IMPERIAL COLLEGE LONDON,
UNIVERSITAT DE VALENCIA

MASTER THESIS

Selection of *t*-channel Single Top Quark Events Using a Bayesian Neural Network to Explore the *Wtb* Vertex with the ATLAS Experiment

Author:
Abigail ALEXANDER

Supervisor:
Susana CABRERA-URBÁN

A thesis submitted in fulfillment of the requirements for the degree of Masters of Physics

in

High Energy Physics
Department of Physics

July 1, 2018

IMPERIAL COLLEGE LONDON,
UNIVERSITAT DE VALENCIA

Abstract

IFIC

Department of Physics

Masters of Physics

Selection of *t*-channel Single Top Quark Events Using a Bayesian Neural Network to Explore the *Wtb* Vertex with the ATLAS Experiment

by Abigail ALEXANDER

Abstract

The use of a multivariate method based on a Bayesian neural network (NeuroBayes) was explored for the first time to provide a selection of t-channel single top quark events with the purpose to explore anomalous couplings in the Wtb vertex. The training and optimisation of this selection has been performed with simulated signal and background samples of Standard Model processes produced in proton-proton collisions at both centre of mass energies of $\sqrt{s} = 8$ and 13 TeV provided by the ATLAS experiment at the LHC. After the fine-tuning of the neural network's parameters and a careful choice of kinematic training variables, a cut in a single output discriminant, O_{NN}, was applied to optimise the samples. For a fixed working significance of ~ 80 at $\sqrt{s} = 13$ TeV, the neural network based method outperformed the previous selections based on sequential optimised cuts, yielding a much larger final value of 1.25 for the ratio of signal to background in comparison to the 0.60 from before.

The Wtb vertex of in t-channel single top quark events, probed in both top quark production and decay, is described by a Lagrangian which parametrises the couplings present in the Standard Model, including the possibility of non Standard Model anomalous couplings. Measurements of these couplings can be extracted from calculations of the forward-backward asymmetries of the relevant angular distributions. To determine whether the new neural network based optimisation will bias the future measurements of potential anomalous couplings, an unfolding procedure to parton level was performed to simulated samples with anomalous couplings for the angular distributions $cos(\theta_l)$ and $cos(\theta_l^N)$ and their unfolded asymmetries calculated. It was found that the analysis maintained sensitivity to the presence of the anomalous couplings, as quantified by small statistical uncertainties in unfolded asymmetries, allowing us to make a consistency check of the Standard Model with these observables. However, the analysis exhibited deviations from the linearity of the response to their measurement. The deviations should be corrected for or prevented in future work prior to the application of the optimisation to real data in the search for anomalous couplings. Prospects for that purpose are discussed.

Resumen

La selección de una muestra de eventos con producción de un solo quark top en el canal t con la cual explorar en detalle la existencia de posibles acoplamientos anómaolos en el vértice Wtb se ha llevado a cabo por primera vez utilizando un método multivariado basado en una red neuronal Bayesiana denominada NeuroBayes. El entrenamiento y la optimización de dicha selección se ha realizado con muestras simuladas de procesos de señal y fondos producidos según el Modelo Estándar en colisiones protón-protón a $\sqrt{s} = 8$ y 13 TeV y proporcionadas por el experimento ATLAS del LHC. Después del ajuste fino de los parámetros de funcionamiento de la red neuronal, y una cuidadosa elección de las variables cinemáticas discriminatorias de entrenamiento, se ha obtenido un único discriminante, O_{NN}, como resultado del entrenamiento de la red y de su aplicación a las muestras de señal y fondos. Más tarde, se optimizó el corte en dicho discriminante para realizar una optimización estadística de la selección. Como resultado de dicha optimización, para una significancia de trabajo fijada en ~ 80 a 13 TeV, el nuevo método ha proporcionado una fracción señal fondo de 1.25, substancialmente mejorada con respecto al valor de 0.6 obtenido en la optimización previa basada en cortes secuenciales.

En eventos con producción de un solo quark top en el canal t, el vértice Wtb puede ser explorado tanto en la producción como en la desintegración del quark top. Dicho vértice está descrito por un Lagrangiano que parametriza los acoplamientos presentes en el Modelo Estándar e incluye términos adicionales con acoplamientos anómalos, más allá del Modelo Estándar. Las medidas de estos acoplamientos se pueden extraer de los cálculos de las asimetrías *forward-backward* de las distribuciones angulares relevantes. Para determinar si la nueva optimización basada en la red neuronal sesgará las medidas futuras de posibles acoplamientos anómalos, se realizó un procedimiento de *Unfolding* a nivel de partones utilizando las muestras simuladas con acoplamientos anómalos para las distribuciones angulares $cos(\theta_l)$ y $cos(\theta_l^N)$ y sus asimetrías *unfolded* fueron calculadas. Se encontró que el análisis mantuvo la sensibilidad a la presencia de los acoplamientos anómalos, tal como se cuantificó mediante pequeñas incertidumbres estadísticas en las asimetrías *unfolded*, lo cual permitirá realizar un test de consistencia del Modelo Estándar utilizando dichos observables. Sin embargo, el análisis mostró desviaciones de linealidad en la medida de asimetrías. Dichas desviaciones han sido cuantificadas y deberán ser corregidas o bien evitadas en futuros análisis del vértice Wtb con datos reales a 13 TeV. Para ello, las posibles lineas de evolución futuras son discutidas.

Acknowledgements

First and foremost, I would like to thank my fantastic project supervisor Susana Cabrera, for her continual guidance, enthusiasm, excellent knowledge and her remarkably kind and caring character. I have learnt so much from my time working at the IFIC and have ultimately enjoyed my project and the subsequent collaboration with others greatly. I could not imagine having a better advisor and mentor for my thesis and for this I am extremely grateful. Muchisimas gracias.

Secondly, I would like to thank my personal tutor at the Universitat de Valencia, José Bordes, and our IFIC team leader Carmen Garcia for their help with the organisation of this project and the final presentation. Furthermore, I would like to thank my year abroad supervisor from Imperial College London, Peter Torok, for his support and visits to Valencia.

Tengo que decir un gran gracias a todos los que trabajan en la oficina del IFIC (Pablo, Marcos, Pepe y Kevin) por su inclusividad, amabilidad y ayuda con física, el idioma español y, a veces, el valenciano también. ¡Gràcies xics!

Finally, I have to express my sincere gratitude for the Erasmus program, all the amazing international friends I have made this year, and my supportive family at home. Without this exchange program, this project and all the adventures I have had in Spain would not be possible.

Contents

Preface

In this preface I will identify which parts of the work detailed in this thesis are mine and which parts were completed by others; those working in the same research group at the IFIC (Universitat de Valencia).

With regards to the current status of physics in this field, this work is a novel attempt to use a a multivariate method, in particular a Bayesian neural network, in the selection of a sample enriched in t-channel events that will be use to explore anomalous couplings in the Wtb vertex. It is also the first evaluation of the bias (the deviations from linearity) introduced by a kinematic selection based on a Bayesian neural network instead of a cut-based selection, in the measurement of angular asymmetries using unfolding methods.

The analysis was conducted using the distributed analysis computing infrastructure of the ATLAS collaboration based on GRID, accessed from the user interfaces machines at the IFIC (the Instituto de Física Corpuscular, part of the Universitat de Valencia) institute. Also, lxplus.cern.ch machines at CERN have been use to train and apply the Bayessian NN.

Chapter 5 describes the experimental work involving the application and training of the Bayesian neural network (NN). The simulated samples used are provided from the Monte Carlo simulation (MC) at the LHC. These samples were converted into the form of ROOT TTrees [1], the input form for the neural network, via a computer program developed by the research group. The neural network itself is provided from the NeuroBayes package. I was not involved with the writing of the NN program nor of the user interface, which is written in the language C++. I did however use and manipulate the user interface to select training variables, choose the NN parameters, conduct overtraining tests, select the input samples and run the Teacher and Expert parts of the program in order to train and apply the NN. With the help of previous codes written by the research group, I wrote the computer programs (the macros) used for post training analysis in Python. These include the programs used to calculate the values for S/B, significance, background rejection, signal efficiency, the performance plots and the O_{NN} and event yields tables. These computer programs incorporated the analysis module ROOT. The computer program used to apply the cut in O_{NN} to the samples and produce the stacked histograms of the kinematic variables was written by the research group but I made use of this program and incorporated new elements into the code - such as the addition of the random Gaussian variable used in the overtraining tests.

Chapter 6 describes the experimental work involving the study of the angular observables. This includes the unfolding procedure of the angular distributions to parton level and the sensitivity and linearity tests. I wrote the computer program to implement the unfolding procedure, from the RooUnfold [2] package, from scratch in python. This includes the generation of all the graphs and calculations of asymmetries. I also wrote the programs to conduct the linearity and sensitivity tests from scratch in python.

Chapter 1

Introduction

The Standard Model (SM) of particle physics, the current theory used to describe the structure and classification of particles and forces in the universe, is a hugely successful in explaining and predicting experimental results but nevertheless remains incomplete with certain inexplicable caveats. One of such caveats is the observed phenomenon of CP violation, a particular focus of this project. Searches for new sources of CP violation are of interest because they offer a potential explanation for the additional observed phenomenon of matter-antimatter asymmetry in the universe, a connected caveat of the SM. Top quark physics, in which the top quark is the heaviest known fundamental particle, offers an interesting and potential probe into such physics beyond the SM via analysis of the Wtb vertex in t-channel single top quark production and decay (a specific production mode described in chapter 2) [3]. The presence of anomalous couplings in the Wtb vertex, non SM couplings expressed explicitly in the Lagrangian of the vertex, would be indicative of and yield information about physics beyond the SM [4]. In particular, a complex value for the coupling g_R may imply that top quark decay has a CP violating component [5]. We are motivated to search for physics beyond the SM in order to further develop our understanding of the universe and build a more complete picture.

The main aim of this research project is to provide a selection criterion that can be applied to a data sample of proton-proton collisions produced by the LHC at centre of mass energies of both 13 and 8 TeV, collected by the ATLAS detector in order to enrich and optimise t-channel single top quark events. This is conducted via the use of a multivariate analysis technique, namely a Bayesian neural network (NN), and idealised MC simulated signal and event samples (detailed in chapter 5). The NN based method will provide a single discriminant, O_{NN}, that can be adjusted to optimize theoretical and experimental systematic uncertainties. The new selection should outperform the previous methods, i.e. cut based optimisations [6], and this is judged by calculations of both the ratio of signal to background and the statistical significance, which should be maximised.

Such a sample enhanced in signal (t-channel) events will subsequently allow for the improved search of anomalous couplings in the Wtb vertex through measurement of the top quark polarisation and W boson spin observables built from angular asymmetries amongst the top-quark decay products and the spectator forward quark from the t-channel production [3]. Since the optimised sample will be used to search for anomalous couplings, a further aim is to ensure the selection process does not bias their detection. Good sensitivity and linearity of response to the presence of anomalous detection is required. This will be assessed via comparisons of relevant angular asymmetries (detailed in chapter 6).

As a summary of the thesis, in chapters 2 and 3 the relevant theory is reviewed and an overview of the LHC and ATLAS detector is provided. Chapter 4 provides a description of the simulated samples and analysis tools used in the work. Chapters 5 and 6 contain the main experimental work conducted, of which, 5 details the training and application of a Bayesian neural network at $\sqrt{s} = 13$ TeV and 6 details the study of the angular observables at $\sqrt{s} = 8$ TeV. Finally, the conclusions and future work are presented in chapter 7.

Chapter 2

Theoretical Framework

2.1 The Standard Model

The Standard Model (SM) of Particle Physics is a "gauge-invariant" quantum field theory that classifies the known elementary particles and three of the four fundamental forces (Electromagnetic, Strong and Weak) that govern their interactions [7], [8]. It states that all matter is composed of fermions; particles with half integer spin. Fermions and their corresponding antiparticles are further sub-categorized into leptons and quarks [9]. Table 2.1 details the classification and properties of such fermions and their division into three generations, yielding 6 different types of leptons and quarks.

Matter Particles					
	Generation				
Fermion	1	2	3	Spin	Charge (units of e)
Lepton	e^-	μ^-	τ^-	1/2	-1
	ν_e	ν_μ	ν_τ	1/2	0
Quark	u	c	t	1/2	+2/3
	d	s	b	1/2	-1/3

TABLE 2.1: Classification and properties of fermions: the "matter particles" [10] [9].

The interactions between these matter particles, the fundamental forces, are mediated by the exchange of bosons; particles with integer spin. The properties of these forces and their corresponding bosons are displayed in Table 2.2. Note that gauge bosons are the bosons associated with a fundamental force and with spin 1 whilst scalar bosons are those with spin 0.

Force Carriers				
Boson Type	Boson	Force	Spin	Charge
Gauge Boson	photon, γ	Electromagnetic	1	0
	gluon, g	Strong	1	0
	$W^\pm$, Z	Weak	1	$\pm$1,0
	?	*Gravity?*	*?*	*?*
Scalar Boson	Higgs, H		0	0

TABLE 2.2: The fundamental forces and Bosons [9],[10].

Since its development in the 1970s, the SM has been largely successful in explaining experimental results and predicting new physical phenomena, such as the discovery of the Higgs Boson in 2012 by the ATLAS and CMS collaborations at the LHC [11], [12], [13]. Nevertheless the SM remains incomplete [8]. As alluded to in Table 2.2, the fundamental force of gravity is yet to be incorporated into the SM [14], but this is just one of the many problems and unanswered questions faced by the SM:

- **Gravity:** The SM does not include the fundamental force of gravity nor explain why it is weaker than the other forces. Gravity has not yet been "unified" nor a consistent quantum theory of gravity defined [15].
- **CP Violation**: CP, a discrete symmetry given by the product of charge conjugation (C) and parity (P), is expected to be a conserved tranformation within the SM when in reality it is observed to be violated for many weak interactions [16]. Within the scope of this project, CP violation can be parametrised by the presence of anomalous couplings in the Wtb vertices of top quark production and decay [17].
- **Matter-Antimatter asymmetry:** The Big Bang should have created equal amounts of matter and antimatter but there is an unexplained dominance of matter over antimatter observed in the universe. CP violation offers one one explanation for the asymmetry in baryonic matter, however, there are not enough current observed sources of CP violation to account for the total matter-antimatter asymmetry. New sources of CP violation are needed [18], [19].
- **Neutrino Oscillations:** the observed quantum-mechanical phenomenon of neutrino oscillations, in which neutrinos spontaneously change flavour in noticeable oscillations , implies Neutrinos have mass, which is contrary to the SM [20].
- **Dark Matter:** Dark matter is the name given to the undetectable matter in the universe that causes galaxies to rotate at speeds not possible without their gravitational presence. The full nature of dark matter is not yet known [8].
- **3 Generations of Particles:** Matter exists in 3 generations, seemingly only differentiated by mass. It is questioned as to whether there is a deeper significance to this arrangement. [15]

The new theoretical physics needed to explain the gaps in the SM is referred to as Physics Beyond the Standard Model (BSM) [14]. In this project, we are motivated to investigate top quark events (specifically, anomalous couplings in the Wtb vertex) in a hope to probe physics BSM and perhaps one day build a more complete picture of the universe.

2.2 Top Quark Physics

Discovered in the Fermilab Tevatron Collider in 1995, the top quark is the heaviest quark (even after the discovery of the Higgs boson m_H =125-126 GeV [13],[12]), with a measured mass of $m_{top} = 172.44 \pm 0.13$ GeV [21], within the SM and hence is of substantial interest and the focus of this research [22]. The top quark forms an isospin doublet with the bottom quark (which is much lighter) and is also strongly coupled to the Higgs boson [9],[23]. Measurements of top quark phenomena offer an interesting probe to understanding the electroweak sector [24] and physics BSM [6].

2.2.1 Top Quark Production

At the LHC, top quarks are produced in proton-proton collisions via two main mechanisms: single top production and top quark pair production [25].

Top quark pair production ($t\bar{t}$) involves the production of top-antitop pairs via the strong interaction. Figure 2.1 illustrates the various sub-processes that produce top quark pairs, of which the most dominant process at the LHC is gluon fusion: $gg \to t\bar{t}$ [23].

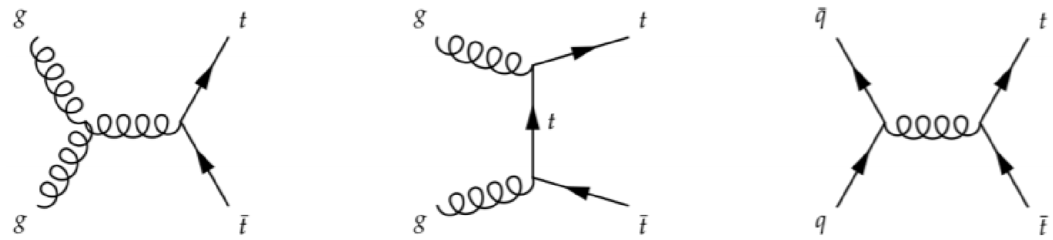

FIGURE 2.1: Leading order[1] Feynman diagrams of the top quark pair production modes [9]

Single top quark production involves the production of a single top quark via the weak interaction. It occurs through three different production channels; illustrated in Figure 2.2 and detailed below:

(a) ***t*-channel**: a light quark (q), and a bottom quark (b) interact via an electroweak scattering process, exchanging between them a virtual space-like W boson. The light quark q is a light flavoured u or d̄ quark, whilst $q^{'}$ is a light flavoured d or ū quark, named the "Spectator Quark" [6]. This channel has the largest cross-section and is hence the dominant production mode of single top quarks [9].

(b) ***s*-channel**: a top quark is produced by quark-antiquark annihilation, with the formation of a virtual time-like W boson that then yields the t and b̄ quarks. At the LHC, the colliding protons supply the initial quarks q, the valence constituent quarks of the protons, and antiquarks $\bar{q}$, the proton "sea quarks". The need for interaction between valence and sea quarks results in a lower production rate via this mode [27],[28], [9].

(c) ***W t*-channel**: also known as the W-associated channel, involves the production of a top quark alongside a real W boson. Also has a low production rate due to the presence of heavy final particles [9].

Overall, top quark pair production has a higher cross section than single top quark production and is the dominant process in proton-proton collisions due to the fact that the weak coupling is suppressed in comparison to the strong interaction [23]. In this research, the signal events are the single top quarks produced via the t-channel

[1] A leading order Feynman diagram is the most simple Feynman diagram that can describe an interaction. This comes from QCD perturbation theory [26]

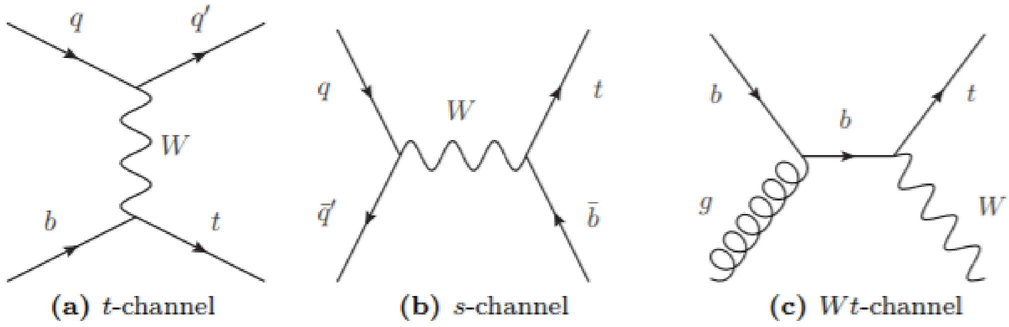

FIGURE 2.2: Leading order Feynman diagrams of the single top production channels [9]

mode. The top quark pair production modes and other single top channels serve as background events that we wish to reduce as much as possible and to separate from signal events. The current theoretical predictions for t-channel single top quark cross sections (in the combined case of both top and antitop production) calculated for a top quark mass of 172.5 GeV at next-to-leading order (NLO) in QCD at 13 TeV and 8 TeV are $\sigma_{t-chan,t+\bar{t}} = 216.99^{+6.62}_{-4.64}$ pb and $\sigma_{t-chan,t+\bar{t}} = 84.69^{+2.56}_{-4.64}$ pb respectively [29], [30].

