GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für
Physik

Master's Thesis

Search for the $t\bar{t}H$ signal with $H \to b\bar{b}$ in pp collisions at 13 TeV with the ATLAS detector at the LHC

prepared by

Ishan Pokharel

from Kathmandu, Nepal

at the II. Physikalisches Institut

Thesis number:	II.Physik-UniGö-MSc-2016/01
Thesis period:	15th October 2015 until 15th April 2016
First referee:	Prof. Dr. Arnulf Quadt
Second referee:	Prof. Dr. Stan Lai

Abstract

A study of discriminating variables for signal and background separation is presented in this thesis. The variables are obtained by training a Neural Network to discriminate between signal and background. The signal considered is the $t\bar{t}H$ process with the Higgs boson decaying into a $b\bar{b}$ pair and $t\bar{t}$ being the main background. This thesis is conducted within the scope of the analysis currently being performed at the ATLAS experiment for the search of the $t\bar{t}H, (H \rightarrow b\bar{b})$ process at 13 TeV. A Neural Network is trained in the most sensitive region which contains at least 6 jets, of which at least 4 are b-tagged. The study uses an integrated luminosity of 3.2 fb^{-1} of proton-proton collision data at $\sqrt{s}$= 13 TeV, collected using the ATLAS detector at the LHC in 2015. Several discriminating variables have been studied for the Neural Network training using $t\bar{t}H$ events as signal and $t\bar{t}$+jets events as background. A NN variable with a separation power between $t\bar{t}H$ and $t\bar{t}$ events of 12.2% is obtained, which represents about a twofold increase compared to discriminating power of normal individual multivariate variables.

Keywords: High Energy Physics, Master thesis, $t\bar{t}H$, Associated production, Neural Network, Higgs boson, Standard Model, Yukawa Coupling, Top quark, jets, electroweak theory, MVA variables.

Contents

1. Introduction **1**

2. Theoretical Background **3**

2.1. The Standard Model of Particle Physics 3

2.1.1. Electroweak Symmetry Breaking 6

2.2. Limitations of the Standard Model 9

3. The Large Hadron Collider and the ATLAS Detector **11**

3.1. The Large Hadron Collider 11

3.2. The ATLAS Detector 13

3.2.1. Inner Detector 14

3.2.2. Calorimeters 16

3.2.3. Muon Spectrometer 17

3.2.4. The Magnet System 17

3.2.5. Trigger and data acquisition 18

4. Monte Carlo Simulated Samples **19**

4.1. $t\bar{t}H$ Signal Modelling 20

4.2. $t\bar{t}$+jets Modelling 20

4.3. Single-Top Modelling 21

4.4. W/Z+jets Modelling 21

4.5. Diboson Modelling 22

4.6. $t\bar{t} + V$ Modelling 22

4.7. Common treatment of MC samples 22

5. The $t\bar{t}H, H \to b\bar{b}$ process **25**

5.1. Motivation 25

5.2. $t\bar{t}H$, $H \to b\bar{b}$ event topology 26

5.3. Object Definition 27

5.3.1. Leptons 27

5.3.2. Jets 28

5.3.3. b-tagging 29
5.4. Event Pre-Selection 31
5.5. Analysis Strategy 32
5.6. Multivariate Analysis 34
5.7. Neural Networks 34
5.7.1. NeuroBayes Algorithm 35

6. Analysis and Results **37**
6.1. Discriminating Variables 38
6.2. Neural Network training 39

7. Conclusion and Outlook **55**

A. List of MC Samples **61**

B. Event Yields for all regions **75**

1. Introduction

–The worthwhile problems are the ones you can really solve or help solve, the ones you can really contribute something to... No problem is too small or too trivial if we can really do something about it.
–The first principle is that you must not fool yourself; and you are the easiest person to fool.

–R.P. Feynman

The following quotes by Richard P. Feynman resonate deeply to how research in particle physics is conducted in the modern age. The complexities of the problems are enormous, both on the theoretical as well as the experimental side. The Standard Model is known to be incomplete, and the validation of all aspects of the theory has taken physicists to technological boundaries. Gone are the days when a handful of researchers with a theory could build an apparatus that can test the said theory effectively.

The search for theoretical particles, the fundamental units that nature has deemed should exist, has taken an enormous effort of imagination as well as technical expertise. In the effort, there lies a comedic notion of how the research is conducted; by constructing enormous machines that ultimately smash two minute masses together to see what they break up into, almost akin to a child trying to break open anything given to them to 'see what is inside'. That curiosity still remains, the slightly child-like behaviour in the particle physics world coupled with an insatiable desire to make sure that the smashing is done in a highly controlled and precise manner so that one now tries to recreate how nature in its infant form would have looked.

The tools in the arsenal of the physics community has also developed by leaps and bounds. Highly complex analysis can be done in record time that allows a physicist to build confidence in their results in a probabilistic world. An example can be seen in the current experiments conducted at the CERN laboratory in Geneva (Switzerland), where the data generated would be too large to ever be completely analysed in the entirety of human existence.

This thesis analyses data generated in one of these experiments, the ATLAS detector at the LHC collider. The final aim is to find a very rare physics signal, the $t\bar{t}H$ process that is

created in the collision of two proton beams. This process provides a direct measurement of the Yukawa coupling between the Higgs boson and the top quark. Since the top quark is the heaviest among the quarks, this coupling is postulated to be close to unity and finding this signal would be a validation of a crucial SM prediction or a window into any deviations in the theory.

2. Theoretical Background

The aim of particle physics is to understand the interactions between the fundamental constituents of nature and their structure. The inception of the field can roughly be regarded as the discovery of the electron about a century ago. In the last 80 years, with the discovery of new particles in cosmic rays, the cumulative knowledge in the field has been combined to form the Standard Model (SM) of particle physics.
The SM is the most comprehensive theory of particles and their interactions in an elegant mathematical framework. It describes all known particles that constitute matter and account for 3 out of 4 fundamental interactions. The last 40 years have yielded discovery and precision measurements of all the fundamental particles postulated in the SM, the last one being the discovery of the Higgs boson in 2012 [1] [2]. The precision measurements are possible mainly due to advancements in accelerator and detector designs. With higher energies and luminosities, the experiments running in the LHC are expected to probe even further into previously unknown energy realms.
This chapter describes the theoretical foundation for the thesis. A brief outline of the Standard Model with a particular emphasis on the Brout-Englert-Higgs mechanism will be presented. Furthermore, a discussion about the discovery of the Higgs boson in 2012 and the limitations of the Standard Model will be discussed.

2.1. The Standard Model of Particle Physics

The Standard Model describes the interactions between the fundamental particles that constitutes matter (*fermions*) via the 'force-carrier' particles (*bosons*). From quantum field theory, the fermions are treated as field quanta and their interactions with bosons follow a given principle: the invariance of the Lagrangian under a local gauge transformation of given symmetry group [3]. Local gauge transformation requires that the fields considered vary differently at any point in space time.
Borrowing the concepts from Classical Mechanics, the invariance in displacement and time lead to the conservation of momentum and energy. Similarly, invariance in gauge fields leads to a conserved charge and the quantisation of the field leads to the gauge boson that

only interact with particles that are defined in this field. This concept can be seen in the case of Quantum ElectroDynamics (QED) [4], where the Lagrangian which is invariant under a local phase transformation of the electromagnetic field leads to the introduction of a boson, the photon and a conserved electric charge. This treatment of the quantisation and the nature of the field and the boson depends on the symmetry group that the gauge transformation belongs to.

In the Standard Model, the gauge field belongs to the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ symmetry group. This gives the description of the strong, weak and the electromagnetic interactions between the fermions and the bosons in the theory.

The $SU(3)_C$ is a non-Abelian group[1] that describes the strong interactions [5]. The 'C' denotes the conserved charge, which in this case is the *colour*, of the interaction of the strong field. The Lie algebra of the group generates eight massless operators known as the *gluons* which are the interacting bosons of this field. The coupling between the gluons and the quarks is denoted by g_s.

The $SU(2)_L$ is a non-Abelian chiral symmetry group that describes the weak interactions. The symmetry operation gives rise to three vector bosons, (W^1, W^2 and W^3) which couple only to left handed fermions (denoted by L). The conserved quantity is the weak isospin ($\vec{I}$) and the coupling is denoted by g.

The $U(1)_Y$ is an unitary Abelian group that describes the electromagnetic interactions. The mediator is a single vector boson denoted by B and the conserved charge (Y) is known as the *hypercharge*.

The unification of the weak and the electromagnetic interactions was achieved by Glashow, Salam and Weinberg [6] [7] in the 60's by looking at the gauge invariance in the $SU(2)_L \otimes U(1)_Y$ symmetry group. As noted above, the electromagnetic interactions are described by the $U(1)_Y$ group, where the force mediating boson is the photon. The photon exhibits chiral symmetry, wherein it does not distinguish between left- or right-handed fermions and interacts with any particle that has charge (Q), the conserved quantity of the electromagnetic field. The chirality determines the interaction of the boson to the fermionic field, which contains spinors. The right handed component is given by the action of the operator $(1-\gamma^5)/2$ and left handed component by $\{(1+\gamma^5)/2\}$ [2]. The charge is related to the *hypercharge* and is given by

$$Q = I_3 + \frac{Y}{2} \tag{2.1}$$

[1] $(\forall\ a, b \in \mathbf{G}\ a \cdot b = b \cdot a)$

[2] γ^5 is the fifth gamma matrix and not γ to the fifth power

where I_3 is the third component of the weak isospin. One can now see the quantised vector bosons of the $SU(2)_L \otimes U(1)_Y$ symmetry group as :

$$B = A\ cos(\theta_W) - Z\ sin(\theta_W) \tag{2.2}$$

$$W^3 = A\ sin(\theta_W) + Z\ cos(\theta_W) \tag{2.3}$$

Here, A represents the neutral boson, i.e. photon and the Z is another neutral boson with a different coupling to left and right handed fermions. The remaining force carriers of the weak interactions can be obtained by a linear combination of the W^1 and W^2 given by

$$W^{\pm} = \frac{1}{\sqrt{2}}(W^1 \mp iW^2) \tag{2.4}$$

This mathematical treatment of the symmetry group representing the Standard Model gives rise to a total of 5 bosons that are responsible for the electromagnetic, weak and strong interactions in the respective fields. The bosons all have a spin of 1, except the Higgs boson which is scalar and thus has a spin of 0. Table 2.1 gives a summary of the properties of the bosons in the Standard Model. The Higgs boson has been included for completeness. A full discussion of the theory behind spontaneous electroweak symmetry breaking will follow in the next section.

Interaction	**boson**	**charge**	**mass (GeV)**	**spin**
Electromagnetic	γ	0	0	1
Weak	$W^{\pm}$	$\pm 1e$	80.385 ± 0.015	1
	Z^0	0	91.1876 ± 0.0021	1
Strong	gluons	0	0	1
Higgs	H	0	125.7 ± 0.4	0

Table 2.1.: Gauge bosons with their properties in the Standard Model. The masses of the $W^{\pm}$, Z^0 and the Higgs boson are taken from [8].

Other than the bosons, all other known elementary particles are called fermions which have a spin of 1/2. The fermions are further classified into leptons, which have an integer charge of 0 or 1 and quarks, which have a charge either of -1/3 or +2/3. The leptons and quarks can be arranged into three families that show a distinct mass hierarchy. Each family contains a neutral and charged lepton and two quarks with fractional charge. For leptons, the mass increases while going from a neutral lepton to a charged lepton within the same family. Table 2.2 presents a summary of the properties of fermions known in the Standard Model.
The fermions from the first generation make up most of the known matter in the universe.

Family	Leptons			Quarks		
	flavour	**Q (e)**	**mass (GeV)**	**flavour**	**Q (e)**	**mass (GeV)**
1^{st}	ν_e	0	$< 2.2 \times 10^{-9}$	u (up)	2/3	$2.3^{+0.7}_{-0.5} \times 10^{-3}$
	e	-1	511×10^{-6}	d (down)	-1/3	$4.8^{+0.5}_{-0.3} \times 10^{-3}$
2^{nd}	ν_μ	0	$< 0.17 \times 10^{-3}$	c (charm)	2/3	1.275 ± 0.025
	μ	-1	105.7×10^{-3}	s (strange)	-1/3	$(95 \pm 5) \times 10^{-3}$
3^{rd}	ν_τ	0	$<15.5 \times 10^{-3}$	t (top)	2/3	173.34 ± 0.76
	τ	-1	1.777	b (bottom)	-1/3	4.18 ± 0.03

Table 2.2.: Fermions with their properties in the Standard model. The masses and charges are taken from [8].

Further, every particle other than neutrinos have a corresponding anti-particle which have the signs in their quantum numbers inverted while having the same mass. The quarks possess an extra quantum number called *colour*, with three possible states: *red*, *blue* and *green*.

The interactions of particles involve scattering, annihilation or creation and radiation. The electromagnetic force is realised by the exchange of photons between charged particles. Similarly, the weak force is realised by the exchange of $W^{\pm}$ or Z^0 bosons. Since the W bosons also have charge, they can interact with the photons themselves in the electric field. The strong force is realised by the exchange of gluons. Gluons can interact only with particles that contain colour charge, i.e. the quarks and other gluons. Finally, all particles in the Standard Model attain their mass by their interaction with the Higgs boson in the Higgs field.

2.1.1. Electroweak Symmetry Breaking

The discussion in the earlier chapters has focused on the construction of the Standard Model on symmetries (gauge symmetries for bosons and chiral symmetries for fermions). One of the prominent features of this concept is the lack of understanding of how the particles acquire mass. This was seen to contradict experimental evidence from the discovery of the W and Z bosons, which were massive. Further, introducing an explicit mass term in the Lagrangian violates the local invariance.

A simple solution was proposed by Brout, Englert and Higgs, the BEH mechanism or so called "Higgs" mechanism, that employed the spontaneous symmetry breaking [9] [10], in which the Lagrangian remains invariant under the symmetry group, but the physical states, in particular the ground state, do not. This spontaneous breaking of symmetry allows the particles to interact with an underlying field called the Higgs field, and thus acquire mass.

For electroweak interactions, let us consider an isospin doublet of a complex scalar field given by:

$$\Phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \tag{2.5}$$

The field can be represented as

$$\Phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \tag{2.6}$$

The contribution of this field in the Lagrangian, in its most general form where we disregard negative terms due to symmetry considerations and higher powers of Φ due to renormalisation, can be expressed as:

$$\mathcal{L} = (D_\mu \Phi)(D^\mu \Phi)^\dagger - \mu^2 \Phi^\dagger \Phi - \lambda(\Phi^\dagger \Phi)^2 \tag{2.7}$$

where the last two terms are the potential of the Lagrangian. The D_μ represents the covariant derivative which implements the electroweak gauge invariance and is given as:

$$D_\mu = (\delta_\mu - igW_\mu^i I_i - ig'\frac{Y}{2}B_\mu) \tag{2.8}$$

The spontaneous symmetry breaking is based on the non-invariane of the vacuum state with respect to the $SU(2)$ symmetry. Looking at equation 2.7, when μ^2 and λ are both positive, the minimum of the potential is found uniquely to be in $\Phi = 0$, which implies $\phi_1 = \phi_2 = 0$. For $\mu^2 < 0$, the minimum of the potential is described by an infinite solutions satisfying :

$$\phi_1^2 + \phi_2^2 = \frac{-\mu^2}{\lambda} \approx \upsilon \tag{2.9}$$

The value of the field in the minimum potential is also known as the *vacuum expectation value* (υ). The choice of any of this ground state configuration, given by equation 2.9, produces the apparent 'symmetry breaking'. Now, the theory can be derived in a perturbative way around the new minimum[3] configuration. Through a suitable rotation and re-parameterisation of the field, Φ can now be defined as:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \upsilon + H(x) \end{pmatrix} \tag{2.10}$$

Here H is a scalar field with 0 vacuum expectation value and corresponds to the physical Higgs boson. The ground state configuration $\Phi = 1/\sqrt{2}(0, \upsilon)^T$ is chosen so as to maintain

[3]Let $\phi_1(x) = \upsilon + \eta(x)$ and $\phi_2(x) = \xi(x)$; where $\eta(x)$ and $\xi(x)$ are new fields.

the $U(1)_Q$ unitary symmetry group and this does not generate a mass term for the photon. The hypercharge is set to 1, and looking at equation 2.1, the Higgs boson does not carry charge.

Substituting equation 2.10 into the Lagrangian equation 2.7, one sees terms proportional to $W^+_\mu W^{-\mu}$ and $Z_\mu Z^\mu$ in the covariant derivative, which introduces explicit mass terms to the corresponding vector bosons. The three initial components of the additional scalar field have been absorbed by the $W^\pm$ and Z bosons creating longitudinal degrees of freedom. The masses of these vector bosons can all be expressed in terms of υ and the $SU(2)_L \otimes U(1)_Y$ coupling constants:

$$m_W = \frac{1}{2}\upsilon g, \quad m_Z = \frac{1}{2}\frac{\upsilon g}{cos\theta_W} \tag{2.11}$$

Also, the mass of the Higgs boson can be derived via the expansion of the scalar potential and is given by:

$$m_H = \sqrt{2\lambda}\upsilon \tag{2.12}$$

The scalar doublet field can also be used to compute the mass of the fermions. The interaction between the scalar doublet and the fermion field can be written as:

$$\mathcal{L}_Y = -\Gamma^{ij}_u \overline{Q^i_L} \epsilon \Phi^* u^j_R - \Gamma^{ij}_d \overline{Q^i_L} \Phi d^j_R - \Gamma^{ij}_e \overline{Q^i_L} \epsilon \Phi e^j_R + h.c \tag{2.13}$$

where Γ^{ij}_u, Γ^{ij}_d and Γ^{ij}_e are 3×3 matrices that mix the contribution of the three fermion families and ϵ is the asymmetric tensor. Now, applying equation 2.10 to the above Lagrangian, we get:

$$\mathcal{L}_Y = -\frac{1}{\sqrt{2}}(\upsilon + H) \cdot [h^i_l(\overline{e^i_L} e^i_R + \overline{e^i_R} e^i_L) + h^i_u(\overline{u^i_L} u^i_R + \overline{u^i_R} u^i_L) + h^i_d(\overline{d^i_L} d^i_R + \overline{d^i_R} d^i_L)] \tag{2.14}$$

which gives rise to the interactions of the fermions with the Higgs field and mass terms for the fermions given by:

$$m_f = \frac{\upsilon h^i_f}{\sqrt{2}} \tag{2.15}$$

Both the interaction and the masses of the fermions are dependent on h^i_f, known as the *Yukawa coupling*, demonstrating that the interaction of the fermions with the Higgs field is proportional to their masses. Equation 2.14 is obtained by a redefinition of the fermion fields that diagonalise the Γ matrices of equation 2.13. The Yukawa couplings are thus, the eigenvalues of these matrices.

For leptons, the transformation applied has no affect due to the absence of right-handed neutrinos. For quarks, this transformation exhibits the mixing between fermionic families,

which manifests itself in the weak interactions. The mixing between the weak eigenstates of the down-type quarks and the corresponding mass eigenstates is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.16}$$

The CKM matrix is fully specified by 4 parameters: 3 mixing angles controlling the mixing between each family pair and 1 complex phase responsible for CP-violation. Thus, the Standard Model is a renormalisable, unitary theory which can be used to perturbatively calculate physics processes in high energy. The theory contains 18 parameters that have to be provided via measurements. These parameters are:

- 3 coupling constants for the electromagnetic, weak and strong fields: g, g' and g_S
- 2 parameters from the electro-weak symmetry breaking : υ and m_H
- 9 Yukawa couplings for the fermion masses.
- 4 parameters from the CKM matrix.

2.2. Limitations of the Standard Model

With the far reaching theoretical predictions and their successive validation by experimental observations, the SM does have some drawbacks from being a complete theory. Certain limitations of the theory are discussed below:

- The gravity problem: The SM does not describe the fourth force seen in nature. A complete and coherent description of gravity within the realm of the SM has not yet been found.
- Mass hierarchy problem: The masses of the fermions and quarks, as shown in table 2.2, show a distinct trend within a family as well as through the families. There is no complete theory to describe why this happens yet. Further, the origins of CP-violation in weak interactions are also not described.
- The mass asymmetry problem: The SM has lacked to explain the existence of a large amount of Dark Matter in the universe.

- The naturalness problem[4]: In the SM, the mass of the Higgs boson discovered in 2012 is unusually low. The average mass of the Higgs boson is expected to be in the TeV range. Only a 'fine-tuning' of the parameters of the SM allows m_H=125 GeV. A complete fundamental idea behind this is not yet been understood. Certain theoretical models have been suggested (example, SuperSymmetry) for heavier symmetric partners to the SM bosons, but no concrete experimental evidence for their existence has been found so far.

- The neutrino oscillation problem: As discussed in Section 2.1, neutrinos are considered massless in the SM, which contradict recent measurements. In the SM Lagrangian, a neutrino with mass can easily be incorporated by also incorporating right-handed neutrinos and then treating them like Dirac-fermions. Further, since neutrinos are defined as particles without a charge or colour, they can be treated as Majorana-fermions while still incorporating the mass as seen in experiments. Also, the mass hierarchy in the SM has not been better understood.

[4]A naturalness of a system assumes that unless a detailed explanation exists, all parameters contributing to an effective Lagrangian that preserve a given symmetry should appear in natural coefficients.

3. The Large Hadron Collider and the ATLAS Detector

This chapter describes the experimental setup that detected the events that were analysed in this thesis. The setup consists of the largest hadronic collider in the world, the LHC and the biggest particle detector ever created, the ATLAS detector. The LHC accelerates protons and collides them in the ATLAS detector, which is set up to detect the physics objects created.
The first part describes the LHC with a general description on the working of the collider. The second part describes the ATLAS detector with a detailed description of sections and components used and their role in detecting the physics objects.

