las direcciones x e y.

La definición de $E_{\mathrm{T}}^{\mathrm{miss}}$ depende del número de colisiones simultáneas que pueden producirse en el mismo evento. Éstas se conocen como interacciones de pile-up. Las interacciones de pile-up producen partículas que pueden depositar su energía en el detector, y por tanto, entran en el cálculo de la medida de $E_{\mathrm{T}}^{\mathrm{miss}}$. El incremento de interacciones de pile-up durante 2012 en el LHC hizo necesarias nuevas investigaciones sobre alternativos métodos de medida de la $\not\!\!E_T$ que mitigaran el efecto producido por interacciones extra. Dos nuevas medidas se desarrollaron, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ y $E_{\mathrm{T}}^{\mathrm{miss,track}}$, las cuales suprimen las contribuciones debidas al pile-up a través de la definición del vértice primario (PV, del inglés *primary vertex*). En ATLAS, el PV

se define como aquel vértice cuya suma del momento de las trazas asociadas a él es mayor.

La definición de $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ se basa en las deposiciones de energía en los calorímetros como la $E_{\mathrm{T}}^{\mathrm{miss}}$ pero escala los términos $E_{x(y)}^{\mathrm{miss,\,SoftTerm}}$ en la Eq. 9.5 por el factor STVF (del inglés, *soft term vertex-fraction*). Esta corrección se define como la fracción de la suma escalar del momento de las trazas asociadas al PV con respecto a la suma escalar del momento de todas las trazas medidas en el ID, incluyendo aquéllas producidas por interacciones de pile-up:

$$\mathrm{STVF} = \frac{\sum\limits_{\mathrm{SoftTerm}}^{\mathrm{PV}} p_{\mathrm{T}}^{\mathrm{track}}}{\sum\limits_{\mathrm{SoftTerm}}^{\mathrm{All}} p_{\mathrm{T}}^{\mathrm{track}}}\,, \tag{9.6}$$

donde $0 \leq \mathrm{STVF} \leq 1$. Además, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ también suprime las contribuciones de jets que son producidos por interacciones pile-up en el término $E_{x(y)}^{\mathrm{miss,\,jets}}$:

$$E_{x(y)}^{\mathrm{miss,\,STVF}} = E_{x(y)}^{\mathrm{miss,\,e}} + E_{x(y)}^{\mathrm{miss,}\,\gamma} + E_{x(y)}^{\mathrm{miss,}\,\tau} + \mathbf{E}_{\mathbf{x(y),\,JVF}}^{\mathbf{miss,\,jets}} + \mathbf{E}_{\mathbf{x(y),\,STVF}}^{\mathbf{miss,\,SoftTerm}} + E_{x(y)}^{\mathrm{miss,}\,\mu}\,, \tag{9.7}$$

donde las contribuciones corregidas con respecto a $E_{\mathrm{T}}^{\mathrm{miss}}$ se han resaltado.

Por último, la definición de $E_{\mathrm{T}}^{\mathrm{miss,track}}$ usa la energía medida a partir de las trazas reconstruidas en el detector interno de ATLAS:

$$E_{x(y)}^{\mathrm{miss,\,track}} = -\sum_{\mathrm{tracks}}^{\mathrm{PV}} p_{x(y)}^{\mathrm{track}}\,, \tag{9.8}$$

donde la suma se extiende únicamente a las trazas asociadas al PV, por lo cual la medida de $E_{\mathrm{T}}^{\mathrm{miss,track}}$ es independiente del número de interacciones pile-up. Sin embargo, la medida de $E_{\mathrm{T}}^{\mathrm{miss,track}}$ también tiene inconvenientes. La limitación espacial del ID y la presencia de partículas neutras de alto momento transverso hacen que la reconstrucción dada por la Eq. 9.8 sea deficiente en procesos con jets en el estado final.

9.3.2 Energía Transversa Perdida: Estudios y Resultados

Los resultados más importantes obtenidos comparando las distintas definiciones de $\not\!\!E_T$ desarrolladas en ATLAS evaluadas en los procesos $Z \to \ell\ell$ y $H \to WW^{(*)} \to \ell\nu\ell\nu$ son presentados y discutidos a continuación.

Figura 9.10 muestra los espectros de las distintas definiciones de $\not\!\!E_T$ obtenidas con eventos consistentes con el proceso $Z \to \ell\ell$ en datos tomados con el detector ATLAS a una energía del centro de masa de 8 TeV y datos simulados en Monte Carlo (MC). Para eventos con jets, la medida de $E_{\mathrm{T}}^{\mathrm{miss,track}}$ presenta mayores colas que los espectros obtenidos con las medidas $E_{\mathrm{T}}^{\mathrm{miss}}$ y $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$, las cuales tienen formas muy parecidas. Esto es debido a que la medida de $E_{\mathrm{T}}^{\mathrm{miss,track}}$ es deficiente en procesos con partículas neutras en el estado final debido a las limitaciones del ID. Sin embargo, en eventos sin jets la $E_{\mathrm{T}}^{\mathrm{miss,track}}$ es la distribución con menores valores medio y anchura.

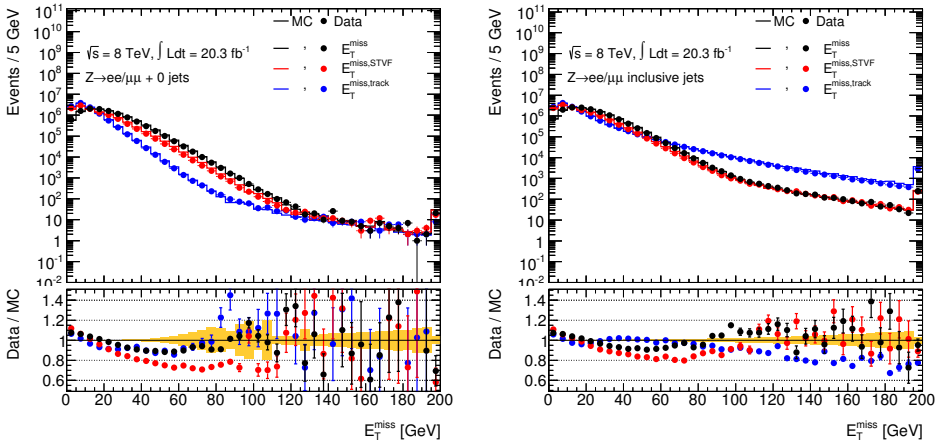

FIGURA 9.10: Distribuciones de las distintas medidas de $\not\!\!E_T$ en datos (puntos) y simulación (líneas) usando eventos consistentes con el proceso $Z \to \ell\ell$. Abajo se muestra el cociente del número de eventos en datos sobre el de la simulación para cada definición de $\not\!\!E_T$. La distribución de la derecha selecciona eventos que no contengan jets con $p_{\mathrm{T}} > 20$ GeV mientras que en la izquierda no se aplica ninguna restricción para el número de jets.

Para evaluar la dependencia de las medidas de $\not\!\!E_T$ con respecto al número de interacciones producidas por pile-up, se define la resolución como la anchura de la componente x e y de la medida de $\not\!\!E_T$ añadidas en cuadratura. La Fig. 9.11 muestra la resolución obtenida para cada una de las distintas definiciones de $\not\!\!E_T$ en función del número medio de interacciones por paquete ($< \mu >$). Las definiciones basadas en las medidas energéticas de los calorímetros, $E_{\mathrm{T}}^{\mathrm{miss}}$ y $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$, muestran una resolución con tendencia positiva. La tendencia y los valores de la resolución obtenidos con dichas medidas es similar en eventos con y sin jets. Sin embargo, la pendiente obtenida con la medida $E_{\mathrm{T}}^{\mathrm{miss,track}}$ es aproximadamente nula, por lo que la resolución de la $E_{\mathrm{T}}^{\mathrm{miss,track}}$ es independiente del número de interacciones pile-up

que tienen lugar en el evento. En eventos sin jets, la resolución de la $E_{\mathrm{T}}^{\mathrm{miss,track}}$ es menor que la obtenida con $E_{\mathrm{T}}^{\mathrm{miss}}$ y $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$. En cambio, cuando los eventos tienen asociados dos o más jets la resolución de la $E_{\mathrm{T}}^{\mathrm{miss,track}}$ es un factor 2 mayor comparando con las definiciones basadas en las medidas proporcionadas por los calorímetros. Esto pone de manifiesto la degradación de la reconstrucción de la $E_{\mathrm{T}}^{\mathrm{miss,track}}$ debido a la presencia de partículas neutras en el proceso físico.

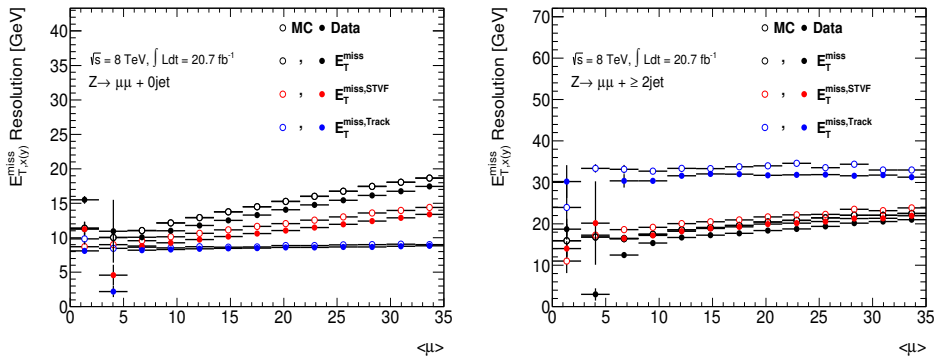

FIGURA 9.11: Resolución de las distintas definiciones de $\not\!\!E_T$ en bines del número de interacciones por paquete $< \mu >$. $Z \rightarrow \mu\mu$ eventos en datos y simulación con 0 y al menos 2 jets se muestran en la izquierda y derecha, respectivamente.

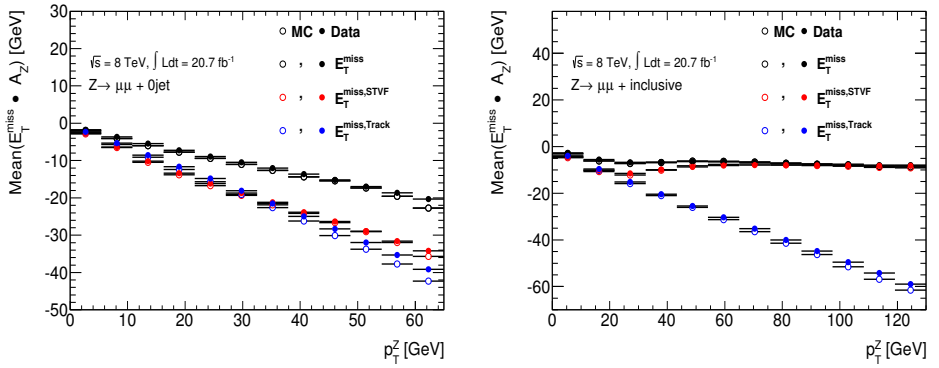

FIGURA 9.12: Proyección de las distintas definiciones de $\not\!\!E_T$ a lo largo de la dirección del bosón Z ($< E_T \cdot A^Z >$) en función del momento transverso del Z (p_{T}^Z). Datos tomados por ATLAS a 8 TeV (círculos) se comparan con datos simulados con MC (circunferencias). $Z \rightarrow \mu\mu$ eventos son separados: 0 jets (izquierda) y al menos dos jets (derecha).

A partir de la topología de eventos del proceso $Z \rightarrow \ell\ell$ se puede definir un eje en el plano transversal en la dirección del bosón Z. La dirección de este eje (A_Z)

viene definido por el momento reconstruido de los leptones:

$$A_Z = \frac{p_{\mathrm{T}}^{\ell^+} + p_{\mathrm{T}}^{\ell^-}}{|\, p_{\mathrm{T}}^{\ell^+} + p_{\mathrm{T}}^{\ell^-} \,|}\,, \tag{9.9}$$

donde los valores p_{T}^{ℓ} representan los vectores de momento transverso del leptón y del anti-leptón. Las componentes de la $\not\!\!E_T$ a lo largo de dicho eje son sensibles a efectos de resolución del detector [59]. La media de la distribución de la $\not\!\!E_T$ proyectada a lo largo de la dirección definida por la Eq. 9.9 ($Mean(E_{\mathrm{T}}^{\mathrm{miss}} \cdot A_Z)$) es sensible a efectos de resolución energética del detector, por lo tanto, mide la escala de la reconstrucción de $\not\!\!E_T$. En la Fig. 9.12 se muestran las distintas definiciones de $\not\!\!E_T$ proyectadas a lo largo del eje A_Z en función del momento transverso del bosón Z.

La linearidad es otra importante magnitud para evaluar la $\not\!\!E_T$ en sucesos donde se espera la presencia de neutrinos. La linearidad evalúa la consistencia entre la medida de $\not\!\!E_T$ reconstruida por distintos algoritmos y el valor esperado ($E_{\mathrm{T}}^{\mathrm{miss,\,True}}$), para el cual el proceso ha sido simulado. La linearidad viene definida por el valor medio del siguiente cociente:

$$\text{Linearidad} = \frac{E_{\mathrm{T}}^{\mathrm{miss}} - E_{\mathrm{T}}^{\mathrm{miss,\,True}}}{E_{\mathrm{T}}^{\mathrm{miss,\,True}}}\,. \tag{9.10}$$

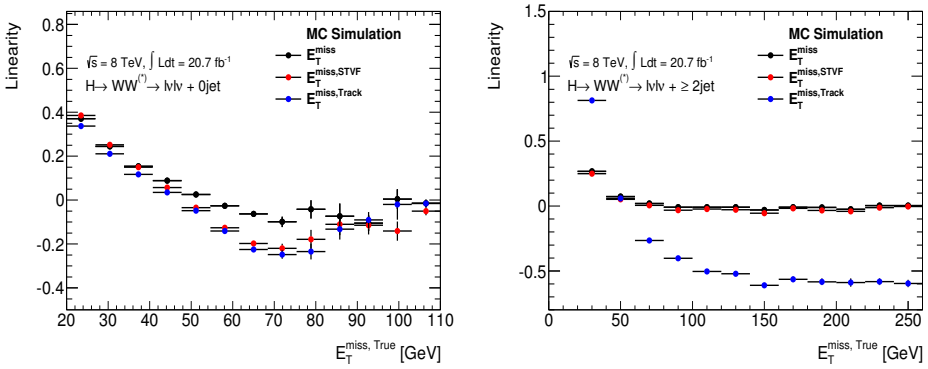

FIGURA 9.13: Distribuciones de la linearidad usando distintas definiciones de $\not\!\!E_T$ como función del valor esperado de la $\not\!\!E_T$ dado por la simulación ($E_{\mathrm{T}}^{\mathrm{miss,\,True}}$). Los eventos corresponden al proceso $H \to WW^{(*)} \to \ell\nu\ell\nu$, simulados con $m_H = 125\,\mathrm{GeV}$, los cuales se muestran para distinto número de jets en el estado final: $H{+}0j$ y $H{+}2j$ en la izquierda y derecha, respectivamente.

La linearidad debe ser nula si la medida de la $\not\!\!E_T$ está reconstruida en la escala correcta. La Fig. 9.13 muestra los resultados obtenidos para cada una de las definiciones de $\not\!\!E_T$ en función de la $E_{\mathrm{T}}^{\mathrm{miss,\,True}}$ obtenida por el generador de MC para la

simulación del proceso $H \to WW^{(*)} \to \ell\nu\ell\nu$. Los resultados están divididos en eventos sin jets y sucesos con alta actividad de jets. De nuevo, la definición de $E_{\mathrm{T}}^{\mathrm{miss,track}}$ tiene una gran diferencia relativa con respecto al valor esperado especialmente en eventos con al menos dos jets energéticos.

9.4 La Búsqueda de $H \to WW^{(*)} \to \ell\nu\ell\nu$ en ATLAS

9.4.1 El Modelo Estándar de Partículas y el Bosón de Higgs

El Modelo Estándar (SM, del inglés *Standard Model*) de Partículas es una teoría cuántica de campos que describe y unifica tres de las cuatro interacciones fundamentales de la naturaleza: la fuerza electromagnética, la fuerza débil y la fuerza fuerte. La Electrodinámica Cuántica (QED) describe las interacciones de las partículas cargadas eléctricamente usando el grupo de simetría $U(1)$ las cuales están mediadas por los fotones. La Cromodinámica Cuántica (QCD) describe las interacciones de las partículas cargadas con color (quarks) en términos del grupo de simetría $SU(3)$ dando lugar a 8 gluones intermediarios. La Teoría Electrodébil (EW) describe las interacciones de partículas cargadas que sienten la fuerza débil en términos del grupo de simetría SU(2) x U(1). Esta descripción da lugar a cuatro bosones intermediarios: $W^{\pm}$, Z y el fotón. Sin embargo, a diferencia de los bosones intermediarios en las QED y QCD los cuales no tienen masa, la interacción débil está mediada por los bosones masivos $W^{\pm}$ y Z. Para que dichos bosones tengan asociados términos de masa en la teoría, la simetría del Lagrangiano electrodébil se debe romper espontáneamente dando lugar a un nuevo campo escalar, el bosón de Higgs.

Las secciones eficaces de los modos de producción más relevantes del bosón de Higgs en función de su masa en colisiones de protones a 8 TeV se muestran en la Fig. 9.14. La Fig. 9.15 muestra los diagramas de Feynman de estos modos de producción. Los modos dominantes de producción del bosón de Higgs en el LHC son fusión de gluones (ggF, en inglés *gluon-gluon fusion*) y fusión de bosones vectoriales (VBF, en inglés *vector boson fusion*) con secciones eficaces de 19,52 pb y 1,58 pb para $m_{\mathrm{H}} = 125\,\mathrm{GeV}$, respectivamente.

9.4.2 Estrategia del Análisis $H \to WW^{(*)} \to \ell\nu\ell\nu$

La desintegración del bosón de Higgs en dos bosones W dando en el estado final dos leptones y dos neutrinos se beneficia de una alta probabilidad de producción (BR,

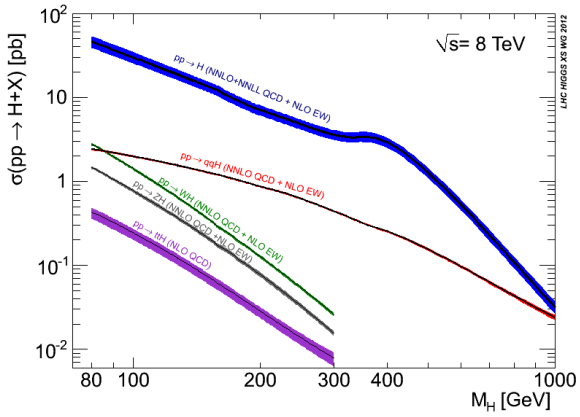

FIGURA 9.14: Secciones eficaces para distintos modos de producción del bosón de Higgs en función de su masa en colisiones de protones a una energía en el centro de masa de 8 TeV.

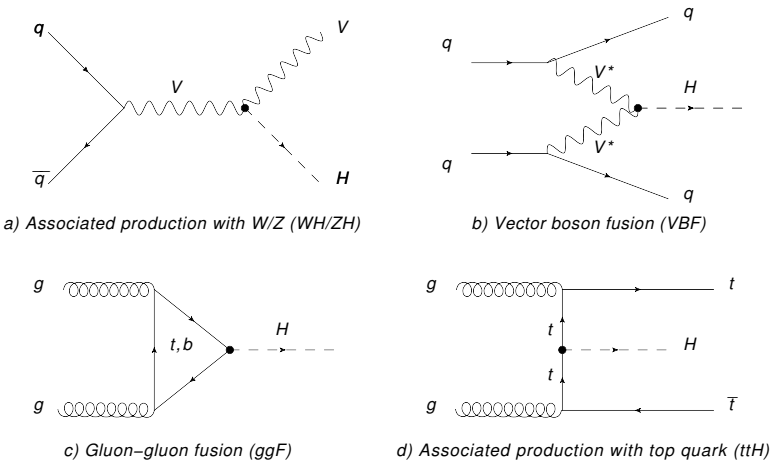

FIGURA 9.15: Diagramas de Feynman de los procesos de producción del bosón de Higgs en colisionadores de protones.

del inglés *branching ratios*) en un amplio rango de masa del bosón de Higgs como muestra la Fig. 9.16.

Así pues, el análisis de $H \to WW^{(*)} \to \ell\nu\ell\nu$ es uno de los canales más importantes para la búsqueda del bosón de Higgs. Sin embargo, este canal de desintegración también tiene inconvenientes. En primer lugar, sufre una gran contaminación procedente de otros procesos físicos que pueden tener reconstruido el mismo estado final. Estos procesos son conocidos como fondos. En segundo lugar, este análisis no es sensible a la masa del bosón de Higgs debido a la presencia de neutrinos en el estado final. Estos dos hechos caracterizan la estrategia del análisis de $H \to WW^{(*)} \to \ell\nu\ell\nu$:

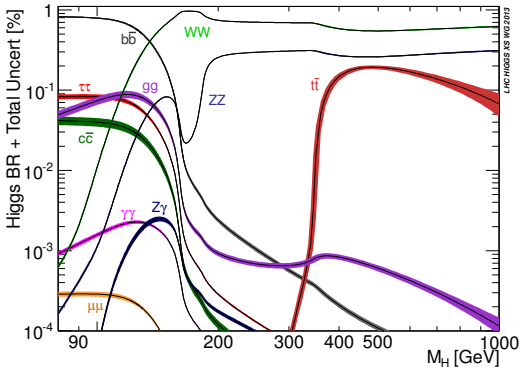

FIGURA 9.16: Probabilidad de desintegración del bosón de Higgs en distintos estados finales en función de la masa del bosón de Higgs.

la selección de eventos debe ser suficientemente dura como para reducir los fondos y, al mismo tiempo, suficientemente permisiva para incluir la señal de Higgs.

CUADRO 9.1: Porcentajes obtenidos de la simulación correspondientes a la producción de un bosón de Higgs con $m_H = 125\,\text{GeV}$ para los modos de producción ggF y VBF en función del número de jets en el estado final.

N_{jets}	ggF	VBF
$= 0$	53.5 %	2 %
$= 1$	25 %	11 %
≥ 2	21.5 %	87 %

La información del modo de producción del bosón de Higgs está relacionada con el número de jets en el estado final. El modo de producción de ggF mayormente origina eventos con cero o un jet. Sin embargo, el bosón de Higgs es mayormente producido vía VBF en eventos con al menos dos jets en el estado final. Tabla 9.1 muestra la contribución esperada de cada modo de producción para eventos con distinto número de jets en el estado final. Por este motivo, el análisis divide los eventos en tres categorías: $H{+}0j$, $H{+}1j$ y $H{+}2j$.

La estrategia del análisis también diferencia el sabor de los leptones dado que la contribución y composición de los fondos es diferente para leptones en el estado final con el mismo sabor ($ee{+}\mu\mu$) y leptones con distinto sabor ($e\mu{+}\mu e$). Figura 9.17 muestra las distribuciones de la multiplicidad de jets para eventos con leptones con mismo y distinto sabor, separadamente. Las contribuciones de fondo dominantes para los estados finales $e\mu{+}\mu e$ son WW en $H{+}0j$ y $t\bar{t}$ en $H{+}1j$ y $H{+}2j$. La contaminación en eventos $ee{+}\mu\mu$ está dominada por el proceso Z/γ^*. La contribución de $t\bar{t}$ toma relevancia en eventos con presencia de jets. La clasificación

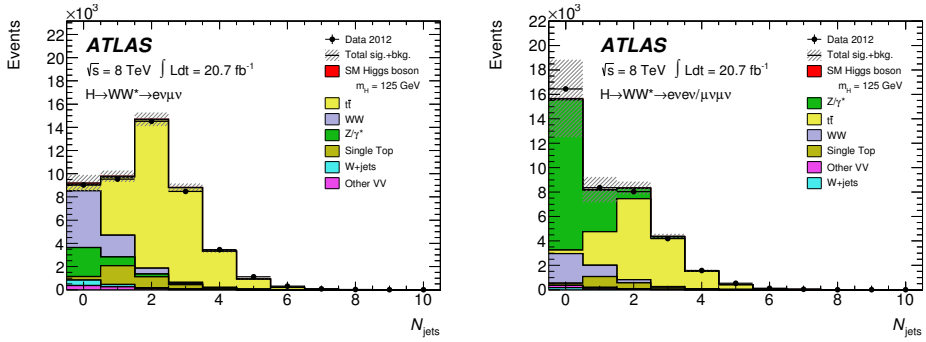

FIGURA 9.17: Número de jets para $e\mu$+μe (izquierda) y ee+$\mu\mu$ (derecha) eventos en datos de ATLAS a 8 TeV. La simulación obtenida mediante MC para la contribución esperada de los fondos del SM también se muestra. La banda amarilla representa la incertidumbre en la simulación.

de eventos debido al número de jets y sabor de los leptones del estado final permite adecuar la selección con el fin de suprimir de forma eficiente los fondos dominantes en cada caso.

9.4.3 Selección de Eventos

El análisis $H \to WW^{(*)} \to \ell\nu\ell\nu$ selecciona eventos con exactamente dos leptones bien reconstruidos de carga opuesta y con $p_{\mathrm{T}} > 25, 15$ GeV, donde los valores representan los umbrales para el leptón de mayor y menor momento transverso, respectivamente. La masa invariante del sistema de los dos leptones seleccionados deben ser mayor que 10, 12 GeV para $e\mu$+μe y ee+$\mu\mu$ estados finales, respectivamente. Para suprimir la contribución de Z/DY, los eventos ee+$\mu\mu$ consistentes con la masa del Z en un intervalo de $\pm$25 GeV son suprimidos y además, se requiere $E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 25$ y 45 GeV en $e\mu$+μe y ee+$\mu\mu$, respectivamente. Figura 9.18 muestra cómo el requisito en $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ mayormente suprime eventos originados por desintegración del bosón Z.

Seguidamente, el análisis categoriza los eventos dependiendo del número de jets en el estado final. En el análisis H+0j la selecciona eventos con ningún jet con $p_{\mathrm{T}} > 25\,(30)$ GeV en $|\eta| < (>)\,2{,}4$. Esto reduce la contribución de fondos producidos por el quark top. Seguidamente se requiere que el ángulo entre el sistema de dileptones y la $E_{\mathrm{T}}^{\mathrm{miss}}$ sea mayor que $\pi/2$. Esto suprime eventos patológicos en los que la $E_{\mathrm{T}}^{\mathrm{miss}}$ apunta en la dirección del par de leptones ($\Delta\phi_{ll,E_{\mathrm{T}}^{\mathrm{miss}}}$), mayormente eventos Z/DY, con un coste muy reducido para la selección de la señal de Higgs como muestra la Fig. 9.19.

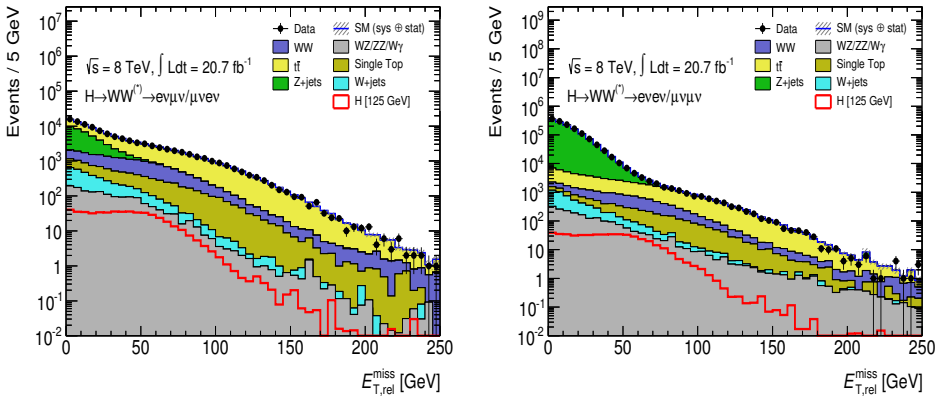

FIGURA 9.18: Distribuciones de $E_{\mathrm{T,rel}}^{\mathrm{miss}}$ para eventos $e\mu{+}\mu e$ y $ee{+}\mu\mu$ obtenidos a 8 TeV que satisfacen los requisitos de preselección del análisis $H\to WW^{(*)}\to \ell\nu\ell\nu$. Los fondos esperados para los distintos procesos del SM son mostrados, así como la señal del bosón de Higgs (línea roja).

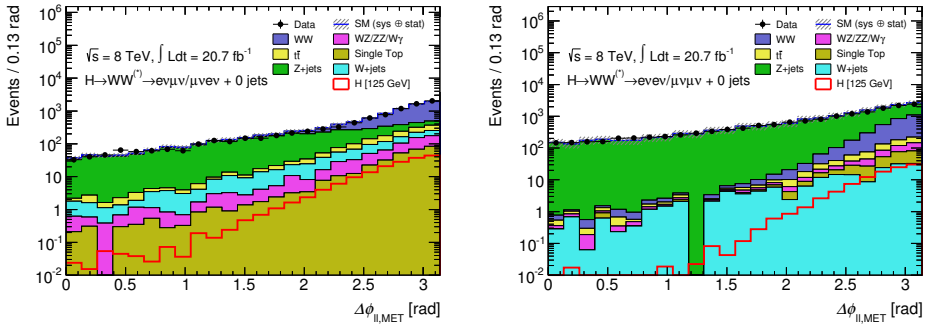

FIGURA 9.19: Distribuciones de $\Delta\phi_{ll,E_{\mathrm{T}}^{\mathrm{miss}}}$ para eventos $H{+}0j$ en datos a 8 TeV. Los plots muestran los estados finales $e\mu{+}\mu e$ en la derecha y $ee{+}\mu\mu$ en la derecha. Las distribuciones esperadas para los fondos del SM obtenidas a partir de simulación de MC se muestran separadamente. La señal esperada del bosón de Higgs con $m_H = 125\,\mathrm{GeV}$ está representada por la línea roja.

El siguiente criterio se basa en las distribuciones del momento transverso del sistema de los dos leptones ($p_{\mathrm{T}}^{\ell\ell}$) mostradas en la Fig. 9.20. La contribución de Z/DY se concentra en la región baja del espectro por lo que los eventos deben satisfacer la condición $p_{\mathrm{T}}^{\ell\ell} > 30\,\mathrm{GeV}$. Después de este requisito, la contaminación debida a Z/DY se reduce un factor ~ 15 y el fondo irreducible de WW representa el $\sim 70\,\%$ de los eventos totales en el estado final $e\mu{+}\mu e$.

Finalmente, los últimos requisitos de la selección se basan en la diferenciación de la topología del sistema de los dos leptones originados por la señal de Higgs. Los

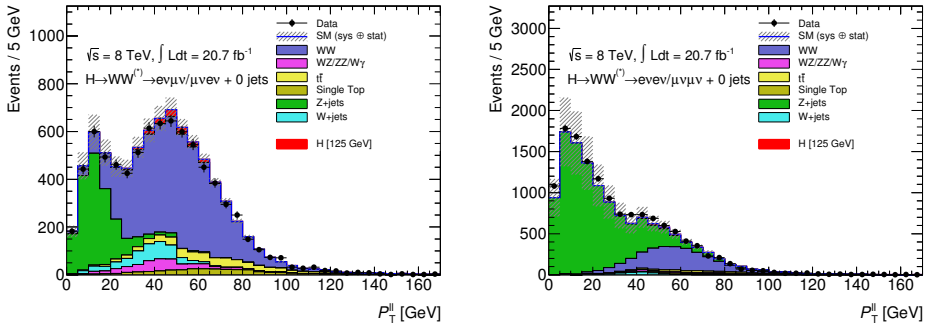

FIGURA 9.20: Distribuciones de $p_{\mathrm{T}}^{\ell\ell}$ para eventos H+0j en datos y simulación a 8 TeV para los estados finales $e\mu$+μe y ee+$\mu\mu$ en la izquierda y derecha, respectivamente. La señal esperada del bosón de Higgs con $m_H = 125\,\mathrm{GeV}$ está representada por el área roja sobre los fondos esperados del SM.

leptones deben emerger en la misma dirección y tener una masa invariante pequeña, por lo tanto: $\Delta\phi_{\ell\ell} < 1{,}8$ radianes y $m_{\ell\ell} < 50\,\mathrm{GeV}$. La distribución de $m_{\ell\ell}$ se muestra en la Fig 9.21 para eventos $e\mu$+μe en la categoría H+0j donde la mayor parte de la señal de Higgs con 125 GeV se concentra en la región más baja del espectro.