2.2.2 Top Quark Decay

The large mass of the top quark leads to a rapid decay via electroweak interactions, with an average lifetime of just $\tau_t \sim 10^{-25}$s [9]. This decay occurs before hadronisation can takes place (a phenomenon unique to only the top quark), allowing the opportunity to study a quasi free quark via its decay products [31], [9].The spin information of the top quark is accessed from its decay products [3]. The top quark decays weakly and almost exclusively (100% branching ratio for $t \rightarrow Wb$ as opposed to decay to other quarks) to a W boson and b quark. This decay vertex is referred to as the Wtb vertex [31]. The W boson produced then decays to either hadronically ($W \rightarrow q\bar{q}$) or leptonically ($W \rightarrow l\nu$), with a relative branching ratio of: $\frac{BR(W \rightarrow q\bar{q})}{BR(W \rightarrow l\nu)} = \frac{6}{3}$ [6]. In this research, only the leptonic decay, where the lepton is specifically an electron as opposed to other leptons, is considered. Figure 2.3 depicts the Feynman diagram of the whole decay and Equation 2.1 describes the whole single top quark t-channel process of production and decay.

$$pp \rightarrow qb \rightarrow q^{'} + t(\rightarrow W(\rightarrow l\nu) + b) \tag{2.1}$$

The final state topology of top quark decay is characterised by the b-jet from the fragmentation and hadronisation of the b quark from top decay, a forward light jet from the fragmentation and hadronisation of the spectator light quark $q^{'}$, a charged lepton from the leptonic decay of the W boson and the missing transverse energy (E_T^{miss})

[2]Note that in general, the charges/signs of the particles are omitted from the diagrams and equations for simplicity. This is because in single top quark t-channel production, the single top quark produced can be either a top or anti-top quark and depending on which, the reactants and decay products will vary. For example, in top decay, t $\rightarrow W^+(\rightarrow l^+\nu) + b$ and $\bar{\text{t}} \rightarrow W^-(\rightarrow l^-\bar{\nu}) + \bar{b}$ are both described within equation 2.1.

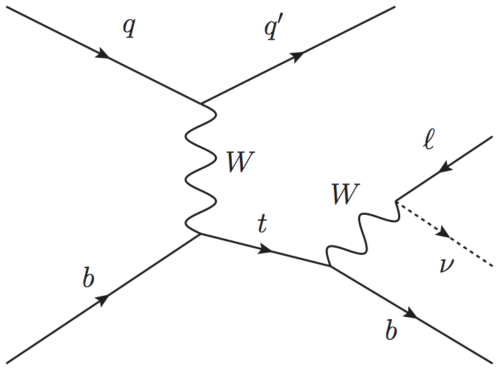

FIGURE 2.3: Leading order Feynman diagram of top quark production and decay via Wtb vertex, including leptonic W boson decay mode [2][32].

that accounts for the escaping neutrino. This is the experimental signature that allows for detection of top quarks produced via the t-channel production mechanism at the LHC [23].

2.3 Polarisation Observables and Asymmetries in the *Wtb* Vertex

This section follows the theoretical framework adopted in reference [3].

2.3.1 Polarisation and Spin Observables

In particle physics, polarisation refers to the extent to which a particle's spin is aligned along a particular direction [33]. The t-channel single top quarks are produced with a strong degree of polarisation along the direction of the momentum of the light spectator quark, whose direction then defines the top quark spin axis [34]. The real W-boson produced from top quark decay is also polarised into one of three helicity states: right, left or zero helicity [9]. Each helicity state has a corresponding partial decay width ($\Gamma_L, \Gamma_0, \Gamma_R$) in which the total decay width is defined as their summation: $\Gamma(t \to Wb) = \Gamma_L + \Gamma_0 + \Gamma_R$ [35]. The degree of polarisation of both the top quark and W boson can be extracted from the angular distributions of their respective decay products. [5].

In the rest frame of the top quark, the general form for the angular distribution of any of its decay products X (= $W, l, \nu, b,$) is given by:

$$\frac{1}{\Gamma}\frac{d\Gamma}{d(cos\theta_X)} = \frac{1}{2}(1 + \alpha_X P cos\theta_X) \tag{2.2}$$

where Γ is the total decay width of the top quark, θ_X is the angle between the top quark spin axis and the direction of motion of X, α_X is the spin analysing power associated with X and P is the degree of polarisation of the top quark [5], [25]. Note that normally $cos(\theta_l)$, the angular observable of the charged lepton (l), is specifically used because it is the most sensitive decay product to top quark polarisation [36],

[6]. The current SM theoretical predictions, calculated at QCD NLO, for the top (anti-top) quark polarisation and lepton spin analysing power are $P_z = 0.91(-0.86)$ [37] and $\alpha_{l\pm} = \pm 0.998$ [38] respectively.

From the spin density matrix for polarised top quarks (described in reference [35]) the W boson helicity components (left, right, zero) can be parametrised by the expectation values of the following 6 independent spin observables: $\langle S_{1,2,3}\rangle$, $\langle A_{1,2}\rangle$ and $\langle T_0\rangle$. This subsequently yields the following equation for the full differential decay width of the W boson:

$$\begin{aligned}\frac{1}{\Gamma}\frac{d\Gamma}{d(\cos\theta_l^*)d\phi_l^*} =& \frac{3}{8\pi}\Big\{\frac{3}{2}+\frac{1}{\sqrt{6}}\langle T_0\rangle(3\cos^2\theta_l^*-1)+\langle S_3\rangle\cos\theta_l^* \\ &+\langle S_1\rangle\cos\phi_l^*\sin\theta_l^*+\langle S_2\rangle\sin\phi_l^*\sin\theta_l^* \\ &-\langle A_1\rangle\cos\phi_l^*\sin 2\theta_l^*-\langle A_2\rangle\sin\phi_l^*\sin 2\theta_l^*\Big\}\end{aligned} \tag{2.3}$$

Here, θ_l^* and ϕ_l^* are the polar and azimuthal angles of the charged lepton momentum in the W boson rest frame [6], [25]. Note that the "$*$" superscript refers to measurements in the W boson rest frame.

Figure 2.4 summarises the entire coordinate system and angles that define the angular observables. Here, the z axis is defined by the momentum $\vec{q}$ of the W boson in the rest frame of the top quark. The x-z plane is defined by the the top quark spin direction $\vec{s}_t$ along the direction of the spectator quark momentum in the top quark rest frame. θ_l^* and ϕ_l^* are the polar and azimuthal angles of the charged lepton's momentum $\vec{p}_l$ in the W boson rest frame. The other axes, $\vec{N}$ (-y) and $\vec{T}$ (x) are defined by the following equations: $\vec{N} = \vec{s}_t \times \vec{q}$, $\vec{T} = \vec{q} \times \vec{N}$. The azimuthal angles ϕ_N^* and ϕ_N^* are those of the charged lepton in the W boson rest frame, defined relative to the $\vec{N}$ and $\vec{T}$ axes ($\phi_T^* = \phi_l^*$). The final angles θ_l^N and θ_l^T, not shown in the figure, are the angles between $\vec{p}_l$ and the axes $\vec{N}$ and $\vec{T}$ respectively [25].

It is also possible to determine the values of the top quark polarisation and the W boson spin observables via calculation of the appropriate asymmetry. These asymmetries are derived by integrating the angular distributions in equations 2.2 and 2.3 [25]. Most of the polarisation and spin observables are connected to a forward-backward asymmetry, defined by the following equation for a particular angular observable $\cos(\theta)$:

$$A_{FB} = \frac{N(cos(\theta) > 0) - N(cos(\theta) < 0)}{N(cos(\theta) > 0) + N(cos(\theta) < 0)} \tag{2.4}$$

Note that N is the number of events [25].

From the wide set of observables analysed in reference [3], due to time constraints, this research focuses solely on two angular observables, $cos\theta_l$ and $cos\theta_l^N$. To summarise, θ_l is the relative angle between the top quark spin axis and the lepton momentum in top quark rest frame, and θ_l^N is the relative angle between the lepton momentum in rest frame of W boson and and the axis $\vec{N}$. These form the focus of the analysis and the angular observables used in the unfolding procedure described in chapter 6. They were chosen because, firstly, A_{FB}^N, the forward-backward asymmetry in the angular distribution $cos(\theta_l^N)$ is very sensitive to the coupling $\mathrm{Im}(g_R)$, in

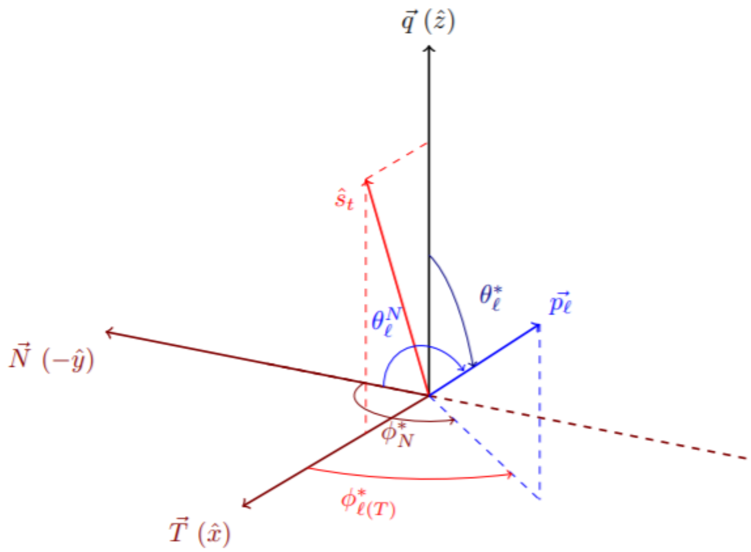

FIGURE 2.4: The coordinate system and angles used to define the W boson angular distributions and spin observables [25].

which a non-zero complex value for this coupling would imply that the top-quark decay has a CP-violating component (see next section for more details). Secondly, A^L_{FB}, the forward-backward asymmetry in the angular distribution $cos(\theta_l)$ is used to measure the top-quark polarization. The relationships between the associated asymmetries of these angles and the polarisation/spin observables are detailed in Table 2.3.

Angular Observable	Asymmetry	Polarisation/Spin Observable	SM Value
$cos(\theta_l)$	A^l_{FB}	$\frac{1}{2}\alpha_l P$	0.45
$cos(\theta^N_l)$	A^N_{FB}	$-\frac{3}{4}\langle S_2 \rangle$	0

TABLE 2.3: The chosen angular observables, their associated asymmetries and their relation to the polarisation/spin observables. The theoretical expected value for the SM is also provided, calculated via QCD NLO and NNLO (next to next to leading order) predictions [25]. Columns 2, 3 and 4 are equal to each other.

2.3.2 Anomalous Couplings in the *Wtb* Vertex

The Wtb vertex refers to the Feynman diagram vertex involving the W boson and top and bottom quarks, found in both t-channel single top quark production, Fig 2.2 (a), and top quark decay, Figure 2.3. The lagrangian for the Wtb vertex, written in effective operator formalism, is:

$$\mathcal{L}_{Wtb} = -\frac{g}{\sqrt{2}}\bar{b}\gamma^{\mu}(V_L P_L + V_R P_R)tW^-_{\mu} - \frac{g}{\sqrt{2}}\bar{b}\frac{i\sigma^{\mu\nu}q_{\nu}}{m_W}(g_L P_L + g_R P_R)tW^-_{\mu} + h.c. \quad (2.5)$$

where g is the weak coupling constant, m_W and q_v are the mass and four-momentum of the W boson, $P_{L,R} = (1 \mp \gamma^5)/2$, γ is the Dirac matrix, $\sigma^{\mu\nu} = i[\gamma^\mu, \gamma^\nu]/2$ and $h.c.$ refers to the higher order corrections [3], [9].

The terms $g_{L,R}$ and $V_{L,R}$ are the left and right handed vector and tensor couplings respectively - the so-called anomalous couplings [9], [39]. In the SM, $V_R = g_L = g_R = 0$ and $V_L = V_{tb} \simeq 1$, where V_{tb} is a quark mixing element in the Cabibbo–Kobayashi –Maskawa (CKM) matrix [39], [40]. Measured deviations from these expected values would indicate the presence of physics BSM. Furthermore, complex values for these couplings, specifically a non 0 value for Im g_R, may imply CP violation [9], [39]. In this research, the real and imaginary values of $g_{L,R}$ are the focus of the analysis.

Considering the whole analysis chain, values for the couplings can be determined via measurements of the asymmetries (and hence polarisation and W boson spin observables) from real LHC data using the TopFit [41] code, which contains the theoretical relationships between the couplings and the asymmetries. Note that these relationships are not detailed here or calculated within the scope of this thesis, but calculations of the asymmetries for the two chosen angular observables are completed in chapter 6.

With regards to the current status of ongoing research, the ATLAS and CMS collaborations at the LHC have set limits on the values of such couplings. Focusing on the anomalous coupling g_R, these limits at a 95% confidence interval for LHC Run I (see chapter 3) are reviewed as follows:

- From reference [3], limits for $\mathrm{Im}(g_R) \in [-0.18, 0.06]$ are set by ATLAS by extraction from the measured A^l_{FB} and A^N_{FB} asymmetries from single top t-channel data at $\sqrt{s} = 8$ TeV, whilst assuming $V_L = 1$ and all other couplings $g_L = V_R = \mathrm{Re}(g_R) = 0$.
- From reference [42], limits for the coupling $\mathrm{Re}(g_R) \in [-0.02, 0.06], [0.74, 0.78]$ are set by ATLAS via measurement of the W boson helicity components from single lepton and dilepton $t\bar{t}$ events at $\sqrt{s} = 8$ TeV.
- From reference [43], limits for $g_R \in [-0.049, 0.048]$ are set by CMS at energies of $\sqrt{s} = 7$ and 8 TeV.

2.4 Statistics of Data and Data Selection Optimisation

Bayes theorem is an approach to statistical analysis that describes the probability of an event occurring based on prior knowledge. The paradigm is expressed mathematically, in its most simple form, according to the following equation:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.6}$$

where P(A | B) is the conditional probability of event A occurring given that B has already occurred, similarly for P(B | A), and P(A) and P(B) are the independent probabilities of events A and B respectively occurring in isolation [44]. Bayesian statistics are important in high energy physics, and generally, because they allow the exclusion of unphysical outcomes by the incorporation of prior knowledge. This approach is implicated in this project via the use of a Bayesian neural network as a multivariate analysis to optimise data selection - *see section 5*.

Throughout this project, two principal indicators are used to assess the extent of data optimisation; the ratio of signal to background, S/B, and Significance, where S and B refer to the number of signal and background events respectively. S/B offers a simple way to quantify the extent to which the signal events have been prioritised over background events. However, it says nothing about the total number of events, $N = S + B$, that are used. The statistical significance of a data set is defined by the following equation:

$$Significance = \frac{S}{\sqrt{S+B}} \tag{2.7}$$

The relationship can be understood as the size of the signal relative to 1 standard deviation ($\sqrt{S+B}$). Significance scales with the amount of statistics used. Cutting the data too tightly to maximise S/B will lead to a decrease in significance [45]. Therefore, significance serves as an important parameter to be used in combination with S/B in order to create an optimisation that seeks to maximise S/B but not so much that it results in a small final sample.

Chapter 3

The LHC and the ATLAS Experiment

3.1 CERN and the LHC

Established in 1954, the European Organisation for Nuclear Research, CERN, is the research organisation that now boasts the largest laboratory and experiments in particle physics in the world. Situated in Geneva and stretching across the French-Swiss border, CERN is an international collaboration of science, with 22 member states and over 10,000 scientists [46]. At CERN, physicists seek to probe at the fundamental structure of the universe by investigating the interactions between subatomic particles. This is done through use of large purpose-built particle accelerators and detectors, which generate millions of gigabytes of data to be analysed every year [46]. An overview plan of the particle accelerators found at CERN is illustrated in Figure 3.1.

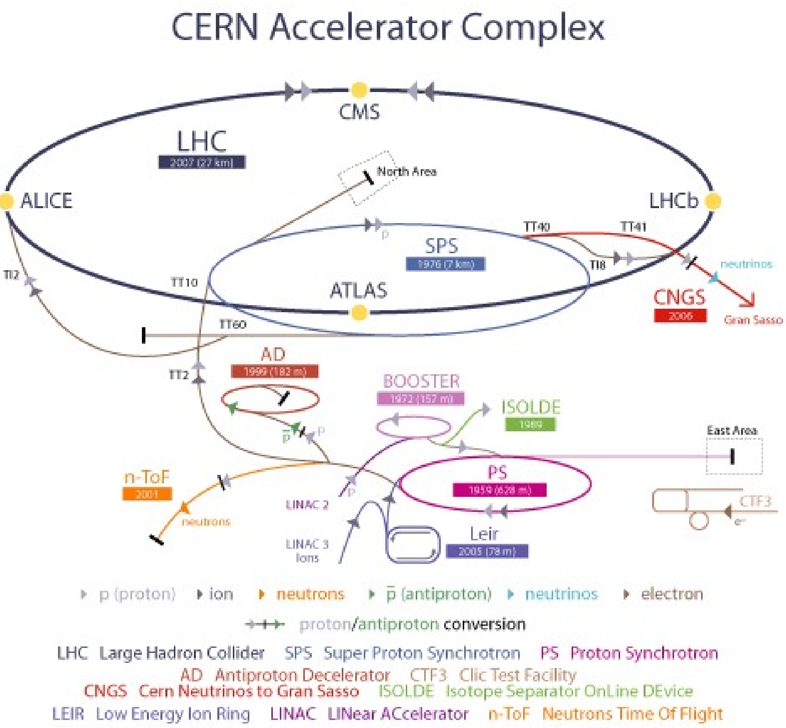

FIGURE 3.1: CERN accelerator complex [47]

The Large Hadron Collider (LHC) was started up in 2008 and is the largest and most powerful of the particle accelerators at CERN and the world to date [48]. It is an underground circular synchotron with a circumference of 27 km, that accelerates high energy hadron beams (primarily protons) to near light speed before making them collide at detector sites [48]. The 2 beams are guided around the accelerator separately in opposite directions through use of strong magnetic fields supplied by superconducting electromagnets located in the ring. Dipoles curve the trajectory of the protons, quadrupoles focus the beams near the colliding points and the RF cavities (Radio Frecuency) accelerate the protons up to the nominal energy and restore the energy losses [49], [50]. The proton-proton collisions at the LHC, pp, involve bunches of up to 10^{11} protons that collide 40 million times per second at various centre of mass energies $\sqrt{s}$ [51].

Along the circumference of the LHC, 4 different interaction points are fixed in which different detectors are placed. Figure 3.1 highlights these 4 detectors; ATLAS, CMS, ALICE and LHCb. ATLAS and CMS are the largest of these purpose built experiments - ATLAS being the detector and collaboration that this research is based in. [31].

3.1.1 Accelerator Performance and Parameters

There are several features and parameters essential for performance of particle accelerators such as the LHC. The energy provided by the colliding beams of a particle accelerator is indicated by the Centre of Mass Energy $\sqrt{s}$ [52]. In general, colliding beams (as opposed to a beam striking a fixed target) are essential to reach the high $\sqrt{s}$ values needed for particle physics [53]. The LHC has operated at different centre of mass energies and collision rates, detailed in Table 3.1 Here, the increase in these energies throughout the years is seen as we progress from Run 1 (Data collected from 2010-2012) to Run 2 (2015 onwards) [54].

In particle physics, a cross section, σ is defined as the probability of a reaction occurring. Events with low probabilities have small cross sections and are generally of more interest. An important parameter for a particle accelerator operation, linked to cross section, is the Luminosity; a quantity that indicates the number of collisions within an accelerator - see equation 3.1. Here we observe that the number of events per second, R, is proportional to the cross section of an event, with the instantaneous luminosity, $\mathcal{L}$, acting as the proportionality constant [53] [6] [9].

$$R = \mathcal{L}\sigma \tag{3.1}$$

For two beams colliding, (collider) instantaneous luminosity is defined more explicitly as:

$$\mathcal{L} = f \cdot \frac{N_1 N_2 N_b}{4\pi\sigma_x\sigma_y} \cdot F(\theta_c, \sigma_x, \sigma_z) \tag{3.2}$$

where N_1 and N_2 are the number of particles in each beam's bunch (a bunch being a packet of protons), N_b is the number of bunches, f is the bunch revolution frequency, σ_x and σ_y are the RMS transverse beam sizes in horizontal and vertical directions, and F is the factor that accounts for luminosity reduction due to the crossing angle

θ_c [55],[53], [56], [6], [9]. Luminosity is hence a function of the beam parameters of an accelerator, such as the LHC, and is subsequently the most important quantity in collider physics besides energy [55]. Additionally, $\mathcal{L}_{int}$, Total Integrated Luminosity, is the integral of $\mathcal{L}$ with respect to time, over an entire year. It is measured in units of inverse femtobarns fb^{-1}, where 1 fb^{-1} corresponds to 10^{12} collisions and 1 barn, b, is 10^{-28}m^2 [57].

Table 3.1 details the values of such parameters for the LHC at different years. Figure 3.2 illustrates the evolution of ATLAS delivered luminosity and energy over time.

Run:		Run I			Run II	
Year:	Design	2010	2011	2012	2015	2016
Beam Energy (TeV)	7.0	3.5	3.5	4.0	6.5	6.5
$\sqrt{s}$ (TeV)	14	7	7	8	13	13
Protons/Bunch (10^{11})	1.15	1.0	1.3	1.5	1.1	1.1
Bunch Spacing (ns)	25	150	50	50	25	25
Max Peak $\mathcal{L}$ (10^{34}cm^{-2}s^{-1})	1.0	0.021	0.35	0.77	0.51	1.01
$\mathcal{L}_{int}$ (fb^{-1})	-	0.048	5.5	22.8	4.2	38.5

TABLE 3.1: Different performance parameters of the LHC at different years. [58], [59]. Note that the centre of mass energy $\sqrt{s}$ is twice that of the beam energy because there are 2 beams colliding.