3.1. The Large Hadron Collider

The LHC is the world's largest hadron accelerator, currently colliding protons at center-of-mass energy ($\sqrt{s}$) of 13 TeV and an instantaneous luminosity in the order of 10^{34} cm^{-2}s^{-1}. It is located 100 m below ground in the Franco-Swiss border in Geneva. It is a 27 km circular collider designed to collide particles up to $\sqrt{s}$= 14 TeV.
The main LHC tunnel is not able to accelerate the protons from an initial rest state to 14 TeV in one run. It requires pre-accelerators that increase the acceleration in stages as shown in Figure 3.1. The hydrogen atoms are ionised and accelerated to 50 MeV at the *LINAC II* and are passed on to the *Booster* to be accelerated to 1.4 GeV. These protons are then injected to the *Proton Synchrotron* (PS) where their acceleration is increased up to 25 GeV. Finally, before injection to the LHC, they are injected into the *Super Proton Sunchrotron* (SPS), which increases the acceleration to a minimal LHC injection acceleration of 450 GeV. The LHC then accelerates the protons to a desired $\sqrt{s}$ value before collision. The LHC is also designed to accelerate heavy ions, primarily used by the ALICE experiment. In the case of heavy ion acceleration, they are first accelerated by the *LINAC* III and the *Low Energy Ion Ring* (LEIR) before being injected to the PS.
The LHC electric fields in cavities for the acceleration and magnets for bending the ion

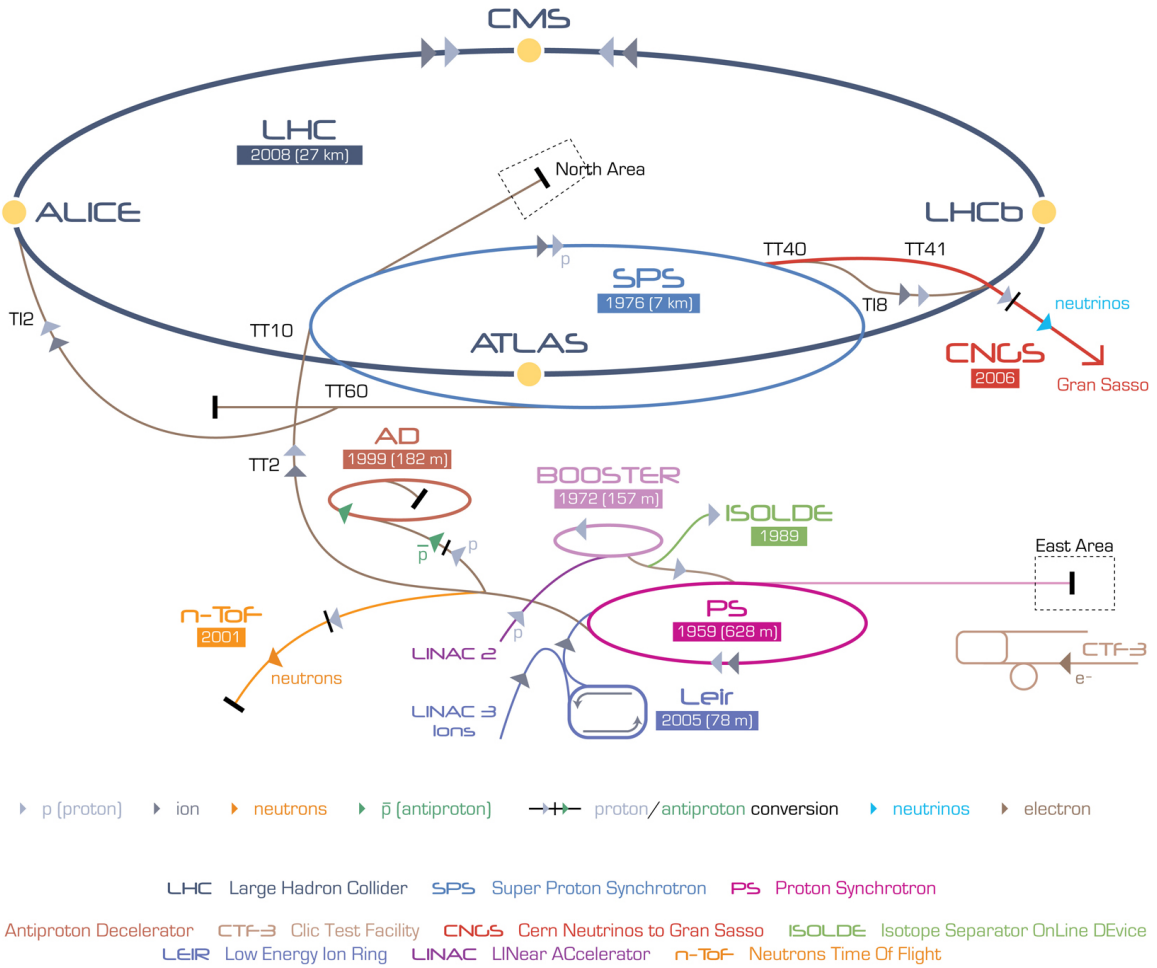


Figure 3.1.: Illustration of the LHC with the pre-accelerators for proton and heavy ion injection in the CERN accelerator complex.

beam and focusing the relativistically accelerated ions. For bending the beam, the LHC is designed with 1232 dipole magnets, each of which is 14.3 m long and provides a magnetic field of 8.33 T [11]. The magnetic field is oriented perpendicular to the direction of acceleration, to allow for the beam to be bent. To achieve such high magnetic fields, the magnets are constructed using superconducting materials, primarily niobium-titanium which is cooled to 1.9 K using supra fluid helium. Further, both beam pipes are fitted in the same enclosing so that one cryostat module and magnets can be used for both beams at the same time, reducing material and cooling costs.

As mentioned earlier, acceleration of the ionised beam is achieved by cavities that provide uniform oscillating electric fields. LHC uses 8 cavities per beam, located at a single point in the circumference, such that the acceleration is increased after every revolution around the accelerator. This design also prevents a continuous beam structure, thereby allowing protons to be grouped into *bunches*, where each bunch contains around 10^{11} protons. In

Run-2, the LHC can accommodate upto 2808 bunches per beam, with a spacing of 25 ns. The acceleration of the protons before collision is in excess of 0.999 c. Quadrupole magnets are used to focus the beams for a better collision probability. After the acceleration and focusing, two beams in the beam pipe travel in opposite directions and are made to collide in one of the 4 main collision points that house different detectors. Looking at Figure 3.1, the 4 main detectors are ATLAS, CMS, ALICE and LHCb. This thesis uses data recorded by the ATLAS detector.

3.2. The ATLAS Detector

The ATLAS (A Toroidal LHC ApparatuS) [12] is one of the two general purpose detectors built about 100 m underground within the LHC. The detector is the biggest of its kind, with physical dimensions encompassing 44 m in length and 25 m in width and weighing around 7000 tons.

The focus for the ATLAS and the CMS detector, was to search for the Higgs boson,

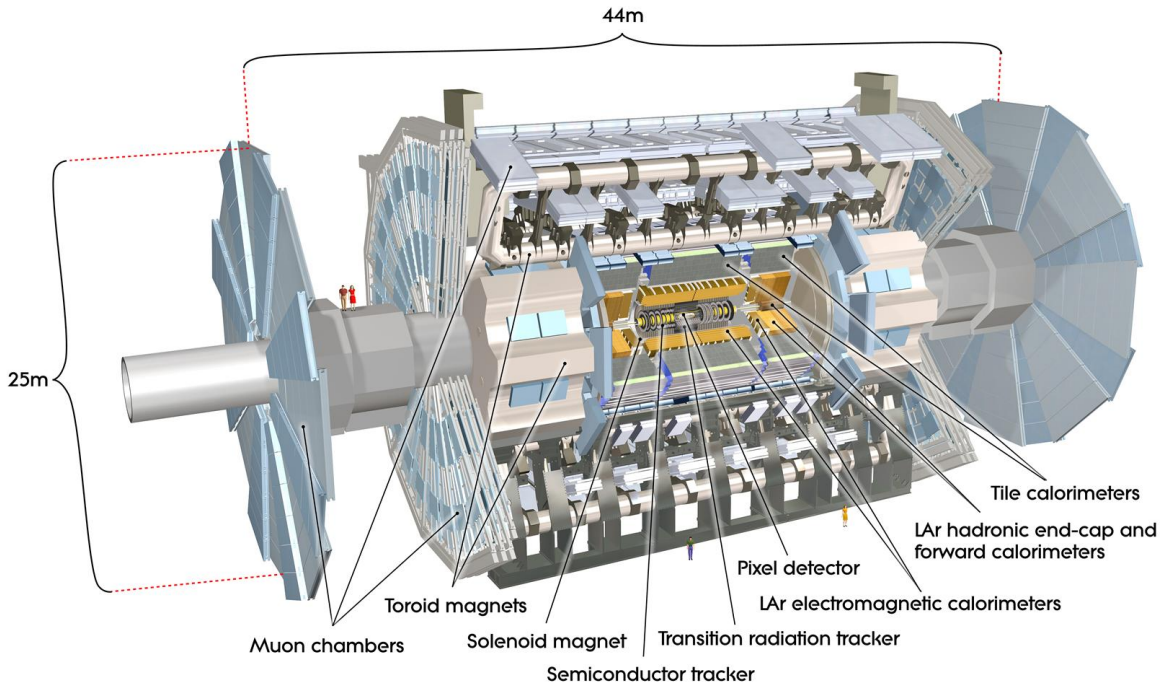


Figure 3.2.: Schematic representation of the ATLAS detector along with the sub-systems and magnets.

including its properties, and also for supersymmetric particles that might provide insight on beyond the SM physics. Thus, the design of the detector was done such that the

kinematic properties of a large spectrum of particles over a wide energy range (few GeV to TeV) could be achieved. These included charged leptons (electrons, muons and tauons), photons, jets produced during the hadronisation of quarks and gluons and particles that can not be directly detected like the neutrinos. The latter particles are identified via the measurement of the imbalance in p_T and energy in the transverse plane for the event.
The ATLAS detector has a cylindrical structure as shown in Figure 3.2. The analysis of particles produced is done in the cylindrical coordinate system with the z-axis oriented along the direction of the beam and z = 0 at the point of collision. The azimuthal angle ϕ is measured along the beam pipe with $\phi = 0$ degree pointing towards the centre of the LHC ring. The polar angle θ is measured from the beam axis towards the azimuthal plane. To make calculations simpler, a quantity called *pseudorapidity* η is chosen, where $\eta = -ln\ tan(\theta/2)$. Since the differences in η are Lorentz invariant, this makes the calculations of various physics properties simple and straightforward.
The ATLAS detector can be thought of an assembly of various sub-detectors, each with a specific function. These sub detectors can be broadly divided into three components, the Inner Detectors, the calorimeters and the muon detectors. A brief description of all the components is given below.

3.2.1. Inner Detector

The ATLAS Inner Detector (ID) is used primarily in tracking particles produced from hard interactions. The detector provides an efficient tracking of the physics objects within $|\eta| < 2.5$ and p_T and an energy resolution from as low as 0.5 GeV up to a few TeV. A precise measurement of the particle curvature in the magnetic field requires high spatial resolution, which is achieved by highly segmented detectors. Further, it provides a very precise reconstruction of primary and secondary vertices, that play a very crucial role in object reconstruction.
The inner detector is further composed of 3 sub-detector layers. In order of their proximity to the beam pipe, they are the *Pixel Detector*, the *Semiconductor Tracker* (SCT) and the *Transition Radiation Tracker* (TRT). All tracking detectors are organised in a concentric cylindrical barrel configuration oriented along the beam axis.

Pixel detector

The Pixel Detector is built using silicon pixel technology, which provides a good pattern recognition performance in dense track environments which occur near the interaction point. This detector is configured to provide at least 3 precise hits near the interaction

point for track reconstruction. Due to its proximity to the beam, this detector is also crucial for the reconstruction of the primary vertex and the track impact parameter[1]. The resolution on the impact parameter is completely dominated by the Pixel Detector. The innermost layer of the detector is the IBL (Insertable B-Layer) [13]. The IBL is located at a distance of 3.2 cm from the beam pipe and is employed for tracking, vertexing and b-tagging at high luminosities. The addition of this new layer is expected to increase the overall tracking resolution of the pixel detector module as a whole and compensate some radiation damage already suffered by other inner detector components.
The high granularity is obtained using a silicon bulk segmented into a matrix of pixels that allow a simultaneous measurement of two spatial coordinates. A single pixel is 50 μm $\times$ 400 μm and has a thickness of 250 μm. The basic unit of the detector is the module, where each module is 6.28 cm $\times$ 2.2 cm, together with all the electronics required for operation and the module is segmented in a matrix of 144 $\times$ 328 pixels.

The SemiConductor Tracker

The SemiConductor Tracker (SCT) layer consists of 4,088 modules in 4 concentric cylindrical barrel layers and 9 end cap layers in each side. The four barrel layers are positioned at 29.9 cm, 37.1 cm, 44.3 cm and 51.4 cm away from the beam pipe, respectively. The SCT uses silicon as the sensor material. The tracking modules are installed as strips, and a single strip electrode yields one dimensional tracking information only. Two-dimensional tracking information can be obtained by a combination of two strips positioned at an "stereo-angle" to each other. The readout of the particles can be obtained at the intersection of these strips, but the efficiency of this model depends only on a few particles providing the hits at a given time. The advantage of this strip model to the pixel detector is that there is less material and electronics in the SCT than in the pixel detector.
An overall resolution of 17 μm is achieved in the direction of the strip pitch and 580 μm is achieved in the direction of the strip crossing.

The Transition Radiation Tracker

The Transition Radiation Tracker (TRT) are composed of gas filled straw tubes interleaved with a matrix of polypropylene fibres. The tubes have a diameter of about 4 mm and a length of 144 cm and 37 cm in the barrel and end cap regions respectively. The combination of these materials is ideally suited for the high radiation operating environment. A signal is generated when a relativistic particle crosses two materials with different

[1]The minimum distance between the reconstructed track and the primary vertex.

dielectric constants. This is ideal for the detection of electrons. The TRT is designed to produce on average, 36 hits per track in the $|\eta|< 2$ region.
The TRT resolves the tracks in the R$-\phi$ region with a spatial resolution of about 130 μm. Two independent threshold signals are applied during signal processing. The high threshold signals are used to identify photons from other transition radiation processes, and a low threshold is used to discriminate electrons from other charged particles.

3.2.2. Calorimeters

The calorimeters of the ATLAS detector surround the ID and are designed to measure the energy of the physics objects. The basic principle employed is for the physics objects to completely deposit their energy by coming to rest within the calorimeter. In order to achieve this, the calorimeters have to be thick and also be fully ϕ-symmetric. They are segmented into a barrel component and two end cap components for a pseudorapidity coverage of up to $|\eta|< 4.9$. The ATLAS calorimeter system consists of an inner electromagnetic calorimeter surrounded by an external hadronic calorimeter. The electromagnetic calorimeters are designed to detect the energies of photons and electrons and have a high resolution. The hadronic calorimeters are used for measuring the energies of hadronic objects, i.e. mostly jets, and have a coarser resolution compared to the former but still allows a precise measurement of jet kinematics and also account for the missing energy calculations.

Electromagnetic Calorimeter

The Electromagnetic CALorimeter (ECAL) is located next to the TRT solenoid system. It is a sampling calorimeter, in the sense that it detects objects by absorbing them. The absorbing material used is lead, whereas liquid Argon (LAr) is used for signal generation. The signal is generated by ionisation and is read out via electrodes. High voltage is applied between the lead plates for signal amplification as well as to collect the ionised electrons from the interaction in the LAr. The coverage of the ECAL is up to $|\eta|< 3.2$.

Hadronic Calorimeter

The Hadronic Calorimeter consists of three different independent sampling calorimeters depending on their resistance to the radiation flux and performance requirements. In the central region, the Tile Calorimeter covers a range from $0 < |\eta| < 1.7$. It is a 11.4 m long cylindrical detector with an inner radius of 2.28 m and an outer radius of 4.25

m. It uses iron as the absorber and plastic scintillators as the active parts. In order to read out the scintillation lights, wavelength shifting fibres are used to direct the light to photomultipliers.
The second component is the Hadronic End-cap Calorimeter (HEC). It covers a region between $1.5 < |\eta| < 3.2$ and uses copper plates as absorber and LAr as the active part. Similar to the operation in the electromagnetic calorimeter, the electric signal produced in the HEC is collected by cathodes on the plates. Finally, for the most forward region of the detector, a Forward CALorimeter (FCAL) is used that covers a region between $3.1 < |\eta| < 4.9$. It consists of tungsten rods as absorbers, embedded in a copper matrix. A thin layer of LAr between the two provides the active part.

3.2.3. Muon Spectrometer

The Muon Spectrometer is the final outer detector layer of the ATLAS detector, covered by the toroidal magnets. The purpose of this detector is to provide a precise measurement of the muon p_T up to very high energies (≈ 3 TeV) and in the pseudorapidity range of $|\eta| < 2.7$. Since almost all the particles are stopped by the calorimeters, the only physics object that are not detected by the internal components are the muons. Thus, the Muon Spectrometer comprises the final layer in the ATLAS detector.
Principally, the Monitored Drift Tubes are used to track the muons. These aluminium drift tube modules consist of 3 to 8 layers of drift tubes and have a diameter of nearly 30 mm with a 50 μm tungsten-rhenium wire. They are mostly filled with Argon gas and mounted cylindrically in three regions around the beam pipe in the barrel region. The end cap consists of three layers mounted perpendicular to the direction of the beam.

3.2.4. The Magnet System

The magnet system in any particle detector plays a crucial role. The tracks and properties of any charged particles are measured via their deflection in a magnetic field. The magnet system of the ATLAS detector consists of four large superconducting magnets that provide a field that is mostly orthogonal to the particle trajectory. Three of the magnets are open air toroids positioned behind the muon spectrometer on the outermost barrel configuration of the detector and a central solenoid enclosing the beam pipe and covered by the calorimeters. The schematic can be seen in Figure 3.2. This design feature has an advantage of extending the pseudorapidity coverage of the muon spectrometer while having no magnetic field effect inside the calorimeters so that it does not adversely affect their performance.

The central solenoid produces a magnetic field that is parallel to the beam axis and bends the particles in the ϕ plane. The solenoid has a length of 5.3 m with a radius of 2.5 meters positioned perpendicular to the beam axis. This solenoid produces a magnetic field of 2 T at the interaction point that is constant in the radial direction and decreases as the distance from the interaction point increases in the z direction.
The toroid magnet system produces a magnetic field for the muon spectrometer that deflects the particles in the η plane. There are two end cap toroids at the two ends enclosing the detector and a barrel toroid located symmetrically around the barrel region of the detector. Each toroid system consists of 8 independent coils arranged in the radial direction to the beam axis. The barrel toroid generates a magnetic field of 3.9 T while the end cap toroids produce a field of 4.1 T. Both magnets are cooled to 4.5 K using liquid helium and behave as superconducting magnets.

3.2.5. Trigger and data acquisition

The selection and recording of events for further analysis in the ATLAS experiment presents a challenge in itself. Each event information is approximately 1.3 Mb in size, and looking at the nominal operating frequency of the LHC at 40 MHz, the total information generated if all data were recorded live would amount to 50 Tb per second. A far more efficient and realistic process had to be implemented, which lead to the development of the trigger and data acquisition system employed by the detector.
The data acquisition system only selects and records events that it deems 'interesting' after certain triggers are satisfied by the events recorded. Three decision levels are implemented depending on the strategies implemented and the physical limitations of the information flow. There are currently two triggers in ATLAS for Run-2 data acquisition, the L1 and the High-Level Trigger (HLT). The L1 trigger only uses information from the muon system and parts of the calorimeters, but does not incorporate any tracking information from the inner detector. The triggering from L1 reduces the effective apparent beam crossing rate to around 40 kHz - 100 kHz. The information is then passed to the HLT, which performs speed optimised full event reconstruction.

4. Monte Carlo Simulated Samples

For any analysis, understanding the theoretical basis of the processes concerned is crucial to model the contributions of background in the search for the signal. To cross check if the simulations are correct, the simulated samples of the signal and background are compared to data in the control regions to see if there is any mismatch. To achieve this, Monte Carlo (MC) generators are exploited as they can model events where the full information about the kinematic properties of the final state objects are retained. Since simulating all events produced at a hadron collider analytically is impossible, a description of the different steps involved in simulating events in a MC is given below.

- Matrix Element Generators: To generate an event of a certain physical process, the matrix element generators calculate all the Feynman diagrams of the process up to a certain order. Current scope of matrix element generators are limited to LO (Leading Order) or NLO (Next to Leading Order) in QCD. The generators calculate a hard process with initial and final state radiation up to one additional emission and account for the hadronisation and underlying event. The end product is a calculation of all possible Feynman diagrams of the event with corresponding final state particles and decays according to the SM. All energy and momenta of the final state particles are calculated and are used in the parton showering afterwards as the hard processes set an energy scale Q^2 up to which the partons can branch into new particles.

- Parton Showering Algorithms: The parton showering algorithms give an approximate perturbative treatment of QCD dynamics at squared momentum transfer (q^2) scales greater than some infra-red cutoff range. This range is typically in the order of 1 GeV. At lower q^2, the interactions between particles become non-peturbative since low energy particles could theoretically go through infinite interactions. The most important part then, for a parton showering algorithm, is to correctly treat the infra-red region of the parton branching. The end result is an event that includes all the physics process, incorporating up to the infra-red cutoff value.

