UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE E TECNOLOGIE

Corso di Dottorato in Fisica, Astrofisica e Fisica Applicata

Ciclo XXVI

Measurement of missing transverse momentum and search for $H \to \tau\tau$ in p-p collisions at $\sqrt{s} = 8$ TeV with ATLAS

Settore Scientifico Disciplinare FIS/01 e FIS/04

Supervisor: Dott. Donatella CAVALLI

Cosupervisor: Dott. Leonardo CARMINATI

Coordinator: Prof. Marco BERSANELLI

Tesi di Dottorato di:

Rosa Simoniello

Anno Accademico 2012/2013

Commission of the final examination:

External Referee:
Prof. Peter LOCH

External Member:
Prof. Marumi KADO

External Member:
Prof. Monica D'ONOFRIO

Internal Member:
Prof. Emanuela MERONI

Final examination:

Date 31/01/2014

Università degli Studi di Milano, Dipartimento di Fisica, Milano, Italy

PACS:

14.80.Bn

07.05.Kf

29.40.Vj

Introduction

The Large Hadron Collider (LHC) at CERN, Geneva, Switzerland, collides protons at unprecedented energies and with the highest luminosities achieved in a collider so far. The proton-proton (pp) collisions provide access to new and far extended kinematic regimes for the two multi-purpose experiments (ATLAS [1] and CMS [2]), and a dedicated experiment analyzing heavy quark production (LHCb [3]). In particular ATLAS and CMS are well suited to not only explore new physics beyond the Standard Model (SM), but also to test SM predictions for elementary particles and their interactions.

Understanding the origin of the mass of elementary particles has been one of the main quests for high energy physics for the last decades. The theory of the mass generating mechanism was independently hypothesized by Brout, Englert and Higgs in 1964 [4, 5]. It has been experimentally confirmed with the discovery of the particle predicted by this theory, the Higgs boson, announced by the ATLAS and CMS experiments on July 4, 2012 [6, 7]. At the time of this announcement only Higgs couplings to bosons could directly be proven. Subsequent crucial tests of the theory require to prove that the Higgs boson also couples to fermions.

The most promising channel for this search is the $H \to \tau\tau$ process that is discussed in this thesis. One of the aspects that makes the analysis of this channel challenging is the need of a proper evaluation of the missing transverse energy ($\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$) introduced by neutrinos in the final state. For this reason, a substantial part of this thesis is devoted to the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, calibration and to techniques for pile-up suppression. More specifically, this thesis reports my work in the last three years developed in the context of the ATLAS collaboration and of the Milano University group, following the full path from low-level detector signals, through $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction and calibration, to the $H \to \tau\tau$ physics analysis. The thesis is organized as follows.

In Chapter 1, an introduction to the theoretical framework of the Standard Model is given together with a description of the mass mechanism and its phenomenology crucial for the analysis.

In Chapter 2, the LHC, the ATLAS experiment, and the experimental conditions during the data taking in 2011 and 2012, which are characterized by a large amount of pile-up interactions introduced by the high luminosity reached with the LHC, are introduced. Particular attention is given to the ATLAS calorimeter system since its importance for the subjects dealt in this thesis.

In Chapter 3, the algorithms for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction used in ATLAS are described. The $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance, which has been extensively studied, is presented for several final state topologies, and the crucial aspects for the $H \to \tau\tau$ analysis are discussed. This chapter represents significant input from my original studies.

In Chapter 4, the degradation of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance due to pile-up is shown and the methods to mitigate these effects are described. I have devoted large part of my activity in the development of innovative methods for pile-up suppression using either tracks or calorimeter information or a combination of both, and to their implementation in the official collaboration software package.

In Chapter 5, the cut-based analysis for the search of $H \to \tau\tau$ in the semileptonic $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ final state is described in all its main aspects, from the selection and categorization of the events, the estimation of the backgrounds, the statistical analysis, to the final signal extraction, where I have been more involved. In particular, it is shown how the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ after pile-up suppression improves the analysis sensitivity. The same pile-up suppressed $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is also used in the ATLAS $H \to \tau\tau$ analysis employing multivariate techniques, which ultimately led to the observation of the Higgs boson decaying into two taus.

Contents

Chapter 1

Introduction to the Standard Model and the Higgs Boson

The Standard Model (SM) is a quantum field theory that describes the matter constituents and their interactions. It provides at present our best understanding of the phenomenology of particle physics. Experiments performed in the last 70 years have successfully tested the model to an impressive level of accuracy.

Special acknowledgments were received with several Nobel prizes assigned to particle physics achievements. The Nobel prize in 1979 to Glashow, Salam and Weinberg for the electroweak unification and the prediction of the neutral currents observed for the first time by the Gargamelle collaboration [8]. The Nobel prize in 1984 to Rubbia and Van Der Meer for their decisive contributions to the large project that led to the discovery of the W and Z bosons by UA1 and UA2 experiments [9, 10]. The Nobel prize in 2008 to Kobayashi and Maskawa for the formulation of the CP (charge-parity) violation extensively studied in B-meson system by the Babar and Belle collaborations. Finally this year (2013) the Nobel prize to Englert and Higgs for the mass mechanism and the prediction of the Higgs boson recently observed by the ATLAS and CMS collaboration [6, 7].

In this chapter, the mathematical formulations of the SM are discussed in Sections 1.1 to 1.3. It includes the main ideas of modern physics: fields, quantum mechanics and special relativity. Local gauge symmetry and spontaneous

symmetry breaking are key concepts to provide dynamics and mass mechanisms for SM particles. Experimentally, the mass mechanism results in an observable massive particle, the Higgs Boson: the status of ATLAS searches and measurements of the Higgs properties are briefly summarized in Section 1.4. Despite the huge success of the SM predictions, there are hints indicating that the SM might not be consider as complete. Conceptual problems and limitations of the SM are highlighted in Section 1.5.

1.1 Basic constituents of the Standard Model

Phenomenologically, four fundamental interactions are observed in nature: electromagnetic, weak, strong and gravity. In the SM framework they are described by vector boson fields (the *gauge fields*) and represented as exchange of particles with spin 1, the force carriers. Gravity is not included in the theory because a consistent gravitational quantum formulation does not yet exist. However, at the energy scales currently accessible in experiments (order 1 TeV) gravity is weaker than the electromagnetic force by 40 orders of magnitude so its effects are negligible. They are expected to become relevant only at a much higher energy scale, the Planck scale ($1.22{\cdot}10^{19}$ GeV).

The observable vector bosons associated respectively to the strong, electromagnetic and weak interaction are the *gluons*, the *photon* and the bosons $W^{\pm}$ and Z^0. The weakness of the weak interaction is due to its massive force mediators. Despite its strength, the weak interaction gives rise to distinctive experimental signatures because it violates parity P, charge conjugation C, their combination CP, time-reversal T, and fermion generation, which are conserved by electromagnetic and strong interactions.

Matter constituents are elementary particles of spin 1/2 (*fermions*) divided into three families with increasing mass. They are represented by complex spinor fields $\psi(x)$ obeying the Dirac equation:

$$(i\gamma^{\mu}\partial_{\mu} - m)\psi(x) = 0 \tag{1.1}$$

where γ^{μ} are the Dirac matrices and m is the fermion mass. Dirac's equation has also a negative energy solution which is interpreted as the corresponding *antiparticle* of each particle, exactly matching the particle but with opposite charge.

Each fermion is associated to two chiralities, *left-handed* and *right-handed*. Chirality is conserved for massless fermions, in which case the chirality coincides

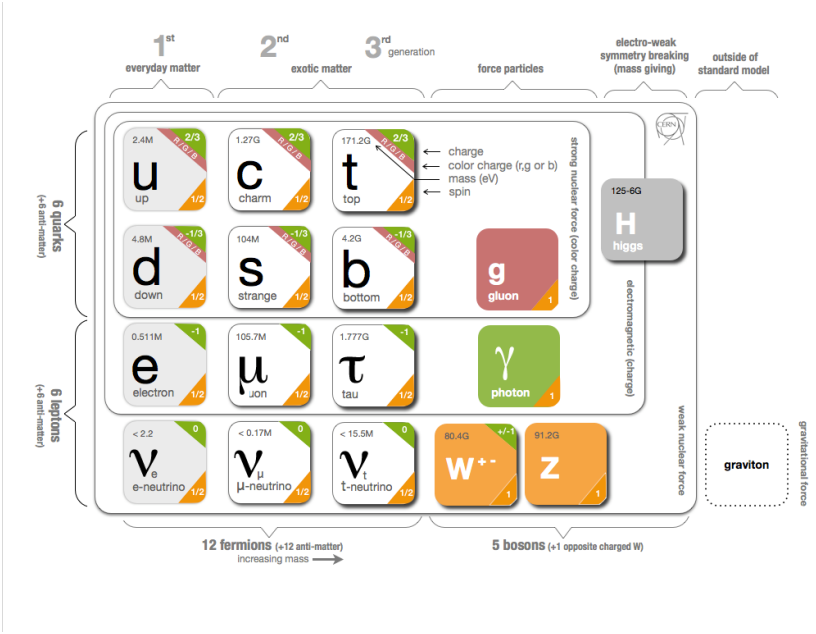


Figure 1.1: List of SM elementary and their interactions [14]. Particles are characterized by their spin, mass, and the quantum numbers (charges) determining their interactions. Moving between fermion generations quantum numbers stay the same while the particle mass increases for higher generations. The heavier generations are unstable and decay into the lightest one, which makes up most of the ordinary matter.

with the helicity. Despite the results from neutrino oscillation experiments [11, 12, 13], in the SM the neutrinos (ν_ℓ) are considered massless, so right-handed neutrinos are not foreseen in the theory.

The coupling with gauge fields provides the interaction between the otherwise free particles. Fermions can be classified according to the interactions experienced. All of them interact by the electroweak force. The fermions which carry also a color quantum number (red, blue, green), the charge of the strong interaction, are classified as *quarks* otherwise they are classified as *leptons*. The elementary particles in the SM with their classification and properties are summarized in Figure 1.1.

Quarks are never observed in isolation but always bound in color-singlet particles, the *hadrons*. Integer spin particles ($q\overline{q}$ states) are called *mesons* while

half-spin particles (qqq states) are called *baryons*.

As shown in the next sections the SM embodies the phenomenological structure described above in an elegant mathematical formulation where the concept of symmetry plays a key role for the construction of the theory.

1.2 Symmetries of the Standard Model

The construction of the Standard Model has been guided by covariance and local gauge symmetry principles [15, 16, 17]. The covariance (invariance under Lorentz transformation) ensures the compatibility of the theory with the special relativity, while symmetries provide information to determine the conservation laws and the dynamics of the interactions in quantum field theory:

Conservation laws are established by the Noether's theorem stating that if an action (integral of the Langrangian over time) is invariant under some group of transformations (symmetry), then there exist conserved quantities which are associated to these transformations.

The dynamics of the SM interaction fields is realized asking for invariance of the theory under local gauge transformations $U(x)$ depending on space and time. Gauge transformations are important because they relate equivalent field configurations accounting for the redundant degrees of freedom in the Lagrangian and associate them with a unique physics observable. Moreover they imply the renormalizability of the theory ensuring the correct unitary behavior at high energy.

Applying the guidelines listed above, the first step is to apply a local gauge transformation $U(x)$ to the fermion field:

$$\psi(x) \rightarrow \psi^{'}(x) = U(x)\psi(x) \tag{1.2}$$

The gauge transformation U(x) is usually expressed, for each symmetry of the theory, using the hermitian generators, H_j, of the Lie group:

$$U(x) = e^{-i\Sigma_{j=1}^{n}\theta_j(x)H_j} \tag{1.3}$$

where $\theta(x)$ is an arbitrary function.

Substituting expressions 1.2 and 1.3 in the Dirac's equation 1.1, the presence of a partial derivative leads to additional terms that break the invariance of the

equation:

$$\begin{aligned}\partial_\mu\psi(x) \to \partial_\mu\psi'(x) &= e^{i\theta(x)\Sigma_{j=1}^{n}H_j}\ \left[\partial_\mu + i\partial_\mu\theta(x)\Sigma_{j=1}^{n}H_j\right]\psi(x) \\ &\neq e^{i\theta(x)\Sigma_{j=1}^{n}H_j}\ \partial_\mu\psi(x).\end{aligned} \tag{1.4}$$

The invariance can be restored by introducing a spin-1 field $A_\mu(x)$, the gauge field, with coupling constant g, to the partial derivative ∂_μ to form the covariant derivative D_μ:

$$\partial_\mu \to D_\mu = \partial_\mu + ig\Sigma_{j=1}^{n}A_\mu(x) \tag{1.5}$$

In this sense one can say that the gauge field responsible for the interaction is brought into the free particle theory by imposing gauge invariance, and the complete dynamic is obtained as follow.

The expression for the transformation of the field $A_\mu(x)$ under $U(x)$ is determined by the invariance requirements of the theory, it must generate terms canceling the non-invariant contributions in Dirac's equation:

$$A_\mu(x) \to A_\mu^{j\prime}(x) = A_\mu^j(x) - \frac{1}{g}\partial_\mu\theta^j(x) - \Sigma_{k,l}h_{jkl}\theta_k(x)A_\mu^l(x) \tag{1.6}$$

where the h_{jkl} are the structure constants of the Lie algebra with generators H_j.

The field strength tensor characterizing the gauge interaction is defined as:

$$F_{\mu\nu}^i = \partial_\mu A_\nu^i - \partial_\nu A_\mu^i - gh_{ijk}A_\mu^j A_\nu^k \tag{1.7}$$

Terms proportional to $F_{\mu\nu}F^{\mu\nu}$ in the Lagrangian are identified with the kinetic terms for the gauge fields.

Theories can be classified as Abelian groups, like $U(1)$, that have not field self-couplings, and non-Abelian groups (also called Yang-Mills theories), like $SU(N)$ that instead imply field self-interaction. This can be understood from the absence of the last term in Equations 1.6 and 1.7 for the Abelian case. Therefore, the Lagrangian has no trilinear or quartic terms in the fields responsible for the self-couplings in this case.

According to the general case discussed above, Quantum Electrodynamics (QED) [18, 19] can be derived from equations 1.2 to 1.6 by asking for local invariance under the $U(1)$ group, $U(1)_{EM}$, of the form $U(x) = e^{i\theta(x)}$. The associated gauge field is identified with the photon and the coupling strength between the photon and a fermion is the electric charge of the fermion. Since $U(1)_{EM}$ is an Abelian group the gauge fields does not experience self-interaction, therefore, the photon does not carry electromagnetic charge.

The QED symmetry group is not directly included in the SM model because can be seen as the low energy theory of a more fundamental symmetry realized at higher energy. In this regime the electromagnetic and the weak interactions are unified in the electroweak force that breaks down in the two separate forces at the electroweak scale, O(100 GeV).

Therefore, in order to match the particle phenomenology described in Section 1.1, the minimal global symmetry group for the SM is:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{1.8}$$

where the $SU(3)_C$ group determines the strong interaction sector, while the $SU(2)_L \otimes U(1)_Y$ group determines the electroweak interaction sector. The independent product of the three symmetry groups results into three independent coupling constants, one for each group. Further details are in the next sections.

1.2.1 The strong interaction sector $SU(3)_C$

The quantum chromodynamics (QCD) [20, 21] describes the strong interaction between quarks. It is based on the $SU(3)_C$ group, where C stands for color. Each quark flavour corresponds to an $SU(3)_C$ quark triplet in a three-dimensional color space with base (red, blue, green).

For the $SU(3)_C$ group there are eight generators expressed by the Gell-Mann matrices and resulting in the same number of massless vector fields, the gluons, G^{μ}_{a}, $a = 1, ..., 8$. Since $SU(3)_C$ is a non-Abelian group, gluon self-interaction occurs implying that gluons carry color charge.

The coupling strength g_S is more commonly expressed in terms of the strong coupling constant $\alpha_S = g_S^2/(4\pi)$. As result of the renormalization, which cancels divergences in the theory, the coupling constant depends on the renormalization scale, μ^2, and on the virtuality (momentum transfer) of the process, Q^2, (running coupling constant) [22]. Even though QCD does not give a prediction for the absolute value of α_S, its form is completely determined by the renormalization group equations (RGE). At the lowest order it can be expressed by the following equation:

$$\alpha_S(Q^2) = \frac{\alpha_S(\mu^2)}{1 + \alpha_S(\mu^2)\beta_0 ln\frac{Q^2}{\mu^2}} \tag{1.9}$$

where $\beta_0 = (33 - 2N_f)/(12\pi) > 0$, and N_f is the number of the accessible quark flavors.

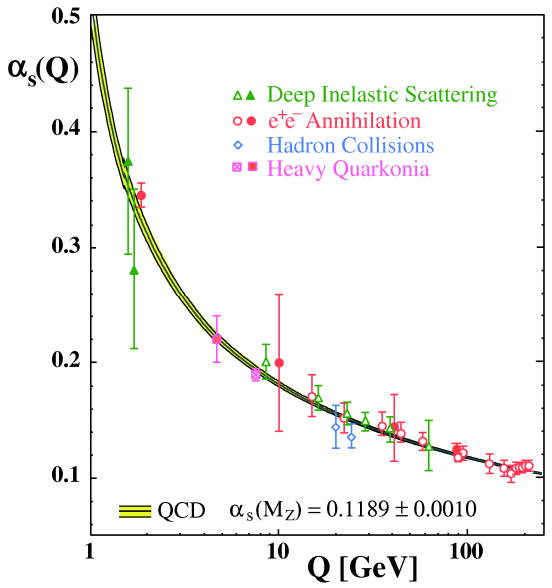


Figure 1.2: Summary of measurements of $\alpha_S(Q)$ as a function of the respective energy scale Q. Open symbols indicate Next-to-Leading Order (NLO), and filled symbols Next-to-next-to-Leading Order (NNLO) QCD calculations used in the respective analysis. The curves are the QCD predictions for the combined world average value of α_S, evaluated at mass M_Z of the Z^0 boson [26].

The prediction of the α_S dependence on Q^2 has been successfully tested by many experiments, as showed in Figure 1.2. $\alpha_S(Q^2)$ decreases with increasing Q^2 of the process. This means that the higher the available energy in the process becomes, the more the quarks can be considered as free particles called partons (*asymptotic freedom*) [23, 24, 25]. On the contrary $\alpha_S(Q^2)$ diverges for low energy values and exceeds unity for $Q^2 < 1$ GeV. In this region the force gets so strong that it is impossible to extract a single quark from a hadron. This is understood from the fact that if the quark receives enough energy to overcome the binding energy of the hadron, it also has sufficient energy to produce quark-antiquark pair until the color charge is neutralized and all quarks are bound into color singlets (*confinement*). Finally, at large α_S, perturbative expansions in α_S are not meaningful so only phenomenological model and numerical computation are presently available.

The cross section for high-energy hadron collisions can be described by a

"hard" contribution given by the parton level interaction, $\hat{\sigma}(ij \to X)$, convoluted with the Parton Distribution Functions (PDFs), $f_i(x, \mu_F^2)$, describing the probability for a parton of flavor i to carry a fraction x of the total momentum of the hadron:

$$\sigma(pp \to X) = \sum_{ij} \int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}(ij \to X). \tag{1.10}$$

PDFs depend on the factorization scale of the process, μ_F^2, that is needed to regularize collinear divergence for incoming partons.

This factorization is the base to describe all interesting processes (hard scatter interactions) at hadrons colliders.

1.2.2 The electroweak sector $SU(2)_L \otimes U(1)_Y$

The electroweak force is described by the symmetry group $SU(2)_L \otimes U(1)_Y$, with $U(1)_Y \neq U(1)_{EM}$ [16]. The three generators T_i, $i = 1, 2, 3$, (Pauli's matrices) of the $SU(2)_L$ group are called weak-isospin generators and act differently on left-handed and right-handed components (in particular, right-handed particles are singlets under the weak isospin, hence the subscript L that stands for left-handed). The generator of the $U(1)_Y$ group is called weak-hypercharge operator Y and it is related to the electric charge (Q) and the weak-isospin generators through the relationship:

$$Q = T_3 + \frac{1}{2} Y \tag{1.11}$$

The generators of the group result in the same number of massless gauge fields consisting of a vector boson triplet under $SU(2)_L$, W_μ^i, $i = 1, 2, 3$, and a vector boson singlet under $SU(2)_L$, B_μ.

The observed weak bosons, $W^\pm$ and Z^0, are massive. This can be generated in the theory through a system of spontaneous symmetry breaking and the Higgs mechanism, described in Section 1.3. The breaking scheme $SU(2)_L \otimes U(1)_Y \to U(1)_{EM}$ lets three independent linear combinations of the four gauge boson fields, identified with the $W^\pm$ and Z^0, to acquire mass while preserving the massless of the photon as required.

The observed states are therefore a mixing of the the gauge bosons:

$$W_\mu^\pm = \frac{W_\mu^1 \mp i W_\mu^2}{\sqrt{2}} \tag{1.12}$$

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \tag{1.13}$$

where θ_W is known as the Weinberg angle or electroweak mixing angle and it is related to the coupling constants by:

$$g \sin\theta_W = g' \cos\theta_W = e \tag{1.14}$$

$$\tan\theta_W = \frac{g'}{g} \tag{1.15}$$

At the time of the formulation of the electroweak interaction only charged weak currents mediated by $W^\pm$ were observed. The prediction of an additional neutral current mediated by the Z^0 boson and its first observation in the Gargamelle neutrino experiment [8] was a huge confirmation of the validity of the theory that was extensively tested by the LEP precision measurements on the Z mass pole [27]. These measurements still provide very stringent constraints on the model and they are useful in testing and possibly excluding new theories.

1.2.3 Gauge invariance and mass terms

As discussed in Section 1.2, the gauge invariance is one of key concepts for building the SM theory. However, invariance requirements also prevent to directly introduce mass terms for fermions and vector bosons in the SM Lagrangian.

For fermions, the mass terms would have the following form:

$$m\overline{\psi}\psi = m\left(\overline{\psi}_R\psi_L + \overline{\psi}_L\psi_R\right) \tag{1.16}$$

that is not $SU(2)_L$ invariant since left-handed and right-handed fermion fields transform in a different way under transformation.

In the same way, the mass term for vector bosons would have the form:

$$m^2 A^\mu A_\mu \tag{1.17}$$

that is not invariant under a gauge transformation of the field A^μ (equation 1.6).

Hence, SM particles should be massless but the mass of the fermions and the vector bosons $W^\pm$ and Z^0 are experimentally measured. The gauge symmetry

must therefore be broken in order to provide these masses. This can be done through the mechanism of spontaneous symmetry breaking. The main concept is that no explicit breaking of the symmetry is introduced: the equations of the dynamics are kept symmetric, but they can accommodate solutions that are not symmetric. Therefore, the system "spontaneously" breaks the symmetry choosing a ground state.

The explicit case of the Higgs mechanism and how the mass of the SM particles are generated are described in next session.

1.3 The mass mechanism and the Higgs boson

The Brout-Englert-Higgs mechanism [4, 5] is a way to provide mass for fermions and vector bosons without spoiling the gauge invariance of the Lagrangian. As argued in Section 1.2.3 this can be achieved with a spontaneous symmetry breaking mechanism (SSB).

In order to have a coherent theory consistent with experimental data, the minimal configuration to realize SSB in the SM Lagrangian is by introducing a complex scalar field, ϕ, with a Lagrangian $\mathcal{L}_{\text{Higgs}}$ defined as:

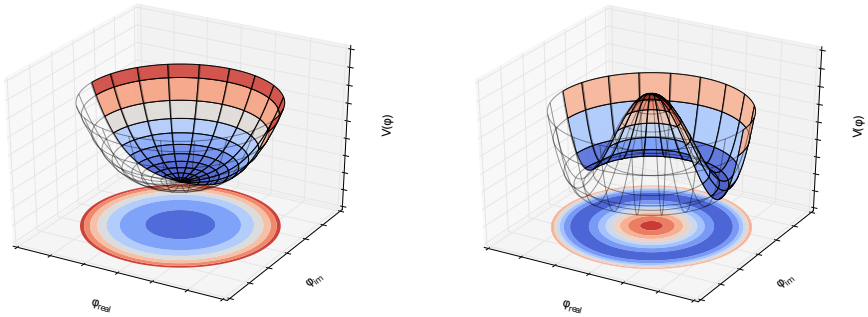
$$\mathcal{L}_{\text{Higgs}} = (\partial^\mu \phi)^\dagger (\partial_\mu \phi) - V(\phi^\dagger \phi), \qquad V(\phi^\dagger \phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \tag{1.18}$$

where the parameter λ must be not negative in order to avoid instability in the theory, while the parameter μ^2 (despite the notation) can be either positive or negative. The shape of the potential for the two cases is shown in Figure 1.3. For $\mu^2 > 0$ there is a unique minimum not allowing any symmetry breaking, whereas for $\mu^2 < 0$ the potential has a whole family of not trivial minima. Therefore, although the Lagrangian remains invariant under $SU(2)_L \otimes U(1)_Y$, the system choosing one of the equivalent ground states breaks the symmetry.

At this point, one can develop the theory around the chosen ground state. The Higgs doublet can be written in terms of four real fields $\theta_1(x)$, $\theta_2(x)$, $\theta_3(x)$ and $H(x)$ as:

$$\phi(x) = \frac{1}{\sqrt{2}} e^{i \frac{\tau_a}{2} \theta_a(x)} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \xrightarrow[\text{transf}]{\text{gauge}} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{1.19}$$

where τ_a are the three Pauli's matrices and v is the vacuum expectation value of the Higgs field. The exponential enclosing the dependence on the three scalar fields $\theta_a(x)$ can be removed with a gauge transformation, leaving only one physical



(a) $\mu^2 > 0$ (b) $\mu^2 < 0$

Figure 1.3: Shape of the potential for a complex scalar field with $\mu^2 > 0$ in (a) and with $\mu^2 < 0$ (b) [28]. For the case with $\mu^2 > 0$ there is only a trivial minimum, instead for $\mu^2 < 0$ a ring of minima is found. Despite the symmetric shape of the potential, the system, choosing one of the equivalent minima (ground state), spontaneously breaks the symmetry.

state, $H(x)$, the Higgs boson. The other three degrees of freedom, removed by the gauge transformation, reappear as mass terms for the the three weak bosons.

In fact, by using the representation in equation 1.19 for the Higgs field and developing its kinematic term using covariant derivative, terms of the form $M_H^2\phi^2$, $M_W^2 W^\mu W_\mu$ and $M_Z^2 Z^\mu Z_\mu$ are generated. They can be interpreted as mass terms respectively for the scalar Higgs boson and for the vector bosons $W^\pm$ and Z^0 with:

$$M_H = \sqrt{2\lambda} v M_W = M_Z \cos\theta_W = \frac{1}{2} g v \tag{1.20}$$

$$\tag{1.21}$$

where θ_W is the Weinberg angle introduced in Section 1.2.2.

Mass terms for fermions can be introduced in the Lagrangian via Yukawa coupling, λ_f, of the left-handed and right-handed fermion fields to the Higgs field that, expressing the Higgs field as in equation 1.19, have the following form for each fermion:

$$\mathcal{L}_{\text{fermion}} = -\frac{(v+H)}{\sqrt{2}} \lambda_f \overline{\psi}\psi \tag{1.22}$$

The term proportional to v has the right form of a mass term for fermions, $m_f \overline{\psi}\psi$,

with:

$$m_f = \lambda_f \frac{v}{\sqrt{2}} \tag{1.23}$$

while the term proportional to H gives the fermion-Higgs coupling. Using equation 1.23 the value of the coupling strength can be expressed as m_f/v, meaning the Higgs field couples to a fermion proportionally to the fermion mass. Since the Yukawa coupling are not a priori set, fermions masses are free parameters of the theory.

1.4 Status of the Higgs boson searches and measurements at LHC

As discussed in Section 1.3, the Brout-Englert-Higgs mechanism addresses the problem of the origin of mass for elementary particles introducing the interaction with a new field, the Higgs field, and predicts the existence of an observable massive particle, the Higgs boson, associated to the field. For many years, particle experiments have searched for this particle able to confirm and complete the SM theory. The difficulty of the search, beyond the small Higgs cross section and the compelling experimental conditions, lies in the a priori unknown mass of the Higgs boson. This means that the search must be performed across a wide range for the Higgs boson mass where the signal is predicted to manifest differently and different background compositions are expected. By consequence, in order to cover this variety of experimental signatures, several analyses have been set up and separately optimized.

The discovery of the Higgs boson on the 4th July 2012 has been a huge success for the physics community and the start of a new physics program. In the SM, all properties of the Higgs boson are defined once its mass is known, making precise measurements of those properties [29, 30, 31] crucial to provide further constraints on the theory. These measurements are also important also to test many alternative theories extending the SM that make different predictions for the properties of one or more Higgs bosons. They include measurement of the Higgs mass, width, quantum numbers (charge, spin, CP), differential cross sections and couplings, and are pursued with high priority at the LHC.

The status of these measurements in ATLAS are reported in the next sections after a description of the main Higgs production and decay modes in high energy proton-proton collisions ($\sqrt{s} = 7/8$ TeV at the LHC).

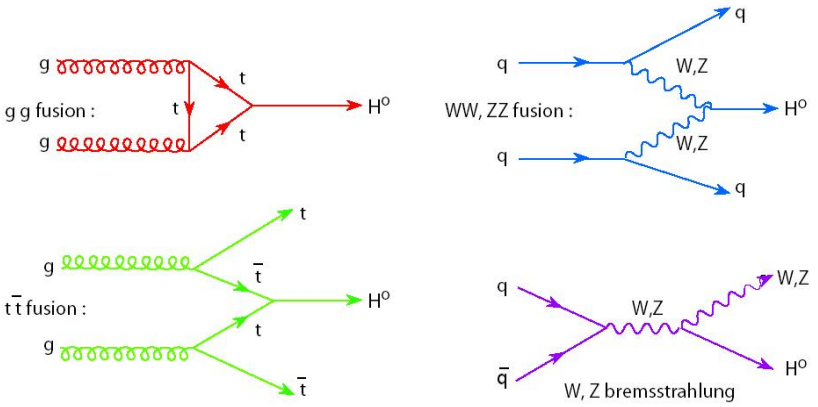


Figure 1.4: Main production modes of the Higgs boson at LHC [32].

1.4.1 Higgs production modes at LHC

At the LHC the Higgs boson is produced through the four main production modes shown in Figure 1.4. The cross sections for the dominant production modes at 8 TeV are compared in Figure 1.5 as a function of the Higgs mass [29] and listed below.

Gluon-gluon fusion is the dominant production mode, mediated by a top quark loop and, to a lesser extend, a b-quark loop. This process receives huge contributions from higher order QCD corrections. The uncertainties on the total values varies from 10% to 40% depending on the prescription used for their calculation.

Vector boson fusion (VBF) is the sub-leading mode at the LHC, with the Higgs produced in association with two quarks. The quarks are expected to give rise to very energetic jets located in the forward regions, with a large rapidity separation between them. Since its a very distinctive experimental signature, it provides a powerful filter against other backgrounds and where possible it is explored in the analysis final state.

Associated production with W or Z bosons initiated by $q\bar{q}$. The decays of the vector bosons to leptons (including neutrinos) provide good trigger efficiency and help in reducing QCD backgrounds. For this reason, in the last year at the increasing of the available luminosity it became more attractive at the LHC and more and more analyses started to explore how to fully

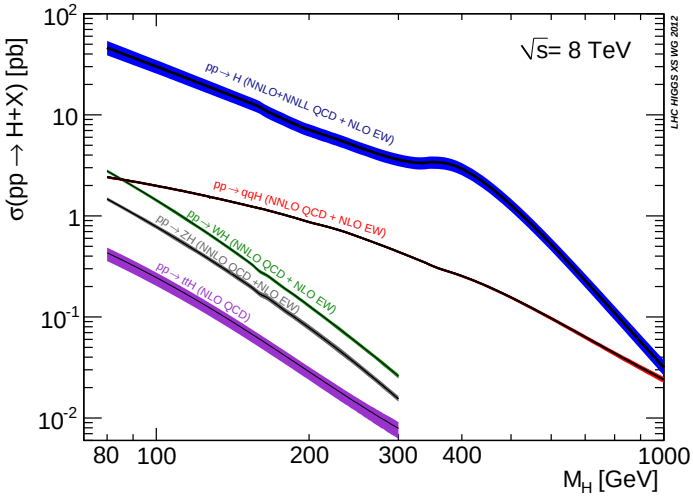


Figure 1.5: Next-to-leading order cross sections for the dominant production modes for a SM Higgs at $\sqrt{s} = 8$ TeV as a function of its mass [33].

maximize its experimental signature.

Associated production with top quarks can be initiated by a pair of gluons or quarks, the Higgs being radiated from a quark line in the latter case. The low yields restrict at the moment the experimental potential of this mode.

1.4.2 Higgs decay modes at LHC

The couplings of the Higgs boson to other SM particles are proportional to the particle mass for fermions and to the square of the mass for the $W^{\pm}$ and Z vector bosons. Hence, the Higgs is favored to decay into the heaviest particles kinematically accessible and the corresponding Higgs decay branching fractions (BR) change as a function of the Higgs mass, as shown in Figure 1.6. Around $m_H \sim 125.5$ GeV, the mass value for the observed Higgs boson, all the decay modes are accessible, and provide a wide spectrum of final states which can be experimentally tested.

In order to select and study Higgs events an efficient trigger is necessary. According to the specific decay channel studied it requires either the presence of leptons or photons or large missing transverse momentum, $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$, and high-

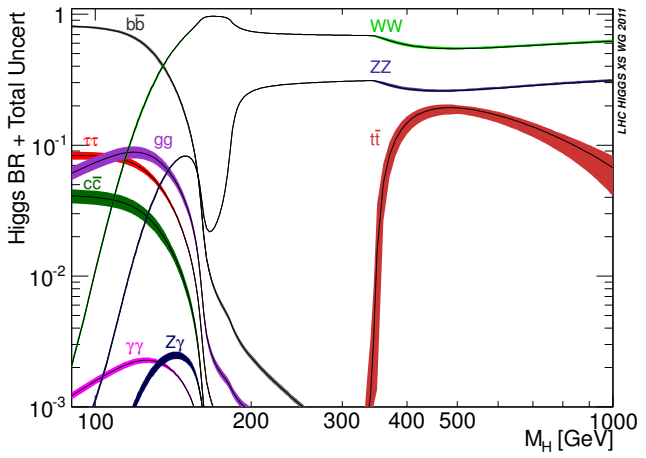


Figure 1.6: Decay branching ratios of the SM Higgs boson as a function of its mass [33]. Decays to massless particles (gluons and photons) proceed through a loop of either fermions or gauge bosons: the major contribution comes from top quark in the case of the gluon channel and from the W boson in the case of the photon channel.

p_{T} jets. Furthermore, experimental conditions and a distinctive signature of the signal with respect to the backgrounds play a crucial role in the tagging of the events and in the final sensitivity of the analysis. Hence, despite the high BR, $H \to b\bar{b}$ and $H \to \tau\tau$ decay modes are not the most sensitive channels due to their complex background composition, the presence of $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ and the not fully efficient b-jet and τ identification. In the same way, despite the low BR, $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ give the highest sensitivity, due to their peculiar signature. Moreover, since in these channels it is possible to fully reconstruct all the physics objects in the final state, high resolution mass measurement can be achieved.

Some more details about the characterizations of the varies decay channels are given in the following.

$\mathbf{H \to b\bar{b}}$ has the highest BR at low masses, but it suffers from huge QCD multijets background, many order of magnitude larger than the signal. The absence of an efficient trigger excludes the identification of gluon-gluon fusion and vector boson fusion productions. Therefore, in order to have a signature for the events, only the associated Higgs production with a leptonically decay vector boson ($W \to \ell\nu$, $Z \to \ell\ell$ and $Z \to \nu\nu$) is considered.

H $\rightarrow \tau\tau$ is the most sensitive channel for H decaying into fermions. A full description for the cut-based analysis is reported in Chapter 5 highlighting the complexity of the mass reconstruction of the $\tau\tau$ system, due to the presence of more than one neutrino and the background composition and estimation.

H $\rightarrow$ **VV**(**V** = **W**, **Z**) provides a good sensitivity in the whole Higgs mass range. In the low Higgs mass range the $H \rightarrow ZZ \rightarrow 4\ell$ and the $H \rightarrow WW \rightarrow \ell\nu\ell\nu$ provide important contributions. The former is the "golden channel" leading to a narrow invariant mass peak on the top of a relatively smooth and small background, the latter has a large BR but does not have a high resolution mass reconstruction due to the presence of missing transverse momentum in the final state. Searches for additional Higgs bosons at higher mass are pursued as well, in this mass region, the most important channel is given by $H \rightarrow WW \rightarrow \ell\nu q\bar{q}$.

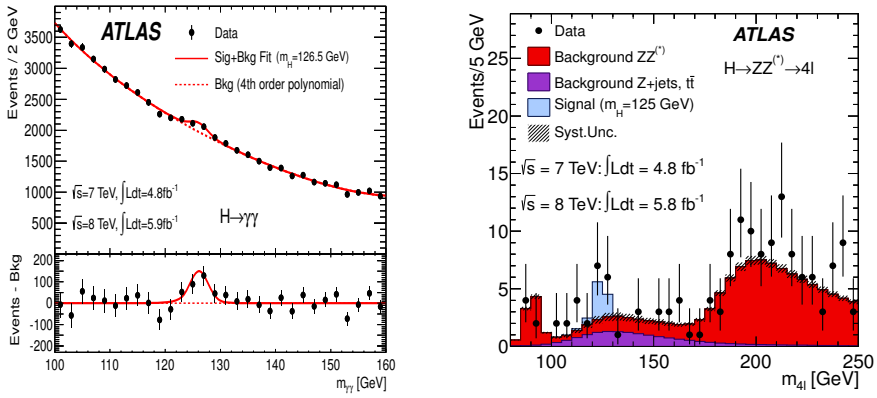
H $\rightarrow \gamma\gamma$ is one of the most important channels in the low mass range because it has a very distinctive signature given by two isolated and energetic photons with a narrow invariant mass peak.

1.4.3 Discovery of the Higgs boson

The ATLAS [6] and CMS [7] experiments independently announced, on the 4th July 2012, the discovery of a new particle consistent with a SM Higgs boson with $m_H \sim 125$ GeV with signal significances of 5.9 and 5.8, respectively. The results were obtained using the data collected by the two experiments in 2011 and 2012.

The discovery was driven by the high resolution mass channels $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ \rightarrow 4\ell$. In both cases the final state can be fully reconstructed: two energetic and isolated photons in the first case and two pairs of same flavor and opposite sign leptons in the latter case. Hence the signal appears in the distribution of the invariant mass as a narrow peak over a quite smooth background, as shown in Figure 1.7, which reports the ATLAS results for the two channels. The channel $H \rightarrow WW \rightarrow \ell\nu\ell\nu$ has a good sensitivity to the Higgs signal as well, but it has a poor mass resolution because of the presence of neutrinos.

To increase the sensitivity to a Higgs boson signal, the analyses exploit different topologies separating the events in mutually exclusive categories having different kinematics distributions and signal-to-background ratios. In particular in the $H \rightarrow \gamma\gamma$ analysis an exclusive category of events containing two jets



(a) $m_{\gamma\gamma}$ (b) $m_{4\ell}$

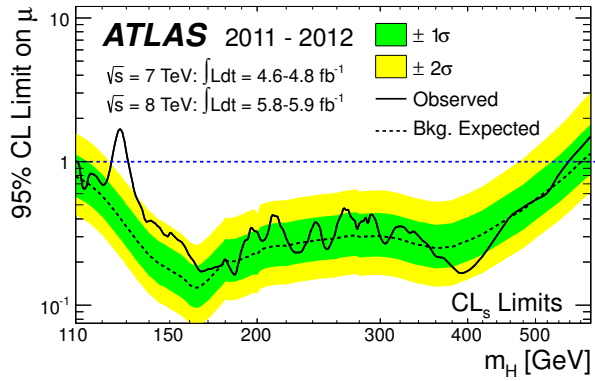
Figure 1.7: The distribution of the invariant mass for the diphoton system in the $H \to \gamma\gamma$ search in (a) and the four leptons system in the $H \to ZZ \to 4\ell$ search in (b) in the ATLAS experiment [6].

improves the sensitivity to VBF.

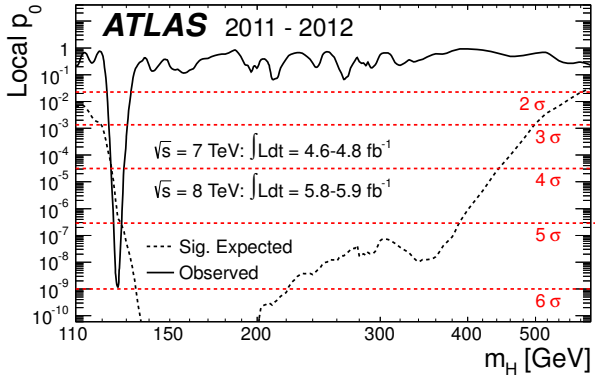
A statistical analysis is performed to quantify the observed excess. Figure 1.8(a) shows the combined 95% CL exclusion limits on the signal strength of the Higgs boson as a function of m_H after the combination of all channels. The mass range accessible with the data ($110 < m_H < 582$ GeV) is excluded, except for the region $122 < m_H < 131$ GeV, where an excess is observed. In order to quantify the significance of this excess the local p_0, i.e. the probability that the background can produce a fluctuation greater than or equal to the excess observed in data, is computed as a function of the Higgs boson mass as shown in Figure 1.8(b): the largest local significance is found for a SM Higgs boson mass hypothesis of $m_H = 126.5$ GeV, where it reaches about 6σ, with an expected value in the presence of a SM Higgs boson signal at that mass of 4.9σ.

1.4.4 Measurements of the Higgs boson properties

After the discovery ATLAS started the study of the properties of the new observed particle, that include mass, coupling constants and spin measurements, to test the compatibility with the SM. Any deviation from the SM predictions, for any of the decay mode, could be a signal of new physics. The study of these properties is performed in the individual channels, then all the measurements are combined



(a) 95% CL limits on the signal strength



(b) Local p_0

Figure 1.8: Exclusion limit on the signal strength of the Higgs boson in (a) and local significance (b) as a function of the SM Higgs boson mass $m_H$in the ATLAS experiment. The results are the combination of the $H \to \gamma\gamma$, $H \to ZZ \to 4\ell$ and $H \to WW \to \ell\nu\ell\nu$ searches [6].

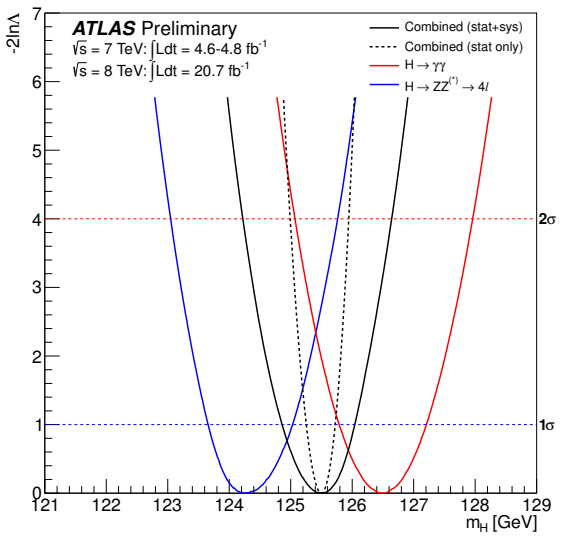


Figure 1.9: The profile likelihood ratio -2 ln $\Lambda(m_H)$ as a function of m_H for the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ channels and their combination. The dashed line shows the statistical component of the mass measurement uncertainty [34].

to maximize the statistical power.

Primarily, the measure of the mass of the observed Higgs boson, m_H, was performed combining the two high-resolution channels, $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ [34]. The most updated result for m_H is obtained by the analysis of the combined full data sets from 2011 and 2012:

$$m_H = 125.5 \pm 0.2(\text{stat})^{+0.5}_{-0.6}(\text{syst})\text{GeV}$$

The individual mass measurements and their combination are shown in Figure 1.9: they are compatible within 2.4σ.

The signal strength, μ, represents the local significance of the signal and it is defined as the ratio of the measured cross section multiplied by the branching ratio and the SM prediction:

$$\mu = \frac{\sigma \times BR}{\sigma_{SM} \times BR_{SM}} \tag{1.24}$$

It is a convenient observable to test the compatibility of the data with the background-only hypothesis, $\mu = 0$, and the SM Higgs hypothesis, $\mu = 1$. Assuming a Higgs mass value of 125.5 GeV, as previously measured, the individual

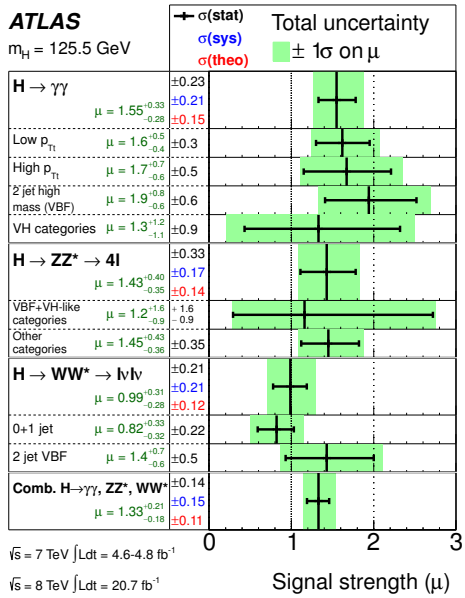


Figure 1.10: The measured production strengths for a Higgs boson of mass $m_H = 125.5$ GeV, for diboson final states and their combination [35].

μ for each diboson channel and for their combination are reported in Figure 1.10. The consistency between this measurement and the SM Higgs boson expectation ($\mu = 1$) is about 7% [35].

The best-fit value for the global signal strength factor μ does not give direct information on the relative contributions from different production modes. Therefore, in addition to the signal strength in different decay modes, the signal strengths of different Higgs production processes contributing to the same final state are determined. A common signal strength scale factor $\mu_{ggF+ttH}$ has been assigned to both gluon fusion production (ggF) and the very low rate $t\bar{t}H$ production mode, as they both scale dominantly with the ttH coupling in the SM. Similarly, a common signal strength scale factor μ_{VBF+VH} has been assigned to the VBF and VH production modes, as they scale with the WH/ZH gauge coupling in the SM [35]. The results are reported in Figure 1.11: no significant deviation from SM couplings is observed, all results are compatible with each other, and with the SM, within 95% CL.

Another important test for the Higgs boson is the determination of its quan-

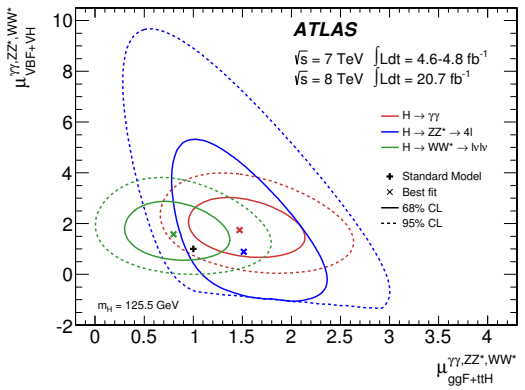


Figure 1.11: Likelihood contours for the diboson final states in the ($\mu_{ggF+ttH}$, μ_{VBF+VH}) plane for a Higgs boson mass hypothesis of $m_H = 125.5$ GeV [35].

tum number. The SM predicts a CP-even, spin-0 particle:

$$J^P = 0^+$$

thus this hypothesis is tested against others: $0^-, 1^\pm, 2^\pm$ [36, 37, 38]. The measurement is performed independently in the diboson decay channels, $H \to \gamma\gamma$, $H \to WW \to \ell\nu\ell\nu$, $H \to ZZ \to 4\ell$, and it is based on combinations of different kinematic observables like the angular distributions of decay products in the resonance rest frame. The Landau-Yang theorem forbids the direct decay of a spin-1 particle into a pair of photons [39, 40], therefore, the spin-1 hypothesis is therefore strongly disfavoured by the observation of the process $H \to \gamma\gamma$. The measurement of the Higgs boson parity is performed in the channel $H \to ZZ \to 4\ell$ and is found to be "positive", so in agreement with the SM expectation. Finally, the 0^+ hypothesis against a 2^+ scenario is tested as well. In this analysis the discriminating observables are dependent on the production mechanism, therefore, multiple scenarios are studied varying the fractions of production processes initiated by qq and gg. The combination of the results from individual channels is performed through a likelihood fit and the 2^+ hypothesis is excluded at a confidence level $> 99\%$.

1.5 Limits of the Standard Model theory

The SM is presently our best description of particle phenomenology, its prediction has been intensively tested leading to impressive constraints of the theory itself. Nevertheless, on the base of principles of universality, elegance of the formulation and experimental consistence, criticism of the SM is rising. The SM is not considered a complete theory because it does not include gravity, does not provide a dark matter candidate and does not accommodate for neutrinos mass terms, all aspects for which there are instead experimental confirmations. Moreover, the SM does not completely explain the baryon asymmetry in the universe and even if it provides mass to the fermions, does not predict their large mass spectrum spanning about 5 orders of magnitude (0.5 MeV - 171 GeV) or more, if neutrinos masses are considered (eV scale or less). These unknowns contribute to the 19 free parameters of the SM, which from a theoretical point of view is considered an excessive number for a fundamental theory. Finally, there is the hierarchy problem: the high energy separation between the electroweak and the Planck scale generates large radiative corrections that ask for a not natural fine-tuning cancellation. This last issue has mostly stimulated the formulation of possible theories extending the SM in the last decades. Further information about some of the conceptual and experimental limits of the SM is given in the following sections.

The gravitational force

Presently the gravitational force is the least understood of all physical forces, and very little is known about is microscopic effects. A gravity mediator, the graviton, can be introduced in the theory in analogy to the other forces but with spin 2, however so far there are no experimental confirmations for its existence. Moreover the gravity may not be a force at all, but only a geometrical effect due to the bending described by General Relativity [41, 42]. Anyway, regardless of the possibility of a quantum formulation for the gravitational force, something is expected to happen at the Planck scale ($1.22 \cdot 10^{19}$ GeV) where gravity would not be negligible anymore and would directly affect particle phenomenology.

Throughout the years, theories extending the Standard Model have been formulated to include gravity in a natural way. Some examples are the "grand unification" theories [43, 44] that aim to merge all forces into a single interaction defined by a larger gauge symmetry, and Supersymmetry theories [45, 46] that provide an additional symmetry allowing to convert fermions into bosons, as

required for a gravity unification, and including a dedicated gravitational sector.

All these theories predict additional particles in the model but so far none has been yet observed. In particular, in the last years LHC experiments highly constrained the parameter space of new models pushing the mass limits for new observable particles to 700-800 GeV [47].

Dark Matter

A proof of the incompleteness of the SM comes from the inability of the SM to provide an explanation for all the mass of the universe, and its phenomenology.

In the last years astrophysics experiments studying the motion of stars in galaxies and the galaxy orbits in galaxy clusters established the presence of Dark Matter, a matter that emits minimal to no light (or other electromagnetic radiation) but has gravitational effects [48]. The Dark Matter should account for a large part of the total mass in the universe but in the SM there is no particle description for it. The total matter described by the SM is only the 4% of the total matter in the universe.

Supersymmetry theories are appealing because they provide a Dark Matter candidate, but still no experimental observation is provided, like in searches looking for a direct Dark Matter pair production [49, 50]. The ATLAS Run 2 will be crucial for these searches, that could rule out some of the models.

Neutrinos masses

In the SM Lagrangian there is not right-hand component for the neutrinos, making them massless particles. This is in contrast with the observation of the neutrinos oscillation phenomena by atmospheric, solar and accelerator-based neutrino experiments [11, 12, 13]. To induce oscillation a mixing mechanism is required: the weak eigenstates must be a mix of the mass eigenstates with different masses. Hence neutrino masses cannot be all the same and certainly not all equal to zero. Presently the upper bound on neutrinos masses is less than 1 eV.

The measurements of the mixing parameters introduced by the oscillation mechanism can provide further information about the neutrino nature. For example, one fundamental point is to determine if the neutrinos are Dirac fermions like the other fermions of the SM, or if they are Majorana fermions, for which particle and anti-particle would coincide [51].

Hierarchy problem

The SM does not explain why the electroweak scale is by far lower than the Planck scale. As an effect of these two highly-separated scales the bare value of the Higgs mass receives high corrections through loop contributions. In fact these corrections are proportional to the next higher scale in the theory ($\delta m_H^2 \propto \Lambda^2$) that, if there is no new physics in between, is the Planck scale. Hence, the Higgs mass should be very large in contrast with the direct and indirect experimental constraints. The Higgs mass is indeed now measured to be 125.5 GeV [34], but also before the Higgs boson's experimental discovery it was known that its mass should have been below 1 TeV in order to unitarize the WW scattering amplitude.

In order to keep the Higgs mass low, a very fine cancellation between the bare value of μ^2 and the radiative corrections is necessary. This fine-tuning process does not seem to be very natural, therefore new theories are developed in order to provide a more natural cancellation of the radiative corrections. There are basically two different approaches. The first is to introduce a new symmetry protecting the Higgs mass, as done in Supersymmetry theories, while the second looks at the Higgs as a composite bound state with strongly interacting dynamics at the TeV scale [52, 53]. Once again no experimental evidence has been found yet, constraining the parameter space for the models beyond the SM.

1.6 Conclusion

In the last 50 years theoretical and experimental successes led to the affirmation of the Standard Model as well-established quantum field theory describing particles and their interactions.

The start of the LHC opened an exciting time for particle physics culminating with the discovery of a new particle consistent with the predicted Higgs boson. The ATLAS and CMS experiments have already started an intense physics program to measure the properties of this particle, and so far everything is found to be compatible with the Standard Model expectations, thus providing further confirmation of the theory. However, the level of precision achieved in the coupling measurements leaves still room for new physics, thus making future LHC runs at higher center-of-mass energies very attractive.

Beyond the Higgs measurements, both ATLAS and CMS have a wide physics program for direct searches of new particles, stimulated mainly by the indirect observation of Dark Matter and by the unexplained large separation between the

electroweak and the Planck scale. Also these searches look with interest at future runs where the higher center of mass energy could increase the sensitivity to new phenomena.

The results so far achieved are just the beginning of a new exciting time during which we expect to improve our understanding of fundamental matter and find answers to the open questions of particle physics.

Chapter 2

The Large Hadron Collider and the ATLAS experiment

The Large Hadron Collider (LHC) is presently the largest and the most powerful particle accelerator in the world. During its first four years of running the data collected and analyzed by the LHC experiments have provided exciting physics results. First of all stands the observation of the Higgs boson [6, 7], followed by the investigation of the fundamental forces of nature up to the TeV scale, an energy regime never explored before, thus opening a new era for particle physics. In 2015, after the long shut-down, LHC will operate at its design conditions and will reach new frontiers of energy and luminosity, further increasing the physics analyses potential and the discovery power for new phenomena.

To completely exploit the physics potential provided by the LHC collisions, the LHC experiments have to face and find solutions for unprecedented technical challenges of complex mechanical structures, radiation tolerant electronics, fast data acquisition and high precision measurements, continuously stimulating engineering innovation and the development of new technologies.

More information about the LHC complexity and the run conditions are reported in Sections 2.1 and 2.2. A brief description of the experiments operating at LHC and their physics reach is given in Section 2.3. Finally, a more detailed discussion about the ATLAS experiments and its sub-detectors is reported in Section 2.4.

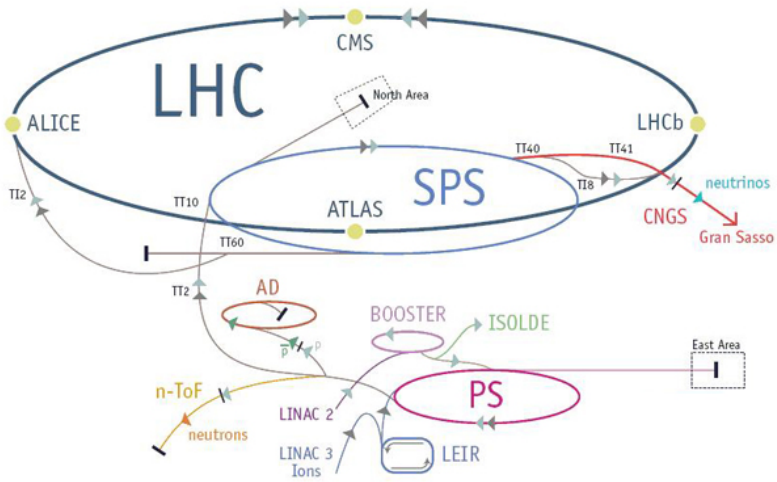


Figure 2.1: The LHC collider: the figure shows the acceleration chain for both proton and heavy ion beams [54].

2.1 The LHC collider

The Large Hadron Collider (LHC) [54, 55, 56] is a proton-proton (p-p), ion-proton (Pb-p) and a ion-ion (Pb-Pb) collider located at CERN, near Geneva, Switzerland. It is installed in the same 27 km long underground tunnel which housed the Large Electron Positron (LEP) [57] until 2000. Since LHC collides particles of the same charge, two separate beams lines with opposite magnetic fields are needed to deflect the particles into circular trajectories. The oppositely running particle beams are finally collided in four dedicated interaction points instrumented with large experiments.

The LHC is the last stage of the acceleration chain shown in Figure 2.1 and composed of a series of particle accelerators that progressively increase the energy of the proton beams [58]. The beams are first accelerated to 50 MeV using a linear accelerator (LINAC2), then they are further accelerated up to 1.4 GeV by a circular booster (BOOSTER) and then up to 26 GeV by the Proton Synchrotron (PS). Before being transferred to the LHC, the proton beams are injected to the Super Proton Synchrotron (SPS), where protons reach the minimum energy at which the LHC can maintain a stable beam, 450 GeV. Finally, in the LHC ring the acceleration is performed by radio-frequency (RF) cavities. A 400 MHz

superconducting system increases the beam energy by 485 keV at each turn until it reaches the designed energy, 3.5 TeV for 2011 runs and 4 TeV for 2012 runs. Because the LHC is a proton collider, energy losses from synchrotron radiation are small: $\sim 10^{-9}$ of the proton's energy.

To allow the acceleration through the RF cavities and the monitoring of the beams, the colliding particles are injected in the acceleration chain in "bunches". Bunch properties like the number, intensity, frequency and collimation have a direct impact on the collider performances. A more complete discussion about these properties and their effects can be found in the next section.

2.2 Luminosity and pile-up conditions

The main purpose of experiments at particle colliders is the search for rare processes like the production of the Higgs boson or new physics beyond the Standard Model, characterized by a small cross section σ. In order to observe these rare processes, it is necessary to maximize the rate dN/dt of the process, that is linearly related to the instantaneous luminosity:

$$\frac{dN}{dt} = \sigma \mathcal{L} \tag{2.1}$$

The luminosity does not depend on the physics process but only on beam-parameters and can be expressed by:

$$\mathcal{L} = \frac{f n N_1 N_2}{\mathcal{A}} \tag{2.2}$$

where n is the number of colliding bunches, N_1 and N_2 is the number of particles in each bunch, f is the accelerator frequency and $\mathcal{A}$ is the inverse of the beam cross section. The values for the main detector parameters during 2010, 2011 and 2012 runs, compared with the design values, are reported in Table 2.1.

The peak luminosity is the maximum value of the instantaneous luminosity. Since during a fill (the period the beams are kept colliding) the instantaneous luminosity drops as the beams lose intensity, the peak luminosity is reached at the beginning of a fill. Figure 2.2 shows the peak luminosity as a function of the data taking time.

Beside holding the luminosity record, LHC is presently the accelerator that collides particles at the highest center-of-mass energy, $\sqrt{s}$ (twice the beam energy). After the first collisions at $\sqrt{s} = 900$ GeV in November 2009, LHC increased the beam energy to reach $\sqrt{s} = 2.36$ TeV in December 2009 (first world

Parameter	2010	2011	2012	design
Beam energy [TeV]	3.5	3.5	4	7
β^* in IP 1 and 5 [m]	2.0/3.5	1.5/1.0	0.6	0.55
Bunch spacing [ns]	150	75/50	50	25
Max. n. of bunches	368	1380	1380	2808
Max. n. of p per bunch	1.2×10^{11}	1.45×10^{11}	1.7×10^{11}	1.15×10^{11}
Peak luminosity [$\mathrm{cm^{-2}s^{-1}}$]	2.1×10^{32}	3.7×10^{33}	7.7×10^{33}	1×10^{34}
Max. $\langle\mu\rangle$	4	17	37	19

Table 2.1: Setup values of the main detector parameters for runs in 2010, 2011, 2012 and comparison with the design values [59]. IP 1 and IP 5 are the interaction points where the two LHC general purpose experiments, ATLAS and CMS, are located. $\langle\mu\rangle$ is the average number of collision per bunch crossing.

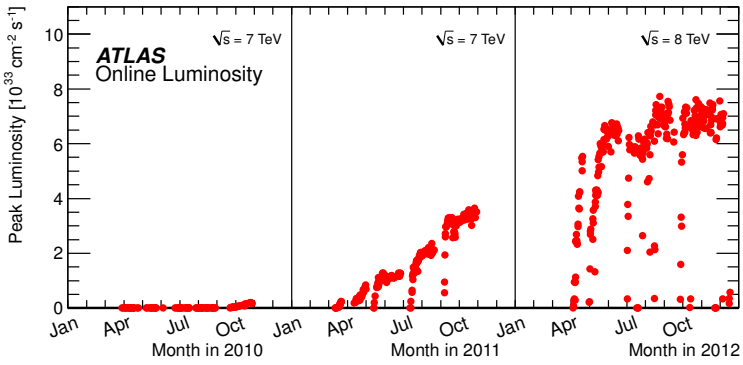


Figure 2.2: Peak luminosity as a function of the data taking time [60].

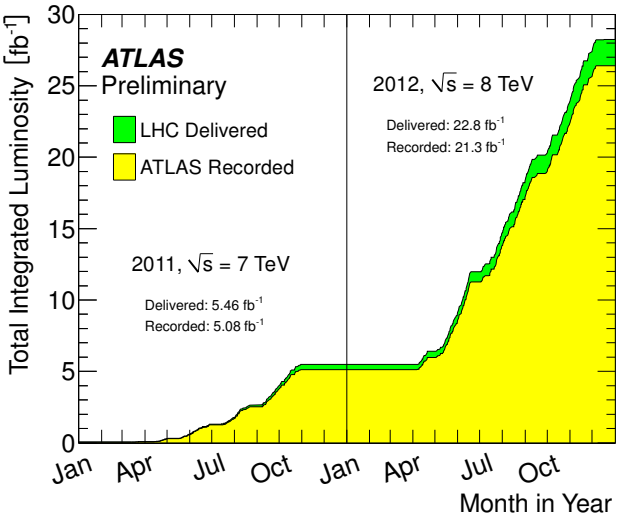


Figure 2.3: Delivered and recorded integrated luminosity as a function of the data taking time [60].

energy record) and further increased it to reach $\sqrt{s} = 7$ TeV for 2010 and 2011 runs and $\sqrt{s} = 8$ TeV for 2012 runs. Currently the machine is in a shut-down phase to be prepared to run at the design $\sqrt{s} = 14$ TeV in 2015.

