GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Fakultät für Physik

Master's Thesis

Simulation und Optimierungsstudie für die Messung des $t\bar{t}H(b\bar{b})$ Prozesses im Semileptonischen Kanal mit $\sqrt{s} = 13$ TeV mit dem ATLAS Detektor

Simulation and optimization studies for the measurement of the $t\bar{t}H(b\bar{b})$ process in the semileptonic channel at $\sqrt{s} = 13$ TeV with the ATLAS detector

prepared by

Gaoyuan Wang

from Jiangsu

at the II. Physikalisches Institut

CERN-THESIS-2017-309

Thesis number:	II. Physik-UniGö-MSc-2017/04
Thesis period:	1st October 2016 until 18th September 2017
First referee:	Prof. Dr. Arnulf Quadt
Second referee:	Prof. Dr. Stan Lai

Zusammenfassung

Die Erzeugung eines Higgs-Bosons zusammen mit einem Top-Quark-Paar ($t\bar{t}H$) bietet die Möglichkeit, die Yukawa-Kopplung eines Top-Quarks direkt zu messen. Diese Arbeit beinhaltet zwei Studien mit dem Ziel, die Sensitivität der Messung und die Genauigkeit der Modellierung des $t\bar{t}H$ Prozesses im ATLAS-Experiment zu verbessern. In einer Optimierungsstudie wurden verschiedene Konfigurationen mit unterschiedlichen Jet-p_T-Schwellen nach ihrer Sensitivität untersucht. Alle Konfigurationen zeigen ein stabiles Fit-Ergebnis und keine verbesserte Sensivititität im Vergleich mit dem ICHEP-Ergebnis 2016. In einer Monte-Carlo-Generator-Studie wurde der Effekt des h_{damp}-Parameters und der Renormierungs- und Faktorisierungsskala in MadGraph untersucht. Die Änderung des h_{damp}-Parameters hat einen vernachlässigbaren Einfluss auf die Modellierung, während die Ereignisse, die mit unterschiedlichen Generatoren und Hadronisationsmodellen erzeugt sind, große Abweichungen voneinander zeigen. Die Studie der Renormierungs- und Faktorisierungsskala zeigt, dass die unterschiedliche Wahl der Skalen in MadGraph und Powheg eine wichtige Quelle für diese Abweichungen ist.

Abstract

The production of a Higgs boson in association with a top quark pair ($t\bar{t}H$) allows a direct measurement of the Yukawa coupling of the top quark. This thesis consists of two studies with the aim of improving the measurement sensitivity and modelling accuracy in the detection of the $t\bar{t}H$ process in the ATLAS experiment. In an optimization study, several sets of configurations for the jet p_T threshold are tested for sensitivity. Different configurations show robust behaviour and no improvement of the sensitivity can be observed compared to the ICHEP result in 2016. In a Monte Carlo generator study, the effect of the h_{damp} parameter in Powheg, and the renormalisation and factorisation scales in MadGraph, are investigated. The variation of the h_{damp} parameter has negligible effect on the events, while events produced with different generators and hadronisation models are found to show large deviations. The study on the renormalisation and factorisation scales shows that the different choices for the value of the scales in MadGraph and Powheg is one of the important origins of the deviation.

Contents

1 Introduction

The Standard Model of elementary particles (SM) is a quantum field theory describing the elementary particles and their interactions. The Standard Model, developed in the 1970s, has been tested through a variety of experiments. However it is known that there is no candidate for dark matter in the SM, which makes up most of the mass of the universe. It lacks also an explanation for other problems like the neutrino oscillations [1], the hierarchy problem [2] and the strong CP problem [3] in the SM. Models beyond the Standard model (BSM) are thus needed. Properties of the SM particles could be sensitive to BSM, so it is possible to find evidence for BSM processes in measurements of the SM particles. Such measurements can be made for instance at the Large Hadron Collider (LHC) at CERN.

With the discovery of the Higgs boson in 2012 by the ATLAS and the CMS collaborations with a mass of approximately 125 GeV [4, 5], the last component of the SM was found. A chain of experiments have been undertaken to study its properties at the LHC.

This project takes part in the ATLAS collaboration, in the search for the Standard Model Higgs boson production associated with a top quark pair. Since the mass of the fermions is determined by its Yukawa coupling to the Higgs boson, the top quark as the heaviest quark has a much larger coupling strength than all the other quarks. This makes the experiment on the top quark the most feasible measurement for the Yukawa coupling of a quark.

This thesis is structured as follows: Chapter 2 presents an introduction of the theoretical background. Chapter 3 gives a short overview of Monte Carlo (MC) simulation procedure. In chapter 4, the LHC and ATLAS experiment are introduced.

Chapter 7 -9 concentrate on the first part - the optimization study based on ICHEP results. The ICHEP result is the first result on the $t\bar{t}H(b\bar{b})$ analysis. It was presented at the ICHEP conference in 2016 [6]. The goal of the optimization study is to optimize the analysis to achieve a better sensitivity by applying different thresholds on the jet p_T. In this study, different sets of configurations are studied and compared to the ICHEP configuration.

In chapter 10 -11, the second part of this thesis, a MC generator study for $t\bar{t}H$ signal

process modelling is presented. In this part of the thesis, some MC generator free parameters are varied and the sensitivity of the events on these free parameters are analysed. The aim of this study is to obtain a better understanding of the performance of MC generators and the differences between various MC generators, for only then an accurate modelling of $t\bar{t}H$ signal process is possible.

2 Theoretical Background

2.1 The Standard Model

In the SM, the elementary particles are divided into two categories according to their spin: fermions and bosons. Fermions are characterised by Fermi-Dirac statistics and obey the Pauli principle. All fermions are particles with half-integer spin, for instance leptons and quarks. Bosons are characterised by Bose-Einstein statistics, so there is no restriction in the number of bosons in the same quantum state. Bosons are particles with integer spin; known bosons are photons, gluons, the Higgs boson, and W and Z bosons.

The quarks are categorised further into three generations and each generation is split again into up-type and down-type quarks. The up-type quarks have the electrical charge $+2/3$ while the down-type quarks have $-1/3$. Quarks participate in all four fundamental interactions, which are gravitation, electromagnetism, weak interactions and strong interactions. Electromagnetism is described by quantum electrodynamics (QED), the mediator of this interaction is the massless photon. The W and Z boson take over the role of mediator in the weak interaction. The quantum chromodynamics (QCD) describes the strong interaction with the gluon as mediator [7].

2.2 The Higgs Boson

The prediction of the existence of the Higgs boson and the Higgs mechanism appeared in 1964 [8, 9], suggested independently by different authors. The SM Higgs boson is a spin zero particle. The scalar potential of the Higgs field is:

$$V(\Phi) = m^2(\Phi^\dagger\Phi) + \lambda(\Phi^\dagger\Phi)^2, \tag{2.1}$$

with λ the Higgs self-coupling parameter.

The Higgs boson has a mass of $m_{higgs} \approx$125 GeV [10], which depends on the vacuum expectation value v (VEV) of the Higgs field via $m_{higgs} = \sqrt{2\lambda}v$. The VEV of the Higgs field depends directly on the Fermi coupling G_F which is known from the muon

lifetime measurement. The VEV is given by $v = (\sqrt{2}G_F)^{-1/2} \approx 246$ GeV [10]. There are four main Higgs production mechanisms at the LHC: gluon fusion, vector-boson fusion (VBF), Higgs-strahlung associated with a gauge boson and Higgs-strahlung associated with a pair of top/antitop quarks. The gluon-gluon fusion is, for the energy range of the LHC, the dominant process.

The dominant decay mode of the Higgs boson is $H \to b\bar{b}$ with a branching ratio of $57.7 \pm 1.9\%$. It is followed by $H \to WW^*\,(21.5 \pm 0.9\%)$, $H \to gg\,(8.57 \pm 0.87\%)$ and $H \to \tau^+\tau^-\,(6.32 \pm 0.36\%)$ [10]. The Higgs boson was observed by ATLAS and CMS decaying into $\gamma\gamma$, WW and ZZ bosons in 2012 [4, 5].

2.3 The Mechanism of Electroweak Symmetry Breaking

Elementary particles gain their masses through their coupling with the Higgs boson in the SM. This mechanism is called the Higgs mechanism [8, 9]. In this way the W and Z gauge bosons acquire mass without breaking the local gauge symmetry. It is very important that the fields are local gauge invariant because only local gauge invariant theories are renormalisable [11]. The Higgs mechanism involves the spontaneous symmetry breaking of a complex scalar field. A simple complex scalar field can be written as

$$\Phi = \frac{1}{\sqrt{2}}(\Phi_1 + i\Phi_2) \tag{2.2}$$

the vacuum at $(\Phi_1, \Phi_2) = (v, 0)$ with v the VEV [7]. Then the field is perturbed around its vacuum $(v, 0)$:

$$\Phi_1(x) = \eta(x) + v\,, \tag{2.3}$$

$$\Phi_1(x) = \xi(x)\,. \tag{2.4}$$

Embedding the perturbed field into the Lagrangian of a scalar field [7], the Lagrangian becomes:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) - \frac{1}{2}m_\eta^2\eta^2 + \frac{1}{2}(\partial_\mu\xi)(\partial^\mu\xi) - V_{int}(\eta, \xi) \tag{2.5}$$

with the interaction term

$$V_{int} = \lambda v\eta^3 + \frac{1}{4}\lambda\eta^4 + \frac{1}{4}\lambda\xi^4 + \lambda v\eta\xi^2 + \frac{1}{2}\lambda\eta^2\xi^2\,, \tag{2.6}$$

$$m_\eta = \sqrt{2\lambda v^2}\,. \tag{2.7}$$

This Lagrangian contains a massless Goldstone Boson ξ. To keep the Lagrangian invariant under the transformation

$$\Phi(x) \to \Phi'(x) = \exp(ig\chi(x))\Phi(x)\,, \tag{2.8}$$

a new gauge field $B_\mu(\partial_\mu \to D_\mu = \partial_\mu + igB_\mu)$ must be introduced in order to cancel out the extra terms [7]. Now the Lagrangian becomes:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) - \frac{1}{2}m_\eta^2\eta^2 + \frac{1}{2}(\partial_\mu\xi)(\partial^\mu\xi) - V_{int}(\eta,\xi) \tag{2.9}$$

$$- \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}g^2v^2B_\mu B^\mu + gvB_\mu(\partial^\mu\xi)\,, \tag{2.10}$$

and the gauge field becomes massive.

Before the symmetry breaking, all the four gauge bosons are massless. After the symmetry breaking, three of the four gauge bosons become massive and only the photon remains massless.

2.4 Top Quark and $t\bar{t}H$ Coupling

The top quark is an up-type quark and has an electric charge of $+2/3$ according to the SM. It is the heaviest quark, with a mass of $173.21 \pm 0.51(\text{stat.}) \pm 0.71$ (syst.) GeV [10], which is much heavier than all the other quarks. The top quark has a very short lifetime because of its high mass. This makes the top quark the only quark that decays before it can hadronise and can be observed as a quasi-free quark.

The $t\bar{t}H$ signal process is shown in figure 2.1 [6], where a Higgs boson produced in association with a top quark pair with a subsequent decay of $H \to b\bar{b}$ [12, 13]. The study in this thesis is performed on the semileptonic channel where the top quarks decay into b quarks and W bosons, one W boson decays into a electron or muon and a neutrino and the other W boson decays hadronically. The characteristic signature of this process in the detector is four b jets, one charged lepton and two (light flavour) jets with a neutrino. The dominant background of this analysis is $t\bar{t}$+jets. They are further categorised into $t\bar{t} + c\bar{c}$, $t\bar{t} + b\bar{b}$ and $t\bar{t}$ + light-jets, with $t\bar{t} + b\bar{b}$ the irreducible background. W/Z+jets,

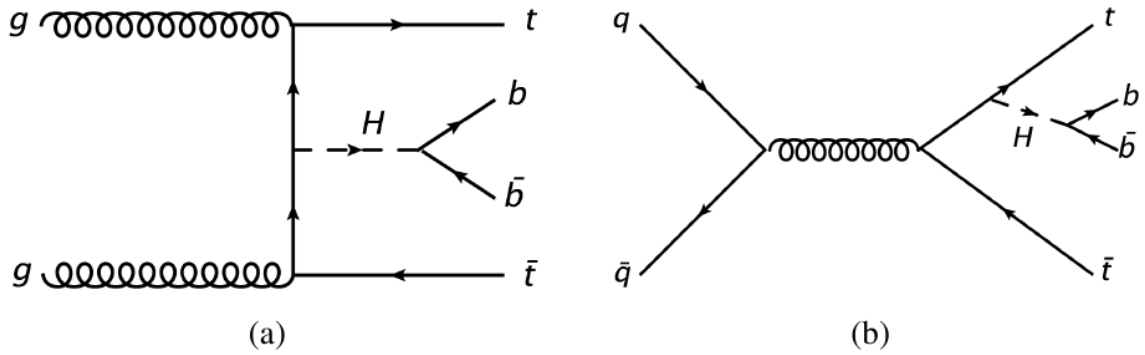

Figure 2.1: Feynman diagram of the $t\bar{t}H$ coupling.

diboson, single top, $t\bar{t}V$ and fake leptons which are objects from semileptonic decays b- and c-hadrons misidentified as prompt leptons are also important backgrounds [6].

3 Monte Carlo Generator

Simulations of collisions at the LHC with MC generators are very useful for both background modelling and signal searches. They are thus important for signal searches in particle physics at large experiments. The analysis of collisions at the LHC is challenging due to the large number of jets and missing energy from neutrinos. So for the $t\bar{t}H$ process, whose signal is much weaker than the (irreducible) background, an accurate simulation is essential for the measurement.

A QCD strong process at the LHC is simulated in different steps. First of all, the matrix element of a hard interaction will be calculated in perturbative QCD up to a certain order. Then QCD Bremsstrahlung will be calculated with a parton shower program and finally the fragmentation and hadronisation processes are simulated according to hadronisation models to match the events in the real experiments. MC generators contain some free parameters in the simulation of hadronisation processes. Next-to-leading order (NLO) perturbative QCD matrix element generator merged with Shower Monte Carlo programs, for instance Powheg [14] and MadGraph [15], interfaced with hadronisation programs such as Pythia [16] and Herwig [17–19], provide a fundamental tool for particle physics.

3.1 Matrix Element Generator

The matrix element generator (ME generator) calculates the matrix elements according to the perturbative QCD and describes the probability of the transition from an initial state into a final state. The QCD cross-section depends on the strong coupling constant which depends inversely on the energy scale. Events with high momentum transfer can be calculated perturbatively while the low energetic processes are simulated with the empirical parton distribution function (PDF). The PDF describes the probability density of finding a parton a carrying the momentum fraction x_a of a hadron i at momentum scale Q^2 [20]. The factorisation scale μ_F is the scale below which the perturbative calculation is no longer validated. For $\mu > \mu_F$, the cross-section is calculated perturbatively while

for $\mu < \mu_F$ the cross-section is calculated with the PDFs [20]:

$$\sigma = \frac{1}{2sx_a x_b} \sum_{a,b} \int_0^1 \mathrm{d}x_a \mathrm{d}x_b \int \mathrm{d}\Phi_n f_{i/a}(x_a, \mu_F) f_{j/b}(x_b, \mu_F) |\mathrm{M}_{ab\to f}|^2. \tag{3.1}$$

Here, s is the squared centre of mass energy, $|\mathrm{M}_{ab\to f}|$ is the matrix element with initial state particle a and b evolving into final state f. Φ_n is the phase space. ME generators usually include contributions up to NLO. The final state is generated accordingly with at most one additional emission (for NLO) carrying a certain energy and momentum. The final state particles can then be decayed with the branching ratios derived from the SM.

MadGraph5 [15] is for instance a general purpose matrix element based event generator. It generates tree-level matrix elements for a process with specified initial and final state particles and calculates cross-sections. The events are generated with e.g. the MadEvent [21] package included in MadGraph5.

Positive weight hard emission generator (Powheg) is a MC generator using the Powheg method [22] which produces positive-weighted events. PowhegBox [23] allows the simulation of a large number of processes at the LHC.

3.2 Parton Shower

Both initial and final states can radiate secondary partons and give branchings. These secondary partons again radiate and this process builds a parton shower [24]. These emissions are higher-order corrections to the matrix element and can be calculated with perturbative QCD approximations. The branching of one single parton into two partons or the radiation of a soft gluon needs to be simulated within perturbative theory before the hadronisation process is computed. The cross-section of an $n+1$ parton state σ_{n+1} can be treated as an n parton state σ_n with an additional emitted parton with momentum fraction z. The differential cross-section of a n parton state can be expressed via the absolute matrix element of this n parton state $|\mathrm{M}_n|$ and the n parton phase space Φ_n [24]:

$$\mathrm{d}\sigma_n \propto |\mathrm{M}_n|^2 \mathrm{d}\Phi_n. \tag{3.2}$$

The phase space of one particle radiation can be described using the azimuthal angle ϕ, the evolution variable t and the splitting variable z. The $n+1$ parton phase space can then be expressed via

$$\mathrm{d}\Phi_{n+1} = \frac{1}{4(2\pi)^3} \mathrm{d}\Phi_n \mathrm{d}t \mathrm{d}z \mathrm{d}\phi. \tag{3.3}$$

After the ϕ dependency is extracted, the $n+1$ parton state differential cross-section can be written as

$$\mathrm{d}\sigma_{n+1} = \mathrm{d}\sigma_n \frac{\mathrm{d}t}{t} \mathrm{d}z \frac{\alpha_S}{2\pi} P(z), \tag{3.4}$$

where $P(z)$ is a probability of an additional parton with momentum fraction z is emitted [24]. The integration of this function can only be evaluated within a certain range. It diverges in the soft limit $z \to 0$ and in the limit of very high energy. A lower momentum cut-off $\mathcal{O}(1\,\mathrm{GeV}^2)$ is implemented in the parton shower algorithm, below which partons do not branch any more [24]. In addition, the upper limit of the integration is set to the energy scale Q^2.

The probability that a parton does not branch from a lower bound t_0 to an upper energy scale t is given by Sudakov form factor [24]:

$$\Delta(t_0, t) = \exp\Big(- \int_{t_0}^{t} \frac{\mathrm{d}t'}{t'} \int \mathrm{d}z \frac{\alpha_s}{\pi} \hat{P}(z) \Big), \tag{3.5}$$

where $\hat{P}(z)$ is the unregularized splitting function. The Sudakov form factor can not be calculated up to an unlimited energy scale since that will lead to divergences. Using the Sudakov form factor, one can compute the evolution of the initial state with given virtual mass scale t_1 and momentum fraction x_1 over time and generate the next state with mass scale t_2 and momentum fraction x_2. The next parton can again radiate and generates a parton shower iteratively until the energy scale Q^2 is reached by t_2. The momentum of the emitted gluons can be computed from the initial state and final state virtual mass scale and momentum fraction [24]. For MadGraph, the real NLO correction is computed directly and for virtual corrections, MadGraph is interfaced with other tools to achieve fully QCD NLO calculation. The MadFKS framework [25] is implemented in MadGraph for FKS subtraction [26], which can be used to compute the cross-section at the NLO in QCD with the exception of one-loop matrix elements. MadLoop [25] computes the NLO one-loop matrix elements cross-section. Matching of the results to parton showers is done by the program MC@NLO [27, 28]. MadSpin [29, 30] calculates the spin correlation for decay and production of particles. After merging the matrix element calculation up to NLO with the parton showers, MadGraph5_aMC@NLO (MG5_aMC) [31] provides a very good event generator at NLO.

Powheg provides another method for NLO event generation. It generates the LO with the Born cross-section. It then calculates the probability of radiation according to the Sudakov form factor for each initial and final state particles to achieve NLO accuracy [22].

