Statistical methods used in the search for singly produced vector-like quarks using the ATLAS detector

Anjishnu Bandyopadhyay

Masterarbeit in Physik
angefertigt im Physikalischen Institut

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

June 2015

I hereby declare that this thesis was formulated by myself and that no sources or tools other than those cited were used.

Bonn, ...
Date Signature

1. Gutachter: Prof. Dr. Ian C. Brock
2. Gutachterin: Prof. Dr. Jochen C. Dingfelder

For my grandmother

Acknowledgements

Contents

CHAPTER 1

Introduction

"*The effort to understand the universe is one of the very few things which lifts human life a little above the level of farce and gives it some of the grace of tragedy.*" -- Steven Weinberg

The field of high energy physics tries to understand the nature of the fundamental building blocks of the Universe and the different interactions that occur between them. The Standard Model of particle physics and the ΛCDM model of cosmology are the best available theories to describe our universe. Experiments like the Large Hadron Collider (LHC) and the Planck Mission have confirmed these theories to high degrees of precision. But these theories can only explain 5% [1] of the Universe. The exact nature of the remaining 95% is still unknown to us. The Standard Model of particle physics along with Einstein's theory of General Relativity is an elegant description of the 5% of the ordinary matter present in our universe. The Standard Model describes the electroweak and strong forces and Generally Relativity describes gravity. But till now physicists have not found an unified theory to describe these forces. Unfortunately, the Standard Model of particle physics is not a complete theory. It presents us with some open questions. One of them is the instability of the mass of the Higgs boson. The loop corrections to the mass of the Higgs boson are quadratically sensitive to the next energy scale of new physics. This problem is called the *hierarchy problem* [2, 3].

The experimentally measured value of the Higgs mass is 125.7 GeV [4, 5] which is clearly lower than the next energy scale. So there has to be a theoretical solution to this problem. There are lot of theories that try to solve this problem, but to date physicists haven't found any experimental evidence of such theories. In this thesis we conduct a search for particles predicted by two such theories. The corrections to the Higgs boson mass comes from its interaction with the fermions of the Standard Model. Since the top-quark, with a mass of 173.2 GeV, is the heaviest fermion of the Standard Model, the major contributions to these corrections come from the top-quark. We consider an effective field theory which adds heavier partners to the top and bottom quark which cancel these corrections. In the second effective field theory we consider, the Higgs boson is assumed to be a composite particle. This model also adds heavier top and bottom partners to solve the hierarchy problem. We expect these theories to have a mass hierarchy like the Standard Model. In the Standard Model the bottom-quark is much lighter than the top-quark. Thus, we expect the heavier partner of bottom-quark to be more accessible to experimental searches. In this thesis we search for the two bottom partners of the aforementioned theories, b^* and B'. We use the proton-proton collision data taken by the ATLAS detector of the Large Hadron Collider for this search. We use the full dataset taken at a centre-of-mass energy $\sqrt{s} = 8$ TeV in 2012. We perform this search in the decay channel $b^*/B' \rightarrow Wt$. The different statistical methods used in this search are presented in this thesis.

The biggest debate in the world of statistics is the validity and accuracy of Bayesian and frequentist statistical methods. Rather than engaging in this debate by taking sides, physicists have tried to compare and combine these methods from an unbiased point of view. There are advantages and disadvantages

of both the methods. Frequentist methods have the advantage that it gets its results from repeatable unbiased experiments, but the definition of probability is not very clear in frequentist methods. Bayesian methods help us to incorporate the prior information about our experiments and also Bayesian methods can use the classical definition of probability. One disadvantage of Bayesian methods is that they can inject a bias with the prior information. In our analysis we use both frequentist and Bayesian methods. We use frequentist methods to search for the presence of a signal in our collision data. We use Bayesian and frequentist methods to set limits on our signal strength. This analysis is under preparation for a publication from the ATLAS collaboration. As decided there, in this thesis also the Bayesian limit is quoted as the final result.

In Chapter 2, we introduce some basic concepts of particle physics and describe the Standard Model. We also present the theoretical motivation behind using our models. In Chapter 3 an overview of the different components of the ATLAS detector is presented. Chapter 4 illustrates the final states of the decay modes we consider in our analysis. We also give an overview of the object reconstruction used in our analysis in this chapter. A detailed description about the statistical methods used in this analysis is given in Chapter 5. The statistical tools used to implement these methods is described in Chapter 6. The details of our statistical analysis is presented in Chapter 7. In Chapter 8 we present the results of our analysis. Finally, we conclude and summarise our findings in Chapter 9.

CHAPTER 2

Theoretical concepts

"*The electron is a theory we use; it is so useful in understanding the way nature works that we can almost call it real.*" -- Richard Feynman

The field of particle physics tries to understand the nature of the electroweak and strong interactions. The Standard Model of particle physics is to date the best theoretical model to describe these forces, but unfortunately it is not a complete theory. In this chapter I would first introduce some basic ideas of particle physics and then go on to describe the details and drawbacks of the Standard Model. In the last section I would describe some theories which try to solve some of the shortcomings of the Standard Model.

2.1 Basic terminology in particle physics

Natural units

In most fields of physics the S.I. system of units is used as a measure of the different quantities used. But in particle physics we use a system of units called *natural units* to describe the different physical variables. In this system of units we consider the Planck's constant, speed of light in vacuum, and gravitational constant to be one i.e. $\hbar = c = G = 1$, and also define all physical quantities in terms of units of energy. We use electronvolt (eV) as the fundamental unit. Most particle physics processes occur at higher energies, thus making MeV or GeV the more commonly used units. Physical quantities like mass and momenta are written in terms of MeV or GeV.

