VNIVERSITAT DE VALÈNCIA

Search for associated production of a Higgs boson and a single top quark in the multi-lepton final state with ATLAS

Jesús Guerrero Rojas

Directors:

Dra. Susana Cabrera Urbán

Dr. Carlos Escobar Ibáñez

Doctorat en Física

Departament de Física Atòmica, Molecular i Nuclear

Maig, 2023

*La **Dra. Susana Cabrera Urbán**, científica titular del Consejo Superior de Investigaciones Científicas (CSIC) y el **Dr. Carlos Escobar Ibáñez**, Contratado Ramón y Cajal del Consejo Superior de Investigaciones Científicas (CSIC)*

Certifican:

Que la presente memoria titulada, "Search for associated production of a Higgs boson and a single top quark in the multi-lepton final state with ATLAS" ha sido realizada bajo su dirección en el Instituto de Física Corpuscular, centro mixto de la Universitat de València y el Consejo Superior de Investigaciones Científicas, por ***Jesús Guerrero Rojas*** y constituye su Tesis para optar al grado de Doctor por la Universitat de València, una vez cursados los estudios en el Doctorado en Física.

Y para que así conste, en cumplimiento de la legislación vigente, firman el presente certificado.

Valencia, a día 12 de mayo de 2023,

Dra. Susana Cabrera Urbán	***Dr. Carlos Escobar Ibáñez***

Con el visto bueno del tutor, el ***Dr. Juan Zúñiga Román***

CONTENTS

Introduction

Researches about the elementary pieces of the universe are one of the most popular today in Physical Science. They involve questions whose answers are partially given by the Standard Model (SM) of Particle Physics. Some responses have not been tested experimentally yet, and others are outside of this theory. The SM provides a set of particles to explain the fundamental components of the matter and its interactions.

Two particles are going to be highlighted in this thesis: the Higgs boson and the top quark, since the main subject is the first search of the associated production of a Higgs boson with a single top quark in the ATLAS detector for multi-lepton final states. Its interest is based on the specific characteristics of both particles and on being sensitive to a possible symmetry violation. The data used in this analysis were collected by the ATLAS detector at the LHC during the Run 2 (from 2015 to 2018), with a luminosity of 139 fb^{-1} and a centre-of-mass energy $\sqrt{s} = 13\text{TeV}$.

The analysis includes a multivariate analysis approach based on several boosted decision trees (BDT) to enhance one specific process each one. The combination of the results of the BDTs allows the definition of regions of interest in the physical phase space of this analysis. Afterwards, these regions are included in a profile likelihood binning fit to provide the final results.

The current thesis is divided in chapters as follows: chapter 1 shows the motivations and introduce the theory behind the search of *tHq*, chapter 2 introduce the LHC and the ATLAS detector during the Run 2, chapter 3 and chapter 4 describe the different event samples and the physical object used in the analysis respectively. Chapter 5 explains the strategy followed in the search of *tHq* process and the different techniques applied to provide the final results. Finally, chapter 6 shows the conclusions from the search of the *tHq* process.

CHAPTER 1

Theoretical framework and motivations

The knowledge of the most basic components of the matter in the universe and their interactions is still one of the hot topics in Physical Sciences. This topic is partially covered by the Standard Model (SM) of Particle Physics. The most basic idea of the SM is the one related to fundamental symmetries of matter. The main goal of the SM is to describe the fundamental particles that make up the matter and to explain their interactions. The SM arises from the Quantum Field Theory; hence it includes both the quantum mechanics and the special relativity.

From a mathematical point of view, the SM is a non-Abelian gauge theory invariant under the transformation of the symmetry groups $\mathrm{SU}(3)_\mathrm{C} \otimes \mathrm{SU}(2)_\mathrm{L} \otimes \mathrm{U}(1)_\mathrm{Y}$, where each group has a physical interpretation: $\mathrm{SU}(3)_\mathrm{C}$ is related to the strong force, and $\mathrm{SU}(2)_\mathrm{L} \otimes \mathrm{U}(1)_\mathrm{Y}$ is related to the unification between the electromagnetic and the weak forces. The meaning of the subscripts is related to different properties of the particles which are involved in each symmetry group: C means the colour charge of the particles, L refers to the left-handed chirality of the particles and Y means the weak hypercharge. More details are given in the following sections.

This chapter is divided as follows: section 1.1 gives a brief overview of the SM, sections 1.2 and 1.3 briefly summarise the physics and the history of the top quark and the Higgs boson, respectively. These particles play a special role in the SM since the top quark is the most massive particle, and the Higgs boson allows to explain the origin of the different particle masses. Last but not least, section 1.4 covers the production of Higgs bosons in association with a single top quark (named *tHq*) and its importance in Particle Physics, which is the main topic of this thesis.

1.1 The Standard Model of Particle Physics

The SM of Particle Physics merges a list of revolutionary and successful theories developed in the 1960s and 1970s. The SM is a gauge theory[1] that describes the fundamental particles of matter and their interactions. The fundamental particles of the SM, highlighted in figure 1.1, are divided into different groups according to their properties. The most general split is done using the spin of the particles. Therefore, two distinct groups are defined using the spin: fermions and bosons.

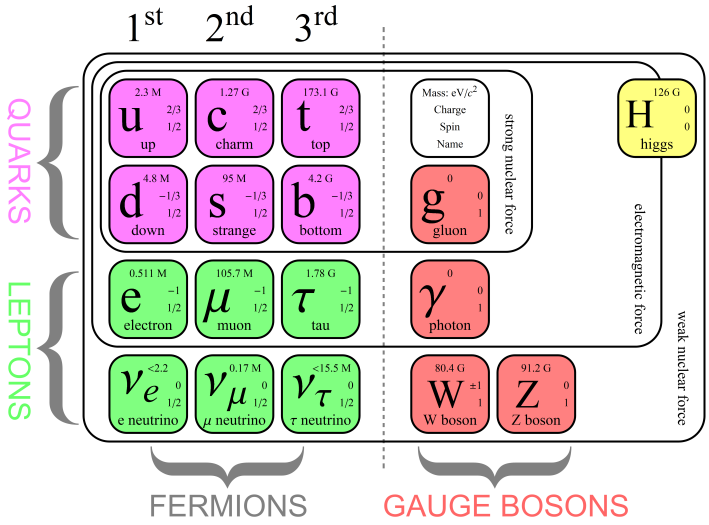


FIGURE 1.1: Fundamental particles of the SM and their characteristics. Particles are split in fermions and bosons. The colours of the squares rely on the spin of the particles. The surrounding squares limit the fundamental force which involve the surrounded particles. The different generations of the particles are also shown [1].

In the first group, the fermions are the particles that have half-integer spin, and they follow the Fermi–Dirac statistics. Inside the SM, there are 12 fundamental fermions with spin 1/2. In addition, these fermions are also divided in two groups depending on whether they have colour charge: quarks and leptons. Quarks are colour-charged

[1] A gauge theory is a type of field theory which is invariant under local transformations of certain Lie groups.

particles and are split in three generations with increasing mass. For each generation a doublet is defined as follows:

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} s \\ c \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}.$$

On the other hand, leptons are colourless-charged particles and are also divided into three generations. Similarly, for each generation of leptons a doublet is defined as:

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix} \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix} \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}.$$

The symbols u, d, s, c, t, b and e, ν_e, ν_μ, ν_τ are shown in figure 1.1.

Despite the existence of three generations of fermions, the usual matter in Nature is only composed by the first generation.

In the second group, the bosons are the particles that have integer spin, and they follow the Bose–Einstein statistics. The fundamental bosons of the SM are four particles. They have either spin 1 (γ, Z, W, and gluon particles) or 0 (Higgs particles). These bosons relied on the interactions between particles at quantum level. More specifically, the exchange of gauge bosons explains the different forces included in the SM.

There are four fundamental forces in Nature: the strong force, the weak force, the electromagnetic force, and the gravity. From them, only the electromagnetic, the weak, and the strong forces are included in the SM. Each of these forces have a direct correspondence with the exchange of one or more gauge bosons at quantum level:

- The strong force appears between particles with different colour charge and its range is limited to the radius-nuclei distance. This interaction is described by the Quantum-Chromodynamic (QCD) theory. The mediator bosons for this interaction are the gluons. A gluon is a massless, electrically-neutral boson that carries colour charge. The strong force explains the stability inside of the atomic nuclei, allowing coexistence of protons and neutrons. This force is the responsible of the colour confinement. In fact, it is the reason why stable particles have neutral colour and quarks, which have colour-charge, are confined inside other particles without colour charge in Nature. These colourless and non-fundamental particles

composed by quarks are known as hadrons. Hadrons are divided in baryons and mesons depending on the number of quarks they are composed of.

- The electromagnetic force occurs between particles with a non-null electric charge and has an infinite range. The mediator particle in this case is the photon (γ). A photon is a massless boson with an electric charge equal to 0.
- The weak force has a range like the radius of the atomic nuclei, and it is the responsible for the radioactive decays. It is mediated by the Z and W bosons. These mediators are massive particles and have electrical charge equal to 0 and $\pm$ 1, respectively.

Nowadays, the gravity is the only force which is not included in the SM. It is not described with a particle mediator, unlike the others fundamental forces, and its effects are negligible at quantum level.

All the particles included in the SM decay into lighter particles only if these decays are allowed by the conservation laws. Therefore, properties of the majority of the particles in the SM are measured through the products of their decays. In the case of the quarks, they join to create hadrons which, if not stable, decay to lighter particles. A special case of quark is the top quark which is the unique quark that directly decays into lighter particles. More details about the top quark are given in section 1.2.

The fermions and bosons have a partner with inverted quantum numbers called anti-particles. In Nature these anti-particles do not exist and when an anti-particle and its partner collide, they are annihilated producing energy and/or other particles, e.g. photons.

From a mathematical point of view, the SM is a non-Abelian gauge theory invariant under the transformations of the groups

$$\mathrm{SU(3)_C} \otimes \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y} \ .$$

where, as mentioned before, C means the colour charge, L the left-handed chirality, and Y the weak hypercharge.

Each of these transformations means a symmetry of the SM. Thus, according to the Noether's theorem when a symmetry exists a physical parameter is preserved [2]. In

other words, each of these symmetries has a physical interpretation. Indeed, the forces included in the SM have a direct relationship with these symmetries.

Firstly, the $\mathrm{SU}(3)_{\mathrm{C}}$ gauge group is on the base of the QCD theory, as mentioned before, which describes the strong force. Consequently, quarks are described inside colour triplets (green, red, blue or RGB) since they are colour particles, and leptons inside colour singlets because of they are colourless. The $\mathrm{SU}(3)_{\mathrm{C}}$ group has eight generators which are represented by the Gell-Mann matrices.

Secondly, the combination of the $\mathrm{SU}(2)_{\mathrm{L}} \otimes \mathrm{U}(1)_{\mathrm{Y}}$ local invariance symmetries is on the base of the electroweak (EW) interaction. Even though the combination of the electromagnetic and weak interactions, i.e. electromagnetic and weak forces are different at macroscopic level, the theory proposed by S.L. Glashow, A. Salam and S. Weinberg [3–5] merges both forces in the EW one at quantum level. The EW theory is a chiral theory, and therefore particles are distributed into left-handed doublets and right-handed singlets. Only particles with left-handed chirality could couple to the weak interaction. The weak hypercharge (Y) is a conserved quantum number, and it is defined by Gell–Mann–Nishijimi formula [6] as:

$$\mathrm{Q} = \mathrm{I}_3 + \frac{\mathrm{Y}}{2} \,,$$

where I_3 is the isospin and Q is the electric charge.

The EW and the QCD interactions do not allow mass terms for fermions and bosons since these kinds of terms do not maintain the local symmetry of the theories. However, the experimental observations indicate that some bosons and fermions have indeed a non-null mass. R. Brout, F. Englert and P. W. Higgs proposed a solution in the 1960s: the EW spontaneous symmetry breaking (SSB) mechanism [7–9].

The SSB introduces a complex scalar field, ϕ, which follows the $\mathrm{SU}(2)$ symmetry, and the Higgs boson in the SM. This mechanism adds a term, $\mathscr{L}_{\mathrm{Higgs}}$, to the SM Lagrangian, $\mathscr{L}_{\mathrm{SM}}$; as follows:

$$\mathscr{L}_{\mathrm{Higgs}} = (\mathrm{D}_\mu \phi)^\dagger (\mathrm{D}^\mu \phi) - \mathrm{V}(\phi) \,,$$

where D_μ is the covariant derivative and $V(\phi)$ is the Higgs-boson potential. This potential is defined as:

$$V(\phi) = \lambda(\phi^\dagger\phi)^2 + \mu^2\phi^\dagger\phi,$$

where λ and μ are free parameters and characterise the potential. If $\mu^2 > 0$, the SSB is not allowed, and the minimum of the potential is equal to zero. However, if $\mu^2 < 0$, the minimum is placed in a circle of radius $v = \sqrt{-\mu^2/\lambda}$, where v is known as vacuum expectation value. The SSB is produced when μ^2 goes from positive value to negative and then the minimum changes from zero to the vacuum expectation value. This potential introduces a new degree of freedom related to the direction of the SSB in the circle with the minimum value. An expansion of the scalar field around the chosen vacuum allows to recover the masses of the SM particles without breaking gauge invariance. This expansion produces an additional scalar field which is identified as the Higgs boson.

The fermions acquire mass through the coupling between these fermions and the Higgs-boson field. These additional terms related to the mass of the fermions are known as Yukawa interactions ($\mathscr{L}_{\text{Yukawa}}$). The Yukawa interactions give rise to the masses of the fermions except for the neutrinos which do not couple to the Higgs field.

To summarise, the SM includes more than 21 free parameters related to the masses of the fundamental particle, the coupling between the particles, etc. The complete SM Lagrangian can be factorised as following:

$$\mathscr{L}_{\text{SM}} = \mathscr{L}_{\text{EW}} + \mathscr{L}_{\text{QCD}} + \mathscr{L}_{\text{Higgs}} + \mathscr{L}_{\text{Yukawa}},$$

where $\mathscr{L}_{\text{EW}}$ and $\mathscr{L}_{\text{QCD}}$ correspond to the EW and strong interaction, respectively.

The SM is extremely successful in describing a wide variety of events in Nature. However, there are open questions that the SM can not yet answer, which seem to indicate that the SM is a part of a more general and complete theory. Some of these open questions are:

- **Neutrino masses**. The neutrinos are massless particles in the SM [10]. However, the flavour oscillation observed in Nature implies that they are massive particle. They can not acquire mass via the methods included in the SM since only left-handed neutrinos are observed in the Nature and right-handed neutrinos should

also exist for explaining the mass within the SM.

- **Matter–antimatter asymmetry**. There is no hint in the SM about why the matter and antimatter should not be created in an equal quantity. However, there is apparently only matter in our known Universe nowadays. The SM does not provide a mechanism to explain the observed asymmetry [11].

- **Dark matter**. The SM only predicts less than 5% of the universe composition according to many cosmological observations [12, 13]. The second minor component of the universe, around 27%, is known as dark matter which up to now has been only detected through its gravitational effects [12, 13]. However, there is no evidence of particle candidates compatible with the SM for explaining dark matter according to the current observations. One of the most promising candidate particles are the weakly interacting massive particles (WIMPs) [14], which would interact weakly with the particles of the SM and have a mass much higher than the ones in the SM (~ 10 GeV $-$ 1 TeV). Nevertheless, there are other possible dark-matter candidates, e.g. axions [15], but none of them have been observed yet.

- **Dark energy** is the major component of the universe, roughly 68% [12, 13]. The accelerated expansion of the universe demonstrates the existence of this component [16], which acts in opposition to the gravity. Currently, there is no clue about the origin of this component of the universe, and of course, it is not described by the SM.

- **Inclusion of gravitational force**. Currently the SM does not include gravity, since there is not a quantum gravity version which can describe the observed events.

- **Hierarchy problem** arises from the huge gap between the EW scale ($\approx 10^2$ GeV) and the Planck scale ($\approx 10^{19}$ GeV) [17]. The fact that the Higgs-boson mass is well defined under the Planck scale could mean that new physics is needed. If the SM still works at the Planck scale, an extremely fine tuning of the parameter related to the Higgs-boson mass would be needed, what is unnatural, to avoid

divergencies. Another possible solution could be the existence of new scalar particles at the TeV scale to cancel the divergencies. The most popular theory which includes these scalar particles is Supersymmetry (SUSY) [18], which it would solve the hierarchy problem.

- **Force unification**. In the SM, only the electromagnetic and weak interactions are merged in the EW interaction. This fact inspired some theories to try to also merge the strong force with these two interactions. These theories are called Grand Unification Theories (GUTs) [19] and they embed the $\mathrm{SU(3)_C \otimes SU(2)_L \otimes U(1)_Y}$ in a larger symmetry group which is broken in a very high energy scale known as GUT scale ($\sim 10^{16}$ GeV). The GUT could also solve some of the points discussed above but their effects have been not yet observed.

In summary, despite its incredibly success, the SM is indeed incomplete as it does not solve any of the points discussed above.

Fortunately, there are a plenty of theories that may solve some of the points discussed above such as SUSY, GUTs, WIMPs, axions, etc. but none of them have been yet observed.

1.2 Top-quark physics

The top quark (t), together with the bottom quark, make up the third family of quarks of the SM. It was predicted by M. Kobayashi and T. Maskawa in 1973 to explain the charge-parity (CP) violation in kaon decays [20]. The first observation of the top quark was done by the D0 [21] and CDF collaborations [22] using proton-antiproton collisions at the Tevatron collider in 1995.

Top quarks are mainly produced in particle-antiparticle pairs ($\mathrm{t\bar{t}}$) at hadron colliders. The $\mathrm{t\bar{t}}$ production cross-section is dominated by gluon–gluon fusion, for proton–proton (pp) collisions. Alternatively, single top-quark production is also allowed but with a smaller cross-section. The single top-quark production occurs via EW interactions in three different modes at hadrons colliders. They are produced through the exchange of a virtual W boson in either the t- or s-channel, and also via the associated production of a top quark and a W boson (named tW) at leading order (LO) in QCD. The t-channel

processes are the dominant at the Large Hadron Collider (LHC) [23]. In the t-channel process, a light-flavour quark q from one of the colliding protons interacts with a b-quark, which can be considered as being emitted directly from the other colliding proton (five-flavour scheme (5FS)) or as originating from a gluon splitting (four-flavour scheme (4FS)). The incoming light-flavour quark exchanges a space-like virtual W boson, producing a top quark t and a recoiling light-flavour quark q', called the spectator quark.

The representative Feynman diagrams at LO for each channel are included in figure 1.2. Figure 1.3 shows the t-channel and tW productions have been measured in ATLAS and CMS experiment, while the s-channel has not been observed yet [24].

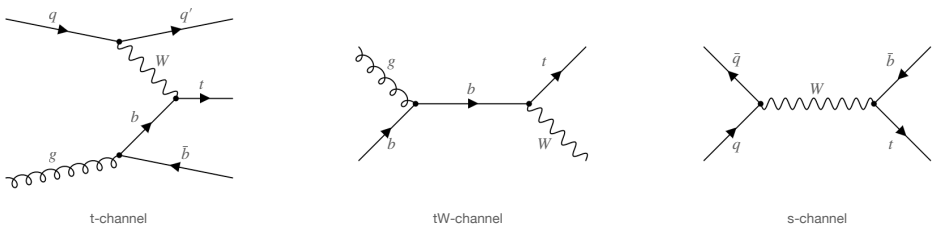


FIGURE 1.2: Representative Feynman diagrams for the main three single top-quark production channels at LO

The top quark almost exclusively decays into a W boson and a b quark (more than 99.83% of the time [25]). Therefore, the final-state decays of the W boson define the possible final states to study the top quark. The W boson decays hadronically (i.e. into a pair of quarks, $\mathrm{W} \to \mathrm{qq}$) and leptonically (i.e. into a lepton and a neutrino, $\mathrm{W} \to \mathrm{l\nu}$), with a branching ratio of 67.4% and 32.6%, respectively [25]. The branching ratios in the $\mathrm{W} \to \mathrm{l\nu}$ decay for each lepton are roughly equal.

The importance of this particle relies on its mass, which is the largest among the fundamental particles in the SM. This fact allows the top quark to decay which gives access to explore the top-quark properties through its decays. Multiple studies are done or are currently in progress to describe the properties of the top quark, e.g. *tWb vertex* studies [27], involving decays of the top quark, production of a single top quark in association with a Z boson (*tZq*) [28, 29] for single top-quark processes, or production of $\mathrm{t\bar{t}}$ in association with a W boson (*ttW*) [30] or a Z boson (*ttZ*) [31, 32] for pairs of top-quark

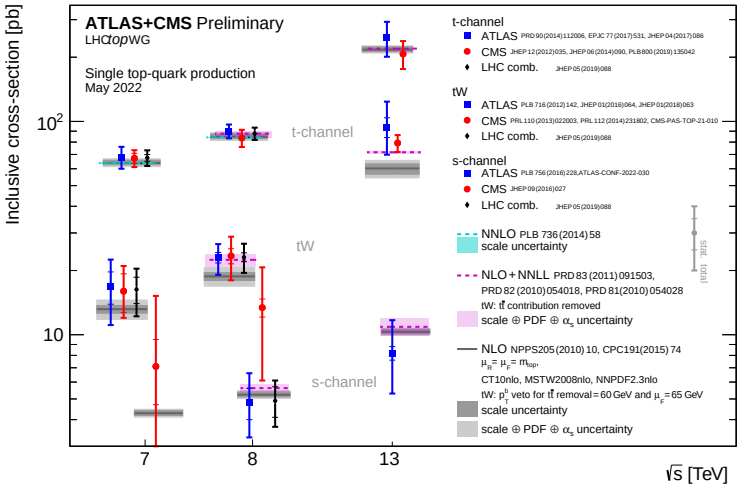


FIGURE 1.3: Summary of measurements for different channels performed by the ATLAS and CMS collaborations. The theoretical calculations are also included for each channel [26].

processes.

The t-channel production and the $W \to l\nu$ decay of the W boson from the top-quark are part of the main topic of this thesis. They are studied in the associated production with the Higgs boson. The motivations and the physics related to this study is summarised in section 1.4.

1.3 Higgs-boson physics

The Higgs boson is a key particle of the SM since it is the necessary particle to explain the SSB. However, though predicted in the 1960s, the Higgs boson was not observed until 2012 by the ATLAS and CMS collaborations in pp collisions [33, 34]. The main characteristics of the Higgs bosons are spin 0, null charge and a mass of 125.09 ± 0.21 GeV [35].

The Higgs boson is mainly produced through four modes, whose representative Feynman diagrams are shown in figure 1.4. In descending frequency order, its production channels at a centre-of-mass- energy of $\sqrt{s} = 13$ TeV are:

- The gluon–gluon fusion (ggF) roughly represents the 87% [25] of the Higgs-boson production. For this mode, two gluons interact through of a triangular loop of quarks to produce a Higgs boson.
- The vector-boson fusion (VBF) covers the 6.8% [25] of the production. In this case, either $W^{\pm}$ or Z bosons fuse to produce a Higgs boson.
- The association of a vector boson with a Higgs boson (VH), also called Higgs-strahlung, is responsible of 4% [25] of the Higgs-boson production. In this production, the process starts by the collision of a quark-antiquark pair which creates an off-shell $W^{\pm}$ or Z boson which irradiates a Higgs boson (WH and ZH, respectively).

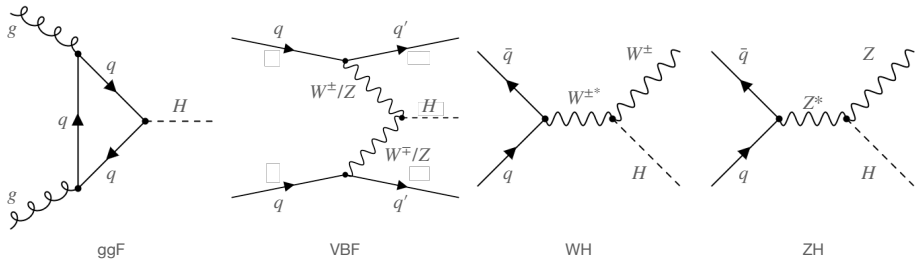


FIGURE 1.4: Feynman diagrams for the main Higgs boson production channels at LO.

In addition to the channels above, the Higgs boson is also produced in association with top or bottom quark. This channel is only responsible of 2.2% [25] of the total Higgs-boson production. The cross-section for each production channel as a function of the centre-of-mass energy is shown in figure 1.5.

The cross-section values normalised to the predicted SM values for the production channel measured by the ATLAS collaboration are shown in figure 1.6. In the case of the production associated to a top or bottom quark, only the case with top quark is measured. The value of the cross-section in the case of the production associated to a top quark was obtained from a combination of a Higgs boson produced in association with a pair of top quarks ($t\bar{t}H$) or in association with a single top quark (*tHq*). Even though the *tHq* production represents a very small fraction of the total production, it is of a special

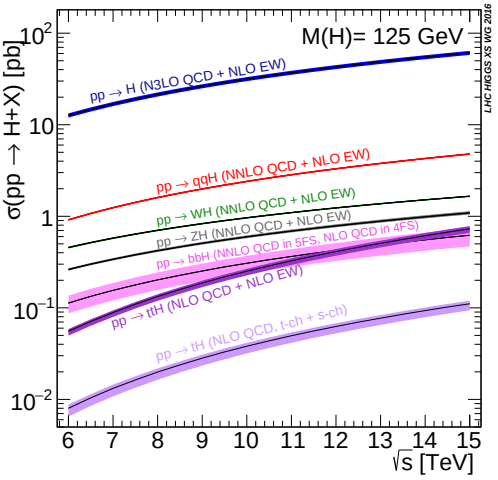


FIGURE 1.5: Theoretical values for the cross-section as a function of the centre-of-mass energy for each process [36].

interest. Therefore, the physics and motivation related to the *tHq* production is covered in section 1.4. This process is indeed the main topic of this thesis.

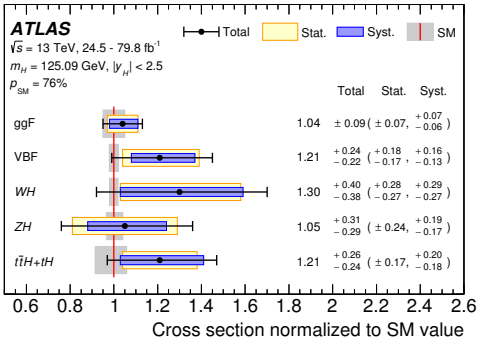


FIGURE 1.6: Summary of the measurements of the cross-section for different processes by the ATLAS collaboration [37].

The study of the different properties of the Higgs boson is done through its decays. The different decays modes as a function of the Higgs-boson mass are shown in figure 1.7. The main decay channel of the Higgs boson with a mass 125 GeV is $b\bar{b}$ [38].

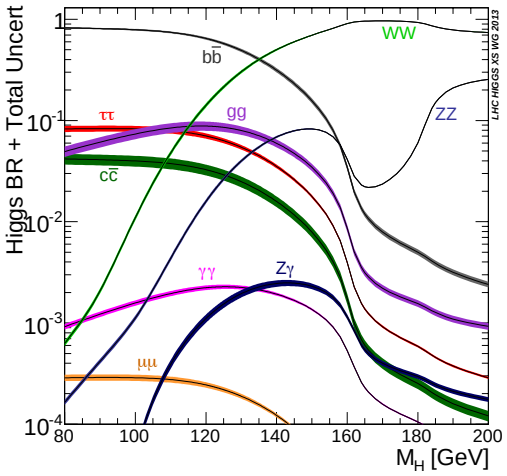


FIGURE 1.7: Theoretical branching ratio for different decays considering their corresponding total uncertainties for the Higgs boson as a function of its mass [39].

1.4 Top-quark–Higgs-boson associated production

After the discovery of the Higgs boson in 2012 [33, 34], the study of the different couplings of the Higgs boson and other elementary particles became an essential test of the SM, either to confirm the nature of the Higgs boson itself or its interaction with other particles. Specifically, the production of Higgs boson in association with a top quark has an especial interest to probe the SM.

The Higgs-boson production in association with a single top quark from pp collisions dominantly occurs via two Feynman diagrams at LO, shown in figure 1.8. The *tHq* production could be considered from two different views: a Higgs-boson production in association with a top quark from the Higgs-boson physics, or a single top-quark production via t-channel with a Higgs-boson radiation from the top-quark physics.

The Yukawa coupling of the Higgs boson to the top quark, $\mathrm{y_t}$, can be determined indirectly from ggF production through a top-quark loop. However, $\mathrm{y_t}$ can be also directly determined from cross-section measurements, either from a pair or a single top-quark production both associated with a Higgs boson.

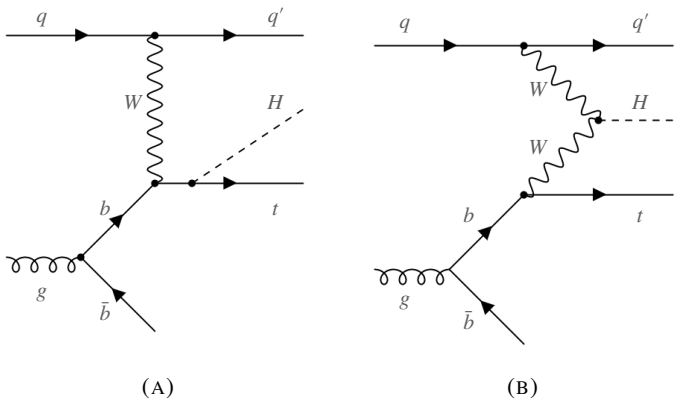


FIGURE 1.8: Two main representative Feynman diagrams for *tHq* production at LO. The Higgs boson is radiated from a top quark (A) and VBF production with a top quark (B).

Nowadays, the $\mathrm{t\bar{t}H}$ is already observed, giving a first measurement of $\mathrm{y_t}$ [40]. Nevertheless, this measurement is only sensitive to the magnitude of $\mathrm{y_t}$ and not to its sign. There are indirect measurements such as Higgs-boson decays to photon pairs [41–44] or some combinations (like combination of *ggH*, *VH* and *VBF* in different decay channels) [45, 46] done by the ATLAS and CMS collaborations, that are also sensitive to the sign of the $\mathrm{y_t}$. These studies disfavour the negative values of the $\mathrm{y_t}$ although they only consider SM particles.

In the case of *tHq*, it is sensitive to the magnitude and sign of $\mathrm{y_t}$, which is of the order of 1 considering the high mass of the top quark. Consequently, the study of the *tHq* process allows to go beyond the SM and explore theories that do not conserve CP. For these theories the cross-section could be up to 10 times higher in the case of complete CP violation ($\mathrm{y_t} = -\mathrm{y_{t,SM}}$), as it is shown in figure 1.9 [47]. In fact, the *tHq* process is especially sensitive to deviations of the SM values due to the interference between the process where the Higgs boson comes from the W boson and the process where comes from the top quark. The interference between both diagrams is destructive in the SM, what causes the *tHq* production cross-section to be very small. In the SM the production cross-section at next to leading order (NLO) in 5FS is 74.25 fb, at $\sqrt{\mathrm{s}} = 13\ \mathrm{TeV}$ considering a mass of Higgs boson equal to 125.0 GeV [48] . Nevertheless, if the interference would not be destructive the cross-section could increase by one order

of magnitude due to the presence of new physics (because of CP violation).

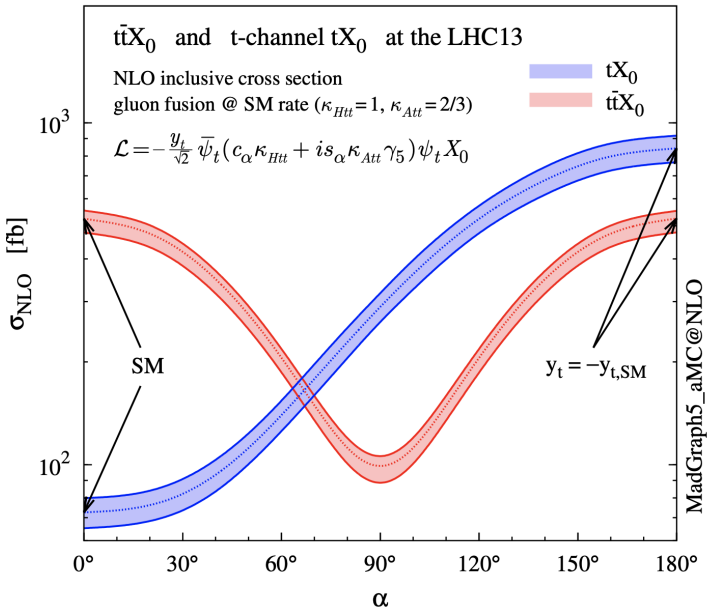


FIGURE 1.9: Cross-section at NLO for $t\bar{t}X_0$ and t-channel tX_0 production from pp collision at $\sqrt{s} = 13$ TeV as a function of the CP-mixing angle α [47].

There are some previous measurements of the *tHq* process by the ATLAS and CMS collaborations which are summarised in table 1.1. They only provide upper limits at 95% confident level (CL), and they are obtained in either combination analyses or with a lower considered luminosity in the case of the *tHq* early analysis by the CMS collaboration. Related to the Yukawa coupling, there are results which give limits to both the SM and the complete inverse hypothesis ($y_t = -y_{t,SM}$) by the CMS collaboration at 95% CL [44]:

$$-0.9 < y_t < -0.7 \text{ or } 0.7 < y_t < 1.1 \, .$$

The study of the *tHq* process using the data collected by the ATLAS detector [51] is the main topic of this thesis. The following chapters describe in a detailed way: the analysis strategy, the signal and background definition, the special background studies,

Analysis	Luminosity (fb^{-1})	Experiment	$\sigma(\mathrm{tHq})/\sigma(\mathrm{tHq})_{\mathrm{SM}}$
tHq (2018) [49]	36	CMS	<14
$\mathrm{t\bar{t}H}$/ *tHq* multilepton (2019) [44]	137	CMS	<5.7
$\mathrm{H} \rightarrow \gamma\gamma$ (2020) [50]	139	ATLAS	<8

TABLE 1.1: Current results for the *tHq* process by the ATLAS and CMS collaborations. The $\sigma(\mathrm{tHq})$ is the cross-section measured in the experiment and the $\sigma(\mathrm{tHq})_{\mathrm{SM}}$ is the reference cross-section given by the SM prediction.

and the fit strategies. The main goal of the analysis is the direct search of the *tHq* production. If this goal were not achieved the analysis would set an upper limit to the production cross-section. This analysis represents one of the first studies of the *tHq* process in the ATLAS collaboration.

In this thesis, only the single top-quark t-channel production in association with a Higgs boson is considered due to two reasons. First, the main goal of the analysis is to perform a direct and dedicated measurement of the *tHq* process in the ATLAS collaboration. Second, the t-channel is the production channel with the largest cross-section.

CHAPTER 2

The LHC and the ATLAS detector

The LHC is located at the CERN laboratory, near the city of Geneva, across the border between France and Switzerland. It is one of the largest scientific collaborative projects in the World. The CERN, acronym derived from *Conseil Européen pour la Recherche Nucléaire*, was founded in 1952, being one of the first European collaborative projects after the World War II.

Nowadays, the LHC is the most powerful and largest circular particle accelerator in the World. This fact makes the LHC an extraordinary place to test physics theories and to study the edge of physics. The LHC started up in September 2008 and it was planned to run over the following 20 years. The schedule of the LHC includes either data-taking periods or long shutdown periods, which are used for the upgrades of the accelerator and the experiments. In fact, the LHC has just started its third data-taking period called Run 3, which is planned until the end of 2025. In addition to Run 3, there were two data-taking periods before: the Run 1 from 2009 to 2013, and the Run 2 from 2015 to 2018. The dataset collected with the ATLAS detector during the last completed data-taking period, i.e. Run 2, is the one used in the analysis presented in this thesis.

This chapter includes a description of the LHC, and a list of the main experiments placed at the LHC in section 2.1. Luminosity and pile-up for the LHC are described in section 2.2. Section 2.3 shows a description of the ATLAS detector during the Run 2. It describes the different subsystems of the detector in subsections 2.3.3-2.3.5, and the performance of the detector, focusing on the alignment of the inner tracker detector in subsections 2.4.1 and 2.4.2, respectively.

The description of both the LHC and the ATLAS detector included in this chapter refers to their status during the Run 2. This data-taking period is used in the analysis presented in chapter 5 of this thesis. The concrete performance of the ATLAS detector explained in this chapter, i.e. alignment of the inner tracker detector, was chosen taking

into account the physics objects measured by the detector that are needed to build the final states of the signal process considered in the analysis.

2.1 The Large Hadron Collider

The LHC is the last accelerator of the CERN accelerator complex, which can be seen in figure 2.1 [52]. The tunnel where the LHC is placed is 100 m underground, composed of several superconducting magnets and radiofrequency cavities which accelerate beams of protons up to more than 99 % of the speed of light. The LHC is hosted in a tunnel of 27 km of circumference where two beams go in opposite directions inside different pipes. Each beam consists of several bunches of protons with a centre-of-mass energy up to 13 TeV, which collides in the hearts of four experiments.

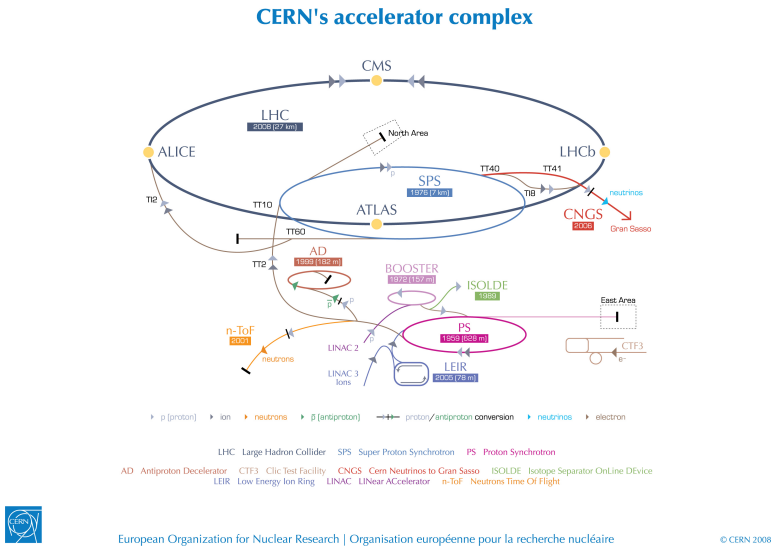


FIGURE 2.1: Schematic representation of the CERN accelerator complex [53].

Protons are extracted from a pressurised tank of hydrogen gas, then are ionised [1] and afterwards are placed at the beginning of the accelerator complex. Firstly, protons are accelerated up to 50 MeV in the Linear Accelerator 2 (LINAC2) [23]. Secondly, the

[1] The hydrogen gas is introduced into the Duoplasmatron [54] where an electrical field breaks down the gas into protons and electrons.

protons are driven to the Proton Synchrotron Booster (PSB) [23], the first circular accelerator, where they reach an energy of 1.4 GeV. Afterwards, the protons are injected into the Proton Synchrotron (PS) [23] and later on into the Super Proton Synchrotron (SPS) [23] which increase the energy of the protons up to 26 GeV and 450 GeV, respectively. These two accelerators are also circular accelerators. Finally, the bunches of protons are injected into the LHC where they reach an energy up to 6.5 TeV and they are ready for collisions in the interaction points (IPs).

In particular, there are four main IPs within the LHC where the following experiments are located:

- TOTEM (*TOTal cross section, Elastic scattering, and diffraction dissociation Measurement at the LHC*) [55]: its goals are precision measurements of the total, elastic, and diffractive pp collisions. The detector uses two tracking telescopes installed on each side of the CMS IP. In particular, the detectors are placed in the CMS forward regions.

- MoEDAL (*Monopole and Exotics Detector at the LHC*) [56]: allows direct searches for magnetic monopole and other highly ionising stable (or pseudo-stable) massive particle at the LHC. It is located at the LHCb IP.

- LHCf (*LHC forward*) [57]: measures the characteristics of particles in the very forward region (nearly zero degrees to the beam). It is installed at 140 m in each side of the ATLAS IP.

- LHCb (*LHC beauty*) [58]: the aim of this detector is the study of flavour physics of the SM, which includes the flavour changing neutral currents, the consistency of the CKM unitary triangle, B-meson decays, mixing and the formation of bound decays. The LHCb is a single-arm forward detector.

- ALICE (*A Large Ion Collider Experiment*) [59]: is a general-purpose heavy ion experiment designed to study the physics of strongly interacting matter and the quark–gluon plasma in nucleus–nucleus collisions at the LHC. For this purpose, heavy nuclei ($^{208}\mathrm{Pb}$) are injected in the LHC to allow the study of hadrons, electrons, muons, and photons produced in its collisions.

- CMS (*Compact Muon Solenoid*) [60]: is a general-purpose detector, designed to study the physics of pp collisions at the LHC. The main aim of CMS is to explore the physics at a very high energy scale to provide precision measurements of parameters of the SM, and to find possible evidences of theories beyond the SM. The CMS detector is embedded in a compact solenoid which provides a magnetic field.

- ATLAS (*A Toroidal LHC ApparatuS*): is another general-purpose detector and shares its goals with CMS. Moreover, the physics measurements in both experiments are expected to be compatible, and in this way, the reproduction of their results is guaranteed. Given that the results shown in this thesis uses the ATLAS detector, a detailed description of the different components of this detector is found in Section 2.3.

2.2 Luminosity and pile-up

In pp collisions a parameter relates the cross-section (σ) and the number of inelastic collisions. This parameter is the instantaneous luminosity (L). It is defined in terms of the average of the number of interactions per bunch (μ), called pile-up, and the frequency of the colliding bunches (f), since the particle beams come in bunches of protons. Thus, the equation for the instantaneous luminosity is:

$$L = \frac{\mu f}{\sigma} .$$

In terms of the beam parameters, the luminosity can be rewritten as:

$$L = \frac{N_1 N_2}{4\pi\sigma_x\sigma_y} f ,$$

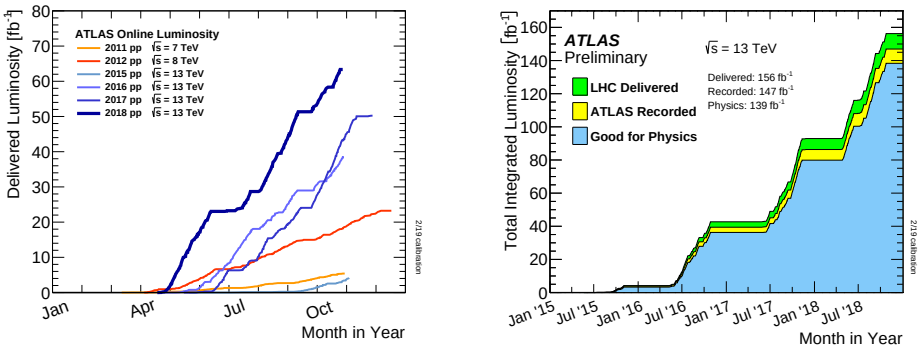
where N_1 and N_2 are the number of particles per bunch in each beam, and σ_x and σ_y are the Gaussian widths in the horizontal and vertical plane per bunch[2]. The total luminosity

[2]Gaussian profile of the transverse particle bunch is assumed.

($\mathscr{L}$) is simply defined as the integral of the instantaneous luminosity over the time:

$$\mathscr{L} = \int \mathrm{L}\mathrm{d}t \ .$$

A particular case is the total luminosity collected by the ATLAS experiment during the Run 2. The total luminosity delivered for physics analyses was 139 fb^{-1} where pp collisions had a $\sqrt{s} = 13$ TeV(see figure 2.2). The analysis presented in this thesis uses the completed dataset of the Run 2.



(A) (B)

FIGURE 2.2: (A) Cumulative luminosity delivered by month in ATLAS during stable beams for high energy pp collisions during Run 1 ($\sqrt{s} = 7$ and 8 TeV) and Run 2 ($\sqrt{s} = 13$ TeV). (B) Cumulative luminosity as a function of the time for delivered, recorded and good for physics, for data periods used in this thesis (i.e. Run 2) [61].

The pile-up quantifies the effects due to the overlap of different events on the detector. It follows the formula:

$$\mu = \frac{N_1 N_2}{4\pi\sigma_x\sigma_y}\sigma \ .$$

These effects appear, for example, when the spacing between bunches is shorter than the response time of the detector or multiple independent interaction occur during one bunch crossing.

The different values of the pile-up during the Run 2 are shown in figure 2.3 in the case of the ATLAS detector. They range from 10 to 70 for the different years. The mean

values of interaction per crossing ($< \mu >$) shown in figure 2.3 corresponds to the mean of the Poisson distribution of the number of interactions per crossing calculated for each bunch.

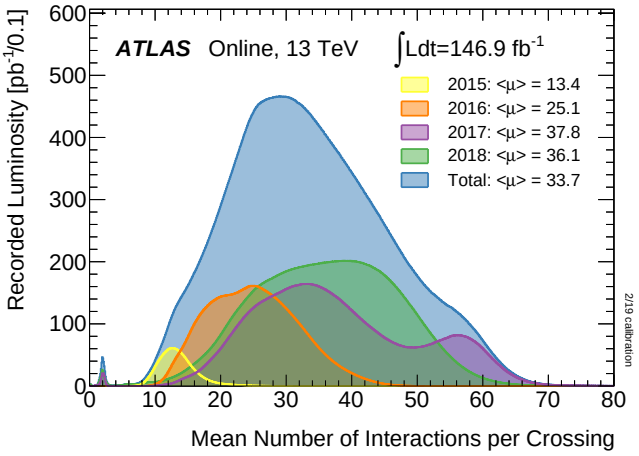


FIGURE 2.3: Luminosity-weighted distribution of pile-up for pp collision data at $\sqrt{s} = 13$ TeV for the entire Run 2 (i.e. 2015–2018). All data recorded by the ATLAS detector during stable beams are shown, and the integrated luminosity and $< \mu >$ are also shown [61].

2.3 The ATLAS detector

The ATLAS detector, shown in figure 2.4, is the largest detector located at the LHC. It is a multipurpose forward/backward-symmetric cylindrical detector with 44 m in length and a diameter of 25 m. The detector is composed of different subdetectors in order to reconstruct and identify all particles emerging from the pp collisions. They also allow measuring the relevant physical properties of particles. Its subdetectors are placed in layers and cover almost the entire solid angle around the IP.

The coordinate system used by the ATLAS detector is a right-handed coordinate system whose origin is at the nominal IP in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the ring of the LHC, and the y-axis points upward. Cylindrical coordinates (r, ϕ) are used in the transverse

plane to the beam, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln \tan(\theta/2)$.

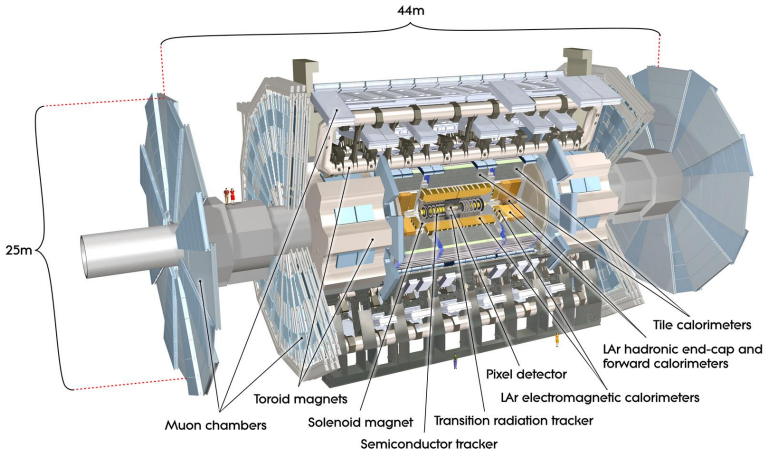


FIGURE 2.4: Computer-generated image of the ATLAS detector, where the dimensions and the subsystems of the detector are shown [62].

From the outside to the inside, the first system is the muon spectrometer, where the muons are detected. The second system is the electromagnetic and hadronic calorimeter used to determine the energy of the interacting particles. Finally, the innermost system is the tracking system, in which charge-particle tracks are reconstructed, and it is embedded in a solenoidal magnetic field. In addition, the whole detector is immersed in a magnetic field that bends the tracks of the charged particles.

2.3.1 Muon spectrometer

The outermost system of the ATLAS detector is the muon spectrometer (MS), see figure 2.5, whose goal is to identify muons and reconstruct their trajectories [51]. The MS covers the complete range of the azimuthal angle and $|\eta| < 2.7$. It is composed of three different layers surrounding the detector and six end-caps placed perpendicular to the beam axis. Moreover, it uses four different technologies to provide a precise tracking and fast triggering. These are:

- **The Monitored Drift Tubes (MDT)** which provide an excellent measurement of the tracks in the principal direction of the bending plane. They are cylindrical layers surrounding the beam pipe in the barrel and circular disks centred at the z-axis of the detector until $|\eta| < 2.7$. The MDTs are made of tubes filled of a mixture of gases, with a diameter of 3 cm and a length varying from 0.9 to 6.2 m.

- **The Cathode Strip Chambers (CSC)** are placed orthogonally to the z-axis of the detector in the innermost layer of the MS in the forward region, $2 < |\eta| < 2.7$, where a higher muon flux is expected. They have a higher granularity than the MDTs. The CSCs are multi-wire proportional chambers filled of a mixture of gases with cathode strip read-out.

- **The Resistive-Plate Chambers (RPC)** are placed in the barrel region at $|\eta| < 1.05$. They are based of gaseous detector made of two parallel resistive Bakelite plates separated by insulating spaces which form a 2 mm gas gap.

- **The Thin Gap Chambers (TGC)** are placed at $2 < |\eta| < 2.7$, in a more forward region. The TGCs are multi-wire proportional chambers filled with a mixture of gases with smaller distance between cathodes and the wire plane compared to the distance between wires.

The main goal of each technology is to record the muon tracking information for the MDT and the CSC systems, and to record trigger and tracking information for the CSC and the TGC systems.

2.3.2 Calorimeters

The goal of the calorimeters (highlighted in figure 2.6) is to identify and measure the energy of the charged and neutral particles [51]. Despite the fact that the calorimeters stop the interacting particles, neutrinos and high-momentum muons escape from the calorimeter material. Nevertheless, the calorimeters can measure the missing transverse momentum associated to the undetected neutrinos as the sum of all the energy deposits in the transverse plane.

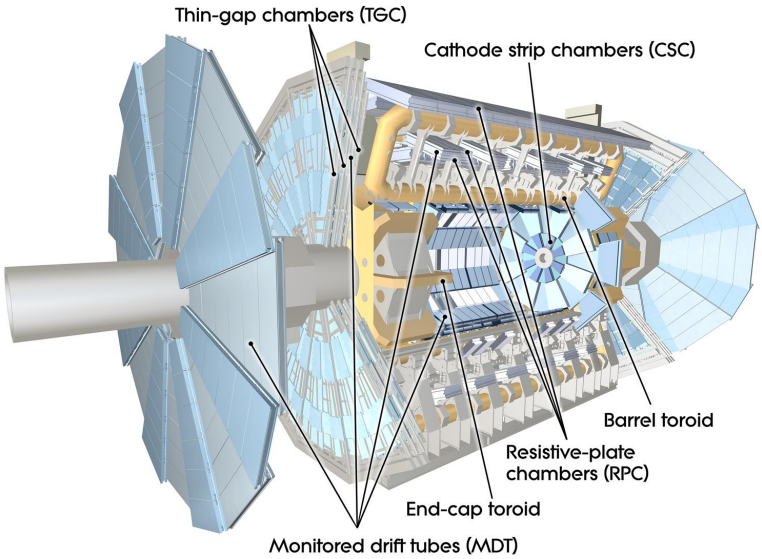


FIGURE 2.5: Computer-generated image of the ATLAS detector, where the MS is highlighted, and its subdetectors are located and labelled [63].

There are two different calorimeters installed in the ATLAS detector: the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). They are both composed of active and passive material, where the passive material causes the showering of the particles and the active material, which is inserted in between the passive material, detects the particles of the shower. A brief description of both calorimeters is given in the following lines:

- **Electromagnetic calorimeter (ECAL)**: it is the innermost calorimeter and is divided in two different parts: barrel ($|\eta| < 1.475$) and two end-caps ($1.375 < |\eta| < 3.2$). The barrel calorimeter is divided in two identical parts separated by a small gap in $z = 0$. In this calorimeter, the passive material is lead (i.e. absorber plates) covering the complete azimuthal range without cracks. Liquid argon (LAr) is used as the active detector medium, chosen for its intrinsic linear behaviour, its stability of response over time and its intrinsic radiation-hardness. The ECAL is housed in a cryostat that ensures the required low temperatures to keep argon in liquid phase.

- **Hadronic calorimeter (HCAL)**: this calorimeter includes two different technologies. In the barrel part ($|\eta| < 1.7$) is the Tile Calorimeter (TileCal), which uses an iron-scintillating technique, and in the end-caps ($1.5 < |\eta| < 4.9$) LAr systems are placed. Apart from its primary goal explained before, this calorimeter tries to avoid that hadronic showers arrive at the muon chambers, for this reason the HCAL needs to be thick enough.

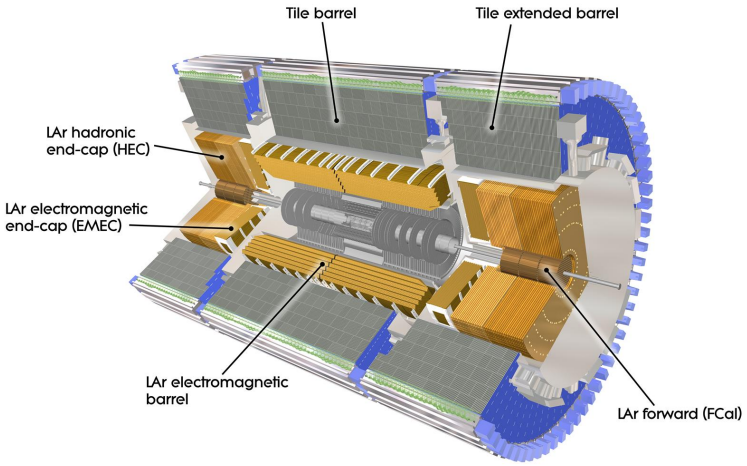


FIGURE 2.6: Schematic view of the ATLAS calorimeter system, where its subdetectors are located and labelled [64].

2.3.3 Inner detector

The innermost system of the ATLAS detector is the Inner Detector (ID), see figure 2.7 [65, 66]. It is embedded in a 2 T superconducting solenoidal-magnetic field to bend the tracks of the charged particles. It has 1.082 m of radius and 6.1 m of length, that means it covers a pseudorapidity range of $|\eta| < 2.5$. The ID provides an excellent reconstruction of charged-particle tracks, as well as primary and secondary vertexes identification. The

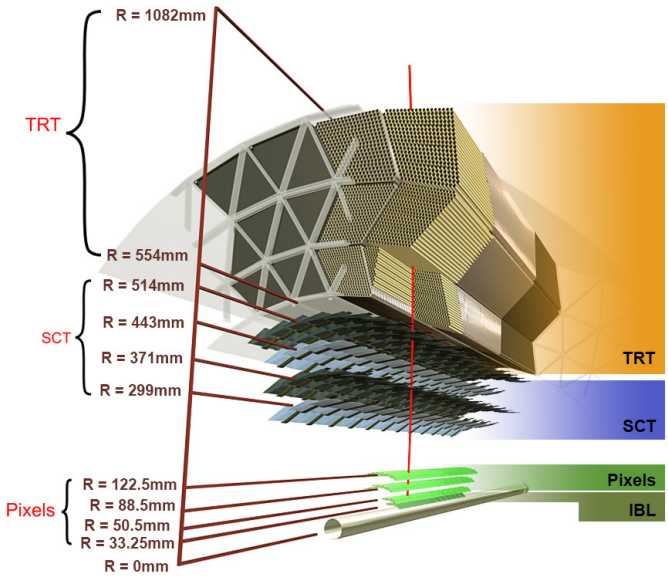


FIGURE 2.7: A 3D visualisation of a section of the barrel of the ID. The different parts and their locations are shown in the picture [67].

great performance of the ID is partially due to a high-performance alignment of all of its detector modules[3]. A detailed description of the alignment is given in section 2.4.1.

The ID is also divided into several subdetectors, and they are briefly described as follows:

- **The Insertable B-Layer (IBL)** is an additional layer added to the Pixel system in the closest place to the IP. It was inserted into the ATLAS detector after Run 1, in 2014, during the long shutdown 1. It is positioned at 33.25 mm in the radial axis. The IBL provides a better resolution of the vertex reconstruction and impact parameter (d_0) information in a higher pile-up environment produced during the Run 2 compared to Run 1. It consists of 280 silicon pixel modules arranged on 14 azimuthal staves. Furthermore, each stave includes a 70 cm long mechanical structure called the bare stave which holds a titanium cooling pipe. There are two

[3] A module is the minimal measuring component of each subdetector.

different kinds of sensors[4] in a stave: 12 planar pixels placed in the central region of the stave and four 3D sensors placed on both extremities of the stave. All in all, a stave mounts 32 pixel in total which are connected to the readout service. The size of the elements of the IBL and its resolution are shown in table 2.1.