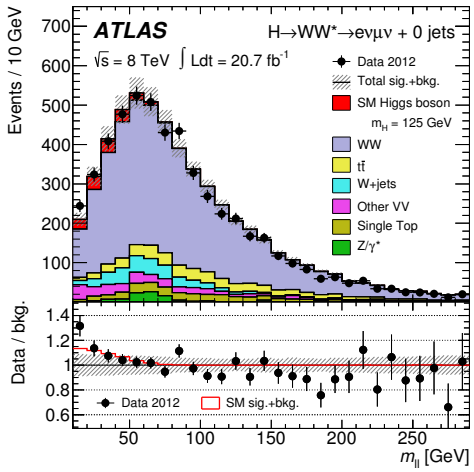

FIGURA 9.21: Distribuciones de $m_{\ell\ell}$ para los estados finales $e\mu$+μe sin jets en datos tomados por ATLAS a 8 TeV y simulación por MC.

Tabla 9.2 muestra los eventos observados y simulados para los distintos fondos y la señal de Higgs con $m_H = 125\,\mathrm{GeV}$ después de aplicar la selección completa para el estado final $e\mu$+μe en la categoría H+0j.

Para eventos con estados finales ee+$\mu\mu$ la contribución de Z/DY es todavía dos veces superior a la del fondo WW . Por ello en estos eventos se aplican requisitos

CUADRO 9.2: Número de eventos observados y esperados para la categoría H+0j con estados finales $e\mu$+μe en datos tomados por ATLAS a 8 TeV y simulación por MC.

	Observed	Total bkg.	Signal [$m_{\mathrm{H}} = 125$]	WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets
$E_{\mathrm{T,Rel}}^{\mathrm{miss}}$>25	46526	46660 ± 80	327.6 ± 2.2	7432 ± 23	778 ± 21	30700 ± 60	3360 ± 25	3356 ± 40	840 ± 10
$N_{\mathrm{jets}} = 0$	9024	9000 ± 40	171.7 ± 1.6	4901 ± 19	368 ± 12	506 ± 8	310 ± 7	2440 ± 32	470 ± 7
$\Delta\phi_{ll,E_{\mathrm{T}}^{\mathrm{miss}}} > \frac{\pi}{2}$	8100	8120 ± 40	170.4 ± 1.6	4839 ± 19	356 ± 12	491 ± 8	305 ± 8	1687 ± 29	437 ± 6
$p_{\mathrm{T}}^{\ell\ell} > 30$	5497	5485 ± 27	156.0 ± 1.5	4048 ± 17	288 ± 12	450 ± 8	280 ± 8	100 ± 14	319 ± 5
$m_{\ell\ell} < 50$	1453	1308 ± 14	124.0 ± 1.3	964 ± 8	110 ± 6	68.5 ± 3.2	45.5 ± 2.8	18 ± 7	101.5 ± 2.4
$\Delta\phi_{\ell\ell} < 1{,}8$	1399	1244 ± 13	119.2 ± 1.3	925 ± 8	107 ± 6	67.2 ± 3.1	44.4 ± 2.8	13 ± 7	87.5 ± 2.3

específicos para reducir la contribución de Z/DY. Debido a la complementariedad de la definiciones de $E_{\mathrm{T}}^{\mathrm{miss}}$ y $E_{\mathrm{T}}^{\mathrm{miss,track}}$, es posible reducir este fondo aplicando un requisito adicional en $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}} > 45\,\mathrm{GeV}$. Además se utiliza el cociente entre la suma escalar del momento de jets con $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ situados en el cuadrante opuesto al sistema de los dos leptones y el momento transverso de éstos, $p_{\mathrm{T}}^{\ell\ell}$. Esta magnitud es conocida como f_{recoil} y distingue entre procesos Z/DY y otros procesos con dos leptones en el estado final. Tabla 9.3 muestra los eventos observados y simulados para los distintos fondos y la señal de Higgs con $m_H = 125\,\mathrm{GeV}$ después de aplicar la selección completa para el estado final ee+$\mu\mu$ en la categoría H+0j.

En el análisis H+1j se seleccionan eventos con exactamente un jet en el estado final. Estos eventos están dominados por el proceso $t\bar{t}$, el cual representa el $\sim$ 70 % en el estado final $e\mu$+μe y es del orden de la contribución de Z/DY en ee+$\mu\mu$, como se observa en las distribuciones de la Fig. 9.17. Para suprimir la contribución de $t\bar{t}$, la selección excluye eventos con jets considerados como productos de un quark b: $N_{b\text{-jet}} = 0$. Para los estados finales $e\mu$+μe en la región de H+1j, el proceso $Z \to \tau\tau$ es suprimido descartando eventos consistentes con la masa del bosón Z: $|m_{\tau\tau} - m_Z| < 25\,\mathrm{GeV}$. Finalmente, los eventos finales deben pasar los mismos requisitos en $m_{\ell\ell}$ y $\Delta\phi_{\ell\ell}$ descritos para el caso H+0j.

Figura 9.22 muestra las distribuciones de $\Delta\phi_{\ell\ell}$ para los eventos que satisfacen la selección para la categoría de H+0j separadamente para los estados finales $e\mu$+μe y ee+$\mu\mu$. Para los eventos ee+$\mu\mu$, también se reduce la contribución del fondo de Z/DY usando $E_{\mathrm{T,Rel}}^{\mathrm{miss,track}} > 45\,\mathrm{GeV}$ y $f_{\mathrm{recoil}} < 0{,}2$. Tablas 9.4 y 9.5 muestran los eventos observados y esperados para la señal del bosón de Higgs con $m_H = 125\,\mathrm{GeV}$

CUADRO 9.3: Número de eventos observados y esperados para la categoría H+0j con estados finales ee+$\mu\mu$ en datos tomados por ATLAS a 8 TeV y simulación por MC.

	Observed	Total bkg.	Signal [$m_H = 125$]	WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets
$E_{\mathrm{T,Rel}}^{\mathrm{miss}}$>45	39330	38430 ± 190	189.2 ± 1.7	3691 ± 16	404 ± 11	15540 ± 50	1776 ± 18	16710 ± 190	306 ± 14
$N_{\mathrm{jets}} = 0$	16446	15550 ± 160	103.7 ± 1.2	2436 ± 13	191 ± 5	281 ± 6	175 ± 6	12300 ± 160	172 ± 11
$\Delta\phi_{ll,E_{\mathrm{T}}^{\mathrm{miss}}} > \frac{\pi}{2}$	13697	12970 ± 140	103.5 ± 1.2	2431 ± 13	188 ± 5	277 ± 6	174 ± 6	9740 ± 140	161 ± 10
$p_{\mathrm{T}}^{\ell\ell} > 45$	5670	5650 ± 70	99.3 ± 1.2	2300 ± 13	172 ± 5	264 ± 6	167 ± 5	2610 ± 70	133.9 ± 3.6
$m_{\ell\ell} < 50$	2314	2393 ± 22	84.0 ± 1.1	759 ± 7	64.1 ± 2.8	53.4 ± 2.9	42.2 ± 2.7	1412 ± 20	62.1 ± 2.5
$E_{\mathrm{T,Rel}}^{\mathrm{miss,track}}$>45	1032	993 ± 10	62.9 ± 0.9	646 ± 7	41.5 ± 2.0	46.7 ± 2.7	38.8 ± 2.5	200 ± 5	19.4 ± 1.5
$\Delta\phi_{\ell\ell} < 1{,}8$	1026	983 ± 10	62.5 ± 0.9	644 ± 7	41.5 ± 2.0	46.0 ± 2.7	38.8 ± 2.5	195 ± 5	18.4 ± 1.5
$f_{\mathrm{recoil}} < 0{,}05$	671	647 ± 7	41.8 ± 0.8	515 ± 6	29.8 ± 1.6	19.4 ± 1.8	21.8 ± 1.9	48.6 ± 2.6	12.0 ± 1.3

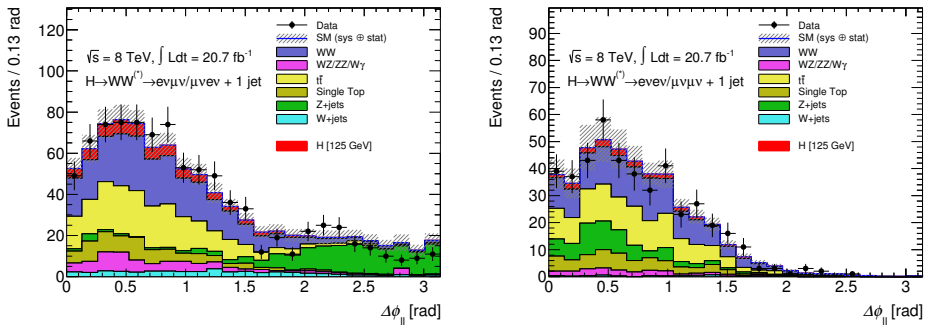

FIGURA 9.22: Distribuciones de $\Delta\phi_{\ell\ell}$ para eventos del tipo H+1j observados en datos y simulación por MC a 8 TeV. Los resultados se muestran para los estados finales $e\mu$+μe y ee+$\mu\mu$ en la izquierda y derecha, respectivamente. La señal esperada del bosón de Higgs con $m_H = 125$ GeV (área roja) se superpone al total de los fondos esperados del SM.

y los distintos fondos del SM después de aplicar la selección completa de la categoría H+1j para estados finales $e\mu$+μe y ee+$\mu\mu$, respectivamente.

La última región en la que el análisis $H \to WW^{(*)} \to \ell\nu\ell\nu$ se divide es la categoría conocidas como H+2j, en la que se seleccionan al menos dos jets en el estado final. Este estado final está dominado por el fondo de $t\bar{t}$ casi por completo por lo que se excluyen eventos con b-jets, como en el caso de H+0j. En la categoría H+2j la señal del bosón de Higgs está mayoritariamente formada a través del modo de

CUADRO 9.4: Número de eventos observados y esperados para la categoría $H+1j$ con estados finales $e\mu+\mu e$ en datos tomados por ATLAS a 8 TeV y simulación por MC.

	Observed	Total bkg.	Signal [$m_H = 125$]	WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets
$E_{T,Rel}^{miss}$>25	46526	46660 ± 80	327.6 ± 2.2	7432 ± 23	778 ± 21	30700 ± 60	3360 ± 25	3356 ± 40	840 ± 10
$N_{jets} = 1$	9527	9460 ± 40	96.7 ± 1.2	1656 ± 10	265 ± 12	4981 ± 25	1604 ± 17	757 ± 15	195 ± 5
$N_{b\text{-jet}} = 0$	4320	4239 ± 25	84.8 ± 1.1	1460 ± 10	224 ± 9	1272 ± 13	457 ± 9	667 ± 14	160 ± 4
$Z \to \tau\tau$ veto	4138	4024 ± 25	84.4 ± 1.1	1417 ± 9	217 ± 9	1220 ± 13	436 ± 9	580 ± 14	155 ± 4
$m_{\ell\ell} < 50$	886	829 ± 11	63.4 ± 0.9	269 ± 4	69 ± 5	216 ± 6	79 ± 4	149 ± 5	45.6 ± 2.3
$\Delta\phi_{\ell\ell} < 1{,}8$	728	650 ± 9	58.8 ± 0.9	247 ± 4	60.5 ± 3.5	204 ± 6	76 ± 4	27.8 ± 3.3	34.5 ± 2.0

CUADRO 9.5: Número de eventos observados y esperados para la categoría $H+1j$ con estados finales $ee+\mu\mu$ en datos tomados por ATLAS a 8 TeV y simulación por MC.

	Observed	Total bkg.	Signal [$m_H = 125$]	WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets
$E_{T,Rel}^{miss}$>45	39330	38430 ± 190	189.2 ± 1.7	3691 ± 16	404 ± 11	15540 ± 50	1776 ± 18	16710 ± 190	306 ± 14
$N_{jets} = 1$	8354	8110 ± 80	54.3 ± 1.0	820 ± 7	137 ± 8	2744 ± 19	885 ± 13	3470 ± 80	60 ± 9
$N_{b\text{-jet}} = 0$	5192	4800 ± 70	47.7 ± 0.9	723 ± 7	120 ± 8	719 ± 10	256 ± 7	2930 ± 70	44 ± 8
$m_{\ell\ell} < 50$	1773	1537 ± 16	37.6 ± 0.7	194.5 ± 3.5	34.8 ± 1.7	166 ± 5	64.9 ± 3.4	1057 ± 14	20.1 ± 1.5
$E_{T,Rel}^{miss,track} > 45$	440	418 ± 7	21.4 ± 0.6	148.1 ± 3.0	20.6 ± 1.3	128 ± 5	51.9 ± 3.1	64 ± 4	5.1 ± 0.8
$\Delta\phi_{\ell\ell} < 1{,}8$	430	407 ± 7	20.4 ± 0.5	143.2 ± 3.0	19.9 ± 1.2	126 ± 4	51.0 ± 3.1	63 ± 4	4.5 ± 0.7
$f_{recoil} < 0{,}2$	346	315 ± 6	16.2 ± 0.5	128.4 ± 2.8	17.4 ± 1.2	97 ± 4	44.3 ± 2.9	25.1 ± 2.1	3.1 ± 0.6

producción VBF, la cual se caracteriza por la presencia de dos jets muy energéticos que emergen en sentidos opuestos. La estrategia del análisis de la categoría $H+2j$ explota estas características para distinguir la señal de Higgs producida vía VBF de los fondos del SM, definiendo observables cinemáticos a partir de los dos jets con mayor p_T en el evento. La masa invariante del sistema de los dos jets debe ser grande: $m_{jj} > 500\,\text{GeV}$, así como la separación entre los mismos: $\Delta y_{jj} > 2{,}8$, donde $\Delta y_{jj} \equiv |y_{j_1} - y_{j_2}|$. Figuras 9.23 y 9.24 muestran las distribuciones de Δy_{jj} y m_{jj}, respectivamente.

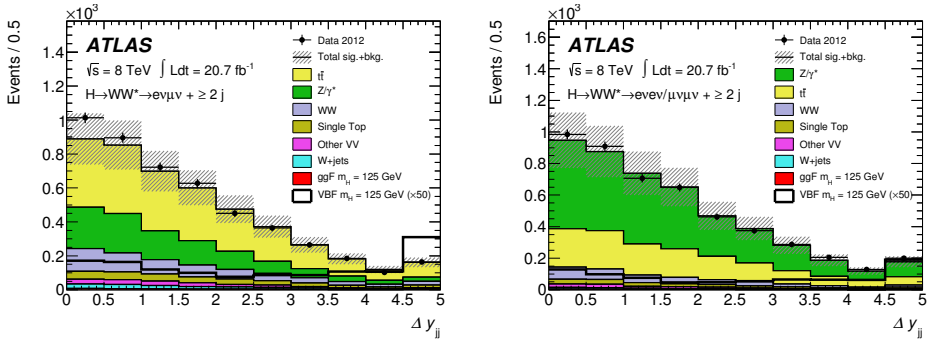

FIGURA 9.23: Distribuciones de Δy_{jj} con $N_{\text{jets}} \geq 2$ en datos de ATLAS tomados a 8 TeV y simulación por MC, después de discriminar b-quarks y $p_{\text{T}}^{\text{tot}} < 45\,\text{GeV}$.

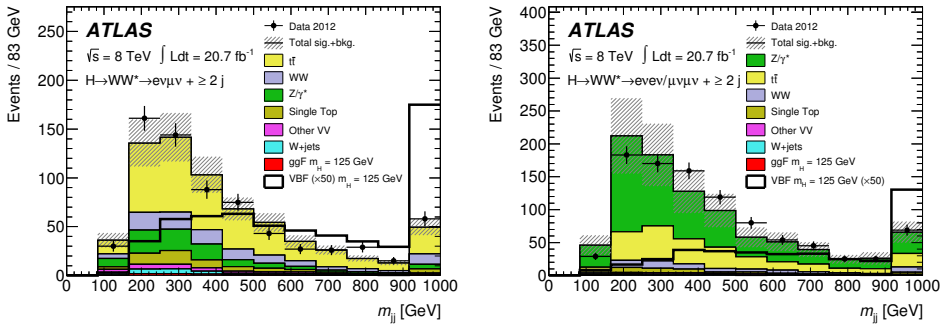

FIGURA 9.24: Distribuciones de m_{jj} en el análisis $H \to WW^{(*)} \to \ell\nu\ell\nu$ con $N_{\text{jets}} \geq 2$ en datos tomados por ATLAS a 8 TeV y simulación de MC, después de requerir $|\Delta y_{jj}| > 2{,}8$.

Además, la actividad en la región definida por los dos jets con mayor p_{T} también permite eliminar procesos procedentes de los fondos. Esta región debe incluir la actividad de los leptones pero no puede contener ningún jet con $p_{\text{T}} > 20\,\text{GeV}$. Después de estos requisitos, los fondos totales del SM disminuyen un factor 3 y la contribución del proceso $t\bar{t}$ se reduce al orden de magnitud del fondo de WW en estados finales con $e\mu$+μe. Finalmente, la selección también excluye eventos con valores altos de $m_{\ell\ell}$ y $\Delta\phi_{\ell\ell}$, cuyas distribuciones para la señal de Higgs producida vía VBF, así como los fondos que satisfacen la selección específica para la categoría H+$2j$ se muestran en las Fig. 9.25 y Fig 9.26, respectivamente.

Tablas 9.6 y 9.7 muestran el número de eventos esperados para la señal de Higgs con $m_H = 125\,\text{GeV}$ producida vía VBF y los fondos del SM después de aplicar cada criterio de selección en la categoría H+$2j$ para los estados finales $e\mu$+μe y ee+$\mu\mu$, respectivamente.

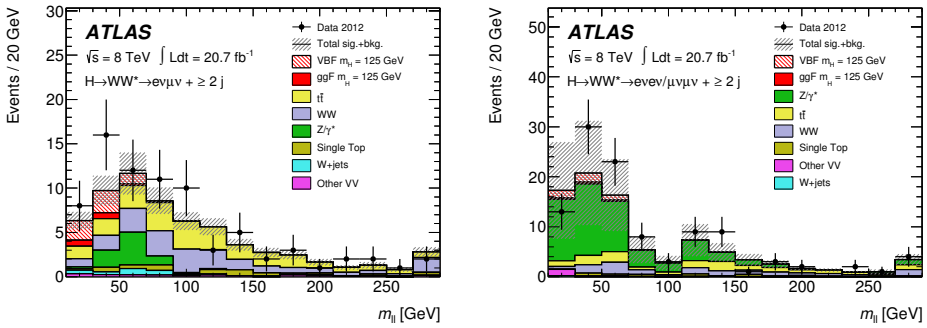

FIGURA 9.25: Distribuciones de $m_{\ell\ell}$ en el análisis de $H\to WW^{(*)}\to \ell\nu\ell\nu$ con $N_{\text{jets}} \geq 2$ en datos tomados por ATLAS y simulación por MC, después de requerir OLV.

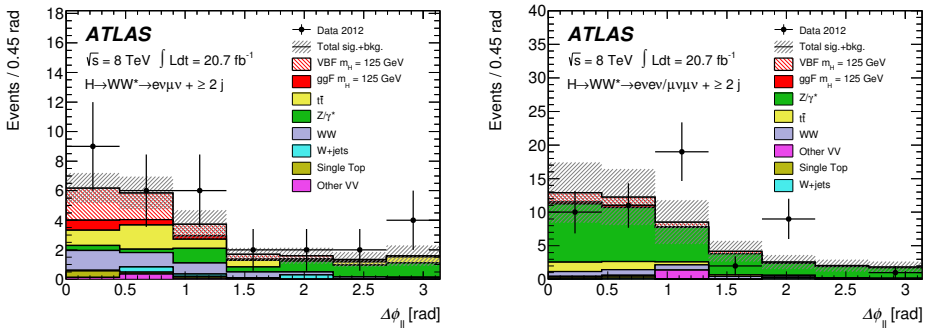

FIGURA 9.26: Distribuciones de $\Delta\phi_{\ell\ell}$ en el análisis $H\to WW^{(*)}\to \ell\nu\ell\nu$ con $N_{\text{jets}} \geq 2$ en datos tomados por ATLAS a 8 TeV y simulación por MC, después de requerir $m_{\ell\ell} < 60$ GeV.

9.4.4 Resultados

Después de la selección de eventos en cada una de las categorías $H{+}0j$, $H{+}1j$ y $H{+}2j$ para los estados finales $e\mu{+}\mu e$ y $ee{+}\mu\mu$ el análisis utiliza la masa transversa definida como:

$$m_{\text{T}} = \sqrt{(E_{\text{T}}^{\ell\ell} + E_{\text{T}}^{\text{miss}})^2 - |\,\vec{p}_{\text{T}}^{\,\ell\ell} + \mathbf{E}_{\text{T}}^{\text{miss}}\,|^2} \tag{9.11}$$

donde $E_{\text{T}}^{\ell\ell} = \sqrt{|\,\vec{p}_{\text{T}}^{\,\ell\ell}\,|^2 + m_{\ell\ell}^2}$ con $|\vec{p}_{\text{T}}^{\,\ell\ell}| = p_{\text{T}}^{\ell\ell}$. Figura 9.27 contiene la distribución de m_{T} para aquellos eventos finales que satisfacen la selección completa del análisis $H\to WW^{(*)}\to \ell\nu\ell\nu$ correspondiente a las categorías $H{+}0j$ y $H{+}1j$.

Equivalentemente, para el modo de producción vía VBF, la distribución de m_{T} en el análisis $H{+}2j$ se muestra en la Fig.9.28 para leptones en los estados finales: $e\mu{+}\mu e$. Para mostrar la sensibilidad del análisis de $H\to WW^{(*)}\to \ell\nu\ell\nu$, en

CUADRO 9.6: Número de eventos observados y esperados para la categoría $H+2j$ con estados finales $e\mu+\mu e$ en datos tomados por ATLAS a 8 TeV y simulación por MC.

	Observed	Total bkg.	Signal $m_H = 125$ GeV		WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets
			VBF	ggF						
E_T^{miss}>20	84324	83810 ± 130	77.5 ± 1.4	376.5 ± 2.2	8801 ± 23	1292 ± 26	49910 ± 80	5051 ± 31	16570 ± 80	1766 ± 16
$N_{jets} \geq 2$	48723	47740 ± 80	42.6 ± 0.9	66.9 ± 0.9	939 ± 7	299 ± 15	41850 ± 70	2368 ± 21	1811 ± 25	437 ± 8
$N_{b\text{-jet}} = 0$	5852	5697 ± 32	30.6 ± 0.7	49.1 ± 0.8	685 ± 6	202 ± 7	2932 ± 21	351 ± 8	1310 ± 18	171 ± 5
$\mathbf{p}_T^{tot} < 45$	4790	4615 ± 29	26.7 ± 0.7	40.8 ± 0.7	591 ± 6	155 ± 7	2319 ± 18	287 ± 8	1168 ± 17	126 ± 4
$Z \to \tau\tau$ veto	4007	3846 ± 26	24.5 ± 0.7	38.0 ± 0.7	544 ± 6	141 ± 7	2148 ± 18	264 ± 7	673 ± 14	108 ± 4
$\|\Delta y_{jj}\| > 2{,}8$	696	678 ± 10	11.9 ± 0.23	9.50 ± 0.34	100.0 ± 2.3	24.8 ± 3.4	377 ± 7	55.1 ± 3.1	95 ± 5	19 ± 2
$m_{jj} > 500$	198	170 ± 4	7.54 ± 0.12	2.93 ± 0.19	33.7 ± 1.2	5.6 ± 0.6	93.4 ± 3.0	11.4 ± 1.2	18.9 ± 2.5	4.4 ± 0.7
CJV	92	77.6 ± 2.4	6.30 ± 0.11	1.74 ± 0.15	25.5 ± 1.0	2.8 ± 0.4	30.2 ± 1.5	5.3 ± 0.8	9.3 ± 1.2	3.1 ± 0.6
OLV	78	58.5 ± 2.1	6.07 ± 0.10	1.57 ± 0.14	18.7 ± 0.8	2.05 ± 0.32	22.5 ± 1.3	4.3 ± 0.7	7.3 ± 1.2	2.4 ± 0.5
$m_{\ell\ell} < 60$	31	16.4 ± 1.3	5.49 ± 0.10	1.48 ± 0.14	3.8 ± 0.4	0.66 ± 0.21	4.48 ± 0.69	0.70 ± 0.31	4.4 ± 0.8	1.0 ± 0.4
$\Delta\phi_{\ell\ell} < 1{,}8$	23	12.3 ± 1.0	5.11 ± 0.09	1.34 ± 0.13	3.5 ± 0.4	0.63 ± 0.21	3.7 ± 0.7	0.70 ± 0.31	1.9 ± 0.5	0.56 ± 0.30

CUADRO 9.7: Número de eventos observados y esperados para la categoría $H+2j$ con estados finales $ee+\mu\mu$ en datos tomados por ATLAS a 8 TeV y simulación por MC.

	Observed	Total bkg.	Signal $[m_H = 125]$		WW	VV	$t\bar{t}$	Single Top	Z+jets	W+jets
			VBF	ggF						
E_T^{miss}>45 $E_T^{miss,\,STVF}$>35	58690	56930 ± 210	45.5 ± 1	198.8 ± 2	3924 ± 20	604 ± 10	29300 ± 60	2863 ± 20	19620 ± 190	463 ± 16
$N_{jets} \geq 2$	32877	32170 ± 110	26.4 ± 0.7	39.7 ± 0.7	537 ± 6	186 ± 9	24540 ± 60	1388 ± 16	5420 ± 90	191 ± 7
$N_{b\text{-jet}} = 0$	65388	6370 ± 80	18.9 ± 0.6	29.6 ± 0.6	394 ± 5	129 ± 7	1747 ± 16	203 ± 6	3810 ± 80	58 ± 4
$p_T^{tot} < 45$	4903	4830 ± 70	16.7 ± 0.5	24.4 ± 0.6	336 ± 4	93 ± 5	1375 ± 14	171 ± 6	2790 ± 70	42.9 ± 2.9
$\|\Delta y_{jj}\| > 2{,}8$	958	926 ± 33	8.06 ± 0.20	6.23 ± 0.28	61.1 ± 1.7	12.7 ± 1.3	253 ± 6	35.3 ± 2.4	552 ± 33	6.2 ± 1.1
$m_{jj} > 500$	298	245 ± 6	5.55 ± 0.10	2.10 ± 0.16	23.5 ± 1.0	4.1 ± 1.1	62.4 ± 2.5	9.3 ± 1.1	139 ± 5	1.4 ± 0.6
CJV	147	119 ± 4	4.65 ± 0.10	1.10 ± 0.12	16.6 ± 0.8	2.8 ± 1.1	19.3 ± 1.3	4.1 ± 0.7	72.6 ± 3.4	0.68 ± 0.38
OLV	108	82.7 ± 3.3	4.45 ± 0.09	0.93 ± 0.11	12.5 ± 0.6	2.3 ± 1.1	14.3 ± 1.1	3.1 ± 0.6	49.0 ± 2.7	0.30 ± 0.30
$m_{\ell\ell} < 60$	52	38.7 ± 2.5	4.03 ± 0.09	0.81 ± 0.10	3.23 ± 0.34	1.7 ± 1.1	3.8 ± 0.6	0.80 ± 0.30	28.3 ± 2.1	0.14 ± 0.24
$\Delta\phi_{\ell\ell} < 1{,}8$	42	33.0 ± 2.4	3.70 ± 0.09	0.72 ± 0.09	2.82 ± 0.32	1.7 ± 1.1	3.3 ± 0.5	0.74 ± 0.30	23.6 ± 2.0	0.06 ± 0.21

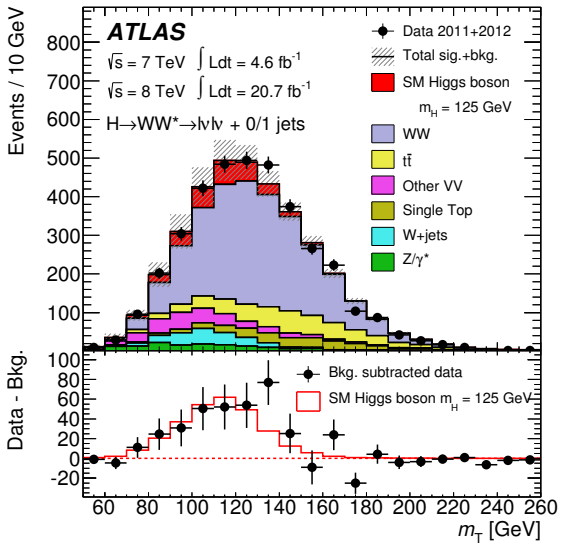

FIGURA 9.27: Distribuciones de m_{T} para los eventos finales en el análisis $H\to WW^{(*)}\to \ell\nu\ell\nu$ usando datos a 8 TeV y simulación por MC. Los resultados muestran las categorías H+0j y H+1j combinadas. La señal esperada del bosón de Higgs se muestra en rojo superpuesta a la suma de los fondos del SM simulados por MC. El área gris representa las incertidumbres de la señal y los fondos debidas a errores estadísticos, experimentales y teóricos. La distribución inferior muestra la diferencia de los datos con respecto a los fondos estimados, comparada con la distribución esperada de m_{T} de la señal del bosón de Higgs con $m_{\mathrm{H}} = 125\,\mathrm{GeV}$.

la Tab. 9.8 se muestran los eventos observados a 8 TeV y esperados por la simulación para los distintos fondos del SM y la señal del bosón de Higgs con $m_H = 125\,\mathrm{GeV}$ en el rango de m_{T} donde hay mayor contribución de señal: m_{T} 0,75 $m_H < m_{\mathrm{T}} < m_H$ para H+0j y H+1j, y $m_{\mathrm{T}} < 1{,}2\, m_H$ for H+2j.

Para comprobar la presencia de señal, el análisis realiza un ajuste estadístico que explota las diferencias en composición y contribución de los distintos fondos usando la distribución de m_{T}. El resultado obtenido a través del ajuste estadístico para el análisis de $H\to WW^{(*)}\to \ell\nu\ell\nu$ combinando los datos tomados por el detector ATLAS a 7 y 8 TeV para todos los estados finales se muestra en la Fig. 9.29. La significancia observada a $m_H = 125\,\mathrm{GeV}$ es de 3.8 σ mientras que la esperada es de 3.7 σ. Para determinar la compatibilidad de los resultados con el SM se define el parámetro μ como el cociente del ajuste estadístico obtenido a partir de los datos usando una masa fija para el bosón de Higgs de $m_H = 125\,\mathrm{GeV}$. El exceso de eventos

CUADRO 9.8: Número de eventos observados en datos a 8 TeV y esperados para la señal del bosón de Higgs con $m_H = 125\,\text{GeV}$ y los fondos del SM en las regiones de m_T $0{,}75\,m_H < m_\text{T} < m_H$ para $H{+}0j$ y $H{+}1j$, y $m_\text{T} < 1{,}2\,m_H$ para $H{+}2j$. Todos los posibles estados finales de los leptones resultantes han sido combinados. El número de eventos totales esperados para los fondos se muestra en la tabla de arriba, mientras que la contribución de cada uno de los procesos dominantes se muestran abajo. Los errores incluyen incertidumbres estadísticas y sistemáticas.