For the $t\bar{t}H, H \to b\bar{b}$ process in the single lepton channel, the main source of background is the $t\bar{t}$ + jets process. Additional background contributions come from the single top production, the production of W or Z bosons in association with jets (W/Z + jets), diboson production (WW, ZZ, WZ) as well as the associated production of a $t\bar{t}$ pair with a vector boson ($t\bar{t}$ + V where V is either a W or Z boson). Multi-jet events also contribute to the background via the misidentification of a jet or photon as electrons or via the contribution of a non-prompt electron. This misidentification, however, is small after requiring the tagging of two jets coming from the b-quarks as in this analysis.
This chapter discusses the simulation of signal and background samples used in this analysis. For all MC samples, the top quark mass is taken as $m_t = 172.5$ GeV, the Higgs mass is taken to be $m_H = 125$ GeV and $\sqrt{s} = 13$ TeV. Comparison of data to MC samples are discussed in detail in Chapter 6.

4.1. $t\bar{t}H$ Signal Modelling

The $t\bar{t}H$ signal process is modelled using the MADGRAPH 5_aMC@NLO [14] generator with next-to-leading-order (NLO) matrix elements. The modelling is inclusive in the Higgs boson decays and use the CT10ME PDF set. The showering is generated by HERWIG ++ [15] which uses the CTEQ6L1 PDF set. The $t\bar{t}H$ cross section and the Higgs branching fractions are taken from theoretical calculations in [16].

4.2. $t\bar{t}$+jets Modelling

The $t\bar{t}$ sample is generated using the POWHEG NLO generator [17] using the CT10 PDF set. Parton shower and hadronisation is modelled using PYTHIA 6.428 [18] with CTEQ6L1 PDF set and Perguia 2012 underlying-event tune. To regulate the high p_T emissions and the recoil of the $t\bar{t}$ system to the emission, the *hdamp* parameter, which controls the p_T of the first additional emission beyond the Born configuration, is set to the mass of the top quark. The bottom and charm hadron decays are modelled using EVTGEN v1.2.0. The events are normalised up to NNLO(Next to-Next to-Leading-Order) in QCD [19–21].
The $t\bar{t}$+jets samples are generated inclusively but the events are categorised depending on the flavour of partons in the jets that do not originate from the $t\bar{t}$ decay. The jets are constructed using the clustering algorithm *anti-k_t* [22] with $R = 0.4$, which clusters stable particles and excludes muons and neutrinos. The particle level jets are required to have $p_T > 15$ GeV and $|\eta| < 2.5$.
Events where at least one particle jet is matched to a b hadron is labelled as $t\bar{t} + b\bar{b}$ event.

Since the $t\bar{t}b\bar{b}$ is an irreducible background for the signal, a separate b-filtered sample using the same generator and showering algorithm was also used to study the $t\bar{t} + b\bar{b}$ events. Similarly, events that were not already classified as $t\bar{t} + b\bar{b}$ and where one of the jets originated from a c-quark were categorised as $t\bar{t} + c\bar{c}$ events. The $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ events are collectively known as $t\bar{t} + HF$ (Heavy Flavoured). The remaining events are classified as $t\bar{t}$+light-jet including events with no additional jets.

4.3. Single-Top Modelling

The single-top quark production in hadron colliders is governed by electro-weak interactions. The t-channel process, wherein a top quark is produced with an additional jet in the final state has the largest cross section. The production of a top quark in association with a W boson (Wt) has the second largest contribution followed by the s-channel single top production with an associated b-jet.
The s-channel and Wt samples are generated with POWHEG 2.0 using the CT10 PDF set. Since the Wt events have a similar final state as that of $t\bar{t}$, any overlap in the events are removed [23]. The t-channel single top production was generated using POWHEG v1 generator with a fixed four flavour CT10f4 PDF set. All single top event samples are interfaced with PYTHIA 6.428 using the CTEQ6L1 PDF set and Perugia 2012 underlying-event tune. The bottom and charm hadron decays, as for the $t\bar{t}$+jets case, are modelled using EVTGEN v1.2.0.
The single-top t- and s-channel samples are normalised to fixed order NLO calculations and the Wt samples are normalised to the approximate theoretical NNLO cross section [24–26].

4.4. W/Z+jets Modelling

The leptonic decay of a W boson produces an isolated lepton with large missing transverse energy. The additional jets can either originate from light quarks (u, d, s), gluons or HF (b- and c-) quarks. Since the production of HF quarks with the W boson is only a few percent of the total cross section, the W+jets contribution in background is suppressed in the analysis by requiring two b-tagged jets in the final state.
The Z boson can decay leptonically ($Z \to l^+l^-$) contributing to the isolated lepton background or hadronically contributing jets to the background and is suppressed by requiring two b-tagged jets in the final state.

Samples of W/Z+jets are generated using the SHERPA [27] 2.1.1 generator. Various samples with W/Z+HF production is also generated and the sum of $W/Z+b\bar{b}$ and $W/Z+c\bar{c}$ processes is adjusted to reproduce the relative rates of Z events with and without b-tags as observed in data. The CT10 PDF set is used in conjunction with the parton shower tuning present in the SHERPA framework. The matrix elements are calculated up to two partons at NLO and four partons in Leading-Order (LO) using COMIX [28] and OpenLoops matrix element generators. COMIX employs Berends-Giele type recursive relations to simplify multi-parton tree level amplitude calculations. Further, the W/Z+jets events are normalised to the NNLO cross sections [29].

4.5. Diboson Modelling

Diboson production (WW, ZZ, WZ) has a minor contribution to the total background due to its smaller cross section. Diboson samples are created using the SHERPA 2.1.1 generator using the CT10 PDF set in conjunction to the parton shower tuning present in the generator. The events are generated at up to one additional parton at NLO and three additional partons at LO. All samples are normalised to their respective NLO cross sections [29].

4.6. $t\bar{t}+V$ Modelling

The associated production of a vector boson (W/Z) with a top quark pair also has a minor contribution to the background due to its smaller cross section compared to the $t\bar{t}$ process. The final state topology of this channel depends on the decay of the W/Z decay, wherein the decay of the $Z \to b\bar{b}$ has the same topology as the signal.
The $t\bar{t}+V$ samples are generated using MADGRAPH 5 and the CTEQ6L1 PDF set. The events are interfaced with PYTHIA 8.1 using the A14 tune for showering. All samples are normalised to their respective NLO cross sections.

4.7. Common treatment of MC samples

All simulated events use PHOTOS 2.15 [30] to simulate photon radiation and TAUOLA 1.20 [31] to simulate τ decays. Both algorithms use the MC generator parameters to simulate their respective particles. All samples generated using HERWIG are also interfaced with JIMMY 4.31 [32] to simulate underlying events.
All simulated MC samples are processed through the same reconstruction software as data

and are processed through a detector geometry and response using either GEANT 4 [33] or fast simulation of calorimetric response. The events are further optimised so that energy resolutions, energy scales and object identification match those determined from data. Table 4.1 gives a summary of the generators used and their basic parameters. A full list of MC samples is presented in Appendix A

Sample	**Matrix Element Generator**	**PDF**	**Parton Shower**	**Normalisation**
$t\bar{t}H$	aMC@NLO	CT10ME	HERWIG ++	NLO
$t\bar{t}$+jets	POWHEG 2.0	CT10	PYTHIA 6.4.28	NNLO
W/Z+jets	SHERPA 2.1.1	CT10	SHERPA 2.1.1	NNLO
Single-top (s-channel, Wt)	POWHEG 2.0	CT10	PYTHIA 6.428	NLO, NNLO
Single-top (t-channel)	POWHEG v1	CT10f4	PYTHIA 6.428	NLO
Diboson	SHERPA 2.1.1	CT10	SHERPA 2.1.1	NLO
$t\bar{t}V$	MADGRAPH 5	CTEQ6L1	PYTHIA 8.1	NLO

Table 4.1.: Summary of MC generators used for signal and background simulated samples.

5. The $t\bar{t}H, H \to b\bar{b}$ process

The $t\bar{t}H$ process is one of the four main production mechanism of the Higgs boson. The search for this channel with the Higgs boson decaying into $b\bar{b}$ pair was carried out with the Run-1 data in the LHC where no excess of signal events above the expected background was found. This search also allowed to determine the ratio of the signal strength to the SM prediction to be $\mu = 1.5 \pm 1.1$ at $\sqrt{s} = 8$ TeV [34].
This chapter discusses the motivation and the procedure related to this search in the ATLAS experiment during the LHC Run-2. The first part of this chapter is dedicated to talking about the motivation and the event selection based on the ATLAS detector. Then, a discussion of the analysis strategy being developed for the Run-2 analysis is presented.

5.1. Motivation

The discovery of the Higgs particle with $m_H = 125$ GeV has opened up avenues to verify the SM predictions in Higgs physics. The focus of the current research has been in finding the precise characteristics of the boson, particularly through the measurements of its couplings. The search for this signal process provides a unique opportunity to test not only the predictions made in the SM, but also deviations from theory if any exist.
The discovery of the particle was driven by the search in the decay of the Higgs boson into $\gamma\gamma$ and a $t\bar{t}$ pair, and the measurement of the spin, mass and parity via its decay into vector bosons (WW, ZZ). Further, the vector boson fusion channel and the coupling of the Higgs to fermions via $H \to \tau\tau$ has also been observed [35].
The $t\bar{t}H$ process also provides the direct measurement for the Yukawa coupling to the top quark. As the $H \to t\bar{t}$ decay is not allowed kinematically since $m_H < 2\ m_t$, the $t\bar{t}H$ is the only process that allows for a direct measurement of $t\bar{t} \to H$ coupling during the production of the Higgs boson. Further, for a Higgs boson with mass 125 GeV, the decay to a pair of b quarks has the highest branching fraction. A search in the associated production with a top pair, with the Higgs decaying into $b\bar{b}$ will also provide direct probing of the Yukawa coupling to the b quark. This decay mode is not accessible to the dominant gluon fusion production process due to the overwhelming QCD multi-jet background. Thus, the

aim of the study presented in this thesis is the search for the $t\bar{t}H$ signal with the Higgs decaying into a b quark pair.

5.2. $t\bar{t}H$, $H \to b\bar{b}$ event topology

Since the top quark almost always decays into a W boson and a b quark, the final state topology of the $t\bar{t}H$ process depends on the decay of the $W^{\pm}$ boson. The W boson can decay either into a lepton and a neutrino or into a pair of quarks.
If both W bosons decay hadronically, the final state will contain 8 jets, 4 of which originate from b quarks. The branching ratio of the hadronic decay of a top quark pair is about 46%. This final state, coupled with the extremely low cross section for the production of the $t\bar{t}H$, is not ideal for the search of the signal due to the presence of an overwhelming background in a hadronic collider.
If both W bosons decay leptonically, then the final state contains 2 isolated leptons, missing transverse energy from the neutrinos and 4 jets originating from 4 b quarks. The dileptonic decay of the top quark pair has a branching ratio of about 9%.
Finally, one W boson can decay hadronically and the other decays leptonically to create the single lepton channel. The final state topology of this channel has one isolated lepton, missing transverse energy from the neutrinos and 6 jets, 4 of which originate from b quarks. The branching ratio of the single leptonic decay of the top quark pair is about 45%. This thesis focuses on the single lepton channel and the event topology is shown in Figure 5.1.

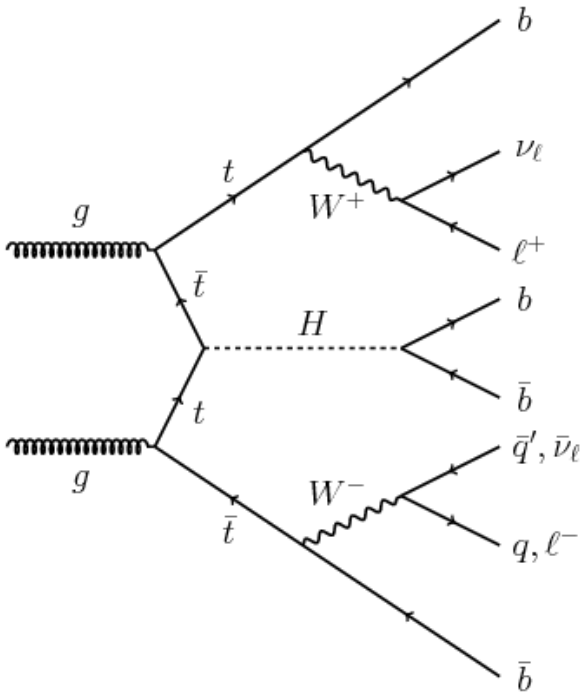


Figure 5.1.: Feynman diagram of the $t\bar{t}H$ event with the final state decay in single lepton channel.

An understanding of the event topology of the dominant background is essential to find an effective strategy for signal identification. The dominant background for the $t\bar{t}H$ with $H \to b\bar{b}$ process are $t\bar{t}$+jets events.

5.3. Object Definition

The events recorded by the ATLAS detector undergo offline reconstruction. The output received from all the detector components are combined to extract information about the tracks and the calorimetric clusters. This is the raw information obtained from the detector which is then used to reconstruct objects of the event. The reconstruction and identification of the $t\bar{t}H$ and $t\bar{t}$ processes then use these objects (leptons, missing transverse energy, jets and b-jets), for which a good performance of all the detector components is necessary. This chapter lists the definition of the various physics objects used in this analysis.

5.3.1. Leptons

The reconstruction of electrons and muons will be discussed in the following section. The reconstruction of tau leptons is not considered since they are not used in this analysis. The taus decay in the detector and are reconstructed from their decay products: electrons, muons or narrow hadronic jets.

Electrons

The electron candidates are reconstructed using energy deposits (clusters) in the electromagnetic calorimeters and the tracks reconstructed in the inner detector. They are required to have $p_T > 25$ GeV and $|\eta_{cluster}| < 2.47$ where $|\eta_{cluster}|$ is the pseudorapidity of the calorimeter clusters associated with the electron candidates. Any electron candidates in the calorimeter transition region of $1.37 < |\eta_{\text{cluster}}| < 1.52$ is excluded.
The electron identification uses an electron likelihood selection using outputs from the tracking detector and calorimeter outputs. The aim is to obtain isolated prompt leptons and discard electron candidates from background processes (electrons from photon conversion, jets faking electrons etc.). For this, three reference selections have been produced based on the identification and isolation working points: *loose*, *medium* and *tight*. The selections are changed by either introducing new variables or changing the cuts on the working points.

Muons

The muon candidates are reconstructed using the tracks seen in the muon spectrometer that are matched to the tracks in the inner detector. All track pairs (between the muon

spectrometer and inner detector) with a small ΔR[1] are considered and identified by a track re-fit to the hits on both detector systems.
Muons are further required to satisfy $p_T > 25$ GeV and $|\eta| < 2.5$. The muons are also required to be isolated to separate contributions from background processes. Similar to the electron selection, the muon selection also employs a likelihood function using the muon identification and isolation working points.

5.3.2. Jets

Jets are collimated sprays of hadrons produced during the hadronisation of quarks and gluons. The jet definition re-combines these spray of hadrons in order to obtain a physics object whose characteristics are as close as possible to the initial partons that created them. Jet reconstruction is achieved by analysing the energy deposits in the calorimeter cells and are called *topo clusters*. These topo-clusters are groups of calorimeter cells that are connected to each other topologically and are calibrated to achieve a significant signal above the noise. The anti-k_T algorithm [22] is used to combine the topo-clusters into jets with a distance parameter $R = 0.4$. It is a sequential recombination algorithm, where

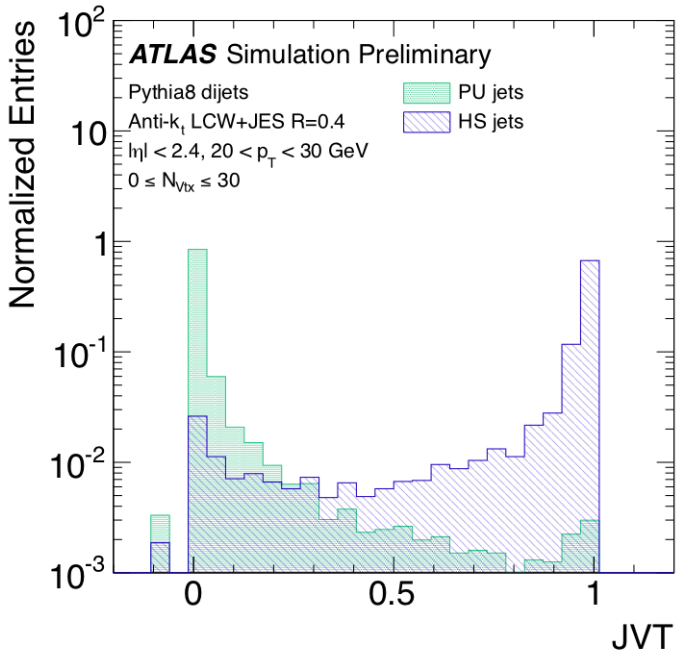


Figure 5.2.: Distribution of JVT for Hard-Scattered (HS) and Pileup (PU) jets [36].

the particles are clustered into jets one particle at a time using a two particle distance

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

parameter given by

$$d_{ij} = min(1/p_{T,i}^2, 1/p_{T,j}^2 \cdot \Delta R_{ij}^2/R^2)$$

where ΔR_{ij} is the angular distance between the particles i and j. This algorithm has been chosen for its infrared and collinear safety and also produces a fairly circular jet in the $\eta - \phi$ plane.

After energy calibrations, the jets are required to have $p_T > 25$ GeV and $\eta < 2.5$. To reduce the contribution of jets from pileup, a so called "*Jet Vertex Tagger (JVT)*" > 0.64 is applied to jets with $p_T < 50$ GeV and $|\eta| < 2.4$ [36]. The JVT algorithm is constructed using two pileup sensitive variables, corrJVF[2] and R_{p_T}[3] into a 2-dimensional likelihood. For each point in the R_{p_T} - corrJVF plane, the relative probability that the jet is a signal type is computed via the ratio of the number of hard-scattered jets divided by the number of hard-scattered plus pileup jets. Figure 5.2 shows the likelihood distribution for the hard scattered versus the pileup jets. A value of $JVT = -0.1$ is assigned to jets with no associated tracks. The choice of an optimal discriminant value minimizes the contribution from pileup jets. The jets selected using this algorithm result in hard-scatter jet efficiencies which are stable within 1%. This pileup stability reduces the dependence on jet reconstruction and other jet tagging techniques since there is no need to re-optimise the JVT cut values depending on pileup conditions.

During jet reconstruction, no distinction is made between identified electrons and jet energy distribution in the calorimeters. To avoid double counting of the electrons, any jet lying within ΔR of 0.2 of the selected electron, the single closest jet is discarded. Following this, any electrons which are within ΔR of 0.4 of the jet are removed. This ensures that any overlap between jets and electrons are accounted.

5.3.3. b-tagging

The identification of jets resulting from the fragmentation of b quarks plays an important role in any top quark physics and $H \to b\bar{b}$ analysis. Under ideal conditions, a 100% identification rate would itself be the best discriminator to identify a $t\bar{t}H$ signal from $t\bar{t}$ background. The relatively long lifetime ($\tau \approx 1.5$ ps) of a B-hadron produced in the hadronization of a b-quark provides a sufficiently large decay length that can be resolved in the detector as a secondary vertex as shown in Figure 5.3. A number of parameters can be exploited to reconstruct the b-jets. If a secondary vertex can be identified, the decay

[2] corrected Jet Vertex Fraction which includes the total number of vertex dependent average scalar sum p_T from pileup tracks associated to jets.

[3] scalar p_T sum of the tracks that are associated to the jet and originate from the hard-scatter vertex divided by the fully calibrated jet p_T.

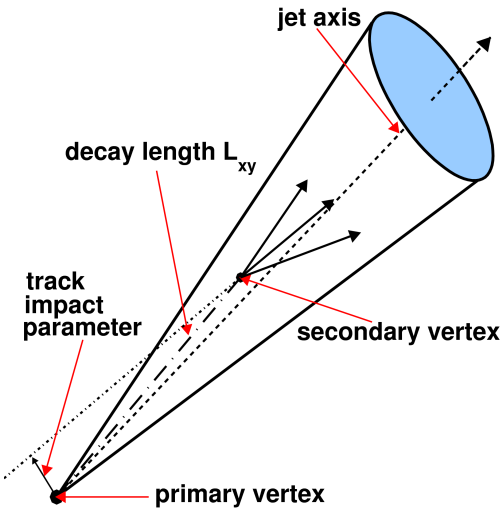


Figure 5.3.: Parameters of interest for the identification of jets from b-quark fragmentation.

length, impact parameters and the mass of the objects in the jet is used. The impact parameter is defined as the minimum distance between the axis of the secondary vertex to that of the primary vertex.

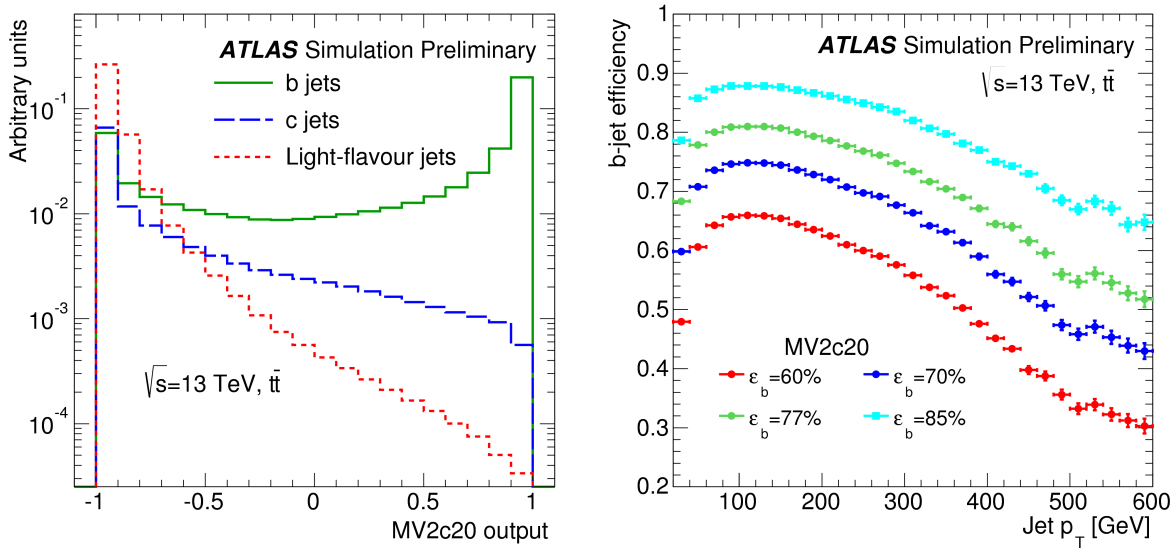
A number of b-tagging algorithms have been developed in ATLAS to exploit the above mentioned properties for the identification of b-jets. During Run 1, a multivariate analysis based b-tagging algorithm, called the MV1 [37] was developed that used the weights associated with various jet parameters to find an optimal b-jet tagging and light + cc-jet rejection discriminant. With the addition of the Insertible B-Layer (IBL) into the beam pipe in the first upgrade of the ATLAS detector for Run-2, the b-tagging algorithms have been revisited and the current algorithm used is called the MV2, which also takes into account the output information from the IBL.