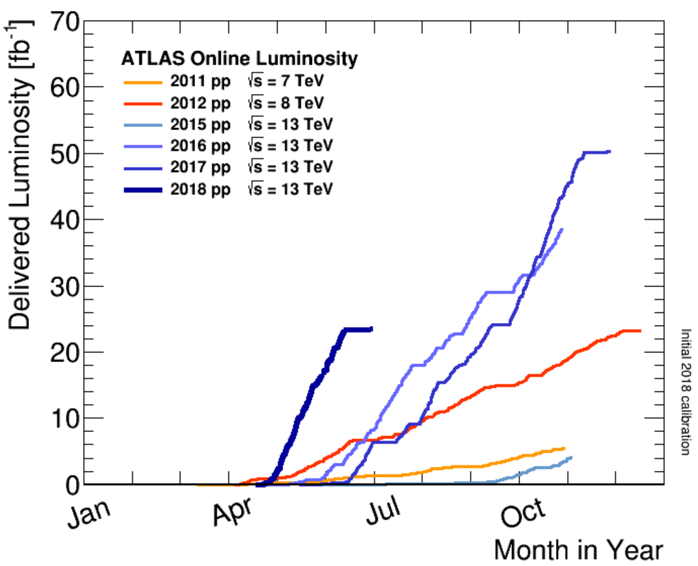

FIGURE 3.2: Delivered luminosity to ATLAS versus time for 2011-2018 during stable beams and for high energy $p-p$ collisions [60]

3.2 The ATLAS Detector

The ATLAS (A Toroidal LHC ApparatuS) detector is a multipurpose particle detector (and experiment) with cylindrical geometry and forward-backward symmetry with respect to the interaction point [51]. As illustrated in Figure 3.3, the ATLAS detector is comprised of 4 major components; the Inner Detector, the Calorimeter, the Muon Spectrometer and the Magnet System. These operate alongside the Trigger and Data Acquisition System and the Computing System [61].

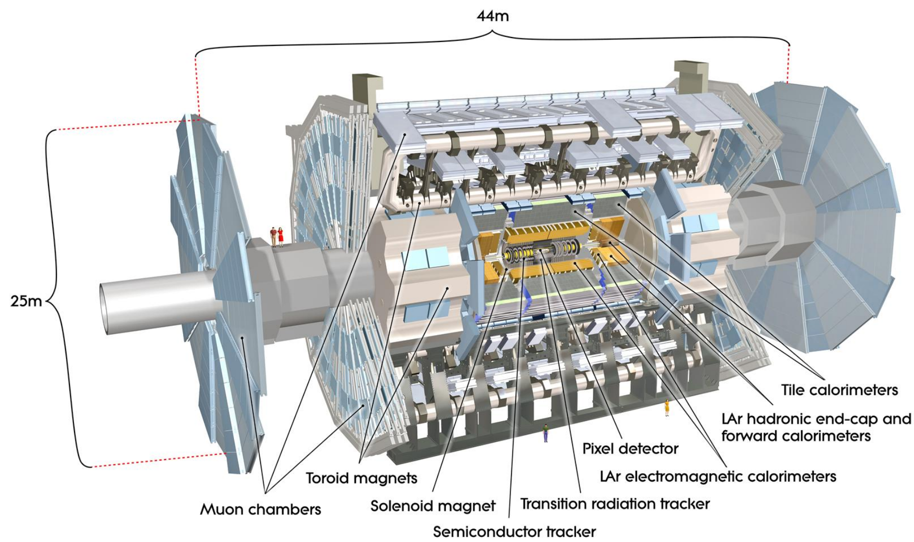

FIGURE 3.3: The ATLAS detector[61]

The coordinate system of ATLAS is defined by setting the origin at the centre of the detector at the interaction point. It is a right-handed coordinate system with the z axis is given by the beam axis and the $x-y$ plane, the transverse plane, running perpendicular to the beam line [6]. Note that the momenta of particles measured in the transverse frame are referred to as transverse momenta, p_T [62]. The azimuthal angle ϕ is measured from the x axis about the beam whilst the polar angle θ is measured from the z axis, as depicted in Figure 3.4 [62]. From the polar angle, the spatial coordinate Pseudorapidity, η, is defined:

$$\eta = -ln\big[tan\Big(\frac{\theta}{2}\Big)\big]. \tag{3.3}$$

The half of the detector with positive z-values is named Side A, the negative half Side C [62].

An axial magnetic field of 2T surrounds the Inner Detector to provide charged particle "tracking" in the pseudorapidity range of $|\eta| < 2.5$ [3]. The combination of a high-resolution silicon pixel detector, a silicon microstrip tracker, and a straw-tube transition radiation tracker allow the measurement of particle momenta and vertices and the identification of particle type [51]. The energies of these particles are measured by the calorimeters. Lead-Liquid Argon (LAr) sampling calorimeters provide electromagnetic energy measurements in the range $|\eta| < 3.2$ whilst hadronic energy measurements in the range $|\eta| < 1.7$ are provided by steel/scintillator-tile calorimeters [51]. The total coverage is extended to $|\eta| = 4.9$ by the presence of LAr calorimeters in the forward region that measure both electromagnetic and hadronic energies [51], [3]. The outermost part of the ATLAS detector, the Muon Spectrometer, surrounds the calorimeter and measures and identifies the muons that pass through the inner regions (Inner Detector and Calorimeter) undetected [61].

Due to the large collision rate at the ATLAS design luminosity of 10^{34} cm^{-2}s^{-1} (roughly 10^9 collisions per second, corresponding to a data volume of more than 60 million megabytes per second [61]) a three-level Trigger System is employed to

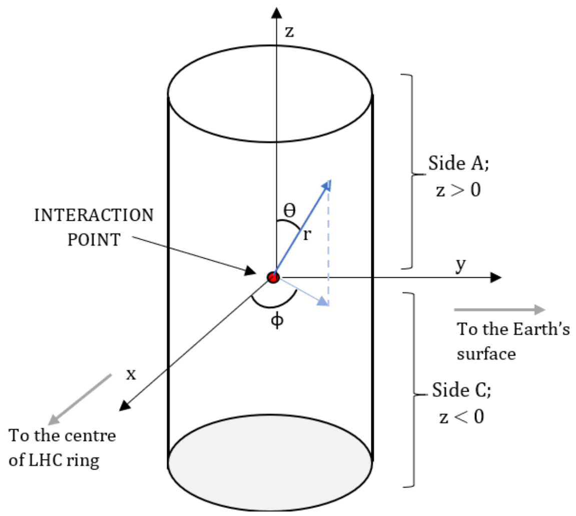

FIGURE 3.4: The coordinate system of the ATLAS detector

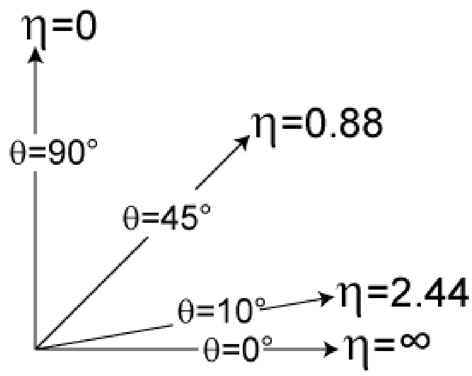

FIGURE 3.5: The relationship between θ and pseudorapidity [63]

reduce the number of events selected and stored. Only events with certain distinguishing physical characteristics are chosen, reducing the stored event rate to a more manageable 400 Hz [3].

Finally, the Grid is the name given to the computing system employed at ATLAS. It consists of a system of 4 "tiers", in which each tier is built from several computer centres and provides a particular set of services [64].

Chapter 4

Simulated Samples and Event Reconstruction

4.1 Monte Carlo Simulation

In this research, the samples used to create the optimised selection criteria are the theoretical simulated event samples of signal and background events of proton-proton collisions from Monte Carlo (MC). The MC samples are produced by accurate models based on the predictions of event characteristics from theory. Within the MC data sets, the identities of events (i.e. signal or background) are known and thus an effective optimisation can be created that can then be applied to real LHC data in the future (for which the identities of measured events will be unknown).

There are 3 steps in MC simulation: Generation - the production of the event from the matrix element of the initial nucleon-nucleon state using an event generator interfaced to a parton shower and hadronisation model; Detector Simulation - the modelling of interaction and propagation of particles through the detector; and Digitization - the conversion of outputs to raw data objects [9] [65]. After this, the same ATLAS trigger and reconstruction algorithms that would occur for real LHC data can be applied to the output, resulting in a data sample of total integrated luminosity $\mathcal{L}_{int}$=66.1 pb^{-1} (20.2 fb^{-1}) at $\sqrt{s}$ = 13 TeV (8 TeV) [9], [3], [6].

The simulation of the ATLAS detector itself is based on the GEANT4 framework [66], [3]. The signal events, t-channel single top quark events, are simulated via the MC Next to Leading Order (NLO) POWHEG-BOX generator [67] and the MC PYTHIA8 generator [68]. The background processes, detailed and classified in Table 4.1 are those that will imitate the experimental signature of the signal events [6]. These must be properly dealt with (i.e. eliminated) in order to make meaningful analyses. The backgrounds of Wt-channel, s-channel and $t\bar{t}$ are also simulated via the MC NLO POWERHEG-BOX and PYTHIA8 generators [3]. The MC Leading Order (LO) SHERPA generator [69] is used to generate the backgrounds from W and Z boson production in association with jets and diboson processes (WW,WZ and ZZ). Multi-jet background production is instead generated using a data driven matrix method (Not MC) [65].

The above deals with MC data for the SM, that is to say, MC samples without anomalous couplings. Such samples are used for the training of the NN and creation of optimised selection criteria. The second half of the analysis, an investigation into the angular observables and unfolding to confirm the potential presence of new

Process Group	Description	Generator
Signal		
t-channel	t-channel single top quark events.	NLO POWERHEG-BOX, PYTHIA8
Background		
Wt-channel and s-channel	Production of single top quarks via other modes, including strong and electroweak interactions.	NLO POWERHEG-BOX, PYTHIA8
$t\bar{t}$	Production of top quarks via top quark pair production ($t\bar{t}$)	NLO POWERHEG-BOX, PYTHIA8
W+jets	Production of real W bosons	LO SHERPA
Z+jets	Production of single Z bosons or Dibosons via electroweak interaction	LO SHERPA
Multijet	QCD Multijet production in which one of the jets will be a lepton, leading to the mismeasurement of the missing transverse energy E_T^{miss}	Data Driven Matrix Method (Non MC)

TABLE 4.1: Signal and background processes for t-channel single top quark events generated by MC generators [6]

physics will not bias the optimisation chosen, deals with MC events with anomalous couplings. MC events with anomalous couplings (in both production and decay vertices) are generated via the LO PROTOS generator [3]. These samples (with the exception of data ID 117788) are non SM samples. The details and coupling configuration of such simulated samples used in this analysis (specifically Chapter 6) are shown in Table 4.2. The coupling constants come from the Wtb vertex lagrangian equation, Equation 2.5, and a range of different coupling configurations have been used so that general trends can be reliably explored. The emphasis of the configuration of the couplings is placed on a complex g_R as this is the coupling linked to CP violation. Note that only signal t-channel events are used, not backgrounds. This is because in this part of the analysis, the aim was only to evaluate the sensitivity and bias to anomalous couplings, not to evaluate the optimisation via S/B etc.

Two centre of mass energies have been used throughout this research: $\sqrt{s}$ = 13 TeV (Run II, 2015 onwards) and $\sqrt{s}$ = 8 TeV (Run I, 2010-2012). The NN was initially set-up, its parameters fine-tuned and trained using MC data without anomalous couplings at 13 TeV. However, since the production of the 13 TeV MC samples with anomalous couplings through the whole official chain of the MC production of the ATLAS collaboration was far beyond the time-line of this research, the 8 TeV MC simulated samples have been considered instead. The NN was retrained using MC samples without anomalous couplings at 8 TeV, maintaining the NN set up parameters obtained from the initial analysis at 13 TeV, and then used to optimise MC samples at 8 TeV with anomalous couplings.

Sample	Dataset ID	Couplings			
		Re(V_L)	Im(g_R)	Re(g_R)	Re(V_R)
t-channel	117788	1.0	0.0	0.0	0.0
t-channel	117789	0.958	0.23	0.0	0.0
t-channel	117790	0.958	-0.23	0.0	0.0
t-channel	117791	0.993	0.094	0.0	0.0
t-channel	117792	0.993	-0.094	0.0	0.0
t-channel	110509	0.982	0.144	0.0	0.0
t-channel	110510	0.992	0.043	0.0	0.0
t-channel	110511	0.982	0.144	0.0	0.0
t-channel	110512	0.992	-0.043	0.0	0.0
t-channel	110502	1.165	0.0	0.18	0.0
t-channel	110503	1.065	0.0	0.07	0.0
t-channel	110504	0.925	0.0	-0.075	0.0
t-channel	110505	0.867	0.0	-0.131	0.0
t-channel	110505	0.867	0.0	-0.131	0.0
t-channel	110506	0.726	0.0	-0.26	0.0

TABLE 4.2: The signal MC samples with anomalous couplings (with the exception of ID 117788 which acts as the SM file) generated with the PROTOS generator.

4.2 The Baseline Selection

As discussed previously, the first "cut" employed with real LHC data is a 3-levelled trigger. For MC events, an equivalent software based trigger is applied, the first of which is a single-muon or single electron trigger [9]. In this research, only W-boson decay modes to an electron (instead of a muon or taon) are considered - the **electron channel** of top quark decay. Hence only the single-electron trigger is needed, which consists of a requirement on transverse energy of E_T > 60 GeV or > 24 GeV and the satisfaction of isolation criteria [70], [71].

After real events at ATLAS filter through the Trigger system and are recorded, this raw data is "reconstructed" via algorithms into meaningful objects; i.e. the signal and background objects [9]. Even after this, not all the events recorded are of physical interest for a certain analysis, and so a series of baseline cuts are applied to the data - the Preselection Region [9]. The same baseline cuts are applied to the MC simulated samples and take place before any further, more specific cuts can be applied later on. The preselection cuts involve a combination of even quality cuts and cuts to reduce the QCD multijet background. These are summarised in table 4.3 .

4.2.1 $m_{l,b}$ requirement

As well as the Preselection region, a further cut is applied to both the 8 and 13 TeV MC samples before the main analysis begins. A requirement in $m_{l,b}$, the invariant mass of the lepton (l) and b-jet (b), is imposed so that the so called "off-shell" region of top-quark decay beyond the kinematic limit of $m_{l,b}^2 = m_l^2 + m_w^2$, where m_w is the mass of the w-boson, is excluded [71]. This off-shell region was not well modelled within the MC generator since such effects were not included in the underlying matrix-element calculation and hence should be excluded [6], [71]. Additionally, the

Preselection Cut	Justification
Single lepton trigger	Software based trigger applied to select only W-boson decay mode to lepton (electron)
1 b-jet	Reconstruction requirement for the t-channel experimental signature
2 jets with $p_T > 30$ GeV and $\|\eta\| < 3.5$	Reconstruction requirement for the t-channel experimental signature
Exactly 1 lepton with $\|\eta\| < 2.47$ and $p_T(l) > 30$ GeV	Reconstruction requirement for the t-channel experimental signature
Primary vertex ≥ 5 tracks	Even quality cut.
No jets badly reconstructed	Even quality cut.
$E_T^{miss} > 30$ GeV	Reduce QCD multijet background events which are characterised by low E_T^{miss}
$m_T(W) > 50$ GeV	Reduce QCD multijet background events which are characterised by low $m_T(W)$
$p_T(l) > 40(\frac{\|\Delta\phi(j_1,l)-1\|}{\pi-1})$ GeV	Reduce QCD multijet background events.

TABLE 4.3: The preselection cuts employed. Note that $m_T(W)$ is the transverse mass of the W boson, $p_T(l)$ is the transverse momentum of the charged lepton (l), $\Delta\phi(j_1, l)$ is the azimuthal angle between l and the jet with the highest p_T (j_1) [70], [6].

uppercut in $m_{l,b}$ was carefully chosen to avoid introducing potential biases in measuring the angular distribution $cos(\theta_l^*)$, which has a clear dependence on $m_{l,b}$ [6]:

$$m_{l,b}^2 = \frac{m_{top}^2 + m_w^2}{2}[1 - cos(\theta_l^*)] \tag{4.1}$$

where θ_l^* is the angle between the charged-lepton momentum in the W-boson rest frame and the W-boson momentum in the top-quark rest frame, as described in Chapter 2. This angle is related to the W-boson helicity fractions F_R, F_L, F_0.

For the 13 and 8 TeV samples, cuts of $m_{l,b} < 140$ GeV and $m_{l,b} < 160$ GeV were employed respectively, as justified in the paper [71]. In figure 4.1, it is observed that the cuts chosen do not bias the angular distributions as they essentially only exclude background events, not signal (and hence also serve as an effective variable to maximise S/B).

4.2.2 Object Reconstruction

Object or event reconstruction involves the conversion of raw data from the LHC into objects with physical meaning. The same reconstruction procedures are applied to the MC samples. The objects that are reconstructed in this analysis are the electrons, the jets (which arise from the hadronisation of the quarks produced in t-channel production and decay) and the E_T^{miss} (which is indicative of the neutrino produced from W boson decay) [6]. These are the experimental signature characteristic of signal t-channel events.

Electrons candidates are reconstructed from clusters of energy deposited in the electromagnetic calorimeter with a well measured track, fulfilling strict requirements to momenta and pseudorapidity [3]. Jets are reconstructed from topological clusters

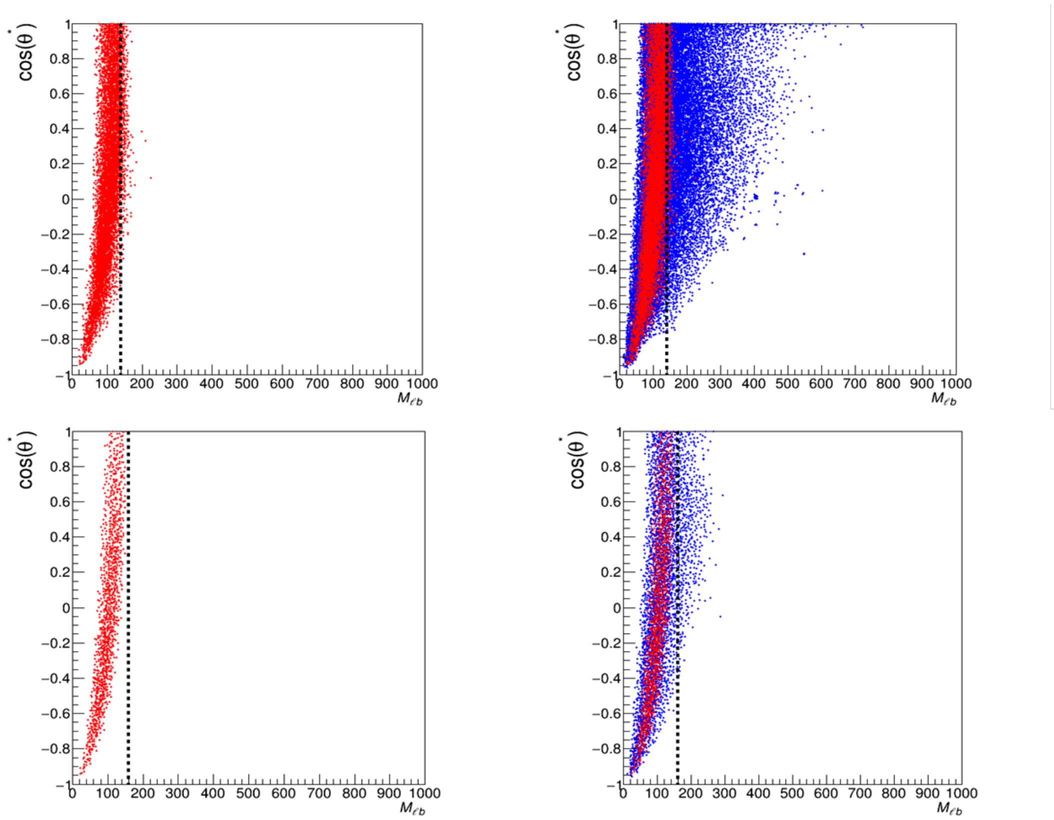

FIGURE 4.1: 2D scatter plots of $cos(\theta_l^*)$ against $m_{l,b}$ for events at 13 TeV (above) and 8 TeV (Below). Signal events are depicted in red and have been scaled by a factor of 0.1 for ease of viewing, background events are shown in blue with scaling of 0.05. The dotted line denotes the uppercut in $m_{l,b}$.

using an ANTI-k_t algorithm, a successive recombination algorithm which takes a radius parameter of 0.4 and inverts the parton shower containing the constituents of a jet according to some distance criteria weighted by relative transverse momenta [72], [6]. In the case of b-jets, a b-tagging algorithm, which also improves rejection of c-jets, based on neural networks or boosted decision trees is used. E_T^{miss} is reconstructed from the vector sum of energy in the transverse plane deposited in the calorimeter [3].