By integrating the rate for a process in a certain period of time, one gets the estimate of the total number of events (N_{tot}) recorded in that period:

$$N_{tot} = \int dt\mathcal{L} \times \sigma \tag{2.3}$$

The quantity $\int dt\mathcal{L}$ is the integrated luminosity, usually it is expressed in inverse of cross section units (i.e. fb^{-1}) and it is a measurement of the collected data size. The integrated luminosity recorded by ATLAS is shown in Figure 2.3 as a function of the data taking time. In this thesis all the datasets corresponding to the total integrated luminosity recorded by ATLAS, 21.3 fb^{-1}, are used.

According to equation 2.2 an increase in the luminosity can be achieved by squeezing the beams and reducing their transverse size[1], or by increasing the number of protons per bunch (up to 1.7×10^{11} at the end of the 2012 run) or increasing the number of bunches (1380 in the 2012 run). However, there are

[1] the beams are squeezed in the transverse plane by magnetic quadrupoles and are confined in an area of O(μm $\times$ μm) at the interaction points

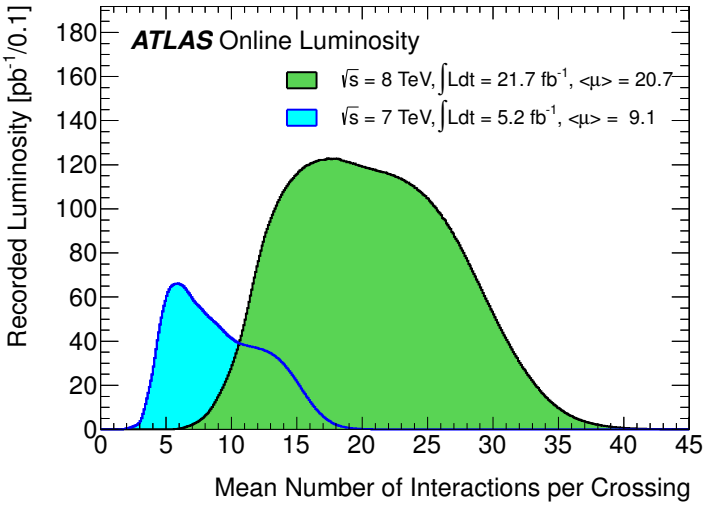


Figure 2.4: $\langle\mu\rangle$ distribution expressing the pile-up level for ATLAS physics runs at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV [60].

limitations on these parameters, such as the beam-beam limit, the long-range beam-beam interactions, the electron cloud effects and machine protection. Furthermore, the high luminosity introduces some experimental difficulties as well, such as the presence of pile-up generated by additional p-p interactions occurring in the same bunch crossing as the hard collision of interest. This pile-up is a consequence of the high overall p-p cross section. The pile-up can be classified as "in-time" if the multiple p-p interactions arise from the same bunch crossing, or as "out-of-time" if the interaction are originated from different bunch crossing during the time taken for the detector to process the signal from a single event.

The in-time pile-up is affected by the focusing of the beams and by the number of protons in each bunch. Since the number of protons is the same for each beam ($N_1 = N_2$ in equation 2.2) this contribution is particularly important because it has a quadratic effect. As one example moving from the low luminosity runs characterized by $N \sim 8 \cdot 10^{10}$ to higher luminosity runs with $N \sim 1.45 \cdot 10^{11}$ the fraction of events with pile-up increases by more than 50%. Experimentally this is evaluated by the number of reconstructed primary vertices, N_{PV}.

The out-of-time pileup is a detector effect due to the integration time of the readout electronics. Its effects are highly dependent on the detector system, and the beam conditions, mainly the bunch spacing time. For example, the

bunch-spacing decreased from 150 ns in 2010 to 75 ns and then to 50 ns in 2012, which strongly increased the contribution of the out-of-time pile-up in the ATLAS Liquid Argon calorimeters. Experimentally it can be evaluated looking at the average number of interactions per bunch crossing $\langle\mu\rangle$ as shown in Figure 2.4 for different physics runs.

2.3 The LHC experiments

The LHC beams collide in four different interaction points where the large experiments are located: ATLAS (A Toroidal LHC ApparatuS) [1], CMS (Compact Muon Solenoid) [2], ALICE (A Large Ion Collider Experiment) [61] and LHCb (LHC beauty experiment) [3]. All of them are aiming to better understand the fundamentals of nature and to answer the open questions concerning particle and astroparticle physics.

ATLAS and CMS are general purpose detectors with broad physics programs including both Standard Model studies and searches for new physics. B-physics and heavy-ions studies are pursued as well. After the observation of the Higgs boson, the particle responsible for the origin of mass and the particles mass spectrum, the study of its properties has become one of the main goals. All these measurements and searches address the open questions of the Standard Model, like the unification of forces (including gravitation) and the presence of the dark matter. Both experiments were designed to operate at the highest luminosity achievable at the LHC but they rely on different detection system and technologies. A more detailed description of the ATLAS detector will be given in the next chapter.

ALICE is specialized in heavy-ion physics and is devoted to the characterization of quark-gluon plasma, a phase present in the early universe and characterized by extremely high temperature and densities. It can help explaining why quarks and gluons are never observed as free particles but always bound together confined inside composite particles, and why only 1% of the proton and neutron mass is given by the quark mass.

LHCb is devoted to b-quark physics and precise CP violation measurements, addressing the question of the asymmetry between matter and antimatter in the universe. Moreover, B-physics analyses can be sensitive as well to

new physics through loop processes. LHCb is designed to operate at a luminosity almost two orders of magnitude lower than the nominal one.

2.4 The ATLAS experiment

The ATLAS detector is optimised for the high-intensity and high-energy LHC collision environment, described in Section 2.2. It provides high precision measurements and the identification of rare processes of physics interest, with cross sections many orders of magnitude below the total p-p cross section dominated by QCD. The comparison between the cross section of Standard Model processes, Higgs production processes and the total cross section is shown in Figure 2.5 as a function of the center-of-mass energy.

As most of the experiments at beam-beam colliders, ATLAS has been designed with a cylindrical layout, a forward-backward symmetry with respect to the interaction point and a nearly hermetic system in order to detect as much as possible all the particles generated by the LHC collisions and to fully reconstruct the physics event.

To measure the particle energy and momentum in a broad p_{T} spectrum (from hundreds of MeV to some TeV), and to have an efficient particle identification, ATLAS is divided into sub-detectors employing different technologies, with different granularity and radiation resistance, that surround the interaction point, as shown in Figure 2.6. The innermost detector is a precision tracking system operating in a solenoidal magnetic field. It uses the measurement of the bending radius to reconstruct the momenta of charged particles and interaction vertices. The middle layer consists of the calorimetric system, divided into an electromagnetic and hadronic calorimeters, which provide the energy measurements of both neutral and charged particles. Finally, the outermost detector is the muon spectrometer that, in combination with a dedicated toroidal magnetic field, measures the muon momenta. Weakly interacting particles as neutrinos or new particles foreseen by Standard Model extensions do not interact with the detector and their contributions can be determined from the energy balance of the event. A summary of the interaction of different particles through the ATLAS detector is sketched in Figure 2.7.

Further information about the ATLAS sub-detectors are given in Sections 2.4.2 to 2.4.4, while the particle reconstruction, identification and calibration are discussed in Section 3.2. Finally, the ATLAS trigger system is described in

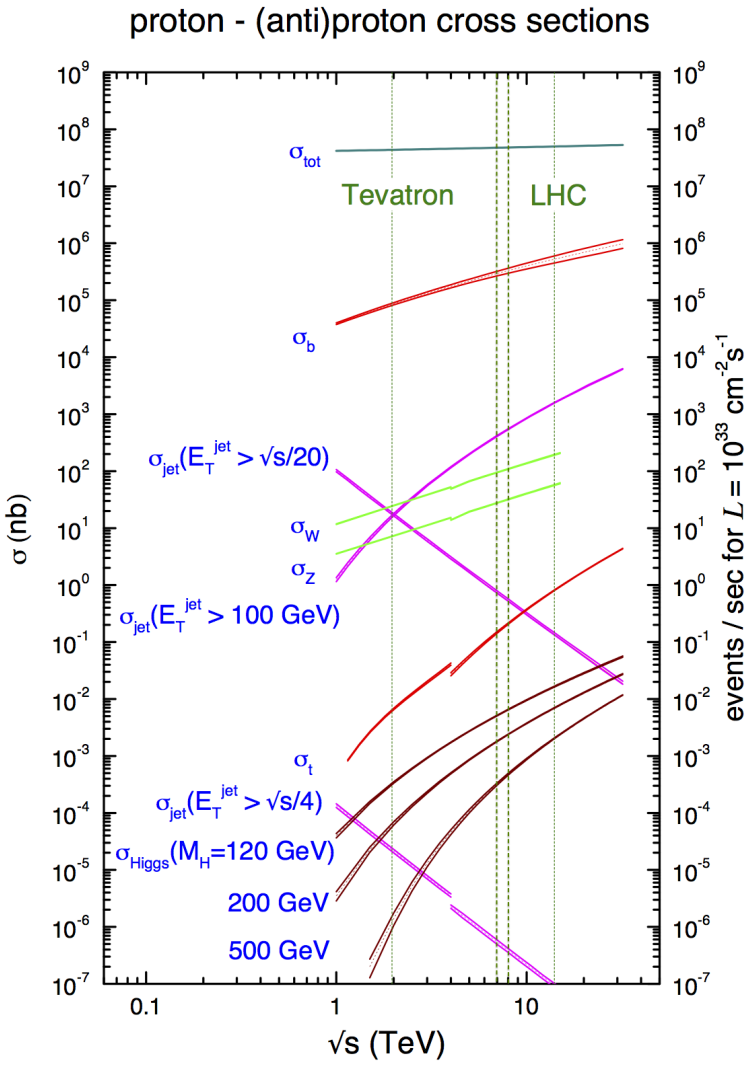


Figure 2.5: Cross sections for some interesting physics processes as a function of the center-of-mass energy. The discontinuity at ∼4 TeV is from the transition from proton-anti-proton collisions at the Tevatron on the left to proton-proton collisions at the LHC on the right. The vertical lines indicate the running energy of the Tevatron (1.96 TeV), the running energy of the LHC in 2011 (7 TeV), in 2012 (8TeV) and the possible future running energy (14 TeV) [62].

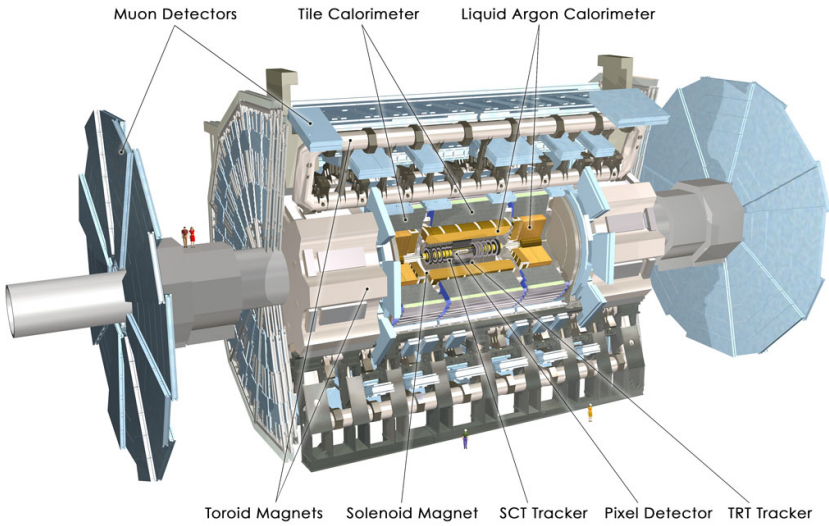


Figure 2.6: The ATLAS detector: the inner tracking detector surrounds the beam pipe and it is followed by the electromagnetic calorimeter, the hadronic calorimeter and the muon spectrometer [1].

Section 2.4.5.

2.4.1 ATLAS coordinate system

ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis coinciding with the axis of the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y axis points upward.

The x and y axes define the transverse plane, where cylindrical coordinates (r, ϕ) are used, ϕ being the azimuthal angle around the beam pipe.

Since the partons that give rise to the signal process of interest carry an a priori unknown fraction of the proton momentum, the overall boost of the collision is not known. For this reason, boost-invariant quantities are preferred.

Therefore, a convenient way of expressing the polar angle, θ, is through the pseudorapidity, η, that transforms additively under boosts in the z-direction and it is defined as:

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

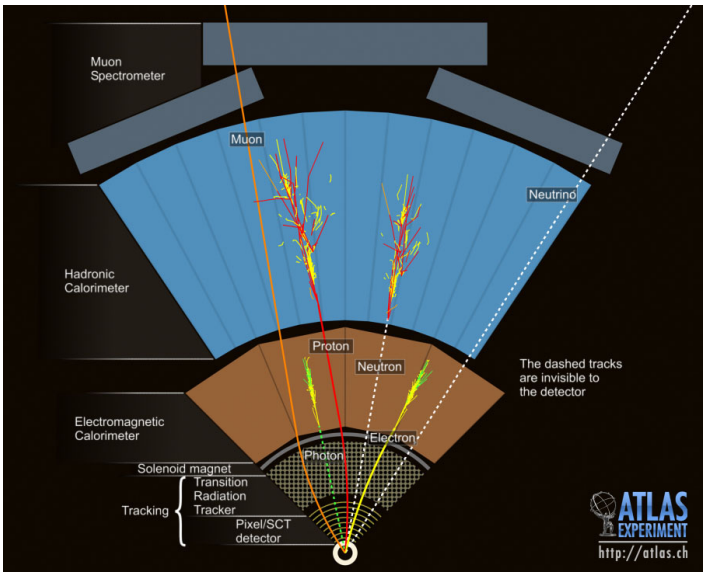


Figure 2.7: Sketch of particles interaction in the ATLAS detector [63].

The pseudorapity is a geometrical quantity and it is the limit case for massless particles of the rapidity, y:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \tag{2.5}$$

The difference in rapidity is as well a Lorentz-invariant but it depends not only by the polar angle but also on the mass of the particle.

Finally, boosts along the beam axis also do not affect the azimuthal angle, hence a Lorentz-invariant distance R can be defined in the $\phi - \eta$ plan:

$$R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \tag{2.6}$$

2.4.2 The tracking system

Tracking systems rely on the measure of the positions of charged particles in different radial layers of the detectors to perform the track reconstruction and on the measure of the curvature radius due to the bending of the magnetic field to compute the track momentum according to: $p_{\mathrm{T}}[\mathrm{GeV}] = 0.3 \times B[\mathrm{T}] \times R[\mathrm{m}]$.

ATLAS has two separate superconducting magnet systems to provide the magnetic fields for the inner detector and the muon spectrometer. The structure is shown in Figure 2.8 and consists in:

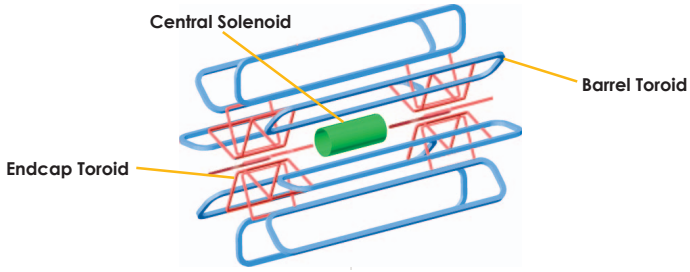


Figure 2.8: Sketch of the ATLAS magnetic system [64].

A central solenoid providing an almost uniform 2 T magnetic field for the inner detector. The field is parallel to the beam axis and bends particles in the transverse plane.

A barrel air-core toroid consisting of 8 independent coils and providing a peak magnetic field of 4 T for the muon spectrometer. The field is mostly perpendicular to the muon trajectories and deflects them in the η direction. Since muon travels mainly through the air the degradation in resolution due to multiple scattering in minimized.

Two endcaps air-core toroids providing a peak magnetic field of 4 T for the muon spectrometer and bending charged particles in the η direction.

The inner detector

The inner detector [65, 66] is the closest sub-detector to the beam pipe and to the interaction point. It is highly granular to allow reconstructing tracks from individual charged particles at high precision in a high particle flux environment. Other features of the inner detector are very precise particle momentum measurements, highly efficient vertex reconstruction and precise and highly efficient primary vertex and secondary vertex (from e.g. b-quark decays) identification.

Precise tracking is achieved by providing few but high-precision space points (track hits) at small radii close to the interaction point, and a larger number of lower precision points at larger radii by combining the measurement from three sub-detectors: the pixels, the SCT and the TRT, that are highlighted in Figure 2.9.

Each of the three sub-detectors is divided in a barrel region, where the detector modules are laid out in cylindrical layers, and an endcap region, where disks are

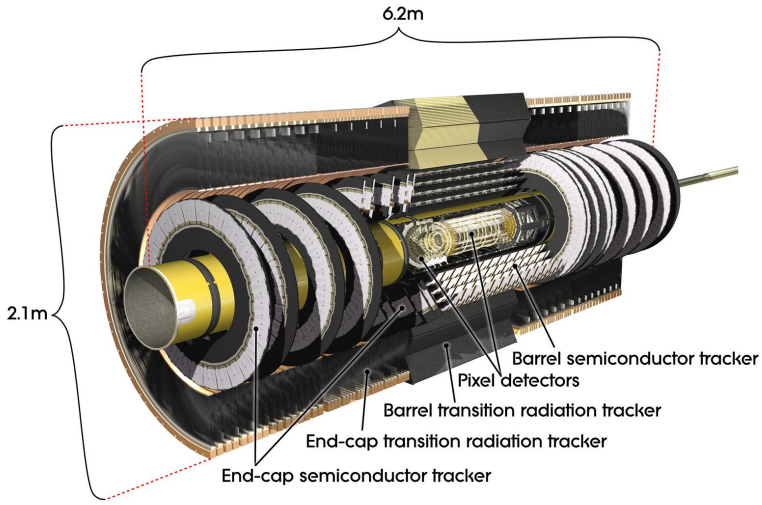


Figure 2.9: The ATLAS inner detector and its components [67].

used to increase the detector coverage in η without a large increase in detector surface.

The Pixel Tracker [68] is the inner-most device, with the first of its three layers positioned at 5 cm from the interaction point. All the layers consist of high-precision and high-efficiency semiconductor modules with a tight bi-dimensional segmentation that provides high granularity. It results in the measurement of three high-resolution points with a spatial resolution of better than 14 μm in the r-ϕ plane.

The SCT (Semi-Conductor Tracker) [66] consists of silicon microstrip layers. It contributes with up to eight high-resolution tracking points with a spatial resolution of better than 20 μm in the r-ϕ plane.

The TRT (Transition Radiation Tracker) [69, 70] is composed by straw tubes filled with gas interleaved with polypropylene fibers and foils. A high-voltage potential is applied to collect the ionization given by the passing charged particles. It provides 36 points with lower resolution with respect to the inner layers (<0.17 mm in the r-ϕ plane). The TRT also helps in discriminating electrons over other particles, since when electrons pass through the material between the straw tubes they generate X-rays.

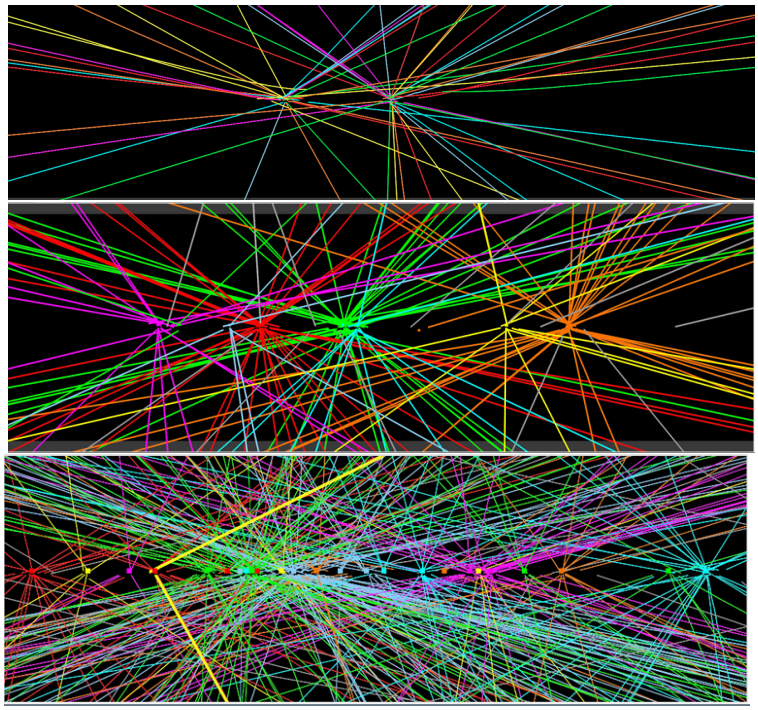


Figure 2.10: From the top, an ATLAS physics event with: 2 vertices, 7 vertices and 25 vertices [71].

In presence of pile-up, the number of primary reconstructed vertices (N_{PV}) increases with the number of additional p-p interaction, making the tracking environment denser and denser, as shown in Figure 2.10. Despite such compelling background the track and the vertex performance in 2012 allowed for a high efficiency for the N_{PV} reconstruction, thus making N_{PV} a stable and unbiased estimator for the in-time pile-up. The first primary collision vertex is chosen as the one with the hardest-scatter contribution, i.e. the maximum $\sqrt{\Sigma_{track} p_{\mathrm{T}}^2}$. The possibility to associate, where possible, energy contributions to their specific vertex, and the insensitivity of the tracker to out-of-time pile-up, makes N_{PV} a powerful variable to identify and suppress the pile-up contribution.

The muon spectrometer

Muons are the most penetrating particles detected by ATLAS and they are able to pass through the inner detector and the calorimeter without being absorbed.

Therefore the last layer of the detector, the muon spectrometer [1, 66], was designed for triggering the muons and for measuring their electrical charge and momenta. Different types of muon chambers are employed to achieve these tasks:

Monitored Drift Tube chambers (MDT) are high-precision chambers, the position of muons can be determined to 80 μm. They are employed in the central region for the measurement of the muon momenta. Since the charge drift time for these chambers is larger than the colliding bunch spacing they need to be integrated by fast trigger chambers tagging the collision event.

Cathode-Strip Chambers (CSC) are multiwire proportional chambers with high resolution muon position can be determined to 60 μm. Since they have high rate capability and time resolution, they are employed in the forward region. They withstand the demanding rate and background conditions, and can cope with the high particle multiplicities in the inner-most tracking layer close to the beam pipe.

Trigger Chambers provide a fast response within 15-25 ns, and can tag the bunch crossing of interest. They are divided into Resistive Plate Chambers (RPC) installed in the barrel region ($|\eta| < 1.05$) and Thin Gap Chambers (TGC) installed in the endcap ($1.05 < |\eta| < 2.4$). They also provide a second coordinate measurement for muons.

2.4.3 The calorimeter system

The calorimeter system [1] measures the energy and position of electrons, photons and hadronic particles, following their shower development and measuring the absorbed energy. For precise measurements of jets and missing transverse momentum ($\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$) the ATLAS calorimeter is built with a full azimuthal coverage and an almost hermetic η coverage extending up to $|\eta| = 4.9$. An optimal performance through the whole η range is provided using different techniques for the barrel and endcap regions, according to the demands of a wide range of physics process and the varying challenges from the radiation environment.

The system is primarily divided into an electromagnetic and a hadronic calorimeter to take into account the difference in the development of electron or photon showers with respect to hadronic ones, and then into seven sub-detectors as shown in Figure 2.11: the presampler barrel (PEMB) in $0 < |\eta| < 1.8$ and endcaps (PEMEC) in $1.5 < |\eta| < 1.8$, the electromagnetic calorimeter barrel (EMB) in $|\eta| < 1.475$ and endcaps (EMEC) in $1.375 < |\eta| < 3.2$, the hadronic

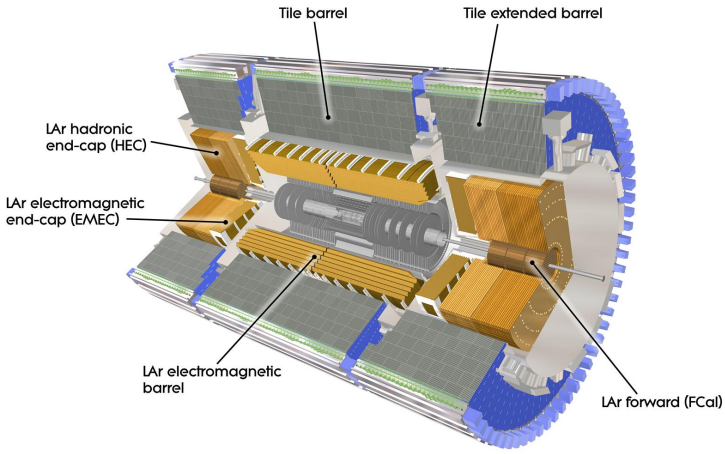


Figure 2.11: ATLAS calorimeter system. It consists of an electromagnetic calorimeter (closer to the beam pipe) and a hadronic calorimeter which use different technologies to contain the different shower development of electromagnetic and hadronic particles, respectively. Then each detector is divided into barrel, endcap and forward according to the η coverage [72].

calorimeter barrel (TILE) in $|\eta| < 1.7$ and endcaps (HEC) in $1.5 < |\eta| < 3.2$, the forward calorimeters (FCAL) in $3.2 < |\eta| < 4.9$. Detailed information about the pseudorapity extensions, segmentation and granularity of each calorimeter are summarized in Table 2.2.

All of them employ a sampling technique alternating layers of absorber material, where the shower is generated, to layers of active medium, where the shower energy is detected. As active medium, Liquid Argon (LAr) [73] has been employed for all the calorimeters, apart the Tile calorimeters, for its intrinsic linear behaviour, radiation-hardness and stability of response over time. On the other hand, the integration time for the electronic pulse is quite slow, about 400 ns making these systems very sensitive to pile-up. The Tile calorimeter, instead, uses a scintillating material characterized by a faster response, therefore, pile-up effects are much less significant.

ATLAS calorimeters are non compensating: the hadron response is lower than the response to electromagnetically interacting particles. Usually, the hadron contribution is corrected applying a proper calibration either before the physics

Calorimeter	**Coverage**		**Granularity** ($\Delta\eta \times \Delta\phi$)
EM calorimeter	**barrel**	**end-cap**	
Presampler	$\|\eta\| < 1.54$	$1.5 < \|\eta\| < 1.8$	0.025×0.1
Sampling 1	$\|\eta\| < 1.475$	$1.375 < \|\eta\| < 3.2$	0.003×0.1[a] 0.025×0.025[b] 0.003 - 0.025×0.1[c] 0.1×0.1[d]
Sampling 2	$\|\eta\| < 1.475$	$1.375 < \|\eta\| < 3.2$	0.025×0.025 0.075×0.025[b] 0.1×0.1[d]
Sampling 3	$\|\eta\| < 1.35$	$1.5 < \|\eta\| < 2.5$	0.05×0.025
Tile calorimeter	**barrel**	**extended barrel**	
Sampling 1 Sampling 2	$\|\eta\| < 1.0$	$0.8 < \|\eta\| < 1.7$	0.1×0.1
Sampling 3			0.2×0.1
Hadronic end-cap calorimeter			
Samplings 1-4	$1.5 < \|\eta\| < 3.2$		0.1×0.1[e] 0.2×0.2[d]
Forward calorimeter			
Samplings 1-3	$3.1 < \|\eta\| < 4.9$		0.2×0.2

[a] $|\eta| < 1.4$, [b] $1.4 < |\eta| < 1.475$, [c] $1.375 < |\eta| < 2.5$, [d] $2.5 < |\eta| < 3.2$, [e] $1.5 < |\eta| < 2.5$

Table 2.2: Pseudo-rapidity coverage, longitudinal segmentation and granularity of the ATLAS calorimeters [68].

object reconstruction at constituent level, or after the reconstruction including this effect in a scale factor applied to recover the correct energy of the reconstructed object.

Electromagnetic calorimeter

The electromagnetic (EM) calorimeters [73, 74] are optimized to detect, contain and measure electron and photon showers. An important parameter expressing the shower containment is the radiation length, X_0, defined as the mean distance over which a high-energy electron on average loses all but $1/e$ of its energy by bremsstrahlung. The EM barrel calorimeter has a thickness of $> 24X_0$ while the endcaps has a thickness of $> 22X_0$ for an almost full containment of the electromagnetic showers generated by electrons and photons. The region between the barrel and the endcaps (crack region), $1.375 < |\eta| < 1.52$, contains additional material to instrument and cool the inner detector, and usually it is excluded from offline analysis for a precise identification and measurement of electrons candidates.

The EM calorimeters employ a lead-LAr technology with accordion-shape absorbers and electrodes, as shown in Figure 2.12. A high-voltage potential is placed through the electrodes across the medium (ionization gap) in order to collect the charges from the ionization produced by the passage of particles through the LAr. The width of this gap varies with the pseudorapity due to the accordion geometry, therefore, the high-voltage potential needs to vary accordingly to maintain a constant calorimeter response (defined as the measured fraction of the energy of the incident particle) as a function of the pseudorapity.

All charged particles produced in the development of an electromagnetic (induced by electrons and photons) or hadronic (induced by neutral and charged hadrons) cascade and entering the LAr generate an ionization signal in form of an electric current across the LAr gap [75]. This current is measured. To reduce and control fluctuations between the energy invested into the ionizations and the resulting current, the LAr is kept at a stable and constantly monitored temperature of 88 K and a stable density, by a dedicated cryogenic system.

Once the ionization charge is collected, it is transformed by the ATLAS electronics in a bipolar pulse shape signal, the amplitude of which gives the energy measurement. The integral of the whole pulse shape is 0, as it is characterized by a positive component and a long negative tail as shown in Figure 2.13. This approach was chosen to have an average cancellation of in-time and out-of-time

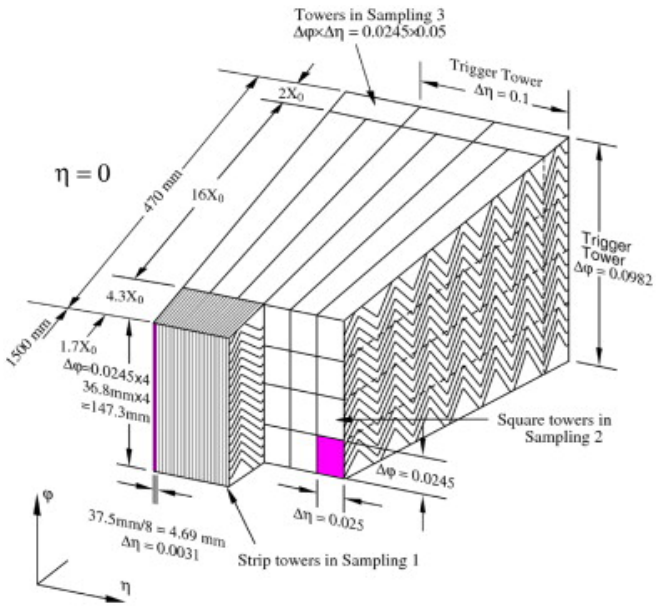


Figure 2.12: The accordion geometry is employed to obtain a hermetic system and a good segmentation [1].

pile-up for collisions happening at the nominal run conditions with a bunch spacing of 25 ns.

The accordion geometry is an innovative design to provide full and symmetric coverage in ϕ without cracks, high granularity and a segmentation in depth. Over the region devoted to precision physics, the EM calorimeter is segmented into three longitudinal sections: strips, middle and back.

The strips is the first layer, it is finely segmented along η, thus providing a high resolution, an accurate position measurement and discriminating power to distinguish photons from decaying neutral pions ($\pi^0 \to \gamma\gamma$).

The middle is the the second layer and collects the largest fraction of the energy of the electromagnetic shower.

The back is the the third layer aiming to collect the tail of the electromagnetic shower.

Finally, the electromagnetic calorimeters, both in the barrel and endcap regions, are integrated with **presamplers**, instrumented with finely segmented

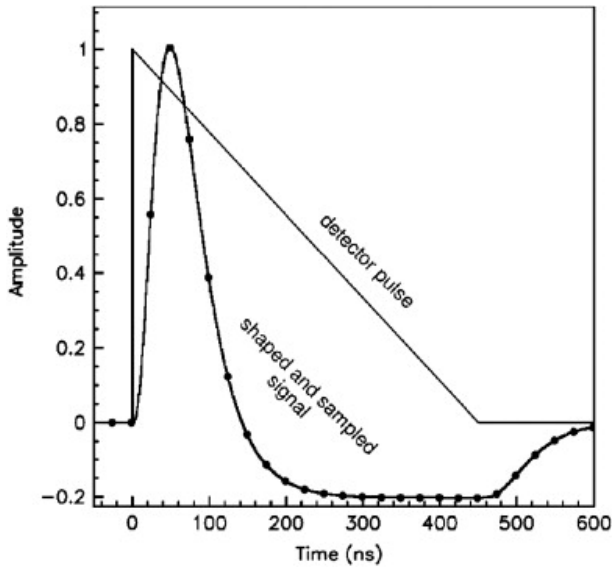


Figure 2.13: The ATLAS bipolar pulse shape [76].

readout structures that provide a measurement of the energy lost in front of the electromagnetic calorimeters.

Hadronic calorimeter

The hadronic calorimeter is designated to measure the energy of jets, a bundle of collimated particles generated by the hadronization of quarks and gluons. Its absorption power is defined in terms of its depth in interaction length, λ, which is defined as the mean distance traveled by a hadron before undergoing an inelastic interaction with matter. The hadronic calorimeter extends up to 9.7 λ in the barrel and 10 λ in the endcaps, thus ensuring good resolution for high-energy jets and limiting particles escaping into the muon spectrometer (punch-through). The material budget for each sub-calorimeter is reported in Figure 2.14 in terms of the interaction length and as a function of pseudorapidity.

The Tile calorimeter [77] (barrel region) is composed by scintillator tiles oriented radially to the beam pipe with steel as absorber material allowing to maximize the radial width while keeping the cost contained. The granularity is coarser with respect to the electromagnetic calorimeter but thigh enough to meet the resolution needs for proper jet reconstruction and energy measurement.

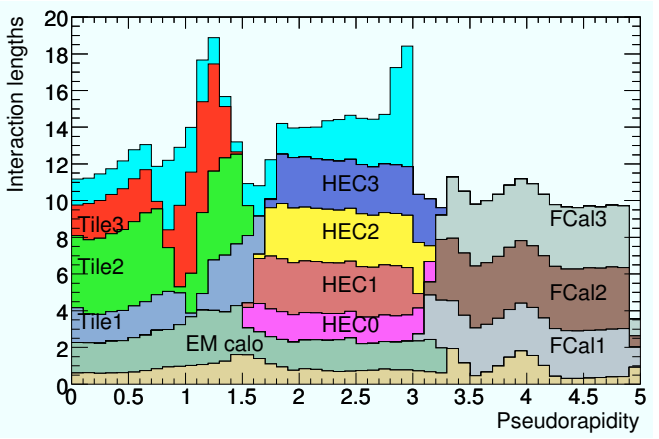


Figure 2.14: Calorimeter material in terms of interaction length as a function of pseudorapidity [1].

The endcap region features the HEC, a copper-LAr parallel plate hadronic calorimeter. The high radiation resistance of copper and LAr, and a high readout granularity, ensure that the calorimeter withstands the high particle fluxes and pile-up characterizing the regions closer to the beam pipe. The HEC consists of two independent wheels per endcap that cover the region $1.5 < |\eta| < 3.1$, overlapping both with the Tile and Forward calorimeters. Each wheel is divided into two longitudinal segments, for a total of four layers per end-cap.

Forward calorimeter

The forward calorimeters increase the calorimeter coverage towards the beam pipe, thus allowing the detection of forward physics objects and improving the resolution for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ measurement.

They are approximately 10 interaction lengths deep, and on each side consist of three modules along the beam axis direction. All of them employ LAr as active medium with different choices for the absorber medium. The first module employs copper and is optimized for electromagnetic shower measurements, while the other two employ tungsten and measure predominantly the energy of hadronic particles.

Since the forward calorimeters are located in the highest pseudorapidity region, they have to cope with very intense particle fluxes. Therefore, in order to deal with pile-up, a special matrix geometry was designed consisting in reg-

Calorimeter	Particle	Energy Resolution	
		a (% $\sqrt{\text{GeV}}$)	c (%)
Electromagnetic	electrons	10.0 ± 0.4	0.4 ± 0.1
Hadronic End-Cap	pions	70.6 ± 1.5	5.8 ± 0.2
Forward	electrons	28.5 ± 1.0	3.5 ± 0.1
	pions	94.2 ± 1.6	7.5 ± 0.4
Tile	pions	56.4 ± 0.4	5.5 ± 0.1

Table 2.3: Resolution of the different calorimeters for pions and electrons evaluated with test beam data, given by the stochastic term a and the constant term c as in equation 2.7. The constant term for the full electromagnetic calorimeter is expected to be around 1% [1].

ularly spaced longitudinal channels filled with an electrode structure parallel to the beam axis.

Calorimeter performance

The energy resolution of each sub-calorimeter is usually evaluated with the following expression:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{2.7}$$

the stochastic term (a) is the contribution arising from stochastic fluctuation in the energy measurements. The noise (b) and the constant (c) term are added in quadrature to account respectively for noise due to the calorimeter electronics and pile-up, and for energy that might be lost in non-instrumented areas of the detector, in addition to effects from uncorrected channel-to-channel signal inefficiencies. The measured resolution performance for different calorimeters is summarized in Table 2.3. Figure 2.15 reports the electronic and pile-up noise for each sub-calorimeter as a function of the pseudorapidity showing an increasing of the noise moving towards forward calorimeters. In particular, at high luminosity the noise in the endcaps and forward region is dominated by the pile-up contribution.

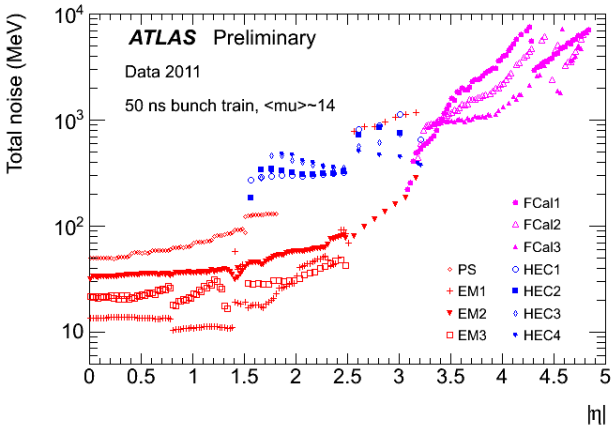


Figure 2.15: Electronic and pile-up noise for LAr subcalorimeters for $\langle\mu\rangle = 14$ [78]. At high luminosity, the noise in the endcaps and forward region is dominated by pile-up.

2.4.4 The forward detectors

The ATLAS experiment has detectors also in the most forward regions to provide inputs about very forward particle flow, including the measurement of the instantaneous luminosity, trigger events and control the general behavior of the experiment [1, 79].

LUCID (LUminosity Cherenkov Integrating Detector) is composed by two symmetric arms deployed at about 17 m from the ATLAS interaction point. The main aim of this detector is to monitor the luminosity delivered by the LHC machine to the ATLAS experiment.

ALFA (Absolute Luminosity For ATLAS) provides a luminosity measurement looking at elastic scattering at small angles (3 μrad). In order to achieve this measurement the two detector stations have to be placed far away from the interaction point (240 m) and as close as possible to the beam.

ZDC (Zero-Degree Calorimeters) has the main aim to detect forward neutrons and photons with $|\eta| > 8.3$, in both proton-proton and heavy-ion collisions. It measures the luminosity recorded by ATLAS, moreover, its inputs are used to reduce backgrounds created by beam-gas and beam-halo

effects, by requiring a tight coincidence from its two arms.

BPTX stations are located along the LHC on both sides of ATLAS, 175 m away from the interaction point. They are used for both L1-trigger and for the monitoring of beams and timing signals.

MBTS (Minimum Bias Trigger Scintillators) consist of two sets of scintillator counters installed in the inner face of the LAr endcap cryostat. They are used to trigger on minimum bias events.

2.4.5 The trigger and the data acquisition system

The high LHC luminosity gives a bunch collision rate of 40 MHz but current technology does not allow recording all detector signals for each collision. However, most of the corresponding interactions are due to soft physics and therefore not of interest for high transverse momentum SM and discovery physics analyses. Hence a three level trigger system [1] is used in ATLAS to discriminate the interesting events due to the hard scattering and to reduce the data flow rate to 200 Hz.

The first level, L1, gets inputs from the trigger muon chambers and the calorimeter to search for high p_{T} muons, electrons, photons, jets, and τ leptons decaying into hadrons, as well as large missing transverse momentum and large total energy. In addition, it is possible to turn on and off the p_{T} thresholds for different objects and set the p_{T} threshold levels, through a configurable trigger menu. In each event, if interesting objects are found, L1 defines one or more Region of Interest (RoI) providing the (η, ϕ) coordinate of the object and the criteria that it has passed. The L1 trigger takes about 2.5 μs to make a decision and reduces the interaction rate from 40 MHz to about 75 kHz. Events which pass the L1 selection are sent to the Data Acquisition system (DAQ) and to the next trigger step: the high level trigger (HLT).

The HLT is divided into Level2 (L2) and Event Filter (EF). The L2 trigger looks at the RoI defined by L1, and takes into consideration signals from the Inner Detector. It takes about 40 ms to make a decision and it is designed to reduce the trigger rate to about 3.5 kHz. Then data goes through the last step of the online selection, the EF, which has a decision time of 1 s and leads to a final rate of about 200 Hz. The EF uses algorithms similar to offline algorithms, including calibrations, alignment and electromagnetic field maps, to record the raw data (RDO).

2.4.6 ATLAS computing and analysis data model

The huge amount of data collected by the ATLAS detector needs a highly efficient and distributed computing system to be recorded, processed, stored and finally analyzed.

The computing toolkit relies on two basic aspects: a degree of hierarchy with distinct roles of the different computer facilities and a high degree of decentralization and sharing of computer resources based on the GRID paradigm [80].

The Analysis Data Model provides common interfaces and data objects which are necessary to ensure easy maintenance and coherence of the experiment software platform over a long period of time for a large collaboration as ATLAS. The data objects are created by the reconstruction program starting from the recorded raw data (RDO) [81]. The software framework used in ATLAS is called ATHENA, a C++ framework based on the GAUDI project [82]. Depending on the different level of information stored, different output formats are used: ESD, DESD, AOD and D3PD.

2.5 Conclusion

LHC is a powerful machine instrumented with large experiments able to achieve an impressive level of precision. Over the first three years of activity the machine, the experiments and the computing facilities performed brilliantly by exceeding all expectations. In particular, the luminosity was continuously increased and the accelerator delivered more than $6 \cdot 10^{18}$ of p-p collisions, allowing the LHC experiments to obtain important results quicker than expected.

During the last few days of activity, the space between bunches has been successfully reduced to 25 ns, in preparation for Run 2. The machine is now in a shut-down phase but an intense activity is under way to upgrade and consolidate the infrastructure to prepare LHC to safely operate at higher energy and luminosity. The major upgrade for the ATLAS detector concerns the tracking system: a fourth layer, the IBL (insertable B-layer), will be added closer to the beam pipe improving the tracking precision and ensuring good performance also for the high pile-up conditions that will characterize the Run 2.

All the valuable experience gained during the Run 1 and the one that will come with Run 2 is a precious starting point to further develop technologies and stimulate the engineer innovation for the subsequent high luminosity era.

Chapter 3

Physics object reconstruction

The proton-proton collisions at LHC produce many particles of different type resulting in a complex final state. In order to resolve this complexity, accurate and efficient particle reconstruction and identification has to be performed.

This is done by exploiting the different interactions of these particles with matter in dedicated ATLAS sub-detectors. The signals generated by these interactions feed into the reconstruction of the final state of a given collision event. The output of the reconstruction are the "physics objects", representing individual particles and particle jets with their respective kinematics, and missing transverse momentum carried by non-interacting particles. The same algorithms used to reconstruct these objects in data are used for the Monte Carlo simulations needed to test physics analysis predictions and performance results. The full ATLAS simulation chain is described in Section 3.1, then the algorithms for particle reconstruction and identification are discussed in Section 3.2. Finally, since its importance for this thesis, special attention is dedicated to the missing transverse momentum reconstruction and calibration in Section 3.3.

3.1 Monte Carlo simulation

Monte Carlo simulations (MC) are widely used in ATLAS to test and extrapolate performance for different run conditions, to derive energy correction factors and to estimate backgrounds for physics processes.

The MC production chain is generally divided into three steps: generation of the event and immediate particle decays, simulation of the particle interactions in the detector and digitization of the energy deposited in the sensitive regions of the detector into electronic signals corresponding to the ones generated in the readout of the ATLAS detector. The output of the simulation chain has a format identical to the output of the ATLAS data acquisition system. Thus, both the simulated and real data from the detector can then be run through the same ATLAS trigger and reconstruction software [83].

The information about stable particles ("truth" information) produced in each physics event is also recorded and can be processed in the reconstruction to measure the performance of the reconstruction software.

3.1.1 Event generation step

Event generators are indispensable tools for the modeling of the complex physics processes that are initiated by a p-p interaction at LHC, potentially leading to the production of hundreds of particles per event. The generator is responsible for any prompt decays and stores any stable particle expected to propagate through a part of the detector. At this level, filter algorithms can be provided to select only interesting event topologies, kinematic phase spaces, or interesting particles for a specific physics channel.

Many event generators are available in ATLAS, like Pythia [84], Alpgen [85], Sherpa [86] and McAtNlo [87]. Pythia has been chosen as the default generator thanks to its easy use, speed, and robustness but it can be supplemented by other generators, either to obtain some estimate of the uncertainties, or when specialized generators are expected to give a better physical description in certain final states.

The description of the proton substructure is encoded in the parton distributions functions (PDFs), which are generally used by all event generators as external inputs. Then, according to the different models used to describe the color coherence effects, fragmentation and confinement, different parton shower and hadronization generators can be employed, such as Pythia or the combination of Herwig+Jimmy [88, 89] specially tuned for the underlying event at ATLAS.

Finally, specialized generators can be run in conjunction with general purpose generators to improve the accuracy for specific decays or specific final states. For example, the Photos and Tauola generators [83, 90, 91] are employed to respectively handle modeling of higher order electromagnetic radiation and tau

decays where particular attention is paid to the tau polarization.

3.1.2 Detector simulation and digitization steps

In the simulation process, each particle provided by the event generator is propagated through a model of the full ATLAS detector. This task performed using the Geant4 (GEometry ANd Tracking) [92, 93], toolkit for the simulation of the passage of particles through matter. In Geant4 it is possible to encode detailed information about the particle interactions and the detector structure, such as the detector geometry including misalignments and distortions, the position and the extension of dead materials, the maps of the electromagnetic fields and the detector response.

Since its complexity, large computing resources are required to accurately model the detector geometry and the detailed physics descriptions in the standard ATLAS detector simulation. This put limits on the available statistics for the Monte Carlo simulation samples, some of which cannot be large enough to meet the requirements of specific physics studies, especially with increasing of the luminosity. This led to the development of some fast simulation strategies which enable faster production of large Monte Carlo samples.

The default fast simulation in ATLAS is Atlfast-II. It reduces the simulation time by one order of magnitude by means of parameterizations of the longitudinal and lateral energy profile of the electromagnetic and hadronic showers in the calorimeters combined with the standard simulation in the Inner Detector and Muon spectrometer. A further order of magnitude in simulation time can be gained using a fast track simulation (Atlfast-IIF) in the Inner Detector and Muon System based on simplified geometry and parameterizations of physics processes [94].

Activity from multiple pile-up interactions per bunch crossing is modeled by overlaying simulated minimum bias events, generated with Pythia and specially tuned for minimum bias interactions at the LHC, over the original hard-scattering event. Recently a pile-up overlay using real zero-bias data events [95] has been also tested providing encouraging results especially for the improved agreement between data and MC simulations.

As a final step, the energies deposited in the sensitive regions of the detector are recorded as "hits" containing the total energy deposition, position, and time. At this level the digitization process [95] is applied to perform the conversion of the energy deposited by particles into electronic signals reproducing the output

given by the readout of the ATLAS detectors. It employs detailed modeling of the signal formation, including the noise, signal shaping and digitalization, in the real detector electronics.

3.2 Physics objects reconstruction

Starting from the recorded electronic detector signals, sophisticated algorithms reconstruct tracks and calorimeter energy clusters. These are the primary elements for physics object reconstruction, particle identification and computation of the particle energy and direction.

The particle reconstruction performance is tested on data and then compared with MC simulation. Since any disagreement between data and MC is propagated directly into physics analyses, where mismodelling are observed, specific parametrized corrections called "scaling factors" are computed to reflect MC particles reconstruction efficiency, isolation, energy resolution and scale, to match the values observed in data. This significantly improves the accuracy of the yields predicted by the simulation in physics analyses.

Systematics uncertainties on the scaling factors are also provided. Their impact on a specific physics analysis depends on the relative importance of the reconstructed physics objects in the final state of the considered analysis.

Calorimeter clustering algorithms

Incoming particles usually deposit their energy in many calorimeter cells, both in the lateral and longitudinal directions (with respect to the particle direction of flight). Clustering algorithms are designed to group these cells scanning the whole calorimeter and to collect the total deposited energy into clusters. These cluster energies are then calibrated to account for the energy deposited outside the cluster and in dead material. The calibration depends on the incoming particle type.

The cell grouping can be either performed with a fixed window size as done in the *sliding window* approach used for electrons and photons, or with a variable size based on the cell signal significance, as done in the *topocluster* approach used for taus, jet and $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ [96]. In particular the topocluster approach is efficient at suppressing noise, and thus improving the energy resolution of the physics objects built from the formed topological clusters.

The topological clustering algorithm usually runs in two steps: the cluster building and the cluster splitting.

The cluster building starts with the identification of a seed cell with significant energy compared to the expected electronic and pile-up noise. This threshold is optimized to be $|E_{\text{cell}}| \geq 4\sigma_{\text{noise}}$ [1]. Then the cluster develops in the three dimensions by adding all the neighbouring cells with $|E_{\text{cell}}| \geq 2\sigma_{\text{noise}}$ and finally all the neighbouring cells of the accumulated ones are added as well, as sketched in Figure 3.1. This 4-2-0 scheme has be found to be the most performant configuration also in busy pile-up environments.

The cluster splitting searches for local maxima with energies larger than 500 MeV and larger than the energy of any neighbouring cells in the clusters built as described above. If a cluster contains more than one maximum it is split.

In order to correct for non-compensating calorimeter effects and energy losses, a calibration can be applied on the clustered cells. The default calibration in ATLAS is based on the local hadronic calibration (LCW) scheme [97, 98], that uses properties of clusters to calibrate them individually. It first classifies calorimeter clusters as electromagnetic or hadronic, according to the cluster topology, and then weights each calorimeter cell signal in clusters according to the cluster energy and the cell energy density. Additional corrections are applied to the cluster energy for the average energy deposited in the non-active material before and between the calorimeters and for unclustered calorimeter energy.

3.2.1 Electrons and photons reconstruction

Electrons and photons reconstruction and identification algorithms are designed to ensure a good discrimination against background objects such as hadron jets and stable performance over the full detector acceptance and over a broad energy range (few GeV to few TeV) [100]. In order to achieve this task, a combination of signals from the ATLAS sub-detectors is used, including electromagnetic calorimeter, inner detector and TRT.

The first step of the procedure consists in an efficient reconstruction of the calorimeter electromagnetic showers based on a "sliding-window" algorithm. The strategy of the algorithm is to group cells moving over the calorimeter a fixed window of size $N_\eta \times N_\phi = 3 \times 5$ [2]. A seed cluster is identified if the energy sum of

[1] σ_{noise} is the Gaussian width of the cell energy distribution measured in randomly triggered events.

[2] N_η, N_ϕ are respectively the number of cells in the middle layer in the η and in the ϕ

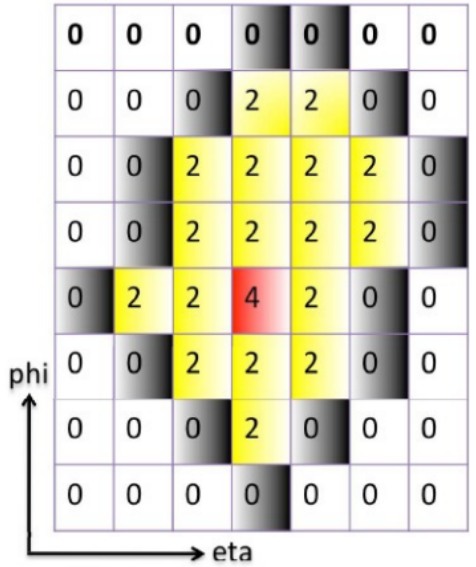


Figure 3.1: Sketch of the topological cluster building [99].

the window cells is above a threshold of 2.5 GeV [96]. The window size and the seed energy threshold are optimized to obtain the best reconstruction efficiency, to collect most of the energy deposited by the particle in the calorimeter, and at the same time to limit the fake rate due to electronic and pile-up noise.

The inner detector information is included to discriminate electrons from photons. Electromagnetic clusters are matched with the tracks extrapolated to the second EM calorimeter layer. If the cluster has no associated track the object is classified as a photon candidate, otherwise as an electron candidate. Then, electrons that are actually from converted photons are tagged looking for secondary vertices. This is of particular importance since the fraction of converted photons is significant, spanning from $\sim 30\%$ in the central region to $\sim 45\%$ in the endcap region [101].

In the identification process, criteria based on shape variables computed from the lateral and longitudinal energy profiles of the shower in the electromagnetic calorimeter and a veto on the energy deposited in the hadronic calorimeter are used to reject charged and neutral hadrons. For electrons, additional criteria are required to ensure a good track quality, strict track-cluster matching and

direction, respectively. Each middle layer cell has size of 0.025×0.025.

high-threshold hits in the TRT. The selection criteria vary as a function of the reconstructed η of the candidate to take into account significant change in the total thickness of the upstream material and variations in the calorimeter geometry or granularity. In particular the very fine granularity of the first EM calorimeter layer allows for a discrimination between single photon shower and two overlapping showers originating from a π^0 decay.

According to the tightness of the selection different working points can be defined [102]. For electrons, there are mainly three working points: loose, medium and tight. The most commonly used working point is the medium one that ensure a high-p_{T} electron efficiency near to 90% at a few percent fake rate. Similarly, for photons two working points are defined: loose and tight. The photon purity is around 90% for isolated high energy photons but it is strongly reduced for non-isolated ones, to about 50% for low energy photons [103].

After the reconstruction, the sliding window cluster energy is calibrated with specific methods based on Monte Carlo simulation [104]. For electrons the energy is computed as a weighted average between the cluster energy and the track momentum. The η and ϕ directions are usually taken from the corresponding track parameters at the vertex. The track refitting is performed with the Gaussian Sum Filter (GSF) algorithm [105], in order to account for the effects of radiative energy by bremsstrahlung, which can give deviations from the original charged particles path especially for high-energy electron.

The standard calibration is individually optimized for electrons, unconverted photons and converted photons and estimates separately four sources of energy loss.

The "front" component is the energy loss due to the amount of material in front of the calorimeter.

The "sampling" or "accordion" component is the energy loss due to dead material inside the calorimeter.

The "out of cluster" component is the energy loss laterally outside of the reconstructed cluster.

The "leakage" component is the energy loss longitudinally behind the electromagnetic calorimeter.

An in-situ calibration with collision data allows to determine the absolute energy scale and inter-calibrates the different regions of the calorimeters. Electrons

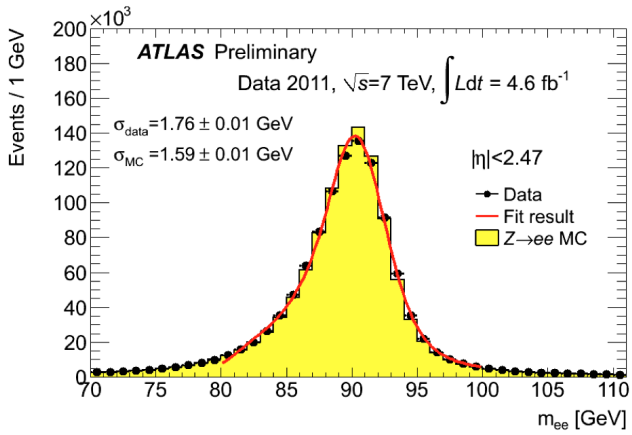


Figure 3.2: Invariant Z mass for electron pairs compared with the Monte Carlo simulation prediction. There is a good agreement between data, MC predictions and the fit results [106].

produced in Z decays are used as shown in Figure 3.2. The derived calibrations and corrections are cross-checked with electrons from $W \to e\nu$ and $J/\psi \to ee$ events.

3.2.2 Muons reconstruction

Mainly two sub-detector systems are involved in muon identification and reconstruction: the Muon Spectrometer (MS) and the Inner Detector (ID). They provide independent momentum measurements that can be combined by specific algorithms to increase the purity and performances of the reconstructed muons. In specific cases also the energy deposited by the muon in the calorimeters, which is on average of about 2-3 GeV, can be used. This allow an optimal reconstruction performance over a large η range and over a broad p_{T} range (from a few GeV up to a few TeV) [100, 107].

In the identification process, according to the available information from different sub-detectors, muons can be classified as follow.

Standalone muons are reconstructed using only the MS information. This extends the coverage up to $|\eta| = 2.7$. The direction of flight and the impact parameter of the muon at the interaction point are determined by extrapolating the spectrometer track back to the beam line taking the energy loss

of the muon in the calorimeters into account.

Combined muons are obtained by the combination of independent MS and ID measurements. The match is performed by a χ^2 quality test. The combination improves the resolution with respect to the single ID and MS measurements and it allows rejection of muons from secondary interactions and from π/K decays in flight.

Segment-tagged muons are identified only by segments in the MS, so the momentum can be reconstructed using only the ID information. The ID track is used as a seed and it is extrapolated to the first station of the MS to be matched, using a χ^2 quality test, with track segments in the precision muons chambers. They are employed to recover a small inactive Muon Spectrometer region around $|\eta| \sim 1.2$.

Calorimeter-tagged muons are identified only by a track in the ID and by calorimetric information. The ID track is used as a seed and the associated energy deposits in the calorimeter are used to check the compatibility with the minimum ionizing particle hypothesis. These muons have lower purity but can help to recover acceptance in the un-instrumented region of the MS around $\eta \sim 0$.

The capability of the ATLAS detector to reconstruct muons on a wide p_{T} range is shown in Figure 3.3, where the di-muon spectrum is shown.

The muon efficiency reconstruction (on average above 97%) and the momentum resolution (1.5-3 GeV) are measured using data-driven techniques employing $Z \to \mu\mu$ or $J/\psi \to \mu\mu$ decays. The comparison with MC simulation allows to derive scale factor corrections as a function of the muon momentum and pseudo-rapidity.

In ATLAS two independent algorithms, *Staco* and $MuID$, are available for the muon reconstruction. A third muon chain called $Muons$ has been recently provided combining the previous approaches and it will be the default for ATLAS Run 2. For the results reported in this thesis only *Staco* muons are used.

3.2.3 Taus reconstruction

Tau leptons are the heaviest known leptons, with a mass of 1.777 GeV [109]. Due to this, taus are the only leptons for which also decays into hadronic particles are allowed. They occur in 64.8% of all tau decays [110]. In this decay mode

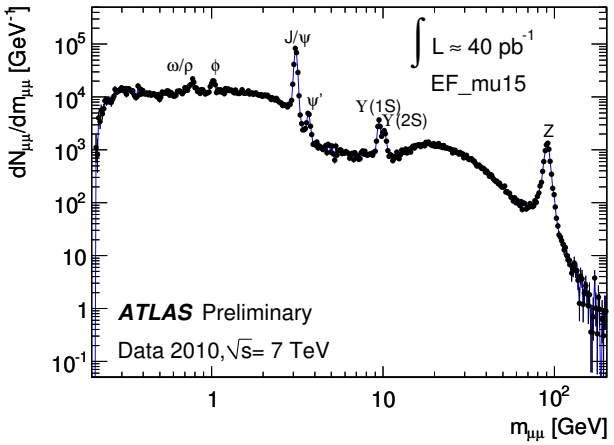


Figure 3.3: Di-muon invariant mass spectrum for data from combined opposite sign muons. Visible peaks correspond to the different resonances reported in the plot [108].

(referred to as τ_{had}), the tau decays to a ν_τ in addition to one or more hadrons (predominantly pions). The relative branching fractions are reported in table 3.1.

With a proper decay length of 87 μm, tau leptons decay before reaching the detector and can only be identified through the reconstruction of their decay products. In the case of leptonic decays ($\tau \to \ell\nu_\ell\nu_\tau$), the decay products cannot be distinguished from prompt electrons or muons, therefore here after only hadronic decays are considered.

τ_{had} candidates are reconstructed using the jet anti-k_t algorithm [111] with a distance parameter $R = 0.4$. They are seeded from jets with $E_{\text{T}} \geq 10$ GeV and $|\eta| \leq 2.5$. The tau candidate can be associated to a different vertex from the one with the highest Σp_{T}^2 (identified as the primary vertex) by a Tau Jet Vertex Association (TJVA) algorithm [112]. Consequently, calorimeter cell and cluster directions are calculated in a coordinate system having the TJVA vertex as origin, and only tracks associated to this vertex are considered. Tracks passing the following quality criteria:

- $p_{\text{T}} \geq 1$ GeV
- $N_{\text{hits}}^{\text{PIXEL}} \geq 2$, $N_{\text{hits}}^{\text{PIXEL}} + N_{\text{hits}}^{\text{SCT}} \geq 7$
- $|d_0| \leq 1.0$ mm, $|z_0 \sin\theta| \leq 1.5$ mm

Decay mode	decay channel	BR
Hadronic decay	$\pi^{\pm}\nu$	11%
	$\pi^{\pm}\pi^{0}\nu$	25%
	$\pi^{\pm}\pi^{0}\pi^{0}\nu$	9%
	$\pi^{\pm}\pi^{\pm}\pi^{\pm}\nu$	9%
	$\pi^{\pm}\pi^{\pm}\pi^{\pm}\pi^{0}\nu$	5%
	other	5%
Leptonic decay	$e\nu\nu$	17%
	$\mu\nu\nu$	17%

Table 3.1: Branching fraction for tau decay modes [109].

are then associated to the candidate if they fall within the core region, defined as $\Delta R < 0.2$, around the jet barycentre. According to the number of associate tracks, τ_{had} candidates are defined either as 1-prong (1 associated track) or multi-prong (mostly 3 associated tracks).

Due to the background from multijet processes, efficient tau identification techniques with large jet rejection are essential. The narrow shower shape and the small number of tracks characteristic of hadronic tau decays are useful in discriminating them from other signatures in the detector. But, since these single variable criteria are not enough to efficiently distinguish taus from jets and electrons, multivariate techniques are employed. The two main algorithms for the tau identification in ATLAS are: a projective likelihood method (LLH) and a boosted decision trees method (BDT) that is used in the $H \to \tau\tau$ analysis reported in this thesis. The BDT is fed with tracking information and calorimeter shape variables properly corrected to mitigate the pile-up effects on the output result.