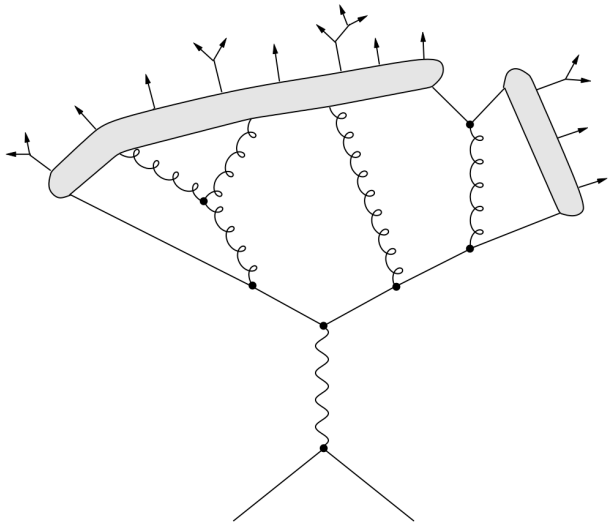

Figure 3.1: This figure demonstrates the basic idea of the string model as used in Pythia for the process $e^+e^- \to q\bar{q}$. Strings between quarks that are moving apart from each other are stretched and new quarks are generated from there. Hadrons are formed at the end of this process [24].

With the Powheg method, the hardest radiation (highest transverse momentum) is generated first with the computed NLO matrix element before soft radiations are generated.

3.3 Hadronisation

Partons are not physical objects and they must form hadrons after the showering process to become colour neutral due to colour confinement. This process can only be treated phenomenologically. The first step is the fragmentation, where partons are produced using different hadronisation models. Then these partons will be combined to form colourless hadrons. There are currently two hadronisation models describing the hadronisation process: the cluster model [32] and the string model [33].

Pythia, for instance, is a program using the string model to describe final state parton showers and fragmentations, where non-perturbative effects are important. This model uses a string to describe the confinement of the quarks as demonstrated in figure 3.1. The self interaction of the gluons keep the field potential within a narrow region like a string.

The string is stretched between the two quarks moving in different directions due to the strong coupling constant increasing linearly with distance. The energy stored in the string becomes larger with stretching until it is large enough for the creation of a $q\bar{q}'$ pair. Then the string breaks and the energy is released, which creates a new $q\bar{q}'$ pair. With the creation of a new quark pair and the distance between the new quark pair increases,

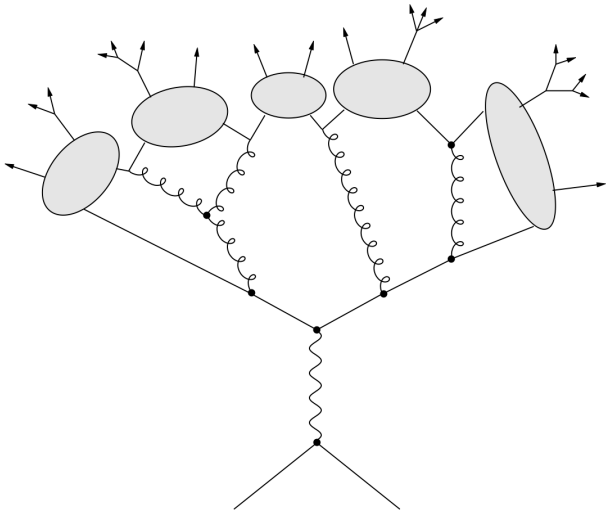

Figure 3.2: The cluster model as used by Herwig is demonstrated for the process $e^+e^- \to q\bar{q}$. Clusters are formed and decay into hadrons [24].

new strings are formed [24]. The strings will be again stretched and form new quark pairs until they all form hadrons and no more energy is available. If there should be a gluon in the system, the quarks will not be connected by the string directly but are connected to the gluon. The gluons are kinks in the string. The string model provides the best data agreement.

Herwig is an example that uses the cluster model as demonstrated in figure 3.2. The basic idea of this model is the preconfinement. Gluons are treated as colour-anticolour pairs. In this model partons will be generated and classified into colour singlet clusters according to their distance in the phase space [24]. Gluons in the system will first be split and form quark pairs using non-perturbative QCD and classified into clusters. The clusters then decay into hadrons as pure two-body decay.

3.4 Free Parameters in Monte Carlo Generator

MC generators work based on phenomenological models, which include a number of approximations. As an artificial effect of the approximations, there are free parameters in the generator whose value are not determined by theory. The value of these parameters are chosen with information drawn from experimental data. Although it is a very complicated process to tune the parameters, systematic tunings of Monte Carlo event generators [34] are quite successful in finding the best value.

Free parameters in MC simulation for a hard process are for instance the choice of

PDFset, the renormalisation and factorisation scales and h_{damp}. Varying these parameters has an effect on the events. Understanding the effects of the free parameters is very important for an accurate modelling of various processes.

4 Experimental Setup

4.1 LHC

The LHC located at CERN is the current largest particle accelerator and collider. It is installed in a 27 km tunnel located about 100 m under ground near Geneva, Switzerland [35]. Inside the two ultrahigh vacuum beam pipes, proton beams travel in opposite directions with very high energy before their collision at specific points around the ring. Superconducting electromagnets guide the beams and keep them focused. The operating temperature for these magnets is about 1.9 K to avoid loss of energy because of electrical resistance [36]. Liquid helium is used to cool the magnets.

There are 7 detectors on the LHC; the two general purpose detectors are ATLAS and CMS. Both detectors were built to look for new physics including signs for extra dimension and dark matter. One of their major tasks is also to study the properties of the Higgs Boson. Although they have the same scientific goals, they use different detection technologies. The acceleration of a proton at CERN is done in several steps as shown in figure 4.1 [37].

1. Protons from hydrogen atoms are sent from the 30 m long linear accelerator Linear Accelerator 2 (Linac2) with an energy of 50 MeV into the Proton Synchrotron Booster (PSB). Inside the synchrotron the protons are accelerated to 1.4 GeV.

2. Then the protons are injected into the proton synchrotron (PS) and will be accelerated to 25 GeV.

3. From the PS the protons are sent to the super proton synchrotron (SPS), which is a 6.9 km long circular accelerator. Here the protons are accelerated to 450 GeV.

4. After it is accelerated by the SPS, the protons will reach the LHC and will be filled in both directions. Protons will be accelerated to the desired nominal centre of mass energy of maximal 7 TeV and circle the LHC beam pipe in bunches.

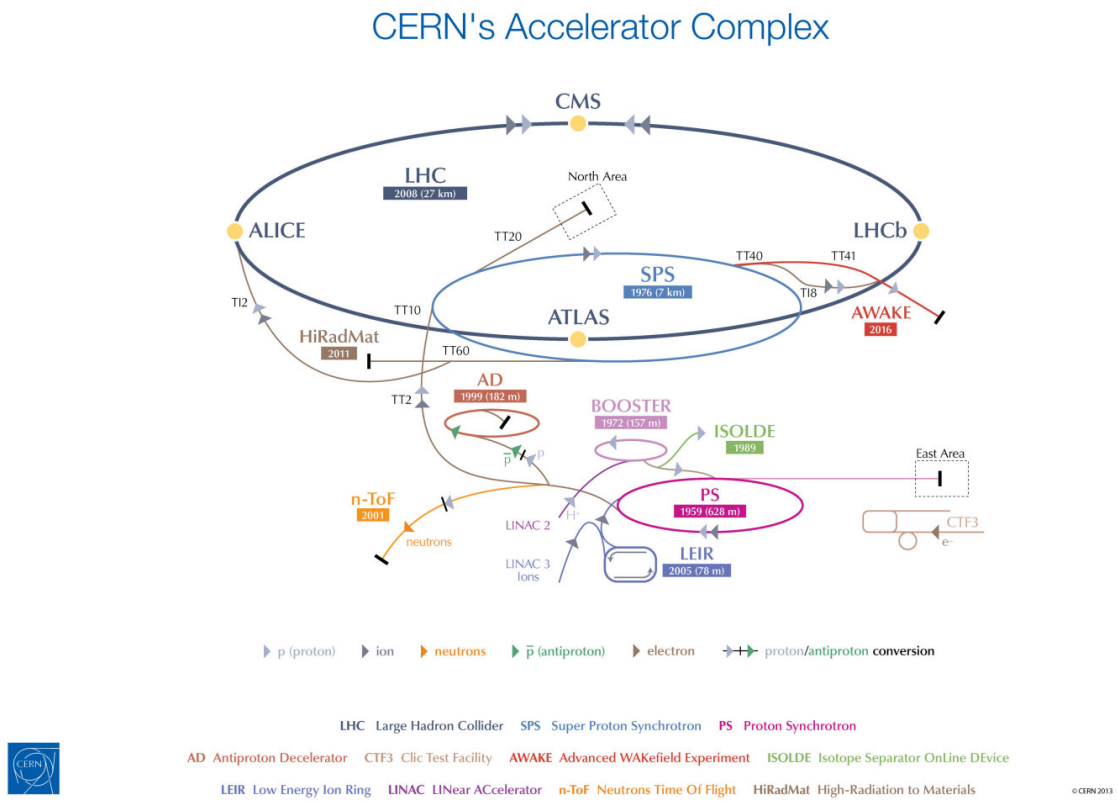

Figure 4.1: Different accelerators at CERN are shown in this plot.

The beams are guided to collide at the locations of the four particle detectors around the accelerator ring: A Toroidal LHC ApparatuS (ATLAS) detector, Compact Muon Solenoid (CMS), A Large Ion Collider Experiment (ALICE) and Large Hadron Collider beauty (LHCb). The other three small experiments at CERN are Large Hadron Collider forward (LHCf), Monopole and Exotics Detector at the LHC (MoEDAL) and Total Elastic and diffractive cross-section Measurement (TOTEM).

The first research run at the LHC started in 2009 and ended in 2013, with a centre of mass energy of $\sqrt{s} = 7-8$ TeV. The second run started in May 2015, after an upgrade, and ran at an increased centre of mass energy of $\sqrt{s} = 13$ TeV. The instantaneous luminosity of a collider is defined as the number of collisions per time per area. The event rate $\dot{N}$, which is the numbers of such event per unit time, can be expressed via

$$\dot{N} = \sigma \cdot L\,, \tag{4.1}$$

$$\sigma = \frac{1}{L}\frac{dN}{dt}\,, \tag{4.2}$$

$$\frac{d\sigma}{d\Omega} = \frac{1}{L}\frac{d^2N}{d\Omega \cdot dt}, \tag{4.3}$$

where σ is the cross-section for this process and L the luminosity.

Since in the LHC, protons are brought to collision in bunches, the luminosity reads:

$$L = \frac{n \cdot N_1 \cdot N_2 \cdot f}{2\pi \Sigma_x \Sigma_y}, \tag{4.4}$$

where N_1 and N_2 is the number of particles in each bunch, n is the number of bunches, the beam crossing frequency is given by f and Σ_x, Σ_y are the horizontal and vertical beam widths [38].

4.2 ATLAS

ATLAS takes the name from its toroidal magnet, Atlas is also the name of the Titan who was punished to hold the Earth on his shoulders. The detector is 44 m long and 25 m high. It has a cylindrical shape with two openings at both sides. The two openings allow protons and ions to enter the detector from both sides and the cylindrical shape ensures a full space coverage of new particles, which fly out in all directions after the collision [39].

Several subdetectors are arranged concentrically in layers to measure the paths, momentum and energy of the particles. The four major components are the inner detector tracking system, the electromagnetic calorimeter, the hadronic calorimeter and the muon spectrometer. Information drawn from these components is used to identify and reconstruct the particles after the collision. The inner detector, embedded in a solenoidal magnetic field is a tracking system used to measure the momenta of charged particles and to identify the interaction vertices. The electromagnetic and hadronic calorimeters are used to measure the energy of neutral and charged particles. All the particles, except for muons and neutrinos are stopped in the calorimeter. Muons have long lifetimes and interact only rarely with the calorimeter. The muon spectrometer identifies them and measures their momentum.

4.2.1 Coordinate System

To describe tracks within the detector, a reference system is needed. ATLAS uses a Cartesian right-handed coordinate system, with the origin set at the nominal interaction point (IP). The three Cartesian coordinates are defined as:

- X-axis is parallel to the ground and perpendicular to the beam direction, the positive direction points from the IP to the centre of the LHC ring

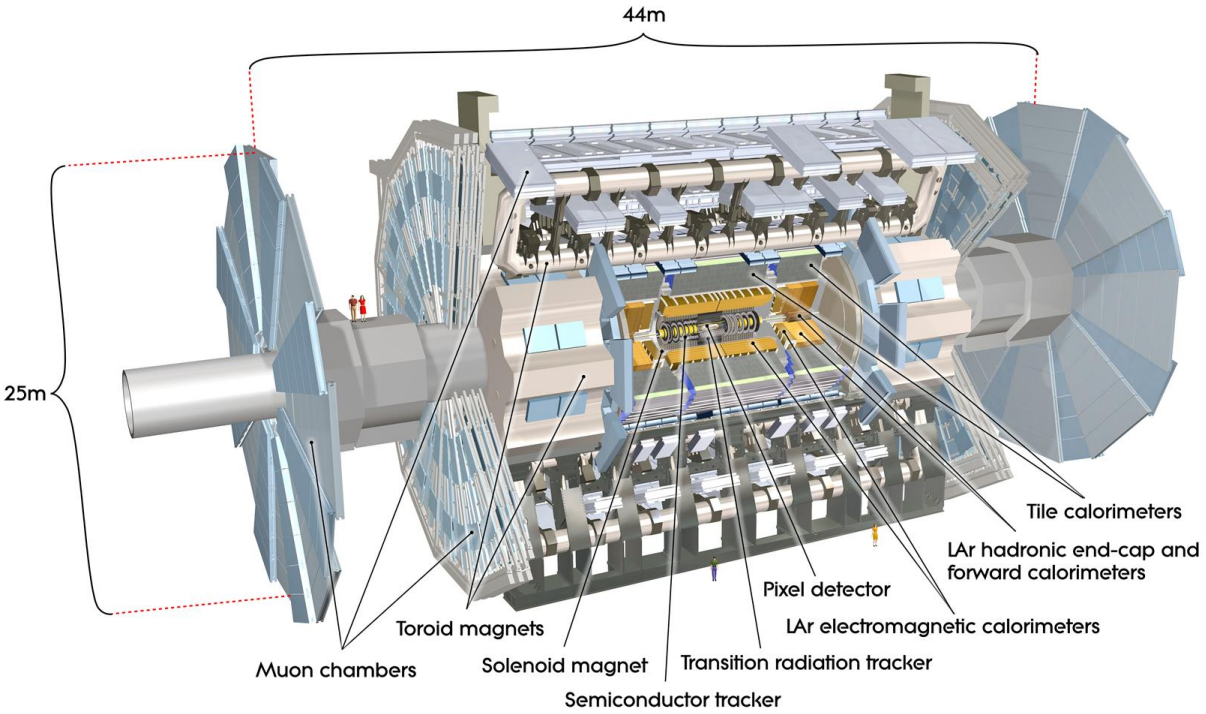

Figure 4.2: ATLAS detector with its main subdetectors is shown in this plot [39].

- Y-axis is perpendicular to the ground and points upwards

- Z-axis is parallel to the ground and parallel to the LHC ring, the positive direction is defined to be along the anti-clockwise beam direction [39].

The polar angle $\theta \in [0, 2\pi)$ is defined with respect to the Z-axis. The horizontal components of the momenta are unknown, while the transverse components of variables like energy and momentum are interesting. The transverse component is defined as the projection on the XY plane. In terms of polar angle $\theta \in [0, 2\pi)$, transverse energy and momentum read:

$$E_T = E\sin(\theta), \quad p_T = p\sin(\theta). \tag{4.5}$$

The rapidity y defined as

$$y \equiv \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right) \tag{4.6}$$

is preferred to the polar angle describing the polar distribution. The pseudorapidity η:

$$\eta \equiv -\ln\left(\tan\frac{\theta}{2}\right) \tag{4.7}$$

approaches the rapidity for vanishing particle mass [39].

4.2.2 Inner Detector

The inner detector measures the momentum of each charged particle. The inner detector consists of three sub-detectors: Pixel Detector, Semiconductor Tracker (SCT) and Transition Radiation Tracker (TRT) [39, 40]. The pixel detector performs a very high-precision measurement, it is the closest sub-detector to the IP. It is designed to determine the impact parameter and measure the particles with short lifetime before they decay.

4.2.3 Calorimeter

The electromagnetic calorimeter in ATLAS is a Liquid Argon calorimeter based on the sampling technique. It uses lead as the absorber material and liquid-Argon as the active medium. It measures the energy of electrons and photons covering the pseudorapidity interval $|\eta| < 3.2$ divided into the barrel region ($|\eta| < 1.475$) and the end-cap region ($1.375 < |\eta| < 3.2$) [40].

The hadronic calorimeter is the Tile Hadronic Calorimeter measuring the energy of hadrons based on sampling techniques. It uses scintillator tiles and Liquid Argon (LAr) as the active medium, and steel, copper, or tungsten as the absorber material. It covers a range of $|\eta| < 4.9$.

The forward calorimeter is a LAr sampling calorimeter located near the incident opening close to the IP. It covers the space around the beam and a rapidity region of $3.2 < |\eta| < 4.9$ to catch high p_T particles from the collisions and is important for missing transverse energy measurement and jet tagging [41].

4.2.4 Muon Spectrometer

The muon spectrometer is used to identify and measure the momentum of muons, which fly through the inner detector and calorimeter almost without energy loss. The spectrometer is made up of 4000 muon chambers separated into further subsections. The muon spectrometer has a fast response trigger system using information from the tracking chambers [39]. It also has tracking chambers for precise measurements of the properties of muons, this includes Monitored Drift Tubes and Cathode Strip Chambers.

4.2.5 The Magnet System

The magnet system bends charged particles in different layers for the momentum measurement. The magnet system consists of three main sections of the magnet system: Central Solenoid Magnet, Barrel Toroid and two end-cap Toroids [40]. The Central Solenoid Magnet uses the superconducting cable to generate a magnetic field of 2 T. It is placed around the inner detector. The Barrel Toroid consists of eight separate coils held by eight rings of struts situated outside the calorimeters. The whole system weighs over 830 tons. It generates a 4 T magnetic field with over 100 km superconducting wire [40]. The working point temperature of the Barrel Toroid is 4.7 K. The end-cap Toroids consist of two modules similar to the Barrel Toroid located around the two openings of ATLAS.

4.2.6 The Trigger System

ATLAS has a powerful data readout system containing the trigger system and the data acquisition system. Trigger-devices are placed around calorimeters and muon chambers. Proton collisions happen with a rate of maximal 40 MHz, the trigger system selects interesting events down to a storage rate of 1000 Hz [42]. This is realized using a multi-level trigger. The first level trigger uses signals from muon chambers and calorimeters to determine regions of interest. Information from the first level trigger is then passed to the second level trigger, it uses full information from all detectors to find physical objects. Then the event filter as the third level trigger uses full information from different detectors to perform a more detailed full event analysis [42]. After the three levels of selection, data are passed by the data acquisition system from the detector to the storage for further offline analysis.

5 Object Reconstruction

ATLAS uses information recorded in the detectors and software frameworks to reconstruct particles passing through the detector. The physical objects that are relevant in this analysis are electrons, muons, jets and jets containing b hadrons (b jets). Also missing transverse momentum is used in the event reconstruction. The missing transverse momentum is defined as the negative sum of transverse momentum of all physical objects in the event.

5.1 Electrons

Electrons are reconstructed from energy deposits in the electromagnetic calorimeter [43]. The recorded energy deposit needs to be associated to reconstructed tracks in the inner detector. The tight likelihood identification criteria [44] are used in this analysis. The candidate must have $p_T > 25\,\text{GeV}$ for the 2015 and 2016 dataset. The transverse and longitudinal impact parameter must fulfil $|z_0 \sin\theta| < 0.5$ mm and $|\frac{d_0}{\sigma(d_0)}| < 5$. Candidates from the region $1.37 < |\eta| < 1.52$ are excluded. The electrons are required to be isolated and they are selected using the method described in [45] [1]. This reduces the background from non-prompt electrons [6].