N_jets	N_obs	N_bkg	N_sig
$= 0$	831	739 ± 39	100 ± 21
$= 1$	309	261 ± 28	41 ± 14
≥ 2	55	36 ± 4	10.9 ± 1.4

N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	$N_{Z/\text{DY}}$	$N_{W+\,\text{jets}}$
551 ± 41	58 ± 8	23 ± 3	16 ± 2	30 ± 10	61 ± 21
108 ± 40	27 ± 6	68 ± 18	27 ± 10	12 ± 6	20 ± 5
4.1 ± 1.5	1.9 ± 0.4	4.6 ± 1.7	0.8 ± 0.4	22 ± 3	0.7 ± 0.2

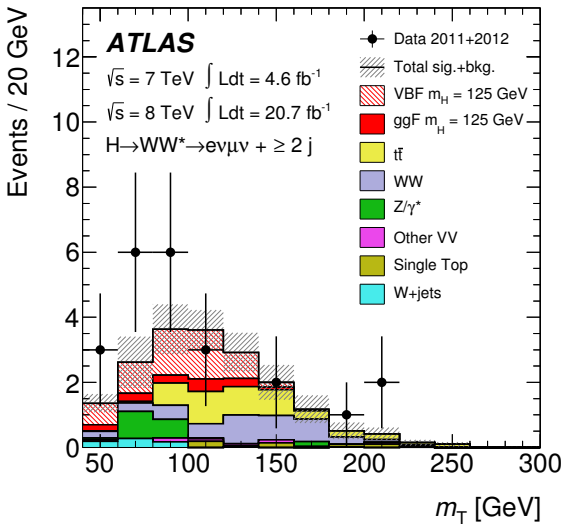

FIGURA 9.28: Distribución de la masa transversa para eventos que satisfacen la selección completa del análisis $H \to WW^{(*)} \to \ell\nu\ell\nu$ con estados finales $e\mu{+}\mu e$ y $N_\text{jets} \geq 2$ en datos de ATLAS a 8 TeV y simulación por MC.

observados es compatible con la predicción del bosón de Higgs con $m_H = 125\,\text{GeV}$:

$$\begin{aligned}\mu_\text{obs} &= 1{,}01 \pm 0{,}21\,(\text{stat.}) \pm 0{,}19\,(\text{th. syst.}) \pm 0{,}12\,(\text{exp. syst.}) \pm 0{,}04\,(\text{lumi.}) \\ &= 1{,}01 \pm 0{,}31.\end{aligned}$$

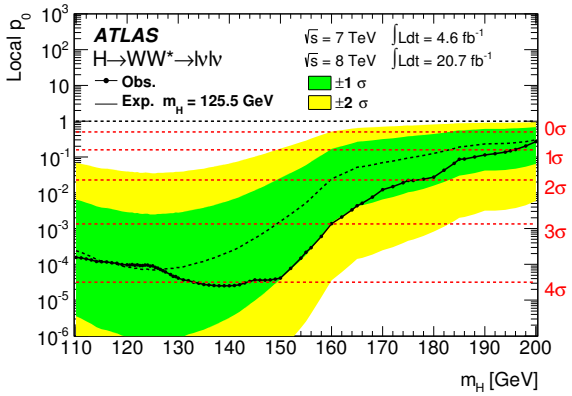

FIGURA 9.29: Valores locales p_0 esperados (línea punteada) y observados (línea continua) en función de la masa del bosón de Higgs en el análisis $H\to WW^{(*)}\to \ell\nu\ell\nu$ combinando los datos tomados en ATLAS a 7 TeV y 8 TeV. La banda verde indica la incertidumbre en la curva del valor esperado de p_0 en el rango $\pm 1\,\sigma$ y la banda amarilla representa la incertidumbre con $\pm 2\,\sigma$.

Para el modo de producción mediante VBF, el ajuste estadístico puede realizarse considerando la señal producida vía ggF como parte del fondo. Figura 9.30 compara la curva p_0 observada con la simulación de la señal del bosón de Higgs con $m_H = 125\,\text{GeV}$. Los valores observados y esperados para dicha masa. La significancia observada para el bosón de Higgs producido a través del mecanismo VBF con $m_H = 125\,\text{GeV}$ es $2{,}5\,\sigma$ mientras que el esperado es de $1{,}6\,\sigma$. Estos valores se encuentran en el rango de compatibilidad comprendido por la banda $\pm 1\,\sigma$ de las incertidumbres

Todos los resultados obtenidos de los datos tomados por el detector ATLAS a $7 + 8\,\text{TeV}$ muestran la compatibilidad de la observación con la predicción del SM de una señal producida por el bosón de Higgs con $m_H = 125\,\text{GeV}$ decayendo a dos bosones W. La combinación de los resultados del análisis $H\to WW^{(*)}\to \ell\nu\ell\nu$ con los resultados de los canales de desintegración $H \to \gamma\gamma$ y $H \to ZZ \to 4\ell$ permitió anunciar el descubrimiento del bosón de Higgs en el detector ATLAS con una signficancia observada de 5 σ el 4 de Julio de 2012 [68].

9.5 Optimización de la $\not{E}_T$ en $H\to WW^{(*)}\to \ell\nu\ell\nu$

Después de los primeros resultados del análisis de $H\to WW^{(*)}\to \ell\nu\ell\nu$, numerosos estudios se centraron en mejorar y optimizar la selección con el fin de aumentar la sensibilidad de la búsqueda. Los resultados finales se benefician principalmente

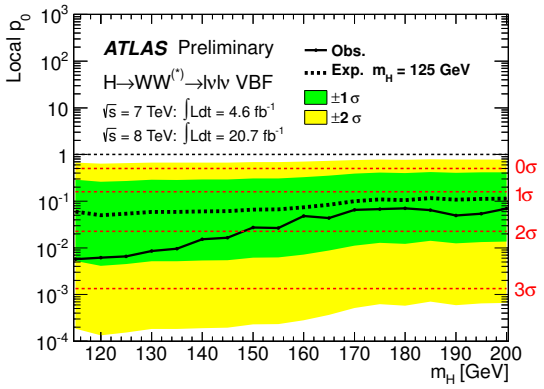

FIGURA 9.30: Valores locales p_0 esperados (línea punteada) y observados (línea continua) en función de la masa del bosón de Higgs producido vía VBF en el análisis $H \to WW^{(*)} \to \ell\nu\ell\nu$ combinando los datos tomados en ATLAS a 7 TeV y 8 TeV. La banda verde indica la incertidumbre en la curva del valor esperado de p_0 en el rango $\pm 1\,\sigma$ y la banda amarilla representa la incertidumbre con $\pm 2\,\sigma$.

de las investigaciones realizadas en la optimización del criterio de $\not{E}_T$ usado con la introducción de una nueva reconstrucción $E_{\mathrm{T}}^{\mathrm{miss,\ track,\ jetCorr}}$.

9.5.1 Definición de $E_{\mathbf{T}}^{\mathbf{miss,\ track,\ jetCorr}}$

La nueva reconstrucción de la medida de la $\not{E}_T$ se basa en la reconstrucción de $E_{\mathrm{T}}^{\mathrm{miss,track}}$ pero reemplazando la energía de las trazas asociadas a objetos por la reconstrucción de energía de los mismos realizada en los calorímetros. Además, jets, reconstruidos con altos criterios de calidad, que no tienen trazas asociadas son añadidos en la nueva reconstrucción $E_{\mathrm{T}}^{\mathrm{miss,\ track,\ jetCorr}}$. La definición de $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ viene dada por la siguiente expresión:

$$E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}} = -\sum_{\text{i trks}}^{PV} \vec{p_{\mathrm{T}}}^{\,i} + \sum_{\ell \text{ leptones}} \left(\vec{p_{\mathrm{T}}}^{\,\ell,\mathrm{trk}} - \vec{p_{\mathrm{T}}}^{\,\ell,\mathrm{calo}} \right) + \sum_{\text{j jets}} \left(\vec{p_{\mathrm{T}}}^{\,j,\mathrm{trk}} - \vec{p_{\mathrm{T}}}^{\,j,\mathrm{calo}} \right), \tag{9.12}$$

donde $\vec{p_{\mathrm{T}}}^{\,j,\mathrm{trk}}$ se refiere a todas las trazas asociadas a jets y $\vec{p_{\mathrm{T}}}^{\,j,\mathrm{calo}}$ es el momento transverso del jet j obtenido en los calorímetros. Aunque la introducción del último término en la Eq. 9.12 puede crear dependencia con el pile-up, la nueva medida $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ aún sigue siendo suficientemente estable ya que contiene la información del PV y además, proporciona una medida más precisa que $E_{\mathrm{T}}^{\mathrm{miss,track}}$ en eventos con partículas neutras en el estado final.

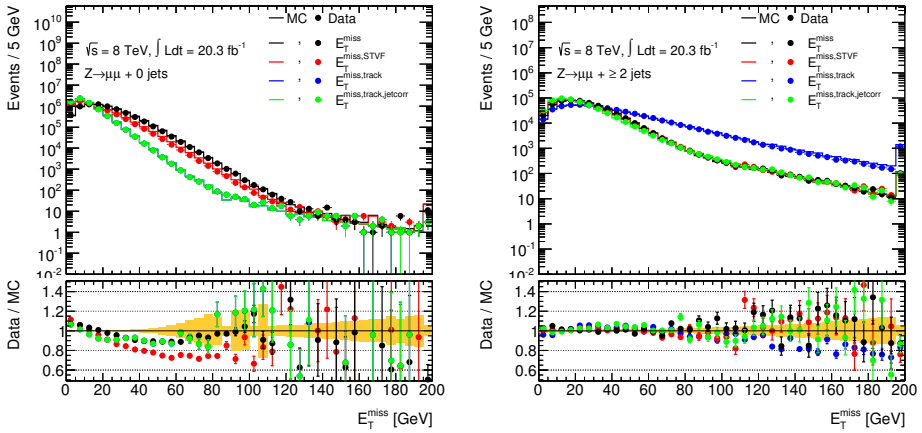

FIGURA 9.31: Distribuciones de las diferentes definiciones de $\not\!\!E_T$ para datos (círculos) y simulación (circunferencias) para eventos consistentes con el proceso $Z \to \mu\mu$. Eventos sin jets (izquierda) y con al menos dos jets (derecha) se muestran separadamente.

Figura 9.31 muestra los espectros para las medidas de $\not\!\!E_T$ realizadas con distintos algoritmos para eventos en datos y simulación consistentes con el proceso $Z \to \mu\mu$. Los resultados separan los eventos en distinto número de jets en el estado final. Para eventos sin jets, la medida de $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ es igual que $E_{\mathrm{T}}^{\mathrm{miss,track}}$, ya que la corrección de jets no se aplica en este caso. En eventos con al menos dos jets reconstruidos en el estado final con $p_{\mathrm{T}} > 25\,\mathrm{GeV}$, se observa que la nueva mediad $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ disminuye considerablemente las colas con respecto a la distribución obtenida para $E_{\mathrm{T}}^{\mathrm{miss,track}}$, estando al nivel de las reconstrucciones $E_{\mathrm{T}}^{\mathrm{miss}}$ y $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$.

La Figura 9.32 muestra la media y anchura de cada medida de $\not\!\!E_T$ en eventos simulados de $Z \to \ell\ell$ en función del número de interacciones por paquete. Los eventos están divididos en distintos número de jets en el estado final. Los resultados muestran la mejora en la medida de $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ con respecto a las definiciones previas, especialmente en eventos con partículas neutras.

Numerosas investigaciones sobre el comportamiento de la $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ han sido desarrolladas en eventos en los que se espera una medida de $\not\!\!E_T$ real debido a la presencia de neutrinos en el estado final. Los estudios de linearidad muestran que la nueva definición de $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ restablece la linearidad de $E_{\mathrm{T}}^{\mathrm{miss,track}}$, proporcionando un valor más cercano al esperado. Además, variables relacionadas con la medida de $\not\!\!E_T$ han sido mejoradas por la introducción de $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$.

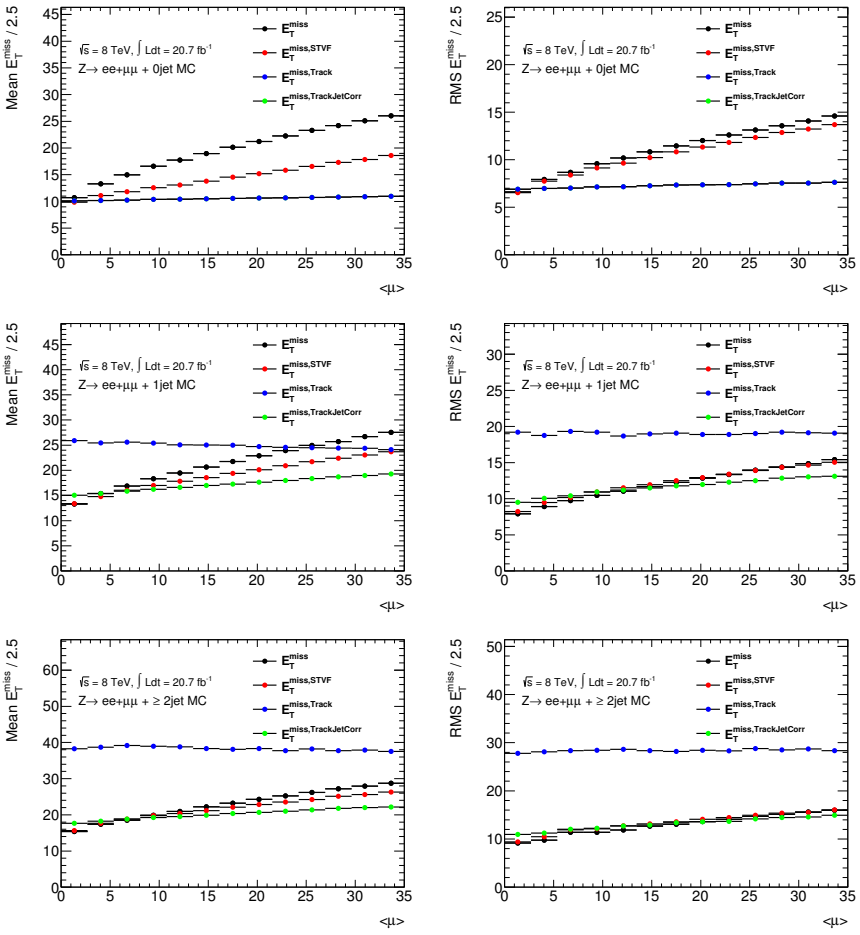

FIGURA 9.32: Valores de la media y RMS para cada definición de $\not{E}_T$ en simulación para eventos del bosón Z usando diferentes número de jets en el estado final: 0 jets (arriba), 1 jet (centro) y ≥ 2 (abajo). Las distribuciones de $p_{\mathrm{T}}^{\mathrm{miss}}$ y $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ son idénticas en eventos sin jets.

La $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ también mejora la resolución de variables como m_{T} y $m_{\tau\tau}$, las cuales son utilizadas en el análisis $H \to WW^{(*)} \to \ell\nu\ell\nu$.

9.5.2 Optimización de la $\not{E}_T$

La estrategia para optimizar el criterio de $\not{E}_T$ en el análisis $H \to WW^{(*)} \to \ell\nu\ell\nu$ se basa en candidatos final simulados por MC los cuales son evaluados a través del ajuste estadístico para determinar su impacto a través de la significancia esperada para cada caso. Debido a que la composición y contribución de los fondos depende del

estado final en el análisis $H \to WW^{(*)} \to \ell\nu\ell\nu$, la optimización de la $\not\!\!E_T$ se evalúa en cada caso. En general para estados finales con $e\mu$+μe, las diferentes reconstrucciones de $\not\!\!E_T$ proporcionan resultados muy similares al final de la selección de eventos. Para todas ellas, el espectro de $\not\!\!E_T$ está prácticamente despoblado en la región de $\not\!\!E_T < 25\,\text{GeV}$ ya que tanto la señal como los fondos remanentes tienen valores esperados para la medida de $\not\!\!E_T$. Figura 9.33 muestra el espectro de $E_{\text{T}}^{\text{miss,track,jetCorr}}$ de cada uno de los fondos y de la señal del bosón de Higgs con $m_{\text{H}} = 125\,\text{GeV}$ para eventos que satisfacen la selección del análisis H+0j con estados finales $e\mu$+μe.

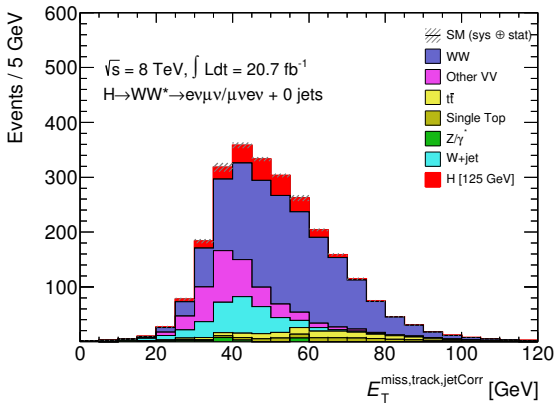

FIGURA 9.33: Distribución de $E_{\text{T}}^{\text{miss,track,jetCorr}}$ obtenido con eventos simulados por MC para cada uno de los fondos y de la señal del bosón de Higgs con $m_{\text{H}} = 125\,\text{GeV}$. Los eventos satisfacen la selección completa del análisis H+0j con estados finales $e\mu$+μe.

Por lo tanto, los resultados obtenidos del ajuste estadístico muestran que no hay un beneficio significativo usando una determinada definición o umbral de $\not\!\!E_T$ ya que cualquier corte suprime tantos eventos de señal como de fondo en semejante porcentaje. Finalmente, se prefiere usar la definición de $E_{\text{T}}^{\text{miss,track,jetCorr}}$ por tener mejor resolución y un valor umbral conservador de 20 GeV se usa para la categorías H+0j y H+1j. Dado que la producción del bosón de Higgs vía VBF está caracterizada por la presencia de quarks, la medida esperada de $\not\!\!E_T$ para la señal es menor que para el modo de producción ggF. Por ello, la aplicación de un corte en $\not\!\!E_T$ suprime más eventos de señal que de fondo, con lo que la significancia de la señal disminuye como muestra la Fig. 9.34. Debido a estos resultados, la nueva estrategia no aplica ningún requisito en $\not\!\!E_T$ en eventos H+2j con estados finales $e\mu$+μe.

Por otro lado, estados finales con leptones de mismo sabor resultan afectados por una gran contribución de Z/γ^* de 4 órdenes de magnitud mayor que la señal,

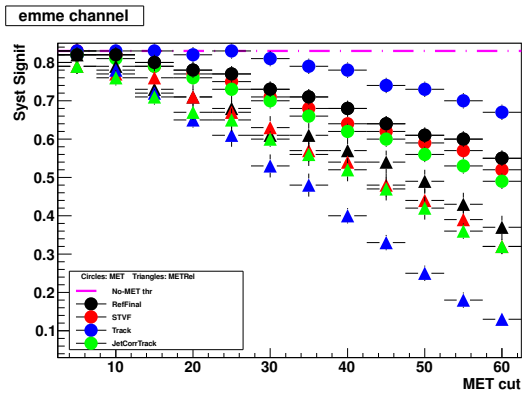

FIGURA 9.34: Significancia esperada para el análisis $H{+}2j$ considerando leptones en el estado final con $e\mu{+}\mu e$ en función de distintos valores usados para el umbral de $\not\!\!E_T$ en pasos de 5 GeV. Las distintas definiciones de $\not\!\!E_T$ son diferenciadas con distintos colores y la línea rosa representa al significancia cuando no se aplica ningún corte en $\not\!\!E_T$.

incluso después de suprimir eventos consistentes con la masa del bosón Z. Este fondo no tiene un valor real de $\not\!\!E_T$ por lo que el requisito que se aplica combina las definiciones de $E_{\mathrm{T}}^{\mathrm{miss}}$ y $E_{\mathrm{T}}^{\mathrm{miss,track}}$ con el fin de reducir este proceso al máximo. Los resultados obtenidos usando $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ concluyen que el poder de supresión de la medida original $E_{\mathrm{T}}^{\mathrm{miss,track}}$ es todavía mayor para estos casos. Esto es debido a que se utilizan las proyecciones de la $\not\!\!E_T$ en la dirección del objeto más cercano ($\not\!\!E_{T,Rel}$) definida como:

$$E_{\mathrm{T,Rel}}^{\mathrm{miss}} = \begin{cases} E_{\mathrm{T}}^{\mathrm{miss}}, \text{ si } \Delta\phi \geqslant \pi/2 \\ \\ E_{\mathrm{T}}^{\mathrm{miss}} \times \sin\Delta\phi, \quad \text{si} \quad \Delta\phi < \pi/2 \end{cases},$$

donde $\Delta\phi$ es el ángulo mínimo entre la dirección de $\not\!\!E_T$ y el leptón o jet más cercanos (obj): $\Delta\phi = \min(|\phi^{\mathrm{miss}} - \phi^{\mathrm{obj}}|)$. Como la definición de $E_{\mathrm{T}}^{\mathrm{miss,track}}$ no considera partículas neutras, la dirección de esta medida tiende a apuntar en la dirección de éstas, lo cual beneficia el cálculo de $\not\!\!E_{T,Rel}$ como se muestra en la Fig. 9.35.

En el caso del análisis de $H{+}2j$ con estado finales $ee{+}\mu\mu$, el cálculo de la $\not\!\!E_{T,Rel}$ puede estar afectado por la gran actividad de jets que caracteriza este proceso. Por este motivo, los resultados apuntan a una mejora de la significancia de la señal combinando cortes en la variables $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ y $E_{\mathrm{T}}^{\mathrm{miss}}$.

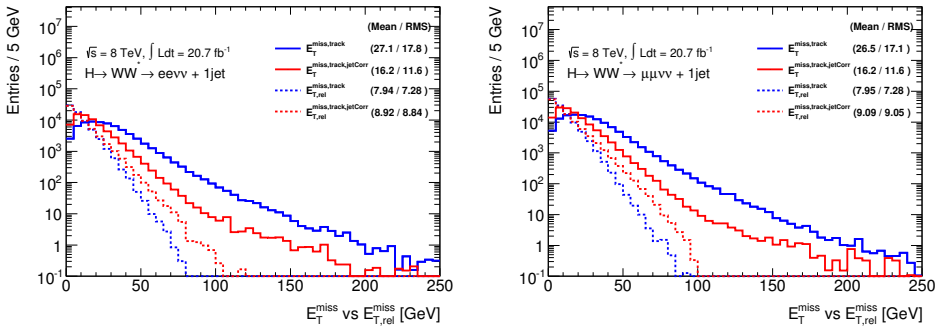

FIGURA 9.35: Comparación de las distribuciones de $E_{\mathrm{T}}^{\mathrm{miss,track}}$ (azul) y $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ (rojo) y sus correspondientes cantidades relativas (líneas discontinuas) para eventos simulados de Z/DY que satisfacen el criterio de selección de $H{+}1j$. En la izquierda y derecha se muestran los estados finales ee y $\mu\mu$, respectivamente.

En general, la introducción de $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ beneficia la búsqueda del bosón de Higgs en el canal de desintegración $H\to WW^{(*)}\to \ell\nu\ell\nu$. Los estudios de optimización aumentan la significancia esperada para la señal de Higgs un 7 % en $H{+}0j$ y $H{+}1j$. Además, la introducción de esta nueva medida en la definición de la m_{T} aumenta el poder de diferenciación del ajuste estadístico, el cual muestra una mejora del $\sim$ 10 % debido a la introducción de $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$. Para el caso del análisis de $H{+}2j$, las mejoras proporcionadas por la optimización de los cortes de $\not\!\!E_T$ y el uso de $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ reflejan un aumento en la significancia del 14 % con respecto a los resultados anteriores.

9.6 Observación de $H\to WW^{(*)}\to \ell\nu\ell\nu$ en ATLAS

9.6.1 Mejoras Introducidas en el Análisis

Con el fin de aumentar la sensibilidad de la búsqueda, un procedimiento completo de optimización del análisis $H\to WW^{(*)}\to \ell\nu\ell\nu$ ha sido desarrollado usando simulación por MC a una energía de centro de masas de 8 TeV. Las mejoras más importantes se basan en la introducción de nuevas variables, como la $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$. Con el fin de incrementar la selección de la señal, se han extendido las regiones definidas por el análisis. Estos estudios concluyen con la disminución de los valores umbrales de la selección , como el p_{T} de los leptones candidatos, y la definición de una nueva región de la señal en el canal $H{+}2j$ para el modo de producción ggF. Por último,

se han mejorado las técnicas de estimación de los distintos fondos en cada uno de los posibles estados finales en los que se divide este análisis.

La significancia esperada después de introducir todos los cambios obtenidos a partir de los diferentes resultados de optimización aumenta de 2,8 σ a 4,36 σ usando únicamente leptones finales $e\mu$+μe en los análisis H+0j y H+1j. Para el análisis del modo de producción vía VBF, la mejora total es del $\sim 70\,\%$ comparando con el resultado anterior, debido a la implantación de técnicas de BDT (en inglés, *Boosted Decision Tree*).

9.6.2 Resultados

La Tab. 9.9 contiene el número de eventos que satisfacen todos los requisitos del análisis $H\to WW^{(*)}\to \ell\nu\ell\nu$ usando datos a 8 TeV y simulación por MC. Se observa un exceso de $\sim$ 500 eventos en datos con respecto a los fondos esperados del SM obtenidos por simulación al final de la selección. Este exceso es compatible con el número de eventos esperados generados por un bosón de Higgs con $m_{\mathrm{H}} = 125\,\mathrm{GeV}$.

CUADRO 9.9: Número de eventos finales observados y esperados para cada estado final consistente con el modo de producción ggF en el análisis $H\to WW^{(*)}\to \ell\nu\ell\nu$ con datos por ATLAS a 8 TeV. N_{sig} y N_{bkg} muestran los eventos esperados para la señal de Higgs y los fondos del SM, respectivamente. Los valores de las incertidumbres incluyen errores teóricos y experimentales.

Selección	Resumen				Composición de N_{bkg}						
	N_{obs}	N_{bkg}	N_{sig}		N_{WW}	N_{top}		N_{misid}		N_{VV}	N_{DY}
			N_{ggF}	N_{VBF}		$N_{t\bar{t}}$	N_t	N_{Wj}	N_{jj}		
$N_{\mathrm{jets}} = 0$	3750	3430 ±90	300 ±50	8 ±4	2250 ±95	112 ±9	195 ±15	360 ±60	16 ±5	420 ±40	78 ±21
$N_{\mathrm{jets}} = 1$	1596	1470 ±40	102 ±26	17 ±5	630 ±50	150 ±10	385 ±20	108 ±20	8.2 ±3,0	143 ±20	51 ±13
$N_{\mathrm{jets}} \geq 2$, $e\mu$+μe ggF	1017	960 ±40	37 ±11	13 ±1,4	138 ±28	56 ±5	480 ±40	54 ±25	62 ±22	56 ±18	117 ±21

Figura 9.36 presenta la distribución de m_{T} para eventos consistentes con el modo de producción del bosón de Higgs vía ggF usando datos tomados por ATLAS durante Run-I.

Para el modo de producción del bosón de Higgs a través de VBF, la Tab. 9.10 muestra el número de eventos finales para datos a 8 TeV y simulación por MC de los distintos procesos del SM y de la señal de Higgs con $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. Los resultados para este modo de producción en el análisis $H\to WW^{(*)}\to \ell\nu\ell\nu$ muestran un exceso de $\sim$ 30 eventos con respecto a los fondos del SM simulados por MC. La

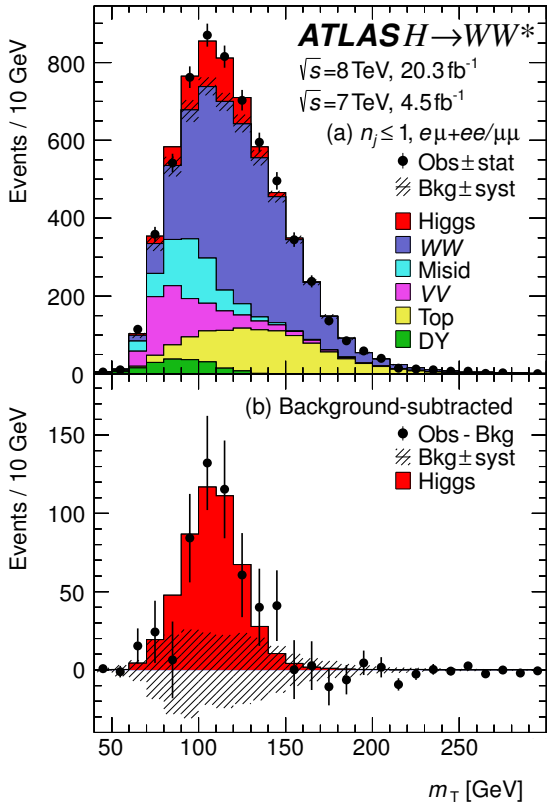

FIGURA 9.36: Distribución de la masa transversa para eventos que satisfacen la selección completa del análisis de $H \to WW^{(*)} \to \ell\nu\ell\nu$ bajo el modo de producción vía ggF usando datos a 7 y 8 TeV y simulación por MC. La distribución de abajo muestra la diferencia de los datos con respecto a los fondos estimados comparando con con la distribución esperada para un bosón de Higgs con $m_{\mathrm{H}} = 125\,\mathrm{GeV}$.

CUADRO 9.10: Número de eventos finales observados y esperados para cada estado final consistente con el modo de producción VBF usando datos de ATLAS a 8 TeV y simulación por MC. N_{sig} y N_{bkg} muestran los eventos esperados para la señal de Higgs y los fondos del SM, respectivamente. Los valores de las incertidumbres incluyen errores teóricos y experimentales.

Selección	Resumen				Composición de N_{bkg}						
	N_{obs}	N_{bkg}	N_{sig}		N_{WW}	N_{top}		N_{misid}		N_{VV}	N_{DY}
			N_{ggF}	N_{VBF}		$N_{t\bar{t}}$	N_t	N_{Wj}	N_{jj}		
$N_{\mathrm{jets}} \geq 2$, VBF	130	99 ±9	7.7 ±2,6	21 ±3	11 ±3,5	5.5 ±0,7	29 ±5	4.7 ±1,4	2.8 ±1,0	4.4 ±0,9	38 ±7

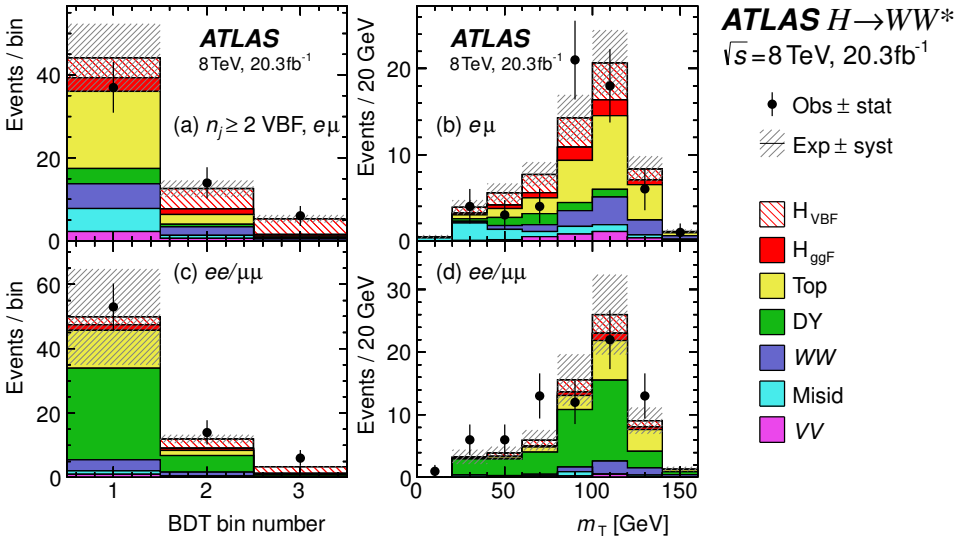

FIGURA 9.37: Distribuciones de BDT (izquierda) y masa transversa (derecha) para eventos que satisfacen la selección completa de la categoría H+2j y son consistentes con el modo de producción VBF. Datos tomados en ATLAS a 8 TeV y simulación son comparados separadamente para los estados finales $e\mu$+μe (arriba) y ee+$\mu\mu$ (abajo), respectivamente.