The key point for any b-tagging algorithm is its capacity to correctly tag a jet coming from a b-quark and the efficiency with which they can discard jets coming from a c-quark or a light flavoured parton (u, d, s-quark and gluons).

The MV2 algorithm uses as input the output weights from the track selection, IP2D and IP3D [4], SV [5] and JetFitter [6] algorithms [38]. The input parameters are combined using a boosted decision tree (BDT) algorithm to discriminate the b-jets from its background. This analysis uses a version of the MV2 algorithm called the MV2c20 where the output

[4]Impact Parameter based Algorithms
[5]Secondary vertex finding algorithm
[6]Decay Chain Multi-Vertex Algorithm

is shown in Figure 5.4(a) for b-, c- and light jets. A cut is done at different working points along the MV2c20 weights that has the highest optimal separation between b-tagged jets from its background. This analysis uses the cut based on 77% b-tagging efficiency working point. The training sample used b-jets as signal and a combination of



(a) Output of the MV2c20 algorithm for jets in $t\bar{t}$ events.

(b) The b-tag efficiency for different working points as a function of jet p_t.

Figure 5.4.: (a)The MV2c20 output for b, c and light-flavour jets in $t\bar{t}$ events. The output is a discriminant from -1 to 1 and the cuts are applied for selection on this scale. (b)The b-jet efficiency as a function of jet p_T for the four operating points of the MV2c20 tagger used in the $t\bar{t}H$ signal search [38].

80% light jets and 20% c-jets as background. Figure 5.4(b) shows the b-tagging efficiency for different working points. For lower b-tagging efficiency, there is a loss in statistics which plays a significant role in the analysis as the $t\bar{t}H$ processes are very rare compared to $t\bar{t}$ background. The working point for the b-tagging algorithm used in this analysis is at 77% compared to 70% for MV1 during Run-1. Optimisation studies are currently ongoing on the yields and efficiency between various working points.

5.4. Event Pre-Selection

The typical final state of $t\bar{t}H$ with $H \to b\bar{b}$ and $t\bar{t}$ single lepton channel consists of one isolated lepton, missing transverse energy and around 6 jets, 4 of which originate from a b quark. Of the 4 b quarks, 2 arise from the $H \to b\bar{b}$ decay and the other 2 from the 2 W boson decays. From the signal topology, the best sensitivity to signal arises when

events with 6 jets and 4 of them b-tagged are selected. Events with lower jet and b-tag multiplicity are included for higher signal acceptance as well as background modelling. This analysis uses data collected by the ATLAS detector at $\sqrt{s} = 13$ TeV during 2015 with an integrated luminosity of 3.2 fb^{-1}. A detailed look at the selection criteria is outlined below.

- **Lepton selection:** Events are required to have exactly one isolated lepton in the final state with $p_T > 25$ GeV.
- **Jet selection:** Events are required to have at least 4 jets in the final state with $p_T > 25$ GeV and $|\eta| < 2.5$. These criteria will be used to obtain more $t\bar{t}$ and $t\bar{t}H$ events that can be used for better understanding of signal and background events.
- **b-tagging requirement:** Events are required to have at least 2 jets that are b-tagged using the MV2c20 algorithm at 77% working point. This requirement helps to considerably reduce non $t\bar{t}$ background contribution.

No selection in the E_T^{miss} and W transverse mass is applied, which is usually done to reduce the non-$t\bar{t}$ background. This reduction is already achieved by requiring at least 2 b-tagged jets.
After the pre-selection, the events are categorised further into 9 independent regions. These regions are separated based on jet and b-tag multiplicities. With consideration to the signal and background final state topologies, these regions are defined by the jet multiplicities given by $4j$, $5j$ and $\geq 6j$. Similarly, the b-jet multiplicities are given by $2b$, $3b$ and $\geq 4b$. The final state topology of the signal is expected to be found mostly in the $\geq 6j$, $\geq 4b$ region.

5.5. Analysis Strategy

With the categorisation of events into regions, the final aim is to adopt a simultaneous template fit method to all the 9 regions so as to incorporate the systematic uncertainties directly in the fit method. This method has the advantage of obtaining a better statistical combination improvement on the background normalisation and to reduce the effect of systematic uncertainties in the analysis.
Figure 5.5 shows the S/B and $S/\sqrt{B}$ for all regions. The three regions highlighted in red have the highest sensitivity to signal. Due to kinematic cuts and limited b-tagging efficiency, we expect the signal to migrate to regions of lower jet and b-tag multiplicities. Thus, additional information on the signal can be obtained from two other regions: ($\geq$

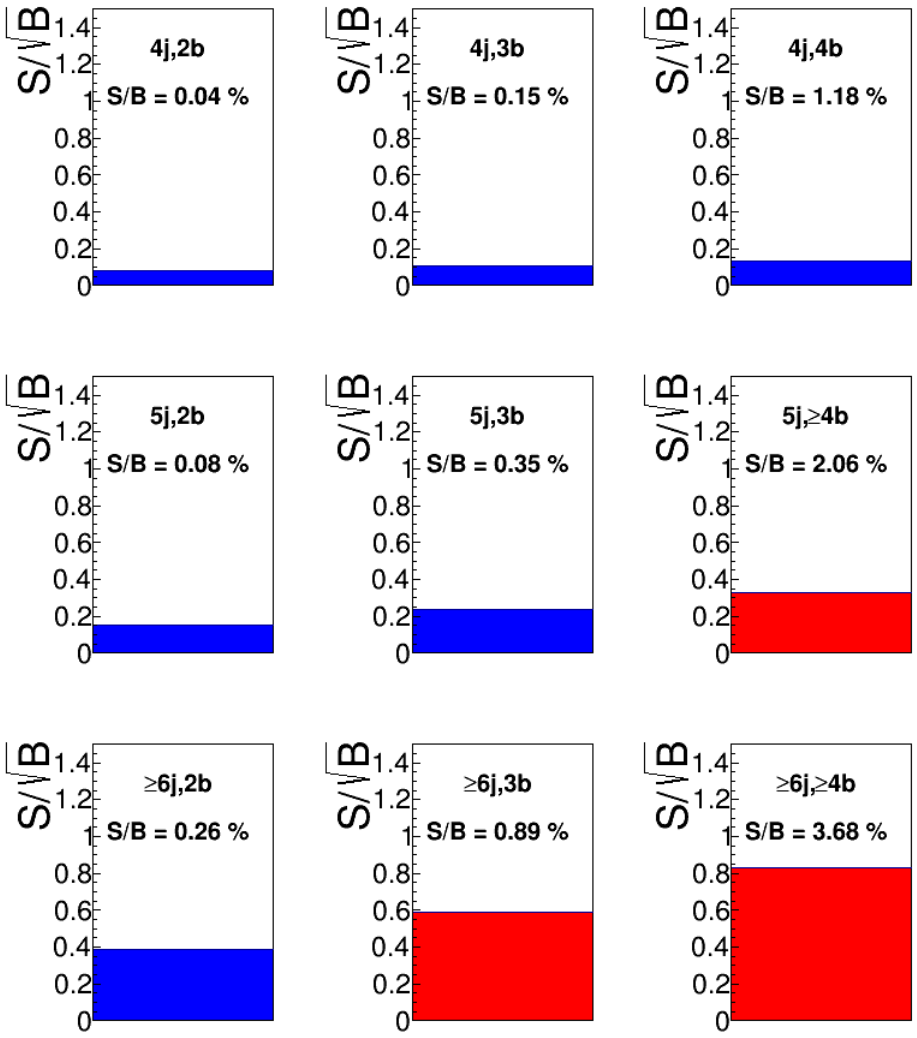


Figure 5.5.: S/B and $S/\sqrt{B}$ for all regions. Regions marked in blue are the background dominated control regions whereas the ones in red are the signal regions. The S/B values are multiplied by 100 and given by %.

$6j, 3b$) and ($5j, \geq 4b$).

For this analysis, any region with an $S/B > 0.8\%$ and $S/\sqrt{B} > 0.2$ is termed as the *signal-rich* region and the others *signal-depleted* regions. The signal depleted regions are used to understand the background contributions. All regions are used in the combined fitting procedure at the end.

The ($\geq 6j, \geq 4b$) region has by far the highest sensitivity among all the regions as was expected. The S/B value of around 0.037 is however very low and thus, a standard cut based approach to signal detection is impossible. Thus, following the Run-1 approach, the use of multivariate analysis, particularly the use of a Neural Network (NN) is implemented for the signal-rich regions to discriminate between signal and background. This report

will present the findings on the NN input variables for the ($\geq 6j, \geq 4b$) region.

5.6. Multivariate Analysis

In high energy particle physics analysis using data from hadron colliders, the detection of $t\bar{t}H$ events from the overwhelming background processes is very complex. With colliders aiming for higher center-of-mass energies, the search further rendered impossible by applying a cut-and-count method based on kinematic properties alone. This led to the development of algorithms that combined various specific kinematic properties of a particular signal, and use a learning algorithm that could be trained to distinguish the signal from background. The 1980s saw a revolution in the advancement of these types of algorithms that one can train, while making sure that this training was not biased and apply the specific weights associated with a typical signal event to identify it in data.
Multi-Variate Analysis (MVA) in high energy physics has seen a tremendous growth since the 1980s' when the complexity of the analysis made the use of the traditional kinematic cut based approach obsolete. The use of MVA shows a marked robustness when no single variable exhibits a clear separation between signal and background. They can be used to extract information about an event, combining information from different input variables to one output discriminant by exploiting the correlations between the said variables. The aim then is to find a list of input variables which allow to discriminate between the signal and background events.
The $t\bar{t}H$ signal, with an extremely low S/B ratio in the signal-rich regions is an ideal place to implement MVA due to the large number of physics objects in its final state and a complex topology. Furthermore, the MVA method can combine different properties of an event so as to not require full event reconstruction which saves time and resources. The use of a Neural Network (NN) method is preferred in this analysis compared to other MVA methods such as Boosted Decision Tree (BDT) particularly because of its robustness against over-training and other features that will be described in the next section.

5.7. Neural Networks

Neural Networks are inspired by the way neurones work in the human brain. Taking in various seemingly random inputs that are connected to a tangible output. Intuitively, one can infer a condition where a neuron 'fires' if the information received from previous neurons exceed a certain threshold. This fundamental idea can be expressed as a function

given by

$$x_j^n = g\{\Sigma_k(w_{jk}^n \cdot x_k^{n-1}) + \mu_j^n\} \tag{5.1}$$

where $g(t)$ is a sigmoid function[7] and μ_j^n defines the threshold [39]. The output of node j in layer n depends on the weighted sum of all nodes in layer $n-1$. Training the network aims at minimising the *loss function* by iteratively readjusting the weights (w_{jk}^n) such that the deviation of the network output is minimised compared to the desired output.
NN is preferred in this analysis because the algorithm can better learn and implement the correlations between the input variables, apply better weights to variables that have a higher separation discriminant between signal and background and can also work if all input nodes are not completely assigned in all instances. This makes it a robust and versatile MVA tool.

5.7.1. NeuroBayes Algorithm

NeuroBayes [40] is a commercially available three-layered feed-forward NN package [39]. Figure 5.6 is a schematic representation of the NeuroBayes algorithm [41].

The input layer contains a separate node for each input variable. In the hidden layer, information from the previous nodes in the input are multiplied by various weights and then summed together. Each output of the hidden layer node is a specific function of the summed weights, called the neural response function. The output is then a mapping of these functions for the n input variables. The output of the previous node is connected to the input in the successive node, so the information can only propagate forward among the layers. It is thus termed as a 'feed-forward' NN. The complexity of the NN can be increased by either adding more input nodes or more hidden layers. For this analysis, only one hidden layer is sufficient. The output of this NN is then uniquely defined by the total number of input nodes, the variables in the input node and the nodes in the hidden layer.
NeuroBayes has three distinctive features: an automated variable preprocessing, ranking of input variables based on their signal-to-background discrimination and a Bayesian regularisation for the training procedure.
Preprocessing is applied on input variables before the correlations are calculated between them with or without the knowledge of the target information. The preprocessing is given by values ij. $i = 1$ de-correlates the input variables and normalises them and $j = 4$ and $j = 9$ are used for continuous and discrete input variables respectively.
The ranking of input variables is done using the correlation matrix between all input

[7] sigmoid function: $f(t) = \frac{1}{1+e^{-t}}$

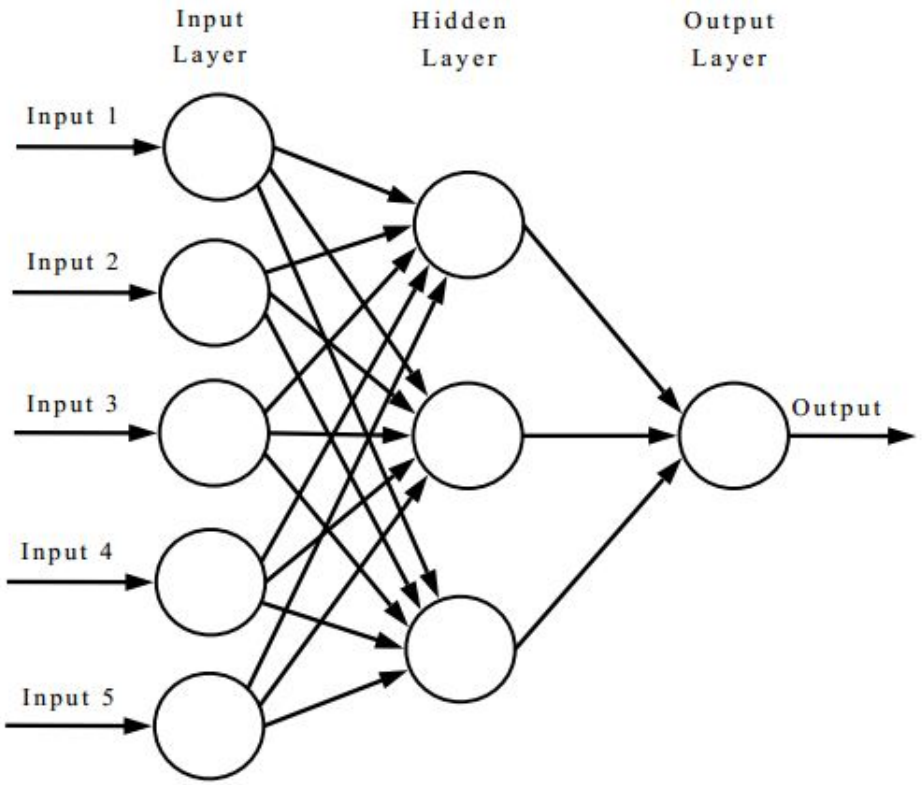


Figure 5.6.: NN architecture for NeuroBayes algorithm for a single discriminating output. The information passes between different layers and is thus called a *feed-forward* architecture.

variables and their correlations to the target after preprocessing. After the correlation matrix is computed, one variable is removed at a time and the correlation to target computed again. They are then sorted based on the loss of correlation to target due to their removal. The least significant variable, i.e. one causing the least loss of information is removed and this process is repeated until $N - 1$ variables remain. This iteration is done for the second least significant variable until only one variable remains, the one with the highest significance. Bayesian regularisation is used to limit overtraining of samples. Statistically insignificant network connections and nodes are removed in this stage to ensure that the network learns real features of data.
NN can be used as a classifier or a shape function. For the purpose of this analysis, since the requirement is a discrimination between signal and background, it is used as a classifier. As a classifier, after training, the network is given weights where signal like events are assigned a value close to 1 and background-like events to -1.

6. Analysis and Results

The aim of this analysis is to find a list of kinematic variables that have the highest discrimination power between signal and background for the ($\geq 6j, \geq 4b$) region in the $t\bar{t}H, H \rightarrow b\bar{b}$ single lepton search. In order to train the network, a set of MC samples has been used based on their contribution as background to the signal. The processes considered for the background include: $t\bar{t}$, single-top, W+jets, Z+jets, $t\bar{t}V$ and diboson production. Furthermore, a parton based heavy-flavour (HF) classification is applied to investigate individual contributions from $t\bar{t}$+light-jets, $t\bar{t}$+$c\bar{c}$ jets and $t\bar{t}$+$b\bar{b}$ jets. The separation of the $t\bar{t}$ decay into light and HF jets aims to study systematics. The $t\bar{t}b\bar{b}$ is an irreducible background for the $t\bar{t}H$ event due to the same final state topology. The event yields for the background, $t\bar{t}H$ signal and data for the ($\geq 6j, \geq 4b$) region is presented in Table 6.1. The event yields for all the nine analysis regions defined is presented in Appendix B. The total uncertainties are reported that include statistical uncertainties before the fitting procedure.

Figure 6.1 presents the relative contribution of different processes to the background.

Samples	**Event yield**
$t\bar{t}$+light jets	90.8 ± 3.0
$t\bar{t} + c\bar{c}$	123.7 ± 2.91
$t\bar{t} + b\bar{b}$	324.0 ± 1.41
W+jets	17.4 ± 2.23
Z+jets	5.1 ± 2.12
Single Top	20.3 ± 0.72
$t\bar{t}V$	6.8 ± 0.05
Diboson	2.3 ± 0.35
Total Background	608.8 ± 5.45
$t\bar{t}H$	18.5 ± 0.20
Data	852.0 ± 0.0

Table 6.1.: Pre-fit event yields for all background processes, $t\bar{t}H$ and data for the (≥6j, ≥4b) analysis region.

For regions of higher b-jet multiplicity, the contribution of $t\bar{t} + b\bar{b}$ increases due to the

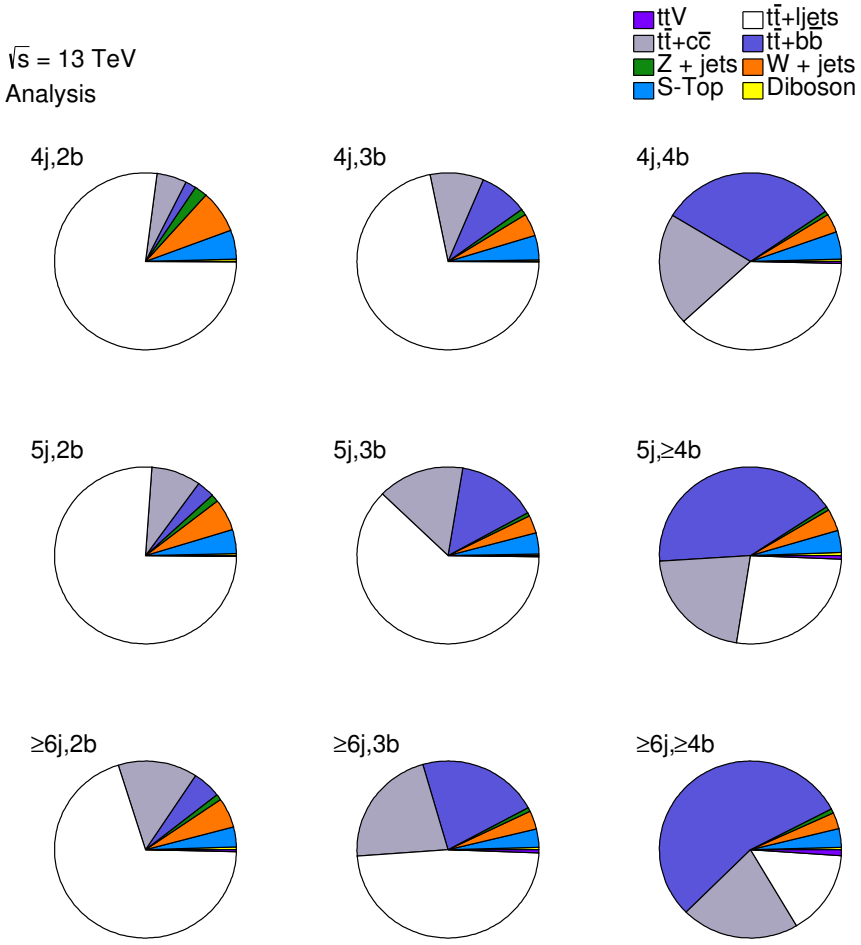


Figure 6.1.: Contribution of different processes to the background in all analysis regions for $\sqrt{s} = 13$ TeV.

higher inclusion of $t\bar{t}b\bar{b}$ samples in the regions. In the $(\geq 6j, \geq 4b)$ region, this is the most dominant background. Similarly, for lower b-jet multiplicity regions, $t\bar{t}$+light-jet are the dominant background. The control regions are used to understand $t\bar{t}$ background due to the large background contribution whereas the signal rich regions are used for training the NN.