5.2 Muons

Muon reconstruction makes use of track segments in the muon spectrometer, this should be associated with recorded tracks in the inner detector [46]. All muon candidates must have an isolation of $\Delta R > 0.4$ from the next selected jet. The muon must have p_T >25 GeV and $|\eta| < 2.5$. They are also required to pass the gradient isolation working point, with $d_0 < 3$ and $|z \sin\theta| < 0.5$ mm.

[1] The current official working points settings can be found under `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/IsolationSelectionTool`.

5.3 Jets

Jets are reconstructed using energy deposits in the calorimeter with the anti-kt algorithm [47]. The radius of the cone for reconstruction is set to 0.4. After the reconstruction, a calibration and correction on jets is applied based on energy and η derived from simulation and data [48, 49]. Slightly different p_T thresholds are applied in various configurations. Jets within $\Delta R < 0.2$ of a selected electron will be removed. Events with jets from detector noise and non-collision sources are removed using the BadLoose operating points [50]. For jets with $p_T < 60$ GeV and $|\eta| < 2.4$, a cut is applied based on the Jet Vertex Tagger in order to reduce the number of pileup jets [51].

Jets are identified as b jets based on the information from the impact parameter and secondary and tertiary decay vertices, which is well developed in several algorithms [52]. They are based on the soft electrons or muons as decay products of hadrons containing a b-quark and on the characteristic lifetime of b hadrons. The lifetime of b hadrons leads to a characteristic mean distance between the vertex of decay and the primary collision point. The algorithm used in this analysis is the multivariate algorithm MV2 [52]. A working point with an efficiency of 70% is used. In the simulation, b jets are identified by matching the jet to particles on the particle level with $\Delta R = 0.3$ [6].

6 Dataset and Event Generation for the Optimization Study

The data used in this analysis is the dataset collected with the ATLAS detector at 13 TeV with a integrated luminosity of 13.2 fb^{-1} consisting of the full dataset from 2015 and a dataset from April to June 2016. The event generation assumes a top quark mass of 172.5 GeV and the Higgs boson mass to be 125 GeV. The $t\bar{t}H$ cross-section is calculated up to NLO in this study [53, 54].

6.1 Signal Modelling

The signal $t\bar{t}H$ modelling uses MG5_aMC [31] for the NLO matrix element calculation. The events are then interfaced with Pythia8 [16] with A14 settings [55] to model parton shower and hadronisation. The PDFset NNPDF3.0 [56] is used. The renormalisation and factorisation scales are set to $H_T/2$, where H_T is defined as the sum of transverse masses $m_T = \sqrt{p_T^2 + m^2}$ over all particles in the final state. The SM branching ratios of the decay of the Higgs boson and signal cross-section are assumed. The uncertainty of the signal modelling is estimated by comparing the nominal samples with the alternative sample using Herwig++ [18].

6.2 Background Modelling

The major backgrounds are introduced in section 2.4. The dominant background $t\bar{t}$+jets is generated with the Powheg-Box v2 NLO generator [57] with the PDFset CT10 [58]. The parton shower and hadronisation modelling is done by Pythia6 [59]. The decay of bottom and charm hadrons are simulated by EvtGen [60] and then normalized to the theoretical next to next to leading order (NNLO) cross-section [61]. The samples are then reweighted. The Wt and single top quark (s-channel) are simulated using Powheg-Box 2.0. A single top quark (t-channel) sample is generated with Powheg-Box v1 [62, 63]. The parton shower and hadronisation for single top quark samples are done using Pythia6.

The single top quark samples are then normalized to the theoretical NNLO cross-section [64], [65]. The W/Z+jets and diboson are simulated using Sherpa 2.1 [66]. The W/Z+jets samples are then normalized to the theoretical NNLO cross-section [67]. $t\bar{t}V$ events are generated using MG5_aMC, the PS and hadronisation is simulated by Pythia8. The fakes background is derived using a data-driven technique with a matrix method [68].

7 Analysis Strategy

Since the characteristic signature of the semileptonic channel is four b jets, one charged lepton and two (light) quarks with a missing neutrino, the event should have one selected lepton with $p_T > 25$ GeV [6] and at least four jets with at least two of them tagged as b jets.

7.1 Signal and Background Regions

To better constrain the backgrounds and distinguish them from the $t\bar{t}H$ signal, three regions are defined with the highest signal to background ratio as the signal regions, and the rest as the control regions. The signal regions should have enhanced signal compared to the background with a signal to background ratio $S/B > 1\%$ and a significance $S/\sqrt{B} > 0.3$ [6]. $S/\sqrt{B}$ is the ratio of the signal and the statistical error [1] on background which grows with growing statistics. The other regions are control regions. They are dominated by backgrounds and are used to constrain the systematic uncertainties on the background prediction. The nine regions are:

- $4j, 2b$: 4 jets and 2 b jets, control region.
- $4j, 3b$: 4 jets and 3 b jets, control region.
- $4j, 4b$: 4 jets and 4 b jets, control region.
- $5j, 2b$: 5 jets and 2 b jets, control region.
- $5j, 3b$: 5 jets and 3 b jets, control region.
- $5j, \geq 4b$: 5 jets and at least 4 b jets, signal region.
- $\geq 6j, 2b$: At least 6 jets and 2 b jets, control region.
- $\geq 6j, 3b$: At least 6 jets and 3 b jets, signal region.

[1] The significance is defined as the signal over standard deviation of background, here the systematic uncertainty on the background is small.

- $\geq 6j, \geq 4b$: At least 6 jets and at least 4 b jets, most important signal region.

The regions are statistically independent and will be combined in the fit to constrain systematic uncertainties and improve the overall sensitivity [6]. The kinematic variable H_T is used in the control regions and multivariate techniques are employed in the signal regions. The three signal regions contain a much higher number of signal events compared to the control regions, but even the regions with the most signal still have a small S/B ratio. This is due to the fact that $t\bar{t}H$ is a very rare process, the yield of the $t\bar{t}H$ process is about a factor of hundred lower than that of the $t\bar{t}$ process in the signal regions. For this reason a Neural Network (NN) is trained independently in each signal region to separate the $t\bar{t}H$ signal from the background.

7.2 Neural Network Training

The NeuroBayes package, developed by Phi-T Physics Information Technologies GmbH is used for the multivariate analysis. This algorithm uses neural network techniques and Bayesian statistics. NeuroBayes consists of two parts: NeuroBayesTeacher and NeuroBayesExpert [69]. The training of the neural network is done by Teacher using Monte Carlo samples or the data-derived background sample. Then it will be applied on real data using Expert to draw a prognosis. NeuroBayes can be directly implemented into the C++ interface. Using the input variables put in, NeuroBayes will assign weights to input variables in the hidden layer and define the new variable NN output in the output layer. The hidden layer has two more nodes than the input layer. The new variable NN output, which in this study is between -1 (background sample) and 1 (signal sample) will have a much larger separation power between signal and background. Ten variables are usually sufficient for a maximal signal-to-background separation. Too many input variables could cause an overtraining problem. $t\bar{t}$+jets is used as background and $t\bar{t}H$ as signal. After the NN is trained, the performance of the training has to be evaluated through cross validation tests and overtraining tests. The sample is split into two samples according to its even/odd event number. The overtraining test ensures that the NN is not biased to the sample used for the training. The cross validation test ensures that the two samples due to the even/odd splitting can both be used when analysing real data. Highly correlated variables will be removed.

8 Statistical Methods, Systematic Uncertainties and Official Result

8.1 Profile Likelihood Fit Procedure

A profile likelihood-based analysis is used in this study. The likelihood function is a product of Poisson measurements in control regions P_{CR} and signal regions P_{SR} times a function describing the systematics C_{syst} [70]

$$L(\mu,\theta) = P_{SR} \times P_{CR} \times C_{syst} \tag{8.1}$$

$$= L_{Pois}(\mu,\theta) \times \prod_{p} f(\theta_p)\,, \tag{8.2}$$

where $\mu = \frac{\sigma}{\sigma^{SM}}$ is the parameter of interest, θ_ps are the nuisance parameters and $f(\theta_p)$ represents the functional form describing the nuisance parameters. Here, $\mu = 0$ corresponds to the background only hypothesis and $\mu = 1$ corresponds to the SM Higgs boson prediction [70]. The three parameters of interest are

- the ratio of the measured $t\bar{t}H$ cross-section to the SM expectation: $\mu_{t\bar{t}H}$,
- $t\bar{t}+ \geq 1b$ and $t\bar{t}+ \geq 1c$ normalisation factors: $k(tt+ \geq 1b)$ and $k(tt+ \geq 1c)$.

The estimation of the uncertainty of μ is done by calculating the likelihood values as a function of μ. The 1σ band corresponds to the point where the logarithm of the likelihood decreases by 0.5 compared to its maximum.

A useful method to estimate the sensitivity is to fit the sample to the Asimov dataset. The Asimov dataset is defined so that the observed number of events are set equal to the predicted one. The simulated dataset is replaced by a single representative one, for which the results after fitting should be $\theta = 0$ and $\mu = 1$. The uncertainty of the fit result to the Asimov dataset of different configurations is compared with each other to draw conclusions on the sensitivity of this configuration.

The nuisance parameters are determined during the fit according to the corresponding

systematic uncertainties to maximize the likelihood function. The fit shifts the nuisance parameters to achieve a better data/Monte Carlo agreement.

8.2 Systematic Uncertainties

Table 8.1 lists the systematic uncertainties that are considered in this analysis. The systematic uncertainties are categorised into experimental uncertainties, which include the uncertainties related to the detector and the object reconstruction and identification, and modelling uncertainties, which are related to the modelling of parton shower and hadronisation processes, cross-section and normalisation. Some uncertainties only effect the overall normalisation, these are denoted with an "N". A "SN" means that this uncertainty affects both the shape of the discriminant distributions and normalisation. One systematic uncertainty could be split into more than one component to achieve a more precise treatment.

8.3 ICHEP Results

An analysis using a 25 GeV cut on all the jets was presented at the ICHEP conference in 2016 [6]. This set of threshold configurations is called the ICHEP configuration. Figure 8.1 shows the signal to background ratio (S/B) and significance $(S/\sqrt{B})$ for the nine regions and the percentage of different background in each region using the ICHEP configuration[1]. The ratio $\mu_{t\bar{t}H} = \frac{\sigma_{t\bar{t}H}}{\sigma^{SM}_{t\bar{t}H}}$ is found to be $2.1^{+1.0}_{-0.9}$ if the Higgs mass is assumed to be 125 GeV [6].

[1]The plots are made by us using the ICHEP configuration, they should be the same as the result published in the official paper [6].

Systematic uncertainty	**Type**	**Components**
Luminosity	N	1
Object reconstruction		
Pileup	SN	1
Jet energy scale	SN	19
Jet energy resolution	SN	1
Jet vertex tagger	SN	1
Missing transverse momentum	SN	3
Electron ID+reco+isolation+trigger	SN	5
Electron energy scale, resolution	SN	2
Muon ID+reco+isolation+trigger	SN	6
Muon momentum, resolution	SN	3
b-tagging efficiency	SN	5
c-tagging efficiency	SN	4
Light jet tagging efficiency	SN	14
High-p_T tagging efficiency	SN	2
Modelling systematics		
$t\bar{t}$ cross-section	N	1
$t\bar{t}$ + heavy flavour normalisation	N	2
$t\bar{t}+ \geq 1b$ NLO shape	SN	10
$t\bar{t}+ \geq 1c$ NLO shape	SN	1
$t\bar{t}$ modelling: residual radiation	SN	3
$t\bar{t}$ modelling: residual NLO generator	SN	3
$t\bar{t}$ modelling: PS + hadronisation	SN	3
$t\bar{t}$ NNLO reweighting	SN	4
$t\bar{t}V$ cross-section	N	4
$t\bar{t}V$ modelling	SN	2
$t\bar{t}H$ cross-section	N	2
$t\bar{t}H$ branching ratios	N	4
$t\bar{t}H$ modelling	SN	2
W+jets normalisation	N	6
Z+jets normalisation	N	6
Single top cross-section	N	2
Single top model	N	2
Diboson normalisation	N	1
Fakes normalisation	SN	7

Table 8.1: The list shows the systematics included in this analysis [6].

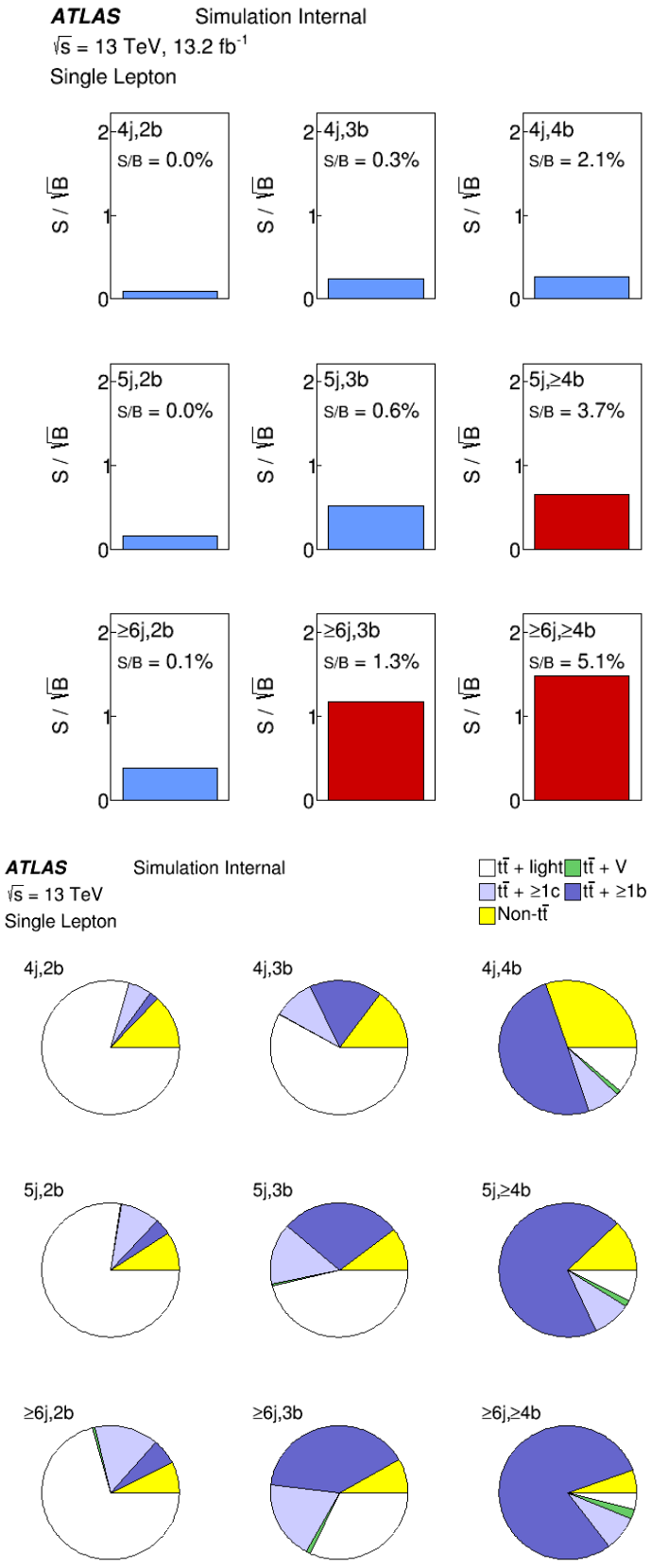

Figure 8.1: The upper plot shows the S/B and $S/\sqrt{B}$ for all nine regions and the lower plot shows the relative contributions of the various backgrounds for the nine regions.

9 Optimization Study on the ICHEP Analysis

The aim of this study is to find a configuration for the jet p_T threshold with improved sensitivity on the basis of the ICHEP results. Five different configurations are analysed in two studies. The first part section 9.1-section 9.5 presents the results for configurations with slightly higher jet p_T threshold for hard jets and in some cases, lower jet p_T threshold for light jets. This analysis will be done with a multivariate analysis. The multivariate technique and the fitting procedure will be described in detail. Then the second part, section 9.6, provides the results on two further configurations with very high p_T cuts on the hard jets.

9.1 Optimization of Selection

The first part studies three configurations,

- CONFIG 1: 35 GeV for the first to the fourth jet (ordered in p_T) and 20 GeV for the rest.
- CONFIG 2: 40 GeV for the first and second jets, 30 GeV for the third and fourth jet and 20 GeV for the rest.
- CONFIG 3: 35 GeV for the first to the fourth jet and 25 GeV for the rest.

These three configurations were suggested because the hard jets from the signal process $t\bar{t}H$ are more energetic than the hard jets from the background processes. So, tightening the jet p_T of the first jets will reject more background. The light jets in the final state are very likely to be decay products of one of the W bosons. Loosening the jet p_T threshold for the last ones will include more jets from the W decay and thus improve the final state reconstruction. This consideration can be justified with the jet p_T distributions from samples where particles are reconstructed from decay products using on tracking, vertexing

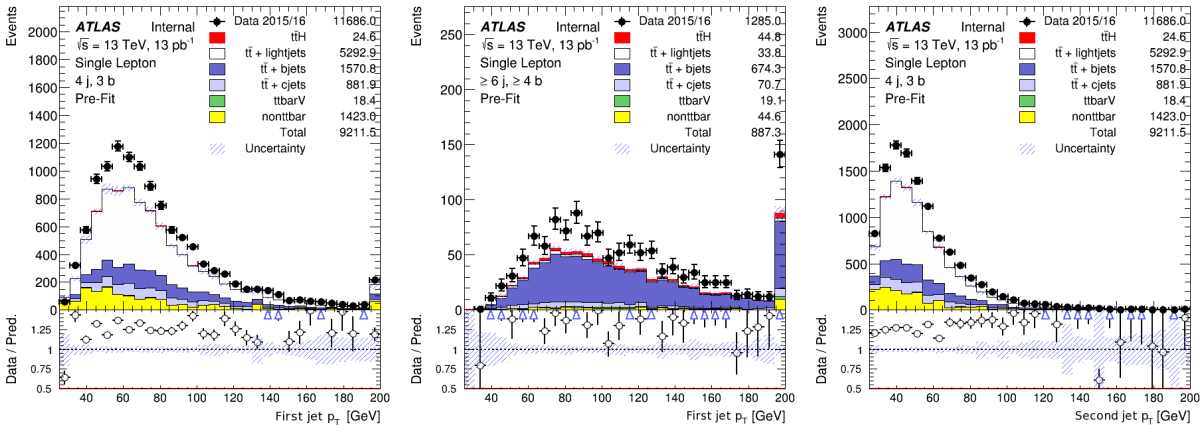

Figure 9.1: The left plot is the jet p_T distribution for the first jet in $(4j, 3b)$ region in ICHEP. The centre one shows the jet p_T distribution for the first jet in $(\geq 6j, \geq 4b)$ region in ICHEP and the right one shows the second jet in $(4j, 3b)$ region in ICHEP.

and particle identification information (reconstruction level), as shown in figure 9.1 and figure 9.2[1].

Six representative jet p_T distributions of CONFIG 1 and ICHEP are shown in figure 9.1 and figure 9.2 to demonstrate the events excluded and included by the new configurations. The S/B and $S/\sqrt{B}$ of the three configurations are shown in figure 9.3.

In figure 9.1 one can see the cross-section as a function of the first jet (left) and second jet (right) p_T in the background dominated region $(4j, 3b)$ for the ICHEP configuration. The peak of the distribution is around 60 GeV and 40 GeV. There are only a few $t\bar{t}H$ signal jets below 60 GeV in the signal region $(\geq 6j, \geq 4b)$ and the distribution has its maximum around 80 GeV (centre). The suggested new configurations have higher jet p_T thresholds for the first jets, this will cut out the jets between 25 GeV and the new threshold around 40 GeV from the ICHEP events. Based on this, it is expected that a threshold at $30 - 40$ GeV for the first jets will exclude, in majority, jets from the backgrounds.