Fig. 9.37 muestra los resultados obtenidos con el procedimiento de BDT (O_{BDT}) y las distribuciones de m_{T} para los eventos candidatos.

Los resultados obtenidos del ajuste estadístico se muestran en la Fig. 9.38 en función de la masa del bosón de Higgs. Se observa un exceso de eventos en datos sobre los fondos esperados del SM de 6,1 σ en el canal de desintegración $H\to WW^{(*)}\to \ell\nu\ell\nu$. Este resultado es consistente con el valor esperado de 5,8 σ procedente de la simulación por MC de un bosón de Higgs de $m_{\mathrm{H}} = 125{,}36\,\mathrm{GeV}$.

Los resultados proporcionados por el test estadístico en función del cociente $\mu_{\mathrm{VBF}}/\mu_{\mathrm{ggF}}$ para $m_{\mathrm{H}} = 125{,}36\,\mathrm{GeV}$ se muestran en la Fig. 9.39. El valor obtenido bajo la hipótesis $\mu_{\mathrm{VBF}}/\mu_{\mathrm{ggF}} = 0$ proporciona la significancia de la producción del bosón de Higgs a través del modo de producción VBF que es de 3,2 σ. El cociente del valor medido con respecto al valor esperado de la sección eficaz de producción del bosón de Higgs con $m_{\mathrm{H}} = 125{,}36\,\mathrm{GeV}$ es $1{,}09\,^{+0,16}_{-0,15}\,(\mathrm{stat.})\,^{+0,17}_{-0,14}(\mathrm{syst.})$. Los correspondientes valores para los mecanismos de producción ggF y VBF son $1{,}02 \pm 0{,}19\,(\mathrm{stat.})\,^{+0,22}_{-0,18}(\mathrm{syst.})$ y $1{,}27\,^{+0,44}_{-0,40}\,(\mathrm{stat.})\,^{+0,21}_{-0,30}(\mathrm{syst.})$, respectivamente. Las observaciones son consistentes con los valores esperados del bosón de Higgs del SM

confirmando la primera observación del bosón de Higgs decayendo a un par de bosones W en el detector ATLAS.

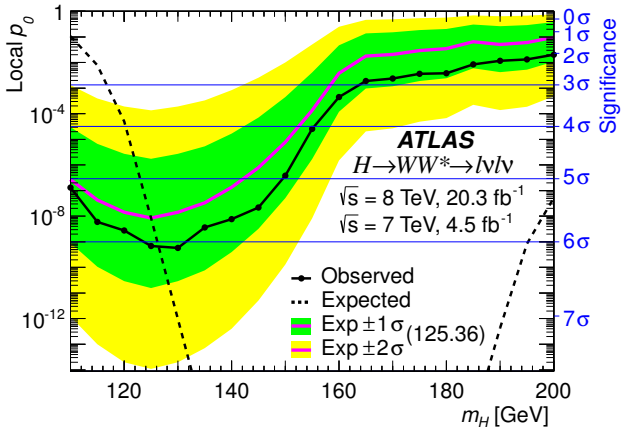

FIGURA 9.38: Valores observados y esperados del valor de p_0 en función de m_{H}. Los valores observados se representan mediante las líneas sólidas con puntos, en los que p_0 es evaluado. La línea punteada muestra los valores esperados debidos a la presencia de la señal para hipótesis de masa. Los valores esperados para $m_{\mathrm{H}} = 125{,}36\,\mathrm{GeV}$ son representados por las línea sólida sin puntos. Las bandas interna y externa representan la incertidumbre dentro de una y dos desviaciones estándar, respectivamente.

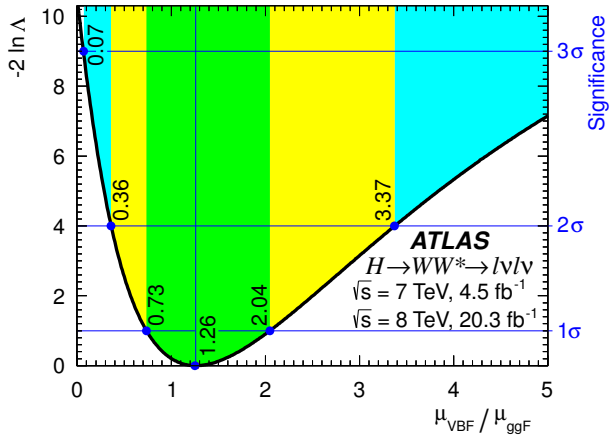

FIGURA 9.39: Resultados del test estadístico en función del cociente $\mu_{\mathrm{VBF}}/\mu_{\mathrm{ggF}}$ para $m_{\mathrm{H}} = 125{,}36\,\mathrm{GeV}$. Las bandas interna, media y externa representan las incertidumbres con una, dos y tres desviaciones estándar con respecto al valor central representado por la línea vertical.

Appendix A

Topocluster Performance Using Different Configurations

This appendix firstly contains the energetic performance for different Topological Clustering algorithm configurations using randomly triggered events. This is covered in Section A.1. The topoclusters is also evaluated through the topoclusters moments and their related-quantities: longitudinal and lateral. They are defined and used as performance testers in Section A.2. These investigations prove that the two-Gaussian description for the TileCal noise constants and the proper treatment of the hot spots, reduce significantly the tails in the MET distribution.

A.1 Energetic Performance

Figure A.1 shows the topoclusters multiplicity distributions for different configurations of (s_s, s_n, s_c) using 2008 randomly triggered cosmic events collected by the ATLAS detector.

The mean values of the number of topoclusters (N_{topo}) for each (s_s,s_n,s_c) configuration are presented in Tab. A.1. It is clear that the dominant threshold in topocluster multiplicity is s_s and differences observed for different s_n and s_c values are relatively small. Furthermore, the $s_n = 2$ and $s_c = 0$ are optimal values as the object formation using these thresholds in general give the minimum number of reconstructed topoclusters.

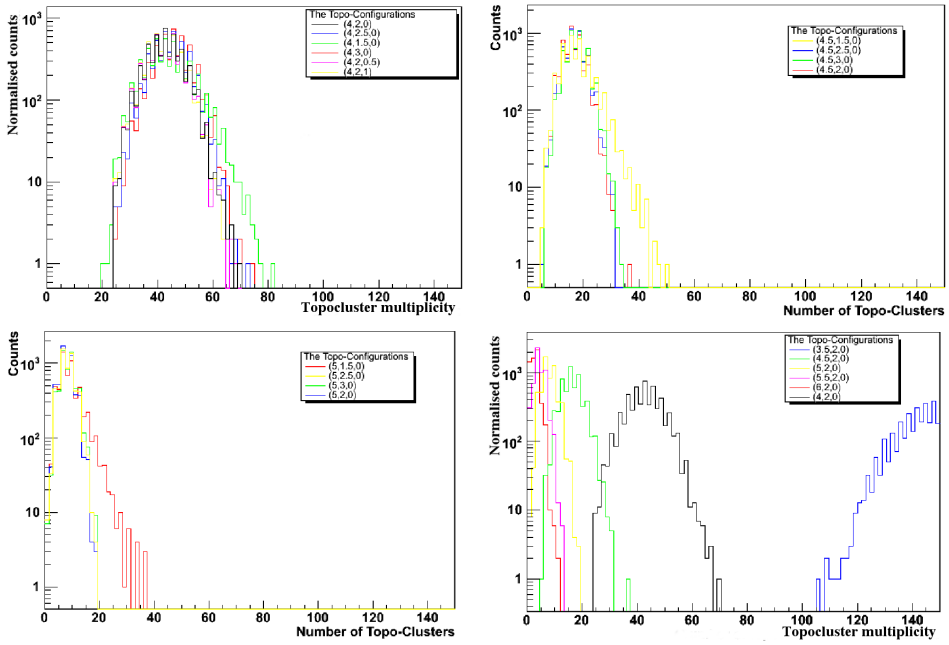

FIGURE A.1: Topocluster multiplicity distributions for different configurations obtained from 2008 ATLAS random triggered events.

$s_s = 4$	N_{topo}	$s_s = 4.5$	N_{topo}	$s_s = 5$	N_{topo}	$s_n = 2$	N_{topo}
(4, 1.5, 0)	43.53	(4.5, 1.5, 0)	17.64	(5, 1.5, 0)	9.02	(3, 2, 0)	> 150
(4, 2, 0)	42.57	(4.5, 2, 0)	16.12	(5, 2, 0)	7.77	(3.5, 2, 0)	140.28
(4, 2.5, 0)	44.24	(4.5, 2.5, 0)	16.86	(5, 2.5, 0)	8.05	(4, 2, 0)	42.57
(4, 3, 0)	45.41	(4.5, 3, 0)	17.36	(5, 3, 0)	8.25	(4.5, 2, 0)	16.12
(4, 2, 0.5)	42.38					(5, 2, 0)	7.77
(4, 2, 1)	42.05					(5.5, 2, 0)	4.30
						(6, 2, 0)	2.54

TABLE A.1: Different topocluster configurations expressed as (s_s, s_n, s_c) -odd columns- and the multiplicity of topoclusters (N_{topo}) -even columns- for each case. Random triggered cosmic events collected by the ATLAS detector during 2008 are used.

The total energy of the topoclusters defined as the sum of the energy of each of the cells forming the topocluster is comparing using different thresholds values in Fig. A.2. The distributions show an asymmetrical tendency which is mainly correlated with the seed's threshold value. The total transverse energy of the topoclusters defined as the sum of the energy of each of the cells forming a topocluster is comparing using different thresholds values in Fig. A.3.

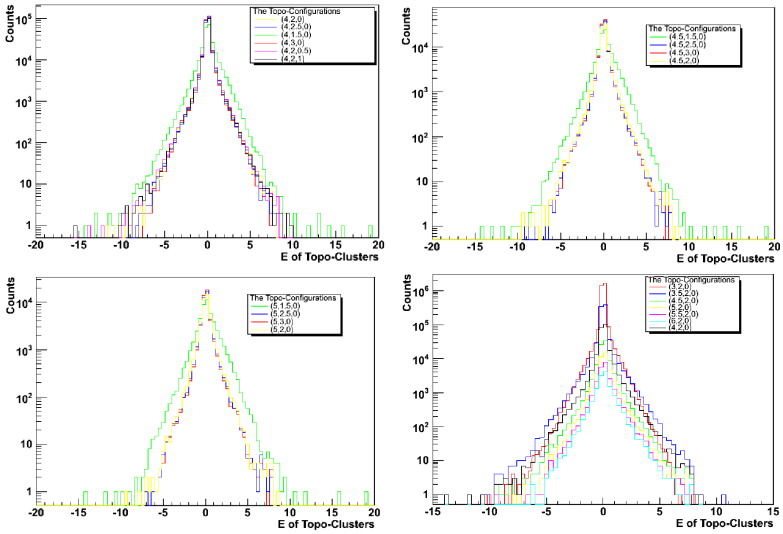

FIGURE A.2: Energy distributions for different configurations of topoclusters obtained from randomly triggered cosmic events collected by the ATLAS detector during 2008. The distributions are normalised to the number of entries of the (4,2,0) configuration and the energy is given in GeV.

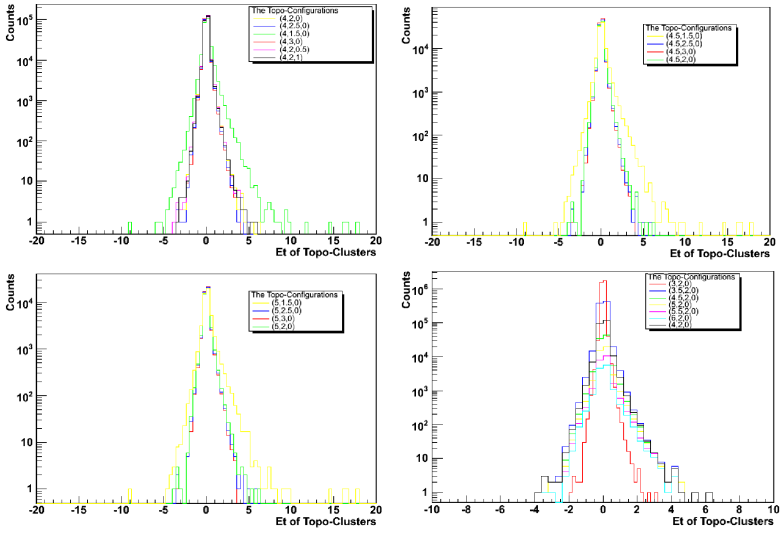

FIGURE A.3: Transverse energy for different configurations of topoclusters obtained from randomly triggered cosmic events collected by the ATLAS detector during 2008. The distributions are normalised to the number of entries of the (4,2,0) configuration and the energy is given in GeV.

A.2 Topoclusters Moments

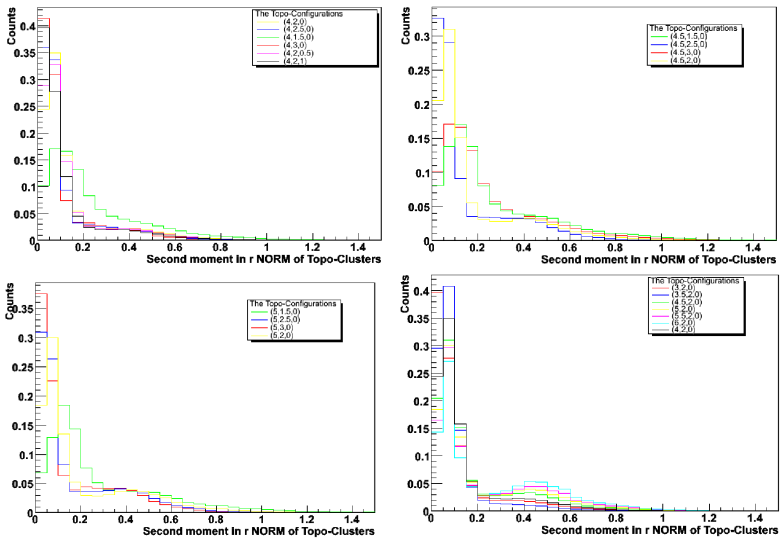

FIGURE A.4: Second moment in r for different configurations of topoclusters obtained from randomly triggered cosmic events collected by the ATLAS detector during 2008. The distributions are normalised to the unity.

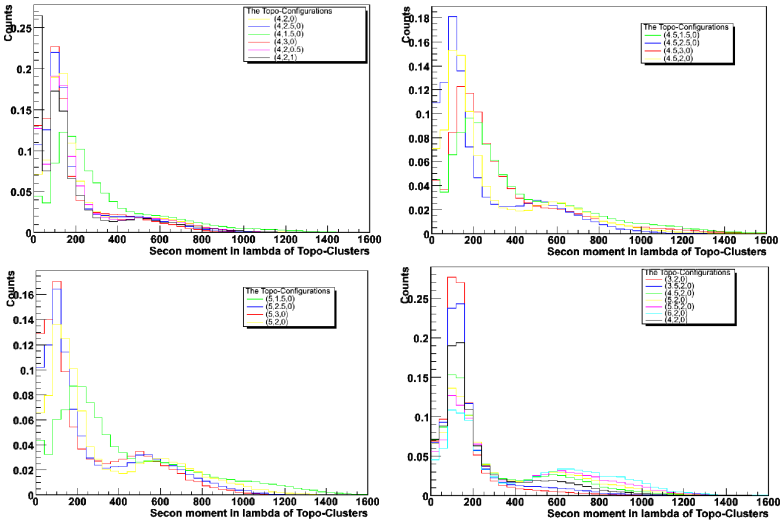

FIGURE A.5: Second moment in λ for different configurations of topoclusters obtained from randomly triggered cosmics events collected by the ATLAS detector during 2008. The distributions are normalised to the unity.

Figure A.4 and Fig. A.5 show the second moment in r and λ, respectively, using different configurations of topoclusters obtained from randomly triggered cosmic events collected by the ATLAS detector during 2008. Figure A.6 shows the contribution of the topoclusters-cells located in different barrels and layers of the TileCal detector with respect to the total energy measured for the topocluster.

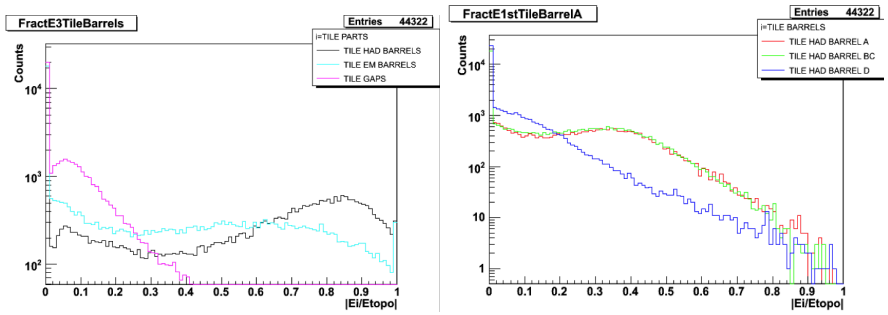

FIGURE A.6: Energy fraction in each of the barrels (left) and layers (right) of the TileCal detector.

In order to measure the contribution of the most energetic cells in the topocluster shape, a new set of quantities are defined: the normalised second lateral and longitudinal moments. These variables are given by the expressions,

$$\text{lateral} = \frac{< r^2 >_{\text{out}}}{< r^2 >_{\text{out}} + < r^2 >_{\text{core}}}, \tag{A.1}$$

$$\text{longitudinal} = \frac{< \lambda^2 >_{\text{out}}}{< \lambda^2 >_{\text{out}} + < \lambda^2 >_{\text{core}}}, \tag{A.2}$$

where the magnitudes $< r^2 >_{\text{out}}$, $< r^2 >_{\text{core}}$, $< \lambda^2 >_{\text{out}}$, and $< \lambda^2 >_{\text{core}}$ are defined as follows,

- $< r^2 >_{\text{out}} = < r^2 >$, with $r = 0\,\text{mm}$ for the two most energetic cells.
- $< \lambda^2 >_{\text{out}} = < \lambda^2 >$, with $\lambda = 0\,\text{mm}$ for the two most energetic cells.
- $< r^2 >_{\text{core}} = < r^2 >$, with $r = 40\,\text{mm}$ for the two most energetic cells and $r = 0\,\text{mm}$ for all other cells.
- $< \lambda^2 >_{\text{core}} = < \lambda^2 >$, with $\lambda = 100\,\text{mm}$ for the two most energetic cells and $\lambda = 0\,\text{mm}$ for all other cells.

Equation A.1 and Eq. A.2 give normalised distributions between 0 and 1 for the two moments. The fixed values for r and λ for the two most energetic cells in a topocluster were obtained from simulation.

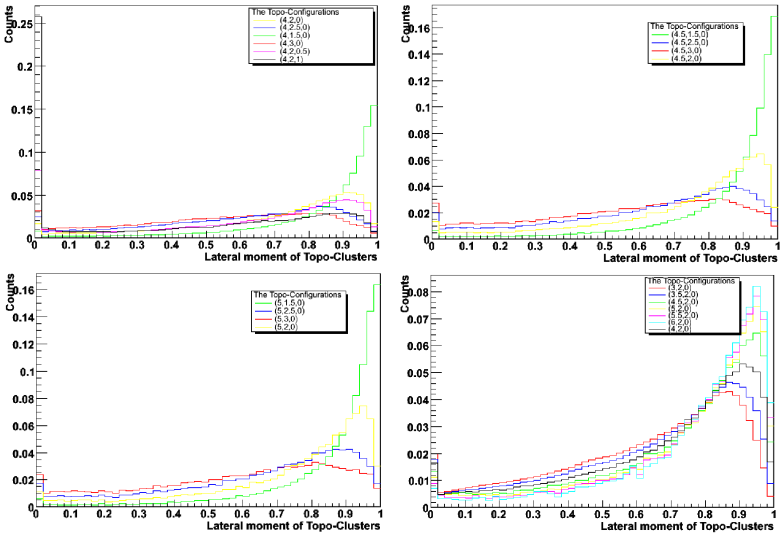

FIGURE A.7: Normalised second lateral moment for different configurations of topoclusters obtained from random triggered cosmic events collected by the ATLAS detector during 2008. Distributions are normalised to the unity.

Figure A.7 and Fig. A.8 show the normalised second lateral and longitudinal moments for different configurations of topoclusters, respectively.

A.2.1 TileCal Contribution to Large Topoclusters

The cell multiplicity in large topoclusters for each ATLAS calorimeter component is illustrated in Fig. A.9. The distribution shows the tendency of large topoclusters in TileCal to contain more cells than EM calorimeter on average. The distribution on the right shows that the energetic contribution from TileCal cells is dominant in large topoclusters.

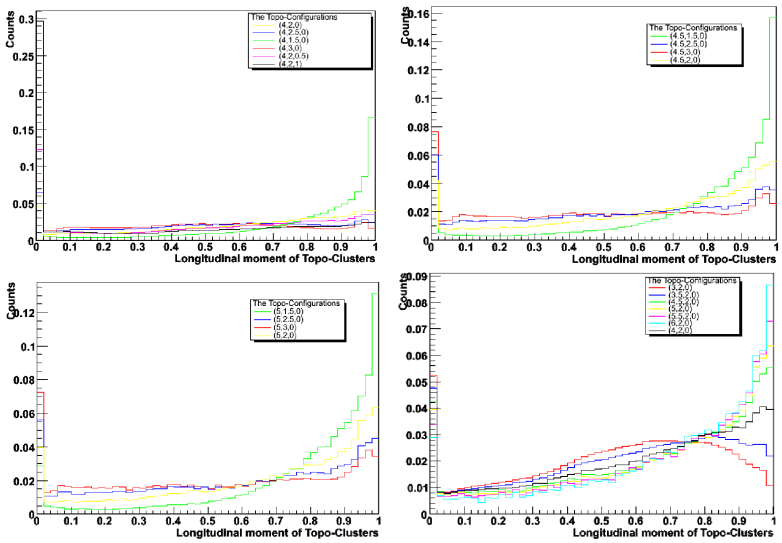

FIGURE A.8: Normalised second longitudinal moment for different configurations of topoclusters obtained from random triggered cosmic events collected by the ATLAS detector during 2008. Distributions are normalised to the unity.

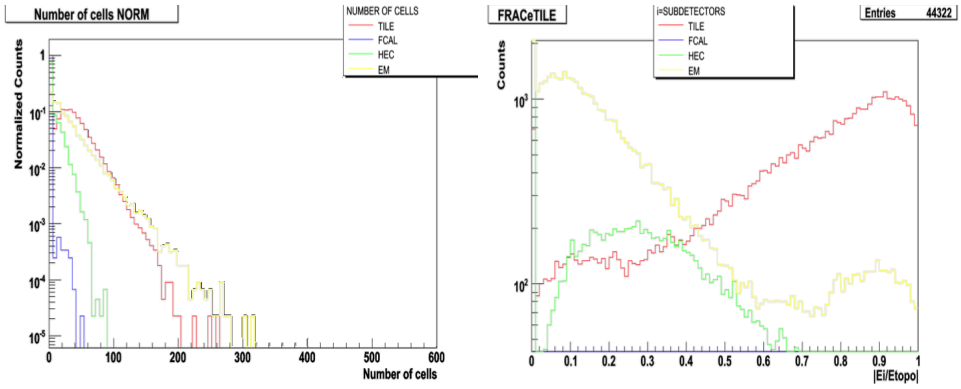

FIGURE A.9: Cell multiplicity from topoclusters with $< \lambda^2 > \in [400, 1200]$ in each of the calorimeters of ATLAS (left) and energetic contribution of each calorimeter to the total topocluster energy (right).

Appendix B

Missing Transverse Momentum Measurements in ATLAS

This appendix firstly contains the selection criteria applied to the $E_{\mathrm{T}}^{\mathrm{miss}}$ input objects definition in Section 4.2

B.1 Details on $E_{\mathrm{T}}^{\mathrm{miss}}$ Reconstruction

In the $E_{\mathrm{T}}^{\mathrm{miss}}$ definition given by Eq. 4.1, calorimeter cells associated to reconstructed physics object are calibrated according to the corresponding physics object. Moreover, the reconstructed energy for each $E_{\mathrm{T}}^{\mathrm{miss}}$ component is scaled individually. The calibration scheme used is the one yielding the best performance in 2010 data, which is described in full detail in Ref. [59] and summarised below. Electrons are calibrated with the default ATLAS electron calibration [154] and photons are used at the electromagnetic scale (EM). The τ-jets, from hadronically decaying τ-leptons, are calibrated with the local cluster weighting (LCW) [155] which involves classifying the energy depositions as electromagnetic or hadronic to weighting them appropriately when computing the topocluster energy. An offset is subtracted to suppress the pile-up effects and the tau energy scale (TES) correction [156] is applied. The jets are reconstructed with the anti-k_t algorithm [114], with distance parameter $R = 0.4$. They are calibrated with the LCW scheme if $10 < p_{\mathrm{T}} < 20\,\mathrm{GeV}$ (known as soft jets, which contribution enters in the soft term) and with the LCW+JES scheme,

where JES is the jet energy scale [137], if $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ (which contribute to the jet term). The contribution from topoclusters not associated to high-p_{T} objects is calculated with LCW calibration combined with tracking information. Finally, the muon term is calculated from the momenta of muon tracks reconstructed with $|\eta| < 2.5$. Only well-reconstructed muons in the MS with a matched track in the ID are considered (combined muons). Outside the ID volume acceptance ($|\eta| > 2.5$), only the momenta from the MS is used for the muon term.

TABLE B.1: The contributions to $E_{\mathrm{T}}^{\mathrm{miss}}$ in Eq. 4.1 from electrons and positrons ($e^{\pm}$), photons (γ), tau leptons ($\tau^{\pm}$), muons ($\mu^{\pm}$), and particle jets. The table is ordered descending in priority for consideration in $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, with the 1 being the highest priority.

Priority	Term	p_{T} [GeV] threshold	Object selection	Calibration and scale	Contribution
1	e	10	ATLAS electron identification: medium++	ATLAS electron calibration	$e^{\pm}$ with reconstruction quality and kinematic criteria
2	γ	10	ATLAS photon identification: tight	EM	γ with reconstruction quality and kinematic cuts, without overlap with 1
3	τ	20	τ-jets from hadronically τ-leptons decays	LCW+TES	$\tau^{\pm}$ with reconstruction quality and kinematic cuts, without overlar with 1 and 2
4	jets	20	Anti-k_t jet algorithm with $R = 0.4$	LCW+JES	Jets with reconstruction quality and kinematic cuts, without overlap with $1-3$
5	SoftTerm	10	Topoclustering algorithm (4,2,0)	LCW	Topoclusters and tracks not associated to high-p_{T} objects without overlap between them
6	μ	6	ID tracks associated to MS ($\|\eta < 2.5\|$)	Combined	$\mu^{\pm}$ with reconstruction quality and kinematic cuts, with adjusted corrections for calorimetric deposits when relevant: (*) muons overlapping with jets
			MS tracks ($2.5 < \|\eta\| < 2.7$)	MS	
			Topoclusters ($\|\eta\| < 2.7$) (*)	LCW	

Table B.1 summarises the selection criteria applied for the high p_{T} object-related terms contained in Eq. 4.1.

B.2 Details on $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ Reconstruction

Tracks entering in the STVF correction, given in Eq. 4.5, should be originated in the PV but not be associated to any high-p_{T} objects accounted in the other $E_{\mathrm{T}}^{\mathrm{miss}}$ terms.

A track is considered to be associated with a hard object when it is the source of the kinematics of a reconstructed particle, or it is used for particle identification, or it overlaps with calorimeter energetic signals representing the particle. These tracks are removed by spacial overlap with the objects defined in Section 4.2. The association to the PV requires perpendicular impact parameter (d_0): $|d_0| < 2\,\text{mm}$ and a longitudinal impact parameter (z_0) with $|z_0 \times \sin\theta| < 2\,\text{mm}$, both with respect to the beam axis. The θ represents the polar angle of the track. Tracks should have $p_{\text{T}}^{\text{track}} > 400\,\text{GeV}$ and a number of hits in the ID system high enough to ensure good reconstruction.

B.3 Details on $E_{\text{T}}^{\text{miss,track}}$ Reconstruction

Tracks considered in the $E_{\text{T}}^{\text{miss,track}}$ calculation need to satisfy a number of requirements in order to ensure a good momentum measurement, an efficient rejection of mis-reconstructed tracks and a very good track to PV association. ? Events must have at least one track satisfying the following requirements:

- $p_{\text{T}} > 500\,\text{MeV}$
- $|\eta| < 2.5$
- at least 1 pixel detector hit
- at least 6 SCT hits
- Transverse impact parameter with respect to the PV: $|d_0| < 1.5\,\text{mm}$
- Longitudinal impact parameters with respect to the PV: $|z_0 \times \sin(\theta)| < 1.5\,\text{mm}$

Besides the good job of the ID system and the high quality requirements listed above, two effects are considered for the mis-reconstructed tracks and for tracks associated to a high-p_{T} object which do not pass the selection. These cases are corrected as follows for improvising the $E_{\text{T}}^{\text{miss,track}}$ reconstruction.

Tracks selected as described above can still have their momentum badly reconstructed. This effect arises from low-p_{T} tracks interacting with the ID material. Those may produce a non-negligible number of secondary particles that leave enough hits in the pixel/SCT detectors to be reconstructed with a much higher momentum. In order to reduce the number of these mis-reconstructed tracks, the following selections are required to isolated tracks:

- the relative uncertainty on the charge to track-momentum ratio (q/p) should be:

$$\frac{\sigma(q/p)}{(q/p)} < 0.4 \tag{B.1}$$

- the energy in calorimeter clusters in a cone of 0.1 radius, in the (η, ϕ) plane, ($E_{0.1}^{\text{calo}}$) around the track should satisfy:

$$\frac{E_{0.1}^{\text{calo}}}{p_{\text{T}}^{\text{track}}} > 0.65\,, \tag{B.2}$$

  to reflect the reconstructed track momentum

$3-4\%$ of the tracks associated to reconstructed electrons and muons fail some of the high quality requirements listed above. The effect is a $E_{\text{T}}^{\text{miss,track}}$ mis-reconstruction since these tracks are not entering in its computation. In order to conserve the physics content of the event, all tracks associated to electrons and muons are used in the $E_{\text{T}}^{\text{miss,track}}$ calculation whether or not they pass the high quality criteria. The selection of the leptons for track association follows the requirements applied for computing the electron and muon terms in $E_{\text{T}}^{\text{miss}}$ calculation, given in Section 4.2.

In 2012, motivated by the increase on pileup events, several studies requiring tighter selection for the tracks showed that they only slightly reduce the number of tracks included in the $E_{\text{T}}^{\text{miss,track}}$ calculation, but make negligible differences overall to its performance. However, the performance improved using the calorimeter energetic reconstruction for electrons instead of the p_{T} of their associated ID tracks. This is due to the reconstruction of the electron energy deposited in the calorimeter system takes into account energetic effects, such as the loose for bremstrahlung radiation, while the ID reconstruction does not. In this light, all selected tracks are replaced by the calorimeter energy reconstruction of the electrons from which they are associated in the $E_{\text{T}}^{\text{miss,track}}$ calculation.

B.4 $\not{E}_T$ Comparisons in $Z \to \ell\ell$ Enriched Region

These events have to satisfy a criterion for being compatible with the $Z \to \ell\ell$ process. The selection requires exactly two isolated high-p_{T} leptons with opposite sign and same flavour in the final state. Moreover, the invariant mass of the two leptons ($m_{\ell\ell}$) can not differ from the mass of the Z boson in more than 25 GeV.