Physical observables

The goal of any theoretical model is to predict the values of some physical observables which can be measured to validate the given model. Most particle physics experiments are scattering or decay experiments. So theoretical models predict the values of physical observables which can be measured in such experiments.

Cross-section and Luminosity Cross-section is the most crucial observable measured in scattering experiments. The cross-section of a process, σ, is the effective area where the process takes place. That gives us a measure of the probability of the occurrence of the given process. It can be computed theoretically using

$$\sigma \propto \int |\mathcal{M}|^2 d\rho \tag{2.1}$$

where $|\mathcal{M}|^2$ is the square of the matrix element of the process and $\int d\rho$ is the integral over the phase space. The matrix element $|\mathcal{M}|$ gives us a measure of the transition amplitude of the transition from the initial to final state. The cross-section is associated with the number of events N as following

$$\frac{dN}{dt} = \sigma \cdot \mathcal{L} \tag{2.2}$$

where $\mathcal{L}$ is called the *instantaneous luminosity* and $\mathrm{L} = \int \mathcal{L}\, dt$ is known as integrated luminosity.

Decay width and branching fraction The particles produced in a scattering experiment decay into different number of particles through various processes. Suppose a particle decays with a lifetime τ then the *decay width* Γ is defined as the reciprocal of the lifetime τ

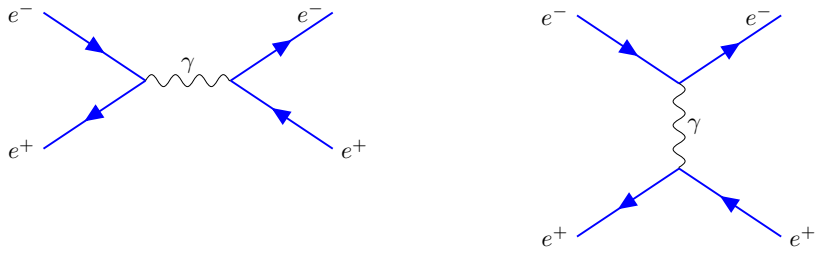
$$\Gamma = \frac{1}{\tau} \tag{2.3}$$

Most particles can decay via different processes. Each of the individual processes have a decay width Γ_i. Then the total decay width Γ_{total} and *branching fraction* $\mathcal{B}_i$ for the i-th process can be defined as

$$\Gamma_{total} = \sum_i \Gamma_i \tag{2.4}$$

$$\mathcal{B}_i = \frac{\Gamma_i}{\Gamma_{total}} \tag{2.5}$$

Feynman diagrams Richard Feynman formulated a very elegant and simple method of representing scattering processes and calculating the corresponding matrix element $|\mathcal{M}|$ under the domain of perturbation theory [6]. Feynman diagrams are space-time diagrams that illustrate the various interactions between the particles and by using Feynman rules the value of $|\mathcal{M}|$ can be calculated from these diagrams. We will use two diagrams of the Bhabha (electron-positron) scattering to present some details about Feynman diagrams. In this thesis we assume that time flows from left to right. In Figures 2.1a



(a) s-channel diagram (b) t-channel diagram

Figure 2.1: Feynman diagrams for Bhabha scattering

and 2.1b we have drawn the scattering of $e^+e^- \to e^+e^-$. While drawing Feynman diagrams, we represent incoming particles (anti-particles) by incoming (outgoing) arrows and outgoing particles (anti-particles) by outgoing (incoming) arrows. The lines between the vertices represent "off-shell" virtual particles. All physical conservation laws are obeyed at these vertices. Diagrams like 2.1a and 2.1b are called s-channel and t-channel diagrams respectively. The matrix element $|\mathcal{M}| \propto 1/s$ in s-channel diagrams, where s is the centre-of-mass energy. $s = (p_{e^+} + p_{e^-})^2$ for Bhabha scattering. In case of t-channel diagrams, $|\mathcal{M}| \propto 1/t$, where $t = (p_A - p_C)^2$ for processes of the form $A + B \to C + D$. This framework is

extremely useful for calculating the theoretical cross-sections and decay rates of complicated physical processes.

2.2 Standard Model

The discovery of the Higgs boson [4, 5] has made sure that the Standard Model (SM) is the best description of the electroweak and strong interactions. The Standard Model predicts the existence of different fundamental particles called **quarks**, **leptons**, **gauge bosons**, and the **Higgs boson**. To our best knowledge, quarks and leptons are the most fundamental building blocks of matter. The different gauge bosons are the force carriers of the electroweak and strong interactions. The Higgs boson helps these particles to gain mass. An overview of the different particles and the allowed interactions is shown in the Figure 2.2 All these particles have been experimentally detected. Unfortunately this elegant theoretical model is incomplete. We describe those problems in Section 2.2.5. In this section we describe the theoretical concepts of Standard Model.

2.2.1 Lagrangian

The Standard Model is a combination of different quantum field theories. Quantum fields are quantities that assume given values for all points in space-time. In quantum field theories, the quantum fields are the fundamental quantities and the space-time coordinates are treated as labels. So instead of using the Lagrangian we use the Lagrangian density. The Lagrangian density $\mathcal{L}$ can be written in terms of the Lagrangian L as

$$L = \int d^4x \, \mathcal{L} \tag{2.6}$$

The quantum field theories that constitute the Standard model are a special type of theory called *gauge theories*. These theories have certain gauge symmetries. The simplest example of a gauge symmetry is the symmetry of A_μ in electromagnetism. The Lagrangian of the Standard Model is invariant under certain gauge transformations. Because of these symmetries, these theories can be embedded into groups called Lie groups. The different particles are the irreducible representations of these groups. And the gauge bosons are the generators of these groups.