Chapter 5

Training and Application of a Bayesian Neural Network

The following chapter deals with the selection of a sample enriched in t-channel single top quark events produced at $\sqrt{s} = 13$ TeV using a Bayesian Neural Network (NN) as a multivariate (MVA) data analysis. MC simulated samples at $\sqrt{s} = 13$ TeV without anomalous couplings have been used.

The errors quoted hereafter are statistical errors that account for the MC limited statistics and their subsequent appropriate propagations. Systematic uncertainties will not be studied within the time scope of this master thesis, especially since real data is not used, although their estimation will be necessary in further optimizations of the whole analysis chain.

5.1 A Bayesian Neural Network

5.1.1 Machine Learning in High Energy Physics

Machine learning, the field of computer science that gives computers the ability to learn from data without being explicitly programmed [73], is used extensively within research domains such as high energy physics where there exists copious amounts of data and a need for efficient yet innovative analyses. With machine learning techniques, such as neural networks and decision trees, accurate models and predictions can be built without the need for the a priori knowledge or programming of the underlying causal links. The computer instead "learns" solely from the data and its correlations and develops its own conclusion/relationship via automated algorithms [74].

Artificial Neural Networks (NN) offer a multivariate approach inspired by the architecture of biological neural networks, such as the human brain, and are of the most commonly used and powerful machine learning methods. Thet consist of an interconnected network of nodes ("neurons"), individual processing units that receive information and send it forward to other nodes provided a threshold activation function is surpassed. The the term "feedforward" refers to a type of NN, such as the package used in this project, in which the connections between the nodes do not form a cycle but are instead arranged into several layers: input, hidden and output [75] [76] [74]. The input layer receives the input data and passes it onto the hidden layers, of which there may be varying numbers, before it finally reaches the output layer where the final response of the network is given. The connections between each node are assigned a weight that, in combination with the threshold activations of each node, determine whether that node will transmit ("fire") information forward. These network parameters are typically determined during the training stage of an analysis, in which the network is exposed to the full data set and "learns" [74]. After training is complete and the network weights are fixed, the NN can be used as a predictive model or to identify events in new data.

5.1.2 NeuroBayes

NeuroBayes [77], the NN package used in this project, combines a neural network approach with Bayesian statistics into the activation functions of the nodes in order to incorporate a priori knowledge into the training to prevent unphysical predictions [78]. The NeuroBayes NN contains between 10-100 nodes, depending on the number of input training variables and hidden layers specified in the interface. In this work, an interface written in C++ has been used. The input samples are incorporated on this interface in the form of ROOT Trees [1].

The package is divided into 2 parts: NeuroBayesTeacher and NeuroBayesExpert. The Teacher uses the training data set (i.e. the idealised simulated MC data set) and its chosen training variables and trains the network [78], exporting the final state of network and fixed connection weights to a saved file that then forms the input to the Expert. Note that the number of input nodes used in the NN set up is equal to the number of training variables + 1 due to the presence of a single bias node, an extra node that always produces a constant value and aids in the network's learning of patterns [76].

The Expert maintains the network set-up from training and can then be used on other data. Within our analysis interface, the Expert generates a single output NN value (O_{NN}) for each event within this data set. This single discriminant output, which takes a value from 0-1, is essentially representative of all the input training variables and is easily manipulated. A single cut in O_{NN} can be applied to optimise the sample selection as an alternative to applying individual cuts in each kinematic variable. This is the fundamental advantage of the MVA NN approach. Figure 5.1 depicts a schematic of the NeuroBayes NN architecture, illustrating the progression of layers from input to output.

Note that the training variables are the relevant kinematic variables used as the input to the NN. The choice of such an initial set of variables for training is done regarding criteria of individual signal discrimination power against background of these variables with the additional requirement that they should not show a clear correlation with the observables used in the subsequent measurements that are intended to be performed with the selected sample via NN.

5.2 Previous Work

Before discussing the current research at hand, the previous work conducted for optimisation of selection criteria of t-channel single top quark events, namely "cut based" optimisation, is summarised so that comparisons can be made later on.

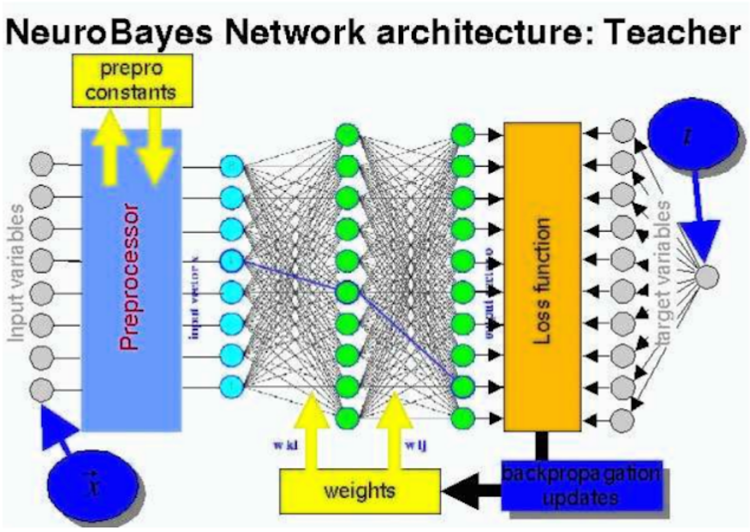

FIGURE 5.1: The NeuroBayes NN architecture [77].

5.2.1 Cut Based Optimisation at 13 TeV

The cut based optimisation, detailed in reference [6], involves a set of selection cuts in various kinematic variables, carefully chosen with the criterion of not being built with kinematic properties of one or part of the decay products of the top quark, that could in principle bias the sensitivity to anomalous couplings subsequently measured from asymmetries in angular observables proposed in chapter 2. These variables have been proven not to show correlations with the angular observables we are interested in measuring. These kinematic variables selected for these cuts (after application of preselection and $m_{l,b}$ cut) include: m_{top}, $m_{\text{j,top}}$, $\Delta\eta_{\text{J,top}}$ and H_{T} (defined in Table 5.2). Note that $\Delta\eta_{\text{J,top}}$ is a linear combination $\Delta\eta_{\text{j,top}} = |\eta_{\text{top}} - \eta_{\text{j}}|$ of the variables η_{top} and η_{j} described in table 5.2.

The final results of this optimisation are shown in Table 5.1, where the overall performance can be judged by the total value for S/B and Significance, as justified in the theory section. Note that event yields refer to the calculated prediction of the number of events for a certain integrated luminosity. The column values are the event yields (with computed statistical error) selected for a particular cut. Note that the event yields are the predictions for the number of events normalised to an integrated luminosity, after a particular selection. This is not the same as the total number of events produced by MC.

5.2.2 Starting Point for NN Training

The current analysis follows on from this previous optimisation by beginning by using the same 5 kinematic variables as the training variables for the NN, detailed in Table 5.2. This is done for 2 reasons. Firstly, the variables were chosen using theoretical knowledge so that they may not be sensitive or connected to the presence of anomalous couplings. I.e. it is desired that the variables are not dependent on the angular distributions in association with anomalous couplings that will be measured and this is fundamental in order to create an optimisation that will hopefully

Process	Event yields. $(H_T(min), m_{j,top}(min))$
	(168, 284)
t-channel	17929 ± 83
Z-jet, diboson	841 ± 85
W+jets	13189 ± 704
$t\bar{t}$	12088 ± 74
Wt s-channel	2052 ± 33
Multijet	1795 ± 45
Total	47895 ± 720
S/B	0.60
Significance	82

TABLE 5.1: Event yields for 13 TeV MC samples after applying the preselection requirements, $m_{l,b} < 140$ GeV, $142 < m_{top} < 418$ GeV, $|\eta_j| > 3.5 - \Delta\eta_{j,top}$ and $m_{j,top}$ and H_T cuts. The column's heading refers to a particular set of lower cuts in $m_{j,top}$ and H_T. These event yields have been normalized to an integrated luminosity of 36.5 fb^{-1} [6].

not bias the detection and measurement of anomalous couplings in the future. Secondly, it is helpful to begin with the same variables as before in order to make an initial comparison and evaluate whether the new method (the NN) has an improved performance over the previous cut based method, following the basic concept of control. After this has been confirmed, the potential for adding further kinematic training variables can be explored.

Kinematic Variable	Definition
m_{top}	The reconstructed mass of the top quark, built from the invariant mass of the neutrino, the electron/lepton and the b-jet.
$m_{\text{j,top}}$	The invariant mass of the four-momentum sum of the four-momenta of the reconstructed top quark and the light forward jet.
η_{top}	The pseudorapidity of the top quark
η_{J}	The pseudorapidity of the light jet
H_{T}	The scalar sum of the transverse energies of all final state particles: $H_T = p_T^l + p_T^j + p_T^b + E_T^{miss}$.

TABLE 5.2: The 5 initial kinematic variables - used in previous cut based optimisation and as the starting training variables for the NN. Note that the superscripts on the transverse momenta - l, j and b - stand for the lepton (electron), the light jet and the b-tagged jet respectively. [6].

The starting point for training the NN involves these kinematic variables after the application of preselection and $m_{l,b}$ cuts.

It is important to add that for NN training, a high statistics signal and background simulation sample is needed to ensure the network is trained correctly. In general this is true for all machine learning or MVA analysis methods. To confirm and demonstrate that sufficient statistics were employed, table 5.3 displays the actual MC statistics used (i.e. the raw number of events generated, not the event yields) after the preselection and $m_{l,b}$ cuts at $\sqrt{s} = 13$ TeV. These form the input to the NN training. As can be seen, there is clearly a sufficient number of events for each signal and background process, with a total number of over 1 million events. Therefore the amount of statistics is not a problem within this analysis.

Process	MC Statistics
t-channel	151264
Z-jet, diboson	62180
W+jets	192163
$t\bar{t}$	254507
Wt s-channel	65982
Multijet	368022
Total	1094118

TABLE 5.3: The actual MC statistics (number of generated events) used as the input to the NN training.

5.3 NN Training with Original 5 Variables

The first stage was to train the NN with the 5 original kinematic variables, Table 5.2, used in previous work to compare whether the use of a NN offers a measurable advantage over old cut-based methods. The adjustable parameters of the NN were fixed to the following:

- Number of Hidden Nodes: 14
- Proportion of Training Sample Used: 0.8

Note that these values were the default values supplied and that they will be explained, explored and optimised properly in the following section (here the aim was simply to provide a quick confirmation that the NN outperforms the cut based method).

After training the NN and running the expert with the same MC simulated samples, output values (O_{NN}) are assigned to every data event. The performance of the NN can be assessed by a pair of performance plots (Figure 5.2): S/B versus Significance and Signal Efficiency versus Background Rejection. These quantities are defined by the following equations:

$$S_{i-bin}^{int} = \sum_{j-bin=i-bin}^{N} S_{j-bin} \tag{5.1}$$

$$\mathrm{Signal-Efficiency_{i-bin}} = \mathrm{S_{i-bin}^{int}/S_{total}^{int}} \tag{5.2}$$

$$B_{i-bin}^{int} = \sum_{j-bin=i-bin}^{N} B_{j-bin} \tag{5.3}$$

$$\mathrm{Background-Rejection_{i-bin}} = 1 - \mathrm{B_{i-bin}^{int}/B_{total}^{int}} \tag{5.4}$$

where S_j (B_j) is the amount of signal (background) in the j-bin of the signal (background) O_{NN} histogram, S_j^{int} (B_j^{int}) is the sum of the bin contents from the j-bin up to the final bin and N is the total amount of bins of the O_{NN} histogram. This is equivalent to an integral from the O_{NN} value equal to the lower edge of the j-bin up to $O_{NN} = 1$. The aforementioned O_{NN} histogram is depicted in figure 5.7.

From the first graph in Fig 5.2, focusing in on the useful part of the graph corresponding to large significance, it is observed that there is a characteristic parabolic shape for the relationship between S/B and significance. At maximum significance, S/B is not at maximum. Although it is desirable to aim to maximise both quantities, there is a compromise to be made here. Table 5.4 details the values of S/B, Significance and O_{NN} for each bin in the O_{NN} histogram, figure 5.7. A value for a lower-cut in O_{NN} must be chosen in order to optimise the selection of the samples. Note that there is no single set way to choose this value, especially given that there is a compromise between significance and S/B. One could, for example, select the point that yields an S/B value of almost 2 (Highest point in Figure 5.2) but this would result in an extremely low significance of 23.66, which is undesirable as too much of the data would have been discarded. In this research, it was decided to choose an O_{NN} value that maintains the same significance as previous research (Significance = 82, Table 5.1) for the sake of comparison, whilst maximising S/B. Due to the parabolic shape of relationship, one value of significance corresponds to two S/B values, the larger of which will be chosen. This led to a chosen cut of $O_{NN} \geq 0.71$.

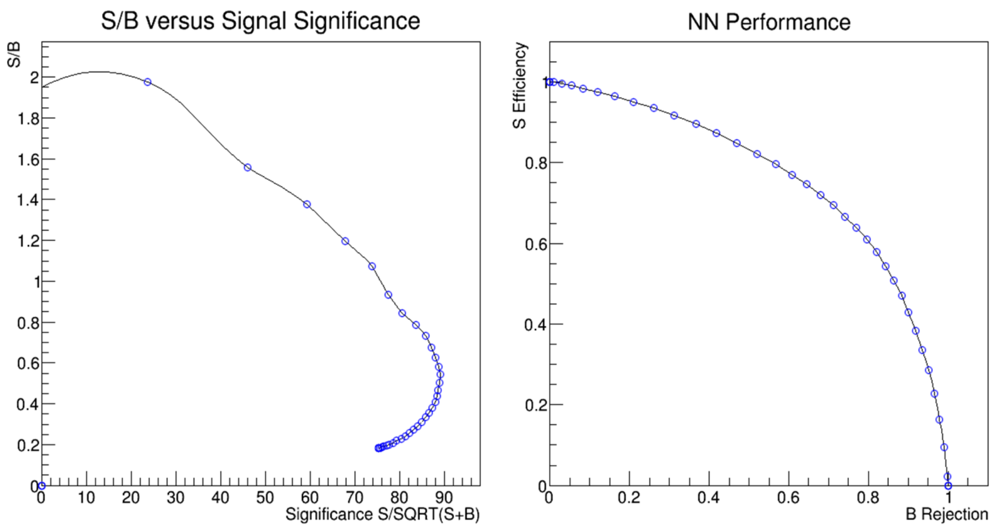

FIGURE 5.2: NN performance plots for training with 5 original kinematic variables.

5.3.1 Cut in the Output of NN

The final event yields after the cut in O_{NN} are calculated and displayed in Table 5.5. Firstly, it can be seen that applying the $m_{l,b}$ cut improves the total value of S/B and significance from the preselection stage - thus justifying the use of this initial cut. Note that the absolute total number of events decreases since some events have been discarded due to the cuts. Secondly, we note that by subsequently applying the O_{NN} cut, both values are optimised even further (significance increasing from 75 (after

S/B	Significance	$O_{NN} \geq$	S/B	Significance	$O_{NN} \geq$
0.18	75.34	0.00	0.41	87.93	0.48
0.18	75.34	0.00	0.44	88.31	0.50
0.18	75.34	0.03	0.47	88.52	0.53
0.18	75.34	0.05	0.50	88.93	0.55
0.18	75.58	0.08	0.54	89.08	0.58
0.19	76.00	0.10	0.58	88.70	0.60
0.19	76.45	0.12	0.63	88.04	0.62
0.19	77.02	0.15	0.68	87.07	0.65
0.20	77.67	0.18	0.73	85.81	0.68
0.21	78.49	0.20	0.79	83.62	0.70
0.22	79.36	0.23	0.84	80.61	0.73
0.23	80.29	0.25	0.93	77.52	0.75
0.24	81.20	0.28	1.07	73.86	0.78
0.26	82.23	0.30	1.19	67.78	0.80
0.27	83.10	0.33	1.38	59.26	0.83
0.29	84.01	0.35	1.56	46.04	0.85
0.31	84.93	0.38	1.98	23.66	0.88
0.33	85.81	0.40	0.00	0.00	0.90
0.36	86.51	0.43	0.00	0.00	0.93
0.38	87.30	0.45	0.00	0.00	0.95

TABLE 5.4: Table of S/B, Significance and O_{NN} values after discretisation of samples into bins. The O_{NN} values are the lower edge values (on x-axis) for the bin. At this stage a cut of $O_{NN} \geq 0.71$ was subsequently applied.

just the application of preselection and $m_{l,b}$ requirements) to 83 (after the application of preselection, $m_{l,b}$ and O_{NN} cut) and S/B increasing from 0.18 to 0.81). Making comparisons against the previous cut-based work (Table 5.1), it is observed that for roughly the same total value for significance of 82/83, the NN notably outperforms the cut-based method, yielding an optimised S/B value of 0.81 as opposed to 0.60. It is concluded that the new method of NN training and implementation, using the same 5 kinematic variables, is a better method for data optimization, quantified by the larger S/B value yielded.

5.4 NN Training with Optimised Parameters

The next stage was to further improve the performance of the NN by investigating and "fine-tuning" its parameters - specifically the following:

- The Training Variables Used
- The Number of Hidden Nodes
- Testing for Overtraining

5.4.1 The Training Variables Used

The choice of kinematic variables used as training variables for NN training were optimised by exploring the addition of the following 2 variables:

Process	Event yields		
	Preselection	$m_{l,b}$	NN with $O_{NN} \geq 0.71$
t-channel	41154 ± 127	36988 ± 120	15259 ± 78
Z-jet, diboson	10217 ± 239	5500 ± 210	519 ± 55
W+jets	151047 ± 1748	87948 ± 1451	8639 ± 606
$t\bar{t}$	119770 ± 230	80409 ± 188	7369 ± 58
Wt s-channel	27580 ± 123	18061 ± 98	1193 ± 25
Multijet	31171 ± 175	12124 ± 107	1025 ± 32
Total	380940 ± 1796	241030 ± 1490	34004 ± 618
S/B	0.12	0.18	0.81
Significance	67	75	83

TABLE 5.5: Evolution of the event yields for 13 TeV MC samples as we add cuts. Each new column considers the cumulative cuts of the previous columns. The NN has been trained with the original 5 variables. These event yields have been normalized to an integrated luminosity of 36.5 fb^{-1}.

- $\Delta P_{\mathrm{T}}^{\mathrm{j,top}}$: (DeltaPtjtop) the absolute difference between the momenta of the reconstructed top quark and light forward jet, j. $\Delta P_{\mathrm{T}}^{\mathrm{j,top}} = |P_T^{top} - P_T^{j}|$.
- $P_{\mathrm{T}}^{\mathrm{top}}$: (PtTop) the momentum of the reconstructed top quark.

$\Delta P_{\mathrm{T}}^{\mathrm{j,top}}$ has been suggested as an NN training variable in reference [70] and is used in combination with $P_{\mathrm{T}}^{\mathrm{top}}$ in order to check if it adds new information to the network. As before, these additional variables were chosen so as not to be sensitive to anomalous couplings since they do not involve kinematics of one or part of the decay products of the top quark. The training of the NN was repeated twice, firstly with only the addition of $\Delta P_{\mathrm{T}}^{\mathrm{j,top}}$ and secondly with the addition of both $\Delta P_{\mathrm{T}}^{\mathrm{j,top}}$ and $P_{\mathrm{T}}^{\mathrm{top}}$ (and of course the original 5 variables). Figure 5.3 shows the pair of performance plots obtained. It is observed that the addition of $\Delta P_{\mathrm{T}}^{\mathrm{j,top}}$ greatly improves the performance, increasing the significance significantly from before for a set S/B. The addition of $P_{\mathrm{T}}^{\mathrm{top}}$ to this however offers a negligible effect to the performance - the graphs of $\Delta P_{\mathrm{T}}^{\mathrm{j,top}}$ and $\Delta P_{\mathrm{T}}^{\mathrm{j,top}}$ plus $P_{\mathrm{T}}^{\mathrm{top}}$ are roughly identical. This conclusion is justified by the calculation of the significance of each training variable - shown in Table 5.6. $\Delta P_{\mathrm{T}}^{\mathrm{j,top}}$ has a large significance of 36.26 σ, the third most significant training variable, whilst $P_{\mathrm{T}}^{\mathrm{top}}$ has a mere 0.14 σ. This can be as expected because the variable $P_{\mathrm{T}}^{\mathrm{top}}$ is essentially accounted for within the variable $\Delta P_{\mathrm{T}}^{\mathrm{j,top}}$ and hence offers negligible new information.

The training variables used were then fixed to include the original 5 and $\Delta P_{\mathrm{T}}^{\mathrm{j,top}}$. $P_{\mathrm{T}}^{\mathrm{top}}$ was not used.

5.4.2 The Number of Hidden Nodes

The number of hidden nodes/layers in the NN is a further parameter that can be adjusted from the NN interface. The NN training was repeated for numbers of hidden nodes of: 7, 14 (default) and 28. Again, the performance plots for these are given in Figure 5.4. It is observed that varying the number of hidden nodes has no

Variable	Significance(σ)
$m_{\mathrm{j,top}}$	83.8
η_{j}	41.21
$\Delta P_{\mathrm{T}}^{\mathrm{j,top}}$	36.26
m_{top}	32.21
H_{T}	20.04
η_{top}	17.07
$P_{\mathrm{T}}^{\mathrm{top}}$	0.14

TABLE 5.6: Variables used as input to the neural network ordered by their importance (significance). These values are generated from the NeuroBayes package.