Three working points for the tau identification are established based on the signal efficiency: loose, medium, and tight. For 1-prong (multi-prong) taus, these efficiencies are 70% (65%), 60% (55%), and 40% (35%), respectively.

Additional fake taus are generated by electrons that mimic the signature of a 1-prong τ_{had}, and by muons if a muon track is associated with a sufficiently energetic calorimeter cluster. In order to reject these backgrounds a further BDT and cut-based approach is used for electrons and muons respectively [113].

In tau reconstruction, calorimeter topoclusters are already used calibrated at the LCW scale to account for the not-compensating calorimeter effects and for

the energy lost in dead materials, yielding an improvement of the τ_{had} energy resolution with respect to the use of topological clusters at the EM scale. Nevertheless, in order to restore the true visible tau energy, a proper calibration (TES) is required to correct for energy lost before the calorimeters, for underlying event and pileup contributions, and for out-of-cone effects [114].

The TES calibration is based on simulated tau decays, with its systematic uncertainty including contributions from single particle response measurements, pile-up and material modeling. The single particle uncertainties are evaluated by in-situ measurements based on the comparison between the calorimeter energy measurements and the momenta measured in the Inner Detector (E/p). This evaluation is then combined with test-beam measurements for $|\eta| < 0.8$ and with an uncertainty estimated comparing different simulated shower models for $0.8 < |\eta| < 2.5$. In-time pile-up effects are corrected with an offset method while out-of-time pile-up effects are found to be less then 1% and thus negligible. The results for 1- and multi-prong taus in the central region of the detector are shown in Figure 3.4: the systematic uncertainty of the TES for $p_{\text{T}}^{\tau} > 20$ GeV and $|\eta| < 2.5$ is found to be less then 3% for the hadronic decay modes with exactly one reconstructed track, and less then 4% for the hadronic decay modes with at least two reconstructed tracks.

The uncertainty on the TES is also evaluated by using in-situ measurements for cross check. The strategy relies on the reconstruction of the visible mass peak[3] for Z bosons decaying semileptonically into $Z \to \tau\tau \to \mu\nu_{\mu}\nu_{\tau}\tau_{\text{had}}$. The results are compatible with the systematic uncertainties determined with the formerly described method.

3.2.4 Jets reconstruction

As result of the strong interaction and hadronization discussed in Section 1.2.1, quarks and gluons materialize into collimated bunches of hadrons flying roughly in the same direction, , the so-called jets. Therefore, their jet reconstruction and identification is crucial to resolve the partonic flow coming from the hard scatter interaction.

Jets in theory and experiment are the result of an algorithm mapping observable final state particles or signals representing them, into one kinematic object if these particles are signals are deemed to come from a common source (parton). This mapping is not deterministic, due limitations in the calculation of

[3] invariant mass of the visible products, no $\boldsymbol{E}_{\text{T}}^{\text{miss}}$ is included.

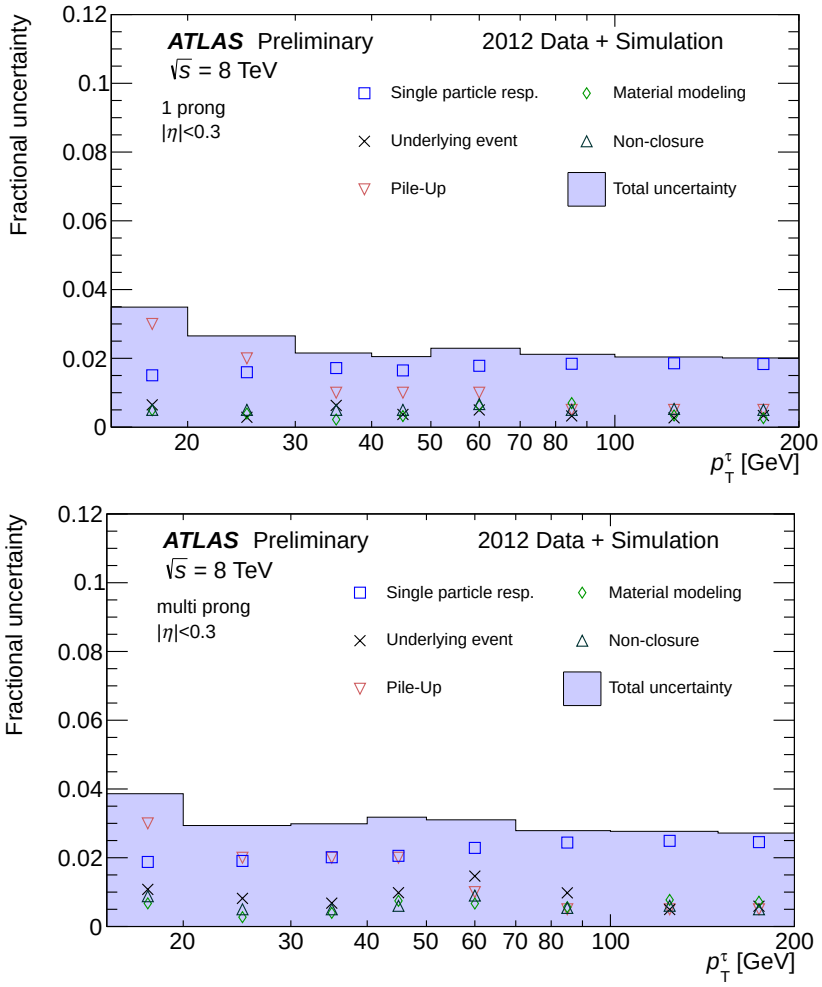


Figure 3.4: TES uncertainty for $\tau_{1-prong}$ (top) and $\tau_{multi-prong}$ (bottom) for a central pseudo rapidity bin ($|\eta| < 0.3$). The individual contributions are shown as points and the combined uncertainty is shown as a filled band. [114].

the parton shower and fragmentation introduced by non-perturbative kinematic domains, and the limitations of the detectors in the reconstruction of all particles generated in the collision. As M. J. Tannenbaum said in [115], "Jets are legal contracts between theorists and experimentalists". This means, that the algorithms defining a jet in the experiment and in calculations need to be completely specified, and follow a few rules to allow for comparisons of measurements with theory. The "Snowmass Agreement" [116] collects these rules:

1. Simple to implement in experimental analysis;
2. Simple to implement in theoretical calculation;
3. Defined at any order of perturbation theory;
4. Yields at a finite cross section at any order of perturbation theory;
5. Yields to a cross section that is relatively insensitive to hadronization;

The fourth requirement directly translates in the infrared and collinear safety (IRC) condition, i.e.: the number of reconstructed jets has to be independent on the emission of a soft (infrared) or collinear particle [117].

Sequential recombination jet algorithms are specifically designed to satisfy the IRC condition and thus to be usable for calculations at any order in perturbation theory. Typically they work by calculating a distance between particles, and then recombine[4] them pairwise according to a given order, until some condition is met. The process terminates when no particles are left.

In order to suppress the calorimeter noise, the topoclusters described in Section 3.2 are used as input objects to jet finding in the experiment. Then, the distance d_{ij} between two objects and the distance d_{iB} between an object and the beam can be defined as:

$$d_{ij} = min\left(k_{\mathrm{T},i}^{2p}, k_{\mathrm{T},j}^{2p}\right) \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2} \tag{3.1}$$

$$d_{iB} = k_{\mathrm{T},i}^{2p} \tag{3.2}$$

where $k_{\mathrm{T},i}$, y_i and ϕ_i are respectively the transverse momentum, the rapidity and the azimuthal angle of the considered object, R is the resolution parameter controlling the extension of the jet, and p is a parameter defining the distance

[4] A four-mometum recombination scheme is used: the combination of two objects is performed via a four-momentum sum.

scales, as discussed later. The R value is a compromise to limit the contribution from underlying event but at the same time to capture all the energy in the direction around the initial quark or gluon generated through parton shower and hadronization effects. In the results reported in this thesis all jets are reconstructed with $R = 0.4$, if not stated differently.

The recombination procedure is defined by the following steps: a list of all the distances d_{ij} defined in Equation 3.1 is compiled, then the minimum distance among all objects, $d_{\min} = \min(d_{ij})$ is computed and compared with the beam distance d_{iB}. If $d_{\min} < d_{iB}$ then the i^{th} and the j^{th} objects are combined and the procedure is repeated from the start. Otherwise the i^{th} object is identified as a jet and removed from the list. The procedure is repeated until all the objects are removed from the input list and classified as jets.

According to the value of the p parameter in Equation 3.1, different algorithms with different properties are defined. For $p = 0$, the algorithm is known as Cambridge-Aachen (C/A) [118] and the object recombination is performed only on the base of the geometrical distance. Shortest distance objects are recombined first, which leads to irregularly shaped jets. The k_t algorithm is obtained by setting $p = 1$. It is characterized by clustering first soft objects, resulting in rather irregular shapes for the final jets [119]. Finally for $p = -1$ the anti-k_t algorithm is obtained [111]. This is the ATLAS default because by clustering the hardest contributions first it effectively removes sensitivity to the internal structure of the parton shower and results in rather regularly shaped jets. Nevertheless, k_t and C/A algorithm can be useful tools for studies looking at the jet substructure and for pile-up suppression methods [120, 121, 122].

A correction factor, the jet energy scale (JES), is required to account for the lower hadron response, calorimeter non-uniformities, pile-up, energy loses and "out-of-cone" effects. The JES is derived using MC-based and in-situ methods according to the sequential procedure sketched in Figure 3.5. The procedure starts with jets at the constituents scale that can be calibrated either with the EM[5] or the LCW calibration. Jets used in the analysis reported in this thesis are LCW calibrated. Then, a pile-up subtraction is performed relying on the Jet Area method [120]. A residual offset correction as a function of N_{PV} and μ is also applied. It is mostly relevant in the high-μ and high-η region where the out-of-time pile-up effects are more important. A MC based correction is then computed

[5]The EM scale is the basic calorimeter signal scale for the ATLAS calorimeters. It provides the correct scale for energy deposited by electromagnetic showers but does not correct for energy leakage or losses in the dead material.

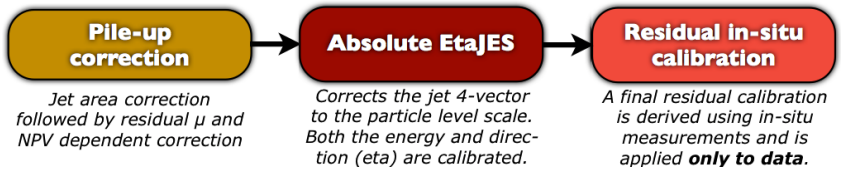


Figure 3.5: Sketch of the JES sequential procedure [125].

to account for all the detector non-uniformites, differences in response for neutral components and for different p_{T} jets. Finally, in-situ methods based on the ratio between the jet p_{T} and a well measured p_{T} reference, are exploited to ensure good uniformity over the whole detector using di-jets events for η intercalibration, and to ensure coherence across a broad jet p_{T} range using Z+jet balance for the low $p_{\mathrm{T}}^{\mathrm{jet}}$ region, γ+jet balance for the intermediated $p_{\mathrm{T}}^{\mathrm{jet}}$ range and multi-jet balance for the high $p_{\mathrm{T}}^{\mathrm{jet}}$ region up to 1 TeV [123, 124].

In-situ methods give a large number (up to 49) contributions to the JES uncertainties grouped in four categories. In addition, systematics for pile-up, flavor, MC non-closure, E/p measurements are considered.

In physics analyses using a profile likelihood approach, systematics uncertainties are treated as nuisance parameters (NP). A good splitting of the different systematics sources is therefore required to ensure a proper combination between different analysis channels or through different periods of data taking with different experimental conditions (e.g., pile-up). Nevertheless, the full systematics scheme typically has too many NPs for a general analysis. Therefore reduction schemes are provided by grouping the in-situ NPs that contain less information about correlations. At the price of loosing only a few % of the correlations a reduction scheme with a total of 24 NPs is used in the final analysis reported in this thesis. A further reduction is then applied at the analysis level in order to include only significative systematics variations in the fit for the limits extraction excluding statistics noise.

The results for the 2012 JES uncertainties split in its main components are shown in Figure in 3.6. In particular the JES uncertainties in the high-η regions are up to 4-7% and have an important contribution in vector boson fusion topologies due the presence of forward jets (see Chapter 5).

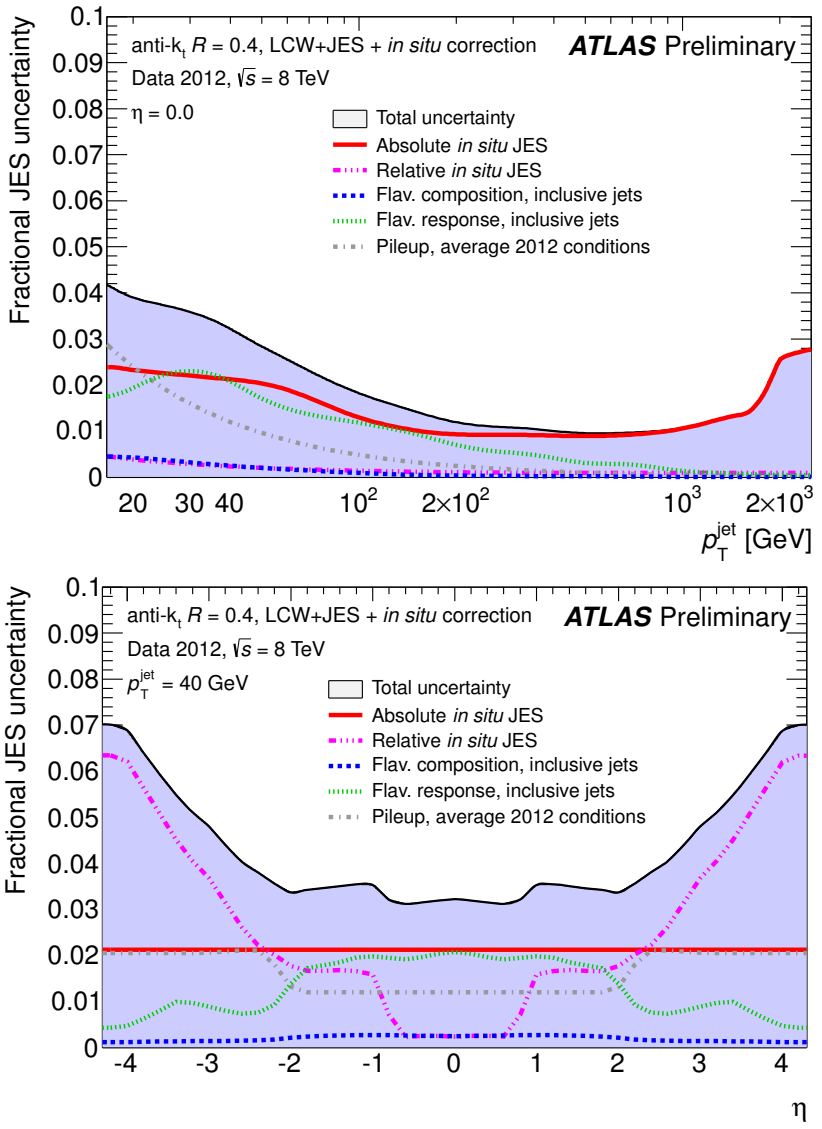


Figure 3.6: Fractional jet energy scale systematic uncertainty components for anti-k_t jets with $R = 0.4$ calibrated using the LCW+JES calibration scheme as a function of the jet p_{T} (top) and ($|\eta|$) (bottom). The total uncertainty (all components summed in quadrature) is shown as a filled blue region topped by a solid black line. Average 2012 pile-up conditions were used, and topology dependent components were taken from inclusive dijet samples [126].

3.3 $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction

In a collider event the missing transverse momentum is defined as the momentum imbalance in the plane transverse to the beam axis, where momentum conservation is expected. Such an imbalance may signal the presence of unseen particles, such as neutrinos or stable, weakly-interacting supersymmetric (SUSY) particles. The vector momentum imbalance in the transverse plane is obtained from the negative vector sum of the momenta of all particles detected in a pp collison and is denoted as missing transverse momentum, $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$. The symbol $E_{\mathrm{T}}^{\mathrm{miss}}$ is used for its magnitude [127, 128, 129].

The precise measurement of $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is essential for physics at LHC. A large $E_{\mathrm{T}}^{\mathrm{miss}}$ is a key signature for searches for new physics processes such as SUSY and extra dimensions. The measurement of $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ also has a direct impact on the quality of a number of measurements of Standard Model (SM) physics, such as the reconstruction of the top-quark mass in $t\bar{t}$ events. Furthermore, it is crucial in the search for the Higgs boson in the decay channels $H \to WW$ and $H \to \tau\tau$, where a good $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ measurement improves the reconstruction of the Higgs boson mass [130].

An important requirement on the measurement of $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is the minimization of the impact of limited detector coverage, finite detector resolution, the presence of dead regions and different sources of noise that can produce fake $E_{\mathrm{T}}^{\mathrm{miss}}$. Such sources can significantly enhance the background from multi-jet events in SUSY searches with large $E_{\mathrm{T}}^{\mathrm{miss}}$ or the background from $Z \to \ell\ell$ events accompanied by jets of high transverse momentum in Higgs boson searches $H \to WW \to \ell\nu\ell\nu$ in final states with two leptons and neutrinos.

The values of $E_{\mathrm{T}}^{\mathrm{miss}}$ and its azimuthal coordinate (ϕ^{miss}) are calculated as:

$$\begin{aligned} E_{\mathrm{T}}^{\mathrm{miss}} &= \sqrt{\left(E_x^{\mathrm{miss}}\right)^2 + \left(E_y^{\mathrm{miss}}\right)^2} \\ \phi^{\mathrm{miss}} &= \arctan(E_y^{\mathrm{miss}}/E_x^{\mathrm{miss}}) \end{aligned} \tag{3.3}$$

where E_x^{miss}, E_y^{miss} are the negative sum of all the momentum components (p_x, p_y) reconstructed with the detector. measured in the detector projected respectively onto the x and y direction. Primarily these include contributions from energy deposits in the calorimeters ($E_{x(y)}^{\mathrm{miss,calo}}$) and muons reconstructed in the muon spectrometer ($E_{x(y)}^{\mathrm{miss},\mu}$):

$$E_{x(y)}^{\mathrm{miss}} = E_{x(y)}^{\mathrm{miss,calo}} + E_{x(y)}^{\mathrm{miss},\mu} \tag{3.4}$$

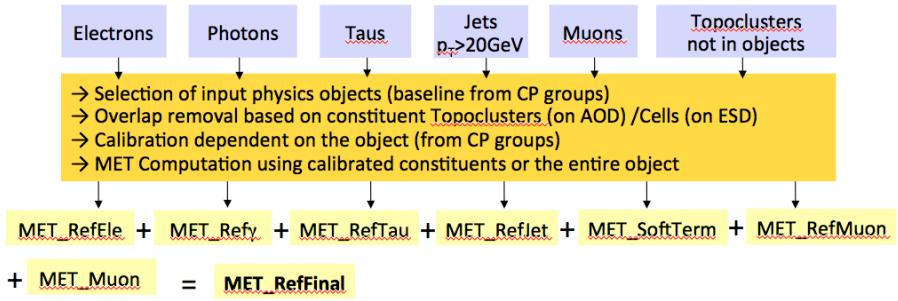


Figure 3.7: Sketch of the *RefFinal* algorithm.

Moreover, information from the inner detector is added as well: low-p_{T} tracks are used to recover low p_{T} particles which do not reach in the calorimeters, and muons reconstructed from the inner detector are used to recover muons in regions poorly covered by the muon spectrometer.

In the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction algorithm, the computation of the calorimeter term uses only the energy in topoclusters (Section 3.2) in order to suppress electronic and pile-up noise and to use only the calorimeter energy deposits that generate a significant signal. Other $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ algorithms, based on a simple cell σ_{noise} cut to suppress the calorimeter noise, have been studied as well [131]. Indeed, despite they have shown to be less performant specially in terms of $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution, they can be employed for tests of the detector performance particularly in very busy environments, like heavy-ions collisions, where the topocluster approach may not be able to ensure anymore infrared safety.

In an early phase of the ATLAS data taking, $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ algorithms that calibrate all the calorimeter cells according to the same calibration scheme (LCW) have been firstly employed [131, 132, 133]. A significant improvement of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance is achieved with a more refined algorithm for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, tagged *RefFinal* [127, 128, 129, 133], where a proper calibration is applied to each physics object, thus providing a $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ computation coherent with the specific choices in the context of a physics analysis. A detailed description of the algorithm is given in Section 3.3.1.

3.3.1 The *RefFinal* algorithm

In this section the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction and calibration based on the *RefFinal* algorithm sketched in Figure 3.7 is described in detail. The $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ calculation

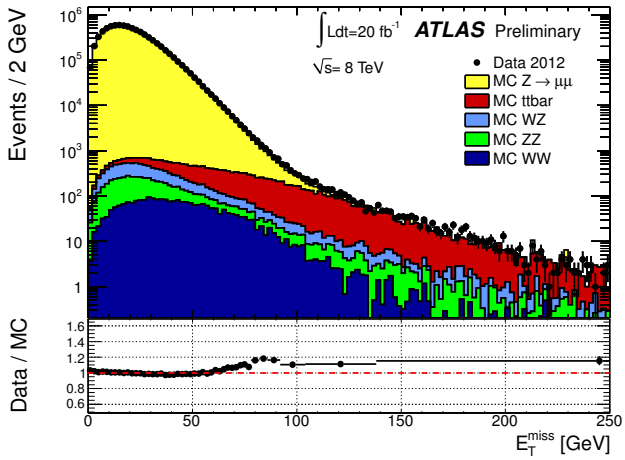


Figure 3.8: Data-MC comparison for the total $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ distribution in $Z \to \mu\mu$ events.

uses reconstructed and well calibrated physics objects. The overlap between the different objects in calorimeters is resolved by associating topoclusters to the reconstructed objects in a defined order: electrons, photons, taus, jets, muons. Topoclusters not associated with any such objects are also taken into account in the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ calculation by collecting them into the soft term $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$. Therefore, the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is calculated as follows:

$$E_{x(y)}^{\mathrm{miss}} = E_{x(y)}^{\mathrm{miss},e} + E_{x(y)}^{\mathrm{miss},\gamma} + E_{x(y)}^{\mathrm{miss},\tau} + E_{x(y)}^{\mathrm{miss,jets}} + E_{x(y)}^{\mathrm{miss,SoftTerm}} + E_{x(y)}^{\mathrm{miss},\mu} \tag{3.5}$$

where each term is calculated as the negative sum of the calibrated reconstructed objects, projected onto the x and y directions. Particular attention is paid in avoiding energy double counting of the various physics objects that could create fake unbalance in the event and result in tails of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ distribution. Figures 3.8 shows that there is a good agreement between data and MC simulation prediction for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ total distribution.

The definition and the calibration of the physics objects entering the *RefFinal* algorithm are described in the following sections for the default configuration used to produce the results reported in this thesis. However, the physics object options are customizable to meet the analysis selection choices.

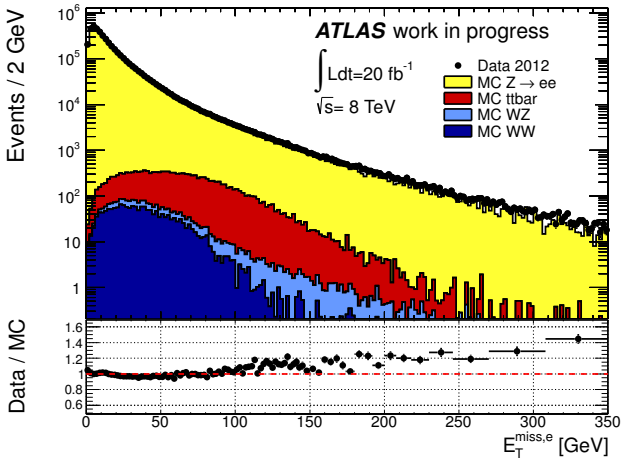


Figure 3.9: Data-MC comparison for the $E_{\mathrm{T}}^{\mathrm{miss},e}$ term in $Z \to ee$ events.

Electron term, $E_{\mathrm{T}}^{\mathrm{miss},e}$

Electrons are required to pass "medium" identification criteria (Section 3.2.1) and have p_{T} greater than 10 GeV. They are treated by the algorithm as full four-momentum objects including the ATLAS standard electron calibration [102]. Since the contribution of topoclusters out of objects is already included in the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ term, the "out-of-cone" correction of the electron calibration is removed to avoid energy double counting. Moreover, since electrons clusters are built with the sliding-window approach (Section 3.2), topoclusters containing more than the 50% of the the sliding window cells are removed in order to avoid energy double counting. Figure 3.9 shows a good agreement for the electron term in $Z \to ee$ events.

Photon term, $E_{\mathrm{T}}^{\mathrm{miss},\gamma}$

Photons are required to pass the "tight" identification criteria (Section 3.2.1) and have p_{T} greater than 10 GeV. If a photon is also reconstructed as an electron, the electron is kept. Since the photon purity is poor for not isolated photons, photons are usually calibrated at the EM scale. For analyses selecting photons in the final state, in order to improve the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, it is possible to customize a specific $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ including the proper calibration for the selected photons. As for electrons, topoclusters containing more than the 50% of the sliding window cells

are removed in order to avoid energy double counting.

Tau term, $E_{\mathrm{T}}^{\mathrm{miss},\tau}$

Taus are required to pass the "medium" identification given by the BDT multivariate algorithm where also muon and electron veto is included (see Section 3.2.3) and have p_{T} greater then 20 GeV. They should not overlap with either electrons or photons, in this case the electron/photon is kept and the remnant topoclusters from the overlap removal are moved to the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ term. Taus are seeded by an anti-k_t jet of $R = 0.4$ and the TES is applied only in the core defined by $R = 0.2$. The full four-momentum τ-object, including calibration and an offset pile-up suppression, is used for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ calculation. To avoid energy double counting the topoclusters are associated to the τ up to $R = 0.3$.

Jet term, $E_{\mathrm{T}}^{\mathrm{miss,jet}}$

The official jet algorithm used for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is the anti-k_t with a distance parameter $R = 0.4$. Jets are treated as four-momentum objects including the full JES calibration described in Section 3.2.4 and they are required to pass a p_{T} cut of 20 GeV evaluated at the full calibrated scale. The default $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction makes use of jets with the LCW+JES calibration scheme but if a physics analysis uses jets calibrated with a different scheme it is possible to customize the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ to include the coherent jets.

Jets are also required to not overlap with previous objects selected by the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ algorithm (electrons, photons and taus) for more then 50% of their energy. If there is an overlap larger than 50%, the overlapping object is kept and the remnant topoclusters[6] of the jet are moved to the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ term. If the overlap is smaller than the 50% both the overlapping object and the jet are kept and, in order to avoid energy double counting, the jet momentum is multiplied by a weight accounting for the percentage of the overlap.

Figure 3.10 shows a good agreement between data and MC simulation for the $E_{\mathrm{T}}^{\mathrm{miss,jet}}$ term in $Z \to \mu\mu$ events. The peak at zero is due to events without jets with $p_{\mathrm{T}} > 20$ GeV, the region below 20 GeV is populated by events with two jets balancing each other, and the peak at $E_{\mathrm{T}}^{\mathrm{miss,jet}} \sim 20$ GeV is a threshold effect.

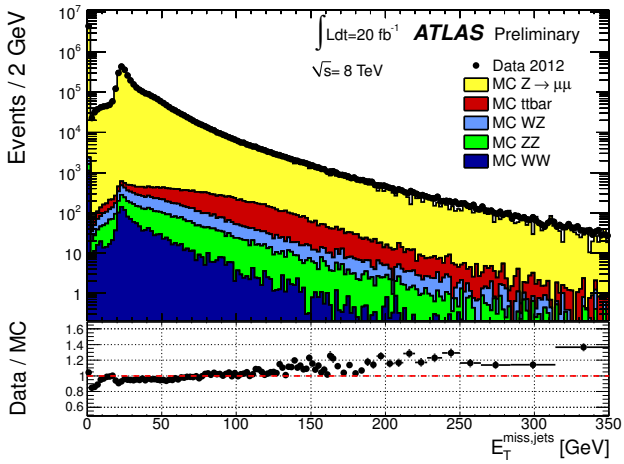


Figure 3.10: Data-MC comparison for the $E_{\mathrm{T}}^{\mathrm{miss,jet}}$ term in $Z \to \mu\mu$ events.

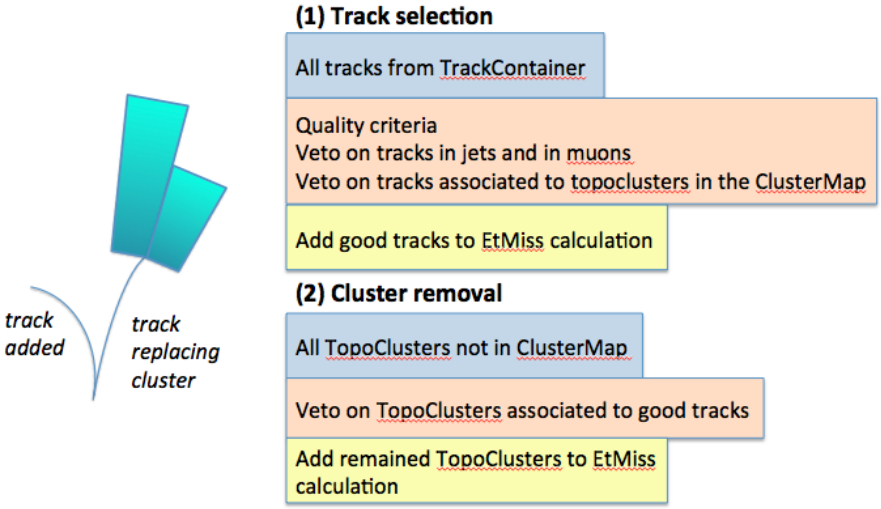


Figure 3.11: Sketch of the *eflow* algorithm.

Soft term, $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$

The soft term is calculated from calorimeter topoclusters and tracks not associated to high-p_{T} objects. Tracks are added to recover the contribution from low-p_{T} particles which do not reach the calorimeter or do not seed a topocluster. The topoclusters are calibrated using the LCW technique. To avoid energy double counting, any overlap between topoclusters and tracks is removed.

Presently also topoclusters with negative energy[7] are used in the soft term computation. Out-of-time pile-up can lead to an increase of the negative energy contribution due to the ATLAS calorimeter signal shape, however, including this negative contribution for the 2012 pile-up condition has a negligible effect (less than 1%) on the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution.

The combination of the track information from the inner detector and the topoclusters is performed by an energy flow (*eflow*) algorithm, sketched in Figure 3.11. The algorithm works in two steps the "track selection" and the "topocluster removal".

Track selection

Reconstructed tracks from any reconstructed vertex with $p_{\mathrm{T}} > 400$ MeV and passing quality selection criteria are used for the calculation of the soft term. A minimum number of hits associated to the reconstructed track in the pixel, SCT and TRT detectors is required.

- Tracks with $p_T > 0.5$ GeV are required to have:
  - $N_{\mathrm{hit}}^{\mathrm{PIXEL}} + N_{\mathrm{hit}}^{\mathrm{SCT}} > 6$
  - $N_{\mathrm{hit}}^{\mathrm{PIXEL}} + N_{\mathrm{hit}}^{\mathrm{SCT}} + N_{\mathrm{hit}}^{\mathrm{TRT}} > 10.$
- To increase the low-p_T acceptance, tracks with $p_T < 0.5$ GeV are required to have:
  - $N_{\mathrm{hit}}^{\mathrm{PIXEL}} + N_{\mathrm{hit}}^{\mathrm{SCT}} > 8$
- For tracks with $p_{\mathrm{T}} > 10$ GeV, the χ^2 probability for the track fit should be:
  - $\chi^2 > 0.01$
- Only tracks with $p_{\mathrm{T}} < 100$ GeV are used to ensure that contribution from high energy objects will not enter in the computation.

[6] Reconstructed jets are larger then reconstructed electrons, photons and taus.
[7] Topoclusters with negative energy are taken at the EM scale.

Track association and topocluster removal

All selected tracks are extrapolated to the second layer of the electromagnetic calorimeter and very conservative criteria are used for the association to reconstructed objects or topoclusters. The association is based on the ratio $\Delta R/\sigma(\Delta R)$. The radial distance ΔR is calculated as:

$$\Delta R = \sqrt{(\eta_{clu} - \eta_{extr})^2 + (\phi_{clu} - \phi_{extr})^2} \tag{3.6}$$

where $\eta_{clu}(\phi_{clu})$ and $\eta_{extr}(\phi_{extr})$ are the topocluster and the track directions at the calorimeter surface respectively and $\sigma(\Delta R)$ is the ΔR resolution parameterized as a function of the track momentum. Tracks are retained if this ratio is greater than 8. Then the p_T resolution, given by the error from the track fit, should be smaller than the expected resolution on the associated cluster.

To avoid energy double counting, the following tracks are vetoed:

- Tracks associated to any high-p_T object used in the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction
- Tracks associated to muons and inside a cone around the reconstructed jets, the dimension of the cone depending on the jet algorithm
- Tracks connected to topoclusters entering the reconstructed objects

Finally the track acceptance and the topocluster removal is performed according to the following criteria:

- If a track is neither associated to a topocluster nor a reconstructed object, its transverse momentum is added to the calculation of the soft term.
- If a topocluster in the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ is associated to a selected tracks, the track momentum is used instead of the topocluster energy for tracks associated to topoclusters, thus exploiting the better calibration and resolution of tracks at low momentum compared to topoclusters.
- If more than one topocluster is associated to a track, only the topocluster with the largest energy in a cone of $\Delta R/\sigma(\Delta R)$ <4 around the selected tracks is excluded from the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ calculation. The remaining topoclusters not associated to tracks are finally added for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ calculation.

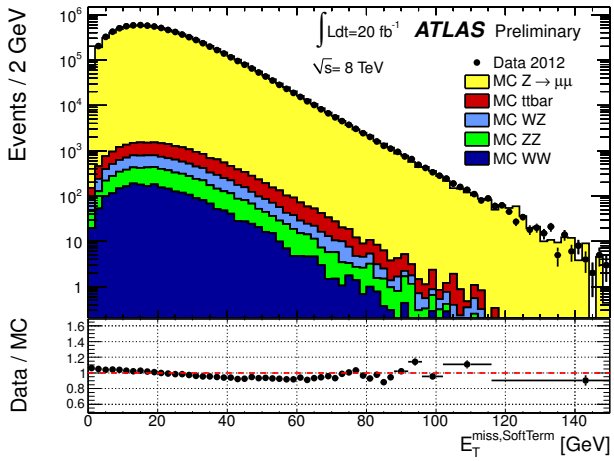


Figure 3.12: Data-MC comparison for the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ term in $Z \to \mu\mu$ events.

Figure 3.12 shows a good agreement between data and MC simulation for the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ in $Z \to \mu\mu$ events. A degradation of the performance and of the efficiency of the algorithm were observed moving to higher pile-up conditions due to the increasing of not-isolated tracks for which selection cuts are not optimal. Therefore, the track selection should be revisited for ATLAS Run 2.

Muon term, $E_{\mathrm{T}}^{\mathrm{miss},\mu}$

The $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ muon term is calculated from the momenta of the reconstructed muons. In order to include in the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ a very well-measured muon contribution, different muon types described in Section 3.2.2 are employed resulting in a rather complex procedure sketched in Figure 3.13.

In the region $|\eta| < 2.5$, mainly combined muons are considered. The matching requirement considerably reduces contributions from fake muons that can be created from high hit multiplicities in the muon spectrometer due to very energetic jets punching through the calorimeter into the muon system. Low-p_{T} muons and muons lost in the small inactive Muon Spectrometer regions can be recovered by the use of tagged muons through the information from inner detector.

In order to properly deal with the energy deposited by the muon in the calorimeters, the muon term is calculated differently for isolated and non-isolated muons[8]:

[8] Muon isolation is defined on the base of the distance $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$: if $\Delta R < 0.3$

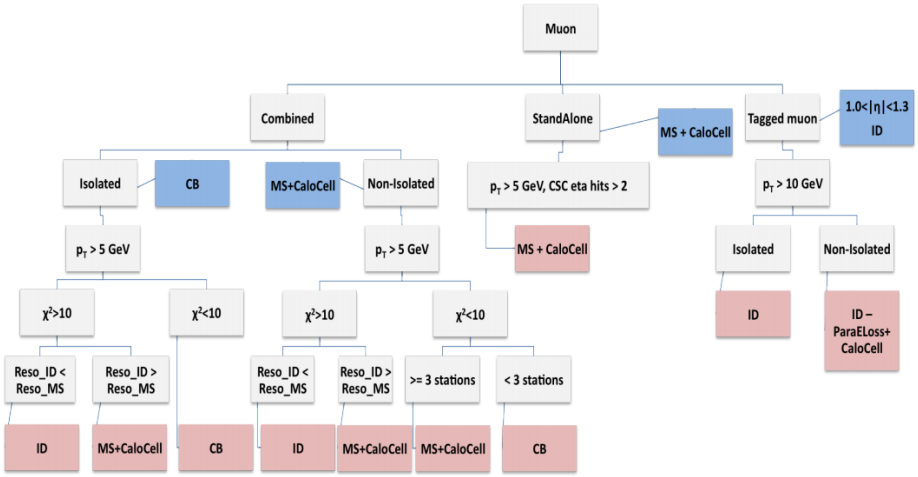


Figure 3.13: Sketch of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ algorithm for the muon term.

Isolated muons: the p_{T} of an isolated muon is determined from the combined measurement of the inner detector and muon spectrometer. The energy lost by the muon in the calorimeter is included in the combined p_{T} so it is not added in the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ computation to avoid double counting of energy. The energy loss in the calorimeter, on average around 2-3 GeV, is not supposed to seed a topocluster, therefore, the energy in the cells contained in a cone around the muon trajectory is used for the computation. Hence this contribution is computed only in those data formats where the cell information in available (ESDs). When the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is performed fro data format contained only the topocluster information (AODs), the parametrized energy loss in calorimeters is subtracted from the p_{T} of the isolated muons entering the $E_{\mathrm{T}}^{\mathrm{miss},\mu}$ term. A reoptimization of the muon energy loss in calorimeter is on-going for Run 2.

Non-isolated muons: for a non-isolated muon, the energy deposited in the calorimeter cannot be resolved from the calorimetric energy depositions of the particles in the jet, so it is already added in the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ computation. Therefore, to avoid energy double counting, the p_{T} measured by the Muon Spectrometer only is used, unless there is a significant mis-match between the spectrometer and the combined measurement. In this case the combined

the muon is classified as non-isolated, otherwise as isolated.

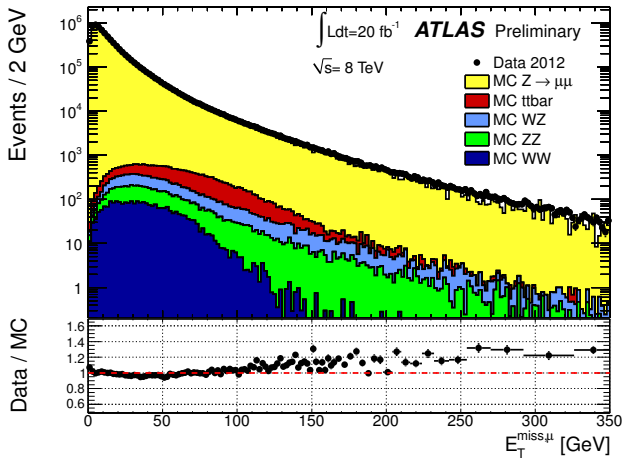


Figure 3.14: Data-MC comparison for the $E_{\mathrm{T}}^{\mathrm{miss},\mu}$ term in $Z \to \mu\mu$ events.

measurement where a parameterized estimation of the muon energy loss in the calorimeter [134] is subtracted is used for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ calculation.

For higher values of pseudorapidity ($2.5 < |\eta| < 2.7$), outside the fiducial volume of the inner detector, there is no matched track requirement and the muon spectrometer p_{T} of standalone muons is used for both isolated and non-isolated muons.

The official $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ configuration uses muons reconstructed starting from the *Staco* muon chain, but it is also possible to customize a case based on the *MuID* muon chain. The performance for the two cases are very similar. The validation of the third muon chain in the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is currently under study.

Figure 3.14 shows a good agreement between data and MC simulation for the $E_{\mathrm{T}}^{\mathrm{miss},\mu}$ term in $Z \to \mu\mu$ events.

3.3.2 Study of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance

The $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance is evaluated in terms of resolution, scale and tails. In order to compare $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance between different event topologies or across different data taking periods, it is useful to define some key-variables to test crucial aspects of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ computation and their impacts on physics analyses. A brief overview of the most used variables and performance plots is reported in the following sections.

Event variables

The total transverse energy in the calorimeters, ΣE_{T}, which includes also the unassociated low-p_{T} tracks used in the soft term but not the track from the muon spectrometer, is an important quantity to parameterize and understand the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance. It is defined as the scalar sum:

$$\sum E_{\mathrm{T}} = \sum E_{\mathrm{T}}^{\mathrm{e}} + \sum E_{\mathrm{T}}^{\gamma} + \sum E_{\mathrm{T}}^{\tau} + \sum E_{\mathrm{T}}^{\mathrm{jets}} + \sum E_{\mathrm{T}}^{\mathrm{SoftTerm}} \tag{3.7}$$

where each contribution is obtained by sum of the transverse energy of the objects reconstructed and calibrated according to the scheme described in Section 3.3.1.

From this quantity the total transverse energy in the event is obtained by summing also the p_{T} of muons:

$$\sum E_{\mathrm{T}}(\mathrm{event}) = \sum E_{\mathrm{T}} + \sum p_{\mathrm{T}}^{\mu} \tag{3.8}$$

This variable is important to have a fair comparison, at the same event activity, between electron and muon channels.

In MC simulation samples, event-by-event comparison between the reconstructed and the "truth" $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ value and direction provide a useful test of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction and calibration performance.

$\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution

The $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolute is expected to depend on the amount of energy measured in the detectors, and in particular on the ΣE_{T} in the calorimeter. The resolution is estimated from the width of the distribution $E_{x,(y)}^{\mathrm{miss}} - E_{x,(y)}^{\mathrm{miss,truth}}$ in bins of the total transverse energy in the event, calculated from Equation 3.8. For minimum bias and $Z \to \ell\ell$ events, where no genuine $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is expected the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution can be estimated also on data events taking the width of the E_x^{miss} and E_y^{miss} distributions. In each ΣE_{T} bin the measure from the two $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ components is combined[9] resulting in two entries for event. The core of each distribution is fitted with a Gaussian over a range spanning twice the expected resolution and the fitted width, σ, is examined as a function of ΣE_{T} (event).

The $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution follows an approximately stochastic behaviour as a function of ΣE_{T}, deviations are expected in the low ΣE_{T} region due to the

[9] Both the $(E_x^{\mathrm{miss}} - E_x^{\mathrm{miss,truth}})$ and $(E_y^{\mathrm{miss}} - E_y^{\mathrm{miss,truth}})$ components are checked to be well centered at zero and with a comparable width. So, no bias is introduced by their combination.

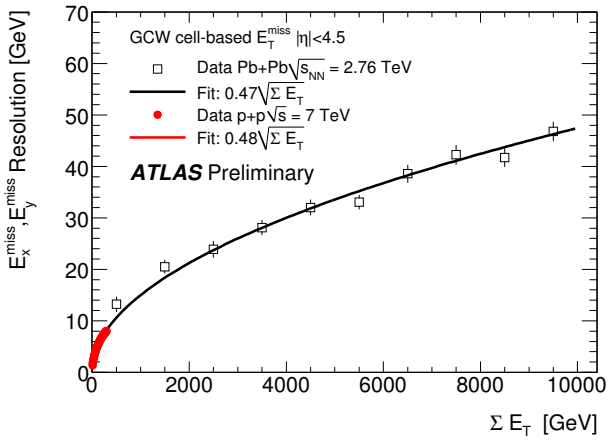


Figure 3.15: $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution curve at the early phase of ATLAS data taking with very low pile-up conditions. Pb-Pb collisions data and p-p collision data are superimposed to show the compatibility between the two measurements.

electronic noise and the presence of pile-up, and in the high ΣE_{T} region where the constant term dominates. The shape of the resolution curve is parametrized and fitted with a good agreement according to the simple function:

$$\sigma = k \cdot \sqrt{\Sigma E_{\mathrm{T}}} \tag{3.9}$$

where the the parameter k quantifies the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution.

In absence of pile-up, this simple law for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution as a function of the ΣE_{T} was firstly tested with p-p collisions and then with Pb-Pb collisions up to the 10 TeV as shown in Figure 3.15 for minimum bias events, demonstrating the excellent behavior of the ATLAS calorimeter. The effects of pile-up on the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution will be largely discussed is Chapter 4.

$\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale: diagnostic plot

A test for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale and bias can be provided for $Z \to \ell\ell$ events exploiting the balance between the leptons from the decaying Z and the hadronic recoil (either jets or soft hadronic contribution) as sketched in Figure 3.16. The importance of this test relies on the possibility to check the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale performance also in data events and not only in MC simulation samples as required by comparisons with the MC "truth" information.

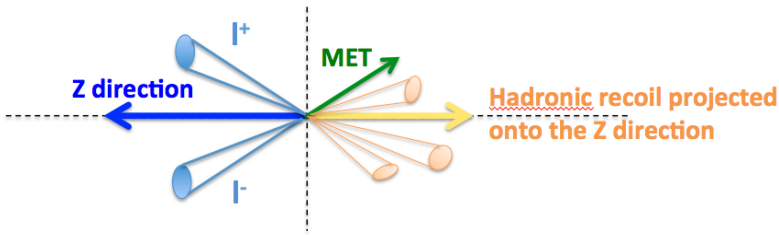


Figure 3.16: Sketch of a $Z \to \ell\ell$ event. The direction of the Z boson in the transverse plane is defined by the momentum of the two leptons. If the system (Z-hadronic recoil) is well-balanced no $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is expected along the Z direction.

The direction of the Z boson in the transverse plance, $\boldsymbol{A_Z}$, can be defined using the momenta of the reconstructed leptons:

$$\boldsymbol{A_Z} = (\ \boldsymbol{p_T}^{\ell^+} + \ \boldsymbol{p_T}^{\ell^\pm})/|\ \boldsymbol{p_T}^{\ell^+} + \ \boldsymbol{p_T}^{\ell^-}|, \tag{3.10}$$

where $\boldsymbol{p_T}^{\ell}$ are the vector transverse momenta of the lepton and anti-lepton.

The distribution of the mean value of the projection of $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ onto the Z direction, $\langle \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}} \cdot \boldsymbol{A_Z} \rangle$, as a function of p_{T}^{Z} is used as a diagnostic plot to validate the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ algorithms, as the distribution should be a straight line through zero if the leptons perfectly balance the hadronic recoil, regardless of the energy of the lepton system. Instead if a negative bias is observed, it suggests either that the lepton system energy is overestimated or the magnitude of the hadronic recoil is underestimated. Since leptons are demonstrated to be well calibrated from the respective combined performance groups, it is interpret as an underestimation of the hadronic recoil.

The effect of the calorimeter coverage up to $|\eta| < 4.9$ gives a small contribution as shown in Figure 3.17. In the same way the increasing of the topocluster noise thresholds for the 2012 run can increase the observed bias in this distribution.

Data and Monte Carlo simulation samples for the study of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance

In order to fully exploit the detector capability in the reconstruction and calibration of different physics objects, several event topologies are explored to test the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction algorithm and its performance.

The minimum bias sample is the first sample where it is possible to test the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance. It is a generic mixture of soft and hard collisions and, apart

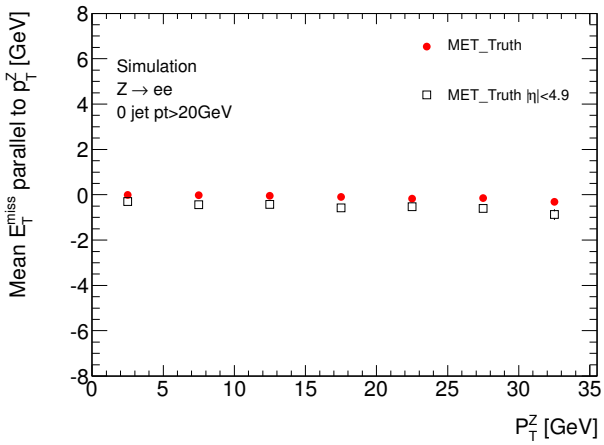


Figure 3.17: $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ projected onto the Z boson direction in MC $Z \to ee$ events. The distribution obtained using the"truth" information from not interacting particles is a straight line as zero as expected, while the one obtained using the"truth" information from all interacting particles inside the ATLAS calorimeter coverage shows a slight negative bias. The effect is considered negligible.

from a small contribution from prompt decays that could be triggered together with the more common QCD events, no genuine $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is expected in these events. Thus the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstructed in these events is a direct result of imperfections in the reconstruction process or in the detector response.

The $Z \to \ell\ell$ channel is well-suited to the study of $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance because of its clean event signature and the relatively large cross-section. In general, apart from a small contribution from the semi-leptonic decay of heavy-flavour hadrons in jets, no genuine $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is expected in these events. Thus, similarly to the minimum bias sample, it can be used to test the detector response. Moreover, compared to minimum bias events, the $Z \to \ell\ell$ events provide an important test for the lepton reconstruction in the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$.

Once the impact of the detector is well understood, it is useful to study the performance of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ measurement also in events containing genuine $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ originating from neutrinos, like in $W \to \ell\nu$ events, where more precise tests on the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale and direction are allowed.

Finally, final states with a dominant presence of jets, taus and photons are also studied in order to have a complete overview of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance.

For all the results reported in this thesis, cleaning criteria are applied to each sample to reduce the impact of instrumental noise and out-of-time energy deposits in the calorimeter from cosmic-rays or beam-induced background. Then, the specific selection criteria required to select the desired topology are also applied on both data and MC simulation samples. They are summarized in Appendix A for the different topologies.

3.3.3 $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ systematics uncertainties

The $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is the sum of several terms corresponding to different types of reconstructed objects. The uncertainty on each individual term is evaluated given the knowledge of the reconstructed objects that are used to build it. The overall systematic uncertainty on the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement is then calculated by combining the uncertainties on each term corresponding to a reconstructed physics object and the uncertainties on the soft term which are discussed in this section.

The relative impact of the uncertainty of the constituent terms on $E_{\mathrm{T}}^{\mathrm{miss}}$ depends on the event topology, i.e. presence of leptons, jet activity, etc. In particular the contribution of the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ is important in Z and W events, while it becomes less important in events with higher jet activity where the $E_{\mathrm{T}}^{\mathrm{miss,jet}}$ term is the dominant contribution.

Different methods for the evaluation of the systematics uncertainties on the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ are developed. In the early phase of the ATLAS data taking (2010-2011 data) a method based on the evaluation and propagation of the cluster energy uncertainties was employed [127]. Since this method leads to largely conservative results, for analyses on 2012 data it was replaced with an evaluation of the systematic uncertainties based on in-situ methods that also include MC modeling and pile-up effects [129**?**]. In particular two methods based on studies of Z events are developed and explained in the following. A brief explanation of the method used for 2010-2011 data is as well reported, highlighting its limitations.

$E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ systematics uncertainty based on the energy cluster uncertainty (for 2010-2011 data)

In an early phase of the ATLAS data taking the systematics uncertainties on the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ were derived combining uncertainties from the detector geometry and MC generator effects, found to be around 3%, with the systematics uncertainties on topoclusters energy derived with an E/p method, that was by far the dominant contribution, of the order of 10%.

This procedure led to very conservative systematics uncertainties for the soft term due to the difficulty to determine the E/p ratio in a busy environment and the overestimation of the cluster energy uncertainty in the forward region where no track information is available.

When moving to the 2011 data, an additional systematic uncertainty accounting for pile-up effects was also needed. At the time it was determined taking the average of the relative discrepancies[10] between the ΣE_{T} distribution in data and MC simulation[11] as a function of the number of primary vertices in the different calorimeter regions and then propagating it to the E_x^{miss} and E_y^{miss} components. The pile-up uncertainty estimated for the 2011 data was 5.6%, reduced to 2.3% after a reoptimization of the procedure.

$E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ systematics uncertainties from the in-situ method based on evaluation from data/MC ratio in $Z \to \mu\mu$ events without jets (for 2011-2012 data)

The systematic uncertainties on both the scale and the resolution of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ soft term are evaluated from the comparison of observables in data with the Monte Carlo prediction for events without jets. In order to isolate the soft term contribution, the subset of $Z \to \mu\mu$ events that do not contain jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ is selected because in these events only the leptons and the soft term contribute to $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$. The muon channel is preferred over the electron channel since on average the muons leave just a small contribution (around 2-3 GeV) in the calorimeter. Nevertheless the results are checked also in $Z \to ee$ events leading to comparable values for the systematics uncertainties.

As discussed in Section 3.3.2, the projection of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ onto the Z boson transverse direction provides a test of the bias on the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale. So, as shown in Figure 3.18 the data-MC ratio of this observable for events without jets is used as a measure of the systematic uncertainty on the scale of the soft term, which is calculated it as the average deviation from unity. A parametrization as a function of ΣE_{T} is chosen to allow an easy extrapolation of the systematic uncertainties also for events that do not contain a Z decay.

The systematic uncertainty on the soft term resolution is determined in a similar manner, using the E_x^{miss} and E_y^{miss} resolution as a function of the ΣE_{T} to quantify the level of data-MC agreement as shown in Figure 3.19.

[10] In this way the effect of the simulation mismodelling was excluded.

[11] effects on MC uncertainties like the μ-scaling were also taken into account.

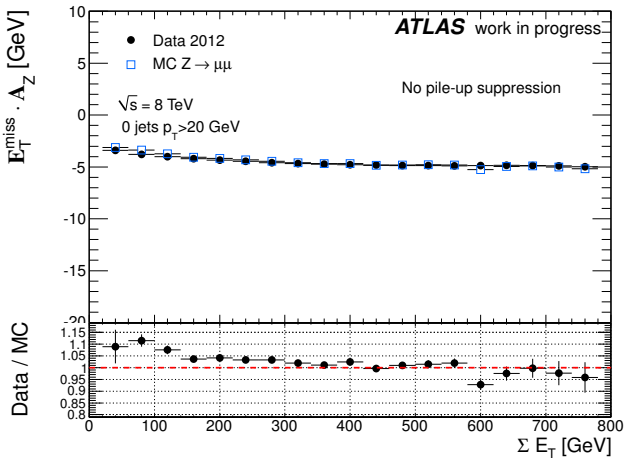


Figure 3.18: Projection of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ onto the Z boson transverse direction as a function of ΣE_{T}, for data and MC simulation $Z \to \mu\mu$ events without jets with $p_{\mathrm{T}} > 20$ GeV. The data-MC ratio used to evaluate the systematics uncertainties is shown in the bottom part of the plot.

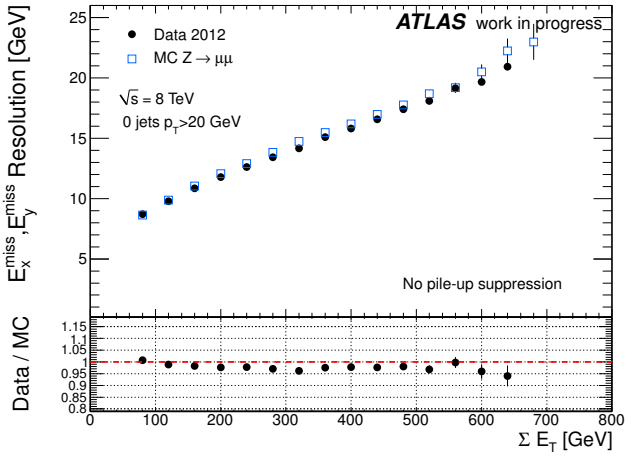


Figure 3.19: $E_{x(y)}^{\mathrm{miss}}$ resolution as a function of ΣE_{T}, for data and MC simulation $Z \to \mu\mu$ events without jets with $p_{\mathrm{T}} > 20$ GeV. The data-MC ratio used to evaluate the systematics uncertainties is shown in the bottom part of the plot.

The N_{PV} and μ dependences of the data-MC ratio for both $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ scale and resolution uncertainties are checked looking at the evolution of the ratio in regions characterized by different pile-up conditions. In order to deal with the limited statistics and to ensure a solid fit procedure, the distributions inclusive in μ are studied to determine the dependence on N_{PV} , and similarly the distributions inclusive in N_{PV} are studied to determine the dependence on μ. The effect of increasing either N_{PV} or μ was determined to be less then the percent level and so negligible, this small dependence is attributed to the almost linear correlation between ΣE_{T} and N_{PV} and μ.

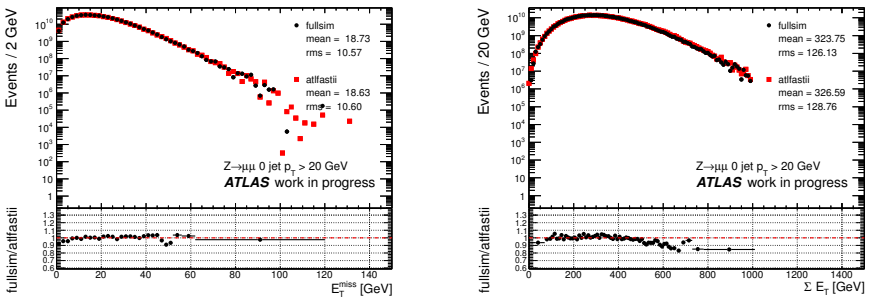
Effects given by possible jet inefficiency reconstruction or either promotion of soft contribution into the $E_{\mathrm{T}}^{\mathrm{miss,jet}}$ term due to the increasing pile-up energy are not examined isolating their specific contribution but their effect is already included in the total value of the uncertainty provided by the data-MC ratio.

The values of the systematics uncertainties evaluated with this method depend on the MC simulation employed. Events simulated with Powheg+Pythia are used to determine the systematic uncertainties on the soft term with this method. Then, it has also been checked that these uncertainties cover the data-MC discrepancies when using $Z \to \mu\mu$ events generated with either Alpgen or Sherpa that are compatible within the 0.5% level. The systematics uncertainties for Atlfast-II (Section 3.1.2) are also evaluated. The results are found to be compatible with the general case because the large pile-up contribution entering the soft term is simulated in the same way in both the Atlfast-II and the full simulation samples resulting in very similar distributions for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ and ΣE_{T} quantities shown in figure 3.20. A large discrepancy up to 20% is only observed in the high ΣE_{T} region that, for the purposes of the systematics uncertainties evaluation, just enters as a small binning effect.

$E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ systematics uncertainties from the in-situ method based on evaluation from the balance between the soft term and the hard objects (for 2011-2012 data)

The method uses inclusive $Z \to \mu\mu$ events and exploits the balance between the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ and the total transverse momentum of the hard objects in the events, defined as:

$$\begin{aligned} p_{x(y)}^{\mathrm{hard}} &= \Sigma_{\mu}\, p_{x(y)}^{\mu} + \Sigma_{\mathrm{e}}\, p_{x(y)}^{\mathrm{e}} + \Sigma_{\mathrm{jets}}\, p_{x(y)}^{\mathrm{jets}} + \Sigma_{\gamma}\, p_{x(y)}^{\gamma} + \Sigma_{\nu}\, p_{x(y)}^{\nu}, \\ p_{\mathrm{T}}^{\mathrm{hard}} &= \sqrt{(p_x^{\mathrm{hard}})^2 + (p_y^{\mathrm{hard}})^2}. \end{aligned} \tag{3.11}$$



(a) (b)

Figure 3.20: Comparison between the full and Atlfast-II simulation for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ and ΣE_{T} distributions for $Z \to \mu\mu$ events without jets with $p_{\mathrm{T}} > 20$ GeV.

$p_{\mathrm{x(y)}}^{\mathrm{hard}}$ is in general calculable only for MC events, since it includes invisible particle momenta which are not known in data. While not an observable, it is nevertheless a useful quantity to characterize events since transverse momentum balance dictates that it ought to be equal to $E_{x,(y)}^{\mathrm{miss,SoftTerm}}$.

The mean and the resolution of the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ components have been studied both with respect to $p_{\mathrm{T}}^{\mathrm{hard}}$ and to N_{PV} to study the effect of pile-up. In these events $p_{x(y)}^{\nu}$ is close to zero and it is assumed to be zero in data. Since the magnitude and direction of the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ depends on the number of jets, leptons and neutrinos in the event, the systematic uncertainties have been derived in bins of $p_{x(y)}^{\mathrm{hard}}$. Therefore the parametrization determined from $Z \to \mu\mu$ events can be used to evaluate the systematic uncertainties on the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ in other samples as well.

To evaluate the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ mean and resolution, the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ is decomposed along the $p_{\mathrm{T}}^{\mathrm{hard}}$ direction and along the orthogonal direction, referred to as longitudinal and perpendicular directions, respectively. The mean longitudinal component is a measure of the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ scale, as the longitudinal direction is sensitive to the balance between the high-p_{T} objects and the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$.

Results and combination of the scale and resolution uncertainties on the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ and closure test

The results for the systematics uncertainties on the scale and resolution of the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$, obtained with the two in-situ methods previously described, are

$E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ uncertainty	data/MC method		balance method		
	scale	resolution	scale		resolution
	(%)	(%)	([GeV])	(%)	(%)
RefFinal	3.6	2.3	< 1 GeV	<13	2.0

Table 3.2: Systematic uncertainties on the scale and on the resolution of the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$, calculated with the two different in-situ methods: based on the data/MC ratio and on the balance between the soft term and the hard objects. The large number quoted in column 4 has to be referred to the small bias of the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ of $\sim$ 3 GeV, therefore its net effect is less then 1 GeV and less important with respect to the resolution uncertainty.

summarized in Table 3.2.

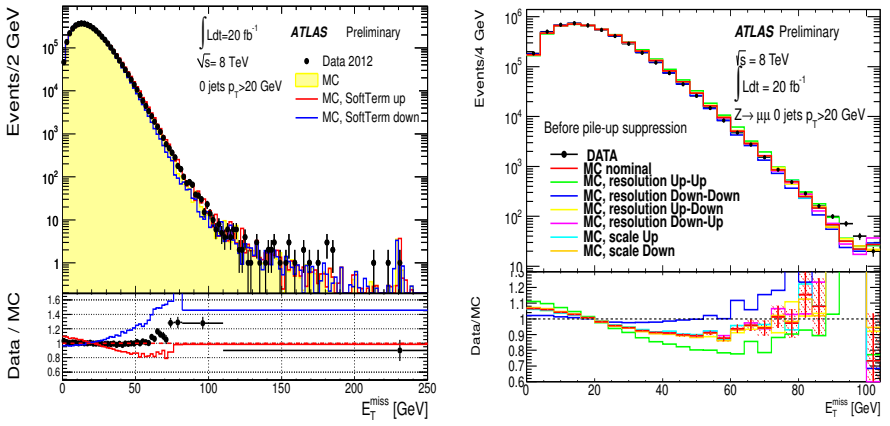
Both the methods are validated performing a closure test in $Z \to \mu\mu$ events without jets with $p_{\mathrm{T}} > 20$ GeV. The results are shown in Figure 3.21(a) for the Poweg+Pythia MC generator using the data-MC ratio method and in Figure 3.21(b) for the Alpgen generator using the balance method. The variation distribution are obtained scaling the soft term up and down according to its scale uncertainty and smearing it according to its resolution uncertainty.