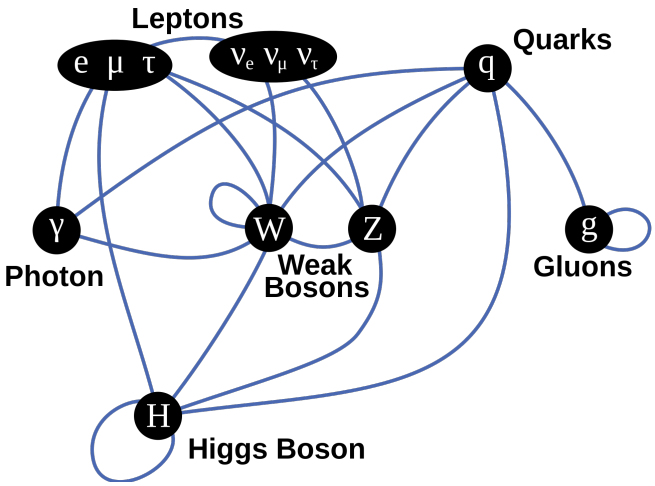


Figure 2.2: An overview of the different particles and the allowed interactions of the Standard Model

The Standard Model can be represented by a $SU(3) \times SU(2) \times U(1)$ group . The $SU(3)$ part describes the strong sector and the $SU(2) \times U(1)$ part represents the electroweak sector. The Standard Model Lagrangian[1] $\mathcal{L}$ can be written as [7]

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_{\text{bosons}} + \mathcal{L}_{\text{fermions}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{Higgs}} \\ &= (-\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu}) + (i\bar{\psi}^i\gamma^\mu D_\mu\psi^i) + (\bar{\psi}^i_L V_{ij}\Phi\psi^j_R + h.c.) + (|D_\mu\Phi|^2 - V(\Phi))\end{aligned} \tag{2.7}$$

The first and second part of the equation represents the different interactions between the gauge bosons, quarks and leptons. In the first term, $F^a_{\mu\nu}$, the stress-energy tensor, gives us a measure of the self-interactions and kinetic energies of the gauge bosons. The second term, the Dirac term, is a measure of the kinetic energies of the leptons and quarks and their interactions with the gauge bosons. The third term is the Yukawa term which represents the mass terms for different particles. The last term represents the contribution from the Higgs field.

2.2.2 Electroweak interactions

The Glashow-Salam-Weinberg theory [8–10] of electroweak interactions is an unified theory describing both electromagnetic and weak interactions. This theory was a further development of the theory of quantum electrodynamics [11–13] developed by Feynman, Schwinger, and Tomonoga. This theory describes the leptons, W and Z bosons. The Glashow-Salam-Weinberg theory can be embedded into a $SU(2)_L \times U(1)_Y$ group. The subscript L denotes that the left-handed particles described by this theory can be embedded into the doublet representation of a $SU(2)$ group. The electroweak theory is a parity violating theory. The subscript Y denotes hypercharge, a linear combination of the electric charge and isospin of the different particles. The different left-handed leptons can be written as doublets and the right-handed leptons can be written as singlets of this group.

$$\psi_L = \begin{pmatrix}\nu_e \\ e\end{pmatrix}_L , \begin{pmatrix}\nu_\mu \\ \mu\end{pmatrix}_L , \begin{pmatrix}\nu_\tau \\ \tau\end{pmatrix}_L \tag{2.8}$$

$$\psi_R = e_R,\ \mu_R,\ \tau_R \tag{2.9}$$

$e_{L,R}$, $\mu_{L,R}$, and $\tau_{L,R}$ denote the left-handed and right-handed electron, muon and tau respectively. ν_e, ν_μ, and ν_τ are the electron, muon and tau neutrinos respectively. The electroweak theory being a parity violating theory doesn't contain right-handed neutrinos. There are four vector fields $W^{1,2,3}_\mu$ and B_μ which contribute to the electroweak interactions in the first and second term of (2.7). A linear combination of W^1_μ and W^2_μ describe the two charged W bosons and two different linear combinations of W^3_μ and B_μ represent the two neutral bosons, the Z boson and the photon.

$$W^\pm_\mu = \sqrt{\frac{1}{2}}(W^1_\mu + iW^2_\mu) \tag{2.10}$$

$$A_\mu = B_\mu \cos\theta_W + W^3_\mu \cos\theta_W \tag{2.11}$$

$$Z_\mu = -B_\mu \cos\theta_W + W^3_\mu \cos\theta_W \tag{2.12}$$

$$\tan\theta_W = \frac{g'}{g} \tag{2.13}$$

[1] From here on Lagrangian means Lagrangian density

where cos θ_W us called the Weinberg angle and it associates the respective coupling strengths of W^i_μ and B^μ, g and g', and the charge via the relation (2.13). These particles gain mass via the Higgs mechanism, the details of which will be given later.

2.2.3 Strong interactions

Quantum chromodynamics

Quantum chromodynamics (QCD) [14] is the theory of the strong interactions based on a gauge symmetry. The quarks and the eight gluons predicted by this theory can be represented by the $SU(3)_c$ group. The quarks and anti-quarks can be represented as colour triplets. There are three types of colour charges, named as red, green and blue. The eight gluons are the generators of the $SU(3)$ group which are the gauge bosons mediating the strong interaction. These gluons also carry colour charges and thus they interact with each other. They contribute to the strong interactions in the first and second term of (2.7). There are three generations of quarks, each generation having two quarks. The six quarks, up, down, charm, strange, top, and bottom can be divided into three generations in the following way.