Figure 9.2 provides the jet p_T distribution of the last jets in CONFIG 1. These are the jets that are included by CONFIG 1 but excluded in ICHEP events. As one can see in the figures, the decreased threshold includes more signal processes. In the background dominated regions, loosening the threshold of the light jets leads to more included $t\bar{t}$+light-jets. Thus, it is helpful for the modelling of $t\bar{t}$ + light-jets and final state reconstruction.

[1]The truth level jet p_T distributions would be an even stronger justification for the suggestion of the new configurations. But the truth level samples do not exist for $t\bar{t}$ process, which is the most important background for the $t\bar{t}H$ signal process. Because of this, the suggestion of the new configurations will be demonstrated based on the reconstruction level jet p_T distributions in this study.

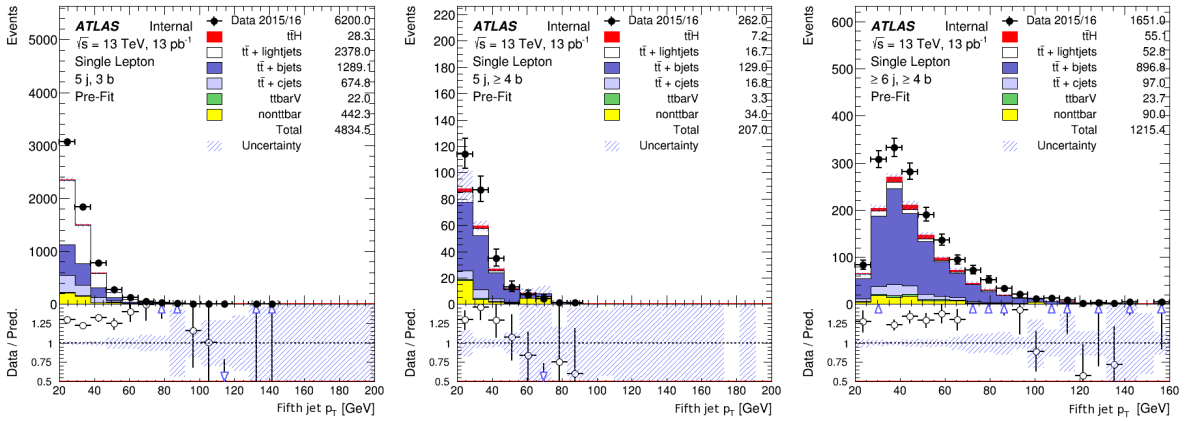

Figure 9.2: Jet p_T distribution for the fifth jet in region $(5j, 3b)$ (left), $(5j, \geq 4b)$ (centre) and $(\geq 6b, \geq 4b)$ (right) for CONFIG 1 is shown.

	ICHEP	CONFIG 1	CONFIG 2	CONFIG 3
$\mu_{t\bar{t}H}$	1.01 +4.12 -4.33	1.01 +4.61 -4.94	1.06 +4.51 -5.05	0.92 +4.65 -4.83
$k(tt+\geq 1b)$	1.00 +0.33 -0.30	1.00 +0.38 -0.33	1.00 +0.37 -0.31	1.01 +0.38 -0.35
$k(tt+\geq 1c)$	1.00 +0.74 -0.63	1.00 +0.66 -0.58	1.01 +0.67 -0.61	0.99 +0.76 -0.65

Table 9.1: The fit result using the Asimov dataset using H_T is shown in this table.

The cuts are applied on the jets instead of the events. This means, although increasing the p_T thresholds for the hard jets will cut out more jets from the backgrounds, the whole background event is not rejected. Instead, the event will contain fewer jets. The cuts do not remove the events, because the available events for this study are limited and the loss of statistics is not wanted. The effect of the variation of jet p_T thresholds is rather complex and needs to be tested empirically. Therefore, the suggested configurations are only partly motivated by the study on the jet p_T distributions but are also estimated based on experience.

9.2 Threshold Study based on H_T

The three configurations were first studied using only the variable H_T. The pre-fit summary plots of signal regions are shown in figure 9.4 and the fit result using the Asimov dataset is shown in table 9.1.

The cuts on CONFIG 1 and CONFIG 2 both tighten the threshold of the first four jets

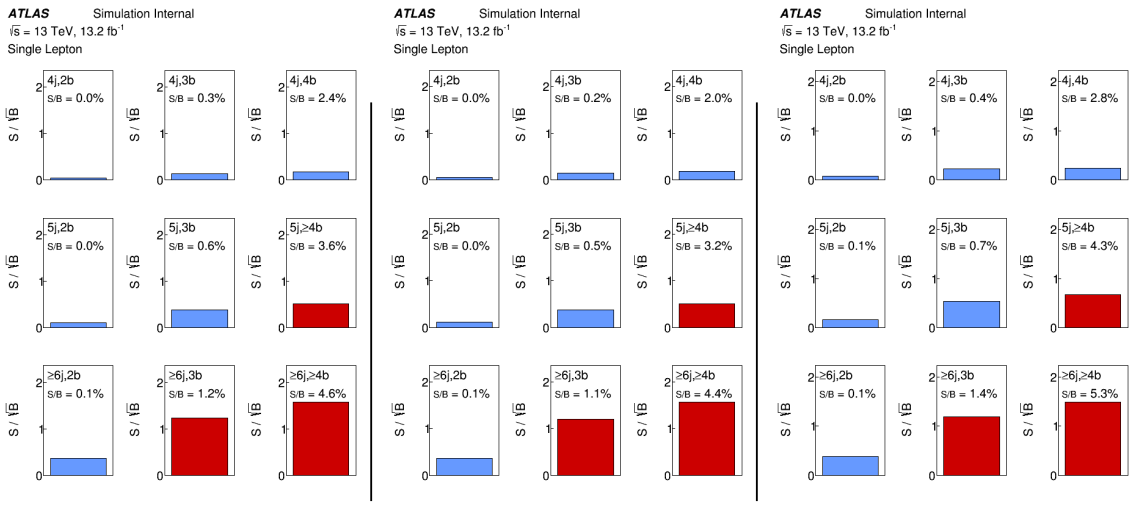

Figure 9.3: S/B and $S/\sqrt{B}$ for ICHEP (left), CONFIG 1 (middle) and CONFIG 3 (right) can be seen here.

and loosen the last two jets, they show similar behaviour in the fit using the variable H_T. Since CONFIG 2 applies three different cuts on the jets and is thus more complicated, only CONFIG 1 and 3 are studied with a multivariate analysis.

9.3 Neural Network Training

A NN is trained for CONFIG 1 and CONFIG 3 for the three signal regions. The result of the NN training for CONFIG 1 for the region $(5j, \geq 4b)$ is shown in figure 9.5 and figure 9.6. The disagreement between the behaviour of odd and even events in the cross-validation test is due to the fact that the separation of the events according to their event number into odd and even sample is not perfect. Since the event numbers are not assigned sequentially, there are more events in one sample than as in the other. This causes different performances of the odd and even samples. Since both samples will be used in the final training of the neural network and the separation power of the two samples are the same, as shown by the ROC curve in figure 9.5, the disagreement between the two samples should have no effect on the result. Figure 9.8 shows the separation power between the signal and background using NN output for CONFIG 1.

The Receiver Operation Characteristics (ROC) curve is an estimator for the separation power of the NN output between signal and background. It is a plot of the background rejection rate against the signal acceptance rate. The ROC curve of the two samples in the cross-validation and overtraining test are supposed to be the same, otherwise the NN training can not pass the test. The Frico-Gini parameter is used to evaluate the separation

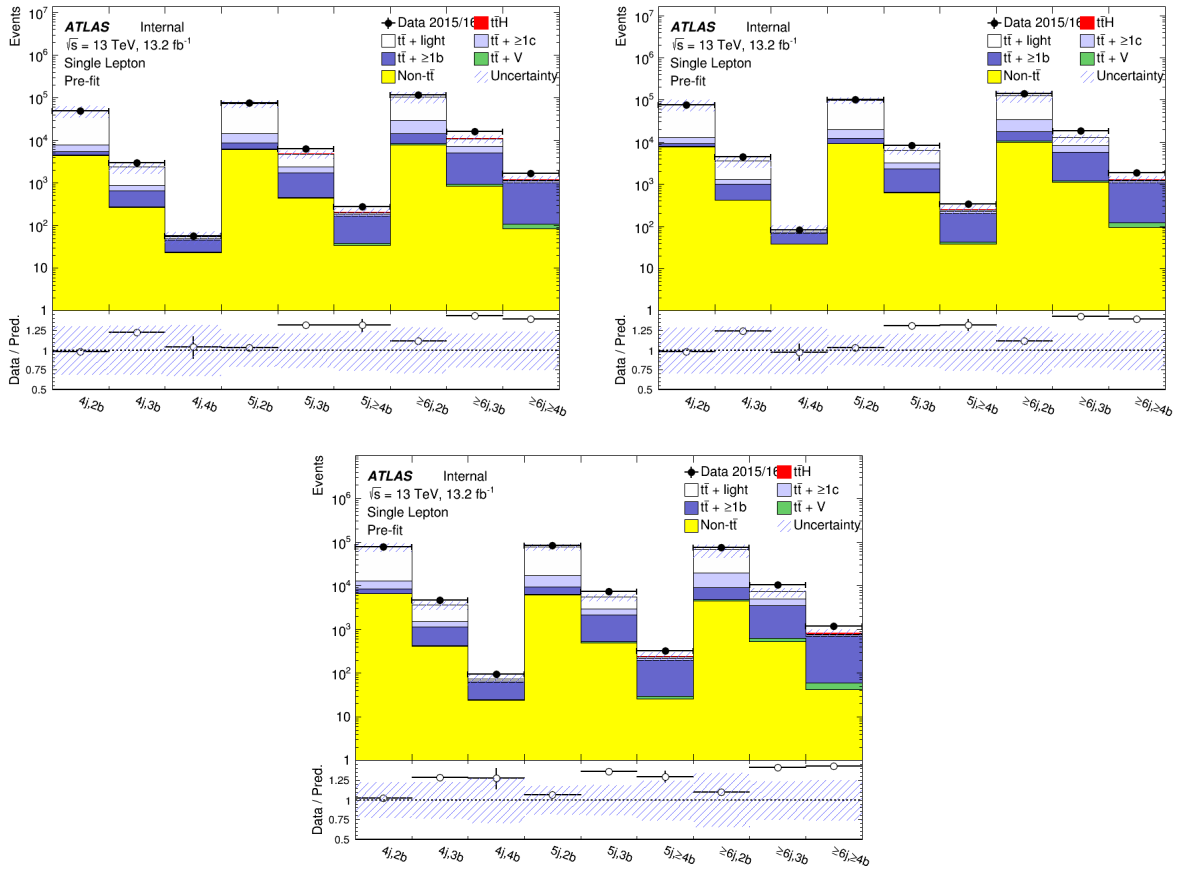

Figure 9.4: This plot shows the prefit summary plots of the three configurations using variable H_T.

power of the NN output in each region. It is defined as twice the area under the ROC curve minus 1 as demonstrated in figure 9.7.

The Frico-Gini parameter for each configuration and for each region is listed in table 9.2. In general, CONFIG 1 and CONFIG 3 show similar NN performance compared to ICHEP.

	ICHEP	CONFIG 1	CONFIG 3
$5j, \geq 4b$	49%	49%	47%
$\geq 6j, 3b$	40%	41%	39%
$\geq 6j, \geq 4b$	46%	47%	46%

Table 9.2: Frico-Gini parameter of the three configurations is listed for all signal regions.

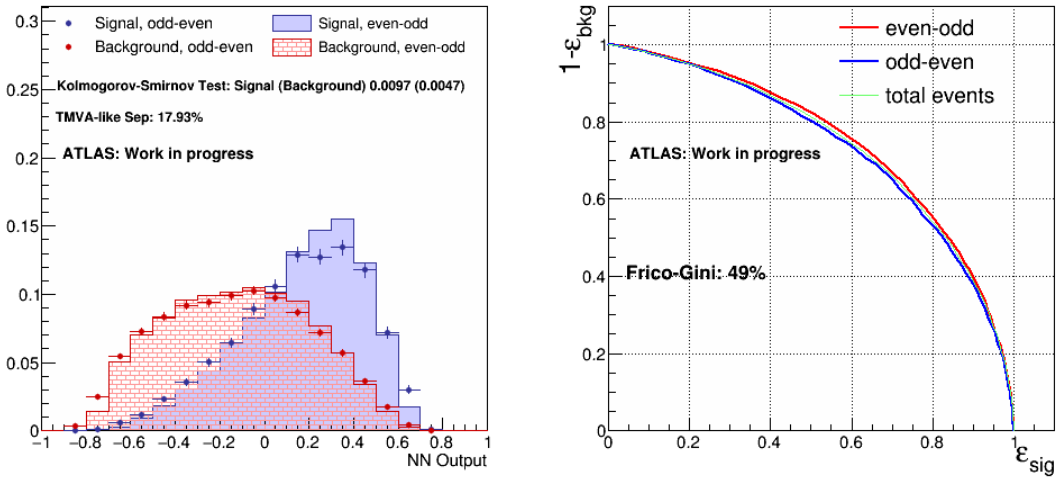

Figure 9.5: This figure shows the result for the cross-validation test for the region $(5j, \geq 4b)$ (left) and the corresponding ROC curve (right).

9.4 Fit

The two samples CONFIG 1 and CONFIG 3 are fitted to the Asimov dataset to test the sensitivity. The fit is performed with the signal plus background hypothesis combining all regions using the H_T variable for control regions and the NN output for the signal regions. To evaluate the systematic uncertainties, a method called tag rate function (TRF) is introduced to enhance MC statistics when modelling the backgrounds in the regimes with high b-jets multiplicity [6]. This method assigns a weight to all events corresponding to the probability of this event containing a certain number of b jets. Events will not be rejected according to their b-tagging count. The agreement of TRF and direct b-tagging is tested for the threshold of 25 GeV.

The yields of ICHEP and CONFIG 3 for all the signal regions of the signal process $t\bar{t}H$ and the major background $t\bar{t}b\bar{b}$ are listed in table 9.3.

9.5 CONFIG 1 and CONFIG 3 Results

CONFIG 1 has a threshold of 20 GeV for the less energetic jets, this configuration thus has a new jet object definition. An effect in CONFIG 1 is the shifting of events from exclusive $(4j)$ regions into inclusive $(\geq 6j)$ regions. A jet with energy between 20 and 25 GeV will be removed from the event in the ICHEP configuration while it will be kept as a jet in CONFIG 1. This allows CONFIG 1 to have more jets than other configurations in

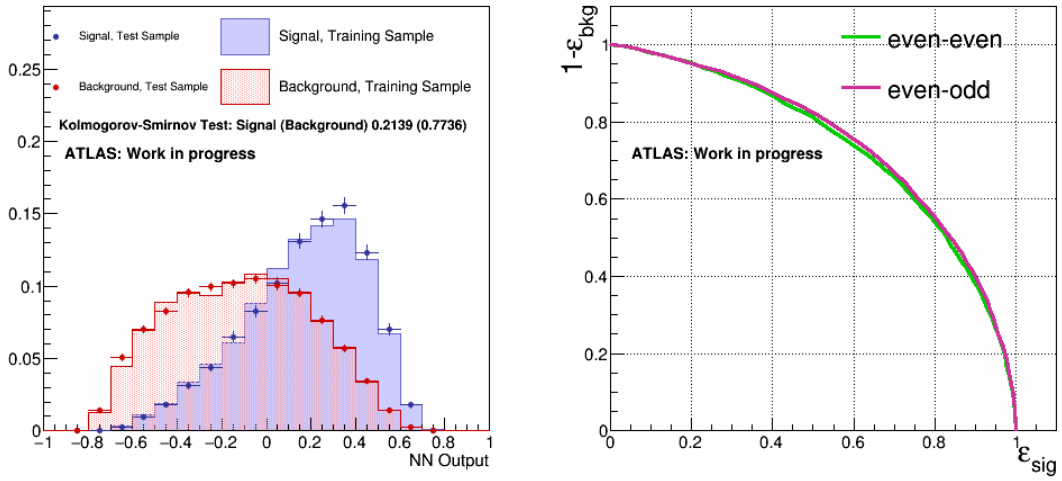

Figure 9.6: This figure shows the results for the overtraining test for the region $(5j, \geq 4b)$ and the corresponding ROC curve.

the same event. Such an event will be categorised in CONFIG 1 into regions with higher jet or b-jets multiplicity. This effect leads to lower statistics and changes the shape of the discriminant distribution in the four and five jets regions, which are control regions important for the background modelling. This effect can be seen in the yields table in table 9.3. This is the major reason for the low sensitivity of CONFIG 1 despite its high overall statistics due to lower jet p_T threshold.

To study which systematic uncertainties cause the worse constraint of CONFIG 1, one can compare the ranking plots of different configurations. The most important systematic uncertainties are shown in figure 9.9.

The most important systematic uncertainty: $tt + b$ generator residual systematics has a much bigger effect on $\mu_{t\bar{t}H}$ in CONFIG 1 (-1, +2.5) than in CONFIG 3 (-0.1, +0.7). This is because in signal regions the change of the shape due to migration in CONFIG 1 becomes large. The shifting effect in CONFIG 1 leads to lower statistics and shape changes in the control regions, which are very important for modelling systematics. The $tt + b$ generator residual uncertainty is highly effected by the shifting. Also other generator systematics, which are influenced by the shifting effect are among the important systematic uncertainties as shown in figure 9.9.

Figure 9.10 shows the nuisance parameter fitted to the Asimov dataset for CONFIG 1 compared to the ICHEP nuisance parameter. The b-tagging nuisance parameter is less constrained for CONFIG 1. This leads to the fact that this nuisance parameter has a bigger impact on $\mu_{t\bar{t}H}$ in CONFIG 1 as shown in the ranking plot of CONFIG 1 in figure 9.9.

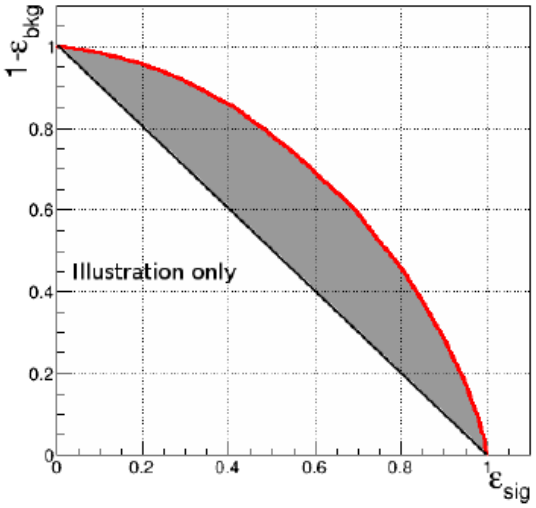

Figure 9.7: The Frico-Gini parameter is defined as the (2× area under the curve (AUC)) -1.

Figure 9.10 also shows different constraining power of the nuisance parameter in CONFIG 1 and ICHEP. Since the most important modelling systematics are less constrained in CONFIG 1, the sensitivity of this configuration is worse.

CONFIG 3 is a subset of the ICHEP configuration. Since the $t\bar{t}H$ signal has more energetic jets than the background process, this threshold configuration was suggested to cut out more background to achieve a better performance. Low statistics cause this configuration to have a bigger statistical and systematic uncertainty. By increasing the jet p_T threshold, the statistics is lower in CONFIG 3. This results not only in bigger statistical uncertainty but also in a bigger modelling systematics due to the shape effect. The comparison between the sensitivity of CONFIG 3 and ICHEP are shown in figure 9.11.

The results for the three parameters fitted to the Asimov dataset are listed in table 9.4. CONFIG 1 and CONFIG 3 do not show improved performance compared to the ICHEP configuration.