- **The Pixel system** is composed of 1774 modules with 4723 silicon pixels on each, which covers the complete azimuthal angle range. The size of each pixel and its intrinsic resolution are shown in table 2.1. The Pixel is distributed in three-barrel layers and six end-cap disks, three in each extremity. The barrel layers are cylindrical layers with different radius placed surrounding the beam axis concentrically, and the end-cap disks have wheel shapes and are installed in the perpendicular plane to the beam. It is designed such that each track creates three hits on average in this system.

- **The SemiConductor Tracker (SCT)** is composed of 4088 silicon strip modules installed in four-barrel layers surrounding the Pixel detector, and nine disks on each of the end-caps. The size of the strips and their resolution are in table 2.1. Each module has two silicon micro-strip sensors glued back-to-back with a stereo angle of 40 mrad to provide a two-dimensional measurement. Each track will create four hits on average at the SCT, thanks to its design.

- **The Transition Radiation Tracker (TRT)** uses a different technology to the previous subsystems. It is roughly composed of 300 k gas-filled tubes (named straw tubes), instead of the silicon detectors. The size and resolution of the straw tubes are in table 2.1. The tubes are placed parallel to the beam in the barrel and radially in the end-caps. The spaces between the straw tubes are filled with polymer fibres in the barrel and foils in the end-caps. This fact allows the TRT to identify electrons since when a particle traverses these spaces a transition radiation, which depends on the particle type and it is much more likely for electrons, is produced and recorded by the detector. The TRT only has sensitivity in the perpendicular plane to the beam and because of its design each track leaves around 30 hits in it.

[4] Sensitive material of each module.

Subdetector	Element size (μm)	Intrinsic resolution (μm)
IBL	50×250	8×40
Pixel	50×400	10×115
SCT	80	17×580
TRT	4000	130

TABLE 2.1: Summary of the main characteristics of the ID subdetectors. The intrinsic resolution of the IBL, the Pixel and SCT is given along $r-\phi$ and z, while for TRT only along $r-\phi$.

2.3.4 Magnetic systems

The ATLAS magnetic systems [51], highlighted in figure 2.8, is composed of two different parts: a central solenoid and a toroidal magnet system. The first is placed in between the ID and the ECAL, and provides a solenoidal-magnetic field of 2 T along the z-axis to the ID. The second is an air-core system composed of a barrel toroid magnet and two end-cap toroids. They provide a magnetic field of 0.5 T in the barrel and of 1 T in the end-caps. The measures of the toroidal magnetic system are 26 m in length and 20 m diameter. Furthermore, the magnetic systems are surrounded by a cooling system which reduces the temperature of the system until 4.5 K. This temperature is necessary due to the fact that the magnetic systems are composed by superconducting magnets.

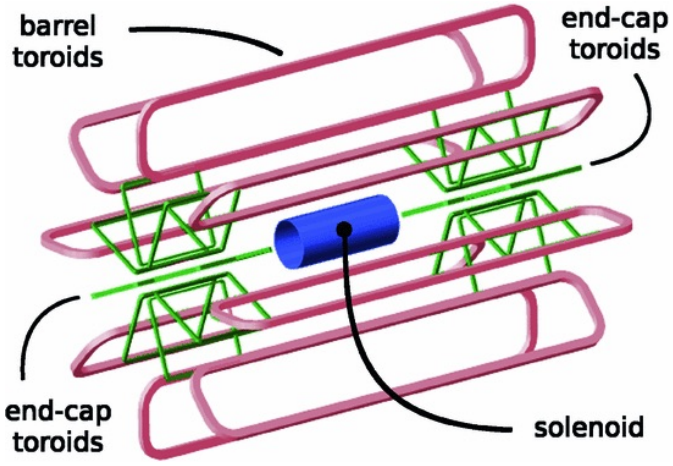


FIGURE 2.8: Diagram of the magnetic systems of the ATLAS detector, where its different magnets are located and labelled [68].

2.3.5 Trigger system

The collision rate at the LHC is 40 MHz as a nominal value. This fact, joint to the pile-up effect, makes unpractical and virtually impossible to store all the information of each event . For this reason, the trigger system [69] was designed to reduce the storage of the bunch crossing rate to a rate at which the data acquisition (DAQ) system can work. In order to do that, the trigger system selects the interesting events for the offline analysis. This is done in two stages.

The Level-1 (L1) trigger is a hardware-based system. The L1 is implemented in custom-built electronics and has access to the partial granularity of the detector. It selects events either using event-level quantities, or multiplicity of objects above thresholds, or topological requirements. In addition, the L1 identifies regions-of-interest (RoIs) in η and ϕ as a region to be investigated for the second trigger stage. Finally, the L1 trigger reduces the collision rate up to approximately 100 kHz within a latency of $2.5\ \mu\mathrm{s}$.

The High-Level Trigger (HLT) is the second stage of the trigger system and is software based. The HLT runs over RoIs or complete granularity of the detector. It includes both fast trigger algorithms and more precise CPU-intensive algorithms, which are similar to the reconstruction ones to select the final events. This process is done on a specific computing farm known as Processing Units, which can evaluate an event within a few hundred milliseconds. The physics output rate, after the HLT, during a run in ATLAS is on average 1.2 kHz with a permanent storage flux of $1.2\ \mathrm{GB/s}$.

2.4 The performance of the ATLAS detector

An essential aspect of any physics analysis is a deep knowledge of the detector itself since this determines its performance and, in the end, the accuracy of the physics results. Therefore, establishing the actual situation and the possible changes of each component of the ATLAS detector is really needed. For this reason, continuous efforts are being made to maintain and upgrade the detector performance in relation to tracking reconstruction, particle identification, jet tagging, etc. In the analysis performed in this thesis, the lepton reconstruction, and the algorithms to identify jets containing b-hadrons play a key role in the final states of the *tHq* process studied which include leptons and jets.

These tasks are extremely dependent on the tracking reconstruction. Thus, a good accuracy of the tracks determines their performance.

As it is mentioned in section 2.3.3, the ID is the main tracking system of the ATLAS detector, and it is extremely precise. However, the high resolution and granularity of the ID are not enough to achieve the maximum accuracy of the physical measurements. Indeed, the knowledge of the actual geometry of the ID determines the accuracy of the track reconstruction and could differ from the nominal geometry. The changes on, for instance, the temperature, powering, or magnetic fields due to the assembly or the operation of the ATLAS detector can affect the knowledge of the nominal geometry. Obviously, the ID is not physically accessible during data-taking periods and direct measurements of the position of the detector are not possible. Therefore, algorithms to estimate and determine the current geometry of the ID are needed. The process to determine the actual geometry of the ID and also its possible changes over time is known as offline alignment.

2.4.1 Alignment of the inner detector

The ID alignment process is performed based on a track-based algorithm, which uses reconstructed tracks of the particles traversing the ID [70]. The alignment uses two different coordinate frames: the global coordinate system (which is indeed the ATLAS coordinate systems), which describes the global position of each detector module, and the local coordinate system, which describes the position of the hits (and cluster) within each sensor in each detector module.

2.4.1.1 Global coordinate system

The global coordinate system $(\mathrm{x},\mathrm{y},\mathrm{z})$ of the ATLAS detector, see figure 2.9, is a right-handed Cartesian system whose origin is the IP within the detector and the axes are defined as follows: the z axis is defined along the beam direction, the x is defined in the radius direction of the ring of the LHC towards its centre, and the y axis is defined in the perpendicular direction to the z and x axes. This is the one already defined in section 2.3.

2.4.1.2 Local coordinate system

The local coordinate system (x', y', z'), see figure 2.9, is also a right-handed Cartesian system and it is defined for each module of the ID. The centre of the coordinate systems is in the middle of each module. The axes are defined as follows: the x' axis points to the most sensitive direction of the module, y' is parallel to the long side of the module and z' is defined with the normal vector of the plane which contains the x' and y' axes. Due to this reason, the axes are defined in different ways depending on the subdetector modules:

- For Pixel and IBL modules, the x' is in the shorter direction of the modules and y' in the longest. The nominal centre is placed in the geometrical centre of each module, and the nominal hit (two-dimensional measurements) refers to the nominal centre.
- For the SCT modules, the axis definition is very similar to the definition in the Pixel and IBL modules except that the SCT modules consist of two micro-strip wafers, one per side, meaning that the x' is in the shorter direction of the modules and y' along the strips. Thus, two local coordinate systems provide two hits per track. In this case, the nominal centre is placed in the geometrical centre along each strip (i.e. along y' axis) per side, and the nominal hit (one-dimensional measurements) refers to the nominal centre.
- For the TRT, the x' is in the direction of radius of the straw tube and y' is along the tube. The nominal hit is placed in the middle point along the tube (i.e. in the geometrical centre along y' axis).

Finally, the hit points are stored by each subsystem in its particular local coordinates to reconstruct the tracks.

2.4.1.3 Formalism of the alignment algorithm

The alignment process uses an approach based on the Newton–Raphson method to determine both the parameters of the trajectories and the alignment parameters. The trajectory of a track in the ATLAS detector is parametrised using the following five track

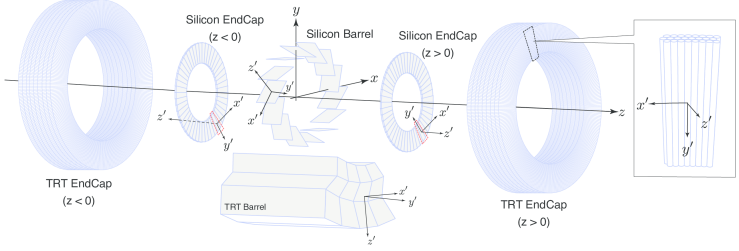


FIGURE 2.9: Schematic representation of the global coordinates (x, y, z) and the local coordinates (x', y', z') for each submodule of the ID.

parameters (π):

$$\pi = (d_0, z_0, \phi_0, \theta_0, q/p)\,,$$

where d_0 and z_0 are the transverse and longitudinal impact parameters, respectively, ϕ_0 is the azimuthal angle and θ_0 the polar angle of the tracks, all defined at the point of closest approach to the $z-$axis of the reference frame. The ratio q/p is the particle charge (q) divided by its momentum (p). The alignment parameters (α) are a set of six

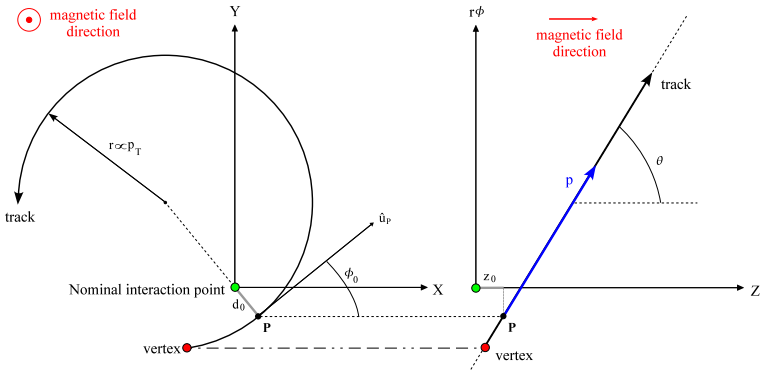


FIGURE 2.10: Schematic representation of the track parameters $(d_0, z_0, \phi_0, \theta_0)$ for a track.

parameters corresponding to the six degrees of freedom of each alignable module of the ID, as follows:

$$\alpha = (T_x, T_y, T_z, R_x, R_y, R_z)\,,$$

where T_x, T_y, T_z are the translations along the three axes (x', y', z') and R_x, R_y, R_z are the three rotations around the axes, both relative to the reference frame of each module.

For each i-th hit, the distance between a nominal hit (m_i) in the plane module, which is part of a track, and an extrapolated hit (e_i) to the module from the fitted tracks, i.e. the reconstructed track, is known as residual (r_i) and is calculated from all the measured hits as:

$$r_i = e_i(\pi, \alpha) - m_i \,,$$

where for each measurement i, m_i is the position of the nominal hit and e_i is the intersection point of the fitted track, described by the parameters π and α, with the surface plane where the hit is measured. Therefore, a track-based χ^2 is calculated using all the residuals, and can be written in vector notation as:

$$\chi^2_t = r^T V^{-1} r \,,$$

where r represents the vector of all the residuals and V is the covariance matrix for all the measured hits. Values of residuals different from zero[5] indicate displacements of the module/plane from its nominal geometry.

2.4.1.4 Global χ^2 alignment algorithm

The alignment process uses a large sample of reconstructed tracks and their hit information in order to implement the track-based Global χ^2 method [71] to determine the alignment parameters using a minimisation with respect to the alignment parameters. First, the χ^2 is written as follows:

$$\chi^2 = \sum_e \sum_{t \in e} \chi^2_t = \sum_e \sum_{t \in e} r^T V^{-1} r \,,$$

where e runs over all the events and t runs over all the tracks of a given event.

[5] Not compatible with the intrinsic resolutions shown in table 2.1.

A minimisation procedure is used to determine the α parameters. The first and second derivatives of the χ^2 with respect to α are considered as follows:

$$\sum_{e}\sum_{t\in e}\frac{d\chi^2_t}{d\alpha}=\sum_{e}\sum_{t\in e}2\left(r^T V^{-1}\frac{dr}{d\alpha}\right)=0\,. \tag{2.1}$$

In order to obtain the alignment parameters, it is considered that around the minimum of the χ^2 the residual can be written using a Taylor expansion in terms of the alignment parameters, in the following way:

$$r=r(\pi_0,\alpha_0)+\left.\frac{dr(\pi_0,\alpha)}{d\alpha}\right|_{\alpha=\alpha_0}\delta\alpha=r(\pi_0,\alpha_0)+\frac{dr(\pi_0,\alpha_0)}{d\alpha_0}\delta\alpha\,, \tag{2.2}$$

where π_0 [6] is a given set of initial reconstructed track parameters, and $\delta\alpha$ is a set of alignment parameter corrections considering an initial set of parameters α_0, which are near the minimum of the χ^2.

The alignment parameters can be written as $\alpha'=\alpha_0+\delta\alpha$. Therefore, if equation 2.2 is evaluated with the condition defined in equation 2.1, the next relation is obtained:

$$\left[\sum_{e}\sum_{t\in e}\left(\frac{dr}{d\alpha_0}\right)^T V^{-1}\left(\frac{dr}{d\alpha_0}\right)\right]\delta\alpha+\sum_{e}\sum_{t\in e}\left(\frac{dr}{d\alpha_0}\right)V^{-1}r(\pi_0,\alpha_0)=0\,. \tag{2.3}$$

Hence, from this equation it is possible to define the alignment matrix and vector as:

$$M_a=\sum_{e}\sum_{t\in e}\left(\frac{dr}{d\alpha_0}\right)^T V^{-1}\left(\frac{dr}{d\alpha_0}\right),$$

$$v_a=\sum_{e}\sum_{t\in e}\left(\frac{dr}{d\alpha_0}\right)V^{-1}r(\pi_0,\alpha_0)\,,$$

Then, the equation 2.3 can be rewritten as follows:

$$M_a\delta\alpha+v_a=0\rightarrow\delta\alpha=-M_a^{-1}v_a\,. \tag{2.4}$$

[6] π_0 is also determined from a Taylor expansion.

The assumption done in the equation 2.2 is not always fully possible. Thus, the alignment process uses an iterative solution. Therefore, the alignment parameters α are iteratively derived until converge is reached as follows:

$$\alpha_N = \alpha_{N-1} + \delta\alpha_N \,,$$

where N means the number of iterations.

The Global χ^2 algorithm involves all the alignable modules and their correlations, what causes the solving of equation 2.4 a hard process. Given the fine granularity and the complexity of the ID, the alignment process can be performed at different levels. They follow the assembly structure of the ID and increase the complexity level sequentially. This is achieved by projecting the residuals computed at module level into larger surfaces, i.e. ID structures. The different parts of the submodules of the ID are sorted in five different levels according to the number of structures from 7 to 351k (see table 2.2).

Level	Description	Number of structures
1	IBL, Pixel, SCT end-caps, TRT barrel and 2 end-caps	7
Si2	IBL layers, Pixel end-cap disks and barrel layers, SCT end-cap disks and barrel layers	32
Si3	IBL modules, Pixel modules and SCT modules	6112
TRT2	TRT barrel modules and end-cap wheels	176
TRT3	TRT straw tubes	351k

TABLE 2.2: Typical alignment configuration split by levels used throughout Run 2 data-taking period.

Additionally, the Global χ^2 algorithm can be extended to add constraints either on the track parameters or on the alignment parameters. The constraints come from external information such as prior knowledge of the geometry of the detector in the case of constraints in the alignment parameter or beam-spot position in the case of constraints in the track parameters. These constraints change the χ^2 definition adding extra terms in

the way:

$$\chi^2 = \sum_i \left[r_i^T(\pi,\alpha) V^{-1} r_i(\pi,\alpha) + R_i^T(\pi) V_i^{-1} R_i(\pi) \right] + R'^T(\alpha) V_\alpha^{-1} R'(\alpha), \tag{2.5}$$

where $R(\pi)_i$ and V_i correspond to the track-parameter constraints and $R'(\alpha)$ and $V'(\alpha)$ correspond to alignment-parameter constraints. In short, these constraints allow the alignment process to avoid large correction values and ensure convergence of the algorithm.

2.4.1.5 Weak modes

Track-based alignment algorithms, like the Global χ^2, are not able to detect some kinds of general geometrical distortions, known as weak modes. In general, a weak mode is such geometrical deformation which leaves the χ^2 formula invariant and can bias the reconstructed track parameters. These weak modes can introduce a bias in the track parameters what could modify the physical measurements. The Global χ^2 algorithm is completely blind to these weak modes even thought their effects can be mitigated by using external constraints (see equation 2.5).

The general geometrical distortions can be related to different movements. They are summed up in figure 2.11 using cylindrical coordinates (R, ϕ, z).

The following subsections are focused on the three main weak modes (sagitta bias, radial distortion, and end-cap expansion) since they directly affect to the reconstruction of the momentum of the particles (which is very important for the analysis presented in this thesis) and how they can be measured through well-known resonances: $Z \to \mu\mu$ and $J/\psi \to \mu\mu$.

A simple way to understand how the weak modes affect tracks is through the transverse momentum (p_T) formula. In the case of a charge particle travelling within a magnetic field B in a cylindrical detector of radius R, the p_T formula in the natural units can be written as:

$$p_T = 0.3 q B \rho = 0.3 q B \left(\frac{R^2}{8s} + \frac{s}{2} \right), \tag{2.6}$$

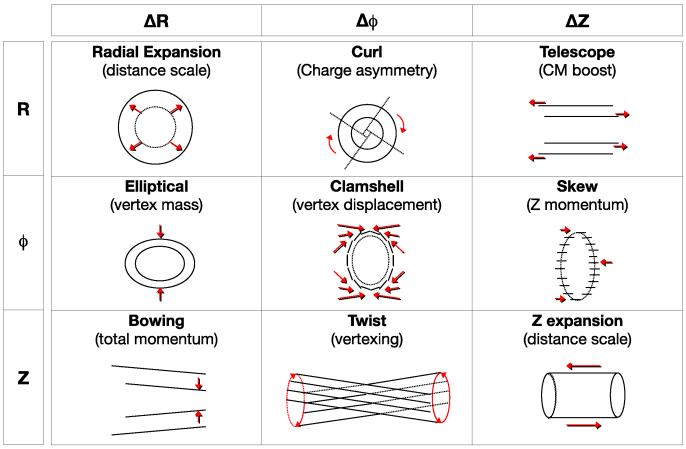


FIGURE 2.11: Representation of all possible distortions along the axes (cylindrical coordinate) and their combinations. These distortions would cause weak modes if they existed.

where q is the charge of the particle, ρ is the radius of the track and s is the sagitta. The sagitta is the distance between the geometrical arc made by the track and the centre of the straight line made from the initial and the final hits included in the track, as shown in figure 2.12.

Cases where $s \ll R$, like those studied, the formula 2.6 can be simplified in the way:

$$p_T = 0.3qB\left(\frac{R^2}{8s}\right) . \tag{2.7}$$

In this way, it is clear how a geometrical deformation in the detector can directly bias the track parameters, in particular, the p_T of the particles.

2.4.1.6 Sagitta bias

A sagitta bias is caused by a geometrical deformation in the bending plane of the tracks and affects in different ways positive- and negative-charged particles. In particular, the momentum changes according to the equation 2.7 in the way:

$$p'_T = p_T(1 + q\, p_T \delta_{sagitta})^{-1},$$

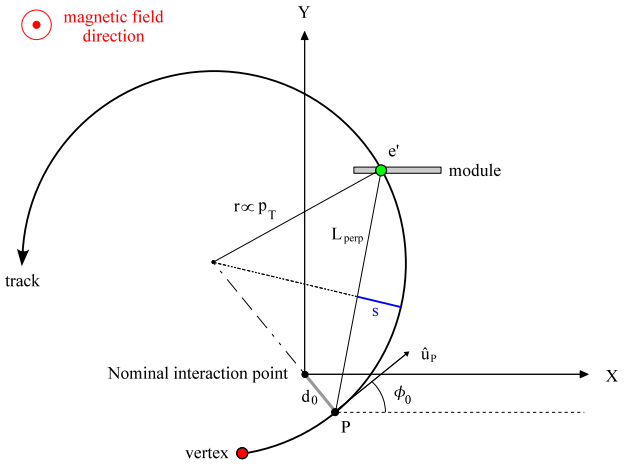


FIGURE 2.12: Schematic representation of the sagitta s for a track. The symbol P is the perigee, which is the path length of the trajectory from the origin to the point of intersection with the module, e'.

where p' corresponds to the reconstructed value, p refers to the true value (which can not be directly measured) and $\delta_{sagitta}$[7] is the value of the sagitta distortion. The change of the momentum allows measuring $\delta_{sagitta}$ through the study of $Z \to \mu\mu$ decays since the decay products have high momentum[8].

The sagitta bias explicitly depends on the region the tracks crossing over. Therefore, the $\delta_{sagitta}$ is a function of η and ϕ (i.e. $\delta_{sagitta}(\eta,\phi)$). Considering the information shown above, the difference between the reconstructed mass of the Z boson (i.e. $m_{\mu\mu}$) and its reference mass (m_Z from reference [25]) as a function of $\delta_{sagitta}(\eta,\phi)$ at first order for each event can be written as:

$$m^2_{\mu\mu} - m^2_Z \approx m^2_Z \left(p'^{+}_T \delta_{sagitta}(\eta^+,\phi^+) - p'^{-}_T \delta_{sagitta}(\eta^-,\phi^-) \right) ,$$

where p'^{+}_T and p'^{-}_T are the reconstructed magnitudes of the transverse momenta for positive and negative electrically charged particles. Even though the value of the bias is

[7] $\delta_{sagitta}$ has units of inverse momentum, e.g. GeV^{-1}.

[8] This fact involves the condition $s \ll R$, and then equation 2.7 is valid.

split for negative and positive particles in the equation above, its value is independent of the charge.

The $\delta_{\text{sagitta}}(\eta, \phi)$ can be calculated using an iterative method where the value for the i-th iteration is:

$$\delta_{\text{sagitta,i}}(\eta, \phi) = -q \frac{m_{\mu\mu}^2 - m_Z^2}{2m_Z^2} \frac{1 + qp'_T \delta_{\text{sagitta,i-1}}(\eta, \phi)}{p'_T} + \delta_{\text{sagitta,i-1}}(\eta, \phi) \, .$$

The bias is determined for each of the two muons of the $Z \to \mu\mu$ decays. The iterations are repeated until convergence is reached.

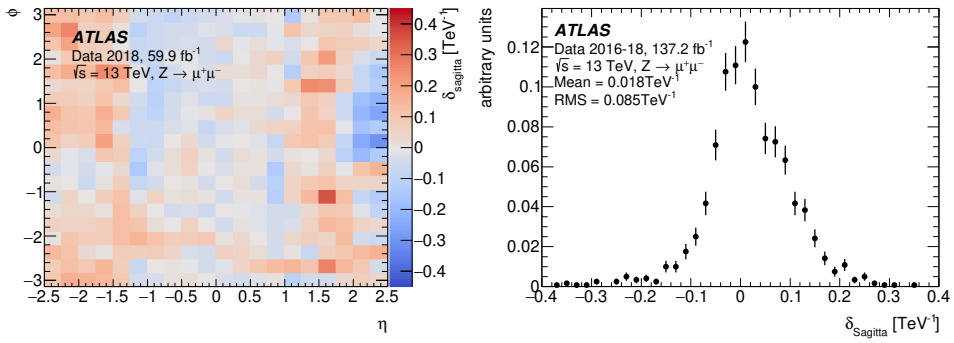
This method is only sensitive to the relative sagitta bias in different sector of the detector. However, alternative methods were also tested, for instance a method which used the E/p ratio. The E/p method is based in the assumption that the calorimeter response is independent of the charge of the particles, and it is perfectly aligned. This second method is also sensitive to global sagitta biases since global displacements of the ID which could affect in the same way to both muons are blinded to the first method.

Measurements of the sagitta bias using the method of the study of the mass of $Z \to \mu\mu$ decays are shown in figure 2.13 and figure 2.14. From this study, it can be concluded that: the average value of the sagitta bias during the Run 2 data period is small (0.018 ± 0.085 TeV^{-1}) and the barrel area of the ID was almost free of bias and some areas out of the barrel ($-2.5 < \eta < 2.5$) showed small effects of it. The shapes of the points of figure 2.14 for the different data-taking periods are compatible, therefore the geometry of the detector was stable during the Run 2.

2.4.1.7 Radial distortion

The radial distortion is caused by a shift along the radial axis of the ID. It could be an expansion or a contraction of the detector and changes the radius of the detector as $\tilde{R} = R_0 + \delta R$, where R_0 is the nominal value of the radius and δR is the displacement which causes the distortion.

This bias is a charge-symmetric alteration, and the momentum changes depending on which approach is used. Two different approaches were studied: the *tower expansion* and the *layer inflation*. For the first case, highlighted in figure 2.15, the p_T, the longitudinal



(A) (B)

FIGURE 2.13: Sagitta biases as a function of η and ϕ for 2018 data period (A), and average of the value of the sagitta bias for the Run 2 data period (B). The uncertainty bars only represent the statistical uncertainty [70].

momentum (p_Z) and the polar angle (θ) are affected at the same time as follows

$$\begin{aligned} p'_T &= p_T(1+2\varepsilon)\,, \\ \cot\theta' &= \cot\theta(1+\varepsilon)^{-1}\,, \\ p'_z &= p_z(1+\varepsilon)\,, \end{aligned} \tag{2.8}$$

where $\varepsilon = \frac{\delta R}{R_0}$ is the value of the distortion. Moreover, the equations above assume that the radial distortion does not affect the sagitta of the tracks.

In order to determine ε is also possible to study the reconstructed invariant mass of the muons from Z, J/ψ and Υ, in a similar way to the sagitta bias. The muons from the decay of three different particles are used to cover a higher range of the p_T. Therefore, the invariant mass of the two muons considering a radial distortion is:

$$\begin{aligned} m^2_{\mu\mu} = {m'_{\mu\mu}}^2 + 2\varepsilon^+ \left[\overrightarrow{p}'^+ \cdot \overrightarrow{p}'^- + \overrightarrow{p_T}'^+ \cdot \overrightarrow{p_T}'^- + \frac{E'^-}{E'^+}\left((E'^+)^2 + (p'^+_T)^2\right)\right] \\ + 2\varepsilon^- \left[\overrightarrow{p}'^+ \cdot \overrightarrow{p}'^- + \overrightarrow{p_T}'^+ \cdot \overrightarrow{p_T}'^- + \frac{E'^+}{E'^-}\left((E'^-)^2 + (p'^-_T)^2\right)\right]\,, \end{aligned}$$

where E' is the energy of the muons, and the positive and negative sign superscripts

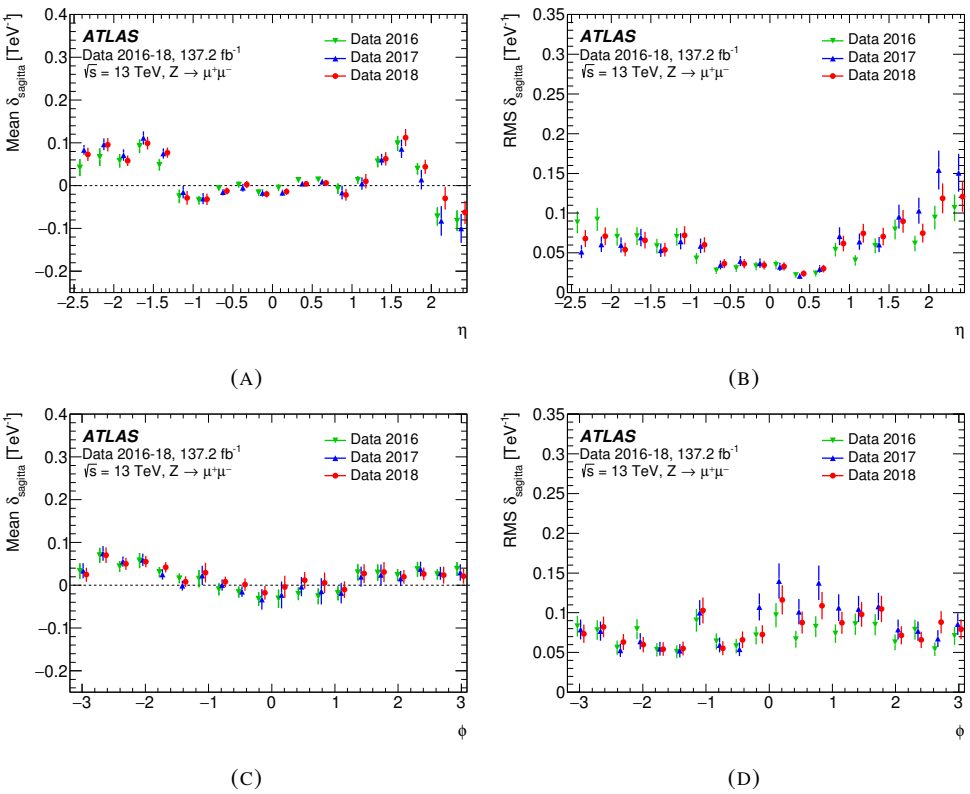


FIGURE 2.14: Projections of the main values of the sagitta bias for η (A,B) and ϕ (C,D) using $Z \to \mu\mu$ decays. The average (A,C) and the RMS (B,D) of the distortion are shown. The uncertainty bars only represent the statistical uncertainty. The markers of the different years are shifted for better visibility [70].

correspond to positive and negative muons. Even though the coefficient for the radial distortion is split for positive and negative muons, it is independent of the charge of the particles. However, the radial bias depends on the region the tracks go through, and thus ε is a function of η and ϕ, i.e. $\varepsilon(\eta, \phi)$.

An iterative process is again done to determine the value of $\varepsilon(\eta, \phi)$, in a similar way to the sagitta bias process. In the end, the difference between the true mass ($m_{\mu\mu}$) and the reconstructed mass of the muons ($m'_{\mu\mu}$) is used to determine the value of $\varepsilon(\eta, \phi)$. The results shown in figure 2.16 are focused on the barrel region of the ID ($|\eta| < 1.07$), where

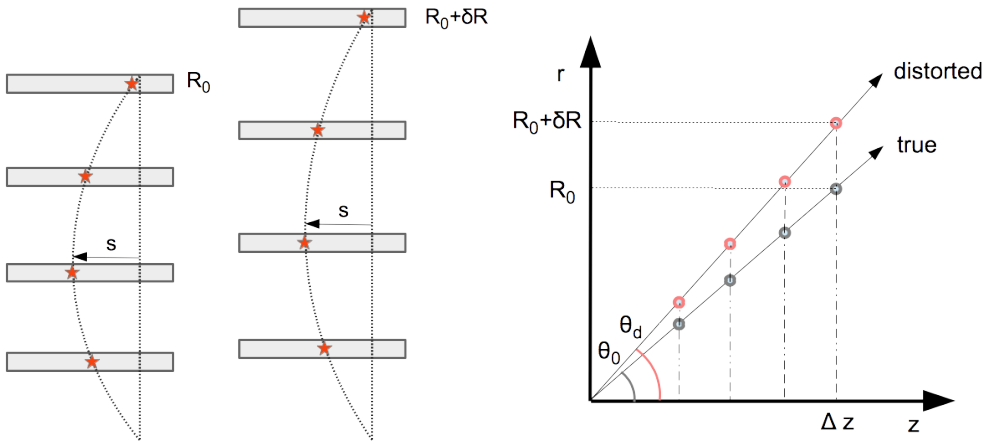


FIGURE 2.15: Schematic representation of the *tower expansion* approach for the radial distortion (A). In this case, the distortion neither modify the value of the sagitta of the tracks nor z, but modifies θ. A simple representation of the layers of the ID where each start represents a hit of the track. In (B), a representation about how this bias affects to the polar angle of the track.

ID is formed by a cylindrical layout and the radial distortion is small. In addition, figure 2.16 shows the radial distortion as a function of the p_T. From this figure, the accuracy of the p_T due to the radial distortion bias is of $\sim 0.1\%$.

In the second case, the *layer inflation* (see figure 2.17), only the p_T and θ are affected but p_z does not. Therefore, they change in the following way:

$$\begin{aligned} p'_T &= p_T(1+\varepsilon)\,, \\ \cot\theta' &= \cot\theta(1+\varepsilon)^{-1}\,, \\ p'_z &= p_z\,, \end{aligned} \tag{2.9}$$

where the main difference to the equation 2.8 is the factor 2 in the p'_T. In this case, the

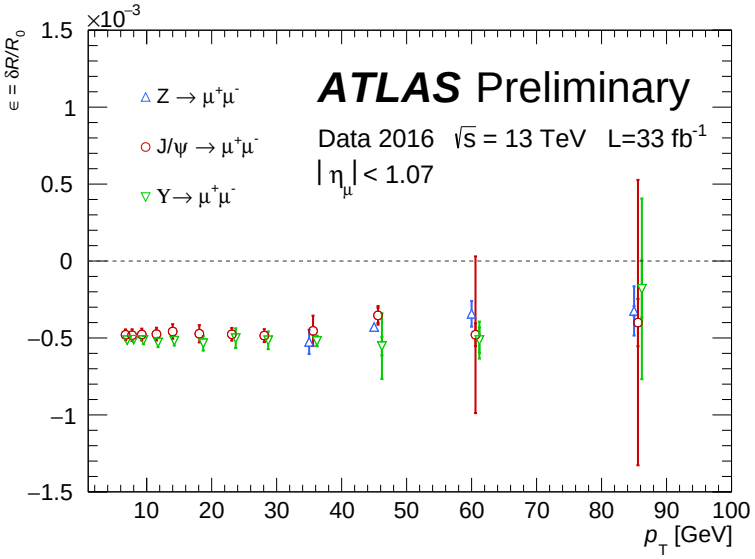


FIGURE 2.16: Radial distortion for the tower expansion approach as a function of the p_T for the 2016 data-taking period for $Z \to \mu\mu$, $J/\psi \to \mu\mu$ and $\Upsilon \to \mu\mu$ decays in the ID barrel zone. The uncertainty bars only represent the statistical uncertainty [67].

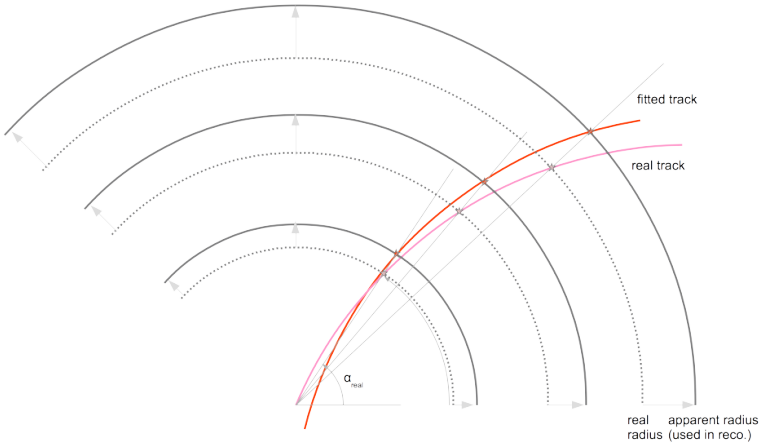


FIGURE 2.17: Simplified representation of the radial distortion for the layer expansion approach. The real and fitted trajectories are shown. Moreover, the real radius (dashed line) and the apparent radius (solid line) are also in the representation.

invariant mass of the pair of muons is affected in the way:

$$m^2_{\mu\mu} = m'^{\,2}_{\mu\mu} + 2\varepsilon^{+}\left[\overrightarrow{p_T}' \cdot \overrightarrow{p_{Td}}' - \frac{E'^{\,-}}{E'^{\,+}}\left(\overrightarrow{p_T}'\right)^2\right] + 2\varepsilon^{-}\left[\overrightarrow{p_T}' \cdot \overrightarrow{p_T}' - \frac{E'^{\,+}}{E'^{\,-}}\left(\overrightarrow{p_T}'\right)^2\right] .$$

Again, an iterative process using the fact that the invariant mass changes according to the value of the bias is done. Now, the radial distortion as a function of the p_T shows a similar behaviour to the one shown in figure 2.16 but the mean value is a half of that value. This reduction is due to the change in equation 2.9 where $p_T \propto \varepsilon$ with respect to equation 2.8 where $p_T \propto 2\varepsilon$.

Other bias can also affect the momentum in a similar way the radial distortion does, e.g. the end-cap expansion which is described in the next subsection 2.4.1.8, and indeed their effects overlap. Therefore, other approaches are studied to try to disentangle the different sources.

2.4.1.8 End-cap expansion

The source of the end-cap expansion is a shift of the end-cap disk along the beam axis (z). In particular, the expansion is produced due to a uniform shift (Δz) along z and it scales with the nominal disk position. Therefore, the z position changes as:

$$z' = z(1+\zeta)\,,$$

where ζ is the parameter of the distortion defined as $\zeta = \frac{\Delta z}{z}$.

This change in the z coordinate only affects p_z and θ. In such a manner, the momentum changes in the following way:

$$\begin{aligned} p'_T &= p_T\,, \\ \cot\theta' &= \cot\theta(1+\zeta)^{-1}\,, \\ p'_z &= p_z(1+\zeta)\,. \end{aligned}$$

If the end-cap expansion exists, it is also possible the study of the invariant mass of muons from Z bosons to determine the ζ value. Thus, in this case, the relationship between the reconstructed and the real invariant mass of the pairs of muons is given by:

$$m_{\mu\mu}^2 = {m'_{\mu\mu}}^2 + 2\left({p'_z}^- - \frac{{E'}^-}{{E'}^+}{p'_z}^+\right){p'_z}^+\zeta^+ + 2\left({p'_z}^+ - \frac{{E'}^+}{{E'}^-}{p'_z}^-\right){p'_z}^-\zeta^- \ . \quad (2.10)$$

For this bias is useful to rewrite the equation 2.10 in terms of η and the local ϕ. If the mass of the particle is neglected, the momentum can be expressed like follows:

$$p = p_T(\cosh\eta, \cos\phi, \sin\phi, \sinh\eta) \ .$$

And the invariant mass can be expressed as:

$$m_{\mu\mu}^2 = {m'_{\mu\mu}}^2 + p_T^+ p_T^- \sinh({\eta'}^+ - {\eta'}^-)(\zeta^+ - \zeta^-) \ . \quad (2.11)$$

From the equation 2.11, it is clear that first, if both tracks of the muons suffer the same distortion this method is not sensitive since $m_{\mu\mu}^2 = {m'_{\mu\mu}}^2$, and second, the higher the difference between η the higher the difference between the masses. Thus, the events selected have one muon whose track passes through the barrel zone and the other one through the end-cap zone.

Again, an iterative process is performed to determine the value of ζ. However, it is observed that the precision of this method is not enough to measure any possible mechanical deformation of the ID. A shift of $1\ \mathrm{mm}$ in the farthest end-cap of the ID is considered like an upper bound of a possible mechanical deformation. This shift means a ζ value around $0.4 \cdot 10^{-3}$. In order to determine the effects of this bias, an ad-hoc value of ζ is introduced to evaluate its impact in the invariant mass of the muons. The sensitivity of the method is limited by the uncertainty on the mass calculation. In figure 2.18 the different between the reconstructed bias mass by the end-cap distortion and the reconstructed mass before the distortion as a function of the value of the distortion ζ is shown.

As mentioned in the previous subsection, this kind of bias can be entangled with

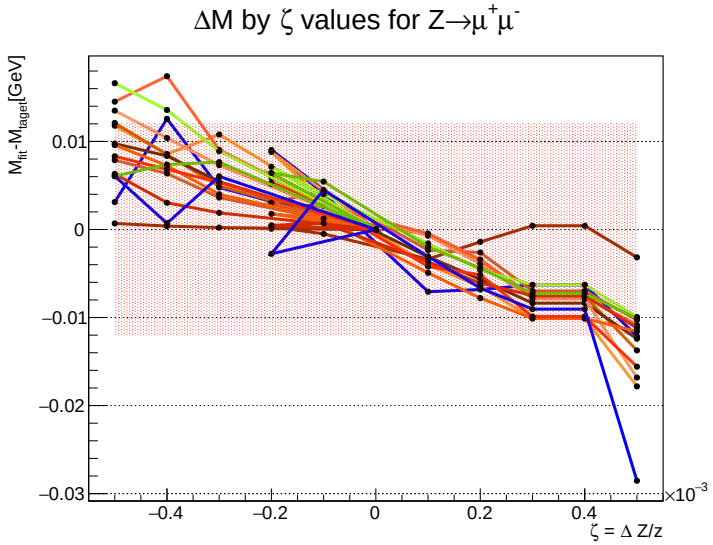


FIGURE 2.18: Different between end-cap bias and non-bias reconstructed masses as a function of ζ. Each line represents a bin for a η value and the dotted box means the area where the method is not sensitive with the current statistics.

other distortions which affect to the momentum in similar ways. Thus, another approach where the biases are considered all together is studied in the next subsection.

2.4.1.9 Length-scale bias

All in all, a displacement of the reconstructed hits can be induced by movements in the radial or in the longitudinal axes of the detector. In the two previous subsections, each method is described independently. However, linear combinations of radial distortion and end-cap expansion are also possible. Moreover, a global-scale bias in the momentum due to a bias in the magnetic field is also entangled to the two biases in the way:

$$p' = p(1+\varepsilon_s)\,,$$

where ε_s is the value of the magnetic bias.

Therefore, if the three biases are in place and assuming all the biases are small and the mass of each muon is negligible, the invariance mass of a particle decaying in two

muons ($m'_{\mu\mu}$) and the true mass ($m_{\mu\mu}$)[9] are related through:

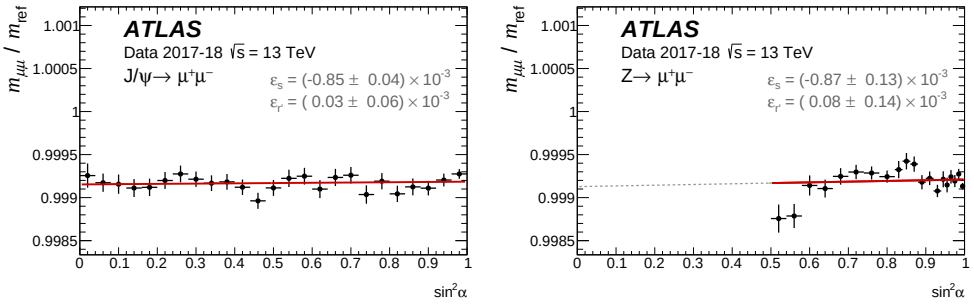
$$m'^{2}_{\mu\mu} \approx m^{2}_{\mu\mu} + 2m^{2}_{\mu\mu}(\varepsilon_s + \varepsilon_{r'} \sin^2(\alpha)), \tag{2.12}$$

where

$$\sin^2(\alpha) = \frac{E^+E^-}{m^2_{\mu\mu}} \left[\frac{p^+_T}{E^+} - \frac{p^-_T}{E^-}\right]^2,$$

and $\varepsilon_{r'}$ is the difference between the radial and longitudinal component of the distortion.

Considering equation 2.12, by measuring the mass as a function of $\sin^2(\alpha)$ is possible to disentangle the radial and the scale bias. Figure 2.19 shows the results from J/ψ and Z-boson decays to a pair of muons. The results show that the value of the radial distortion (ε_r) is negligible since the mass is constant as a function of $\sin^2(\alpha)$. Nonetheless, they show a clear dependence on the momentum-scale bias (ε_s).



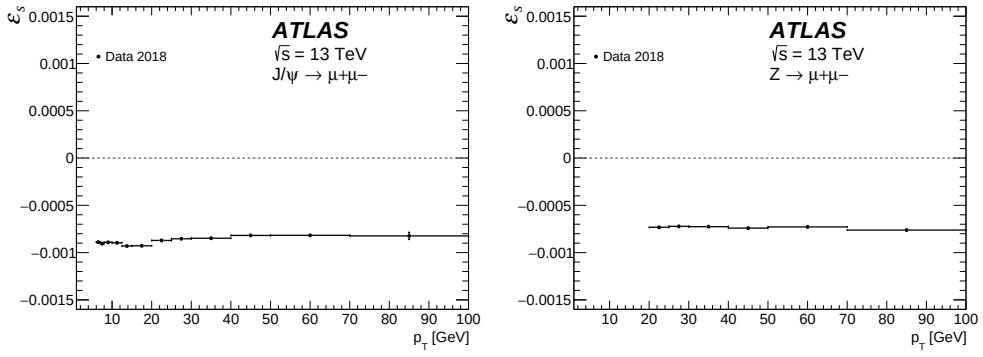
(A) (B)

FIGURE 2.19: Ratio of the measured mass and the reference as a function of $\sin^2\alpha$. The range of data points are different due to the event kinematic between $J/\psi \to \mu\mu$ events (A) and $Z \to \mu\mu$ events (B). The red lines show the fit to data from which the value of ε_s and $\varepsilon_{r'}$ are extracted. The uncertainty bars only represent statistical uncertainty in both figures [70].

The magnitude of the momentum-scale bias for both decays suggests that exists a global-scale bias. The value of the momentum-scale bias can be also measured as a function of the p_T as it is shown in figure 2.20. It is observed the bias is independent of the p_T as expected from a length-scale bias. As highlighted, the source of a global-scale

[9] It is the mass used as reference, usually from the PDG [25].

bias is unclear since it could be either an end-cap expansion or magnetic field bias. This can not be solved for this study and further studies are needed, which are unfortunately out of the scope of this thesis.



(A) (B)

FIGURE 2.20: Momentum scale bias ε_s as a function of the p_T of the tracks, for $J/\psi \to \mu\mu$ events (A) and $Z \to \mu\mu$ (B) for 2018 data. The uncertainty bars only represent statistical uncertainty in both figures [70].

2.4.2 Alignment of the Run 2 dataset

All the techniques described above were used during the alignment of the ID for the Run 2 data-taking period. As already mentioned, the alignment consists of a track-based algorithm that minimises the track hit residuals. As an example, the residuals for the barrel of the Pixel detector are shown in figure 2.21. The alignment process follows a hierarchical level of complexity of the different structures of the ID.

For each data-taking period a set of baseline alignment constants are determined and therefore considered later on in the processing of the data. They are used as initial estimates for the time-dependent alignment refinements, which is done for every new LHC fill. In order to determine the baseline alignment constants a large amount of data are used ($\sim 2\ \mathrm{fb}^{-1}$).

During the Run 2 data-taking period several studies were done aiming to mitigate the time-dependent corrections, which are included in the time-dependent alignment. The

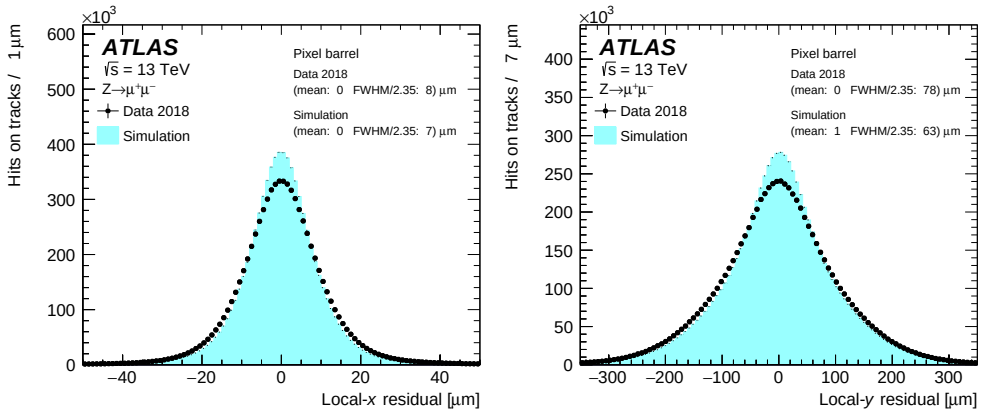


FIGURE 2.21: The Pixel local-x (A) and local-y (B) residual distributions for the $Z \to \mu\mu$ data sample for the 2018 period compared to simulated data. The distributions are integrated over all the hits and tracks [70].

two main time-dependent corrections are the *Temperature dependent IBL distortion* and the *Vertical movements of the Pixel detector*.

The first correction, *Temperature dependent IBL distortion*, was noticed during the commissioning of the IBL. It was soon observed that the IBL staves were displaced a hundred of μm from the nominal geometry in the azimuthal direction. Moreover, these displacements were related with the operating temperature. The size of the distortion was measured using a track-based alignment and fitted with an appropriate model. The value of the IBL distortion for different temperatures, using 2015 and 2016 pp collision data, is shown in figure 2.22. The source of this distortion is the asymmetric mechanical coupling of material with different thermal expansion coefficients.

The second correction, *Vertical movements of the Pixel package*, consists of a vertical displacement in the global-y axis of the Pixel detector by up to 8 μm at the start of a LHC fill. The source of this movements is related to the operation of the Pixel detector. At the beginning of the LHC fill when the Pixel detector is switched on and the temperature of the modules increases almost immediately, them the occupancy of the modules and the instantaneous luminosity decrease over the course of the fill and the temperature

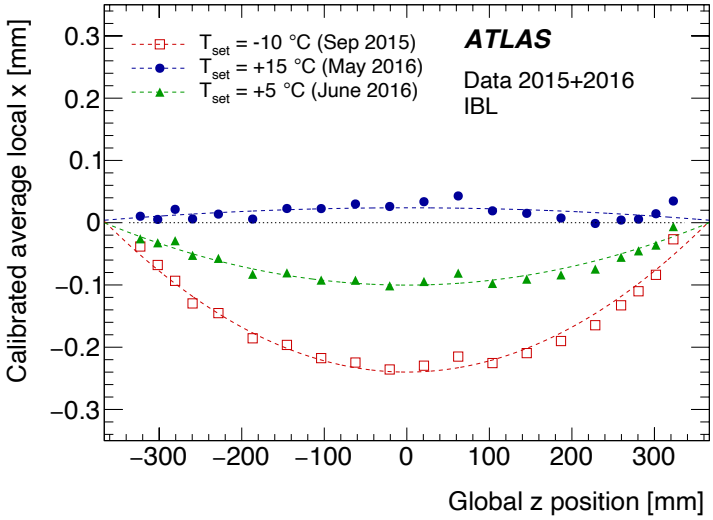


FIGURE 2.22: Average of local-x position in the global transverse plane (global-z) over all the IBL staves. The position only represents the movements of the IBL due to the *Temperature dependent IBL distortion*. Events from 2015 and 2016 are shown for different temperatures without uncertainty bars associated to the data points. The *Temperature dependent IBL distortion* was constant during the all the LHC fills [70].

decreases gradually. That fact origins an additional movement in the opposite direction of the initial one. Figure 2.23 shows the Pixel detector movements for a fill.

The alignment process, which corrects these kinds of relative quick movements of the Pixel and the IBL and the relative position with respect to the baseline alignment of all the other sub-detectors, is executed for each LHC fill automatically. Moreover, dedicated alignment campaigns are done to perform a detailed alignment of all the structures. Additionally, extra studies to mitigate the effects of the weak modes were done during these campaigns.

To sum up, the accuracy achieved after the alignment of the ID allowed the ATLAS detector to reconstruct particles without any loss in efficiency. The analysis in which this thesis is focused is really sensitive to this fact for two main reasons. First, the final states studied involve two or three leptons. Second, the *tHq* production would be clearly affected by any loss in efficiency of the reconstruction of charged particles due to its low

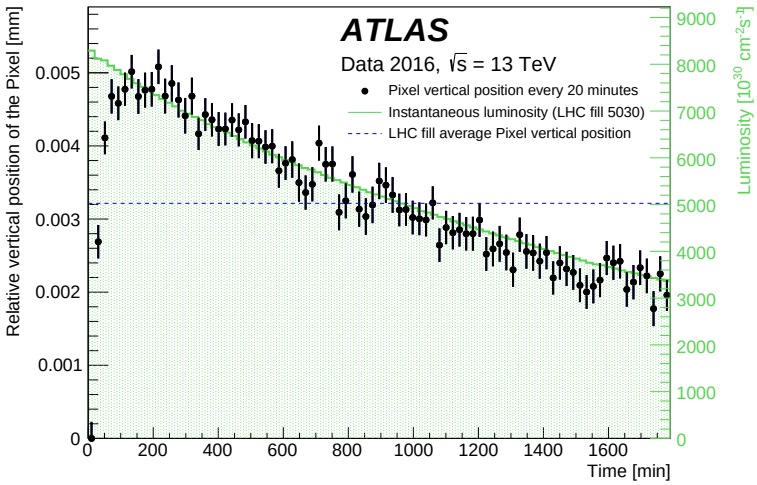


FIGURE 2.23: Vertical movements of the Pixel detector as a function of the time since the start of one LHC fill. The average of the displacement during the fill (dashed line) is compared with its evolution and its instantaneous luminosity. The uncertainty bars only represent statistical uncertainty [70].

cross-section production.

CHAPTER 3

Data and simulated events

The goal of this section is to describe data event samples collected by the ATLAS detector and the Monte Carlo (MC) simulated event samples, both used in the analysis presented in chapter 5. Either events from data or MC simulation are processed by the same reconstruction software in the ATLAS experiment, called ATHENA [72]. In the case of data events, the information from the ATLAS detector, which is described in section 2.3, is used to reconstruct the physical objects. They are described in chapter 4 and they make up the final states. In the case of MC simulated events, they are generated by MC generators, and they are passed through the detector simulation before reconstructing the final-state objects. Data were collected from pp collisions at the LHC, from 2015 to 2018, by the ATLAS detector.

The description of the data event samples is done in section 3.1. An overview of all the different steps performed in the MC simulation is presented in section 3.2. Finally, a list of the MC simulated samples used in this thesis is shown for the signal and the background processes in section 3.3.1 and 3.3.2.

3.1 Data event samples

The analysed data event samples were from $25\,\mathrm{ns}$ pp collision delivered by the LHC Run 2, i.e. from 2015 to 2018 , at $\sqrt{\mathrm{s}} = 13\,\mathrm{TeV}$ and collected by the ATLAS detector. Events were selected from a common data stream using unprescaled single-lepton triggers as described in Refs. [73–75]. Events that fired single-electron triggers in the data stream from single-muon triggers were not selected in order to avoid double counting of events.

The registered data events were filtered at the level of small portions of luminosity, called luminosity blocks. These events were stored in good-run lists. This fact means they were registered where the LHC beams were stable as well as a proper performance

of all the detector and trigger components. The amount of data used by this analysis corresponds to an integrated luminosity of 139.0 fb^{-1}. The total uncertainties on the integrated luminosities for each individual year of data-taking range from $2.0\,\%$ to $2.4\,\%$ and are partially correlated between years [76]. These uncertainties are derived from the calibration of the luminosity scale using x-y beam-separation scans, following a methodology similar to that detailed in Ref. [77], and using the LUCID-2 detector for the baseline luminosity measurements [78]. The uncertainty in the combined 2015–2018 integrated luminosity is 1.7% [76], that means 2.4 fb^{-1}, obtained using the LUCID-2 detector for the primary luminosity measurements. The partial and total integrated luminosities together with their uncertainties and some additional details are given in table 3.1. The explanation about the luminosity and its cumulative value delivered by the ATLAS experiment is shown in section 2.2.

TABLE 3.1: Integrated luminosity per year with their relative uncertainties. Additionally, run numbers per year are shown.

Year	Periods	Run numbers	Number of events (10^6)	Integrated luminosity $[pb^{-1}]$
2015	D–J	276262–284484	220.58	$3219.56 \pm 2.1\%$
2016	A–L	297730–311481	1057.84	$32988.1 \pm 2.2\%$
2017	B–K	325713–340453	1340.80	$44307.4 \pm 2.4\%$
2018	B–Q	348885–364292	1716.77	$58450.1 \pm 2.0\%$
2015–2018	All	276262–364292	4335.99	$138965.16 \pm 1.7\%$

3.2 Simulation event samples

This process is performed in several steps from the calculation of the parton-level cross-section to the simulation of the parton cascade and non-perturbative effects, and the simulation of the detector. The MC generation provides a set of final-state particles. They describe the fundamental physics within the SM, and allow to model the signals and background processes of the collisions. A schematic view of a pp collision is shown in figure 3.1, where the different steps are highlighted. Given the importance and relevance in this thesis, this section includes a general explanation about the techniques used in the simulation and a brief description of the most common generators.