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix} \tag{2.14}$$

Asymptotic freedom The existence of self-interaction between gluons leads to an "anti-screening" effect. This effect is larger than the screening due to virtual gauge bosons by a factor of $11N_c - 3N_f$, where N_c and N_f are the number of colours and flavours of the interacting quarks respectively. Due to this, the strong coupling constant, α_s, decreases with increasing energies or decreasing distances.

$$\alpha_s(Q) \propto \frac{1}{\ln(\frac{Q}{\Lambda})} \tag{2.15}$$

This implies that at distances larger than $\sim \frac{1}{\Lambda}$ the quarks interact strongly [15]. Experiments show that these distances are close to the sizes of light hadrons. So beyond these distances the quarks are not free and they produce different bound states called mesons (quark anti-quark pair) and baryons (three quarks or anti-quarks). This phenomenon is called *confinement*. Since the strong coupling is weak at distances $< \frac{1}{\Lambda}$ the quarks are asymptotically free inside the hadrons. This feature of the strong force is known as *asymptotic freedom* [16].

Partons In the Large Hadron Collider beams of protons are collided together. Due to the asymptotic freedom of quarks, the inelastic scattering between protons can be thought of as collision between different constituents of the proton. The different quarks and gluons that make up a proton are called partons. When two protons scatter the partons carry a fraction of its momentum denoted by $x = \frac{p_{parton}}{p_{proton}}$. In that case we can write the scattering cross-section as

$$\sigma(p_1\ p_2 \to X) = \int_0^1 \int_0^1 \sum_{i,j=q,\bar{q},g} f_i(x_i)\ f_j(x_j)\ \sigma_{i,j}(ij \to X) \tag{2.16}$$

where $f_i(x_i)$ and $f_j(x_j)$ are the respective parton density functions (PDF) which describe the probabilities of finding the partons with momentum fraction x_i inside the proton. Here $\sigma_{i,j}(ij \to X)$ denotes the cross-section of the different parton scattering processes. These individual partonic cross-sections have been extensively measured in experiments as they cannot be calculated by theory. Collaborations like CTEQ

use this data and perform global fits to it to find the precise values of PDFs which can be used for cross-section calculation. The values of the PDFs are used as inputs in our experiments. There is a factorisation scheme which decides which partons contribute to the scattering cross-section of a given process.

2.2.4 Higgs mechanism

The Higgs mechanism [17, 18] is a way of giving masses to the different particles of the Standard Model via "spontaneous symmetry breaking". We add an extra term to the Lagrangian $\mathcal{L}_{\text{Higgs}}$ given by

$$\begin{aligned} \mathcal{L}_{Higgs} &= |D_\mu \Phi|^2 - V(\Phi) \\ &= |D_\mu \Phi|^2 + \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2 \end{aligned} \tag{2.17}$$

where Φ is a complex doublet. The minimum of this potential is at

$$v = \left(\frac{\mu^2}{\lambda}\right)^{1/2} \tag{2.18}$$

Subsequently, the Higgs field acquires a vacuum expectation $\begin{pmatrix} 0 \\ v \end{pmatrix}$ value (vev). If we expand the Lagrangian around this vev the vector bosons gather a mass term. In addition to this we get some mass terms for a new scalar particle ϕ called the *Higgs boson*. The mass of the gauge bosons and the Higgs boson are given by

$$m_W = g \cdot \frac{v}{2} \tag{2.19}$$

$$m_Z = \sqrt{g^2 + g'^2} \cdot \frac{v}{2} \tag{2.20}$$

$$m_\phi = \sqrt{2\lambda} \cdot v \tag{2.21}$$

where $v \approx 246$ GeV [19]. The measured properties of the Higgs boson, discovered by ATLAS and CMS in 2012 [4, 5] imply the fact that this Higgs boson is most probably the Standard Model Higgs boson.

Yukawa mass terms The third term in (2.7) contains the Yukawa mass terms for the quarks and leptons. This term describes the interaction of the Higgs boson with all the fermions of the Standard Model. The gauge invariance condition of the Standard Model only allows these kind of terms. We replace the Higgs field by its vev to get the corresponding mass terms.

$$m_{l,q} = \frac{1}{\sqrt{2}} \lambda_{l,q} \cdot v \tag{2.22}$$

The term V_{ij} in (2.7) represents the quark-mixing matrix. It is known as the Cabibo-Kobayashi-Maskawa (CKM) matrix [20, 21]. The elements of this matrix represent the strength of the interactions between the different quarks. The experimental values are such that the diagonal elements of the CKM matrix are close to unity and the off-diagonal terms are one or two orders of magnitude smaller.

2.2.5 Drawbacks of the Standard Model

Although being a fairly complete model, the Standard model has some drawbacks and limitations. Theorists have suggested many theories which try to solve most of the problems of Standard Model, but unfortunately there have been no experimental evidence of these theories. Experiments like the Large Hadron Collider (LHC), Large Underground Xenon (LUX) experiment etc. will be looking for evidences of such theories in the near-future. Some of the major drawbacks of the Standard Model are:

- **Hierachy problem** The particles of the Standard Model gain their masses via their interaction with the Higgs boson. However the one-loop corrections to the Higgs-boson mass arising from the SM fermions are quadratically sensitive to Λ, the next scale of new physics. The top-quark being the heaviest fermion in the SM, contributes most to these corrections. This problem is known as the *hierachy problem* [2, 3].

$$\delta m_H^2 \propto |y_t|^2 \Lambda^2 + c \ln\Lambda \tag{2.23}$$

  y_t being the Yukawa coupling between the top-quark and the Higgs boson. There are theories which try to cancel out these corrections, some of which will be discussed later. Since the experimental value of the Higgs mass is 125.7 GeV, these corrections have to be cancelled theoretically. energy scales.