6.1. Discriminating Variables

In the $(\geq 6j, \geq 4b)$ region, due to a variety of physics objects in the final state and the nature of the final state topology of signal and background, a variety of discriminating

variables is considered. They can be broadly classified into 4 distinct groups:

- *Object kinematics* : p_T and η of each jet and lepton in the event.
- *Global event variables* : scalar sum of p_T of all jets in the event (H_T^{had}), number of jets and b-tagged jets, number of jets above a given p_T threshold (40, 60 and 80 GeV) etc.
- *Event shape variables* : variables defined using the combinations of eigenvalues of the linear momentum tensor (circularity, sphericity, aplanarity), centrality and Fox-Wolfram moments. Event shape variables, as the name suggests, exhibit a unique shape of the distribution depending on the physics process. For example, centrality is defined as the ratio of the total scalar sum of the p_T of all the jets to that of the energy of the jets and might be different for $t\bar{t}$ and $t\bar{t}H$. The $t\bar{t}H$ signal is expected to be more energetic and central in the detector compared to the $t\bar{t}$ as a higher $\hat{s}$ is needed to produce a $t\bar{t}$ pair in association with a Higgs boson.
- *Object pair variables* : kinematic properties of objects in pairs (p_T, invariant mass) and certain separation between them (ΔR) are combined to look at the average effects ($\_avg$), objects with maximum p_T ($\_MaxPt$), largest invariant mass ($\_MaxM$) and minimum separation between objects ($\_Min\Delta R$) of various objects in the event.

All global and event shape variables are also defined exclusively for all jets, only b-jets and all untagged jets. Since the analysis region considers events with 6 or more jets with at least 4 of them b-tagged the remaining jets that are not tagged as b-jets are ordered based on their p_T. Object pair variables defined using jets are further classified into b-jet pair (b,b), jet pair (j,j), untagged jet pair (u,u) and jet and b-jet pair (b,j).
Variables displaying a poor agreement between data and MC in the control regions are discarded. A further look into the separation between signal and background for each variable is also considered for variable selection. At the end, more than 120 discriminating variables are implemented and studied.

6.2. Neural Network training

The training of Neural Networks is performed for the ($\geq$6j, $\geq$4b) region. The NN is trained by using the $t\bar{t}H$ sample as signal and $t\bar{t}$+light jets, $t\bar{t} + c\bar{c}$ and $t\bar{t} + b\bar{b}$ as backgrounds. Initially, all variables considered are provided to the NN for training. Table 6.2

lists the details of all the options provided for the configuration of the NeuroBayes program.
The parameters *learning speed factor* and *max. learning speed* control the rate at which weights given to each node evolve during the training procedure. The *Bayesian* regularisation procedure divides the weights into three different classes: one class for the bias node in the input layer, one class for all input nodes except the bias node, and finally a class for all weights from the hidden layer to the output layer. The number of iterations represents the number of times the training is evaluated for the set of events in order to minimize the loss function.

nodes in hidden layer	$n_{variables}$+2
update weights interval	50
learning speed factor	1
max. learning speed	100
regularisation	Bayesian
loss function	Entropy
number of iterations	100
training algorithm	BFGS

Table 6.2.: Details of NeuroBayes settings for NN analysis in (≥6j, ≥4b) region.

In order to obtain a final list of the most effective MVA variables, the NN is run by considering all events in the samples. Initially the NN lists all the input variables with their respective rankings. It also provides variables that have no contribution to the discrimination between signal and background, and these variables are excluded in successive iterations. Following the first iteration, variables are dropped based on three criteria: the overall ranking of the variables, the expected separation between signal and background in the (≥6j, ≥4b) region and the data-MC agreement for each variable in control regions. Table 6.3 lists the separation values between $t\bar{t}$ and $t\bar{t}H$ for the three signal regions: (5j, ≥4b), (≥6j, 3b) and (≥6j, ≥4b). The separation is a measure of the discriminating power of each variable between signal and background events and is defined as:

$$separation = \frac{1}{2} \cdot \sum_{i}^{bin} \frac{(N_i^{signal} - N_i^{background})^2}{(N_i^{signal} + N_i^{background})} \tag{6.1}$$

where N_i^{signal} and $N_i^{background}$ are the total events in each bin after the histograms have been normalised to unity. The average ΔR between two b-jets in a event has the highest separation power between a $t\bar{t}H, H \to b\bar{b}$ and $t\bar{t}$ event. This can be understood for the

Variable	(5j, ≥4b)	(≥6j, 3b)	(≥6j, ≥4b)
ΔR_{bb}^{avg}	**5.77 %**	**3.04 %**	**6.38 %**
$M_{bb}^{\Delta R,min}$	2.63 %	0.40 %	1.97 %
N_{30}^{Higgs}	3.35 %	0.31 %	3.72 %
$p_T^{5^{th}\ jet}$	3.42 %	3.27 %	2.45 %
$M_{uu}^{p_T,max}$		0.51 %	0.46 %
Centrality	3.55 %	2.16 %	3.17 %
$M_{bj}^{p_T,max}$	2.38 %	0.75 %	1.38 %
H_{bjets}^{1}	0.542 %	0.59 %	0.8 %
$\Delta R_{lep-bjet_3}$	2.15 %	0.86 %	1.82 %
$\Delta R_{uu}^{Max\ M}$		0.58 %	0.31 %
$M_{bj}^{\Delta R,\ min}$	3.15 %	0.54 %	0.87 %
$p_{T,\ uu}^{M,\ max}$		0.24 %	0.26 %
$p_{T,\ jj}^{avg}$	0.5 %	0.17 %	0.14 %

Table 6.3.: Separation values for the final set of variables obtained from NN for (5j, ≥4b), (≥6j, 3b) and (≥6j, ≥4b) regions, respectively. The variables are presented in the order of the ranking obtained from NN for the (≥6j, ≥4b) region. Since the separation value is too small, the values obtained are multiplied by 100 and presented as a %. The highest separation value for each region is presented in bold.

(≥6j, ≥4b) region because the $t\bar{t}H$ event has an extra pair of b-quarks that originate from the decay of a Higgs boson whereas the extra pair of b-jets for the $t\bar{t}$ event originate from gluon splitting. Thus the average ΔR for a b-jet pair will be markedly different.

Table 6.3 also presents separation values for all three signal regions for completeness. One has to note that the list of variables and the rankings presented is based only on the (≥6j, ≥4b) region and a separate NN training has to be applied to each of the other two regions. Thus, for the (5j,≥4b) and (≥6j, 3b) regions, the list of variables chosen and their rankings will be different than for the current analysis.

The data-MC comparison plots for the control regions are presented in Figures 6.8 to 6.13. The comparison is done between data and the MC expectations. Contributions from multijet events are missing in this analysis, but is expected to be negligible due to the requirement of b-jets in the events. The labels denote the samples used and their respective yields in the given region. There is a data-MC mismatch of around 10%, and this will be discussed in detail in Chapter 7.

Figure 6.2 represents the correlation to target for an increasing number of input variables

in the NN training. The correlation is calculated using the Frico-Gini Index which is given by the equation:

$$Frico - Gini\ Index = \left(\int ROC - \frac{1}{2} \right) \times 2 \tag{6.2}$$

The correlation increases sharply and then exhibits a plateau shape, where the addition of any extra variables does not appreciably change the correlation to output. Thus a total of 13 variables are chosen to train the NN, as the addition of any extra variables does not appreciably change the correlation to output. These variables with their definitions and rankings in the NN output is given in Table 6.4. Figure 6.3 presents the shapes of

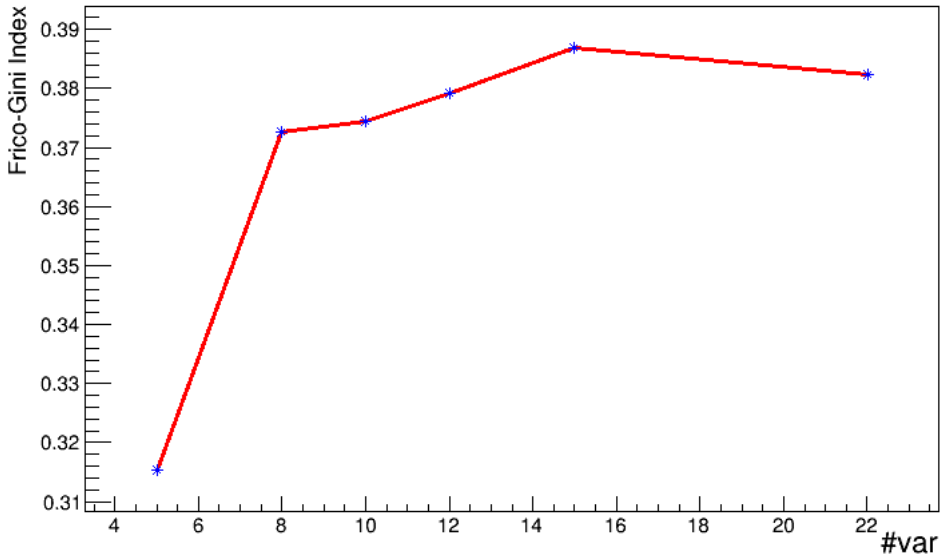


Figure 6.2.: Correlation to target as a function of number of variables used in the NN training for (≥6j, ≥4b) region.

the chosen discriminating variable in the (≥6j, ≥4b) region and the separation calculated using equation 6.1. The variable with the highest separation is also the highest ranked in the NN output. Comparison plots between data and total MC is also considered for the choice of the discriminating variables. A correlation matrix is generated to check for the dependency of variables amongst each other and is shown in Figure 6.4. Each variable considered is ensured to have less than 40% correlation between all other variables.

After obtaining the discriminating variable, a cross-validation test is conducted to check for consistencies in the choice of the training and validation samples. The sample is divided into *odd* and *even* samples based on the event number. The network is trained using the even sample and validated on the odd samples. This process is repeated again, by interchanging the training and validation samples and the two distributions are plotted

Variable	Rank	Definition
ΔR_{bb}^{avg}	1	Average ΔR between two b-tagged jets.
$M_{bb}^{\Delta R,min}$	2	Mass of two b-tagged jets with minimum ΔR separation between them.
N_{30}^{Higgs}	3	Number of Higgs boson candidates with their mass within 30 GeV of the Higgs boson mass.
$p_T^{5^{th}\ jet}$	4	p_T of the 5^{th} jet.
$M_{uu}^{p_T,max}$	5	Invariant mass of two untagged jets with maximum p_T.
Centrality	6	Centrality (ratio of scalar sum p_T and the energy of all objects) in the event.
$M_{bj}^{p_T,max}$	7	Invariant mass of a jet and b-jet pair, each with maximum p_T.
H_{bjets}^{1}	8	Second Wolfram moment of all the b-jets in each event.
$\Delta R_{lep-bjet_3}$	9	ΔR between the lepton and the 3^{rd} b-jet ordered by jet p_T.
$\Delta R_{uu}^{Max\ M}$	10	ΔR between two untagged jets each with maximum invariant mass.
$M_{bj}^{\Delta R,\ min}$	11	Invariant mass of a jet-b-jet pair with minimum ΔR between them.
$p_{T,\ uu}^{M,\ max}$	12	p_T of an untagged jet pair each with maximum invariant mass.
$p_{T,\ jj}^{avg}$	13	Average p_T of a jet pair in an event.

Table 6.4.: Summary of variables used, the NN ranking and variable definitions for the ($\geq$6j, $\geq$4b) analysis region.

to check for consistency. Figure 6.5(b) represents the cross validation applied in this analysis. There is a clear agreement between the odd-even and even-odd training using the $t\bar{t}H$ and $t\bar{t}$ samples.

Furthermore, a test to check for overtraining of the network is performed. In this test, the training of the network is done for the $t\bar{t}H$ and $t\bar{t}$ sample. This training is then checked on the same sample. Any deviation in the overlapped plots is a clear indication of overtraining. Figure 6.5(a) shows the overtraining test performed in this analysis. This is a clear indication that the network training is free of bias since the overlap of the testing and training conducted within each sample is nicely matched.

One can observe that in Figure 6.5(b), the total separation between signal and background using the NN is about 11%. Referring back to Table 6.3, the highest value for separation between any single variable (ΔR_{bb}^{avg}) is around 6.4%. This represents the power of using MVA techniques for signal-background separation, when the signal is a rare event.

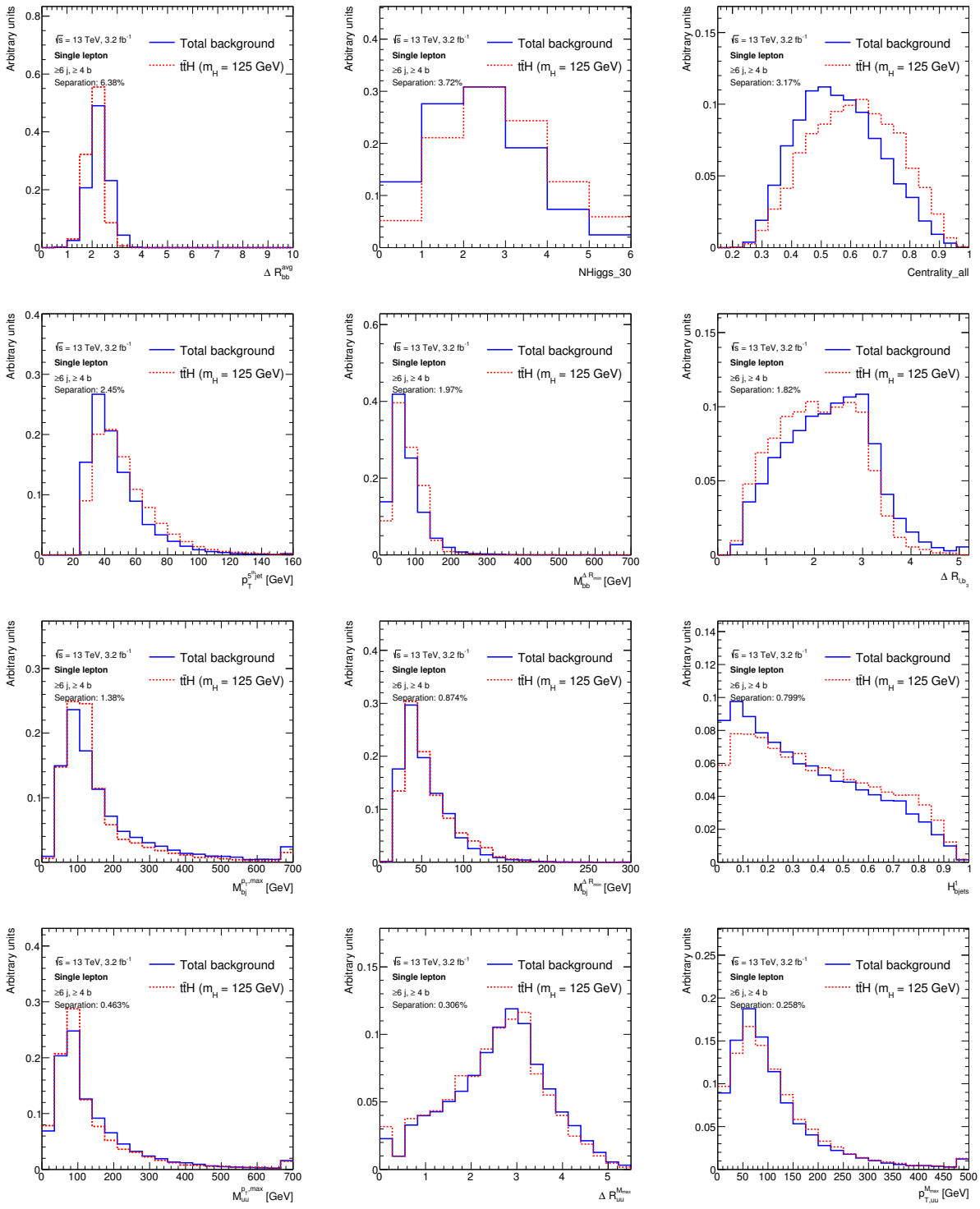


Figure 6.3.: Separation plots for the highest discriminating variables in the (≥6j, ≥4b) region obtained from the ranking of the NN output.

Figure 6.6 shows the ROC curves for the training sample. The validation between even-odd and odd-even training samples agree with the total events in the training. This

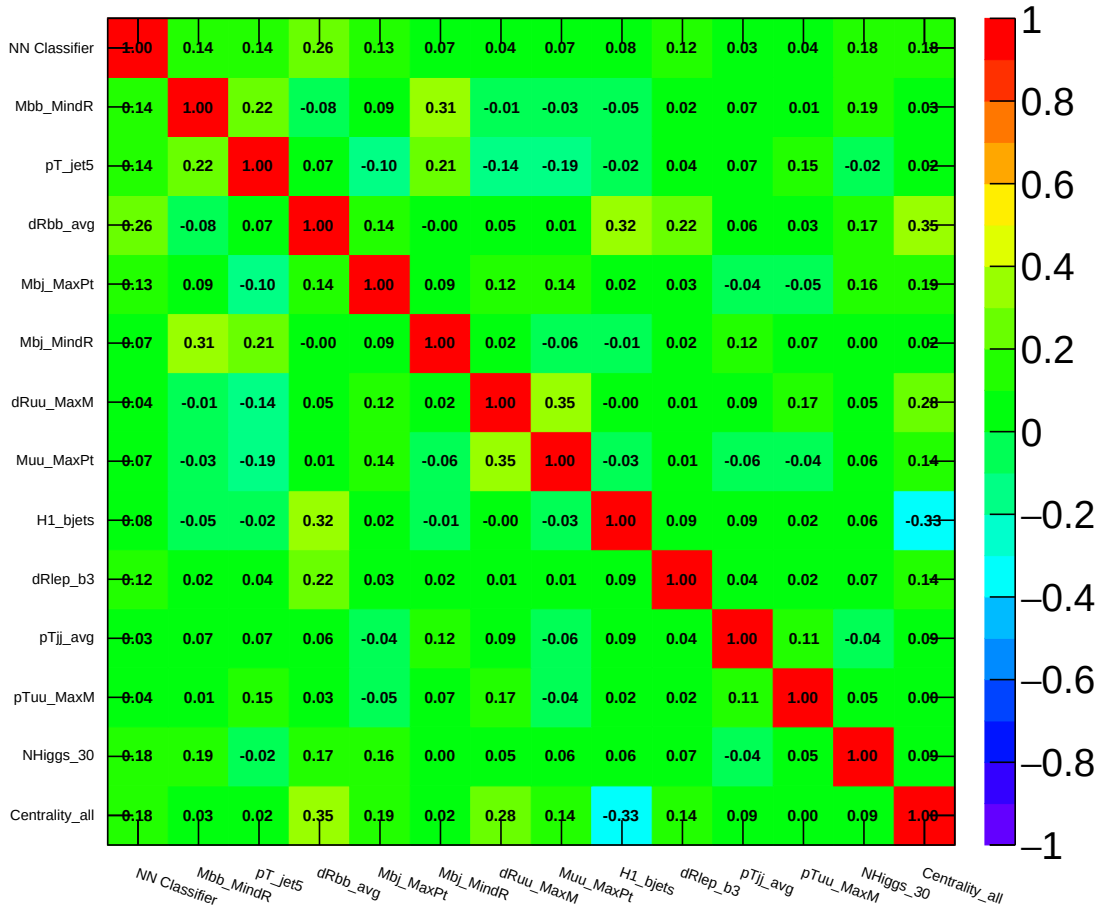
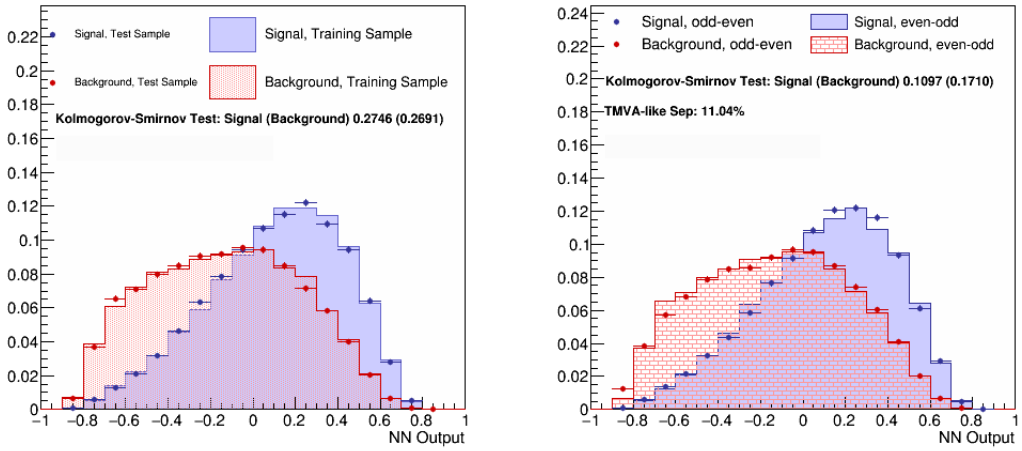


Figure 6.4.: Correlation matrix for the variables obtained from NN in the ($\geq$6j, $\geq$4b) region. This matrix is used to ascertain that the input variables are not highly correlated to each other, which would affect the weights obtained for discrimination and ranking in the training.

provides further proof that overtraining has been avoided in the training process.

The weights from the training of the NN using the 13 variables are used to obtain the NN discriminant variable. The training weights are derived from the $t\bar{t}H$ and $t\bar{t}$ MC samples used as signal and background, respectively. Figure 6.7 shows the new variable created from the NN training that has a separation value of 13.2%.



(a) Overtraining test for NN training.

(b) Cross-validation plot for even-odd and odd-even training.

Figure 6.5.: Figure 6.5(a) represents the agreement between the test and training sample for signal and background. The agreement between the curves and the data points show that overtraining is avoided in the analysis. Figure 6.5(b) shows the crossvalidation between odd and even numbered samples.