For the first method the scale and resolution uncertainties are assumed uncorrelated and must be both scaled up and down and then added in quadrature. For the second method both longitudinal and perpendicular components of the resolution uncertainties must be varied and the correlated and anticorrelated cases uncertainties must be added in quadrature.

The first thing to notice is that while for the Powheg+Pythia generator the nominal $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ value is quite centered around the unity for the Alpgen generator the ratio distribution shows a systematic trend suggesting that the MC $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ distribution is larger then the one observed in data. This observation is also confirmed by the fact that the down variation for the resolution gives a data-MC ratio around unity. Anyway both methods are able to cover the deviation of data-MC ratio from unity in the full range.

Prospects for the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ systematics uncertainties evaluation

In order to further improve the evaluation of the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ systematics uncertainties, a lot of work is currently on-going on different fronts.



(a) (b)

Figure 3.21: Comparison between data, nominal and variated MC simulation for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ distribution in $Z \to \mu\mu$ events without jets with $p_{\mathrm{T}} > 20$ GeV for the Powheg+Pythia MC generator for the first method in (a) and for the second method in (b). In the bottom part of the plots are shown the respective ratio between data and the nominal and the variated MC.

A more precise evaluation can be achieved with a split of the resolution and scale $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ uncertainties into components originated by different physics sources. This would allow each physics analysis to constrain more the specific components of the systematics uncertainties to which the analysis is more sensitive, and it would also give more flexibility in the combination of different physics analyses. In order to perform a complete splitting of the systematics uncertainties a common metric scale not depending on pile-up effects and on the specific topology examined would be very useful, hence there is a large activity in this direction.

The splitting of the systematics uncertainties in physics sources could also help in the evaluation of the correlation between the different terms of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$. In particular, a correlation between the jets and the soft term is expected, since only a p_{T} threshold establishes the separation between a jet and a soft term contribution. Using a splitting scheme for the soft term uncertainties might give the opportunity to identify a relation between the components of the systematics uncertainties of different physics objects. These studies should also include possible migration effects between the different $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ terms and the evaluation of

"close-by" effects between objects of different type, like electrons and jets, that could not negligibily affect boosted topologies[12].

Finally the systematics uncertainty on the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale might benefit of the use of an in-situ evaluation of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale using $W \to \ell\nu$ events [127] and $Z \to \tau\tau$ events, so far never explored due to the limited statistics after the hard selection required to highly reject the background in this analysis.

3.4 Conclusion and prospects

Since the beginning of the data taking a lot of work was done for a good understanding of the detector system and the development of high precision reconstruction algorithm exploiting information from the full detector. High efficient and performant physics object reconstruction, identification and calibration and an accurate evaluation of their systematics uncertainties have a direct impact on physics analyses allowing to achieve high precision measurements and searches for rare events.

In particular, the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is a complex and refined procedure that relies on the reconstruction and calibration performance of all the other physics objects. A careful work was done to improve the treatment of each contribution and to guarantee a good flexibility to have coherence between the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ and the analysis selection choices.

The use of more sophisticated and refined techniques is required to cope with the increasing of pile-up condition resulting in a very busy environment. The case of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ algorithm will be extensively discussed in the next chapter.

[12] In these final states the decay products are close to each other.

Chapter 4

Pile-up suppression methods for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction

As discussed in Section 2.2, the high luminosity provided by LHC (with an instantaneous luminosity peak close to $8 \cdot 10^{33}$ cm^{-2} s^{-1}) combined with a bunch crossings time of 50 ns, lead to unprecedented backgrounds from additional proton-proton collisions.

The $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is affected not only by the multiple interactions in the same bunch crossing (in-time pile-up), but also by the bunch crossing signal history in the ATLAS calorimeters (out-of-time pile-up). Several approaches to suppress the pile-up signals, especially for the soft term contribution, have been developed and evaluated, making use of reconstructed tracks, calorimeter signals, or a combination of both.

The impact of pile-up in the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is discussed in Section 4.1. The pile-up suppression methods are described in Sections 4.2-4.4. The performance improvements for various final states are evaluated and compared in Section 4.5. Finally, the evaluation of the systematic uncertainties after applying the corrections are presented in Section 4.6.

4.1 Pile-up effects in $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance

The large pile-up in the 2012 data taking has significant effects on the detector signals relevant for $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, in particular on the calorimeter energy

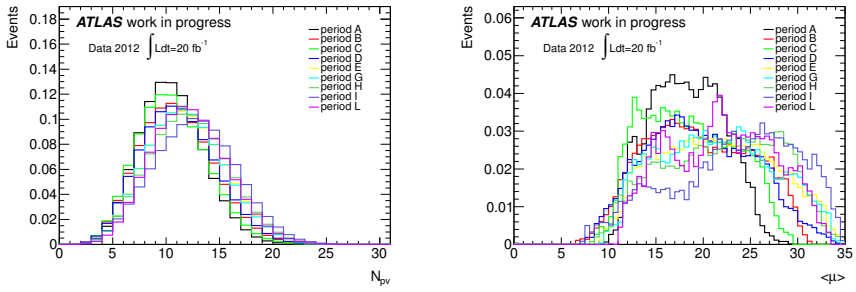
clusters: energy fluctuations in these observables directly translate in fluctuation in the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ measurement not only worsening the resolution but also affecting the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale.

In order to mitigate the effects from the survival of pile-up contributions in the clustering formation or either in creating additional clusters, an accurate optimization of the LAr optimal filtering, and a re-evaluation of the noise topocluster thresholds and of the LCW weights are provided for the processing of 2012 data. Yet, a large pile-up contribution is still included in the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction.

The degradation in performance is quantified studying the stability of the basic $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction observables as a function of the number of the reconstructed primary vertices, N_{PV}, that is a good estimator for the in-time pile-up, and as a function of the average number of collisions in a given time window around the recorded event, μ, that is sensitive to the effects of the out-of-time pile-up. The evolution of the pile-up conditions described by the N_{PV} and μ observables during the 2012 data taking is shown in Figure 4.1. The effects of pile-up for the various ΣE_{T} terms and for the total ΣE_{T} in different pseudorapidity calorimeter regions are respectively shown in Figure 4.2(a) and 4.2(b) as a function of N_{PV}. Figure 4.2(a) shows that the soft term is largely dominated by pile-up, therefore, a dedicated correction is needed, especially in low multiplicity final states, such as Z, W and Higgs events, where the importance of the soft term contribution to the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is enhanced. A simple omission of this term is not the right approach because the true soft event associated with the triggered hard scattering signal gives an important contribution to $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction performance especially with respect to the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution, and depending on the final state topology, also to $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale. In the following sections, techniques for the pile-up suppression in the soft term are described in detail. It is important to notice that a correction for the soft term is not straight forward because the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ is reconstructed from tracks and calorimeter clusters not associated with the hard objects, and thus lacking a universal and stable calibration reference especially for the individually calorimeter cluster signals.

4.2 Pile-up suppression methods

As discussed in the previous section the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is hardly challenged by the presence of pile-up, especially the soft term that receives a huge additional energy contribution. This pile-up contribution needs to be removed from the



(a) N_{PV} (b) μ

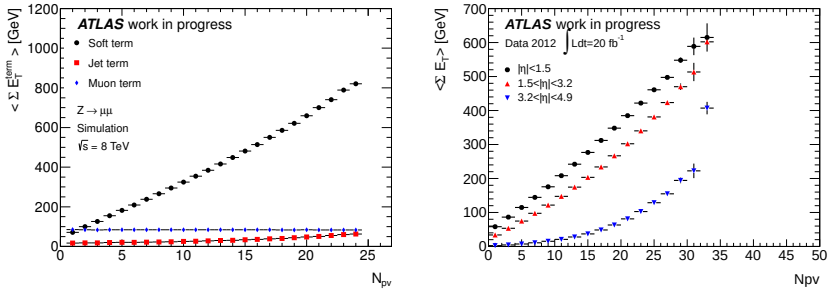
Figure 4.1: The N_{PV} and μ distribution are shown respectively in (a) and (b) as a function of the ATLAS data taking period in 2012.

true signals as much as possible in order to restore comparable reconstruction performances to the ones achieved in the low luminosity running periods of the ATLAS data taking. It is important to notice that, in order to not create a fake unbalanced in the $E_{\mathrm{T}}^{\mathrm{miss}}$ computation, the pile-up contribution must be removed from all the $E_{\mathrm{T}}^{\mathrm{miss}}$ terms in Equation 3.5. Therefore, the pile-up corrections for each physics object, described in the previous chapter, are automatically included in the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction.

Concerning jets, since the pile-up not only affects the jet energy reconstruction but can also create additional jets and since the jet area correction described in Section 3.2.4 only captures event-by-event fluctuations and not local fluctuations in the same event, an additional track based filter (JVF) is applied in order to enhance the likelihood for a particular jet to be generated by the hard scattering vertex. A description of the JVF filter is given in Section 4.3.1.

In this section methods for the pile-up suppression for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ soft term are presented. They are mainly divided into two categories: track based methods and jet area based methods. The main concepts on which these methods are based are described in the following while a full description of each method is reported in Sections 4.3-4.4.

Tracking based pile-up correction (STVF, TST). These methods exploit the tracks association with the primary vertex to exclude from the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ computation pile-up contributions. Moreover the tracker detector, having a faster response then the calorimeter, is not sensitive to the out-of-time pile-up contribution making tracks a powerful tool for pile-up suppression.



(a) ΣE_{T} terms (b) ΣE_{T} in η regions

Figure 4.2: The dependence on pile-up of the ΣE_{T} terms and of the ΣE_{T} distribution in different pseudorapidity regions as a function of N_{PV} are shown respectively in (a) and (b).

Two methods are explored in this thesis and are described in details in Section 4.3:

STVF The Soft Term Vertex Fraction (STVF) method makes use of both tracks and clusters entering the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$, and employs the ratio of the scalar sum of the soft event track p_{T} associated with the hard scatter vertex to the sum of all the soft event track p_{T} from all reconstructed vertices in the event. This ratio is used to scale all soft event contributions to $E_{\mathrm{T}}^{\mathrm{miss}}$ and ΣE_{T} in the given event.

TST The Track-based Soft Term (TST) method uses only tracks not included in high-p_{T} physics objects and associated to the primary vertex. No cluster information is used neglecting the contribution of neutral and forward particles soft enough to enter the soft term.

Jet area based pile-up suppression (EJA, EJAF, JAF). The common aspect of these methods is the use of an event-by-event estimator for the transverse momentum density, ρ, of the soft event, which is then used to compute the pile-up contribution in each jet, $\rho \times A_{\mathrm{jet}}$, where A_{jet} is the jet area. The procedure involves the decomposition of the soft event into soft jets, down to $p_{\mathrm{T}} = 0$, with typically two different definitions for these jets: one for the measurement of the transverse momentum density ("ρ-jets"), and another one as a basis for applying the p_{T} threshold ("filter-jets"). Three implementations are studied in this thesis:

EJA The Extrapolated Jet Area (EJA) method measures the ρ using the soft event in the central part of ATLAS ($|\eta| < 2$, approximately) only. It is then extrapolated to the forward region using transverse momentum flow profiles measured with minimum bias data.

EJAF The Extrapolated Jet Area with Filter method (EJAF) uses a similar configuration as EJA to measure the ρ, including the extrapolation. It then applies a JVF based selection on filter-jets in addition.

JAF The Jet Area with Filter (JAF) method uses a ρ calculated from the soft event over the full acceptance of ATLAS ($|\eta| < 5$), without extrapolation. It also applies a JVF based selection on the filter-jets.

4.3 Track-based methods

As discussed in the previous section, the tracking based pile-up correction exploits the p_{T} activity from charged tracks not associated with physics objects but linked to a given hard scatter primary vertex, hereafter indicated with the symbol V_{primary}. The track reconstruction is as well challenged by the increasing of pile-up conditions, in particular the effects of splitting and merging between vertices has a large impact on the vertex reconstruction efficiency.

In this section the computation of the JVF filter and the two track based methods for the pile-up suppression in the soft term, STVF and TST, are described in detail.

4.3.1 Jet Vertex Fraction (JVF) filter

Jets entering the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ computation still have a pile-up contamination, improvements in the performance are achieved by requesting an association of jets contributing to $E_{\mathrm{T}}^{\mathrm{miss,jet}}$ term to the hard scatter vertex. For $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, central jets not even weakly associated with the primary event vertex can be safely interpreted as originating from one of the additional pile-up interactions and should therefore be omitted from $E_{\mathrm{T}}^{\mathrm{miss}}$.

The filter applied to the accepted jets is based on the jet vertex fraction JVF, which measures the amount of p_{T} carried by reconstructed tracks associated[1] with the jet and coming from the primary vertex V_{primary} relative to the p_{T}

[1] Tracks are associated to calorimeter jets following the ghost association procedure.

carried by all tracks associated with the jet:

$$\mathrm{JVF} = \frac{\sum_{i=1}^{N_{\mathrm{trk}}^{\mathrm{jet}}(V_{\mathrm{primary}})} p_{\mathrm{T,trk},i}^{\mathrm{jet}}(V_{\mathrm{primary}})}{\sum_{k=1}^{N_{\mathrm{vtx}}} \sum_{i=1}^{N_{\mathrm{trk}}^{\mathrm{jet}}(V_k)} p_{\mathrm{T,trk},i}^{\mathrm{jet}}(V_k)} \tag{4.1}$$

where $N_{\mathrm{trk}}^{\mathrm{jet}}(V_k)$ is the number of tracks at vertex V_k pointing to the jet, and N_{vtx} is the total number of collision vertices in the event. $p_{\mathrm{T,trk},i}^{\mathrm{jet}}(V_k)$ is the p_{T} of track i associated with vertex V_k and pointing to the jet.

JVF is a quantity assigned jet-by-jet. It can only be calculated for jets within the ID acceptance ($|\eta_{\mathrm{jet}}| < 2.4$), and for jets with tracks associated at all:

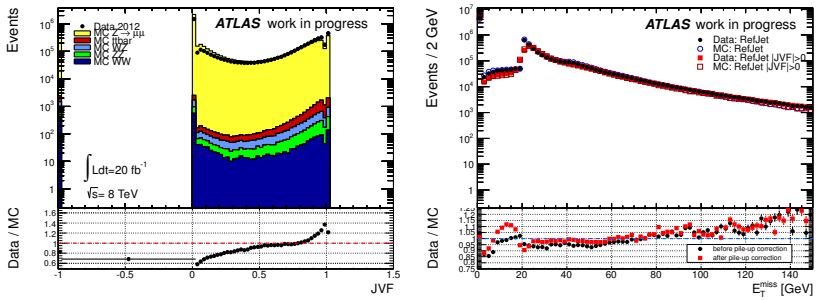
$$\mathrm{JVF} = \begin{cases} -1 & \text{no tracks associated with jet} \\ 0 \ldots 1 & \text{all central jets with tracks} \end{cases}$$

In particular, JVF = 0 corresponds to the case in which the jet has no tracks associated to the V_{primary}.

Jets within $|\eta_{\mathrm{jet}}| < 2.4$ and with $p_{\mathrm{T}} < 50$ GeV are accepted for $E_{\mathrm{T}}^{\mathrm{miss}}$ only if JVF $\neq 0$ (weak association with V_{primary}). The selection applied is efficient for letting hard scattering jets survive, with a significant rejection of pile-up jets, see Ref. [135] for details.

The JVF distribution, shown in Figure 4.3(a), is not well modelled specially for low values, therefore, cutting on this variable, a discrepancy in the jet selection efficiency for $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ can be introduced. This effect can be observed in the comparison between data and MC simulation for the $E_{\mathrm{T}}^{\mathrm{miss,jet}}$ term distribution before applying the JVF cut and after the cut, shown in Figure 4.3(b). The region below 20 GeV is dominated by jets balancing each other and a clear disagreement is introduced after applying the JVF cut since for the MC simulation is more likely[2] to filter out one of the jets reducing the number of the entries in this region with respect to the ones observed in data. Instead in the region defined by $E_{\mathrm{T}}^{\mathrm{miss,jet}} > 20$ GeV, a better agreement is observed after the application of the JVF cut probably due to an overestimation in the number of jets in the MC simulation before the JVF pileup suppression. This hypothesis is also confirmed by the improvement in the data-MC agreement for the jet multiplicity studied before and after the JVF filter [135].

[2]From the JVF plot one can observe that MC simulation has more jets with JVF=0 with respect to the data.



(a) JVF (b) $E_{\mathrm{T}}^{\mathrm{miss,jet}}$

Figure 4.3: The comparison between data and MC simulation in $Z \to \mu\mu$ events is shown for the JVF distribution in (a) and for the $E_{\mathrm{T}}^{\mathrm{miss,jet}}$ distribution before and after applying the JVF filter cut in (b).

4.3.2 Soft Term Vertex Fraction (STVF)

The STVF method provides an event-by-event average correction to be applied to both the tracks and the topoclusters contributing to the soft term as described in Section 3.3.1. In particular, it relies on the evaluation of the pile-up activity within an event via the Soft Term Vertex Fraction (STVF), defined as the ratio of the scalar p_{T} sum of tracks from the primary (hard scatter) event vertex V_{primary} to the total summed reconstructed track p_{T} from all $k = 1 \ldots N_{\mathrm{vtx}}$ event vertices V_k:

$$\mathrm{STVF} = \frac{\sum_{i=1}^{N_{\mathrm{trk}}(V_{\mathrm{primary}})} p_{\mathrm{T},i}^{\mathrm{trk}}(V_{\mathrm{primary}})}{\sum_{k=1}^{N_{\mathrm{vtx}}} \sum_{i=1}^{N_{\mathrm{trk}}(V_k)} p_{\mathrm{T},i}^{\mathrm{trk}}(V_k)} \tag{4.2}$$

where $p_{\mathrm{T},i}^{\mathrm{trk}}(V_k)$ is the p_{T} of soft event track i coming from vertex V_k, $N_{\mathrm{trk}}(V_k)$ is the total number of reconstructed tracks not associated with any hard object from this vertex, and N_{vtx} is the total number of the collision vertices in the event. The tracks used are all the reconstructed tracks in the soft event and the track selection is the same used by the *eflow* algorithm introduced in Section 3.3.1.

The pile-up suppression is then applied by scaling the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ compo-

nent and the $\sum E_{\mathrm{T}}^{\mathrm{SoftTerm}}$ by the STVF factor, with $0 \leq \mathrm{STVF} \leq 1$:

$$\begin{aligned} E_{x(y),\mathrm{corr}}^{\mathrm{miss,SoftTerrm}} &= \mathrm{STVF} \cdot E_{x(y)}^{\mathrm{miss,SoftTerm}} \\ E_{\mathrm{T,corr}}^{\mathrm{miss,SoftTerrm}} &= \mathrm{STVF} \cdot E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}} \\ \Sigma E_{\mathrm{T,corr}}^{\mathrm{SoftTerrm}} &= \mathrm{STVF} \cdot \Sigma E_{\mathrm{T}}^{\mathrm{SoftTerrm}} \end{aligned}$$

Applying the STVF factor on the total $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ and ΣE_{T} implicitly stands on two strong assumptions. The first is that the fraction of the charged pile-up component is the same as the fraction of the neutral pile-up component and the second is that the pile-up estimation in the central region may be used also in the forward region. Attempts applying the STVF factor only in the central region where the track information is available lead to discouraging results since the contribution in the forward region are then included in the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ computation without any correction bringing to a large pile-up dependence.

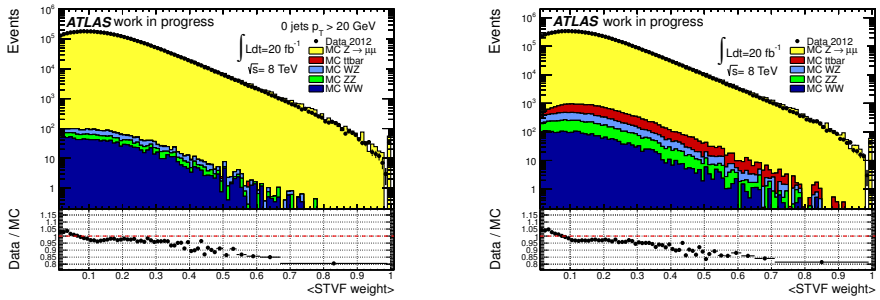
The corrected $E_{\mathrm{T}}^{\mathrm{miss}}$ components, $E_{x(y),\mathrm{corr}}^{\mathrm{miss,SoftTerrm}}$, are then combined following Equation 3.5, together with a $E_{x(y)^{\mathrm{miss,jet}}}$ term using JVF filtered jets, to finally calculate the total corrected $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$.

Figure 4.4 shows the STVF distribution in $Z \to \mu\mu$ events without and with hard jets ($p_{\mathrm{T}} >$ 20 GeV). The agreement between data and MC is good in the regions of low STVF, but deteriorates at larger STVF. This is due to more significant effects from mis-modeling of the track activity in MC for small N_{PV} values (large STVF) While for large N_{PV} (small STVF) the overall track distribution in space and the overall track p_{T} spectrum agrees more with data (due to mixing of many individually simulated interactions), the basic differences between simulated and measured track distributions are more enhanced for individual vertices (p-p interactions).

4.3.3 Track-based Soft Term (TST)

This method completely neglects the cluster contribution and computes the soft term using only tracks associated with the hard scatter vertex. The soft term so built is expected to be resilient to pile-up effects with limitation arising from the omission of the soft neutral and forward particle contributions. It is then combined with the contribution from the high-p_{T} physics objects properly calibrated according to Equation 3.5.

Former studies [128], showed a worse performance for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstructed with a soft term calculated from tracks associated to the primary vertex and selected as for the *eflow* algorithm (see Section 3.3.1). The TST method is now



(a) No jets with $p_{\mathrm{T}} > 20$ GeV

(b) Inclusive final states

Figure 4.4: Comparison of the Soft Term Vertex Fraction, STVF, as defined in Equation 4.2, for a $Z \to \mu\mu$ sample without any jets with $p_{\mathrm{T}} > 20$ GeV in(a), and for the inclusive sample from the same final state in (b).

optimized to use the same track selection employed for the track based $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ estimation [136].

- The tracks are required to have:
  - $p_{\mathrm{T}} > 500$ MeV, $|\eta| < 2.5$
  - $N_{\mathrm{hits}}^{\mathrm{PIXEL}} \geq 1$, $N_{\mathrm{hits}}^{\mathrm{SCT}} \geq 6$
- The association with the primary vertex is obtained asking for:
  - transverse impact parameter with respect to the primary vertex $|d_0| < 1.5$ mm
  - longitudinal impact parameters with respect to the primary vertex $|z_0 \sin(\theta)| < 1.5$ mm.
- Finally, in order to reduce the number of mis-reconstructed tracks[3], isolated tracks, excluding muon tracks, with $p_{\mathrm{T}} > 200$ GeV and $|\eta| < 1.5$, and with $p_{\mathrm{T}} > 120$ GeV and $|\eta| > 1.5$ are used if:
  - the relative uncertainty on the charge to track-momentum ratio (q/p) $\frac{\sigma(q/p)}{(q/p)} < 0.4$

[3]Tracks can have their momentum badly mis-reconstructed due to low p_T tracks that interacting with the ID material may produce a large number of secondary particles leaving enough hits in the pixel/SCT subdetectors to be reconstructed with a much higher momentum.

– the energy in calorimeter clusters in a cone of radius 0.1, in the (η, ϕ) plane, around the track should reflect the reconstructed track momentum $E_{\mathrm{T}}^{0.1}/p_{\mathrm{T}}^{\mathrm{track}} > 0.65$

4.4 Jet Area based methods

The track based pile-up suppression methods discussed in the previous Section 4.3 have the advantage of using well reconstructed tracks from the central detector region to correct $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$, that can rely on the correlation between the central and forward transverse momentum flow, and the identification of the actual hard scattering vertex in a given event. To use more direct measures of the p_{T} flow across the full ATLAS detector acceptance ($|\eta| < 5$), an alternative approach largely exploiting calorimeter signals from the soft event has been developed. It is based on the p_{T} density of the soft event, in a variation of the originally suggested pile-up suppression strategy in Ref. [137], and its application to jets in ATLAS discussed in Ref. [135].

The common procedure for the jet area techniques described in the following sections is based on three steps:

1. The determination event-by-event of the event transverse momentum density, $\rho_{\mathrm{evt}}^{\mathrm{med}}$, described in Section 4.4.1.

2. The computation jet-by-jet of the jet's susceptibility to pile-up, A_{jet}, and the subtraction of pile-up contribution $\rho_{\mathrm{evt}}^{\mathrm{med}} \times A_{\mathrm{jet}}$, described in Section 4.4.2.

3. An additional and optional JVF filter on the previously corrected soft term jets.

4.4.1 Determination of the transverse momentum density

The transverse momentum density, $\rho_{\mathrm{evt}}^{\mathrm{med}}$, is designed to capture the event-by-event fluctuations in pile-up.

In the original suggestion, all particles within the full detector acceptance are clustered into jets using a recursive recombination algorithm like the original k_t [119, 138] or the Cambridge-Aachen [118, 139] flavoured version, both with small ($R = 0.4$) distance parameters. All jets with $p_{\mathrm{T}}^{\mathrm{jet}} \geq 0$ are formed and

their catchment (active) area A_{jet} [140] is calculated[4]. This allows to measure a transverse momentum density $\rho_{\mathrm{jet},i}$ for any soft event jet i with $p_{\mathrm{T},i}^{\mathrm{jet}}$ and $A_{\mathrm{jet},i}$, and to determine a median p_{T} density from the soft event ($\rho_{\mathrm{evt}}^{\mathrm{med}}$) from all N_{jets} soft event jets within a given range $\eta_{\mathrm{min}} < \eta_{\mathrm{jet}} < \eta_{\mathrm{max}}$:

$$\rho_{\mathrm{jet},i} = \frac{p_{\mathrm{T},i}^{\mathrm{jet}}}{A_{\mathrm{jet},i}} \quad \text{and} \quad \rho_{\mathrm{evt}}^{\mathrm{med}} = \mathrm{median}\,\{\rho_{\mathrm{jet},i}\} \\ \text{for } i = 1 \ldots N_{\mathrm{jets}} \text{ in } \eta_{\mathrm{min}} < \eta_{\mathrm{jet}} < \eta_{\mathrm{max}} \tag{4.3}$$

The evaluation range $[\eta_{\mathrm{min}}, \eta_{\mathrm{max}}]$ for $\rho_{\mathrm{evt}}^{\mathrm{med}}$ can be the full detector acceptance or, as for the correction discussed below, any restricted region of sufficient size. The specific use of the median[5] $\rho_{\mathrm{evt}}^{\mathrm{med}}$ of all ρ_{jet} in any given η region emphasizes the contribution of the soft event signals to the event p_{T} density, which is most sensitive to pile-up.

Studies of pile-up suppression for jets in ATLAS found that $\rho_{\mathrm{evt}}^{\mathrm{med}}$ is an appropriate estimator of the in-time pile-up activity, especially if determined in the central detector region only (about $|\eta| < 2$) [135], but it also has some sensitivity to the out-of-time pile-up contribution. According to this indication, two of the methods studied in this thesis evaluate the $\rho_{\mathrm{evt}}^{\mathrm{med}}$ in the central detector region and then extrapolate it to the forward region using transverse momentum flow profiles measured with minimum bias data, as described in the following section, while a third method employs a $\rho_{\mathrm{evt}}^{\mathrm{med}}$ estimated in the whole detector acceptance ($|\eta| < 4.9$). Including the whole event plane in (η, ϕ) into the $\rho_{\mathrm{evt}}^{\mathrm{med}}$ reconstruction yields a smaller estimate of the pile-up activity than the one obtained from the central detectors. This is due to the particularities of the ATLAS calorimeter and to the reconstruction of its (cluster) signal. The readout granularity in the more forward regions of the calorimeter system is significantly reduced, leading to a more sparsely populated event plane even at the level of calorimeter cells. Applying the topological cell clustering and its implicit noise suppression, which is necessary to reduce local signal fluctuations to acceptable levels, leads to even more sparser spatial occupancy, as cell signals are typically collected into only a few (η, ϕ) barycenters. Analyzing the event plane with e.g. k_t jets with $p_{\mathrm{T}}^{\mathrm{jet}} \geq 0$

[4] Note that jets with $p_{\mathrm{T}} = 0$ are not actually clustered, rather they reflect the unclustered area in the rapidity/azimuth plane ($\Delta y \times \Delta\phi \approx 10 \cdot 2\pi$ for an approximate y-range of $|y| < 5$ in ATLAS) after all jets with $p_{\mathrm{T}} > 0$ are removed. The number of $p_{\mathrm{T}} = 0$ jets is then this unclustered area divided by the most probable expected active jet area $A_{\mathrm{jet}}(p_{\mathrm{T}} = 0) \approx \pi R^2/2$ for the k_t algorithm, if no particles are present (*active ghost area*, see Ref. [140]).

[5] A median evaluation is indeed less biased then an average evaluation by the few hard scatter contributions entering the ρ_{evt} computation.

leads to a larger number of $p_{\mathrm{T}}^{\mathrm{jet}} = 0$ jets, thus reducing the median transverse momentum density of the event significantly. While some drop of the transverse momentum density is expected with increasing η [141], the observed drop of the local density is much steeper due to the instrumental effects discussed above. This can be partly mitigated by e.g. increasing the jet distance parameter to $R = 0.8$, and thus decreasing the number of $p_{\mathrm{T}} = 0$ jets.

Extrapolation of the transverse momentum density into the forward regions

The jet area based pile-up corrections implemented for $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ using $\rho_{\mathrm{evt}}^{\mathrm{med}}$ employ filter-jets with $R = 0.4$, built from the soft event tracks and calorimeter clusters with the k_t algorithm implemented in FastJet [141]. The contribution $p_{\mathrm{T}}^{\mathrm{SoftJet}}$ of these *soft event jets* (with transverse momentum $p_{\mathrm{T}}^{\mathrm{jet}}$, area A_{jet}, and at direction η_{jet}) is defined by the following filter:

$$E_{x(y)}^{\mathrm{miss,SoftTerm}} = -\sum_{i=1}^{N_{\mathrm{jets}}} p_{x(y),i}^{\mathrm{SoftJet}} \text{ ,with}$$

$$p_{\mathrm{T},i}^{\mathrm{SoftJet}} = \begin{cases} 0 & p_{\mathrm{T},i}^{\mathrm{SoftJet}} < f_{\mathrm{scale}} \cdot \rho_{\mathrm{evt}}^{\mathrm{med}}(\eta_{\mathrm{jet},i}) \cdot A_{\mathrm{jet},i} \\ p_{\mathrm{T},i}^{\mathrm{SoftJet}} - \rho_{\mathrm{evt}}^{\mathrm{med}}(\eta_{\mathrm{jet},i}) \cdot A_{\mathrm{jet},i} & p_{\mathrm{T},i}^{\mathrm{SoftJet}} \geq f_{\mathrm{scale}} \cdot \rho_{\mathrm{evt}}^{\mathrm{med}}(\eta_{\mathrm{jet},i}) \cdot A_{\mathrm{jet},i} \end{cases} \tag{4.4}$$

The scale factor f_{scale} can be optimized, but it has been found that $f_{\mathrm{scale}} = 1$ delivers good performance. The median transverse momentum density $\rho_{\mathrm{evt}}^{\mathrm{med}}(\eta_{\mathrm{jet}})$ in this case is determined from ρ-jets, with various configurations as described in Section 4.4.2. To avoid the already discussed occupancy issues, it is determined event by event in the central detector region only (typically $|\eta| < 1.8 - 2.0$), and then extrapolated to higher η. The extrapolation function is measured with minimum bias events using a sliding window of total width $\Delta\eta = 1.6$ such that the mean $\langle p_{\mathrm{T}} \rangle(\eta)$ at any direction $\eta \in [-5, 5]$ is the average of the event by event summed p_{T} from calorimeter signal clusters reconstructed using the local hadronic calibration, and located within this window $(\eta - \Delta\eta/2, \eta + \Delta\eta/2)$.

The average amount of energy scattered into any $\Delta\eta$ window at a given direction η by these minimum bias events depends on the in-time and out-of-time pile-up. Thus the reconstructed $\langle p_{\mathrm{T}} \rangle(\eta)$, which is exposed to these influences, depends on the experimental observables measuring the in-time pile-up activity (N_{PV}) and the out-of-time pile-up effect (μ), and needs to be determined as $\langle p_{\mathrm{T}} \rangle(\eta, N_{\mathrm{PV}}, \mu)$ for all run conditions (N_{PV}, μ) occurring in 2012 ATLAS data taking. This is done by collecting $\langle p_{\mathrm{T}} \rangle(\eta, N_{\mathrm{PV}}, \mu)$ profiles in bins of $N_{\mathrm{PV}}(\Delta N_{\mathrm{PV}} = 1)$ and μ

($\Delta\mu = 2$) and converting them to average density profiles $\langle\rho\rangle(\eta, N_{\mathrm{PV}}, \mu)$, using the effective width of the η windows in the area calculation by taking into account detector boundaries.

As the central $\rho_{\mathrm{evt}}^{\mathrm{med}}$ determined in the central region is a sensitive event by event estimator of the pile-up signal activity, the $\langle\rho\rangle(\eta, N_{\mathrm{PV}}, \mu)$ profiles have been normalized such that:

$$P^{\rho}(\eta, N_{\mathrm{PV}}, \mu) = \frac{\langle\rho\rangle(\eta, N_{\mathrm{PV}}, \mu)}{\langle\rho\rangle_{\mathrm{central}}(N_{\mathrm{PV}}, \mu)}. \tag{4.5}$$

η_{plateau} defines the boundary of the central profile region within which $\langle\rho\rangle(\eta, N_{\mathrm{PV}}, \mu)$ can be approximated by its average:

$$\langle\rho\rangle_{\mathrm{central}}(N_{\mathrm{PV}}, \mu) = \frac{1}{2\eta_{\mathrm{plateau}}} \int_{-\eta_{\mathrm{plateau}}}^{+\eta_{\mathrm{plateau}}} \langle\rho\rangle(\eta, N_{\mathrm{PV}}, \mu)\, d\eta$$

therefore the normalization scale $1/\langle\rho\rangle_{\mathrm{central}}(N_{\mathrm{PV}}, \mu)$ in Equation 4.5 only depends on N_{PV} and μ. For the normalized profiles this means that $P^{\rho}(\eta, N_{\mathrm{PV}}, \mu) = 1$ for $|\eta| < \eta_{\mathrm{plateau}}$ for all pile-up conditions (N_{PV}, μ). For simplicity of the extrapolation function and supported by the experimental observations, $P^{\rho}(\eta, N_{\mathrm{PV}}, \mu)$ can safely be assumed as symmetric around $\eta = 0$. $P^{\rho}(\eta, N_{\mathrm{PV}}, \mu)$ drops from its plateau value starting at $|\eta| \approx \eta_{\mathrm{plateau}}$. The shape of these drops are well described by Gaussian shaped shoulders with width σ_{center}, and mean η_{plateau} for the $\eta > 0$ hemisphere, and $-\eta_{\mathrm{plateau}}$ for the $\eta < 0$ hemisphere. A wide baseline underlines the plateau and the Gaussian slopes, it follows a Gaussian form with a mean of $\eta = 0$, a width σ_{base}, and a peak amplitude $\mathcal{A}_{\mathrm{base}}$ and it is constraint by the measured averaged densities at high $|\eta|$. The sum of the central (G_{center}) and base (G_{base}) shapes is normalized such that the total amplitude peaks at 1 at $\eta = \pm\eta_{\mathrm{plateau}}$, to smoothly connect to the normalized plateau value described above. The overall functional form describing the complete shape is then

$$P_{\mathrm{fct}}^{\rho}(\eta, N_{\mathrm{PV}}, \mu) = \begin{cases} 1 & |\eta| < \eta_{\mathrm{plateau}} \\ (1 - G_{\mathrm{base}}(\eta_{\mathrm{plateau}})) \cdot G_{\mathrm{center}}(\eta) + G_{\mathrm{base}}(\eta) & |\eta| \geq \eta_{\mathrm{plateau}} \end{cases}. \tag{4.6}$$

The Gaussian shapes in P_{fct}^{ρ} are defined as

$$G_{\mathrm{center}}(\eta) = \begin{cases} \exp\left[-(\eta - \eta_{\mathrm{plateau}})^2/(2\sigma_{\mathrm{center}}^2)\right] & \eta \geq \eta_{\mathrm{plateau}} \\ \exp\left[-(\eta + \eta_{\mathrm{plateau}})^2/(2\sigma_{\mathrm{center}}^2)\right] & \eta \leq -\eta_{\mathrm{plateau}} \end{cases} \tag{4.7}$$

$$G_{\mathrm{base}}(\eta) = \mathcal{A}_{\mathrm{base}} \cdot \exp\left[-\eta^2/(2\sigma_{\mathrm{base}}^2)\right] \tag{4.8}$$

The measured $P^{\rho}(\eta, N_{\mathrm{PV}}, \mu)$ shapes have been carefully studied in all the available (N_{PV}, μ) bins. The dependence on N_{PV} and μ of the shape parameters, η_{plateau}, σ_{center}, $\mathcal{A}_{\mathrm{base}}$, σ_{base}, in Equations 4.7 and 4.8 are shown in Figures 4.5 and 4.6. An iterative fitting procedure of the functional form $P^{\rho}_{\mathrm{fct}}(\eta, N_{\mathrm{PV}}, \mu)$ from Equation 4.6 with a basic polynomial ansatz for the N_{PV} and μ dependence of the parameters in G_{base} and G_{center}, and thus capturing a possible dependence of these parameters on the pile-up environment, yields:

$$\begin{aligned}
\eta_{\mathrm{plateau}}(N_{\mathrm{PV}}, \mu) &= \eta_{\mathrm{plateau}} = \text{const} \\
\sigma_{\mathrm{center}}(N_{\mathrm{PV}}, \mu) &= \sigma_{\mathrm{center}}(N_{\mathrm{PV}}) = \alpha_0 + \alpha_1 N_{\mathrm{PV}} + \alpha_2 N_{\mathrm{PV}}^2 \\
\mathcal{A}_{\mathrm{base}}(N_{\mathrm{PV}}, \mu) &= \mathcal{A}_{\mathrm{base}}(N_{\mathrm{PV}}) = \beta_0 + \beta_1 N_{\mathrm{PV}} + \beta_2 N_{\mathrm{PV}}^2 \\
\sigma_{\mathrm{base}}(N_{\mathrm{PV}}, \mu) &= \gamma_0(N_{\mathrm{PV}}) + \gamma_1(N_{\mathrm{PV}})\,\mu + \gamma_2(N_{\mathrm{PV}})\,\mu^2 \\
&= \gamma_{0,0}\left(1 + \gamma_{0,1}\exp(\gamma_{0,2} N_{\mathrm{PV}}) + \gamma_{0,3}\exp(\gamma_{0,4} N_{\mathrm{PV}}^2)\right) \\
&+ \gamma_{1,0}\left(1 + \gamma_{1,1} N_{\mathrm{PV}}\right)\mu \\
&+ \gamma_{2,0}\left(1 + \gamma_{2,1} N_{\mathrm{PV}}\right)\mu^2
\end{aligned} \tag{4.9}$$

As shown in figure 4.6, all μ dependence of $P^{\rho}_{\mathrm{fct}}(\eta, N_{\mathrm{PV}}, \mu)$ can be collected in $\sigma_{\mathrm{base}}(N_{\mathrm{PV}}, \mu)$, reflecting that in general the μ dependence of the calorimeter signal is largest in the ATLAS forward calorimeters, the region which constraints σ_{base} most.

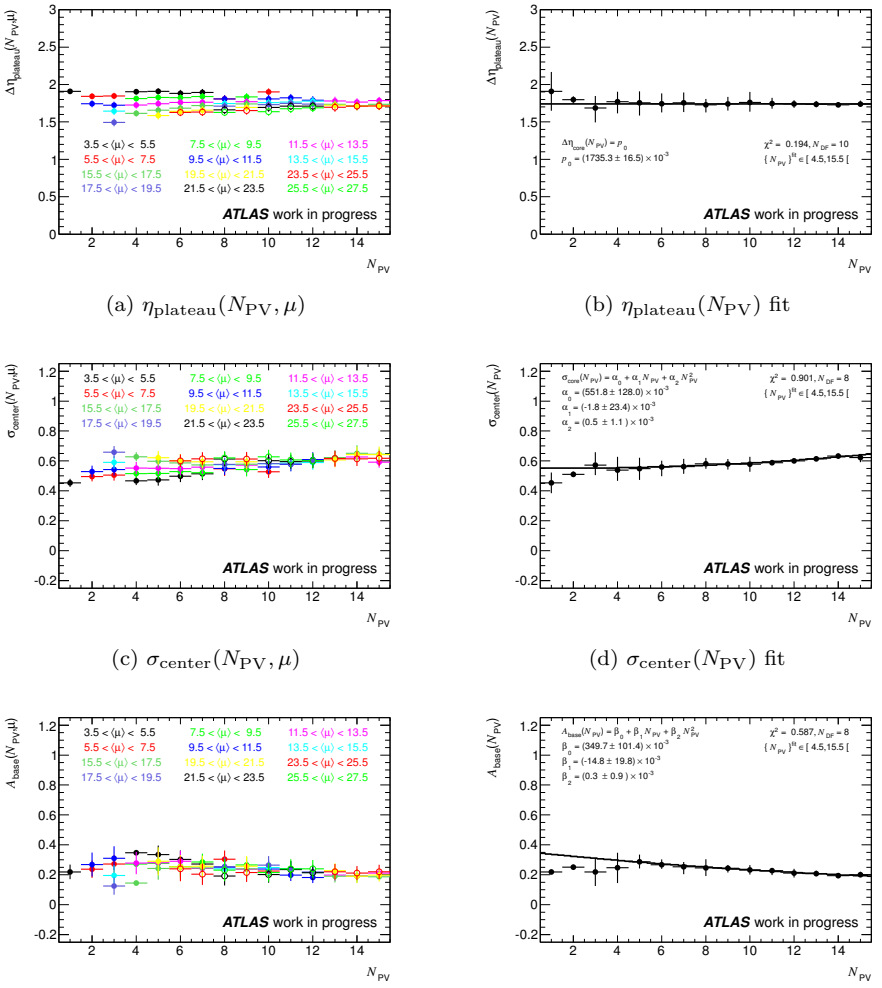
The final set of 16 parameters $\{\eta_{\mathrm{plateau}}, \alpha_i, \beta_i, \gamma_{i,k}\}$ is universal and valid for the whole 2012 data taking period. It has been exclusively derived from data, and the resulting $P^{\rho}(\eta, N_{\mathrm{PV}}, \mu)$ shapes are used for both data and MC. The η, N_{PV} and μ dependent transverse momentum density from minimum bias events is then

$$\rho^{\mathrm{med}}_{\mathrm{evt}}(\eta) = \rho^{\mathrm{med}}_{\mathrm{evt}} \cdot P^{\rho}_{\mathrm{fct}}(\eta, N_{\mathrm{PV}}, \mu), \tag{4.10}$$

where $\rho^{\mathrm{med}}_{\mathrm{evt}}$ is determined within $\eta_{\mathrm{min}} = -\eta_{\mathrm{plateau}}$ and $\eta_{\mathrm{max}} = \eta_{\mathrm{plateau}}$ from the already discussed soft event k_t jets with $R = 0.4$, see Equation 4.3. This is expected to be a good representation of the pile-up activity and effect on the soft event calorimeter signals for any pile-up condition in ATLAS running in 2012. The normalized shapes $P^{\rho}_{\mathrm{fct}}(\eta, N_{\mathrm{PV}}, \mu)$ for selected pile-up conditions expressed by combinations of N_{PV} and μ are shown in Figure 4.7.

4.4.2 Applying a jet area based pile-up suppression

To apply pile-up suppression to the soft term, the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ components are re-summed using only the soft event jets passing the filter using the transverse

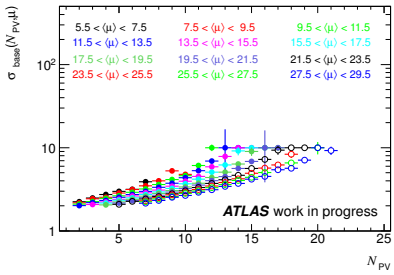


(a) $\eta_{\mathrm{plateau}}(N_{\mathrm{PV}}, \mu)$ (b) $\eta_{\mathrm{plateau}}(N_{\mathrm{PV}})$ fit

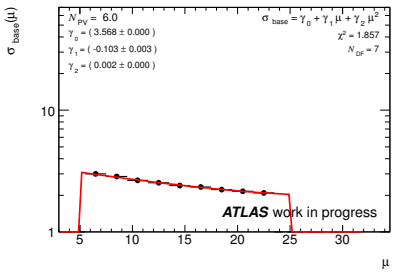
(c) $\sigma_{\mathrm{center}}(N_{\mathrm{PV}}, \mu)$ (d) $\sigma_{\mathrm{center}}(N_{\mathrm{PV}})$ fit

(e) $\mathcal{A}_{\mathrm{base}}(N_{\mathrm{PV}}, \mu)$ (f) $\mathcal{A}_{\mathrm{base}}(N_{\mathrm{PV}})$ fit

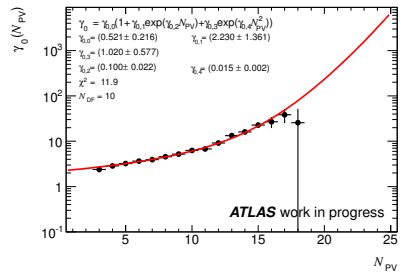
Figure 4.5: The dependence on N_{PV} and μ and the relative fit function of the η_{plateau} parameter are shown in (a) and (b) respectively. The same dependencies are shown for the σ_{center} parameter in (c) and (d), and for the $\mathcal{A}_{\mathrm{base}}$ parameter in (e) and (f).



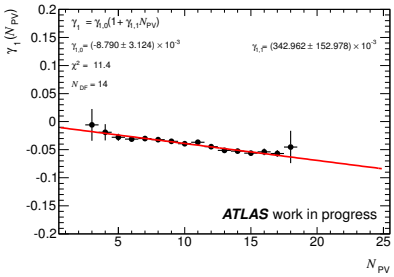
(a) $\sigma_{\text{base}}(N_{\text{PV}}, \mu)$



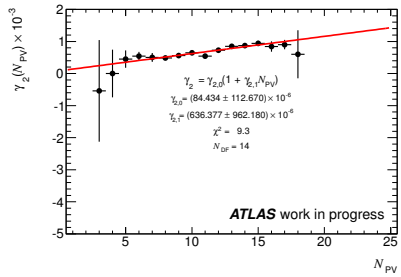
(b) $\sigma_{\text{base}}(\mu)$ fit



(c) $\gamma_0(N_{\text{PV}})$ fit

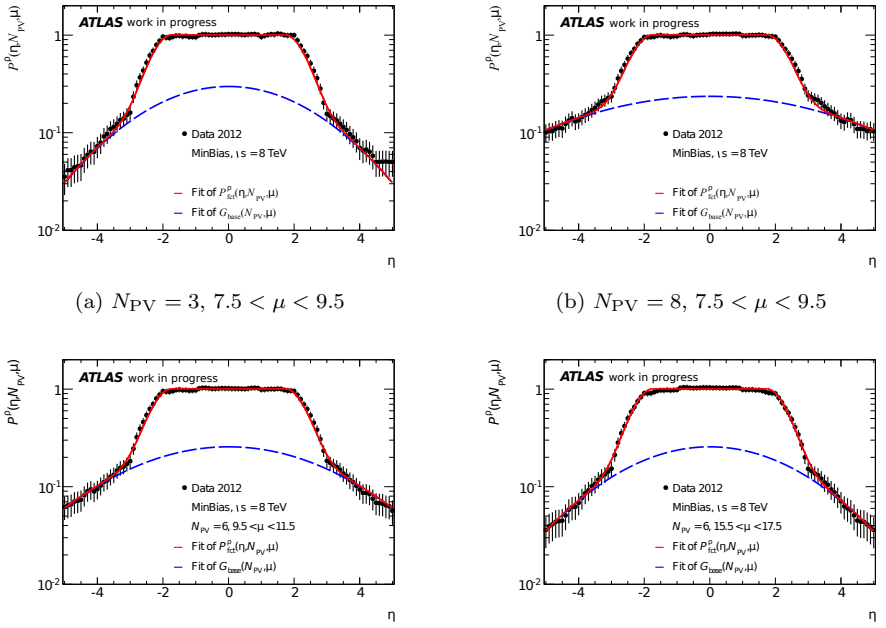


(d) $\gamma_1(N_{\text{PV}})$ fit



(e) $\gamma_2(N_{\text{PV}})$ fit

Figure 4.6: The dependence on N_{PV} and μ of the shape parameter σ_{base} is shown in (a). The fit of the μ dependence according to a three-parameter function is shown in (b). The N_{PV} dependence of the three parameters γ_0, γ_1, γ_2 is shown respectively in (c), (d) and (e).



(a) $N_{\mathrm{PV}} = 3$, $7.5 < \mu < 9.5$

(b) $N_{\mathrm{PV}} = 8$, $7.5 < \mu < 9.5$

(c) $N_{\mathrm{PV}} = 6$, $9.5 < \mu < 11.5$

(d) $N_{\mathrm{PV}} = 6$, $15.5 < \mu < 17.5$

Figure 4.7: The average transverse momentum density shape $P^{\rho}(\eta, N_{\mathrm{PV}}, \mu)$ for $N_{\mathrm{PV}} = 3$ and $7.5 < \mu < 9.5$, determined from 2012 ATLAS minimum bias data, as a function of η (a). The relative increase of the forward activity due to increasing of in-time pile-up ($N_{\mathrm{PV}} = 8$) for the same level of out-of-time pile-up can seen in (b). The signal modulation in the forward region by the out-of-time pile-up effects, as introduced by the particularities of the signal shaping functions in this calorimeter region, can be seen by comparing plot (c) at lower μ with the plot in (d), which shows $P^{\rho}(\eta, N_{\mathrm{PV}}, \mu)$ at higher μ but the same in-time pile-up ($N_{\mathrm{PV}} = 6$). The solid curves show the results for the fully parameterized extrapolation shape $P^{\rho}_{\mathrm{fct}}(\eta, N_{\mathrm{PV}}, \mu)$ given in Equation 4.6, while the dashed curves indicate the shapes of the underlying $G_{\mathrm{base}}(\eta)$ model given in Equation 4.8.

momentum densities, computed coherently for each method, according to Equation 4.4 with $f_{\text{scale}} = 1$. The various configurations considered are:

Extrapolated Jet Area (EJA) $\rho_{\text{evt}}^{\text{med}}(\eta_{\text{jet}})$ is measured and extrapolated as explained in Section 4.4.1. The ρ-jets and filter-jets are identical, and formed with the k_t algorithm with $R = 0.4$.

Extrapolated Jet Area Filtered (EJAF) $\rho_{\text{evt}}^{\text{med}}(\eta_{\text{jet}})$ is measured and extrapolated as explained in Section 4.4.1. The ρ-jets and the filter-jets are k_t jets with $R = 0.6$. After being selected according to Equation 4.4, an additional filter is applied by requiring $|\text{JVF}| > 0.25$ on these jets.

Jet Area Filtered (JAF) $\rho_{\text{evt}}^{\text{med}}$ is determined within the whole pseudo rapidity range $|\eta| < 5$ with k_t jets with $R = 0.8$. The filter-jets are clustered with k_t with $R = 0.4$, and corrected according to Equation 4.4. They are then further filtered by requiring $|\text{JVF}| > 0.25$.

Note that the filter-jets used for defining the pile-up corrected signal contribution to $E_{\text{T}}^{\text{miss,SoftTerm}}$ do not need to be constructed by the same jet definition (algorithm and algorithm parameters) as the ρ-jets used to measure $\rho_{\text{evt}}^{\text{med}}$. Any jet definition using the soft event signals as input and providing a consistent jet area measurement can be used. The details of the JVF filter are given in Section 4.3.1, with a different jet definition and a requirement for a stronger association ($|JVF| > 0.25$) to the primary event vertex.

4.4.3 Transverse momentum density in data and MC

As discussed in the previous sections, the median transverse momentum density $\rho_{\text{evt}}^{\text{med}}$ is the basic observable for the jet area based pile-up corrections of the soft event contribution to $E_{\text{T}}^{\text{miss}}$. It has been measured for the three ρ-jet sizes under considerations, $R = \{0.4, 0.6, 0.8\}$ and the corresponding η ranges. The resulting densities are shown in Figure 4.8 as a function of N_{PV} and μ, respectively for data and MC simulation. Note that the differences in the $\rho_{\text{evt}}^{\text{med}}$ value ranges are primarily not introduced by the ρ-jet size, but rather by the η range used to determined $\rho_{\text{evt}}^{\text{med}}$.

A reasonable agreement is found between data and MC simulations for all three considered configurations with some discrepancies that can be due to pile-up modeling. Also, using different $\rho_{\text{evt}}^{\text{med}}$ for the same pile-up conditions in data and MC is in principal not a problem because $\rho_{\text{evt}}^{\text{med}}$ is determined event by event

and therefore automatically consistent with the real or modeled pile-up activity. The $\rho_{\mathrm{evt}}^{\mathrm{med}}$ extrapolation shapes displayed in Figure 4.7 and used by the EJA and EJAF methods, are found to be universal as well, so that common data derived extrapolation functions $P^{\rho}(\eta, N_{\mathrm{PV}}, \mu)$ are used in both data and MC.

4.5 Performance results and comparison

The $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction performance are extensively tested and evaluated separately in many final states with different event topologies.

In this section the effects of the various pile-up correction methods on the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution, scale and direction are estimated and compared as a function of a reference p_{T} or the global event activity measured by ΣE_{T}. The stability of these performance as a function of the pile-up activity (parametrized by N_{PV} and μ) is also studied. Since this thesis is focused on the methods for the pile-up suppression in the soft term, the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance are firstly studied in events without jets in order to see the direct impact of the various corrections, then the performance are also studied in inclusive events to understand the effects of the pile-up and of the pile-up suppression methods in the presence of a hard scattering interaction.

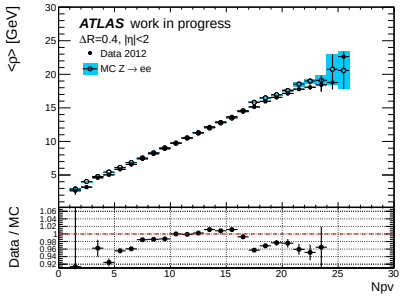
These large variety of studies is not fully available for the TST case, since the method was recently developed and only tests on $Z \to \ell\ell$ simulated samples are performed so far.

4.5.1 Effects of pile-up corrections in $Z \to \mu\mu$ events

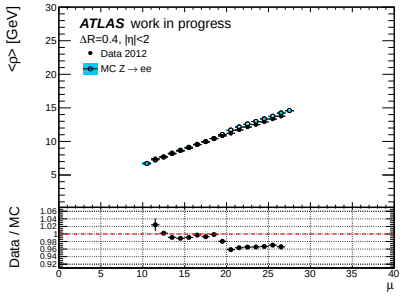
$\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ distribution in $Z \to \mu\mu$ inclusive events

The principal validation of the pile-up suppression methods is based on the data-MC comparisons of the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ and its components. The effect of the various corrections on the distribution of the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ in inclusive $Z \to \mu\mu$ events is shown in Figure 4.9. The observed increased disagreement in the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft term spectra after pile-up corrections using tracks, as can be seen for the STVF based correction in Figure 4.9(b), and the two corrections applying JVF based filters (EJA in Figure 4.9(d) and JAF in Figure 4.9(e)), indicates the already mentioned mis-modeling of reconstructed tracks in Pythia (see discussion of Figure 4.4 in Section 4.3.2).

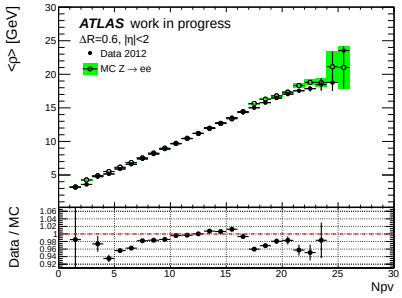
The overall $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions for this sample are presented in Figure 4.10.



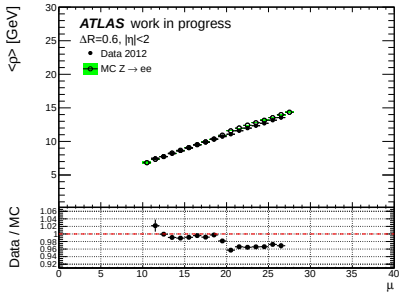
(a) EJA, $\Delta R = 0.4$, $|\eta| < 2$



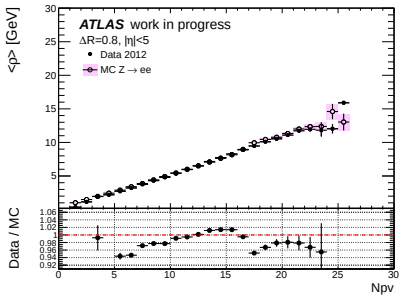
(b) EJA, $\Delta R = 0.4$, $|\eta| < 2$



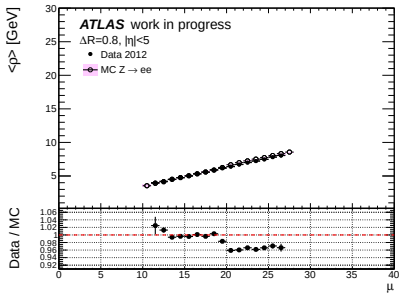
(c) EJAF, $\Delta R = 0.6$, $|\eta| < 2$



(d) EJAF, $\Delta R = 0.6$, $|\eta| < 2$

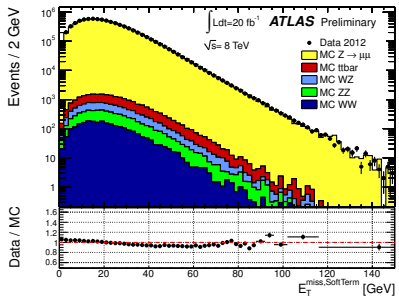


(e) JAF, $\Delta R = 0.8$, $|\eta| < 5$

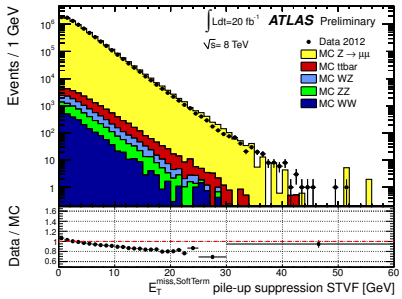


(f) JAF, $\Delta R = 0.8$, $|\eta| < 5$

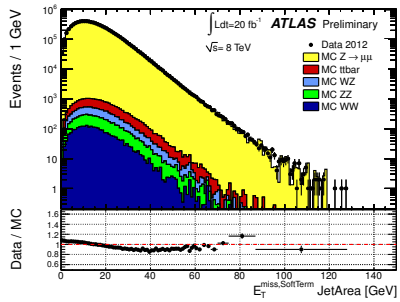
Figure 4.8: Data-MC comparisons of the mean $\langle\rho\rangle = \langle\rho_{\text{evt}}^{\text{med}}\rangle$, as used in the ρ configurations for the jet area based pile-up correction methods. The central $\rho_{\text{evt}}^{\text{med}}$ used in the ρ extrapolation for EJA is shown as a function of N_{PV} in (a) and as a function of μ in (b). Similarly, the central $\rho_{\text{evt}}^{\text{med}}$ used in the extrapolation and filter method EJAF is shown as a function of N_{PV} in (c) and as a function of μ in (d). The $\rho_{\text{evt}}^{\text{med}}$ within the full ATLAS calorimeter coverage, as used by the JAF method, is shown as a function of N_{PV} in (e) and as a function of μ in (f).



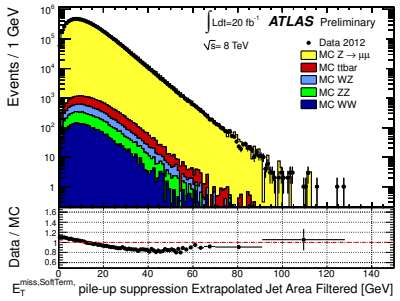
(a) $Z \to \mu\mu$, no pile-up correction



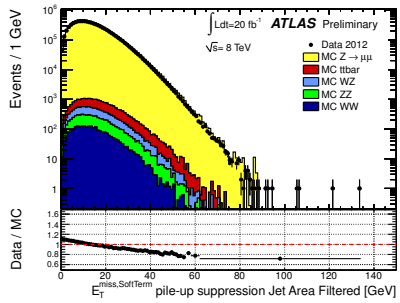
(b) STVF



(c) EJA



(d) EJAF



(e) JAF

Figure 4.9: Data-MC comparisons of the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ distribution in inclusive $Z \to \mu\mu$ events: the uncorrected spectrum in (a), after the STVF based pile-up suppression in (b), after the EJA in (c), after the EJAF in (d) and after the JAF in (e).

The improved data-MC agreement with respect to the soft term comparisons can be understood because the relative contribution of the soft term to the total $E_{\mathrm{T}}^{\mathrm{miss}}$ is strongly reduced after pile-up suppression, moreover, the total $E_{\mathrm{T}}^{\mathrm{miss}}$ includes also jets and leptons that are the same for all the soft term corrections.

$E_{\mathrm{T}}^{\mathrm{miss}}$ and ΣE_{T} dependence on N_{PV} in exclusive $Z \to \mu\mu$ events without jets with p_{T} 20 GeV

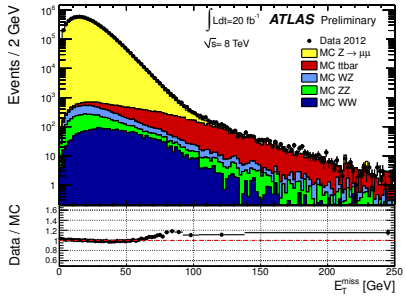
A very good test of the pile-up suppression methods is the study of the stability of some observables as a function of N_{PV}.

The dependence of ΣE_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ on N_{PV} is shown in Figure 4.11 for a $Z \to \mu\mu$ sample without jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 20$ GeV. Before applying any pile-up suppression the mean value of both the ΣE_{T} and the $E_{\mathrm{T}}^{\mathrm{miss}}$ in each N_{PV} bin is largely affected by pile-up and linearly increases with N_{PV}. Large improvements are observed for all the pile-up suppression methods in particular for the STVF method that has very stable performance as a function of N_{PV}. Examining the Jet Area techniques, a residual pile-up dependence is observed in the ΣE_{T} distribution for the EJA method but not for EJAF and JAF indicating that the track-based JVF filter actually helps to identify the contributions associated to the hard scatter signal. Both definitions largely mitigate the pile-up dependence also for the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution even though they are not completely stable as a function of N_{PV}.