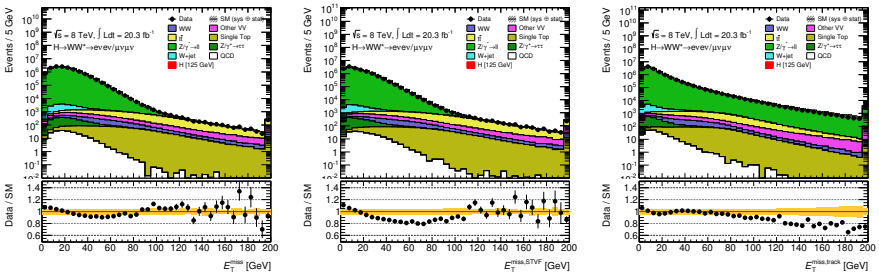

FIGURE B.1: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ distributions for $Z \to \ell\ell$ events in 8 TeV ATLAS data and MC simulation for $ee + \mu\mu$ final states and inclusive number of jets.

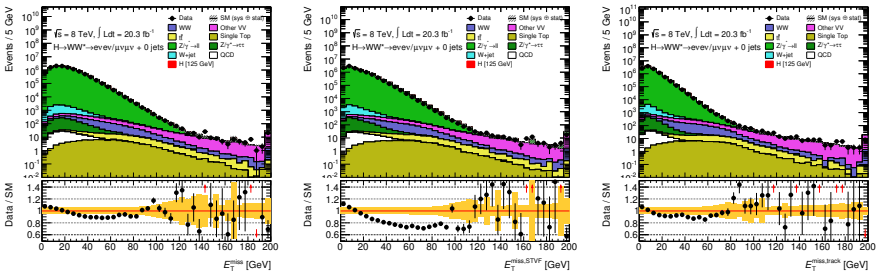

FIGURE B.2: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ distributions for $Z \to \ell\ell$ events in 8 TeV ATLAS data and MC simulation for $ee + \mu\mu$ final states with 0 jets.

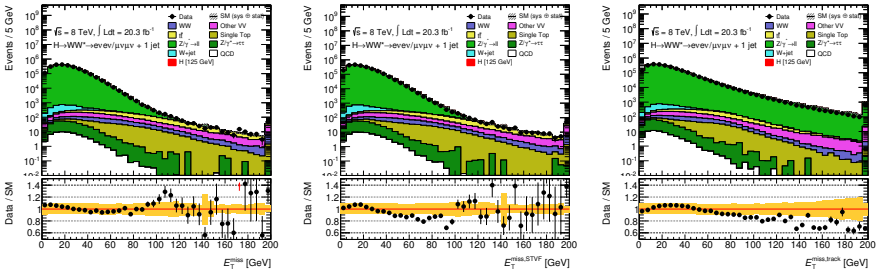

FIGURE B.3: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ distributions for $Z \to \ell\ell$ events in 8 TeV ATLAS data and MC simulation for $ee + \mu\mu$ final states with exactly one jet.

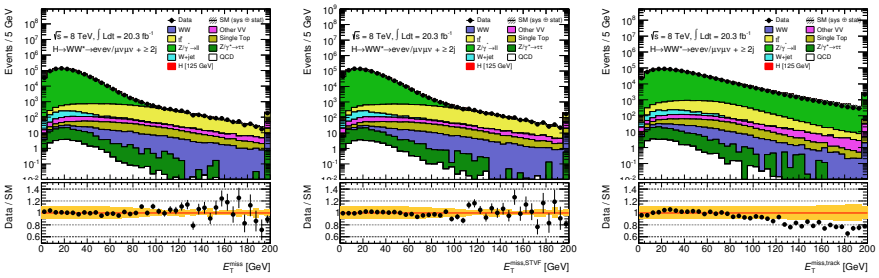

FIGURE B.4: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ distributions for $Z \to \ell\ell$ events in 8 TeV ATLAS data and MC simulation for $ee + \mu\mu$ final states and inclusive number of jets with at least two jets.

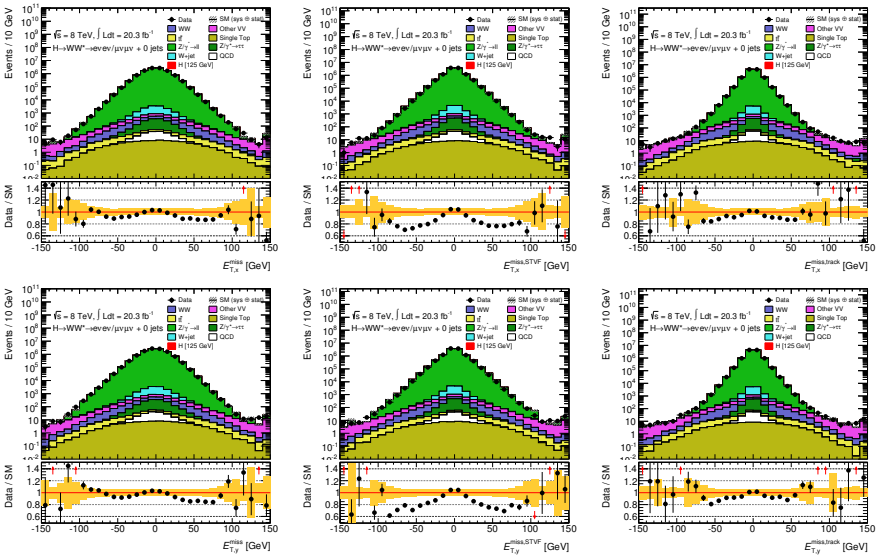

FIGURE B.5: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ x and y components for $Z \to \ell\ell$ events in 8 TeV ATLAS data and MC simulation for $ee + \mu\mu$ final states with 0 jets.

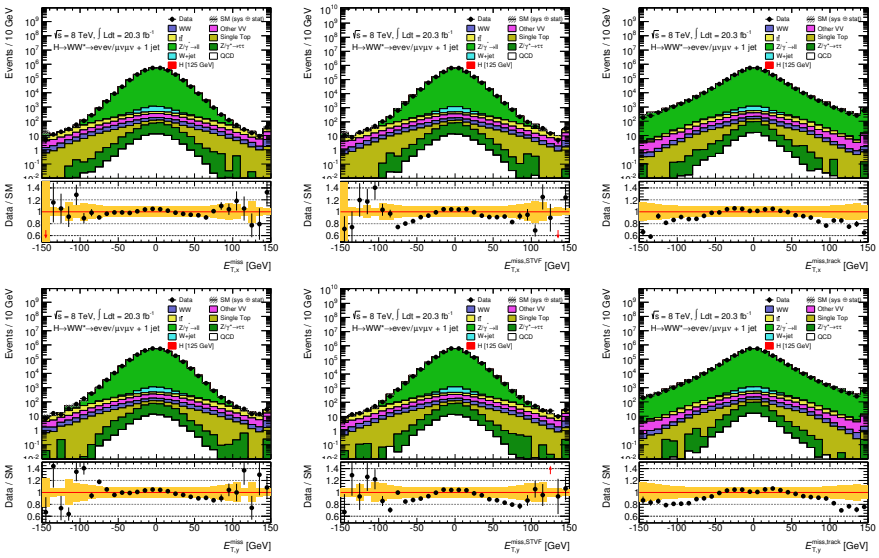

FIGURE B.6: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ x and y components for $Z \to \ell\ell$ events in 8 TeV ATLAS data and MC simulation for $ee + \mu\mu$ final states and inclusive number of jets with exactly 1 jet.

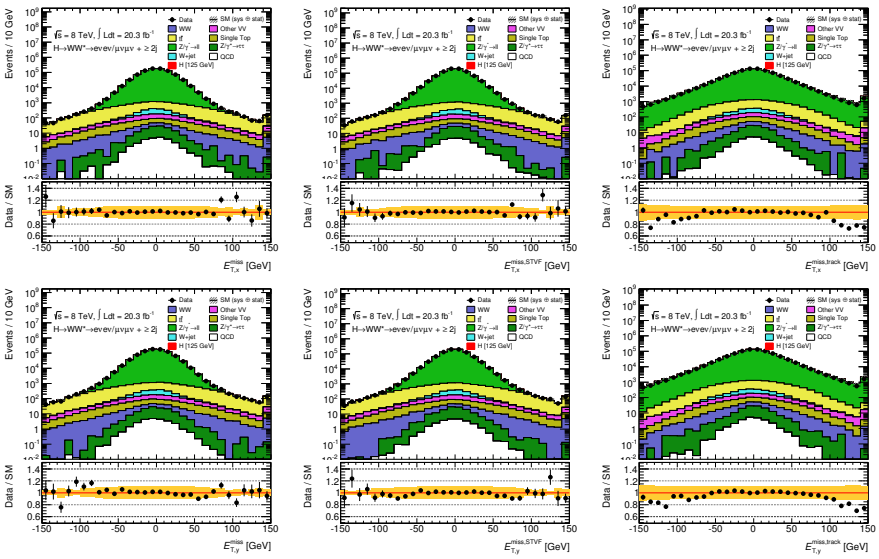

FIGURE B.7: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track}}$ distributions for $Z \to \ell\ell$ events in 8 TeV ATLAS data and MC simulation for $ee + \mu\mu$ final states with at least two jets.

Appendix C

$H \to WW^{(*)} \to \ell\nu\ell\nu$ Event Yields for 7 TeV Data

Table C.1 shows the expected event yield for the Higgs boson, with $m_H = 125\,\text{GeV}$, the background estimation and the number of observed events at each state of the pre-selection requirements and the E_T^{miss} for all lepton channels combined using the 2011 data.

TABLE C.1: Observed and expected event yields after the pre-selection and E_T^{miss} requirements in the 2011 period are shown for all lepton channels combined. The signal is shown for $m_H = 125\,\text{GeV}$ Higgs boson. All the MC predictions are from simulation except the $W+$ jets background, which is estimated entirely from data. Only statistical uncertainties associated with the numbers of events in the MC samples are shown.

Selection	N_{obs}	N_{bkg}	N_{sig}	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\text{jets}}$
lepton p_T	2740091	2720000 ± 1000	159.2 ± 0.4	3625 ± 7	1715 ± 9	14660 ± 30	1558 ± 10	2691000 ± 1000	8110 ± 50
OS leptons	2734190	2712000 ± 1000	156.9 ± 0.4	3611 ± 6	1238 ± 7	14590 ± 30	1550 ± 10	2684000 ± 1000	6600 ± 50
$m_{\ell\ell} > 12, 10$	2727098	2708000 ± 1000	154.9 ± 0.4	3606 ± 6	1230 ± 7	14580 ± 30	1550 ± 10	2682000 ± 1000	5310 ± 50
Z veto	278330	265300 ± 400	153.0 ± 0.4	3266 ± 6	583 ± 6	14020 ± 30	1490 ± 10	244200 ± 400	1700 ± 30
E_T^{miss} cuts	15068	14990 ± 50	93.2 ± 0.3	1897 ± 5	210 ± 4	7440 ± 20	860 ± 10	4320 ± 50	266 ± 7

Table C.2 shows the expected signal and background yields, and the number of observed events in each cut stage of the $H+0j$ analysis using the 7 TeV data.

Table C.3 shows the expected signal and background yields, and the number of observed events in each cut stage of the $H+1j$ analysis using the 7 TeV data.

The expected signal and background yields, and the number of observed events in each cut stage of the $H+2j$ analysis using the 7 TeV data are shown in Tab. C.4.

TABLE C.2: Selection table for H+0j in 7 TeV data. The observed, $N_{\rm obs}$, and expected, $N_{\rm exp}$, yields for the signal, $N_{\rm sig}$, and background, $N_{\rm bkg}$, processes are shown for all channels combined: ee+$\mu\mu$ + $e\mu$+μe. The composition of $N_{\rm bkg}$ is given on the right. The requirements are imposed sequentially from top to bottom. Energies, masses, and momenta are in units of GeV. All uncertainties are statistical.

Selection	$N_{\rm obs}$	$N_{\rm bkg}$	$N_{\rm sig}$	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\rm jets}$
H+0j	5231	4970 ± 40	60.1 ± 0.2	1330 ± 4	112 ± 3	180 ± 3	92 ± 3	3090 ± 40	164 ± 5
$\Delta\phi_{ll,E_{\rm T}^{\rm miss}} > \frac{\pi}{2}$	4123	3940 ± 30	59.8 ± 0.2	1318 ± 4	109 ± 3	165 ± 3	90 ± 3	2110 ± 30	152 ± 4
$p_{\rm T}^{\ell\ell} > 30, 40$	1774	1730 ± 10	51.9 ± 0.2	1084 ± 4	87 ± 2	148 ± 3	80 ± 3	230 ± 10	100 ± 3
$m_{\ell\ell} < 50$	543	532 ± 8	43.2 ± 0.2	291 ± 2	36 ± 2	27 ± 1	18 ± 1	125 ± 7	35 ± 2
$\Delta\phi_{\ell\ell} < 1.8$	532	521 ± 8	42.2 ± 0.2	285 ± 2	35 ± 2	27 ± 1	18 ± 1	125 ± 7	32 ± 2

TABLE C.3: Selection table for H+1j in 7 TeV data. The observed, $N_{\rm obs}$, and expected, $N_{\rm exp}$, yields for the signal, $N_{\rm sig}$, and background, $N_{\rm bkg}$, processes are shown for all channels combined: ee+$\mu\mu$ + $e\mu$+μe. The composition of $N_{\rm bkg}$ is given on the right. The requirements are imposed sequentially from top to bottom. Energies, masses, and momenta are in units of GeV. All uncertainties are statistical.

Selection	$N_{\rm obs}$	$N_{\rm bkg}$	$N_{\rm sig}$	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\rm jets}$
H+1j	2721	2631 ± 20	20.9 ± 0.1	307 ± 2	64 ± 2	1357 ± 8	397 ± 6	447 ± 14	59 ± 4
$N_{b\text{-jet}} = 0$	1270	1150 ± 10	18.3 ± 0.1	270 ± 2	56 ± 2	304 ± 4	99 ± 3	370 ± 10	45 ± 3
$m_{\ell\ell} < 50$	323	271 ± 6	14.1 ± 0.1	57.4 ± 0.7	17 ± 1	54 ± 2	21 ± 1	108 ± 6	13 ± 2
$\Delta\phi_{\ell\ell} < 1.8$	205	186 ± 5	12.6 ± 0.1	52.5 ± 0.7	15 ± 1	50 ± 2	20 ± 1	39 ± 4	11 ± 1

TABLE C.4: Selection table for H+2j in 7 TeV data. The observed, $N_{\rm obs}$, and expected, $N_{\rm exp}$, yields for the signal, $N_{\rm sig,VBF}$, and background, $N_{\rm bkg}$, processes are shown for all channels combined: ee+$\mu\mu$ + $e\mu$+μe. In this table, the $N_{\rm sig,ggF}$ is included in $N_{\rm bkg}$; the $N_{\rm sig,VH}$ is included in $N_{\rm sig,VBF}$, but the contributions are negligible after the VBF-related criteria. The composition of $N_{\rm bkg}$ is given on the right. The requirements are imposed sequentially from top to bottom. Energies, masses, and momenta are in units of GeV. All uncertainties are statistical.

Selection	$N_{\rm obs}$	$N_{\rm bkg}$	$N_{\rm sig,VBF}$	$N_{\rm sig,ggF}$	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\rm jets}$
H+2j	11546	11600 ±100	8.5± 0.1	10.1± 0.1	218 ± 2	66± 3	9700 ±100	550 ± 10	990 ±10	79 ± 6
$N_{b\text{-jet}} = 0$	1568	1560 ±20	6.1 ± 0.1	7.3± 0.1	158± 1	49± 2	540 ±10	67± 3	710±10	26± 3
$\mathbf{p}_{\rm T}^{\rm tot} < 45$	1131	1130±10	5.5± 0.1	6.1± 0.1	135± 1	39± 2	395± 9	54± 3	480±10	18± 3
$Z \to \tau\tau$ veto	1003	1010±10	5.2± 0.1	5.8± 0.1	128± 1	37± 2	373± 9	50± 3	395± 9	18± 2
$\lvert\Delta y_{jj}\rvert > 2.8$	143	152± 5	2.80± 0.02	1.40± 0.03	24.4± 0.5	5.4± 0.7	64± 4	9± 1	43± 3	4± 1
$m_{jj} > 500$	29	43± 3	1.75± 0.02	0.42± 0.01	9.1± 0.2	1.4± 0.3	15± 2	2.3± 0.6	13± 2	1.3± 0.5
No jets in y gap	18	24± 2	1.55± 0.01	0.28± 0.01	7.2± 0.2	0.9± 0.3	7± 1	1.4± 0.5	7± 1	0.6± 0.4
Both ℓ in y gap	13	18± 2	1.49± 0.01	0.24± 0.01	5.6± 0.2	0.7± 0.3	5± 1	1.1± 0.5	5± 1	0.5± 0.4
$m_{\ell\ell} < 60$	3	7± 1	1.35± 0.01	0.22± 0.01	1.3± 0.1	0.08± 0.02	1.0± 0.4	0.1± 0.2	4± 1	0.4± 0.3
$\Delta\phi_{\ell\ell} < 1.8$	2	5.2± 0.9	1.24± 0.01	0.20± 0.01	1.1± 0.1	0.05± 0.01	0.9± 0.3	0.1± 0.2	2.5± 0.8	0.2± 0.2

Appendix D

Data and Simulation $\not\!\!E_T$ Comparisons in 8 TeV Data

In this appendix, all the different missing transverse momentum flavours studied in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ re-analysis are show for the following enriched regions,

- $Z \to \ell\ell$ events,
- WW control region,
- Top control region,
- $Z \to \tau\tau$ control region.

The distributions show each of the final lepton channels in which the analysis is divided: ee, $\mu\mu$, $e\mu$ and μe, where the first lepton represents the one with the highest transverse momentum. All the distributions show the data over MC ratio and using 20.3 fb^{-1} data at 8 TeV and MC12 simulation samples.

D.1 $Z \to \ell\ell$ Events

In this section all the different missing transverse momentum reconstructions, and their x and y components, are show for $ee + \mu\mu$ events in the 15 GeV window around the Z mass.

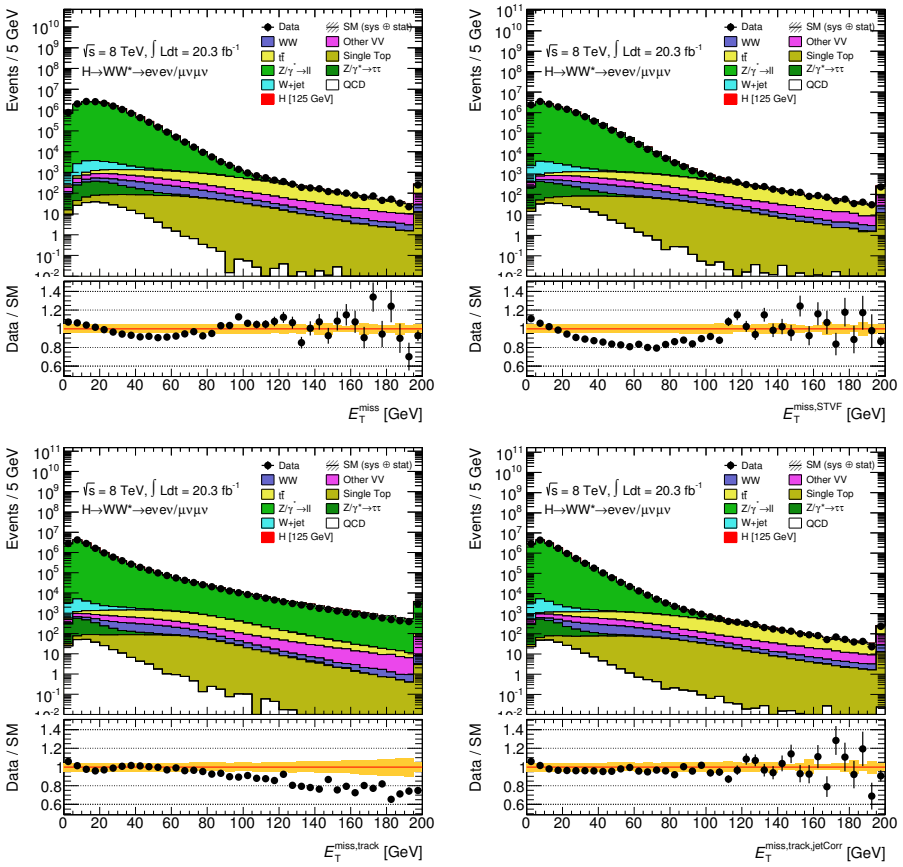

FIGURE D.1: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$, $E_{\mathrm{T}}^{\mathrm{miss,track}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ for $Z \to \ell\ell$ events in 8 TeV ATLAS data and MC simulation for the ee+$\mu\mu$ final states and inclusive in jet multiplicity.

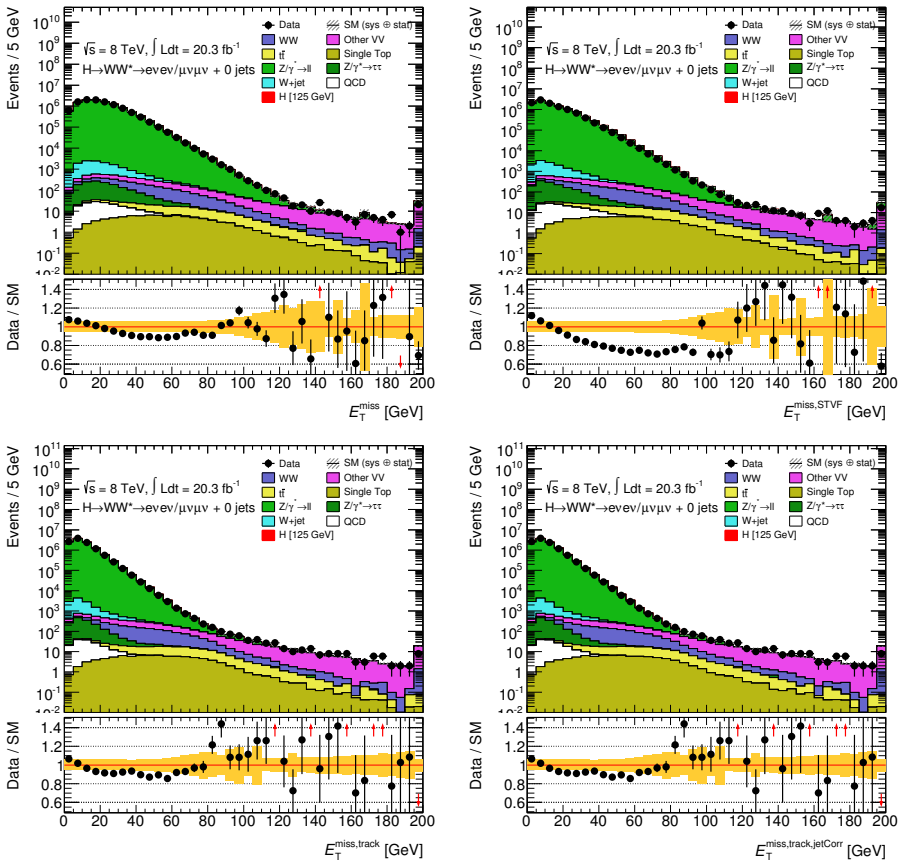

FIGURE D.2: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$, $E_{\mathrm{T}}^{\mathrm{miss,track}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ for $Z \to \ell\ell$ events in 8 TeV ATLAS data and MC simulation for the ee+$\mu\mu$ final state with 0 jets.

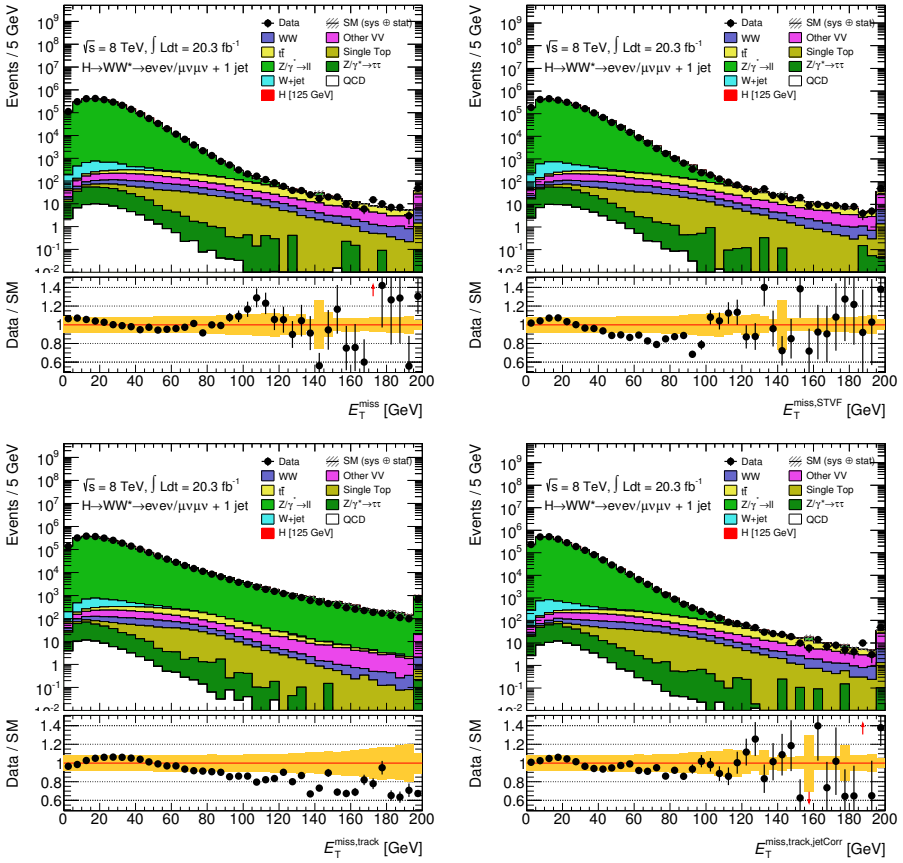

FIGURE D.3: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$, $E_{\mathrm{T}}^{\mathrm{miss,track}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ for $Z \to \ell\ell$ events in 8 TeV ATLAS data and MC simulation for the ee+$\mu\mu$ final state with exactly 1 jet.

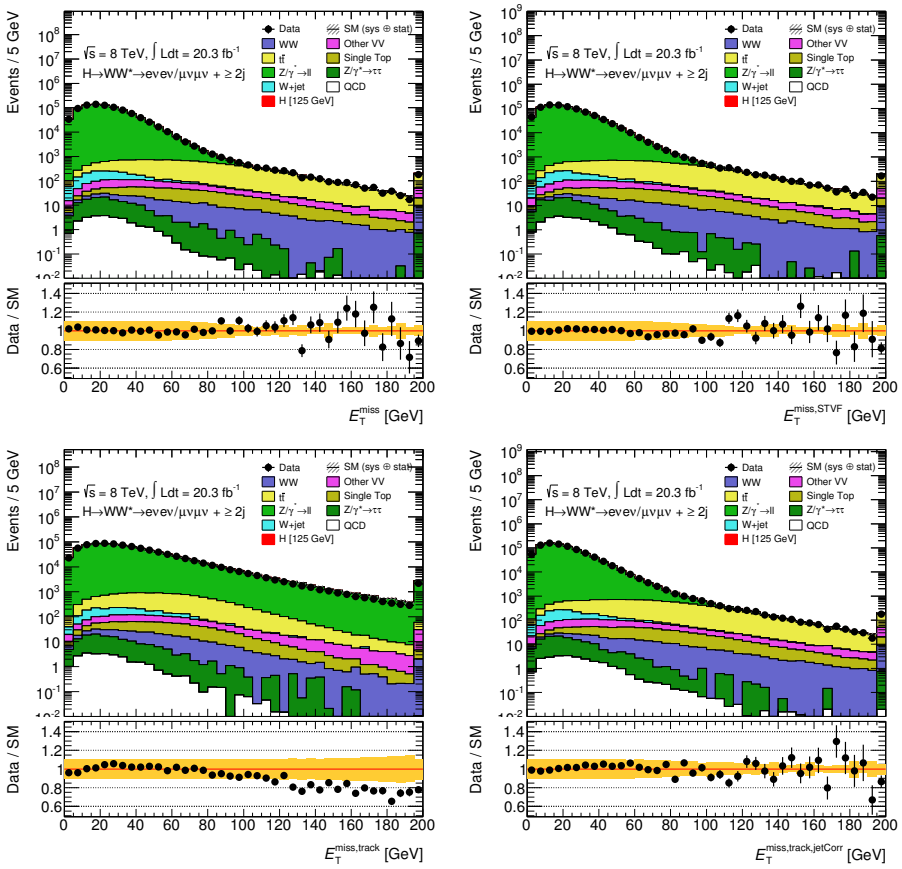

FIGURE D.4: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$, $E_{\mathrm{T}}^{\mathrm{miss,track}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ for $Z \to \ell\ell$ events in 8 TeV ATLAS data and MC simulation for the ee+$\mu\mu$ final state with at least 2 jets.

D.2 WW Control Region

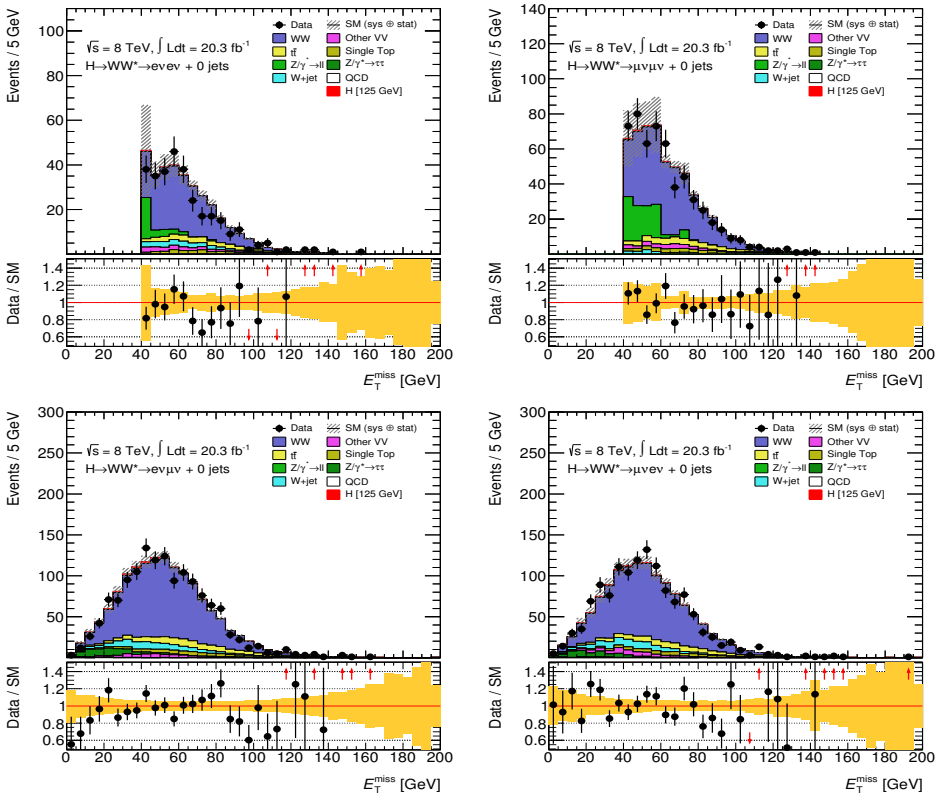

FIGURE D.5: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the WW control region in 0-jet for the final states of ee, $\mu\mu$, $e\mu$ and μe.