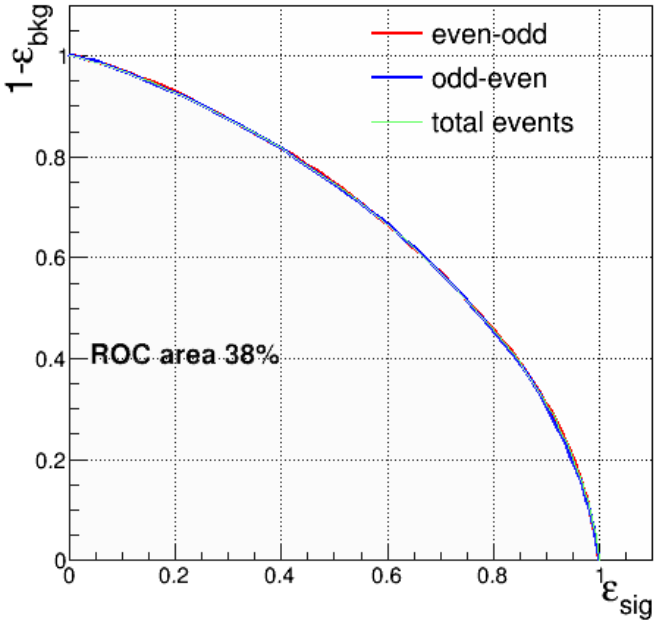


Figure 6.6.: Twofold validation for even-odd and odd-even testing and training done on $t\bar{t}H$ and $t\bar{t}$ MC samples.The function is plotted against the acceptance vs. the rejection and overlaid with the training done in the total samples.

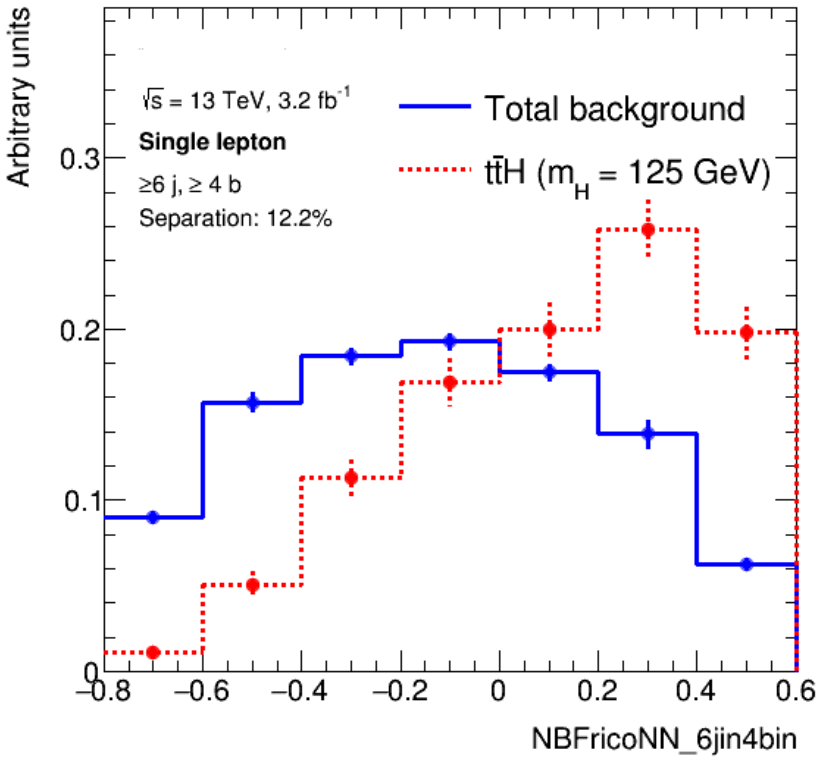


Figure 6.7.: The NN output discriminating variable obtained from the training of the network. The separation between signal and background is clearly seen and the separation value is highest among all variables considered.

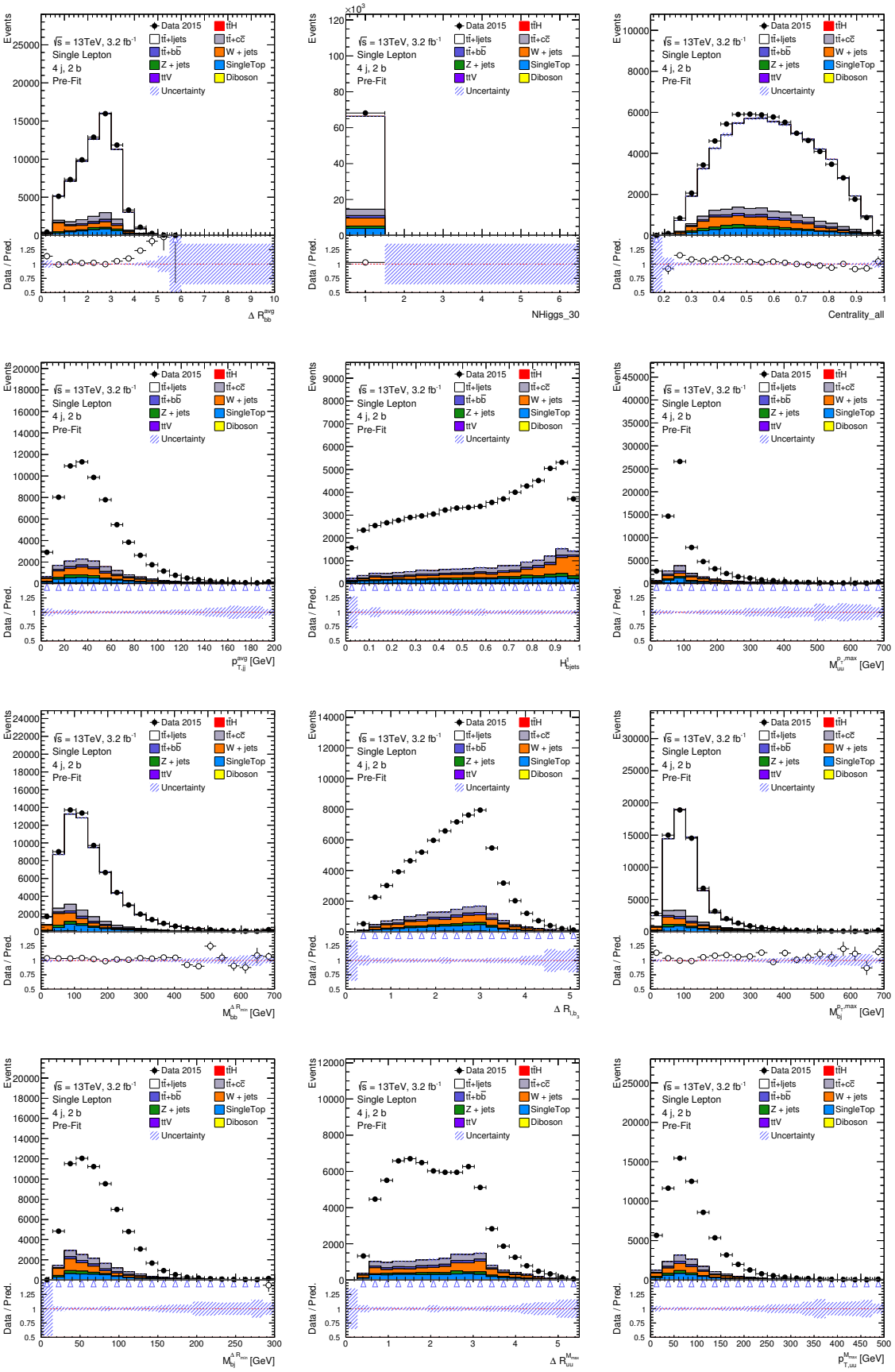


Figure 6.8.: Data-MC comparison for the discriminating variables used in the NN training in the exclusive (4j, 2b) control region.

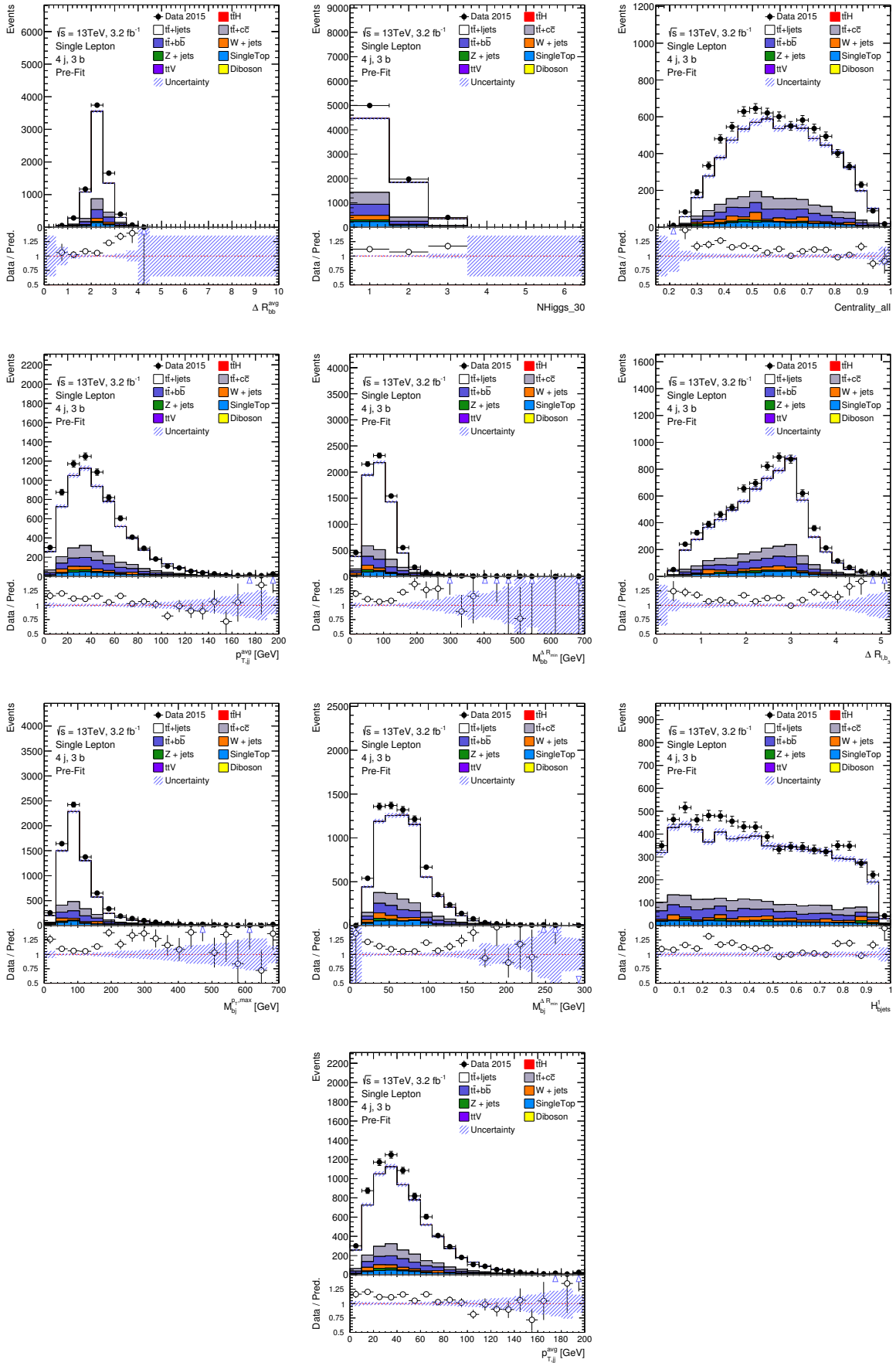


Figure 6.9.: Data-MC comparison for the discriminating variables used in the NN training in the exclusive (4j, 3b) control region.

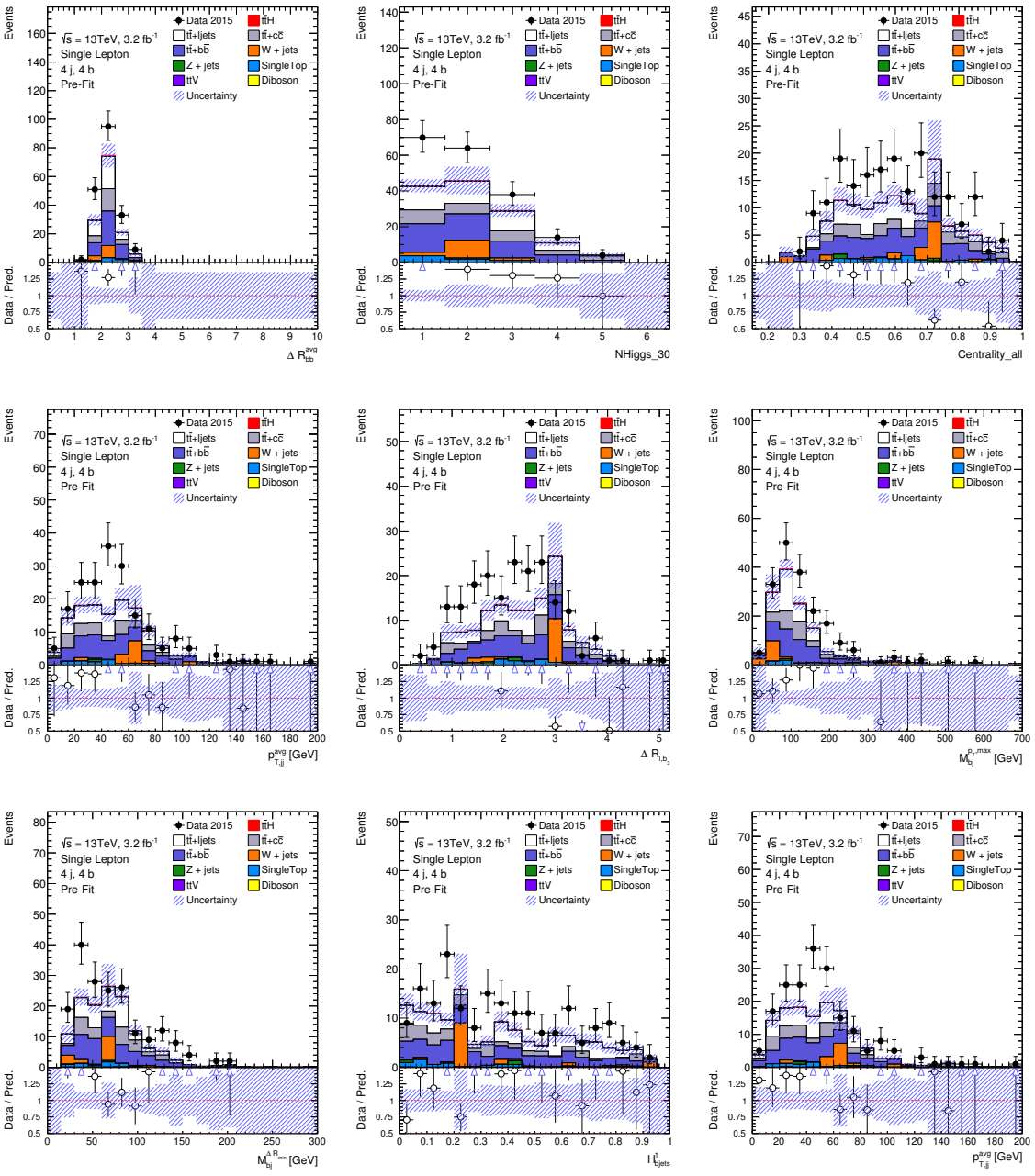


Figure 6.10.: Data-MC comparison for the discriminating variables used in the NN training in the exclusive (4j, 4b) control region.

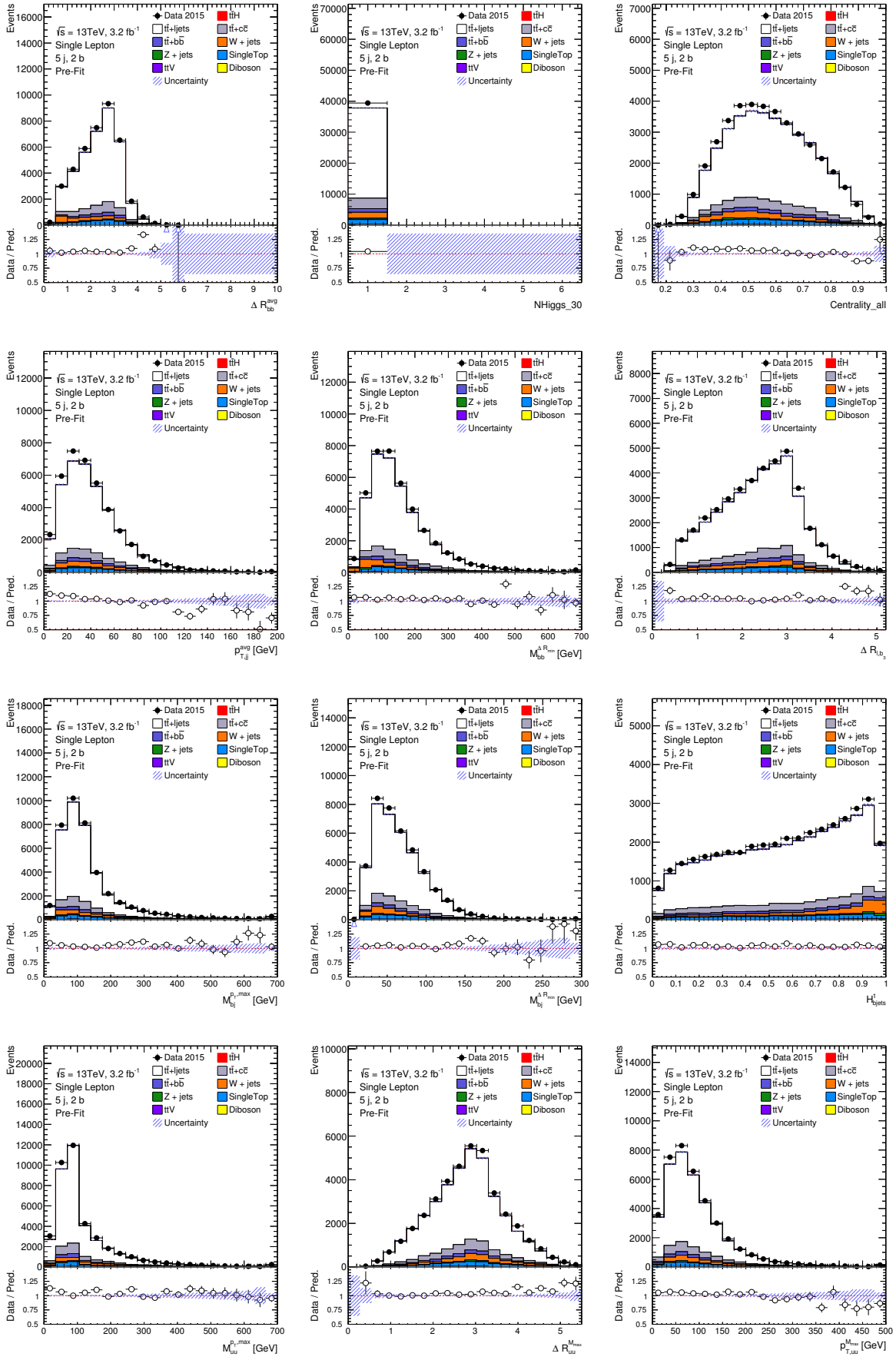


Figure 6.11.: Data-MC comparison for the discriminating variables used in the NN training in the exclusive (5j, 2b) control region.

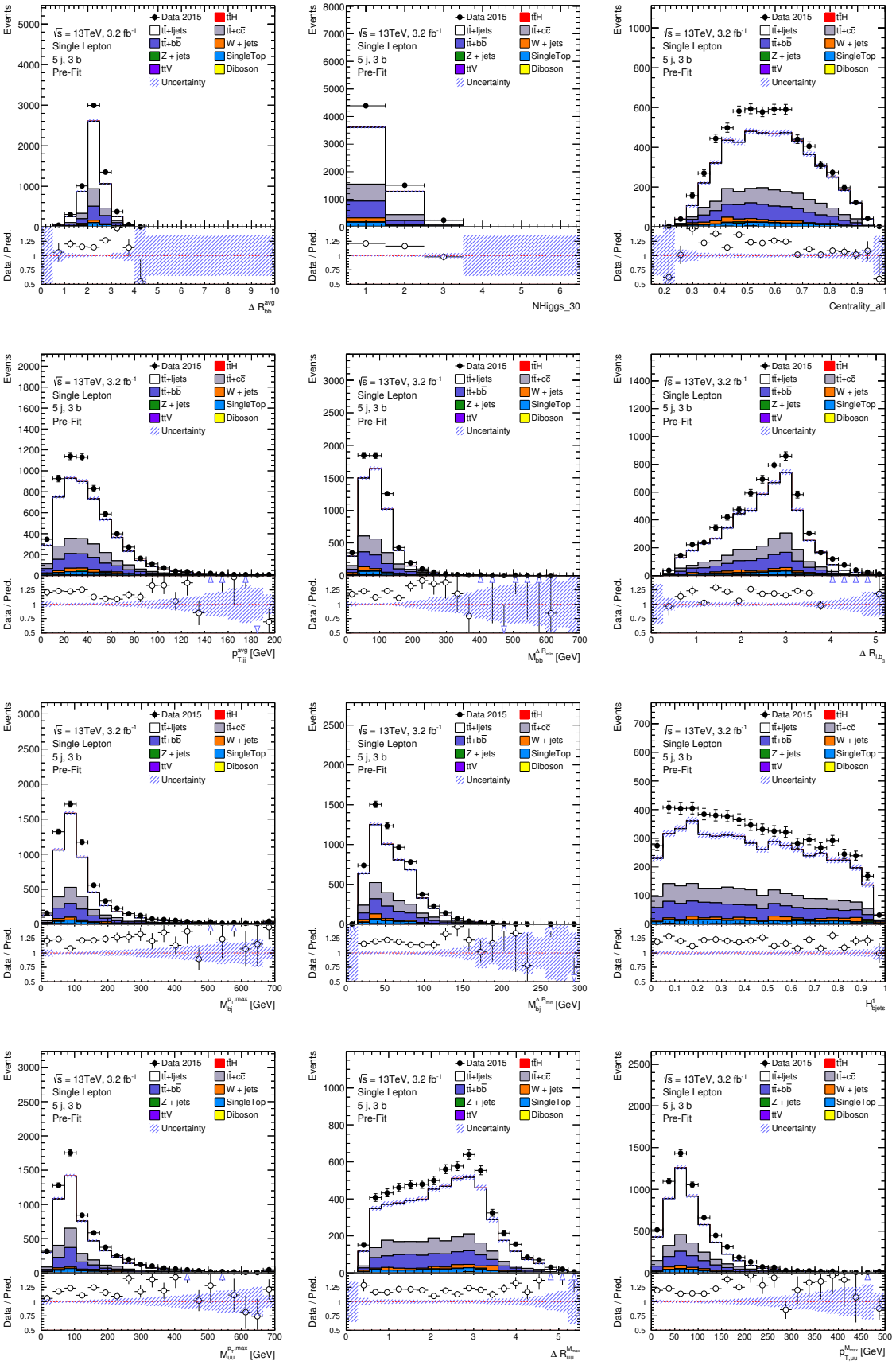


Figure 6.12.: Data-MC comparison for the discriminating variables used in the NN training in the exclusive (5j, 3b) control region.