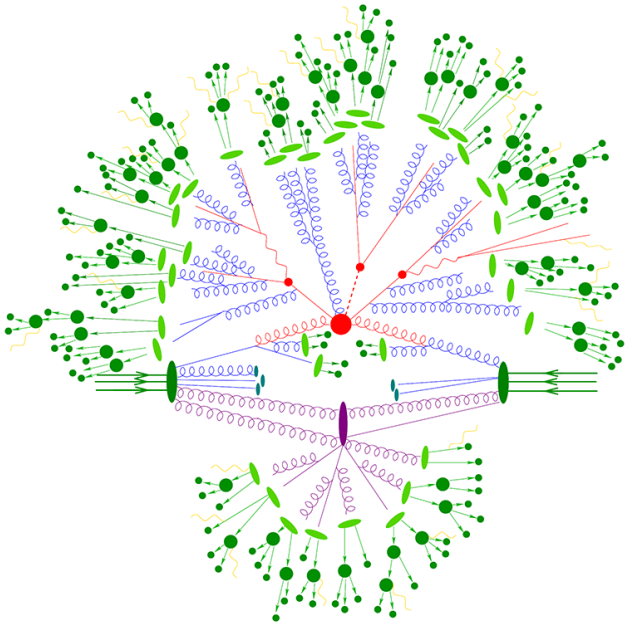


FIGURE 3.1: A sketch of the structure of a pp collision. The centre represents the hard scattering process (red). The red blob in the centre is surrounded by a tree-like structure representing Bremsstrahlung radiation by parton shower. The blue blobs indicate the initial-state partons. The secondary hard scattering (underlying event) is shown in violet. Finally, the hadronisation (light green) and the hadronical final states (green) are also shown. The yellow lines represent the soft photon radiation [79].

3.2.1 Monte Carlo simulation

The MC simulations involve different physic processes: hard scattering also known as matrix element (ME), parton shower (PS), hadronisation, underlying event (UE), hadron decays, and pile-up. These processes are simulated in different ways, and they use different input information.

3.2.1.1 Hard scattering

The first process to simulate is the hard scattering or ME. This step involves the interaction of simulated initial particles, e.g. pp, at high momentum. The ME provides the

physics objects from the hard-scattering process ,i.e. pp collisions in the presented analysis, up to the parton level [80]. These are used as starting point in the following steps of the simulation.

3.2.1.2 Parton-shower simulation

The second process to simulate is the PS [81]. The simulation starts with the computation of the hard-scattering cross-section at some given order in perturbation theory. All the incoming or outgoing partons are involved in the PS simulation. The PS simulation involves large momentum transfers and radiations since electric and colour charge particles can emit either QED (i.e. photons) or QCD (i.e. gluons) radiation. The PS simulation computes the terms of the perturbative expansion in the strong coupling constant considering the gluons emissions.

3.2.1.3 Hadronization simulation

The simulated quarks and gluons from the hard-scattering simulation, the PS and multiple-scattering simulation must become in colourless final states after the PS simulation. This process is known as hadronisation [81]. It involves phenomenological models to describe the mechanisms in which partons are joined in hadrons. The hadronisation is based on the parton–hadron duality hypothesis [82]. For this reason, the interchange of momentum and quantum numbers at the hadron level must follow the same rules as at parton level. Consequently, partons are joined among them to create hadrons according to their distances in their own phase spaces.

Nowadays, there are two main algorithms to compute the hadronisation. They are called the Lund string model [83, 84] and the cluster model [85]. They are briefly described as follows:

- The Lund string model describes the colour dynamics between quarks in terms of strings. The model assumes a linear confinement potential. If the distance between a pair of partons increases, the energy of the string increases through the confinement potential. In case of the energy arrives up to the mass threshold of a new $q\bar{q}$ pair, the string is broken and the new $q\bar{q}$ pair causes the formation

of hadrons. Moreover, additional strings are created when a gluon perturbatively splits, whereas the remaining gluons at the end of the PS lead to kinks in the string segments that connect them.

- The cluster model is based on the characteristics of the PS. The adjacent colour connected particles have an asymptotic mass distribution that steeply falls at high masses and is asymptotically independent. The model starts with a non-perturbative splitting of a gluon in a pair of quarks. Later, the pairs of quarks are merged into colour singlet combinations, which form clusters. Finally, these clusters decay into pairs of hadrons following an isotropic pattern.

3.2.1.4 Underlying event simulation

There are some extra hadron productions which are not linked to any showering from the coloured partons participating in the subprocesses from the primary process. They are known as underlying event (UE) [81]. The UE arises from the collision of partons which are not included in the hard subprocesses. In other words, the UE includes all simulated objects which are not coming from the primary hard-scattering process. The parameters of the model involved in the UE simulation need to be tuned using experimental data.

3.2.1.5 Hadron decay simulation

The last step of the generation chain is the decay of unstable hadrons. The experimental data indicate that a large fraction of the observed final-state particles come from the decays of excited hadronic states. Therefore, the majority of the known excited hadrons, and their decay modes, needs to be included in the simulation.

3.2.1.6 Pile-up simulation

The effect produced by multiple interactions per bunch crossing are computed by overlaying the original hard-scattering events, i.e. pp collisions. The simulated events were weighted to reproduce the distribution of the average number of interactions per bunch

crossing ($< \mu >$) observed in data. The $< \mu >$ value in data was rescaled by a factor of 1.03 ± 0.04 to improve the agreement between data and simulation in the visible cross-section of inelastic pp collisions as measured in data [86].

3.2.2 Monte Carlo generators

There are several event generators programs in the market to produce the MC event simulation. All of them use features described in section 3.2 and its subsections. The goal of the MC event generator is to describe experimental data for physics processes. The most common are described as follows:

- POWHEG BOX [87–90] is a NLO generator. It produces a hard-scattering ME for each event following the $2 \to 2$ or the $2 \to 3$ schemes.
- MADGRAPH is also a generator based on ME [91]. It produces a hard-scattering ME for each event following $2 \to 1$, $2 \to 2$ or $2 \to 3$ schemes. Later, the information of all the generated event is passed to PYTHIA or HERWIG for the PS step. MADGRAPH can simulate process either to LO for any Lagrangian defined by the user or to NLO in the case of QCD corrections to the SM.
- SHERPA is a generator [92] offers a complete set of hadronic final states in simulation for high-energy particle collisions. It includes both ME and PS simulation at LO or NLO.
- PYTHIA 8 shower generator is a MC event generator based on MEs at LO, and it implements the calculation for $2 \to 1$ and $2 \to 2$. Furthermore, the initial state radiation (ISR) and the final state radiation (FSR) are matched in $\mathrm{p_T}$-ordered in the PS. In this case, the Lund model is on the base of the fragmentation simulation. The UE uses a multiple-interaction model [93].
- HERWIG 7 [94] generator includes a huge diversity of QCD processes which includes ME and PS simulation at LO and NLO. Contrary to PYTHIA 8, the PS is ordered using either angular distributions or dipole distribution.

All the different MC generators described in the items above are used in this thesis. Some of them are used to produce the nominal simulation samples, and others to produce

the alternative simulation samples to evaluate the systematic uncertainties due to the election of the MC event or PS generator.

3.2.3 Detector simulation

The last step of the simulation is the simulation of the ATLAS detector. The simulation is performed with the dedicated ATLAS software infrastructure [95] in two different ways: either including a detailed physics description simulation of all subdetectors with the GEANT4 [96] framework, i.e. full-simulated (FS) detector response or considering a parametric cell response of the ATLAS calorimeter and a complete GEANT4 description and detector response for the rest [95], i.e. fast-simulated (AFII).

In the analysis presented in chapter 5 the FS event samples are always used as baseline samples unless not available. The AFII event samples are mainly used for evaluating most of the systematic effects.

3.3 Simulated event sample

As it is mentioned before the simulated event samples were produced using one or a combination of the MC generator listed on section 3.2.2. After the event generator step, the detector simulation is performed as it is explained in section 3.2.3.

The pile-up simulation, explained in section 3.2.1.6, was modelled by overlaying the simulated hard-scattering event with inelastic pp events generated with PYTHIA 8.186 [97] using the NNPDF2.3LO set of parton distribution functions (PDFs) [98] and the third ATLAS set of tuned parameters for minimum-bias events (A3 tune) [99] over the original hard-scattering event.

In the analysis shown in chapter 5, samples of event generated using MC simulations were produce for the *tHq* signal process and most of the background processes. These event samples are used to evaluate models of efficiency and resolution. Moreover, some systematic sources are estimated using alternative simulated event samples. Table 3.2 summarises the simulated and background event samples used in the analysis as baseline. Details about simulation samples for the most important processes are given in sections 3.3.1 and 3.3.2.

TABLE 3.2: Summary of the baseline simulated signal and background event samples used in the *tHq* multi-lepton analysis.

Process	Generator	ME order	PDF set	Parton shower	PDF set (tune)
			Signal		
tHq	MADGRAPH5_AMC@NLO 2.6.2	NLO (4FS)	NNPDF3.0NLO nf4	PYTHIA 8.230	NNPDF2.3LO (A14 tune)
			Backgrounds		
$t\bar{t}$	POWHEG BOX v2	NLO (5FS)	NNPDF3.0NLO	PYTHIA 8.230	NNPDF2.3LO (A14 tune)
V+jets	SHERPA 2.2.1	NLO+LO	NNPDF3.0NNLO	-	-
Diboson	SHERPA 2.2.1-2	NLO+LO	NNPDF3.0NNLO	-	-
Triboson	SHERPA 2.2.2	NLO+LO	NNPDF3.0NNLO	-	-
$t\bar{t}V$	MADGRAPH5_AMC@NLO 2.3.3	NLO	PYTHIA 8.210	NNPDF2.3LO (A14 tune)	
$t\bar{t}H$	POWHEG BOX v2	NLO (5FS)	PYTHIA 8.230	NNPDF2.3LO (A14 tune)	
t-channel	POWHEG BOX v2	NLO (4FS)	NNPDF3.0NLO nf4	PYTHIA 8.230	NNPDF2.3LO (A14 tune)
Wt	POWHEG BOX v2	NLO (5FS, DR)	NNPDF3.0NLO	PYTHIA 8.230	NNPDF2.3LO (A14 tune)
s-channel	POWHEG BOX v2	NLO	NNPDF3.0NLO	PYTHIA 8.230	NNPDF2.3LO (A14 tune)
tZq	MADGRAPH5_AMC@NLO 2.3.3	NLO	NNPDF3.0NLO	PYTHIA 8.230	NNPDF2.3LO (A14 tune)
tWH	MADGRAPH5_AMC@NLO 2.8.1	NLO (5FS, DR)	NNPDF3.0NLO	PYTHIA 8.245p3	NNPDF2.3LO (A14 tune)
tWZ	MADGRAPH5_AMC@NLO 2.3.3	NLO	NNPDF3.0NLO PYTHIA 8.212	NNPDF2.3LO (A14 tune)	
ttt	MADGRAPH5_AMC@NLO 2.2.2	NLO	NNPDF3.1NLO PYTHIA 8.186	NNPDF2.3LO (A14 tune)	
tttt	MADGRAPH5_AMC@NLO 2.3.3	NLO	NNPDF3.1NLO PYTHIA 8.230	NNPDF2.3LO (A14 tune)	
ggH	POWHEG BOX v2	NLO	CT10	PYTHIA 8.210	CTEQ6L1 (AZNLO tune)
qqH	POWHEG BOX v1	NLO	CT10	PYTHIA 8.186	CTEQ6L1 (AZNLO tune)
WH	PYTHIA 8.186	LO	NNPDF2.3LO	-	-
ZH	PYTHIA 8.186	LO	NNPDF2.3LO	-	-

3.3.1 Simulated signal sample

The event samples for the *tHq* process were simulated using the MADGRAPH5_AMC@NLO 2.6.2 [91] generator at NLO with the NNPDF3.0NLO nf4 [100] PDF set. PYTHIA 8.230 [101] was used to interfaced the events with the A14 tune parameters (A14 tune) [102] and the NNPDF2.3LO [100] PDF set. The top quark was decayed at LO using MADSPIN [103, 104] to preserve spin correlations, whereas the Higgs boson was decayed by PYTHIA in the PS.

Alternative samples of simulated *tHq* signal events with a different PS generator were produced using the MADGRAPH5_AMC@NLO 2.8.1 at NLO with the NNPDF3.0NLO nf4 PDF set, interfaced with HERWIG 7.1.6 with the MMHT2014NNLO [105] PDF set, using the HERWIG 7.1 default set of tuned parameters.

All these samples were generated in the 4FS. A filter at parton level of at least two leptons was used. The efficiencies of the filter were estimated using the Rivet package [106] and they are shown in table 3.3

The functional form of the renormalisation and factorisation scales, i.e. μ_r and μ_f, was set to the default scale $0.5 \times \sum_i \sqrt{m_i^2 + p_{T,i}^2}$, where the sum runs over all the particles

TABLE 3.3: Efficiencies of the filter requiring at least two leptons. The leptons have $p_T > 5$ GeV and $|\eta| < 10$ at ME generation level.

Higgs boson decay	$\varepsilon_{2\ell}$
Inclusive	0.38
ZZ^*,WW^*,$\tau\tau$	0.50
$b\bar{b}$	0.35
$c\bar{c}$	0.25
gg	0.17
$\gamma\gamma$	0.22

generated from the ME calculation. The decays of bottom and charm hadrons were simulated using the EVTGEN 1.6.0 or 1.7.0 program [107]. The simulation samples only include $H \to \tau\tau/H \to ZZ^*/H \to WW^*$.

To simulate higher order contribution, μ_r and μ_f scales were reduced independently by a factor of 0.5. For lower parton radiation, μ_r and μ_f were increased independently by a factor of two. In both cases, these variations are included as additional event weights in the nominal *tHq* event sample.

3.3.2 Simulated background event samples

A list of simulated samples is used to reproduce the kinematic distribution of the SM background processes. All the background simulated event samples included in this analysis are described in this section. The most important background event samples: top-quark pair ($t\bar{t}$) process and top-quark pair association of a single-boson ($t\bar{t}V$) process, are explained in detail. The other background event samples are listed at the end of this section and also summarised in table 3.2.

Top-quark pair process The production of a pair of top quarks ($t\bar{t}$) events was modelled using the POWHEG BOX v2 [87–90] generator, which provided ME at NLO in the strong coupling constant (α_S), and the NNPDF3.0NLO set of PDF. The h_{damp} parameter, which controls the matching in POWHEG and effectively regulates the high-p_T radiation against which the $t\bar{t}$ system recoils, was set to 1.5 m_{top} [108]. The functional form of the μ_r and μ_f was set to the default scale $\sqrt{m_{top}^2 + p_T^2}$. The events were interfaced with PYTHIA 8.230 for the PS and hadronisation, using the A14 tune and the NNPDF2.3LO

set of PDFs. The decays of bottom and charm hadrons were simulated using the EVTGEN 1.6.0 program. The analysis uses a non-all-hadronic filtered simulation sample.

To assess the uncertainty in the matching of NLO MEs to the PS, a POWHEG BOX sample was compared with an event sample generated by MADGRAPH5_AMC@NLO. The first sample is the nominal $\mathrm{t\bar{t}}$ sample just described above while the second sample used MADGRAPH5_AMC@NLO 2.6.0 with the NNPDF3.0NLO PDF set for the calculation of the hard scattering. The PS starting scale had the functional form $\mu_\mathrm{q} = \mathrm{H_T}/2$ [109], where $\mathrm{H_T}$ is defined as the scalar sum of the $\mathrm{p_T}$ of all outgoing partons. The events from both generators were interfaced with HERWIG 7.13, using the HERWIG 7.1 default set of tuned parameters and the MMHT2014LO PDF set [105]. The μ_r and μ_f choice in the MADGRAPH5_AMC@NLO set-up was the same as for the POWHEG BOX set-up.

The impact of using a different PS and hadronisation model was evaluated by comparing the nominal $\mathrm{t\bar{t}}$ sample with an event sample also produced with the same generator but interfaced with HERWIG 7.13, using the HERWIG 7.1 default set of tuned parameters and the MMHT2014LO PDF set. POWHEG BOX provided MEs at NLO in the α_S and used the NNPDF3.0NLO PDF set and an $\mathrm{h_{damp}}$ parameter value of $1.5\,\mathrm{m_{top}}$.

Alternative samples using Var3c up and down variations from the A14 tune, where the Var3c A14 tune variation largely corresponds to the variation of α_S for ISR in the A14 tune to estimate the uncertainty due to ISR. To simulate higher order contribution, μ_r and μ_f scales were reduced independently by a factor of 0.5 while simultaneously increasing the $\mathrm{h_{damp}}$ value to $3.0\,\mathrm{m_{top}}$ and using the Var3c up variation from the A14 tune. For lower ISR, μ_r and μ_f were increased by a factor of two while keeping the $\mathrm{h_{damp}}$ value set to $1.5\,\mathrm{m_{top}}$ and using the Var3c down variation in the PS. The impact of FSR was evaluated by varying the renormalisation scale for emissions from the PS up and down by a factor of two. All these variations were implemented in the nominal $\mathrm{t\bar{t}}$ simulation sample as alternative weights.

A variation of the $\mathrm{h_{damp}}$ parameter was considered by comparing nominal with alternative event samples with $\mathrm{h_{damp}}$ parameter set to $3.0\,\mathrm{m_{top}}$.

All these samples were generated in the 5FS, and top quarks were decayed at LO using MADSPIN to preserve spin correlations. The decays of bottom and charm hadrons

were simulated using the EvtGen 1.6.0 program.

The $t\bar{t}$ sample was normalised to the cross-section prediction at next-to-next-to-leading order (NNLO) in QCD including the resummation of next-to-next-to-leading logarithmic (NNLL) soft-gluon terms calculated using Top++ 2.0 [110–116]. For pp collisions at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$, this cross-section corresponds to $\sigma(t\bar{t})_{\text{NNLO+NNLL}} = 832 \pm 51$ fb using a top-quark mass of $m_{\text{top}} = 172.5\,\text{GeV}$. The uncertainties in the cross-section due to the PDF and α_S were calculated using the PDF4LHC15 prescription [117] with the MSTW2008nnlo [118, 119], CT10nnlo [120, 121] and NNPDF2.3lo PDF sets in the 5FS and were added in quadrature to the effect of the scale uncertainty.

Top-quark pair + Single-boson process The production of $t\bar{t}V(V = W/Z)$ events was modelled using the MadGraph5_aMC@NLO 2.3.3 generator, which provided MEs at NLO in the α_S with the NNPDF3.0nlo PDF. The functional form of the μ_r and μ_f was set to the default of $0.5 \times \sum_i \sqrt{m_i^2 + p_{T,i}^2}$, where the sum runs over all the particles generated from the ME calculation. Top quarks were decayed at LO using MadSpin to preserve spin correlations. The events were interfaced with Pythia 8.210 for the PS and hadronisation, using the A14 tune and the NNPDF2.3lo PDF set. The decays of bottom and charm hadrons were simulated using the EvtGen 1.2.0 program. The $t\bar{t}W$ event sample also includes EW corrections.

The cross-sections were calculated at NLO QCD and NLO EW accuracy using MadGraph5_aMC@NLO as reported in Ref. [122]. In the case of $t\bar{t}\ell\ell$ the cross-section was scaled by an off-shell correction estimated at one-loop level in α_S. The predicted values at $\sqrt{s} = 13\,\text{TeV}$ are $0.88^{+0.09}_{-0.11}$ pb and $0.60^{+0.08}_{-0.07}$ pb for $t\bar{t}Z$ and $t\bar{t}W$, respectively, where the uncertainties were estimated from variations of α_S and the μ_r and μ_f.

Additional $t\bar{t}V$ samples were produced with the Sherpa 2.2.10 [123] generator at LO accuracy, using the MEPS@LO set-up with up to one additional parton for the $t\bar{t}\ell\ell$ sample and two additional partons for the others. A dynamic μ_r scale was used and is defined similarly to that of the nominal $t\bar{t}V$ samples. The CKKW matching scale of the additional emissions was set to 30 GeV. The default Sherpa 2.2.10 PS was used along with the NNPDF3.0nnlo PDF set.

Two additional samples were generated with same settings as the nominal one but employed the Var3c up or down variation of the A14 tune, which corresponds to the variation of α_S for ISR in the A14 tune to estimate the uncertainty due to ISR. Uncertainties due to missing higher-order corrections were evaluated by simultaneously varying the μ_r and μ_f by factors of 2.0 and 0.5. These variations are included as additional event weights in the nominal $t\bar{t}V$event sample.

Non-dominant processes

Non-dominant background processes considered in the current analysis are: single boson ($V + jets$), diboson (VV), triboson (VVV), Higgs boson (ggF, VBF, VH), top-quark pair in association with a Higgs boson ($t\bar{t}H$), single top-quark ($t-channel$, $s-channel$, $tW-channel$), single top-quark in association with a Z boson (*tZq*), with both a Z and a W boson (tWZ) and with a W boson and a Higgs boson (tWH), three top quark (ttt) and four top quark ($tttt$). The baseline event samples used for all of these processes are summarised in table 3.2.

CHAPTER 4

Object definition and event reconstruction

The information of all interactions of the particles in the different systems of the ATLAS detector is combined to convert electronic pulses in physical objects. This process is known as reconstruction. The goal of this chapter is briefly described the physical objects used in this thesis and the methods to reconstruct them. All the reconstruction methods follow the latest recommendation for either simulation data or real data collected by the ATLAS detector between 2015 and 2018.

This chapter is divided as follows: sections 4.1 and 4.2 briefly explain the track and vertex reconstruction and the trigger selection, respectively. The following sections describe how physical objects are defined in the analysis included in this thesis: section 4.3 for electrons and muons, section 4.5 for jets and section 4.6 for missing transverse momentum. Finally, section 4.7 describes the criteria to avoid the overlap between physical objects.

4.1 Tracking and vertex

The tracks, which are defined as the trajectories of charged particles, as mentioned in chapter 2, are mainly reconstructed with the information given by the ID. The particles leave a path of pulses when they pass through the detector. They interact with the different materials of each sub-detector according to its characteristics, as it is shown in figure 4.1. Then, these pulses are stored as space points in 3D coordinates of the detector. Finally, these points are used by the track reconstruction algorithms for the pattern recognition and to build the tracks [124].

In addition to the track reconstruction algorithm a neural network is used to identify tracks in jets [125]. A Kalman filter [126] is used to improve the performance of the reconstructed tracks. This algorithm is based on the use of consecutive Gaussian filters

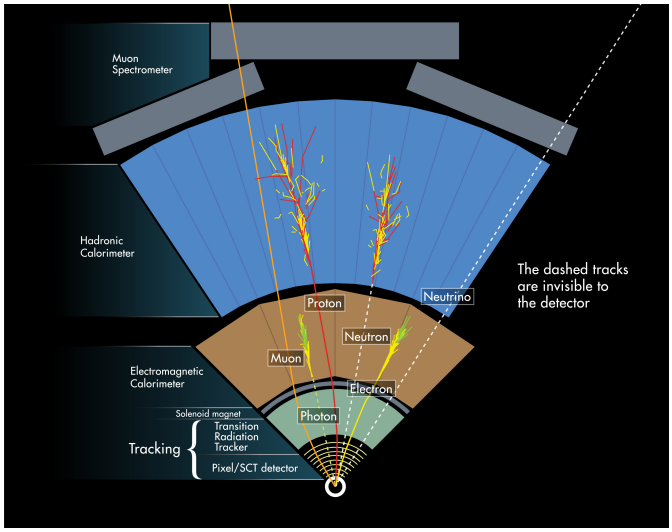


FIGURE 4.1: Diagram of different particle paths in the ATLAS detector. Different particle interactions with the sub-detectors are highlighted. Electrons and photons produce showers in the ECAL, and protons and neutrons produce showers in the HCAL. Muons tracks interact with all the sub-detectors. The dashed tracks are invisible to the detector, and solid tracks are visible to the detector.

[127]. The performance of the track reconstruction of the ATLAS ID at $\sqrt{s}$ =13 TeV is shown in Ref. [128].

The vertices are defined as the origin point of the tracks and are identified through a χ^2 function. The algorithm to identify the vertices uses clustering position measurements and the physical position of the modules of the detector. In particular, the primary vertex is the vertex with the highest scalar sum of the squared p_T from the associated tracks with $p_T > 400$ MeV [129].

4.2 Trigger selection

The ATLAS system trigger, explained in section 2.3.5, consists in two different levels, L1 and HLT, and also provides a list of triggers for physical analyses. In the analysis presented in this thesis different single-lepton un-prescaled triggers are used due to the change of the pile-up conditions during the Run 2 data-taking period [130, 131]. They

are combined using a logical OR. There are different single-lepton triggers for electrons and muon in the following way:

- The electron triggers identification criteria are based in a multivariate likelihood algorithm, the electron identification, and the electron isolation. Moreover, several lower transverse energy (E_T) thresholds, from 20 GeV to 26 GeV, were applied for electrons with low p_T. For electrons with high p_T, two additional complementary triggers were used with different lower E_T thresholds from 60 GeV to 140 GeV. The different electrons triggers are selected according to the data-taking year and the trigger level.
- The muon triggers are based in the matching of tracks reconstructed in the MS and in the ID and the muon isolation. Moreover, in this case, several lower p_T thresholds were applied from 20 GeV to 50 GeV according to the data-taking year and the muon isolation. Each one of these lower thresholds correspond to a different single-muon trigger.

Finally, MC event samples are rescaled to include the effects of trigger selection in data event samples.

4.3 Electrons and muons

Electron candidates are reconstructed from energy deposits in the ECAL associated with a track in the ID. The selected energy deposits are also used as clusters for the electron identification [132, 133]. The identification of prompt electrons relies on a likelihood-based algorithm which includes the measurements from the ID, the ECAL, and the combination of both. Several levels of identification are given by this method: *tight*, *medium* or *loose* according to the identification efficiency [134]. This algorithm is optimised to achieve the maximum discrimination between prompt and non-prompt leptons. Therefore, the *loose* category has a more significant acceptance of electrons but lower purity in the identification of prompt lepton than the *tight* category.

Electrons candidates in the presented analysis are required to satisfy $p_T > 10$ GeV, $|\eta_{\text{cluster}}| < 2.47$[1] and the *tight* level of identification electron. There is an exclusion

[1] $|\eta_{\text{cluster}}|$ is the η of the energy deposit in the ECAL.

region, $1.37 < |\eta_{\text{cluster}}| < 1.52$, where the electrons are not accepted due to the transition of the barrel and the end-cap sections of the ECAL. The track associated with the electron must also pass the requirements: $z_0 \cdot \sin\theta < 0.5$ mm and $d_0/\sigma(d_0) < 5$, where $\sigma(d_0)$ is the uncertainty of d_0.

Moreover, in the analysis presented in this thesis, specific algorithms are applied to electrons to improve the rejection of electrons with mis-identified electrical charge and to suppress the contribution from electrons originating from γ-conversions. For the first case, an algorithm called *Electron Charge ID Selector Tool* (ECIDS) is used. It is based on a boosted-decision tree (BDT) whose input variable are related to the electron characteristics, e.g. $q \times d_0$ or E/p. For the second case, a set of tags called ambiguity requirements is used. The different tags use information related to the origin of the electrons provided by the MC even sample generator. The ambiguity requirement is only applied in some of the region defined in chapter 5 for the analysis.

Muon candidates are reconstructed using the information of tracks either from the ID and from the MS. In this analysis, muons are reconstructed with independent information of tracks from the ID and the MS, with is finally combined. Therefore, these muons must be within the ID acceptance region. These tracks require a number of hits in the ID and in the MS, and the corresponding muon must have a minimum value of charge-to-momentum ratio q/p [135, 136]. In addition, the algorithm in Ref.[137] is used to identify prompt muons. It also provides three different levels of identification in increasing purity and decreasing efficiency: *tight*, *medium* and *loose*.

In this analysis, muon candidates satisfy the requirements: $p_T > 10$ GeV, $|\eta| < 2.5$ and the *medium* level of identification. Moreover, the track associated with the muon candidate must also satisfy $z_0 \cdot \sin\theta < 0.5$ mm and $d_0/\sigma(d_0) < 3$.

Isolation criteria are defined using the isolation level provided by a multivariate likelihood algorithm. They are applied to muons and electrons with $p_T > 10$ GeV. The algorithm involves the combination of the electromagnetic-shower shapes and track information from the ID. It is optimised to distinguish prompt leptons from fake/non-prompt leptons from hadronic jets, γ-conversions and heavy-flavour hadron decays. Two levels of isolation are given by the algorithm: *tight* and *very-tight*, increasing purity and decreasing efficiency. In the analysis presented in this thesis, the *tight* level of isolation is

required for both electrons and muons.

Finally, all the requirements for muons and electrons are summarised in table 4.1. Leptons are also required to satisfy a process called *overlap removal*. This method is explained in section 4.7, and it is requested after passing the selection criteria explained above except the isolation, identification and ECIDS criteria.

TABLE 4.1: Summary of the electron and muon object definitions used in the analysis presented in this thesis.

	Electrons	Muons
Identification	`tight`	`medium`
Acceptance	$p_T > 10\,\text{GeV}$, $\lvert\eta^{\text{cluster}}\rvert < 2.47$ except $1.37 < \lvert\eta^{\text{cluster}}\rvert < 1.52$	$p_T > 10\,\text{GeV}$, $\lvert\eta\rvert < 2.5$
Impact parameter	$\lvert d_0/\sigma(d_0)\rvert < 5.0$ $\lvert z_0 \sin(\theta)\rvert < 0.5$ mm	$\lvert d_0/\sigma(d_0)\rvert < 3.0$ $\lvert z_0 \sin(\theta)\rvert < 0.5$ mm
Isolation	`tight`	`tight`
Extra selection	ECIDS, ambiguity-cuts	
Overlap removal	See section 4.7	

4.4 Taus

Hadronic taus (τ_{had}) are vetoed in the analysis presented in this thesis. The selection criteria for the τ_{had} are: $p_T > 20$ GeV, $\lvert\eta^{\text{cluster}}\rvert < 2.5$ except $1.37 < \lvert\eta^{\text{cluster}}\rvert < 1.52$ and the number of associated tracks must be one or three. Afterwards, multivariate algorithms are applied to the objects in order to discriminate τ_{had} against other objects, mainly jets.

4.5 Jets

Jets are reconstructed using the anti-k_t algorithm [138] with a distance of the cone size set to $\Delta R = 0.4$[2]. The algorithm uses a four-momentum recombination schema, and it calibrates the jet energy to the hadronic scale with the effect of pile-up removed. The

[2] $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

clusters inside the cone in the calorimeters are weighted to add corrections due to the non-compensating nature of the calorimeters [139]. The jets outside the cone have a dedicated calibration method which involves the suppression of the jet area pile-up and different weights, which are based on the jet p_T to the particle level from MC simulations [140]. The jets are required to satisfy: $p_T > 20$ GeV and $|\eta| < 4.5$.

In addition to the anti-k_t algorithm two extra algorithm are applied to jets: one is used to select central jets[3] with a $p_T < 60$ GeV called *Jet Vertex Tagger* (JVT) [141, 142], the other is used to select forward-jets[4] with a $p_T < 120$ GeV called *forward Jet Vertex Tagger* (fJVT) [143].

The requirements of the selected jets are summarised in table 4.2. Moreover, the *overlap removal* explained in section 4.7 is also applied to selected jets.

4.5.1 b-tagged jets

The jets originated from the hadronisation of a b-quark are known as b-jets and its identification as b-tagging. A specific tagger algorithm called `DL1r` is used as b-tagging algorithm [144]. This tagger distinguishes between b-jets and other sources of jets like c-jets (jets originated from c-quarks), light-flavour jets or simply named as *light*-jets (jets originated from gluons or other flavours quarks). The `DL1r` is a multivariate algorithm which combines the information from the impact parameters of the displaced tracks and topological properties of secondary and tertiary decay vertices reconstructed within the jets to identify b-jets [145–147]. The b-jets are jets whose values of the `DL1r` are above a certain threshold, hereafter referred to as WPs. Four WP are defined for the `DL1r`, the 70% of the b-jets being selected in $t\bar{t}$ simulated event WP is used to define the b-jets. The efficiency of the `DL1r` algorithm is measured in collision data.The detector can only identify b-jets in its central region ($|\eta| < 2.5$) since information from the tracks, i.e. from the ID, is needed for the discrimination. Moreover, an extra requirement is applied to b-jets: $p_T > 20\,\text{GeV}$. The requirements for the b-jets are also summarised in table 4.2.

[3] In this analysis, a central jet is defined with $|\eta| < 2.5$.
[4] In this analysis, a forward jet is defined with $2.5 < |\eta| < 4.5$.

TABLE 4.2: Summary of selection criteria for jets and b-jets.

Jet	
Acceptance	$p_T > 20\,GeV, \|\eta\| < 4.5$
Jet Vertex Tagger	JVT > 0.5 if $\|\eta\| < 2.4$ and $p_T < 60\,GeV$
Forward Jet Vertex Tagger	fJVT < 0.4 if $2.5 < \|\eta\| < 4.5$ and $p_T < 120\,GeV$
Overlap removal	See 4.7
b-tagging jet	
Acceptance	$p_T > 20\,GeV$, $\|\eta\| < 2.5$
b-tagging	`DL1r` algorithm

4.6 Missing transverse momentum

The sum of the p_T of all the products of a collision must be zero in an ideal case due to the conservation of the momentum in the transverse plane to the beam. Thus, the negative vector sum of the p_T of the reconstructed and calibrated objects is known as missing transverse momentum [148, 149], and its magnitude is written in the following as E_T^{miss}. Besides the vector components associated with the final-state particles, the transverse momentum deposited in the detector, which is not associated with any hard process, is also considered (this term is named as *soft term*). The E_T^{miss} is related to undetected particles, such as neutrinos, and to the energy lost due to detector inefficiencies.

4.7 Overlap removal

Objects could satisfy different selection at the same time. Therefore, the overlap between objects is resolved in order to avoid double-counting of physics objects using a specific procedure called *overlap removal* as mentioned above. In the special case of leptons, the *overlap removal* is applied over leptons without the ECIDS requirement and with the loosest level of isolation and identification. The *overlap removal* requirements are:

1. Any electron found to share a track with a muon is removed. Since the electron is very likely to correspond to the reconstructed muon.

2. Any jet found within a ΔR of 0.2 of an electron is removed due to the fact that the jet is very possible that corresponds to the electron.

3. In order to reduce the impact of non-prompt electrons, if any electron subsequently found within ΔR of 0.4 of a jet is removed.

4. Any jet with less than three tracks associated to it and separated from a muon by $\Delta R < 0.2$ is removed to avoid fake jets from muons depositing a large fraction of their energy in the calorimeter.

5. Any jet with less than three tracks associated to it, which has a muon ID track ghost-associated to it, is removed to reduce the number of fake jets from muons depositing energy in the calorimeters.

6. Muons subsequently found within ΔR of 0.4 of a jet are removed to reduce the contribution from muons from heavy-flavour decays inside a jet.

The criteria of the *overlap removal* are applied in the specific order followed in the list above.

CHAPTER 5

Search of *tHq*

As already mentioned, the search of the *tHq* process using the data collected by the ATLAS experiment during the Run 2 is the main topic of this thesis. In this analysis only two final states of the *tHq* process are considered: three final-state light-flavour leptons (3ℓ) and two final-state light-flavour leptons with the same charge (2ℓSS).

The goal of this analysis is to perform the first direct search of the *tHq* process towards its first observation at the LHC. Moreover, an upper limit to the production cross-section will be set using a statistical algorithm. The analysis is already theoretically motivated in section 1.4. The experimental configuration of the ATLAS detector for the data-taking periods used in this analysis is discussed in section 2.3. Finally, the different simulated processes as well as the physical-object definitions and the techniques for the simulations used are described in chapter 3 and 4, respectively.

This chapter shows a detailed view of the analysis following its different steps together with its results. The set of requirements which describes the pre-selection region, and the multivariate analysis (MVA) algorithm are explained in sections 5.1 and 5.2. Moreover, the set of input variables to the MVA and how it is optimised are discussed in sections 5.1 and 5.2 for the 3ℓ and the 2ℓSS channel, respectively. Section 5.3 gives an overview of the different sources of backgrounds and how special backgrounds are estimated. The regions of interest used in the analysis are defined in section 5.4. The different sources of systematic uncertainties are listed in section 5.5. Finally, the results of the complete analysis are given and discussed in section 5.6.

5.1 Event selection for the 3ℓ final state

This section covers the selection strategy for the 3ℓ channel. First, an initial set of requirements is defined, named pre-selection region. This region is used as starting point

of MVA techniques to further enhance the signal to background separation. Second, the techniques and the strategies for MVA followed are explained in a detailed way.

5.1.1 Pre-selection requirements

Pre-selection criteria are requested before applying the MVA technique to the 3ℓ final state. The pre-selection region is defined with the following requirements:

- Exactly three light-flavour leptons (electrons or muons, as defined in section 4.3).
- The sum of the charge of the leptons must be ± 1.
- The three leptons, which are ordered by their $\mathrm{p_T}$, with the leading lepton having $\mathrm{p_T} > 27$ GeV, sub-leading lepton having $\mathrm{p_T} > 20$ GeV,and the softest lepton having $\mathrm{p_T} > 10$ GeV.
- The events with jets originated by hadronically decaying from tau leptons, as defined in section 4.4, are vetoed.
- The number of jets is required to be between one and six, as defined in section 4.5.
- The number of b-jet is required to be between one and three, as defined in section 4.5.1.
- The $\mathrm{E_{Tmiss}}$ is required to be between 5 GeV and 800 GeV.

The goal of these criteria is to maximise the signal acceptance while minimising the background contamination. Table 5.1 shows the event yields of the different processes contributing to the pre-selection region as predicted by the MC simulation, together with data events.

The minor-background composition is done regarding the negligible MC simulation samples in the signal region (defined in table 5.9). The MC simulation samples in minor backgrounds are: triboson, single top-quark t-channel and s-channel, *W+jets*,*tttt*,*ttt*, ggF, VBF and VH.

TABLE 5.1: The 3ℓ channel pre-selection region yields as predicted by the MC simulation and data events. The uncertainties include statistical and all the systematic sources.

Process	Yields
tHq	2.53 ± 0.11
tWH	3.12 ± 0.21
tWZ	80 ± 42
$t\bar{t}$	322 ± 7
Z+jets	135 ± 25
$t\bar{t}W$	173.3 ± 5.8
$t\bar{t}Z$	563 ± 125
$t\bar{t}H$	74 ± 12
tZq	271 ± 35
tW	19.5 ± 8.6
Diboson	571 ± 143
Minor backgrounds	15.7 ± 9.0
Total background	2231 ± 216
Data	2457

5.1.2 Multivariate analysis

The results obtained in the 3ℓ channel depend on the discrimination power provided by a given MVA method using a particular set of input variables (also known as features). The response variable of the MVA algorithm is used to define enriched regions in either the signal process or a particular background process. In the 3ℓ channel, the *XGBoost* python library [150] is used to develop three independent boosted decision trees (BDTs), which target the *tHq* signal process, the $t\bar{t}$ and the $t\bar{t}W$ background processes, respectively. The BDTs are trained using events passing the pre-selection criteria described in section 5.1.1. Roughly 30% of the *tHq* simulated events have negative MC weights. Such events cannot be used in the training as their presence would bias the BDT response. Different options are explored to deal with negative-weighted events. The exclusion of the negative-weighted events from the training process is found to be the best solution for the analysis of the 3ℓ channel. More details about the tested options are given in appendix A. The list of input variables used in each BDT is optimised as described in the incoming section 5.1.2.1. The optimisation of the BDT parameters together with the final performance and resulting scores are reported in section 5.1.2.2.

5.1.2.1 Input variables and their importance

The sets of discriminant variables used as input for each BDT training are described in this section. Different ordering of the three leptons in the final state are tested to build some of the variables. Leptons are ordered in three different ways:

- *Ordering based on transverse momentum*: the lepton with the highest $\mathrm{p_T}$ is called ℓ_A, the one with the second highest $\mathrm{p_T}$ is called ℓ_B, while the lepton with the softest $\mathrm{p_T}$ is called ℓ_C.
- *Ordering based on charge and* $\Delta\mathrm{R}$ *between leptons*: the lepton with the opposite charge to the sum of the three lepton charges is called ℓ_0. The lepton with the smallest $\Delta\mathrm{R}$ to ℓ_0 is called ℓ_1, and the remaining lepton is called ℓ_2.
- *Ordering based on charge and* $\Delta\mathrm{R}$ *with respect the leading b-jet*: the lepton with the opposite charge to the sum of the three lepton charges is called $\hat{\ell}_0$, for consistency. The lepton with the smallest $\Delta\mathrm{R}$ to leading b-jet is called $\hat{\ell}_1$, while the remaining lepton is called $\hat{\ell}_2$.

The jet passing the b-tagging requirement and having the highest $\mathrm{p_T}$ is called the leading b-jet. The jet that fails the b-tagging requirement and maximises the invariant mass with the leading b-jet is called the spectator jet, labelled as $\mathrm{jet_{spect}}$ in the following. Every jet, not b-tagged, having $|\eta| > 2.5$ is called a forward jet, and it is labelled as $\mathrm{jet_f}$. A forward jet is defined as leading if it has the highest $\mathrm{p_T}$ among all the forward jets. Every jet with $|\eta| < 2.5$ is called a central jet, it can be either b-tagged or not. Using the aforementioned lepton and jet definitions, several variables are constructed.

The list of variables used in each BDT is optimised with an iterative approach based on the impact they have on the BDT performance (parametrised by the figure of merit *Gain*[1] value) and on their correlations with the other variables. More details are given in appendix B.1.

[1] The *Gain* value, as calculated by the *XGBoost* package, represents the accuracy brought by the variable to the BDT branches where it is used.

As it is mentioned before, three independent BDTs are used targeting different processes: the signal process *tHq* (named BDT(*tHq*)) and the $\mathrm{t\bar{t}}$ and $\mathrm{t\bar{t}}W$ background processes (named BDT($\mathrm{t\bar{t}}$) and BDT($\mathrm{t\bar{t}}W$), respectively). The three BDTs are binary classifiers whose goals are to identify their target processes against all the other processes. The final set of input variables used in the three BDTs is summarised in table 5.2, where BDT(*tHq*), BDT($\mathrm{t\bar{t}}$) and BDT($\mathrm{t\bar{t}}W$) use 22, 18 and 30 variables, respectively. That means 42 different variables among the three BDTs.

TABLE 5.2: List of variables considered in the training of the BDTs in the 3ℓ channel. The $\times$ symbol marks in which BDTs the variable is used.

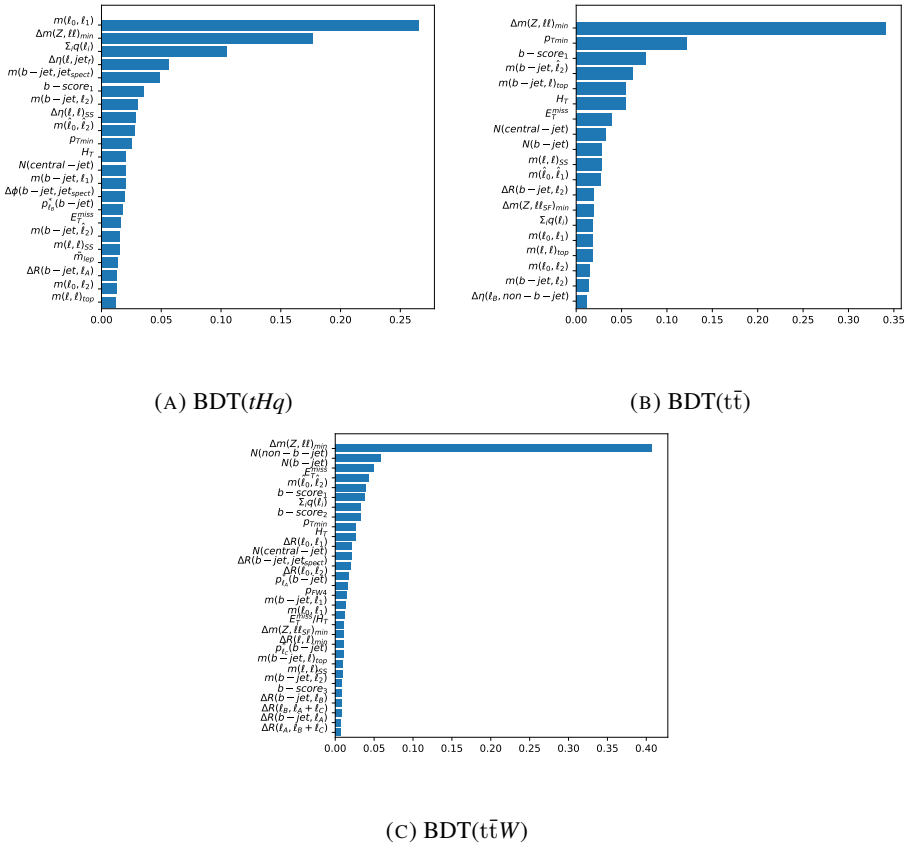
Variable name	BDT(*tHq*)	BDT($\mathrm{t\bar{t}}$)	BDT($\mathrm{t\bar{t}}W$)	Description
$\mathrm{p}^{*}_{\ell_\mathrm{C}}$(b-jet)	-	-	$\times$	Momentum of lepton with respect to the
$\mathrm{p}^{*}_{\ell_\mathrm{B}}$(b-jet)	$\times$	-	-	leading b-jet rest frame.
$\mathrm{p}^{*}_{\ell_\mathrm{A}}$(b-jet)	-	-	$\times$	
$\Delta\mathrm{R}(\ell_\mathrm{A}, \ell_\mathrm{B}+\ell_\mathrm{C})$	-	-	$\times$	ΔR between lepton and the system of the other
$\Delta\mathrm{R}(\ell_\mathrm{B}, \ell_\mathrm{A}+\ell_\mathrm{C})$	-	-	$\times$	two leptons.
$\Delta\mathrm{R}(\ell_\mathrm{A}, \mathrm{b\text{-jet}})$	$\times$		$\times$	ΔR between lepton and leading b-jet.
$\Delta\mathrm{R}(\ell_\mathrm{B}, \mathrm{b\text{-jet}})$	-	-	$\times$	
$\mathrm{m}(\ell,\ell)$	$\times$	$\times$	$\times$	Invariant mass and $\Delta\eta$ of same-sign leptons.
$\Delta\eta(\ell,\ell)$	$\times$	-	-	
$\mathrm{m}(\ell_0,\ell_2)$	$\times$	$\times$	-	Invariant mass and ΔR between leptons and/or
$\mathrm{m}(\ell_0,\ell_1)$	$\times$	$\times$	$\times$	leading b-jet. The lepton ordering is charge
$\mathrm{m}(\ell_1, \mathrm{b\text{-jet}})$	$\times$	-	$\times$	and ΔR based.
$\mathrm{m}(\ell_2, \mathrm{b\text{-jet}})$	$\times$	$\times$	-	
$\Delta\mathrm{R}(\ell_0,\ell_1)$	-	-	$\times$	
$\mathrm{m}(\hat{\ell}_0,\hat{\ell}_1)$	-	$\times$	-	Invariant mass between $\hat{\ell}_0$ and $\hat{\ell}_1$.
$\mathrm{m}(\hat{\ell}_0,\hat{\ell}_2)$	$\times$	-	$\times$	Invariant mass and ΔR between $\hat{\ell}_0$ and $\hat{\ell}_2$.
$\Delta\mathrm{R}(\hat{\ell}_0,\hat{\ell}_2)$	-	-	$\times$	
$\mathrm{m}(\hat{\ell}_2, \mathrm{b\text{-jet}})$	$\times$	-	$\times$	Invariant mass and ΔR between leading b-jet
$\Delta\mathrm{R}(\hat{\ell}_2, \mathrm{b\text{-jet}})$	-	$\times$	-	and $\hat{\ell}_1$ and $\hat{\ell}_2$.
$\mathrm{m}(\ell, \mathrm{b\text{-jet})_{top}}$	-	$\times$	$\times$	Invariant mass between lepton and b-jet giving best top-quark visible mass.

TABLE 5.2: List of variables included in the training of the BDTs in the 3ℓ channel. The $\times$ symbol marks in which BDT the variable is used.

Variable name	BDT(*tHq*)	BDT($t\bar{t}$)	BDT($t\bar{t}W$)	Description
$m(\ell,\ell)_{\text{top}}$	$\times$	$\times$	-	Invariant mass between two leptons giving best top-quark visible mass.
p_{FW4}	-	-	$\times$	Fourth Fox–Wolfram moment [151]
$p_{T,\text{min}}$	$\times$	$\times$	$\times$	p_T of the softest lepton.
E_T^{miss}/H_T	-	-	$\times$	Ratio between E_T^{miss} and H_T.
N(non-b-jet)	-	-	$\times$	Number of non-b-tagged jets.
N(central-jet)	$\times$	$\times$	$\times$	Number of central jets.
N(b-jet)	-	$\times$	$\times$	Number of b-tagged jets.
$\Delta m(Z,\ell\ell_{SF})_{\text{min}}$	-	$\times$	$\times$	Minimum difference between the reconstructed invariant mass of two leptons with the same flavour and Z boson mass.
$\Delta m(Z,\ell\ell)_{\text{min}}$	$\times$	$\times$	$\times$	Minimum difference between the reconstructed invariant mass of two leptons and Z boson mass.
E_T^{miss}	$\times$	$\times$	$\times$	E_T^{miss}.
$m(\text{b-jet},\text{jet}_{\text{spect}})$	$\times$	-	-	Invariant mass, $\Delta\phi$ and ΔR between leading b-jet and spectator jet.
$\Delta\phi(\text{b-jet},\text{jet}_{\text{spect}})$	$\times$	-	-	
$\Delta R(\text{b-jet},\text{jet}_{\text{spect}})$	-	-	$\times$	
$\Sigma_i q(\ell_i)$	$\times$	$\times$	$\times$	Sum of lepton charges.
H_T	$\times$	$\times$	$\times$	H_T.
$\tilde{m}_{\text{lep}}$	$\times$	-	-	$\sqrt{\Sigma_i(E_{\ell_i}^2+p_{T,\ell_i}^2)}$.
$\Delta R(\ell,\ell)_{\text{min}}$	-	-	$\times$	Minimum ΔR between two leptons.
$\Delta\eta(\ell_B,\text{non-b-jet})$	-	$\times$	-	$\Delta\eta$ between ℓ_B and closest non-b-tagged jets.
$\Delta\eta(\ell,\text{jet}_f)$	$\times$	-	-	$\Delta\eta$ between leading forward jet and closest lepton.
b-score_1	$\times$	$\times$	$\times$	Binned DL1r score (calibrated) of leading b-jet
b-score_2	-	-	$\times$	Binned DL1r score (calibrated) of second b-jet
b-score_3	-	-	$\times$	Binned DL1r score (calibrated) of third b-jet

The variables are ranked based on their *Gain* value, where higher rankings corresponds to higher *Gain* values. The relevant ranking plots for the three BDTs are shown

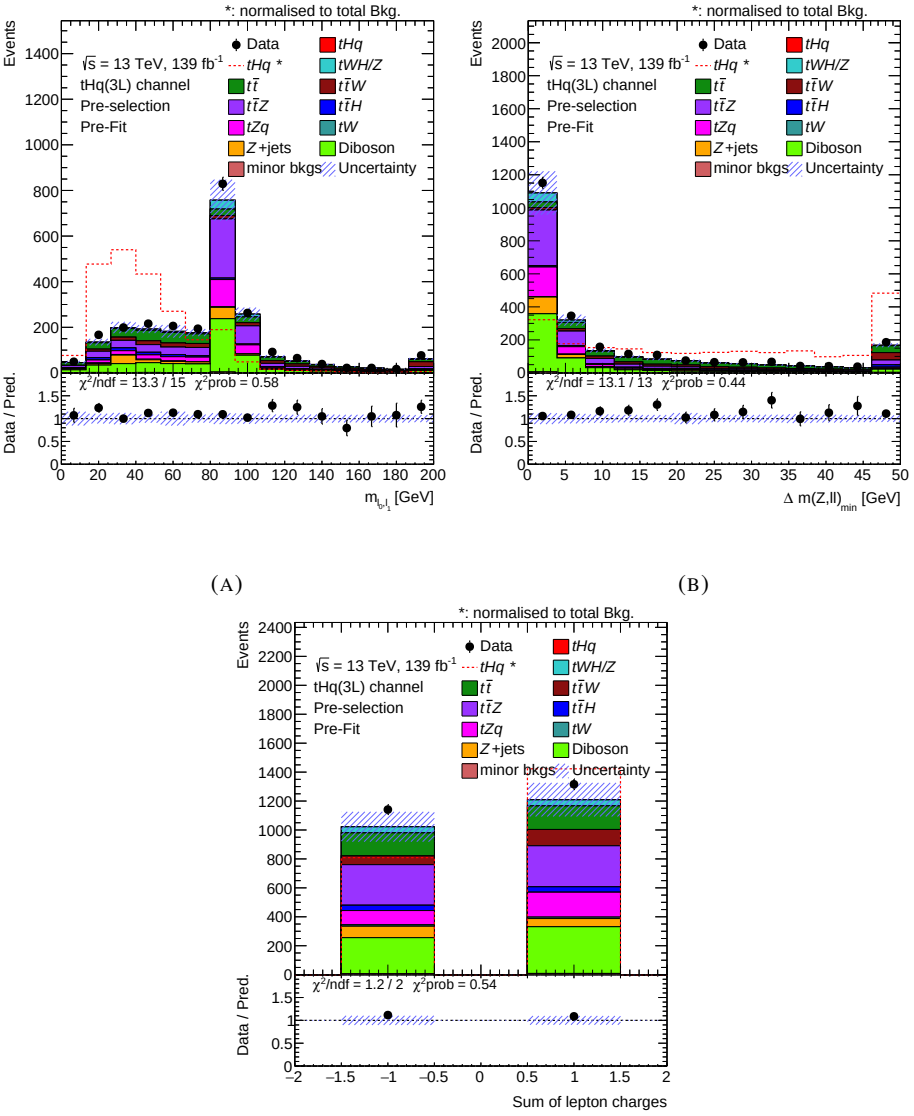
in figure 5.1. The distributions of the three input variables with the highest *Gain* values are shown in figures 5.2-5.4 for the three trained BDTs. The dashed lines on these figures represent the target process normalised to the total background for each BDT, that means all the processes except the target process. These distributions show a high separation power for these variables what could explain their positions on their rankings. Despite these large separation powers a MVA is still needed.



(A) BDT(*tHq*) (B) BDT($t\bar{t}$)

(C) BDT($t\bar{t}W$)

FIGURE 5.1: Ranking distributions for the three BDTs of the 3ℓ channel for (A) BDT(*tHq*), (B) BDT($t\bar{t}$) and (C) BDT($t\bar{t}W$) value. The x-axis corresponds to the value given by the *Gain* value.



(A) (B)

(C)

FIGURE 5.2: The three input variables with the highest *Gain* values in the 3ℓ channel for BDT(*tHq*). The distributions show data and simulation samples in the pre-selection region for (A) $\mathrm{m}_{\ell_0,\ell_1}$, (B) $\Delta\mathrm{m}(\mathrm{Z},\ell\ell)_{\mathrm{min}}$ and (C) sum of lepton charges. The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated data events. Moreover, the χ^2 over the number degree of freedom (ndf) and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.

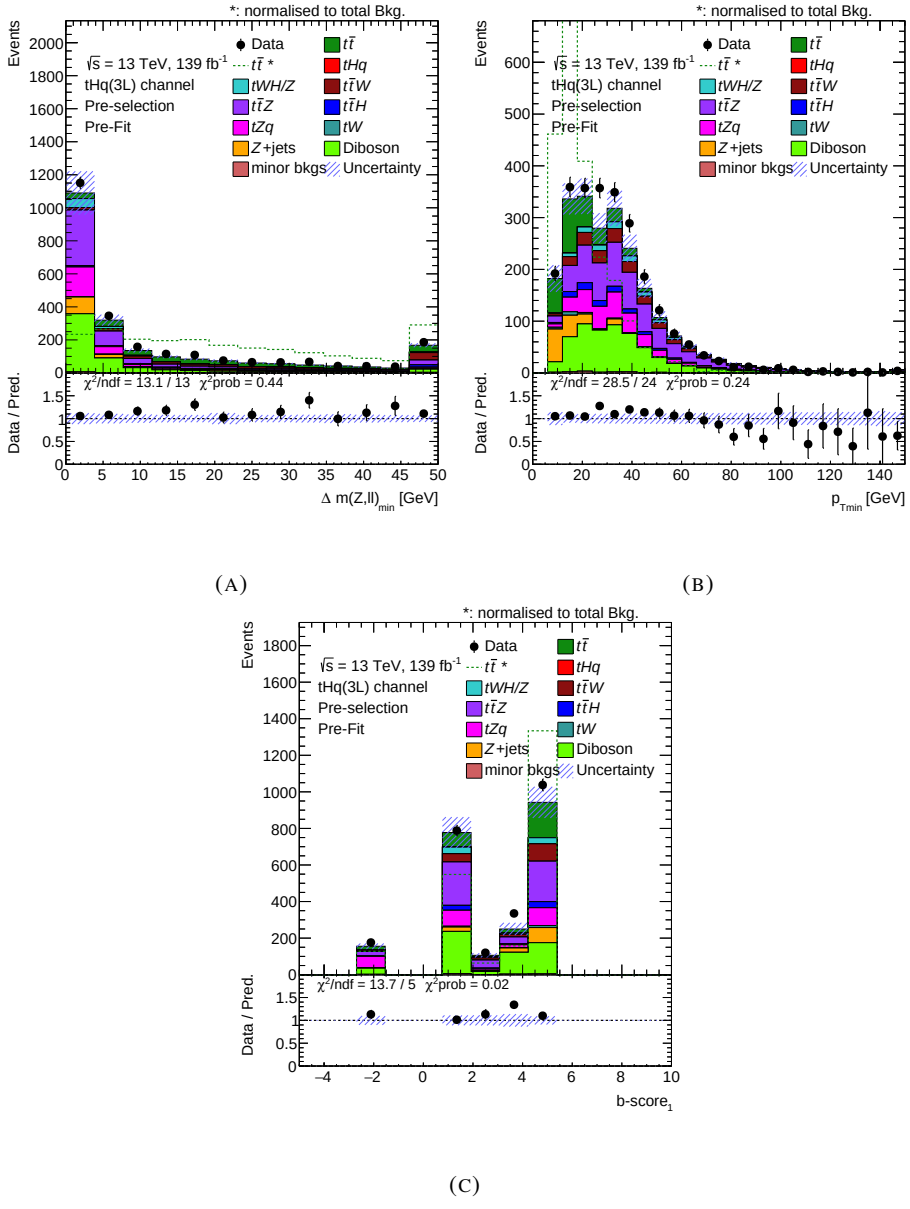


FIGURE 5.3: The three input variables with the highest *Gain* values in the 3ℓ channel for BDT($t\bar{t}$). The distributions show data and simulation samples in the pre-selection region for (A) $\Delta m(Z,\ell\ell)_{min}$, (B) $p_{T min}$ and (C) $b-score_1$. The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.