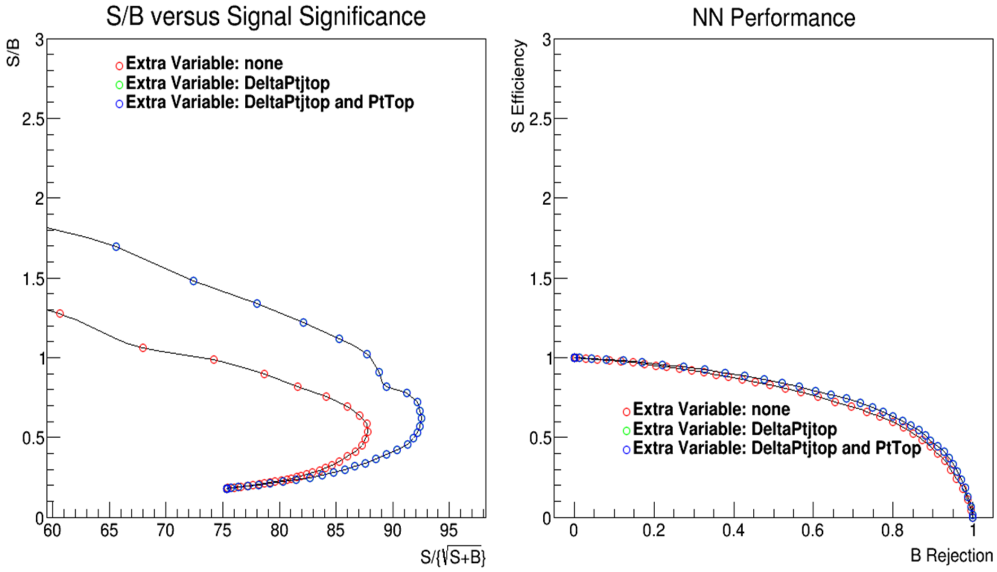

FIGURE 5.3: Comparison of NN performance when adding new variables.

effect on the final performance of the NN. This was as expected theoretically, since the NeuroBayes manual [77] states that, in general, the number of hidden nodes is not critical due to the presence of "pruning" - an NN computational phenomena that causes unused nodes to exponentially decay away. Despite all values yielding essentially the same result, the final fixed value chosen was 14. This was because 28 significantly increased the length of computing time, and 7 ran a risk of limiting the network's learning capabilities.

5.4.3 Testing for Overtraining

Overtraining is the name given to the phenomenon that occurs within machine learning (and more generally within data modelling) where the outcome of a model too closely matches a single given data set and hence fails to accurately fit or predict further data sets. The statistical model may describe the random error/noise

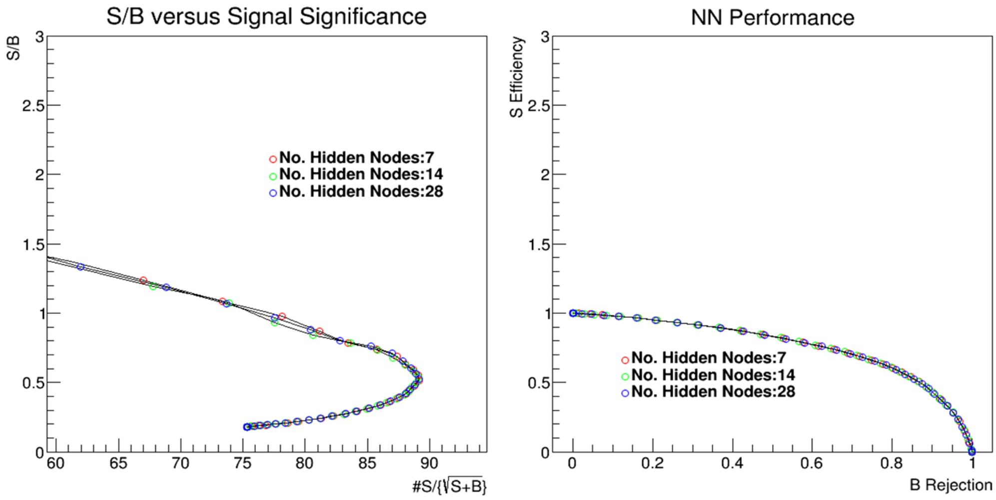

FIGURE 5.4: Comparison of NN performance when varying the number of hidden nodes in the NN.

of a data set instead of finding the general underlying relationship [79]. It is important to ensure that the trained NN has not been overtrained. 2 statistical tests are performed to check for the presence of overtraining.

Firstly the performance of the NN is evaluated when using varying proportions of the training sample. The parameter RTRAIN, a value between 0-1, is adjusted from the NN interface and training is redone for values of 0.25, 0.5, 0.8 (default) and 1.0. Note that although training is only conducted for the proportion RTRAIN, the expert is always run on 100% of the samples. The performance plots obtained are shown in Figure 5.5. It is clearly seen that adjusting the proportion of the training sample used has no effect on the performance of the NN. This implies there is no overtraining since the same result is obtained regardless of the amount of data needed to train the network. A fixed value of RTRAIN=0.8 was arbitrarily chosen for the final set up of the NN.

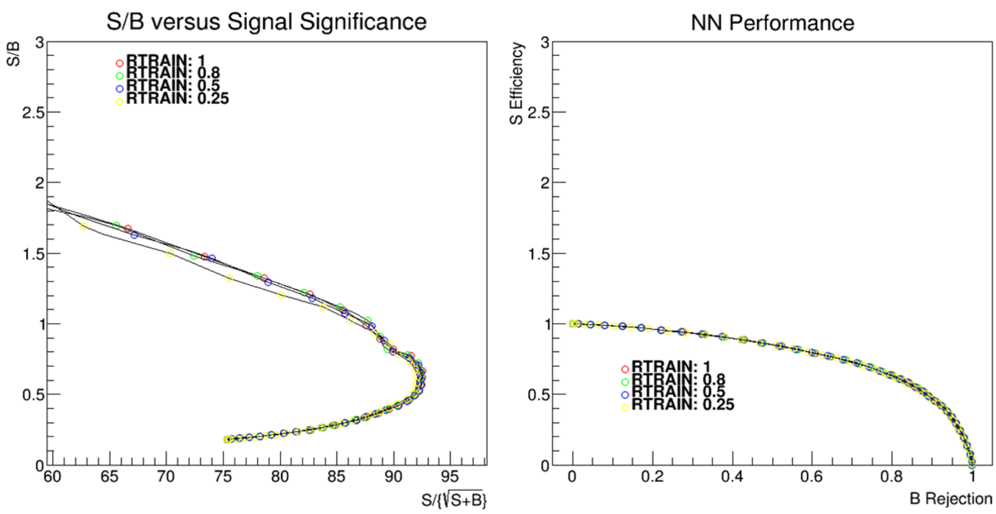

FIGURE 5.5: Comparison of NN performance when adjusting the proportion of training sample used, RTRAIN.

The second test was to randomly split the data sample into 2 halves and train separately on each half, run the expert on the corresponding other half and compare relative performance. The random splitting was conducted by the creation of a random variable with a Gaussian distribution, between -1 and 1, assigned to each event. This random Gaussian variable was generated via the ROOT "TRandom" function [1]. The network was first trained only using events with a positive value for the "RandomGauss" variable, then trained with the negative half. Figure 5.6 shows the performance plots obtained for each half as well as for both halves - "Data: Normal", red points. It is observed that the performance for each half is the same and hence the choice of which events are used for training does not influence the outcome. Again, no ovetraining is observed. Note that the full sample set (red) points yield a higher significance by roughly a factor of $\sqrt{2}$. This is as expected since the equation for significance (Equation 2.7) will scale by a factor of the square root of the fraction of sample used. It is concluded that there is no overtraining present and that the trained NN will be a reliable model of general data trends. Again, this can be as expected as the presence of "pruning" techniques within the NeuroBayes package reduce the risk of overtraining [78].

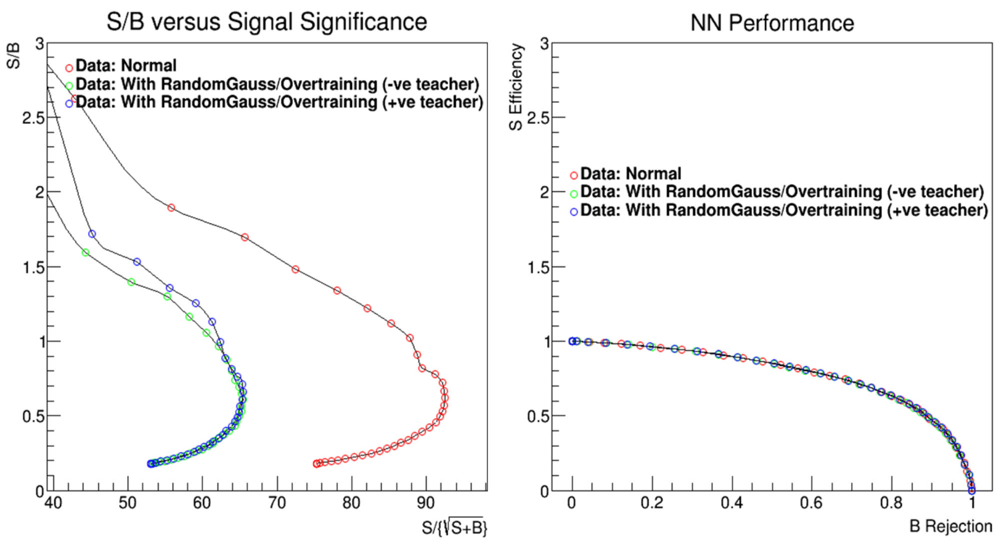

FIGURE 5.6: Comparison of NN performance with overtraining tests.

The final set-up for the NN is detailed below and will be fixed for the remainder of the research (including at 8 TeV):

- Training Variables: m_{top}, $m_{j,top}$, η_J , $\eta_{J,top}$, H_T **and** $\Delta P_{\mathrm{T}}^{\mathrm{j,top}}$
- Number of Hidden Nodes: 14
- Proportion of Training Sample Used, RTRAIN: 0.8

5.4.4 Cut in the Output of NN

As done in the previous section, a value for a lower-cut in O_{NN} is chosen by studying the produced O_{NN} table (shown in appendix A) and choosing a value corresponding to a significance of roughly 82. This led to a chosen cut of $O_{NN} \geq 0.78$.

Figure 5.7 shows the histogram for O_{NN} before and after the cut in O_{NN}. Figures 5.8 to 5.13 show the stack plots (histograms divided into separate components for each signal and background part, stacked one on top of the other) for all the kinematic training variables used, before and after cut in O_{NN}. It can be seen that the application of the cut maximises the ratio S/B (signal events are shown in blue). The black points in the stack plots refer to the SM predictions (these are not real data points) and these are simply the total sum of the signal and background contributions. Note that figure 5.11 shows a symmetric plot about 0, as expected, because the t-channel signal events have the forward light jet peaking towards high values of pseudorapidity (as seen for the spectator quark in the Feynman diagram 2.3).

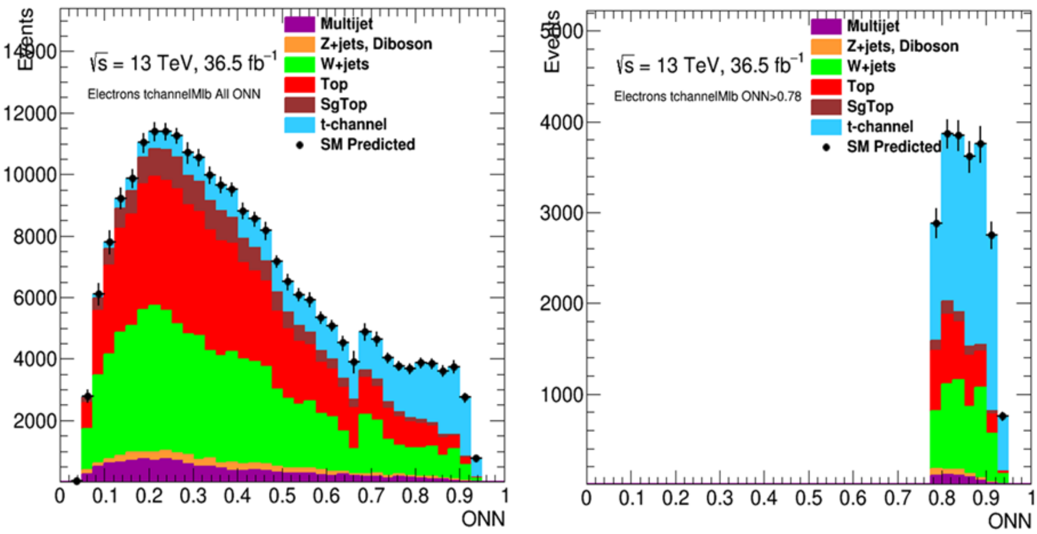

FIGURE 5.7: Histograms of ONN distributions, before (left) and after (right) the cut $O_{NN} \geq 0.78$ is applied.

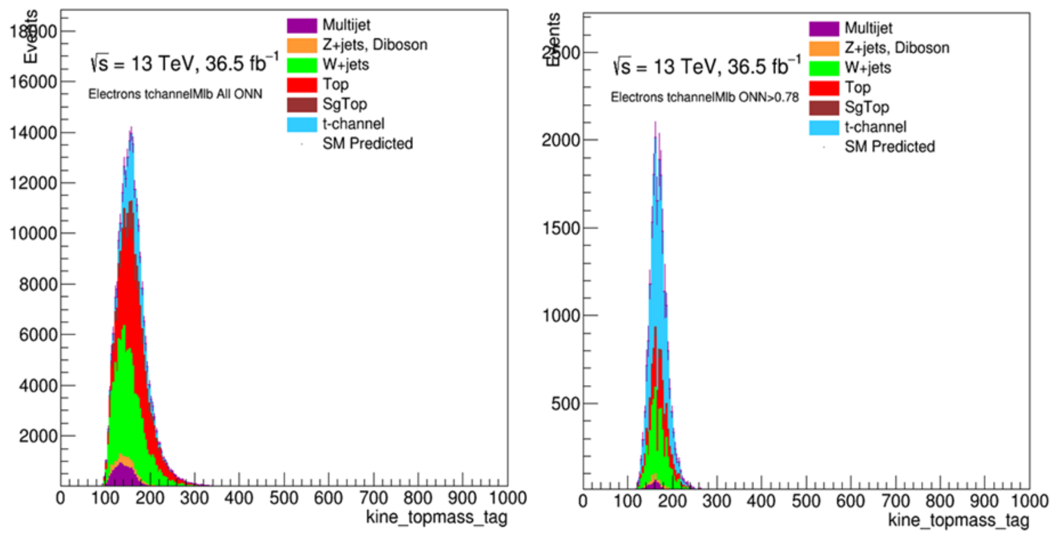

FIGURE 5.8: Stack plot of the kinematic training variable m_{top} before (left) and after (right) the cut $O_{NN} \geq 0.78$ is applied.

Table 5.7, the table for the final event yields obtained with the new optimised O_{NN} cut, shows that the NN has greatly improved the event selection. For a significance of around 81/82/83, the cut in O_{NN} after optimised training yields a total S/B of 1.25. This is notably greater than both the cut based method (S/B = 0.60, Table 5.1)

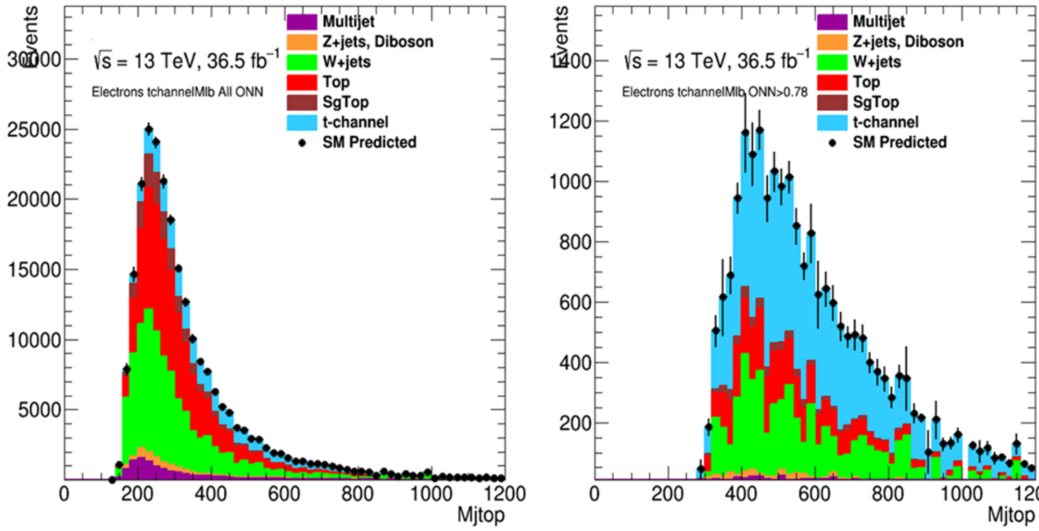

FIGURE 5.9: Stack plot of the kinematic training variable $m_{\mathrm{j,top}}$ before (left) and after (right) the cut $O_{NN} \geq 0.78$ is applied.

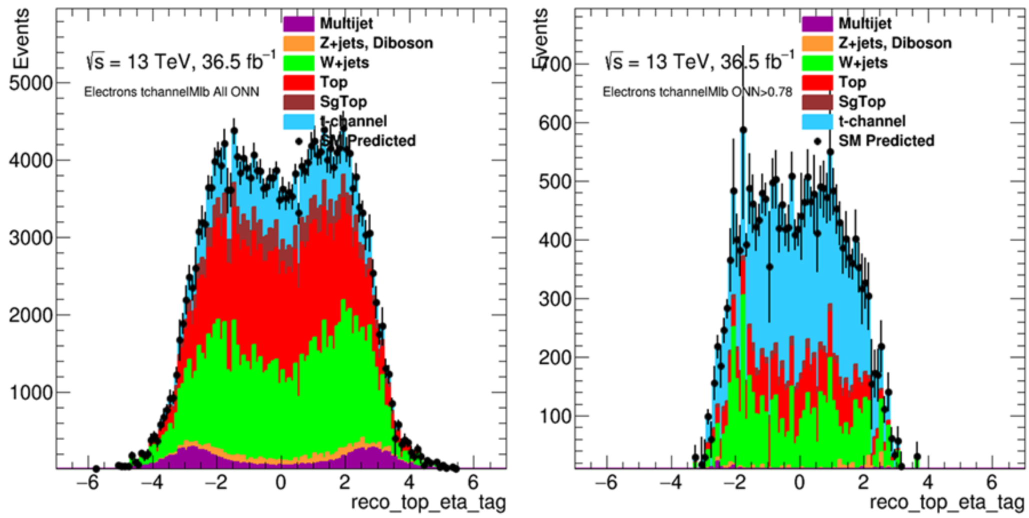

FIGURE 5.10: Stack plot of the kinematic training variable η_{top} before (left) and after (right) the cut $O_{NN} \geq 0.78$ is applied.

and the previous O_{NN} cut (S/B = 0.81, Table 5.5). Therefore it can be concluded that the fine-tuning of the NN parameters, most importantly so the addition of the $\Delta P_{\mathrm{T}}^{\mathrm{j,top}}$ training variable, has led to an improved selection and that generally NN application has greatly surpassed that of the previous cut-based optimisation. Not only this, but the application of the Bayesian NN was in many ways simpler and more efficient since many separate kinematic variables (the training variables) were incorporated into a single discriminant output (O_{NN}) that could then be easily manipulated. This would also serve to minimise systematic errors, should real data be used in the future. This is another advantage of the NN based optimisation.

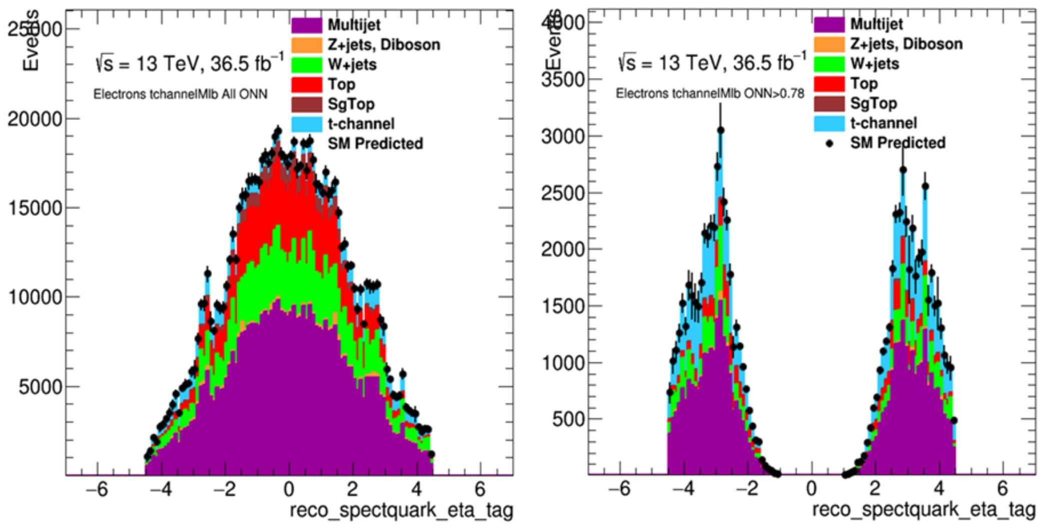

FIGURE 5.11: Stack plot of the kinematic training variable η_{J} ("reconstructed spectator light jet") , before (left) and after (right) the cut $O_{NN} \geq 0.78$ is applied.