9.6 Optimization Study using Drastic Cuts

In addition to the three configurations studied above, two configurations using drastic cuts, where very high jet p_T thresholds are required, were also studied. The two configurations have jet p_T threshold as specified below:

- CONFIG 4: 60 GeV for the first to the third jet and 25 GeV for the rest,

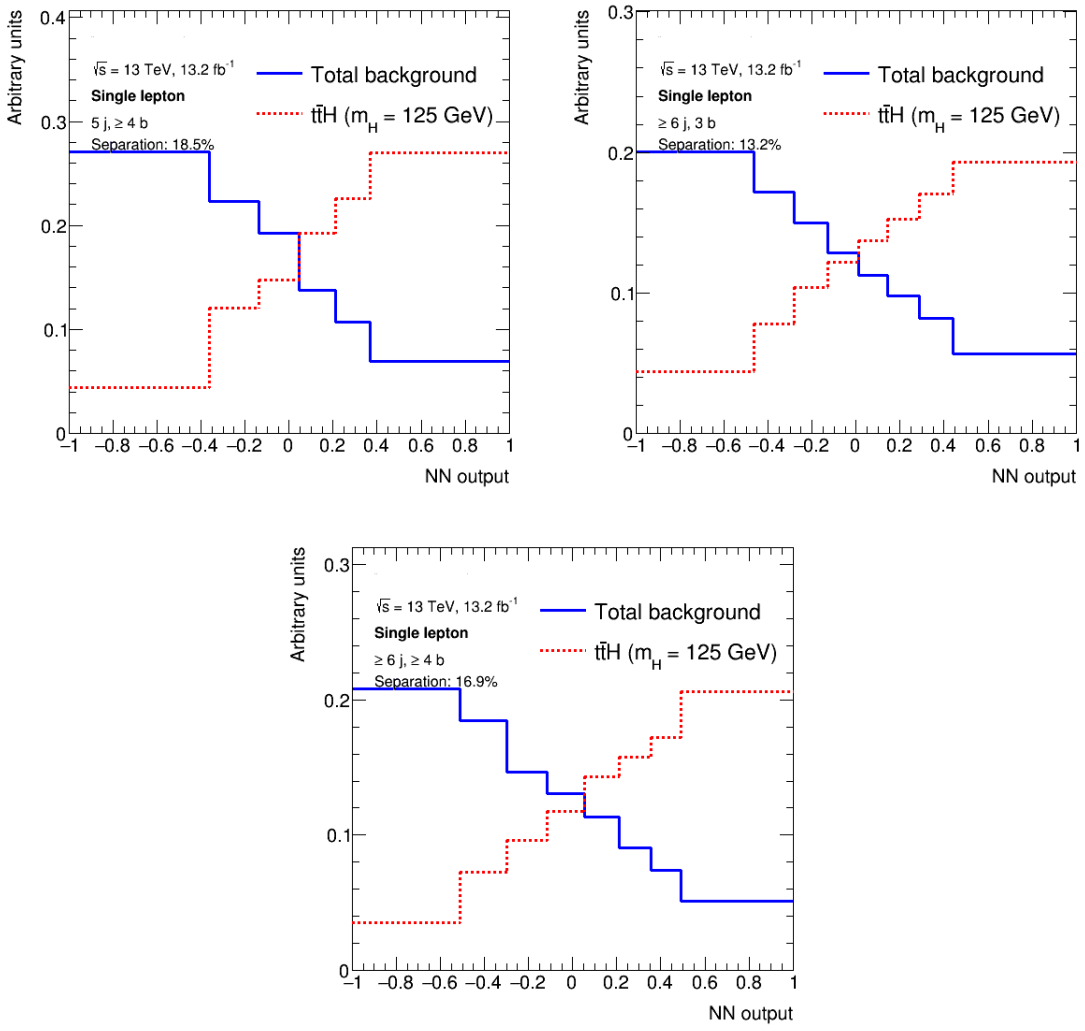

Figure 9.8: Separation plots using NN output for CONFIG 1 in the signal regions ($5j, \geq 4b$) (top left), ($\geq 6j, 3b$) (top right) and ($\geq 6j, \geq 4b$) (bottom) is shown.

- CONFIG 5: 100 GeV for the first jet, 80 GeV for the second jet and 25 GeV for the rest.

CONFIG 4 is motivated by the proposed new jet trigger threshold (L1 trigger) for Run2 at $\sqrt{s} = 13$ TeV. This high jet p_T threshold is due to the high centre of mass energy in Run2 and thus higher jet p_T for both signal and backgrounds. In order to reject more background, higher thresholds need to be applied. CONFIG 5 is proposed to test the robustness of the fit process. The result of optimization studies show a relatively stable sensitivity of different configurations, with the exception of CONFIG 1 which has a new jet definition. So in CONFIG 5, the stability of the fit result for drastic cuts is tested.

The samples for the two new configurations were not generated again but directly

	ICHEP	CONFIG 1	CONFIG 3
$t\bar{t}H(bb)$ $5j, \geq 4b$	11.66 ± 2.61	7.16 ± 1.71	10.18 ± 2.29
$t\bar{t}H(bb) \geq 6j, 3b$	107.78 ± 16.66	127.45 ± 19.14	103.85 ± 16.09
$t\bar{t}H(bb) \geq 6j, \geq 4b$	43.54 ± 9.46	53.45 ± 11.62	42.36 ±9.21
$t\bar{t}bb$ $5j, \geq 4b$	223.90 ± 62.044	128.99 ± 54.81	170.527 ± 46.85
$t\bar{t}bb \geq 6j, 3b$	3235.56 ± 802.79	4120.28 ± 1387.03	2994.83 ± 734.06
$t\bar{t}bb \geq 6j, \geq 4b$	674.23 ± 191.28	896.84 ± 341.80	639.807 ± 182.46

Table 9.3: Yields table for the signal process $t\bar{t}H(bb)$ and the dominant background process $t\bar{t}bb$ in the signal regions $(5j, \geq 4b)$, $(\geq 6j, 3b)$ and $(\geq 6j, \geq 4b)$ is shown.

	ICHEP	CONFIG 1	CONFIG 3
$\mu_{t\bar{t}H}$	1.00 +0.95 -0.93	1.00 +1.27 -1.19	1.00 +1.48 -1.34
$k(tt+ \geq 1b)$	1.00 +0.18 -0.16	1.00 +0.20 -0.17	1.00 +0.20 -0.17
$k(tt+ \geq 1c)$	1.00 +0.65 -0.51	1.00 +0.60 -0.50	1.00 +0.60 -0.48

Table 9.4: Fit result of the three configurations to the Asimov dataset is shown in this table.

subtracted from the ICHEP samples, because they are subsets of the ICHEP configuration. The neural network is not optimized for CONFIG 4 and CONFIG 5, the NN output for ICHEP is inherited. Although the sensitivity can be increased with optimized neural network, the result should be rather similar. This is because the separation power of the ICHEP neural network on CONFIG 4 and CONFIG 5 is only about 1% worse than on the ICHEP configuration. And due to the limited statistics of CONFIG 4 and CONFIG 5, it is very hard to achieve a better separation power by retraining the NN.

The S/B and $S/\sqrt{B}$ for CONFIG 4 and CONFIG 5 are shown in figure 9.12. The ICHEP configuration has the highest significance $S/\sqrt{B}$. The signal to background ratios S/B are improved slightly compared to the ICHEP configuration. To compare the sensitivity of the two configurations with the ICHEP configuration, the samples are fitted using the Asimov dataset. The fitting results are listed in table 9.5. The result of

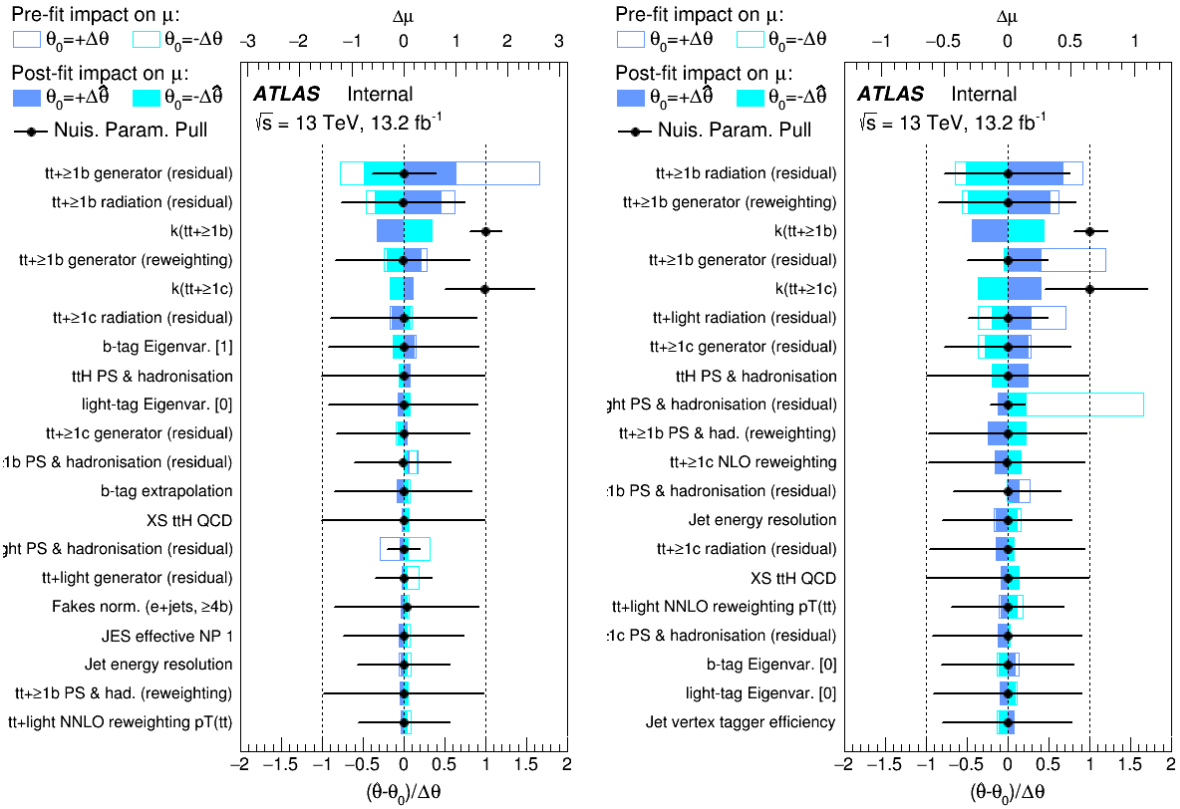

Figure 9.9: The ranking plot of CONFIG 1 (left) and CONFIG 3 (right) shows the most important systematic uncertainties.

CONFIG 4 and CONFIG 5 are similar to the ICHEP results. No improved performance is observed.

	ICHEP	CONFIG 4	CONFIG 5
$\mu_{t\bar{t}H}$	1.00 +0.95 -0.93	1.00 +1.01 -0.94	1.00 +1.00 -0.94
$k(tt+\geq 1b)$	1.00 +0.18 -0.16	1.00 +0.23 -0.19	1.00 +0.18 -0.16
$k(tt+\geq 1c)$	1.00 +0.65 -0.51	1.00 +0.63 -0.50	1.00 +0.67 -0.54

Table 9.5: Fit results of the three configurations using the Asimov dataset.

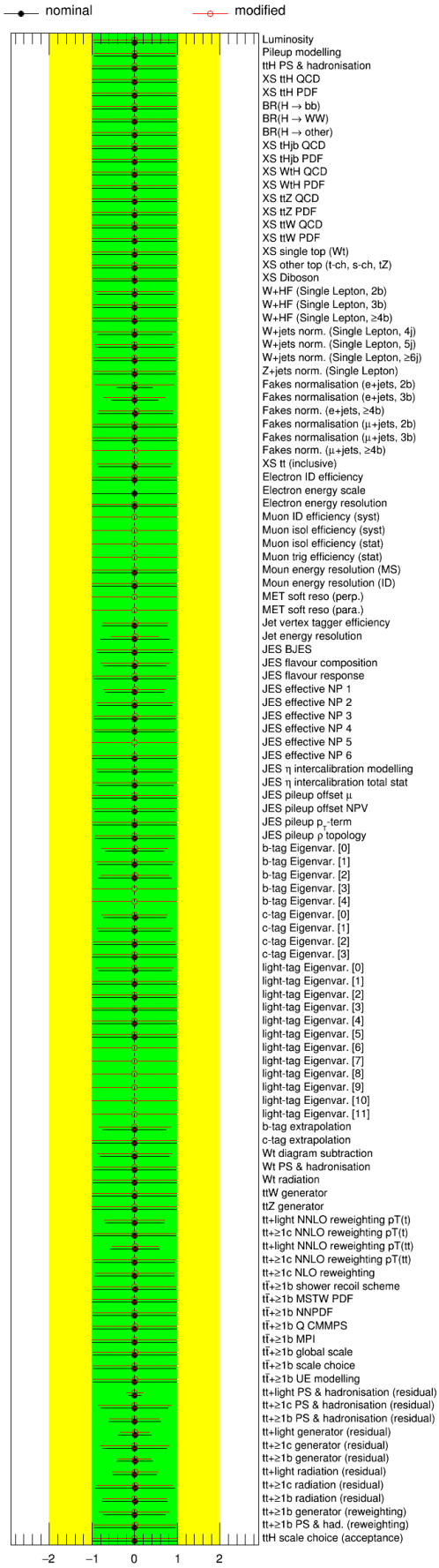

Figure 9.10: The nuisance parameters fitted using Asimov dataset of CONFIG 1 (modified) are compared with the ICHEP configuration (nominal).

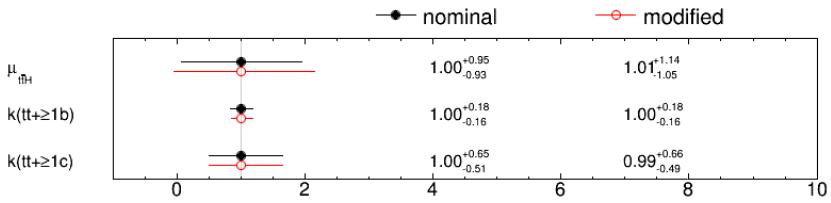

Figure 9.11: Fit results of the parameters of interest are shown. The nominal entry demonstrates ICHEP result and modified entry presents CONFIG 3.

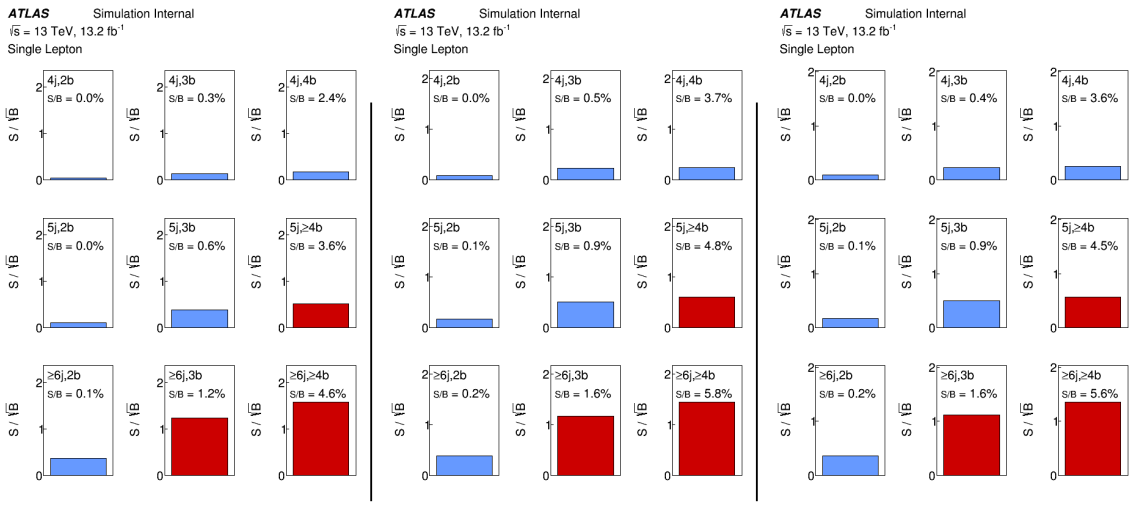

Figure 9.12: The S/B and $S/\sqrt{B}$ of ICHEP (left) CONFIG 4 (middle) and CONFIG 5 (right) are shown for comparison.

10 Monte Carlo Events Generation

In chapter 3, an overview of MC event generation for hard processes was given. There are several MC generators using different methods to simulate hard processes. Events generated with different generators show deviations from each other. However, because different MC generators describe the same physical theory, there is no physical origin of this deviation.

The current $t\bar{t}H(b\bar{b})$ process at 13 TeV for Run2 is modelled using MG5_aMC+Pythia8 (MG5_aMC+P8) in ATLAS. This process is also implemented in Powheg+Pythia8 (PP8). This study aims at providing information to support a future transition from MG5_aMC+P8 to PP8 for $t\bar{t}H$ modelling.

The first study will test the sensitivity of Powheg samples on the parameter h_{damp}. The role of free parameters in MC generators was explained in section 3.4. These free parameters are introduced in MC generators as an effect of approximation and are not physical. Thus there are no preferential values for them. Choosing values for the free parameters is done systematically with a tuning procedure [71]. The impact of the parameter choice on the event kinematic distributions needs be studied, because a modelling of the $t\bar{t}H$ process is only then reliable. The generator performance is evaluated partly based on how strongly the event properties depend on the choice of these free parameters.

The second study compares two MC generators: MG5_aMC and Powheg with respect to renormalisation and factorisation scales. More precisely, the events generated by these MCs will be compared and the origin of the deviations are studied. The deviations should be accounted for in the estimation of the modelling systematic uncertainty. A quantification of the modelling systematic uncertainty is not part of this thesis. The comparison is done by exchanging the default renormalisation and factorisation scales of the two generators, because this is one of the most important differences between them. This study will provide the answer to the question whether the deviation between the two generators is caused by the different choices for renormalisation and factorisation scales.

Basic kinematic distributions of the $t\bar{t}H$ process will be analysed because h_{damp} and renormalisation and factorisation scales have large impact on the kinematics. The kinematic distributions are computed at parton level unless it is stated otherwise. The current

$t\bar{t}H(b\bar{b})$ signal process modelled with MG5_aMC+P8 will be used as the nominal sample. Only statistical uncertainty is included for the h_{damp} and the scale study. The generator systematic uncertainty for the analysis is currently estimated using the alternative sample MG5_aMC+Herwig++. This sample will be used as a reference sample in this study.

In this chapter, the three parameters that are of interest will be described in detail. The samples generated for this study and their settings will be summarised. In addition to this, event selections and observables that are sensitive to the parameters studied will be introduced.

10.1 Resummation Scale: h_{damp}

The resummation scale h_{damp} is Powheg specific because of the unique Powheg method used. The Powheg method calculates the hardest radiation before the softer ones according to the Sudakov formula Δ_R. The hardest emission is computed according to the cross-section [72]:

$$\begin{aligned} \mathrm{d}\sigma &= \bar{B}(\Phi_N)\mathrm{d}\Phi_N\Big[\Delta_R(p_T^{min}) + \frac{R^s(\Phi_N,\Phi_{rad})}{B(\Phi_N)}\Delta_R\Big(p_T(\Phi_{rad})\Big)\mathrm{d}\Phi_{rad}\Big] && (10.1)\\ &+ R^f\mathrm{d}\Phi_N\mathrm{d}\Phi_{rad} + R_{ref}\Phi_N\mathrm{d}\Phi_{rad}, && (10.2)\end{aligned}$$

with $\bar{B}(\Phi_N) = B(\Phi_N) + V(\Phi_N) + \int \mathrm{d}\Phi_{rad} R^s(\Phi_N,\Phi_{rad})$. $B(\Phi_N)$ is the Born contribution and $V(\Phi_N)$ is the virtual contribution. Here, Φ_{rad} in equation 10.1 is the radiation phase space, Φ_N is the n parton leading-order phase space and p_T^{min} is the infra-red cut-off. The real emission contribution can be split into the divergent collinear part R_{div} and the regular part R_{reg} [72]. The divergent part can again be expressed as the sum of the term containing the singularity and the final part $R_{div} = R^s + R^f$. By introducing a damping function

$$D = \frac{h_{damp}^2}{p_T^2 + h_{damp}^2}, \tag{10.3}$$

the singularity term can be separated from the final part with:

$$R^s = DR_{div}\,, \quad R^f = (1-D)R_{div}\,. \tag{10.4}$$

The Sudakov form factor can be then limited under some transverse momentum scale without loosing the NLO accuracy. h_{damp} describes the damping of radiation with a high transverse momentum in Powheg. It separates the low and the high transverse

momentum regions and controls the hardest radiation [73]. D is close to one if the transverse momentum is much smaller than h_{damp}. In this case there is only very small correction. D is close zero if the transverse momentum is much larger than h_{damp}, in this region the cross-section diverges if there is no damping function.