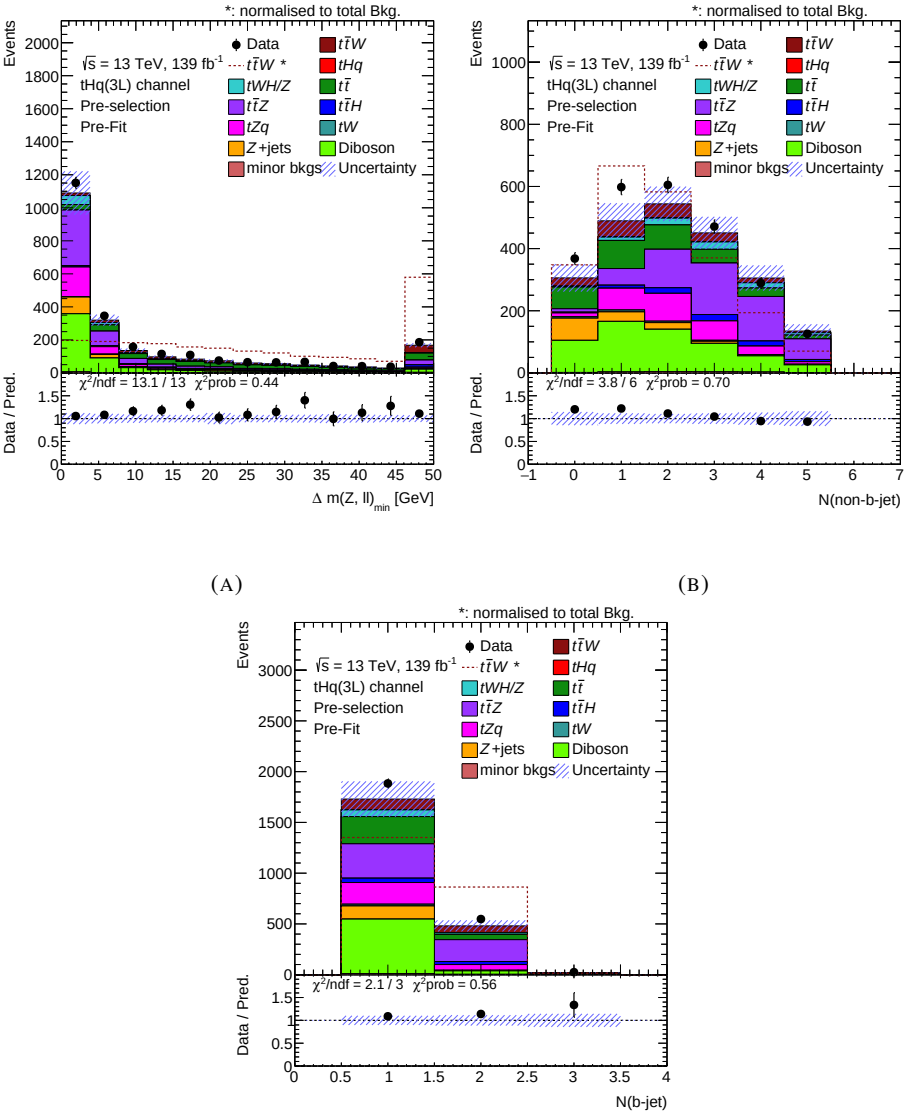


FIGURE 5.4: The three input variables with the highest *Gain* values in the 3ℓ channel for BDT($t\bar{t}W$). The distributions show data and simulation samples in the pre-selection region for (A) $\Delta m(Z,\ell\ell)_{min}$, (B) $N(non-b-jet)$ and (C) $N(b-jet)$. The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.

5.1.2.2 Optimisation of the BDT parameters and obtained performance

Once the lists of input variables are defined the BDT parameters are optimised using the Genetic Algorithm (GA)[2] method [152]. These parameters characterise the BDT architecture and influence and determine its performance. After the optimisation of the BDT parameters, the BDTs are trained and evaluated to get the score of each BDT. The resulting BDT scores are shown in figure 5.5. The comparison between data and prediction for the same distributions shows a good agreement within the total uncertainty. Overall, a good separation between the target process and the total background is achieved, as shown in figure 5.6.

The stability of the results of the BDTs is evaluated using the k-fold cross-validation method[3]. The k-fold cross-validation method allows to verify that the training performed over a sub-sample of events is representative of the full sample and identify the possible presence of overtraining. In this analysis, the k-fold cross-validation method consists of five steps: in each step $80\,\%$ of the simulation events are used as training while the remaining $20\,\%$ is used as test sample.

The performance of the trained BDTs is measured using three figures of merit: the receiver operating characteristic (ROC) curve, the area under the ROC curve (AUC_ROC), and the logarithmic loss function (log_loss). The ROC curve of a BDT is defined as the true positive rate as a function of the false positive rate. The ROC curve quantifies both the separation power and the accuracy of a BDT. The AUC_ROC quantifies the separation power of the BDT (a perfect separation corresponds to AUC_ROC=1) while the log_loss estimates the correctness of the labelling of the target or non-target processes (a perfect labelling corresponds to log_loss=0). Since the k-fold cross-validation method uses five steps, the figures of merit are evaluated five times, once for each fold. The five pairs of ROC curves[4] resulting from the k-fold cross-validation steps over the three BDTs are shown in figure 5.7. The five values of the AUC_ROC and the log_loss values and their average of the BDTs are reported in tables 5.3 and 5.4, respectively. All the ROC curves show a similar behaviour, as can be inferred from figure 5.7, and from the

[2] More details about the GA method are given in appendix B.2.
[3] More details on the cross-validation method are given in appendix B.3.
[4] Each pair corresponds to a fold where test and train ROC curves are included.

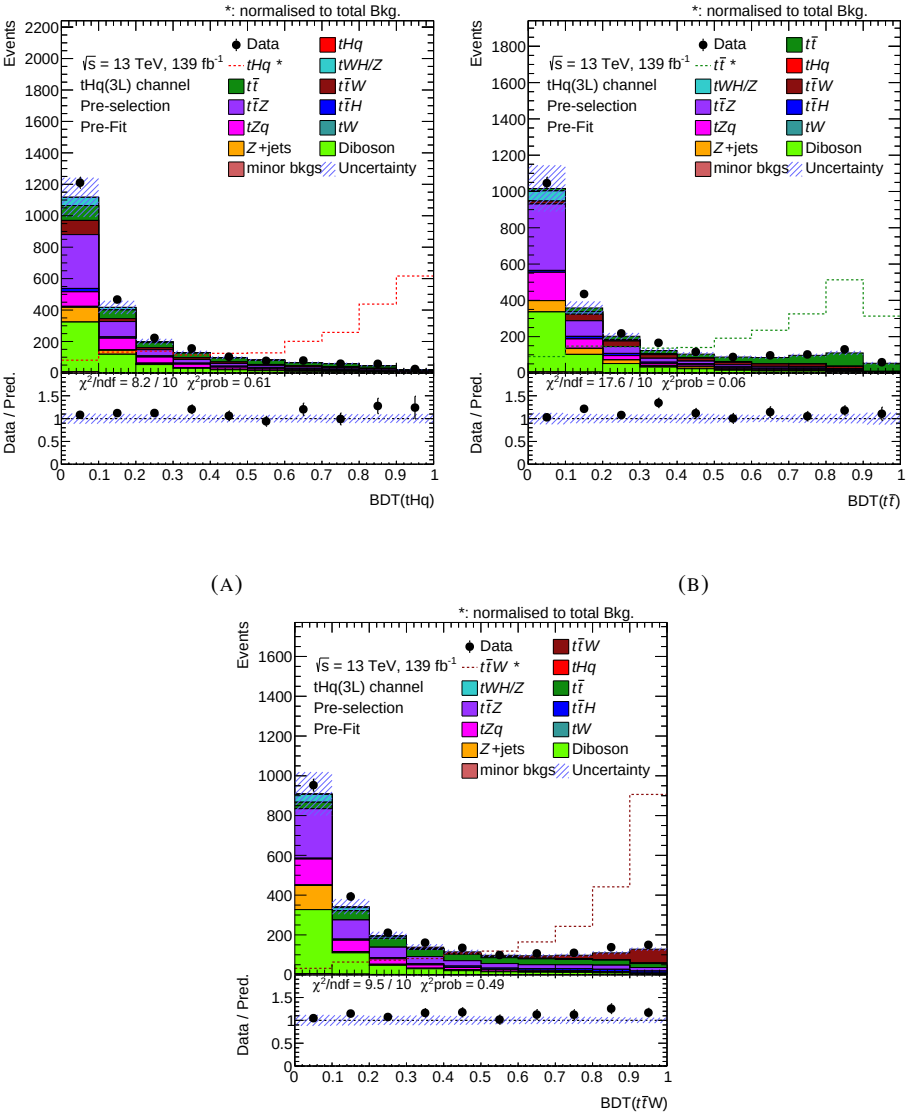


FIGURE 5.5: Distributions of the BDT scores for (A) BDT(*tHq*), (B) BDT($t\bar{t}$) and (C) BDT($t\bar{t}W$) in the 3ℓ channel. The dashed line represents the target process of each BDT. The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.

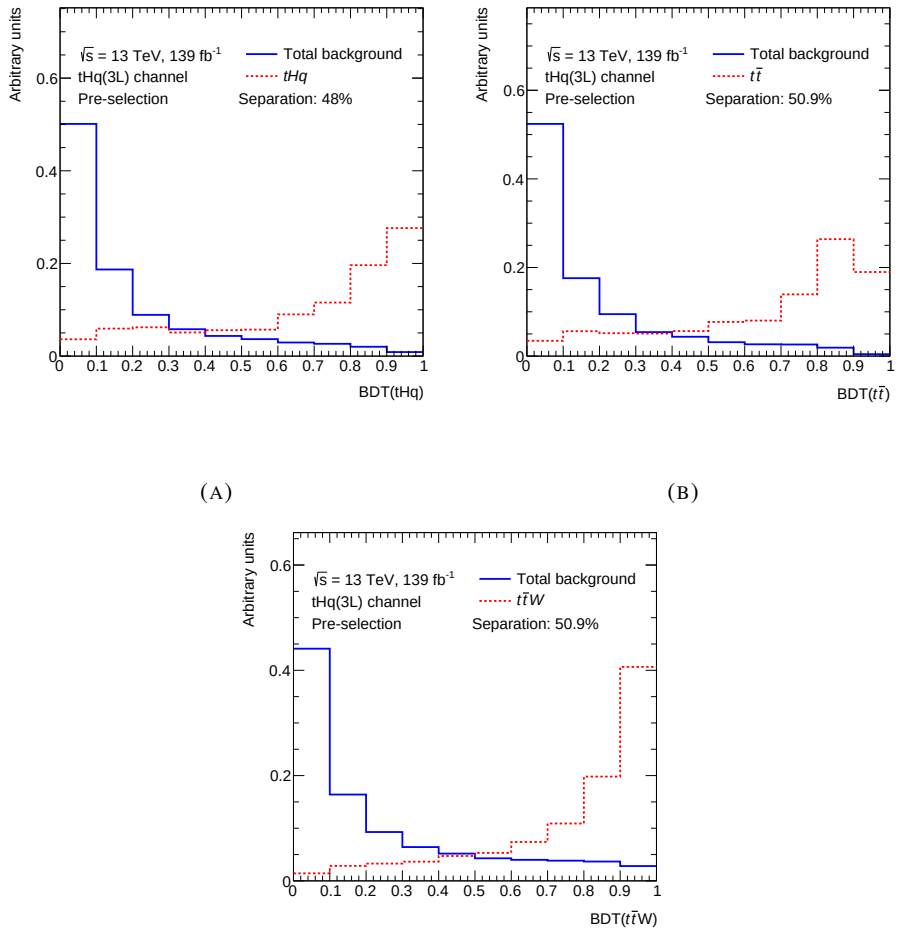
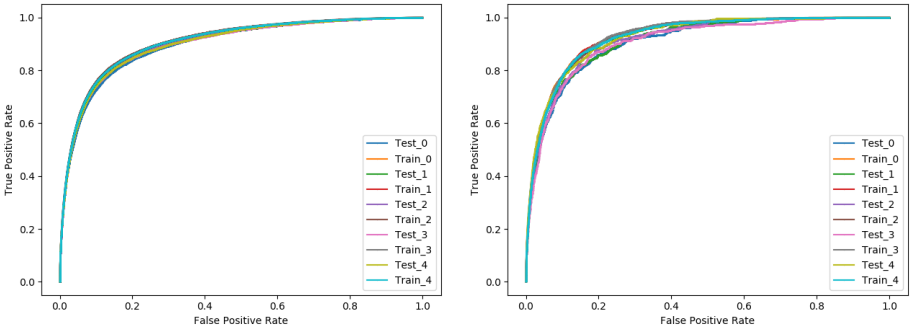
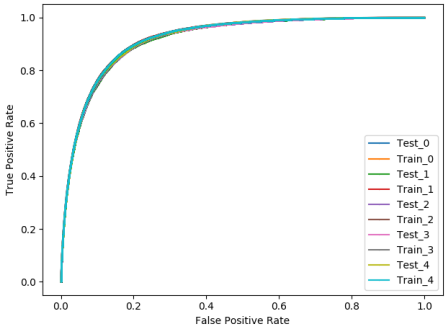


FIGURE 5.6: Normalised distributions of the BDT score for (A) BDT(tHq), (B) BDT($\mathrm{t\bar{t}}$) and (C) BDT($\mathrm{t\bar{t}}W$) in the 3ℓ channel. The target process (red dashed line) and the background samples (blue solid line) are normalised to the same area. In each case the background sample is defined as any sample which is not target. The separation is computed using the formula 1 in Ref. [153].

tables 5.3. Moreover, the values on the tables are compatible within its uncertainties for each BDT. Therefore, the results of the k-fold cross-validation method show that the BDTs are robust and that each training is representative of the full sample.



(A) BDT(*tHq*) (B) BDT($t\bar{t}$)



(C) BDT($t\bar{t}W$)

FIGURE 5.7: ROC curves for each BDT in the 3ℓ channel: (A) BDT(*tHq*), (B) BDT($t\bar{t}$) and (C) BDT($t\bar{t}W$). Five pairs of ROC curves, i.e. for test and train samples are shown where each one corresponds to one fold X, where X $\in [0, 4]$, which represents the number of the fold.

5.2 Event selection for the 2ℓSS final state

Similarly to the previous section for the 3ℓ channel, this section covers the selection strategy for the 2ℓSS channel, and the requirements which define the pre-selection region

TABLE 5.3: Values of the ROC_AUC and the log_loss given by each BDT in the 3ℓ channel. The value of the ROC_AUC for each fold X is shown in each row. In the last row, the mean values and their statistical uncertainties are given for each fold.

	BDT(*tHq*)	BDT($t\bar{t}$)	BDT($t\bar{t}W$)
ROC_AUC_fold_1	0.811	0.81	0.848
ROC_AUC_fold_2	0.819	0.82	0.842
ROC_AUC_fold_3	0.820	0.81	0.842
ROC_AUC_fold_4	0.818	0.81	0.846
ROC_AUC_fold_5	0.817	0.83	0.844
ROC_AUC_average	0.817 ± 0.003	0.81 ± 0.01	0.844 ± 0.002

TABLE 5.4: Values of log_loss given by each BDT in the 3ℓ channel. The value of the log_loss for each fold X is shown in each row. In the last row, the mean value and its statistical uncertainty are given for each fold.

	BDT(*tHq*)	BDT($t\bar{t}$)	BDT($t\bar{t}W$)
log_loss_fold_1	0.267	0.223	0.354
log_loss_fold_2	0.268	0.225	0.354
log_loss_fold_3	0.264	0.225	0.356
log_loss_fold_4	0.267	0.222	0.357
log_loss_fold_5	0.266	0.228	0.353
log_loss_average	0.266 ± 0.002	0.225 ± 0.002	0.354 ± 0.001

in the 2ℓSS channel. Moreover, the details of the MVA algorithm used in this particular case are also discussed in this section.

5.2.1 Pre-selection requirements

Pre-selection criteria are requested before applying the MVA technique to the 2ℓSS final state. The pre-selection region is defined as follows:

- Exactly two light-flavour leptons (electrons and muons, see 4.3).
- The sum of the charge of the leptons must be $\pm$ 2.
- The two leptons are ordered by their $\mathrm{p_T}$, with the leading lepton having $\mathrm{p_T} > 27$ GeV, and the sub-leading lepton having $\mathrm{p_T} > 20$ GeV.
- The events including hadronically decays of taus, as defined in section 4.4, are vetoed.
- The number of jets is required to be between one and six, as defined in section 4.5.
- The number of b-jet is required to be between one and three, as defined in section 4.5.1.
- The $\mathrm{E_{Tmiss}}$ is required to be between 5 GeV and 800 GeV.

The goal of this set of requirements is to maximise the signal acceptance while minimising the background contamination. Table 5.5 shows the event yields of the different processes contributing to the pre-selection region as predicted by the MC simulation together with data events.

The minor backgrounds composition is done regarding the negligible MC samples in the Signal Region (defined in table 5.11). The MC samples inside minor backgrounds are: *tWZ*, *Z+jets*, triboson, tW, single top-quark s-channel, *W+jets*, *tttt*, *ttt*, ggF, VBF and VH.

TABLE 5.5: The 2ℓSS channel pre-selection region yields as predicted by the MC simulation and data events. The uncertainties include statistical and all the systematic sources.

Process	Yields
tHq	9.96 ± 0.34
tWH	5.24 ± 0.52
$t\bar{t}$	1420 ± 108
$t\bar{t}W$	726 ± 27
$t\bar{t}Z$	164 ± 38
$t\bar{t}H$	126 ± 21
tZq	88 ± 11
Diboson	295 ± 74
Single top-quark t-channel	44 ± 19
Minor backgrounds	678 ± 344
Total background	3546 ± 384
Data	3841

5.2.2 Multivariate analysis

The results presented in the 2ℓSS channel depend on the discrimination power provided by MVA techniques using a set of input variables. A set of particular input variables is selected for the 2ℓSS channel. The response of the MVA algorithm is used to define several enriched regions for the signal process and the main background processes. The strategy followed in the 2ℓSS channel is similar to the one followed in the 3ℓ channel. The *XGBoost* python library is also used to develop several independent binary-classifier BDTs. For the 2ℓSS channel, there are four BDTs: one targets the *tHq* signal process, and three target the $t\bar{t}$, the $t\bar{t}W$ and the diboson background processes. The negative-weighted simulated events also appear in a similar percentage for *tHq* simulated events to the 3ℓ channel. Thus, the same solution as the one for the 3ℓ channel is followed for the 2ℓSS channel. The list of input variables used in each BDT is optimised as described in section 5.2.2.1. The optimisation of the BDT parameters together with the final performance and resulting scores are reported in section 5.2.2.2.

5.2.2.1 Input variables optimisation and their importance

The set of discriminant variables used as input for each BDT training are described in this section. Using the aforementioned lepton and jet definition several variables are constructed similarly to the 3ℓ channel.

The list of variables used in each BDT is optimised using an iterative approach based on the impact they have in the BDT performance, parametrised by the *Gain* value and on the correlations with other variables. As already mentioned, the details about this process are given in appendix B.1.

The final set of input variables used in the four BDTs is summarised in table 5.6. The four discriminant BDTs are: BDT(*tHq*), which targets the signal process, and BDT($t\bar{t}$), BDT($t\bar{t}W$) and BDT(diboson), which target the $t\bar{t}$, the $t\bar{t}W$ and the diboson background processes, respectively. The BDT(*tHq*), BDT($t\bar{t}$), BDT($t\bar{t}W$) and BDT(diboson) use 28, 19, 29 and 27 variables, respectively, what means a total number of 46 variables among the four BDTs.

TABLE 5.6: List of variables included in the training of the BDTs for 2ℓSS channel. The $\times$ indicates in which BDTs the variable is used.

Variable name	BDT(*tHq*)	BDT($t\bar{t}$)	BDT($t\bar{t}W$)	BDT(diboson)	Description
$p_{T,min}$	$\times$	$\times$	$\times$	-	p_T of softest lepton.
E_T^{miss}/H_T	-	-	-	$\times$	Ratio between E_T^{miss} and H_T.
$\eta(\text{no-b-jet})_{max}$	-	-	$\times$	-	Maximum η value among non-b-tagged jets.
E_T^{miss}	$\times$	$\times$	$\times$	$\times$	E_T^{miss}.
$\Sigma_i q(\ell_i)$	$\times$	$\times$	$\times$	$\times$	Sum of lepton charges.
$\tilde{m}_{lep}$	$\times$	$\times$	$\times$	$\times$	$\sqrt{\Sigma_i(E_{\ell_i}^2 + p_{T,\ell_i}^2)}$.
$m_H(WW)$	$\times$	$\times$	$\times$	-	Mass of a candidate Higgs Boson when decays to a pair of W bosons.
$\Delta m(Z,\ell\ell)$	-	$\times$	$\times$	$\times$	Difference between the invariant mass of $\ell\ell$ and the mass of the Z boson.

TABLE 5.6: List of variables included in the training of the BDTs for 2ℓSS channel. The $\times$ indicates in which BDTs the variable is used.

Variable name	BDT(*tHq*)	BDT($t\bar{t}$)	BDT($t\bar{t}W$)	BDT(diboson)	Description
N(non-b-jet)	$\times$	-	-	$\times$	Number of non b-tagged jets.
N(central-jet)	$\times$	$\times$	$\times$	$\times$	Number of central jets.
N(b-jet)	-	$\times$	$\times$	$\times$	Number of b-tagged jets.
$m(jet_{spect})$	$\times$	-	-	-	Mass of the spectator jet.
$E(jet_{spect})$	$\times$	-	-	-	Energy of the spectator jet.
$p_T(jet_{spect})$	$\times$	-	-	-	p_T of the spectator jet.
$\chi^2(m(\ell_i,\ell_j)_{top})_{min}$	-	$\times$	$\times$	$\times$	Minimum χ^2 value between the masses of top candidates given by the combination of leptons.
$m(\ell,\ell)_{top}$	-	-	-	$\times$	Invariant mass between two leptons giving best top visible mass.
H_T	$\times$	-	-	$\times$	H_T.
$H_T(\ell\ell)$	$\times$	$\times$	$\times$	-	Sum of the p_T of the two leptons.
H_T(jets)	-	$\times$	$\times$	-	Sum of the p_T of the jets.
$m(jet,jet)_W$	$\times$	-	$\times$	$\times$	Mass of the best two jet candidates for the W boson.
$m(jet,jet)_{top}$	$\times$	$\times$	$\times$	$\times$	Mass of the best top quark candidate from jets.
$m(\ell_2,jet_1)$	$\times$	-	-	$\times$	Invariant mass of ℓ_2 and jet_1.
$m(\ell_2,jet_2)$	$\times$	$\times$	$\times$	$\times$	Invariant mass of ℓ_2 and jet_2.

TABLE 5.6: List of variables included in the training of the BDTs for 2ℓSS channel. The $\times$ indicates in which BDTs the variable is used.

Variable name	BDT(*tHq*)	BDT($t\bar{t}$)	BDT($t\bar{t}W$)	BDT(diboson)	Description
$m(\ell_1+\ell_2, \text{jet}_1)$	-	$\times$	-	-	Invariant mass of both lepton and the jet_1.
$m(\ell_1+\ell_2, \text{jet}_1+\text{jet}_2)$	-	-	$\times$	-	Invariant mass of both leptons, the leading jet and the second jet.
$m(\ell_1+\ell_2)$	$\times$	-	-	-	Invariant mass of both leptons.
$m(\text{jet}_1+\text{jet}_2)$	$\times$	-	-	$\times$	Invariant mass of the leading jet and the second jet.
p_{FW1}	-	-	$\times$	-	First Fox–Wolfram moment [151].
p_{FW4}	-	$\times$	-	$\times$	Fourth Fox–Wolfram moment [151].
Lepton flavours	-	$\times$	-	-	Identification of the flavour of the leptons.
eμ-events	-	$\times$	-	$\times$	Identification of the events eμ.
ee-events	$\times$	-	$\times$	$\times$	Identification of the events ee.
$\Delta R(\ell_1, \text{b-jet})$	-	$\times$	-	$\times$	ΔR and $\Delta\eta$ between
$\Delta\eta(\ell_1, \text{b-jet})$	-	$\times$	-	$\times$	ℓ_1 and its closest b-jet.
$\Delta R(\ell_2, \text{b-jet})$	-	$\times$	-	$\times$	ΔR and $\Delta\eta$ between ℓ_2
$\Delta\eta(\ell_2, \text{b-jet})$	-	$\times$	-	$\times$	and the its closest b-jet.
$m(\text{b-jet}, \text{jet}_{\text{spect}})$	$\times$	-	$\times$	-	Invariant mass, ΔR, $\Delta\eta$
$\Delta R(\text{b-jet}, \text{jet}_{\text{spect}})$	$\times$	-	$\times$	-	between the spectator jet
$\Delta\eta(\text{b-jet}, \text{jet}_{\text{spect}})$	-	$\times$	-	-	and the leading b-jet.
$\Delta R(\ell, \text{jet}_{\text{f}})$	-	$\times$	-	-	ΔR and $\Delta\eta$ between the leading forward

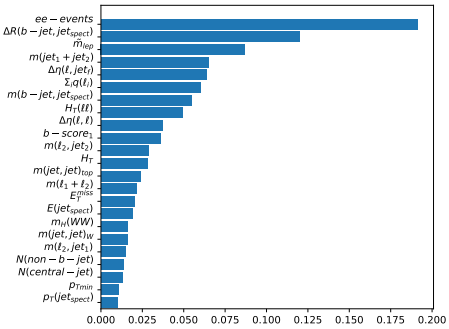
TABLE 5.6: List of variables included in the training of the BDTs for 2ℓSS channel. The $\times$ indicates in which BDTs the variable is used.

Variable name	BDT(*tHq*)	BDT($t\bar{t}$)	BDT($t\bar{t}W$)	BDT(diboson)	Description
$\Delta\eta(\ell, \text{jet}_f)$	$\times$	-	-	-	jet and its closest lepton.
$\Delta\phi(\ell,\ell)$	-	-	$\times$	-	$\Delta\phi$, ΔR and $\Delta\eta$.
$\Delta R(\ell,\ell)$	-	$\times$	$\times$	-	between the two leptons.
$\Delta\eta(\ell,\ell)$	$\times$	-	-	-	
b-score$_1$	$\times$	$\times$	$\times$	$\times$	b-tagging (from `DL1r` algorithm) score of leading b-jet
b-score$_2$	-	$\times$	$\times$	$\times$	b-tagging (from `DL1r` algorithm) score of second b-jet
b-score$_3$	-	$\times$	$\times$	$\times$	b-tagging (from `DL1r` algorithm) score of third b-jet

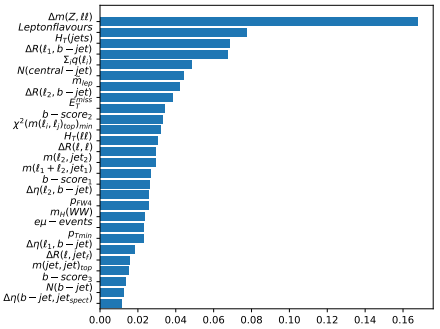
Variables are ranked based on their *Gain* values, with higher rankings corresponding to higher *Gain* values. The relevant ranking distributions for the four BDTs are shown in figure 5.8. The three input variables with the highest *Gain* values are shown in figures 5.9-5.12 for the four BDTs, to understand why these variables are highly ranked.

5.2.2.2 Optimisation of the BDT parameters and obtained performance

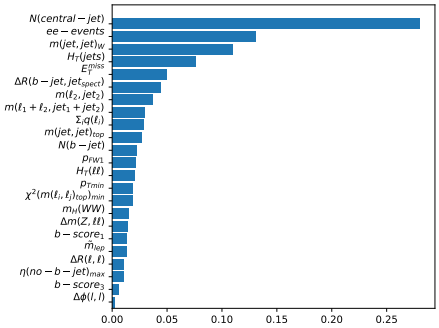
Once the list of input variables is defined, the BDT parameters are optimised using the GA method, as mentioned in section 5.1.2.1 (more details in appendix B.2). After the optimisation process, the BDTs are trained and evaluate to obtain their predictions like the probability/score of an event to be the target process. The resulting BDT scores are shown in figure 5.13 and good agreement between data and simulation can be observed. Overall a good separation between the target process and the total background is achieved, as it is shown in figure 5.14.



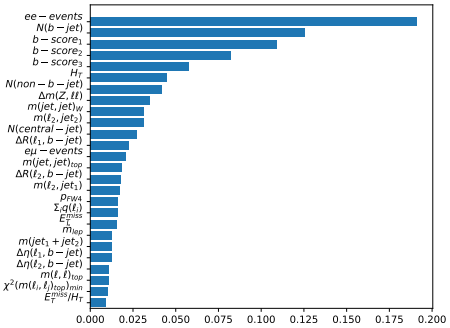
(A) BDT(tHq)



(B) BDT($t\bar{t}$)



(C) BDT($t\bar{t}W$)



(D) BDT(diboson)

FIGURE 5.8: Ranking distribution of variables for each BDT for the 2ℓSS channel for (A) BDT(tHq), (B) BDT($t\bar{t}$), (C) BDT($t\bar{t}W$) and (D) BDT(diboson). The x-axis corresponds to the value given by the *Gain* value.

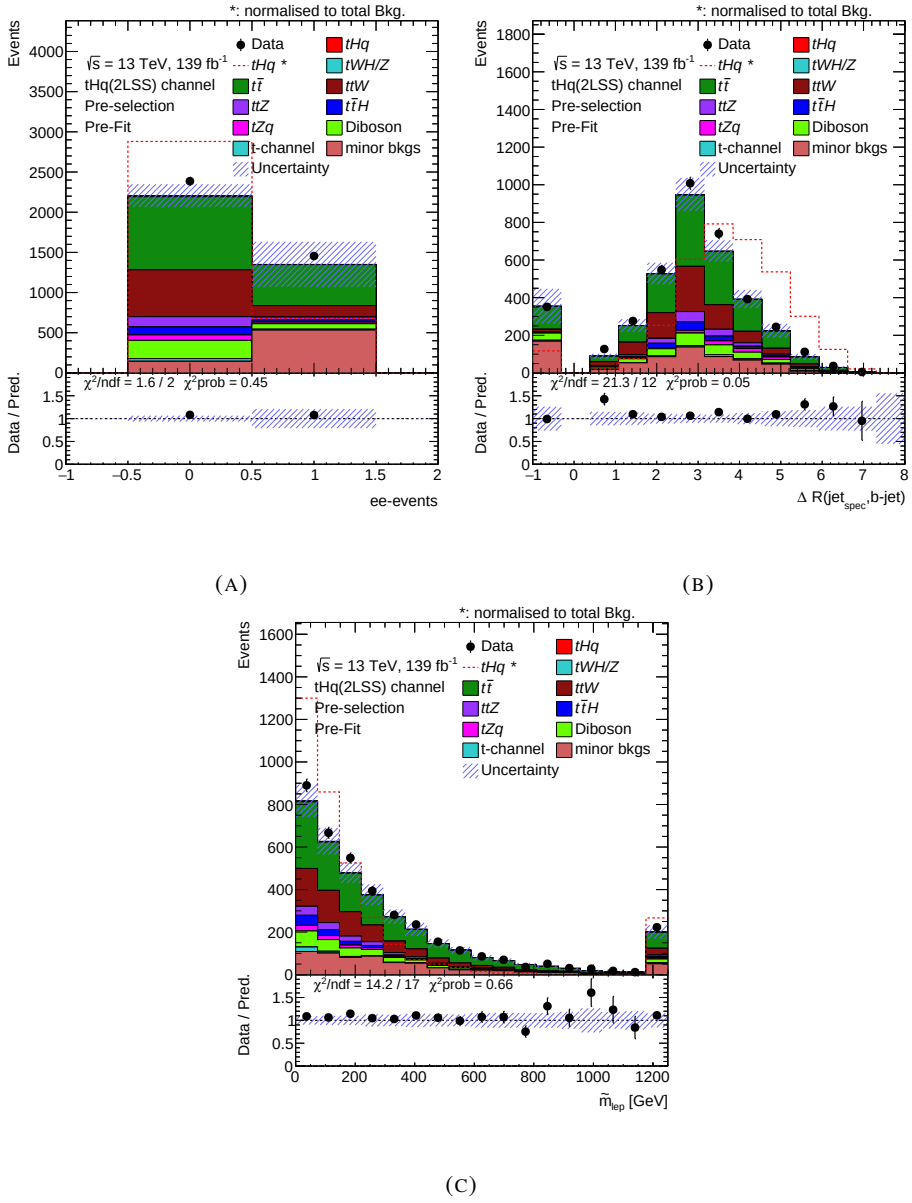
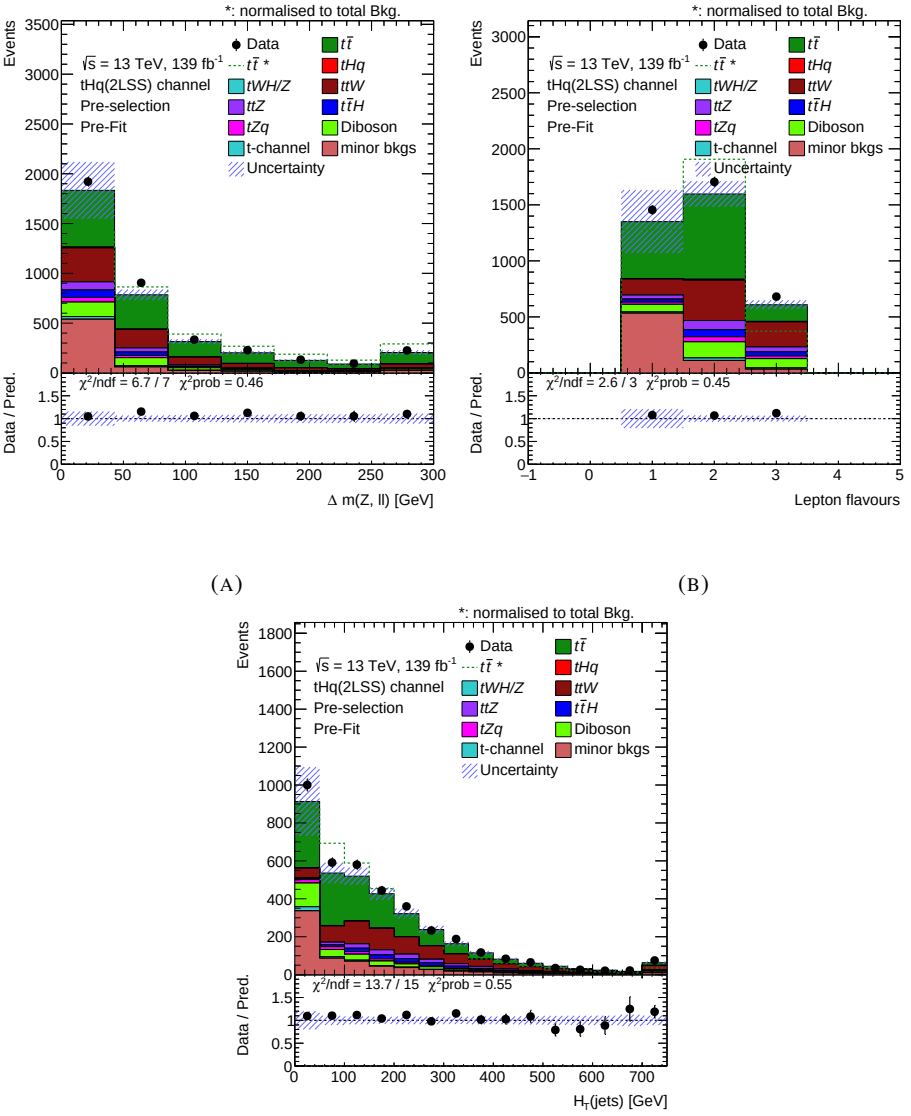


FIGURE 5.9: The three input variables with the highest *Gain* values in the 2ℓSS channel for BDT (*tHq*). The distributions show data and simulation samples in the pre-selection region for (A) $\mathrm{ee-events}$, (B) $\Delta\mathrm{R}(\mathrm{jet_{spect}}, \mathrm{b-jet})$ and (C) $\tilde{\mathrm{m}}_{\mathrm{lep}}$. The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated background data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.



(A) (B)

(C)

FIGURE 5.10: The three input variables with the highest *Gain* value in the 2ℓSS channel for BDT ($t\bar{t}$). The distributions show data and simulation samples in the pre-selection region for (A) $\Delta m(Z, \ell\ell)$, (B) Lepton flavours and (C) $H_T(\text{jets})$. The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated background data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.

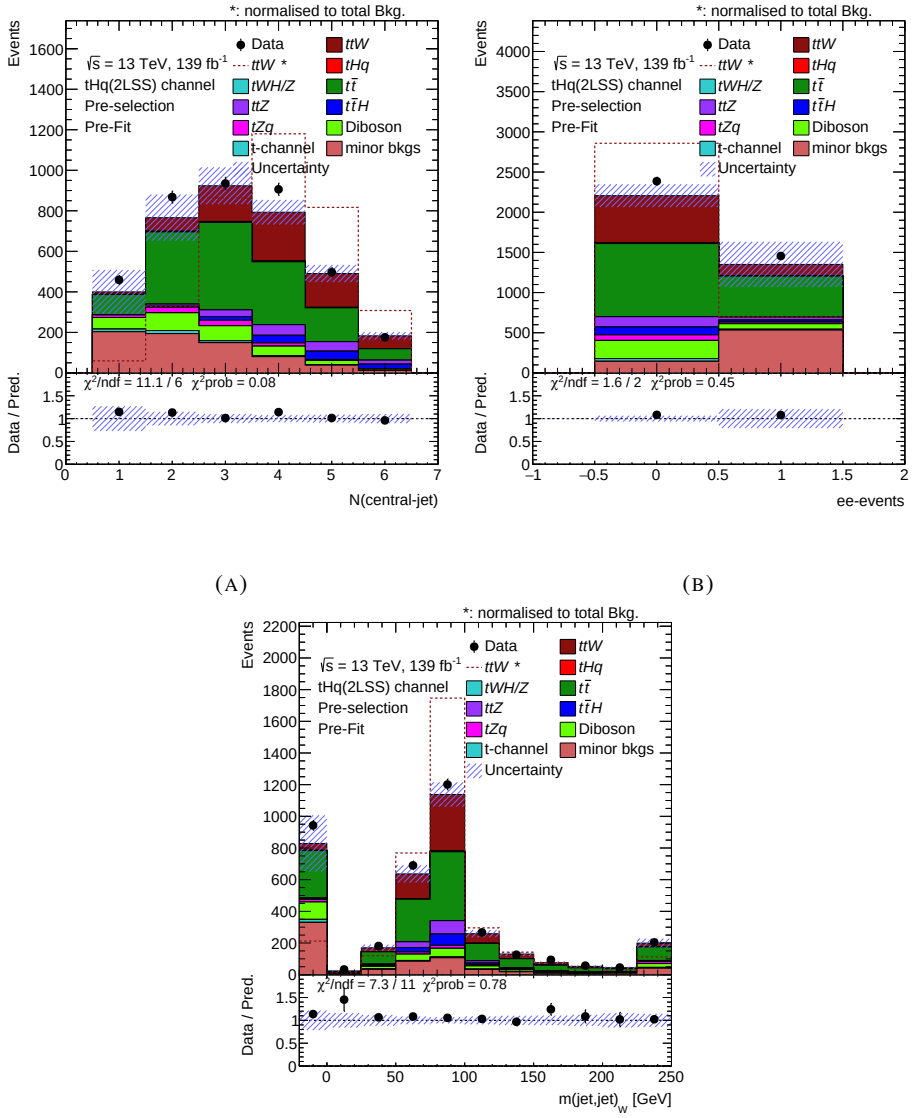


FIGURE 5.11: The three input variables with the highest *Gain* value in the 2ℓSS channel for BDT ($t\bar{t}W$). The distributions show data and simulation samples in the pre-selection region for (A) $\mathrm{N(central-jet)}$, (B) $\mathrm{ee-events}$ and (C) $\mathrm{m(jet,jet)_W}$. The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated background data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.

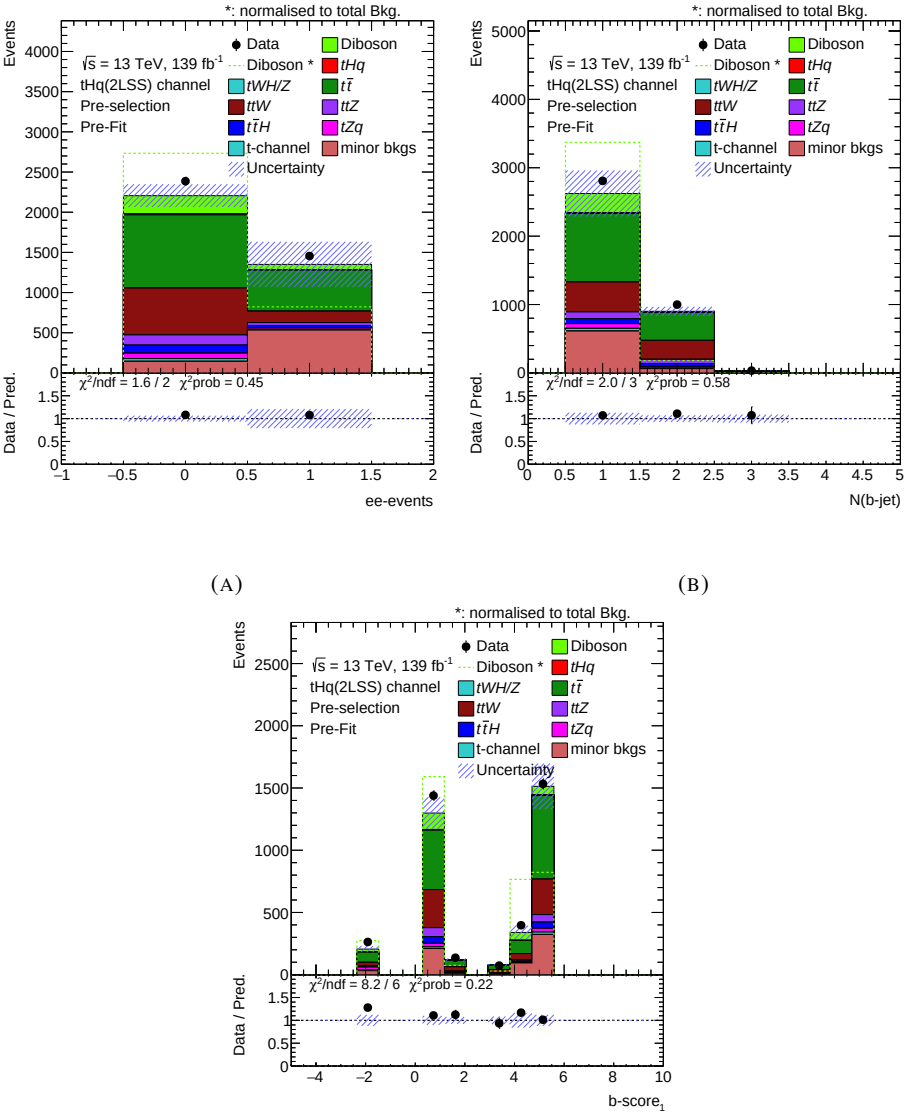
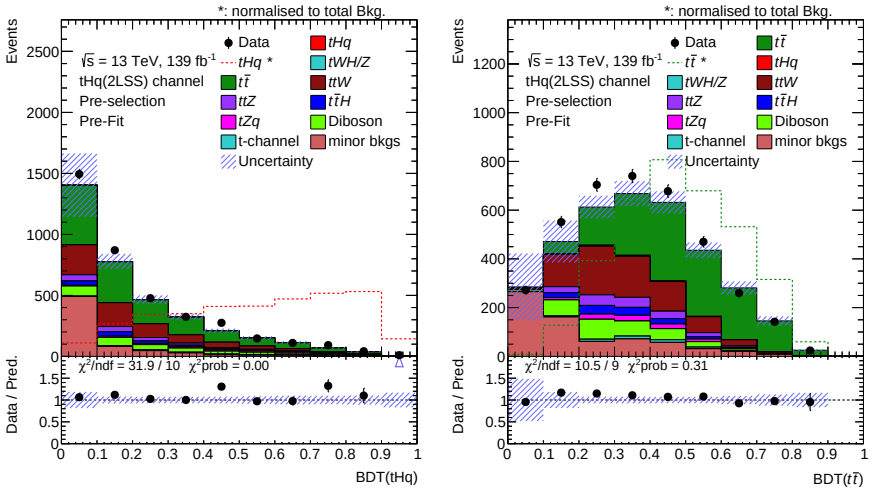
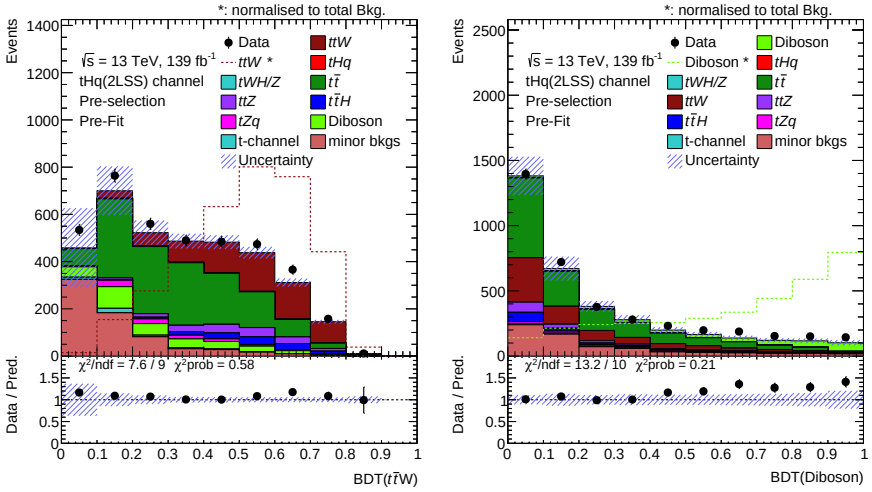


FIGURE 5.12: The three input variables with the highest *Gain* value in the 2ℓSS channel for BDT(diboson). The distributions show data and simulation samples in the pre-selection region for (A) $\mathrm{ee-events}$, (B) $\mathrm{N(b-jet)}$ and (C) $\mathrm{b-score_1}$. The uncertainty bands include the statistic and all the systematic sources. The lower panels show the ratio between real and simulated background data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.



(A) BDT(*tHq*)

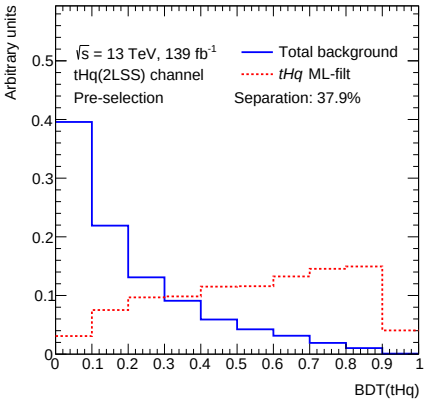
(B) BDT(t$\bar{\text{t}}$)



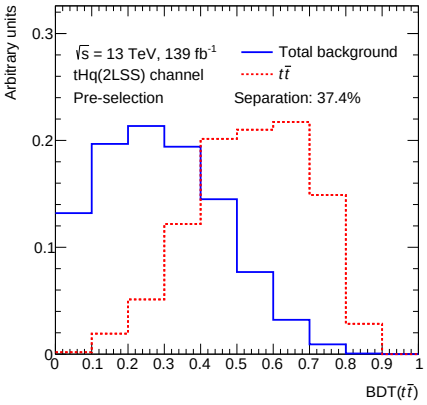
(C) BDT(t$\bar{\text{t}}W$)

(D) BDT(diboson)

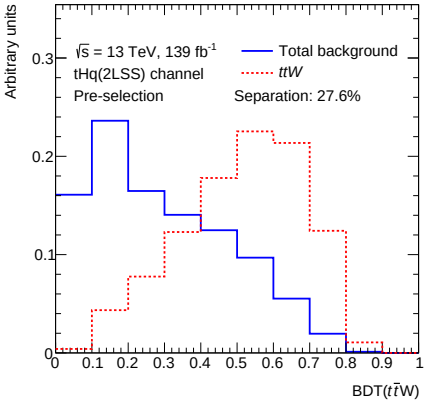
FIGURE 5.13: Distributions of the BDT scores for each of the BDT for (A) BDT(*tHq*), (B) BDT(t$\bar{\text{t}}$), (C) BDT(t$\bar{\text{t}}W$) and (D) BDT(diboson) in the 2ℓSS channel. The distributions show data and simulation samples in the pre-selection region. The dashed lines represent the target process of each BDT. The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated background data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.



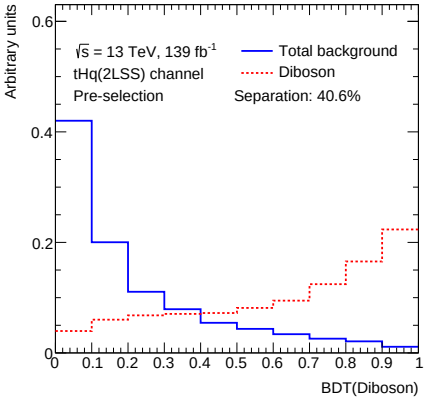
(A) BDT(*tHq*)



(B) BDT($t\bar{t}$)



(C) BDT($t\bar{t}W$)



(D) BDT(diboson)

FIGURE 5.14: Distribution of the BDT scores for each BDT for (A) BDT(*tHq*), (B) BDT($t\bar{t}$), (C) BDT($t\bar{t}W$) and (D) BDT(diboson) in the 2ℓSS channel. The target process (red dashed lines) and the background samples (blue solid line) are normalised to the same area. In each case the background sample is defined as any sample which is not target. The separation is computed using the formula 1 in Ref. [153].

As for the 3ℓ channel, three figures of merit are used for the 2ℓSS channel: the ROC curve, the ROC_AUC and the log_loss function. In this case the k-fold cross-validation also uses five folds, thus the figures of merits are measured five times, one per fold. The five pairs of the ROC curves[5]resulting from the k-fold cross-validation steps for the four BDTs are shown in figure 5.15. All the ROC curves show a similar behaviour, demonstrating the robustness of the BDTs. The five values, together with their means, of the ROC_AUC, and the log_loss of the four BDTs are shown in table 5.7 and 5.8, respectively.

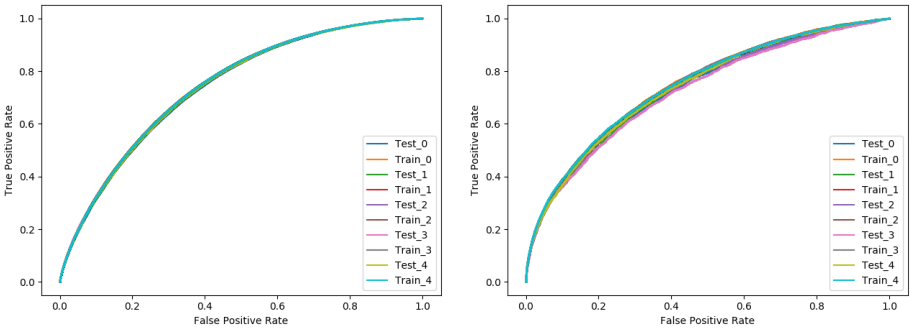
TABLE 5.7: Values of ROC_AUC given by each BDT in the 2ℓSS channel. The value of the ROC_AUC for each fold X is shown in each raw. In the last raw, the mean value and its statistical uncertainty are given by the k-fold cross-validation method.

	BDT (*tHq*)	BDT ($t\bar{t}$)	BDT ($t\bar{t}W$)	BDT (diboson)
ROC_AUC_1	0.660	0.656	0.664	0.730
ROC_AUC_2	0.664	0.645	0.662	0.734
ROC_AUC_3	0.665	0.653	0.658	0.736
ROC_AUC_4	0.666	0.644	0.662	0.734
ROC_AUC_5	0.662	0.653	0.664	0.731
ROC_AUC	0.663 ± 0.002	0.650 ± 0.005	0.662 ± 0.002	0.733 ± 0.002

TABLE 5.8: Values of log_loss given by each BDT in the 2ℓSS channel. The value of the log_loss for each fold X is shown in each raw. In the last raw, the mean value and its statistical uncertainty are given from the k-fold cross-validation method.

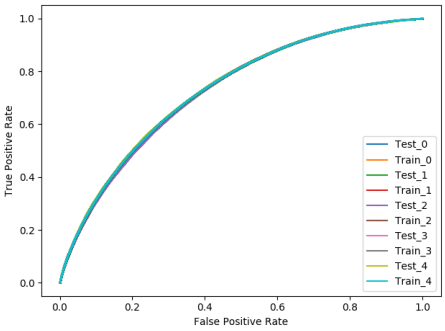
	BDT (*tHq*)	BDT ($t\bar{t}$)	BDT ($t\bar{t}W$)	BDT (diboson)
log_loss_1	0.509	0.400	0.539	0.415
log_loss_2	0.504	0.404	0.542	0.417
log_loss_3	0.506	0.405	0.541	0.415
log_loss_4	0.505	0.403	0.542	0.415
log_loss_5	0.505	0.449	0.540	0.416
log_loss	0.506 ± 0.002	0.402 ± 0.002	0.541 ± 0.002	0.416 ± 0.001

[5]Each pair corresponds to a fold where test and train ROC curves are included.

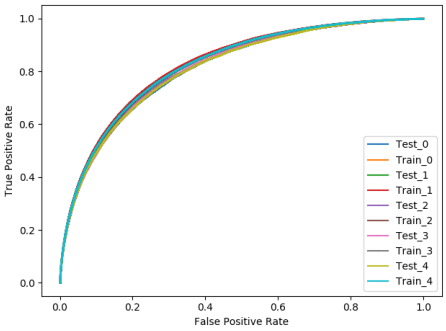


(A) BDT(tHq)

(B) BDT($t\bar{t}$)



(C) BDT($t\bar{t}W$)



(D) BDT(diboson)

FIGURE 5.15: ROC curves for each BDT in the 2ℓSS channel: (A) BDT(tHq), (B) BDT($t\bar{t}$), (C) BDT($t\bar{t}W$) and (D) BDT(diboson). Five pairs of ROC curves, i.e. for test and train samples, are shown where each one corresponds to one fold X.

5.3 Background estimation

Background processes can be classified regarding their sources as irreducible and reducible. The irreducible backgrounds are processes whose final states can contain the same particles as the signal of interest. Hence, these kind of backgrounds are harder to reduce. The reducible backgrounds are the results of experimental inefficiencies. In other words, they arise from an incorrect experimental performance, such as the lepton charge mis-identification. Consequently, improving the experimental techniques would reduce the effects of this kind of background. In addition to the charge mis-identification, which is important for the 2ℓSS channel, these backgrounds can also include mis-identified (also known as fake) or non-prompt leptons in the final state. A description about these backgrounds and how they are estimated is given in section 5.3.1.

In the case of the 3ℓ channel, the most important irreducible backgrounds are diboson, $t\bar{t}Z$, $t\bar{t}W$, *tZq*, $t\bar{t}H$ and *tWZ*, and the most important reducible backgrounds are $t\bar{t}$, *Z+jets* and single top-quark processes. Other minor background processes are also considered as triboson, three and four top quarks and Higgs boson productions (VBF,ggH and VH).