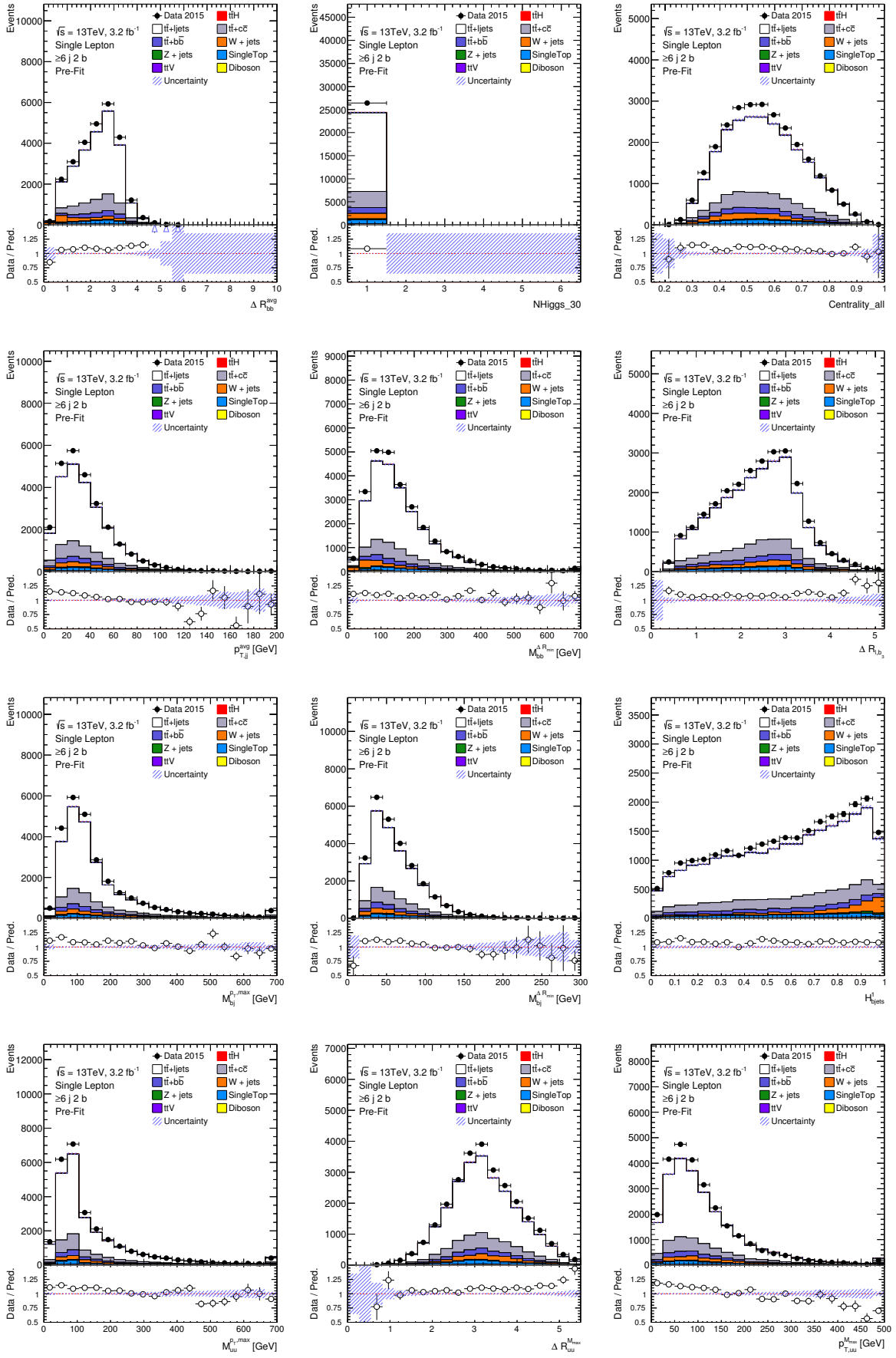


Figure 6.13.: Data-MC comparison for the discriminating variables used in the NN training in the (≥6j, 2b) control region.

7. Conclusion and Outlook

The aim of this thesis is to find variables that have the best discrimination between signal and background for inclusive ($\geq$6j, $\geq$4b) region of the single lepton channel of $t\bar{t}H, H \rightarrow b\bar{b}$ analysis currently being performed in the ATLAS experiment. The implementation of MVA techniques in the search for the rare $t\bar{t}H$ signal has been found to be very effective as a discriminant of signal from an overwhelming background. As discussed in the beginning of the thesis, a successful implementation of the NeuroBayes software has helped to set limits on the signal search performed with Run-1 data of the ATLAS detector. This thesis presents preliminary work done in implementing a new set of MVA variables in the Run-2 data of the ATLAS detector using the NeuroBayes algorithm within the $t\bar{t}H, H \rightarrow b\bar{b}$ analysis working group in their search for the elusive signal. The analysis strategy implemented was in lieu to that done in Run-1.
The NN discriminating variable has definitely shown a high separation power. This validates the careful process followed in choosing the final set of MVA variables used in the testing of the NN. There are however, certain irregularities that are seen in the MC simulations performed and Figures 6.8 to Figure 6.13 clearly show this mismatch in modelling to data in the control regions. Further look at the event yields in the Appendix Chapter B shows a clear decrease in event yields for high b-jet multiplicity regions. With low event yields, the statistical uncertainty per bin in these regions are high, as seen in Figure 6.10.

A careful look at all the comparison plots for all variables implemented is done in the control regions. Due to the unavailability of MC samples for the QCD processes, a slight disagreement in the data-MC was expected in the lower jet multiplicity regions as the QCD processes produce soft jets i.e. jets with low p_T. A first hint of modelling issues was seen with the plots of H_T, the scalar sum of p_T of all jets in an event. Figure 7.1 show the H_T plots for control regions with exactly 2 b-tagged jets and 4, 5 and $\geq$6 jets. There is a clear slope in the ratio of data to MC, and this indicates that the modelling of jets from the top quarks are not properly done by the MC generators. This is one of the main issues in the present analysis and a dedicated task force to improve $t\bar{t}$ modelling is ongoing.

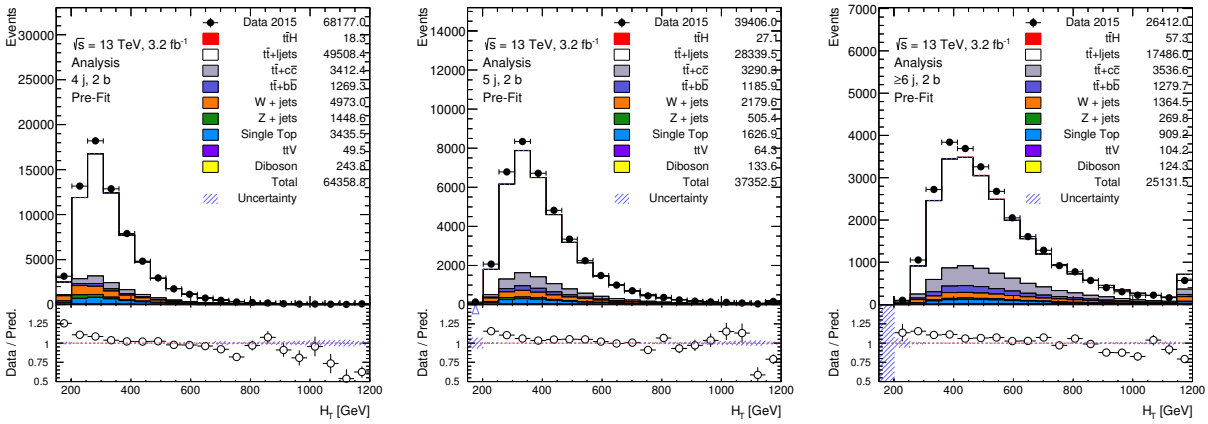


Figure 7.1.: Data-MC comparison plots for the scalar sum of the p_T of all jets (H_T) in regions with four to at least 6 jets and exactly 2 b-jets in all events.

Outlook

The Run-1 analysis was limited by statistics for the discovery of the $t\bar{t}H$ signal. Due to higher energies, higher luminosity and an increased cross section of the $t\bar{t}H$ process, the signal detection with Run-2 data is expected. The analysis is also being conducted in the dileptonic channel, where the final state consists of two opposite charged leptons and 4 jets for the $t\bar{t}H$ event. The sensitivity from both channels will be considered for a combined search for the signal.

$t\bar{t}$ modelling studies are currently underway to see if a better data-MC agreement can be obtained. Furthermore, b-tagging studies are also being developed for different working points, as to obtain the best possible sensitivity to the $t\bar{t}H$ event.

NN analysis will also be conducted independently in all signal rich regions to get a better discriminant and information from all regions will be used in a final fitting to data. Testing of more MVA variables for each of the three regions is currently underway and the variables will be implemented in the final analysis. Effort is also being done for final state reconstruction of $t\bar{t}H$ and $t\bar{t}$ events.

The prospect for the search of the $t\bar{t}H$ signal is very good within the realm of experiences gained from Run-1 analysis and improvements in analysis techniques. A culmination of this search will be either a discovery, or an insight into other deeper, more fundamental processes of nature.

Bibliography

[1] The ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 01.

[2] The CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30.

[3] F. Halzen and A. D. Martin, *Quarks and Leptons: An introductory course in Modern Particle Physics*. John Wiley & Sons, Inc., 1984.

[4] D. Griffiths, *Introduction to Elementary Particles*. Wiley-VCH Verlag GmbH & Co. KGaA, 2 ed., 2008.

[5] D. H. Perkins, *Introduction to High Energy Physics*. Cambridge University Press, 4 ed., 2005.

[6] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519** (1968) 367.

[7] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (Nov, 1967) 1264.

[8] Particle Data Group, Olive, K. A. and others, *Review of Particle Physics*, Chin. Phys. C **38** (2014) 090001.

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

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

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

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

[13] The ATLAS Collaboration, *The ATLAS Pixel Insertable B-Layer (IBL)*, Nucl. Instrum. Meth. A **650** (Dec, 2010) 45.

[14] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079.

[15] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639.

[16] S. Dawson et al., *Associated Higgs boson production with top quarks at the CERN Large Hadron Collider: NLO QCD corrections*, Phys. Rev. D **68** (Aug, 2003) 034022.

[17] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **11** (2004) 040.

[18] T. Sjöstrand et al., *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026.

[19] M. Cacciari et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. B **710** (2012) 612.

[20] P. Bärnreuther et al., *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, Phys. Rev. Lett. **109** (2012) 132001.

[21] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP **1** (2013) 1.

[22] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04** (2008) 063.

[23] S. Frixione et al., *Single-top hadroproduction in association with a W boson*, JHEP **07** (2008) 029.

[24] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. D **81** (2010) 054028.

[25] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. D **83** (2011) 091503.

[26] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a $W-$ or $H-$*, Phys. Rev. D **82** (2010) 054018.

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

[28] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039.

[29] J. Butterworth et al., *Single Boson and Diboson Production Cross Sections in pp Collisions at $\sqrt{s}$ = 7 TeV*, Tech. Rep. ATL-COM-PHYS-2010-695, CERN, Geneva, Aug, 2010.

[30] P. Golonka and Z. Was, *PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays*, Eur. Phys. J. C **45** (2006) no. 1, 97.

[31] S. Jadach, J. H. Kuhn, and Z. Was, *TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons*, Comput. Phys. Commun. **64** (1990) 275.

[32] J. M. Butterworth et al., *Multiparton interactions in photoproduction at HERA*, Z. Phys. C **72** (1996) no. 4, 637.

[33] S. Agostinelli et al., *Geant4 - A Simulation Toolkit*, Nucl. Instr. Meth. **A** (2003) 250.

[34] The ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Eur. Phys. J. C **75** (2015) 349.

[35] The ATLAS Collaboration, *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector*, JHEP **04** (2015) 117.

[36] The ATLAS Collaboration, *Tagging and suppression of pileup jets*, Tech. Rep. ATLAS-CONF-2014-018, 2014.

[37] The ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, Tech. Rep. ATLAS-CONF-2014-046, CERN, Geneva, Jul, 2014.

[38] The ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2*, Tech. Rep. ATL-PHYS-PUB-2015-022, CERN, Geneva, Jul, 2015.

[39] F. M. and K. U., *The NeuroBayes neural network package*, Nucl. Instr. Meth. A **559** (2006) 190.

[40] *NeuroBayes*, 2016. `http://neurobayes.phi-t.de`.

[41] M. Feindt, *A Neural Bayesian Estimator for Conditional Probability Densities*, `arXiv:physics/0402093`.

A. List of MC Samples

NOTE: All MC datasets begin with "mc15_13TeV.". The list below contains the names of the datasets with the DataSetID (DSID).

$t\bar{t}H$

341177.aMcAtNloHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_ttH125_dil.merge.DAOD_TOPQ1.e4277_s2608_s2183_r6869_r6282_p2454

341270.aMcAtNloHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_ttH125_semilep.merge.DAOD_TOPQ1.e4277_s2608_s2183_r6869_r6282_

341271.aMcAtNloHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_ttH125_allhad.merge.DAOD _TOPQ1.e4277_s2608_s2183_r6869_r6282_p2454

Diboson

361063.Sherpa_CT10_llll.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_r6282_p2454

361064.Sherpa_CT10_lllvSFMinus.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_r6282_p2454

361065.Sherpa_CT10_lllvOFMinus.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_r6282_p2454

361066.Sherpa_CT10_lllvSFPlus.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_r6282_p2454

361067.Sherpa_CT10_lllvOFPlus.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_r6282_p2454

361068.Sherpa_CT10_llvv.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_r6282_p2454

361070.Sherpa_CT10_llvvjj_ss_EW6.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_r6282_p2454

361071.Sherpa_CT10_lllvjj_EW6.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_r6282_p2454

361072.Sherpa_CT10_llljj_EW6.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_r6282_p2454

361073.Sherpa_CT10_gglll.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_r6282_

p2454
361077.Sherpa_CT10_ggllvv.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_r6282_
p2454
361081.Sherpa_CT10_WplvWmqq.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_
r6282_p2454
361082.Sherpa_CT10_WpqqWmlv.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_
r6282_p2454
361083.Sherpa_CT10_WlvZqq.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_r6282_
p2454
361084.Sherpa_CT10_WqqZll.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_r6282_
p2454
361085.Sherpa_CT10_WqqZvv.merge.DAOD_TOPQ1.e3836_s2608_s2183_r6869_r6282_
p2454
361086.Sherpa_CT10_ZqqZll.merge.DAOD_TOPQ1.e3926_s2608_s2183_r6869_r6282_
p2454
361600.PowhegPy8EG_CT10nloME_AZNLOCTEQ6L1_WWlvlv.merge.DAOD_TOPQ1.e4054_
s2608_s2183_r6869_r6282_p2454
361601.PowhegPy8EG_CT10nloME_AZNLOCTEQ6L1_WZlvll_mll4.merge.DAOD_TOPQ1.e4054_
s2608_s2183_r6869_r6282_p2454
361602.PowhegPy8EG_CT10nloME_AZNLOCTEQ6L1_WZlvvv_mll4.merge.DAOD_TOPQ1.e4054_
s2608_s2183_r6869_r6282_p2454
361603.PowhegPy8EG_CT10nloME_AZNLOCTEQ6L1_ZZllll_mll4.merge.DAOD_TOPQ1.e4054_
s2608_s2183_r6869_r6282_p2454
361604.PowhegPy8EG_CT10nloME_AZNLOCTEQ6L1_ZZvvll_mll4.merge.DAOD_TOPQ1.e4054_
s2608_s2183_r6869_r6282_p2454
361606.PowhegPy8EG_CT10nloME_AZNLOCTEQ6L1_WWlvqq.merge.DAOD_TOPQ1.e4054_
s2608_s2183_r6869_r6282_p2454
361607.PowhegPy8EG_CT10nloME_AZNLOCTEQ6L1_WZqqll_mll20.merge.DAOD_TOPQ1.e4054_
s2608_s2183_r6869_r6282_p2454
361608.PowhegPy8EG_CT10nloME_AZNLOCTEQ6L1_WZqqvv.merge.DAOD_TOPQ1.e4054_s2608_
s2183_r6869_r6282_p2454
361609.PowhegPy8EG_CT10nloME_AZNLOCTEQ6L1_WZlvqq_mqq20.merge.DAOD_TOPQ1.e4054_
s2608_s2183_r6869_r6282_p2454

W**+jets**

361300.Sherpa_CT10_Wenu_Pt0_70_CVetoBVeto.merge.DAOD_TOPQ1.e3651_s2608_s2183_
r6869_r6282_p2454

361301.Sherpa_CT10_Wenu_Pt0_70_CFilterBVeto.merge.DAOD_TOPQ1.e3651_s2586_s2174_
r6869_r6282_p2454
361302.Sherpa_CT10_Wenu_Pt0_70_BFilter.merge.DAOD_TOPQ1.e3651_s2586_s2174_r6869_
r6282_p2454
361303.Sherpa_CT10_Wenu_Pt70_140_CVetoBVeto.merge.DAOD_TOPQ1.e3651_s2586_s2174_
r6869_r6282_p2454
361304.Sherpa_CT10_Wenu_Pt70_140_CFilterBVeto.merge.DAOD_TOPQ1.e3651_s2586_s2174_
r6869_r6282_p2454
361305.Sherpa_CT10_Wenu_Pt70_140_BFilter.merge.DAOD_TOPQ1.e3651_s2608_s2183_r6869_
r6282_p2454
361306.Sherpa_CT10_Wenu_Pt140_280_CVetoBVeto.merge.DAOD_TOPQ1.e3651_s2586_s2174_
r6869_r6282_p2454
361307.Sherpa_CT10_Wenu_Pt140_280_CFilterBVeto.merge.DAOD_TOPQ1.e3651_s2586_
s2174_r6869_r6282_p2454
361308.Sherpa_CT10_Wenu_Pt140_280_BFilter.merge.DAOD_TOPQ1.e3651_s2586_s2174_
r6869_r6282_p2454
361309.Sherpa_CT10_Wenu_Pt280_500_CVetoBVeto.merge.DAOD_TOPQ1.e4133_s2608_
s2183_ r6869_r6282_p2454
361310.Sherpa_CT10_Wenu_Pt280_500_CFilterBVeto.merge.DAOD_TOPQ1.e4133_s2608_
s2183_r6869_r6282_p2454
361311.Sherpa_CT10_Wenu_Pt280_500_BFilter.merge.DAOD_TOPQ1.e4133_s2608_
s2183_r6869_r6282_p2454
361312.Sherpa_CT10_Wenu_Pt500_700_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361313.Sherpa_CT10_Wenu_Pt500_700_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361314.Sherpa_CT10_Wenu_Pt500_700_BFilter.merge.DAOD_TOPQ1.e4133_s2608_
s2183_r6869_r6282_p2454
361315.Sherpa_CT10_Wenu_Pt700_1000_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361316.Sherpa_CT10_Wenu_Pt700_1000_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361317.Sherpa_CT10_Wenu_Pt700_1000_BFilter.merge.DAOD_TOPQ1.e4133_s2608_
s2183_r6869_r6282_p2454
361318.Sherpa_CT10_Wenu_Pt1000_2000_CVetoBVeto.merge.DAOD_TOPQ1.e4133_s2608_
s2183_r6869_r6282_p2454

361319.Sherpa_CT10_Wenu_Pt1000_2000_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361320.Sherpa_CT10_Wenu_Pt1000_2000_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361321.Sherpa_CT10_Wenu_Pt2000_E_CMS_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361322.Sherpa_CT10_Wenu_Pt2000_E_CMS_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361323.Sherpa_CT10_Wenu_Pt2000_E_CMS_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361324.Sherpa_CT10_Wmunu_Pt0_70_CVetoBVeto.merge.DAOD_TOPQ1.e3651_
s2608_s2183_r6869_r6282_p2454
361325.Sherpa_CT10_Wmunu_Pt0_70_CFilterBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361326.Sherpa_CT10_Wmunu_Pt0_70_BFilter.merge.DAOD_TOPQ1.e3651_s2608_
s2183_r6869_r6282_p2454
361327.Sherpa_CT10_Wmunu_Pt70_140_CVetoBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361328.Sherpa_CT10_Wmunu_Pt70_140_CFilterBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361329.Sherpa_CT10_Wmunu_Pt70_140_BFilter.merge.DAOD_TOPQ1.e3651_s2586_
s2174_r6869_r6282_p2454
361330.Sherpa_CT10_Wmunu_Pt140_280_CVetoBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361331.Sherpa_CT10_Wmunu_Pt140_280_CFilterBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361332.Sherpa_CT10_Wmunu_Pt140_280_BFilter.merge.DAOD_TOPQ1.e3651_s2586_
s2174_r6869_r6282_p2454
361333.Sherpa_CT10_Wmunu_Pt280_500_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361334.Sherpa_CT10_Wmunu_Pt280_500_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361335.Sherpa_CT10_Wmunu_Pt280_500_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361336.Sherpa_CT10_Wmunu_Pt500_700_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454