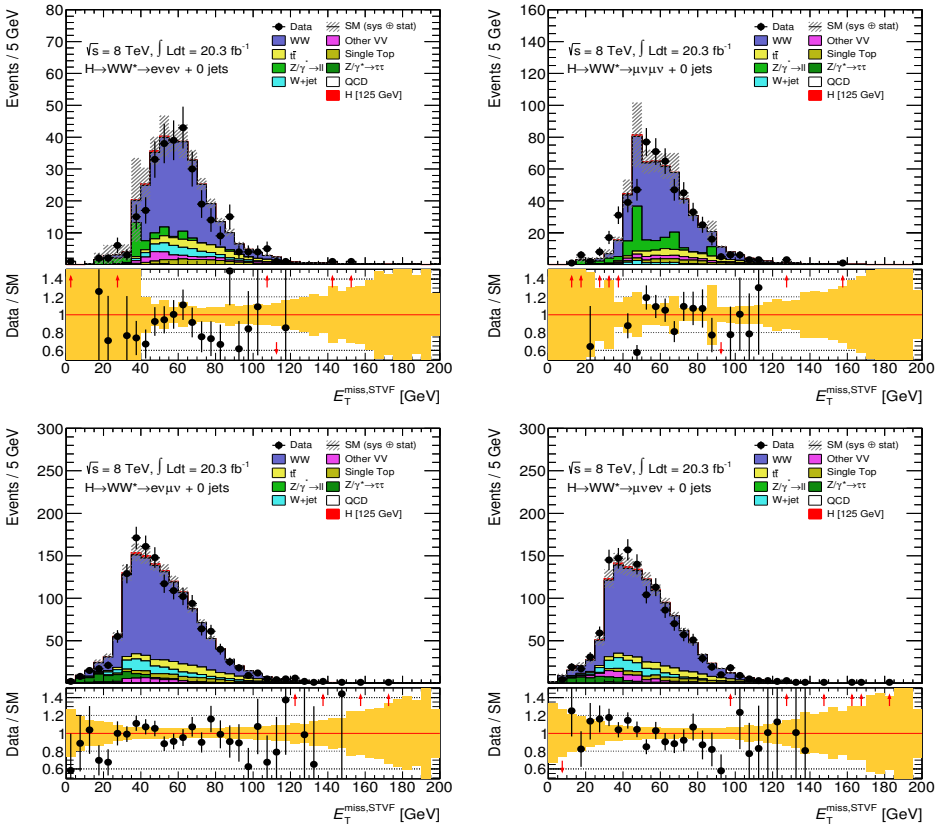

FIGURE D.6: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ in the WW control region in 0-jet for the final states of ee, $\mu\mu$, $e\mu$ and μe.

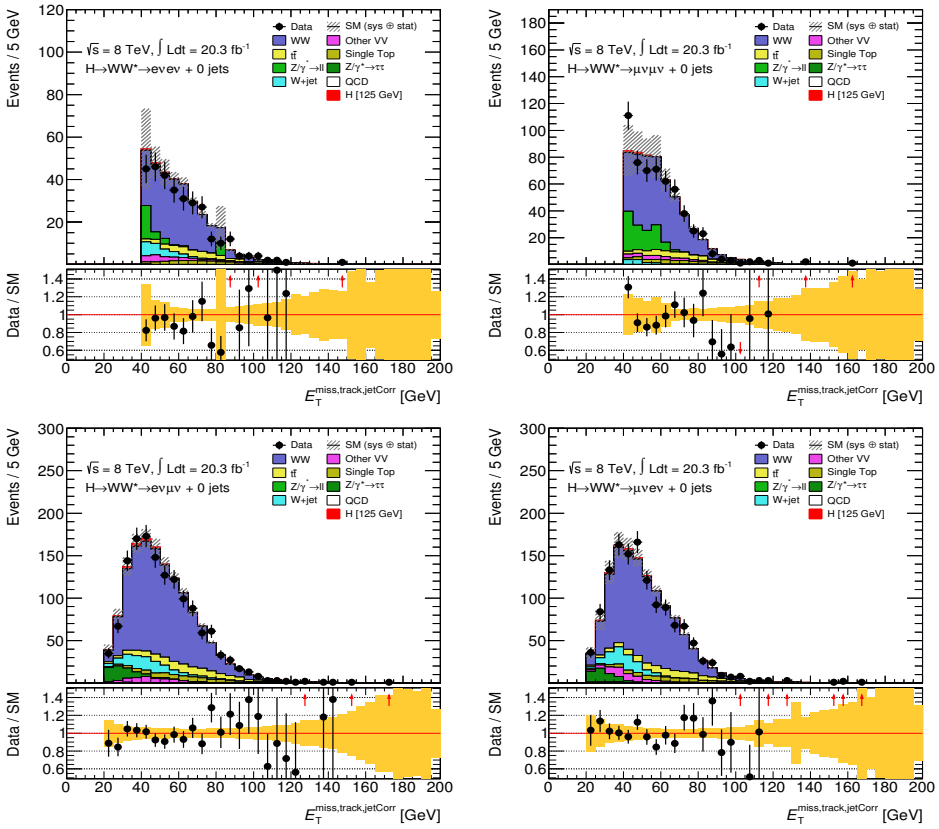

FIGURE D.7: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ in the WW control region in 0-jet for the final states of ee, $\mu\mu$, $e\mu$ and μe.

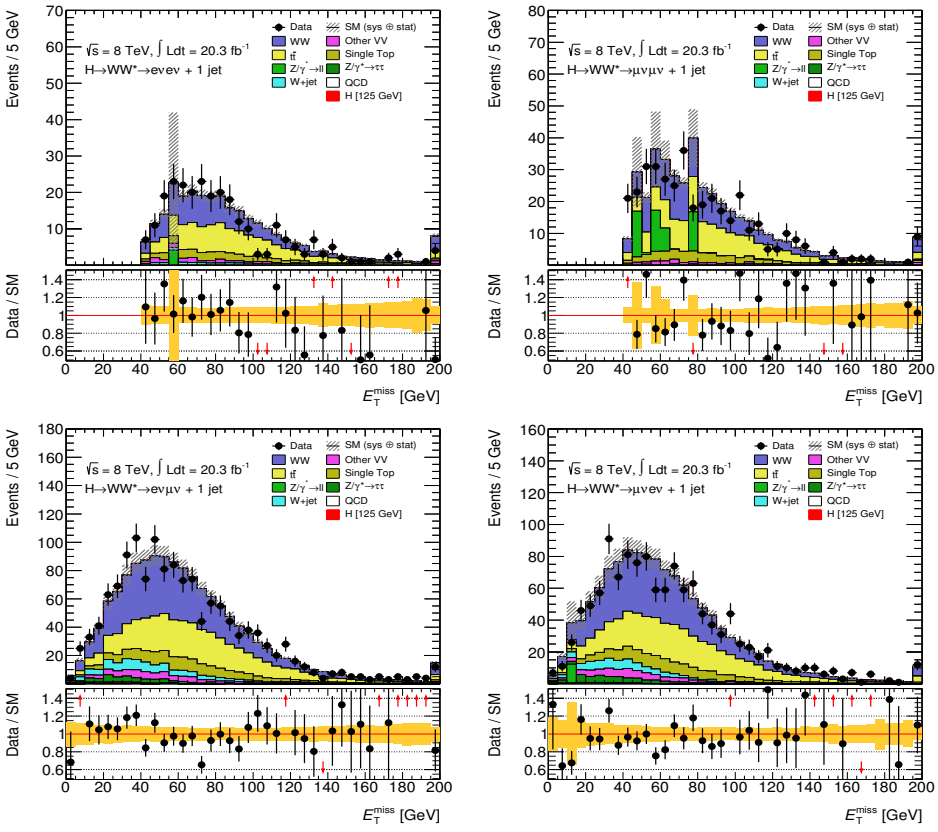

FIGURE D.8: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the WW control region in 1-jet for the final states of ee, $\mu\mu$, $e\mu$ and μe.

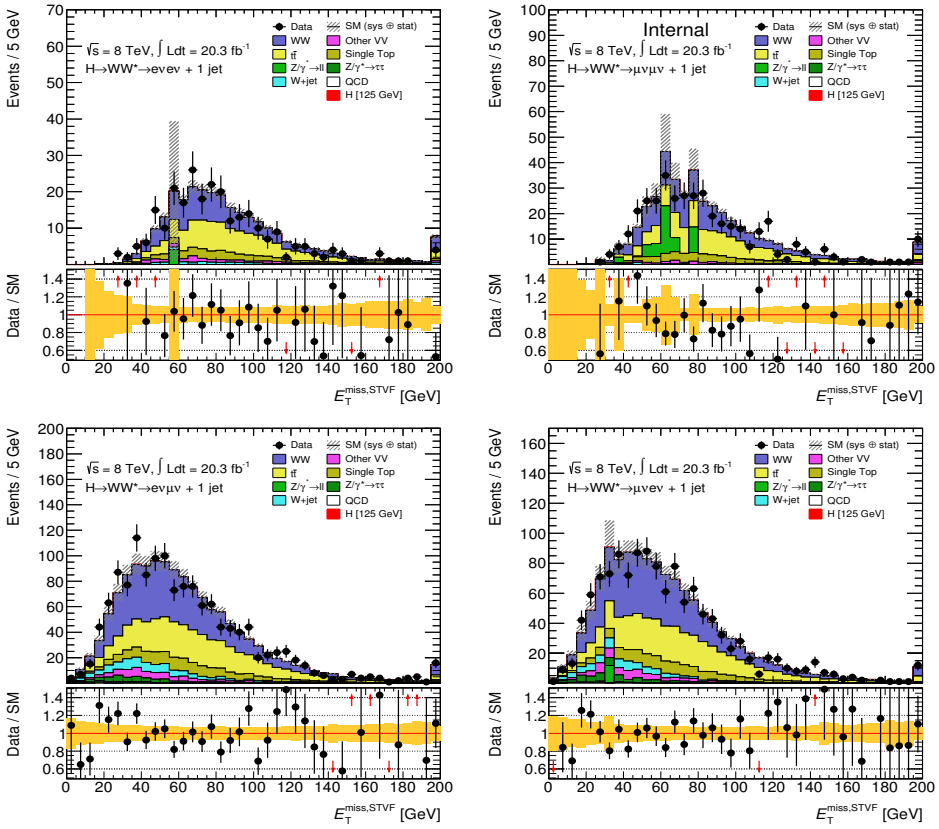

FIGURE D.9: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ in the WW control region in 1-jet for the final states of ee, $\mu\mu$, $e\mu$ and μe.

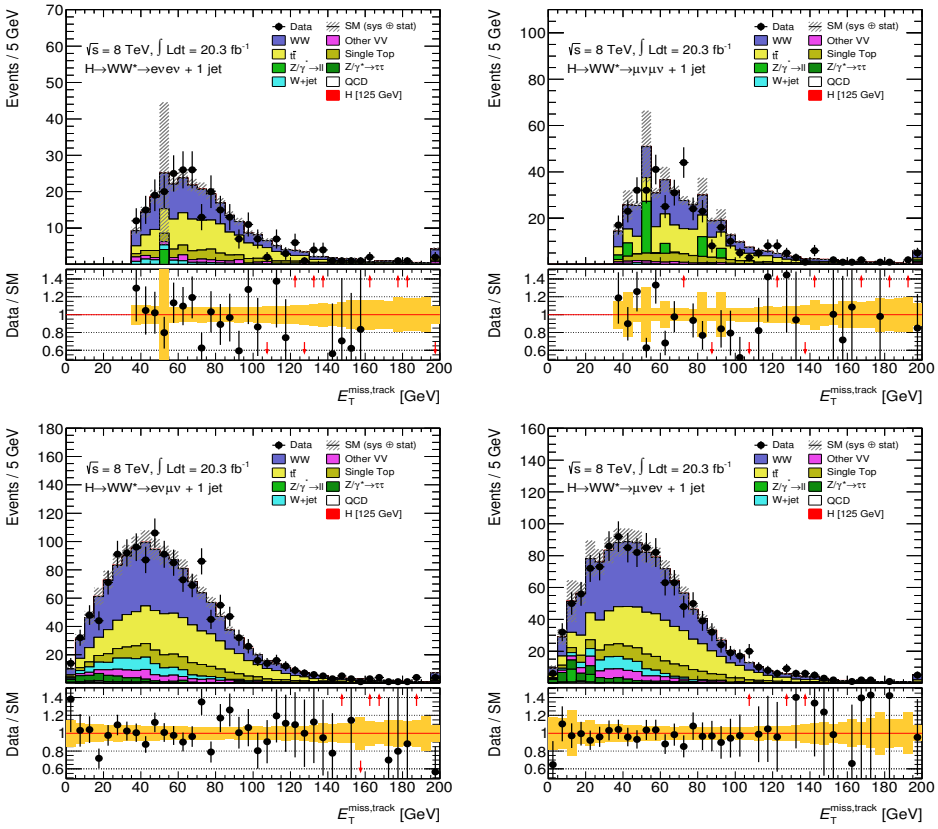

FIGURE D.10: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ in the WW control region in 1-jet for the final states of ee, $\mu\mu$, $e\mu$ and μe.

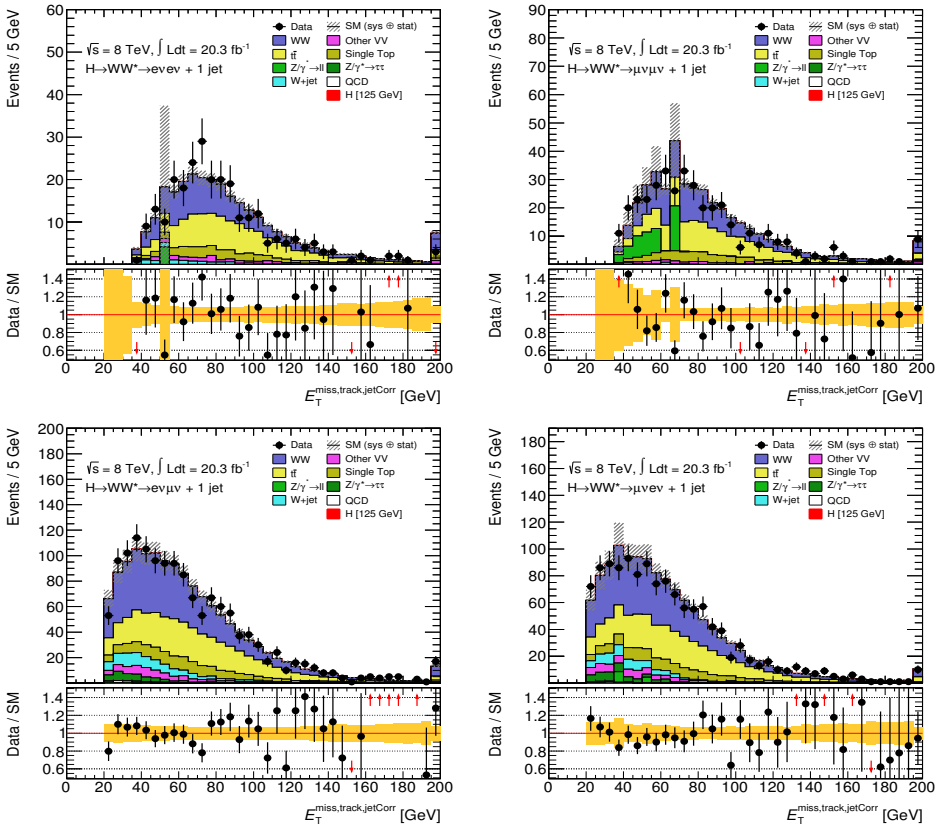

FIGURE D.11: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ in the WW control region in 1-jet for the final states of ee, $\mu\mu$, $e\mu$ and μe.

D.3 Top Control Region

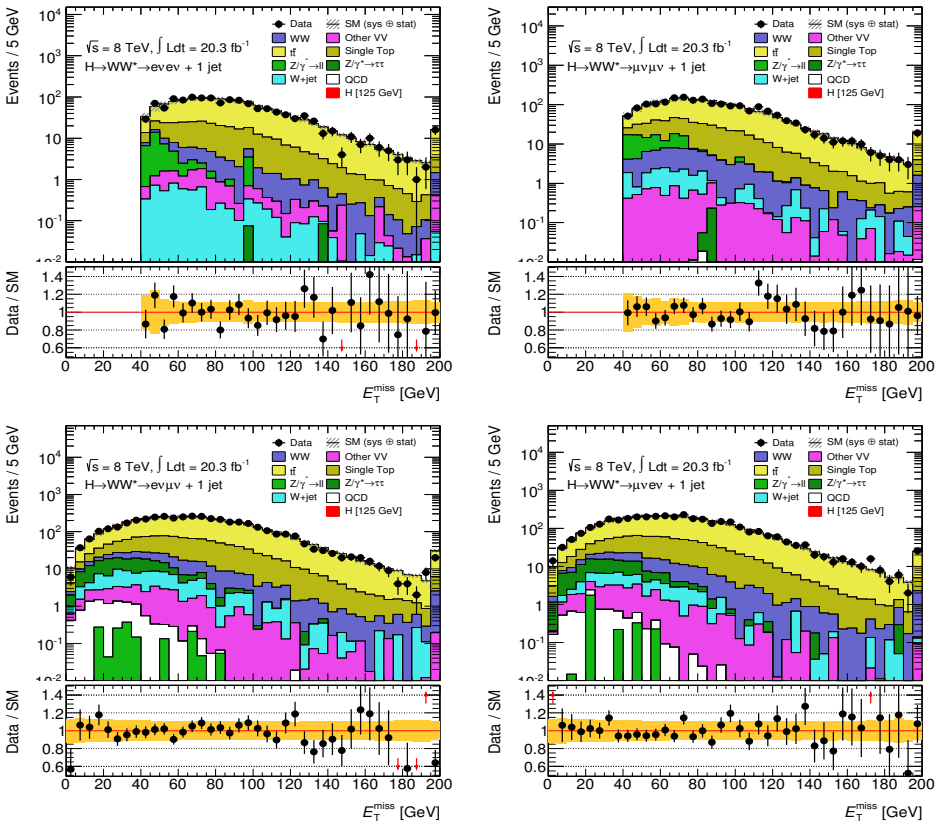

FIGURE D.12: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the top control region for the final states of ee, $\mu\mu$, $e\mu$ and μe plus 1 jet.

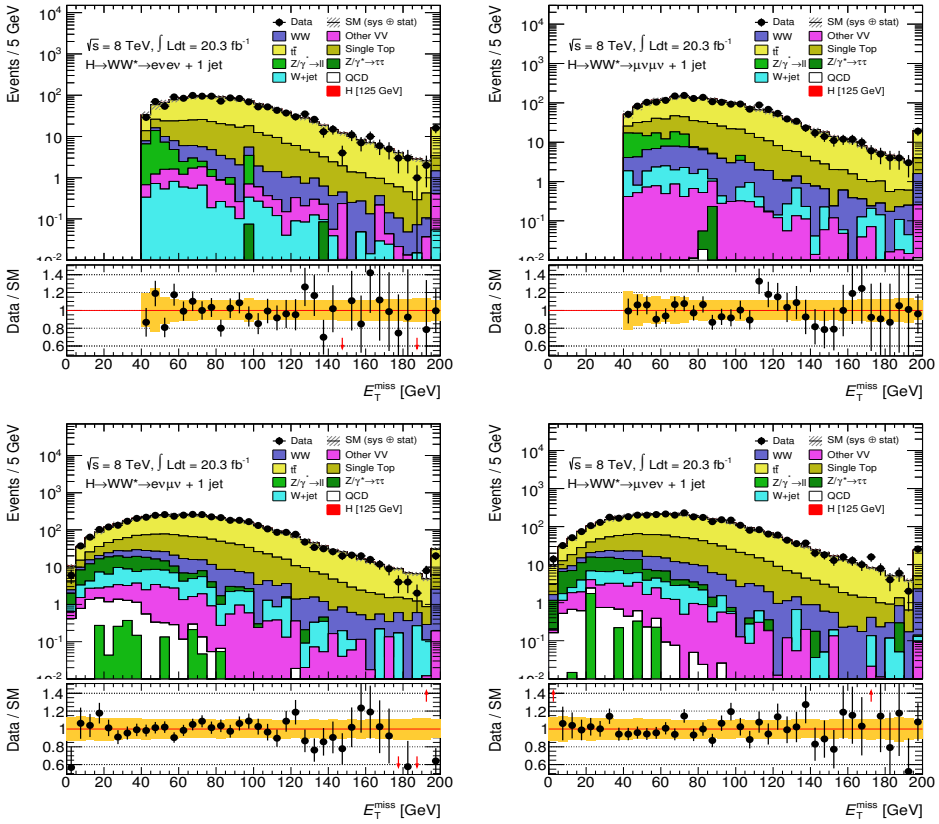

FIGURE D.13: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the top control region for the final states of ee, $\mu\mu$, $e\mu$ and μe plus 1 jet.

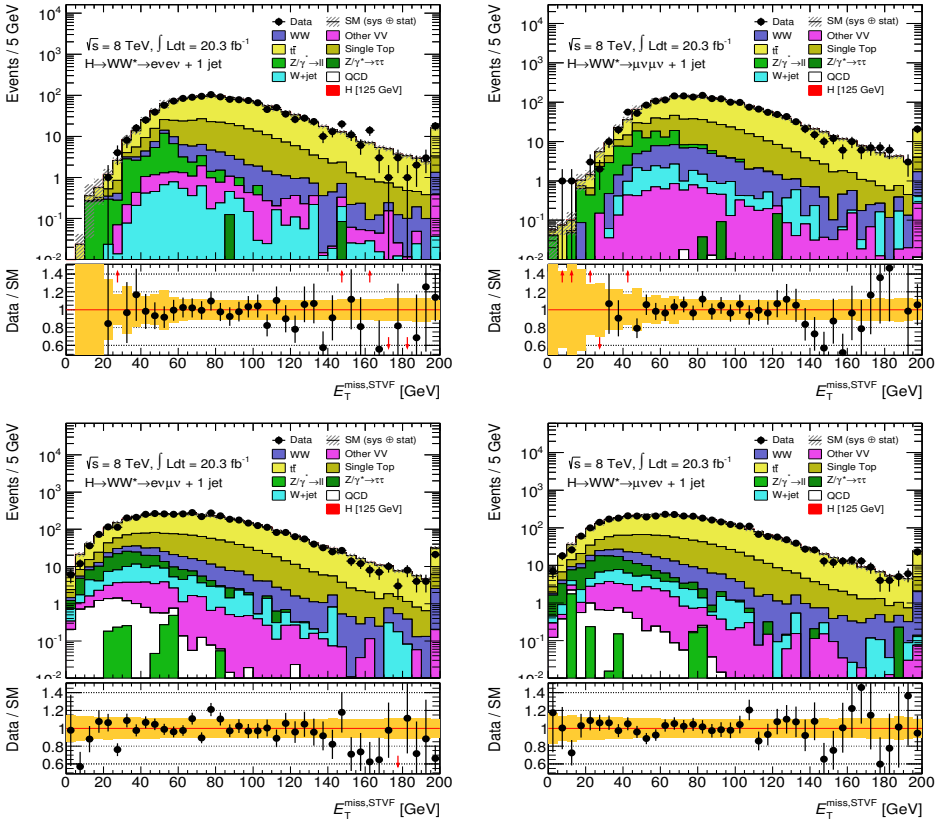

FIGURE D.14: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ in the top control region for the final states of ee, $\mu\mu$, $e\mu$ and μe plus 1 jet.

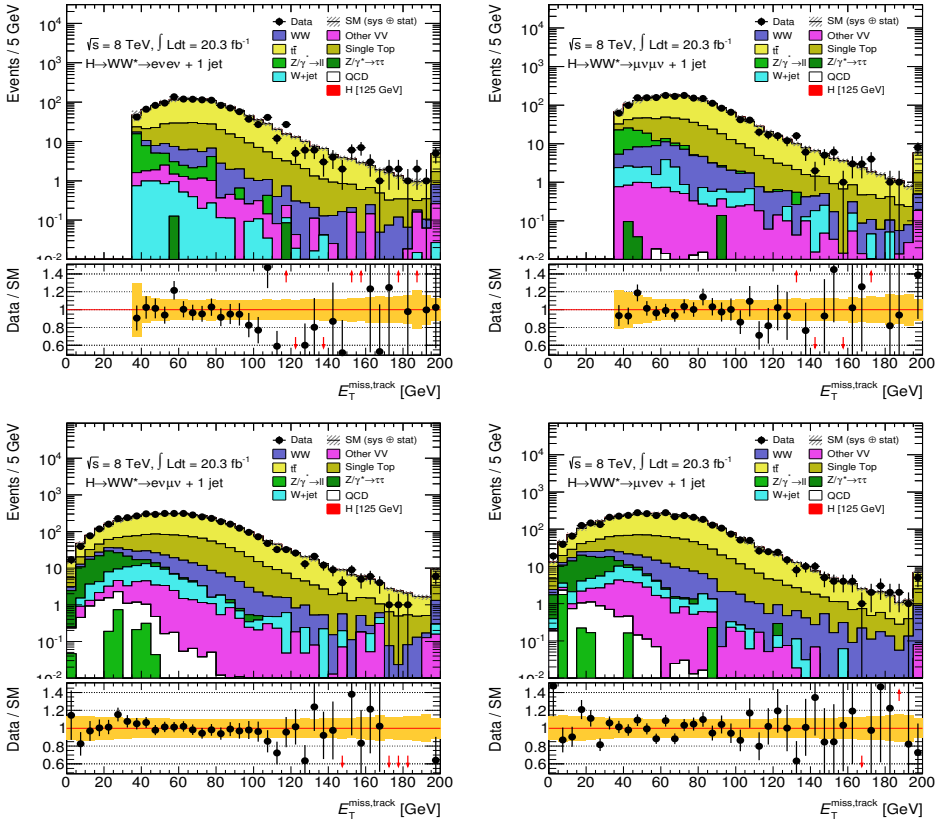

FIGURE D.15: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ in the top control region for the final states of ee, $\mu\mu$, $e\mu$ and μe plus 1 jet.

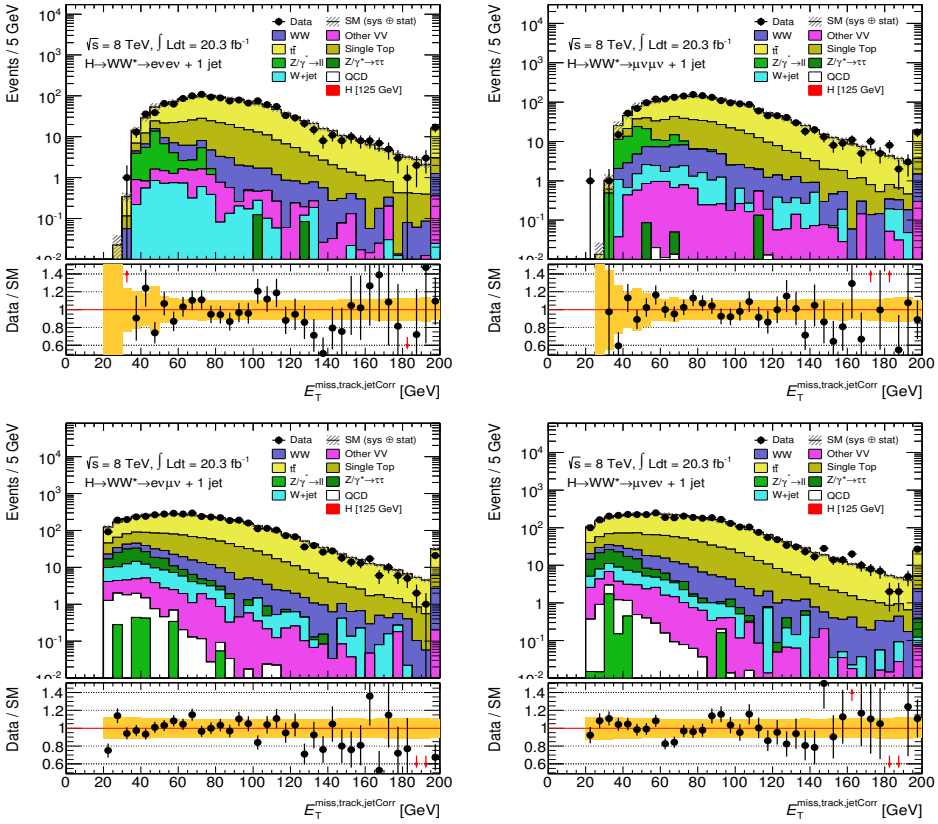

FIGURE D.16: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ in the top control region for the final states of ee, $\mu\mu$, $e\mu$ and μe plus 1 jet.

D.4 $Z \to \tau\tau$ Control Region

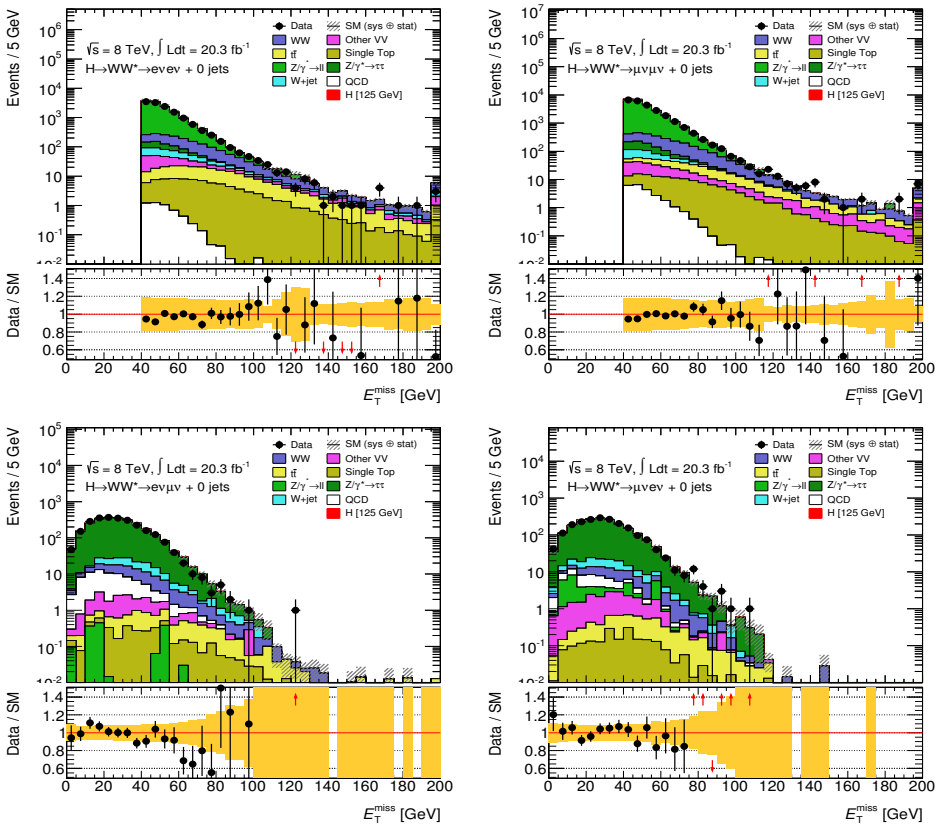

FIGURE D.17: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the $Z \to \tau\tau$ control region in 0-jet for the final states of ee, $\mu\mu$, $e\mu$ and μe .