In the case of the 2ℓSS channel, the most important irreducible and reducible background processes are similar to the 3ℓ channel. However, in this case the mis-identification of the electric charge of the electrons also plays an important role in the reducible backgrounds. It is directly estimated from the information giving by the MC simulation, in particular using the tools provided by the ATLAS Isolation and Fake Forum group. An uncertainty of $10\,\%$ is applied to this category, and it is provided from a data-driven method in similar analysis. Moreover, a specific data-driven method is also applied for the 2ℓSS channel, but it is unfortunately out of the scope of this thesis.

5.3.1 Fakes and non-prompt estimation with the template-fit method

Fakes and non-prompt leptons are one important source of reducible background for the 3ℓ and the 2ℓSS channel. They are objects mis-identified as electrons or muons. Usually, the origins of fakes are electrons or muons from meson decays, electrons from γ–conversions or a light-flavour jet with a similar signature (at detector level) to a lepton.

The non-prompt leptons appear from heavy-flavour (HF) hadron decays produced in association with either a W or Z boson decaying to leptons.

Since fake/non-prompt leptons are an important background for both final states and they are not perfectly estimated, a semi-data-driven method called *template fit method* (TFM) is used to estimate the contribution of this background. The TFM uses the different tools provided by the ATLAS Isolation and Fake Forum group to identify the fake/non-prompt leptons from the MC simulation. The fake/non-prompt leptons are split in different categories: muons from HF hadron decays (μ_{HF}), electrons from HF hadron decays (e_{HF}) and electrons from γ–conversions (e_{conv}). In addition to these fake/non-prompt categories, a category called *Other fake* is defined to merge the leptons categorised as fake/non-prompt but not as HF hadron decays or γ–conversions.

After the classification, normalisation factors for fake/non-prompt leptons are obtained from a profile likelihood binned fit to the data which includes special regions enriched with fake/non-prompt. These regions (also called control regions (CRs)) are defined for each one of the categories mentioned before for the 3ℓ and the 2ℓSS channel, they are defined in section 5.4. In the case of *Other fake*, a specific region is not defined due to its minor contribution to the fakes and thus, a normalisation factor is not needed (nor provided). The normalisation factors are $\mathrm{k}(\mu_{\text{HF}})$, $\mathrm{k}(e_{\text{HF}})$ and $\mathrm{k}(e_{\text{conv}})$ attending the three categories of fake/non-prompt leptons, i.e μ_{HF}, e_{HF} and e_{conv}, respectively.

The variables used in the TFM are chosen to have a different shape distribution between the specific fake category in the region enriched on it and the other backgrounds. The binning is optimised to obtain a statistical uncertainty per bin smaller than $20\,\%$. Moreover, a conservative uncertainty of $50\,\%$ is employed to the total number of events identified as *Other fake*.

In the case of the 3ℓ channel, the fake/non-prompt leptons roughly represent $28\,\%$ of the background processes in the signal region. Only *Z+jets* and $\mathrm{t\bar{t}}$ processes are considered as possible fake/non-prompt because they are the main reducible backgrounds. The variable used in the TFM are shown in the figures 5.16 (before the fit, i.e. pre-fit) and 5.17 (after the fit, i.e. post-fit). Overall, a good agreement between data and MC distributions is found in the three control regions. The normalisation factors $\mathrm{k}(\mu_{\text{HF}})$, $\mathrm{k}(e_{\text{HF}})$ and $\mathrm{k}(e_{\text{conv}})$ are shown in the figure 5.20a, in the third column labelled as *All CR*. They

are compatible with the unity within their uncertainties (though $\mathrm{k(e_{HF})}$ is compatible at 2σ).

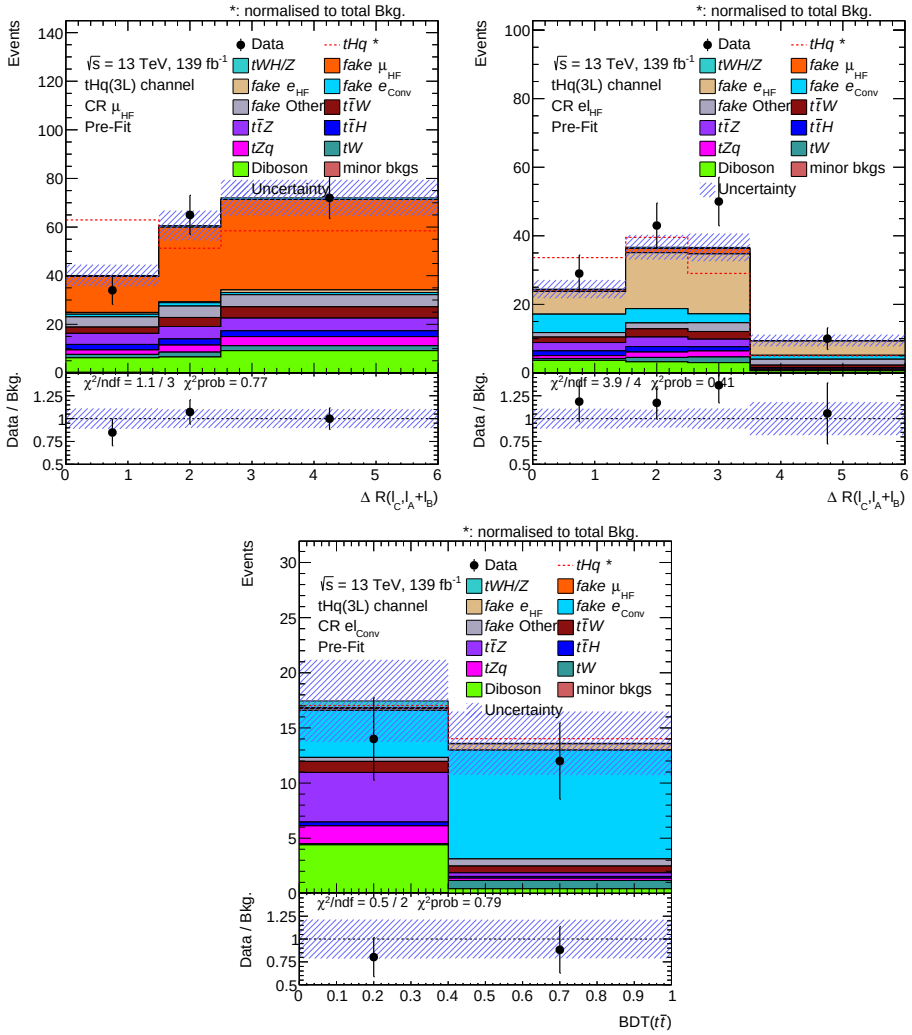


FIGURE 5.16: Pre-fit distributions in the three dedicated control regions used in the TFM to calculate the normalisation factor for the different fake/non-prompt categories for the 3ℓ channel. The real and simulated data events are shown for: (A) $\Delta\mathrm{R}(\ell_\mathrm{C}, \ell_\mathrm{A}+\ell_\mathrm{B})$, (B) $\Delta\mathrm{R}(\ell_\mathrm{C}, \ell_\mathrm{A}+\ell_\mathrm{B})$ and (C) $\mathrm{BDT(t\bar{t})}$. Variables are defined in table 5.2. The uncertainties bands include statistical and all the systematic sources.

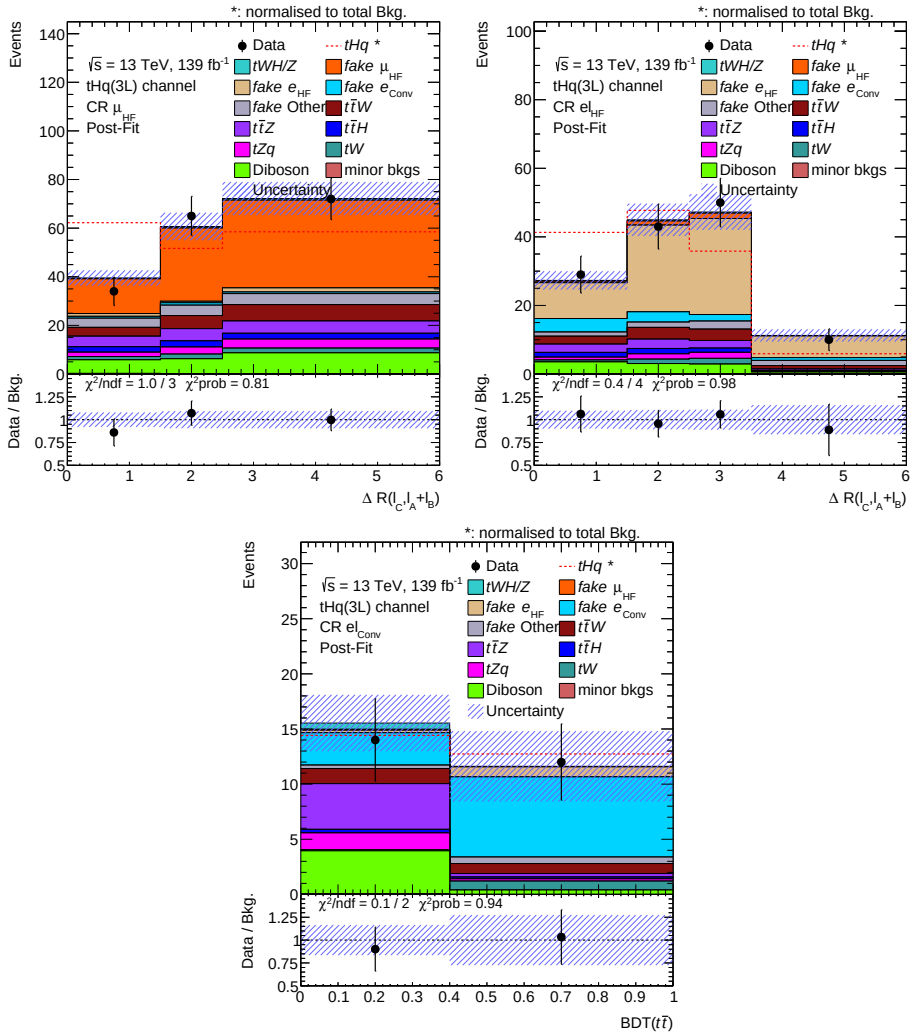


FIGURE 5.17: Post-fit distributions in the three dedicated control regions used in the TFM to calculate the normalisation factor for the different fake/non-prompt categories for the 3ℓ channel. The real and simulated data events are shown for: (A) $\Delta R(\ell_C, \ell_A + \ell_B)$, (B) $\Delta R(\ell_C, \ell_A + \ell_B)$ and (C) $\mathrm{BDT}(\mathrm{t\bar{t}})$. Variables are defined in table 5.2. The uncertainty bands include the statistical and all the systematic sources.

In the case of the 2ℓSS channel, the fake/non-prompt leptons roughly represent $38\,\%$ of the background processes in the signal region. Only $\mathrm{t\bar{t}}$ process is considered as possible fake/non-prompt because it is the main reducible backgrounds. The variables used in the TFM are shown in the figures 5.18 (pre-fit) and 5.19 (post-fit). A good agreement is found in all the control regions. The normalisation factors $\mathrm{k(e_{HF})}$, $\mathrm{k(e_{conv})}$ and $\mathrm{k(\mu_{HF})}$ are shown in figure 5.20b, in the column labelled as *All CR*. They are compatible with the unity within their uncertainties ($1\,\sigma$ for this channel).

Moreover, two tests of the stability of the method are performed. The goal of these tests is to show that the normalisation factors are independent to the shape of the distributions and to other normalisation factor involved in the analysis, i.e. $\mathrm{k(t\bar{t}}W)$. For the first test, the TFM is made using the binning shown in the figures 5.16 and 5.18 but dropping the control region for the $\mathrm{t\bar{t}}W$ process and fixing the value of it normalisation to one (i.e. $\mathrm{k(t\bar{t}}W) = 1$). The values of the normalisation factors for this test are shown in 5.20a, column *No* $\mathrm{t\bar{t}}W$ *CR*, for the 3ℓ channel, and in 5.20b, column *No* $\mathrm{t\bar{t}}W$ *CR* for the 2ℓSS channel. For the second test, the TFM is made using one bin in all regions, dropping the control region for the $\mathrm{t\bar{t}}W$ process and fixing the value of $\mathrm{k(t\bar{t}}W)$ to 1. The results of the second test are shown in figure 5.20a and 5.20b, column *One bin*, for the 3ℓ and 2ℓSS channel, respectively. All in all, the results show that the values of the normalisation factors are robust against the shape and the normalisation factor $\mathrm{k(t\bar{t}}W)$ since they are compatible in the three scenarios within the uncertainties.

In addition, comparing the values of the normalisation factors for both channels, each one of them is compatible within their uncertainties.

5.4 Definition of the signal, control and validation regions

Several regions of interest are defined after the process of optimisation and evaluation of the best BDTs for each analysis, as it shows in section 5.1.2 and 5.2.2. There are three different kind of regions regarding their goals: the signal region (SR) is defined to maximise the contribution of the signal process, i.e. *tHq*, the control region (CR) aims to estimate the mis-modelling of backgrounds with less accurate simulation, and the validation region (VR) is defined to evaluate the resulting model in different backgrounds.

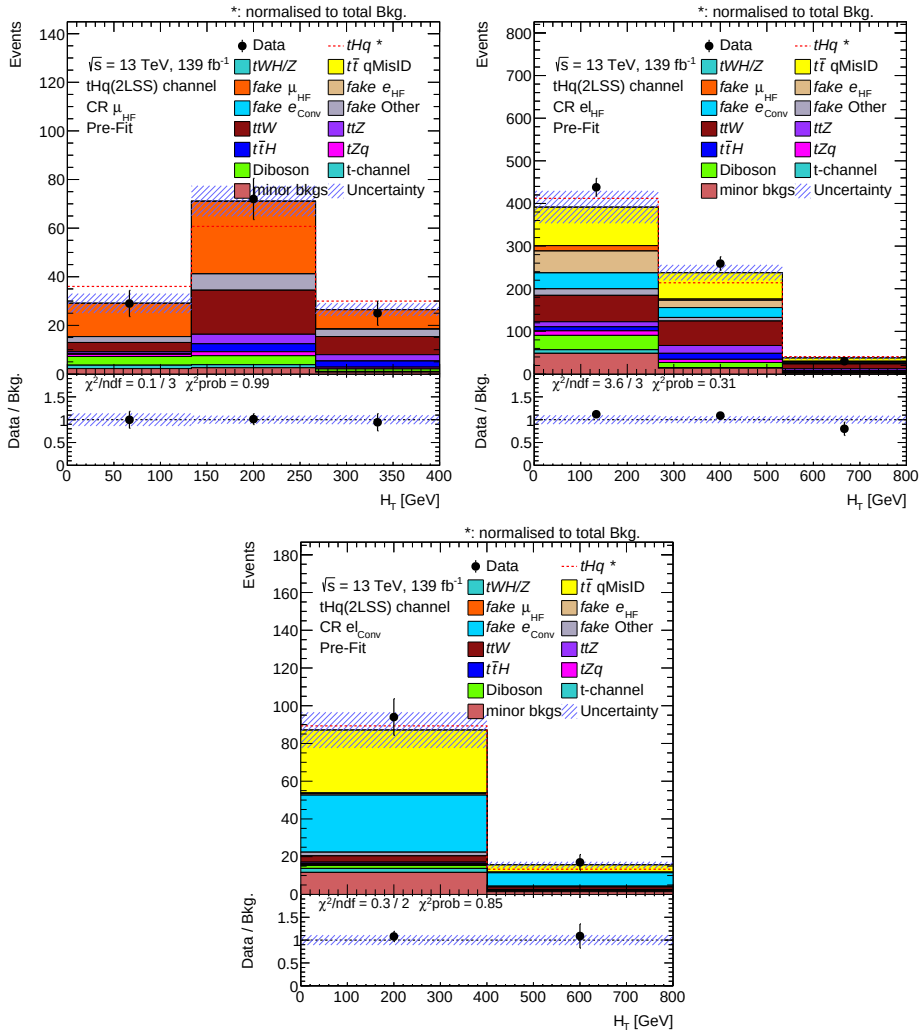


FIGURE 5.18: Pre-fit distributions in the three dedicated control regions used in the TFM to calculate the normalisation factor for the different fake/non-prompt categories for the 2ℓSS channel. The real and simulated data events are shown for: (A) H_T, (B) H_T and (C) H_T. Variables are defined in table 5.6. The uncertainties bands include all the systematic sources. The uncertainty bands include the statistical and all the systematic sources.

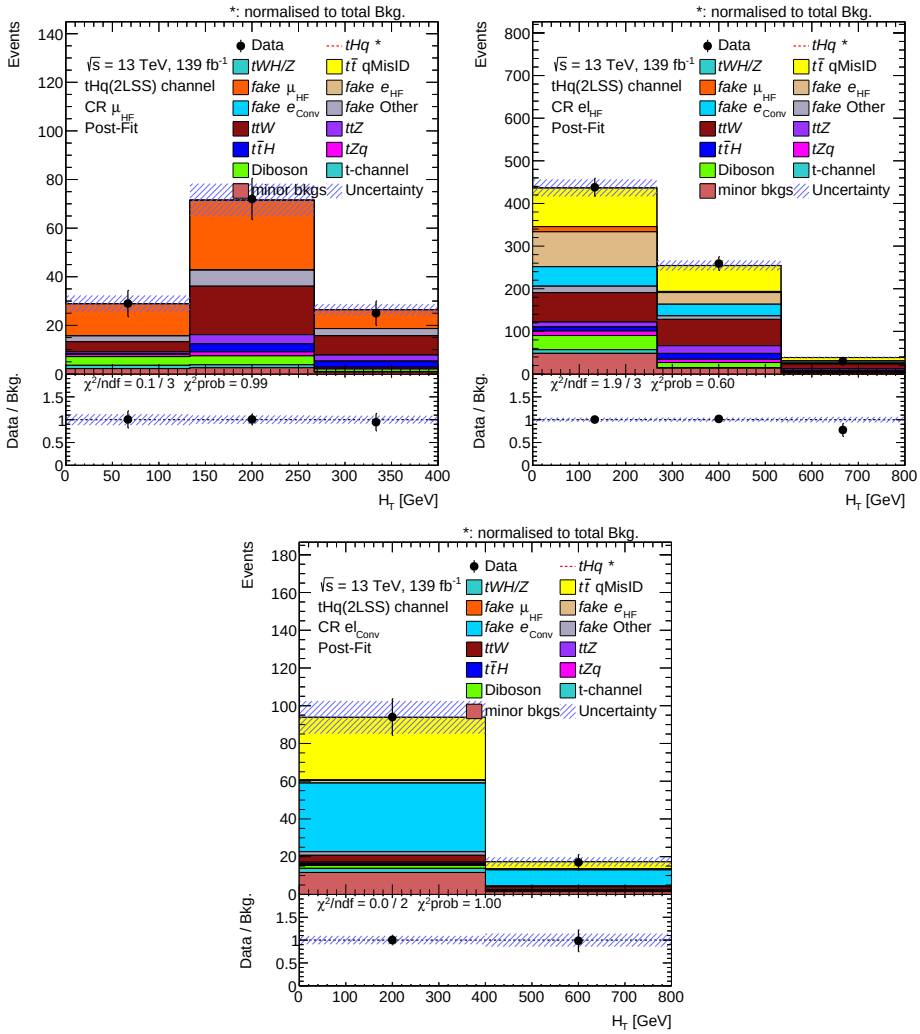
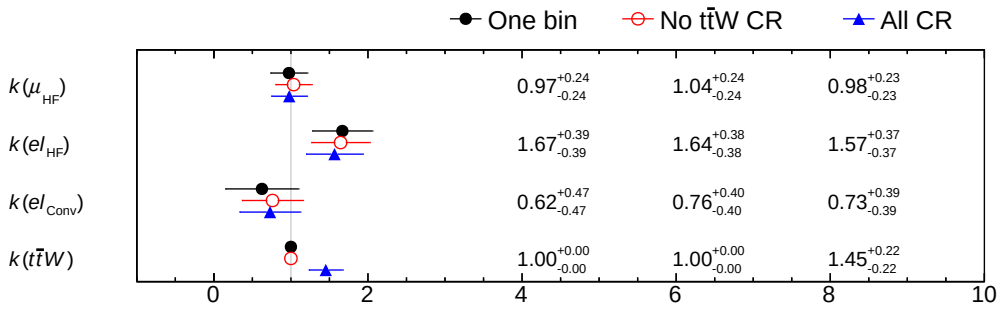
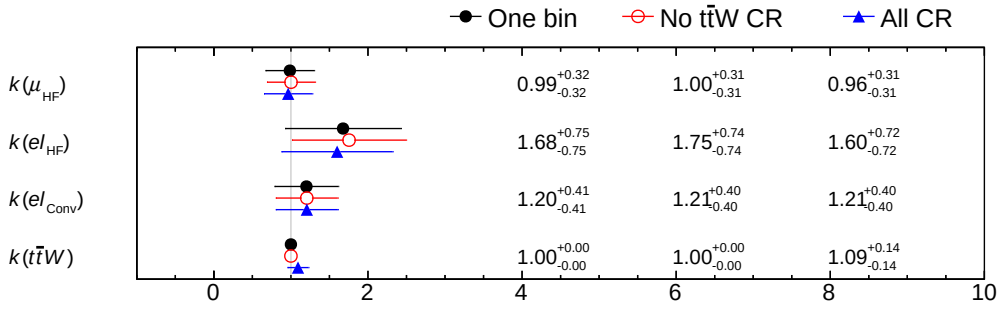


FIGURE 5.19: Post-fit distributions in the three dedicated control regions used in the TFM to calculate the normalisation factor for the different fake/non-prompt categories for the 2ℓSS channel. The real and simulated data events are shown for: (A) H_T, (B) H_T and (C) H_T. Variables are defined in table 5.6. The uncertainties bands include all the systematic sources. The uncertainty bands include the statistical and all the systematic sources.



(A) 3ℓ channel



(B) 2ℓSS channel

FIGURE 5.20: Normalisation factors for the μ_{HF}, e_{HF} and e_{conv} categories. The uncertainty includes statistical and systematic effects. The black circles show the results of the fit without relying on the templates shape, and only considering the mis-identified lepton CRs and fixing the other normalisation factors. The red circles show the result fitting over binned distributions in the mis-identified lepton CRs but still fixing the other normalisation factors. The blue triangles show the results when the CR($t\bar{t}W$) and the normalisation factors for $t\bar{t}W$ are also included in the fit.

Moreover, the regions must be statistically independent, even though the CR is defined in a region of the phase space similar to the one in the SR.

All the regions are defined using a mixture of requirements on the BDT output, the jet multiplicity, and also invariant masses. In addition to the criteria mentioned before, an extra flag called *Ambiguity Requirement* is used. It is related to the identification of the electron, in particular if a conversion electron candidate can be reconstructed from an ID track, the extra tracks close to the ID tracks to this electron are not considered and the flag is activated.

The table 5.9 summarises the definition of SR, CRs and VRs for the 3ℓ channel. In this case, three CRs are defined to extract the mis-identified leptons backgrounds: CR(μ_{HF}), CR(e_{HF}) and CR(e_{conv}) and one CR to measure the modelling of the $t\bar{t}W$ process. Moreover, three VRs are defined to evaluate the model in the main irreducible backgrounds the $t\bar{t}Z$, the *tZq* and the diboson process. The composition predicted by the simulated samples and data event for all the regions is shown in table 5.10.

TABLE 5.9: Selection requirements for the analysis region definitions in the 3ℓ channel.

Region	BDT score	Ambiguity requirement	Jets	Softest lepton	Invariant mass	Other
SR	$\mathrm{BDT}(tHq) > 0.7$ $\mathrm{BDT}(t\bar{t}) < 0.9$ $\mathrm{BDT}(t\bar{t}W) < 0.8$	yes	-	-	-	-
CR(μ_{HF})	$\mathrm{BDT}(tHq) < 0.7$ $\mathrm{BDT}(t\bar{t}) > 0.5$ $\mathrm{BDT}(t\bar{t}W) < 0.8$	yes	-	muon	-	-
CR(e_{HF})	$\mathrm{BDT}(tHq) < 0.7$ $\mathrm{BDT}(t\bar{t}) > 0.5$ $\mathrm{BDT}(t\bar{t}W) < 0.8$	yes	-	electron	-	-
CR(e_{conv})	-	inverted	-	electron	-	-
CR($t\bar{t}W$)	$\mathrm{BDT}(t\bar{t}W) > 0.8$	yes	-	-	-	-
VR($t\bar{t}Z$)	$\mathrm{BDT}(tHq) < 0.7$ $\mathrm{BDT}(t\bar{t}) < 0.5$ $\mathrm{BDT}(t\bar{t}W) < 0.8$	yes	$N^{\text{b-jet}} \geq 2$ $N^{\text{jet}}_{\text{central}} \geq 4$	-	$\Delta m(Z, \ell\ell_{SF})_{\text{min}} < 10\,\text{GeV}$	$E_T^{\text{miss}} > 50\,\text{GeV}$
VR(*tZq*)	$\mathrm{BDT}(tHq) < 0.7$ $\mathrm{BDT}(t\bar{t}) < 0.5$ $\mathrm{BDT}(t\bar{t}W) < 0.5$	yes	$N^{\text{jet}}_{\text{central}} < 4$ $N^{\text{jet}}_{\text{forward}} \geq 1$	-	-	-
VR(Diboson)	$\mathrm{BDT}(tHq) < 0.7$ $\mathrm{BDT}(t\bar{t}) < 0.5$ $\mathrm{BDT}(t\bar{t}W) < 0.5$	yes	$N^{\text{jet}}_{\text{central}} < 4$ $N^{\text{jet}}_{\text{forward}} = 0$	-	-	

TABLE 5.10: Yields as predicted by the MC simulation and data for the different regions in the 3ℓ channel. Each MC simulation is normalised to the cross-section of the process. The uncertainties include statistical and all the systematic sources.

Process	SR	CR(μ_{HF})	CR(e_{HF})	CR(e_{conv})	CR($t\bar{t}W$)	CR($t\bar{t}Z$)	CR(tZq)	VR(Diboson)
tHq	1.28 ± 0.06	0.162 ± 0.020	0.102 ± 0.015	0.020 ± 0.004	0.211 ± 0.023	0.027 ± 0.010	0.227 ± 0.027	0.140 ± 0.023
tWH	0.44 ± 0.07	0.31 ± 0.05	0.185 ± 0.032	0.025 ± 0.009	0.84 ± 0.07	0.072 ± 0.022	0.14 ± 0.04	0.180 ± 0.034
tWZ	1.0 ± 0.5	1.4 ± 0.7	0.8 ± 0.5	0.64 ± 0.34	3.5 ± 1.9	6 ± 4	13 ± 7	17 ±10
fake μ_{HF}	11.0 ± 2.3	83 ± 6	3.4 ± 1.1	0.08 ± 0.12	9.7 ± 2.1	0.6 ± 0.6	22 ± 7	72 ±22
fake e_{HF}	7.2 ± 2.4	2.2 ± 0.9	45 ± 5	0.7 ± 0.9	7.1 ± 1.5	0.02 ± 0.04	6 ± 4	30 ± 8
fake e_{Conv}	4.3 ± 1.3	3.3 ± 0.8	13.3 ± 2.8	14 ± 4	10.7 ± 1.6	0.7 ± 0.4	2 ± 4	3 ± 6
fake Other	6 ± 4	14 ± 7	7 ± 4	1.0 ± 0.7	10 ± 5	0.8 ± 0.6	1.9 ± 2.8	16 ±13
$t\bar{t}W$	7.7 ± 0.8	10.9 ± 0.7	6.9 ± 0.5	1.62 ± 0.13	100 ± 4	2.69 ± 0.30	5.0 ± 0.6	7.2 ± 0.7
$t\bar{t}Z$	17 ± 4	14.9 ± 3.5	7.9 ± 1.9	4.8 ± 1.1	37 ± 8	84 ±19	62 ±16	61 ±22
$t\bar{t}H$	9.1 ± 1.5	7.2 ± 1.2	4.4 ± 0.8	0.50 ± 0.12	20.9 ± 3.4	3.5 ± 0.7	2.1 ± 0.4	2.4 ± 0.5
tZq	13.2 ± 1.9	8.5 ± 1.4	4.5 ± 0.7	1.83 ± 0.28	5.9 ± 0.9	10.6 ± 2.8	117 ±17	57 ±12
tW	1.4 ± 0.8	5 ± 4	3.5 ± 2.1	0.9 ± 0.8	1.6 ± 1.4	0.0 ± 0.0	1.3 ± 1.6	1.2 ± 1.2
Diboson	16 ± 4	21 ± 6	10.3 ± 2.8	4.7 ± 1.2	12.0 ± 3.4	6.5 ± 2.1	112 ±33	250 ±70
Minor backgrounds	0.50 ± 0.27	0.5 ± 0.4	0.23 ± 0.12	0.08 ± 0.04	2.6 ± 1.4	0.59 ± 0.30	4.0 ± 3.0	3.3 ± 2.2
Total background	94 ±10	173 ±15	107 ±10	31 ± 5	222 ±15	116 ±20	350 ±50	520 ±90
Data	107	171	132	26	266	107	421	599

In the case of the 2ℓSS channel, the regions are summarised in table 5.11. Similar to the 3ℓ channel, there are also three CRs that aim to evaluate the mis-identified leptons backgrounds: CR(μ_{HF}), CR(e_{HF}) and CR(e_{conv}) and one CR to measure the modelling of the $t\bar{t}W$ process. Only one VR is defined targeting the diboson process for the 2ℓSS channel analysis. The composition predicted by the simulated samples and data events for all the regions is shown in table 5.12.

5.5 Systematics uncertainties

The different sources of systematic uncertainties involved in this analysis are explained in this section.

5.5.1 Monte Carlo statistical uncertainty

The statistical uncertainty arises from the finite number of MC simulated events for the signal and background processes. It contributes to the overall uncertainty through the likelihood fit expression explained in section 5.6 with the other systematic uncertainties

5.5.2 Experimental uncertainties

Experimental (i.e. detector-related) uncertainties can arise from the reconstruction of physics objects in the detector. These uncertainties are applied as recommended by

TABLE 5.11: Selection requirements for the analysis region definitions in the 2ℓSS channel.

Region	BDT score	Ambiguity requirement	Jets	ℓ_1-ℓ_2 flavour	Other
SR	$\mathrm{BDT}(tHq) > 0.65$ $\mathrm{BDT}(t\bar{t}) < 0.5$ $\mathrm{BDT}(t\bar{t}W) < 0.6$ $\mathrm{BDT}(\mathrm{VV}) < 0.8$	yes	-	-	-
CR(μ_{HF})	$\mathrm{BDT}(tHq) < 0.65$ $\mathrm{BDT}(t\bar{t}) > 0.3$ $\mathrm{BDT}(t\bar{t}W) < 0.6$ $\mathrm{BDT}(\mathrm{VV}) < 0.9$	yes	-	μ-μ	-
CR(e_{HF})	$\mathrm{BDT}(tHq) < 0.65$ $\mathrm{BDT}(t\bar{t}) > 0.3$ $\mathrm{BDT}(t\bar{t}W) < 0.6$ $\mathrm{BDT}(\mathrm{VV}) < 0.9$	yes	-	μ/e-e	$\mathrm{H_T}(\ell) < 225\,\mathrm{GeV}$
CR(e_{conv})	$\mathrm{BDT}(t\bar{t}) > 0.3$	inverted	-	μ/e-e	$\mathrm{m}(\ell_1 + \ell_2) < 150\,\mathrm{GeV}$
CR($t\bar{t}W$)	$\mathrm{BDT}(t\bar{t}W) > 0.6$ $\mathrm{BDT}(t\bar{t}) < 0.3$	yes	-	-	-
VR(VV)	$\mathrm{BDT}(tHq) < 0.65$ $\mathrm{BDT}(t\bar{t}) < 0.5$ $\mathrm{BDT}(t\bar{t}W) < 0.6$ $\mathrm{BDT}(\mathrm{VV}) > 0.9$	yes	-	-	-

TABLE 5.12: Yields as predicted by the MC simulation and data for the different regions in the 2ℓSS channel. Each MC simulation is normalised to cross-section of the process. The uncertainties include statistical and all the systematic sources.

Process	SR	CR(μ_{HF})	CR(e_{HF})	CR(e_{conv})	CR($t\bar{t}W$)	VR(VV)
tHq	3.1 ± 0.6	0.43 ± 0.09	1.75 ± 0.30	0.067 ± 0.008	0.41 ± 0.05	0.149 ± 0.032
tWH	0.14 ± 0.13	0.23 ± 0.04	1.11 ± 0.24	0.021 ± 0.010	0.89 ± 0.10	0.047 ± 0.021
$t\bar{t}$ mis-ID	7.5 ± 1.7	0.0 ± 0.0	159 ±20	37 ± 5	6.8 ± 1.8	1.9 ± 1.2
fake μ_{HF}	15.3 ± 3.4	52 ± 4	15.9 ± 2.4	0.24 ± 0.26	9.4 ± 1.5	4.3 ± 1.5
fake e_{HF}	5.4 ± 2.0	0.0 ± 0.0	71 ± 5	1.2 ± 0.4	1.6 ± 0.6	1.1 ± 0.9
fake e_{Conv}	5.0 ± 1.5	0.0 ± 0.0	63 ± 5	37.3 ± 3.1	6.6 ± 1.5	0.7 ± 0.6
fake Other	4.0 ± 2.4	12 ± 6	25 ±13	2.4 ± 1.3	8 ± 4	0.4 ± 0.9
$t\bar{t}W$	17.6 ± 1.0	29.4 ± 1.8	131 ± 6	4.63 ± 0.29	163 ± 6	3.0 ± 0.5
$t\bar{t}Z$	4.9 ± 1.3	7.2 ± 1.7	34 ± 8	1.11 ± 0.29	25 ± 6	0.68 ± 0.26
$t\bar{t}H$	3.1 ± 0.9	6.2 ± 1.1	26 ± 5	0.77 ± 0.15	26 ± 4	0.26 ± 0.14
tZq	19.3 ± 3.3	3.3 ± 0.6	19.5 ± 2.6	0.68 ± 0.11	1.97 ± 0.28	1.48 ± 0.32
Diboson	9.8 ± 2.7	8.3 ± 2.7	47 ±12	1.8 ± 0.5	13.3 ± 3.4	60 ±17
Single top-quark t-channel	5 ± 6	3 ± 5	10 ± 7	2.3 ± 3.4	0.6 ± 0.9	2.1 ± 2.4
Minor backgrounds	3 ± 5	5.6 ± 3.5	64 ±34	13 ± 7	7 ± 4	11 ±10
Total background	100 ±12	127 ±12	670 ±50	103 ±11	269 ±15	87 ±21
Data	121	126	727	111	288	123

different dedicated performance groups inside the ATLAS collaboration. The sources of uncertainty that are considered are listed in the following.

Luminosity: The 2015–2018 luminosity estimate of 139.0 fb^{-1} has a relative uncertainty of $1.7\,\%$ [76]. This uncertainty is obtained using the LUCID-2 detector [78] for the baseline luminosity measurements. This uncertainty is applied to all processes modelled using MC simulations.

Pile-up re-weighting: The events of the MC simulation samples are re-weighted to match the observed distribution of the average number of interactions per bunch-crossing in data [154], as shown in figure 2.3. An uncertainty related this difference, applied MC events to account for differences in pile-up distributions between MC and data, is applied. This uncertainty is obtained by re-scaling the $\langle\mu\rangle$ value in data by $1/0.99$ and $1/1.07$ around the nominal scale factor of $1/1.03$ [155].

Jet energy scale: The jet energy scale (JES) and its uncertainty are derived combining information from test-beam data, LHC collision data and simulation [156, 157]. On the one hand, events with a vector boson and additional jets are used to calibrate jets in the central region. On the other hand, di-jets events are exploited to calibrate forward jets against the jets in the central region of the detector. Finally, multi-jet events are used to calibrate high p_T. All in all, there are 30 independent nuisance parameters, each with an up/down variation, regarding pile-up, jet flavour composition, single-particle response, and effects of jet not contained within the calorimeter.

Jet energy resolution: For the jet energy resolution (JER), a smearing model corresponding to 13 nuisance parameters is used. The JER is measured separately for data and MC using two in-situ techniques [156, 157]. A systematic uncertainty is defined as the quadratic difference between the JER for data and MC. To evaluate the associated systematic uncertainty, the energies of the jets in MC are smeared by their residual differences and the changes in shapes and normalisations of the final discriminant are compared to the default predictions. In order to propagate the

uncertainty in the p_T resolution, for each jet in MC, a random number (r) is generated from a Gaussian PDF with mean of zero and sigma equal to the quadratic difference between the fractional p_T resolution with the tool and the nominal one. The four-momentum of the jet is then scaled by a factor $1 + r$. Since jets in MC cannot be under-smeared, by definition the resulting uncertainty on the normalisation and shape of the final discriminant is one-sided. This uncertainty is then symmetrised.

Jet vertex tagger: Uncertainties associated to the JVT (see section 4.5) take into accounts for the residual contamination from pile-up jets after pile-up suppression and the MC generator choice [158].

Heavy- and light-flavour tagging: The efficiency of the flavour-tagging algorithm is measured for each jet flavour using control samples in data and in simulation. From these measurements, correction factors are derived to correct the tagging rates in the simulation. In the case of b-tagged jets, the correction factors and their uncertainties are estimated from data using di-leptonic $t\bar{t}$ events [159, 160]. In the case of c-jets, they are derived from jets arising from W boson decays in $t\bar{t}$ events [161]. In the case of light-flavour jets, the correction factors are derived using di-jet events [162]. Sources of uncertainty affecting the b- and c-tagging efficiencies are evaluated as a function of jet p_T, including bin-to-bin correlations. The `BTaggingEfficiencyTool` [160] is used to apply uncertainties in the efficiency. Additional uncertainties are assigned to account for the extrapolation of the b-tagging efficiency measurement from the p_T region used to determine the correction factors to regions with higher p_T. In total, 20 nuisance parameters are considered for the light-, and c-jets, and 50 nuisance parameters for the b-jets. All this set of nuisance parameters model this experimental uncertainty in the final fit.

Electron and Muon reconstruction, identification, isolation, and trigger: The performance differs between data and MC simulation for electron and muon reconstruction, identification, isolation, and trigger. Therefore, scale factors are applied to correct them. They are measured with a "tag-and-probe" method in

$Z \to e^+e^-$and $J/\psi \to e^+e^-$events using similar methods to Ref. [163]. The uncertainties are evaluated by varying up and down by $1\,\sigma$ the predicted event yields and re-applying the event selection to the signal and backgrounds.

Electron energy scale and resolution: The accuracy of the electron momentum scale and resolution in MC is checked using reconstructed distributions of the $Z \to e^+e^-$and $J/\psi \to e^+e^-$masses. E/p studies using $W \to e\nu$ events are also used. Small discrepancies are observed between data and MC and corrections for the electron energy scale and resolution are implemented using the tools provided by the E/gamma combined performance group of the ATLAS collaboration. The number of different nuisance parameters (3) is reduced using a simplified correlation model. In this scheme all the effects are considered fully correlated in η and they are summed in quadrature to provide up/down variation for a reduced set of uncertainties [132, 133].

Muon momentum scale and resolution: Momentum scale and resolution corrections are applied to muons in MC simulated events. Uncertainties on both the momentum scale and resolutions in the MS and the ID tracking systems are considered and varied separately (using 4 nuisance parameters). Additional uncertainties are considered to account for the charge-dependent scale correction ("sagitta bias") applied on data. A more detailed description can be found in Refs. [135, 136].

E_T^{miss} **soft term:** Uncertainties are applied to the scale and resolution of the soft-track component ("soft term") on the E_T^{miss}, which cannot be associated to any of the reconstructed and calibrated physics objects ("hard term"). They are derived from the level of agreement between data and MC of the p_T balance between the hard and soft E_T^{miss} components [148]. The scale and resolution uncertainties of E_T^{soft} are treated as separate nuisance parameters.

5.5.3 Theoretical uncertainties

In this section the definition of the background modelling uncertainties is given. Modelling uncertainties are evaluated in three ways: comparing the nominal prediction to

an alternative prediction, varying the internal parameters of the nominal simulation, or varying the predicted cross-section within the theoretical uncertainty.

$\mathrm{t\bar{t}}$ **modelling:** Four independent variations affecting ISR are defined. The uncertainty due to the choice of the renormalisation and factorisation scales, respectively, μ_{R} and μ_{F}, in the hard-scatter and in the showering is evaluated using a prediction obtained with the POWHEG +PYTHIA 8 simulation. The μ_{R} and μ_{F} are varied independently by a factor 0.5 and 2. The scales in the showering are varied changing the Var3c eigentune of the A14 tune [109]. Finally, the $\mathrm{h_{damp}}$ parameter, which is set to $1.5 \cdot \mathrm{m_{top}}$ in the nominal simulation, is set to $3 \cdot \mathrm{m_{top}}$ in a dedicated POWHEG +PYTHIA 8 sample whose prediction is compared to the nominal.

The POWHEG +PYTHIA 8 sample where μ_{R} used in the final-state shower is varied by a factor 0.5 and 2 with respect to the nominal value is used to evaluate the uncertainty due to the FSR simulation.

An uncertainty is also attributed to the choice of the POWHEG approach to perform the matching between the hard-scatter, and the parton shower is estimated comparing the POWHEG +HERWIG 7.1.3 prediction with
the AMC@NLO +HERWIG 7.1.3 simulation.

The uncertainty due to the choice of the hadronisation model and the other non-perturbative aspects of the parton shower is evaluated comparing the nominal sample with POWHEG +HERWIG 7.2.1.

The theoretical prediction of the $\mathrm{t\bar{t}}$ cross-section is affected by the scale uncertainty, the PDF+α_{S} uncertainty and the uncertainty on the top-quark mass. For a top-quark mass of $172.5\,\mathrm{GeV}$ the $\mathrm{t\bar{t}}$ cross-section is

$$834^{+21}_{-30}(\text{scale})^{+21}_{-21}(\text{PDF+}\alpha_{\mathrm{S}})\text{pb}.$$

The total uncertainty, corresponding to the 6%, is used to vary the $\mathrm{t\bar{t}}$ cross-section.

$\mathrm{t\bar{t}H}$ **modelling:** To estimate the impact of the $\mathrm{t\bar{t}H}$ modelling in this analysis ISR, FSR, parton-shower and hadronisation model and parton shower to hard-scatter matching uncertainties are evaluated.

Similarly, to the $t\bar{t}$ case, the ISR uncertainty is evaluated using the POWHEG +PYTHIA 8 simulation and varying independently the μ_R and μ_F scale by a factor 0.5 and 2 and changing the Var3c eigentune of the A14 tune. The FSR uncertainty is also derived using the POWHEG +PYTHIA 8 simulation by varying by a factor of 0.5 and 2 the renormalisation scale used in the final-state shower.

The uncertainty due to the choice of the hadronisation model and the other non-perturbative aspects of the parton shower is evaluated comparing the nominal sample with POWHEG +HERWIG 7.

The uncertainty attributed to the choice of the POWHEG approach to perform the matching between the hard-scatter and the parton shower is estimated comparing the POWHEG +PYTHIA 8 prediction with the AMC@NLO +PYTHIA 8 simulation.

The predicted $t\bar{t}H$ cross-section uncertainty is $^{+5.8\%}_{-9.2\%}(\text{scale}) + 3.6\%(\text{PDF+}\alpha_S)$. The two uncertainty components are considered uncorrelated in this analysis.

Single top modelling The same method used for the $t\bar{t}H$ cases is implied to evaluate the modelling of ISR (where μ_R, μ_F and Var3c parameters are varied in the nominal simulation) and FSR in the three single top-quark processes: t-channel, tW and s-channel.

The uncertainty due to the choice of the hadronisation model and the other non-perturbative aspects of the parton shower is evaluated comparing the nominal prediction with that provided by the POWHEG+HERWIG 7.1.6 simulation.

An uncertainty is also attributed to the choice of the POWHEG approach to perform the matching between the hard-scatter and the parton shower is estimated comparing:

- For the tW, the nominal POWHEG+PYTHIA 8 prediction with the AMC@NLO+PYTHIA 8 simulation.
- For the t-channel and s-channel the POWHEG+HERWIG 7.1.6 and AMC@NLO+HERWIG 7.1.6 prediction.

An additional uncertainty is quoted for the tW-channel to estimate the difference between the diagram subtraction (DS) and diagram removal (DR) schemes used to deal with the overlap between the $\mathrm{t\bar{t}}$ and tW simulations. The nominal tW POWHEG+PYTHIA 8 DR simulation is compared to the tW POWHEG+PYTHIA 8 DS simulation.

Finally, a 5% uncertainty on the theoretical cross-section of single top-quark t-channel, tW and s-channel is evaluated in this analysis.

$\mathrm{t\bar{t}}W$ and $\mathrm{t\bar{t}}Z$ modelling: The comparison of SHERPA 2.2.10 (including EW corrections) with the nominal prediction is used to evaluate the $\mathrm{t\bar{t}}W$ modelling uncertainty. The predicted $\mathrm{t\bar{t}}W$ cross-section uncertainty is $^{+12.9\%}_{-11.5\%}(\text{scale})$ and $+3.4\%(\text{PDF+}\alpha_{\mathrm{S}})$. The two uncertainty components are considered uncorrelated in this analysis.

The nominal simulation of the $\mathrm{t\bar{t}}Z$ process is compared to the SHERPA prediction to evaluate the $\mathrm{t\bar{t}}Z$ modelling uncertainty. The predicted $\mathrm{t\bar{t}}Z$ cross-section uncertainty is $^{+9.6\%}_{-11.3\%}(\text{scale}) + 4\%(\text{PDF+}\alpha_{\mathrm{S}})$. The two uncertainty components are considered uncorrelated in this analysis.

Other background modelling: For minor backgrounds such as *Z+jets*, *W+jets*, *tZq*, *tWZ*, diboson, rare top-quark processes (three and four top-quark production) and other Higgs production (VH, VBF and ggF) only the theoretical uncertainty on the predicted cross-section is considered.

A 35% and 40% uncertainty is applied to the *Z+jets* and *W+jets* predicted cross-sections, respectively.

The predicted *tZq* cross-section uncertainty is $^{+7.7\%}_{-7.9\%}(\text{scale}) + 0.9\%(\text{PDF+}\alpha_{\mathrm{S}})$. The two uncertainty components are considered uncorrelated in this analysis.A 50% uncertainty is applied to the *tWZ* cross-section prediction.

The uncertainty on the diboson predicted cross-section is 24.5 %, which corresponds to the quadratic sum of 5 % from 0 b-jet uncertainty and 24 % from 1 b-jet uncertainty [164].

A 50% uncertainty is considered for rare top-quark processes and other Higgs production processes.

5.6 Results

The presence of *tHq* for 3ℓ and 2ℓSS channel is tested using a profile likelihood binned fit (PLBF) method in the SR and CRs defined in the tables 5.9 and 5.11, respectively using the TRExFitter package. The effect of systematic uncertainties on the signal and background expectations are introduced in the fit using nuisance parameters (NPs).

5.6.1 Profile likelihood binned fit

Firstly, the definition of the expected number of events in the *i*-th bin is needed before defining the fitted distribution. In a general case for both channels, it is given by:

$$\mathbb{E}_i(\mu, k_j, \theta) = \mu \, \mathscr{N}_{sig}(\theta) + \sum_{j=0}^{M} k_j \, \mathscr{N}_{bkg}(\theta)\,, \tag{5.1}$$

where $\mathscr{N}_{sig}$ and $\mathscr{N}_{bkg}$ are the expected number of events in each *i*-bin for signal and background, respectively, and M is the total number of background processes considered. The θ symbol is a given NP, and μ is the signal strength and k_j is the normalisation factor for a given background sample j. The value of k is equal to 1 for the majority of the backgrounds, and it is let to float only for specific background processes. For this analysis, k is floating for the $t\bar{t}W$ process and the three sources of fake leptons evaluated (e_{HF}, e_{conv} and μ_{HF}). Moreover, the agreement between the observed number of data events and the simulated data events for the different SM processes can be evaluated using two hypotheses: the Asimov hypothesis where μ is set to 1 and k_j is set to 1 for all the background processes, and the background-only hypothesis where μ is set to 0.

The likelihood expression is the product of Poisson distributions for each bin $\mathscr{P}(N, \mathbb{E}_i)$ in the SR and the CRs. Moreover, the NPs are included in the likelihood expression using a Gaussian distribution for each NP, $\mathscr{G}(\theta; \nu, \sigma)$, where the standard deviation $\sigma = 1$ and mean ν, with expected value $\nu = 0$. The NPs affect in a coherent way either $\mathscr{N}_{sig}$ or $\mathscr{N}_{bkg}$ through an interpolation or an extrapolation of three discrete values, where $\theta = 0$, it is the nominal value , and $\theta = \pm 1$, given by their $\pm 1\sigma$ variation. Finally, the likelihood expression is given by:

$$\mathscr{L}(N, \mu, k_j, \theta) = \prod_{\forall i} \mathscr{P}(N_i, \mathbb{E}_i) \times \prod_{\forall l} \mathscr{G}(\theta_l; \nu, \sigma)\,, \tag{5.2}$$

where N_i is the observed data of each bin, the i-index runs over all the bins in the SR and the CRs, and l-index runs over all the NPs. In short, the fit procedure becomes in a maximisation problem of a multidimensional likelihood expression whose results is a set of values for the parameters μ, k_j and θ_l. The results regarding the NPs could change in two different ways: its central value could be different from zero, which is named pulled, or its uncertainty could be lower than $\pm 1\sigma$, which is called constraint.

In addition to the signal strength, an upper limit is also provided by a one side statistical test. The test depends on the value of μ, and the upper limit is extracted using the CLs method [165] in order to set a 95% of confidence level (CL). The interpretation of the signal strength, as well as the upper limit, can be also written in term of the production cross-section of the signal process in the following way:

$$\sigma^{\text{obs}} = \mu \times \sigma^{\text{theory}}\,,$$

where σ^{theory} is the value of the production cross-section predicted by the theory.

5.6.2 Treatment of uncertainties

As mentioned before, the systematic uncertainties are included in the fit through NPs, which in turn are included in the likelihood expression 5.2 using Gaussian distributions. In general, there are two variations for each source of systematic named 'up' and 'down'. The uncertainties are symmetrised before including in the fit to avoid divergences in the maximisation process of the likelihood. In this case, a two-sided symmetrisation is applied when the two variations exist. Therefore, the variations are replaced by:

$$\text{NP}_{\text{sym}} = \frac{\text{up} - \text{down}}{2}\,,$$

where the NP_{sym} is computed bin-by-bin, and it is used as up/down variations around the nominal value.

In the case of uncertainties on the luminosity and the theoretical cross-section of the processes included in the analysis, they are computing with up/down variation directly and without considering the binning of the distribution used in the fit.

A pruning procedure is applied over the NPs in order to avoid instabilities in the fit due to the large number of NPs introduced. The goal of this procedure is to reject the NPs with negligible impact in the final results. The criteria followed to reject the NPs are the shape or the normalisation impact. In the first case, the shape impact consists of the maximum difference of bin entries between the nominal and the varied distribution, considering the normalisation is the same. In the second case, the normalisation impact is computed by integrating over the nominal distribution and the varied ones. In the current analysis, the thresholds are 0.5% for the shape impact and 1% for the normalisation impact in both the 3ℓ and the 2ℓSS channels.

There are another two procedures in order to avoid instabilities in the fit, but in this case due to the low statistics. The first one, if a systematic uncertainty for a sample has fewer simulated predicted events than $1 \cdot 10^{-4}$, the systematic is dropped for that samples. Moreover, if a systematic uncertainty for a sample in a region has low statistics and the variations is large, that region is excluded for this uncertainty and for the particular sample, e.g in figure 5.21 shows an excluded region. The second one is a smoothing procedure. It consists in a re-binning procedure of the distribution which merges adjacent bins until statistical uncertainty of the NP is lower than a tolerance, i.e. 0.08, and if after re-binning the number of variations (number of bins minus one) was larger than the maximum allowed, i.e. 4, the tolerance would be divided by a factor two and re-binning would re-start. In addition, the algorithm "353QH twice" [166] is used to avoid artificially flat uncertainties due to re-binning.

Finally, the normalisation impact of all the NPs regarding to the signal strength ($\mu(tHq)$) or the normalisation factors, i.e. $\mathrm{k}(\mathrm{t\bar{t}}W)$, $\mathrm{k}(\mathrm{e_{HF}})$, $\mathrm{k}(\mathrm{e_{conv}})$ and $\mathrm{k}(\mu_{\mathrm{HF}})$, are dropped to avoid overestimate the uncertainties of the results of the fit.

5.6.3 Asimov hypothesis

One of the usual hypotheses to check the agreement between observed data and MC simulation is the Asimov hypothesis. In this hypothesis, the signal strength and the

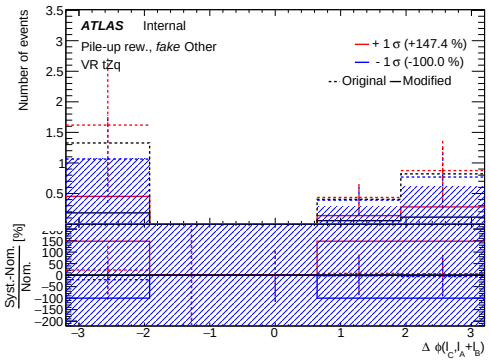


FIGURE 5.21: Example of excluded systematic, i.e. pile-up systematic, for the VR(*tZq*) and *fake Other* sample in the 3ℓ channel. The dashed lines represent the input values to the fit before any uncertainties treatment. The solid lines show the values after the any uncertainties treatment. The black lines correspond to the nominal samples, and the red and blue lines correspond to the up and down variations. The blue hatched area shows the statistical uncertainty.

background normalisation factors written in the formula 5.1 are equal to 1. Moreover, the NPs should take their nominal values ($\theta = 0$). In the end, the effects of all the NPs can be observed using this hypothesis, and if their variations directly affect to the final result.

The information given by this hypothesis is related to the uncertainties since μ and k_i are fixed. In other words, the effects in the results due to the different uncertainties can be differentiated and highlighted from the others. This fact allows to estimate the uncertainty contributions in the final results of the fit, and find out possible issues.

The pre-fit distribution of the variables used in the fit are shown in figures 5.22 and 5.23 for the 2ℓSS and the 3ℓ channel, respectively. Each distribution is considered in a different SR, CR or VR, and all of them show, in general, a good agreement between data and simulation distributions within the uncertainties. These regions correspond to the ones explained in section 5.4. The variables are selected according to the shape of the distribution for the main process that contributes to that region. The binning used in each distribution is optimised to reduce the uncertainty in each bin.

The NPs considered in each fit model are shown in figure 5.24 and 5.25 for the 2ℓSS and the 3ℓ channel, respectively. It is displayed how the uncertainty treatment,

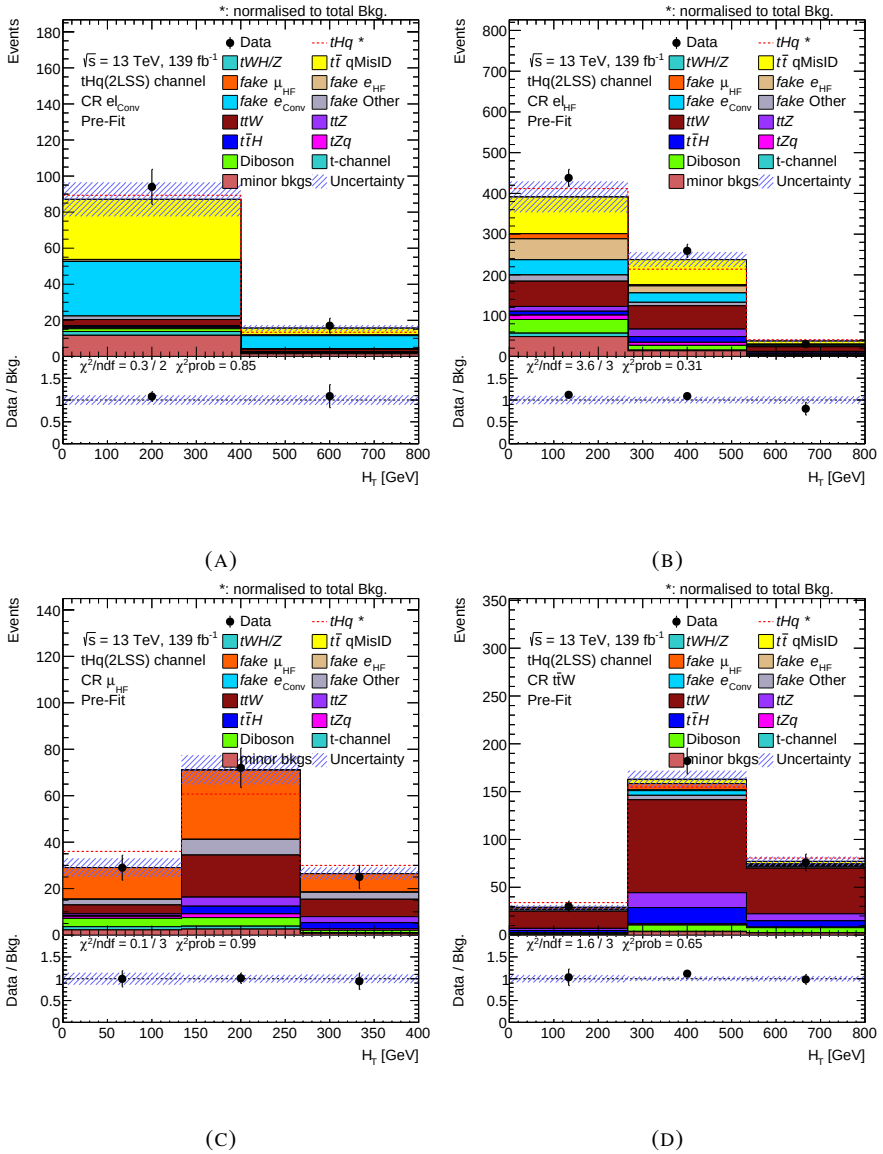
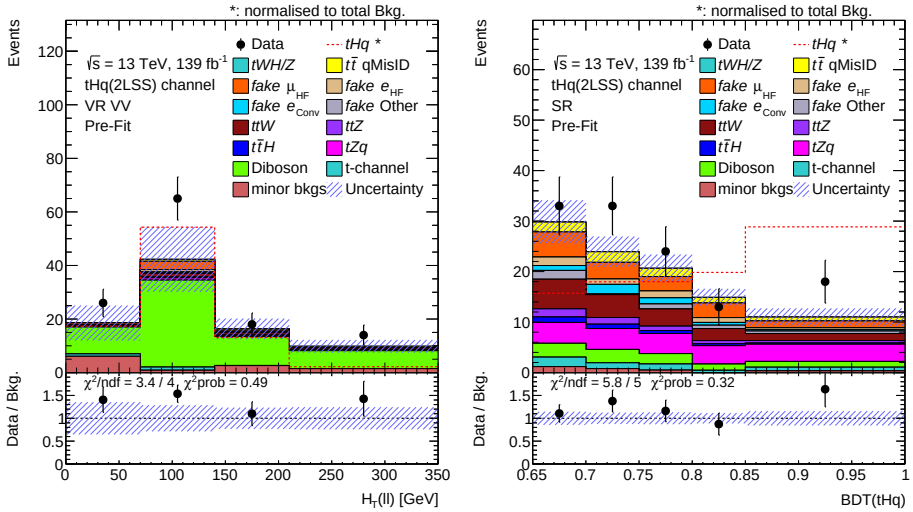


FIGURE 5.22: Pre-fit distributions in different CRs used in the PLBF for the 2ℓSS channel. The real and simulated data events are shown using the following distributions: (A) H_T in the CR(e_{conv}), (B) H_T in the CR(e_{HF}), (C) H_T in the CR(μ_{HF}), and (D) H_T in the CR($t\bar{t}W$). The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated background data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.