- **Matter-antimatter asymmetry** The Standard Model doesn't predict any matter-antimatter asymmetry. In our universe the ratio of antimatter to matter is negligibly small. To achieve this asymmetry theories have to be CP (charge-parity) violating and they also have to violate baryon and lepton number conservation. Theories like leptogenesis try to achieve these conditions, but these theories are experimentally constrained.

- **Dark matter and energy** The Standard Model is a theory that can only describe $\sim 5\%$ of the Universe. It cannot accommodate any description about dark matter and energy. Calculations from the relic densities of cold dark matter suggest that they should interact at the weak scale. Some extensions of the SM predict very good candidates, but they haven't been detected experimentally.

- **Naturalness problem and free parameters** The different parameters of the Standard Model have values ranging across six orders of magnitude. The SM has 19 free parameters, the values of which are experimentally determined. There is no theoretical justification of these features of the SM. The hierarchy of the different parameters is also a strong hint of the fact that there might be new physics at the TeV scale.

2.2.6 Theoretical solutions

There are a lot of Beyond the Standard Model (BSM) theories which try to address some of the issues mentioned above. Since Hierarchy problem is one of the major drawbacks of the SM, we will discuss some theories which have predicted some solutions to it. In this thesis we have tried to find experimental evidence for some of these theories.

Supersymmetry Supersymmetry (SUSY) [22] is a class of theories that solves some of the drawbacks of the Standard Model. It solves the hierarchy problem by adding bosonic partners to the different fermions of the SM and vice-versa. This ensures that the quadratic term mentioned in (2.23) gets cancelled by the corresponding bosonic partners of the fermions. These theories also predict promising

candidates for dark matter. There are different supersymmetric theories like MSSM, CMSSM, NMSSM etc. These theories have different assumptions, thus having different properties of the predicted particle spectra. The number of free parameters also vary in each of these theories. The existence of supersymmetric particles have not been experimentally proved until now. Experimental constraints have been set on the properties of the supersymmetric particle spectra.

Effective field theories Effective field theories [23] are quantum field theories that are low-energy approximation of some more general physical theory. Although effective field theories are not exact theories, but they provide us with prospective solutions to the different problems. In these theories higher dimensional operators are added to the Standard Model Lagrangian. These terms give rise to interactions which cancel the quadratic divergences. These theories are valid up to the scale of the more general theories. Since these broader theories are not known, we are free to choose this scale. We can use experimental data to see in what energy ranges these theories are valid or not. In this thesis we search for the existence of new particles predicted by two such effective field theories.

2.3 Vector-like quarks

There is no theoretical threshold on the number of quark generations in nature. Experiments have excluded the existence of a fourth generation SM-like chiral quark [24], but these experiments do not set any constraints on models of vector-like quarks (VLQ), fermions that can have equal left and right-handed couplings to the W-boson. These particles help us to solve the Higgs mass hierarchy problem. Theories like SUSY [25], composite Higgs models [26, 27], little Higgs models [28] etc. predict these VLQs, thus making them an exciting tool for looking for new physics. In this thesis we consider two VLQ models, b^* and B', the first being predicted by an effective field theory and the second by an effective composite Higgs model.

2.3.1 b^*

Effective field theories which try to solve the hierarchy problem generally add an extra top-quark doublet which can cancel the quadratic divergences. So in these models we have a lighter bottom-quark partner which is more accessible to experiments. Here an effective field theory is considered where b^*, a down-type quark, [29] is the only state below the energy cutoff scale of the theory. In addition to being a VLQ, the production of b^* occurs through an anomalous chromomagnetic coupling. We consider the decay of b^* into a W boson and a top-quark. The Feynman diagram of this process is shown in Figure (2.3). The bgb^* coupling is a flavour changing vertex, but since flavour changing neutral currents (FCNC) are assumed to be absent in this theory, the production cross-section is FCNC suppressed. The decay width is proportional to f_g^2 and $(f_L^2 + f_R^2)$, where we assume $f_g = f_L = f_R = 1$. Here f_g quantifies the bgb^* coupling and f_L and f_R denotes the coupling of b^* with the W boson.

2.3.2 B'

B' is a VLQ predicted by a composite Higgs model [27, 30]. In such effective theories the electroweak symmetry breaking is mediated by a strong-interaction-like dynamics. The breaking is achieved via couplings between elementary and composite fermions. This model is derived from a TS-10 model (two-site), the two sites denoted by the composite and the elementary sector. The composite fermions are embedded in a 10-dimensional representation of $SO(5) \times U(1)_X$ which decomposes as $SU(2)_L \times SU(2)_R \times U(1)_X$. The elementary sector is written as the SM-group without the Higgs boson. B' can

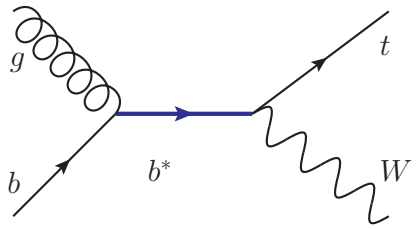


Figure 2.3: Feynman diagram for $pp \to b^* \to Wt$

decay into Zb, Hb (H is the SM Higgs), and Wt with branching fractions having a ratio of 2:1:1. Here we consider a t-channel production via a Z boson decaying to Wt. The final state also contains an extra light quark. The Feynman diagram is shown in Figure 2.4. The cross-section is proportional to λ^2, λ being the coupling of B' to the electroweak sector.