361337.Sherpa_CT10_Wmunu_Pt500_700_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361338.Sherpa_CT10_Wmunu_Pt500_700_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361339.Sherpa_CT10_Wmunu_Pt700_1000_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361340.Sherpa_CT10_Wmunu_Pt700_1000_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361341.Sherpa_CT10_Wmunu_Pt700_1000_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361342.Sherpa_CT10_Wmunu_Pt1000_2000_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361343.Sherpa_CT10_Wmunu_Pt1000_2000_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361344.Sherpa_CT10_Wmunu_Pt1000_2000_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361345.Sherpa_CT10_Wmunu_Pt2000_E_CMS_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361346.Sherpa_CT10_Wmunu_Pt2000_E_CMS_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361347.Sherpa_CT10_Wmunu_Pt2000_E_CMS_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361348.Sherpa_CT10_Wtaunu_Pt0_70_CVetoBVeto.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454
361349.Sherpa_CT10_Wtaunu_Pt0_70_CFilterBVeto.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454
361350.Sherpa_CT10_Wtaunu_Pt0_70_BFilter.merge.DAOD_TOPQ1.e3733_s2608_
s2183_r6869_r6282_p2454
361351.Sherpa_CT10_Wtaunu_Pt70_140_CVetoBVeto.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454
361352.Sherpa_CT10_Wtaunu_Pt70_140_CFilterBVeto.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_ r6282_p2454
361353.Sherpa_CT10_Wtaunu_Pt70_140_BFilter.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454
361354.Sherpa_CT10_Wtaunu_Pt140_280_CVetoBVeto.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454

361355.Sherpa_CT10_Wtaunu_Pt140_280_CFilterBVeto.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454
361356.Sherpa_CT10_Wtaunu_Pt140_280_BFilter.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454
361357.Sherpa_CT10_Wtaunu_Pt280_500_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361358.Sherpa_CT10_Wtaunu_Pt280_500_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361359.Sherpa_CT10_Wtaunu_Pt280_500_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361360.Sherpa_CT10_Wtaunu_Pt500_700_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361361.Sherpa_CT10_Wtaunu_Pt500_700_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361362.Sherpa_CT10_Wtaunu_Pt500_700_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361363.Sherpa_CT10_Wtaunu_Pt700_1000_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361364.Sherpa_CT10_Wtaunu_Pt700_1000_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361365.Sherpa_CT10_Wtaunu_Pt700_1000_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361366.Sherpa_CT10_Wtaunu_Pt1000_2000_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361367.Sherpa_CT10_Wtaunu_Pt1000_2000_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361368.Sherpa_CT10_Wtaunu_Pt1000_2000_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361369.Sherpa_CT10_Wtaunu_Pt2000_E_CMS_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361370.Sherpa_CT10_Wtaunu_Pt2000_E_CMS_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361371.Sherpa_CT10_Wtaunu_Pt2000_E_CMS_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361520.MadGraphPythia8EvtGen_A14NNPDF23LO_Wenu_Np0.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454

361521.MadGraphPythia8EvtGen_A14NNPDF23LO_Wenu_Np1.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361522.MadGraphPythia8EvtGen_A14NNPDF23LO_Wenu_Np2.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361523.MadGraphPythia8EvtGen_A14NNPDF23LO_Wenu_Np3.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361524.MadGraphPythia8EvtGen_A14NNPDF23LO_Wenu_Np4.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361525.MadGraphPythia8EvtGen_A14NNPDF23LO_Wmunu_Np0.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361526.MadGraphPythia8EvtGen_A14NNPDF23LO_Wmunu_Np1.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361527.MadGraphPythia8EvtGen_A14NNPDF23LO_Wmunu_Np2.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361528.MadGraphPythia8EvtGen_A14NNPDF23LO_Wmunu_Np3.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361529.MadGraphPythia8EvtGen_A14NNPDF23LO_Wmunu_Np4.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361530.MadGraphPythia8EvtGen_A14NNPDF23LO_Wtaunu_Np0.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361531.MadGraphPythia8EvtGen_A14NNPDF23LO_Wtaunu_Np1.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361532.MadGraphPythia8EvtGen_A14NNPDF23LO_Wtaunu_Np2.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361534.MadGraphPythia8EvtGen_A14NNPDF23LO_Wtaunu_Np4.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454

Z+jets

361372.Sherpa_CT10_Zee_Pt0_70_CVetoBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361373.Sherpa_CT10_Zee_Pt0_70_CFilterBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361374.Sherpa_CT10_Zee_Pt0_70_BFilter.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361375.Sherpa_CT10_Zee_Pt70_140_CVetoBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361376.Sherpa_CT10_Zee_Pt70_140_CFilterBVeto.merge.DAOD_TOPQ1.e3651_

s2586_s2174_r6869_r6282_p2454
361377.Sherpa_CT10_Zee_Pt70_140_BFilter.merge.DAOD_TOPQ1.e3651_s2586_
s2174_r6869_r6282_p2454
361378.Sherpa_CT10_Zee_Pt140_280_CVetoBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361379.Sherpa_CT10_Zee_Pt140_280_CFilterBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361380.Sherpa_CT10_Zee_Pt140_280_BFilter.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361381.Sherpa_CT10_Zee_Pt280_500_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361382.Sherpa_CT10_Zee_Pt280_500_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361383.Sherpa_CT10_Zee_Pt280_500_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361384.Sherpa_CT10_Zee_Pt500_700_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361385.Sherpa_CT10_Zee_Pt500_700_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361386.Sherpa_CT10_Zee_Pt500_700_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361387.Sherpa_CT10_Zee_Pt700_1000_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361388.Sherpa_CT10_Zee_Pt700_1000_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361389.Sherpa_CT10_Zee_Pt700_1000_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361390.Sherpa_CT10_Zee_Pt1000_2000_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361391.Sherpa_CT10_Zee_Pt1000_2000_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361392.Sherpa_CT10_Zee_Pt1000_2000_BFilter.merge.DAOD_TOPQ1.e4133_s2608_
s2183_r6869_r6282_p2454
361393.Sherpa_CT10_Zee_Pt2000_E_CMS_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361394.Sherpa_CT10_Zee_Pt2000_E_CMS_CFilterBVeto.merge.DAOD_TOPQ1.e4133_

s2608_s2183_r6869_r6282_p2454
361395.Sherpa_CT10_Zee_Pt2000_E_CMS_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361397.Sherpa_CT10_Zmumu_Pt0_70_CFilterBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361398.Sherpa_CT10_Zmumu_Pt0_70_BFilter.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361399.Sherpa_CT10_Zmumu_Pt70_140_CVetoBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361400.Sherpa_CT10_Zmumu_Pt70_140_CFilterBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361401.Sherpa_CT10_Zmumu_Pt70_140_BFilter.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361402.Sherpa_CT10_Zmumu_Pt140_280_CVetoBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361402.Sherpa_CT10_Zmumu_Pt140_280_CVetoBVeto.merge.DAOD_TOPQ1.e3651_
s2608_s2183_r6869_r6282_p2454
361403.Sherpa_CT10_Zmumu_Pt140_280_CFilterBVeto.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361404.Sherpa_CT10_Zmumu_Pt140_280_BFilter.merge.DAOD_TOPQ1.e3651_
s2586_s2174_r6869_r6282_p2454
361405.Sherpa_CT10_Zmumu_Pt280_500_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361406.Sherpa_CT10_Zmumu_Pt280_500_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361407.Sherpa_CT10_Zmumu_Pt280_500_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361409.Sherpa_CT10_Zmumu_Pt500_700_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361410.Sherpa_CT10_Zmumu_Pt500_700_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361411.Sherpa_CT10_Zmumu_Pt700_1000_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361412.Sherpa_CT10_Zmumu_Pt700_1000_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361413.Sherpa_CT10_Zmumu_Pt700_1000_BFilter.merge.DAOD_TOPQ1.e4133_

s2608_s2183_r6869_r6282_p2454
361414.Sherpa_CT10_Zmumu_Pt1000_2000_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361415.Sherpa_CT10_Zmumu_Pt1000_2000_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361416.Sherpa_CT10_Zmumu_Pt1000_2000_BFilter.merge.DAOD_TOPQ1.e4133_s2608_
s2183_r6869_r6282_p2454
361417.Sherpa_CT10_Zmumu_Pt2000_E_CMS_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361418.Sherpa_CT10_Zmumu_Pt2000_E_CMS_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361419.Sherpa_CT10_Zmumu_Pt2000_E_CMS_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361420.Sherpa_CT10_Ztautau_Pt0_70_CVetoBVeto.merge.DAOD_TOPQ1.e3733_
s2586_s2174_r6869_r6282_p2454
361420.Sherpa_CT10_Ztautau_Pt0_70_CVetoBVeto.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454
361421.Sherpa_CT10_Ztautau_Pt0_70_CFilterBVeto.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454
361422.Sherpa_CT10_Ztautau_Pt0_70_BFilter.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454
361423.Sherpa_CT10_Ztautau_Pt70_140_CVetoBVeto.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454
361424.Sherpa_CT10_Ztautau_Pt70_140_CFilterBVeto.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454
361425.Sherpa_CT10_Ztautau_Pt70_140_BFilter.merge.DAOD_TOPQ1.e3733_s2608_
s2183_r6869_r6282_p2454
361426.Sherpa_CT10_Ztautau_Pt140_280_CVetoBVeto.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454
361427.Sherpa_CT10_Ztautau_Pt140_280_CFilterBVeto.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454
361428.Sherpa_CT10_Ztautau_Pt140_280_BFilter.merge.DAOD_TOPQ1.e3733_
s2608_s2183_r6869_r6282_p2454
361429.Sherpa_CT10_Ztautau_Pt280_500_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361430.Sherpa_CT10_Ztautau_Pt280_500_CFilterBVeto.merge.DAOD_TOPQ1.e4133_

s2608_s2183_r6869_r6282_p2454
361431.Sherpa_CT10_Ztautau_Pt280_500_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361432.Sherpa_CT10_Ztautau_Pt500_700_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361433.Sherpa_CT10_Ztautau_Pt500_700_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361434.Sherpa_CT10_Ztautau_Pt500_700_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361435.Sherpa_CT10_Ztautau_Pt700_1000_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361436.Sherpa_CT10_Ztautau_Pt700_1000_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361437.Sherpa_CT10_Ztautau_Pt700_1000_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361438.Sherpa_CT10_Ztautau_Pt1000_2000_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361439.Sherpa_CT10_Ztautau_Pt1000_2000_CFilterBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361441.Sherpa_CT10_Ztautau_Pt2000_E_CMS_CVetoBVeto.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361443.Sherpa_CT10_Ztautau_Pt2000_E_CMS_BFilter.merge.DAOD_TOPQ1.e4133_
s2608_s2183_r6869_r6282_p2454
361468.Sherpa_CT10_Zee_Mll10to40_Pt0_70_BVeto.merge.DAOD_TOPQ1.e4198_
s2608_s2183_r6869_r6282_p2454
361486.Sherpa_CT10_Ztautau_Mll10to40_Pt70_140_BVeto.merge.DAOD_
TOPQ1.e4198_s2608_s2183_r6869_r6282_p2454
361487.Sherpa_CT10_Ztautau_Mll10to40_Pt70_140_BFilter.merge.DAOD_
TOPQ1.e4198_s2608_s2183_r6869_r6282_p2454
361488.Sherpa_CT10_Ztautau_Mll10to40_Pt140_400_BVeto.merge.DAOD_
TOPQ1.e4198_s2608_s2183_r6869_r6282_p2454
361489.Sherpa_CT10_Ztautau_Mll10to40_Pt140_400_BFilter.merge.DAOD_
TOPQ1.e4198_s2608_s2183_r6869_r6282_p2454
361490.Sherpa_CT10_Ztautau_Mll10to40_Pt400_E_CMS_BVeto.merge.DAOD_
TOPQ1.e4198_s2608_s2183_r6869_r6282_p2454
361491.Sherpa_CT10_Ztautau_Mll10to40_Pt400_E_CMS_BFilter.merge.DAOD_

TOPQ1.e4198_s2608_s2183_r6869_r6282_p2454
361500.MadGraphPythia8EvtGen_A14NNPDF23LO_Zee_Np0.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361501.MadGraphPythia8EvtGen_A14NNPDF23LO_Zee_Np1.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361502.MadGraphPythia8EvtGen_A14NNPDF23LO_Zee_Np2.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361506.MadGraphPythia8EvtGen_A14NNPDF23LO_Zmumu_Np1.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361507.MadGraphPythia8EvtGen_A14NNPDF23LO_Zmumu_Np2.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361508.MadGraphPythia8EvtGen_A14NNPDF23LO_Zmumu_Np3.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361509.MadGraphPythia8EvtGen_A14NNPDF23LO_Zmumu_Np4.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361510.MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_Np0.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361511.MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_Np1.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361512.MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_Np2.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361513.MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_Np3.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454
361514.MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_Np4.merge.DAOD_TOPQ1.e3898_
s2608_s2183_r6869_r6282_p2454

$t\bar{t}$ nominal

410000.PowhegPythiaEvtGen_P2012_ttbar_hdamp172p5_nonallhad.merge.DAOD_
TOPQ1.e3698_s2608_s2183_r6765_r6282_p2454
410120.PowhegPythiaEvtGen_P2012_ttbar_hdamp172p5_nonallhad_bfilter.merge.DAOD_
TOPQ1.e4373_s2608_r6869_r6282_p2454

Single Top

410011.PowhegPythiaEvtGen_P2012_singletop_tchan_lept_top.merge.DAOD_
TOPQ1.e3824_s2608_s2183_r6869_r6282_p2454
410012.PowhegPythiaEvtGen_P2012_singletop_tchan_lept_antitop.merge.DAOD_
TOPQ1.e3824_s2608_s2183_r6869_r6282_p2454
410013.PowhegPythiaEvtGen_P2012_Wt_inclusive_top.merge.DAOD_

TOPQ1.e3753_s2608_s2183_r6869_r6282_p2454

410014.PowhegPythiaEvtGen_P2012_Wt_inclusive_antitop.merge.DAOD_TOPQ1.e3753_s2608_s2183_r6869_r6282_p2454

410025.PowhegPythiaEvtGen_P2012_SingleTopSchan_noAllHad_top.merge.DAOD_TOPQ1.e3998_s2608_s2183_r6869_r6282_p2454

410026.PowhegPythiaEvtGen_P2012_SingleTopSchan_noAllHad_antitop.merge.DAOD_TOPQ1.e3998_s2608_s2183_r6869_r6282_p2454

$t\bar{t}V$

410066.MadGraphPythia8EvtGen_A14NNPDF23LO_ttW_Np0.merge.DAOD_TOPQ1.e4111_s2608_s2183_r6869_r6282_p2454

410067.MadGraphPythia8EvtGen_A14NNPDF23LO_ttW_Np1.merge.DAOD_TOPQ1.e4111_s2608_s2183_r6869_r6282_p2454

410068.MadGraphPythia8EvtGen_A14NNPDF23LO_ttW_Np2.merge.DAOD_TOPQ1.e4111_s2608_s2183_r6869_r6282_p2454

410073.MadGraphPythia8EvtGen_A14NNPDF23LO_ttZnnqq_Np0.merge.DAOD_TOPQ1.e4111_s2608_s2183_r6869_r6282_p2454

410074.MadGraphPythia8EvtGen_A14NNPDF23LO_ttZnnqq_Np1.merge.DAOD_TOPQ1.e4143_s2608_s2183_r6869_r6282_p2454

410075.MadGraphPythia8EvtGen_A14NNPDF23LO_ttZnnqq_Np2.merge.DAOD_TOPQ1.e4111_s2608_s2183_r6869_r6282_p2454

410111.MadGraphPythia8EvtGen_A14NNPDF23LO_ttee_Np0.merge.DAOD_TOPQ1.e4265_s2608_s2183_r6869_r6282_p2454

410112.MadGraphPythia8EvtGen_A14NNPDF23LO_ttee_Np1.merge.DAOD_TOPQ1.e4265_s2608_s2183_r6869_r6282_p2454

410113.MadGraphPythia8EvtGen_A14NNPDF23LO_ttmumu_Np0.merge.DAOD_TOPQ1.e4265_s2608_s2183_r6869_r6282_p2454

410114.MadGraphPythia8EvtGen_A14NNPDF23LO_ttmumu_Np1.merge.DAOD_TOPQ1.e4265_s2608_s2183_r6869_r6282_p2454

B. Event Yields for all regions

Samples	**4j, 2b**	**4j, 3b**	**4j, 4b**
$t\bar{t}$+light jets	49498.5 ± 56.4	4614.4 ± 17.6	49.9 ± 2.21
$t\bar{t} + c\bar{c}$	3412.4 ± 14.8	609.9 ± 6.21	27.2 ± 1.33
$t\bar{t} + b\bar{b}$	1269.3 ± 2.74	555.2 ± 1.81	42.0 ± 0.50
W+jets	4973.0 ± 77.8	261.8 ± 11.1	4.5 ± 0.93
Z+jets	1448.7 ± 31.5	77.0 ± 5.3	0.8 ± 0.31
Single Top	3435.5 ± 8.91	290.2 ± 2.67	6.7 ± 0.42
$t\bar{t}V$	49.5 ± 0.12	6.7 ± 0.04	0.51 ± 0.01
Diboson	243.8 ± 3.59	14.9 ± 0.98	0.52 ± 0.17
Total Background	64349.0 ± 103.0	6438.4 ± 22.6	133.4 ± 2.85
$t\bar{t}H$	18.3 ± 0.18	8.2 ± 0.11	1.4 ± 0.04
Data	68177.0 ± 0.0	7369.0 ± 0.0	189.0 ± 0.0

Table B.1.: Event yields for the 4-jet region.

Samples	**5j, 2b**	**5j, 3b**	**5j, ≥4b**
$t\bar{t}$+light jets	28339.0 ± 42.6	3200.3 ± 14.9	73.8 ± 2.73
$t\bar{t} + c\bar{c}$	3290.3 ± 14.5	795.6 ± 7.13	59.2 ± 1.98
$t\bar{t} + b\bar{b}$	1185.9 ± 2.65	744.2 ± 2.1	114.9 ± 0.83
W+jets	2179.6 ± 34.3	167.7 ± 7.56	11.0 ± 1.89
Z+jets	505.4 ± 14.3	31.9 ± 2.49	2.0 ± 0.53
Single Top	1626.9 ± 6.08	199.8 ± 2.2	11.1 ± 0.55
$t\bar{t}V$	64.3 ± 0.14	12.7 ± 0.06	1.8 ± 0.02
Diboson	133.6 ± 2.62	14.8 ± 0.93	1.5 ± 0.31
Total Background	37352.0 ± 58.9	5182.7 ± 18.6	279.9 ± 4.04
$t\bar{t}H$	27.1 ± 0.23	15.7 ± 0.17	4.7 ± 0.09
Data	39406.0 ± 0.0	6140.0 ± 0.0	347.0 ± 0.0

Table B.2.: Event yields for the 5-jet region

Samples	**≥6j, 2b**	**≥6j, 3b**
$t\bar{t}$+light jets	17486.0 ± 33.5	2378.9 ± 13.1
$t\bar{t} + c\bar{c}$	3536.6 ± 15.1	1074.3 ± 8.36
$t\bar{t} + b\bar{b}$	1279.7 ± 2.75	1066.6 ± 2.51
W+jets	1364.5 ± 24.5	164.5 ± 12.5
Z+jets	269.8 ± 9.41	36.0 ± 2.99
Single Top	909.2 ± 4.54	166.6 ± 2.02
$t\bar{t}V$	104.2 ± 0.19	29.1 ± 0.1
Diboson	124.3 ± 2.44	18.8 ± 1.0
Total Background	25131.5 ± 45.5	4973.9 ± 20.4
$t\bar{t}H$	57.3 ± 0.36	39.2 ± 0.29
Data	26412.0 ± 0.0	6020.2 ± 0.0

Table B.3.: Event yields for the 6-jet region.

Acknowledgement

This previous year spent being involved in an analysis that is searching for a signal of a physics process has been one of the most rewarding experiences of my life. The learning experience alone, for a person from Nepal, was virtually unattainable 2 years ago.
All of what I have achieved, and what I hope to achieve in the future in particle physics, can almost exclusively be attributed to one man, Prof. Dr. Arnulf Quadt. He has been a source of constant support, inspiration and guidance for me in developing not only my skills academically, but being active in fulfilling my responsibility towards my society. He's expressed support and confidence in me when at times, I have lost that in myself. The last 1.5 years working for him, and the communication I had before I started here, will be something I will always remember and cherish.
I would also like to thank Dr. Maria Moreno Illacer and Matteo Mantoani; two incredible people that have helped me develop my skills in physics analysis and have been the most patient and supportive mentors that one can ask for. I will eternally be thankful to Matteo, for being on the other end, helping me through my problems. Maria has been an incredible teacher, someone I can depend on whenever I have any issues with understanding the physics behind my analysis.
I would also like to thank Dr. Elizaveta Shabalina, who has been there for me, guiding me in my research and providing crucial feedback on my work. I hope to be able to learn more from her in the future. I would also like to thank the II. Institute for Physics for all the help and support they have provided me in the course of my research here.
I would finally like to thank all my friends and family, a pillar of support for me emotionally, and the distraction I needed at times to clear out my head.

Ishan Pokharel
June 10, 2016

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 10. Juni 2016

(Ishan Pokharel)