(E) (F)

FIGURE 5.22: Pre-fit distributions in the VR(Diboson) and in the SR used in the PLBF for the 2ℓSS channel. The real and simulated data events are shown using the following distributions: (E) $H_T(\ell\ell)$ in the VR(Diboson), (F) BDT(*tHq*) in the SR. The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated background data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.

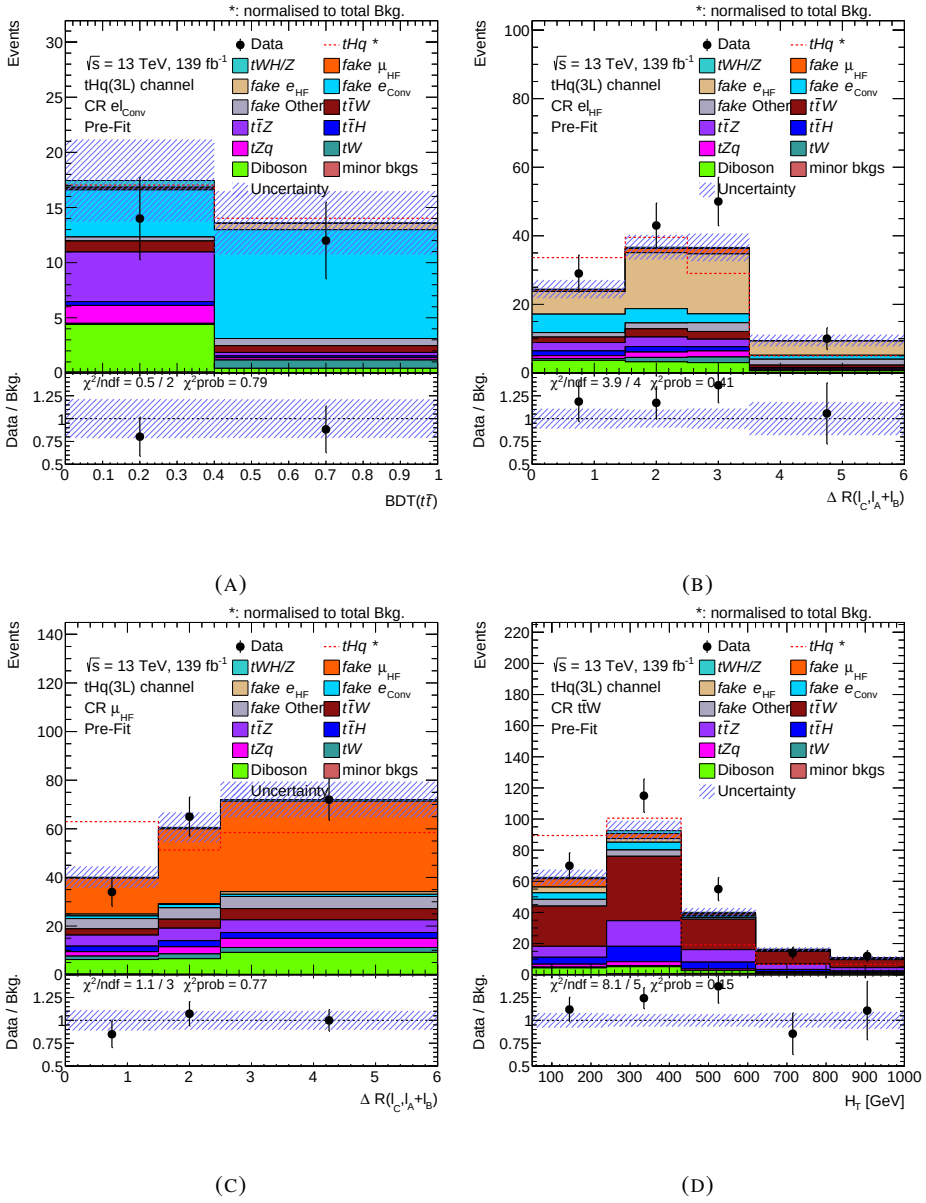


FIGURE 5.23: Pre-fit distributions in different CRs used in the PLBF for the 3ℓ channel. The real and simulated data events are shown using the following distributions: (A) BDT($t\bar{t}$) in the CR(e_{conv}), (B) $\Delta R(\ell_C, \ell_A + \ell_B)$ in the CR(e_{HF}), (C) $\Delta R(\ell_C, \ell_A + \ell_B)$ in the CR(μ_{HF}), and (D) H_T in the CR($t\bar{t}W$). The lower panels show the ratio between real and simulated background data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.

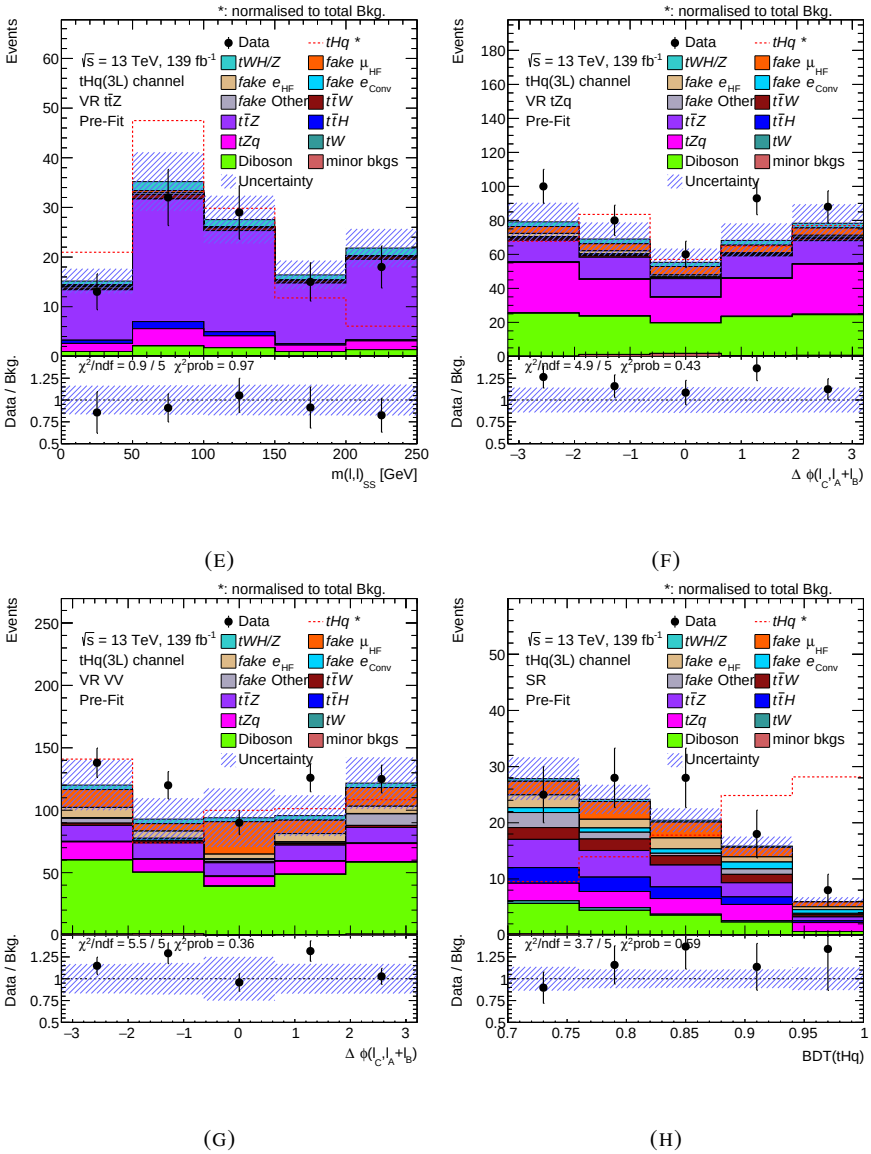


FIGURE 5.23: Pre-fit distributions in different VRs and in the SR used in the PLBF for the 3ℓ channel. The real and simulated data events are shown using the following distributions: (E) $\mathrm{m}(\ell\ell)_{\mathrm{SS}}$ in the VR($\mathrm{t\bar{t}}Z$), (F) $\Delta\phi(\ell_{\mathrm{C}}, \ell_{\mathrm{A}} + \ell_{\mathrm{B}})$ in the VR(tZq), (G) $\Delta\phi(\ell_{\mathrm{C}}, \ell_{\mathrm{A}} + \ell_{\mathrm{B}})$ in the VR(Diboson) , and (H) BDT(tHq) in the SR. The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated background data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.

which is explained in section 5.6.2, affects the different NPs before being considered in the fit. The NPs can be included in the fit or be partially or be completely dropped. There are some NPs that are not present because they have been removed following the criteria explained in section 5.6.2. The NPs which are kept for the fit also depend on the distributions shown in figures 5.22 and 5.23.

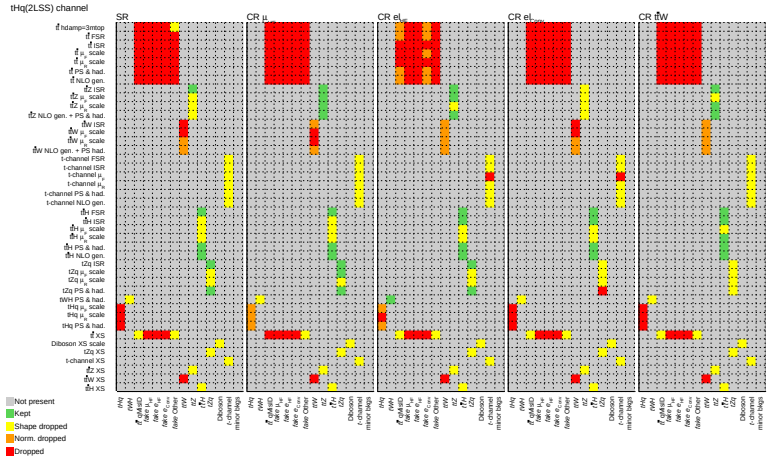
The NPs obtained from the fit are shown in figure 5.26 for both channels. In this hypothesis, the central values of the NPs should be zero and the constrained of the variation should be small. From the results of the fit, there are not significance constraints in any NP, and the NPs are not pulled as it assumed in the current hypothesis.

The values of $\mathrm{k_i}$ are shown in figure 5.27 for both the 2ℓSS and the 3ℓ channel. Their central values are equal to one since the prior assumption is the Asimov hypothesis. The uncertainties are similar in both channel except for $\mathrm{k}(\mathrm{e_{HF}})$. This different is explained by the different phase space of both channels even though the source of this background is the same.

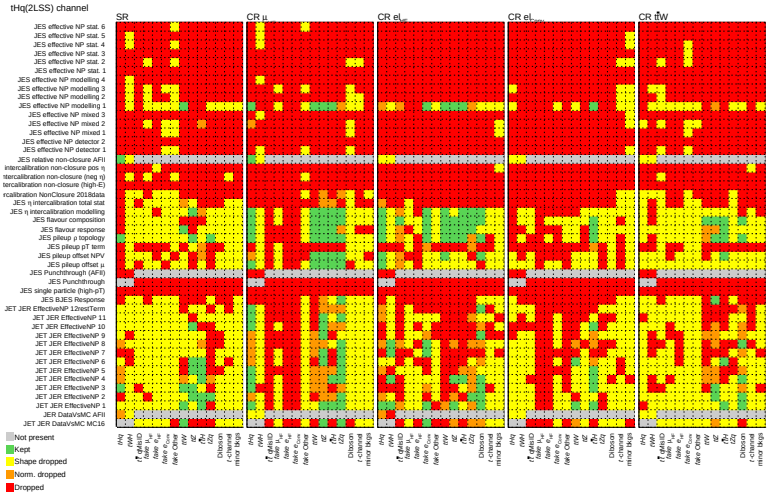
In figure 5.28, the $\mu(tHq)$ values for both channels are shown. Their central values are also 1, as expected. The uncertainties are split in total and statistical uncertainties from the data events. In both cases the statistical uncertainty is the main component of the total uncertainty (analysis statistically dominated). The impact of the different groups of systematic uncertainties on the $\mu(tHq)$ is shown in table 5.13. The NPs with the largest impact are those on the JES and JER.

In addition to the impact values are shown in table 5.13, while figure 5.29 shows the impact of the 20 NPs which are ranked by their impact on the result of the fit. Moreover, the value of the fitted NPs and their variations are also included.

In the case of the ranking of the NP for the 3ℓ channel 5.29, the highest parameter in the ranking corresponds to the fourth bin of the SR (γ (SRBDTtHq3L bin 4)). Even though, the impact of γ (SRBDTtHq3L bin 4) is not high, roughly ± 1.2, it is the first ranked. This fact is due to the yield composition of this bin, shown in table 5.14. The main background is the fake/non-prompt leptons, i.e. 40%, and in addition, their relative uncertainties are between 40% and the 90%. Moreover, there are three normalisation factors related to the fake/non-prompt leptons what causes that the yields of fake/non-prompt predicted by the MC samples have a high impact on the fit.

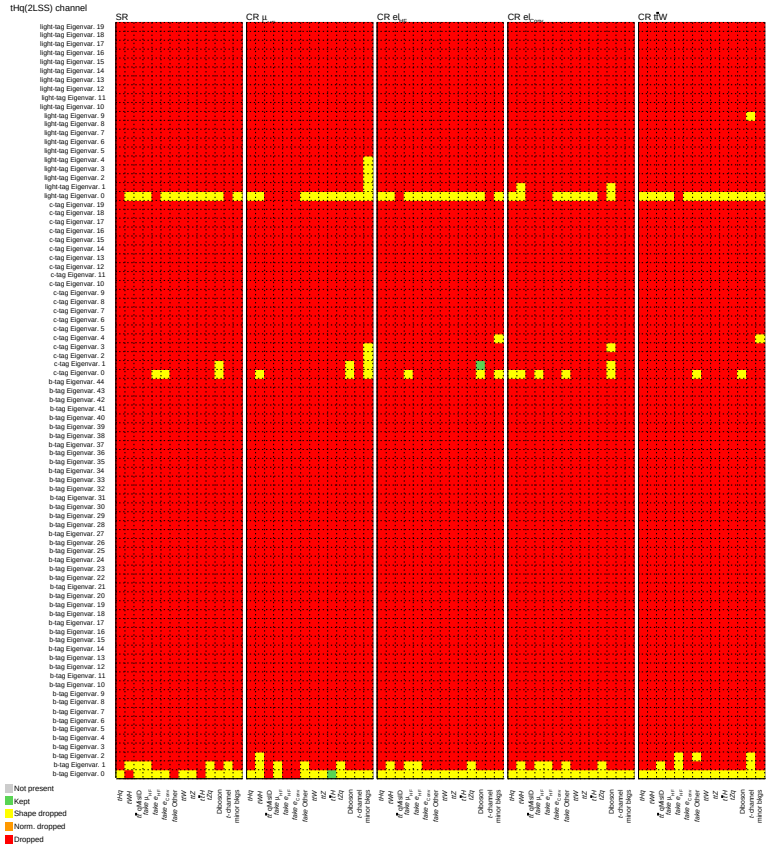


(A)



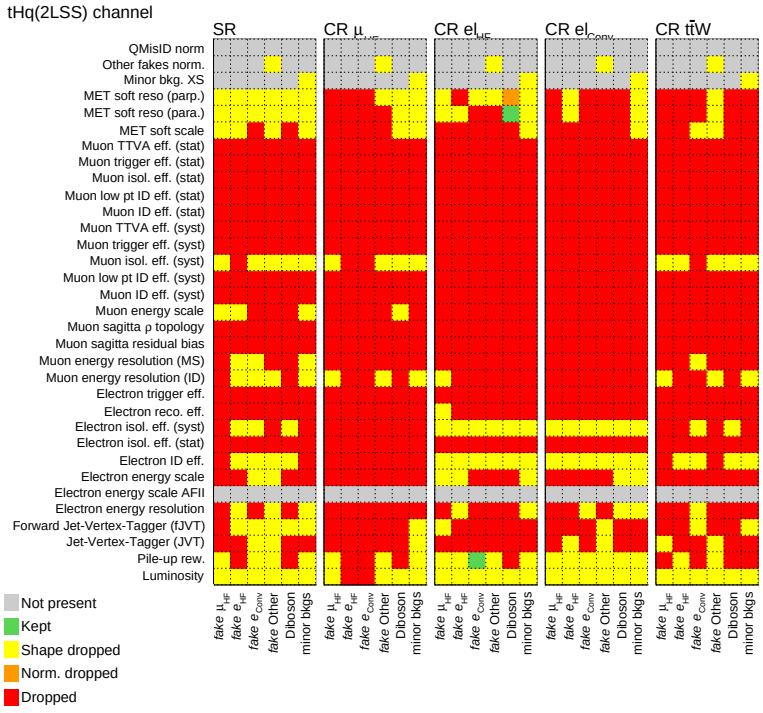
(B)

FIGURE 5.24: List of NPs included in PLBF as input in the 2ℓSS channel. It is shown with a colour code how the uncertainty treatment, which is explained in section 5.6.2, affects the different NPs before being considered in the fit model. Additionally, the list of NPs is split by regions.



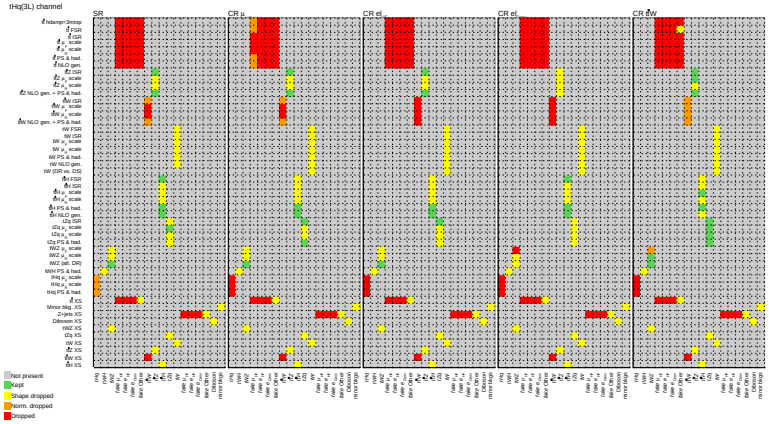
(c)

FIGURE 5.24: List of NPs included in PLBF as input in the 2ℓSS channel. It is shown with a colour code how the uncertainty treatment, which is explained in section 5.6.2, affects the different NPs before being considered in the fit model. Additionally, the list of NPs is split by regions.

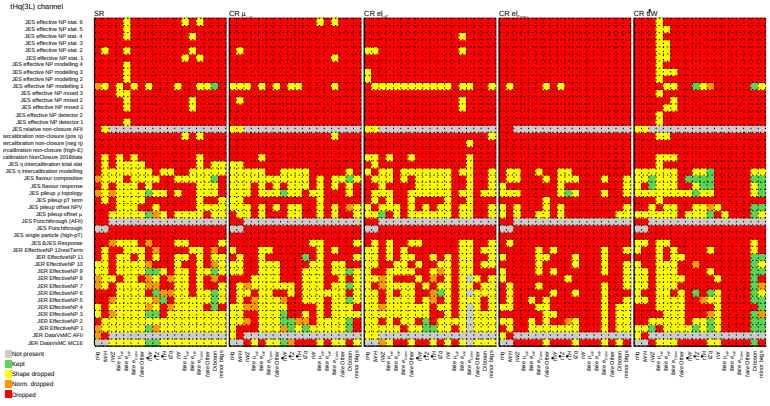


(D)

FIGURE 5.24: List of NPs included in PLBF as input in the 2ℓSS channel. It is shown with a colour code how the uncertainty treatment, which is explained in section 5.6.2, affects the different NPs before being considered in the fit model. Additionally, the list of NPs is split by regions.

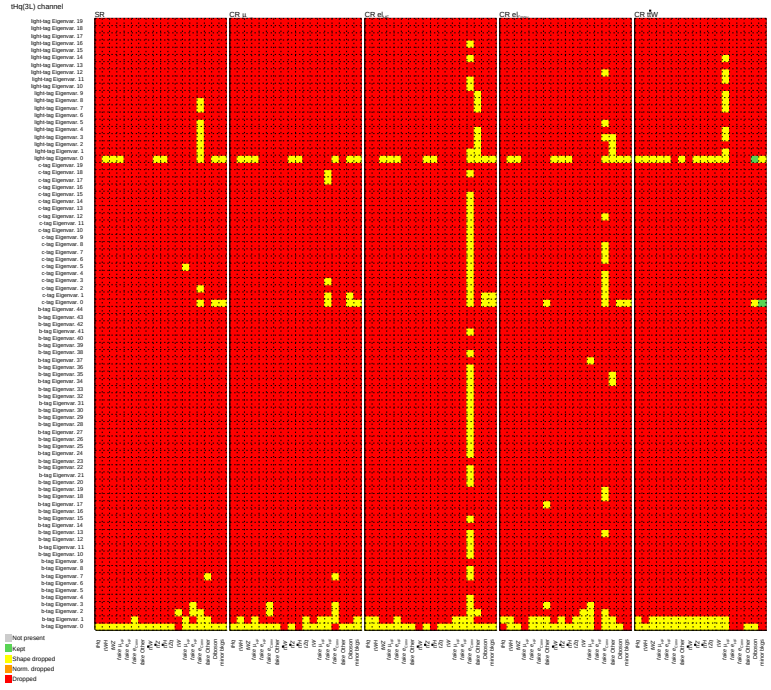


(A)



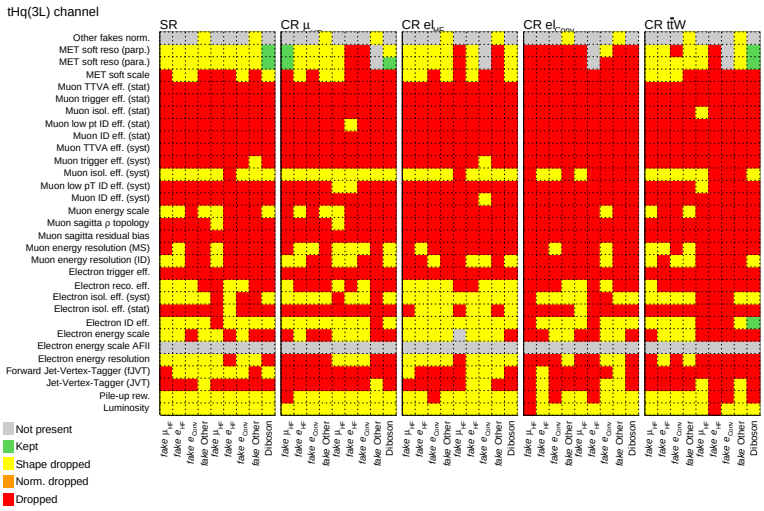
(B)

FIGURE 5.25: List of NPs included in PLBF as input in the 3ℓ channel. It is shown with a colour code how the uncertainty treatment, which is explained in section 5.6.2, affects the different NPs before being considered in the fit model. Additionally, the list of NPs is split by regions.



(c)

FIGURE 5.25: List of NPs included in PLBF as input in the 3ℓ channel. It is shown with a colour code how the uncertainty treatment, which is explained in section 5.6.2, affects the different NPs before being considered in the fit model. Additionally, the list of NPs is split by regions.

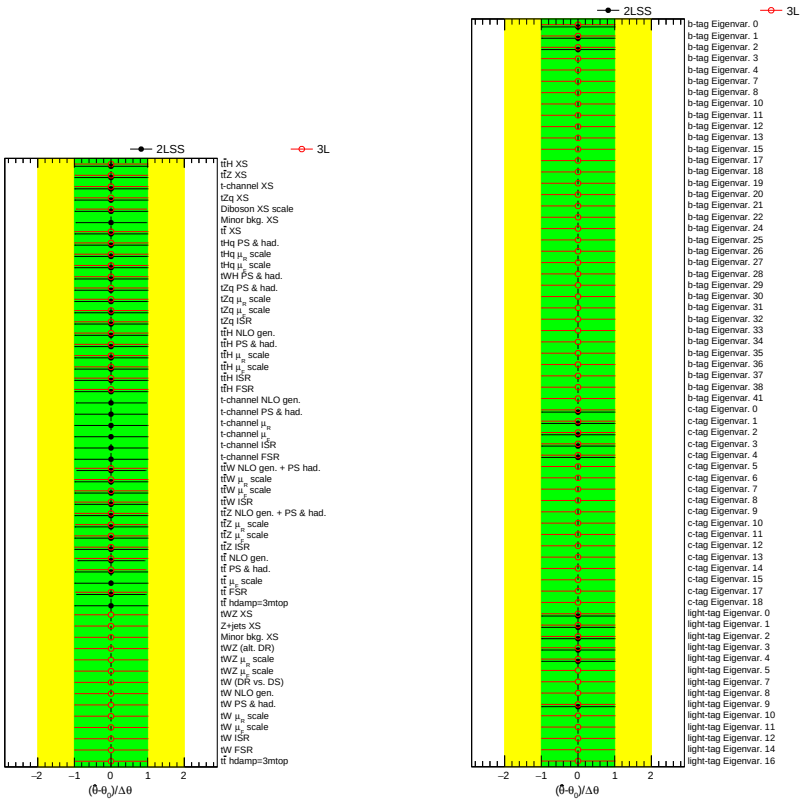


(D)

FIGURE 5.25: List of NPs included in PLBF as input in the 3ℓ channel. It is shown with a colour code how the uncertainty treatment, which is explained in section 5.6.2, affects the different NPs before being considered in the fit model. Additionally, the list of NPs is split by regions.

TABLE 5.13: Systematic uncertainties in the measurement of $\mu(tHq)$ for both the 2ℓSS and the 3ℓ channel in the Asimov hypothesis. The impact of each group of uncertainties is computed by performing a fit where the NPs in the group are fixed to their best-fit values, and then subtracting the resulting uncertainty on the $\mu(tHq)$ in quadrature from the nominal fit.

Uncertainty source	2ℓSS	3ℓ
Modelling		
Theoretical uncertainties	$\pm$ 1.307	$\pm$ 1.17
Experimental		
Jet energy scale/resolution	$\pm$ 1.798	$\pm$ 1.64
Jet flavour tagging	$\pm$ 0.087	$\pm$ 0.13
Mis-identified lepton	$\pm$ 0.072	$\pm$ 0.10
Other experimental uncertainties	$\pm$ 0.233	$\pm$ 0.69
Simulation statistics	$\pm$ 0.865	$\pm$ 1.31
Normalisation factors	$\pm$ 1.241	$\pm$ 0.56
Total systematic uncertainty	$\pm$ 2.906	$\pm$ 2.77



(A) (B)

FIGURE 5.26: Post-fit NPs obtained from the fit under the Asimov hypothesis for the 2ℓSS channel in black filled circles and the 3ℓ channel in red empty circles. Each NP is shown as the relative change from its nominal value. The green and yellow areas represent the $\pm1\sigma$ and $\pm2\sigma$ deviations from the nominal value of the NP, respectively. The points represent the best-fit value for the NP and the uncertainty bars represent the post-fit uncertainty.

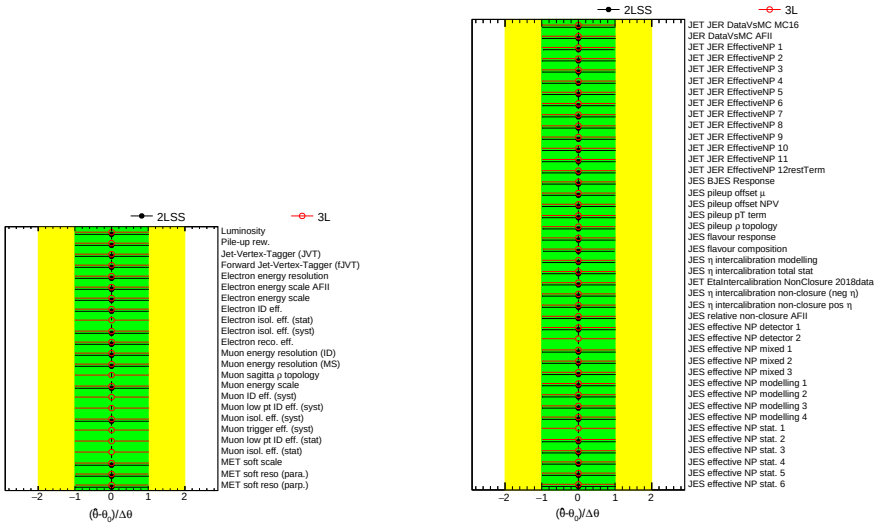


FIGURE 5.26: Post-fit NPs obtained from the fit under the Asimov hypothesis for the 2ℓSS channel in black filled circles and the 3ℓ channel in red empty circles. Each NP is shown as the relative change from its nominal value. The green and yellow areas represent the $\pm1\sigma$ and $\pm2\sigma$ deviations from the nominal value of the NP, respectively. The points represent the best-fit value for the NP and the uncertainty bars represent the post-fit uncertainty.

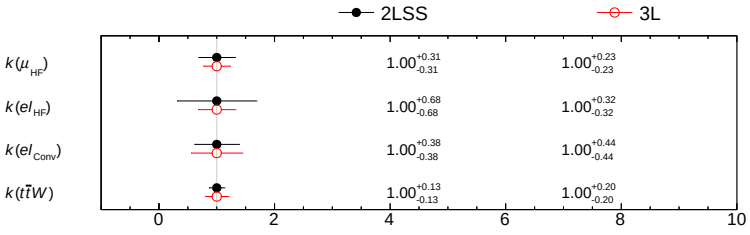


FIGURE 5.27: Normalisation factors, i.e. $k(\mu_{HF})$, $k(e_{HF})$, $k(e_{conv})$ and $k(t\bar{t}W)$, in the Asimov hypothesis for the 2ℓSS channel in black filled circles and the 3ℓ channel in red empty circles. The uncertainties include statistical and all the systematic source.

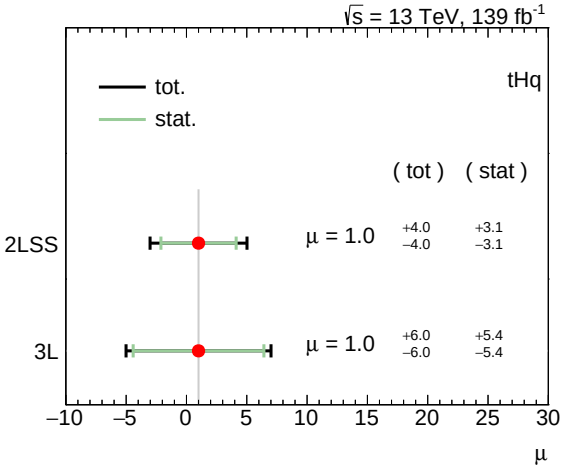
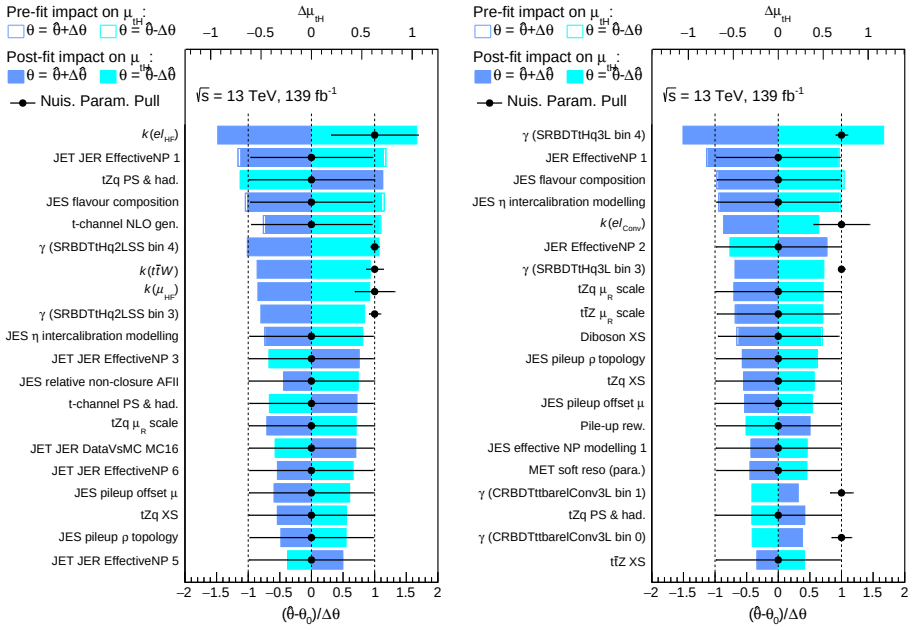


FIGURE 5.28: Signal strength values for the for the 2ℓSS and the 3ℓ channel in the Asimov hypothesis. The total uncertainty (tot) includes statistical and systematic effects. Additionally, the statistical uncertainty (stat) is also shown.

TABLE 5.14: Event yields predicted by the MC simulation and data in the fourth bin of the SR for the 3ℓ channel. The uncertainty includes statistical and all the systematic sources.

Process	Yields
tHq	0.384 ± 0.032
tWH	0.019 ± 0.014
tWZ	0.043 ± 0.032
Fake μ_{HF}	0.9 ± 0.6
Fake e_{HF}	0.51 ± 0.33
Fake e_{Conv}	0.69 ± 0.3
Fake Other	0.27 ± 0.25
$t\bar{t}W$	0.27 ± 0.13
$t\bar{t}Z$	0.7 ± 0.24
$t\bar{t}H$	0.31 ± 0.07
tZq	1.59 ± 0.29
Single top tW	0.0 ± 0.0
Diboson	0.59 ± 0.25
Minor backgrounds	0.04 ± 0.04
Total background	6.0 ± 1.0
Data	8



(A) 2ℓSS (B) 3ℓ

FIGURE 5.29: Ranking of the NPs sorted by their impacts on $\mu(tHq)$ for the 2ℓSS channel (A) and the 3ℓ channel (B). The blue boxes refer to the upper x-axis and show the impact on $\mu(tHq)$. This impact is preformed fixing to one the specific NP and varying the others upwards or downwards by its pre-fit, in the case of pre-fit impact, or post-fit, in the case of post-fit impact, uncertainty to compute the fit in each configuration. Then, the value of the impact is the $\mu(tHq)$ obtained in each of these four configurations minus the $\mu(tHq)$ obtained in the nominal fit. Moreover, the NPs values and their uncertainties are also included as dots and lines, respectively. They refer to the lower x-axis and are also shown in figure 5.26.

The correlation matrices of the NPs and normalisation factors are presented in figures 5.30 and 5.31 for the 2ℓSS and the 3ℓ channel, respectively. From these figures any high correlation has been measured, in the case of the 2ℓSS channel the highest correlation is $-43.6\,\%$, and it is $-41.4\,\%$ for the 3ℓ channel.

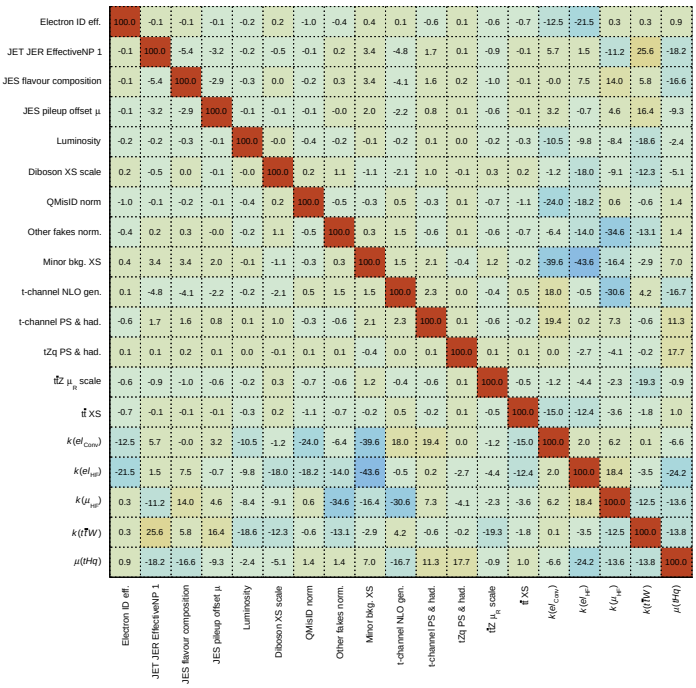


FIGURE 5.30: Correlation matrix of the NPs and normalisation factors for the 2ℓSS channels in the Asimov hypothesis. Only the correlations of the NPs and normalisation factors with at least one correlation greater than 15% are shown.

All the results from the Asimov hypothesis are presented. They show a stable fit without any spurious results either in the NPs, the normalisation factors, or in the signal strength for the 2ℓSS and the 3ℓ channels. Once the fit model in the Asimov hypothesis is well established and it is understood, the following step is to solve the likelihood fit equation without fixing any parameters to obtain the final results.

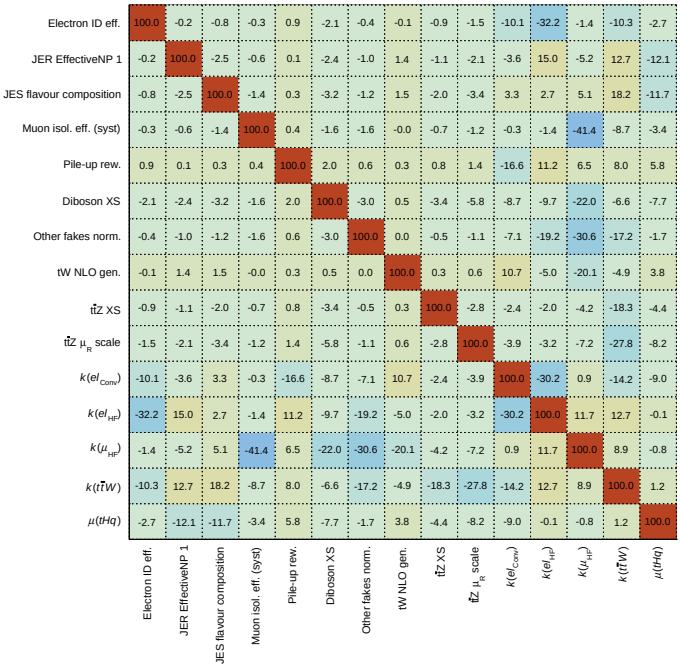


FIGURE 5.31: Correlation matrix of the NPs and normalisation factors for the 3ℓ channels in the Asimov hypothesis. Only the correlations of the NPs and normalisation factors with at least one correlation greater than 15% are shown.

5.6.4 Final results of profile likelihood fit

In this case, the signal strength, the normalisation factors, and the NPs in equation 5.2 do not have any ad-hoc condition. That means their mean values could be different from one.

The distribution of the variables used in the fit, at pre-fit level, are the same as in the Asimov hypothesis since they are the input distribution to the fit. Figures 5.32 and 5.33 show the distribution of the same variables at post-fit level. In this case, the simulated samples have been scaled using the results of the fit. The uncertainty bands have been also changed by the fitted NPs. The results from the fit have been independently obtained for each channel from the CRs and SR, and they are applied in these regions and in the VRs. The value of the probabilistic χ^2 increases or keeps similar between the pre-fit and post-fit level what means that the results of the fit are consistent since they improve the agreement between simulated and data samples.

The set of NPs involves in this fit is identical to the one in the Asimov hypothesis, and they are affected in the same way by the uncertainties treatment explained in section 5.6.2. The different ways in which the NPs are considered in the fit are shown in figures 5.24 and 5.25.

The fitted NPs are shown in figure 5.34 for both channels. In this case some of the NPs are not centred in 1, they are slightly pulled and all of them are inside the $\pm 1\sigma$ deviation. The uncertainties for these NPs are also pulled in a coherent way with respect the central value, and some of them are slightly constrained. From these results, one can conclude that there are not significance pull nor constraint in any NPs.

The impact of the NPs in different groups of systematics is shown in table 5.15. The NPs with the largest impact are those related to the JES and JER, as in the Asimov hypothesis. In addition to the table, figure 5.35 shows the impact of the twenty highest ranked NPs. They are ranked by their impact on the μ of the fit and the values of NPs are also included.

In the case of the 3ℓ channel, the first position in the ranking is the fourth bin of the SR. This fact is also observed in the Asimov hypothesis, and it is due to the same reason that has been discussed in the previous section.

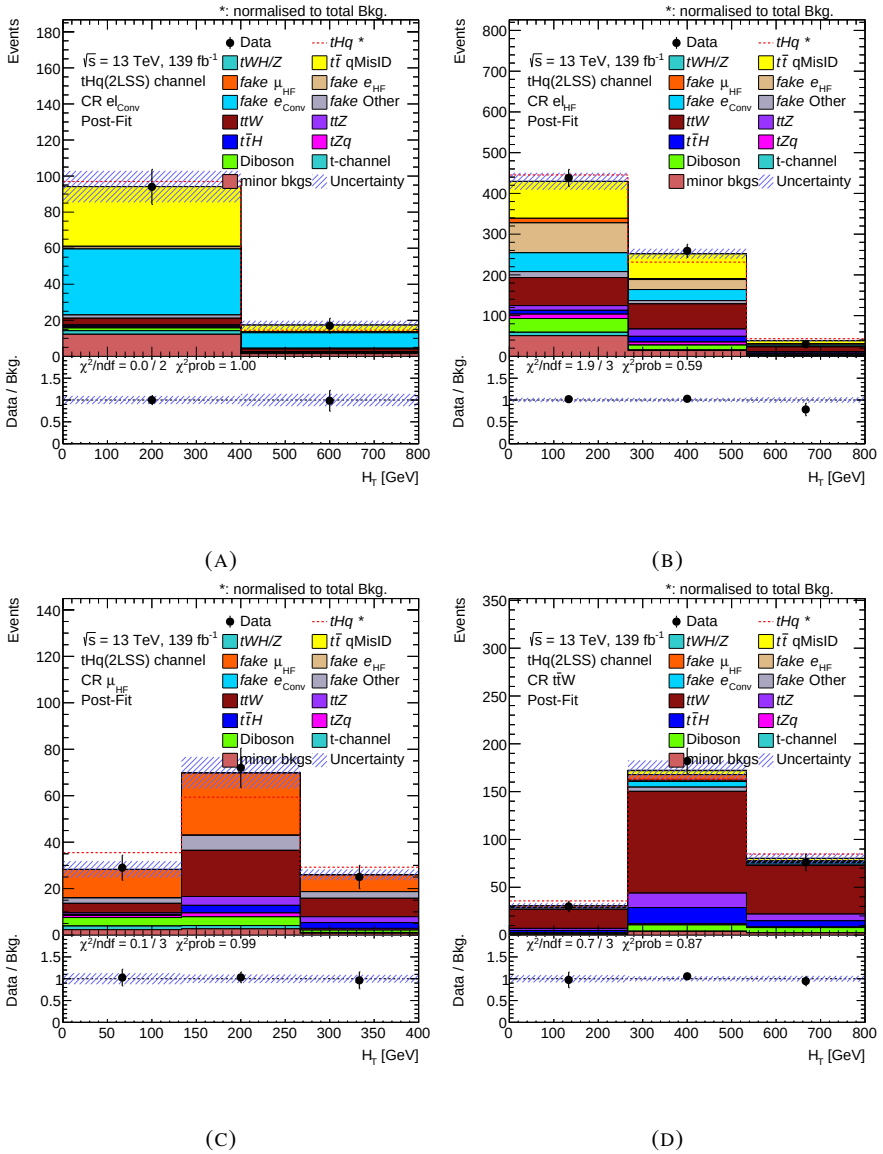
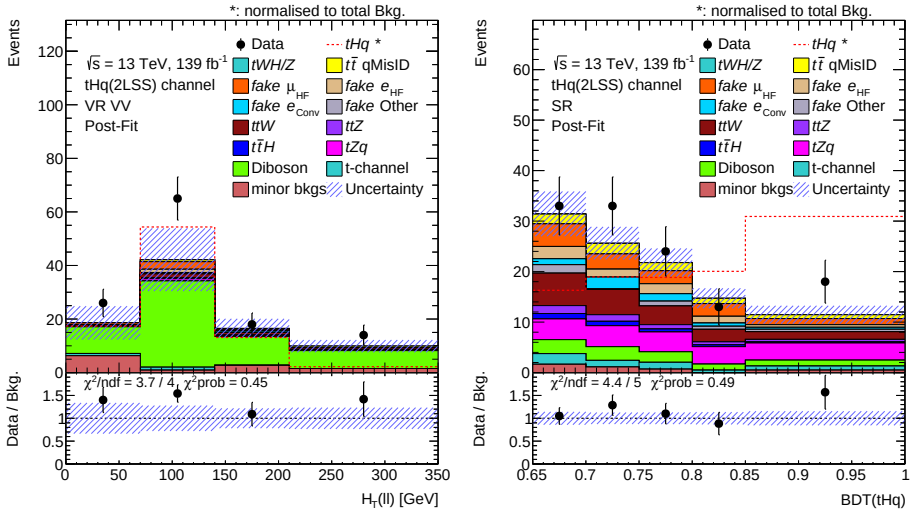


FIGURE 5.32: Post-fit distributions in different CRs used in the PLBF for the 2ℓSS channel. The real and simulated data events are shown using the following distributions: (A) H_T in the CR(e_{conv}), (B) H_T in the CR(e_{HF}), (C) H_T in the CR(μ_{HF}), and (D) H_T in the CR($t\bar{t}W$). The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated background data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.



(E) (F)

FIGURE 5.32: Post-fit distributions in the VR(Diboson) and in the SR used in the PLBF for the 2ℓSS channel. The real and simulated data events are shown using the following distributions: (E) $H_T(\ell\ell)$ in the VR(Diboson), (F) BDT(*tHq*) in the SR. The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated background data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.

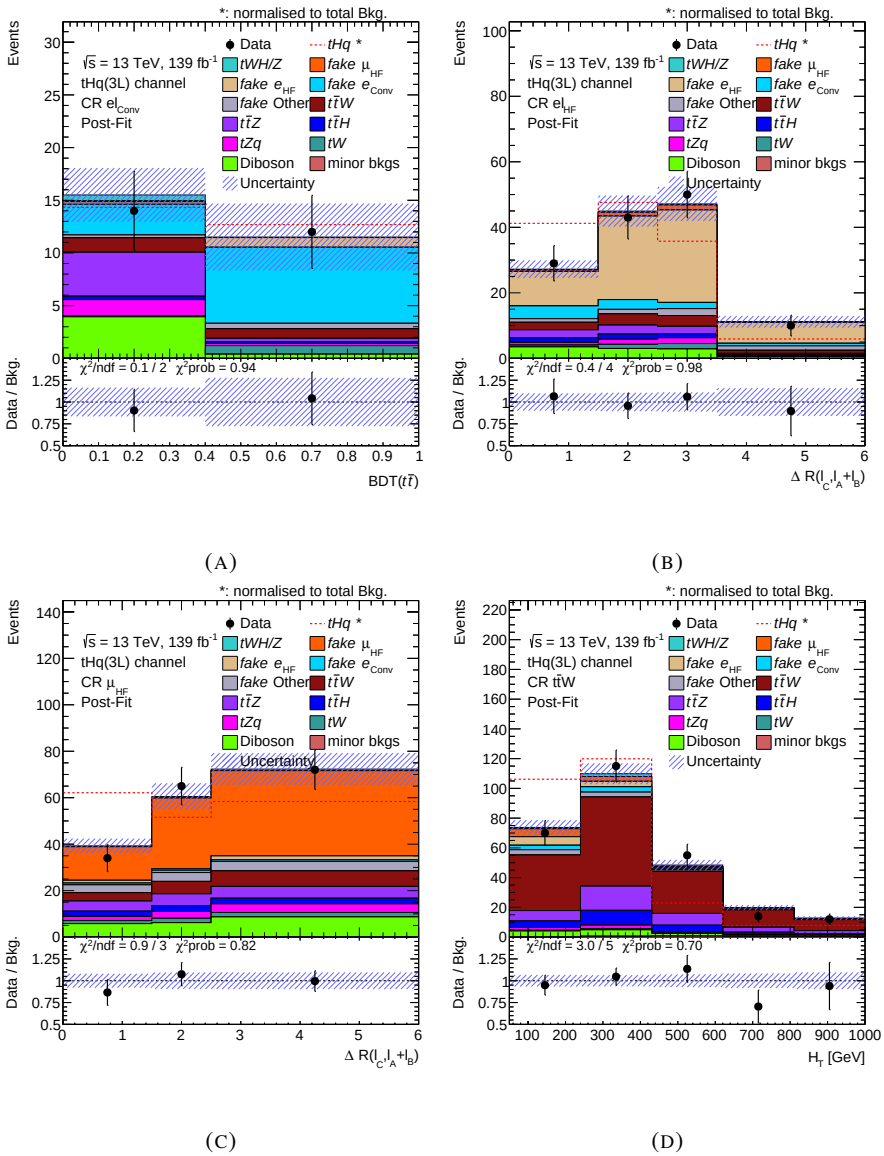


FIGURE 5.33: Post-fit distributions in different CRs used in the PLBF for the 3ℓ channel. The real and simulated data events are shown using the following distributions: (A) BDT($t\bar{t}$) in the CR(e_{conv}), (B) $\Delta R(\ell_C, \ell_A + \ell_B)$ in the CR(e_{HF}), (C) $\Delta R(\ell_C, \ell_A + \ell_B)$ in the CR(μ_{HF}), and (D) H_T in the CR($t\bar{t}W$). The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated background data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.

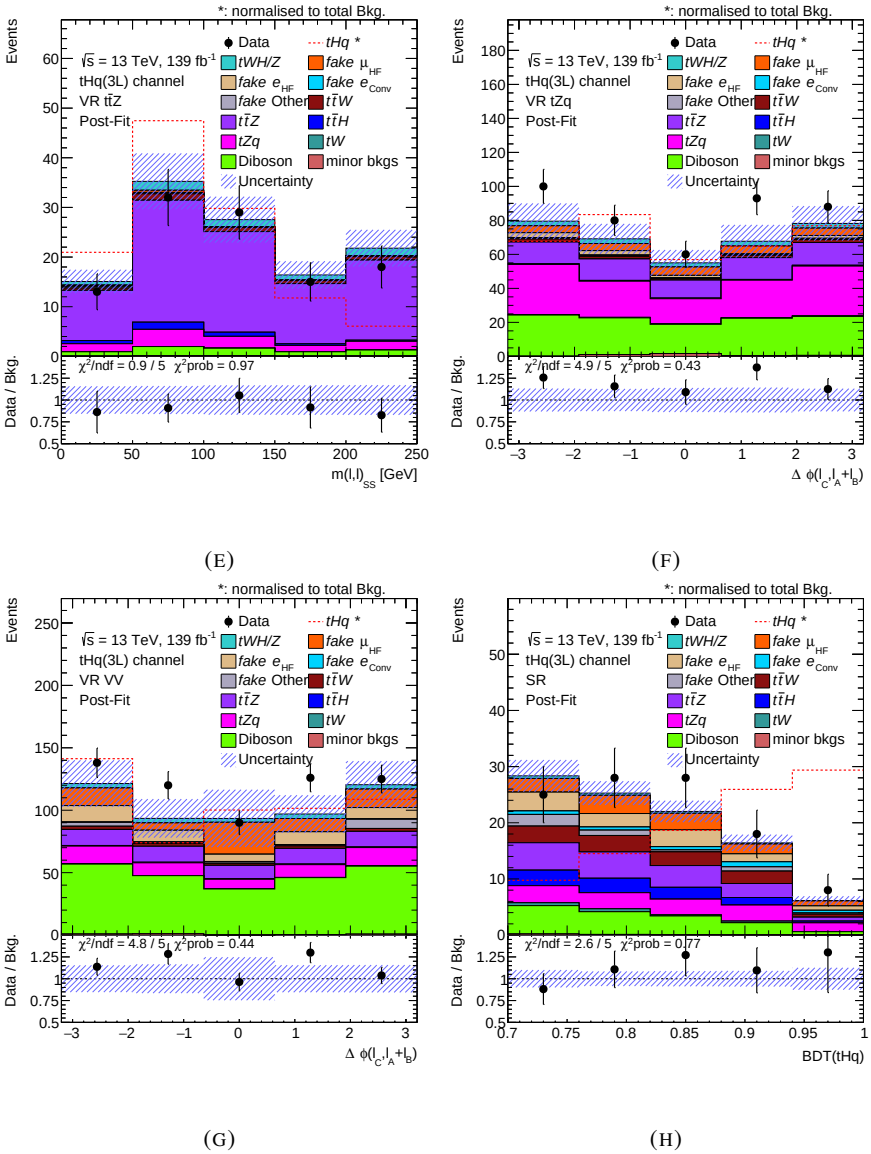
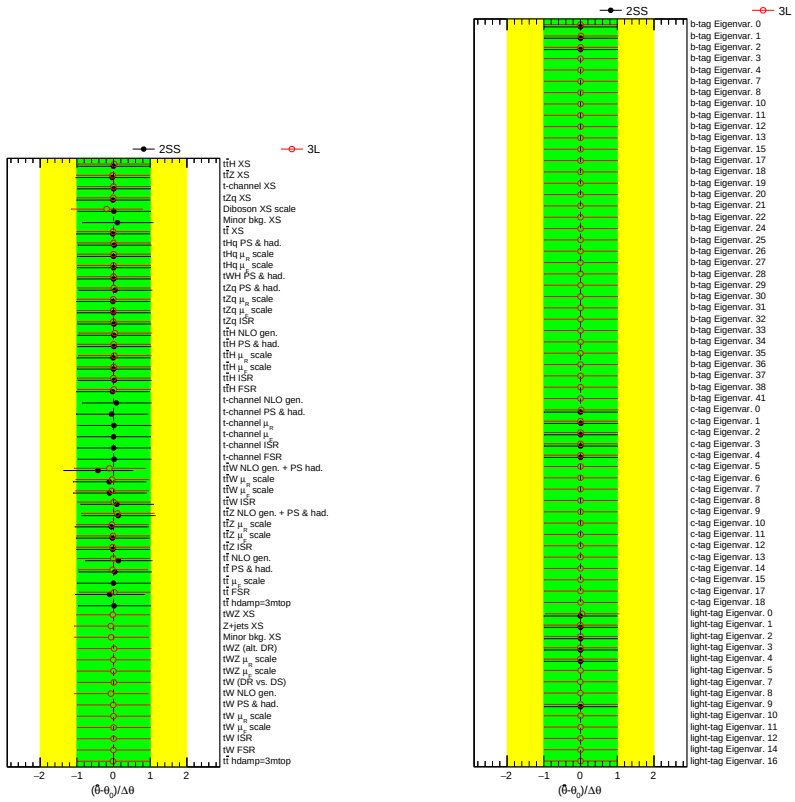


FIGURE 5.33: Post-fit distributions in different VRs and in the SR used in the PLBF for the 3ℓ channel. The real and simulated data events are shown using the following distributions: (E) $\mathrm{m}(\ell\ell)_{\mathrm{SS}}$ in the VR($\mathrm{t\bar{t}}Z$), (F) $\Delta\phi(\ell_{\mathrm{C}},\ell_{\mathrm{A}}+\ell_{\mathrm{B}})$ in the VR(tZq), (G) $\Delta\phi(\ell_{\mathrm{C}},\ell_{\mathrm{A}}+\ell_{\mathrm{B}})$ in the VR(Diboson) , and (H) BDT(tHq) in the SR. The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated background data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.