The sensitivity of the distribution of kinematic variables on h_{damp} will be checked, since this parameter was tuned to a $t\bar{t}$ sample. The PP8 samples with different h_{damp} choices will then be compared to the $t\bar{t}H$ nominal and reference sample. The setting h_{damp} = 258.5 GeV = $1.5 \times m_{top}$ as used in the $t\bar{t}$ samples will be compared to h_{damp} = 352.5 GeV = $1.5 \times (m_{top} + m_{higgs}/2)$, which also includes the mass of the Higgs boson. Furthermore, a setting of h_{damp}=∞ will be studied. This corresponds to the case without damping.

10.2 Renormalisation and Factorisation Scales

Further free parameters in the MC event generation are the renormalisation and factorisation scales. In perturbative QCD, scales enter the calculation for the proper treatment of infra-red (IR) or ultraviolet (UV) divergences. The renormalisation scale μ_R is used for dimensional regularisation of UV divergence [10, 74]. μ_R is the cut-off introduced to control the divergences appearing when integrating up to very high energy level. The strong coupling constant α_s is evaluated at the renormalisation scale. The factorisation scale μ_F introduced in section 3.2 is used as IR cut-off to control the divergences when additional jets with very small energy fraction are radiated in strong processes [75]. This is the scale above which a PDF is used for the parton shower calculation.

renormalisation and factorisation scales are connected directly to the additional radiation and are important for a systematic uncertainty estimation. This motivates one reason to study the effect of the variation of the scales. In addition, Powheg and MG5_aMC developers recommend different choices for renormalisation and factorisation scales and the default values are claimed to be best choice. One would automatically expect the two generators to generate similar events when the default values are chosen. The scale study is partly motivated by the results of the h_{damp} study as introduced in section 10.1. As the results of h_{damp} will show, the deviation between samples generated with different generators is much bigger than the deviation between samples with different h_{damp} choices. Since one of the biggest differences between the two generators is the renormalisation and factorisation scales choice, a test on the samples with modified scale settings will be performed. A sample generated using Powheg with modified scales is not present, because it requires changes in the Powheg source code and could lead to other changes that are not foreseen.

ME Generator	MG5_aMC@NLO	MG5_aMC@NLO	Powheg
Hadronisation	Herwig++	Pythia8	Pythia8
Cross-section	0.447 ± 0.007	0.457 ± 0.001	0.511 ± 0.003

Table 10.1: An overview of $t\bar{t}H$ samples at NLO and their cross-section [pb] at a centre of mass energy of $\sqrt{s} = 13$ TeV is given. The top mass is set to m_{top} =172.5 GeV and the Higgs mass is set to m_{higgs} =125 GeV.

10.3 Settings and Generated Samples

The samples used for this study are generated either with MG5_aMC or Powheg at NLO, they are then interfaced to Pythia8 or Herwig++ for hadronisation. The settings used in the PP8 samples follow the standard setup recommendation for the comparison of $t\bar{t}H$ simulations[1]. The default central values for the renormalisation μ_R and factorisation scale μ_F is set to $H_T/2$ for MG5_aMC samples and $\Big(m_T(\text{Higgs}) \times m_T(\text{top}) \times m_T(\text{antitop})\Big)^{1/3}$ for Powheg samples. A centre of mass energy of $\sqrt{s} = 13$ TeV is used, the Higgs mass is set to m_{higgs} =125 GeV and the top mass is set to $m_{top} = 172.5\,\text{GeV}$. The cross-sections at NLO in QCD at $\sqrt{s}$ =13 TeV of different samples are listed in table 10.1.

10.3.1 Generated Samples for the h_{damp} Study

Five $t\bar{t}H$ samples are used for the h_{damp} study. Two sets of samples are generated with MG5_aMC with NLO matrix elements (ME). One of them is interfaced to Pythia8, the other one with Herwig++. In the MG5_aMC samples, the decay of the top quarks is simulated using MadSpin and the decay of the Higgs boson is calculated with the parton shower program. μ_F and μ_R are set to $H_T/2$ in MG5_aMC. The MG5_aMC+Herwig++ sample uses the PDFset CT10 [58] with UE-EE-5 tune [2]. The MG5_aMC+P8 sample uses the PDFset NNPDF3.0 [56] with the A14 tune [3] [55]. MG5_aMC+Herwig++ sample serves as a reference in the h_{damp} study since the current modelling systematic uncertainty is estimated using this sample.

[1]The recommendation can be found under `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/ProposaltTH.`

[2]The input EVNT file used for this sample is {mc15_13TeV.341270.aMcAtNloHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_ttH125_semilep.merge.DAOD_TRUTH1.e4277_p2514/}

[3]The input EVNT file used for this sample is {mc15_13TeV.343366.aMcAtNloPythia8EvtGen_A14_NNPDF23_NNPDF30ME_ttH125_semilep.merge.DAOD_TRUTH1.e4706_p2419/}

	Generator	Shower	PDF	h_{damp}	scale
MG5_aMC+P8	MG5_aMC	Pythia8	NNPDF3.0	-	$H_T/2$
PP8_ hdamp352.5	Powheg	Pythia8	NNPDF3.0	$1.5 \times (m_{top} + m_{higgs}/2)$	$(m_T(\text{Higgs}) \times m_T(\text{top}) \times m_T(\text{antitop}))^{1/3}$
PP8_ hdamp258.5	Powheg	Pythia8	NNPDF3.0	$1.5 \times m_{top}$	$\left(m_T(\text{Higgs}) \times m_T(\text{top}) \times m_T(\text{antitop})\right)^{1/3}$
PP8_ hdampinfinity	Powheg	Pythia8	NNPDF3.0	∞	$\left(m_T(\text{Higgs}) \times m_T(\text{top}) \times m_T(\text{antitop})\right)^{1/3}$
MG5_aMC+Herwig++	MG5_aMC	Herwig++	CT10	-	$H_T/2$

Table 10.2: The settings of the samples used for the h_{damp} study.

Three sets of sample are generated with Powheg. They are then interfaced to Pythia8 for showering and hadronisation. The central value for the scales μ_F and μ_R in Powheg is $\left(m_T(\text{Higgs}) \times m_T(\text{top}) \times m_T(\text{antitop})\right)^{1/3}$. All three samples use the PDFset NNPDF3.0 with A14 tune. The Powheg specific parameter h_{damp} is set to: h_{damp}=258.5 GeV = $1.5 \times m_{top}$ as used in the $t\bar{t}$ samples, h_{damp}=352.5 GeV = $1.5 \times (m_{top} + m_{higgs}/2)$ and h_{damp}=∞ in order to investigate the sensitivity of the $t\bar{t}H$ samples to h_{damp}. The settings are summarised in table 10.2.

10.3.2 Generated Samples for the Scale Study

Four $t\bar{t}H$ samples are used for the scale study. Three sets of samples are generated using MG5_aMC. The MG5_aMC+Herwig++ is the same as used in the h_{damp} study and is also used as a reference sample. Two samples are generated with MG5_aMC and interfaced with Pythia8. One of the two MG5_aMC+P8 samples is the same one already used in the h_{damp} study, where the default MG5_aMC scale $H_T/2$ is used. The other MG5_aMC+P8 sample has modified renormalisation and factorisation scales. The modified renormalisation and factorisation scales are set to $\left(m_T(\text{Higgs}) \times m_T(\text{top}) \times m_T(\text{antitop})\right)^{1/3}$, equal to the renormalisation and factorisation scales in the Powheg samples. The PDFset NNPDF3.0 is used for this sample. One sample is generated with Powheg, this is the PP8 sample with $h_{damp} = 1.5 \times (m_{top} + m_{higgs}/2)$ from the h_{damp} study. A summary of the settings of the samples used for the scale study can be found in table 10.3.

10.4 Event Selection

The $t\bar{t}H(b\bar{b})$ process has in the final state a $b\bar{b}$ pair, a $t\bar{t}$ pair and the decay products of the two W bosons. The W bosons can decay either leptonically into a lepton and a neutrino or hadronically into two quarks. The dilepton channel is difficult to analyse due to the

	Generator	Shower	PDF	h_{damp}	scale
MG5_aMC+P8	MG5_aMC	Pythia8	NNPDF3.0	-	$H_T/2$
MG5_aMC+P8 newscale	MG5_aMC	Pythia8	NNPDF3.0	-	$\left(m_T(\text{Higgs}) \times m_T(\text{top}) \times m_T(\text{antitop})\right)^{1/3}$
PP8_ hdamp352.5	Powheg	Pythia8	NNPDF3.0	$1.5 \times \left(m_{top} + (m_{higgs})/2\right)$	$\left(m_T(\text{Higgs}) \times m_T(\text{top}) \times m_T(\text{antitop})\right)^{1/3}$
MG5_aMC+Herwig++	MG5_aMC	Herwig++	CT10	-	$H_T/2$

Table 10.3: An overview of the generator and shower program used to generate the samples in the scale study. The PDFset, h_{damp} value and renormalisation and factorisation scales are also listed.

presence of two neutrinos, which appear as missing energy in the detector. The study on the full hadronical channel is also challenging because of the difficulties in the assignment of the four jets to their mother particles during the reconstruction. The study will be performed on the semileptonic channel, because on the reconstruction level it is clear that the two jets are from the same W boson and the charged lepton and the neutrino are from the other W boson. Cuts on the top and Higgs boson decay mode are applied on the truth level to guarantee semileptonic decays and $H \to b\bar{b}$. Only electrons and muons are defined as leptons. Events with tau leptons are not analysed in this study.

In all the samples, the $t\bar{t}H$ events are selected using the same cuts. The cuts used in this study are motivated by the cuts used by the analysis group for the ICHEP result as mentioned in section 8.3. However, for the ICHEP results, the cuts are applied on the reconstruction level while in this study only parton level events are present. The transverse momentum p_T of the first three jets are required to be at least 25 GeV. The absolute value of the pseudorapidity of the first three jets should fulfil $|\eta| < 2.5$.

10.5 Observable Definitions

To investigate the effect of different parameter choices, one needs to perform tests on observables which are sensitive to these parameters. For this MC generator study, the $t\bar{t}H(b\bar{b})$ semileptonic channel is analysed. Six jets, out of which four are $b/\bar{b}$ quarks, one lepton and one neutrino are expected in the final state. One $b\bar{b}$ pair is the decay product of the Higgs boson. The $t/\bar{t}$ quark decays via W bosons into a $b/\bar{b}$ quark, one W boson decays leptonically and the other W boson decays hadronically into two (light) quarks.

The renormalisation scale, factorisation scale and the h_{damp} are parameters affecting the kinematics of the hadrons in the system. Observables related to hadron momentum or momentum of a system consisting of several hadrons are sensitive to these parameters. The observables studied are as following.

- Transverse momentum of the $t\bar{t}H$ system: $t\bar{t}H$ system p_T, the vector sum of the transverse momentum of the top, the anti-top and the Higgs boson. For leading order events this sum should be zero.
- Transverse momentum of the $t\bar{t}$ system: $t\bar{t}$ system p_T, the vector sum of the transverse momentum of the top and the anti-top in the system.
- Scalar sum of $\sqrt{p_T^2 + m^2}$ of all final state particles: H_T.
- Jet multiplicity in the final state: Jet multiplicity, this observable is evaluated on the particle level. For the $t\bar{t}H(b\bar{b})$ semileptonic channel, a maximum at 6 jets is expected.
- Transverse momentum of the leading jet: Leading jet p_T, the transverse momentum of the most energetic jet.
- Transverse momentum of the second leading jet: Second leading jet p_T, the transverse momentum of the second most energetic jet.
- The transverse momentum of the top quark: Top p_T, the top quark is defined as the last top quark after radiation and before decay.
- The transverse momentum of the Higgs boson: Higgs p_T, the transverse momentum of the Higgs boson.

11 Monte Carlo Generator Study Result

In the previous chapter, three free parameters in the MC generator: h_{damp}, renormalisation and factorisation scales together with their roles in the strong process simulation were introduced. This chapter presents the effect of the parameters on various kinematic observables. The results, more precisely the cross-section as a function of these observables, will be compared. This chapter is arranged as follows: in section 11.1 results of the h_{damp} study will be presented, then in section 11.2 the results of the scale study will be discussed.

11.1 h_{damp} Study Result

Samples generated for the h_{damp} study were summarised in the last chapter in table 10.2. The first part of the h_{damp} study concentrates on the comparison of the two PP8 samples with h_{damp}=352.5 GeV and h_{damp}=258.5 GeV. These two samples will be compared with the nominal sample: MG5_aMC+P8, which is the baseline generator in ATLAS for $t\bar{t}H$ simulation, and MG5_aMC+Herwig++, which is the reference sample for the systematic uncertainty estimation in the analysis [6]. The $t\bar{t}$ system p_T and $t\bar{t}H$ system p_T are found to be most sensitive to the choice of h_{damp}. This is because h_{damp} regulates the hardness of the hardest radiation. This emission recoils against the $t\bar{t}H$ and the $t\bar{t}$ system and has an effect on the $t\bar{t}H$ and $t\bar{t}$ system momenta in the transverse direction [73]. The sensitivity of event distribution on the h_{damp} choice can be seen in figure 11.1 and figure 11.2. All cross-sections are normalized to one in order to compare them. The ratio plot shows the difference computed against the nominal sample MG5_aMC+P8.

For all observables, the deviation is large in the high energetic region due to the low statistics. There is thus a large statistical uncertainty in this region. Therefore, the conclusions of all studies are drawn based on the behaviour of the samples in the low and middle energetic regions where the statistical uncertainty is small.

Only small deviations can be seen between the two PP8 samples with different h_{damp}

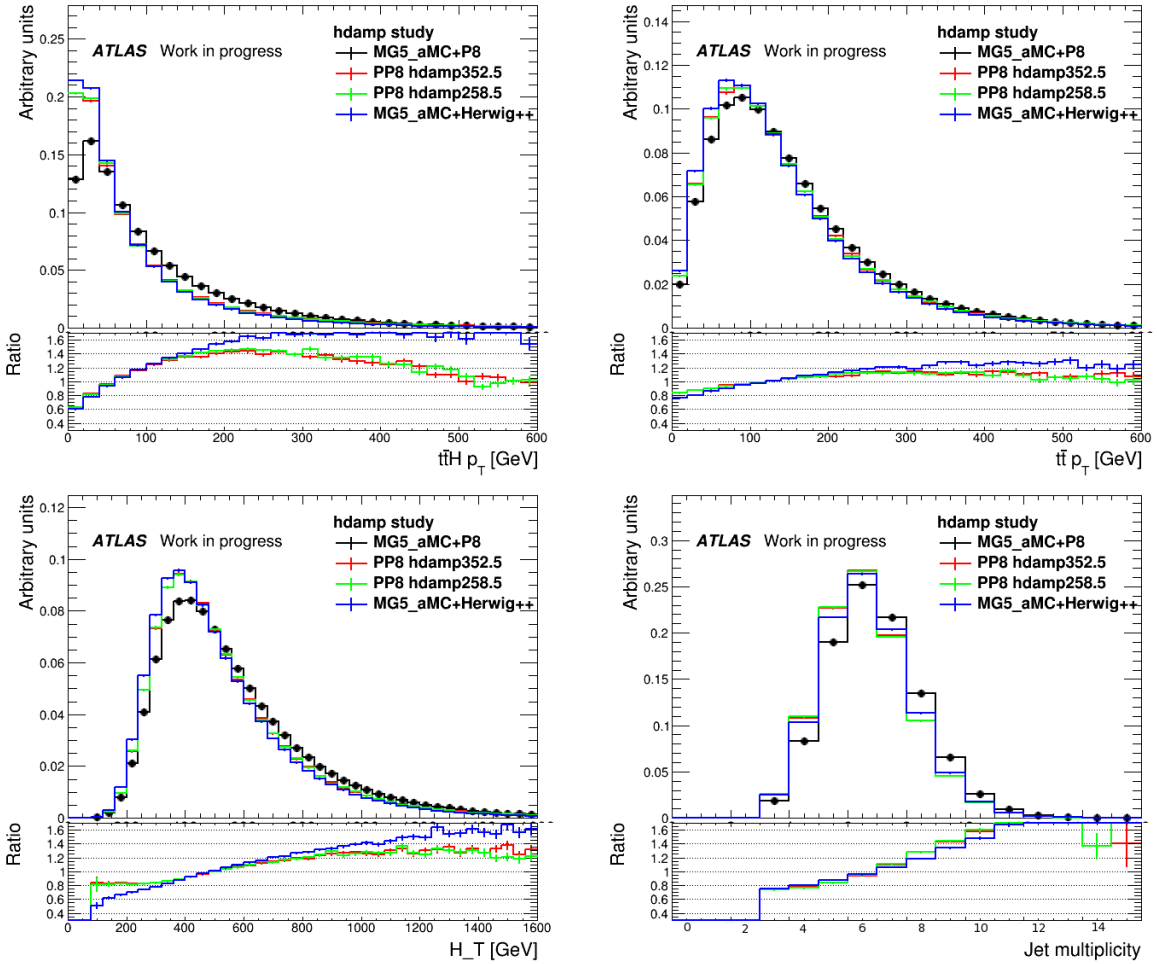

Figure 11.1: The normalized differential cross-section as a function of $t\bar{t}H$ system (top left) and $t\bar{t}$ system p_T (top right), as a function of H_T (bottom left) and jet multiplicity (bottom right) for the $t\bar{t}H(b\bar{b})$ process generated with various generators or settings at NLO.

parameter choices. Even for the most sensitive observables $t\bar{t}$ and $t\bar{t}H$ system p_T, the deviation between the two PP8 samples with different h_{damp} choices is less than 5%. The deviation between different ME generators and hadronisation models as shown in figure 11.1 on the other hand is huge: it is 40% in the $t\bar{t}H$ system p_T and 20% in the $t\bar{t}$ system p_T.