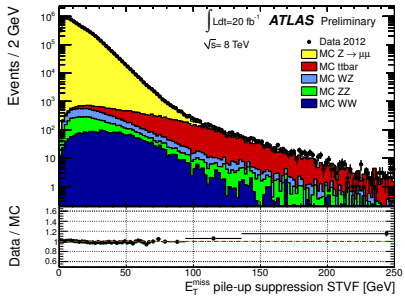
It can be noticed that the EJAF methods performs better for $N_{\mathrm{PV}} < 15$ while the JAF method performs better for $N_{\mathrm{PV}} > 15$. This is due to the limited phase space coverage in (N_{PV}, μ) available for the determination of the extrapolation functions described in Section 4.4.1. In particular, Figure 4.2(b) shows a faster rise of the ΣE_{T} in the forward calorimeter region beyond $N_{\mathrm{PV}} \approx 15$ than the one reflected by the extrapolation.

$E_{\mathrm{T}}^{\mathrm{miss}}$ resolution in $Z \to \mu\mu$ events

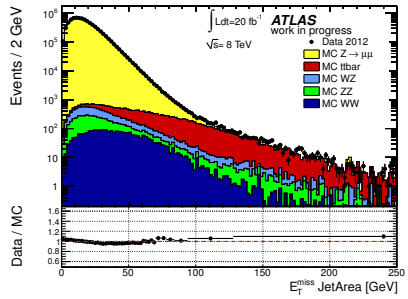
Details about the resolution evaluation are given in Section 3.3.2. The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution in events without genuine missing transverse momentum is measured by the fluctuations in the $E_{\mathrm{T}}^{\mathrm{miss}}$ components E_x^{miss} and E_y^{miss}, which are expected to have a gaussian distribution centered at zero. These fluctuations are increased by the presence of the in-time and out-of time pile-up. One of the aims of the pile-up corrections is to reduce any dependence of the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ on



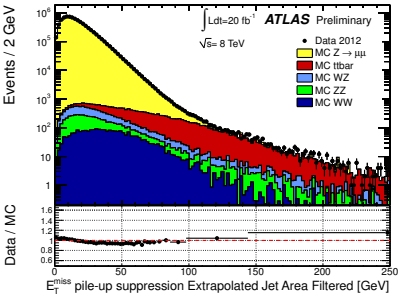
(a) $Z \to \mu\mu$, no pile-up correction



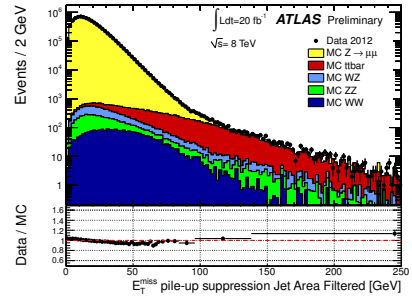
(b) STVF



(c) EJA

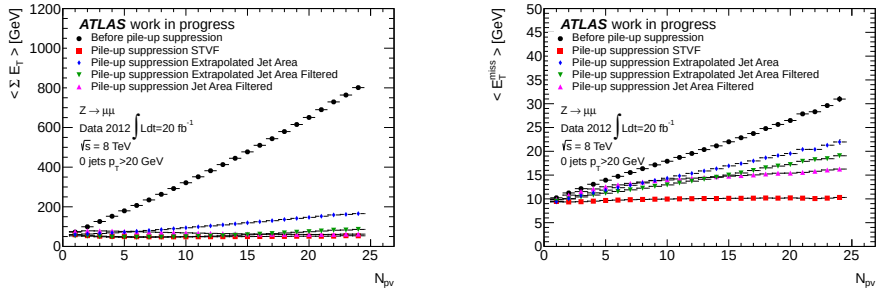


(d) EJAF



(e) JAF

Figure 4.10: Data-MC comparisons of the total $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in inclusive $Z \to \mu\mu$ events: the uncorrected spectrum in (a), after the STVF based pile-up suppression in (b), after the EJA in (c), after the EJAF in (d) and after the JAF in (e).



(a) ΣE_{T} (b) $E_{\mathrm{T}}^{\mathrm{miss}}$

Figure 4.11: Mean of the ΣE_{T} in (a) and of the $E_{\mathrm{T}}^{\mathrm{miss}}$ in (b) as a function of N_{PV} in exclusive $Z \to \mu\mu$ data events without jets with $p_{\mathrm{T}} > 20$ GeV. The performance before and after the various pile-up suppression methods are compared.

N_{PV} and μ and to restore the resolution to values as much as possible similar to the ones measured in the absence of pile-up.

Figure 4.12 shows the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution as a function of N_{PV}. The same trends discussed for the $E_{\mathrm{T}}^{\mathrm{miss}}$ and ΣE_{T} distribution in the previous section are observed: the corrections using track based information (STVF based, and EJAF and JAF) have a better performance. In particular, the STVF method looks very stable at the increasing of the pile-up activity. Nevertheless no conclusion should be drawn at this level: the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance should be evaluated in all its aspects including the study of the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale, direction and tails. Since the STVF method relies on an overall reduction of the soft term signal in general, the effect on the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale, discussed in detail in Section 4.5.3, can be particularly crucial.

4.5.2 Effects of pile-up corrections in inclusive hard scattering events

In this section the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance are studied in presence of a hard scattering final states with and without genuine $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$. The inclusive $Z \to \mu\mu$ sample provides a reference for a sample lacking genuine $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$, while the evaluation of the performance for the soft term pile-up corrections in final states with genuine $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is mainly based on an inclusive $W \to e\nu$ sample, where the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is generated by the neutrino (p_{T}^{ν}). For both the Z and W samples employed for these studies, about 40% of the events are reconstructed with at least one jet with $p_{\mathrm{T}}^{\mathrm{jet}} >$

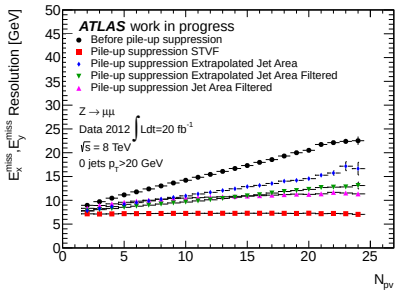


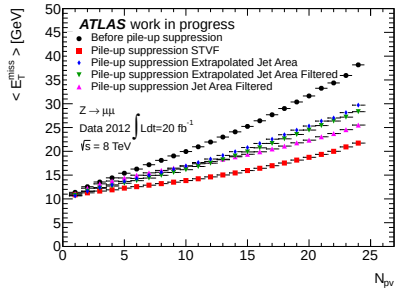
Figure 4.12: $E_{x(y)}^{\mathrm{miss}}$ resolution as a function of N_{PV} in exclusive $Z \to \mu\mu$ data events without jets with $p_{\mathrm{T}} > 20$ GeV. The performance before and after the various pile-up suppression methods are compared.

20 GeV. Simulated Higgs samples with the Higgs decaying to a τ-pair are also used to test the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance since their importance for this thesis, the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is given by the neutrinos produced in the τ decays. In all these samples the soft term gives an important contribution to the total $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$, therefore, the effects of applying a pile-up suppression on this term can be appreciably observed. Topologies with a larger jet activity such as $t\bar{t}$ and simulated SUSY events, are also investigated in Ref. [129] and show no clear benefit by applying of a soft term pile-up suppression in these final states.

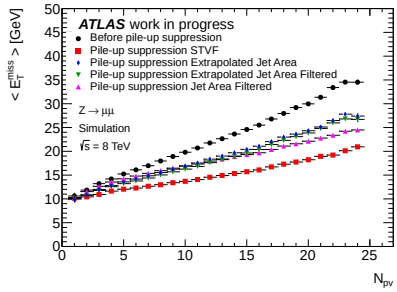
$E_{\mathrm{T}}^{\mathrm{miss}}$ and ΣE_{T} dependence on N_{PV} in inclusive Z/W events

The mean value of the reconstructed $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ as a function of N_{PV} are shown in Figure 4.13 for the $Z \to \mu\mu$ and $W \to e\nu$ inclusive samples before and after the different pile-up suppression corrections. In the comparison between Z and W events it can be noticed that also before applying a pile-up suppression for the soft term the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ in W events is more stable as a function of N_{PV} due to the presence of a genuine $E_{\mathrm{T}}^{\mathrm{miss}}$ than the one reconstructed in Z events. Indeed in events without a genuine $E_{\mathrm{T}}^{\mathrm{miss}}$ the measured mean value of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$(different from 0) is given by the finite E_x^{miss} and E_y^{miss} resolution.

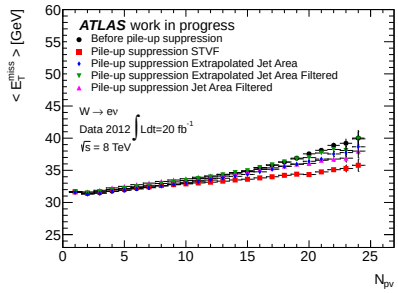
Comparing Figure 4.13(a) and 4.13(b), a good data-MC agreement is observed in Z events, whereas, comparing Figure 4.13(c) and 4.13(d), a disagreement of about 3 GeV is observed in W events for the whole N_{PV} spectrum. This is due to the QCD and electroweak background that are contained in data but that are not considered in the MC simulation. Indeed all these backgrounds, expect for $t\bar{t}$,



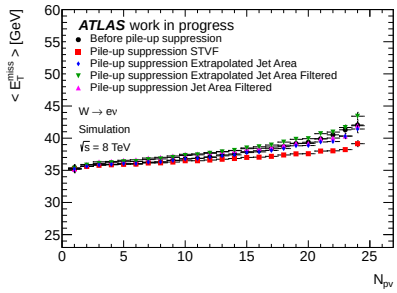
(a) $Z \to \mu\mu$ data, inclusive



(b) $Z \to \mu\mu$ MC, inclusive

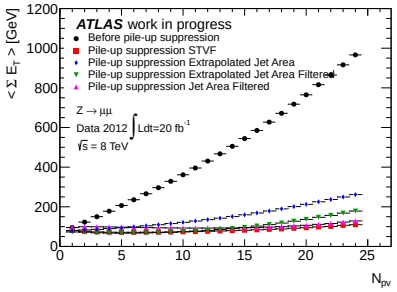


(c) $W \to e\nu$ data, inclusive

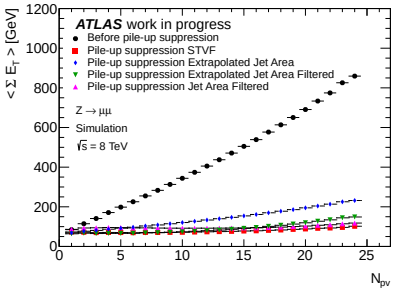


(d) $W \to e\nu$ MC, inclusive

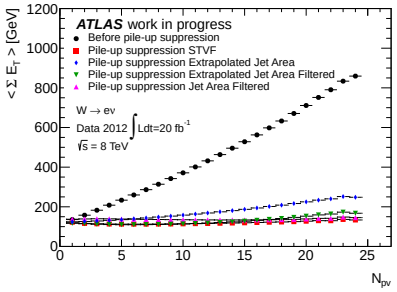
Figure 4.13: Mean of the $E_{\mathrm{T}}^{\mathrm{miss}}$ as a function of N_{PV} in inclusive $Z \to \mu\mu$ data events in (a), in MC events in (b), in inclusive $W \to e\nu$ data events in (c) and MC events in (d). The performance before and after the various pile-up suppression methods are compared.



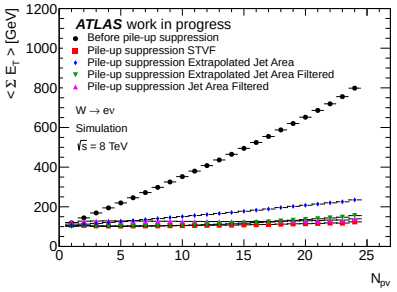
(a) $Z \to \mu\mu$ data, inclusive sample



(b) $Z \to \mu\mu$ MC, inclusive sample



(c) $W \to e\nu$ data, inclusive sample

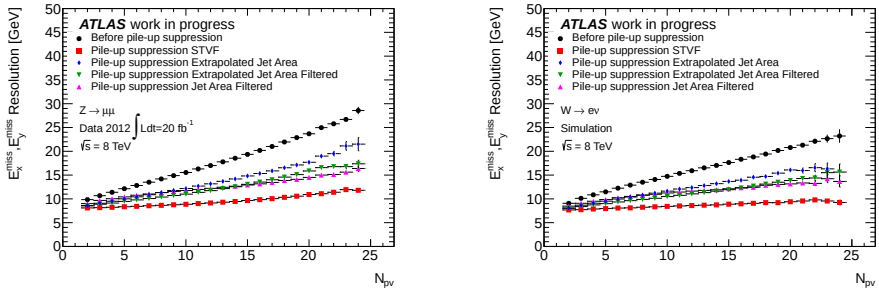


(d) $W \to e\nu$ MC, inclusive sample

Figure 4.14: Mean of the ΣE_{T} as a function of N_{PV} in inclusive $Z \to \mu\mu$ data events in (a), in MC events in (b), in inclusive $W \to e\nu$ data events in (c) and MC events in (d). The performance before and after the various pile-up suppression methods are compared.

contribute mostly to the low $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ region lowering the mean value of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$.

The dependence of the uncorrected ΣE_{T} and of the ΣE_{T} after the different pile-up corrections on N_{PV} can be seen in Figure 4.14. In general the corrections have the same effect in the Z and the W final states, as expected. The soft term pile-up corrections yield slightly worse pile-up suppression in inclusive events with respect to the events without jets shown in Figure 4.11(a) and Figure 4.11(b). This indicates a residual pile-up dependence in the jet reconstruction and a small incoherence between the pile-up corrections applied to jets and to the soft term (that is stronger corrected).



(a) $Z \to \mu\mu$ data, inclusive sample

(b) $W \to e\nu$ MC, inclusive sample

Figure 4.15: $E_{x(y)}^{\text{miss}}$ resolution as a function of N_{PV} evaluated in inclusive $Z \to \mu\mu$ data events in (a) and in inclusive $W \to e\nu$ MC events in (b). The performance before and after the various pile-up suppression methods are compared.

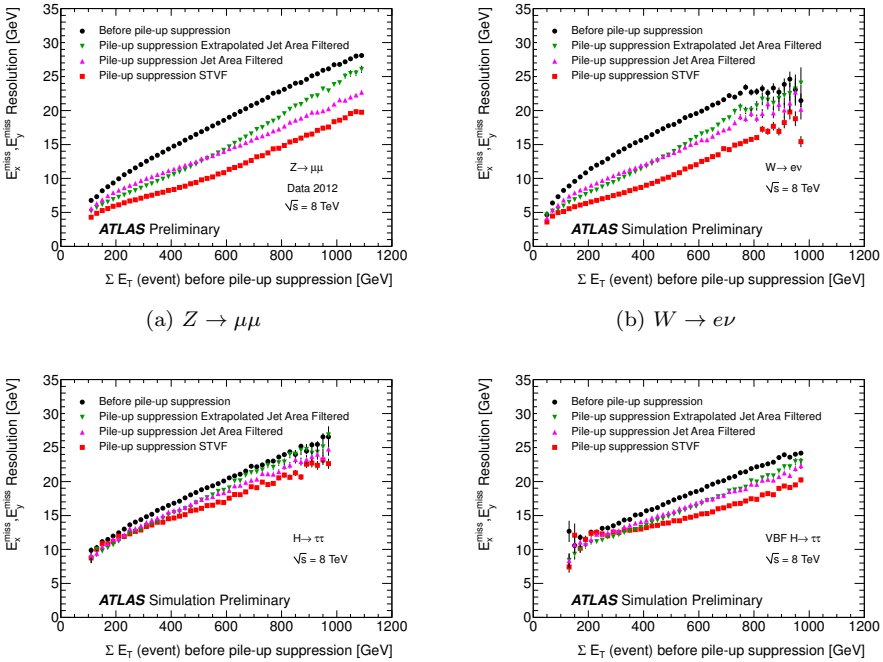
$E_{\text{T}}^{\text{miss}}$ resolution as a function of N_{PV} for inclusive Z/W events

The $E_{\text{T}}^{\text{miss}}$ resolution is calculated as described in Section 3.3.2. It can be measured only in MC sample for $W \to e\nu$ because the $E_{x(y)}^{\text{miss,True}}$ is needed, while it can be measured in both data and MC for $Z \to \mu\mu$ events because $E_{x(y)}^{\text{miss,True}}$ is expected to be zero. The $E_{\text{T}}^{\text{miss}}$ resolution as a function of N_{PV} is shown in Figure 4.15(a) for the inclusive final state with a Z boson in data, and in Figure 4.15(b) for the inclusive W boson in MC. All the pile-up suppression methods improves the $E_{\text{T}}^{\text{miss}}$ resolution with respect to the not corrected case. As already observed in Figure 4.11 and Figure 4.12, the figures indicate that the efficiency of the pile-up correction methods using the extrapolation of the transverse momentum density into the forward region drops for $N_{\text{PV}} > 15$, approximately. This is due to the limitations in the fitting of the extrapolation function, which are already discussed for the the exclusive $Z \to \mu\mu$ sample in Section 4.5.1.

In the comparison with the exclusive case in Figure 4.12, it can be noticed that also for the resolution, slightly worse performance are achieved in the inclusive case due to residual pile-up contamination in the jet reconstruction.

$E_{\text{T}}^{\text{miss}}$ resolution as a function of ΣE_{T} for inclusive hard scattering events

As discussed in Section 3.3.2, the $E_{\text{T}}^{\text{miss}}$ resolution performance can be studied as a function of the total event ΣE_{T}. In this representation an increasing of the resolution proportional to $\sqrt{\Sigma E_{\text{T}}}$ is expected. The results, before and after



(a) $Z \to \mu\mu$ (b) $W \to e\nu$

(c) $gg \to H \to \tau\tau$ (d) VBF $H \to \tau\tau$

Figure 4.16: $E_{x(y)}^{\mathrm{miss}}$ resolution as a function of the uncorrected total event ΣE_{T} evaluated in inclusive $Z \to \mu\mu$ data events in (a), in inclusive $W \to e\nu$ MC events in (b) and in inclusive MC $H \to \tau\tau$ events produced via a $g-g$ fusion mechanism in (c) and via a VBF mechanism in (d). The performance before and after the various pile-up suppression methods are compared.

applying a pile-up suppression, are shown in Figure 4.16 for data $Z \to \mu\mu$ events, and simulated $W \to e\nu$ and $H \to \tau\tau$ events. For the Higgs events the two main production modes, gluon-gluon fusion and vector boson fusion (VBF), are considered.

Particular attention has to be paid in the definition of the x-axis: the choice of employing the not corrected ΣE_{T} as a common axis provides a direct comparison between the various pile-up suppression methods. Similar conclusions as before can be drawn: all the pile-up suppression methods improve the resolution with respect to the not corrected case, in particular, the STVF method performs best. The improvement in the resolution after pile-up suppression is smaller in

Higgs events because of the higher jet activity in these topologies that makes the correction on the soft term less significant. It also important to notice that the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution is estimated according to the width of a Gaussian fit performed on the core of the $E_{x(y)}^{\mathrm{miss}}$ distributions, thus tails given by the presence of backgrounds events in data or either by possible inefficiency of the pile-up suppression methods are disentangled from the resolution measurements.

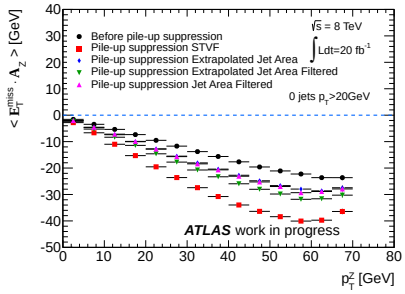
4.5.3 Effects of pile-up corrections on the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale

The $E_{\mathrm{T}}^{\mathrm{miss}}$ scale is given by the residual p_{T} mis-balance between all objects contributing to the $E_{\mathrm{T}}^{\mathrm{miss}}$ signal in final states without genuine missing transverse momentum. The evaluation of the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale in events with $E_{\mathrm{T}}^{\mathrm{miss,True}} > 0$ requires MC, as only here the expectation value for $E_{\mathrm{T}}^{\mathrm{miss}}$ is available (e.g., $E_{\mathrm{T}}^{\mathrm{miss,True}} = p_{\mathrm{T}}^{\nu}$). In both cases it is expected that the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale is independent of any other hard scale in the event, such as p_{T}^{Z} or p_{T}^{ν}. Any absolute systematic and constant deviation from the expectation value for the examined final state is not very important, in particular when this deviation can be well modeled in MC.

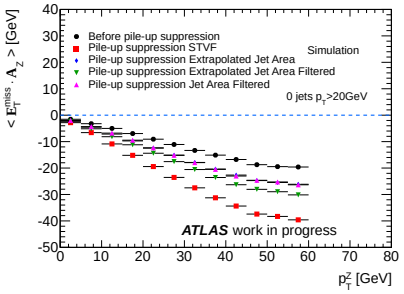
For low reference p_{T} scales, a non-linear deviation from signal linearity is expected for $E_{\mathrm{T}}^{\mathrm{miss}}$, as by construct this observable suffers from an observation bias introduced by the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution. This is discussed further in the following section.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ scale from the diagnostic plot in $Z \to \ell\ell$ events

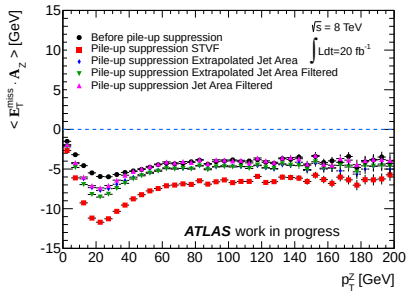
As discussed in Section 3.3.2, for collision events without genuine p_{T}, the projection of the $E_{\mathrm{T}}^{\mathrm{miss}}$ components onto the Z transverse momentum direction is indicative of the features of the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale as function of a (stable) hard scale in the event, like p_{T}^{Z}. This reference is particularly attractive as it is very little to not at all affected by pile-up. In addition, the particular deviation from a linear $E_{\mathrm{T}}^{\mathrm{miss}}$ response, which depends on the composition of the final state and therefore on the inter-calibration between all contributions to $E_{\mathrm{T}}^{\mathrm{miss}}$, is very visible in this diagnostic observable. The effects of pile-up and the applied corrections in the soft term can be seen in the diagnostic plot for the exclusive $Z \to ee$ sample presented in Figure 4.17(a) for data, and in Figure 4.17(b) for MC. The indications from these figures are that the STVF suppresses pile-up but it also removes a larger part of the momentum recoil to the Z boson in the transverse plane and thus leads to a worse $E_{\mathrm{T}}^{\mathrm{miss}}$ response than the jet area based methods,



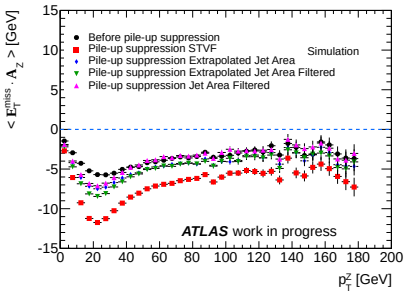
(a) Data, no jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 20$ GeV



(b) MC, no jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 20$ GeV



(c) Data inclusive



(d) MC inclusive

Figure 4.17: Projection of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ onto the Z direction as a function of p_{T}^{Z} in exclusive $Z \to ee$ events without jets with $p_{\mathrm{T}} > 20$ GeV for data in (a) and of MC in (b), and in inclusive $Z \to ee$ events for data in (c) and for MC in (d). The performance before and after the various pile-up suppression methods are compared.

in particular the EJA and JAF methods. These two methods lead to about the same (accidental) loss of recoil signal, and the additional application of JVF in JAF, does not remove any more signals than EJA. Using EJAF, which uses larger jets ($R = 0.6$ instead of $R = 0.4$) for the determination of the central transverse momentum density $\rho_{\text{evt}}^{\text{med}}$ and applies a JVF based selection, removes slightly more recoil than EJA and JAF, mainly due to the fact that the larger soft jet size more likely collects recoil signals and pile-up signals into the same jet. These jets then have lower overall transverse momentum density and are more likely removed from $E_{\text{T}}^{\text{miss,SoftTerm}}$ due to the $\rho_{\text{evt}}^{\text{med}}$ based selection in Equation 4.4.

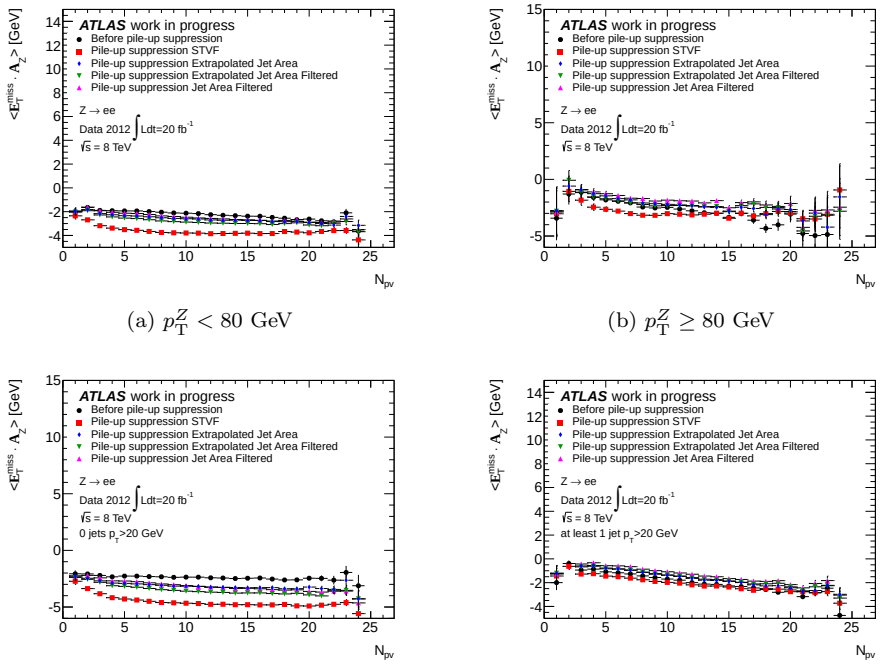
Both data and MC show very similar effects with respect to the uncorrected and the various pile-up corrected projections for the two $Z \to ee$ event selections. This is revisited in the discussion of the systematic uncertainties related to the pile-up correction methods in Section 4.6.

For the inclusive $Z \to ee$ sample, the $E_{\text{T}}^{\text{miss}}$ response is already partly recovered by the (corrected) hard jet response, especially at higher p_{T}^{Z}. As can be seen in Figure 4.17(c) and (d), the STVF method does not fully restore the scale of the not pile-up corrected case neither for high p_{T}^{Z}, while the jet area based methods are performing better with this respect. This is yet another indication of the already mentioned observation that STVF suppresses too much recoil signal. Instead applying the JVF based filters on the soft term jets in addition to the jet area based selections enhances the recoil signal in the inclusive sample by removing a sufficient amount of pile-up everywhere, while not affecting the now harder signals in the non-jet recoil of the event.

The $E_{\text{T}}^{\text{miss}}$ scale as a function of N_{PV} in $Z \to ee$ events

The deviation from linearity of the $E_{\text{T}}^{\text{miss}}$ scale as function of N_{PV} is displayed in Figure 4.18 for both inclusive and exclusive $Z \to ee$ samples. As expected, the pile-up affects less the scale of $\boldsymbol{E}_{\text{T}}^{\text{miss}}$ than its resolution, resulting in a less pronounced dependence of the $\boldsymbol{E}_{\text{T}}^{\text{miss}}$ scale on N_{PV} also before any pile-up suppression.

For the inclusive final state, two transverse momentum regimes have been selected asking for $p_{\text{T}}^{Z} < 80$ GeV and $p_{\text{T}}^{Z} \geq 80$ GeV, respectively. This is motivated by the features visible in the diagnostic projection shown in Figure 4.17. The performance of the pile-up suppression methods affects the scale very differently at low p_{T}^{Z} when compared to the higher p_{T}^{Z} regime. For low p_{T}^{Z} all the physics



(a) $p_{\mathrm{T}}^{Z} < 80$ GeV

(b) $p_{\mathrm{T}}^{Z} \geq 80$ GeV

(c) No jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 20$ GeV

(d) At least one jet with $p_{\mathrm{T}}^{\mathrm{jet}} > 20$ GeV

Figure 4.18: Projection of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ onto the Z direction as a function of N_{PV} in $Z \to ee$ data events with $p_{\mathrm{T}}^{Z} < 80$ GeV in (a), with $p_{\mathrm{T}}^{Z} \geq 80$ GeV in (b), in events without jets with $p_{\mathrm{T}} > 20$ GeV in (c) and in events with at least one jet with $p_{\mathrm{T}} > 20$ GeV (d). The performance before and after the various pile-up suppression methods are compared.

objects are less boosted making harder a completely clean separation between true low-energy signals arising from the hadronic recoil and pileup energy deposits, therefore, together with a large amount of pile-up also some signal is removed. Whereas, for the high p_{T}^{Z} regime the pile-up suppression methods seem to work better, in particular, jet area based methods are even able to slightly improve the performance of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale of the not pile-up corrected case.

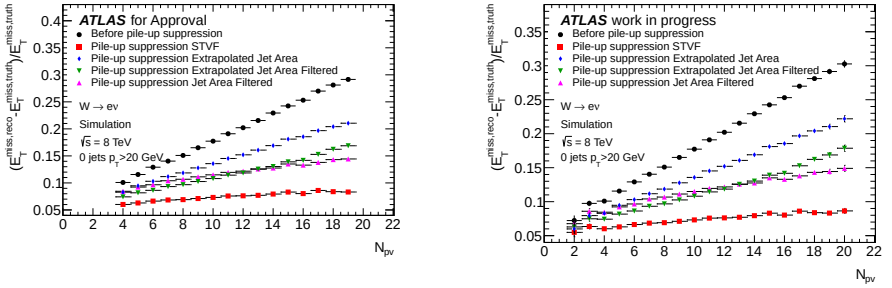
The exclusive plots show that most of the worsening in the scale of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is observed in events with no jets, that are indeed less boosted. For events with at least one jet with $p_{\mathrm{T}}^{\mathrm{jet}} > 20$ GeV, since the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale is driven now by the hard objects, also the STVF case approaches the performance given by other methods.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ scale in events with genuine $E_{\mathrm{T}}^{\mathrm{miss}}$

The effect of the pile-up corrections on the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale in events with genuine missing transverse momentum can only be evaluated in MC, as already discussed before. The signal linearity here is then evaluated for the uncorrected and the various corrected $E_{\mathrm{T}}^{\mathrm{miss}}$ by the relative deviation of the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ from the true missing transverse momentum taken from the neutrino,

$$\frac{E_{\mathrm{T}}^{\mathrm{miss}} - E_{\mathrm{T}}^{\mathrm{miss,True}}}{E_{\mathrm{T}}^{\mathrm{miss,True}}}, \qquad \text{with} \qquad E_{\mathrm{T}}^{\mathrm{miss,True}} = p_{\mathrm{T}}^{\nu}.$$

The results are shown in Figure 4.19 as a function of N_{PV} for events that have no jet with $p_{\mathrm{T}}^{\mathrm{jet}} > 20$ GeV. The vast majority of the events are around $E_{\mathrm{T}}^{\mathrm{miss,True}} = p_{\mathrm{T}}^{\nu} \approx 40$ GeV. The low $E_{\mathrm{T}}^{\mathrm{miss}}$ region, defined as $E_{\mathrm{T}}^{\mathrm{miss,True}} < 40$ GeV, is dominated by a bias introduced by the finite $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution. Hence, for these events the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ would be more probably larger then the true $E_{\mathrm{T}}^{\mathrm{miss}}$ value. Therefore, in each N_{PV} bin the distribution of the difference of the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ and its true value, $E_{\mathrm{T}}^{\mathrm{miss}} - E_{\mathrm{T}}^{\mathrm{miss,True}}$, is not symmetric around zero, but shows a positive tail biasing the overall result towards higher values. Moreover, this also brings in the plot the large $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution dependence on N_{PV}, as shown in Figure 4.19(a). In order to decouple the resolution effects from the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale in Figure 4.19(b) only the peak value of the linearity distribution is fitted. The points in these plots show larger error bars since they reflect the quality of the fit. The results show a reduced dependence of the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale on N_{PV} and it is now visible that the STVF method gives slightly negative values, again indicating an overcorrection of the $E_{\mathrm{T}}^{\mathrm{miss}}$ soft term by this method.



(a) Mean value of the $E_{\mathrm{T}}^{\mathrm{miss}}$ linearity

(b) Fitted peak value of the $E_{\mathrm{T}}^{\mathrm{miss}}$ linearity

Figure 4.19: $E_{\mathrm{T}}^{\mathrm{miss}}$ linearity as a function of N_{PV} in exclusive MC $W \to e\nu$ events without jets with p_{T} 20 GeV: the mean value in (a), the position of the fitted peak in (b). The performance before and after the various pile-up suppression methods are compared.

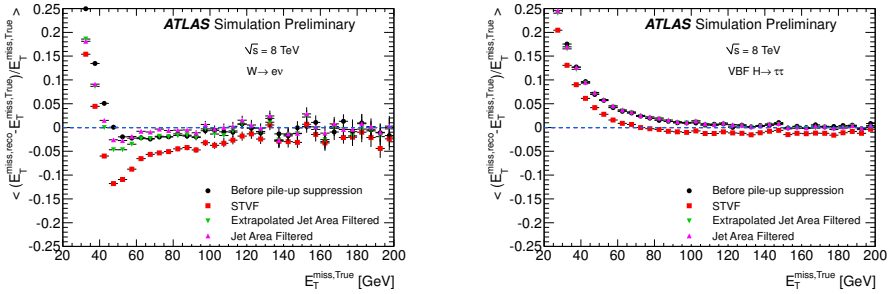
The jet area methods show a correct behavior for the scale improving a bit the performance of the not corrected case.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ linearity can also be studied as a function of the $E_{\mathrm{T}}^{\mathrm{miss,True}}$, as shown in Figure 4.20(a) for an inclusive MC $W \to e\nu$ sample. In this plot the positive bias for $E_{\mathrm{T}}^{\mathrm{miss,True}} < 40$ GeV described previously is directly visible. This plot confirms the conclusions drawn so far: despite the STVF method is more stable at the increasing of the pile-up condition it also tends to overcorrect the soft term with an evaluation of the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement in general lower than the true value, mostly driven by events without jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 20$ GeV. All the jet area techniques perform better with this respect.

The same study is also performed on MC simulated VBF Higgs events decaying to a τ-pair and shown in Figure 4.20(b). The linearity is good (within 1%) for all the $E_{\mathrm{T}}^{\mathrm{miss}}$ definitions. The STVF method, thanks to its better resolution, manages to better restrict the positive bias observed at low $E_{\mathrm{T}}^{\mathrm{miss,True}}$ values and achieves a better linearity mainly in the region $E_{\mathrm{T}}^{\mathrm{miss,True}} < 100$ GeV.

Effect of pile-up corrections on the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ azimuth measurement

The $E_{\mathrm{T}}^{\mathrm{miss}}$ direction in the transverse plane (ϕ^{miss}) which is calculated from the components of $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$, has no specific truth expectation value in events without genuine missing transverse momentum. It is subject to rather large fluctuations and is very sensitive to small signal and intercalibration features, misalignment



(a) MC $W \to e\nu$ events

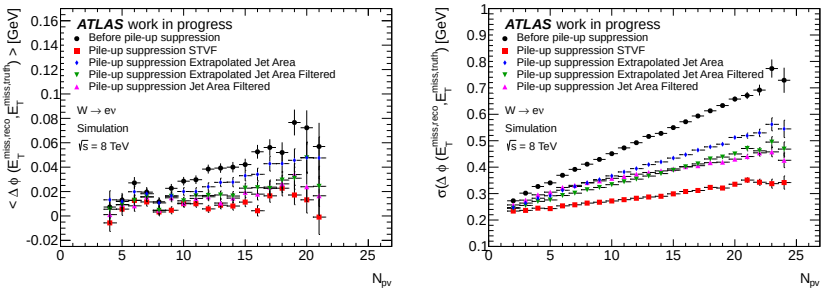
(b) MC VBF $H \to \tau\tau$ events

Figure 4.20: $E_{\mathrm{T}}^{\mathrm{miss}}$ linearity as a function of $E_{\mathrm{T}}^{\mathrm{miss,True}}$ in inclusive MC $W \to e\nu$ events in (a), and MC VBF $H \to \tau\tau$ events in (b). The performance before and after the various pile-up suppression methods are compared.

of the detector with respect to the interaction vertex, and any uncorrected misalignment between detector subsystems. In events with $E_{\mathrm{T}}^{\mathrm{miss,True}} > 0$, though, this angle represents the neutrino scattering azimuth, and the performance of its reconstruction can be evaluated using MC simulations.

Figure 4.21(a) shows the average angular deviation $\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, E_{\mathrm{T}}^{\mathrm{miss,True}})$ between the neutrino azimuth and the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ azimuth, in the inclusive $W \to e\nu$ sample, as function of N_{PV}. The central value is very close to zero, which indicates a good reconstruction of the neutrino direction in the transverse event plane. Neither pile-up itself, nor the methods applied to correct for it affect the angular reconstruction in a significant way. A slight pull can be observed, in particular for the uncorrected and, to a lesser extent for the EJA corrected $\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, E_{\mathrm{T}}^{\mathrm{miss,True}})$ at higher N_{PV}. All corrections using tracking and jet area can re-establish the expected performance on average.

The azimuthal resolution, shown in Figure 4.21(b) , is measured by the fluctuations of $\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, E_{\mathrm{T}}^{\mathrm{miss,True}})$ around its central value in each bin of N_{PV}. It shows sensitivity to pile-up for the uncorrected $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, which can be reduced by applying the pile-up corrections. The methods combining tracking and calorimetry (EJAF and JAF) perform better than the calorimeter only based EJA method, yet the tracking only based STVF performs best with respect to the azimuthal resolution.



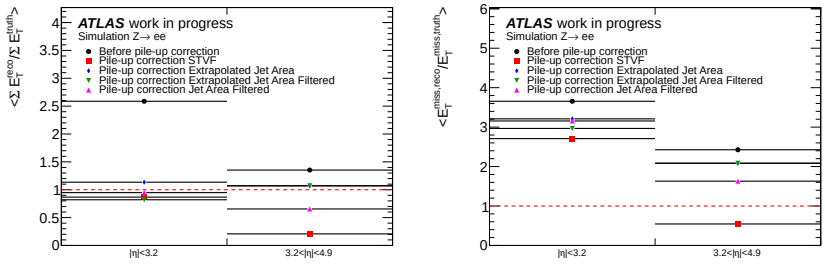
(a) $\langle \Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, E_{\mathrm{T}}^{\mathrm{miss,True}}) \rangle (N_{\mathrm{PV}})$ (b) $\sigma(\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, E_{\mathrm{T}}^{\mathrm{miss,True}}))(N_{\mathrm{PV}})$

Figure 4.21: Average angular dispersion between the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ azimuth and the neutrino azimuth as a function of N_{PV} in the inclusive $W \to e\nu$ MC sample in (a), and the fluctuation around this average in (b). The performance before and after the various pile-up suppression methods are compared.

4.5.4 Effects of pile-up corrections in calorimeter regions

The effect of pile-up on the calorimeter signal formation is different in the various calorimeter regions defined by the pseudorapdity η_{det} with respect to the nominal vertex in ATLAS. It depends on the sensitivity in these regions to smaller energy deposits, mostly limited by signal fluctuations introduced by electronic and pile-up noise, and by reconstruction thresholds (in terms of smallest safely measurable energy) in the topological cluster formation in the calorimeter. Additional limitations in sensitivity to small energies in the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ are introduced by the changing of the calorimeter readout granularity, which may sparsify the signal in the event plane in $(\eta_{\mathrm{det}}, \phi)$ space, as already seen in the discussion of the $\rho_{\mathrm{evt}}^{\mathrm{med}}$ measurement and extrapolation in Section 4.4.1. As for the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale, the signal modulations by out-of-time pile-up are also changing with calorimeter regions in ATLAS, which is demonstrated in Ref. [142]. These region dependencies are addressed differently in the correction methods, with STVF and JAF using no particular regional information, while EJA and EJAF use an η_{det}-dependent transverse momentum density.

It is possible to define a truth expectation for $E_{\mathrm{T}}^{\mathrm{miss}}$ for a given calorimeter region using MC simulations. Contrary to the usual $E_{\mathrm{T}}^{\mathrm{miss}}$ truth, which is given by the transverse momentum of the final state neutrino in interactions with genuine missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss,True}} = p_{\mathrm{T}}^{\nu}$), the regional truth reference



(a) Regional ΣE_{T} reconstruction

(b) Regional $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction

Figure 4.22: Comparison of the reconstructed ΣE_{T} in (a) and $E_{\mathrm{T}}^{\mathrm{miss}}$ in (b) with the corresponding expectations from the particle level truth event in inclusive $Z \to ee$ events. The effects of the various pile-up corrections on these two quantities are shown in two regions, $|\eta_{\mathrm{det}}| < 3.2$ and $3.2 < |\eta_{\mathrm{det}}| < 4.9$.

has to be calculated using the generated flow of interacting particles in MC. It is therefore a pure detector performance related reference, rather than one for physics analysis, with the advantage of providing a truth reference also for samples without neutrinos.

The pile-up correction methods are expected to at least reinstate the interacting particle truth on average, overall and in given regions of the detector. Figure 4.22 shows the average ratio of reconstructed over particle level ΣE_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ in two larger regions of the ATLAS calorimeters ($|\eta_{\mathrm{det}}| < 3.2$ and $3.2 < |\eta_{\mathrm{det}}| < 4.9$), for an inclusive $Z \to ee$ MC sample. The results are presented inclusive in N_{PV}, with an average $\langle N_{\mathrm{PV}} \rangle \approx 11$ in the sample. The already indicated overcorrection by the STVF approach is confirmed, especially for ΣE_{T} in the more forward calorimeter region. The methods using an extrapolated transverse momentum density into this region, EJA and EJAF, produce the same corrected ΣE_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ in the forward region, as the JVF based selection is not applied to soft term jets beyond $|\eta_{\mathrm{det}}| > 2.4$. They reconstruct the true ΣE_{T} very well in the forward region, and also perform best for the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction.

In the central region all corrections reconstruct ΣE_{T} well, with a slight under-correction for EJA and a slight over-correction for the STVF method, both expected. The combined jet area/tracking based corrections EJAF and JAF perform best for ΣE_{T}. On the other, $E_{\mathrm{T}}^{\mathrm{miss}}$ in both central and forward directions is not reconstructed as well on average, which is related to the already mentioned

observation bias introduced by $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution for this quantity. As most of the events populate phase space with low p_{T}^{Z}, $E_{\mathrm{T}}^{\mathrm{miss}}$ is typically too large compared to the truth expectation, with STVF introducing the smallest ratio to truth due to the fact that it shows the best performance in $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution for this sample, compare Figure 4.12 in Section 4.5.1.

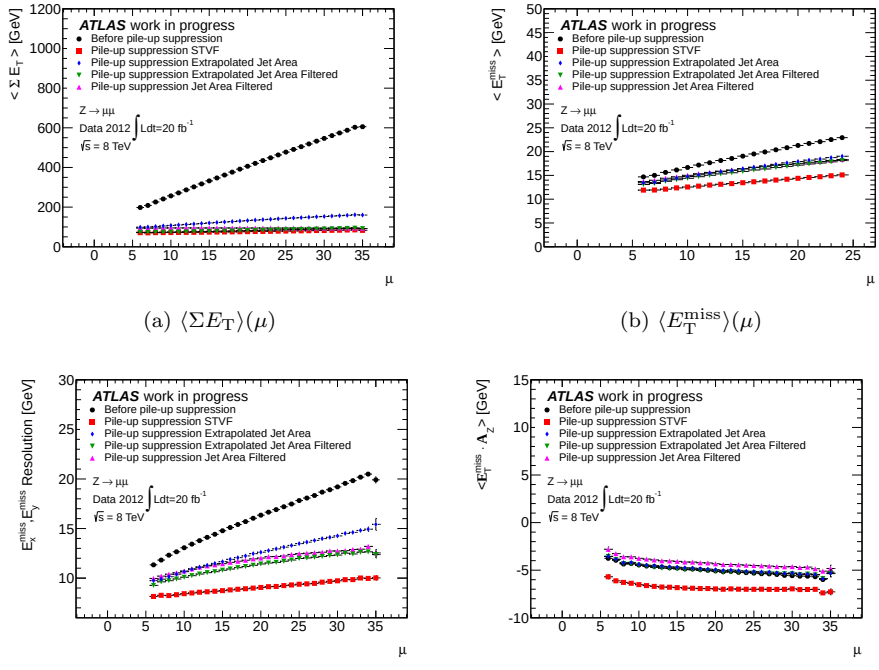
4.5.5 Effects of pile-up corrections with respect to μ

The out-of-time activity in ATLAS relates to the average number of interactions μ measured in a given time window around the triggered collision event, see e.g. Ref. [142] for more details. Unfortunately μ is only slowly changing and the overall accessible range of μ values allowing a statistically meaningful evaluation of the pile-up correction features is limited. Nevertheless, Figure 4.23 shows the dependence on μ of ΣE_{T} in (a), of $E_{\mathrm{T}}^{\mathrm{miss}}$ in (b), of the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution in 4.23(c), and of the $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ projection in (d), for the inclusive $Z \to \mu\mu$ sample. The pile-up corrections work very well also with respect to μ. In particular the EJAF method does not show the particular performance degradation seen in the N_{PV} dependence of some different observables (see e.g. Figures 4.12 and 4.15). This indicates a sufficiently good description of the out-of-time pile-up signal features by the extrapolation functions, which are mostly functions of N_{PV} in the regions closest to η_{plateau}(see Equation 4.9). The μ-dependence is explicitly included only in the width of the Gaussian baseline that mostly contributes in the forward region. In particular at an increasing of μ corresponds an effective decreasing of the transverse energy, partially compensating[6] the overall increasing given by the in-time pile-up [142].

4.5.6 Effects of pile-up corrections on $E_{\mathrm{T}}^{\mathrm{miss}}$ tails

Instrumental effects and badly measured contributions can create fake high momentum unbalance in the evaluation of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$. These tails of the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution lead to an additional background in searches for new undetected particles. It is, therefore, important that the methods used to narrow the bulk of the resolution function do not increase the size of those tails. Figure 4.24 compares the number of events which have $E_{\mathrm{T}}^{\mathrm{miss}}$ above a fixed threshold before and after the pile-up suppression with the various pile-up suppression methods considered in

[6] Due to the particularities of the signal shaping function in the ATLAS forward calorimeters, a large μ can even lead to overcompensation of the in-time pile-up signal contribution and therefore increasing true signal loss in the region $3.2 < |\eta_{\mathrm{det}}| < 4.9$

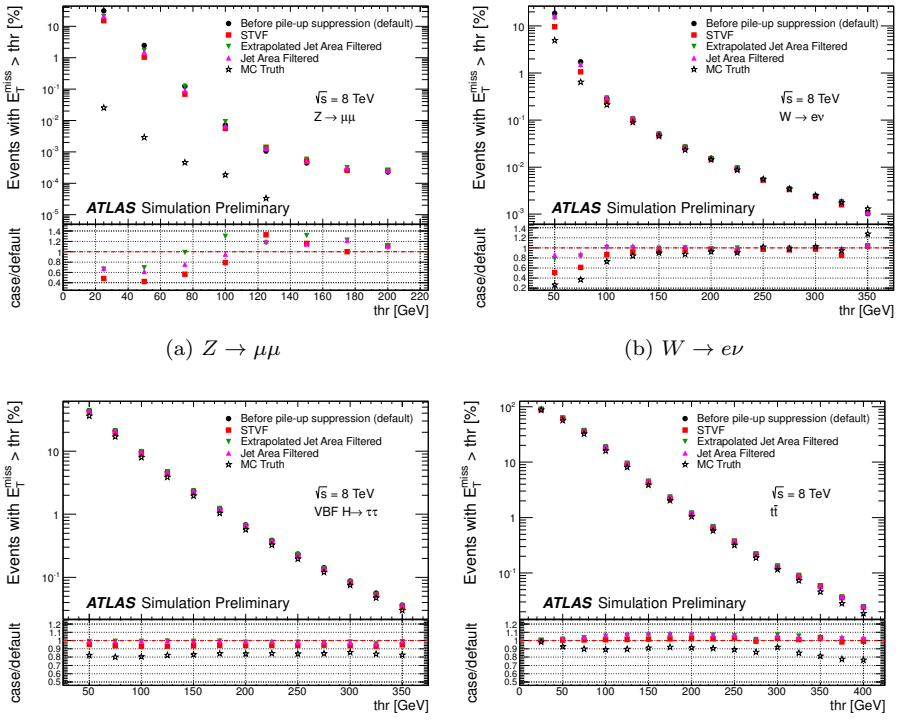


(a) $\langle \Sigma E_{\mathrm{T}} \rangle(\mu)$ (b) $\langle E_{\mathrm{T}}^{\mathrm{miss}} \rangle(\mu)$

(c) $E_{x(y)}^{\mathrm{miss}}(\mu)$ resolution (d) $\langle \mathbf{E}_{\mathrm{T}}^{\mathrm{miss}} \cdot \mathbf{A}_{\mathbf{Z}} \rangle(\mu)$

Figure 4.23: μ-dependence in inclusive $Z \to \mu\mu$ data events of the mean ΣE_{T} in (a), of the mean $E_{\mathrm{T}}^{\mathrm{miss}}$ in (b), of the $E_{x(y)}^{\mathrm{miss}}$ resolution in (c) and of the projection of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ onto the Z direction, averaged over the full p_{T}^{Z} spectrum, in (d). The performance before and after the various pile-up suppression methods are compared.

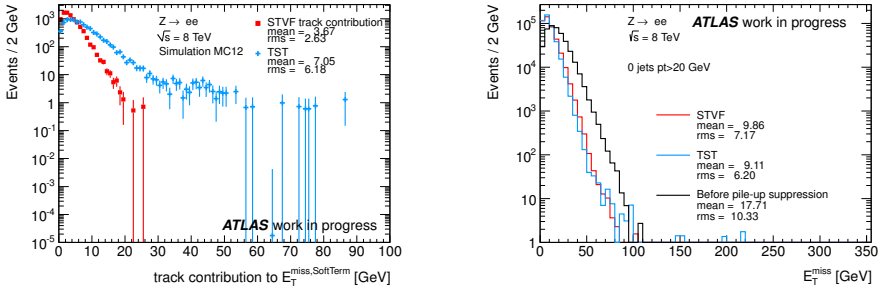
this note for MC events in different samples. It can be seen that the pile-up mitigation techniques do not significantly increase the tails of the MET distributions. Few additional tails are created in events with jets filtered by the JVF cut and are currently under study. Since all the methods presented here share the same treatment for the jet term, the size of the additional tails is similar across all methods, while no tails are observed in events with no jets.



(a) $Z \to \mu\mu$ (b) $W \to e\nu$

(c) VBF $H \to \tau\tau$ (d) $t\bar{t}$

Figure 4.24: Percentage of events with $E_{\mathrm{T}}^{\mathrm{miss}}$ above a varying threshold in inclusive MC $Z \to \mu\mu$ events in (a), $W \to e\nu$ events in (b), VBF $H \to \tau\tau$events in (c), $t\bar{t}$ events in (d). The performance before and after the various pile-up suppression methods are compared. The $E_{\mathrm{T}}^{\mathrm{miss,True}}$ is also shown for comparison. The lower parts of the figures show the ratio of each case (pile-up suppression method or Truth) over the default (before pile-up suppression).



(a) soft track term

(b) $E_{\mathrm{T}}^{\mathrm{miss}}$

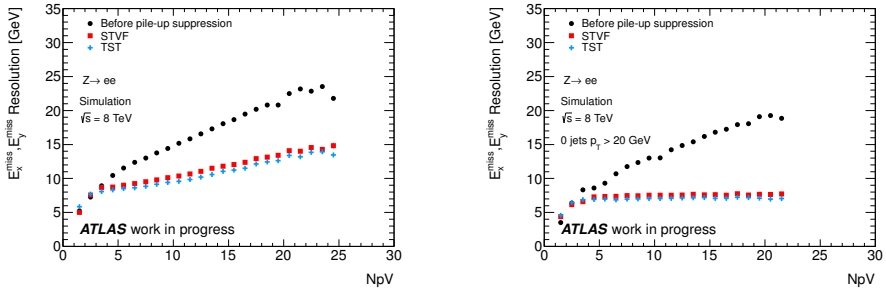
Figure 4.25: The soft track term calculated with the two track selections employed by the STVF and TST methods is compared in (a) for inclusive $Z \to ee$ MC events, the $E_{\mathrm{T}}^{\mathrm{miss}}$ spectrum before the pile-up suppression and after the STVF and TST methods is compared in (b) for exclusive $Z \to ee$ MC events without jets with $p_{\mathrm{T}} > 20$ GeV.

4.5.7 Study of the performance of the TST method

The TST method, described in Section 4.3.3, was recently developed and it is currently under validation. In this section first preliminary studies based on MC simulation samples are discussed and the TST method is compared with the other track-based method, STVF, and with the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ before applying a pile-up correction for the soft term.

The first observation is that the different track quality criteria employed by the STVF and TST methods have a large impact on the soft term computation from tracks from primary vertex only. This effect can be quantified from Figure 4.25(a), where the track soft term is shown for $Z \to ee$ inclusive events for the two methods. In particular the tighter quality criteria used by the STVF result in a lower estimation of the track soft term. On the other hand the TST track selection, even after applying the set of cuts to reject mis-reconstructed tracks listed in Section 4.3.3, still creates some tails in the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in comparison with the STVF method as shown in Figure 4.25(b).

The $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution and the diagnostic plot for the TST method are respectively shown in Figures 4.26 and 4.27 for inclusive and exclusive $Z \to ee$ events. The resolution as a function of N_{PV} demonstrates the good stability of both TST and STVF methods with comparable performance. The diagnostic plot clearly



(a) inclusive $Z \to ee$ (b) $Z \to ee$ without jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 20$ GeV

Figure 4.26: $E_{x(y)}^{\mathrm{miss}}$ resolution as a function of N_{PV} for inclusive MC $Z \to ee$ events in (a) and exclusive MC $Z \to ee$ events without jets with $p_{\mathrm{T}} > 20$ GeV in (b). The performance before pile-up suppression and after the STVF and TST methods are compared.

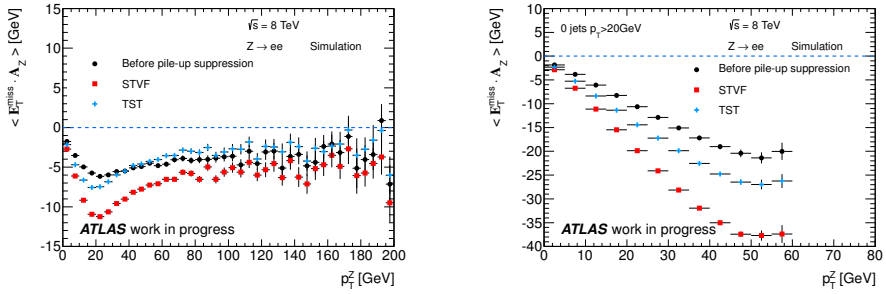
indicates that the TST method performs better, restoring the performance of the not-corrected case at high-p_{T}^{Z} and strongly reducing the negative bias also for events with low p_{T}^{Z}.

4.6 Systematics uncertainties

In the following sections, the total systematics uncertainties on the soft term scale and resolution are provided for all the pile-up suppression methods previously discussed in Section 4.2. A specific evaluation of the contribution on the uncertainties introduced by applying a pile-up suppression method is discussed in Section 4.6.2.

4.6.1 Total systematics uncertainties for the pile-up corrected $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ cases

To evaluate the systematics uncertainties on the soft term for each of the pile-up suppressed version of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$, the same in-situ methods explained in Section 3.3.3 are employed. The results are summarized in Table 4.1, for comparison also the systematics uncertainties on the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution before applying any pile-up suppression are reported. It should be noticed that also if the fractional uncertainties are larger for the pile-up suppressed $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ in some cases,



(a) inclusive $Z \to ee$ (b) $Z \to ee$ without jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 20$ GeV

Figure 4.27: Projection of the $E_{\mathrm{T}}^{\mathrm{miss}}$ onto the Z direction as a function of p_{T}^{Z} for inclusive MC $Z \to ee$ events in (a) and exclusive MC $Z \to ee$ events without jets with $p_{\mathrm{T}} > 20$ GeV in (b). The performance before pile-up suppression and after the STVF and TST methods are compared.

they have a smaller impact on the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ global uncertainty because the soft term is much smaller after pile-up suppression.

The systematics uncertainties for the TST $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ case are not discussed here since they are currently under derivation using a variation of the balance method (see Section 3.3.3).

4.6.2 Systematics uncertainties introduced by pile-up suppression methods

The evaluation of the systematic uncertainties, both on the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution, introduced by any of the studied pile-up correction methods for the soft term is based on the quality of the modeling in MC for the relative variations between uncorrected and corrected observables, O, as a function of a hard scale in the collision event. This relative effect $\mathcal{R}$ of any of the corrections is measured using the ratio of a specific observable O after and before a given correction is applied:

$$\mathcal{R}(\Sigma E_{\mathrm{T}}) = \frac{O_{\mathrm{corrected}}(\Sigma E_{\mathrm{T}})}{O_{\mathrm{uncorrected}}(\Sigma E_{\mathrm{T}})} \,. \tag{4.11}$$

The hard scale at which $\mathcal{R}$ is measured is given by the (uncorrected) ΣE_{T} of the event. The event sample used is the exclusive $Z \to \mu\mu$ sample without jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 20$ GeV, because the events in this sample are dominated by the soft

$E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$ uncertainty	data/MC method		balance method		
	scale	resolution	scale		resolution
	(%)	(%)	([GeV])	(%)	(%)
RefFinal	3.6	2.3	< 1 GeV	<13	2.0
STVF	7.9	4.8	< 1 GeV	<12	4.5
EJA	3.3	1.5	–	–	–
EJAF	4.7	2.0	< 1 GeV	< 18	3.0
JAF	5.8	2.5	< 1 GeV	< 16	2.0

Table 4.1: Systematic uncertainties on the scale and on the resolution of the $E_{\mathrm{T}}^{\mathrm{miss,SoftTerm}}$, calculated with the two different in-situ methods: based on the data/MC ratio and on the balance between the soft term and the hard objects (see Section 3.3.3).

term. The observables O are the same variables employed for the determination of the total uncertainties described in Section 4.6.1, the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ projected onto the Z boson direction $< \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}} \cdot \boldsymbol{A}_{\mathbf{Z}} >$ for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale evaluation and the gaussian width of the E_x^{miss} and E_y^{miss} components for the resolution evaluation.

The scoring variable used to measure the systematic uncertainty is the ratio R_{syst} of the relative correction effect in MC ($\mathcal{R}^{\mathrm{MC}}$) to the one in data ($\mathcal{R}^{\mathrm{data}}$), both calculated for each pile-up correction as given in Equation 4.11,

$$\begin{aligned} \mathcal{F}_{\mathrm{scale}}(\Sigma E_{\mathrm{T}}) &= \frac{\mathcal{R}^{\mathrm{MC}}(\Sigma E_{\mathrm{T}})}{\mathcal{R}^{\mathrm{data}}(\Sigma E_{\mathrm{T}})} \\ &\equiv \frac{O^{\mathrm{MC}}_{\mathrm{corrected}}(\Sigma E_{\mathrm{T}})/O^{\mathrm{MC}}_{\mathrm{uncorrected}}(\Sigma E_{\mathrm{T}})}{O^{\mathrm{data}}_{\mathrm{corrected}}(\Sigma E_{\mathrm{T}})/O^{\mathrm{data}}_{\mathrm{uncorrected}}(\Sigma E_{\mathrm{T}})} \\ &\equiv \frac{O^{\mathrm{MC}}_{\mathrm{corrected}}(\Sigma E_{\mathrm{T}})/O^{\mathrm{data}}_{\mathrm{corrected}}(\Sigma E_{\mathrm{T}})}{O^{\mathrm{MC}}_{\mathrm{uncorrected}}(\Sigma E_{\mathrm{T}})/O^{\mathrm{data}}_{\mathrm{uncorrected}}(\Sigma E_{\mathrm{T}})} \,. \end{aligned} \tag{4.12}$$

The systematic uncertainties for each pile-up suppression method are then given by the weighted standard deviation of $\mathcal{F}_{\mathrm{scale}}(\Sigma E_{\mathrm{T}})$,

$$\begin{aligned} \Delta\mathcal{F}^{\mathrm{syst}}_{\mathrm{scale}} &= \sqrt{\frac{1}{\mathcal{N}} \sum \mathcal{F}^2_{\mathrm{scale},i}/\Delta\mathcal{F}^2_{\mathrm{scale},i} - \langle\mathcal{F}\rangle^2_{\mathrm{scale}}} \\ \text{with } \mathcal{N} = \sum 1/\Delta\mathcal{F}^2_{\mathrm{scale},i} \;&\text{ and } \; \langle\mathcal{F}\rangle_{\mathrm{scale}} = \frac{1}{\mathcal{N}} \sum \mathcal{F}_{\mathrm{scale},i}/\Delta\mathcal{F}^2_{\mathrm{scale},i} \end{aligned} \tag{4.13}$$

from their central value $\langle\mathcal{F}\rangle_{\mathrm{scale}}$ averaged over the full accessible ΣE_{T} range. The

Table 4.2: Summary of the fractional systematic uncertainty contribution from applying the pile-up correction methods to the exclusive $Z \to \mu\mu$ sample.