(I) (J)

FIGURE 5.33: Post-fit NPs obtained from the fit for the 2ℓSS channel in black filled circles and the 3ℓ channel in red empty circles. Each NP is shown as the relative change from its nominal value. The green and yellow areas represent the $\pm1\sigma$ and $\pm2\sigma$ deviations from the nominal value of the NP, respectively. The points represent the best-fit value for the NPs and the uncertainty bars represent the post-fit uncertainty.

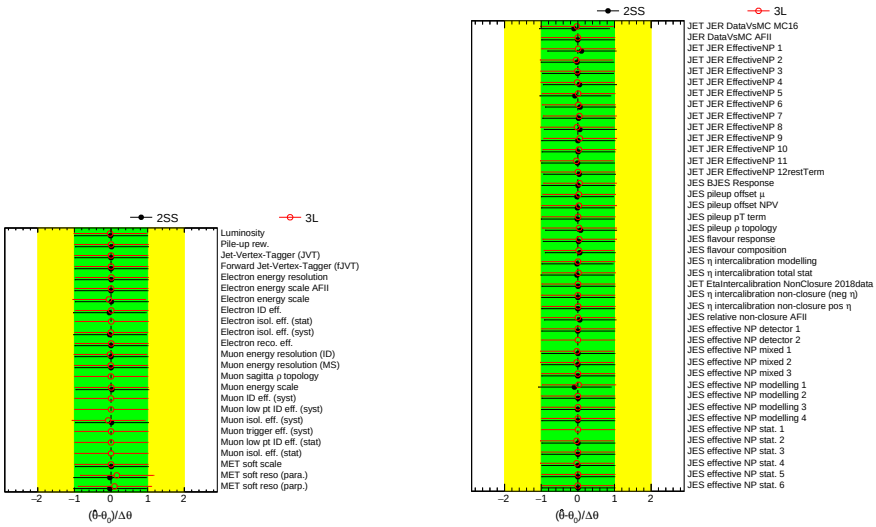
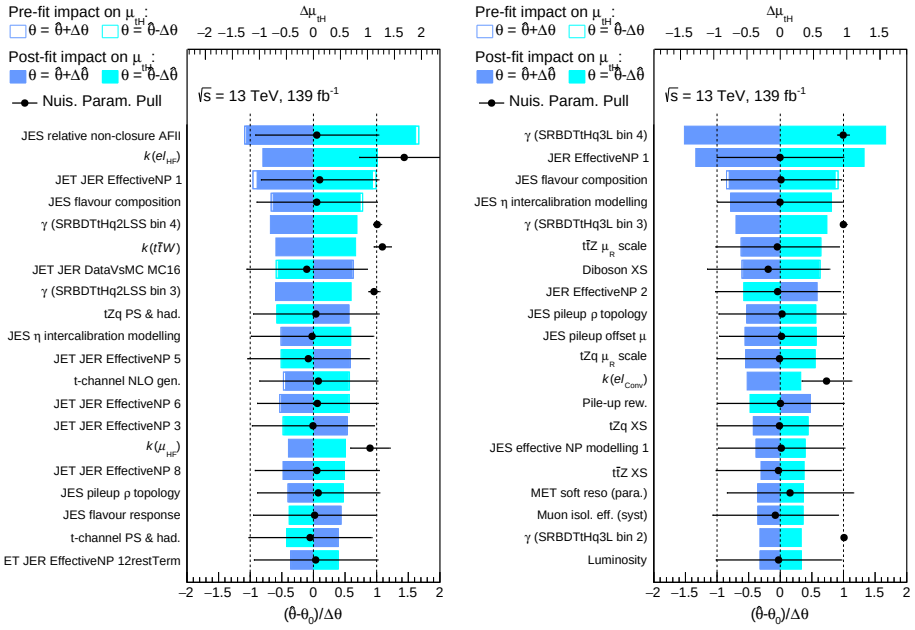


FIGURE 5.34: Post-fit NPs obtained from the fit for the 2ℓSS channel in black filled circles and the 3ℓ channel in red empty circles. Each NP is shown as the relative change from its nominal value. The green and yellow areas represent the $\pm 1\sigma$ and $\pm 2\sigma$ deviations from the nominal value of the NPs, respectively. The points represent the best-fit value for the NPs and the uncertainty bars represent the post-fit uncertainty.

TABLE 5.15: Systematic uncertainties in the measurement of $\mu(tHq)$ for the 2ℓSS and the 3ℓ channel. The impact of each group of uncertainties is computed by performing a fit where the NPs in the group are fixed to their best-fit values, and then subtracting the resulting uncertainty on the $\mu(tHq)$ in quadrature from the nominal fit.

Uncertainty source	2ℓSS	3ℓ
Modelling		
Theoretical uncertainties	$\pm$ 1.23	$\pm$ 1.39
Experimental		
Jet energy scale/resolution	$\pm$ 2.76	$\pm$ 2.1
Jet flavour tagging	$\pm$ 0.14	$\pm$ 0.20
Mis-identified lepton	$\pm$ 0.12	$\pm$ 0.20
Other experimental uncertainties	$\pm$ 0.39	$\pm$ 0.95
Simulation statistics	$\pm$ 1.08	$\pm$ 1.76
Normalisation factors	$\pm$ 1.38	$\pm$ 0.92
Total systematic uncertainty	$\pm$ 3.68	$\pm$ 3.76



(A) 2ℓSS (B) 3ℓ

FIGURE 5.35: Ranking of the NPs sorted by their impacts on $\mu(tHq)$ for the 2ℓSS channel (A) and the 3ℓ channel (B). The blue boxes refer to the upper x-axis and show the impact on $\mu(tHq)$. This impact is preformed fixing to one the specific NP and varying the others upwards or downwards by its pre-fit, in the case of pre-fit impact, or post-fit, in the case of post-fit impact, uncertainty to compute the fit in each configuration. Then, the value of the impact is the $\mu(tHq)$ obtained in each of these four configurations minus the $\mu(tHq)$ obtained in the nominal fit. Moreover, the NPs and their uncertainties values are also included as dots and lines, respectively. They refer to the lower x-axis and are also shown in figure 5.34.

The correlation matrices for all the NPs are shown in figures 5.36 and 5.37 for the 2ℓSS and 3ℓ channel, respectively. There are no shown of high correlations, in the case of the 2ℓSS channel the highest values is –37.5 %, and for the 3ℓ channel it is –40.3 %.

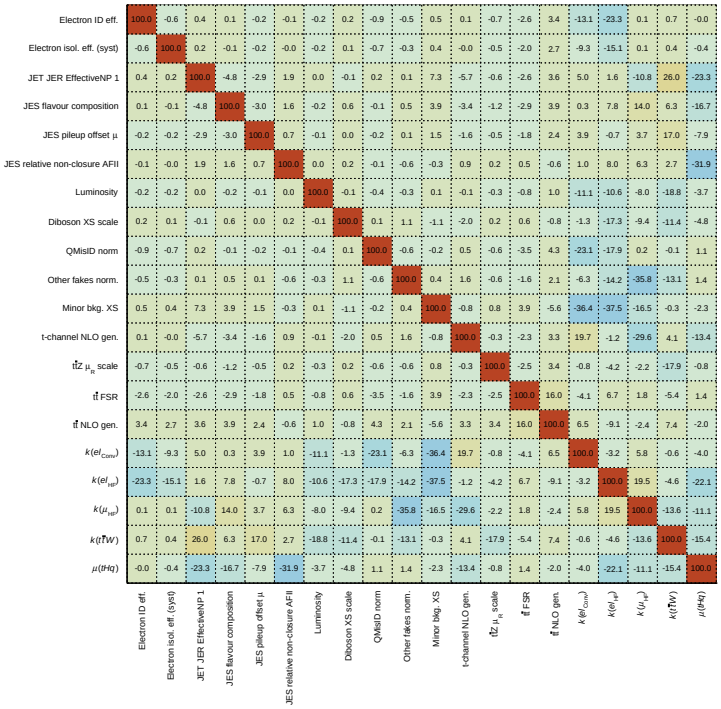


FIGURE 5.36: Correlation matrix of the NPs and normalisation factors for the 2ℓSS channels. Only the correlations of the NPs and normalisation factors with at least one correlation greater than 15% are shown.

The normalisation factors, i.e $\mathrm{k}(\mu_{\mathrm{HF}})$, $\mathrm{k}(\mathrm{e}_{\mathrm{HF}})$, $\mathrm{k}(\mathrm{e}_{\mathrm{conv}})$, and $\mathrm{k}(\mathrm{t\bar{t}}W)$, are shown in figure 5.38 for both channels. They are compatible within their uncertainties, and with the SM in at least one of the channels.

In the case of $\mathrm{k}(\mathrm{t\bar{t}}W)$, the values for both channels are in the limit of the compatibility within their uncertainties. This difference is due to the different phase space for the CR($\mathrm{t\bar{t}}W$) in both channels. Considering the distributions of number of central jets for both channels at pre-fit level, in figure 5.39, the number of central jets is lower in the

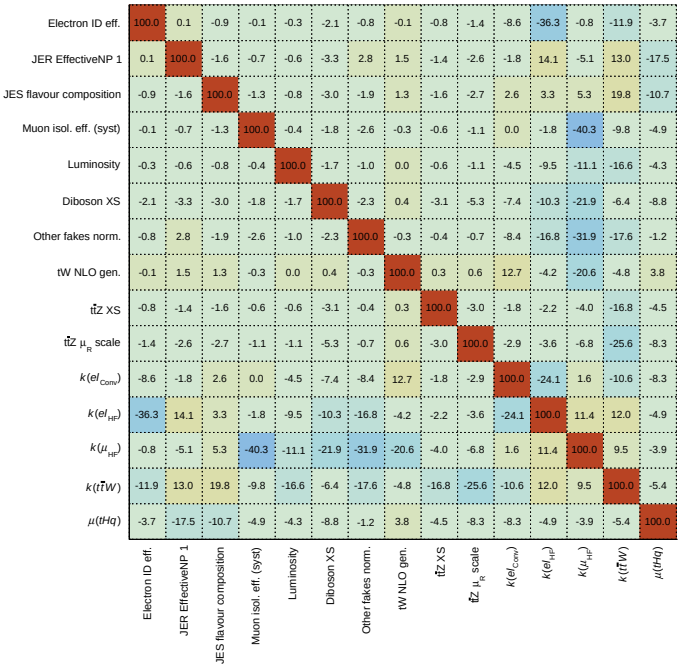


FIGURE 5.37: Correlation matrix of the NPs and normalisation factors for the 3ℓ channels. Only the correlations of the NPs and normalisation factors with at least one correlation greater than 15% are shown.

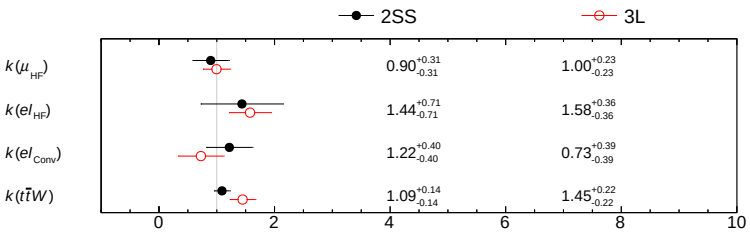
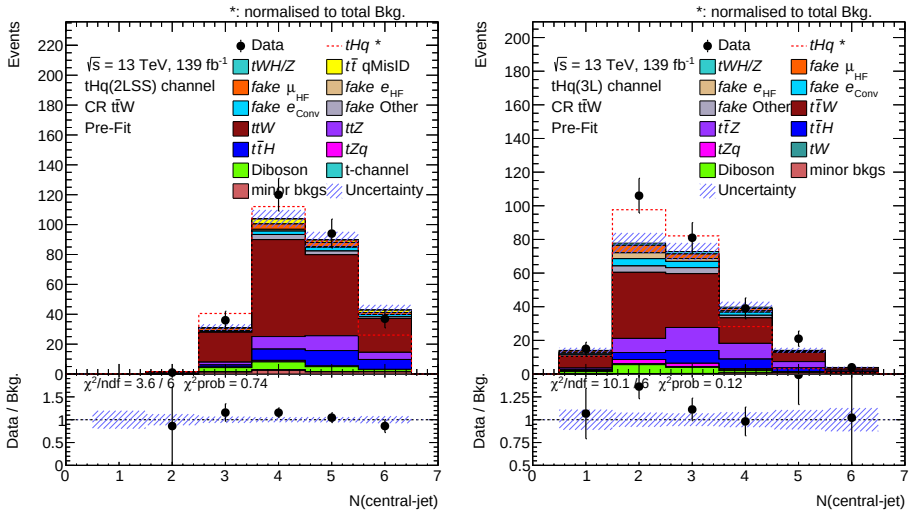


FIGURE 5.38: Normalisation factors, i.e. $\mathrm{k}(\mu_{\mathrm{HF}})$, $\mathrm{k}(\mathrm{e}_{\mathrm{HF}})$, $\mathrm{k}(\mathrm{e}_{\mathrm{conv}})$ and $\mathrm{k}(\mathrm{t\bar{t}}W)$, for the 2ℓSS channel in black dots and the 3ℓ channel in red circles. The uncertainties include statistical and all the systematic sources.

case of the 3ℓ channel than in the case of the 2ℓSS channel. On the other hand, the mis-modelling is higher when the number of central jets is lower, and the general agreement measured by the probabilistic χ^2 is lower for the 3ℓ channel. Therefore, these facts explain the difference values of $\mathrm{k}(\mathrm{t\bar{t}}W)$ in both channels. Moreover, the $\mathrm{t\bar{t}}W$ simulation is currently under study to improve the agreement with the data.



(A) (B)

FIGURE 5.39: Pre-fit distributions on the number of central jets in CR($\mathrm{t\bar{t}}W$) for (A) the 2ℓSS and (B) the 3ℓ channel. The uncertainty bands include the statistical and all the systematic sources. The lower panels show the ratio between real and simulated background data events. Moreover, the χ^2 over the ndf and the probabilistic χ^2 are included in order to measure the agreement between real and simulated data events.

In figure 5.40, the $\mu(tHq)$ for both channels are shown. Their values are compatible between them and with the SM within their uncertainties. The uncertainties are split in total and statistical uncertainties from the data events. In both cases the statistical uncertainty is the main component of the total uncertainty.

As it was mentioned in section 5.6.1, in addition to the $\mu(tHq)$ it is possible to obtain a limit of this value using the CL method. The upper limits for both channels are shown in figure 5.41. Two different limits are shown: the expected and the observed upper

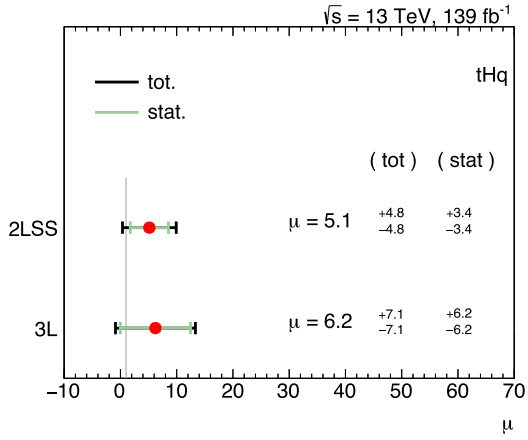


FIGURE 5.40: Signal strength values for the both the 2ℓSS and the 3ℓ channel. The total uncertainty (tot) includes statistical and systematic effects. The statistical uncertainty (stat) is also shown, separately.

limits. The expected upper limit uses the values of the normalisation factors and the NPs from the fit and it assumes the $\mu(tHq)$ is equal to zero. In addition to the upper limit, $\pm 1\sigma$ and $\pm 2\sigma$ variations are also shown. The observed upper limit is computed only with data samples.

The results shown in figure 5.40 and 5.41 represent the final result of the direct search of *tHq*. That means they are the results of the current thesis. They have been computed using a PLBF with all the available sources of statistical and systematic and considering four normalisation factors.

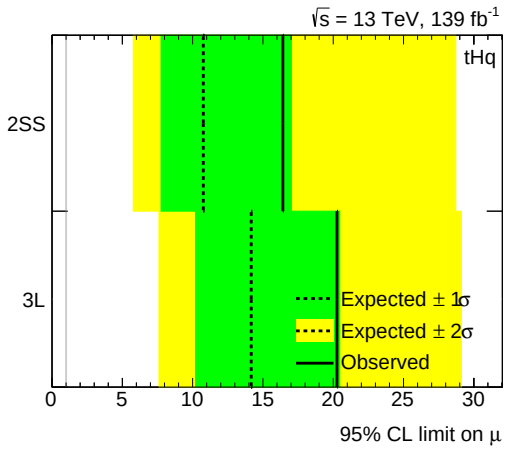


FIGURE 5.41: Upper limits of the signal strength $\mu(tHq)_{\mathrm{CL95}}$ for the 2ℓSS and the 3ℓ channel. Expected and observed upper limits are included. The green and yellow areas represent the $\pm 1\sigma$ and $\pm 2\sigma$ variation of the expected upper limit.

CHAPTER 6

Conclusion

The analysis presented in this thesis is focused on the first search for the direct production of a Higgs boson in association with a top quark in final states with three light-flavour leptons and two light-flavour leptons with the same charge using the ATLAS detector.

This analysis is motivated by the interaction of the two important particles which are involved, the Higgs boson and the top quark. On the one hand the Higgs boson is the particle which allows to explain how particles acquire mass through the SSB mechanism and the Yukawa couplings. On the other hand, the top quark is the most massive quark in the SM and the only one which directly decays without hadronising. These two facts clearly indicate the special interest of the study of the interaction between these two particles. Moreover, the Yukawa coupling of this interaction also allows a possible CP violation that would affect to the value of the production cross-section of this process.

The analysis is performed using the pp collisions at $\sqrt{s} = 13$ TeV recorded during the Run 2 by the ATLAS detector at the LHC. The status and the characteristics of the detector during this period are reviewed in section 2. The total integrated luminosity used in the analysis is 139 fb^{-1}, which allows the study of processes with a low cross-section as in the case of the *tHq* production. Furthermore, the extraordinary performance achieved by the ATLAS detector also allows us to explore such processes. The alignment of the ID is included in this thesis as an example of work related to the performance of the detector.

MVA approach based on several BDTs is implemented to define the SRs, VRs and CRs. The fake/non-prompt leptons are estimated with a dedicated TFM method, whose results are obtained as normalisation factors. They are split in three categories according to the physical process which causes the mis-identification of leptons. Moreover, possible mis-modelling of the $t\bar{t}W$ process is also considered through an additional normalisation factor. The values of these normalisation factors given by the PLBF are shown in table 6.1, summarising results already shown in figure 5.38.

TABLE 6.1: Normalisation factors, i.e. $k(\mu_{HF})$, $k(e_{HF})$, $k(e_{conv})$ and $k(t\bar{t}W)$, for both channels. The uncertainties include statistical and all the systematic sources.

	$k(\mu_{HF})$	$k(e_{HF})$	$k(e_{conv})$	$k(t\bar{t}W)$
2ℓSS	0.90 ± 0.31	1.44 ± 0.71	1.22 ± 0.40	1.09 ± 0.14
3ℓ	1.00 ± 0.23	1.58 ± 0.36	0.73 ± 0.39	1.45 ± 0.22

The values of all the normalisation factor are compatible between them, and their values are compatible with the SM at least for one of the channel within their uncertainties. The normalisation factor of the $k(t\bar{t}W)$ process is on the limit of their compatibility. The source of this tension is the different phase space for the $t\bar{t}W$ process of each channel.

The values of $\mu(tHq)$ for both chanel are shown in table 6.2 and the limits for $\mu(tHq)_{CL95}$ are shown in table 6.3.

TABLE 6.2: Signal strength values for the for the 2ℓSS and the 3ℓ channel. The uncertainty includes statistical and systematic effects.

	$\mu(tHq)$
2ℓSS	5.1 ± 4.8
3ℓ	6.2 ± 7.1

TABLE 6.3: Values of the upper limits of the signal strength $\mu(tHq)_{CL95}$ for the 2ℓSS and the 3ℓ channel.

	$\mu(tHq)_{CL95}$	
	Expected	Observed
2ℓSS	< 11	< 16
3ℓ	< 14	< 20

The values of $\mu(tHq)$ are compatible between them, as well as with the SM predictions within their uncertainties. The values of the upper limits are not directly comparable with the upper limits shown in table 1.1 in section 1.4 for the $t\bar{t}H$/*tHq* multilepton and $H \to \gamma\gamma$ analyses. In the first case, the signal is composed for both processes $t\bar{t}H$ and *tHq* instead of only the *tHq* process. In the second case, the value of the limit is

given after fixing the other processes included using a simplified template cross-section method instead of evaluating the normalisation factor for the *tHq* process at the same time. The results given for both processes are going to be combined within the ATLAS collaboration to provide a single set of results, but unfortunately it is out of the scope of this thesis.

The results of this analysis could be improved in the future due to several factors, such as:

- The increment of the luminosity during the Run 3, since the limited data statistic is the main source of uncertainties in both channels. Thus, an increment in the number of data events clearly will benefit the study of the *tHq* process.
- Reduction of experimental systematic uncertainties due to a better objects reconstruction, identification and calibration.
- New studies and results about the Yukawa coupling of the top quark could provide new information which would directly affect to the cross-section production of the tHq and its direct search. They could come from specific analyses about this coupling, or other processes which also involved this coupling.
- The improvement of the simulation and knowledge about the $\mathrm{t\bar{t}}W$ process.

In the longer term, the improvements expected for an upgraded ATLAS detector and after the Run 3 is finished and the later High-Luminosity phase, will allow to deeply investigate the *tHq* process. For instance, they would allow not only a direct observation but also differential measurements of its cross-section.

APPENDIX A

Dealing with negative weights in MVA techniques

This appendix discusses the issue of having negative event weights in MVA methods.

Negative weights are present in the event sample to correct redundant events in MC ME generation. In general, event weights are used to represent histograms or to calculate expected event yields. However, these event weights in MVA techniques are used as a multiplicative factor of another internal weight given by the model to one event. Afterwards, the internal weight is used during the minimisation process of the loss function, known as training.

Therefore, negative weights can bias the training causing a prejudice in the MVA response since they change the sign of the internal weights of the model. If there were only a small fraction of negative weights in the event samples or they were not in the targeting samples, they would not cause any issue. Unfortunately, negative weights represent about 30 % of the *tHq* signal simulation sample in both final-state channels. Thus, dealing with negative weights during the training becomes a key point in the MVA training for the current analysis.

Several approaches are tested in order to achieve the best performance, in particular, the following three:

- *Absolute values:* absolute value of event weights is used as input weights in the MVA methods for the training.
- *Events with positive weights:* only events with positive weights are considered in the training.
- *Redefining weights:* the reason of an event weight to be negative is because of MC generator event weight is negative. Therefore, weights become strictly positive by

only removing the MC generator event weight from the weight definition. Thus, redefined weights are used as input weights in the MVA methods for the training.

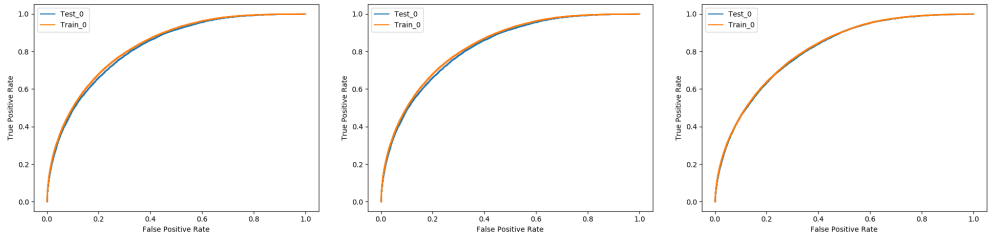
Other strategies could be taken as for example re-scaling the weights to make them all positive, but they are not tested due to limitation on time.

A.1 Negative-weight strategy for the BDTs

These three approaches are tested for the 3ℓ channel only. In this case, the MVA method used involved several binary BDTs using the *XGBoost* package. The ROC curves produced in each approach, shown in figure A.1, is used as the figure of merit to try to choose the best one.

In the light of the result the three approaches have a similar performance. Thus, the easiest one, i.e. *Events with positive weights*, is selected for the training for both the 3ℓ and the 2ℓSS channel.

Independently of the approach selected to deal with negative weights, that approach is only used in the training step. Predictions and validations for the MVA models are always done with the complete set of events including negative and positive weights.



(A) *Absolute values* (B) *Events with positive weights* (C) *Redefining weights*

FIGURE A.1: ROC curves for the 3ℓ channel for the three difference approaches: (A) *Absolute values*, (B) *Events with positive weights* and (C) *Redefining weights*. As can be seen, the performance achieved in the three cases is similar.

APPENDIX B

Optimisation and evaluation of the BDTs

As it is mentioned in chapter 5, the BDTs play a key role in the analysis presented in this thesis. The results presented depend on the discrimination power provided by the BDT classifiers. That discrimination power is a consequence of a strategy of optimisation and evaluation of the BDT classifiers.

The optimisation involves two different processes: the optimisation of the list of input variables, in section B.1, and the optimisation of the hyperparameter, in section B.2. Both processes are intrinsically linked, even though they are done independently. The order in which are done is: first the optimisation of the list of input variables, and second the optimisation of the hyperparameters. In any case, the first process is again evaluated at the end of the second one to check if the optimisation is different and is needed to restart the process with the set of hyperparameter values given by the optimisation. These final sets of hyperparameters and input variables is used in the analysis. The optimisation is performed by splitting the event simulated sample in five sub-samples where one containing the 20% of the events is used as test sub-sample, and the other four sub-samples are merged, which contain the 80% of the events, to be used as training. A schema of the division of the simulated event sample is shown in figure B.1.

Once the optimisation provides the optimal input variables and the optimal hyperparameters, the evaluation of the BDT is produced. It consists in training the BDT classifier to build the model, and storing the score provided by the BDT for each simulated event. The event score from the BDT can be only used when the event is located in the test sub-sample. The evaluation of the BDT is done using the k-fold cross-validation method, in section B.3, in a particular way. This method makes it possible to use the entire simulated event samples since it guarantees that all the simulated events are being used inside the test sub-samples and that the score from BDT can be used for that event. Moreover, this method provides a test of the existence of bias due to the optimisation, that must be avoided.

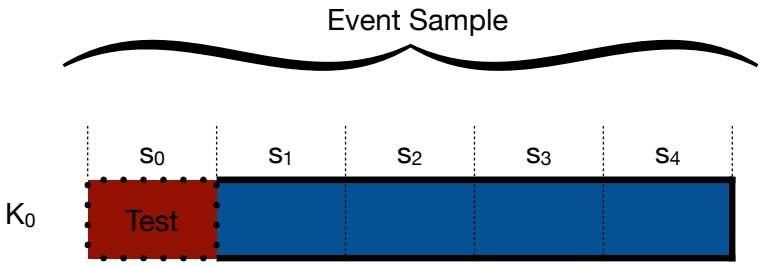


FIGURE B.1: Schematic view of the sub-samples from the full simulated event sample. The red box represents the test sub-sample, and the blue boxes represent the training sub-sample. The $\mathrm{s_k}$ stands for the sub-sample k, in this case $\mathrm{k} = 4$. The $\mathrm{K_0}$ indicates that the election of the test and train sub-sample is only done once.

B.1 Optimisation of the list of input variables of a BDT through the ranking

The optimisation of the list of input variables is split in two different steps. First, an iterative method based on the ranking of features given by the *XGBoost* package is used. After defining an initial set of input variables, the following steps are done in each iteration:

- Training the BDT.
- The input variables are ranked using the *Gain* value and different metrics as the accuracy, the log-loss function and the AUC of the ROC curve are stored.
- The last ranked feature is drop and the BDT is restarted.

All these steps are highlighted schematically in figure B.2 , where each iteration can be easily followed.

The iterative process finishes when the last feature is dropped. The list of features is done using the metrics stored, only the list of features with the best performance remains. This process is done independently for each BDT of each final-state channel.

The second step of the process removes the input variables with high linear correlations. This step avoids instabilities in the BDT response and reduces the number of features without changing the performance of the method. In the case of two input variables that are high linear correlated, i.e. higher than 0.9, the lower ranked is dropped. Figures B.3 and B.4 show the linear correlations between all the variables which are used as initial input variables in the BDTs for the 3ℓ and the 2ℓSS channels, respectively.

B.2 The Genetic algorithm for the hyperparameter optimisation of a BDT

A genetic algorithm (GA) is a general optimisation process inspired by the concept of natural selection [152], and its implementation is known as an evolutionary algorithm. At the beginning of the GA, a set of values for a given set of parameters, which are going to be optimised, are defined randomly within a domain. These sets of values are known as the initial population. Moreover, a function to classify the different set of values,

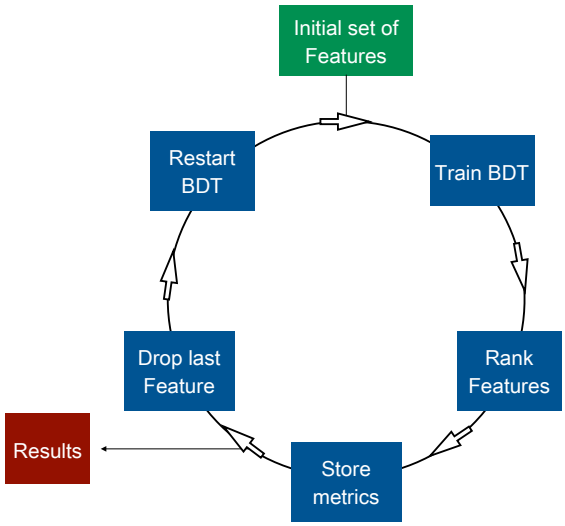
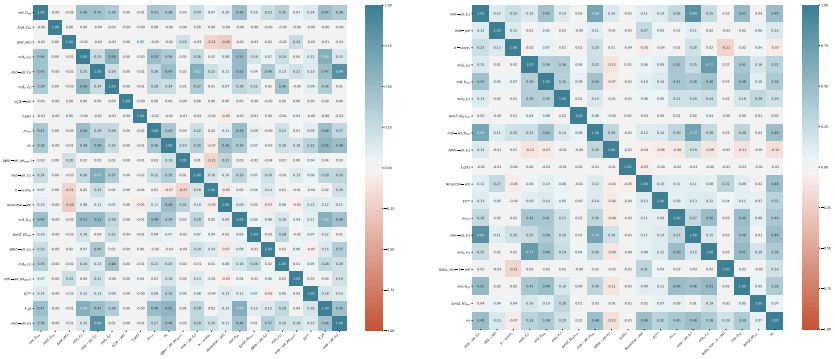
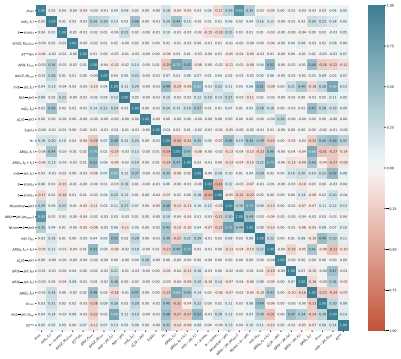


FIGURE B.2: Schematic view of the iterative process to reduce the number of features of a BDT through the ranking and metrics.

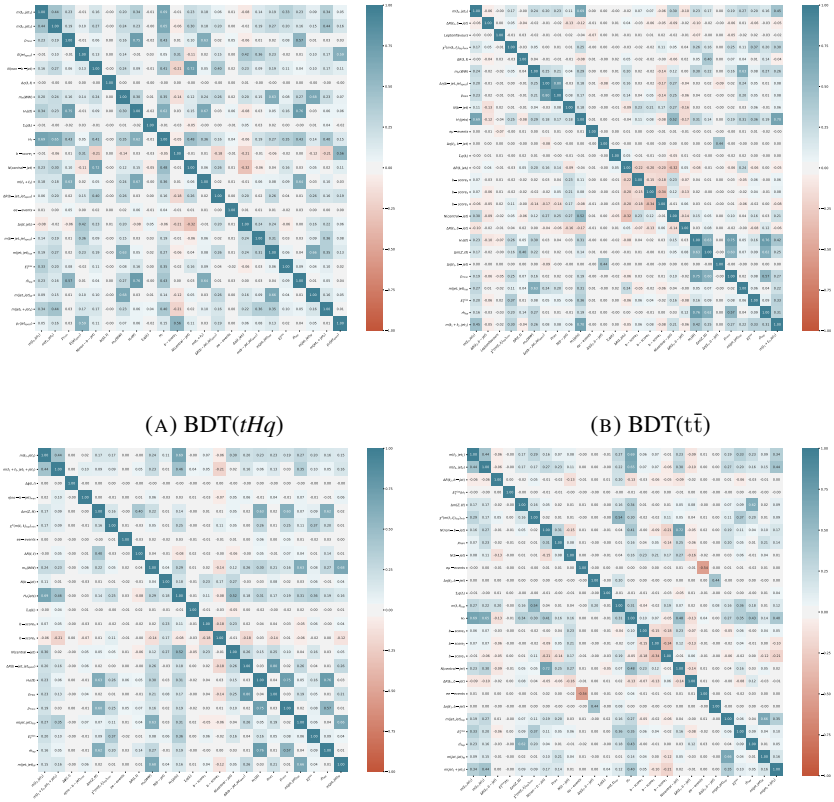


(A) BDT(tHq) (B) BDT($t\bar{t}$)



(C) BDT($t\bar{t}W$)

FIGURE B.3: Linear correlations between the list of variables after the first step described above for the BDTs of the 3ℓ channel: (A) BDT(tHq), (B) BDT($t\bar{t}$) and (C) BDT($t\bar{t}W$). The size and the colours of the squares represent the different correlation values between two variables. If two variables are high correlated, they are not used in the same BDT.



(A) BDT(tHq) (B) BDT($t\bar{t}$)

(C) BDT($t\bar{t}W$) (D) BDT(Diboson)

FIGURE B.4: Linear correlations between the list of input variables after the first step described above for the BDTs of the 2ℓSS channel: (A) BDT(tHq), (B) BDT($t\bar{t}$), (C) BDT($t\bar{t}W$) and (D) BDT(Diboson). The size and the colours of the squares represent the different correlation values between two variables. If two variables are high correlated, they are not used in the same BDT.

called *fitness function*, is defined. At the end, the goal of the GA is to maximise this function. Later on, an iterative process is started with the following steps:

- *Selection and Drop:* the GA ranks each set following the fitness function and removing one half of the initial population. Then, the method duplicates the half remained to continue with the same number of sets.
- *Cross pair:* There is the possibility that a specific value of a parameter is exchanged between two rows. The goal of this process is generating a new set from the old one.
- *Mutate:* There is a probability that the algorithm modifies the value of a parameter to avoid local minimum. In the current case, the value could be multiplied by a random value from a normal distribution.
- *Drop duplicate and renew population:* the GA removes duplicates if they exist, and it adds new sets until it arrives to the same initial population.

The iteration finishes when the values convergence. In the current analysis, the goal of the GA is to optimise the hyperparameters of the BDTs. A BDT is trained for each set of values, and it is restarted at the beginning of each iteration. A schematic view of this iterative process is shown in figure B.5. The *fitness function* (named as Zn) used is:

$$\mathrm{Zn} = \frac{1}{1-\mathrm{AUC}} - \log(\mathrm{LogLoss}),$$

where AUC is the area under the ROC curve and LogLoss is the value of the log-loss function for each set of values.

The hyperparameters of the BDT involved in the optimisation are:

- *scale_pos_weight*: controls the balance between the positive (signal event samples) and negative (background event samples) weights inside the BDT. This parameter is really important in the case of this analysis since the simulated signal and background event samples are unbalanced.
- *min_child_weight*: minimum sum of the internal weights assigned by the BDT needed to produce a new split in the tree.

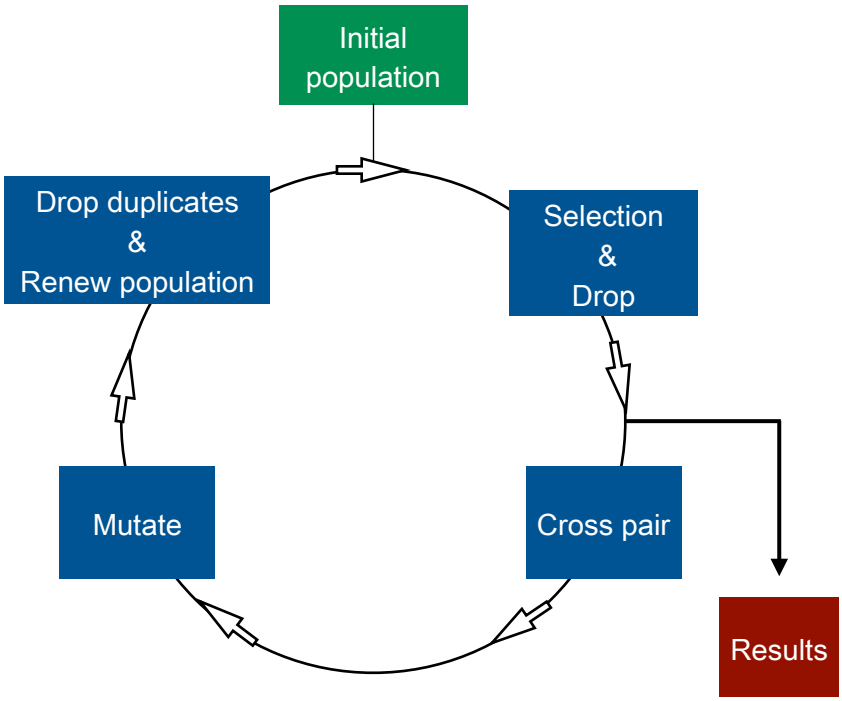


FIGURE B.5: Schematic view of the GA procedure.

- *learning_rate*: Step size shrinkage used between one tree and the following one, inside the BDT.

The convergence of the GA is achieved after roughly 15 iterations for each BDT independently. In the analysis presented in this thesis, the values given by the GA are used in the BDTs to achieve the results shown in the 3ℓ and the 2ℓSS channels.

B.3 The k-fold cross-validation method

In general, the goal of a cross-validation method is to estimate the performance that would achieve an algorithm in a more general situation, that it should be similar to the accessible one. In the current analysis, the k-fold cross-validation method shown in figure B.6 has the advantage of allowing to use the full statistics of the simulated event samples since the event score can be only stored if the event has been used inside the test sub-sample.

The method has a single parameter called k which represents the number of sub-samples of a given sample in which is going to be split. The k-fold consists of two steps. In the first one, shuffling the events and splitting the event sample into k sub-samples. In the last one, taking one of the sub-samples as test and the others as training for the corresponding BDT. These two steps must be repeated k times defining k BDT classifiers. This fact allows storing a BDT score for all the events since they are used inside the test sub-sample once.

The k-fold cross-validation is used in the both 3ℓ and the 2ℓSS channel. The value of k used for both channels is five. The BDT scores are stored for each simulated event sample come from the corresponding fold where this event was used as test. The uncertainties related to the performance of the BDTs come from the different folds.

The sub-samples splitting used in the k-fold method is the same as in the optimisation step. As it shows in figure B.1 the optimisation is done using the fold K_0 in figure B.6. Consequently, if the BDT responses were different considering statistical fluctuations among the different folds, it means that the BDT classifiers would be biased by the optimisation processes given that the only different is the test sub-sample election. Figures B.7 and B.8 show the event score distributions when the event is part of the

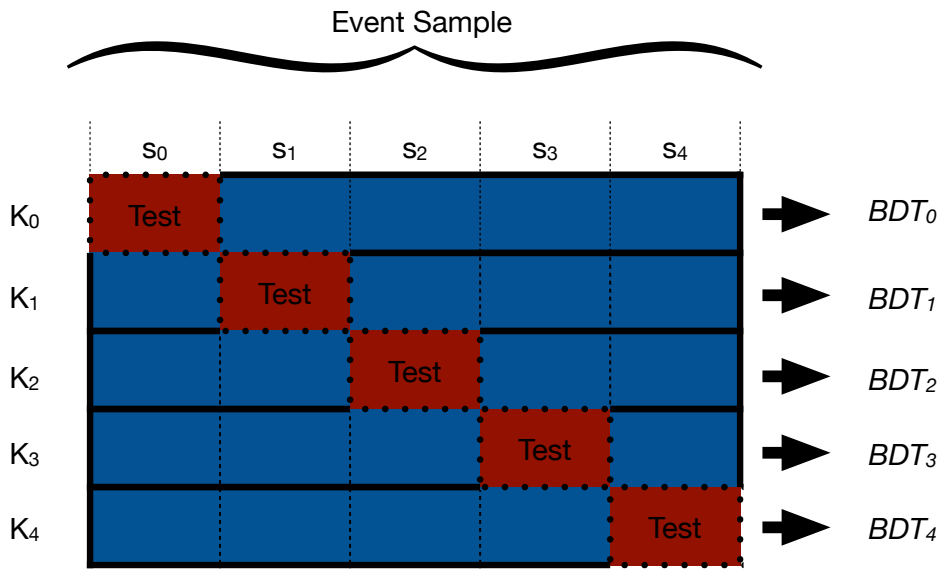
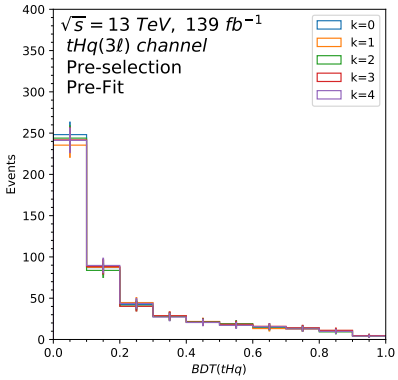
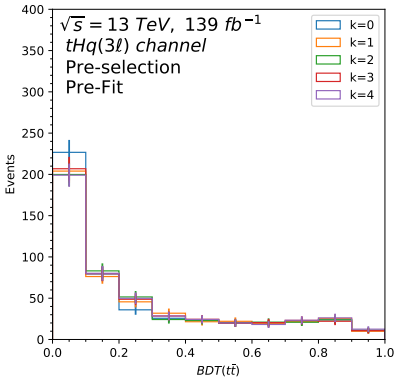


FIGURE B.6: Schematic view of the k-fold process. In this case $k = 5$. The red boxes represent the test sub-samples, and the blue boxes represent the training sub-sample. The s_k stands for the sub-sample k. The K_k indicates the fold k whose BDT trained is stored in BDT_k. Note the event sample for each k is exactly the same. The scores from the BDT_k are storing for the events inside the sub-sample s_k.

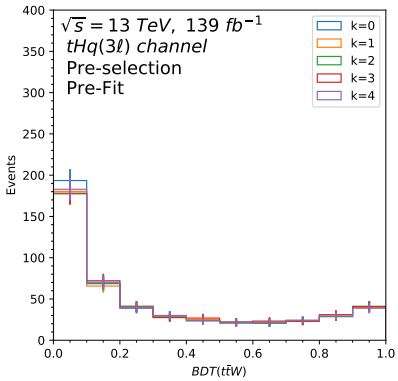
test sub-sample for all the BDT classifiers used in the analysis split by k-fold for both channels. That means that only 20% of the full statistics is represented in each k-fold, and that all the scores shown are used in the analysis. The full statistics is shown since the five folds are on the figures. These distributions do not reveal any biases due to the optimisation since all the folds are compatible among them within their statistical uncertainty. Moreover, the same conclusion is also arrived at using the performances of the BDT classifiers since they are similar for all the folds in both channels. They are shown in chapter 5 in tables 5.3, 5.4 and figure 5.7 for the 3ℓ channel, and in tables 5.7, 5.8 and figure 5.15 for the 2ℓSS channel.



(A) BDT(tHq)

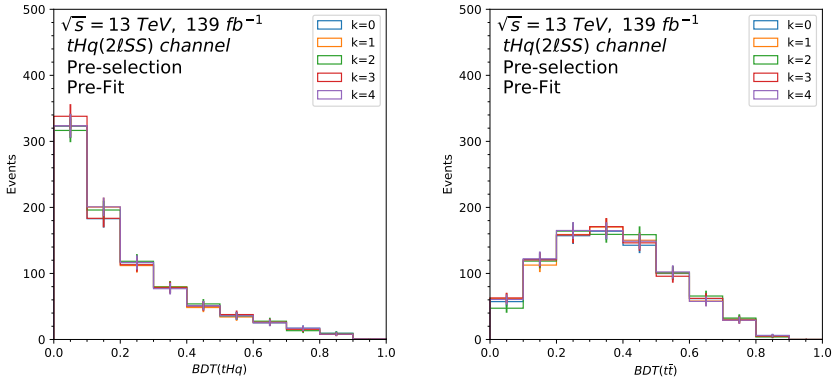


(B) BDT($\mathrm{t\bar{t}}$)

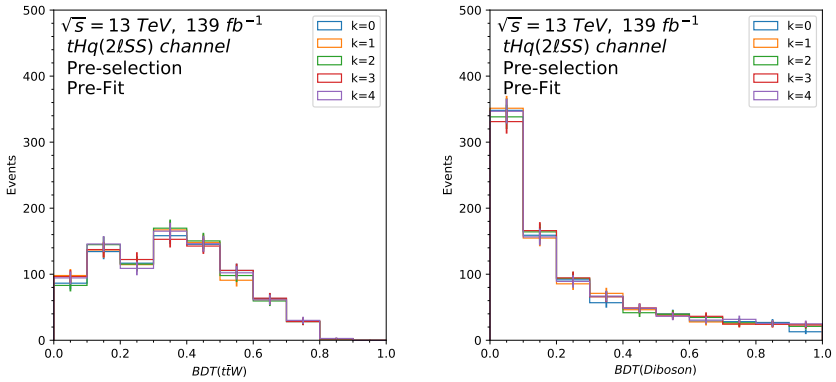


(C) BDT($\mathrm{t\bar{t}}W$)

FIGURE B.7: BDT score distributions of test sub-samples for each k-fold for the BDT classifiers of the 3ℓ channel: (A) BDT(tHq), (B) BDT($\mathrm{t\bar{t}}$) and (C) BDT($\mathrm{t\bar{t}}W$). The colours represent different k-fold. Each line only represents the 20% of total statistic. The uncertainty lines only include the statistical uncertainty.



(A) BDT(tHq) (B) BDT($\mathrm{t\bar{t}}$)



(C) BDT($\mathrm{t\bar{t}}W$) (D) BDT(Diboson)

FIGURE B.8: BDT score distributions of test sub-samples for each k-fold for the BDT classifiers of the 2ℓSS channel: (A) BDT(tHq), (B) BDT($\mathrm{t\bar{t}}$), (C) BDT($\mathrm{t\bar{t}}W$) and (D) BDT(Diboson). The colours represent different k-fold. Each line only represents the 20% of total statistic. The uncertainty lines only include statistical uncertainty.

Resumen

1 Fundamentos teóricos

El conocimiento sobre los componentes más básicos de la materia en el universo y cómo estos componentes interaccionan entre sí es, a día de hoy, una de las líneas de investigación principales en física de partículas. La teoría que explica parcialmente cuales son estos componentes y sus interacciones fundamentales es el Modelo Estándar (SM) de física de partículas.

El SM de física de partículas agrupa una serie de revolucionarias teorías desarrolladas durante los años setenta del siglo XX. En él, se incluyen tanto la mecánica cuántica de campos como la relatividad especial. Desde un punto de vista matemático, el SM se basa en una combinación de distintos grupos de simetría fundamentales, donde cada grupo tiene una interpretación en física de partículas.

Las partículas fundamentales descritas por el SM, mostradas en la figura 1.1, se pueden dividir en dos grupos teniendo en cuenta sus propiedades físicas: fermiones y bosones. Los fermiones son partículas con espín semientero que siguen una estadística de Fermi-Dirac y están subdivididos en dos grupos dependiendo de si tienen o no carga de color: los quarks que tienen carga de color y los leptones, sin ella. Los bosones son partículas con espín entero y siguen una estadística de Bose-Einstein. Los bosones están relacionados con las interacciones fundamentales entre las partículas a nivel cuántico.

Existen cuatro interacciones fundamentales en la naturaleza: la fuerza fuerte, la fuerza débil, la fuerza electromagnética y la gravedad. De estas cuatro solo las tres primeras están incluidas en el SM. Cada una de estas fuerzas tiene una correspondencia directa con el intercambio de uno o más bosones a nivel cuántico:

- La fuerza fuerte aparece entre partículas con diferente carga de color y su rango de acción está limitado al interior del núcleo atómico. La teoría que describe esta fuerza es la cromodinámica cuántica (QCD). En este caso, el bosón mediador de

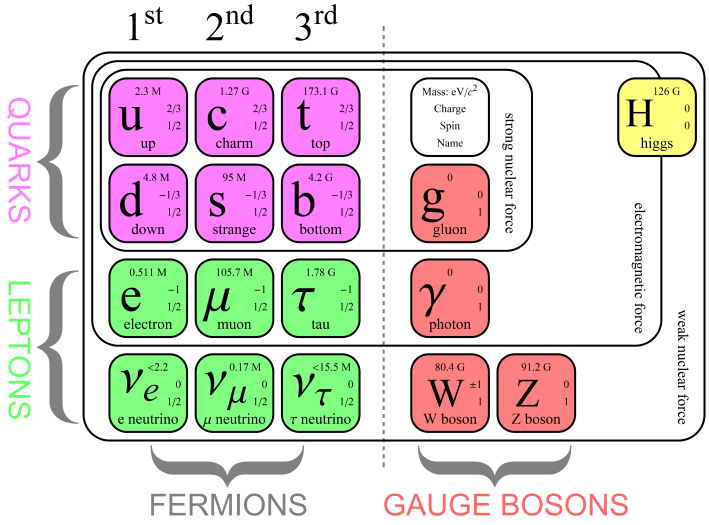


FIGURA 1.1: Partículas fundamentales del SM y sus características físicas. Las partículas están divididas en fermiones y bosones. El color de los cuadrados hace referencia al espín de las partículas. Las líneas que rodean los cuadrados limitan a través de que fuerza fundamental interaccionan las partículas [1].

esta interacción es el gluon. Esta fuerza explica la estabilidad del núcleo atómico, y la coexistencia en él del protón y el neutrón.

- La fuerza electromagnética es la que ocurra entre partículas con carga eléctrica. Su rango de acción es infinito y el bosón mediador es el fotón.

- La fuerza débil es la responsable del decaimiento radioactivo de las partículas y su rango de acción es similar al radio atómico. Los bosones que median esta interacción son los bosones Z y W.

Aunque la fuerza electromagnética y la fuerza débil sean distintas a nivel macroscópico, a nivel cuántico ambas se combinan en la interacción electrodébil (EW) propuesta por S.L. Glashow, A. Salam and S. Weinberg [3–5]. La única interacción fundamental que no está incluida en el SM es la gravedad. En la actualidad, no existen evidencias de una partícula mediadora para esta fuerza de forma similar a las que existen para las otras fuerzas. De todas formas, los efectos de la gravedad se consideran despreciables a escalas propias de la física cuántica.

Todas las partículas incluidas en el SM decaen en partículas más ligeras si este proceso de desintegración está permitido por las leyes de conservación. Los quarks se unen entre sí para formar partículas llamadas hadrones que, en el caso de no ser estables, decaerán en partículas más ligeras. Debido a esto, las propiedades de la mayoría de las partículas se miden a través del estudio de sus decaimientos. Un caso especial de quark es el quark top. Este quark decae directamente en partículas más ligeras debido a tener una vida media menor que el resto.

Teniendo en cuenta solo las interacciones EW y QCD ni los fermiones ni los bosones pueden adquirir masa. Sin embargo, existen medidas experimentales que indican que algunos bosones y fermiones tienen masa. Para explicar estas masas R. Broght, F. Englert y P. W. Higgs propusieron en los años sesenta del siglo XX un mecanismo llamado ruptura espontánea de simetría EW (SSB) [7–9]. El mecanismo SSB introduce un campo escalar complejo cuyo valor mínimo se encuentra localizado en un círculo para una cierta configuración de sus parámetros, el radio de ese círculo es conocido como valor esperado del vacío. Esto introduce un nuevo grado de libertad relacionado con el punto del círculo de valor mínimo en el que se encuentra el potencial. La expansión del potencial entorno a un mínimo elegido en ese círculo produce un campo escalar que se identifica con el bosón de Higgs. Los fermiones del SM adquieren masa a través de interacciones entre el fermión y el campo del bosón de Higgs. Estas interacciones son conocidas como interacciones de Yukawa.

El SM es una teoría de extraordinario éxito que describe gran variedad de sucesos en la naturaleza. Sin embargo, no se trata de una teoría completa ya que no consigue responder a todas las preguntas planteadas en física de partículas. Algunas de estas preguntas son: la adquisición de masa de los neutrinos, que no puede ser explicada usando el SM pero se ha comprobado experimentalmente que son partículas masivas; la explicación de por qué solo existe materia en el universo actual, cuando no existe ningún motivo por el cual la materia y la antimateria no sean creadas a partes iguales en el SM; la inclusión de una posible partícula candidata a materia oscura; una posible explicación de la energía oscura; la inclusión de la gravitación como el resto de fuerzas fundamentales de la naturaleza; la unificación de todas las fuerzas fundamentales a una muy alta escala de energías, etc. Afortunadamente existen teorías que pueden explicar algunas de estas

preguntas abiertas y son compatibles con las predicciones ya observadas del SM, pero ninguna de ellas ha podido ser comprobada experimentalmente.

Existen dos partículas de especial interés dentro del SM: el quark top, por ser el único quark que directamente decae en partículas más ligeras, y el bosón de Higgs, que permite que las partículas adquieran masa.

El quark top fue predicho en el año 1973 por M. Kobayashi y T. Maskawa [20], la primera observación experimental fue realizada por la colaboración D0 [21] y la colaboración CDF [22] en el colisionador de protón-antiprotón Tevatron en el año 1995. Su decaimiento principal (más del 99.83% de las veces) es en un bosón W y en un quark botton. Los posteriores decaimientos del bosón W en otras partículas estables van a definir las partículas que van a ser estudiadas en las investigaciones que involucren al quark top.

El bosón de Higgs juega un importante papel en el SM, como se ha mencionado anteriormente. Su existencia fue postulada junto al mecanismo SSB, pero su observación experimental no se produjo hasta el año 2012 por las colaboraciones ATLAS y CMS en el acelerador LHC. Los decaimientos del bosón de Higgs se muestran en la figura 1.2 en función de su masa, para la masa del Higgs observada $\sim 125\,\text{GeV}$. La principal desintegración del bosón se produce en un par de quark-antiquark bottom.

Tras la observación del bosón de Higgs el estudio del acoplamiento de las demás partículas a él se convirtió en una prueba esencial del SM. En especial, el acoplamiento del bosón de Higgs con el quark top ($\mathrm{y_t}$) adquiere un especial interés debido a las singularidades de estas partículas. El valor de $\mathrm{y_t}$ puede ser calculado mediante el estudio de la producción del bosón de Higgs acompañado de un par de quark antiquark top o un único quark top. Este último caso es la principal línea de investigación que se incluye en esta tesis.

La búsqueda de la producción asociada de un bosón de Higgs y un único quark top, *tHq*, es sensible tanto al valor como al signo de $\mathrm{y_t}$. El valor de su sección eficaz de producción puede incrementarse hasta en un orden de magnitud dependiendo el valor de $\mathrm{y_t}$. Existen algunos resultados previos de las colaboraciones ATLAS y CMS sobre el proceso de *tHq*, aunque el primer caso no son medidas directas y en el segundo con un menor número de datos.

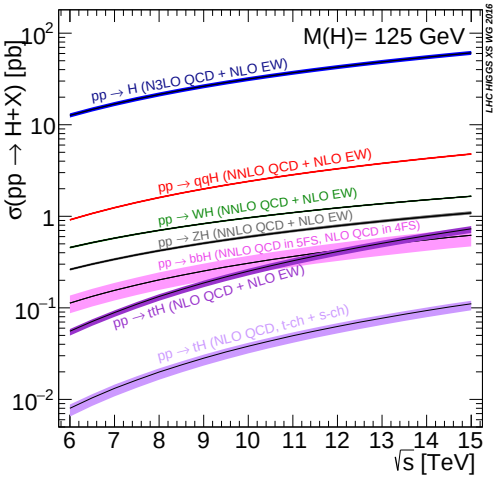


FIGURA 1.2: Valores teóricos para la sección eficaz en función de la energía del centro de masas para cada proceso [36].

El estudio del proceso *tHq* incluido en esta tesis es la primera búsqueda directa de este proceso en la colaboración ATLAS. El principal objetivo de este análisis es proporcionar un valor de la sección eficaz de producción de *tHq*, además de proporcionar un límite superior a este valor. Los pasos seguidos en este análisis están detallados en esta tesis: estrategia general del análisis, definición de señal y fondo, estudio de fondos especiales y la estrategia seguida en el ajuste para obtener los resultados finales.