Varying the h_{damp} has very small effect on the observable H_T, jet multiplicity and jet kinematic distributions. For H_T, first and second leading jet p_T distributions, the deviation between different ME generators and hadronisation models is large. An effect of nearly 30% can be seen in figure 11.1 and figure 11.2 in the comparison between MG5_aMC+P8 and MG5_aMC+Herwig++ in the region between 50 GeV and

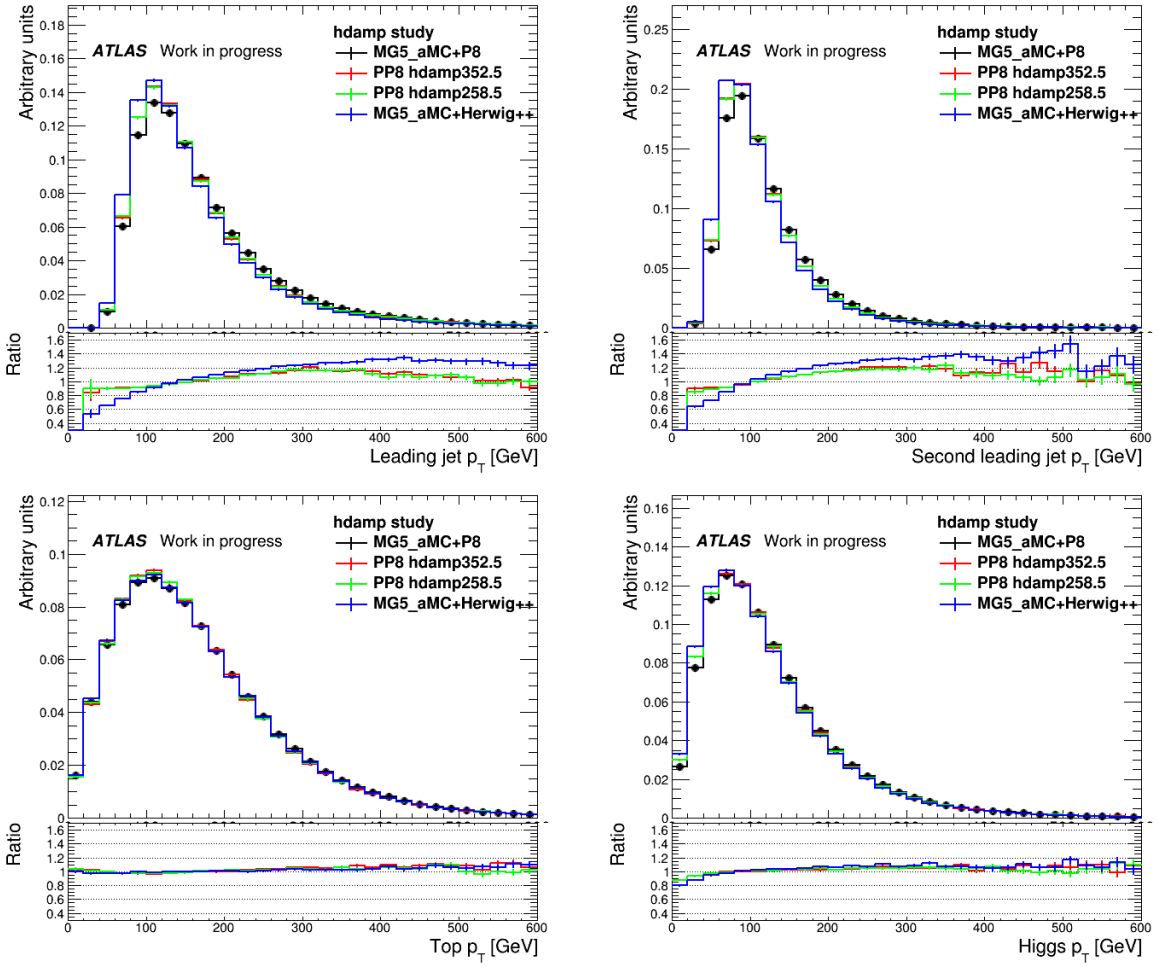

Figure 11.2: The normalized differential cross-section as a function of the jet kinematics of the leading jet (top left) and second leading jet (top right), as well as of the top quark p_T (bottom left) and the Higgs boson p_T (bottom right) for $t\bar{t}H(b\bar{b})$ process generated with various generators or settings at NLO.

400 GeV for H_T and jet kinematic distributions. For the observable jet multiplicity in figure 11.1 (bottom right), the differential cross-section has a maximum at 6 jets. This corresponds to the final state expectation for the semileptonic channel. For the jet multiplicity, the MG5_aMC+Herwig++ sample is closer to the PP8 samples than the MG5_aMC+P8 sample, although in comparison between the PP8 samples and the MG5_aMC+Herwig++ sample, ME generator and hadronisation programs are both changed and in the comparison with the MG5_aMC+P8 sample only the hadronisation program is changed. The observables Top p_T and Higgs p_T are relatively insensitive to variations of h_{damp}. For these two variables, as one can see in figure 11.2 (bottom left and bottom right), the deviation between samples generated with different ME generators

and hadronisation models is less than 10%.

From the figures, one can conclude that negligible deviations can be seen when varying the h_{damp} parameter. The deviation between samples generated with different generators and hadronisation models is much bigger.

Another interesting setting is $h_{damp} = \infty$ corresponding to the case without damping. In the study on $t\bar{t}$ samples [73] generated with PP8, a deviation between events with h_{damp}=∞ and h_{damp}=$2 \times m_{top}$ was observed. A similar study with $h_{damp} = \infty$ is performed for $t\bar{t}H$ samples. The three samples used here are generated using PP8, with h_{damp}=258.5 GeV, h_{damp}=352.5 GeV and h_{damp}=∞ as listed in table 10.2. All cross-sections are normalized to one for the comparison. The ratio plot shows the difference computed against the sample with h_{damp}=352.5 GeV.

One can see in figure 11.3 and figure 11.4, the cross-section corresponding to the parameter choice h_{damp}=352.5 GeV lies between the cross-section of h_{damp}=258.5 GeV and h_{damp}=∞ for all observables. This is consistent with the gradually changing of h_{damp}. For the most sensitive observables $t\bar{t}H$ and $t\bar{t}$ system p_T, $h_{damp} = \infty$ sample and h_{damp} =258.5 GeV deviate around 2% from the nominal sample h_{damp} =352.5 GeV in two opposite directions in the low energetic region. In the case of the observable H_T, the deviation between $h_{damp} = \infty$ and the other two samples reaches 5% around the maximum. For the jet kinematic distributions, a deviation of the $h_{damp} = \infty$ sample from the other two samples of about 2% around the maximum region is also observed. In the case of jet multiplicity, the h_{damp} =258.5 GeV sample and the h_{damp} =352.5 GeV sample are very close to each other, while the $h_{damp} = \infty$ sample deviates about 1% from the other two samples in the central region. For the two insensitive observables Top p_T and Higgs p_T, only very small deviations can be observed.

One can conclude based on figure 11.1 - figure 11.4 that the effect of variation of h_{damp} in Powheg is very small and the kinematics of the events are not very sensitive to h_{damp} choice. It is also observed that there are large deviations between samples produced with different ME generators, different hadronisation models and showering programs. The MG5_aMC+Herwig++ sample is the one that deviates from all other samples the most for nearly all observables.

The result from the h_{damp} study motivates the second study, where MG5_aMC and Powheg are compared with respect to the renormalisation and factorisation scales.

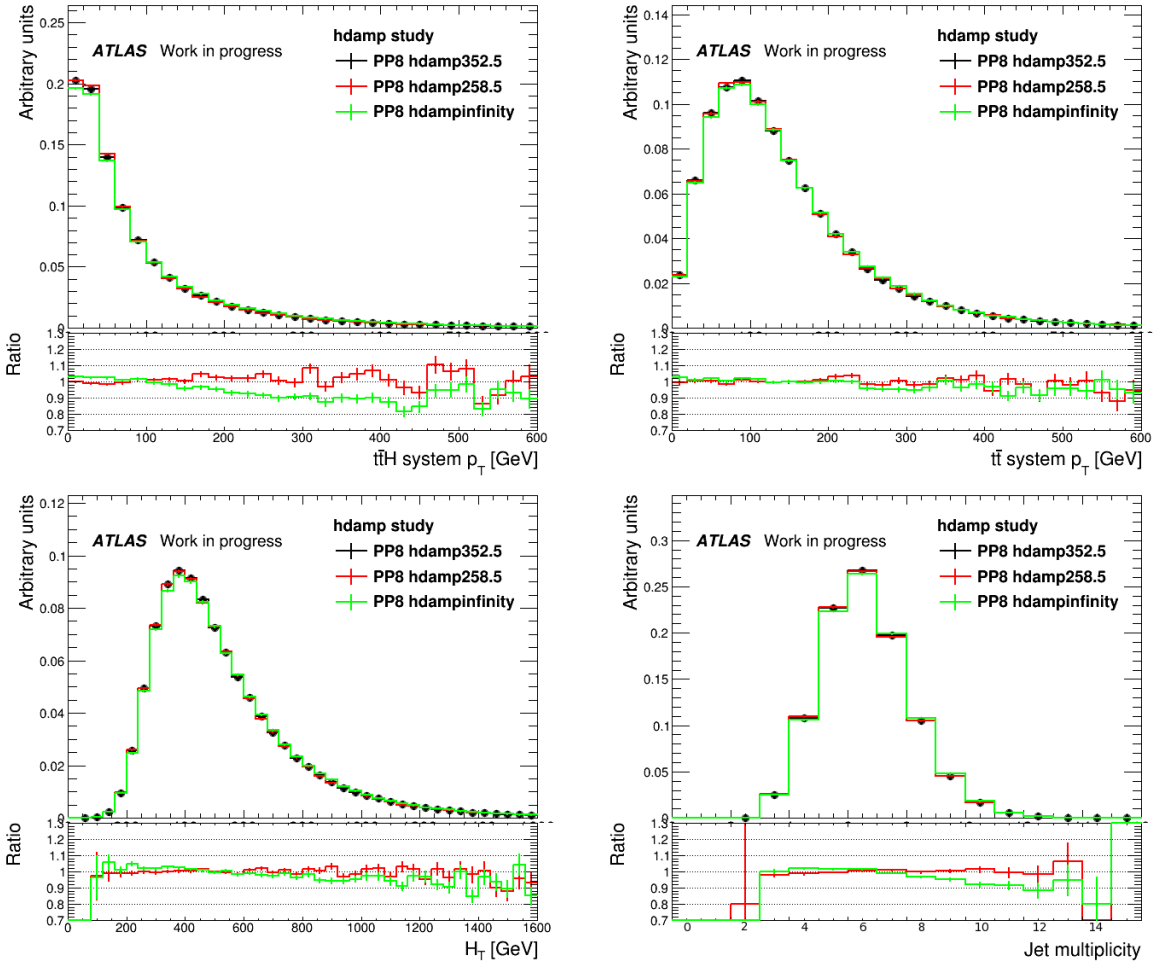

Figure 11.3: The normalized differential cross section is shown as a function of the $t\bar{t}H$ system (top left) and the $t\bar{t}$ system transverse momentum (top right), as a function of H_T (bottom left) and jet multiplicity (bottom right) for the $t\bar{t}H(b\bar{b})$ process generated using PP8 with different h_{damp} choices at NLO.

11.2 Scale Study Result

Large deviations between MG5_aMC+P8, PP8 and MG5_aMC+Herwig++ samples were observed in the h_{damp} study. Other $t\bar{t}H$ modelling studies have also discussed this problem. This can be an overlapping effect of different ME generators, different PDFsets, different renormalisation and factorisation scales used by the generator, as well as the choice of parton shower and different hadronisation models.

An earlier study [76] discussed the deviation between different hadronisation models. Samples generated with MG5_aMC+P8 and MG5_aMC+Herwig++ were compared. In reference [76], a deviation of approximately 40% was observed for the observable $t\bar{t}H$

system p_T and a deviation of approximately 20% for the observable $t\bar{t}$ system p_T between MG5_aMC+P8 and MG5_aMC+Herwig++. Another study [77] concentrated on the comparison of different ME generators. In reference [77], a deviation of about 25% was observed between MG5_aMC+P8 and PP8 for $t\bar{t}H$ system p_T. It is worth mentioning that the PDFsets used in [76, 77] are not kept identical. So the different behaviours can also be partly due to the different PDFset choices.

This study concentrates on the effect of the renormalisation and factorisation scales on the events. Since Powheg and MG5_aMC have different default value for the renormalisation and factorisation scales, it is worthwhile to compare samples with interchanged renormalisation and factorisation scales. However, it is very critical to change the renormalisation and factorisation scales in Powheg because a change in the Powheg source code is needed. To avoid a change to the source code, the renormalisation and factorisation scales in Powheg are not modified in this study. This modification is easier in MG5_aMC and the procedure for doing this is known. One sample is generated with MG5_aMC and the renormalisation and factorisation scales are changed to the central value of renormalisation and factorisation scales in Powheg, namely $\mu_F = \mu_R = \Big(m_T(\text{Higgs}) \times m_T(\text{top}) \times m_T(\text{antitop})\Big)^{1/3}$. This sample is interfaced with Pythia8 and will be referred to as MG5_aMC+P8 with modified scales in the following. The sample MG5_aMC+P8 with modified scales is compared with the MG5_aMC+P8 sample with default renormalisation and factorisation scales $\mu_F = \mu_R = H_T/2$ and the PP8 sample with renormalisation and factorisation scales $\mu_F = \mu_R = \Big(m_T(\text{Higgs}) \times m_T(\text{top}) \times m_T(\text{antitop})\Big)^{1/3}$. The comparison between the three samples along with the reference sample MG5_aMC+Herwig++ is shown in figure 11.5 and figure 11.6. All cross-sections are normalized to one. The ratio plot shows the difference computed against the nominal sample MG5_aMC+P8 with the default renormalisation and factorisation scales.

One can see in figure 11.5, the $t\bar{t}H$ system p_T and $t\bar{t}$ system p_T are the most sensitive variables to the scale variation. For $t\bar{t}H$ system p_T, $t\bar{t}$ system p_T, H_T and jet multiplicity, the curve of the differential cross-section of the MG5_aMC+P8 sample with modified scales is between the PP8 sample and the MG5_aMC+P8 sample with default scale. For leading jet and second leading jet p_T (in figure 11.6), the MG5_aMC+P8 sample with modified scales overlaps with the PP8 sample and they deviate from the MG5_aMC+P8 sample with default scale. Top p_T and Higgs p_T (figure 11.6) are the two variables that are most insensitive to the radiation parameter. For these two variables, all samples are close to each other and no remarkable deviation is observed. For the variable jet multiplicity and jet kinematics in figure 11.6, the MG5_aMC+P8 sample with modified scales is closer to the MG5_aMC+Herwig++ sample.

From this study, one can conclude that the events generated with MG5_aMC+P8 become closer to the events generated using PP8 with modified scales $\Big(m_T(\text{Higgs}) \times m_T(\text{top}) \times m_T(\text{antitop})\Big)^{1/3}$. This means that the scale is one important origin of the deviation between the Powheg and the MG5_aMC samples but not the only one. The remaining differences can be due to different numerical implementations of the generators, PDFsets and other reasons that are not discussed here. Except for Top p_T and Jet multiplicity, the MG5_aMC+Herwig++ sample shows the biggest deviation from all other samples. The size of the difference between the MG5_aMC+Herwig++ sample and the nominal sample MG5_aMC+P8 is consistent with the systematic uncertainty estimation used for the analysis of the $t\bar{t}H$ signal process [6].

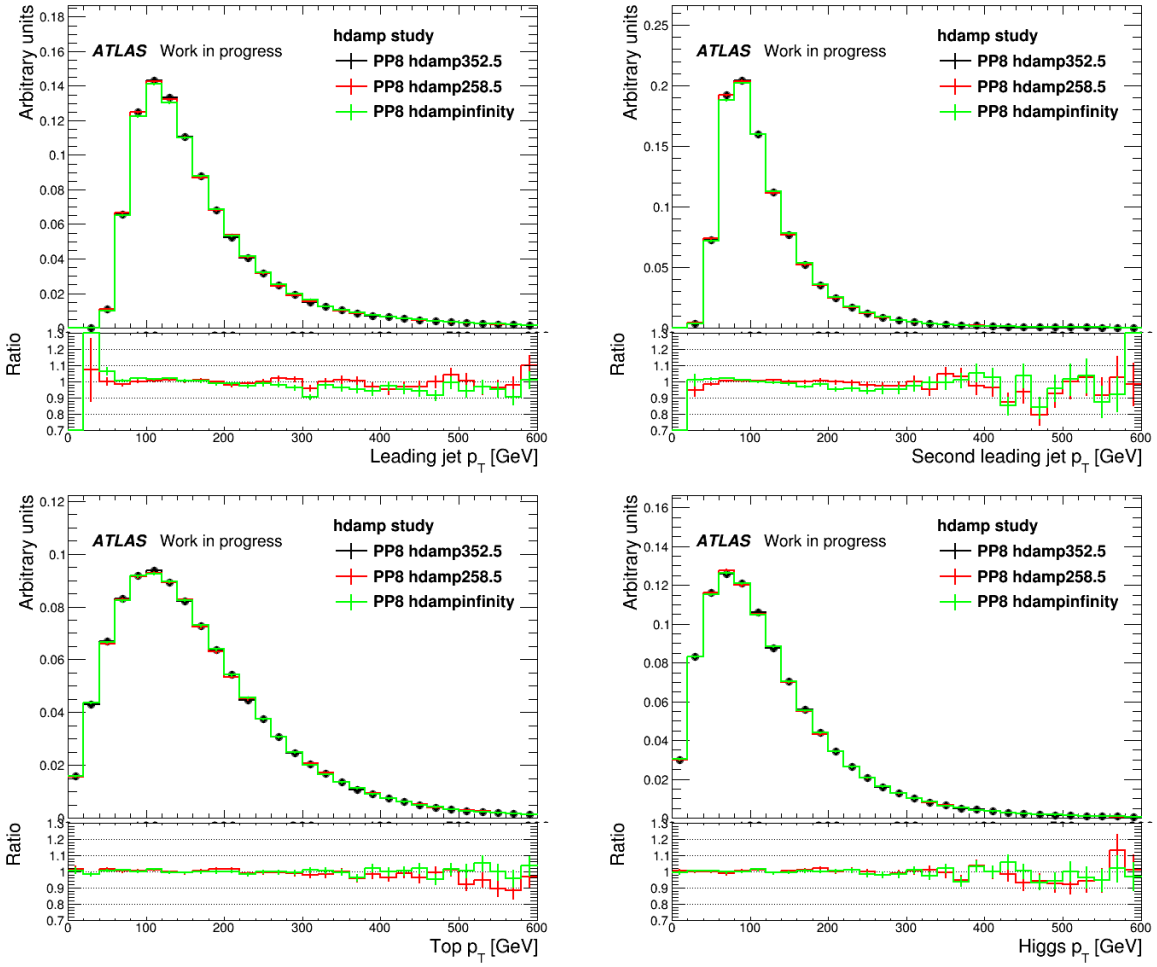

Figure 11.4: The normalized differential cross section is shown as a function of jet kinematics of the leading jet (top left) and second leading jet (top right), as well as of the transverse momentum of the top quark (bottom left) and the Higgs boson (bottom right) for the $t\bar{t}H(b\bar{b})$ process generated using PP8 with different h_{damp} choices at NLO.

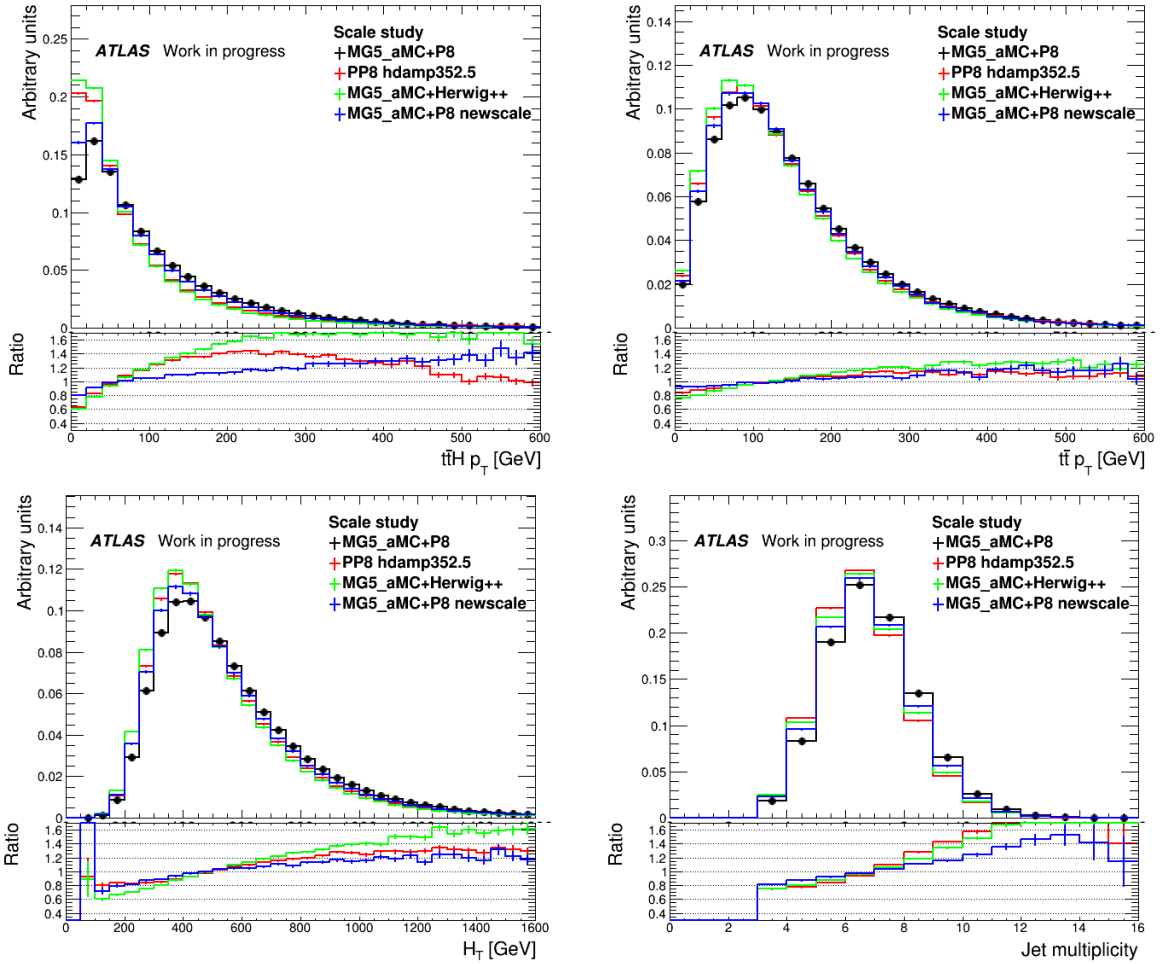

Figure 11.5: The normalized differential cross section as a function of the $t\bar{t}$H system (top left) and the $t\bar{t}$ system transverse momentum (top right), of H_T (bottom left) and jet multiplicity (bottom right) for the $t\bar{t}H(b\bar{b})$ process generated using MG5_aMC and Powheg with different scale choices at NLO.