Pile-up correction method	Fractional systematic uncertainties	
	$E_{\mathrm{T}}^{\mathrm{miss}}$ scale [%]	$E_{\mathrm{T}}^{\mathrm{miss}}$ resolution [%]
STVF	1.4	0.7
EJA	1.0	0.5
EJAF	1.6	1.1
JAF	1.1	0.7

sums run over ΣE_{T} bins such that

$$\mathcal{F}_{\mathrm{scale},i} = \langle \mathcal{F}_{\mathrm{scale}} \rangle (\Sigma E_{\mathrm{T}}) \quad \text{for} \quad \Sigma E_{\mathrm{T}} \in [(\Sigma E_{\mathrm{T}})_i, (\Sigma E_{\mathrm{T}})_{i+1}[\,,$$

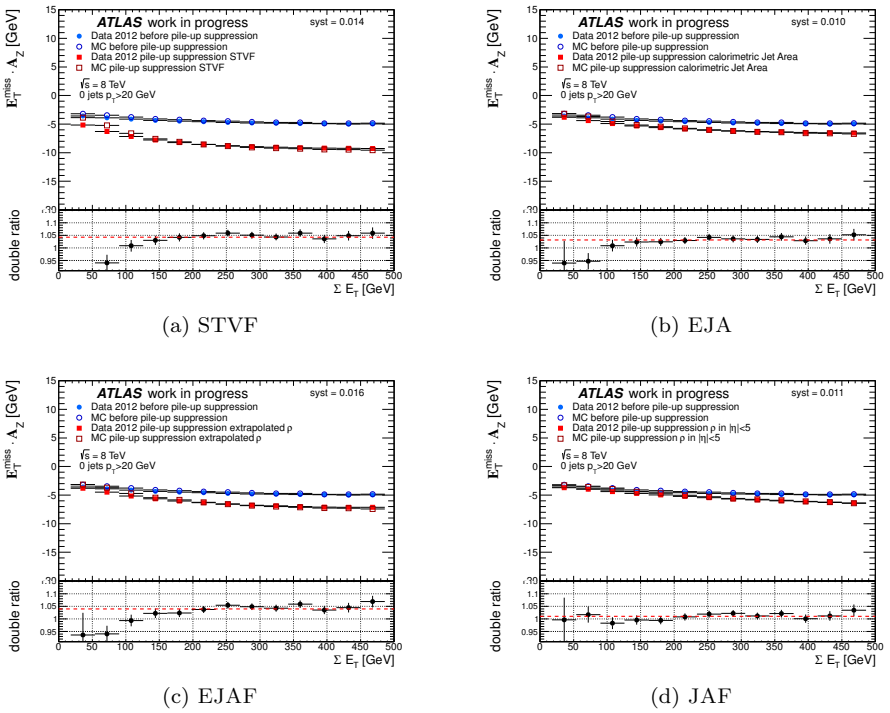
and the corresponding (bin-by-bin) uncertainty $\Delta\mathcal{F}_{\mathrm{scale},i}$ determined by the uncertainty associated with this bin average. The definition decouples the estimated systematic uncertainty of the $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement after corrections in data from the absolute agreement with MC simulations reflected by $\langle \mathcal{F} \rangle_{\mathrm{scale}}$, as this average value can be otherwise constraint in the context of a full systematic uncertainty determination, reported in Section 4.6.1.

The scoring variable, discussed in Equation 4.13, for the evaluation of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution uncertainties is shown respectively in Figure 4.28 and in 4.29 for the four soft term pile-up correction methods, STVF, EJA, EJAF, JAF. The TST method is not considered here since the TST is more properly a different reconstruction approach for the soft term instead of a pile-up correction method.

The systematic scale and resolution uncertainties introduced by the various pile-up correction methods and determined as described in Equation 4.13 are summarized in Table 4.2. These uncertainties have a fractional contribution to the overall uncertainties presented in Section 4.6.1, which, due to other constraints on the overall $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, may be reduced compared to the uncertainties quoted here.

Systematics uncertainties from instabilities in the jet area method

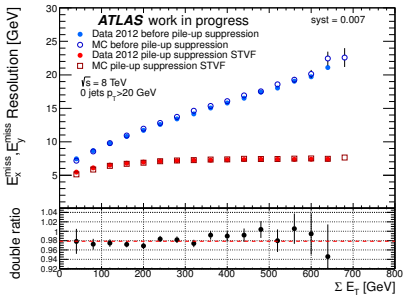
The determination of the median transverse momentum density $\rho_{\mathrm{evt}}^{\mathrm{med}}$ in FastJet uses an "active ghost area" approach, which allows the consistent measurement



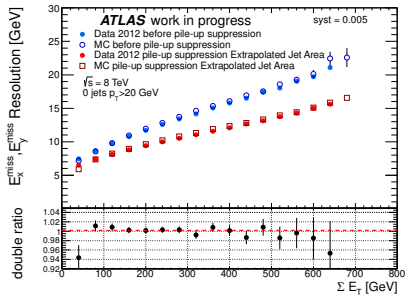
(a) STVF (b) EJA

(c) EJAF (d) JAF

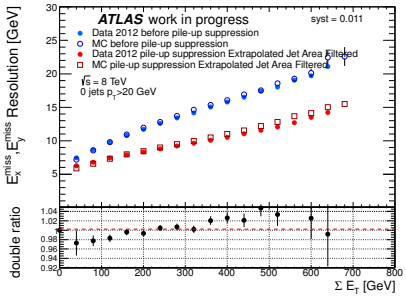
Figure 4.28: Projection of the $E_{\mathrm{T}}^{\mathrm{miss}}$ onto the Z direction as a function of the uncorrected ΣE_{T} in exclusive $Z \to \mu\mu$ events without jets with p_{T} 20 GeV for data and MC simulation before and after applying the STVF method in (a), the EJA method in (b), the EJAF method in (c), and the JAF method in (d). The double ratios $\mathcal{F}_{\mathrm{scale}}(\Sigma E_{\mathrm{T}})$, as described in Equation 4.13, are shown under the curves, respectively.



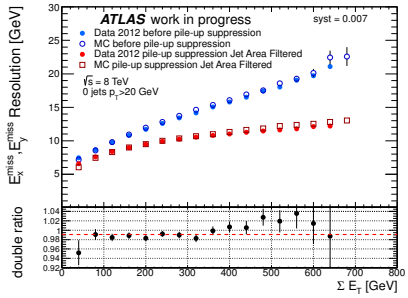
(a) STVF



(b) EJA



(c) EJA



(d) JAF

Figure 4.29: $E_{x(y)}^{\text{miss}}$ resolution as a function of the uncorrected ΣE_{T} in exclusive $Z \to \mu\mu$ events without jets with p_{T} 20 GeV for data and MC simulation before and after applying the STVF method in (a), the EJA method in (b), the EJAF method in (c), and the JAF method in (d). The double ratios $\mathcal{F}_{\text{scale}}(\Sigma E_{\text{T}})$, as described in Equation 4.13, are shown under the curves, respectively.

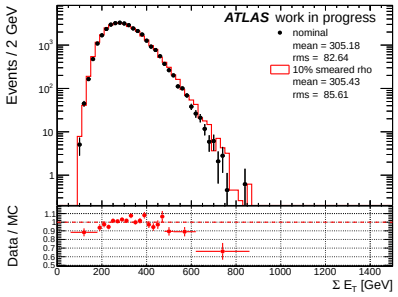
of the jet area for irregular shaped jets as well as regular (cone shaped) jets. The approach uses random seed to cluster "ghost particles" with tiny but finite p_{T} into the jets formed by real particles or detector signals. The jet area determination shows a small instability of at most 10% for very low p_{T} soft term jets in this jet area measurement approach. While the ability to reproduce the $\rho_{\mathrm{evt}}^{\mathrm{med}}$ measurement is improved in ATLAS by controlling the random seeding in sequences of multiple jet area calculations, the effect of a residual Gaussian fluctuation of $\rho_{\mathrm{evt}}^{\mathrm{med}}$ with $\sigma = 10\%\rho_{\mathrm{evt}}^{\mathrm{med}}$ is evaluated for the ΣE_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions for a $Z \to ee$ MC sample. The results are shown in Figure 4.30.

4.7 Conclusion and prospects

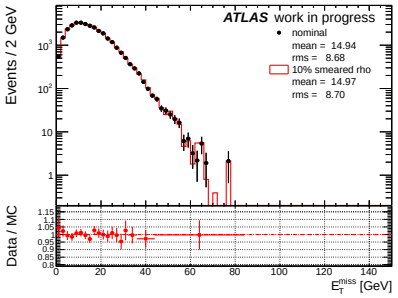
The high level of pile-up reached during 2012 introduced a large amount of additional energy in the ATLAS detector, in particular in the calorimeter system. This increases the fluctuation in the energy measurements with a direct and negative impact to the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, largely worsening the resolution and scale performance. The degradation in performance is mostly driven by the soft energy contributions entering the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ computation. Therefore, several dedicated methods are developed to suppress the pile-up contribution and to restore the good $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance achieved for the low-luminosity data taking.

The most performant methods are the track based methods, STVF and TST, that, exploiting the track association with the identified primary vertex, result in very stable computation. Presently, the STVF method suffers the inefficiency of the used tight track quality selection that causes an underestimation of the soft term scale, while the TST method mostly suffers the inclusion in the computation of mismeasured tracks that can create large tails. These results point out the need of a careful optimization for the track selection.

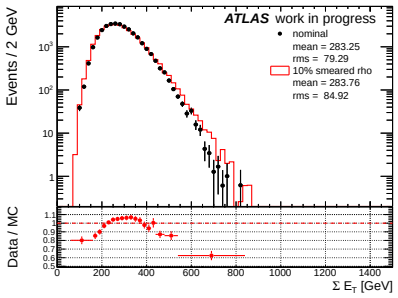
The jet area techniques show impressive improvements in the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance but are not completely able to remove the pile-up contamination. These methods can be promising in combination with track based corrections in order to re-integrate the calorimeter information from soft neutral and forward particles. One of the current limits spotted by these studies is the employment of the same topocluster noise thresholds for both the central and the forward calorimeter regardless to the different geometries and technologies employed in these regions. In particular, it was proved that the use of the current values for the noise thresholds in the forward region suppresseses not only pile-up but also



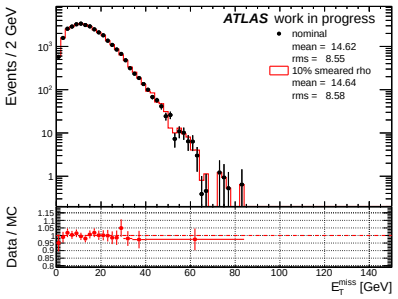
(a) ΣE_{T}, EJA pile-up correction



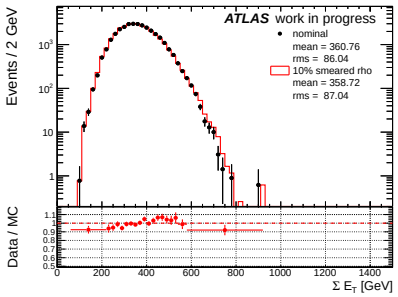
(b) $E_{\mathrm{T}}^{\mathrm{miss}}$, EJA pile-up correction



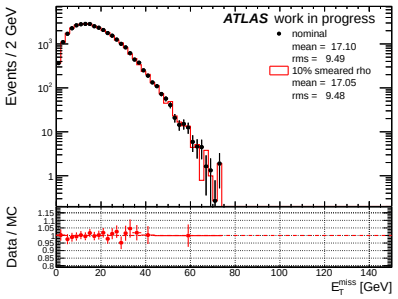
(c) ΣE_{T}, EJAF pile-up correction



(d) $E_{\mathrm{T}}^{\mathrm{miss}}$, EJAF pile-up correction



(e) ΣE_{T}, JAF pile-up correction



(f) $E_{\mathrm{T}}^{\mathrm{miss}}$, JAF pile-up correction

Figure 4.30: The effect of $\rho_{\mathrm{evt}}^{\mathrm{med}}$ fluctuations on the ΣE_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions in $Z \to ee$ MC events for EJA respectively in (a) and (b), for EJAF respectively in (c) and (d) and for JAF respectively in (e) and (f). The variated spectra are obtained with a 10% $\rho_{\mathrm{evt}}^{\mathrm{med}}$ Gaussian smearing. The ratio between the nominal and the variated spectrum is reported in the lower part of the figures.

part of the signal. A re-optimization of the noise topocluster thresholds in the forward region is currently on-going.

In general, all these studies on calorimeter-based and track-based pile-up suppression methods increased our understanding about pile-up effects on primary inputs, on the response of different subdetectors and on clustering algorithms, that is a fundamental starting point to further achieve improvements in performance exploiting the best and most complete information from each subdetector.

Chapter 5

$H \to \tau\tau$ search in the semileptonic channel: cut-based analysis

The search for a SM Higgs boson decaying into a pair of taus, $H \to \tau\tau$, is one of the most relevant channels in the SM Higgs low-mass region: thanks to the high branching ratio and the experimental ability to suppress fake tau leptons, it is the channel with the highest sensitivity for a SM Higgs decaying to fermions. Therefore, after the recent discovery of the Higgs boson in the diboson channels $\gamma\gamma$, ZZ and WW, this search is particularly interesting because it can provide the first direct proof and measurement of the Higgs couplings to fermions.

According to the decay mode of the two τ leptons, the final state can be classified in three different channels: lepton-lepton ($\tau_{\text{lep}}\tau_{\text{lep}}$) if both taus decay into either an electron or a muon, lepton-hadron ($\tau_{\text{lep}}\tau_{\text{had}}$) if one of the two taus decays leptonically and the other one hadronically, hadron-hadron ($\tau_{\text{had}}\tau_{\text{had}}$) if both taus decay into hadrons. Due to the different signal sensitivity and background composition, separate analyses are optimized for each channel. In this chapter, only the $\tau_{\text{lep}}\tau_{\text{had}}$ final state, which contributes for the 46% of the total branching ratio, is considered. Two independent analyses are developed for this search: the first makes use of multivariate techniques (MVA) [143] and the second is based on a cut-based approach [144, 145]. The refined MVA analysis is able to provide a better sensitivity, on the over hand the cut-based analysis, thanks

to its robustness and since it selects a different phase space from the MVA can provide an important cross check for the Higgs discovery in this channel and the mass measurement.

In this chapter, the cut-based analysis is discussed in all its aspects. The main strategy including the event selection and categorization is discussed in Section 5.1, the background composition and the techniques employed for its estimation are described in Section 5.2. In Sections 5.3 and 5.4 the treatment of the main systematics uncertainties and the statistics analysis for the limit derivation are presented. A brief description of the multivariate analysis together with the main results obtained are finally summarized in Section 5.5.

5.1 Analysis strategy

The search for the $H \to \tau\tau$ is a rather complex analysis due to the variety of background sources, therefore, to maximize the analysis sensitivity an accurate selection and categorization procedure is needed.

Firstly general cleaning criteria are applied to ensure a good detector functionally and to reject non-collision events such as cosmic-rays and beam-induced background. In addition, only events with a good reconstructed vertex with at least four tracks with $p_{\mathrm{T}} > 500$ MeV are considered. For the so selected events, the object reconstruction and identifications (see Section 3.2) are performed according the the following criteria to ensure an efficient event reconstruction: *combined* muons, *medium* electrons, *medium* taus, pile-up corrected $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ with the STVF variant (see Section 4.3.2), anti-k_t ($R = 0.4$) jets calibrated LCW+JES and passing the $|\mathrm{JVF}| > 0.5$ requirement to remove residual dependences in the jet multiplicity due to pile-up. In order to solve possible overlap, defined following a geometrical criterium $\Delta R < 0.2$, between the selected objects, only the first object is kept according to the following priority order: μ, e, τ_{had}, jet. For this procedure, the identification for leptons considered for overlap removal with τ_{had} candidates is lowered to *loose* in order to reduce the misidentification of τ_{had} from leptons.

At this level, a preliminary set of criteria are required to increase the purity of the data sample by selecting the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ topology and partially rejecting background processes. Finally, exclusive analysis categories are defined to full exploit the specific experimental signatures of the different Higgs production modes, and additional selection criteria are applied to each category to suppress specific

	$gg \to H \to \tau\tau$	VBF $H \to \tau\tau$	W $H \to \tau\tau$	Z $H \to \tau\tau$
$\sigma \times BR$ [pb]	0.55552503	0.04549584	0.02031456	0.01197365

Table 5.1: Summary of the $\sigma \times BR$ of a SM Higgs boson with $m_H = 125$ GeV decaying in the $\tau_{\text{lep}}\tau_{\text{had}}$ final state for the main Higgs production modes.

background processes.

5.1.1 Signal and background processes

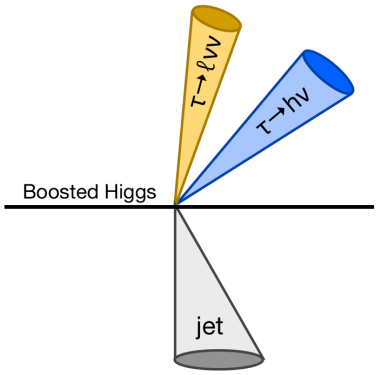
The $\tau_{\text{lep}}\tau_{\text{had}}$ final state is defined as follow. The hadronic τ decay gives in the final states a τ-jet and a neutrino, while the leptonic τ decay gives a lepton (either an electron or a muon) and two neutrinos. The three neutrinos result in a $\boldsymbol{E}_{\text{T}}^{\text{miss}}$ signature for the final state, with the $\boldsymbol{E}_{\text{T}}^{\text{miss}}$ vector mainly expected between the directions of the visible τ decay products and in general more aligned with the lepton since in the leptonic τ decay two neutrinos are produced. Jets can be also present in the final state.

In this section the peculiar jet topologies for the Higgs signal for the main production modes are described together with the main background processes that can mimic these topologies.

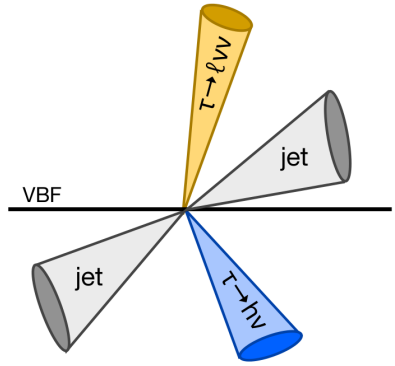
Higgs signal

As discussed in Section 1.4.1, in the SM the Higgs boson is produced mainly via gluon-gluon fusion ($gg \to H$), vector-boson fusion (VBF) or in association with a vector boson $V = Z, W$ (VH). For the different production modes, a brief description of the expected topology (sketched in Figure 5.1) is given in the following. The production cross section at $\sqrt{s} = 8$ TeV multiplied by the branching ratio ($\sigma\times$ BR) for a SM Higgs boson with $m_H = 125$ GeV decaying in the $\tau_{\text{lep}}\tau_{\text{had}}$ final state are summarized in Table 5.1.

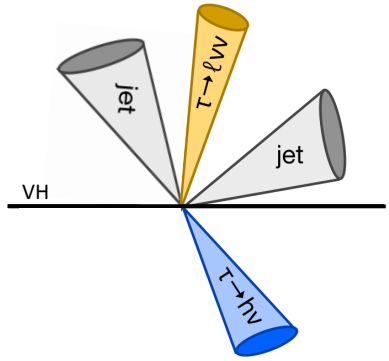
gg $\to$ H $\to \tau\tau$. At leading order, events are approximately back-to-back, resulting in a low $\boldsymbol{E}_{\text{T}}^{\text{miss}}$ measurements. At the next-to-leading oder, through an initial state radiation (ISR), the Higgs boson can be produced in association with jets and receive a boost in the transverse plane. This allows a better separation of the signal from the background processes for two main reasons:



(a) $gg \to H \to \tau\tau$



(b) VBF $H \to \tau\tau$



(c) V $H \to \tau\tau$

Figure 5.1: The expected topology for the $gg \to H \to \tau\tau$ is sketched in (a), for the VBF $H \to \tau\tau$ in (b) and for the V $H \to \tau\tau$ in (c).

- the enhance of the $E_{\mathrm{T}}^{\mathrm{miss}}$ mean value
- the larger jet multiplicity for gg initiated processes, such as $gg \to H$, compared to the qq initiated process such as the main $qq \to Z$ background.

VBF $\mathbf{H \to \tau\tau}$. As already stated, the VBF production provides a very distinct experimental signature of two high-p_{T} jets (already from processes at leading order) with a large separation in pseudorapidity and large dijet invariant mass. The Higgs boson decay products in such events are expected to be found mainly in the central region between the tagging jets, with little additional jet activity.

V $\mathbf{H \to \tau\tau}$. In the VH production mode, the Higgs boson tends to be boosted in the transverse plane with either one or two additional leptons or two jets from the W/Z vector boson decay. In the analysis reported in this thesis the additional production of leptons is not exploited, but a parallel analysis optimized various categories to take full advantage by this particular topology.

Background processes

There is a large variety of background processes which can mimic the signal events:

$Z/\gamma^* \to \tau\tau + jets$ **production:** these events are the major irreducible background since they have the same final state and a similar kinematics as the $H \to \tau\tau$.

$Z/\gamma^* \to \ell\ell + jets$ **production ($\ell = e, \mu$):** these events can contribute as background if either one of the leptons or the jets is misidentified as a τ-jet ($\ell \to \tau_{\mathrm{had}}, jet \to \tau_{\mathrm{had}}$). This is not a large background but especially the $\ell \to \tau_{\mathrm{had}}$ case needs a careful estimation because it tends to peak exactly where a signal for a Higgs boson with a mass of 120-125 GeV is expected.

$W \to \ell\nu + jets$ **production:** these events give a significant background because of the large cross section and the similar event topology to the signal final state with a lepton, genuine $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ and a jet that can fake the τ-jet.

WW, ZW, ZZ **production:** these events contribute as irreducible background if both the vector bosons decay leptonically in the $\tau\tau, e\tau, \mu\tau$ final states. In

case of a vector boson decaying hadronically these events can still contribute as background if one jet is misidentified as a τ-jet.

QCD multijets process: these events are an important source of background because of the large cross section. One of the jet can be misidentified as a τ-jet and at the same time another jet can be misidentified as an electron or a muon. Genuine leptons can also be produced in semileptonic decays of a B or D hadrons.

$t\bar{t}$ production: these events decaying according to $tt \to WbWb$ can produce genuine $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ and leptons (e, μ, τ) if at least one of the W decays leptonically. One of the b-jet or a jet from a hadronic decay of the W can be misidentified as a τ-jet.

single top production (via t- or s-channel or in association with a W): these events decaying according to $t \to Wb$ contribute as background if the W decays leptonically and the τ-jet is either due to a misidentified jet or, for Wt production, comes from the decay of the second W boson.

5.1.2 Data and simulated samples

Data samples

The analysis reported in this thesis is performed on the full 2012 available statistics at $\sqrt{s} = 8$ TeV collected with the ATLAS detector and corresponding to a total integrated luminosity of $\mathcal{L} = 20.3$ fb^{-1}. A combination of the single lepton trigger (SLT) and the lepton+tau trigger (LTT) is used in order to increase the signal yield, since the p_{T} lepton threshold of the LTT is consistently lower than the equivalent threshold for the SLT as summarized in Table 5.2. Therefore, different ATLAS data streams are employed according to the trigger fired by the event: the *Egamma* and the *Muon* streams include respectively events triggered by the electron and muon SLT, while the *JetTauEtmiss* includes events triggered by the LTT.

MC simulation samples

The Higgs signal samples are generated for 11 mass points from $m_H = 100$ GeV to $m_H = 150$ GeV with a step of 5 GeV. The signal simulations are based on the NLO perturbative calculation using Powheg [87, 146, 147] event generator for the gluon fusion [148] and the VBF production modes [149]. The Powheg

	SLT	LTT	
	p_{T}^{ℓ} [GeV]	p_{T}^{τ} [GeV]	p_{T}^{ℓ} [GeV]
electron	24	20	18
muon	24	20	15

Table 5.2: lepton p_{T} thresholds for single lepton trigger and lepton+tau trigger.

gluon fusion samples were generated with the parameter *hfact* set to $m_H/1.2$ in order to better match the Higgs transverse momentum spectrum predicted by the NNLO+NNLL calculations from the HqT program [150, 151]. The parton shower, hadronization and underlying event simulations are interfaced with Pythia [152]. The associated production samples are based on the lowest order and simulated using Pythia.

Concerning the backgrounds samples, the MC samples for $W/Z + jets$ events are generated with Alpgen [153]. This generator employs the MLM matching scheme [154] between the hard process, calculated with leading-order matrix elements for up to five jets, and the parton shower. The $t\bar{t}$ and diboson (WW, WZ, ZZ) samples are produced with McAtNlo [87] with NLO accuracy. Single-top (t/s-channel, Wt) events are simulated with AcerMC [154]. In all the background MC samples the parton shower and the hadronization are simulated with Herwig [155] and the underlying event with Jimmy [156] The loop-induced $gg \to WW$ processes are generated using gg2WW [157]. Tauola [90] and Photos [91] are used to simulate the τ decay and additional photon radiation from charged leptons to fit the data, respectively.

The set of parton distribution functions (PDF) CT10 [158] is used for the Powheg and McAtNlo samples, while CTEQ6L1 [159] is used for the Pythia and Alpgen samples. The detector simulation in all the MC samples is performed with Geant4 [160].

5.1.3 Mass reconstruction

In the process:

$$H \to \tau\tau \to \ell\nu_\ell\nu_\tau\tau_{\mathrm{had}}\nu_\tau \tag{5.1}$$

the presence of multiple neutrinos increases the complexity of the $\tau\tau$ mass reconstruction because the resulting $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is the sum of their contributions and, in

order to compute the invariant mass, the energy and the direction of each neutrino are required. The information can be obtained from a kinematic equations system, assuming that the total $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ comes from the neutrinos only. Nevertheless, the number of unknowns exceeds the number of constraints:

$$\begin{cases} E_x^{\mathrm{miss}} & = p_{\mathrm{mis}_1} \sin\theta_{\mathrm{mis}_1} \cos\phi_{\mathrm{mis}_1} + p_{\mathrm{mis}_2} \sin\theta_{\mathrm{mis}_2} \cos\phi_{\mathrm{mis}_2} \\ E_y^{\mathrm{miss}} & = p_{\mathrm{mis}_1} \sin\theta_{\mathrm{mis}_1} \sin\phi_{\mathrm{mis}_1} + p_{\mathrm{mis}_2} \sin\theta_{\mathrm{mis}_2} \sin\phi_{\mathrm{mis}_2} \\ m_{\tau_1}^2 & = m_{\mathrm{mis}_1}^2 + m_{\mathrm{vis}_1}^2 + 2\sqrt{p_{\mathrm{vis}_1}^2 + m_{\mathrm{vis}_1}^2}\sqrt{p_{\mathrm{mis}_1}^2 + m_{\mathrm{mis}_1}^2} \\ & \quad -2p_{\mathrm{vis}_1} p_{\mathrm{mis}_1} \cos\Delta\theta_{vm_1} \\ m_{\tau_2}^2 & = m_{\mathrm{vis}_2}^2 + 2\sqrt{p_{\mathrm{vis}_2}^2 + m_{\mathrm{vis}_2}^2}\sqrt{p_{\mathrm{mis}_2}^2} - 2p_{\mathrm{vis}_2} p_{\mathrm{mis}_2} \cos\Delta\theta_{vm_2} \end{cases} \tag{5.2}$$

where m_{mis_1} p_{mis_1}, θ_{mis_1}, ϕ_{mis_1} are respectively the invariant mass, the momentum, the polar and the azimuth angle of the $\nu_\tau\nu_\ell$ system from the leptonic τ decay; p_{mis_2}, θ_{mis_2}, ϕ_{mis_2} are the momentum, the polar and the azimuth angle of the ν_τ from the hadronic τ decay; $p_{\mathrm{vis}_{1,2}}$, $\theta_{\mathrm{vis}_{1,2}}$, $\phi_{\mathrm{vis}_{1,2}}$ are the invariant mass, the polar and azimuthal angle for the visible τ decay products (ℓ, τ_{had}); finally, $\Delta\theta_{vm_{1,2}}$ is the angle between the visible and invisible decay products for each of the two τ leptons.

Presently, there are two approaches to face this problem: the collinear approximation and the missing mass calculator (MMC).

$\tau\tau$ mass reconstruction with the collinear approximation

The collinear approximation makes the hypothesis that, being the τ-leptons very boosted ($m_\tau \ll m_H$), the direction of the neutrinos produced in the τ decays is the same as the direction of the visible decay products, and it assumes a null mass for the τ-leptons. Hence the system in Equation 5.2 reduces to:

$$\begin{cases} E_x^{\mathrm{miss}} & = p_{\mathrm{mis}_1} \sin\theta_{\mathrm{vis}_1} \cos\phi_{\mathrm{vis}_1} + p_{\mathrm{mis}_2} \sin\theta_{\mathrm{vis}_2} \cos\phi_{\mathrm{vis}_2} \\ E_y^{\mathrm{miss}} & = p_{\mathrm{mis}_1} \sin\theta_{\mathrm{vis}_1} \sin\phi_{\mathrm{vis}_1} + p_{\mathrm{mis}_2} \sin\theta_{\mathrm{vis}_2} \sin\phi_{\mathrm{vis}_2} \end{cases} \tag{5.3}$$

and the $m_{\tau\tau}$ is computed as:

$$m_{\tau\tau} = \sqrt{2(p_{\mathrm{vis}_1} + p_{\mathrm{mis}_1})(p_{\mathrm{vis}_1} + p_{\mathrm{mis}_2})(1 - \cos\theta)} \tag{5.4}$$

where θ is the angle between the directions of the decay products of the two τ decays.

In this approach the mass reconstruction is not possible when the two taus are back-to-back or the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is poorly reconstructed. In VBF $H \to \tau\tau$ ($Z \to \tau\tau$)

events, the mass reconstruction efficiency decreases from 97(93)% when using the $E_{\mathrm{T}}^{\mathrm{miss,True}}$ to 80(65)% when using the reconstructed $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ before pile-up suppression. It is $\sim$ 80(55)% when using the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ after pile-up suppression with STVF and $\sim$ 80(60)% when using the Jet Area pile-up suppression. The efficiency of the $Z \to \tau\tau$ mass reconstruction is smaller even using the $E_{\mathrm{T}}^{\mathrm{miss,True}}$ because the two τ-leptons are more back-to-back (the Z is less boosted than the H produced through VBF) and when using the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ before pile-up suppression because the $Z \to \tau\tau$ events have a smaller $E_{\mathrm{T}}^{\mathrm{miss}}$. It decreases after the pile-up suppression because the $E_{\mathrm{T}}^{\mathrm{miss}}$ becomes very small, mainly in the events with no jets, so the probability to have a negative solution for one or both neutrino momenta increases. These differences in the reconstruction efficiency after the pile-up suppression in the two samples improve the signal significance.

The mass so reconstructed is extremely sensitive to the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ evaluation: the peak position depend on the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ scale and the mass resolution is completely dominated by the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution. Hence it is an optimal variable for testing the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance. Some studies are shown in Section 5.1.7.

$\tau\tau$ mass reconstruction with MMC

The MMC approach [161] does not assume the strict collinearity of the visible and invisible τ decay products of the previous method. The tau is assumed massive with $m_\tau = 1.777$ GeV and, in order to solve the equation system reported in Equation 5.2, a scan in a grid of points for the$(\phi_{\mathrm{mis}_1}, \phi_{\mathrm{mis}_2})$ parameter space is performed. Then, an event weight is assigned to the mass evaluated in each point to enhance more probable solutions. This information is derived exploiting additional knowledge about the τ decay kinematics, in particular, the 3-dimensional angle $\Delta\theta_{3D}$ between the directions of visible and invisible τ decay products. This distribution is fitted with a linear combination of Gaussian and Landau functions and parametrized as a function of the initial tau mometum p_τ, $\mathcal{P}(\Delta\theta, p_\tau)$. Figure 5.2 shows the $\Delta\theta_{3D}$ distribution for different τ decay modes. The event probability is thus obtain as:

$$\mathcal{P}_{\mathrm{event}} = \mathcal{P}(\Delta\theta_1, p_{\tau_1}) \times \mathcal{P}(\Delta\theta_2, p_{\tau_2}) \tag{5.5}$$

As explained in the previous section the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution strongly affects the mass resolution. In order to mitigate this effect, the MMC increases the dimensionality of the scan parameter space including also the E_x^{miss} and E_y^{miss}

components and varying them inside their experimental resolution. The event probability thus becomes:

$$\mathcal{P}_{\text{event}} = \mathcal{P}(\Delta\theta_1, p_{\tau_1}) \times \mathcal{P}(\Delta\theta_2, p_{\tau_2}) \times \mathcal{P}(\Delta E_x^{\text{miss}}) \times \mathcal{P}(\Delta E_y^{\text{miss}}) \tag{5.6}$$

where the probability functions $\mathcal{P}(\Delta E_{x(y)}^{\text{miss}})$ are defined as:

$$\mathcal{P}(\Delta E_{x(y)}^{\text{miss}}) = \exp\left(-\frac{(\Delta E_{x(y)}^{\text{miss}})^2}{2\sigma^2}\right) \tag{5.7}$$

where σ is the $E_{\text{T}}^{\text{miss}}$ resolution and $\Delta E_{x(y)}^{\text{miss}}$ are the differences between the measured values for $E_{x(y)}^{\text{miss}}$ and the values in the parameter space during the scan.

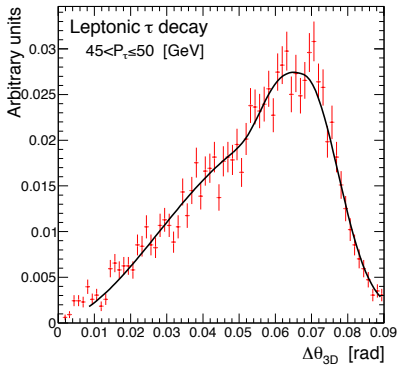
According to this procedure, a very high efficiency ($\sim$ 99%) is achieved for signal and $Z \to \tau\tau$ events, while a low efficiency is observed for background samples, providing an extra handle against them. The reason why MMC rejects $\sim$ 1% of signal events is related to the fact that the $\Delta\phi$ of visible and invisible τ decay products can be outside the scan range, or that $E_{\text{T}}^{\text{miss}}$ fluctuates more than 3σ in resolution.

5.1.4 Blind prescription

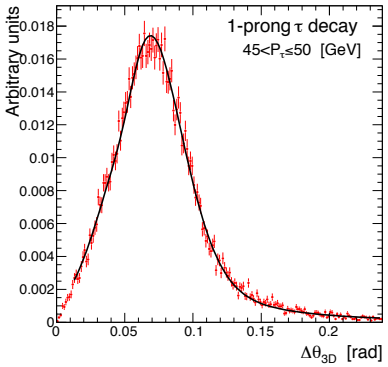
In order to not bias the results during the optimization of the analysis, a blind prescription is applied on the dataset under use. The data can be looked at in the background control regions, where the signal contamination is expected to be negligible. Since the small separation between the $Z \to \tau\tau$ background and the expected Higgs signal, the $Z \to \tau\tau$ control region, described in Section 5.2.1, has some signal contamination but anyway with a signal efficiency below 20%. On this basis, also for this control region, all distributions can be safely examined also for data. Regarding signal regions, all kinematics distributions, except the MMC mass, can be safely compared between data and expectations since all these variables have a low discriminating power for the Higgs signal.

5.1.5 Preselection

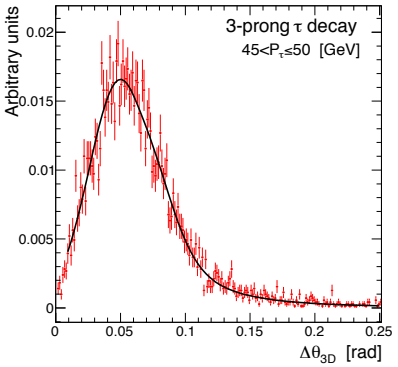
In order to select the $\tau_{\text{lep}}\tau_{\text{had}}$ topology previously discussed, the following criteria are required:



(a) leptonic decay



(b) 1-prong hadronic deacay



(c) 3-prong hadronic deacay

Figure 5.2: Example of the probability distribution functions $\mathcal{P}(\Delta\theta, p_\tau)$ for $45 < p_\tau \leq 50$ GeV and for leptonic τ decays in (a), for 1-prong hadronic decay in (b) and for 3-prong decay in (c) [161].

	p_{T}^{τ} [GeV]	p_{T}^{ℓ} [GeV]
SLT (e and μ)	20	26
LTT (e)	25	20-26
LTT (μ)	25	17-26

Table 5.3: Offline p_{T} thresholds for τ, μ and e.

τ-lepton: exactly one τ-jet is required. The values of the p_{T} threshold on the τ-lepton change according to the trigger passed by the event and are summarized in Table 5.3.

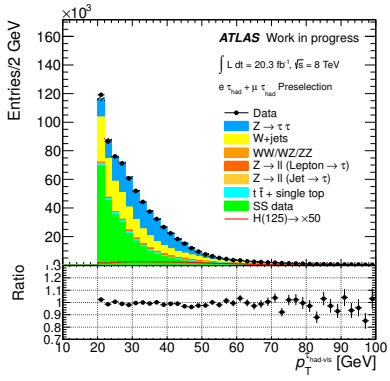
e-, μ-lepton: exactly one electron or muon are required. The values of the p_{T} threshold on the light lepton change according to the trigger passed by the event and are summarized in Table 5.3. Events with more than one lepton are vetoed to reject the $Z \to \ell\ell$ and top processes.

charge correlation: the charges of the light lepton and the τ_{had} are required to have opposite sign.

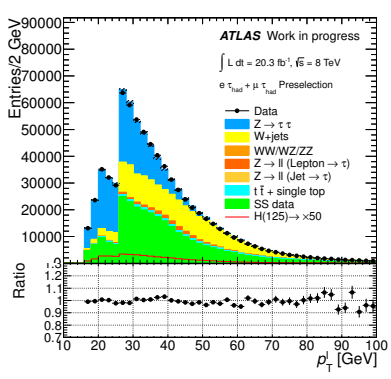
Figure 5.3 shows basic kinematics distributions which will be used in the analysis. The background processes are estimated as described in Section 5.2, they look well modeled and in agreements with the observed data. The MMC mass in Figure 5.3(e) shows a systematic trend in the region $80 < m_{\tau\tau}^{\mathrm{MMC}} < 150$ GeV, where the $Z \to \tau\tau$ contribution is the dominant. This can be attributed to a shift of about 1.6% of the MC based TES employed in this analysis with respect to the in-situ measurement, resulting in a small overestimation of the $Z \to \tau\tau$ background. As shown in Figure 5.3(f) this effect is covered by the systematic uncertainties assigned to the TES, in particular for a downwards variation of 1σ a good agreement is found between data and expectations.

5.1.6 Analysis categories

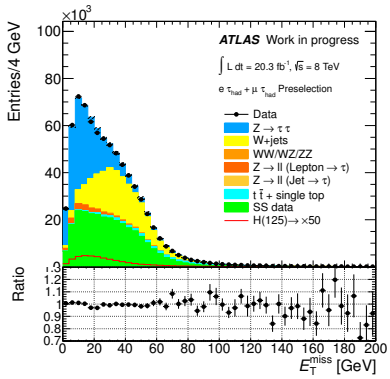
As discussed in Section 5.1.1 the different Higgs production modes, in particular the VBF, offer peculiar jet signatures that can be exploited to optimize separeted categories to increase the overall signal sensitivity of the analysis. Each category aims to target a specific jet topology and it is characterized by a different background composition and signal sensitivity, therefore, after the categorization cri-



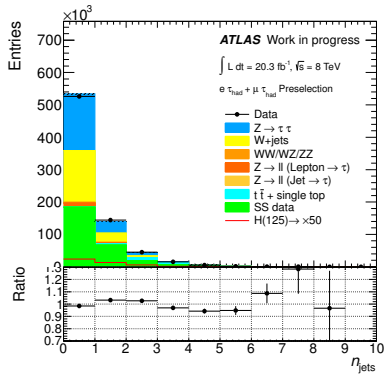
(a) p_{T}^{τ}



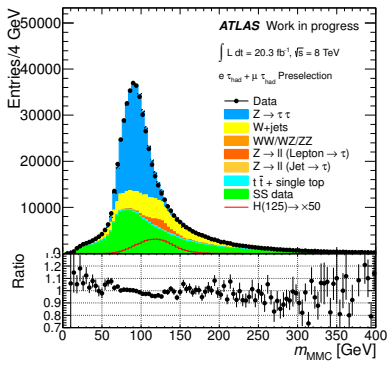
(b) p_{T}^{ℓ}



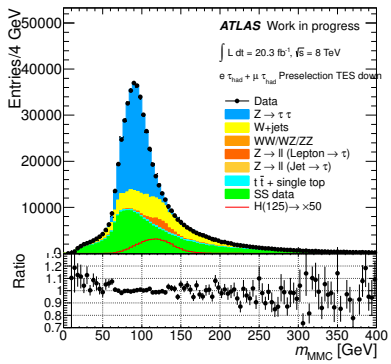
(c) $E_{\mathrm{T}}^{\mathrm{miss}}$



(d) n_{jets}



(e) MMC mass (nominal)



(f) MMC mass (1σ TES variation)

Figure 5.3: Distributions at the preselection level: p_{T}^{τ} in (a), p_{T}^{ℓ} in (b), $E_{\mathrm{T}}^{\mathrm{miss}}$ in (c), n_{jets} in (d), MMC mass in (e) and MMC mass with 1σ TES variation in (f).

teria required to classified the events, additional selection criteria are optimized for rejecting specific backgrounds. A full description is given in the following.

Common variables for non-resonant background rejection

A set of common variables can be useful across different categories to reject non-resonant backgrounds with a fake tau, such as QCD, diboson, $W + jets$, $t\bar{t}$ and single top processes. These variables are defined as follow:

- The transverse mass of the lepton and the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ system defined as:

$$m_{\mathrm{T}} = \sqrt{2p_{\mathrm{T}}^{\ell} E_{\mathrm{T}}^{\mathrm{miss}}(1 - \cos\Delta\phi)} \tag{5.8}$$

  where p_{T}^{ℓ} is the lepton p_{T} and $\Delta\phi$ is the angle between the lepton and the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$. This variable is an efficient discriminant between W and $t\bar{t}$ events, that tend to have large values for m_{T}, and the signal that tends to have low value of m_{T} (since the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ vector usually points between the visible τ-lepton decay products).

- The $\Sigma\Delta\phi$ variable defined as:

$$\Sigma\Delta\phi = |\phi^{\ell} - \phi^{\mathrm{miss}}| + |\phi^{\tau} - \phi^{\mathrm{miss}}| \tag{5.9}$$

  where $\phi^{\ell}, \phi^{\tau}, \phi^{\mathrm{miss}}$ are respectively the azimuthal directions of the light lepton, the τ-lepton and the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$. As the m_{T}, also this variable provides rejection against the W background, since signal events usually have the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ pointing between the visible τ-lepton decays products resulting in $\Sigma\Delta\phi < \pi$ while this is not general true for the W background.

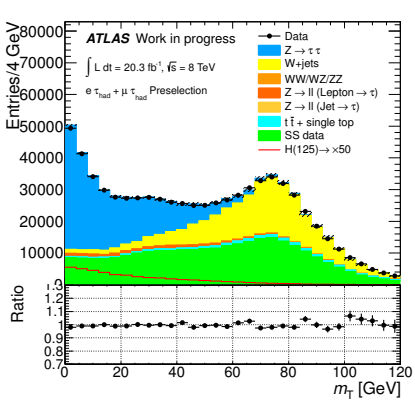
- The $\Delta\Delta R$ variable defined as:

$$\Delta\Delta R = |\Delta R^{\ell\tau} - \Delta R_{\mathrm{pred}}^{\ell\tau}| \tag{5.10}$$

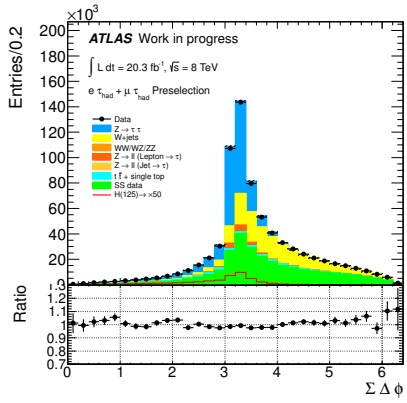
  where $\Delta R^{\ell\tau}$, $\Delta R_{\mathrm{pred}}^{\ell\tau}$ are respectively the measured angular separation between the lepton and the τ_{had} and its predicted value as a function of $p_{\mathrm{T}}^{\ell\tau}$ of the lepton-τ system. The predicted value is computed parametrizing with a Landau function the correlation between $\Delta R^{\ell\tau}$ and the $p_{\mathrm{T}}^{\ell\tau}$ from simulated Higgs events.

- The b-veto (70% efficiency point) is applied to reject the $t\bar{t}$ background.

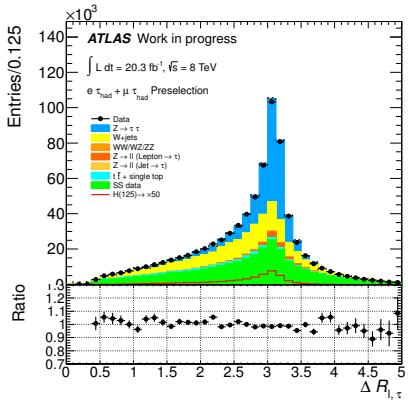
  Figure 5.4 shows a good agreement between data and expectations for the variables described above at the preselection level.



(a) m_{T}



(b) $\Sigma\Delta\phi$



(c) $\Delta R^{\ell\tau}$

Figure 5.4: Distributions of discriminating variables at the preselection level: m_{T} in (a), $\Sigma\Delta\phi$ in (b) and $\Delta R^{\ell\tau}$ in (c).

Definition of the analysis categories

Two mutually exclusive categories are defined according to the jet multiplicity and topology: the VBF category and the Boosted category. The categorization procedure starts with the requirement for an even to pass the VBF classification criteria, if the event is discarded it can be recycled by the Boosted category but it cannot be classified in both categories at the same time, so avoiding double counting. Then, additional selection criteria are optimized for both categories based on the discriminating variables presented in the previous section to mitigate specific background contributions. The categorization requirements for the VBF and Boosted category are defined as follow:

VBF category aims to select the experimental VBF production modes, described in Section 5.1.1, asking for:

- at least two high-p_{T} jets:
  - in opposite halves of the detector and with a large pseudo rapidity separation $\Delta\eta_{j_1 j_2}$
  - with a large invariant mass $m_{j_1 j_2}$
- tau decay products between the two tagged jets
- small total transverse momentum of the full system, $p_{\mathrm{T}}^{\mathrm{tot}}$, defined as:

$$p_{\mathrm{T}}^{\mathrm{tot}} = \mid \mathbf{p}_{\mathrm{T}}^{\ell} + \mathbf{p}_{\mathrm{T}}^{\tau} + \mathbf{p}_{\mathrm{T}}^{j_1} + \mathbf{p}_{\mathrm{T}}^{j_2} + \mathbf{E}_{\mathrm{T}}^{\mathrm{miss}} \mid \tag{5.11}$$

- only events that pass the SLT[1]

All the optimal values for the selections described above are summarized together with the background rejection criteria in Table 5.4.

Boosted category aims to select Higgs boson candidates with a significant boost in the transverse plane and balanced by one or more jets, in particular the $gg \to H \to \tau\tau$ with ISR described in Section 5.1.1, asking for:

- high reconstructed Higgs transverse momentum, p_{T}^{H}, defined as the vector sum of its expected decay products in the hypothesis that the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ comes from the neutrinos contribution only:

$$p_{\mathrm{T}}^{H} = \mid \mathbf{p}_{\mathrm{T}}^{\ell} + \mathbf{p}_{\mathrm{T}}^{\tau} + \mathbf{E}_{\mathrm{T}}^{\mathrm{miss}} \mid \tag{5.12}$$

[1] The LTT requirements on the τ_{had} at trigger level completely deplete the corresponding region for background estimation (anti-τ region), explained and defined in Section 5.2.3, therefore it is impossible to obtain a reliable and accurate estimate of the background shape.

This variable is particularly attractive, because it allows to classify boosted events without explicitly relying on the jet multiplicity avoiding a direct dependence on the jet energy scale, resolution and modeling. Moreover it helps to suppress $t\bar{t}$ and QCD multijets non-resonant backgrounds.

– requirements on the energy fraction of the visible τ decay products defined as:

$$x_\ell = \frac{p_x^{\tau_{\text{had}}} \cdot p_y^\ell - p_y^{\tau_{\text{had}}} \cdot p_x^\ell}{p_x^{\tau_{\text{had}}} \cdot E_y^{\text{miss}} - p_y^{\tau_{\text{had}}} \cdot E_x^{\text{miss}} + p_x^{\tau_{\text{had}}} \cdot p_y^\ell - p_y^{\tau_{\text{had}}} \cdot p_x^\ell} \tag{5.13}$$

$$x_{\tau_{\text{had}}} = \frac{p_x^{\tau_{\text{had}}} \cdot p_y^\ell - p_y^{\tau_{\text{had}}} \cdot p_x^\ell}{p_y^{\tau_{\text{had}}} \cdot E_x^{\text{miss}} - p_x^{\tau_{\text{had}}} \cdot E_y^{\text{miss}} + p_x^{\tau_{\text{had}}} \cdot p_y^\ell - p_y^{\tau_{\text{had}}} \cdot p_x^\ell} \tag{5.14}$$

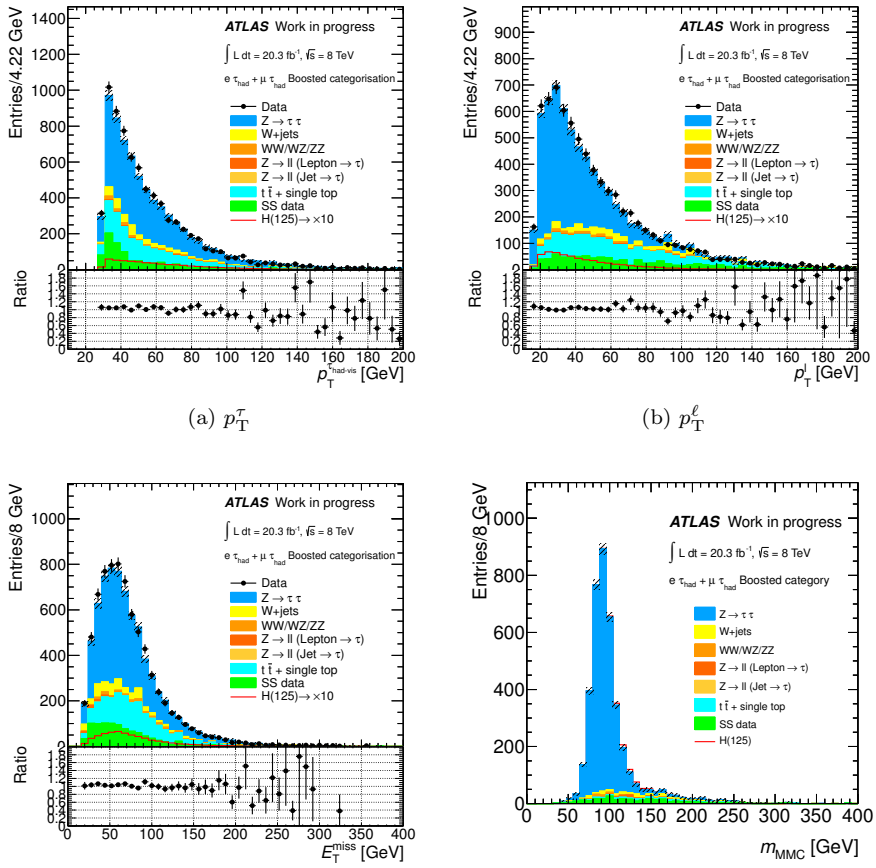
computed under the hypothesis of the collinear approximation that assumes collinear direction for visible and invisible (neutrinos) τ decay products and that the $\boldsymbol{E}_{\text{T}}^{\text{miss}}$ come from neutrinos only. This variable is particularly efficient to enhance the signal contribution over non-resonant backgrounds.

All the optimal values for the selections described above are summarized together with the background rejection criteria in Table 5.4.

Figures 5.5 and 5.6 show an acceptable agreement between data and expectations for kinematics distribution at the level of the Boosted and VBF categorization respectively. For the MMC mass distribution data are not shown according to the blind prescription described in Section 5.1.4.

In the analysis presented in Ref. [144] also the additional 1Jet and 0Jet categories were included asking respectively for the presence in the final state of one and zero jet with $p_{\text{T}} > 30$ GeV. It was observed that these categories add just a small contribution to the total expected sensitivity, their main benefit relying on the constrains of some nuisance parameters. Therefore, studies are on-going either to convert these categories in more proper control regions or to re-optimize them.

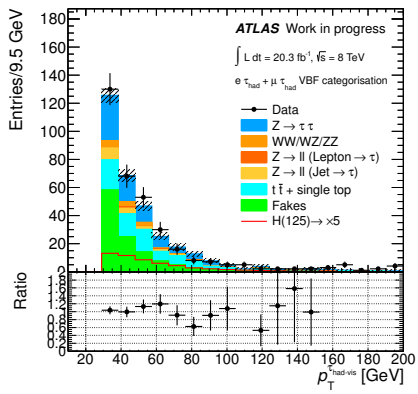
[1] The optimal value was found to be at $p_{\text{T}}^H > 150$ GeV, but the theoretical uncertainties on signal production cross section cannot be reliably estimated at such value of p_{T}^H, so the the cut is lowered to 100 GeV.



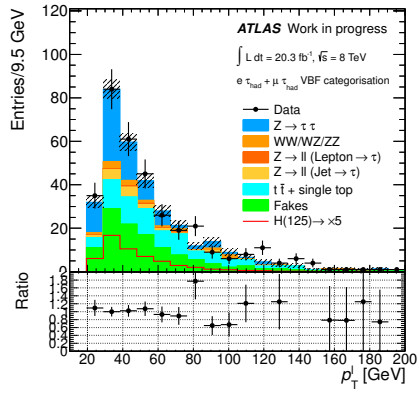
(a) p_{T}^{τ} (b) p_{T}^{ℓ}

(c) $E_{\mathrm{T}}^{\mathrm{miss}}$ (d) MMC mass

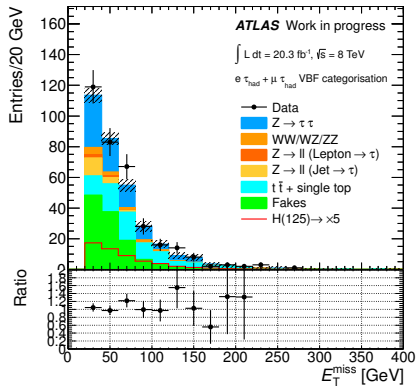
Figure 5.5: Kinematics variables distribution at the Boosted categorization level: p_{T}^{τ} in (a), p_{T}^{ℓ} in (b), $E_{\mathrm{T}}^{\mathrm{miss}}$ in (c), MMC mass in (d).



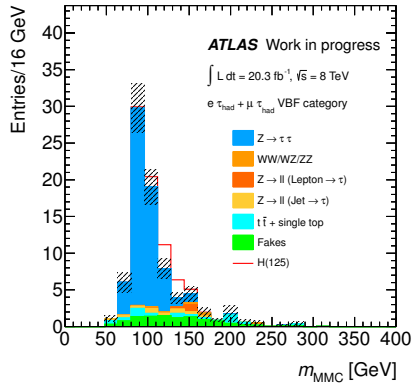
(a) p_{T}^{τ}



(b) p_{T}^{ℓ}



(c) $E_{\mathrm{T}}^{\mathrm{miss}}$

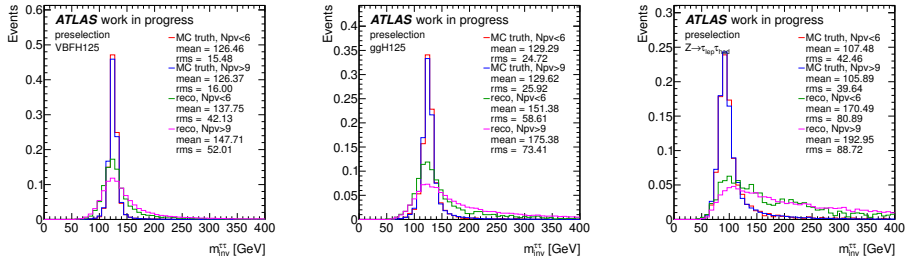


(d) MMC mass

Figure 5.6: Kinematics variables distribution at the VBF categorization level: p_{T}^{τ} in (a), p_{T}^{ℓ} in (b), $E_{\mathrm{T}}^{\mathrm{miss}}$ in (c), MMC mass in (d).

	VBF	Boosted
Categorization criteria	$E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV $p_{\mathrm{T}}^{\tau} > 30$ GeV $p_{\mathrm{T}}^{\mathrm{tot}} < 30$ GeV $m_{\mathrm{vis}} > 40$ GeV $p_{\mathrm{T}}^{j_1} > 40$ GeV $p_{\mathrm{T}}^{j_2} > 30$ GeV $\eta_{j_1} \times \eta_{j_2} < 0$ $\Delta\eta_{j_1 j_2} > 3.0$ $m_{j_1 j_2} > 500$ GeV $\min(\eta_{j_1}, \eta_{j_2}) < \eta_{\ell}$ $\max(\eta_{j_1}, \eta_{j_2}) > \eta_{\tau}$	$E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV $p_{\mathrm{T}}^{\tau} > 30$ GeV $p_{\mathrm{T}}^{H} > 100$ GeV [1] $0 < x_{\ell} < 1$ $0.2 < x_{\tau_{\mathrm{had}}} < 1.2$
Selection criteria	$m_{\mathrm{T}} < 50$ GeV $\Delta\Delta R < 0.8$ b-jet veto $\Sigma\Delta\phi < 2.8$	$m_{\mathrm{T}} < 50$ GeV $\Delta\Delta R < 0.8$ b-jet veto

Table 5.4: Categorization and selection criteria to define the VBF and the Boosted categories.



(a) VBFH (b) $gg \to$H (c) $Z \to \tau\tau$

Figure 5.7: Comparison at the preselection level of the invariant mass computed with the MC truth $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ and with the reconstructed $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ in the low pile-up conditions ($N_{\mathrm{PV}} < 6$) and high pile-up conditions ($N_{\mathrm{PV}} > 9$) for the signal vector boson fusion events in (a) and gluon gluon fusion events in (b) and for the main background $Z \to \tau\tau$ events in (c).

5.1.7 $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ studies in the $H \to \tau\tau$ analysis

As explained in Section 5.1.1, the $Z \to \tau\tau$ process is the main background for this analysis because it has an event kinematics very similar to the signal Higgs events and a mass peak close to an expected light Higgs mass around 125 GeV. A good mass resolution is therefore crucial to distinguish signal events from background events increasing the analysis sensitivity. Figure 5.7 shows the effect of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution on the $\tau\tau$ invariant mass reconstructed with the collinear approximation (see Section 5.1.3): as it is clearly visible the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution completely dominates the mass reconstruction. Hence, providing the best $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ evaluation as possible, especially in term of resolution performance, is essential for this analysis.

As extensively shown in Chapter 4, the pile-up largely worsens the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution, with a subsequent effect on the mass resolution, as shown in Figure 5.7 which compares the reconstruction in low pile-up conditions ($N_{\mathrm{PV}} < 6$) with the reconstruction in high pile-up conditions ($N_{\mathrm{PV}} > 9$). Improvements in the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution are expected for Z and Higgs events by the use of the pile-up suppression methods described in details in Section 4.2 and shown in Figure 4.16. In the following, the pile-up suppression methods for the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ are tested in the specific context of the $H \to \tau\tau$ analysis at the preselection and categorization level. For this purpose, the figures of merit considered are:

$\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ **resolution:** estimated from the distribution $E_{x(y)}^{\mathrm{miss}} - E_{x(y)}^{\mathrm{miss,True}}$. Beyond being crucial for the mass reconstruction, the resolution performance is also important for the background rejection. Indeed the rejection efficiency of the requirements on $E_{\mathrm{T}}^{\mathrm{miss}}$ and on variables computed using the $E_{\mathrm{T}}^{\mathrm{miss}}$ (like the transverse mass), depends on the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution.

$\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ **linearity:** estimated from the distribution $E_{\mathrm{T}}^{\mathrm{miss}} - E_{\mathrm{T}}^{\mathrm{miss,True}}$. A good $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ evaluation ensures the proper reconstruction also of more complex quantities, in particular of the invariant mass. Though the MMC reconstruction partially compensates for $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ effects, providing a good $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ measurement is still essential because it would improve the final mass evaluation.

$\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ **direction measurement:** estimated from the distribution $\phi^{\mathrm{miss}} - \phi^{\mathrm{miss,True}}$. The $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction enters in the invariant and transverse mass computation. Moreover it can also affect categorization variables like $p_{\mathrm{T}}^{\mathrm{tot}}$, where a vector sum is involved, or x_{ℓ}, $x_{\tau_{\mathrm{had}}}$ variables, where the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ is projected onto visible τ decay products.

$m_{\tau\tau}$ **reconstruction:** computed as described in Section 5.1.3. The MMC mass is the final discriminant for this analysis. However since the complexity of the MMC procedure[2], the studies in this section are performed on the invariant mass reconstructed with the collinear approximation. The effect of the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ on this variable are direct and well visible, thus it provides an optimal quantity for the performance studies.

Figures 5.8-5.10 show the comparison at the preselection level before and after applying a $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ pile-up suppression methods for the testing variables described above, respectively for the vector boson fusion and the gluon gluon fusion Higgs events and for $Z \to \tau\tau$ events. For all the distributions a clear improvement is observed for the pile-up corrected quantities, in particular for the STVF.

Figures 5.11-5.15 show the same comparison at the level of the Boosted and VBF categorization level, dropping the requirements involving $E_{\mathrm{T}}^{\mathrm{miss}}$ in order to not bias the distributions for the methods under test. The same conclusion can be drawn. After applying all the selection criteria of the VBF category, the gluon gluon fusion signal and the $Z \to \tau\tau$ samples suffer of a low statistics,

[2] In order to have a fair comparison, the MMC scanning and weighting procedure should be reoptimized for each of the $E_{\mathrm{T}}^{\mathrm{miss}}$ pile-up suppression method.

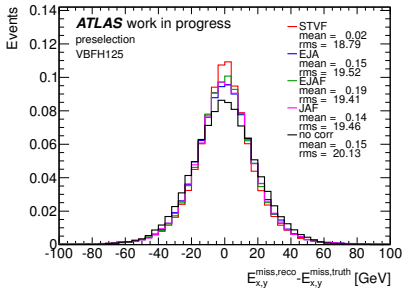
therefore in order to benefits of a larger statistics for the performance studies the requirement on the invariant mass of the jet system, $m_{j_1 j_2}$, is released (denoted as "VBF category loose" in the figures). Anyway, since these distributions provide a less significant result due to the limited statistics, only the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution distribution and the invariant mass distribution are shown in Figure 5.15.

In Section 4.5.3, an underestimation of the soft term contribution was observed for the STVF method. This can have an effect on the reconstructed mass, therefore, the MMC mass is tested as a function of the expected truth Higgs mass, as shown in Figure 5.16 for VBF and Boosted categories. A good match between the reconstructed and the expected values is found for both the categories except for a constant small underevaluation of about 1.5 GeV observed for the gluon gluon fusion Higgs events in the Boosted category. An optimization of the pile-up suppression methods, can hopefully correct for this effect. Anyway for the purpose of a first observation of the Higgs boson in the $\tau\tau$ channel this small effect is not crucial. For this reason, the STVF method that provides the best resolution is preferred and employed. The STVF method also provides a better stability of the reconstruction performance as a function of N_{PV} and by consequence as a function of the data taking time, as visible in Figure 5.17 which shows the invariant mass reconstructed with the collinear approximation as a function of N_{PV} before and after applying the STVF pile-up suppression method. The stability requirement is important since the analysis selection and optimization is performed on the whole datasets without dividing for periods.

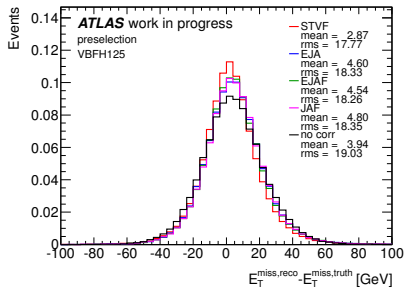
5.2 Background estimation

The selection criteria described above are aimed to select signal events and to suppress backgrounds. Anyways estimation procedures for the residual backgrounds are needed. The are described in detail in this section. The main backgrounds are evaluated as much as possible with data driven techniques or if either the MC is employed it is then normalized to data in dedicated control regions. In particular for the main $Z \to \tau\tau$ background a hybrid semi-data driven technique is employed and described in Section 5.2.1. Only for the small diboson background[3] a full MC based estimation is performed.