2 El LHC y el detector ATLAS

El LHC es el último acelerador de un complejo sistema de varios aceleradores de partículas, como se muestra en la figura 1.3, situado en las instalaciones del CERN que se localiza en la ciudad de Ginebra en la frontera entre Suiza y Francia. Se trata de un acelerador circular de 27 km de circunferencia situado a 100 m bajo tierra. En él se aceleran paquetes de partículas hasta velocidades superiores al 99% de la velocidad de la luz mediante el uso de imanes superconductores y cavidades de radiofrecuencia. En concreto, se aceleran dos haces de partículas en direcciones opuestas alcanzando una energía de hasta 6.8 TeV cada uno. Dichos haces se cruzan en cuatro puntos en los que

se hacen colisionar los paquetes de partículas que componen cada haz. La mayoría del tiempo de operación en el LHC se destina a la aceleración de protones. En torno a esos cuatro puntos donde se hacen colisionar los haces se localizan los distintos experimentos que analizarán las colisiones con diversos objetivos.

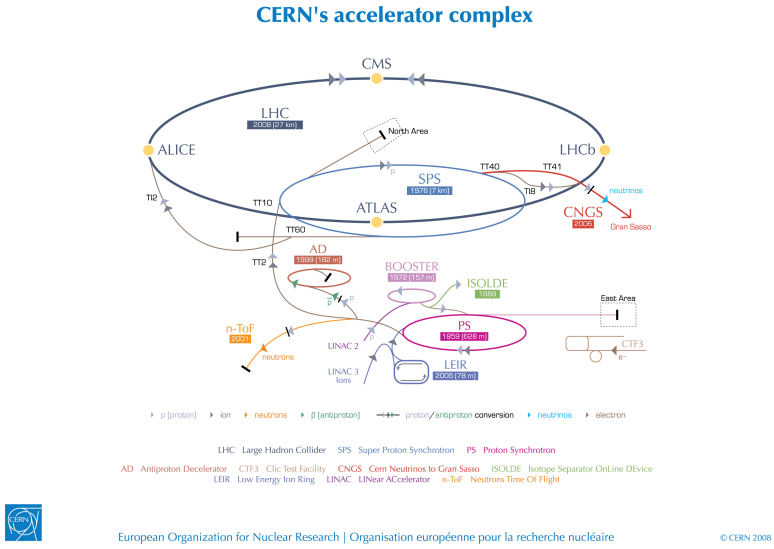


FIGURA 1.3: Representación esquemática del sistema de aceleradores del CERN [53].

Los datos usados en el análisis incluido en esta tesis son los recogidos por el detector ATLAS, esquemáticamente representado en la figura 1.4, durante el periodo de operación del LHC denominado Run 2, que comprende desde el año 2015 al 2018. El detector ATLAS es el más grande de los detectores instalados en el LHC, se trata de un detector cilíndrico, multipropuesta y simétrico a ambos lados del punto donde se produce la colisión. El detector se compone de varios subdetectores especializados en distintos objetivos, situados formando capas de tal forma que al combinar toda la información recogida se identifiquen y reconstruyan todas las partículas que provienen de la colisión.

De dentro a fuera los subdetectores que componen el detector ATLAS son:

- El espectrómetro de muones, cuyo principal objetivo es identificar los muones y reconstruir sus trayectorias. La información obtenida por este subdetector también es usada para la selección de eventos de interés de forma temprana.

- Los calorímetros, de los que existen dos tipos distintos: el electromagnético, que es el más interno, y el hadrónico. EL objetivo principal de los calorímetros es identificar y medir la energía de partículas neutras y cargadas parando el recorrido de las partículas que interaccionan con ellos.

- El detector interno, se encuentra inmerso en un campo magnético solenoidal de 2 T y es el principal sistema de detección de trazas del detector, que son las trayectorias de las partículas cargadas. Proporciona una excelente reconstrucción de las partículas cargadas, así como la identificación de vértices primarios y secundarios producidos en la colisión.

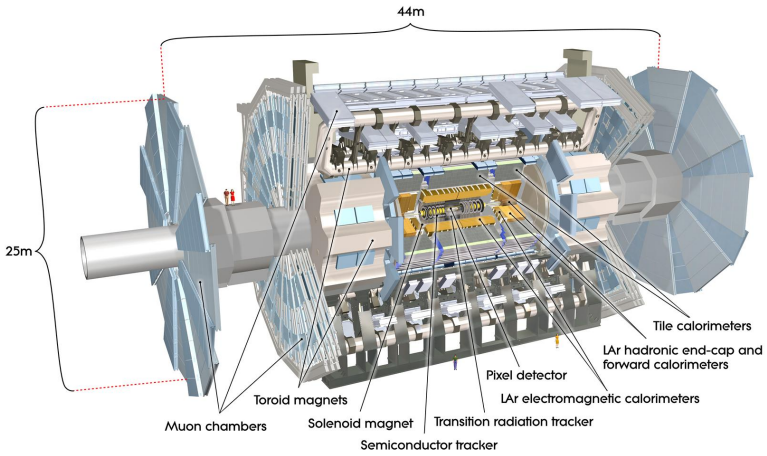


FIGURA 1.4: Imagen virtual del detector ATLAS, donde se muestran tanto las dimensiones como los subsistemas que componen el detector [62].

Además de estos subdetectores, todo el detector se encuentra inmerso en un gran campo magnético generado por imanes toroidales situados en la parte externa del detector. Estos imanes proporcionan un campo magnético de 0.5 T en el centro del detector y de 1 T en sus extremos.

Debido al gran volumen de información producida en cada una de las colisiones es necesario el uso de un sistema de selección temprana que reduzca la cantidad de información que posteriormente será usada en los análisis. Este sistema está compuesto

por dos subsistemas: el primero está basado en herramientas de hardware y solo usa información de algunas partes del detector para reducir el número de eventos e identificar regiones del detector de especial interés; el segundo está basado en herramientas de software, usa información obtenida en el conjunto del detector y utiliza las regiones de interés identificadas por el primer subsistema para reducir el número de eventos hasta el volumen adecuado para poder ser almacenados para su posterior uso en los análisis físicos.

Un aspecto esencial de cualquier análisis físico es el conocimiento en profundidad del detector ya que esto determinará el rendimiento del detector, y en definitiva, la precisión de cualquier resultado. Por este motivo, las tareas de mantenimiento y actualización de sistemas del detector son continuas, como por ejemplo las relacionadas con la reconstrucción de trazas o la identificación de partículas. En el análisis incluido en esta tesis tanto la reconstrucción de los leptones como la identificación de chorros de partículas que contienen un quark botton son fundamentales, ya que el análisis de *tHq* incluye estos objetos en los estados finales que serán utilizados. Estas tareas son extremadamente sensibles a la reconstrucción de las trazas, luego un buen rendimiento del detector interno es fundamental. Una de las tareas que tiene como objetivo la mejora del rendimiento del detector interno es, sin duda, su correcto alineamiento.

El alineamiento del detector interno tiene como objetivo determinar la geometría actual del detector, así como sus posibles cambios en el tiempo. La geometría actual puede diferir de la nominal debido al proceso de montaje o a la operación del detector. El detector interno no es accesible físicamente durante los periodos de tomas de datos por lo que son necesarios métodos indirectos para conocer su geometría. El alineamiento usa una gran muestra de trazas para implementar un método denominado Global χ^2, con el objetivo de conocer la geometría actual del detector y aumentar la precisión de las trazas reconstruidas. El Global χ^2 involucra todos los módulos que componen el detector y sus correlaciones, lo que convierte su resolución en un proceso difícil. Dada esta dificultad, el método se resuelve mediante un proceso iterativo en el que se van incluyendo distintas partes del detector interno secuencialmente atendiendo a su complejidad.

Existen ciertos tipos de movimientos generales del detector interno que el Global χ^2 no es capaz de detectar, estos movimientos son conocidos como deformaciones débiles.

En general, una deformación débil es aquella que deja invariante la fórmula del Global χ^2 y puede desviar las trazas reconstruidas.

Las principales deformaciones débiles estudiadas son: la desviación de la sagita, la expansión radial y la expansión de los end-caps. Estas desviaciones afectan directamente a la reconstrucción del momento de las partículas, y son medidas a través de dos resonancias bien conocidas: $Z \to \mu\mu$ y $J/\psi \to \mu\mu$.

La desviación de la sagita consiste en una desviación de la traza en el plano en el que se encuentra, cambiando el arco de circunferencia formado por la traza en ese plano. El valor de esta desviación depende explícitamente de la zona del detector por la que pasa la traza. La estimación del valor de esta desviación se realiza usando un proceso iterativo, usando la diferencia ente la masa del bosón Z reconstruida a partir de dos muones ($Z \to \mu\mu$) y su masa de referencia. Las medidas de esta desviación durante el Run 2 tienen un valor medio pequeño igual a 0.018 ± 0.085 TeV^{-1} en la parte central del detector interno y casi inexistente en sus extremos.

La expansión radial consiste en una expansión o contracción en el radio del cilindro que forma el detector interno. Siguiendo una técnica similar a la usada para determinar la desviación de la sagita se encuentra que esta expansión puede modificar el momento transverso de las partículas entre un $0.5-1$ %.

La expansión de los end-caps consiste en un movimiento en los extremos del detector interno en el eje del haz, pudiendo ser tanto una expansión como una contracción. Con los datos obtenidos durante el Run 2 y siguiendo un método similar a los anteriores, se comprueba que esta desviación no es observable con el número actual de datos considerando un límite mecánico de esta deformación de 1 mm en la parte más alejada del centro del detector interno.

Combinaciones lineales de estas dos últimas deformaciones débiles también son posibles por lo que se ha implementado un método para poder separar y medir el efecto de ambas, la expansión radial y de los end-caps, al mismo tiempo. Los resultados obtenidos con este segundo método muestran que la expansión radial es despreciable, pero que existe una distorsión global que puede estar relacionada con la expansión de los end-caps. La distorsión global calculada a través de este método también puede ser causada por una desviación en el módulo del campo magnético, no obstante los estudios

necesarios para separar ambos efectos quedan fuera del contenido de esta tesis.

Todas las técnicas descritas sobre el alineamiento del detector interno han sido usadas durante el Run 2, esto ha permitido, entre otras tareas de mantenimiento y actualización, al detector ATLAS la reconstrucción de partículas con una menor perdida de precisión. El análisis incluido en esta tesis es realmente sensible a este hecho debido a dos razones principalmente: los estados finales estudiados involucran dos o tres leptones y parte de su identificación depende del detector interno, y la búsqueda del proceso *tHq* se vería claramente perjudicada por una pequeña perdida de eficiencia del detector debido a su baja sección eficaz de producción.

3 Simulación y adquisición de datos

La simulación y la adquisición de datos es también en sí una parte fundamental del análisis, ya que es con ellos con los que posteriormente se realizará el estudio del proceso *tHq*.

Los datos usados para el análisis incluido en esta tesis son los recogidos por el detector ATLAS durante el Run 2, i.e. de 2015 a 2018. Los datos fueron producidos durante colisiones pp en el LHC con una frecuencia de 25 ns y una energía en el centro de masas de $\sqrt{\mathrm{s}} = 13$ TeV. La luminosidad total integrada alcanzada durante este periodo fue de 139 fb^{-1}, cuyo error varía entre el 2.0% y el 2.4%.

La simulación de eventos, también conocida como simulación Monte Carlo (MC), se divide en varias etapas que abarcan desde el cálculo de la sección eficaz a nivel de partones (quark y gluones) hasta las cascadas de partones y los efectos no perturbativos del proceso, y por último la simulación del paso de las partículas por el detector. Los diferentes pasos de una colisión de pp a tener en cuenta se muestran de forma esquemática en la figura 1.5.

La simulación de MC comprende varios procesos físicos dentro de la colisión: la dispersión fuerte, también conocida como la creación de los elementos de matriz; la cascada de partones, la hadronización, los procesos de dispersión fuerte secundarios, el decaimiento hadrónico y el pile-up. Cada uno de estos procesos se simula de forma indeopendiente y usan diferente información de entrada.

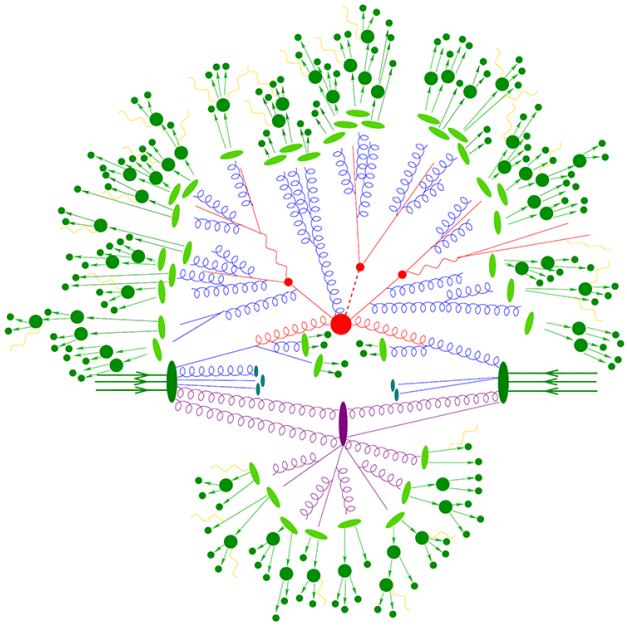


FIGURA 1.5: Esquema de la estructura de una colisión de pp. En el centro se representa el proceso de dispersión fuerte (en rojo). Los círculos rojos en el centro están rodeados por estructuras tipo árbol que representan la radiación de Bremsstrahlung producida por la cascada de partones. Los puntos azules indican los partones en su estado inicial. Los procesos de dispersión fuerte secundarios se muestran en violeta. Por último, se muestra el proceso de hadronización (en verde claro) y los estados finales hadrónicos (en verde). Además, aparecen unas líneas amarillas que representan la radiación de fotones.

Existen diversos programas en el mercado para realizar una o varias partes de la simulación. Los usados en alguna de las simulaciones en el análisis incluido en esta tesis son: POWHEG BOX, MADGRAPH, SHERPA, PYTHIA 8 y HERWIG. Algunos de ellos han sido usados para la simulación nominal de un proceso, mientras que otros se han usado para obtener el error derivado del uso de uno u otro programa.

El último paso de la simulación de eventos es la simulación del paso de las partículas por el detector. Esta se realiza a través de un software dedicado en la colaboración ATLAS [95] de dos formas distintas: incluyendo una descripción física detallada en la simulación usando GEANT4 [96], denominada simulación completa, o considerando

únicamente una descripción parametrizada del calorímetro y GEANT4 para el resto del detector, denominada simulación rápida. En el análisis expuesto en esta tesis, las simulaciones completas son usadas mayoritariamente como muestras nominales, a menos que no esten disponibles, y las simulaciones parciales se usan para estimar distintas fuentes de error.

Las distintas combinaciones de software usadas tanto para el proceso de señal, i.e. *tHq*, como para todos los fondos considerados se muestran en la tabla 1.1. Además de esto, para el proceso de señal se ha añadido un filtro intermedio al nivel de la dispersión fuerte para seleccionar eventos con al menos dos leptones, pudiendo ser electrones, muones o taus, con el fin de aumentar el número de eventos simulados en los estados finales que se estusiarán en el análisis reduciendolos del flujo de la simulación.

TABLA 1.1: Resumen de las muestras simuladas nominales tanto para los procesos de señal como para los procesos de fondos usadas en el análisis.

Proceso	Generador	ME orden	conjunto de PDF	Cascada partonica	parámetros PDF
			Señal		
tHq	MADGRAPH5_AMC@NLO 2.6.2	NLO (4FS)	NNPDF3.0NLO nf4	PYTHIA 8.230	NNPDF2.3LO (A14 tune)
			Fondos		
$t\bar{t}$	POWHEG BOX v2	NLO (5FS)	NNPDF3.0NLO	PYTHIA 8.230	NNPDF2.3LO (A14 tune)
V+jets	SHERPA 2.2.1	NLO+LO	NNPDF3.0NNLO	-	-
Diboson	SHERPA 2.2.1-2	NLO+LO	NNPDF3.0NNLO	-	-
Triboson	SHERPA 2.2.2	NLO+LO	NNPDF3.0NNLO	-	-
$t\bar{t}V$	MADGRAPH5_AMC@NLO 2.3.3	NLO	PYTHIA 8.210	NNPDF2.3LO (A14 tune)	
$t\bar{t}H$	POWHEG BOX v2	NLO (5FS)	PYTHIA 8.230	NNPDF2.3LO (A14 tune)	
t-channel	POWHEG BOX v2	NLO (4FS)	NNPDF3.0NLO nf4	PYTHIA 8.230	NNPDF2.3LO (A14 tune)
Wt	POWHEG BOX v2	NLO (5FS, DR)	NNPDF3.0NLO	PYTHIA 8.230	NNPDF2.3LO (A14 tune)
s-channel	POWHEG BOX v2	NLO	NNPDF3.0NLO	PYTHIA 8.230	NNPDF2.3LO (A14 tune)
tZq	MADGRAPH5_AMC@NLO 2.3.3	NLO	NNPDF3.0NLO	PYTHIA 8.230	NNPDF2.3LO (A14 tune)
tWH	MADGRAPH5_AMC@NLO 2.8.1	NLO (5FS, DR)	NNPDF3.0NLO	PYTHIA 8.245p3	NNPDF2.3LO (A14 tune)
tWZ	MADGRAPH5_AMC@NLO 2.3.3	NLO	NNPDF3.0NLO PYTHIA 8.212	NNPDF2.3LO (A14 tune)	
ttt	MADGRAPH5_AMC@NLO 2.2.2	NLO	NNPDF3.1NLO PYTHIA 8.186	NNPDF2.3LO (A14 tune)	
tttt	MADGRAPH5_AMC@NLO 2.3.3	NLO	NNPDF3.1NLO PYTHIA 8.230	NNPDF2.3LO (A14 tune)	
ggH	POWHEG BOX v2	NLO	CT10	PYTHIA 8.210	CTEQ6L1 (AZNLO tune)
qqH	POWHEG BOX v1	NLO	CT10	PYTHIA 8.186	CTEQ6L1 (AZNLO tune)
WH	PYTHIA 8.186	LO	NNPDF2.3LO	-	-
ZH	PYTHIA 8.186	LO	NNPDF2.3LO	-	-

4 Definición de objetos y reconstrucción de eventos

La información de todas las interacciones de las partículas con las distintas partes del detector es recopilada y combinada para definir los objetos físicos que luego serán usados en el análisis. Un esquema de las distintas interacciones de cada una de las partículas con los distintos subdetectores atendiendo a sus características se muestra en la figura

1.6. La reconstrucción de eventos incluye tanto la definición de las trazas, de los vértices como de los diferentes objetos físicos tras pasar un proceso de selección temprana.

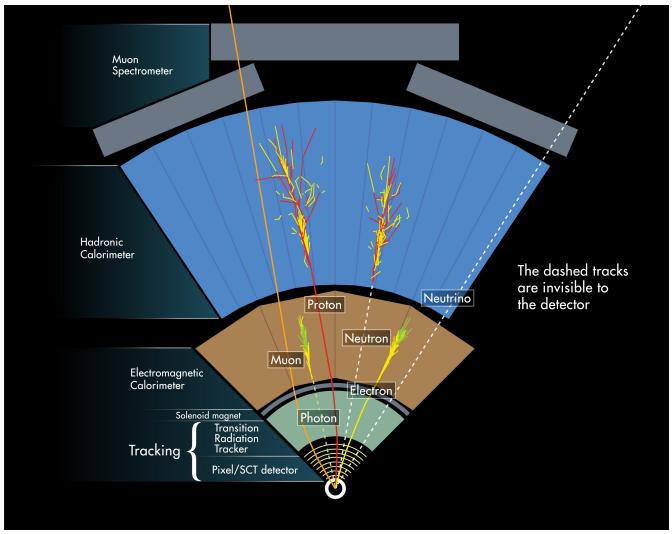


FIGURA 1.6: Diagrama de las trazas de las partículas en el detector ATLAS. Se destacan diferentes partículas y sus interaccionnes con el detecor [167].

Los objetos físicos definidos son:

- Electrones y muones: los primeros son reconstruidos a partir de información recogida en el detector interno y en los calorímetros, la reconstrucción de los segundos usa la información del detector interno y del espectrómetro de muones. Ambas partículas deben cumplir criterios de aislamiento.

- Jets: son reconstruidos usando un algoritmo específico llamado anti-k_t [138], en él se usa la información de distintas cascadas de partículas producidas en el calorímetro en el interior de un cono y la información de las trazas dadas por el detector interno. Los jets que provienen de la hadronización de un quark botton se denominan b-jet, ellos forman una categoría especial dentro de los jets y se utiliza un algoritmo de identificación específico para definirlos.

- Energía transversa perdida: es la magnitud del vector suma de todos los momentos transversos de los objetos reconstruidos y calibrados. Está relacionada con las

partículas que no son detectadas por el detector e ineficiencias de este, ya que en el caso ideal la suma de todos los momentos en el plano transverso al haz es cero.

Además de los procesos de identificación de cada uno de los objetos, se utiliza un último proceso de selección que corre sobre todos los objetos ya definidos para evitar que un posible objeto sea identificado como dos objetos físicos al mismo tiempo.

5 Búsqueda del proceso *tHq*

El análisis del proceso *tHq* usando los datos obtenidos por el detector ATLAS durante el Run 2 es el tema principal de esta tesis. La motivación teórica y la configuración del detector en este periodo se han mencionado con anterioridad. En este análisis solo se han tenido en cuenta dos estados finales del proceso *tHq*: tres leptones ligeros (canal 3ℓ) y dos leptones ligeros con la misma carga (canal 2ℓSS). El objetivo principal de este análisis es contribuir a la primera medida del proceso *tHq* en la colaboración ATLAS. Para alcanzar este objetivo se ha definido una clara estrategia de selección de eventos usando un algoritmo multivariable (MVA) y se han definido varias regiones de interés y, por último, se ha seguido una estrategia para el uso de un ajuste probabilístico que incluye las regiones de interés mencionadas. Estos pasos se siguen de forma independientemente para ambos canales.

Antes de aplicar el algoritmo MVA se define una región llamada de preselección, en la cual se aplicará el algoritmo. Esta región es diferente para cada uno de los canales y se define usando criterios generales en el número de leptones, el número de jets y b-jets, imponiendo límites inferiores al momento transverso de los leptones, requiriendo que la suma de las cargas sea igual al valor adecuado y usando una horquilla de valores atendiendo a la energía transversa perdida. Dichas regiones preselección se definen usando la misma simulación para ambos procesos y tras su definición se usan para determinar las regiones de interes de cada canal. Las tablas 1.2 y 1.3 muestran la composición de estas regiones para cada uno de los canales.

Una vez definida la región de preselección se aplica un método de MVA. En este caso se usa el paquete de Python *XGBoost* para desarrollar varios árboles de decisión impulsados (BDT) con una clasificación binaria de cada evento, es decir el evento pertenece

TABLA 1.2: Eventos predichos en la región de preselección para el canal 2ℓSS por la simulación de MC. Todas las fuentes de incertidumbre han sido incluidas.

Proceso	Yields
tHq	9.96 ± 0.34
tWH	5.24 ± 0.52
$t\bar{t}$	1420 ± 108
$t\bar{t}W$	726 ± 27
$t\bar{t}Z$	164 ± 38
$t\bar{t}H$	126 ± 21
tZq	88 ± 11
Diboson	295 ± 74
Single top t-channel	44 ± 19
Fondos menores	678 ± 344
Total de fondos	3546 ± 384
Datos	3841

TABLA 1.3: Eventos predichos en la región de preselección para el canal 3ℓ por la simulación de MC. Todas las fuentes de incertidumbre han sido incluidas.

Proceso	Yields
tHq	2.53 ± 0.11
tWH	3.12 ± 0.21
tWZ	80 ± 42
$t\bar{t}$	457 ± 72
$t\bar{t}W$	173.3 ± 5.8
$t\bar{t}Z$	563 ± 125
$t\bar{t}H$	74 ± 12
tZq	271 ± 35
Single top tW	19.5 ± 8.6
Diboson	571 ± 143
Fondos menores	15.7 ± 9.0
Total de fondos	2231 ± 216
Datos	2457

a una muestra objetivo o no, de forma independiente. El caso del canal 3ℓ se han usado tres BDTs para diferentes procesos objetivo: el proceso de la señal, i.e. *tHq*, y para los dos fondos de mayor interés, i.e. $\mathrm{t\bar{t}}$ y $\mathrm{t\bar{t}}W$. En el caso del canal 2ℓSS, se han utilizado cuatro BDTs para diferentes procesos: el proceso de la señal, i.e. *tHq*, y para los fondos de mayor interés en este caso, i.e. $\mathrm{t\bar{t}}$, $\mathrm{t\bar{t}}W$ y diboson. Todas las BDTs han sido sometidas a un proceso de optimización tanto de la lista de variables de entrada como de los hiperparámetros que definen la arquitectura de la BDT.

Los procesos considerados como fondos pueden ser clasificados en dos categorías atendiendo a su fuente: irreducible y reducibles. Los irreducibles son aquellos procesos cuyos estados finales en el detector son idénticos al proceso señal y se determinan directamente con muestras simuladas. Los reducibles son los fondos que resultan de ineficiencias experimentales, es decir de un incorrecto rendimiento, algunos ejemplos son la mala identificación de las cargas o la mala identificación de los muones o los electrones. Para el caso de una mala identificación de los muones o los electrones se ha utilizado una técnica específica para estimar estos fondos llamada método de ajuste del modelo (TFM). Este método se basa en realizar un ajuste de probabilidad en regiones donde este tipo de fondos es mayoritario.

Las respuestas de las distintas BDTs son usadas para definir todas las regiones necesarias en ambos canales. Estas regiones se subdividen en tres tipos: la región de señal (SR) se define para maximizar la contribución de la señal en esa región, las regiones de control (CR) tienen como objetivo estimar el desacuerdo entre la simulación y los datos para los fondos cuya precisión es menor y se definen de tal forma que la contribución de ese fondo en esa región sea importante y, por último, las regiones de validación se usan para evaluar el resultado obtenido por el ajuste en diferentes fondos. En el canal de 3ℓ se define una única SR; cuatro CRs, tres relacionadas con el TFM y una con el proceso $\mathrm{t\bar{t}}W$; y tres VR para evaluar el modelo en regiones donde los procesos $\mathrm{t\bar{t}}Z$, *tZq* y diboson son importantes. En el canal de 2ℓSS, se definen las mismas SR y CRs adecuando sus definiciones al canal y una única VR para el proceso de diboson.

La presencia del proceso *tHq* en ambos canales se estudia de forma independiente usando un ajuste de probabilidad por bines que incluye las regiones ya definidas y todas las fuentes de error disponibles en el momento de redacción de esta tesis. Para la

realización del ajuste se define una función de probabilidad que contiene el número de datos en cada bin. Además, se incluye, un factor de normalización para la señal y otro por cada fondo de interés; en este caso son tres relacionados con la mala identificación de electrones y muones, y uno relacionado con el proceso $t\bar{t}W$; con el objetivo de medir el acuerdo entre simulación de MC y datos. Por último, también se incluyen todas las fuentes de error en la función de probabilidad. Una vez definida la función de probabilidad el proceso de ajuste se traduce en un problema de maximización de esta función, cuyos resultados finales serán los distintos factores de normalización.

Además de los factores de normalización, también se proporciona un límite superior a la normalización de la señal. Para obtener este límite se realiza una prueba estadística que depende del valor de la normalización conocido como CLs [165], con el objetivo de alcanzar un nivel de confianza del 95% en el resultado.

Se han realizado distintas pruebas sobre la función de probabilidad antes de liberar la totalidad de los parámetros que la componen con el objetivo de probar la validez de la función. Una vez la función de probabilidad ha superado las distintas pruebas, se ha realizado el ajuste completo.

Los resultados para los factores de normalización relacionados con los fondos se muestran en la figura 1.7. De donde se puede concluir que todos son compatibles entre sí teniendo en cuenta su error. En el caso del factor de normalización relacionado con el proceso $t\bar{t}W$ los valores se encuentran en el límite de ser compatibles. Esto se debe a que se han obtenido en dos espacios de fases distintos entre sí.

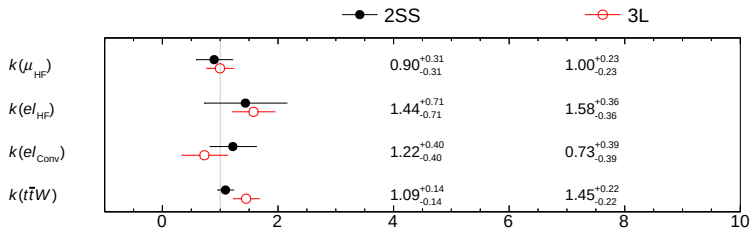


FIGURA 1.7: Factores de normalización para los fondos, donde $k(\mu_{HF})$, $k(e_{HF})$, $k(e_{conv})$ son los relacionados con la mala identificación de leptones ligeros y $k(t\bar{t}W)$ con el proceso $t\bar{t}W$. Se muestran los valores para el canal 2ℓSS en negro y para el canal 3ℓ en rojo. Las incertidumbres incluyen todas las fuentes de error.

En el caso de la normalización de la señal, los valores obtenidos se muestran en la figura 1.8, ambos son compatibles entre sí y con el SM. Además, la figura 1.9 muestra el resultado de los límites superiores para el factor de normalización en ambos canales.

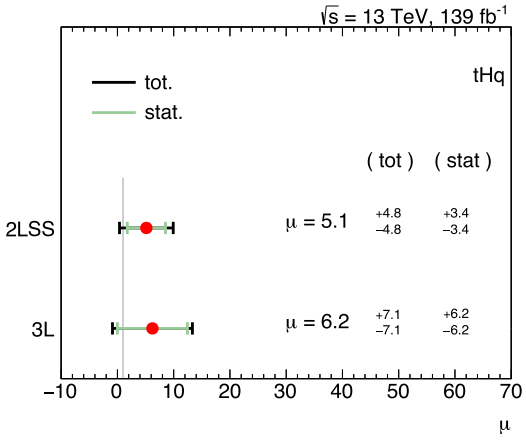


FIGURA 1.8: Factor de normalización de la señal (μ) el canal de 2ℓSS y 3ℓ. La incertidumbre total incluye los efectos tanto estadísticos como sistemáticos. La incertidumbre estadística también se muestra por separado.

Los resultados mostrados en las figuras 1.8 y 1.9 representan el resultado final de la búsqueda directa del proceso *tHq* y por lo tanto de esta tesis.

6 Conclusión

El análisis presentado en esta tesis representa la primira búsqueda directa de la producción de un bosón de Higgs asociado con un quark top dentro de la colaboración ATLAS. En concreto, se han usado los estados finales compuestos por tres leptones ligeros y dos leptones ligeros con la misma carga. El análisis se ha realizado usando colisiones pp con $\sqrt{s} = 13$ TeV recogidas por el detector ATLAS durante el Run 2. La luminosidad integrada usada en el análisis es de 139 fb^{-1}, lo que permite el estudio de procesos con una baja sección eficaz como es el caso de la producción *tHq*.

Se ha implementado una estrategia de selección de eventos a través de MVA, en concreto usando BDTs para definir las diferentes regiones de interes usadas en el posterior

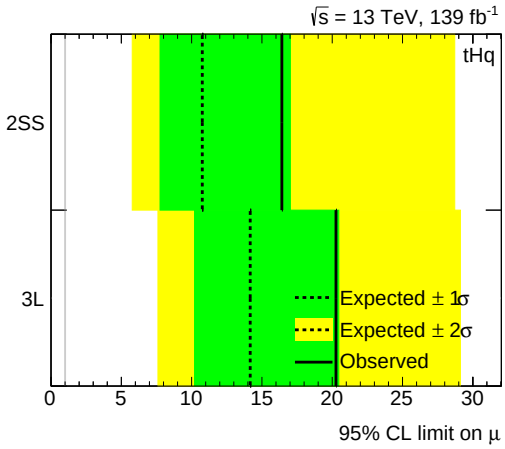


FIGURA 1.9: Límites superiores del factor de normalización de la señal para ambos canales. Se muestran tanto el límite esperado como el observado. Las áreas verde y amarilla representan las variaciones $\pm 1\sigma$ y $\pm 2\sigma$ respectivamente, del límite superior esperado.

ajuste. Se ha realizado un ajuste de probabilidad por bines para cada región de interés, con el objetivo de medir la normalización de los fondos y la señal. El ajuste de probabilidad se realiza de forma conjunta para todas las regiones de interés, con el objetivo de medir la normalización de los fondos y la señal de forma simultánea para ambos canales. La tabla 1.4 muestra los resultados obtenidos para los factores de normalización de los fondos de interés, se puede observar que todos los valores son compatibles entre sí y son compatibles con el SM para todos los fondo en al menos uno de los canales. Los factores de normalización $\mathrm{k}(\mathrm{t\bar{t}}W)$ se encuentran en el límite de ser compatibles entre sí, esto se debe a la diferencia entre los espacios de fase en los que se ha obtenido.

TABLA 1.4: Factores de normalización, i.e. $\mathrm{k}(\mu_{\mathrm{HF}})$, $\mathrm{k}(\mathrm{e}_{\mathrm{HF}})$, $\mathrm{k}(\mathrm{e}_{\mathrm{conv}})$ and $\mathrm{k}(\mathrm{t\bar{t}}W)$, para ambos canales. Todas las fuentes de incertidumbre han sido incluidas.

	$\mathrm{k}(\mu_{\mathrm{HF}})$	$\mathrm{k}(\mathrm{e}_{\mathrm{HF}})$	$\mathrm{k}(\mathrm{e}_{\mathrm{conv}})$	$\mathrm{k}(\mathrm{t\bar{t}}W)$
2ℓSS	0.90 ± 0.31	1.44 ± 0.71	1.22 ± 0.40	1.09 ± 0.14
3ℓ	1.00 ± 0.23	1.58 ± 0.36	0.73 ± 0.39	1.45 ± 0.22

Los factores de normalización para la señal se muestran en la tabla 1.5 y la tabla 1.6

muestra el valor de los límites superiores para ambos canales. Los valores de los factores de normalización de la señal son compatibles entre sí y con el SM teniendo en cuenta su incertidumbre.

TABLA 1.5: Factores de normalización para la señal en los canales 2ℓSS y 3ℓ. La incertidumbre incluye incluye efectos estadísticos y sistemáticos.

	$\mu(tHq)$
2ℓSS	5.1 ± 4.8
3ℓ	6.2 ± 7.1

TABLA 1.6: Límites superiores del factor de normalización de la señal ($\mu(tHq)_{\mathrm{CL95}}$) para ambos canales.

	$\mu(tHq)_{\mathrm{CL95}}$	
	Esperado	Observado
2ℓSS	< 10.8	< 16.4
3ℓ	< 14.2	< 20.3

Los resultados obtenidos en este análisis podrían ser mejorados en el futuro debido a varias razones como el incremento de la luminosidad durante el Run 3, la la reducción de incertidumbres sistemáticas, la mejor comprensión de las simulaciones de procesos como $\mathrm{t\bar{t}}W$, etc. Además, a más largo plazo, este tipo de análisis se verá muy beneficiados por la mejoras en el detector ATLAS que se realizarán después del Run 3 así como la posterior fase de actividad del LHC llamad a High-Luminosity. Esta fase permitirá investigar la producción de *tHq* con una precisión sin precedentes, lo que permitirá ya no solo una observación directa del proceso sino también otras medidas como la sección eficaz diferencial del proceso.

BIBLIOGRAPHY

[1] M. Lubej, *A pretty diagram of the Standard Model*, 2015, URL: https://www-f9.ijs.si/~lubej/SM.pdf.

[2] E. Noether, *Invariant variation problems*, Transport Theory and Statistical Physics **1**.3 (1971) 186–207.

[3] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22**.4 (1961) 579–588.

[4] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266.

[5] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) 367–377.

[6] T. Nakano and K. Nishijima, *Charge Independence for V-particles*, Progress of Theoretical Physics **10**.5 (1953) 581–582.

[7] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13**.9 (1964) 321–323.

[8] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13**.16 (1964) 508–509.

[9] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13**.20 (1964) 585–587.

[10] P. F. de Salas et al., *Status of neutrino oscillations 2018: 3σ hint for normal mass ordering and improved CP sensitivity*, Phys. Lett. **B782** (2018) 633–640, arXiv: 1708.01186 [hep-ph].

[11] A. D. Sakharov, *Violation of CP in variance,Casymmetry, and baryon asymmetry of the universe*, Soviet Physics Uspekhi **34**.5 (1991) 392–393.

[12] G. Hinshaw et al., *Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results*, Astrophys. J. Suppl. **208** (2013) 19, arXiv: 1212.5226 [astro-ph.CO].

[13] Planck Collaboration, *Planck 2018 results. I. Overview and the cosmological legacy of Planck*, Astron. Astrophys. **641** (2020) A1, arXiv: 1807.06205 [astro-ph.CO].

[14] L. Roszkowski, E. M. Sessolo, and S. Trojanowski, *WIMP dark matter candidates and searches—current status and future prospects*, Reports on Progress in Physics **81**.6 (2018) 066201.

[15] L. Di Luzio et al., *The landscape of QCD axion models*, Physics Reports **870** (2020) 1–117.

[16] A. G. Riess, *The Case for an Accelerating Universe from Supernovae*, Publications of the Astronomical Society of the Pacific **112**.776 (2000) 1284–1299.

[17] H.C. Cheng and I. Low, *TeV symmetry and the little hierarchy problem*, JHEP **09** (2003) 051, arXiv: hep-ph/0308199.

[18] F. Cooper, A. Khare, and U. Sukhatme, *Supersymmetry and quantum mechanics*, Physics Reports **251**.5-6 (1995) 267–385.

[19] W. De Boer, *Grand unified theories and supersymmetry in particle physics and cosmology*, Progress in Particle and Nuclear Physics **33** (1994) 201–301.

[20] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652–657.

[21] D0 Collaboration, *Observation of the top quark*, Phys. Rev. Lett. **74** (1995) 2632–2637, arXiv: hep-ex/9503003.

[22] CDF Collaboration, *Observation of top quark production in* $\bar{\mathrm{p}}\mathrm{p}$ *collisions*, Phys. Rev. Lett. **74** (1995) 2626–2631, arXiv: hep-ex/9503002.

[23] M. Benedikt et al., *LHC Design Report.V3*, CERN Yellow Reports: Monographs, 2004, URL: https://cds.cern.ch/record/823808.

[24] ATLAS Collaboration, *Measurement of single top-quark production in the s-channel in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, JHEP **06** (2023) 191, arXiv: 2209.08990 [hep-ex].

[25] Particle Data Group, *Review of Particle Physics*, Progress of Theoretical and Experimental Physics **2022**.8 (2022).

[26] ATLAS Collaboration, *Top working group cross-section summary plots June 2022*, ATL-PHYS-PUB-2022-031, 2022, URL: https://cds.cern.ch/record/2812502/.

[27] ATLAS Collaboration, *Probing the W tb vertex structure in t-channel single-top-quark production and decay in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, JHEP **04** (2017) 124, arXiv: 1702.08309 [hep-ex].

[28] CMS Collaboration, *Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at* $\sqrt{s}$ *=13 TeV*, Phys. Rev. Lett. **122**.13 (2019) 132003, arXiv: 1812.05900 [hep-ex].

[29] ATLAS Collaboration, *Observation of the associated production of a top quark and a Z boson in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, JHEP **07** (2020) 124, arXiv: 2002.07546 [hep-ex].

[30] CMS Collaboration, *Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at* $\sqrt{s} =$ *13 TeV*, JHEP **08** (2018) 011, arXiv: 1711.02547 [hep-ex].

[31] CMS Collaboration, *Measurement of top quark pair production in association with a Z boson in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **03** (2020) 056, arXiv: 1907.11270 [hep-ex].

[32] ATLAS Collaboration, *Measurements of the inclusive and differential production cross sections of a top-quark–antiquark pair in association with a Z boson at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Eur. Phys. J. C **81**.8 (2021) 737, arXiv: 2103.12603 [hep-ex].

[33] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv: 1207.7214 [hep-ex].

[34] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: 1207.7235 [hep-ex].

[35] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in* pp *Collisions at* $\sqrt{s} = 7$ *and* $8\,TeV$ *with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114** (2015) 191803, arXiv: 1503.07589 [hep-ex].

[36] D. De Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 2016, arXiv: 1610.07922 [hep-ph].

[37] ATLAS Collaboration, *Combined measurements of Higgs boson production and decay using up to* 80 fb^{-1} *of proton-proton collision data at* $\sqrt{s} = 13$ TeV *collected with the ATLAS experiment*, Phys. Rev. D **101** (1 2020) 012002.

[38] CMS Collaboration, *Observation of Higgs Boson Decay to Bottom Quarks*, Phys. Rev. Lett. **121** (12 2018) 121801.

[39] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, 2013, arXiv: 1307.1347 [hep-ph].

[40] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. B **784** (2018) 173, arXiv: 1806.00425 [hep-ex].

[41] CMS Collaboration, *Search for associated production of a Higgs boson and a single top quark in proton–proton collisions at* $\sqrt{s} = 13\,TeV$, Phys. Rev. D (2019).

[42] ATLAS Collaboration, *Study of the* CP *properties of the interaction of the Higgs boson with top quarks using top quark associated production of the Higgs boson and its decay into two photons with the ATLAS detector at the LHC*, Phys. Rev. Lett. **125** (2020) 061802, arXiv: 2004.04545 [hep-ex].

[43] CMS Collaboration, *Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at* $\sqrt{s} = 13\,TeV$, JHEP **07** (2021) 027, arXiv: 2103.06956 [hep-ex].

[44] CMS Collaboration, *Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **81**.4 (2021) 378, arXiv: 2011.03652 [hep-ex].

[45] ATLAS Collaboration, *Combined measurements of Higgs boson production and decay using up to* $139\,fb^{-1}$ *of proton-proton collision data at* $\sqrt{s} = 13$ *TeV collected with the ATLAS experiment*, ATLAS-CONF-2021-053, 2021, URL: https://cds.cern.ch/record/2789544.

[46] CMS Collaboration, *Measurement and interpretation of differential cross sections for Higgs boson production at* $\sqrt{s} = 13\,TeV$, Phys. Lett. B **792** (2019) 369, arXiv: 1812.06504 [hep-ex].

[47] F. Demartin et al., *Higgs production in association with a single top quark at the LHC*, Eur. Phys. J. C **75**.6 (2015) 267, arXiv: 1504.00611 [hep-ph].

[48] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, **2/2017** (2016), arXiv: 1610.07922 [hep-ph].

[49] CMS Collaboration, *Search for the associated production of a Higgs boson and a single top quark in pp collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ (2018), URL: https://cds.cern.ch/record/2628662/.

[50] ATLAS Collaboration, *Measurement of the properties of Higgs boson production at* $\sqrt{s}$*=13 TeV in the* $H \to \gamma\gamma$ *channel using 139* fb^{-1} *of* pp *collision data with the ATLAS experimentr*, ATLAS-CONF-2020-026, 2020, URL: https://cds.cern.ch/record/2725727/.

[51] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[52] L. Evans and P. Bryant, *LHC Machine*, JINST **3**.08 (2008) S08001.

[53] C. Lefèvre, *The CERN accelerator complex. Complexe des accélérateurs du CERN* (2008), URL: https://cds.cern.ch/record/1260465.

[54] M. A. Hone, *The duoplasmatron ion source for the new CERN Linac preinjector*, Geneva, 1979, URL: http://cds.cern.ch/record/2640736.

[55] TOTEM Collaboration, *Total cross-section, elastic scattering and diffraction dissociation at the Large Hadron Collider at CERN : TOTEM Technical Design Report*, TOTEM-TDR-001, 2004, URL: https://cds.cern.ch/record/704349.

[56] MoEDAL Collaboration, *Technical Design Report of the MoEDAL Experiment*, MoEDAL-TDR-001, 2009, URL: https://cds.cern.ch/record/1181486.

[57] LHCf Collaboration, *LHCf experiment : Technical Design Report*, LHCF-TDR-001, 2006, URL: https://cds.cern.ch/record/926196.

[58] LHCb Collaboration, *Road map for selected key measurements from LHCb*, LHCB-PUB-2009-029, 2010, URL: https://cds.cern.ch/record/1224241.

[59] ALICE Collaboration, *ALICE physics performance : Technical Design Report*, J. Phys. G 32 (2006).

[60] CMS Collaboration, *CMS Physics Technical Design Report, Volume II: Physics Performance*, J. Phys G **34** (2007) 995.

[61] ATLAS Collaboration, *Run 2 Luminosity Public Plots*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ElectronChargeFlipTaggerToolhttps://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[62] J. Pequenao, *Computer generated image of the whole ATLAS detector*, CERN-GE-0803012, 2008.

[63] J. Pequenao, *Computer generated image of the ATLAS Muons subsystem* (2008), URL: https://cds.cern.ch/record/1095929.

[64] J. Pequenao, *Computer Generated image of the ATLAS calorimeter* (2008), URL: https://cds.cern.ch/record/1095927.

[65] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1*, CERN-LHCC-97-016, 1997.

[66] B. Abbott et al., *Production and Integration of the ATLAS Insertable B-Layer*, JINST **13**.05 (2018) T05008, arXiv: 1803.00844 [physics.ins-det].

[67] ATLAS Collaboration, *Studies of radial distortions of the ATLAS Inner Detector* (2018), URL: http://cds.cern.ch/record/2309785.

[68] ATLAS Collaboration, *ATLAS magnet system: Technical Design Report, 1*, 1997, URL: https://cds.cern.ch/record/338080.

[69] ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2*, JINST 15, 10, P10004 (2020).

[70] ATLAS Collaboration, *Alignment of the ATLAS Inner Detector in Run-2*, Eur. Phys. J. C **80**.12 (2020) 1194, arXiv: 2007.07624 [hep-ex].

[71] P. Brückman, A. Hicheur, and S.J. Haywood, *Global chi2 approach to the Alignment of the ATLAS Silicon Tracking Detectors*, ATL-INDET-PUB-2005-002, 2005, URL: https://cds.cern.ch/record/835270.

[72] ATLAS Collaboration, *ATLAS Computing: technical design report*, Technical Design Report ATLAS, 2005.

[73] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. **C77** (2017) 317, arXiv: 1611.09661 [hep-ex].

[74] ATLAS Collaboration, *2015 start-up trigger menu and initial performance assessment of the ATLAS trigger using Run-2 data*, ATL-DAQ-PUB-2016-001, 2016.

[75] ATLAS Collaboration, *Trigger Menu in 2016*, ATL-DAQ-PUB-2017-001, 2017.

[76] ATLAS Collaboration, *Luminosity determination in* pp *collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021 (2019), URL: https://cds.cern.ch/record/2677054.

[77] ATLAS Collaboration, *Luminosity determination in* pp *collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector at the LHC*, Eur. Phys. J. **C76** (2016) 653, arXiv: 1608.03953 [hep-ex].

[78] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13**.07 (2018) P07017.

[79] J. M. Katzy, *QCD Monte-Carlo model tunes for the LHC*, Progress in Particle and Nuclear Physics **73** (2013) 141 –187.

[80] T. Martini and P. Uwer, *The Matrix Element Method at next-to-leading order QCD for hadronic collisions: Single top-quark production at the LHC as an example application*, JHEP **05** (2018) 141, arXiv: 1712.04527 [hep-ph].

[81] S. Höche, *Introduction to parton-shower event generators* (2015) 235–295, arXiv: 1411.4085 [hep-ph].

[82] Y. I. Azimov et al., *Similarity of Parton and Hadron Spectra in QCD Jets*, Z. Phys. C **27** (1985) 65–72.

[83] B. Andersson et al., *Parton fragmentation and string dynamics*, Physics Reports **97**.2 (1983) 31 –145.

[84] T.Sjöstrand, *The Lund Monte Carlo for jet fragmentation and e+e- physics - jetset version 6.2*, Compt. Phys. Commun. **39**.3 (1986) 347 –407.

[85] I. Borozan and M.H. Seymour, *An Eikonal model for multiparticle production in hadron hadron interactions*, JHEP **09** (2002) 015, arXiv: hep-ph/0207283.

[86] ATLAS Collaboration, *Measurement of the Inelastic Proton–Proton Cross Section at $\sqrt{s} = 13\,TeV$ with the ATLAS Detector at the LHC*, Phys. Rev. Lett. **117** (2016) 182002, arXiv: 1606.02625 [hep-ex].

[87] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146.

[88] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph].

[89] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph].

[90] S. Frixione, P. Nason, and G. Ridolfi, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126, arXiv: 0707.3088 [hep-ph].

[91] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph].

[92] T. Gleisberg, S. Hoeche, M. Schoenherr, F. Siegert, J. Winter, *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph].

[93] T. Sjöstrand, M. van Zijl, *A multiple-interaction model for the event structure in hadron collisions*, Eur. Phys. J. **C36**.7 (1987).

[94] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. **C76**.4 (2016) 196, arXiv: 1512.01178 [hep-ph].

[95] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det].

[96] S. Agostinelli et al., *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[97] T. Sjostrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852–867, arXiv: 0710.3820 [hep-ph].

[98] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) 244, arXiv: 1207.1303 [hep-ph].

[99] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016.

[100] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph].

[101] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph].

[102] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014.

[103] S. Frixione et al., *Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations*, JHEP **04** (2007) 081, arXiv: hep-ph/0702198.

[104] P. Artoisenet et al., *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP **03** (2013) 015, arXiv: 1212.3460 [hep-ph].

[105] L.A. Harland-Lang et al., *Parton distributions in the LHC era: MMHT 2014 PDFs*, Eur. Phys. J. C **75**.5 (2015) 204, arXiv: 1412.3989 [hep-ph].

[106] C. Bierlich et al., *Robust Independent Validation of Experiment and Theory: Rivet version 3*, SciPost Phys. **8** (2020) 026, arXiv: 1912.05451 [hep-ph].

[107] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152–155.

[108] ATLAS Collaboration, *Studies on top-quark Montecarlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, 2016.

[109] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling with Sherpa and MG5aMCatNLO*, ATL-PHYS-PUB-2017-007, 2017.

[110] M. Beneke et al., *Hadronic top-quark pair production with NNLL threshold resummation*, Nucl. Phys. **B855** (2012) 695–741, arXiv: 1109.1536 [hep-ph].

[111] M. Cacciari et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. **B710** (2012) 612–622, arXiv: 1111.5869 [hep-ph].

[112] P. Bärnreuther, M. Czakon, and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to* $q\bar{q} \to t\bar{t} + X$, Phys. Rev. Lett. **109** (2012) 132001, arXiv: 1204.5201 [hep-ph].

[113] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **12** (2012) 054, arXiv: 1207.0236 [hep-ph].

[114] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP **01** (2013) 080, arXiv: 1210.6832 [hep-ph].

[115] M. Czakon, P. Fiedler, and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through* $O(\alpha_S^4)$, Phys. Rev. Lett. **110** (2013) 252004, arXiv: 1303.6254 [hep-ph].

[116] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930, arXiv: 1112.5675 [hep-ph].

[117] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph].

[118] A. D. Martin et al., *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189–285, arXiv: 0901.0002 [hep-ph].

[119] A. D. Martin et al., *Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections*, Eur. Phys. J. C **64** (2009) 653–680, arXiv: 0905.3531 [hep-ph].

[120] H.L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) 074024, arXiv: 1007.2241 [hep-ph].

[121] J. Gao et al., *The CT10 NNLO Global Analysis of QCD*, Phys. Rev. D **89** (2014) 033009, arXiv: 1302.6246 [hep-ph].

[122] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, CERN Yellow Reports: Monographs **2** (2017), arXiv: 1610.07922 [hep-ph].

[123] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7**.3 (2019) 034, arXiv: 1905.09127 [hep-ph].

[124] T. Cornelissen et al., *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, ATL-SOFT-PUB-2007-007, 2007, URL: https://cds.cern.ch/record/1020106.

[125] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, JINST (2014), URL: https://doi.org/10.1088/1748-0221/9/09/p09009.

[126] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262**.2 (1987) 444 –450.

[127] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047, 2012.

[128] ATLAS Collaboration, *Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-018, 2015, URL: https://cds.cern.ch/record/2037683.

[129] F. Meloni, *Primary vertex reconstruction with the ATLAS detector*, JINST **11**.12 (2016) C12060.

[130] ATLA Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, Eur. Phys. J. C **80**.1 (2020) 47, arXiv: 1909.00761 [hep-ex].

[131] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, JINST **15**.09 (2020) P09015, arXiv: 2004.13447 [physics.ins-det].

[132] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s} = 13\,TeV$, Eur. Phys. J. C **79** (2019) 639, arXiv: 1902.04655 [hep-ex].

[133] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, JINST **14** (2019) P12006, arXiv: 1908.00005 [hep-ex].

[134] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, JINST **14**.12 (2019) P12006, arXiv: 1908.00005 [hep-ex].

[135] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s} = 13\,TeV$, Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex].

[136] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2* pp *collision data set at* $\sqrt{s} = 13\,TeV$, Eur. Phys. J. C **81** (2020) 578, arXiv: 2012.00578 [hep-ex].

[137] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2* pp *collision data set at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **81**.7 (2021) 578, arXiv: 2012.00578 [hep-ex].

[138] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-*k_t *jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph].

[139] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at* $\sqrt{s} = 7$ TeV, Eur. Phys. J. C **73**.3 (2013), arXiv: 1112.6426 [hep-ex].

[140] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton–proton collisions at* $\sqrt{s} = 7$ TeV *with the ATLAS detector*, Eur. Phys. J. C **75** (2015) 17, arXiv: 1406.0076 [hep-ex].

[141] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014.

[142] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in* pp *collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector*, Eur. Phys. J. **C76** (2016) 581, arXiv: 1510.03823 [hep-ex].

[143] ATLAS Collaboration, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, Eur. Phys. J. C **77**.9 (2017) 580, arXiv: 1705.02211 [hep-ex].

[144] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2* pp *collision dataset* (2022), arXiv: 2211.16345 [physics.data-an].

[145] ATLAS Collaboration, *Commissioning of the ATLAS* b*-tagging algorithms using* $t\bar{t}$ *events in early Run 2 data*, ATL-PHYS-PUB-2015-039, 2015, URL: https://cds.cern.ch/record/2047871.

[146] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS* b*-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, 2017, URL: https://cds.cern.ch/record/2273281.

[147] ATLAS Collaboration, *Identification of Jets Containing* b*-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003, 2017, URL: https://cds.cern.ch/record/2255226.

[148] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **78** (2018) 903, arXiv: 1802.08168 [hep-ex].

[149] ATLAS Collaboration, E_T^{miss} *performance in the ATLAS detector using 2015–2016 LHC* pp *collisions*, ATLAS-CONF-2018-023, 2018, URL: https://cds.cern.ch/record/2625233.

[150] T. Chen and C. Guestrin, "XGBoost", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2016, URL: https://doi.org/10.1145%2F2939672.2939785.

[151] F. C. Geoffrey and S. Wolfram, *Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes*, Phys. Rev. Lett. **41** (23 1978) 1581–1585.

[152] J. H. Holland, *Genetic Algorithms*, Scientific American **267**.1 (1992) 66–73, URL: http://www.jstor.org/stable/24939139.

[153] A. Hocker et al., *TMVA - Toolkit for Multivariate Data Analysis with ROOT: Users guide.* CERN-OPEN-2007-007, 2007, arXiv: physics/0703039.

[154] Z. Marshall, *Simulation of Pile-up in the ATLAS Experiment*, J. Phys. Conf. Ser. **513** (2014) 022024.

[155] W. Buttinger, *Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data*, ATL-COM-SOFT-2015-119, 2015, URL: https://cds.cern.ch/record/2014726.

[156] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. **D96** (2017) 072002, arXiv: 1703.09665 [hep-ex].

[157] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2020) 689, arXiv: 2007.02645 [hep-ex].

[158] ATLAS Collaboration, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, Eur. Phys. J. C **77** (2017) 580, arXiv: 1705.02211 [hep-ex].

[159] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, JHEP **08** (2018) 089, arXiv: 1805.01845 [hep-ex].

[160] ATLAS Collaboration, *ATLAS* b*-jet identification performance and efficiency measurement with* $\mathrm{t\bar{t}}$ *events in* pp *collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **79** (2019) 970, arXiv: 1907.05120 [hep-ex].

[161] ATLAS Collaboration, *Measurement of* b*-tagging efficiency of* c*-jets in* $\mathrm{t\bar{t}}$ *events using a likelihood approach with the ATLAS detector*, ATLAS-CONF-2018-001, 2018, URL: https://cds.cern.ch/record/2306649.

[162] ATLAS Collaboration, *Calibration of light-flavour* b*-jet mistagging rates using ATLAS proton–proton collision data at* $\sqrt{s} = 13$ *TeV*, ATLAS-CONF-2018-006, 2018, URL: https://cds.cern.ch/record/2314418.

[163] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, ATLAS-CONF-2016-024, 2016.

[164] F.A. Berends et al., *On the production of a W and jets at hadron colliders*, Nuclear Physics B **357**.1 (1991) 32–64.

[165] ATLAS Collaboration, *The CLs method: information for conference speakers* (2011), URL: https://www.pp.rhul.ac.uk/~cowan/stat/cls/CLsInfo.pdf.

[166] *Proceedings of the 1974 CERN School of Computing: Godøysund, Norway 11 - 24 Aug 1974. 3rd CERN School of Computing* (1974), URL: https://cds.cern.ch/record/186223.

[167] J. Pequenao and P. Schaffner, *How ATLAS detects particles: diagram of particle paths in the detector*, CERN-EX-1301009, 2013.