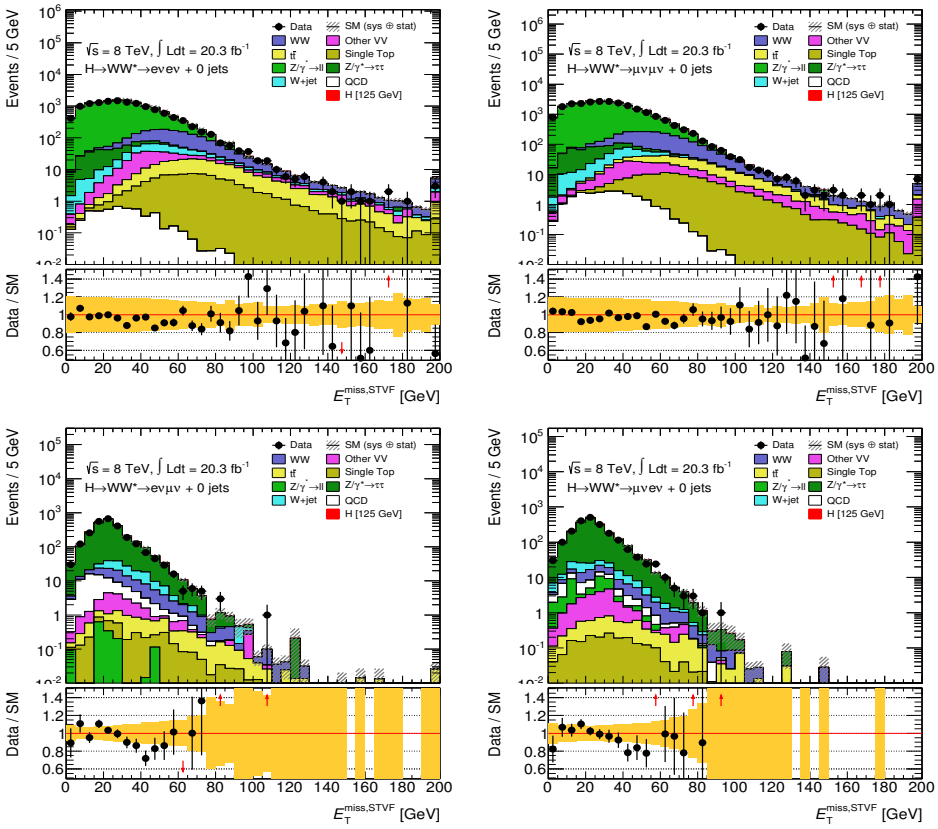

FIGURE D.18: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ in the $Z \to \tau\tau$ control region in 0-jet for the final states of ee, $\mu\mu$, $e\mu$ and μe .

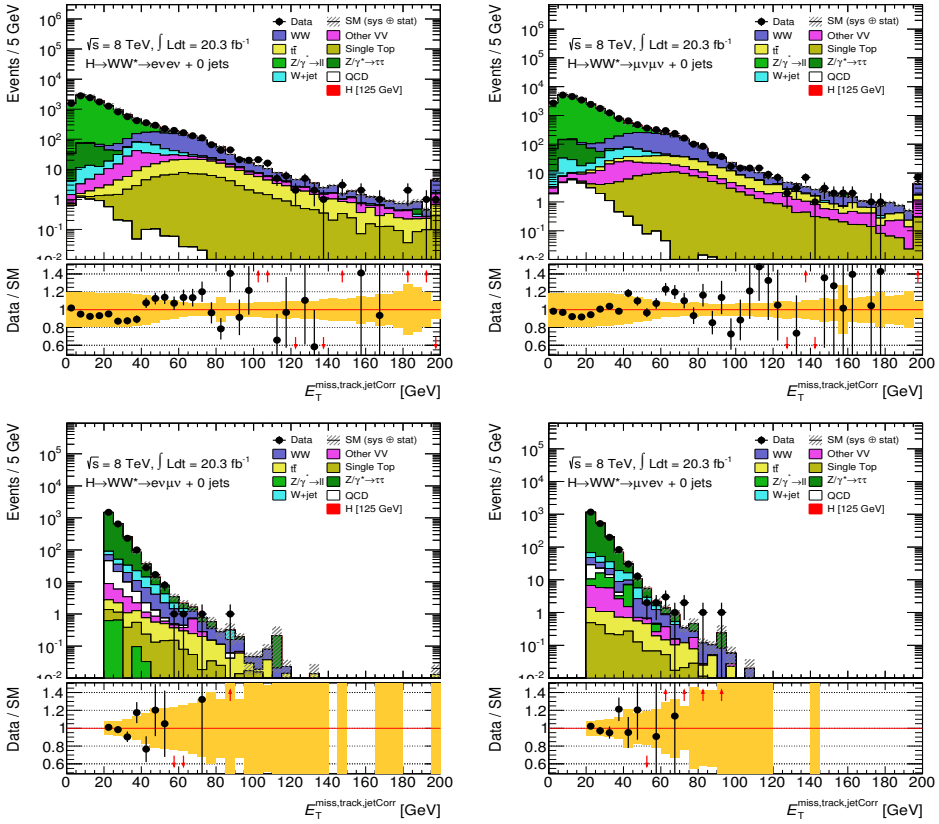

FIGURE D.19: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ in the $Z \rightarrow \tau\tau$ control region in 0-jet for the final states of ee, $\mu\mu$, $e\mu$ and μe .

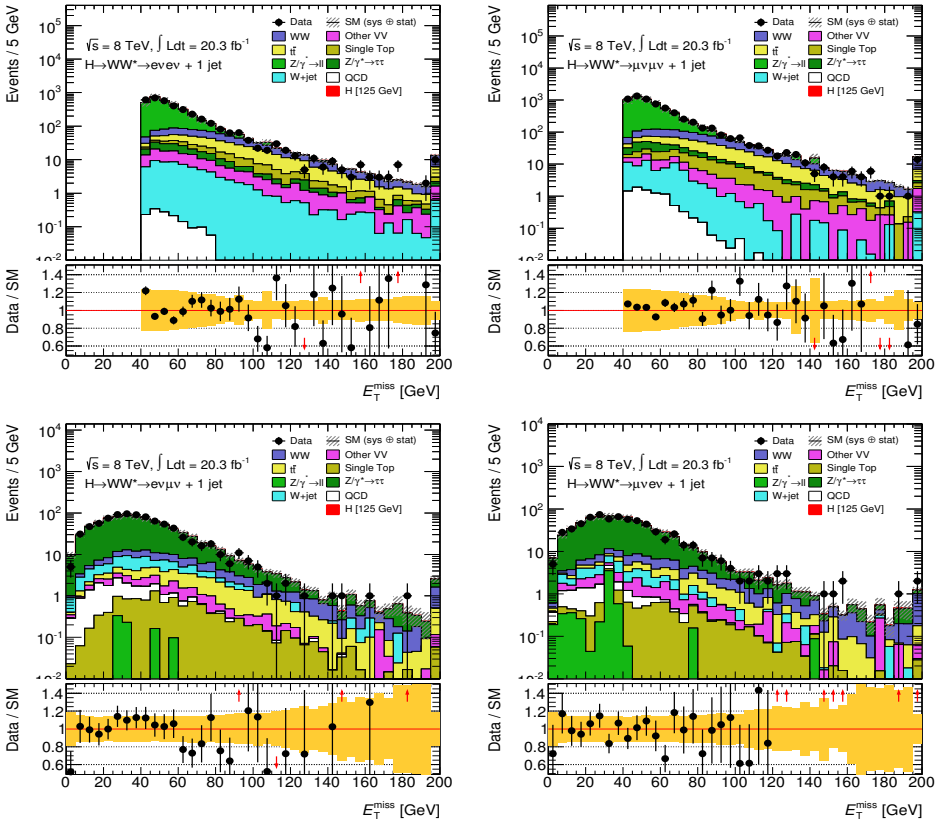

FIGURE D.20: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the $Z \to \tau\tau$ control region in 1-jet for the final states of ee, $\mu\mu$, $e\mu$ and μe .

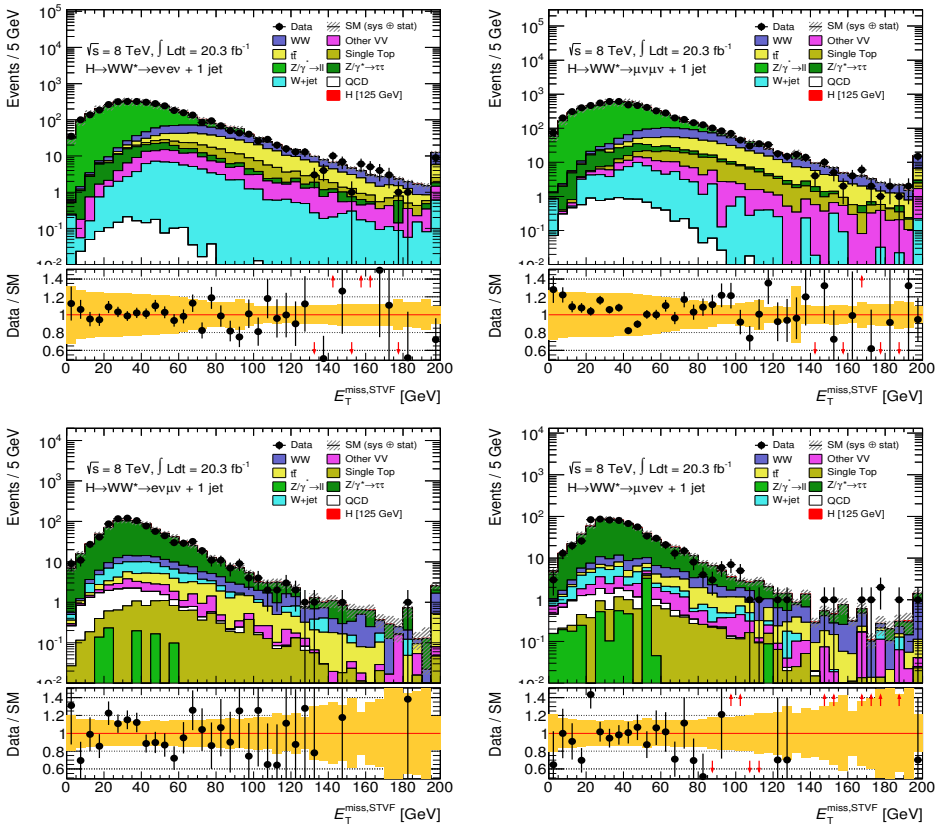

FIGURE D.21: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ in the $Z \to \tau\tau$ control region in 1-jet for the final states of ee, $\mu\mu$, $e\mu$ and μe .

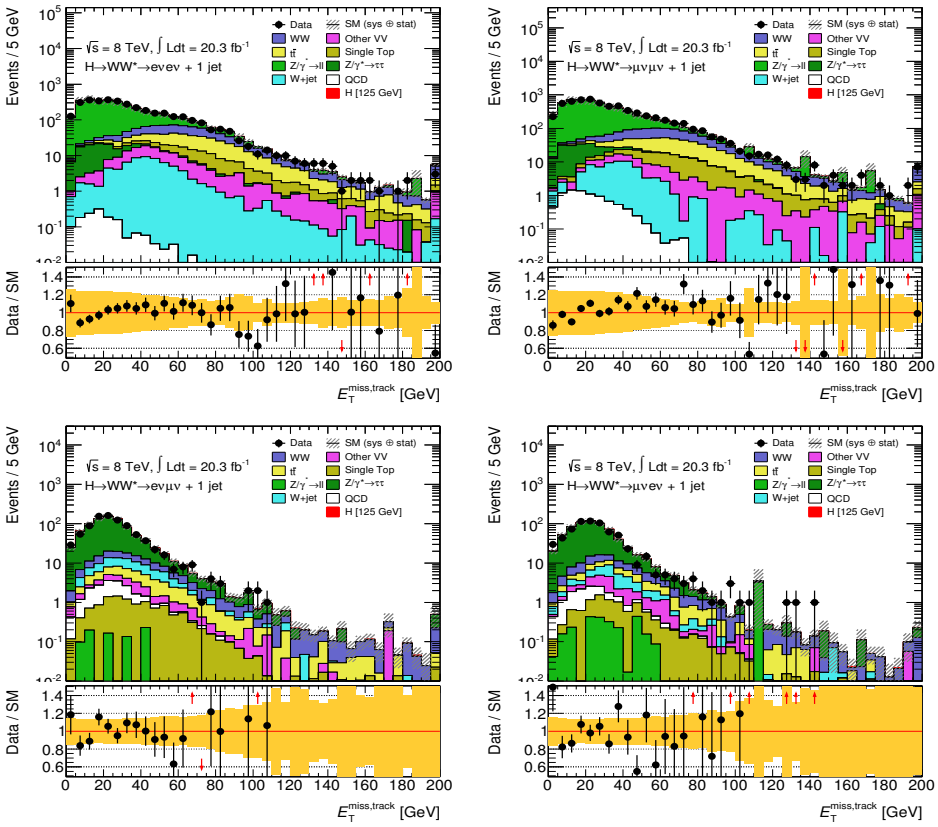

FIGURE D.22: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss,track}}$ in the $Z \to \tau\tau$ control region in 1-jet for the final states of ee, $\mu\mu$, $e\mu$ and μe .

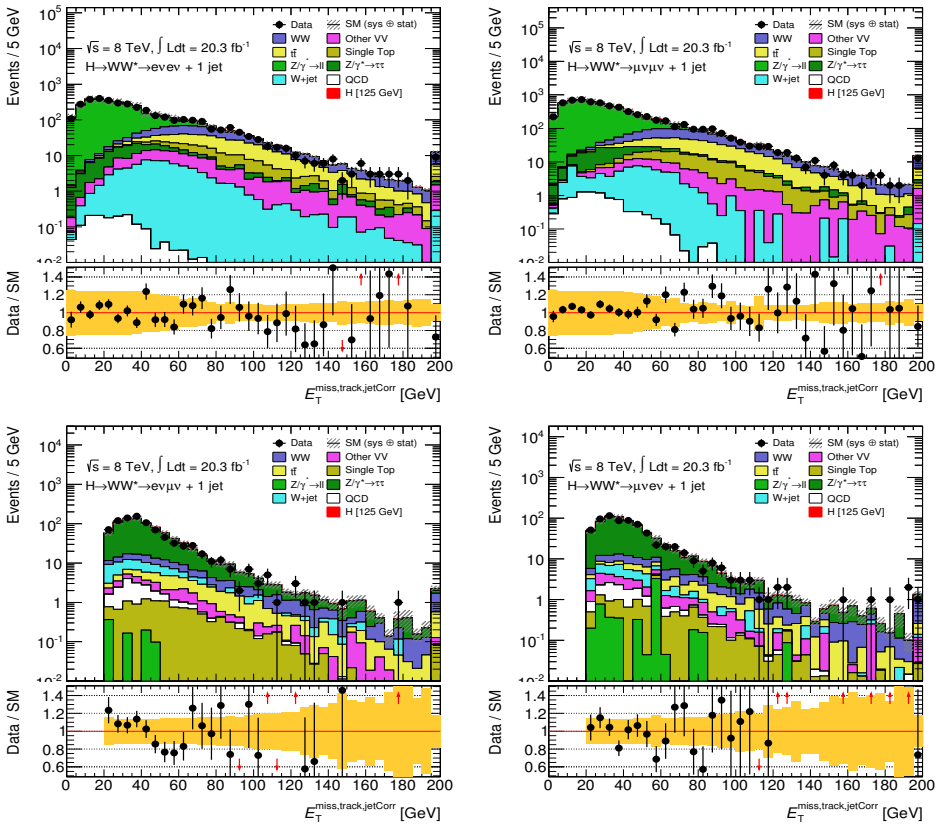

FIGURE D.23: Distributions of the $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$ in the $Z \rightarrow \tau\tau$ control region in 1-jet for the final states of ee, $\mu\mu$, $e\mu$ and μe .

Appendix E

MC Comparisons of $\not{E}_T$ in $H \to WW^{(*)} \to \ell\nu\ell\nu$ Candidates

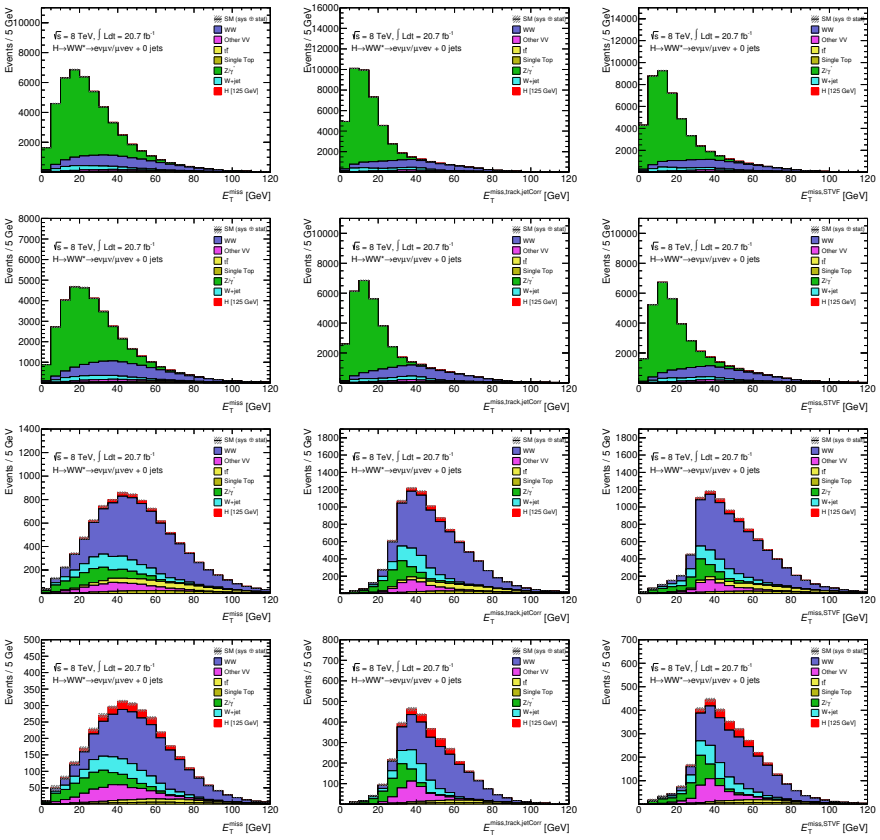

FIGURE E.1: Evolution of $\not\!\!E_T$ shape for $e\mu$+μe final states with 0 jets from MC simulation through different H+0j selections. The plots follow the selection in Section 5.7.2: jet veto (top left), $\Delta\phi_{ll,MET} > 1.57$ (top right), $P_T^{ll} > 30GeV$ (bottom left), and $M_{ll} < 55\,\text{GeV}$ (bottom right). From left to right the order of the $\not\!\!E_T$ varieties is: $E_{\text{T}}^{\text{miss}}$, $E_{\text{T}}^{\text{miss,track,jetCorr}}$, and $E_{\text{T}}^{\text{miss, STVF}}$.

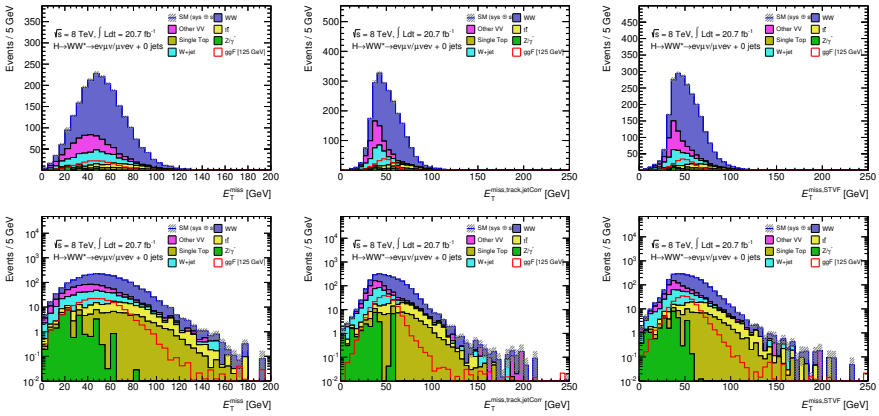

FIGURE E.2: $\not\!\!E_T$ distributions from MC simulation after all event selections applied for eμ+μe final states in the H+0j category. On top, the plots show linear scale and on the bottom, logaritmic scale. From left to right the order of the $\not\!\!E_T$ varieties is: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$, and $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$.

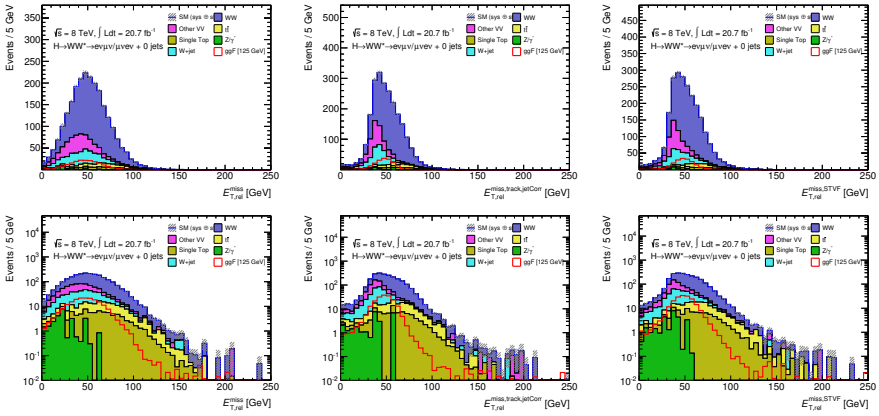

FIGURE E.3: $\not\!\!E_{T,Rel}$ distributions from MC simulation after all event selections applied for $e\mu$+μe final states in the H+0j category. On top, the plots show linear scale and on the bottom, logaritmic scale. From left to right the order of the $\not\!\!E_T$ varieties is: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss,\,STVF}}$ and $E_{\mathrm{T}}^{\mathrm{miss,track,jetCorr}}$.

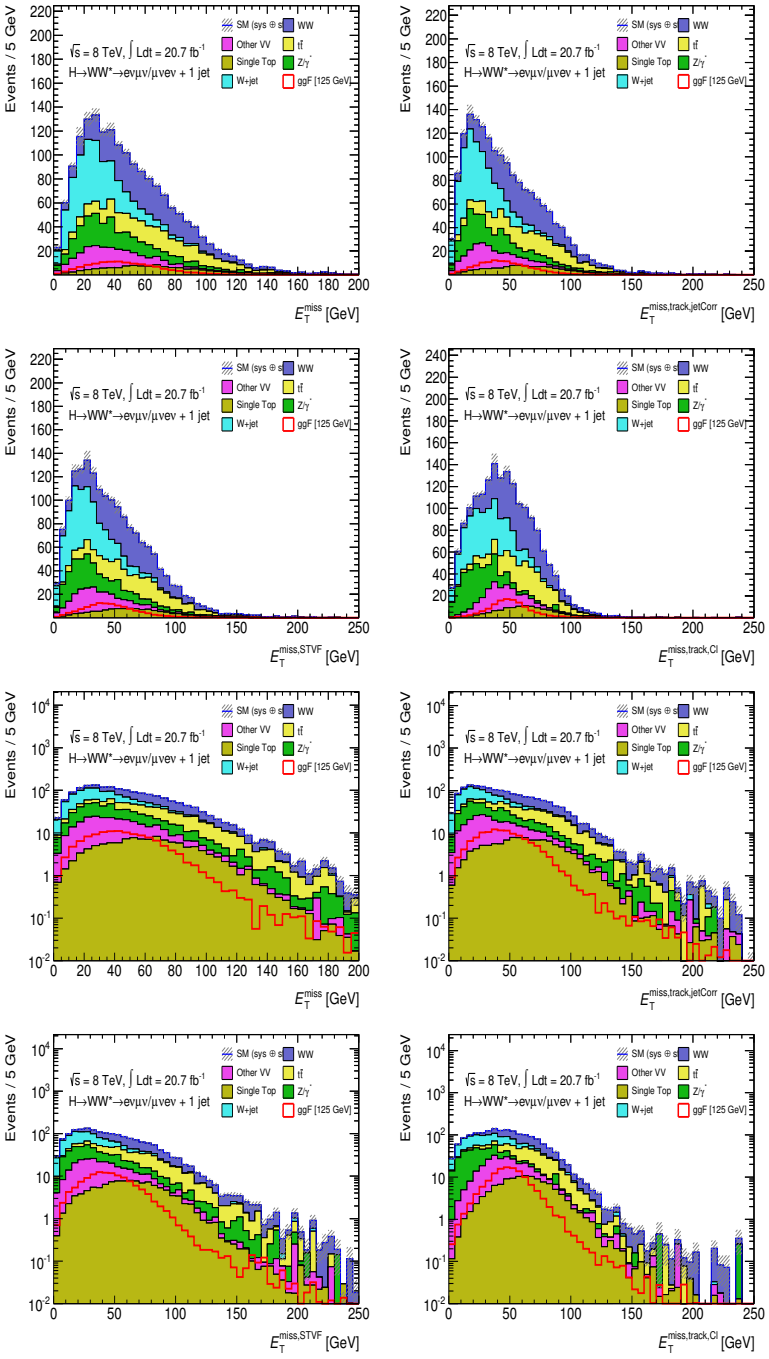

FIGURE E.4: $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions after all event selection applied for $e\mu$+μe final states with exactly 1 jet. Two top rows are in linear scale and the two bottom rows in logarithmic.

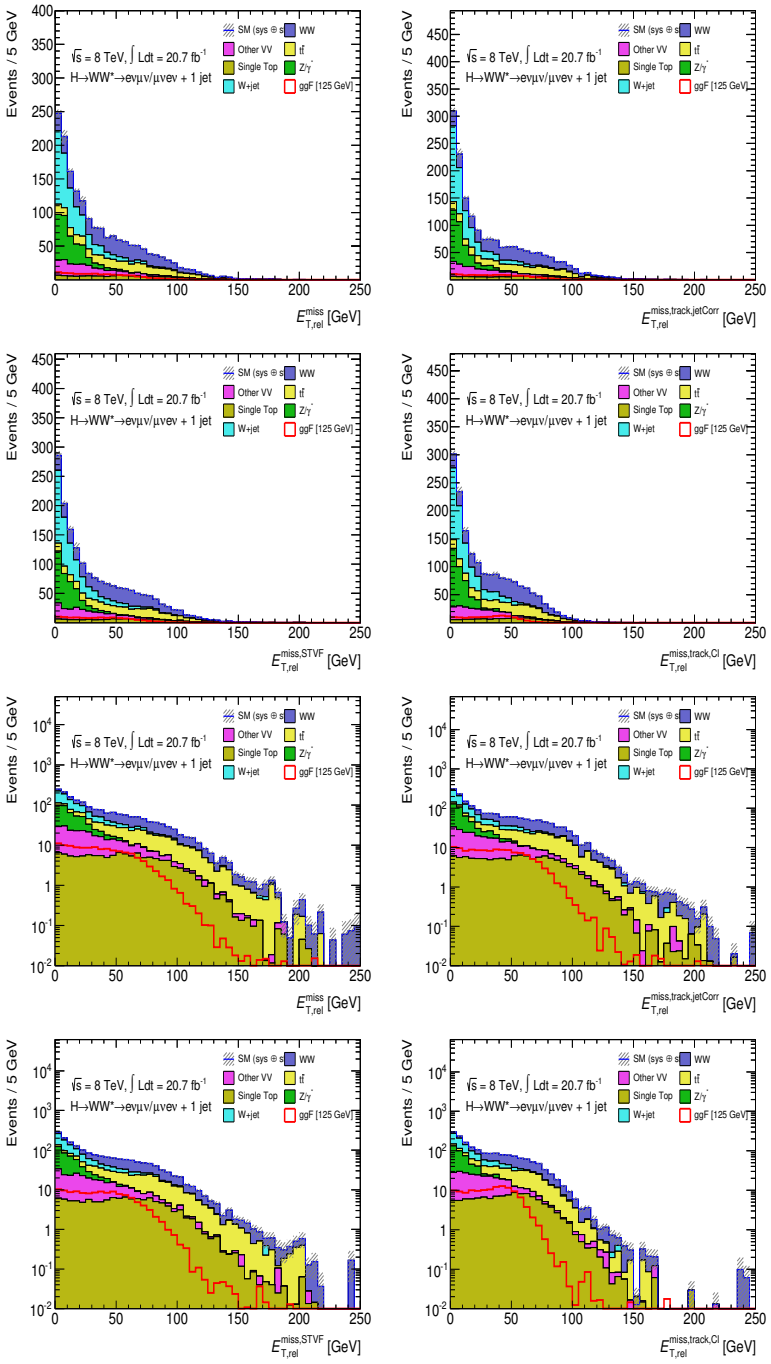

FIGURE E.5: $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ distributions after all event selection applied for $e\mu$+μe final states in the H+1j category. Two top rows are in linear scale and the two bottom rows in logaritmic.

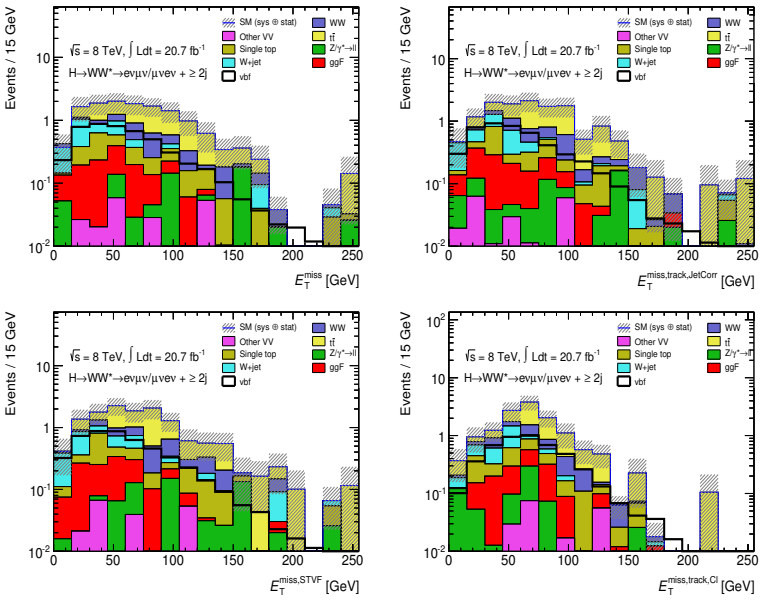

FIGURE E.6: $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions from MC simulation after all event selection applied for $e\mu$+μe final states in the H+$2j$ category.

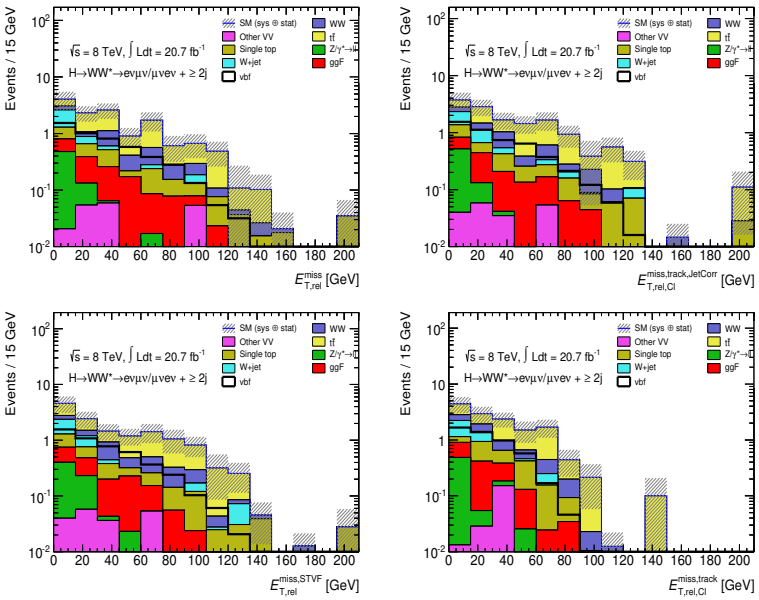

FIGURE E.7: $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ distributions from MC simulation after all event selection applied for $e\mu$+μe final states in the H+$2j$ category.

List of Acronyms

ADC	Analog to Digital Converter.
ALICE	A Large Ion Collider Experiment.
ATLAS	A Toroidal ApparatuS.
BDT	Boosted Decision Tree.
BEH	Brout-Englert-Higgs.
CERN	Conseil Européen pour la Recherche Nucléaire.
CJV	Central Jet Veto.
CL	Confidence Level.
CMS	Compact Muon Solenoid.
CR	Control Region.
CSC	Cathode Strip Chamber.
DY	Drell-Yan.
EB	Event Builder.
EBA	Extended Barrel A side.
EBC	Extended Barrel C side.
EM	ElectroMagnetic.
EMEC	ElectroMagnetic End-Cap.
EF	Event Filter.
EW	ElectroWeak.
FSR	Final State Radiation.
GSF	Gaussian Sum Fitter.
HEC	Hadronic End-cap Calorimeter.
ID	Inner Detector.