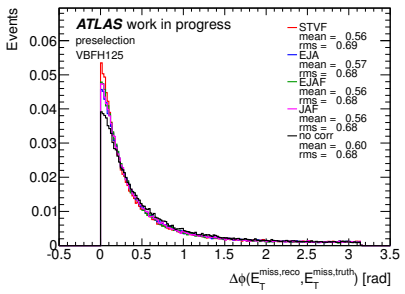
[3] After preselection it contributes only for the 0.5%.



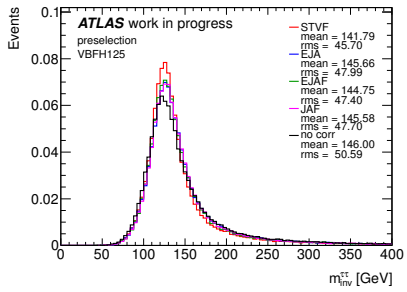
(a) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution



(b) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ linearity

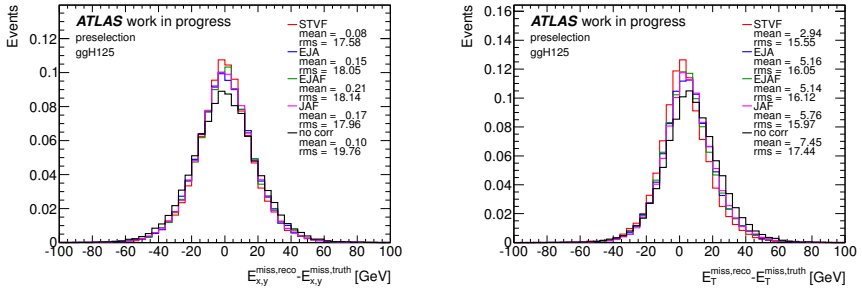


(c) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction



(d) $m_{\tau\tau}$

Figure 5.8: Comparison at the preselection level before and after applying the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ pile-up suppression methods in MC vector boson fusion Higgs events at $m_H = 125$ GeV: $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution in (a), $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ linearity in (b), $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction in (c), and reconstructed invariant mass in (d).



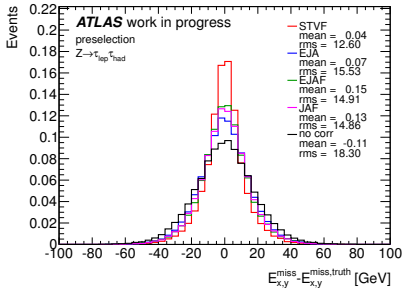
(a) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution

(b) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ linearity

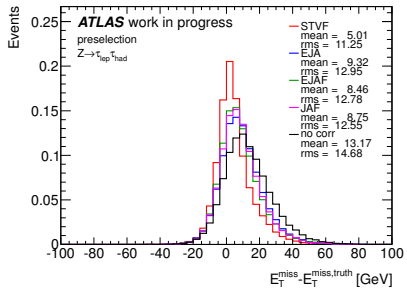
(c) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction

(d) $m_{\tau\tau}$

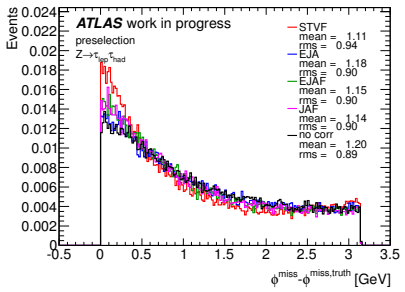
Figure 5.9: Comparison at the preselection level before and after applying the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ pile-up suppression methods in MC gluon gluon fusion Higgs events at $m_H = 125$ GeV: $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution in (a), $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ linearity in (b), $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction in (c), and reconstructed invariant mass in (d).



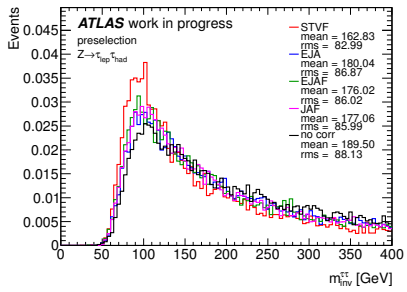
(a) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution



(b) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ linearity

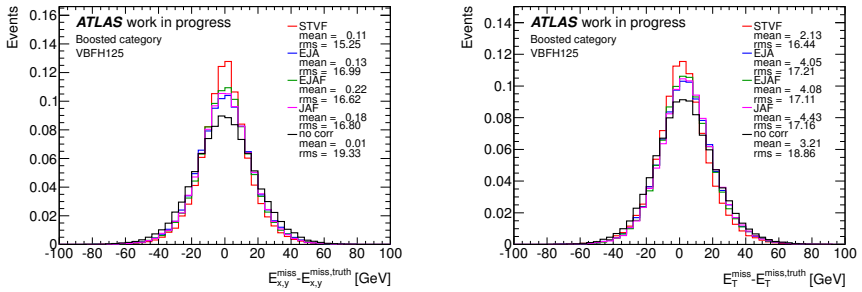


(c) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction



(d) $m_{\tau\tau}$

Figure 5.10: Comparison at the preselection level before and after applying the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ pile-up suppression methods in MC $Z \to \tau\tau$ events: $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution in (a), $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ linearity in (b), $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction in (c), and reconstructed invariant mass in (d).



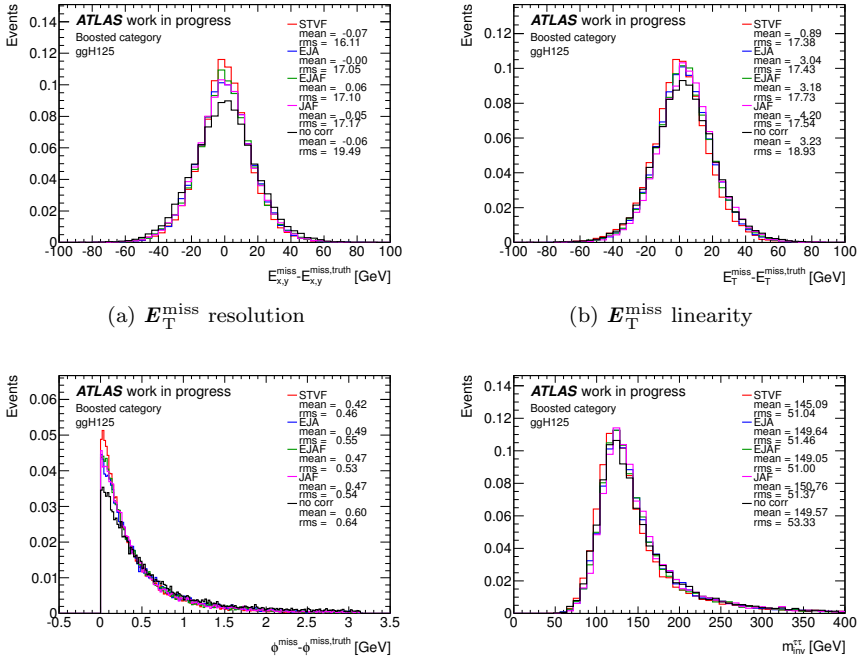
(a) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution

(b) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ linearity

(c) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction

(d) $m_{\tau\tau}$

Figure 5.11: Comparison at the Boosted categorization level before and after applying the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ pile-up suppression methods in MC vector boson fusion Higgs events at $m_H = 125$ GeV: $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution in (a), $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ linearity in (b), $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction in (c), and reconstructed invariant mass in (d).



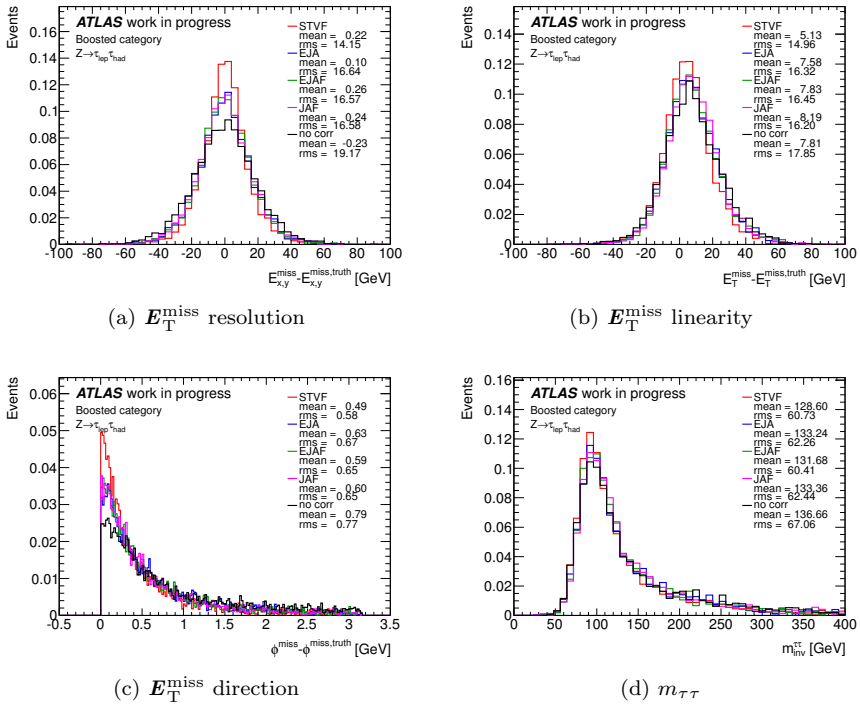
(a) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution

(b) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ linearity

(c) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction

(d) $m_{\tau\tau}$

Figure 5.12: Comparison at the Boosted categorization level before and after applying the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ pile-up suppression methods in MC gluon gluon fusion Higgs events at $m_H = 125$ GeV: $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution in (a), $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ linearity in (b), $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction in (c), and reconstructed invariant mass in (d).



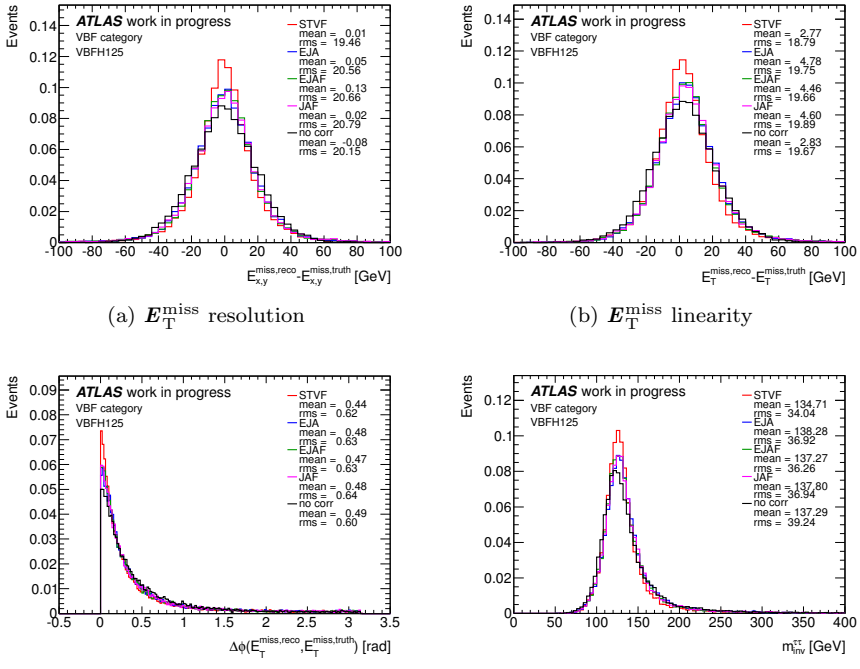
(a) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution

(b) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ linearity

(c) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction

(d) $m_{\tau\tau}$

Figure 5.13: Comparison at the Boosted categorization level before and after applying the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ pile-up suppression methods in MC $Z \to \tau\tau$ events: $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution in (a), $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ linearity in (b), $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction in (c), and reconstructe invariant mass in (d).



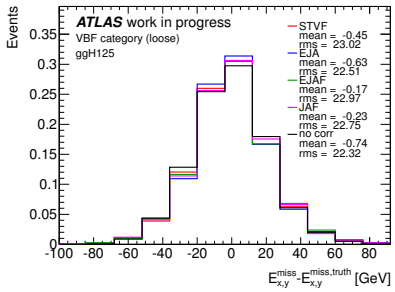
(a) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution

(b) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ linearity

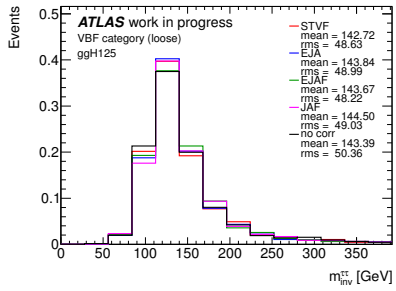
(c) $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction

(d) $m_{\tau\tau}$

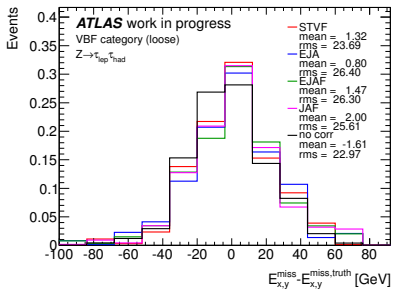
Figure 5.14: Comparison at the VBF categorization level before and after applying the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ pile-up suppression methods in MC vector boson fusion Higgs events at $m_H = 125$ GeV: $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution in (a), $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ linearity in (b), $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ direction in (c), and reconstructed invariant mass in (d).



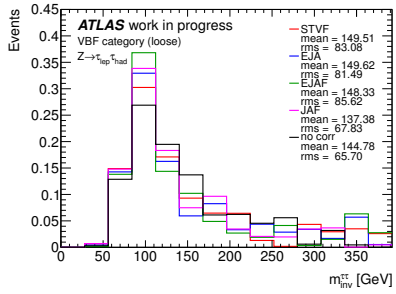
(a) $gg \to$H, $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution



(b) $gg \to$H, $m_{\tau\tau}$

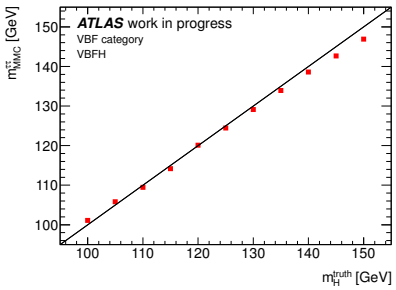


(c) $Z \to \tau\tau$, $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution

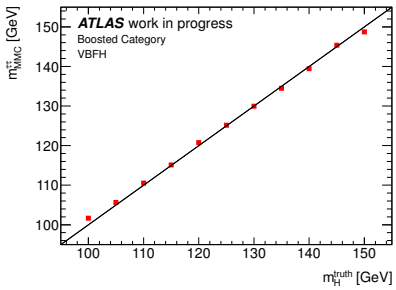


(d) $Z \to \tau\tau$, $m_{\tau\tau}$

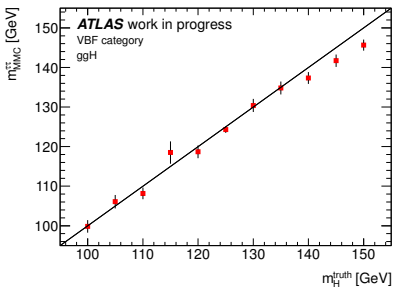
Figure 5.15: Comparison at the VBF categorization level before and after applying the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ pile-up suppression methods in MC gluon gluon fusion Higgs events at $m_H = 125$ GeV: $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution in (a) and reconstructed invariant mass in (b). The same comparison is shown in MC $Z \to \tau\tau$ events: $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution in (c) and reconstructed mass in (d).



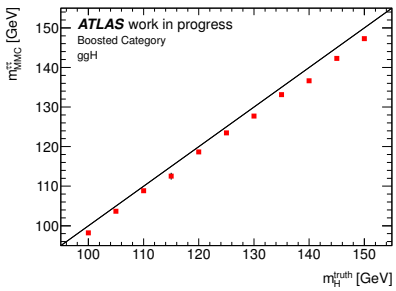
(a) VBFH in VBF



(b) VBFH in Boosted

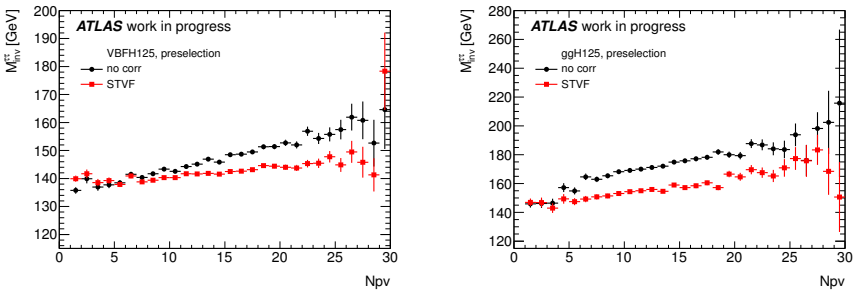


(c) $gg \rightarrow$H in VBF



(d) $gg \rightarrow$H in Boosted

Figure 5.16: The reconstructed value of the MMC mass is plotted as a function of the true m_H value for MC vector boson fusion Higgs events in the VBF category in (a) and in the Boosted category in (b), for MC gluon gluon fusion Higgs events in the VBF category in (c) and in the Boosted category in (d).



(a) VBFH (b) $gg \to$H

Figure 5.17: The invariant mass reconstructed with the collinear approximation is plotted at the preselection level as a function of N_{PV} before and after applying the STVF pile-up suppression method. MC vector boson fusion Higgs events with $m_H = 125$ GeV are shown in (a) and MC gluon gluon fusion Higgs events with $m_H = 125$ GeV are shown in (b).

5.2.1 $Z \to \tau\tau$ estimation with embedded sample

The $Z \to \tau\tau$ process is the main dominant irreducible background for the $H \to \tau\tau$ analysis. A full data driven estimation for this background is not possible for two main reasons:

- the τ identification that cannot ensure the selection of a sufficiently pure high statistics $Z \to \tau\tau$ data sample
- the presence of signal contamination from $H \to \tau\tau$ events that cannot be completely separated from $Z \to \tau\tau$ events

However, in order to not have a full MC based estimation for this important background, a hybrid sample, obtained embedding simulated τ leptons in $Z \to \mu\mu$ data, is employed [162]. This technique allows to exploit the main event kinematics, such as the underlying event, the pile-up and the jet activity, from data and to rely on MC only for the τ decay process that is well modeled in simulation, including the τ polarization and spin correlation.

The $Z \to \mu\mu$ data sample has be chosen as starting point for the embedding procedure for the following reasons:

- the good purity and the available high statistics for the $Z \to \mu\mu$ sample

- the small Higgs signal contamination due to the small muon mass that has a small coupling to the Higgs boson
- an event kinematics extremely similar to $Z \to \tau\tau$ events except for the mass difference between muons and taus. This difference is taken into account and corrected through a rescaling of the muon four momentum.

Before adding the simulated τ-leptons to the considered event, the tracks associated to the muons are removed from the data event and the corresponding estimated contribution in the calorimeter is subtracted at cell level to avoid energy double counting.

Since the complexity of the procedure, the embedded samples are extensively tested and few additional corrections are provided to ensure a kinematic spectrum in good agreement with the expected $Z \to \tau\tau$ spectrum. These correction includes:

- a weight to account separately for the muon trigger and reconstruction efficiencies affecting the starting $Z \to \mu\mu$ data sample.
- a p_{T} and η dependent weight to account for a non-operational pixel detector module that is not included in simulation. This correction is derived separately for electrons and muons.
- a weight to account for trigger efficiency effects that are not applied during the embedding procedure.

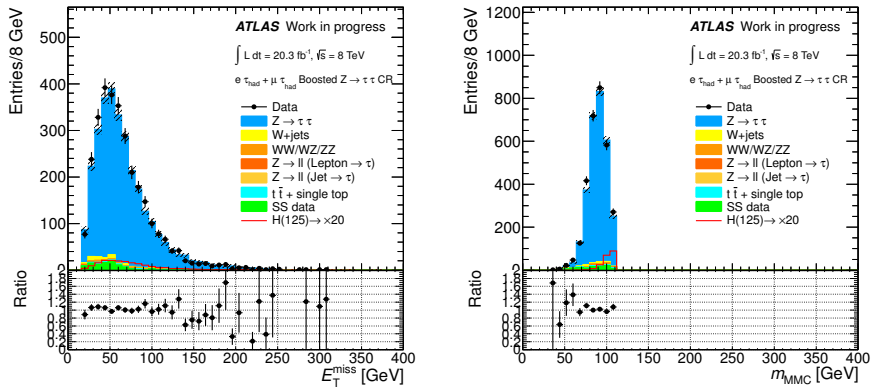
The embedded sample so obtained is normalized to data (with the non $Z \to \tau\tau$ backgrounds subtracted) in the visible mass[4] region: $40 < m_{\mathrm{vis}} < 70$ GeV.

Embedded validation

Inside the analysis the embedded $Z \to \tau\tau$ sample, described in the previous section, is validated comparing the background expectation with data in dedicated control regions for the VBF and Boosted categories defined as:

- category requirements (see Section 5.1.6)
- $m_{\mathrm{T}} < 40$ GeV
- $m_{\tau\tau}^{\mathrm{MMC}} < 110$ GeV

[4]The visible mass is defined as the invariant mass computed using only the visible τ decay products. Therefore the neutrinos contribution is excluded.



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ (b) MMC mass

Figure 5.18: Kinematics variables distribution in the $Z \to \tau\tau$ Boosted control region: $E_{\mathrm{T}}^{\mathrm{miss}}$ in (a) and MMC mass in (b). For $Z \to \tau\tau$ the embedded sample is used.

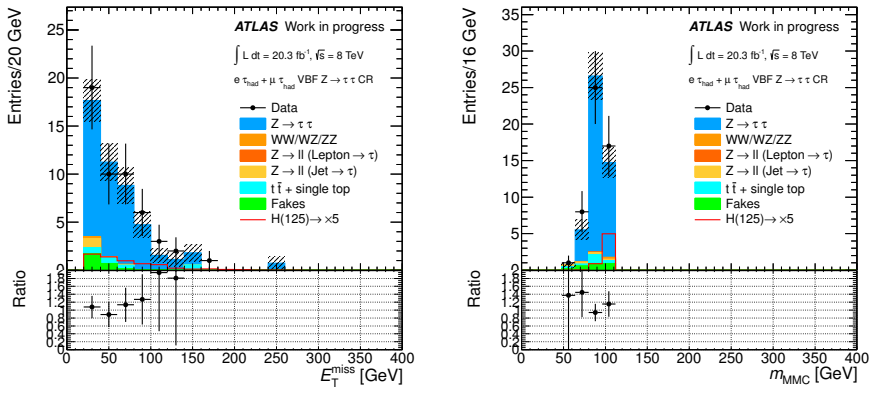
The $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ and the MMC mass distributions are shown in Figures 5.18 and 5.19 respectively for the Boosted and the VBF category control region. The agreement between data and the expectation is good.

5.2.2 Boosted category: OS-rSS method

At the event preselection, one of the requirements to identify $H \to \tau\tau$ signal events is the opposite charge correlation for the visible τ decay products, since they are expected to come from the decay of a neutral particle into two taus of opposite charge. This condition is not true for all backgrounds, so it can be used for the background estimation. For the Boosted category, the number of background events in the opposite sign signal region (OS events) is estimated making use of data where the lepton and the τ_{had} have the same charge sign (SS events). The main backgrounds show the following charge correlations:

Irreducible background with a true $\ell\tau_{\mathrm{had}}$ signature: dominated by $Z/\gamma^* \to \tau\tau$ events and with some contributions from diboson and top events. They are characterized by a strong charge correlation $N_{OS} \gg N_{SS}$, where N_{OS} and N_{SS} are respectively the number of opposite and same sign events.

Background with a jet misidentified as τ_{had}: dominated by QCD multijets,



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ (b) MMC mass

Figure 5.19: Kinematics variables distribution in the $Z \to \tau\tau$ VBF control region: $E_{\mathrm{T}}^{\mathrm{miss}}$ in (a) and MMC mass in (b).

W+jets and top[5] events. They are characterized by a moderate charge correlation $N_{OS} > N_{SS}$. For $Z \to \ell\ell$+jets events $N_{OS} \sim N_{SS}$

Background with a lepton misidentified as τ_{had}: dominated by $Z \to \ell\ell$ events. They are characterized by a strong charge correlation $N_{OS} \gg N_{SS}$.

The total background prediction is given by the number of same sign data events N_{SS}^{data}, dominated by QCD multijets events, with additional contributions, referred as *add-on* terms N_{add-on}^{X}, from other backgrounds estimated from MC or embedded samples:

$$\begin{aligned} N_{OS}^{bkg} = r_{QCD} \cdot N_{SS}^{data} + N_{add-on}^{Z\to\tau\tau} + N_{add-on}^{Z\to\ell\ell(\to\tau)} + N_{add-on}^{Z\to\ell\ell+jet(\to\tau)} \\ + N_{add-on}^{W+jets} + N_{add-on}^{top} + N_{add-on}^{VV} \end{aligned} \tag{5.15}$$

where r_{QCD} is a factor that accounts for potential differences in flavor composition of final state jets. The underlying assumption of this method is that the shape of the discriminant distribution, the MMC mass, in the signal region is the same for OS and SS events passing all kinematic selection cuts (except for charge requirements) for a given analysis category. In order to address possible mismodelling of the $\ell \to \tau_{\mathrm{had}}$ and $jet \to \tau_{\mathrm{had}}$ fake rate in MC simulation, the *add-on* terms are normalized to data in dedicated control regions, so k-factors are

[5]The top background includes both $t\bar{t}$ and single top processes

	$Z \to \tau\tau$	$Z \to \ell\ell$ ($\to \tau_{\text{had}}$)	$Z \to \ell\ell$+ jets($\to \tau_{\text{had}}$)	$W + jets$ ($e\tau_{\text{had}}$)	$W + jets$ ($\mu\tau_{\text{had}}$)	top	VV
k_{OS}	1	1	0.783±0.005	0.84±0.05	0.84±0.05	0.86±0.01	1
k_{SS}	1	1	0.783±0.005	0.99±0.08	0.99±0.08	0.99±0.03	1

Table 5.5: Summary of the k-factors for each background in the Boosted category.

defined as $k = N(data)/N(MC)$ for each background. The k-factor for a given background X can be different for OS and SS events, to account for differences in the tau misidentification rate for jets from quark and gluon hadronization, but it is assumed to be the same between a signal region and the corresponding control region.

Each *add-on* term can be expressed as:

$$N^X_{add-on} = k^{OS}_X N^X_{OS} - r_{QCD} k^{SS}_X N^X_{SS} \tag{5.16}$$

where the k^{OS}_X, k^{SS}_X are the k-factors respectively for OS and SS events and the $r_{QCD} k^{SS}_X N^X_{SS}$ term accounts for the contribution of each non QCD background already included in the N^{data}_{SS} counting in Equation 5.15. The values for the different k-factors are summarized in Table 5.5: when the k-factor is assumed equal to unity, as for the $Z \to \tau\tau$, $Z \to \ell\ell(\to \tau_{\text{had}})$ and diboson backgrounds, no error is quoted. Since the diboson (VV) background is fully estimated from MC, the N_{OS} and the N_{SS} are taken from MC. For the $Z \to \ell\ell + jet(\to \tau_{\text{had}})$ background, no statistically significant difference between OS and SS events is expected, hence only N_{OS} is used:

$$N^{Z\to\ell\ell+jet(\to\tau)}_{add-on} = k_{Z\to\ell\ell+jet(\to\tau)} N^{Z\to\ell\ell+jet(\to\tau)}_{OS} (1 - r_{QCD}) \tag{5.17}$$

r_{QCD} determination

As explained in the previous section, the QCD multijets contribution, with one of the jets misidentified as a prompt lepton and second jet misidentified as a τ_{had}, is estimate according to Equations 5.15 and 5.16:

$$N^{QCD} = r_{QCD} \cdot \left(N^{data}_{SS} - \sum_X k^{SS}_X N^X_{SS} \right) \tag{5.18}$$

At LO order, QCD dijets events include qq, $qq^{'}$, qg and gg pairs in the final state. No charge correlation between a fake lepton and a fake τ_{had} is expected

	r_{QCD}
$e\tau_{\text{had}}$	1.00±0.05(stat)±0.12(sys)
$\mu\tau_{\text{had}}$	1.10±0.06(stat)±0.13(sys)

Table 5.6: Value of r_{QCD} for $e\tau_{\text{had}}$ and $\mu\tau_{\text{had}}$ events.

in events with jets from qg and gg parton pairs. However, significant charge correlation is expected in events with jets from qq and $qq^{'}$ pairs. Therefore, the charge correlation requirement can alter the flavor composition of partons in the final state and lead to differences in N_{OS}^{QCD} and N_{SS}^{QCD} and in kinematic distributions. This effect is corrected introducing the r_{QCD} factor in Equations 5.15 and 5.18, defined as the ratio $N_{OS}^{QCD}/N_{SS}^{QCD}$ measured in an enriched QCD data control region. The QCD control region is defined as:

- $E_{\text{T}}^{\text{miss}} < 15$ GeV
- $m_{\text{T}} < 30$ GeV
- no lepton isolation requirement
- loose tau identification

The results for the r_{QCD} determination are reported in Table 5.6 separately for $e\tau_{\text{had}}$ and $\mu\tau_{\text{had}}$ events.

Systematic uncertainties are computed varying the lepton isolation requirement, the tau identification working point, the fit range. These systematics are finally summed in quadrature and simmetrized to obtain a total uncertainty of 11.9% and 11.6% respectively for the $e\tau_{\text{had}}$ and $\mu\tau_{\text{had}}$ channels.

k-factor determination

In the Boosted category, the W+jets, $t\bar{t}$, $Z \to \ell\ell + jet(\to \tau_{\text{had}})$ backgrounds are evaluated from MC samples and normalized to data in separate control regions to correct for the MC mismodeling of the $jet \to \tau_{\text{had}}$ fake rate. Each control region is enriched of a specific background and the small contamination from other backgrounds is subtracted from data before performing the normalization. The requirements used to select the control regions are reported in the following.

W+jets control region:

- Category requirements
- $m_{\mathrm{T}} > 70$ GeV
- $\Sigma\Delta\phi$, $\Delta\Delta R$, x_ℓ and $x_{\tau_{\mathrm{had}}}$ requirements are not applied

Separate k-factor for OS and SS events are determined. Further corrections need to be applied on the shape of the W contribution, to account for observed shape differences between MC and data in the W control region. Two shape-dependent correction functions are applied: the first depends on the ratio between the lepton and the tau momentum $p_{\mathrm{T}}^{\ell}/p_{\mathrm{T}}^{\tau}$, while the second depends on the pseudorapidity separation between the lepton and the tau $\Delta\eta(\tau, \ell)$. After this procedure a good data-MC agreement is found as shown for the $E_{\mathrm{T}}^{\mathrm{miss}}$ and MMC mass distributions in Figure 5.20.

top control region:

- at least two jets with $p_{\mathrm{T}} > 30$ GeV
- at least a b-tagged jet
- $E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV
- $m_{\mathrm{T}} > 70$ GeV

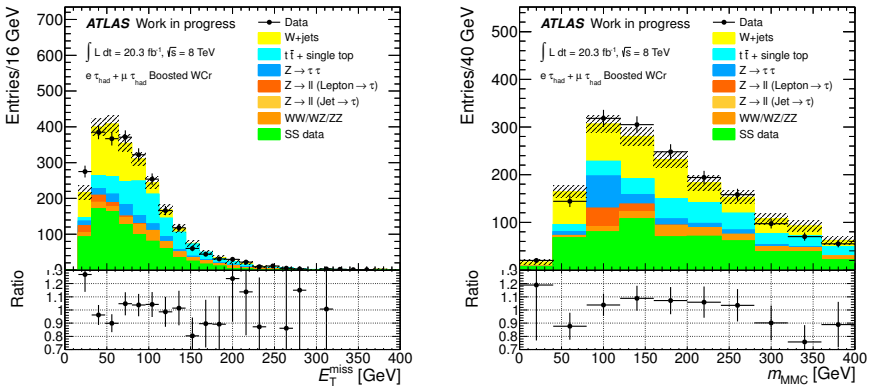
Separate k-factor for OS and SS events are determined. A good data-MC agreement is found after normalization as shown for the $E_{\mathrm{T}}^{\mathrm{miss}}$ and MMC mass distributions in Figure 5.21.

$Z \to \ell\ell$ control region:

- two leptons with opposite sign and same flavor
- standard isolation requirements on the leading lepton
- no lepton associated with the τ_{had} at truth level for the $Z \to \mu\mu$ MC sample
- $61 < m_{\ell\ell} < 121$ GeV

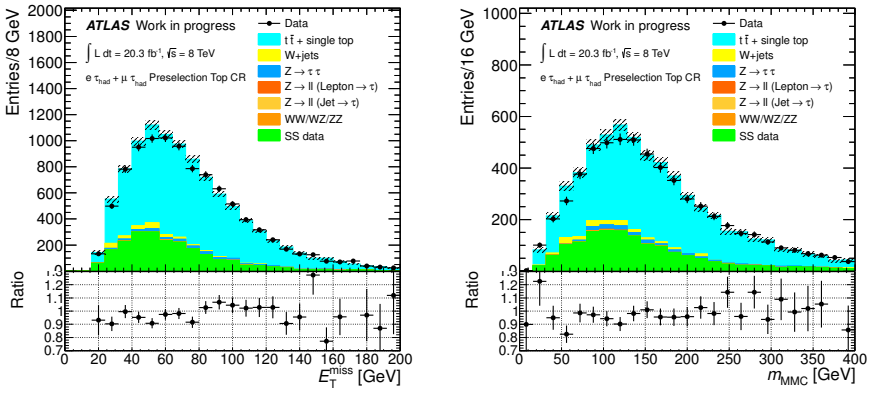
A unique k-factor is derived for OS and SS events.

Possible mismodeling in the $\ell \to \tau_{\mathrm{had}}$ fake rate, important for the $Z \to \ell\ell(\to \tau_{\mathrm{had}})$, are also considered. Scale factors, computed with a tag-and-probe technique using $Z \to ee$ events, are applied to the MC simulation to correct for the electron misidentification rate. No correction is instead applied on the $Z \to \mu\mu$ background, since its small contribution, and a conservative systematic uncertainty of 15% is propagated through the analysis.



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ (b) MMC mass

Figure 5.20: Kinematics variables distribution in the W Boosted control region: $E_{\mathrm{T}}^{\mathrm{miss}}$ in (a) and MMC mass in (b).



(a) $E_{\mathrm{T}}^{\mathrm{miss}}$ (b) MMC mass

Figure 5.21: Kinematics variables distribution in the top control region: $E_{\mathrm{T}}^{\mathrm{miss}}$ in (a) and MMC mass in (b).

5.2.3 VBF category: fake factor method

In the VBF category, the $Z \to \tau\tau$ and the MC based backgrounds are estimated and normalized to data as described in the previous sections. Since the low available statistics in this category for SS events, a different background estimation for the QCD multijets and W+jets contribution is provided through the Fake Factor method [163, 164].

The method uses a control sample defined by applying the same event selection used for the VBF category described in Section 5.1.6, with the exception that the τ_{had} candidates are required to fail the τ_{had} identification (referred to as anti-τ in the following). Fake backgrounds, N^{est}_{bkg}, can be estimated from anti-τ events, $N_{\text{anti}-\tau}$, and fake factors (FF) according to the following formula:

$$N^{est}_{bkg} = N_{\text{anti}-\tau} \times FF \tag{5.19}$$

$$FF = \frac{N_{\text{identified}-\tau}}{N_{\text{anti}-\tau}} \tag{5.20}$$

where $N_{\text{identified}-\tau}$ is the number of events passing the tau identification criteria.

The fake factor is split into two separate components, FF_{QCD} and FF_{W+jets}, for samples dominated by gluon and quark jets, respectively. Thus, Equations 5.19 and 5.20 become:

$$N^{est}_{bkg} = \left(N^{data}_{\text{anti}-\tau} - N^{Z\to\tau\tau}_{\text{anti}-\tau} - N^{others}_{\text{anti}-\tau}\right) \times FF_{\text{MIX}} \tag{5.21}$$

$$FF_{\text{MIX}} = R_{W+jets} \cdot FF_{W+jets} + (1 - R_{W+jets})\, FF_{QCD} \tag{5.22}$$

where $N^{others}_{\text{anti}-\tau}$ is the number of other electroweak components predicted with MC simulation, and R_{W+jets} is the fraction of W+jets events in anti-τ events.

The R_{W+jets} is computed according to the quark and gluon fractions in the events passing the selection criteria:

$$R_{W+jets} = \frac{N^{est}_{W+jets}}{N^{est}_{W+jets} + N^{est}_{QCD}} \tag{5.23}$$

where the estimated number of W+jets events N^{est}_{W+jets} is obtained from the number of data events in the W+jets control region (as described in the previous section, it includes $m_{\text{T}} > 70$ GeV) and by the MC acceptance:

$$N^{est}_{W+jets} = N^{data}_{\text{anti}-\tau,WCR} \times \frac{N^{W+jets,MC}_{\text{anti}-\tau}}{N^{W+jets,MC}_{\text{anti}-\tau,WCR}} \tag{5.24}$$

while the estimated number of QCD multijets events is obtained from the number of data events with all the other backgrounds subtracted:

$$N_{QCD}^{est} = N_{\text{anti}-\tau}^{data} - \left(N_{W+jets}^{est} + N_{\text{anti}-\tau}^{Z\to\tau\tau} + N_{\text{anti}-\tau}^{others}\right) \tag{5.25}$$

FF_{W+jets} and FF_{QCD} are measured in separate control regions. The $m_{\text{T}} > 70$ GeV region is used for FF_{W+jets}, while the loose lepton region is used for measuring FF_{QCD}. Here, loose leptons are non-isolated muons for the muon channel, or non-tight identified and non-isolated electrons for the electron channel.

After applying this procedure, a residual mismodeling is observed between the background estimate and the data, due to the different treatment of τ_{had} candidates, that are classified τ, and anti-τ, that are classified mainly as jets in the $E_{\text{T}}^{\text{miss}}$ calculation. Therefore, in order to correct for this effect, the fake background estimate was reweighted with weights derived in the W control region where all signal selections were applied, except for a requirement of $m_{\text{T}} > 70$ GeV. These weights are binned in the variable Ξ, defined as:

$$\Xi = E_{\text{T}}^{\text{miss}} \cos\left(\Delta phi(\tau_{\text{had}}, E_{\text{T}}^{\text{miss}})/p_{\text{T}}^{\tau}\right) \tag{5.26}$$

A recomputation of the $\boldsymbol{E}_{\text{T}}^{\text{miss}}$ using a calibration for the anti-τ as τ objects, that is coherent with the analysis choice in this particular control region, has showed to improve the agreement between the background estimate and the data with no need for a reweighting. A fully coherent $\boldsymbol{E}_{\text{T}}^{\text{miss}}$ recomputation is under development and it will be integrated in the next analysis for this search.

5.3 Systematics uncertainties

In this section the several sources of systematics uncertainties for this analysis are described and quantified. The description of the ATLAS detector responses and the modeling of various physics processes like pile-up conditions, trigger efficiencies and physics objects reconstruction and identification lead to separate systematics uncertainties that are propagated through the whole analysis. Theory uncertainties on the Higgs production cross section and on the modeling of the underlying event are as well considered. Finally, systematics uncertainties that arise from the employment of the data driven methods for the background estimation and background modeling are also estimated and propagated. All these systematics uncertainties are then included as nuisance parameters in the profile likelihood global fit discussed in Section 5.4. The effects of the systematics uncertainties is considered on both the normalization and on the shape of the MMC distribution. More details for each source of systematic uncertainties are given in the following.

5.3.1 Detector and physics object uncertainties

The following systematic uncertainties on detector performance and physics object reconstruction are considered:

Luminosity: the uncertainty on the integrated luminosity is $\pm 2.8\%$. It is derived following the same methodology as that detailed in Ref. [165], from a preliminary calibration of the luminosity scale derived from beam-separation scans performed in November 2012.

Tau identification: the correction is obtained using tag-and-probe measurement with $Z \to \tau\tau$ events. The correction factors, to be applied in MC to ensure agreement with data, and the upward and downward corresponding variation are applied only on truth-matched hadronic taus. Furthermore, the correction factors on the rate of misidentification of electrons as τ_{had} candidates (applied on τ candidates matched with a true electron) are also varied within their uncertainties.

Tau energy scale: the upwards and downwards variations are applied to the tau according to the recommendations [166]. The TES uncertainty is decorrelated for true and fake τ candidates, but not split into further sources of uncertainty, and considered also as a shape systematic.

Lepton reconstruction/identification/isolation/trigger efficiency: the corrections are obtained for each of these types of efficiencies using tag-and-probe measurements. These are then applied to MC samples to ensure agreement with data. The uncertainties are propagated to the final result by varying the corrections by one standard deviation.

Lepton energy/momentum resolution: the electron energy and the muon momentum are smeared according to the resolutions measured in data. This systematic uncertainty plays a very minor role on the final result.

Jet energy resolution (JER) and scale (JES): concerning the JER, an upwards variation of 1σ is obtained by smearing each jet with a factor accounting for the uncertainty in the resolution in-situ measurement. The final effect of the variation is symmetrized in order to have a two-sided uncertainty in the fit. This uncertainty is also included as shape systematic in the fit. As discussed in Section 3.2.4, the total JES systematic is split into several contributions according to different physics sources. However, due to the limited MC statistics, a reduction scheme is derived to

not introduce statistics noise in the global fit. Starting from the official reduction scheme with 24 nuisance parameters only the ones relevant for the analysis are retained: *JESFlavResp, JESFlavComp, JESModelling1, JESEtaModelling, JESEtaIntercalibrationModelling, JESEtaIntercalibrationStatMethod, JESStatistical1, JESDetector1, JESBJet, JESPUNPV, JESPURho.* The first four are also considered as shape systematics. Since it is known that background and signal components can be dominated more by quark (VBFH, VH, top, diboson) or by gluon ($gg \to H$, V+jets), two different nuisance parameters are considered for the flavor composition systematic. The uncertainty components are fully correlated between categories and analysis channels, except for the *JESPURho* that is still considered correlated across channels but only correlated by process within three groups differing by initial state: qq initiated (VH, VBF, diboson), qg initiated (V+jets) and gg initiated ($gg \to H$, top).

$E_{\mathrm{T}}^{\mathrm{miss}}$ **energy scale and resolution:** for each of the physics object systematic uncertainty listed above the corresponding variation is also propagated to the $E_{\mathrm{T}}^{\mathrm{miss}}$ that is therefore recomputed accordingly. Independent systematics on the scale and resolution of the soft term, derived as described in Sections 3.3.3 and 4.6, are also considered and propagated through the analysis. For the soft term $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution, the final effect of the variation is symmetrized in order to have a two-sided uncertainty in the fit.

5.3.2 Theory uncertainties

The Higgs cross section is used to normalize the signal MC samples, therefore, uncertainties on the computation like neglecting high-order perturbative corrections or including the mass of the quarks or choosing a specific PDF set, need to be taken into account. These theory uncertainties must be applied only on the signal samples and more specifically include:

QCD scale uncertainty: it provides a realistic estimate of higher order contributions to analyses which classify events by the number of jets. An exclusive requirement on the number of jets introduces complexity in the QCD perturbative computation and in systematic uncertainties on the scale of QCD processes. The prescription considers the QCD scale uncertainties on inclusive multijet cross sections ($\sigma \geq 0$ jet, $\sigma \geq 1$ jets, $\sigma \geq 2$ jets) and assumes that they are uncorrelated and propagated to the exclusive

jet bins. The scale uncertainties for the inclusive multijet cross sections are computed with the MCFM 6.3 [167] program and the jet bin fractions are determined by subtracting the appropriate inclusive cross section. The uncertainties are estimated using MCFM by varying the renormalization μ_R and factorization μ_F scale between $m_H/2$ and $2m_H$. While the inclusive cross section uncertainty for the gluon gluon fusion process is only 8%, the uncertainty in the exclusive bins used in this analysis ranges from 22% to 26% and can be as high as 74% for boosted events with at least two jets.

Modelling of the differential cross section $d\sigma/dp_{\mathrm{T}}^{H}$: the default $gg \to H$ simulation in ATLAS is performed using the Powheg generator. The McAtNlo generator implements the same matrix element but quark mass effects on $d\sigma/dp_{\mathrm{T}}^{H}$ of the two generators are known to be quite different. The difference between the output of these two generators is assigned as an additional systematic uncertainty. A variation of 29% is observed for the Boosted category and a variation of 18% is observed for the VBF category.

Modeling of the underlying event: underlying event effects have been computed comparing the Perugia 2011C underlying event tuning with the AUET2B tune [168]. Discrepancies of ∼6% and ∼30% for the signal yields in vector boson fusion and gluon gluon fusion production in the VBF category have been found.

Parton Distribution Functions: MCFM was also used to verify that the variation of the differential cross sections associated with each analysis category due to different PDF sets (comparing MSTW [169], NNPDF [170] and CT10 [158]) are smaller than or equal to the inclusive variation. A PDF uncertainty of 7.5% is assigned in all categories for gluon fusion production, and 2.8% is assigned for VBF and VH production.

The small diboson background is evaluated completely from MC and it is normalized using its predicted cross section, thus theory uncertainties on the cross section evaluation and on the use of a specific PDF set are provided as well, they are respectively 1% and 4%.

5.3.3 Background estimation uncertainties

The background estimation described is Section 5.2 makes use of data driven and semi data driven techniques that are additional sources of systematic uncertain-

ties.

Embedding technique: two separate systematic uncertainties are considered. The first concerns the muon isolation criteria in the selection of the $Z \to \mu\mu$ data sample. To address this uncertainty the embedded events with the nominal selection are compared with events selected using no isolation for muons or with a tighter isolation. The second source of systematic uncertainty comes from the replacement of the muons with the simulated taus, in particular in the muon energy subtraction in the calorimeter. To evaluate this systematic uncertainty the energy of each cell is conservatively scaled upward and downward by 20% before the subtraction. It is also considered as a shape systematic uncertainty.

W+jets MC reweighting: the two separate shape corrections as a function of the $\Delta\eta(\tau, \ell)$ and of the ratio $p_{\mathrm{T}}^{\ell}/p_{\mathrm{T}}^{\tau}$, described in Section 5.2.2, applied after the W+jet background estimation, are obtained by a fitting procedure. The systematic uncertainty on this method reflects the statistical uncertainty of the fit and it is considered as a shape systematic.

OS-rSS technique: several source of systematic are associated to this method. A systematics uncertainty is derived for the r_{QCD} correction as described in Section 5.2.2. Then, each of the k-factor, introduced for the non resonant backgrounds to normalize the MC to data, has a statistical uncertainty associated that depends on the control region requirements and that is assigned as systematics uncertainty and propagated to the analysis.

Fake factor: a conservative 50% uncertainty is assigned on W+jet and QCD multijets background processes because of imperfect knowledge of the quark and gluon jet flavor composition in the signal region.

5.4 Statistical analysis and signal extraction

In the search for the Higgs boson, testing statistical hypothesis plays an important role. In particular, the sensitivity of an analysis is given by reporting the expected significance, and the final results for either the exclusion or the discovery of a new process are quoted in terms of probability [145, 171].

Given a hypothesis H, the p-value is defined as the probability, under assumption of H, of finding data of equal or greater incompatibility with the predictions

of H. The hypothesis is excluded if its p-value is observed below a specified threshold. The p-value can be also converted into an equivalent significance:

$$Z = \Phi^{-1}(1 - p) \tag{5.27}$$

where Φ^{-1} is the quantile of the normal Gaussian, representing the number of standard deviations for a normal Gaussian function above which the area of the Gaussian function itself is equal to p.

Both the limit and the discovery procedures involve the rejection of a null hypothesis, H_0, in the comparison with an alternative hypothesis, H_1. For purposes of discovering a new signal process, H_0 is defined as describing only the known background processes, and it is tested against the alternative H_1, which includes both background and signal. The discovery is claimed when the background only hypothesis is rejected with a significance of at least 5σ ($p = 2.87 \cdot 10^7$). For purposes of setting the exclusion limits for a process, the model with signal plus background plays the role of H_0, which is tested against the background only hypothesis, H_1. The exclusion is set when the signal hypothesis is rejected at the 95% confidence level (CL). It has become customary to express results of the SM Higgs searches according to a signal strength modifier μ, defined in Equation 1.24, that is taken to change the SM Higgs boson cross sections of all production mechanisms by exactly the same scale μ. In the absence of a Higgs boson (background only hypothesis), $\mu = 0$, and for the Standard Model expectation, $\mu = 1$.

For the $H \to \tau\tau$ search the statistical analysis of the data employs the profile likelihood method [172] with the MMC mass as the discriminating observable. The binned likelihood function is constructed as the product of Poisson probability terms as an estimator for μ. In addition to parameters of interest such as the rate (cross section) of the signal process, the signal and background models will contain nuisance parameters whose values are not taken as known a priori but must be fitted from the data. The flexibility of the fit for each of these additional terms can be used to model associated systematic uncertainties, or to provide additional constraints on the background estimate. In particular in this analysis, the $Z \to \tau\tau$ background rate is allowed to float freely in the global fit and the impact of systematic uncertainties on the signal and background expectations is described by nuisance parameters, $\overrightarrow{\theta}$, which are parametrized by a Gaussian or

log-normal constraint.

$$\mathcal{L}(\mu, \overrightarrow{\beta}_{\text{samp}}, \overrightarrow{\theta}_{s}, \overrightarrow{\theta}_{b}, \overrightarrow{\theta}_{\text{global}}) = \text{Pois}(n|\mu_T)\text{Pois}(n_{\text{samp}}|\beta_{\text{samp}}) \mathcal{L}(\overrightarrow{\beta}_{\text{samp}}, \overrightarrow{\theta}_{s}, \overrightarrow{\theta}_{b}, \overrightarrow{\theta}_{\text{global}}) \tag{5.28}$$

where:

- n is the number of events in the signal region
- $\overrightarrow{\beta}_{\text{samp}}$ are the statistical uncertainties of the MC or data driven control sample events, using the initial event numbers (n_{samp}), before scaling to the cross section
- $\overrightarrow{\theta}_{s,b}$ are the specific nuisance parameters related to the signal and the background, such as the efficiency and the cross section uncertainties
- $\overrightarrow{\theta}_{\text{global}}$ represent the common nuisance parameters which are correlated between channels, such as the luminosity uncertainty
- μ_T is the total number of expected events given by

$$\mu_T = \sum_{l=1}^{4} \mu L \sigma_l(m_H) f_s(\overrightarrow{\theta}_{s}) f_g(\overrightarrow{\theta}_{\text{global}}) + \sum_{j} L\beta_j f_b(\overrightarrow{\theta}_{b}) f_g(\overrightarrow{\theta}_{\text{global}}) \tag{5.29}$$

  where:

  - L is the nominal integrated luminosity
  - μ is the one parameter of interest, the scaling factor for the expected signal rate (signal strength)
  - $\sigma_l(m_H)$ is the effective cross section (in pb) for signal events in channel l ($gg \rightarrow$H, VBFH, WH, ZH)
  - β_j is the nominal effective cross section (in pb) for background j (including β_{samp})
  - $f_{s,b,\text{global}}$ represent the dependence of the expected number of events on the various nuisance parameters.

The likelihood is used to construct a statistical test based on the profile likelihood ratio and asymptotic formulae are used when appropriate. The statistical

test is given by:

$$\tilde{q}_\mu = \begin{cases} -2\ln\dfrac{L(\mu, \hat{\hat{\vec{\theta}}}(\mu))}{L(0, \hat{\hat{\theta}}(0))} & \hat{\mu} < 0 \\ -2\ln\dfrac{L(\mu, \hat{\hat{\vec{\theta}}}(\mu))}{L(\hat{\mu}, \hat{\vec{\theta}})} & 0 \leq \hat{\mu} \leq \mu \\ 0 & \hat{\mu} \geq \mu \end{cases} \tag{5.30}$$

where:

- $\hat{\vec{\theta}}$ represent the nuisance parameters evaluated at μ
- $\hat{\hat{\vec{\theta}}}(\mu)$ are the Maximum Likelihood Estimators (MLE) of μ, $\vec{\theta}$

Toy MC experiments are generated to construct the pdf $f(\tilde{q}_\mu|\mu, \hat{\hat{\theta}}(\mu))$ under an assumed signal strength μ. From this, the p-value for μ is calculated using:

$$p_\mu = \int_{\tilde{\mu}_\mu^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\hat{\theta}}(\mu))d\tilde{q}_\mu \tag{5.31}$$

The procedure to compute exclusion limits is based on the modified frequentist method, often referred to as CL_s [173], for which the exclusion is not quoted simply as p-value under the $s+b$ (signal plus background) hypothesis:

$$p_\mu = CL_{s+b} = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}}|\mu s + b) \tag{5.32}$$

but it is quoted in terms of CL_s:

$$CL_s = \frac{p_\mu}{1 - p_b} \tag{5.33}$$

where

$$p_b = CL_b = P(\tilde{q}_\mu \leq \tilde{q}_\mu^{\text{obs}}|\mu = 0) = \int_{-\infty}^{\tilde{\mu}_\mu^{\text{obs}}} f(\tilde{q}_\mu|\mu = 0, \hat{\hat{\theta}}(0))d\tilde{q}_\mu \tag{5.34}$$

The CL_s method is introduced to reduce the exclusion of region where the sensibility is very small. In particular the CL_{s+b} method in Equation 5.32 excludes regions where $p_\mu < 0.05$ also when the expected number of signal events is much less than that of background. In the more conservative modified approach, using the CL_s, the p-value is effectively penalized by dividing by $1 - p_b$. If the two distributions $f(\tilde{q}_\mu|\mu = 0, \hat{\hat{\theta}}(0))$ and $f(\tilde{q}_\mu|\mu, \hat{\hat{\theta}}(\mu))$ are widely separated, then

Category	MMC binning [GeV]
Boosted	0,80,90,95,100,105,110,115,118,121,124,127,130,135,140,145,160,180,200,400
VBF	0,60,80,90,100,110,120,130,140,200,400

Table 5.7: Summary of the binning employed for the Boosted and VBF categories.

$1 - p_b$ is only slightly less than unity, the penalty is small, and thus exclusion based on CL_s is similar to that obtained from the usual CL_{s+b}. If, however, one has little sensitivity to the signal model, then the two distributions are close together, $1 - p_b$ becomes small, and thus the p-value is penalized more. In this way one is prevented from excluding signal models in cases of low sensitivity.

5.4.1 Settings of the fit model

In this section the main settings that have been optimized for the fit model are summarized.

Choice of binning for the MMC mass discriminant: since the limit extraction uses a binned likelihood, the binning that gives the highest sensitivity is chosen separately for each category as summarized in Table 5.7.

$Z \to \tau\tau$ **normalization:** this normalization is kept as an additional free parameter in the fit because of the mismodelling observed in Figure 5.3(e). In this way the the normalization can be extracted directly from data during the profile likelihood procedure.

Treatment of the systemics uncertainties: this is an important point for the fit model and the minimization procedure. A full description is given in the next section.

With these optimizations the limit results are obtained using a histogram-based fitting machinery HistFactory in the RooStat package [174] giving as input the shape of the nominal distributions of the MMC mass in each category and the corresponding systematic variations optimized as previously described.

5.4.2 Treatment of systematics uncertainties

The systematics uncertainties are an important aspect of the analysis. It must be noticed that for some samples, after all the selections have been applied, small

upwards and downwards systematic variations can be dominated by the statistical error without providing useful information. In order to avoid the possible introduction of statistical noise in the global ft and in the limit extraction, only systematic uncertainties on the normalization with an effect larger than 0.5% are kept. Moreover, for each shape systematic considered, the following smoothing and pruning criteria are applied to test the level of significance and thus decide if the systematic can enter in the global fit:

Pruning 1: the Kolmogorov-Smirnov test between the upwards and the downwards fluctuated shape with respect to the nominal is performed, for each potential shape systematic NP and for each sample. In this calculation the statistical uncertainty that enters is only the largest of the nominal or varied one. The shape systematic is retained if the result of the Kolmogorov-Smirnov test is less than 95%, for either of the upwards or downwards fluctuated shape. If neither the upward nor the downward variation's probability is lower than that threshold, the shape variation is considered to not be significant for the given background sample, and the shape NP is not used in the fit (the corresponding normalization uncertainty is still kept however).

Smoothing: the ratio of variation to nominal (separately for upwards and downwards variations) is smoothed, using the TH1::Smooth(1) method [175] of ROOT. The smoothed varied shape is then obtained by multiplying the nominal with the smoothed ratio. The reason for smoothing the ratio, rather than the varied shape directly, is that the ratio is expected to be a relatively smooth and non-rapidly changing function in the absence of noise, possible over-smoothing is thus avoided.

Pruning 2: for each systematic and for each background sample, the maximum bin-by-bin variation significance of the MMC discriminating variable, $\max_i S_i$, should be at least 0.1. The variation significance S_i is defined as $S_i = |u_i - d_i|/\sigma_i^{\text{tot}}$, with u_i, d_i being respectively the upwards and downwards variation in bin i for a given background sample, and with σ_i^{tot} being the statistical uncertainty for the total background estimation (i.e. for all samples) in bin i. If this condition is not verified, the shape variation is considered non significant and not considered further.

The effect of $\pm 1\sigma$ variation is shown in Figures 5.22, 5.23, 5.24 and 5.25 for the main shape systematics on the MMC mass distribution (binned as reported

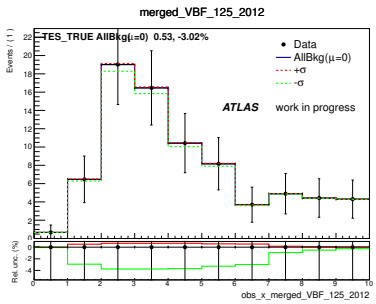
in Table 5.7) in the Boosted and VBF categories for the backgrounds and signal samples with $m_H = 125$ GeV. Only the distributions for the dominant signal contribution (vector boson fusion in the VBF category and gluon gluon fusion in the Boosted category) are reported in the figures, but all the significant systematics for each signal sample are considered in the fit. Systematics uncertainties on the TES are important for all the samples and, as expected, for the backgrounds the "true" component is more important for the central mass region where the $Z \to \tau\tau$ (with genuine τ-leptons) is the dominant background, while the "fake" component contributes most in the low mass and high mass bins where the contribution of QCD multijets and of the W+jets backgrounds (where a jet is misidentified as a τ_{had}) is dominant. In general, uncertainties on the JES gives significative shape variations, while other shape uncertainties like the ones on the data driven background estimation techniques can be less significative and occasionally pruned away.

5.4.3 Tests of the fit model

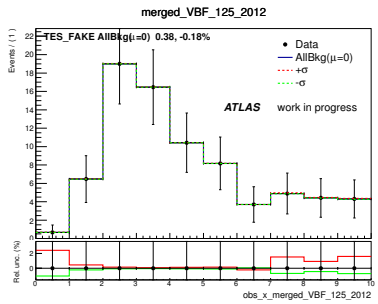
Since the large number of degrees of freedom introduced in the fit and the complexity of the fit procedure itself, before proceeding with the limit extraction, it is useful to test the stability and the sanity of the fit, examining possible over constraints or ill-behaviors of the nuisance parameters. A full list of all the nuisance parameters entering in the fit together with a brief explanation is reported in Appendix B.

Figure 5.26(a) shows the correlation matrix between the nuisance parameters. In the building of the likelihood the distinct nuisance parameters are treated as uncorrelated, possible correlation after fit can lead to undesired behavior, but in from the figure the results look sane and solid. No strong correlation between the nuisance parameters are observed, proving an overall good behavior of the fit machinery.

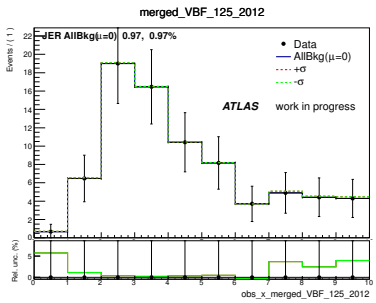
Another test consists in checking if some of the nuisance parameters are significantly different after the fit from their nominal value before the fit. This can be symptomatic of an important mismodeling which is artificially absorbed by the fit and must be carefully scrutinized. Figure 5.26(b) shows the difference between the ML estimator for the nuisance parameters and their nominal value in unit of the uncertainty of the nuisance parameters themselves, $(\theta_{\text{fit}} - \theta_0)/\Delta\theta$. No suspicious behavior is observed: no nuisance parameter is pulled in a significant way.



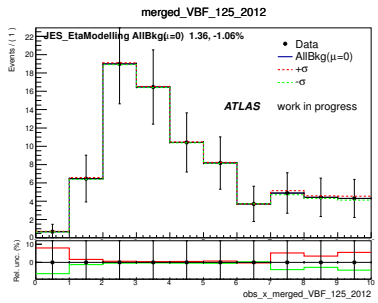
(a) TES true



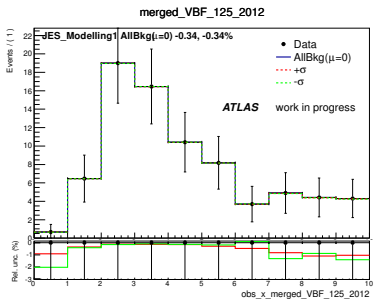
(b) TES fake



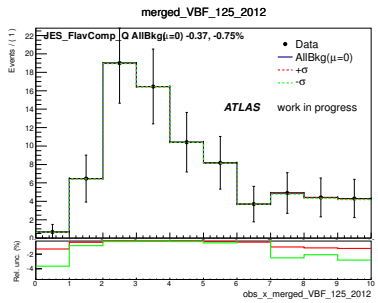
(c) JER



(d) JESEtaModelling

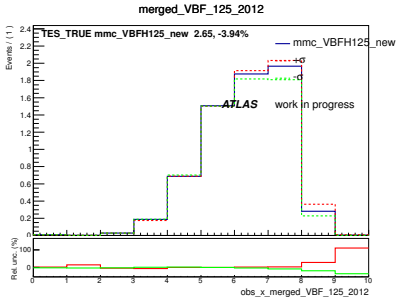


(e) JESModelling1

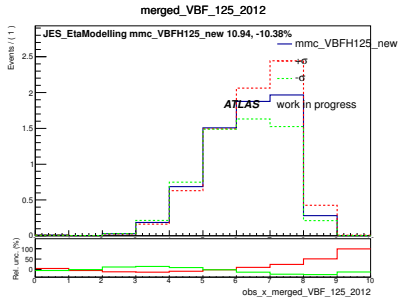


(f) JESFlavComp

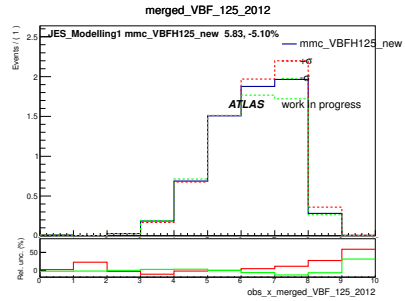
Figure 5.22: Effect of $\pm 1\sigma$ systematic shape variation on the MMC mass distribution for the total background in the VBF category: the TES components are shown in (a) and (b), the JER is shown in (c), the JES components are shown in (d), (e) and (f). The bottom part of the figures shows the relative variation in percent. In these plots "data" refers to Asimov data [171, 176] and coincide with the non variated distributions.