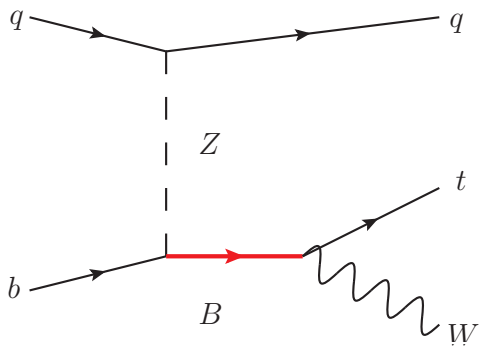


Figure 2.4: Feynman diagram for $pp \to B' \to Wt$

CHAPTER 3

LHC and ATLAS

"*Experimentation is the least arrogant method of gaining knowledge. The experimenter humbly asks a question of nature.*"-- Issac Asimov

The Large Hadron Collider (LHC) was built to cement our understanding of the Standard Model and explore the possibilities of the existence of Beyond the Standard Model (BSM) physics. It has operated from 2009 to 2013 for Run I at centre-of-mass energies, $\sqrt{s}$ = 7 and 8 TeV. It was shut down for the next two years for an upgrade. After the discovery of Higgs boson, the LHC will now operate for Run II from 2015 to 2018 at centre-of-mass energies of 13 and 14 TeV to look for BSM physics. In the first part of this chapter we will give an overview of the LHC. In the second part, we illustrate the working of the ATLAS detector, the one which was used for our analysis.

3.1 The Large Hadron Collider

The LHC [31] is a circular proton-proton collider built in the former Large Electron Positron (LEP) collider tunnel. The 26.7 km long collider is housed 100 m underground at the European Center for Nuclear Research (CERN) in Geneva. This grand machine was designed to operate at $\sqrt{s}$ = 14 TeV with instantaneous luminosity of 10^{34} cm^2 s^{-1}. The LHC has taken 20.3 fb^{-1} of data in 2012 at $\sqrt{s}$ = 8 TeV [32] . In this analysis we have used data from this dataset.

The proton beams are collided at four interaction points. The LHC hosts four detectors at these points. The two general purpose detectors, ATLAS [33] and CMS [34] are used for detecting different final states. LHCb [35] and ALICE [36] are two special purpose detectors located at the two other collision points. LHCb studies B-physics whereas ALICE is designed to study heavy-ion collisions at lower energies. Besides these four experiments there are also three smaller experiments: TOTEM [37], LHCf [37], and MoEDAL [38]. TOTEM will measure the total proton-proton cross-section and other phenomena like the elastic scattering and diffraction dissociation. LHCf measures neutral particle at low angles. The results of this experiment will be used as an input in studying cosmic-ray-induced particle showers in our atmosphere. MoEDAL will look for magnetic monopoles and other exotic particles. A schematic diagram showing all the major experiments is shown in Figure 3.1

The protons used are accelerated in multiple steps before being injected into the LHC ring. First they are accelerated by radio-frequency (RF) cavities up to 750 keV. They are then sent to the proton Linear Accelerator to gain energies of 50 MeV. Subsequently the Proton Synchotron Booster (PSB) boosts the protons to 1.4 GeV and injects them into the Proton Synchrotron (PS). The PS increases the energy of the protons to 25 GeV before sending them to the Super Proton Synchrotron which finally injects them into the LHC ring after accelerating them to energies of 450 GeV. The LHC is designed in a manner such that it can collide 2808 proton bunches with 1.15×10^{11} protons per beam with a bunch spacing of 25 ns.

.