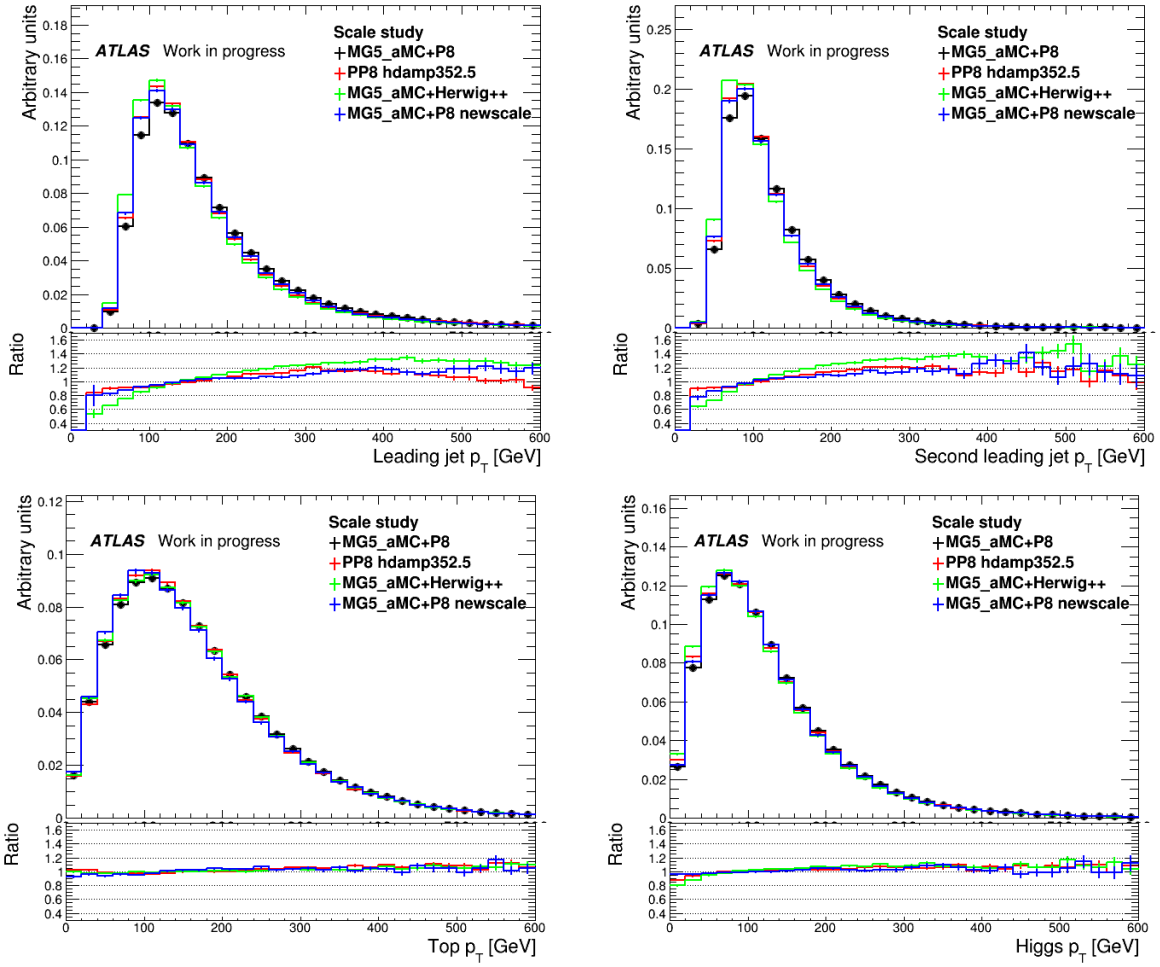

Figure 11.6: The normalized differential cross section as a function of the jet kinematics such as the leading jet (top left) and second leading jet (top right), and of the transverse momentum of the top quark (bottom left) and the Higgs boson (bottom right) for $t\bar{t}H(b\bar{b})$ process generated using MG5_aMC and Powheg with different scale choices at NLO.

12 Conclusion and Outlook

In this thesis, the theoretical background of the Standard Model, QCD and Higgs mechanism was introduced. Then an overview of MC event generation, parton shower concept and different hadronisation models was given. The LHC and ATLAS experimental setups were presented.

The results of this thesis consist of an optimization study for the search of the Standard Model Higgs boson produced in association with a top quark pair and a MC generator study for the $t\bar{t}H$ process modelling.

The first part of this thesis concentrated on the optimization study. In this study, five different sets of configurations for the jet p_T threshold were studied based on the result of the ICHEP analysis in 2016 [6] to achieve a better sensitivity and a better constraint on the systematic uncertainty. The p_T thresholds for the hard jets were increased to reject more background and the p_T thresholds of the light jets were decreased to include more jets from W decay, which is useful for the final state reconstruction.

Three configurations as listed in section 9.1 were studied first. The samples were fitted using the variable H_T. Since CONFIG 1 and CONFIG 2 show very similar behaviour, the study was continued with CONFIG 1 and CONFIG 3. A neural network was trained separately for CONFIG 1 and CONFIG 3 in each signal region and a neural network output with a separation power similar to the ICHEP configuration was reached as shown in section 9.3. The fit based on the neural network output using CONFIG 1 and CONFIG 3 does not improve the performance compared to ICHEP. CONFIG 3 shows no improved sensitivity because of the lower statistics compared to ICHEP which leads to bigger systematic uncertainties and statistical uncertainties. CONFIG 1 includes more events but still shows no improved sensitivity due to the shifting of events from exclusive regions to inclusive regions caused by the new jet object definition. CONFIG 1 has much lower statistics in the control regions and thus less constraints on the systematic uncertainties. Thus one can conclude, although increasing the p_T thresholds of the hard jets will reject more background, this also leads to lower statistics. Lowering the p_T thresholds of the light jets on the one hand increases the statistics, but also causes the migration of events from low jet multiplicity regions to high jet multiplicity regions and thus worsens the

background modelling. In both cases, the negative effect overcomes the positive effects and the sensitivity shows no better performance in both configurations.

Then a study on two configurations with drastic cuts was performed in section 9.6. The motivation of CONFIG 4 is the proposed new jet trigger threshold for Run2 at $\sqrt{s} =$ 13 TeV. CONFIG 5 was suggested in order to test the robustness of the fit process. The neural network was not retrained for CONFIG 4 and 5. The events used for CONFIG 4 and 5 were directly drawn from the ICHEP samples. The drastic cuts reject a lot of the background, but also leads to lower statistics. Taking both effects into account, the results of CONFIG 4 and CONFIG 5 as in table 9.5 show similar performance as the ICHEP result.

Combining the results of all four configurations, one can conclude that in the optimization study different configurations show a relatively stable sensitivity and the ICHEP configuration shows the best performance.

The second part of this thesis concentrated on the MC generator study. The sensitivity of events on the MC generator free parameter h_{damp}, and the factorisation and renormalisation scales was studied. Several sets of samples were generated with different generators and settings. After the generation of the samples, cuts were applied to ensure the quality of events. The studies were performed on the semileptonic channel. PP8 samples with different h_{damp} choices, the nominal sample MG5_aMC+P8 and the reference sample MG5_aMC+Herwig++ were compared first. Three h_{damp} choices were tested, they are h_{damp} =258.5 GeV, h_{damp} =352.5 GeV and $h_{damp} = \infty$. The result in section 11.1 shows that the deviation between the PP8 samples with different h_{damp} choices is very small. This means the events generated with PP8 are not sensitive to the h_{damp} choice. In the scale study, one sample was generated with MG5_aMC and the renormalisation and factorisation scales were modified to $\mu_F = \mu_R = \Big(m_T(\text{Higgs}) \times m_T(\text{top}) \times m_T(\text{antitop})\Big)^{1/3}$, which is the default value used by Powheg. This sample was compared to another MG5_aMC sample, where the MG5_aMC default value for renormalisation and factorisation scales $\mu_F = \mu_R = H_T/2$ is used. They were compared with a PP8 sample with Powheg default renormalisation and factorisation scales and h_{damp} =352.5 GeV. The sample MG5_aMC+Herwig++ was also shown in the plots as a reference. It can be seen in the scale study in section 11.2 that MG5_aMC+P8 with modified scales lies between PP8 and MG5_aMC+P8 with default scale. Thus one can conclude, the scales are one important origin of the deviation between Powheg and MG5_aMC generator but not the only one. Different numerical implementations of the generators, PDFset could be other factors that lead to the observed deviation.

Further studies on the MC simulation of $t\bar{t}H$ process could be done with respect to

other free parameters. For instance PDFsets were not kept identical during this study. It would be interesting to test the sensitivity of the samples with respect to the PDFset choice. One could also estimate the $t\bar{t}H$ modelling systematic uncertainty for PP8 caused by scale variation and PDFset variation with the samples generated for this study.

Bibliography

[1] Y. Fukuda, et al., *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81**, 1562 (1998)

[2] D. Buttazzo, et al., *Investigating the near-criticality of the Higgs boson*, JHEP **12** (2013)

[3] R. D. Peccei, *The Strong CP problem and axions*, Lect. Notes Phys. **741**, 3 (2008)

[4] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716**, 1 (2012)

[5] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716**, 30 (2012)

[6] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector* (2016), ATLAS-CONF-2016-080

[7] M. Peskin, D. Schroeder, *An Introduction to Quantum Field Theory*, Advanced book classics, Addison-Wesley Publishing Company (1995)

[8] F. Englert, R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13**, 321 (1964)

[9] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13**, 508 (1964)

[10] C. Patrignani, et al., *Review of Particle Physics*, Chin. Phys. **C40(10)**, 100001 (2016)

[11] I. Aitchison, A. Hey, *Gauge Theories in Particle Physics: QCD and the Electroweak Theory, Third Edition*

[12] J. N. Ng, P. Zakarauskas, *QCD-parton calculation of conjoined production of Higgs bosons and heavy flavors in $p\bar{p}$ collisions*, Phys. Rev. **D29**, 876 (1984)

[13] Z. Kunszt, *Associated production of heavy Higgs boson with top quarks*, Nucl. Phys. **B247**, 339 (1984)

[14] S. Frixione, P. Nason, G. Ridolfi, *A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09**, 126 (2007)

[15] J. Alwall, et al., *MadGraph 5 : Going Beyond*, JHEP **06**, 128 (2011)

[16] T. Sjostrand, et al., *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178**, 852 (2008)

[17] G. Corcella, et al., *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP **01**, 010 (2001)

[18] M. Bahr, et al., *Herwig++ Physics and Manual*, Eur. Phys. J. **C58**, 639 (2008)

[19] S. Gieseke, et al., *Herwig++ 1.0: an event generator for e^+e^- annihilation*, JHEP **02**, 005 (2004)

[20] A. Buckley, et al., *General-purpose event generators for LHC physics*, Phys. Rept. **504**, 145 (2011)

[21] F. Maltoni, T. Stelzer, *MadEvent: Automatic event generation with MadGraph*, JHEP **02**, 027 (2003)

[22] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11**, 040 (2004)

[23] C. Oleari, *The POWHEG-BOX*, Nucl. Phys. Proc. Suppl. **205-206**, 36 (2010)

[24] R. Ellis, W. Stirling, B. Webber, *QCD and Collider Physics*, Cambridge University Press (2003)

[25] V. Hirschi, et al., *Automation of one-loop QCD corrections*, JHEP **05**, 044 (2011)

[26] R. Frederix, *Automation of next-to-leading order computations in QCD: The FKS subtraction*, JHEP **10**, 003 (2009)

[27] S. Frixione, P. Nason, B. R. Webber, *Matching NLO QCD and parton showers in heavy flavor production*, JHEP **08**, 007 (2003)

[28] S. Frixione, B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP **06**, 029 (2002)

[29] S. Frixione, et al., *Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations*, JHEP **04**, 081 (2007)

[30] P. Artoisenet, et al., *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP **03**, 015 (2013)

[31] 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**, 079 (2014)

[32] B. R. Webber, *A QCD Model for jet fragmentation including soft gluon interference*, Nucl. Phys. **B238**, 492 (1984)

[33] B. Andersson, et al., *Parton fragmentation and string dynamics*, Phys. Rep. **97**, 31 (1983)

[34] A. Buckley, et al., *Monte Carlo tuning and generator validation*, in *Proceedings, 1st International Workshop on Multiple Partonic Interactions at the LHC (MPI08)* (2009), `arXiv:0906.0075[hep-ph]`

[35] O. S. Bruning, et al., *LHC Design Report Vol.1: The LHC Main Ring* (2004), CERN-2004-003

[36] L. Evans, P. Bryant, *LHC Machine*, JINST **3(08)**, S08001 (2008)

[37] C. Lefèvre, *The CERN accelerator complex. Complexe des accélérateurs du CERN* (2008), CERN-DI-0812015

[38] M. Lamont, *Status of the LHC*, J. Phys. Conf. Ser. **455(1)**, 012001 (2013)

[39] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3**, S08003 (2008)

[40] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 1* (1999), CERN-LHCC-99-015

[41] A. Artamonov, et al., *The ATLAS Forward Calorimeter*, JINST **3**, P02010 (2008)

[42] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. **C77(5)**, 317 (2017)

[43] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data*, Eur. Phys. J. **C74(7)**, 2941 (2014)

[44] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data* (2016), ATLAS-CONF-2016-024

[45] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data* (2016), ATLAS-CONF-2016-024

[46] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at* $\sqrt{s}$ *=13 TeV*, Eur. Phys. J. **C76(5)**, 292 (2016)

[47] M. Cacciari, et al., *The anti-*k_t *jet clustering algorithm*, JHEP **04**, 063 (2008)

[48] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at* $\sqrt{s} = 13$ *TeV* (2015), ATL-PHYS-PUB-2015-015

[49] ATLAS Collaboration, *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS* (2015), ATLAS-CONF-2015-037

[50] ATLAS Collaboration, *Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector* (2015), ATLAS-CONF-2015-029

[51] M. Testa, *Performance of pile-up mitigation techniques for jets in pp collisions with the ATLAS detector* (2015), ATL-PHYS-PROC-2015-035

[52] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run* (2016), ATL-PHYS-PUB-2016-012

[53] R. Raitio, W. Wada, *Higgs-boson production at large transverse momentum in quantum chromodynamics*, Phys. Rev. **D19**, 941 (1979)

[54] W. Beenakker, et al., *NLO QCD corrections to* $t\bar{t}H$ *production in hadron collisions*, Nucl. Phys. **B653(1)**, 151 (2003)

[55] ATLAS Collaboration, *ATLAS Run 1 Pythia8 tunes* (2014), ATL-PHYS-PUB-2014-021

[56] R. D. Ball, et al., *Parton distributions for the LHC Run II*, JHEP **04**, 040 (2015)

[57] S. Alioli, et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06**, 043 (2010)

[58] H. Lai, et al., *New parton distributions for collider physics*, Phys. Rev. **D82**, 074024 (2010)

[59] T. Sjostrand, S. Mrenna, P. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05**, 026 (2006)

[60] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462**, 152 (2001)

[61] M. Cacciari, et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. **B710**, 612 (2012)

[62] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. **C71**, 1547 (2011)

[63] S. Alioli, et al., *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP **09**, 111 (2009)

[64] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. **D81**, 054028 (2010)

[65] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. **D83**, 091503 (2011)

[66] S. Schumann, F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **03**, 038 (2008)

[67] ATLAS Collaboration, *Measurement of W and Z Boson Production Cross Sections in pp Collisions at* $\sqrt{s}$*= 13 TeV in the ATLAS Detector* (2015), ATLAS-CONF-2015-039

[68] ATLAS Collaboration, *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton-proton collisions at* $\sqrt{s}$*=8 TeV with the ATLAS detector* (2014), ATLAS-CONF-2014-058

[69] M. Feindt, U. Kerzel, *The NeuroBayes neural network package*, Nucl. Instrum. Meth. **559(1)**, 190 (2006)

[70] K. Cranmer, et al., *HistFactory: A tool for creating statistical models for use with RooFit and RooStats* (2012), CERN-OPEN-2012-016

[71] J. Collins, *Choosing the renormalization/factorization scale (QCD)*, J. Phys. G: Nucl. Part. Phys. **17(10)**, 1547 (1991)

[72] E. Bagnaschi, A. Vicini, *The Higgs transverse momentum distribution in gluon fusion as a multiscale problem*, JHEP **01**, 056 (2016)

[73] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions from Powheg and Sherpa to ATLAS measurements of top pair production at 7 TeV* (2015), ATL-PHYS-PUB-2015-011

[74] T. Plehn, *LHC Phenomenology for Physics Hunters*, in *Proceedings of Theoretical Advanced Study Institute in Elementary Particle Physics on The dawn of the LHC era (TASI 2008)* (2008), `arXiv:0810.2281[hep-ph]`

[75] B. R. Webber, *Fragmentation and hadronization*, Int. J. Mod. Phys. **A15S1**, 577 (2000)

[76] ATLAS Collaboration, *Modelling of the $t\bar{t}H$ and $t\bar{t}V$ ($V = W, Z$) processes for $\sqrt{s} =$ 13 TeV ATLAS analyses* (2016), ATL-PHYS-PUB-2016-005

[77] C. Reißel, *Monte Carlo simulation and analysis of the $t\bar{t}H$ process with the ATLAS experiment at $\sqrt{s} = 13$ TeV*, Bachelor's thesis, Georg-August-Universität Göttingen (2016), II.Physik-UniGö-BSc-2016/11

Danksagung

This Master's course at the University of Goettingen took two years. I want to thank Arnulf Quadt for the delightful lectures on experimental particle physics. He encouraged me to participate the DESY summerschool 2016, during which my research skills improved a lot. I also want to thank him for the open discussions and support during the Master's thesis period.

I want to thank Boris Lemmer, Lisa Shabalina, María Moreno Llácer, Royer Edson Ticse Torres, Clara Nellist and Matteo Mantoani. Matteo supervised the first part of this thesis. Lisa and María guided my Master's thesis project through the whole year. Boris, Royer and Clara were always there for any questions and discussions. Also thanks to all members of the II. Institute of Physics for their support.

Finally, I could not have done the Master's without a lot of support from my family and close friends. Thank you so much.

Erklärung nach §17(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen:

Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.
Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den January 25, 2018

(Gaoyuan Wang)