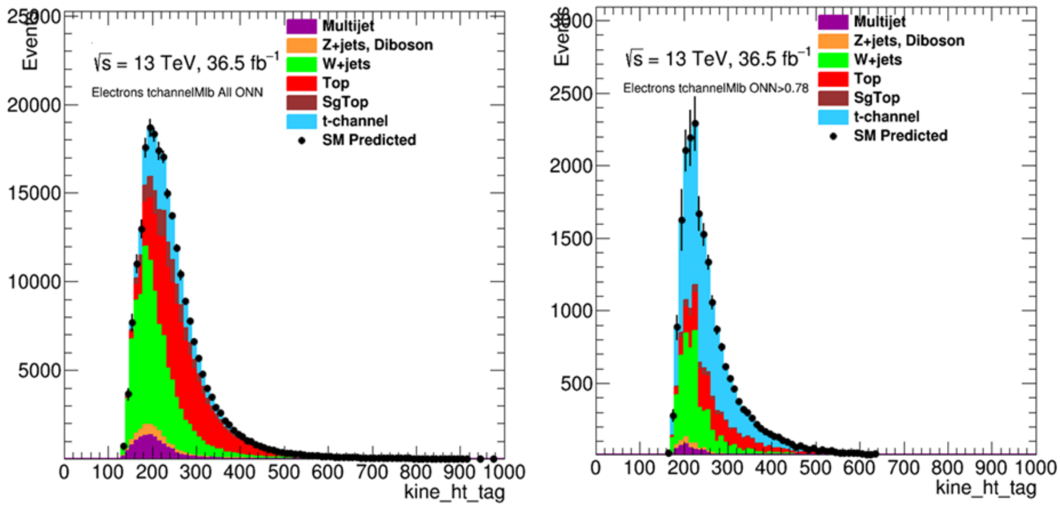

FIGURE 5.12: Stack plot of the kinematic training variable H_{T} before (left) and after (right) the cut $O_{NN} \geq 0.78$ is applied.

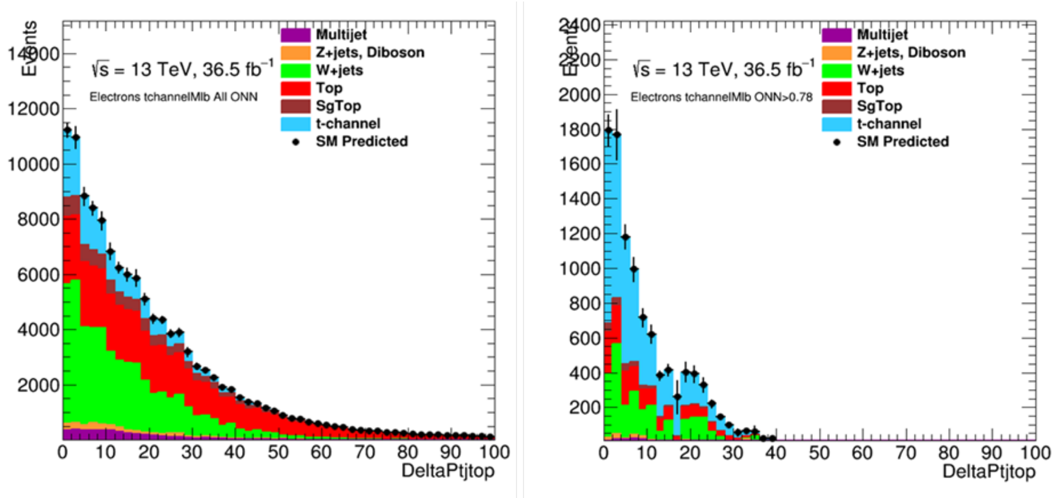

FIGURE 5.13: Stack plot of the kinematic training variable $\Delta P_T^{j,top}$ before (left) and after (right) the cut $O_{NN} \geq 0.78$ is applied.

Process	Event yields		
	Preselection	$m_{l,b}$	NN with $O_{NN} \geq 0.78$
t-channel	41154 ± 12	36988 ± 120	11929 ± 69
Z-jet, diboson	10217 ± 239	5500 ± 210	269 ± 39
W+jets	151047 ± 1748	87948 ± 1451	4966 ± 404
$t\bar{t}$	119770 ± 230	80409 ± 188	3265 ± 38
Wt s-channel	27580 ± 123	18061 ± 98	587 ± 17
Multijet	31171 ± 175	12124 ± 107	452 ± 22
Total	380940 ± 1796	241030 ± 1490	21468 ± 414
S/B	0.12	0.18	1.25
Significance	67	75	81

TABLE 5.7: Evolution of the event yields for 13 TeV MC samples as we add cuts. Each new column considers the cumulative cuts of the previous columns. The NN has been trained with optimised parameters. These event yields have been normalized to an integrated luminosity of 36.5 fb^{-1}.

Chapter 6

Study of Top Quark Polarisation and W Boson Spin Observables

The following chapter deals with a study of two of the observables proposed in references [5] and [35] in order to evaluate the impact of a selection based on a Bayesian neural network in the analysis of the Wtb vertex reported at [3]. The NN based optimisation from chapter 5 is first revisited for samples at $\sqrt{s} = 8$ TeV and then applied to the MC samples with anomalous couplings. An investigation into the sensitivity and linearity of the analysis to the presence of anomalous couplings is then conducted via the computation of forward-backward asymmetries after an unfolding procedure is performed.

6.1 Repeating Optimisation at $\sqrt{s} = 8$ TeV

The NN was retrained, maintaining the same optimised parameters from chapter 5, using the $\sqrt{s} = 8$ TeV MC samples of the SM signal and background events (from the Wtb vertex analysis in reference [3]). The preselection cuts and $m_{l,b} < 160$ GeV requirement (Chapter 4) were applied before training. Note that there are fewer total event yields for the samples at 8 TeV in comparison to those at 13 TeV, matching the smaller total integrated luminosity of 20.2 fb^{-1} in the LHC Run I. The full table for the O_{NN}, significance and S/B values produced are detailed in Appendix A.2.

The previous cut based optimisation for 8 TeV is described in the references [9] and [3], where the final values (for the electron channel) for Significance and S/B quoted are 39 and 1.32 respectively. Hence, following the same method as before and choosing to work at a comparable significance of roughly 39, a value of $O_{NN} \geq 0.82$ was chosen for the uppercut in O_{NN}. The final event yields with this O_{NN} cut are detailed in Table 6.1. The yields after preselection and the $m_{l,b}$ cut for 8 TeV are also detailed. As with 13 TeV, the use of the NN has greatly optimised the data, yielding a final value for S/B of 1.95, which is notably higher than the value of 1.32 yielded from the previous cut based method whilst working at the same significance.

6.2 Optimisation of Samples with Anomalous Couplings

The next stage was to use the pre-trained NN to apply the same optimisation as above to the 8 TeV MC PROTOS signal samples with anomalous couplings, Table

Process	Event yields		
	Preselection	$m_{l,b}$	NN with $O_{NN} \geq 0.82$
t-channel	8685 ± 26	8363 ± 25	2472 ± 14
Z-jet, diboson	2888 ± 45	1997 ± 39	49 ± 4
W+jets	19600 ± 142	13544 ± 120	619 ± 33
$t\bar{t}$	19138 ± 34	15324 ± 31	390 ± 5
Wt s-channel	3192 ± 43	2578 ± 39	64 ± 6
Multijet	4315 ± 37	3154 ± 32	146 ± 8
Total	57819 ± 165	44960 ± 141	3741 ± 38
S/B	0.18	0.23	1.95
Significance	36	39	40

TABLE 6.1: Evolution of the event yields for 8 TeV MC samples (without anomalous couplings) as we add cuts. Each new column considers the cumulative cuts of the previous columns. The NN has been trained with optimised parameters.

4.2. That is to say, the same training file (i.e. the same network weights and structure) produced from the training of 8 TeV events without anomalous couplings was maintained as the input for the NeuroBayes Expert. The Expert was then run using the samples with anomalous couplings in order to produce O_{NN} values for each event. The same cut of $O_{NN} \geq 0.82$ was subsequently applied to these events. Note that the purpose of this cut was not to optimise the samples, as the MC background events are not considered in this exercise, but to reproduce the selection for the samples with anomalous couplings in order to then conduct tests to determine bias and sensitivity of entire analysis.

6.3 Unfolding of Angular Observables to Parton Level

At ATLAS (or indeed any detector), the recorded events do not properly correspond to those of the expected underlying events from theory and this is due to the distorting of the data recorded via detector effects. Such effects include the detector resolution, selection inefficiencies as well as hadronisation and parton showering of the particles before detection [9]. For this reason, it is necessary to perform what is known as an "Unfolding" procedure to the recorded data before comparing the distributions to theory. The unfolding procedure involves deconvoluting the detector effects from the recorded data in order to extract back the original parton level distributions [1].

The simulated MC sample files used in this research contain two main types of histograms generated from PROTOS: reconstruction histograms which display the information at detector level and truth or generator level histograms which display the information at parton level. The same detector effects that convolute real data are also modelled by the MC simulation and applied to the distributions generated at truth level in order to produce the distributions at reconstruction level. Note that

[1]In particle physics, partons refer to quarks and gluons. Parton level considers to the level at which such particles interact, as opposed to particle or detector level - which refer to the effects and products that occur and are detected after the raw interaction. Parton showers are the cascades of secondary particles produced when high energy particles interact with matter [80]

the reconstructed histograms contain the kinematic cuts applied from the optimisation (preselection requirements, $m_{l,b} < 160$ GeV, $O_{NN} \geq 0.82$) whilst the truth level histograms do not. Within the scope of this project, the truth histograms in combination with the reconstructed histograms were used to create an unfolding procedure, that in principle could then be applied to real data in the future. This unfolding procedure was subsequently applied to the reconstruction histograms so that comparisons can be made and asymmetries can be calculated from the relevant unfolded angular distributions.

The unfolding was performed on the two angular observables, described in the Chapter 2, that will be used to calculate the forward backward asymmetries: $cos(\theta_l)$ (named X in the code) and $cos(\theta_l^N)$ (named LN in the code). Figure 6.1 shows these initial angular distributions for the SM MC samples after the preselection, $m_{l,b}$ and $O_{NN} \geq 0.82$ cuts are applied, prior to unfolding. The histograms are distributed into just 2 bins in order to aid in the subsequent calculations of forward-backward asymmetries A_{FB} (equation 2.4), which involves grouping the events into those above and below 0.

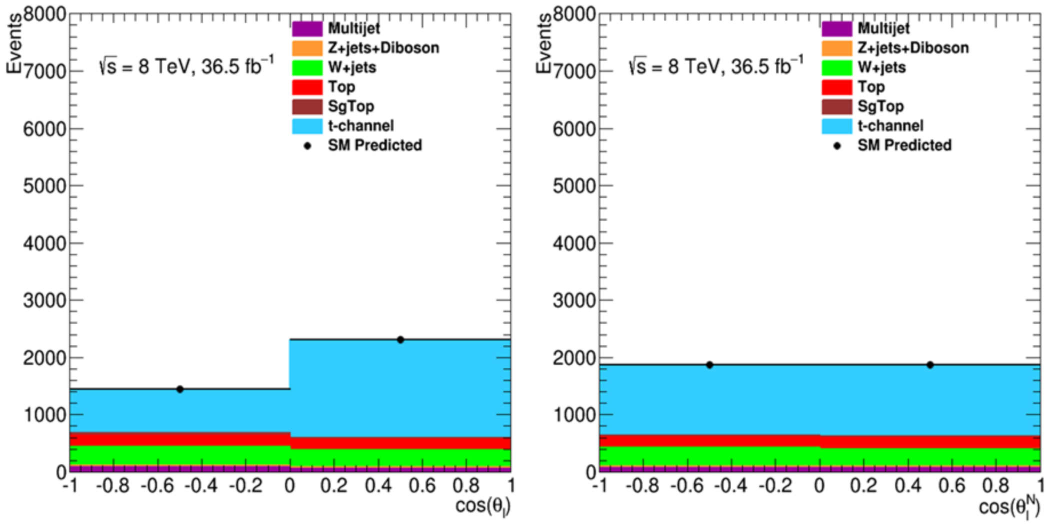

FIGURE 6.1: Distributions of the 2 chosen angular observables, $cos(\theta_l)$ (left) and $cos(\theta_l^N)$ (right), at reconstruction level for the SM MC samples after cut $O_{NN} \geq 0.82$ is applied and prior to unfolding. The histograms have just 2 bins.

The unfolding procedure itself involves a matrix inversion combined with an efficiency correction and is implemented using the ROOUNFOLD package [2] which uses an iterative bayesian method [3]. This operation takes as an input the reconstructed (or measured in the case of real data) histogram (N_j^{reco}), the SM migration matrix ($M_{i,j}$) and the event selection efficiency (ϵ_j) in order to yield the final unfolded histogram ($N_j^{unfolded}$) in accordance with equation 6.1 [3].

$$N_j^{unfolded} = \frac{\sum_i M_{i,j}^{-1} N_i^{reco}}{\epsilon_j} \tag{6.1}$$

Note that N_i^{histo} refers to the number of signal events within the bin i for the corresponding histogram. The migration matrices and event selection efficiencies are computed from the PROTOS generation of the samples [9]. The unfolding process also compensates for the optimisation criteria (i.e. the kinematic cuts implemented by the cut in O_{NN} to the reconstructed histograms etc) via these event selection efficiencies. The migration matrices, also named resolution histograms, are a third type of histogram provided in the MC sample files: 2-D histograms of the reconstructed angular distributions against the truth angular histograms. Figure 6.2 shows the 2 migration matrices for each angular observable used in the unfolding procedure. Only the resolution histograms for the SM files (those without anomalous couplings) are used as the migration matrices. This is because the unfolding procedure is SM based but performed onto non SM files (those with anomalous couplings).

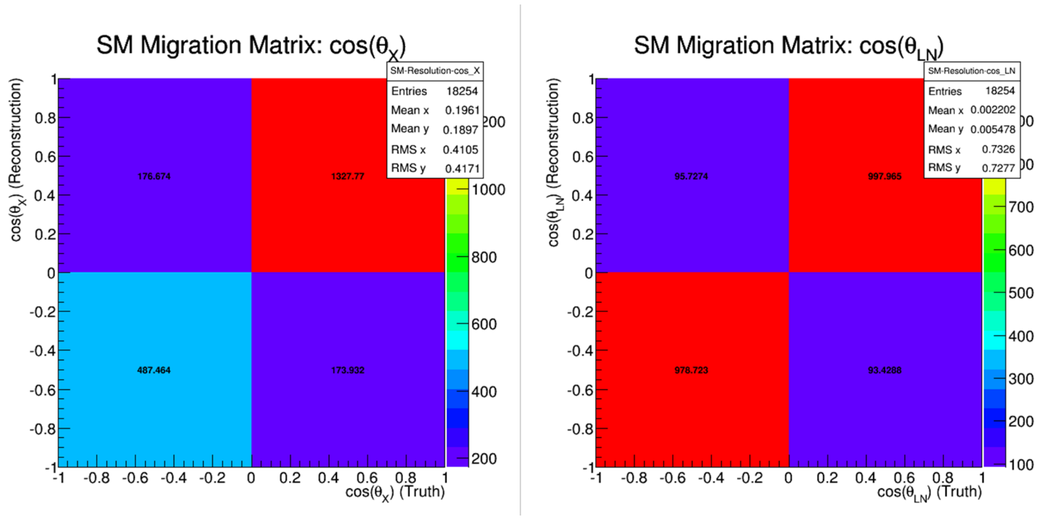

FIGURE 6.2: Migration matrices (SM resolution histograms) for the 2 chosen angular observables, $cos\theta_l$ (left) and $cos\theta_l^N$ (right). The histogram has been plotted in a style such that the block colour of the bin (as well as the explicit number of entries stated in the centre) represents the number of events in that bin.

The number of iterations is a further parameter of the unfolding procedure that must be specified. The optimum number of iterations was found by investigating the effects of varying the number of iterations from 0-20 for each angle. The unfolded forward-backward asymmetry, A_{FB} (unfolded), was computed for each unfolded histogram for each iteration per angle, with the results displayed in Figures 6.3, 6.4 and 6.5. The optimum number of iterations is seen at the point at which the graph reaches convergence: the minimum number of iterations required for a stable value for A_{FB} (unfolded) such that it does not vary with subsequent iterations. This is first conducted for the single SM sample, figure 6.3, and it is observed that convergence and agreement with the truth asymmetry is achieved immediately. The graphs also display agreement with the SM values for A_{FB}^l and A_{FB}^N of 0.45 and 0.00 respectively (Table 2.3). This is as expected and offers an initial confirmation the unfolding procedure works correctly for samples without anomalous couplings. The same is then repeated for the samples with anomalous couplings. These samples have been divided into 2 graphs; one for the samples with a non 0 real component for the coupling g_R and one for those with a non 0 imaginary g_R. This was done because it was expected that the results yielded would differ depending on the nature of this coupling. Here, convergence occurs after **8** and **5** iterations for $cos(\theta_l)$ and $cos(\theta_l^N)$ respectively. The number of iterations for the non SM samples was subsequently fixed to these optimum numbers and a final unfolding performed.

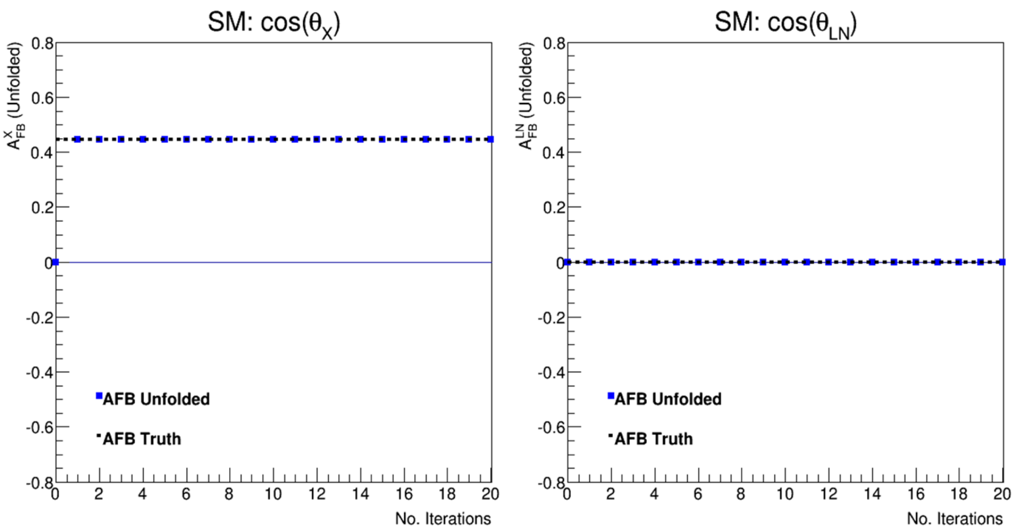

FIGURE 6.3: Unfolded asymmetries for the SM angular distribution $cos(\theta_l)$ as a function of the number of iterations. The samples with Re(g_R) are shown on the left, those with Im(g_R) to the right. Convergence and agreement with the SM truth values are immediately observed.

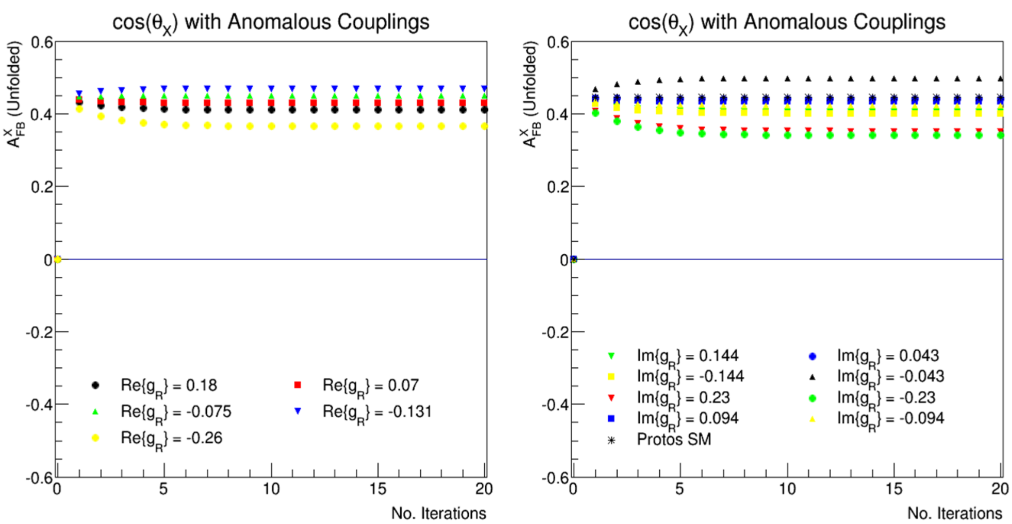

FIGURE 6.4: Unfolded asymmetries for the angular distribution $cos(\theta_l)$ as a function of the number of iterations. The samples with Re(g_R) are shown on the left, those with Im(g_R) to the right. Convergence is observed after roughly 8 iterations.