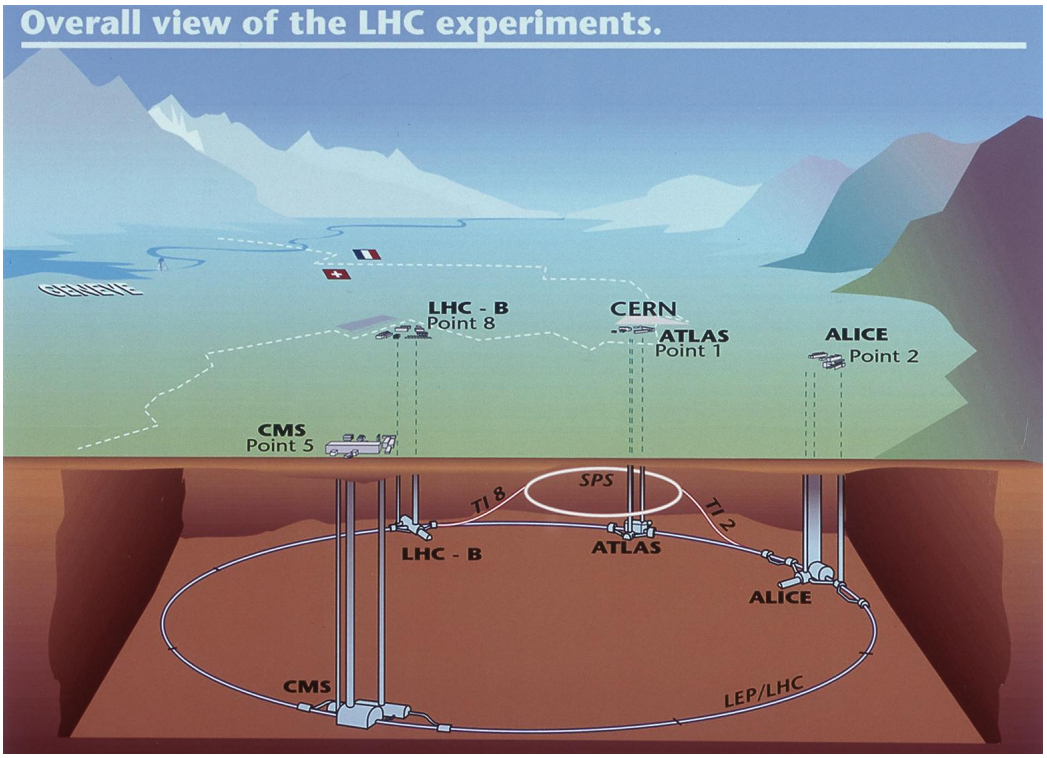


Figure 3.1: A schematic view of the major experiments at LHC

3.2 ATLAS detector

The ATLAS detector is a general purpose detector built to observe all possible final states. The 44 m long, 25 m high, and 7000 ton heavy detector is the largest detector at LHC. The detector is managed by the ATLAS collaboration which also performs analysis on the data acquired by ATLAS. The detector comprises the inner detector, the electromagnetic and hadronic calorimeters, the muon spectrometer, and the magnet systems [39]. An overview of the detector is shown in Figure 3.2

3.2.1 Kinematic observables

Before going into the details of the detector, we would like to describe the kinematic observables used in ATLAS for the various measurements. For measurements, ATLAS uses a right-handed coordinate system. The direction of the beam is assumed to be the z-axis. The transverse plane is taken to be the xy plane. The y direction is in the vertical direction pointing upwards. The x axis is defined in the horizontal direction pointing toward the center of the LHC ring. In reality partons, the different constituents of the proton collide with each other. Each of those partons carries a part of the proton, thus giving us an uncertainty while measuring the total momentum in the z direction. So we use the transverse momentum p_{T} which is defined as

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} \tag{3.1}$$

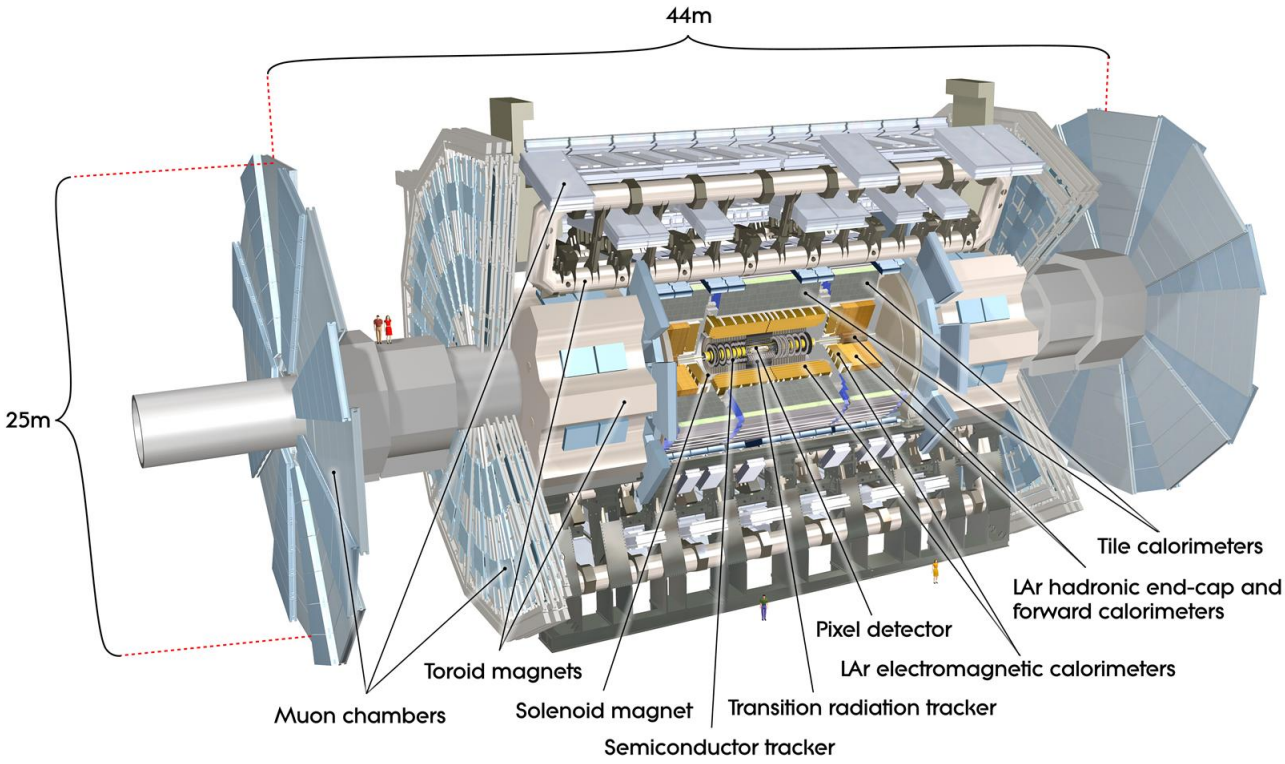


Figure 3.2: An overview of the ATLAS detector

The neutrinos are not detected by the detector due to their weak interactions. So we define a similar variable in the transverse plane $E_{\mathrm{T}}^{\mathrm{miss}}$ to account for that. $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated as the magnitude of the vector imbalance of the momentum in the transverse plane. We use a cylindrical detector, thus making cylindrical co-ordinate system the natural choice inside the detector.