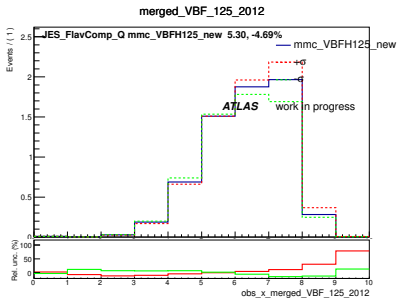
(a) TES true



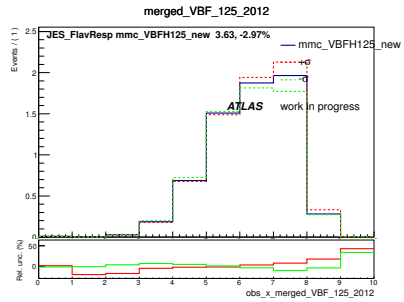
(b) JESEtaModelling



(c) JESModelling1

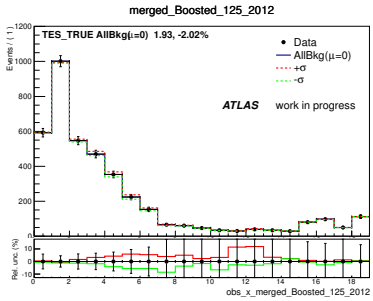


(d) JESFlavComp

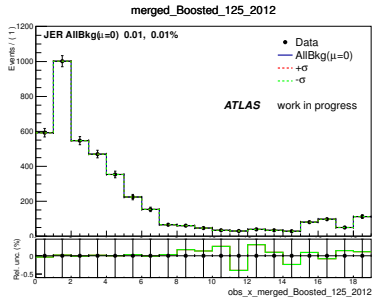


(e) JESFlavResp

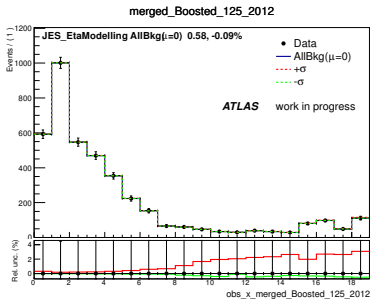
Figure 5.23: Effect of $\pm 1\sigma$ systematic shape variation on the MMC mass distribution for the vector boson fusion signal in the VBF category: the TES true component is shown in (a), the JES components are shown in (b), (c), (d), and (e). The bottom part of the figures shows the relative variation in percent.



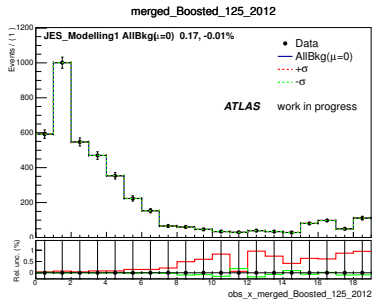
(a) TES true



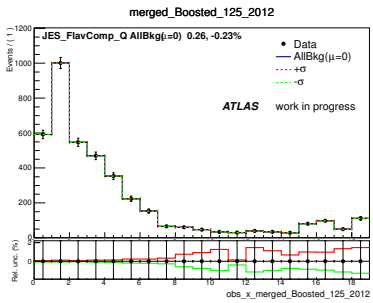
(b) JER



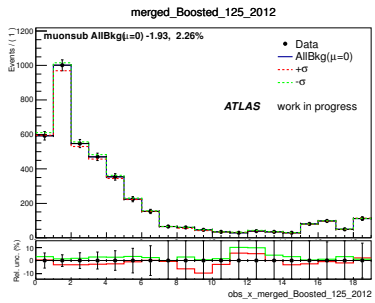
(c) JESEtaModelling



(d) JESModelling1

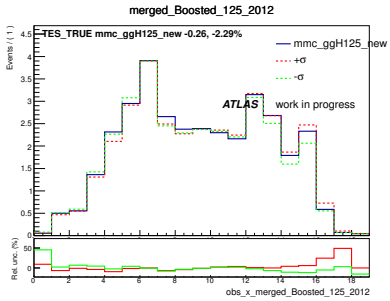


(e) JESFlavComp

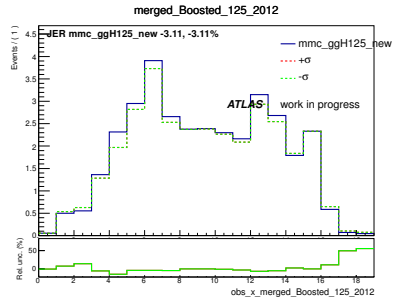


(f) muon subtraction in embedding

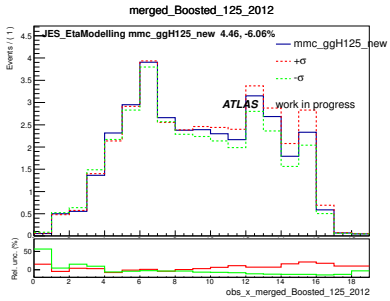
Figure 5.24: Effect of $\pm 1\sigma$ systematic shape variation on the MMC mass distribution for the total background in the Boosted category: the TES true component is shown in (a), the JER is shown in (b), the JES components are shown in (c), (d) and (e), the muon subtraction systematic on the embedded procedure is shown in (f). The bottom part of the figures shows the relative variation in percent. In these plots "data" refers to Asimov data [171, 176] and coincide with the non variated distributions.



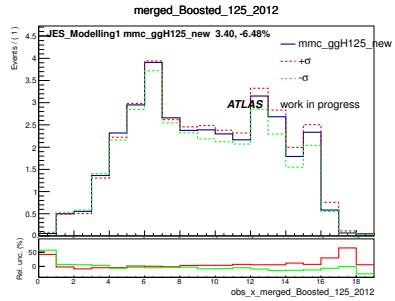
(a) TES true



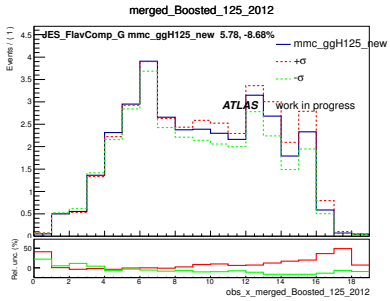
(b) JER



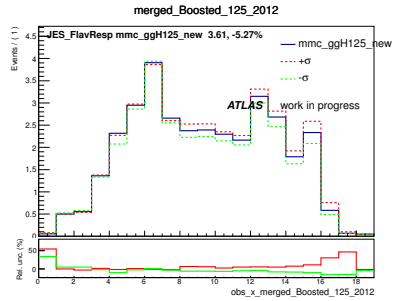
(c) JESEtaModelling



(d) JESModelling1



(e) JESFlavComp



(f) JESFlavResp

Figure 5.25: Effect of $\pm 1\sigma$ systematic shape variation on the MMC mass distribution for the gluon gluon fusion signal in the Boosted category: the TES true component is shown in (a), the JER is shown in (b), the JES components are shown in (c), (d), (e), and (f). The bottom part of the figures shows the relative variation in percent.

Component	Boosted region yield	VBF region yield
$Z \to \tau\tau$	3412 ± 36	56 ± 4
SS/fakes	270 ± 17	11 ± 1
W+jets	148 ± 18	
Rest	191 ± 11	12 ± 2
Total background	4022 ± 364	79 ± 16
VBFH(125)	10.78 ± 0.09	6.55 ± 0.07
$gg \to$ H(125)	34.2 ± 0.5	1.5 ± 0.1
WH(125)	3.85 ± 0.07	0.05 ± 0.004
ZH(125)	1.91 ± 0.04	0.04 ± 0.002
Total signal	50.7 ± 0.5	8.1 ± 0.2

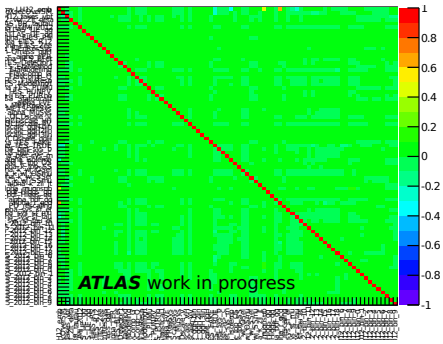
Table 5.8: Expected yields in the signal regions for $H \to \tau_{\text{lep}}\tau_{\text{had}}$, the quoted error is statistics only. The "Rest" contribution includes all the small backgrounds summed together.

Also the errors on nuisance parameter should be carefully examined. Before the fit they are set to the systematic error itself, after the likelihood fit, the error on the nuisance parameter is obtained from $\partial \log \mathcal{L}/\partial\theta$. If the error after the minimization is much smaller compared to the systematic uncertainty assigned to the initial parameter, it may indicate problem in the fit. The uncertainties on the TES, on the fake factor method and on the muon subtraction in the embedding procedure, zoomed in Figures 5.26(c), 5.26(d) and 5.26(e), are slightly constraint but not in a pathological way.

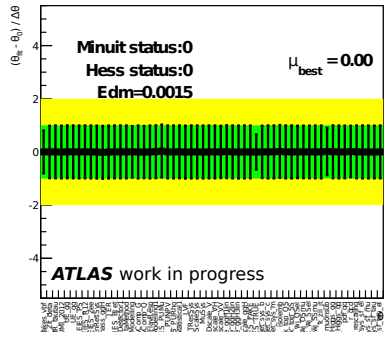
5.4.4 Results

In this section the expected results for the $H \to \tau_{\text{lep}}\tau_{\text{had}}$ search are quoted and discussed. The expected yields in the signal region are quoted in Table 5.8.

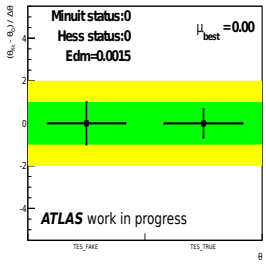
Figure 5.27 shows the exclusion limit on the SM Higgs cross section normalised to the theory cross section for the Boosted and VBF categories separately, in order to estimate their relative contribution to the final analysis sensitivity, and their combination. The expected 95% confidence level limit is shown as a dashed line, then, the green and yellow bands correspond to the 1σ and 2σ error bands. No observed limit is quoted, since the cut-based analysis presented in this thesis is still under approval by the ATLAS collaboration and the datasets are subjected



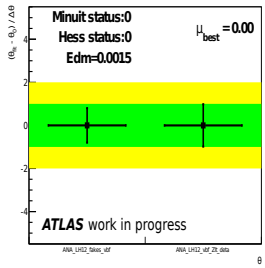
(a) matrix correlation



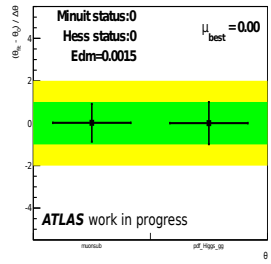
(b) $(\theta_{\text{fit}} - \theta_0)/\Delta\theta$



(c) TES



(d) fakes in VBF



(e) muon subtraction in embeddeding

Figure 5.26: Test of the nuisance parameters in the global fit: the correlation between the nuisance parameters is shown in (a) and the pull and the constraints after the fit are shown for each nuisance parameter in (b). The zoom for the TES uncertainty is shown in (c), for the uncertainty of the fakes background in the VBF category in (d), and for the uncertainty on the muon subtraction in the embedded procedure in (e).

Exp. Lim. m_H [GeV]	100	105	110	115	120	125	130	135	140	145	150
VBF	3.34	2.49	2.11	1.80	1.70	1.82	1.99	1.94	2.63	3.12	4.77
Boosted	6.27	4.13	3.12	2.61	2.21	2.15	2.21	2.35	3.17	4.05	5.66
Combined	1.98	1.73	1.48	1.33	1.25	1.30	1.40	1.42	1.92	2.35	3.47

Table 5.9: Expected exclusion limits separately for the VBF and Boosted categories and their combination for the cut-based $H \to \tau_{\text{lep}}\tau_{\text{had}}$ analysis.

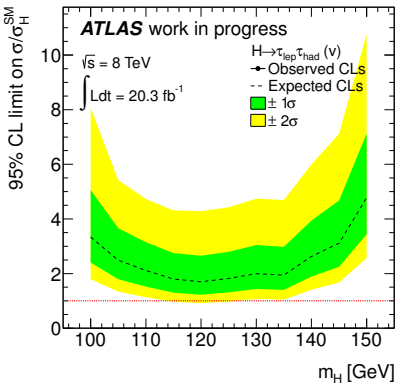
Exp. Sign. m_H [GeV]	100	105	110	115	120	125	130	135	140	145	150
VBF	0.63	0.82	0.99	1.23	1.37	1.29	1.20	1.24	0.90	0.76	0.50
Boosted	0.35	0.57	0.74	0.96	1.06	1.11	1.09	1.03	0.80	0.62	0.45
Combined	1.18	1.32	1.52	1.74	1.86	1.76	1.63	1.61	1.19	0.97	0.67

Table 5.10: Expected significance separately for the VBF and Boosted categories and their combination for the cut-based $H \to \tau_{\text{lep}}\tau_{\text{had}}$ analysis.

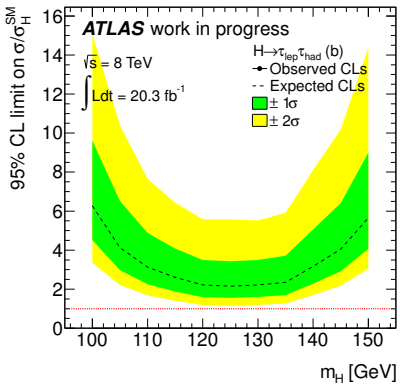
to a blind prescription. The combined expected limit vary between 1.25 and 3.47 times the SM Higgs cross section in the Higgs mass range between 100 and 150 GeV, and it is 1.30 at $m_H = 125$ GeV. As expected this limit is mostly driven by the VBF category that by itself give a limit of 1.82 times the SM Higgs cross section at $m_H = 125$ GeV. All the expected values are quoted in Table 5.9.

The expected local p_0 is also compute and shown in Figure 5.28 separately for the Boosted and the VBF categories and their combination. The expected p-value at $m_H = 125$ GeV is 1.76, all the values are quoted in Table 5.10. The dominion uncertainties on the measurement of the signal strength parameter include: the experimental uncertainty on the TES, the statistical uncertainties in the signal regions and the uncertainties on the QCD multijets background estimate.

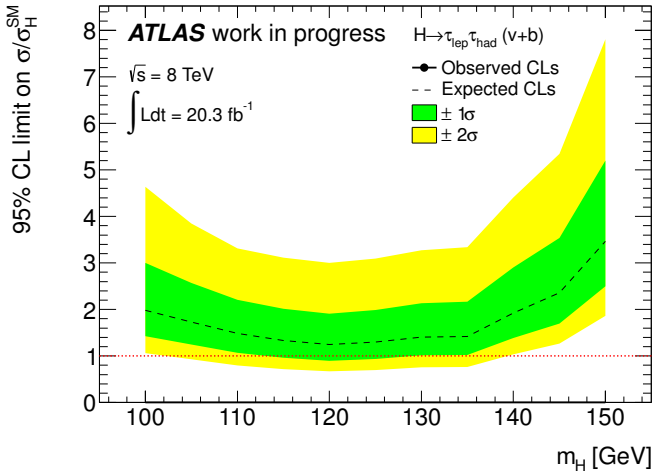
This analysis has to be combined with the other τ decay channels $\tau_{\text{lep}}\tau_{\text{lep}}$ and $\tau_{\text{had}}\tau_{\text{had}}$, but also after this combination the expected sensitivity to the Higgs boson would not reach the needed 3σ level to claim an observation. For this reason, an analysis based on multivariate techniques (MVA) is also developed for this search, and it led to the evidence for the $H \to \tau\tau$ in ATLAS. For completeness, a brief description is given in Section 5.5, a full description can be found in Ref. [143].



(a) VBF



(b) Boosted



(c) Combination

Figure 5.27: Expected exclusion limits for Standard Model Higgs boson production cross section normalised to the theoretical SM prediction as a function of m_H for the VBF category in (a) and for the Boosted categories in (b) and their combination in (c), for the cut-based $H \to \tau_{\text{lep}}\tau_{\text{had}}$ analysis.

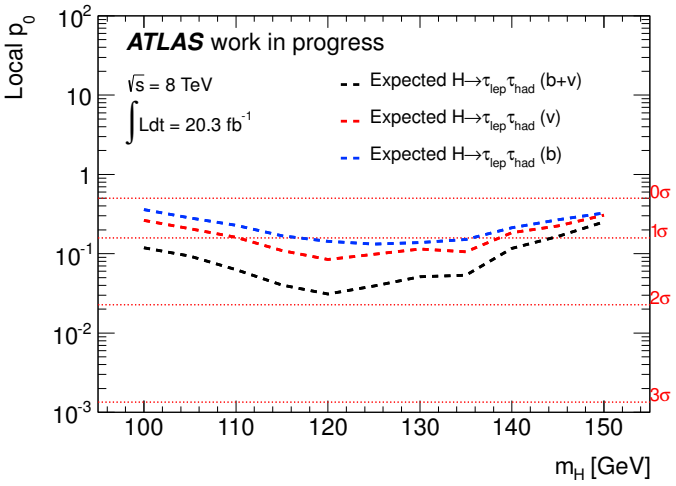


Figure 5.28: Expected local probability as a function of m_H obtained from the combined fit in all categories for the cut-based $H \to \tau_{\text{lep}}\tau_{\text{had}}$ analysis. The breakdown of the sensitivities from VBF and Boosted categories is also shown.

5.5 Multivariate analysis and results

As for the cut-based approach, separate analyses for the three channels $\tau_{\text{lep}}\tau_{\text{lep}}$, $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ are optimized due to the different background compositions and sensitivity to the Higgs boson. Anyway, where possible common criteria and techniques are employed in order to harmonize the analyses for the final combination, common procedures with the cut-based analysis are also used where applicable.

Each channel is split into two mutually exclusive categories, Boosted and VBF, that are defined with looser criteria, reported in Table 5.11, with respect to the ones applied in the cut-based analysis (see Section 5.1.6). To this purpose, Boosted Decision Trees (BDT) are used in each category to extract the Higgs boson signal from the large number of background events. Decision trees [177] recursively partition the parameter space into multiple regions where signal or background purities have been enhanced. Boosting is a method which improves the performance and stability of decision trees and involves the combination of many trees into a single final discriminant [178, 179].

The input discriminating variables for the BDT are the same employed for the category selections in Section 5.1.6, including also the MMC mass. Few

Categroy	Selection	$\tau_{\text{lep}}\tau_{\text{lep}}$	$\tau_{\text{lep}}\tau_{\text{had}}$	$\tau_{\text{lep}}\tau_{\text{had}}$
VBF	$p_{\text{T}}^{j_1}$ [GeV]	40	50	50
	$p_{\text{T}}^{j_2}$ [GeV]	30	50	30/35
	$\Delta\eta_{j_1 j_2}$	2.2	3.0	2.0
	p_{T}^{H} [GeV]	–	–	40
	b-jet veto for jet $p_{\text{T}}^{\text{jet}}$ [GeV]	25	30	–
	m_{vis} [GeV]	–	40	–
Boosted	$p_{\text{T}}^{j_1}$ [GeV]	40	–	–
	p_{T}^{H} [GeV]	100	100	100
	b-jet veto for jet $p_{\text{T}}^{\text{jet}}$ [GeV]	25	30	–

Table 5.11: Selection criteria applied in each analysis category for each channel.

additional variables are also defined:

- the $\boldsymbol{E}_{\text{T}}^{\text{miss}}$ ϕ centrality: quantifies the relative angular position of the $\boldsymbol{E}_{\text{T}}^{\text{miss}}$ with respect to the τ decay products in the transverse plane
- the sphericity: describes the isotropy of energy flow
- the τ_{had}/ℓ η centrality: quantified the η position of either a τ_{had} or an isolated lepton with respect to the two leading jets in the event

A special technique that trains a first BDT on one half of the data sample for then evaluating it on the second half of the events, and that at the same time trains a second separate BDT on the second half of the events for then evaluating it on the first half, ensures that all the available statistics is used without the employment of the same BDT for both training and testing.

Regarding the background estimate, the same techniques described in Section 5.2 are employed also for this analysis, with the exception that the W+jets and QCD multijets backgrounds are evaluated also in the Boosted category using the fake factor method.

Finally, the same signal extraction procedure, described in Section 5.4, with a profile likelihood ratio using the BDT as a discriminating variable, and the same smoothing and pruning procedure for shape systematics, described in Section 5.4.2, are employed to obtain the final results. The fitted value of the signal strength parameter from the likelihood fit is $\mu = 1.43^{+0.31}_{-0.29}(\text{stat.})^{+0.41}_{-0.30}(\text{syst.})$ for

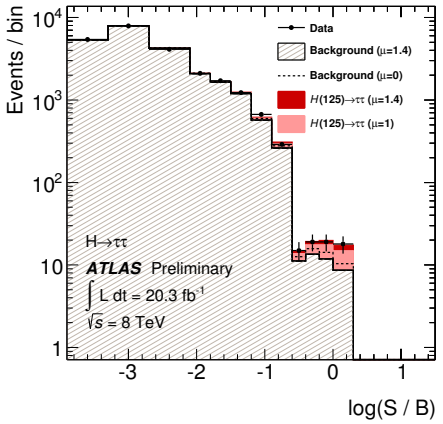


Figure 5.29: Event yields as a function of $\log(S/B)$ for the ATLAS MVA $H \to \tau\tau$. The predicted background is obtained from the global fit (with $\mu = 1.4$), and signal yields are shown for $m_H = 125$ GeV, at $\mu = 1$ and $\mu = 1.4$ (the best-fit value). The background only distribution (dashed line), is obtained from the global fit, but fixing $\mu = 0$ [143].

$m_H = 125$ GeV. The observed p_0 value at $m_H = 125$ GeV is $2.0 \cdot 10^{-5}$, which corresponds to a deviation from the background-only hypothesis of 4.1σ, compared to an expected p_0 value of $6.6 \cdot 10^{-4}$ (3.2σ). This is direct evidence for $H \to \tau\tau$ decays.

Figure 5.29 shows the expected and observed data, in bins of $\log(S/B)$, for all signal region bins. Here, S/B is the signal-to-background ratio calculated assuming $\mu = 1$ for each BDT bin in the signal regions. The expectation is shown for signal yields for both $\mu = 1$ and the best-fit value $\mu = 1.4$ for $m_H = 125$ GeV, are shown on top of the background prediction taken also from the best-fit values. The background expectation where the signal strength parameter has been fixed to $\mu = 0$ is also shown for comparison. A clear excess of events is observed in the most sensitive bins optimized for a VBF Higgs signal. An event containing a Higgs candidate produced via vector boson fusion is shown in Figure 5.30.

Figure 5.31 shows the two-dimensional contours in the plane of $\mu_{ggF} \times B/B_{SM}$ and $\mu_{VBF+VH} \times B/B_{SM}$ for $m_H = 125$ GeV, where B and B_{SM} are the hypothesised and the SM branching ratios for $H \to \tau\tau$. The best-fit values are $\mu_{ggF} \times B/B_{SM} = 1.1^{+1.3}_{-1.0}$ and $\mu_{VBF+VH} \times B/B_{SM} = 1.6^{+0.8}_{-0.7}$.

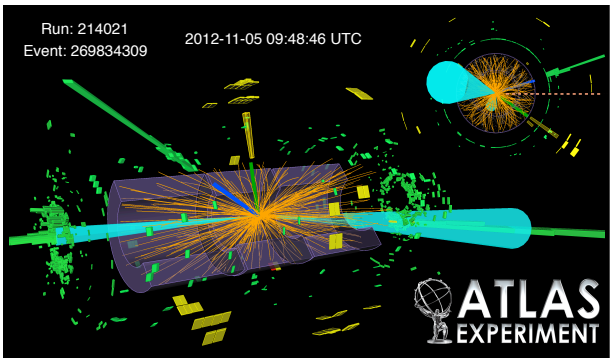


Figure 5.30: Display of an event selected by the ATLAS MVA $H \to \tau_{\text{lep}}\tau_{\text{had}}$ channel in the VBF category, where one τ decays to an electron (the blue track match to the green cluster). The hadronically decaying τ lepton (1 prong decay) is indicated by a green track and the yellow cluster. The two jets are marked with turquoise cones. $p_{\text{T}}^{\ell} = 56$ GeV, $p_{\text{T}}^{\tau} = 27$ GeV, $E_{\text{T}}^{\text{miss}} = 113$ GeV, $m_{j_1 j_2} = 1.53$ TeV, $m_{\tau\tau}^{\text{MMC}} = 129$ GeV, and the BDT score is 0.99. The S/B ratio in the BDT score bin of this event is 1.0 [143].

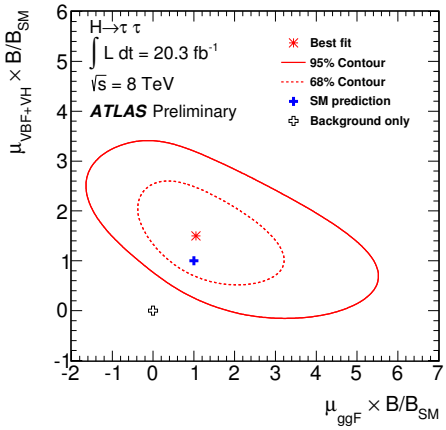


Figure 5.31: Likelihood contours for the ATLAS MVA $H \to \tau\tau$ in the ($\mu_{ggF} \times B/B_{SM}$, $\mu_{VBF+VH} \times B/B_{SM}$) plane are shown for the 68% and 95% CL by dashed and solid lines, respectively, for $m_H = 125$ GeV. The SM expectation and the one corresponding to background-only hypothesis are shown by a filled plus and an open plus symbol, respectively. The best-fit to the data is shown for the case when both the μ_{ggF} and μ_{VBF+VH} are unconstrained [143].

5.6 Conclusions and prospects

Separate cut-based and multivariate analyses were developed for the search for a Higgs boson of mass 125 GeV decaying into the $\tau\tau$ final state using the full ATLAS 2012 dataset, corresponding to an integrated luminosity of $\mathcal{L} = 20.3$ fb^{-1} of p-p collisions at a center of mass energy of 8 TeV.

In this chapter the cut-based analysis was described in full details for the semileptonic decay channel. A brief overview of the multivariate analysis is also given.

The multivariate analysis measured a signal with a significance of 4.1 standard deviations, compared with an expected significance of 3.2 standard deviations. This constitutes direct evidence of the decay of the Higgs boson to fermions: the observed signal strength $\mu = 1.4^{+0.5}_{-0.4}$ is compatible with the Standard Model expectation. The next important step consists in the measurement of the Higgs properties in this channel, first of all the mass measurement.

The cut-based analysis provides a lower sensitivity but has to be maintained because it can be an important confirmation for the observation of the Higgs boson in this channel. In order to correctly interpret the cut-based result an accurate study about the correlation between the MVA and cut-based analysis should be performed. Moreover the current cut-based analysis could be further optimized redesigning the categorization procedure with more subdivisions in order to isolate more specifically each Higgs production mode and peculiar properties that can increase the sensitivity of the analysis. Improvements in the background estimation techniques are as well under investigation.

Chapter 6

Conclusion

The observation of the decay of the Higgs to a τ-pair in ATLAS, with a significance of 4.1σ, gives the first direct evidence of the Higgs coupling to fermions. It therefore provides an important confirmation of the Standard Model and of the mass generating mechanism for fermionic particles.

The result achieved has been made possible by the excellent performance during the whole data taking period in 2011 and 2012 (Run1), of the ATLAS detector and of the LHC which has provided p-p collisions at energies and luminosities never reached before. In these high-luminosity conditions the biggest experimental challenge is the control of the effects introduced by pile-up on the detector signals. In particular, $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction includes soft energy contributions that need to be extracted from pile-up. The pile-up suppression techniques studied and implemented in ATLAS and described in this thesis allowed to restore the $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ resolution to values close to the ones observed in the absence of pile-up.

These $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ pile-up suppression techniques can be used in the $H \to \tau\tau$ search, with the benefit of a better $\tau\tau$ mass resolution and of a stronger rejection of the main $Z \to \tau\tau$ background, thus increasing the analysis sensitivity. The ATLAS final results are obtained with the employment of a $H \to \tau\tau$ analysis based on multivariate techniques, which provides the optimal sensitivity. However, the cut-based analysis discussed in this thesis, which has the advantage to be simpler and robust, can provide not only an important confirmation for the observation of the Higgs boson in this channel, but can also be useful in the future for a mass measurement. In particular, it is shown that an expected significance of 1.76σ can be reached by the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel at $m_H = 125$ GeV.

The obvious extension of this analysis is the measurement of the Higgs properties in the $\tau\tau$ channel, because any deviation from the Standard Model predictions, especially in the coupling measurement, can hint to the presence of new physics. In the measurement of the Higgs mass, improvements in the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale can be important to achieve more precise results. Therefore, ongoing activity is devoted to the optimization of the pile-up suppression methods for the re-analysis of the whole 2011 and 2012 dataset (including collisions at $\sqrt{s} = 7$ TeV). These optimization efforts are even more important for the preparation for LHC Run 2, where higher energy and luminosities will make the reconstruction more challenging. In the expected more hostile environments, pile-up suppression methods will be essential to fully exploit the physics potential of ATLAS.

Appendix A

Data selection and MC simulation samples for $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ studies

The $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ performance are studies in a large variety of topologies in order to fully test the reconstruction capability for different physics objects. The selection criteria and the MC simulation samples used for the different topologies are listed in the following.

Minimum bias

Minimum bias events were selected both by a random trigger and by the minimum bias trigger scintillators (MBTS), which are mounted at each end of the detector in front of the LAr end-cap calorimeter cryostats 2.4.4. For each event, at least one good primary vertex is required with a z displacement from the nominal p-p interaction point of less than 200 mm and with at least five associated tracks.

$Z \to \ell\ell$ selection

Candidate $Z \to \ell\ell$ events, where ℓ is an electron or a muon, are required to pass an electron, photon or muon trigger with a transverse momentum, p_{T}, threshold between 15 and 20 GeV, where the exact trigger selection varies depending on the data period analysed. For each event, at least one good primary vertex is

required with a z displacement from the nominal pp interaction point of less than 200 mm and with at least three associated tracks.

The selection of $Z \to \mu\mu$ events requires the presence of exactly two good muons. A good muon is defined to be a muon reconstructed in the muon spectrometer with a matched track in the inner detector with transverse momentum above 25 GeV and $|\eta| < 2.5$. Additional requirements on the number of hits used to reconstruct the tracks in the inner detector are applied. The z displacement of the muon tracks from the primary vertex is required to be less than 10 mm. Isolation criteria are applied around the muon track.

The selection of $Z \to ee$ events requires the presence of exactly two identified electrons with $|\eta| < 2.47$, which pass the"medium" identification criteria (section 3.2.1), optimized for 2012 data, and have trans- verse momenta above 25 GeV. Electron candidates in the electromagnetic calorimeter transition region, $1.37 < |\eta| < 1.52$, are not considered for this study.

In both the $Z \to ee$ and the $Z \to \mu\mu$ selections, the two leptons are required to have opposite charge and the reconstructed invariant mass of the dilepton system, $m_{\ell\ell}$, is required to be consistent with the Zmass, $66 < m_{\ell\ell} < 116$ GeV.

$W \to \ell\nu$ selection

Lepton candidates are selected with lepton identification criteria similar to those used for the Z selection. An isolation cut is applied around the electron energy deposits in the calorimeter to reduce contamination from jets. The event is required to contain exactly one reconstructed lepton (electron or muon). The $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$, is required to be greater than 25 GeV. The reconstructed mass of the transverse momentum of the lepton, $\boldsymbol{p}_{\mathrm{T}}^{\ell}$, and $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$:

$$m_{\mathrm{T}} = 2p_{\mathrm{T}}^{\ell} E_{\mathrm{T}}^{\mathrm{miss}}(1 - \cos\phi) \tag{A.1}$$

where ϕ is the azimuthal angle between the lepton momentum and $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}$ directions, must satisfy $m_{\mathrm{T}} > 50$ GeV.

Monte Carlo simulation samples

Monte Carlo (MC) samples of $Z \to \ell\ell$ and $W \to \ell\nu$ production are generated with the next-to-leading (NLO) order Powheg [87] model, with the final state partons showered by the Pythia8 program [152, 180], using the CT10 next-to-leading order (NLO) parton distribution function (PDF) [158] and the ATLAS

AU2 tune [181]. Samples of $Z \to \ell\ell$ generated with Alpgen [153] are also used for some additional data-MC comparison.

Additional inelastic pp interactions are generated using the Pythia8 program with the ATLAS MC12 A2M tune [181] and the MSTW08 leading order (LO) PDF [169]. The proton-proton bunches are organized in trains, with 50 ns spacing between bunches, closely matching the bunch structure of the LHC. The MC simulation samples are weighted such that the distribution of the average number of interactions per bunch crossing matches that observed in the 2012 data sample, to ensure that the pile-up interactions are accurately described. When the pile-up conditions are not specified for a given figure, they should be assumed to be matched to those observed in the 2012 data sample used.

Appendix B

Summary of the systematics uncertainties used in the fit

In the following tables the systematics uncertainties used in the fit model are summarized. A brief description is provided for each uncertainty.

Workspace name	description
ANA_LH12_Fake	uncertainty on Fake Factor method in VBF category
BTag_BEFF	uncertainty on b-tagging efficiency (b-jets)
BTag_CEFF	uncertainty on b-tagging efficiency (c-jets)
BTag_LEFF	uncertainty on b-tagging efficiency (light-jets)
EL_EFF	electron identification efficiency uncertainty
EL_EFF_Emb	uncertainty on difference in embedding normalization for SLT and LTT events
EL_RES	electron energy resolution uncertainty
EL_SCALE	electron energy scale uncertainty
JVF	systematics on jet-vertex-fraction cut
LUMI_2012	systematics on measured integrated luminosity in 2012
TAU_EFAKE	uncertainty on $e \to \tau$ misidentification probability
TAU_ID	uncertainty for tau identification efficiency
TAU_MFAKE	uncertainty on $\mu \to \tau$ misidentification probability
TES_FAKE	uncertainty on tau energy scale for fake candidates
TES_TRUE	uncertainty on tau energy scale for true candidates
TRIGGER_LH_2012	uncertainty on the lep-had trigger efficiencies
TRIGGER_LH_2012_Emb	uncertainty on the lep-had trigger efficiencies for embedding
TRIGGER_LH_2012_Fake	uncertainty on the lep-had trigger efficiencies on fake sample

Table B.1: Summary of systematic uncertainties used in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel.

Workspace name	description
JER	uncertainty on jet energy resolution
JES_Detector1	JES uncertainty component
JES_Eta_StatMethod	JES uncertainty component
JES_Modelling1	JES uncertainty component
JES_PileRho_TAU_GG	JES uncertainty component, applies to Top and $gg \to H$
JES_PileRho_TAU_QG	JES uncertainty component, applies to $W \to l\nu$ and Z+jets
JES_PileRho_TAU_QQ	JES uncertainty component, applies to VBF Higgs, VV, WH, ZH
JES_Statistical1	JES uncertainty component
JES_Eta_Modelling	JES uncertainty component
JES_FlavComp_TAU_G	JES uncertainty component, applies to $gg \to H$, $W \to l\nu$ and Z+jets
JES_FlavComp_TAU_Q	JES uncertainty component, applies to Top, VBF Higgs, VV, WH, ZH
JES_FlavResp	JES uncertainty component
JES_Flavb	JES uncertainty component
JES_Mu	JES uncertainty component
JES_NPV	JES uncertainty component
MET_RESOSOFT	MET resolution uncertainty on the soft term
MET_SCALESOFT	MET scale uncertainty on the soft term
MU_EFF	muon identification efficiency uncertainty
MU_EFF_Emb	uncertainty on difference in embedding normalization for SLT and LTT events
MU_SCALE	muon momentum scale uncertainty
PU_RESCALE	pileup reweighting uncertainty
BR_tautau	uncertainty on $H \to \tau\tau$ BR
Gen_Qmass_ggH	b- and top quark mass effect on $p_T(H)$ spectrum in $gg \to H$
QCDscale_V	uncertainty for W/Z+jets acceptance from QCD scale
QCDscale_VH	uncertainty for VH acceptance from QCD scale
QCDscale_ggH	uncertainty for $gg \to H$ (≥ 0 jet) acceptance from QCD scale
QCDscale_ggH1in	uncertainty for $gg \to H$ (≥ 1 jet) acceptance from QCD scale
QCDscale_ggH2in	uncertainty for $gg \to H$ (≥ 2 jet) acceptance from QCD scale
QCDscale_ggH3in	uncertainty for $gg \to H$ (≥ 3 jet) acceptance from QCD scale
QCDscale_qqH	uncertainty for VBF acceptance from QCD scale
pdf_Higgs_gg	PDF uncertainty on ggF production
pdf_Higgs_qq	PDF uncertainty on VBF/VH production
pdf_qq	PDF uncertainty on MC-based background samples
UE_gg	Underlying event uncertainty on ggF
UE_qq	Underlying event uncertainty on VBF

Table B.2: Summary of systematic uncertainties used in the $\tau_{\text{lep}}\tau_{\text{had}}$ channel.

Bibliography

[1] The ATLAS Collaboration, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[2] CMS Collaboration Collaboration, S. Chatrchyan et al., *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[3] LHCb Collaboration Collaboration, J. Alves, A. Augusto et al., *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

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

[5] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509. http://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[6] The ATLAS Collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys.Lett. **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex].

[7] CMS Collaboration Collaboration, S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys.Lett. **B716** (2012) 30–61, arXiv:1207.7235 [hep-ex].

[8] Gargamelle Neutrino Collaboration Collaboration, F. Hasert et al., *Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment*, Phys.Lett. **B46** (1973) 138–140.

[9] UA1 Collaboration Collaboration, G. Arnison et al., *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV*, Phys.Lett. **B122** (1983) 103–116.

[10] UA2 Collaboration Collaboration, M. Banner et al., *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider*, Phys.Lett. **B122** (1983) 476–485.

[11] Super-Kamiokande Collaboration Collaboration, Y. Ashie et al., *Evidence for an oscillatory signature in atmospheric neutrino oscillation*, Phys.Rev.Lett. **93** (2004) 101801, arXiv:hep-ex/0404034 [hep-ex].

[12] SNO Collaboration Collaboration, Q. Ahmad et al., *Measurement of the rate of nu/e + d → p + p + e- interactions produced by B-8 solar neutrinos at the Sudbury Neutrino Observatory*, Phys.Rev.Lett. **87** (2001) 071301, arXiv:nucl-ex/0106015 [nucl-ex].

[13] OPERA Collaboration Collaboration, N. Agafonova et al., *Observation of a first ν_τ candidate in the OPERA experiment in the CNGS beam*, Phys.Lett. **B691** (2010) 138–145, arXiv:1006.1623 [hep-ex].

[14] CERN Collaboration, "Standard model standard infographic.". https://cds.cern.ch/journal/CERNBulletin/2012/35/News%20Articles/1473657.

[15] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*,.

[16] C. Becchi and G. Ridolfi, *An introduction to relativistic processes and the standard model of electroweak interactions*,.

[17] S. Novaes, *Standard model: An Introduction*, arXiv:hep-ph/0001283 [hep-ph].

[18] R. P. Feynman, *Quantum Electrodynamics*,.

[19] N. Gribov and J. Nyiri, *Quantum Electrodynamics*,.

[20] R. K. Ellis, W. J. Stirling, and B. Webber, *QCD and collider physics*, Camb.Monogr.Part.Phys.Nucl.Phys.Cosmol. **8** (1996) 1–435.

[21] P. Nason, *Introduction to QCD*, Conf.Proc. **C9705251** (1997) 94–149.

[22] G. Prosperi, M. Raciti, and C. Simolo, *On the running coupling constant in QCD*, Prog.Part.Nucl.Phys. **58** (2007) 387–438, arXiv:hep-ph/0607209 [hep-ph].

[23] D. Gross and F. Wilczek, *Asymptotically Free Gauge Theories. 1*, Phys.Rev. **D8** (1973) 3633–3652.

[24] D. Gross and F. Wilczek, *ASYMPTOTICALLY FREE GAUGE THEORIES. 2.*, Phys.Rev. **D9** (1974) 980–993.

[25] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys.Rev.Lett. **30** (1973) 1346–1349.

[26] S. Bethke, *Experimental tests of asymptotic freedom*, Prog.Part.Nucl.Phys. **58** (2007) 351–386, arXiv:hep-ex/0606035 [hep-ex].

[27] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group Collaboration, S. Schael et al., *Precision electroweak measurements on the Z resonance*, Phys.Rept. **427** (2006) 257–454, arXiv:hep-ex/0509008 [hep-ex].

[28] R. Turra, *Energy calibration and observation of the Higgs boson in the diphoton decay with the ATLAS experiment*,.

[29] LHC Higgs Cross Section Working Group Collaboration, S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, arXiv:1101.0593 [hep-ph].

[30] S. Dittmaier, S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka, et al., *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, arXiv:1201.3084 [hep-ph].

[31] LHC Higgs Cross Section Working Group Collaboration, S. Heinemeyer et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, arXiv:1307.1347 [hep-ph].

[32] CERN Collaboration, "Higgs production diagrams.". http://sites.uci.edu/energyobserver/2012/11/26/higgs-production-and-decay-channels.

[33] ATLAS Collaboration, "Standard model higgs production cross-section and branching ratios.". https://twiki.cern.ch/twiki/bin/view/AtlasPublic/HiggsTheoryPlots.

[34] The 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.". https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2013-014.

[35] The ATLAS Collaboration, G. Aad et al., *Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC*, Phys.Lett. **B726** (2013) 88–119, arXiv:1307.1427 [hep-ex].

[36] The ATLAS Collaboration, G. Aad et al., *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, Phys.Lett. **B726** (2013) 120–144, arXiv:1307.1432 [hep-ex].

[37] The ATLAS Collaboration, *Study of the spin of the Higgs-like boson in the two photon decay channel using 20.7 fb-1 of pp collisions collected at $\sqrt{s}$ = 8 TeV with the ATLAS detector*,.

[38] The ATLAS Collaboration, *Study of the spin properties of the Higgs-like particle in the $H \to WW^{(*)} \to e\nu\mu\nu$ channel with 21 fb^{-1} of $\sqrt{s} = 8$ TeV data collected with the ATLAS detector.*,.

[39] L. D. Landau Collaboration, *The moment of a 2-photon system.*, Dokl. Akad. Nauk SSSR **60** (1948) 207.

[40] C.-N. Yang, *Selection Rules for the Dematerialization of a Particle Into Two Photons*, Phys.Rev. **77** (1950) 242–245.

[41] A. Einstein, *The Formal Foundation of the General Theory of Relativity*, Sitzungsber.Preuss.Akad.Wiss.Berlin (Math.Phys.) **1914** (1914) 1030–1085.

[42] A. Einstein and N. Rosen, *The Particle Problem in the General Theory of Relativity*, Phys.Rev. **48** (1935) 73–77.

[43] A. Buras, J. R. Ellis, M. Gaillard, and D. V. Nanopoulos, *Aspects of the Grand Unification of Strong, Weak and Electromagnetic Interactions*, Nucl.Phys. **B135** (1978) 66–92.

[44] L. E. Ibanez, *Grand Unification, Supersymmetry, Superstrings: An Introduction to Physics Beyond the Standard Model*, NATO Adv.Study Inst.Ser.C.Math.Phys.Sci. **20** (1990) 59–115.

[45] S. P. Martin, *A Supersymmetry primer*, arXiv:hep-ph/9709356 [hep-ph].

[46] I. J. Aitchison, *Supersymmetry and the MSSM: An Elementary introduction*, arXiv:hep-ph/0505105 [hep-ph].

[47] The ATLAS Collaboration, "Susy public results.". https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SUSY/.

[48] Planck Collaboration Collaboration, P. Ade et al., *Planck 2013 results. I. Overview of products and scientific results*, arXiv:1303.5062 [astro-ph.CO].

[49] J. L. Feng, *Dark Matter Candidates from Particle Physics and Methods of Detection*, Ann.Rev.Astron.Astrophys. **48** (2010) 495–545, arXiv:1003.0904 [astro-ph.CO].

[50] The ATLAS Collaboration, "Exotics public results.". https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/EXOTICS.

[51] P. B. Pal, *Dirac, Majorana and Weyl fermions*, Am. J. Phys. **79** (2011) 485–498, arXiv:1006.1718 [hep-ph].

[52] K. Agashe, R. Contino, and A. Pomarol, *The Minimal composite Higgs model*, Nucl.Phys. **B719** (2005) 165–187, arXiv:hep-ph/0412089 [hep-ph].

[53] T. Han, H. E. Logan, B. McElrath, and L.-T. Wang, *Phenomenology of the little Higgs model*, Phys.Rev. **D67** (2003) 095004, arXiv:hep-ph/0301040 [hep-ph].

[54] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[55] O. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, et al., *LHC Design Report. 1. The LHC Main Ring*,. http://cds.cern.ch/record/782076.

[56] O. Buning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, et al., *LHC Design Report. 2. The LHC infrastructure and general services*,. http://cds.cern.ch/record/815187.

[57] S. Myers and E. Picasso, *The Design, construction and commissioning of the CERN Large Electron Positron collider*, Contemp.Phys. **31** (1990) 387–403.

[58] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl, *LHC Design Report. 3. The LHC injector chain*,. http://cds.cern.ch/record/823808.

[59] CERN Collaboration, "Machine parameters.". http://cerncourier.com/cws/article/cern/54381.

[60] The ATLAS Collaboration, "Luminosity public results.". https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[61] ALICE Collaboration Collaboration, K. Aamodt et al., *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[62] J. Stirling, "Cross sections as a function of the center-of-,ass energy.". http://mstwpdf.hepforge.org/plots/plots.html.

[63] The ATLAS Collaboration, "Particles interaction in the atlas detector.". http://www.interactions.org/imagebank/images/CE0155M.jpg.

[64] The ATLAS Collaboration, "Atlas fact sheet.". https://cds.cern.ch/record/1457044/.

[65] ATLAS Collaboration, G. Aad et al., *The ATLAS Inner Detector commissioning and calibration*, Eur.Phys.J. **C70** (2010) 787–821, arXiv:1004.5293 [physics.ins-det].

[66] The ATLAS Collaboration, "Atlas inner detector.". http://www.atlas.ch/photos/inner-detector-combined.html.

[67] The ATLAS Collaboration, "Atlas inner detector.". http://www.atlas.ch/photos/inner-detector-combined.html.

[68] ATLAS Collaboration, M. Alam et al., *ATLAS pixel detector: Technical design report*, 1998.

[69] ATLAS TRT Collaboration, E. Abat et al., *The ATLAS TRT barrel detector*, JINST **3** (2008) P02014.

[70] E. Abat, E. Arik, M. Arik, N. Becerici, O. Dogan, et al., *The ATLAS TRT end-cap detectors*, JINST **3** (2008) P10003.

[71] The ATLAS Collaboration, "Event displays.". https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayStandAlone.

[72] The ATLAS Collaboration, "Calorimeter system.". http://www.fsp101-atlas.de/e197881/e200194/.

[73] The ATLAS Collaboration, G. Aad et al., *Readiness of the ATLAS Liquid Argon Calorimeter for LHC Collisions*, Eur.Phys.J. **C70** (2010) 723–753, arXiv:0912.2642 [physics.ins-det].

[74] CERN, *ATLAS: Detector and physics performance technical design report. Volume 2*, 1999.

[75] R. Wigmans, *Calorimetry: Energy measurement in particle physics*, Int.Ser.Monogr.Phys. **107** (2000) 1–726.

[76] R. Brown and D. Cockerill, *Electromagnetic calorimetry*, Nucl.Instrum.Meth. **A666** (2012) 47–79.

[77] ATLAS Collaboration, G. Aad et al., *Readiness of the ATLAS Tile Calorimeter for LHC collisions*, Eur.Phys.J. **C70** (2010) 1193–1236, arXiv:1007.5423 [physics.ins-det].

[78] The ATLAS Collaboration, "Lar public results.". https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LArCaloPublicResultsDetStatus.

[79] The ATLAS Collaboration, G. Aad et al., *Performance of the ATLAS Detector using First Collision Data*, JHEP **1009** (2010) 056, arXiv:1005.5254 [hep-ex].

[80] The ATLAS Collaboration, G. Duckeck et al., *ATLAS computing: Technical design report*, 2005.

[81] F. Akesson and E. Moyse, *Event data model in ATLAS*, 2005.

[82] G. Barrand, I. Belyaev, P. Binko, M. Cattaneo, R. Chytracek, et al., *GAUDI - A software architecture and framework for building HEP data processing applications*, Comput.Phys.Commun. **140** (2001) 45–55.

[83] The ATLAS Collaboration, G. Aad et al., *The ATLAS Simulation Infrastructure*, Eur.Phys.J. **C70** (2010) 823–874, arXiv:1005.4568 [physics.ins-det].

[84] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput.Phys.Commun. **178** (2008) 852–867, arXiv:0710.3820 [hep-ph].

[85] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **0307** (2003) 001, arXiv:hep-ph/0206293 [hep-ph].

[86] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al., *Event generation with SHERPA 1.1*, JHEP **0902** (2009) 007, arXiv:0811.4622 [hep-ph].

[87] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **0711** (2007) 070, arXiv:0709.2092 [hep-ph].

[88] The ATLAS Collaboration, *New ATLAS event generator tunes to 2010 data*, 2011.

[89] The ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, 2012.

[90] Z. Was, *TAUOLA the library for tau lepton decay, and KKMC / KORALB / KORALZ /... status report*, Nucl.Phys.Proc.Suppl. **98** (2001) 96–102, arXiv:hep-ph/0011305 [hep-ph].

[91] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS Interface in C++: Technical and Physics Documentation*, arXiv:1011.0937 [hep-ph].

[92] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl.Instrum.Meth. **A506** (2003) 250–303.

[93] J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Dubois, et al., *Geant4 developments and applications*, IEEE Trans.Nucl.Sci. **53** (2006) 270.

[94] The ATLAS Collaboration, W. Lukas, *Fast Simulation for ATLAS: Atlfast-II and ISF*, J.Phys.Conf.Ser. **396** (2012) 022031.

[95] J. Chapman, K. Assamagan, P. Calafiura, D. Chakraborty, D. Costanzo, et al., *The ATLAS detector digitization project for 2009 data taking*, J.Phys.Conf.Ser. **219** (2010) 032031.

[96] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, et al., *Calorimeter clustering algorithms: Description and performance*,.

[97] The ATLAS Collaboration, T. Barillari et al., *Local hadronic calibration*,.

[98] ATLAS Liquid Argon Endcap Collaboration Collaboration, J. Pinfold et al., *Evaluation of the local hadronic calibration with combined beam-test data for the endcap and forward calorimeters of ATLAS in the pseudorapidity region* $2.5 < |\eta| < 4.0$, Nucl.Instrum.Meth. **A693** (2012) 74–97.

[99] N. Boelaert, *Jet Reconstruction*, in *Dijet Angular Distributions in Proton-Proton Collisions*, pp. , 83–98. Springer Berlin Heidelberg, 2012. http://dx.doi.org/10.1007/978-3-642-24597-8_7.

[100] ATLAS Collaboration Collaboration, G. Aad et al., *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, arXiv:0901.0512 [hep-ex].

[101] ATLAS Collaboration Collaboration, *Measurements of the photon identification efficiency with the ATLAS detector using 4.9 fb?1 of pp collision data collected in 2011*,.

[102] The ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur. Phys. J. C **72** (2012) 1909, arXiv:1110.3174.

[103] ATLAS Collaboration Collaboration, G. Aad et al., *Measurement of the inclusive isolated prompt photon cross section in pp collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector using 4.6 fb-1*, arXiv:1311.1440 [hep-ex].

[104] D. Banfi, L. Carminati, and L. Mandelli, *Calibration of the ATLAS electromagnetic calorimeter using calibration hits*,.

[105] ATLAS Collaboration Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*,.

[106] The ATLAS Collaboration, "Electron public results.". https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ElectronGammaPublicCollisionResult

[107] T. A. collaboration, *Preliminary results on the muon reconstruction efficiency, momentum resolution, and momentum scale in ATLAS 2012 pp collision data*,.

[108] The ATLAS Collaboration, "Muon public results.". https://twiki.cern.ch/twiki/bin/view/AtlasPublic/MuonPerformancePublicPlots.

[109] J. B. et al., *Particle Data Group*, Phys. Rev (2012), 0901.0512. http://pdg.lbl.gov/.

[110] T. A. collaboration, *Identification of the Hadronic Decays of Tau Leptons in 2012 Data with the ATLAS Detector*,.

[111] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **0804** (2008) 063, arXiv:0802.1189 [hep-ph].

[112] ATLAS Collaboration Collaboration, *Performance of the Reconstruction and Identification of Hadronic Tau Decays in ATLAS with 2011 Data*,.

[113] ATLAS Collaboration Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons*,.

[114] T. A. collaboration, *Determination of the tau energy scale and the associated systematic uncertainty in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector at the LHC in 2012*,.

[115] M. Tannenbaum, *Hard-scattering and Jets from RHIC to LHC: A Critical review*, PoS **LHC07** (2007) 004, arXiv:0707.1706 [nucl-ex].

[116] J. E. Huth, N. Wainer, K. Meier, N. Hadley, F. Aversa, et al., *Toward a standardization of jet definitions*,.

[117] G. P. Salam, *Towards Jetography*, Eur.Phys.J. **C67** (2010) 637–686, arXiv:0906.1833 [hep-ph].

[118] Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, *Better jet clustering algorithms*, JHEP **9708** (1997) 001, arXiv:hep-ph/9707323 [hep-ph].

[119] S. Catani, Y. L. Dokshitzer, M. Seymour, and B. Webber, *Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*, Nucl.Phys. **B406** (1993) 187–224.

[120] T. A. collaboration, *Pile-up subtraction and suppression for jets in ATLAS*,.

[121] T. A. collaboration, *Performance of pile-up subtraction for jet shapes*,.

[122] A. Altheimer, A. Arce, L. Asquith, J. Backus Mayes, E. Bergeaas Kuutmann, et al., *Boosted objects and jet substructure at the LHC*, arXiv:1311.2708 [hep-ex].

[123] ATLAS Collaboration Collaboration, *In situ jet pseudorapidity intercalibration of the ATLAS detector using dijet events in sqrts=7 TeV proton-proton 2011 data*,.

[124] ATLAS Collaboration Collaboration, *TeV-scale jet energy calibration using multijet events including close-by jet effects at the ATLAS experiment*,.

[125] ATLAS Collaboration Collaboration, *Jet energy scale and its systematic uncertainty in proton-proton collisions at sqrt(s)=7 TeV with ATLAS 2011 data*,.

[126] The ATLAS Collaboration, "Jet public results.". https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetEtmissApproved2013JESUncertainty.

[127] The ATLAS Collaboration, *Performance of missing transverse momentum reconstruction in proton-proton collisions at 7 TeV in ATLAS*, Eur. Phys. J. C **72** (2012) 1844, arXiv:1108.5602.

[128] The ATLAS Collaboration, "Performance of missing transverse momentum reconstruction in atlas with 2011 proton-proton collisions at $\sqrt{s} = 7$ tev." ATLAS-CONF-2012-101.

[129] The ATLAS Collaboration, "Performance in missing transverse momentum reconstruction in atlas in 2012 proton-proton collisions at $\sqrt{s} = 8$ TeV." In prepration.

[130] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[131] The ATLAS Collaboration, "Performance of the missing transverse energy reconstruction in minimum bias events at $\sqrt{s}$ of 7 tev with the atlas detector." ATLAS-CONF-2010-039.

[132] The ATLAS Collaboration, "Performance of the missing transverse energy reconstruction in minimum bias events at $\sqrt{s}$ of 900 gev and 2.36 tev with the atlas detector." ATLAS-CONF-2010-008.

[133] The ATLAS Collaboration, "Performance of the missing transverse energy reconstruction and calibration in proton-proton collisions at $\sqrt{s}$ of 7 tev with the atlas detector." ATLAS-CONF-2010-057.

[134] The ATLAS Collaboration, "Muon reconstruction performance." ATLAS-CONF-2010-064.

[135] The ATLAS Collaboration, "Pileup subtraction and suppression for jets." In preparation.

[136] The ATLAS Collaboration, "Measurement of track-based missing 1 transverse momentum in proton-proton collisions at $\tilde{A}s = 8$ tev centre-of-mass energy with the atlas detector." ATL-COM-PHYS-2013-1577.

[137] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys.Lett. **B659** (2008) 119–126, arXiv:0707.1378 [hep-ph].

[138] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys.Rev. **D48** (1993) 3160–3166, arXiv:hep-ph/9305266 [hep-ph].

[139] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, arXiv:hep-ph/9907280 [hep-ph].

[140] M. Cacciari, G. P. Salam, and G. Soyez, *The Catchment Area of Jets*, JHEP **0804** (2008) 005, arXiv:0802.1188 [hep-ph].

[141] M. Cacciari, G. P. Salam, and G. Soyez. http://fastjet.fr/.

[142] The ATLAS Collaboration, *Pile-up corrections for jets from proton-proton collisions at $\sqrt{s} = 7$ TeV in ATLAS in 2011*, 2012.

[143] ATLAS Collaboration, *Evidence for Higgs Boson Decays to the $\tau^+\tau^-$ Final State with the ATLAS Detector*,.

[144] ATLAS Collaboration, G. Aad et al., *Search for the Standard Model Higgs boson in the H to $\tau^+\tau^-$ decay mode in $\sqrt{s} = 7$ TeV pp collisions with ATLAS*, JHEP **1209** (2012) 070, arXiv:1206.5971 [hep-ex].

[145] ATLAS Collaboration, *Search for Standard Model $H \to \tau_{\rm lep}\tau_{\rm had}$ with the ATLAS Detector in 8 TeV Proton-Proton Collisions*,.

[146] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **0411** (2004) 040, arXiv:0409146.

[147] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **1006** (2010) 043, arXiv:1002.2581 [hep-ph].

[148] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO Higgs boson production via gluon fusion matched with shower in POWHEG*, JHEP **0904** (2009) 002, arXiv:0812.0578 [hep-ph].

[149] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, JHEP **1002** (2010) 037, arXiv:0911.5299 [hep-ph].

[150] G. Bozzi, S. Catani, D. de Florian, and M. Grazzini, *Higgs boson production at the LHC: Transverse-momentum resummation and rapidity dependence*, Nucl.Phys. **B791** (2008) 1–19, arXiv:0705.3887 [hep-ph].

[151] D. de Florian, G. Ferrera, M. Grazzini, and D. Tommasini, *Transverse-momentum resummation: Higgs boson production at the Tevatron and the LHC*, JHEP **1111** (2011) 064, arXiv:1109.2109 [hep-ph].

[152] R. Corke and T. Sjostrand, *Interleaved Parton Showers and Tuning Prospects*, JHEP **032** (2011) 1103, arXiv:1011.1759.

[153] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **07** (2003) 001, arXiv:0206293 [hep-ph].

[154] J. Alwall, S. Hoche, F. Krauss, N. Lavesson, L. Lonnblad, et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, Eur.Phys.J. **C53** (2008) 473–500, arXiv:0706.2569 [hep-ph].

[155] G. Corcella et al., *HERWIG 6.5 release note*, arXiv:0210213 [hep-ph].

[156] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Multiparton interactions in photoproduction at HERA*, Z. Phys. C **72** (1996) 637–646, arXiv:9601371 [hep-ph].

[157] T. Binoth, M. Ciccolini, N. Kauer, and M. Kramer, *Gluon-induced W-boson pair production at the LHC*, JHEP **0612** (2006) 046, arXiv:hep-ph/0611170 [hep-ph].

[158] H.-L. Lai et al., *New parton distributions for collider physics*, Phys.Rev. **D82** (2010) 074024, arXiv:1007.2241 [hep-ph].

[159] J. Pumplin, D. Stump, J. Huston, H. Lai, P. M. Nadolsky, et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **0207** (2002) 012, arXiv:hep-ph/0201195 [hep-ph].

[160] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. **A 506** (2003) 250–303.

[161] A. Elagin, P. Murat, A. Pranko, and A. Safonov, *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, Nucl.Instrum.Meth. **A654** (2011) 481–489, arXiv:1012.4686 [hep-ex].

[162] The ATLAS Collaboration, *Search for the Standard Model Higgs boson in $H \to \tau\tau$ decays in proton-proton collisions with the ATLAS detector*,.

[163] The ATLAS Collaboration, *Measurement of the total ZZ production cross section in the four-lepton channel using 5.8 $fb^{?1}$ of ATLAS data at $\sqrt{s}$ = 8 TeV*,.

[164] ATLAS Collaboration, *Z+jets background estimate using a data driven fake factor method for a search for the Higgs boson in the associated mode whwwlnlnln and zhwwlllnln with the ATLAS detector at 8 TeV*,.

[165] ATLAS Collaboration, G. Aad et al., *Improved luminosity determination in pp collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector at the LHC*, Eur.Phys.J. **C73** (2013) 2518, arXiv:1302.4393 [hep-ex].

[166] The ATLAS Collaboration, "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.

[167] J. M. Campbell, R. K. Ellis, and C. Williams, *Vector boson pair production at the LHC*, JHEP **1107** (2011) 018, arXiv:1105.0020 [hep-ph].

[168] The ATLAS Collaboration, *ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11*,.

[169] A. Martin, W. Stirling, R. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur.Phys.J. **C63** (2009) 189–285, arXiv:0901.0002 [hep-ph].

[170] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, et al., *Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology*, Nucl.Phys. **B849** (2011) 296–363, arXiv:1101.1300 [hep-ph].

[171] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur.Phys.J. **C71** (2011) 1554, arXiv:1007.1727 [physics.data-an].

[172] G. Ranucci, *The Profile likelihood ratio and the look elsewhere effect in high energy physics*, Nucl.Instrum.Meth. **A661** (2012) 77–85, arXiv:1201.4604 [physics.data-an].

[173] A. L. Read, *Presentation of search results: The CL(s) technique*, J.Phys. **G28** (2002) 2693–2704.

[174] ROOT Collaboration Collaboration, K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*,.

[175] J. Friedman, *Data analysis techniques for high energy particle physics*,.

[176] I. Asimov, *Franchise*, Isaac Asimov, The Complete Stories **1** (1990).

[177] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*,.

[178] J. H. Friedman, *Stochastic gradient boosting*, Comput.Stat.Data Anal. **35** (2002) 367–378.

[179] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, Journal of Computer and System Sciences **55** no. 1, (1997) 119 – 139. http://www.sciencedirect.com/science/article/pii/S002200009791504X.

[180] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852–867, arXiv:0710.3820 [hep-ph].

[181] ATLAS Collaboration, "Summary of atlas pythia 8 tunes." ATL-PHYS-PUB-2012-003.