The full angular distributions for these final unfolded histograms, with a fixed number of iterations, are shown in figure 6.6. In comparison to the initial distributions prior to unfolding, figure 6.1, the same general shapes of the asymmetries is observed for both. Specifically, $cos(\theta_l)$ contains significantly more events in the range $0 < cos(\theta_l) < 1$ (bin 2) than $-1 < cos(\theta_l) < 0$ (bin 1) whilst $cos(\theta_l^N)$ displays a more even distribution across the 2 bins.

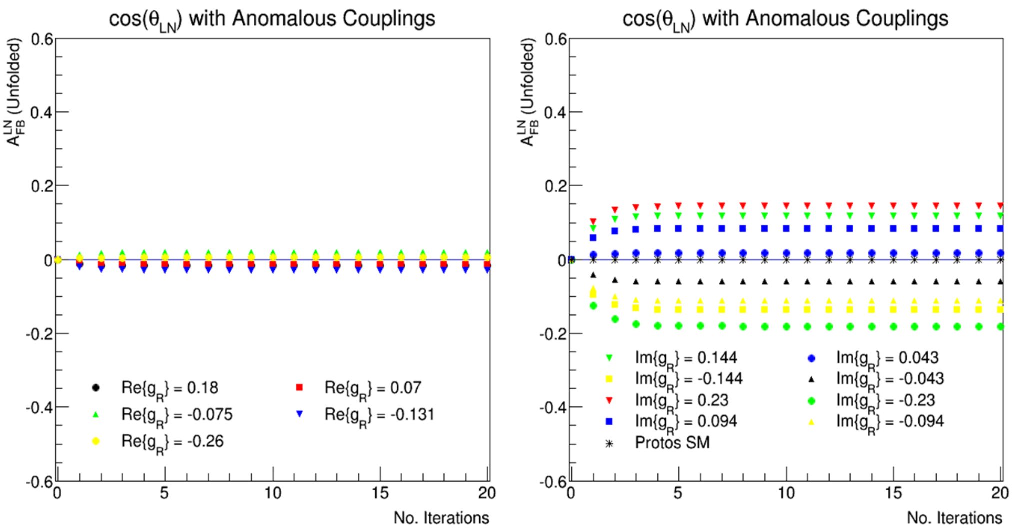

FIGURE 6.5: Unfolded asymmetries for the angular distribution $cos(\theta_l^N)$ as a function of the number of iterations. The samples with $\mathrm{Re}(g_R)$ are shown on the left, those with $\mathrm{Im}(g_R)$ to the right. Convergence is observed after roughly 5 iterations.

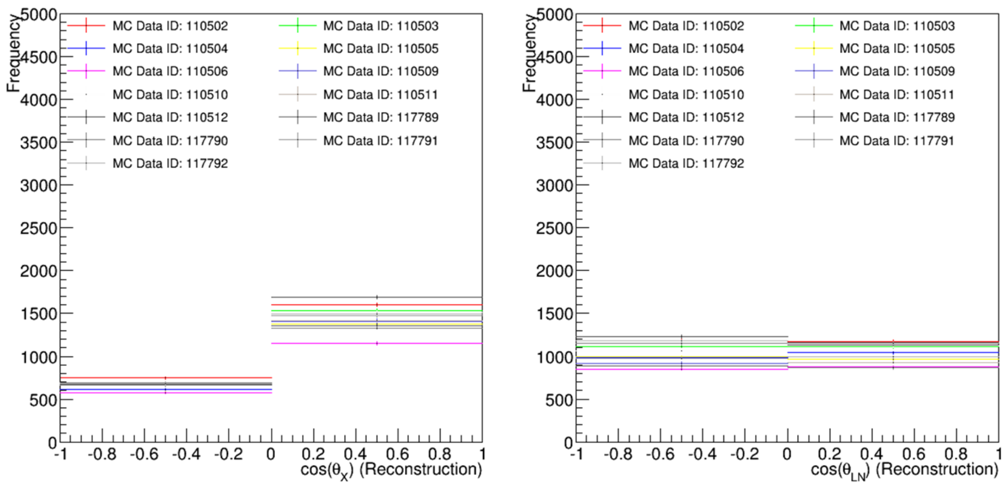

FIGURE 6.6: Final unfolded histograms of the 2 chosen angular observables, $cos\theta_l$ (left) and $cos\theta_l^N$ (right) for a fixed number of iterations (5 and 8 respectively). The graphs of each unfolded sample are overlayed with the data IDs shown in the legend corresponding to the IDs given in table 4.2

6.4 Sensitivity and Linearity to Anomalous Couplings in the *Wtb* Vertex

After completing the unfolding of the reconstructed histograms with anomalous couplings and asymmetry calculations, the final stage was to conduct tests to investigate both the sensitivity and the linearity of the response of the entire analysis to anomalous couplings. This is an important consideration for when the optimisation and unfolding processes are to be performed onto real data in the future.

Firstly, the sensitivity of the response refers to the ability of the analysis to detect new physics, specifically anomalous couplings. It is important to ensure that the application of the optimisation to real data will not hinder or prevent the detection of potential anomalous couplings. The sensitivity can essentially be quantified by the uncertainties of the unfolded asymmetries. Small uncertainties are desirable because large uncertainties would result in it being impossible to distinguish between a measured asymmetry that is in accordance with the SM and one which is indicative of BSM anomalous couplings. Figures 6.7 and 6.8 illustrate the relative errors per sample ($|\frac{\sigma_{A_{FB}}}{A_{FB}}|$). Note that here the errors provided are solely statistical and that in the case of real data, systematic errors would have to be incorporated into the error calculations as well. Therefore this method is an initial estimation. As seen in the graphs, the relative errors are in general fairly small, which confirms a good sensitivity to anomalous couplings. For the angular distribution $cos(\theta_l^N)$, only the graph with the samples with Im(g_R) are displayed and this is because there is no dependency (and hence no sensitivity) between Re(g_R) and A_{FB}^N, as explained in reference [5]. $cos(\theta_l^N)$ is primarily dependent only on Im(g_R).

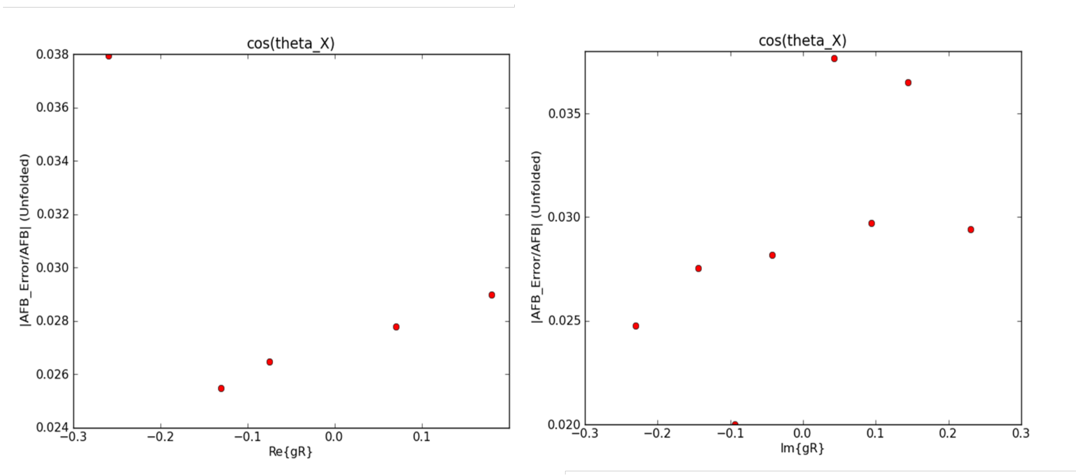

FIGURE 6.7: Relative statistical uncertainties ($|\frac{\sigma_{A_{FB}}}{A_{FB}}|$) in A_{FB} unfolded for Re(g_R) (left) and Im(g_R) (right) for the angular observable $cos(\theta_l)$.

Secondly, the linearity of the response refers to the relationship between the amount of new physics (anomalous couplings) measured and the amount of new physics actually present. The desired linear response would mean that the detection and measurement of anomalous couplings (after the applied optimisation via use of NN) returns the exact quantity and configuration of anomalous couplings present. For this it is important that the training of the NN is independent of the presence of anomalous couplings and will not result in a bias of their detection. This linearity can be investigated for the non SM MC samples by plotting graphs of their unfolded asymmetries against truth asymmetries, as shown in figure 6.9. Perfect linearity occurs when the unfolded asymmetries correspond exactly to the truth events - as represented by the addition of the green line, A_{FB} (unfolded) = A_{FB} (truth). Although a similar linear trend (with a positive correlation) is observed for both angular observables, there are clear deviations from the perfect linearity in which the error bars of multiple points do not overlap the green line. Note that for $cos(\theta_l^N)$, the samples with a non 0 value for Im(g_R) (the red points) are the samples that display a similar linear trend, whilst those with non 0 Re(g_R) (the blue points) show no variation. This is because, again, there is not a theoretical dependency between Re(g_R) and

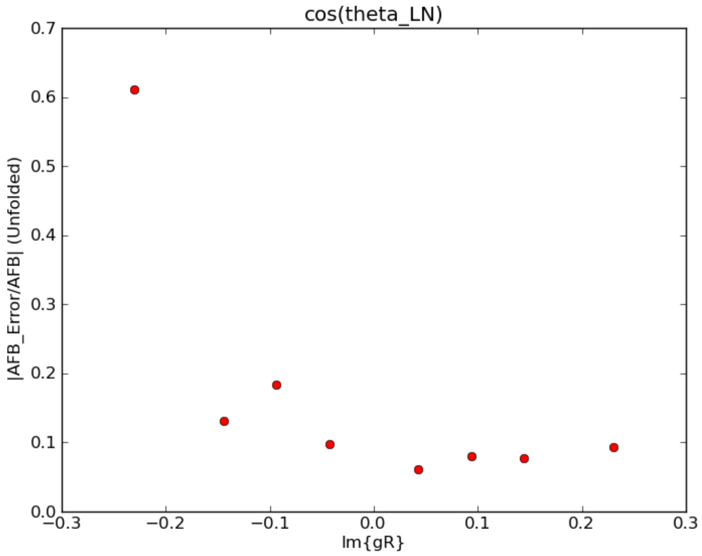

FIGURE 6.8: Relative statistical uncertainties ($|\frac{\sigma_{A_{FB}}}{A_{FB}}|$) in A_{FB} unfolded for Im(g_R) for the angular observable $cos(\theta_l^N)$.

$cos(\theta_l^N)$ but there is for Im(g_R).

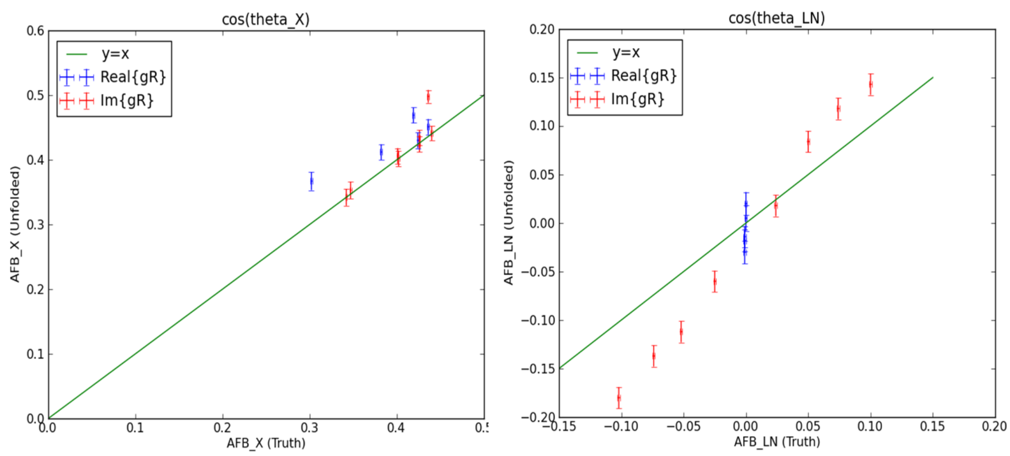

FIGURE 6.9: Linearity tests, graphs of A_{FB} (unfolded) Vs. A_{FB} (truth), for the angular observables $cos(\theta_l)$ (left) and $cos(\theta_l^N)$ (right). The errors in A_{FB} were calculated via propagation of the statistical errors provided for the bin contents of the relevant histograms. The blue points represent the samples with Re(g_R) and the red Im(g_R). The green line, A_{FB} (unfolded) = A_{FB} (truth), represents perfect linearity.

These deviations from linearity are further illustrated in figures 6.10 and 6.11, graphs representing the relative difference between the truth and unfolded A_{FB} values, $|\frac{A_{FB}(unfolded)-A_{FB}(truth)}{A_{FB}(truth)}|$ for each sample. It is observed that many of the samples yield notable deviations (the relative difference in A_{FB} is not 0) but that the largest deviations from linearity come from the samples with Re(g_R) for $cos(\theta_l^N)$.

It can be concluded that the optimisation and analysis used in this research is sensitive to the presence of anomalous couplings but exhibits deviations from a linear

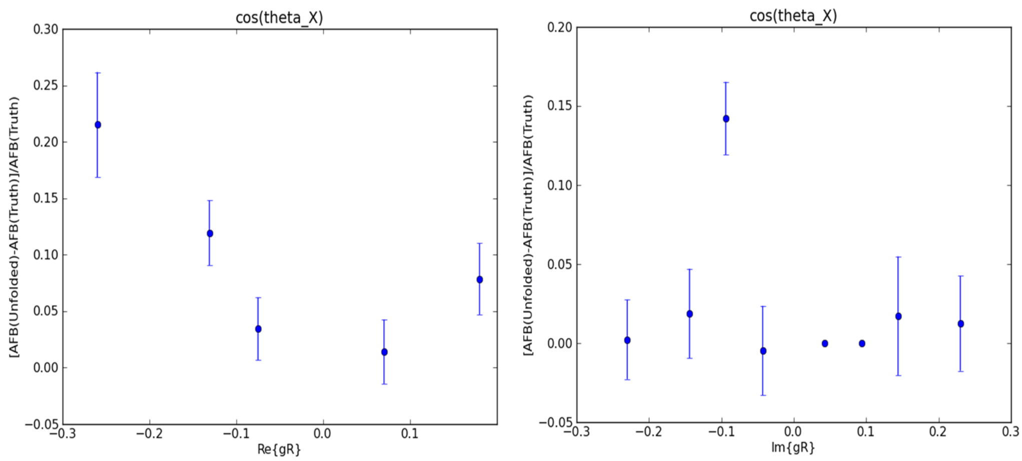

FIGURE 6.10: Deviations from linearity for the angular observable $cos(\theta_l)$. The graphs show $|\frac{A_{FB}(unfolded)-A_{FB}(truth)}{A_{FB}(truth)}|$ vs the coupling value, $\mathrm{Re}(g_R)$ (left) and $\mathrm{Im}(g_R)$ (right).

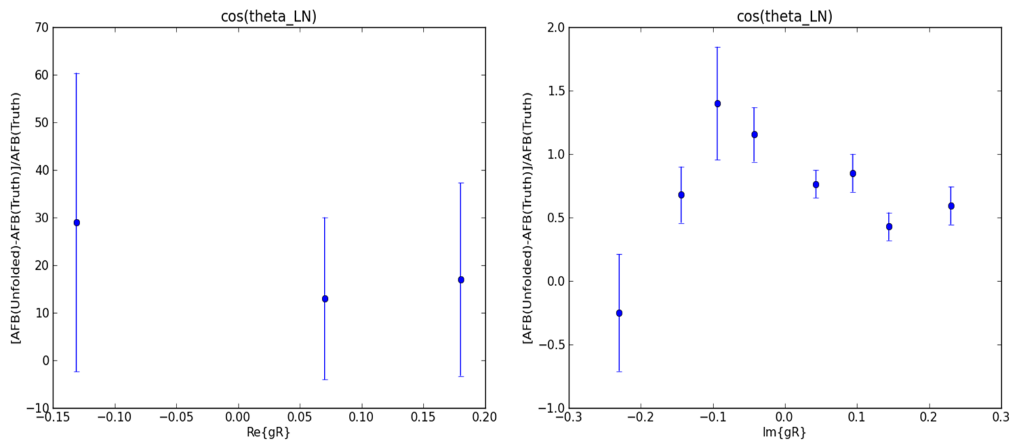

FIGURE 6.11: Deviations from linearity for the angular observable $cos(\theta_l^N)$. The graphs show $|\frac{A_{FB}(unfolded)-A_{FB}(truth)}{A_{FB}(truth)}|$ vs the coupling value, $\mathrm{Re}(g_R)$ (left) and $\mathrm{Im}(g_R)$ (right).

response. Although not desirable, there are steps that can be taken to correct for these deviations, as described in chapter 7.

Chapter 7

Conclusions

7.1 Conclusions

In this project, a review of the relevant theoretical framework, the current status of research, the analysis tools used and the experimental analysis conducted has been provided. An evaluation of the extent to which the aims of the project detailed in Chapter 1 have been met is provided below.

Firstly, it is concluded that a new and improved optimised selection of t-channel single top quark events has been achieved via use of a Bayesian neural network: the NeuroBayes package. The parameters of the NN were explored to obtain an optimum NN performance, as evaluated by the corresponding performance plots (e.g. figure 5.3) and a single discriminant output, O_{NN}, was correctly manipulated to apply a single cut to the sample in order to maximise S/B and significance. The final event yields of the optimised selection greatly outperform those yielded from the previous cut based method, as quantified by a greater value of S/B for a similar working significance. For example, at $\sqrt{s} = 13$ TeV (chapter 5), after a cut of $O_{NN} \geq 0.78$, a value for S/B of 1.25 was obtained for a significance of 81 and this was notably larger than the S/B value of 0.60 yielded for the previous cut based method at a comparable significance of 82. Furthermore, not only does the NN based method offer the advantage of an improved optimisation, but also the use of the single cut in the discriminant O_{NN} allows for an easier and simpler procedure as opposed to applying a series of separate cuts in multiple variables.

Secondly, it is concluded that, at first glance, this new optimised selection provided maintains the sensitivity of the analysis to the presence of anomalous couplings, as justified in a quantifiable way by small the relative statistical errors of the unfolded asymmetries. This evaluation will be updated when real data is used via the additional incorporation of experimental systematic errors. However, perfect linearity of the response to anomalous couplings was not found and instead clear deviations were observed. This is a problem that must be corrected for in order to make accurate measurements of the exact quantities of potential anomalous couplings in the future.

7.2 Future Work

There are various opportunities for future work, in both the short and long term, that can be completed, as discussed below. Note that these strategies are outside the time scale of this thesis.

Firstly, the use of other multivariate analysis or machine learning techniques for the optimisation of the selection of t-channel single top quark events could be investigated. Primarily, the use of a Boosted Decision Tree (BDT). This computational approach involves a succession of binary decisions (i.e. yes/no, left/right), using a particular discriminating variable per decision, that continuously divide the events into branches until a final outcome is reached [81]. This process classifies the events into signal or background. The discriminating variables are comparable to the kinematic training variables used for the NN. Not only could this approach potentially yield better results, i.e. a larger value for S/B, but could also perhaps yield angular distributions that, when unfolded, maintain the linearity of the response to the detection of anomalous couplings. Additionally there are other neural network packages that could be explored.

The use of the neural network could be further developed by exploring different methods for choosing the cut in O_{NN}. This thesis has only presented one such way, of maintaining a comparable working significance to that of the previous work and choosing an O_{NN} cut that maximised S/B. Instead, a method of fixing S/B and maximising significance could be adopted and compared. Regarding the whole picture of the analysis, once real data is used and systematic uncertainties introduced, the choice of O_{NN} must also be explored to minimise these uncertainties. As stated previously, there is not just one obvious way to make this choice, and the evaluation of the final result also depends on how one decides to measure the performance.

In order to solve the deviations from linearity, there are different strategies that could be followed. Firstly, an investigation into whether a particular training variable (or variables) is causing a bias can be conducted via the systematic elimination of each NN training variables, the retraining of the network and the comparison of the result. If, after this, linearity is obtained, the kinematic variable(s) responsible can be permanently excluded from the NN training (although this may come at the expense of a less optimised selection). The elimination of the preselection cuts and the $m_{l,b}$ requirement should also be investigated in case these are causing a bias. Secondly, instead of preventing the deviations from linearity in unfolded A_{FB} values, they could be corrected for via an iterative method with Lagrangian interpolation using the MC simulated samples with anomalous couplings. This method is described in reference [3]. Thirdly the unfolding procedure could be performed to particle level instead of to parton level, as done in the fiducial t-channel cross section measurements at 8 TeV of reference [71].