IP	Interaction Point.
ITC	Intermediate Tile Calorimeter cell.
ISR	Initial State Radiation.
JVF	Jet Vertex Fraction.
LEIR	Low Energy Ion Ring.
LEP	Large Electron-Positron machine.
LHC	Large Hadron Collider.
LVPS	Low Voltage Power Supplies.
MC	Monte Carlo.
MDT	Monitored Drift Tube.
ML	Maximum Likelihood.
MS	Muon Spectrometer.
OF	Optimal Filtering.
OLV	Outside Lepton Veto.
PMT	PhotoMultiplier Tube.
PS	Parton Shower.
PSB	Proton Synchrotron Booster.
PV	Primary Vertex.
QCD	Quantum ChromoDynamics.
QED	Quantum ElectroDynamics.
QFT	Quantum Field Theory.
RMS	Root Mean Square.
ROB	Read-Out Buffer.
ROI	Regions Of Interest.
ROD	Read-Out Driver.
RPC	Resistive Plate Chamber.
SCT	SemiConductor Tracker.
SM	Standard Model.
SPS	Super Proton Synchrotron.
SSB	Spontaneous Symmetry Breaking.
STVF	Soft-Term Vertex Fraction.
TGC	Thin Gap Chamber.

TRT	Transition Radiation Tracker.
UE	Underlaying Event.
VBF	Vector Boson Fusion.
VEV	Vacuum Expectation Value.
VR	Validation Region.

Bibliography

[1] K Nakamura. Review of particle physics. *J. Phys.*, G37:075021, 2010. doi: 10.1088/0954-3899/37/7A/075021.

[2] M. Fierz. Über die relativistische theorie kräftefreier teilchen mit beliebigem spin. *Helvetica Physica Acta*, 12:3–37, 1939.

[3] KamLAND collaboration. Measurement of neutrino oscillation with kamland: Evidence of spectral distortion. *Phys. Rev. Lett.*, 94:081801, Mar 2005. doi: 10.1103/PhysRevLett.94.081801. URL `http://link.aps.org/doi/10.1103/PhysRevLett.94.081801`.

[4] W. Pauli. The connection between spin and statistics. *Phys. Rev.*, 58:716–722, Oct 1940. doi: 10.1103/PhysRev.58.716. URL `http://link.aps.org/doi/10.1103/PhysRev.58.716`.

[5] S. L. Glashow. Partial symmetries of weak interactions. *Nucl. Phys. B*, 22: 579, 1961.

[6] S. Weinberg. A model of leptons. *PRL*, 19:1264, 1967.

[7] A. Salam. *Elementary particle theory.* Almqvist and Wiksells, Stockholm, 1968.

[8] G.'t Hooft. Renormalizable lagrangians for massive yang-mills fields. *Nuclear Physics B*, 35(1):167 – 188, 1971. ISSN 0550-3213. doi: http://dx.doi.org/10.1016/0550-3213(71)90139-8. URL `http://www.sciencedirect.com/science/article/pii/0550321371901398`.

[9] G.'t Hooft and M. Veltman. Regularization and renormalisation of gauge fields. *Nuclear Physics B*, B44:189 – 213, 1972.

[10] J. Goldstone. Field theories with « superconductor » solutions. *Il Nuovo Cimento*, 19(1):154–164, 1961. ISSN 0029-6341. doi: 10.1007/BF02812722. URL `http://dx.doi.org/10.1007/BF02812722`.

[11] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *PRL*, 13:321–323, 1964. doi: 10.1103/PhysRevLett.13.321.

[12] P. W. Higgs. Broken symmetries and the masses of gauge bosons. *PRL*, 13: 508–509, 1964. doi: 10.1103/PhysRevLett.13.508.

[13] G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble. Global conservation laws and massless particles. *Phys.Rev.Lett.*, 13:585–587, 1964. doi: 10.1103/PhysRevLett.13.585.

[14] The CERN Council. Hundred and Sixty-fourth Session of CERN Council, 2012. URL `http://council.web.cern.ch/council/en/Governance/Decisions164Sept12.html`.

[15] LEP Collaboration. *LEP design report.* CERN, Geneva, 1984. Copies shelved as reports in LEP, PS and SPS libraries.

[16] TeVI Group. Design Report Tevatron 1 project. 1984.

[17] LEP Working Group for Higgs boson searches. Search for the Standard Model Higgs boson at LEP. *PLB*, 565:61, 2003. doi: 10.1016/S0370-2693(03)00614-2.

[18] Campbell, J. Huston, and W. Stirling. Hard interactions of quarks and gluons: A primer of LHC physics. *Rept.Prog.Phys*, 70:89, 2007. doi: 10.1103/PhysRevD.81.074023.

[19] Jakobs, Karl. Physics at the LHC – From Standard Model measurements to Searches for New Physics. *arXiv*, pages 309–358, 2014. doi: 10.5170/CERN-2014-003.309.

[20] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur. Phys. J. C*, 63:189, 2009. doi: 10.1140/epjc/s10052-009-1072-5.

[21] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. *CERN-2012-002*, CERN, Geneva, 2012.

[22] S. Dawson. Radiative corrections to Higgs boson production. *Nucl. Phys. B*, 359:283, 1991. doi: 10.1016/0550-3213(91)90061-2.

[23] A. Djouadi, M. Spira, and P.M. Zerwas. Production of Higgs bosons in proton colliders: QCD corrections. *Phys.Lett.*, B264:440–446, 1991. doi: 10.1016/0370-2693(91)90375-Z.

[24] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. *CERN-2011-002*, CERN, Geneva, 2011.

[25] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 3. Higgs properties. *CERN-2013-004*, CERN, Geneva, 2013.

[26] Bruce Mellado, William Quayle, and Sau Lan Wu. Feasibility of searches for a higgs boson using $h \to ww + \ell\ell$+met and high p_t jets at the tevatron. *Phys.Rev. D*, 76:093007, 2007. doi: 10.1103/PhysRevD.76.093007.

[27] The ATLAS Collaboration. The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08):S08003, 2008. URL `http://stacks.iop.org/1748-0221/3/i=08/a=S08003`.

[28] The CMS Collaboration. The cms experiment at the cern lhc. *Journal of Instrumentation*, 3(08):S08004, 2008. URL `http://stacks.iop.org/1748-0221/3/i=08/a=S08004`.

[29] The LHCb Collaboration. The lhcb detector at the lhc. *Journal of Instrumentation*, 3(08):S08005, 2008. URL `http://stacks.iop.org/1748-0221/3/i=08/a=S08005`.

[30] The ALICE Collaboration. The alice experiment at the cern lhc. *Journal of Instrumentation*, 3(08):S08002, 2008. URL `http://stacks.iop.org/1748-0221/3/i=08/a=S08002`.

[31] Thomas Sven Pettersson and P Lefèvre. The Large Hadron Collider: conceptual design. Technical Report CERN-AC-95-05 LHC, CERN, Geneva, Oct 1995.

[32] Michael Benedikt, Paul Collier, V Mertens, John Poole, and Karlheinz Schindl. *LHC Design Report*. CERN, Geneva, 2004.

[33] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, 2008. URL `http://stacks.iop.org/1748-0221/3/i=08/a=S08001`.

[34] Bajko, M *et al.* Report of the Task Force on the Incident of 19th September 2008 at the LHC. Technical Report LHC-PROJECT-Report-1168. CERN-LHC-PROJECT-Report-1168, CERN, Geneva, Mar 2009.

[35] ATLAS Collaboration. Luminosity public results, March 2011. URL `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults`.

[36] ATLAS Collaboration. Luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC. *Eur.Phys.J.C71:1630,2011*, 2011.

[37] ATLAS Collaboration. Detector and physics performance technical design report. Technical Report CERN/LHCC 99-14, CERN, Geneva, 1999.

[38] ATLAS Collaboration. Detector and physics performance technical design report. *CERN/LHCC 99-15*, Vol II, 1999.

[39] The ATLAS Collaboration. Studies of the performance of the atlas detector using cosmic-ray muons. *The European Physical Journal C*, 71(3):1593, 2011. ISSN 1434-6044. doi: 10.1140/epjc/s10052-011-1593-6. URL `http://dx.doi.org/10.1140/epjc/s10052-011-1593-6`.

[40] The ATLAS Collaboration. Performance of the atlas detector using first collision data. *Journal of High Energy Physics*, 2010(9):56, 2010. doi: 10.1007/JHEP09(2010)056. URL `http://dx.doi.org/10.1007/JHEP09%282010%29056`.

[41] Joao Pequenao and Paul Schaffner. An computer generated image representing how ATLAS detects particles. Jan 2013. URL `http://www.atlas.ch/photos`.

[42] ATLAS Collaboration. *ATLAS Tile Calorimeter Detector Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996. URL `https://cds.cern.ch/record/331062`. CERN/LHCC 96-42.

[43] ATLAS Collaboration. *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1997. distribution.

[44] ATLAS Collaboration. *ATLAS magnet system: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1997.

[45] ATLAS Collaboration. *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. CERN, Geneva, 2003. ATLAS-TDR-016; CERN-LHCC-2003-022.

[46] ATLAS Collaboration. Readiness of the atlas tile calorimeter for lhc collisions. *The European Physical Journal C*, 70(4), 2010. ISSN 1434-6044. doi: 10.1140/epjc/s10052-010-1508-y. URL `http://dx.doi.org/10.1140/epjc/s10052-010-1508-y`.

[47] ATLAS Collaboration. Signal reconstruction of the atlas hadronic tile calorimeter: implementation and performance. *J. Phys.: Conf. Ser.*, 293 (012056), 2011. doi: doi:10.1088/1742-6596/293/1/012056.

[48] E. Fullana, J. Castelo, V. Castillo, C. Cuenca, A. Ferrer, E. Higon, C. Iglesias, J. Poveda, A. Munar, A. Ruiz-Martinez, B. Salvachua, C. Solans, R. Teuscher, and J. Valls. Optimal filtering in the atlas hadronic tile calorimeter. *CERN, Geneva*, (ATL-TILECAL-2005-001), 2005.

[49] A. Valero Biot. The back-end electronics for the atlas hadronic tile calorimeter at the large hadron collider. http://hdl.handle.net/10261/112262.

[50] E. Starchenko et al. Cesium monitoring system for atlas tile hadron calorimeter. *Nucl. Instrum. Methods A*, 494(281-284), 2002.

[51] N. Shalanda et al. Radiative source control and electronics for the atlas tile calorimeter cesium calibration system. *Nucl. Instrum. Methods A*, 508(276-286), 2003.

[52] E. Starchenko et al. Laser monitoring system for atlas tile hadron calorimeter. *Nucl. Instrum. Methods A*, 617(120-122), 2010.

[53] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajogopolan, D. Rousseau, S. Snyder, and G. Unal. Calorimeter Clustering Algorithms: Description and Performance. *ATL-LARG-PUB-2008-002*, page 063, CERN, Geneva, May 5, 2008. doi: 10.1088/1126-6708/2008/04/063.

[54] Peter Loch, Sven Menke, Pierre-Antoine Delsart, Christopher John Young, Zach Marshall, and Bradley Axen. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run I. Technical Report ATL-COM-PHYS-2014-1439, CERN, Geneva, Nov 2014. URL `https://cds.cern.ch/record/1967028`.

[55] Peter Loch, Sven Menke, Pierre-Antoine Delsart, Christopher John Young, Zach Marshall, Bradley Axen, and Michael Bender. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run I: Backup document for TopoCluster paper. Technical Report ATL-COM-PHYS-2015-868, CERN, Geneva, Aug 2015. URL `https://cds.cern.ch/record/2042236`.

[56] Performance of the Fast ATLAS Tracking Simulation (FATRAS) and the ATLAS Fast Calorimeter Simulation (FastCaloSim) with single particles. Technical Report ATL-SOFT-PUB-2014-001, CERN, Geneva, Mar 2014. URL `http://cds.cern.ch/record/1669341`.

[57] A. Artamonov, L. Fiorini, B. T. Martin, J. Novakova, A. Solodkov, I. Vichou. Description of the tile calorimeter electronic noise. *CERN, Geneva*, (ATL-TILECAL-INT-2011-002), March 2 2011.

[58] S Amor Santos, J P Araque, N Castro, V Drugakov, J Novakova, A Onofre, C Santoni, P Starovoitov, and S Yanush. Tile Calorimeter description of the noise with increasing Pile-up. Technical Report ATL-TILECAL-INT-2014-001, CERN, Geneva, Feb 2014. URL `https://cds.cern.ch/record/1647848`.

[59] ATLAS Collaboration. Performance of missing transverse momentum reconstruction in proton-proton collisions at 7 TeV with ATLAS. *Eur. Phys. J. C*, 72:1844, 2012. doi: 10.1140/epjc/s10052-011-1844-6.

[60] ATLAS Collaboration. Performance of the missing transverse energy reconstruction and calibration in *pp* collisions at center-of-mass energy of $\sqrt{s} = 7$ tev with the atlas detector. *ATLAS-CONF-2010-057*, 2010.

[61] ATLAS Collaboration. Performance of missing transverse momentum reconstruction in atlas with 2011 proton-proton collisions at $\sqrt{s} = 7$ tev. *ATLAS-CONF-2012-101*, 2012.

[62] ATLAS Collaboration. Pile-up correction in missing transverse momentum reconstruction in the atlas experiment in proton-proton collisions at $\sqrt{s} =$ 8 tev. *ATLAS-CONF-2014-019*, 2014.

[63] ATLAS Collaboration. Measurement of the missing transverse momentum based on tracks in proton-proton collisions at $\sqrt{s} = 900$ gev center-of-mass energy with the atlas detector. *ATLAS-CONF-2010-020*, 2010.

[64] ATLAS Collaboration. Measurement of track-based missing transverse momentum in proton-proton collisions at $\sqrt{s} = 8$ tev center-of-mass energy with the atlas detector. *ATLAS-CONF-2013-1577*, 2013.

[65] Tevatron New Physics Higgs Working Group, CDF Collaboration, D0. Updated Combination of CDF and D0 Searches for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data. *FERMILAB-CONF-12-318-E*, 2012.

[66] ATLAS Collaboration. Combined search for the Standard Model Higgs boson up to 4.9 fb^{-1} of pp collision data at sqrt(s)=7 TeV with the ATLAS detector at the LHC. 2012.

[67] CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Phys. Lett. B*, 716:30–61. 59 p, 2012.

[68] 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*, 716:1–29. 39 p, 2012.

[69] ATLAS Collaboration. Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC. *Physics Letters B*, 726(1 - 3):88 – 119, 2013. ISSN 0370-2693. doi: http://dx.doi.org/10.1016/j.physletb.2013.08.010. URL `http://www.sciencedirect.com/science/article/pii/S0370269313006369`.

[70] The ATLAS Collaboration. Measurements of the properties of the Higgs-like boson in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ decay channel with the ATLAS detector using 25 fb^{-1} of proton-proton collision data. *ATL-COM-PHYS-2013-030*, 2013.

[71] A. Bredenstein, Ansgar Denner, S. Dittmaier, and M.M. Weber. Precise predictions for the Higgs-boson decay $H \to WW/ZZ \to 4$ leptons. *Phys.Rev.*, D74:013004, 2006. doi: 10.1103/PhysRevD.74.013004.

[72] A. Bredenstein, A. Denner, S. Dittmaier, and M. Weber. Radiative corrections to the semileptonic and hadronic Higgs-boson decays H $\to$ WW/ZZ $\to$ 4 fermions. *JHEP*, 0702:080, 2007.

[73] A. Djouadi, J. Kalinowski, and M. Spira. HDECAY: A program for Higgs boson decays in the Standard Model and its supersymmetric extension. *Comput. Phys. Commun.*, 108:56, 1998. doi: 10.1016/S0010-4655(97)00123-9.

[74] S. Dawson. Radiative corrections to Higgs boson production. *Nucl.Phys.*, B359:283–300, 1991. doi: 10.1016/0550-3213(91)90061-2.

[75] M. Spira, A. Djouadi, D. Graudenz, and P.M. Zerwas. Higgs boson production at the LHC. *Nucl.Phys.*, B453:17–82, 1995. doi: 10.1016/0550-3213(95)00379-7.

[76] R. Harlander and W. B. Kilgore. Next-to-next-to-leading order Higgs production at hadron colliders. *PRL*, 88:201801, 2002. doi: 10.1103/PhysRevLett.88.201801.

[77] Charalampos Anastasiou and Kirill Melnikov. Higgs boson production at hadron colliders in NNLO QCD. *Nucl.Phys.*, B646:220–256, 2002. doi: 10.1016/S0550-3213(02)00837-4.

[78] V. Ravindran, J. Smith, and W. L. van Neerven. NNLO corrections to the total cross-section for Higgs boson production in hadron hadron collisions. *Nucl.Phys.*, B665:325–366, 2003. doi: 10.1016/S0550-3213(03)00457-7.

[79] U. Aglietti, R. Bonciani, G. Degrassi, and A. Vicini. Two loop light fermion contribution to Higgs production and decays. *Phys.Lett.*, B595:432–441, 2004. doi: 10.1016/j.physletb.2004.06.063.

[80] Stefano Actis, Giampiero Passarino, Christian Sturm, and Sandro Uccirati. NLO electroweak corrections to Higgs boson production at hadron colliders. *PLB*, 670:12, 2008. doi: 10.1016/j.physletb.2008.10.018.

[81] S. Catani, D. de Florian, M. Grazzini, and P. Nason. Soft-gluon re-summation for Higgs boson production at hadron colliders. *JHEP*, 0307:028, 2003. doi: 10.1088/1126-6708/2003/07/028.

[82] Charalampos Anastasiou, Stephan Buehler, Franz Herzog, and Achilleas Lazopoulos. Inclusive Higgs boson cross-section for the LHC at 8 TeV. *JHEP*, 1204:004, 2012.

[83] D. de Florian and M. Grazzini. Higgs production at the LHC: updated cross sections at $\sqrt{s} = 8$ TeV. 2012.

[84] J. Baglio and A. Djouadi. Higgs production at the LHC. *JHEP*, 1103:055, 2011. doi: 10.1007/JHEP03(2011)055.

[85] Paolo Bolzoni, Fabio Maltoni, Sven-Olaf Moch, and Marco Zaro. Higgs production via vector-boson fusion at NNLO in QCD. *Phys.Rev.Lett.*, 105:011801, 2010. doi: 10.1103/PhysRevLett.105.011801.

[86] M. Ciccolini, Ansgar Denner, and S. Dittmaier. Strong and electroweak corrections to the production of Higgs+2 jets via weak interactions at the LHC. *Phys. Rev. Lett.*, 99:161803, 2007. doi: 10.1103/PhysRevLett.99.161803.

[87] Mariano Ciccolini, Ansgar Denner, and Stefan Dittmaier. Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC. *Phys. Rev. D*, 77:013002, 2008. doi: 10.1103/PhysRevD.77.013002.

[88] K. Arnold et al. VBFNLO: A parton level Monte Carlo for processes with electroweak bosons. *Comput. Phys. Commun.*, 180:1661, 2009. doi: 10.1016/j.cpc.2009.03.006.

[89] Tao Han and S. Willenbrock. QCD correction to the $pp \to WH$ and ZH total cross-sections. *Phys.Lett.*, B273:167–172, 1991. doi: 10.1016/0370-2693(91)90572-8.

[90] Oliver Brein, Abdelhak Djouadi, and Robert Harlander. NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders. *Phys.Lett.*, B579:149–156, 2004. doi: 10.1016/j.physletb.2003.10.112.

[91] M. L. Ciccolini, S. Dittmaier, and M. Krämer. Electroweak radiative corrections to associated WH and ZH production at hadron colliders. *Phys. Rev. D*, 68:073003, 2003. doi: 10.1103/PhysRevD.68.073003.

[92] P. Nason and C. Oleari. NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG. *JHEP*, 1002:037, 2010. doi: 10.1007/JHEP02(2010)037.

[93] T. Sjöstrand, S. Mrenna, and P. Z. Skands. A brief introduction to PYTHIA 8.1. *Computer Physics Communications*, 178(11):852–867, 2008.

[94] T. Sjostrand, S. Mrenna, and P. Z. Skands. PYTHIA 6.4 physics and manual. *JHEP*, 0605:026, 2006. doi: 10.1088/1126-6708/2006/05/026.

[95] T. Gleisberg et al. Event generation with SHERPA 1.1. *JHEP*, 0902:007, 2009. doi: 10.1088/1126-6708/2009/02/007.

[96] Nikolas Kauer and Giampiero Passarino. Inadequacy of zero-width approximation for a light Higgs boson signal. *JHEP*, 1208:116, 2012. doi: 10.1007/JHEP08(2012)116.

[97] T. Binoth, M. Ciccolini, N. Kauer and M. Krämer. Gluon-induced w-boson pair production at the lhc. *JHEP*, 0612:046, 2006. doi: 10.1088/1126-6708/2006/12/046.

[98] G. Corcella et al. HERWIG 6: An event generator for hadron emission reactions with interfering gluons (including super-symmetric processes). *JHEP*, 0101:010, 2001. doi: 10.1088/1126-6708/2001/01/010.

[99] Stefano Frixione and Bryan R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 0206:029, 2002. doi: 10.1088/1126-6708/2002/06/029.

[100] B. P. Kersevan and E. Richter-Was. The Monte Carlo event generator AcerMC version 2.0 with interfaces to PYTHIA 6.2 and HERWIG 6.5. TPJU-6/2004, 2004.

[101] Johan Alwall et al. MadGraph/MadEvent v4: The new web generation. *JHEP*, 0709:028, 2007. doi: 10.1088/1126-6708/2007/09/028.

[102] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : going beyond. *JHEP*, 1106:128, 2011. doi: 10.1007/JHEP06(2011)128.

[103] Richard C. Gray, Can Kilic, Michael Park, Sunil Somalwar, and Scott Thomas. Backgrounds to Higgs boson searches from $W\gamma^* \to l\nu l(l)$ asymmetric internal conversion. 2011.

[104] J. M. Butterworth, Jeffrey R. Forshaw, and M. H. Seymour. Multiparton interactions in photoproduction at HERA. *Z. Phys.*, C72:637, 1996. doi: 10.1007/s002880050286.

[105] J. Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J. C*, 53:473, 2008. doi: 10.1140/epjc/s10052-007-0490-5.

[106] John M. Campbell and R. Keith Ellis. An update on vector boson pair production at hadron colliders. *Phys. Rev. D*, 60:113006, 1999. doi: 10.1103/PhysRevD.60.113006.

[107] H.-L. Lai et al. New parton distributions for collider physics. *Phys. Rev. D*, 82:074024, 2010. doi: 10.1103/PhysRevD.82.074024.

[108] P. M. Nadolsky et al. Implications of CTEQ global analysis for collider observables. *Phys. Rev. D*, 78:013004, 2008. doi: 10.1103/PhysRevD.78.013004.

[109] ATLAS Collaboration. The ATLAS simulation infrastructure. *Eur. Phys. J. C*, 70:823, 2010. doi: 10.1140/epjc/s10052-010-1429-9.

[110] S. Agostinelli et al. Geant 4, a simulation toolkit. *Nucl. Instrum. Meth.*, A506: 250, 2003. doi: 10.1016/S0168-9002(03)01368-8.

[111] HWW working group. Object selection for the $H \to WW^{(*)} \to \ell\nu\ell\nu$ search with the ATLAS detector at $\sqrt{s} = 8$ TeV. *ATL-COM-PHYS-2013-147*, Feb 2013.

[112] ATLAS Collaboration. Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung. *ATLAS-CONF-2012-047*, CERN, Geneva, May 2012.

[113] ATLAS Collaboration. Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector. *ATLAS-CONF-2011-063*, 2011. https://cdsweb.cern.ch/record/1345743.

[114] M. Cacciari, G. P. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 0804:063, 2008. doi: 10.1088/1126-6708/2008/04/063.

[115] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys.Lett.*, B659:119–126, 2008. doi: 10.1016/j.physletb.2007.09.077.

[116] Jet energy scale and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ tev with atlas 2011 data. *ATLAS-CONF-2013-004*, 2013.

[117] ATLAS Collaboration. Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data. *ATLAS-CONF-2011-102*, 2011. https://cdsweb.cern.ch/record/1369219.

[118] ATLAS Collaboration. Measurement of the b-tag efficiency in a sample of jets containing muons with 5 fb^{-1} of data from the ATLAS detector. *ATLAS-CONF-2012-043*, 2012. https://cdsweb.cern.ch/record/1435197.

[119] Calibrating the b-tag and mistag efficiencies of the sv0 b-tagging algorithm in 3 pb^{-1} of data with the atlas detector. Technical Report ATLAS-CONF-2010-099, CERN, Geneva, Dec 2010.

[120] Calibrating the b-tag efficiency and mistag rate in 35 pb^{-1} of data with the atlas detector. *ATLAS-CONF-2011-089*, 2011.

[121] ATLAS Collaboration. VBF $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis of $20.7 fb^{-1}$ at $\sqrt{s} = 8$ TeV: Moriond 2013 update: cut-based em+me and ee+mm. Technical Report ATL-COM-PHYS-2013-156, CERN, Geneva, Feb 2013. URL `https://cds.cern.ch/record/1514562`.

[122] R.K. Ellis et al. Higgs decay to $\tau^+\tau^-$: A possible signature of intermediate mass Higgs bosons at the SSC. *Nucl. Phys. B*, 297:221, 1988. doi: 10.1016/0550-3213(88)90019-3.

[123] HSG3 team group. Analysis of $H \to WW^{(*)} \to \ell\nu\ell\nu$ in the same-flavour channels. Technical Report ATL-COM-PHYS-2012-1445, CERN, Geneva, Sep 2012. URL `https://cds.cern.ch/record/1482155`.

[124] ATLAS Collaboration. Search for the Standard Model Higgs boson in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ decay mode with 4.7 fb^{-1} of ATLAS data at $\sqrt{s} = 7$ TeV. *Phys.Lett.*, B716:62–81, 2012. doi: 10.1016/j.physletb.2012.08.010.

[125] HSG3 team group. Background estimates in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis with 21 fb^{-1} of data collected with the ATLAS detector at $\sqrt{s} = 8$ TeV. Technical Report ATL-COM-PHYS-2013-139, CERN, Geneva, Feb 2012. URL `https://cds.cern.ch/record/1514189`.

[126] B. Mellado, X. Ruan, and Z. Zhang. Extraction of top backgrounds in the Higgs boson search with the $H \to WW^\star \to \ell\ell + E_{\mathrm{T}}^{\mathrm{miss}}$ decay with a full-jet veto at the LHC. *Phys. Rev. D*, 84:096005, 2011. doi: 10.1103/PhysRevD.84.096005.

[127] I. Stewart and F. Tackmann. Theory uncertainties for Higgs mass and other searches using jet bins. *Phys. Rev. D*, 85:034011, 2012. doi: 10.1103/PhysRevD.85.034011.

[128] Stefano Catani and Massimiliano Grazzini. An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC. *Phys. Rev. Lett.*, 98:222002, 2007. doi: 10.1103/PhysRevLett.98.222002.

[129] Massimiliano Grazzini. NNLO predictions for the Higgs boson signal in the $H \to WW \to l\nu l\nu$ and $H \to ZZ \to 4l$ decay channels. *JHEP*, 0802:043, 2008. doi: 10.1088/1126-6708/2008/02/043.

[130] ATLAS and CMS Collaborations. Procedure for the lhc higgs boson search combination in summer 2011. *ATL-PHYS-PUB-2011-011, CMS-NOTE-2011-005*, 2011. https://cdsweb.cern.ch/record/1375842.

[131] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Hadronic production of a Higgs boson and two jets at next-to-leading order. *Phys.Rev.*, D81:074023, 2010. doi: 10.1103/PhysRevD.81.074023.

[132] M. Botje et al. The PDF4LHC working group interim recommendations. 2011.

[133] R. D. Ball et al. Impact of heavy quark masses on parton distributions and LHC phenomenology. *Nucl. Phys. B*, 849:296, 2011. doi: 10.1016/j.nuclphysb.2011.03.021.

[134] R. Frederix, S. Frixione, F. Maltoni and T. Stelzer. Automation of next-to-leading order computations in QCD: The FKS subtraction. *JHEP*, 0910:003, 2009.

[135] V. Hirschi, R. Frederix, S. Frixione, M. V. Garzelli, F. Maltoni and R. Pittau. Automation of one-loop QCD corrections. *JHEP*, 1105:044, 2011.

[136] ATLAS Collaboration. Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC. 2013.

[137] Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV. 2011. Submitted to Eur. Phys. J. C.

[138] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. doi: 10.1140/epjc/s10052-011-1554-0.

[139] A. Stuart, J.K. Ord, and S. Arnold. Advanced Theory of Statistics. *Classical Interference and the Linear Model*, Vol. 2A, 1999.

[140] Virgil L. Highland. Estimation for Upper Limits from Experimental Data. *Template University Report*, July 1986, Revised Frebruary 1987.

[141] A.L. Read. Presentation of search results: the CL_s technique. *J. Phys. G*, 28: 2693, 2002. doi: 10.1088/0954-3899/28/10/313.

[142] ATLAS Collaboration. Update of the $H \to WW^{(*)} \to e\nu\mu\nu$ analysis with 13 fb^{-1} of $\sqrt{s} = 8$ data collected with the ATLAS Detector. ATLAS-CONF-2012-158, 2012. https://cdsweb.cern.ch/record/1493601.

[143] ATLAS Collaboration. Combined measurements of the mass and signal strength of the Higgs-like boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton collision data. Technical Report ATLAS-CONF-2013-014, CERN, Geneva, Mar 2013. URL `https://cds.cern.ch/record/1523727`.

[144] ATLAS Collaboration. Measurements of the properties of the Higgs-like boson in the four lepton decay channel with the ATLAS detector using 25 fb^{-1} of proton-proton collision data. *ATLAS-CONF-2013-013*, 2013. https://cdsweb.cern.ch/record/1523699.

[145] ATLAS Collaboration. Measurements of the properties of the Higgs-like boson in the two photon decay channel with the ATLAS detector using 25 fb^{-1} of proton-proton collision data. *ATLAS-CONF-2013-012*, 2013. https://cdsweb.cern.ch/record/1523698.

[146] S. Dittmaier, C. Mariotti, G. Passarino and R. Tanaka. The Higgs Cross Section Working Group web page. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections, 2012.

[147] ATLAS Collaboration. Pile-up Reweighting in ATLAS, 2011. URL `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/PileupReweighting`.

[148] HWW working group. Object selections in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis with 20.3 fb^{-1} of data collected with the ATLAS detector at $\sqrt{s} = 8$ TeV. *ATL-COM-PHYS-2013-1504*, July 2014.

[149] ATLAS Collaboration. Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data. *CERN*, January 2015. URL `https://cds.cern.ch/record/1706245`.

[150] L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. *Chapman and Hall, New York*, January 1, 1984.

[151] Yoav Freund and Robert E Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1):119 – 139, 1997. ISSN 0022-0000. doi: http://dx.doi.org/10.1006/jcss.1997.1504. URL `http://www.sciencedirect.com/science/article/pii/S002200009791504X`.

[152] Jerome H. Friedman. Stochastic gradient boosting. *Computational Statistics And Data Analysis*, 38(4):367 – 378, 2002. ISSN 0167-9473. doi: http://dx.doi.org/10.1016/S0167-9473(01)00065-2. URL `http://www.sciencedirect.com/science/article/pii/S0167947301000652`. Nonlinear Methods and Data Mining.

[153] ATLAS Collaboration. Measurement of the higgs boson mass from the $h \to \gamma\gamma$ and $h \to zZ^* \to 4\ell$ channels in pp collisions at center-of-mass energies of 7 and 8 tev with the atlas detector. *Phys. Rev. D*, 90:052004, Sep 2014. doi: 10.1103/PhysRevD.90.052004. URL `http://link.aps.org/doi/10.1103/PhysRevD.90.052004`.

[154] ATLAS Collaboration. Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data. *Eur. Phys. J. C*, 72:1909, 2012. doi: 10.1140/epjc/s10052-012-1909-1.

[155] T. Barillari et al. Local Hadronic Calibration. 2009.

[156] Determination of the tau energy scale and the associated systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ tev with the atlas detector at the lhc in 2011. *ATLAS-CONF-2012-054*, 2012.

List of Figures

List of Tables