After the deviations from linearity are corrected for, the analysis can finally be applied to real data, which is the overall aim of the entire analysis. Using real data of proton-proton collisions from the LHC, collected at ATLAS at an energy of 13 TeV (Run II, ongoing currently), the optimisation via the bayesian neural network can be applied to the data to select t-channel events. To do this, the same training file/network configuration will be maintained from the training of the simulated MC events but the Expert (NeuroBayes package) alone will be run on the real data. Then, after the production of O_{NN} values for the data, the same optimised cut used for the MC samples, i.e. $O_{NN} \geq 0.78$ for 13 TeV, will be applied to the real data. Once optimised, the unfolding procedure can be applied to the data and the asymmetries calculated. From these, the values for anomalous couplings can be determined (i.e. via use of the TopFit program [41]), and hence the search for physics beyond the SM is facilitated.

Appendix A

O_{NN} Tables

S/B	Significance	$O_{NN} \geq$	S/B	Significance	$O_{NN} \geq$
0.18	75.34	0.00	0.49	91.82	0.50
0.18	75.34	0.00	0.53	92.18	0.53
0.18	75.34	0.03	0.57	92.40	0.55
0.18	75.35	0.05	0.62	92.55	0.58
0.18	75.70	0.08	0.67	92.40	0.60
0.19	76.41	0.10	0.73	92.18	0.62
0.19	77.26	0.12	0.78	91.21	0.65
0.20	78.20	0.15	0.82	89.47	0.68
0.21	79.18	0.18	0.91	88.79	0.70
0.22	80.30	0.20	1.02	87.75	0.73
0.24	81.44	0.23	1.12	85.27	0.75
0.25	82.66	0.25	1.22	82.09	0.78
0.26	83.74	0.28	1.34	78.01	0.80
0.28	84.76	0.30	1.48	72.41	0.83
0.30	85.76	0.33	1.70	65.60	0.85
0.32	86.64	0.35	1.89	55.74	0.88
0.34	87.58	0.38	2.63	42.87	0.90
0.36	88.51	0.40	4.02	22.03	0.93
0.39	89.47	0.43	0.00	0.00	0.95
0.42	90.41	0.45			
0.46	91.30	0.48			

TABLE A.1: O_{NN} values for optimised NN Training at 13 TeV

S/B	Significance	$O_{NN} \geq$
0.23	39.44	0.00
0.23	39.44	0.00
0.23	39.44	0.03
0.23	39.45	0.05
0.23	39.60	0.08
0.24	39.98	0.10
0.25	40.45	0.12
0.26	40.97	0.15
0.27	41.54	0.18
0.29	42.18	0.20
0.30	42.77	0.23
0.32	43.38	0.25
0.34	43.90	0.28
0.37	44.42	0.30
0.39	44.98	0.33
0.42	45.55	0.35
0.45	46.08	0.38
0.49	46.66	0.40
0.53	47.12	0.43
0.58	47.56	0.45
0.62	47.94	0.48
0.67	48.22	0.50
0.72	48.39	0.53
0.77	48.44	0.55
0.83	48.55	0.58
0.90	48.65	0.60
0.97	48.46	0.62
1.05	48.19	0.65
1.13	47.80	0.68
1.23	47.27	0.70
1.37	46.69	0.73
1.50	45.64	0.75
1.64	44.15	0.78
1.79	42.36	0.80
1.99	39.85	0.83
2.25	36.68	0.85
2.52	32.23	0.88
3.25	26.24	0.90
5.96	15.84	0.93
0.00	0.00	0.95

TABLE A.2: O_{NN} values for optimised NN Training at 8 TeV

Bibliography

[1] R. Brun and F. Rademakers. "ROOT: An object oriented data analysis framework". In: *Nucl. Instrum. Meth.* A389 (1997), pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.

[2] T. Adye. "Unfolding algorithms and tests using RooUnfold". In: *ArXiv e-prints* (May 2011). arXiv: 1105.1160 [physics.data-an].

[3] M. etal Aaboud. "Probing the W tb vertex structure in t-channel single-top-quark production and decay in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2017.4 (Apr. 2017), p. 124. ISSN: 1029-8479. DOI: 10.1007/JHEP04(2017)124. URL: https://doi.org/10.1007/JHEP04(2017)124.

[4] ATLAS Collaboration. "Search for anomalous couplings in the Wtb vertex from the measurement of double differential angular decay rates of single top quarks produced in the t-channel with the ATLAS detector". In: *ArXiv e-prints* (Oct. 2015). arXiv: 1510.03764 [hep-ex].

[5] J. A. Aguilar-Saavedra and J. Bernabéu. "W polarisation beyond helicity fractions in top quark decays". In: *Nuclear Physics B* 840 (Nov. 2010), pp. 349–378. DOI: 10.1016/j.nuclphysb.2010.07.012. arXiv: 1005.5382 [hep-ph].

[6] P.Martínez-Agulló. "Optimisation of selection criteria of t-channel single-top-quark events at $\sqrt{s} = 13$ TeV for studies of anomalous couplings in the Wtb vertex". MA thesis. University of Valencia, 2017. URL: CERN-THESIS-2017-156.pdf.

[7] W. Hollik. "Quantum field theory and the Standard Model". In: *ArXiv e-prints* (Dec. 2010). arXiv: 1012.3883 [hep-ph].

[8] CERN. *The Standard Model*. URL: https://home.cern/about/physics/standard-model.

[9] S.Pedraza-López. "Measurement of top quark and W boson polarisation observables with t-channel single-top-quark events in the ATLAS experiment." PhD thesis. University of Valencia, 2017. URL: CERN-THESIS-2017-159.pdf.

[10] D. Sen. "Horizons in World Physics (Vol. 275)". In: ed. by A. Reimer. Nova Science Publishers Inc, 2011. Chap. Of Mass, Charge and Spin, the basic attributes of matter –Their physical origin, pp. 275–291. ISBN: 978-1-61324-015-1. URL: https://www.researchgate.net/publication/233813051_Of_Mass_Charge_and_Spin_the_basic_attributes_of_matter_--Their_physical_origin.

[11] U. F. Nauk E. E. Boos and P N Lebedev. "Standard Model and predictions for the Higgs boson". In: *Physics-Uspekhi* 57.9 (2014). URL: http://iopscience.iop.org/article/10.3367/UFNr.0184.201409h.0985/pdf.

[12] G. Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716 (Sept. 2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[13] S. Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716 (Sept. 2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[14] J. Womersley. "Beyond the Standard Model". In: *Symmetry Magazine* 2.1 (2011). URL: https://www.symmetrymagazine.org/sites/default/files/legacy/pdfs/200502/beyond_the_standard_model.pdf.

[15] John Ellis. "Outstanding questions: physics beyond the Standard Model". In: *Phil. Trans. R. Soc. A* 370 (2012), pp. 818–830. DOI: doi:10.1098/rsta.2011.0452.

[16] Gabriella Sciolla. "The Mystery of CP Violation". In: *MIT Physics Annual* 6 (2006). URL: http://web.mit.edu/physics/news/physicsatmit/physicsatmit_06_sciollafeature.pdf.

[17] S. K. Gupta. "CP violating anomalous top-quark couplings at the Large Hadron Collider". In: *ArXiv e-prints* (Sept. 2009). arXiv: 0910.0068 [hep-ph].

[18] L. Canetti, M. Drewes, and M. Shaposhnikov. "Matter and antimatter in the universe". In: *New Journal of Physics* 14.9, 095012 (Sept. 2012), p. 095012. DOI: 10.1088/1367-2630/14/9/095012. arXiv: 1204.4186 [hep-ph].

[19] J. H. Christenson et al. "Evidence for the 2π Decay of the K_2^0 Meson". In: *Phys. Rev. Lett.* 13 (4 July 1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.138.

[20] Y. Fukuda et al. "Evidence for oscillation of atmospheric neutrinos". In: *Phys. Rev. Lett.* 81 (1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. arXiv: hep-ex/9807003 [hep-ex].

[21] CMS Collaboration. "Measurement of the top quark mass using proton-proton data at sqrt(s) = 7 and 8 TeV". In: *ArXiv e-prints* (Sept. 2015). arXiv: 1509.04044 [hep-ex].

[22] C. Campagnari and M. Franklin. "The discovery of the top quark". In: *Reviews of Modern Physics* 69 (Jan. 1997), pp. 137–211. DOI: 10.1103/RevModPhys.69.137. eprint: hep-ex/9608003.

[23] Hendrik Esch and Reiner Klingenberg. "Measurement of the top quark mass in topologies enhanced with single top-quarks produced in the t-channel using flavour-tagging and a neural network with ATLAS data at $\sqrt{s}8$ TeV". Presented 24 Oct 2014. Jan. 2015. URL: http://cds.cern.ch/record/1980097.

[24] F. Larios, E. Malkawi, and C.-P. Yuan. "Probing the Electroweak Symmetry Breaking Sector with the Top Quark". In: *ArXiv High Energy Physics - Phenomenology e-prints* (Sept. 1996). eprint: hep-ph/9609482.

[25] Susana Cabrera Urban. *Anomalous couplings in single top quark and searches for rare top quark couplings with the ATLAS detector*. Tech. rep. ATL-PHYS-PROC-2017-175. Geneva: CERN, Oct. 2017. URL: https://cds.cern.ch/record/2286439.

[26] D. E. Soper. "Basics of QCD Perturbation Theory". In: *ArXiv High Energy Physics - Phenomenology e-prints* (Jan. 1997). eprint: hep-ph/9702203.

[27] U. Husemann. "Top-quark physics: Status and prospects". In: *Progress in Particle and Nuclear Physics* 95 (July 2017), pp. 48–97. DOI: 10.1016/j.ppnp.2017.03.002. arXiv: 1704.01356 [hep-ex].

[28] CMS Collaboration. "Search for s channel single top quark production in pp collisions at sqrt(s) = 7 and 8 TeV". In: *ArXiv e-prints* (Mar. 2016). arXiv: 1603.02555 [hep-ex].

[29] N. Faltermann. "Measurement of the t-channel single-top quark production cross section at 13 TeV with the CMS detector". In: *ArXiv e-prints* (Nov. 2016). arXiv: 1611.08443 [hep-ex].

[30] P. Kant et al. "HATHOR for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions". In: *Computer Physics Communications* 191 (June 2015), pp. 74–89. DOI: 10.1016/j.cpc.2015.02.001. arXiv: 1406.4403 [hep-ph].

[31] Oscar E. Pastor. "Probing the Wtb Structure at the LHC". MA thesis. University of Valencia, 2015. URL: https://cernbox.cern.ch/index.php/s/SZFU2YzE7ubbzam#pdfviewer.

[32] Joseph Boudreau et al. "Single top quark differential decay rate formulae including detector effects". In: (2013). arXiv: 1304.5639 [hep-ex].

[33] J. Kessler. *Polarized Electrons*. Berlin Heidelberg: Springer Verlag, 1976, pp. 7–19. DOI: ISBN0-387-07678-6.

[34] Gregory Mahlon and Stephen Parke. "Improved spin basis for angular correlation studies in single top quark production at the Fermilab Tevatron". English (US). In: *Physical Review D - Particles, Fields, Gravitation and Cosmology* 55.11 (Jan. 1997), pp. 7249–7254. ISSN: 0556-2821. DOI: 10.1103/PhysRevD.55.7249.

[35] J.A. Aguilar-Saavedra and Bernabeu, J. "Breaking down the entire W boson spin observables from its decay". In: *Phys. Rev* 93.1 (Jan. 2016). DOI: 10.1103/PhysRevD.93.011301. arXiv: 1508.04592 [hep-ph].

[36] *Measurement of top quark polarization in t-channel single-top production*. Tech. rep. CMS-PAS-TOP-13-001. Geneva: CERN, 2013. URL: https://cds.cern.ch/record/1601800.

[37] Reinhard Schwienhorst et al. "Single top quark production and decay in the t channel at next-to-leading order at the LHC". In: *Phys. Rev. D* 83 (3 Feb. 2011), p. 034019. DOI: 10.1103/PhysRevD.83.034019. URL: https://link.aps.org/doi/10.1103/PhysRevD.83.034019.

[38] A. Brandenburg, Z. G. Si, and P. Uwer. "QCD-corrected spin analysing power of jets in decays of polarized top quarks". In: *Physics Letters B* 539 (July 2002), pp. 235–241. DOI: 10.1016/S0370-2693(02)02098-1. eprint: hep-ph/0205023.

[39] M. Aaboud and etal. "Probing the W tb vertex structure in t-channel single-top-quark production and decay in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *Journal of High Energy Physics* 2017.4 (Apr. 2017), p. 124. ISSN: 1029-8479. DOI: 10.1007/JHEP04(2017)124. URL: https://doi.org/10.1007/JHEP04(2017)124.

[40] Z. Ligeti (LBNL) A. Ceccucci (CERN) and Y. Sakai (KEK). "THE CKM QUARK-MIXING MATRIX". In: (2014). URL: http://pdg.lbl.gov/2014/reviews/rpp2014-rev-ckm-matrix.pdf.

[41] J. A. Aguilar-Saavedra et al. "TopFit program, Probing anomalous Wtb couplings in top pair decays". In: *European Physical Journal C* 50 (Apr. 2007), pp. 519–533. DOI: 10.1140/epjc/s10052-007-0289-4. eprint: hep-ph/0605190.

[42] M. Aaboud and etal. "Measurement of the W boson polarisation in $t\bar{t}$ events from pp collisions at $\sqrt{s} = 8$ TeV in the lepton + jets channel with ATLAS". In: *The European Physical Journal C* 77.4 (Apr. 2017), p. 264. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-4819-4. URL: https://doi.org/10.1140/epjc/s10052-017-4819-4.

[43] CMS Collaboration. "Search for anomalous Wtb couplings and flavour-changing neutral currents in t-channel single top quark production in pp collisions at $\sqrt{s}$ = 7 and 8 TeV". In: *ArXiv e-prints* (Oct. 2016). arXiv: 1610.03545 [hep-ex].

[44] L. Lyons. "Bayes and Frequentism: a particle physicist's perspective". In: *Contemporary Physics* 54 (Feb. 2013), pp. 1–16. DOI: 10.1080/00107514.2012.756312. arXiv: 1301.1273 [physics.data-an].

[45] G. Punzi. "Sensitivity of Searches for New Signals and Its Optimization". In: *Statistical Problems in Particle Physics, Astrophysics, and Cosmology*. Ed. by L. Lyons, R. Mount, and R. Reitmeyer. 2003, p. 79. eprint: physics/0308063.
[46] CERN. *About Cern*. URL: https://home.cern/about.
[47] CERN. *The accelerator complex*. 2008. URL: http://public-archive.web.cern.ch/public-archive/en/research/AccelComplex-en.html.
[48] CERN. *The Large Hadron Collider*. URL: https://home.cern/topics/large-hadron-collider.
[49] CERN. *How an Accelerator Works*. 2018. URL: https://home.cern/about/how-accelerator-works.
[50] CERN. *Radiofrequency Cavities*. 2018. URL: https://home.cern/about/engineering/radiofrequency-cavities.
[51] The ATLAS Collaboration et all. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (2008), S08003. URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08003.
[52] Prof. C. Anastasiou et all. "Particle Physics Phenomenology 1 – HS 2010". In: ed. by A. Reimer. 2010. Chap. Accelerators and collider experiments. URL: http://edu.itp.phys.ethz.ch/hs10/ppp1/PPP1_4.pdf.
[53] Werner Herr and B Muratori. "Concept of luminosity". In: (2006). URL: http://cds.cern.ch/record/941318.
[54] J G Panduro Vazquez. *The ATLAS Data Acquisition System: from Run 1 to Run 2*. Tech. rep. ATL-DAQ-PROC-2014-035. Geneva: CERN, Oct. 2014. URL: https://cds.cern.ch/record/1954156.
[55] A. Hoecker. "Physics at the LHC Run-2 and Beyond". In: *ArXiv e-prints* (Nov. 2016). arXiv: 1611.07864 [hep-ex].
[56] C. Patrignani and Particle Data Group. "Review of Particle Physics". In: *Chinese Physics C* 40.10 (2016), p. 100001. URL: http://stacks.iop.org/1674-1137/40/i=10/a=100001.
[57] CERN. *The Inverse Femtobarn*. URL: http://writing-guidelines.web.cern.ch/entries/inverse-femtobarn.
[58] Roderik Bruce et al. "LHC Run 2: Results and challenges". In: CERN-ACC-2016-0103 (July 2016), MOAM5P50. 7 p. URL: https://cds.cern.ch/record/2201447.
[59] ATLAS. *ATLAS Experiment - Public Results*. 2018. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.
[60] ATLAS. *ATLAS Experiment - Public Results*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.
[61] ATLAS. *Detector and Technology*. URL: https://atlas.cern/discover/detector.
[62] John Alison. "The Road to Discovery: Detector Alignment, Electron Identification, Particle Misidentification, WW Physics, and the Discovery of the Higgs Boson". PhD thesis. University of Pennsylvania, 2012. URL: https://hep.uchicago.edu/~johnda/thesis/alison_thesis.pdf.
[63] JabberWok Wikipedia-en. *A graphic showing the relationship between angle and pseudorapidity*. 2007. URL: https://commons.wikimedia.org/wiki/File:Pseudorapidity2.png.
[64] CERN. *The Grid: A system of tiers*. 2018. URL: https://home.cern/about/computing/grid-system-tiers.
[65] G. Aad et al. "Measurement of the top quark pair production cross-section with ATLAS in the single lepton channel". In: *Physics Letters B* 711 (May 2012), pp. 244–263. DOI: 10.1016/j.physletb.2012.03.083. arXiv: 1201.1889 [hep-ex].

[66] S. Agostinelli et.all. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. URL: http://www.sciencedirect.com/science/article/pii/S0168900203013688.

[67] Simone Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. In: *Journal of High Energy Physics* 2010.6 (June 2010), p. 43. ISSN: 1029-8479. DOI: 10.1007/JHEP06(2010)043. URL: https://doi.org/10.1007/JHEP06(2010)043.

[68] T. Sjöstrand, S. Mrenna, and P. Skands. “PYTHIA 6.4 physics and manual”. In: *Journal of High Energy Physics* 5, 026 (May 2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. eprint: hep-ph/0603175.

[69] T. Gleisberg et al. “Event generation with SHERPA 1.1”. In: *Journal of High Energy Physics* 2, 007 (Feb. 2009), p. 007. DOI: 10.1088/1126-6708/2009/02/007. arXiv: 0811.4622 [hep-ph].

[70] ATLAS Collaboration. “Measurement of the inclusive cross-sections of single top-quark and top-antiquark t-channel production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *ArXiv e-prints* (Sept. 2016). arXiv: 1609.03920 [hep-ex].

[71] M. Aaboud and etal. “Fiducial, total and differential cross-section measurements of t-channel single top-quark production in pp collisions at 8 TeV using data collected by the ATLAS detector”. In: *The European Physical Journal C* 77.8 (Aug. 2017), p. 531. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5061-9. URL: https://doi.org/10.1140/epjc/s10052-017-5061-9.

[72] *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS*. Tech. rep. ATLAS-CONF-2015-037. Geneva: CERN, Aug. 2015. URL: https://cds.cern.ch/record/2044941.

[73] John R. Koza et al. “Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming”. In: *Artificial Intelligence in Design ’96*. Ed. by John S. Gero and Fay Sudweeks. Dordrecht: Springer Netherlands, 1996, pp. 151–170. DOI: 10.1007/978-94-009-0279-4_9. URL: https://doi.org/10.1007/978-94-009-0279-4_9.

[74] Pushpalatha C. Bhat. “Advanced Analysis Methods in Particle Physics”. In: *FermiLab Techincal Publications* 10.425 ().

[75] Andreas Zell. *Simulation Neuronaler Netze [Simulation of Neural Networks] (in German) (1st ed.)*. Addison-Wesley, 1994, p. 73. ISBN: 3-89319-554-8.

[76] “NeuroComputing”. In: chap. Neural Network Basics. URL: http://www.uta.fi/sis/tie/neuro/index/Neurocomputing2.pdf.

[77] *The NeuroBayes User’s Guide*. Physics Information Technologies. 2013.

[78] M. Feindt and U. Kerzel. “The NeuroBayes neural network package”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 559.1 (2006). Proceedings of the X International Workshop on Advanced Computing and Analysis Techniques in Physics Research, pp. 190–194. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2005.11.166. URL: http://www.sciencedirect.com/science/article/pii/S0168900205022679.

[79] Gerard Nico. *Machine Learning - (Overfitting | Overtraining | Robust | Generalization) (Underfitting)*. 2018. URL: https://gerardnico.com/data_mining/overfitting.

[80] Z. Nagy and D. E. Soper. “What is a parton shower?” In: *ArXiv e-prints* (May 2017). arXiv: 1705.08093 [hep-ph].

[81] A. Hoecker et al. "TMVA - Toolkit for Multivariate Data Analysis". In: *ArXiv Physics e-prints* (Mar. 2007). eprint: physics/0703039.