The azimuthal angle ϕ and the polar angle θ are both measured around the beam axis. In our system, we use a variable η as a measure of the polar angle.

$$\eta = -\ln \tan\frac{\theta}{2} \tag{3.2}$$

In the region around the beam pipe, particle densities are roughly flat as a function of η. We also use the distance, ΔR in the $\eta - \phi$ coordinates for our measurements. It is defined as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{3.3}$$

3.2.2 Constituents of the ATLAS detector

Magnets

ATLAS uses both toroidal and solenoid magnets. The inner magnet is a central solenoid located around the beam pipe. It provides a field of 2 T in the inner tracker. The second magnet system is toroidal and is split into three parts, the barrel part is installed around the central calorimeter and two end caps are

installed at the two ends of the detector. It produces a field of 0.5 T in the barrel region and a field of 0.15-2.5 T in the end-cap regions.

Inner detector

The inner detector (ID) of the ATLAS detector is the closest detector component to the beam pipe. The 3.5 m long ID which has a radius of 1.1 m starts at a distance of 5 cm from the interaction point. It is embedded in the solenoid and covers a volume corresponding to $|\eta| < 2.5$. The three main components of the ID are the pixel detector, semiconductor tracker (SCT), and the transition radiation tracker (TRT).

Pixel detector The pixel detector used in ATLAS is a silicon detector. The pixel sensors in this detector have a size of $50 \times 400\ \mu\text{m}^2$. Three cylindrical barrel layers and three discs in each of the end caps make up this detector. The cylindrical structure is made up of pixel modules. These pixel layers have a design accuracy of 10 μm in the $R - \phi$ direction and 115 μm in the z-direction. Such high degree of precision allows us to distinguish between particles with lifetimes of the order of picoseconds (e.g. B hadrons and tau leptons) decaying via a primary vertex and those decaying via a secondary vertex.

SCT THE SCT is the next sub-detector in the ID. In the barrel region, the SCT is made up of four cylindrical double-layers of silicon microstrips and in the end-cap regions it is built of nine layers of discs. The SCT has a spatial resolution of 16 μm in the $R - \phi$ plane and 580 μm in the z direction.

TRT The TRT is the largest part of the ID. It is built from straws-tubes having 4 mm diameter. It has a coverage up to $|\eta| \leq 2$. Each of these straw-tubes acts as a drift chamber. The straws are placed along the z-axis. The TRT has 351,000 readout channels. It has the lowest tracking precision among the sub-components of the ID, but it can combine precision measurements from the pixel detector and the SCT at larger radius. Due to the longer track length of the TRT it can contribute significantly to momentum measurements. It also helps us to do a robust pattern recognition and charged particle identification.

Calorimeters

Calorimeters are that part of the detector which contribute towards energy measurements and particle-track identification. ATLAS has two calorimeters, namely the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). The ATLAS calorimeters have a coverage of $|\eta| < 5$.

ECAL The ECAL records electromagnetic showers from electrons, positrons, and photons. It consists of a central region covering the barrel covering $|\eta| < 1.475$ and two co-axial wheels at the end-caps which have a coverage of $1.375 < |\eta| < 3.2$. The ECAL uses liquid argon (LAr) [40] as its primary material and accordion-shaped lead plates as absorbers. The ECAL has a thickness of more than 22 radiation lengths, X_0 (X_0 is the distance needed to factorise the electron energy by $1/e$ of its original value) in the barrel and more than 24 X_0 at the end-caps. This ensures negligible leakage to the HCAL. The resolution of the ECAL is

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\% \tag{3.4}$$

where E is the energy of a particle and σ_E is the RMS of the reconstructed energy probability.

HCAL The HCAL [41] records showers from strongly interacting particles and jets. It occupies the outer barrel region with $|\eta| < 1.6$ and end-cap regions with $1.5 < |\eta| < 3.2$. The barrel region uses scintillating tiles as active material and steel as absorbers whereas the end-caps are LAr calorimeters with copper absorbers. The energy resolution of the HCAL is given by

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\% \tag{3.5}$$

Muon spectrometer

The muon spectrometer is situated at the outer periphery of the ATLAS detector. As the name suggests, it is used to track and measure the momentum of muons. Since muons have a relatively longer lifetime they escape the ECAL and are measured by the muon spectrometer. They also do not make electromagnetic or hadronic showers, thus needing a spectrometer for detection. The measurements in this region strongly rely on the structure of the magnetic field. The central region of $|\eta| < 1$ is dominated by the field of barrel toroids and the end-cap region of $1.4 < |\eta| < 2.7$ is affected by the end-cap toroids. The intermediate region $1.0 < |\eta| < 1.4$ is affected by an interplay of both the fields. The thin gap or the resistive plate chambers is used for switching on the muon trigger.

Trigger system

The ATLAS detector uses three levels of trigger system. The Level 1 (L1) trigger is a completely hardware based trigger. Objects are selected by using information from regions with low granularity. This trigger looks for leptons with high p_{T}, photons and hadronically decaying taus. This trigger has a latency of 2 μs and a maximum acceptance rate of 75 kHz. Muons are triggered for $p_{\mathrm{T}} > 10$ GeV at L1. The central trigger towers ($\eta < 3.2$) are used for triggering electron. L1 uses a $p_{\mathrm{T}} > 14 - 16$ GeV electron trigger. The Level 2 (L2) filter reduces the acceptance rate from 75 kHz to ~ 3 kHz. In this trigger, information from regions located around indicated objects is used. The third trigger or the event filter is used to select events for future analysis. At this level we use an acceptance rate of 300 Hz. This corresponds to data acquisition of 300 MB s^{-1} i.e. 3000 TB per year, thus making storage demanding. The L2 and the event filter comprises the High-Level trigger (HLT).

CHAPTER 4

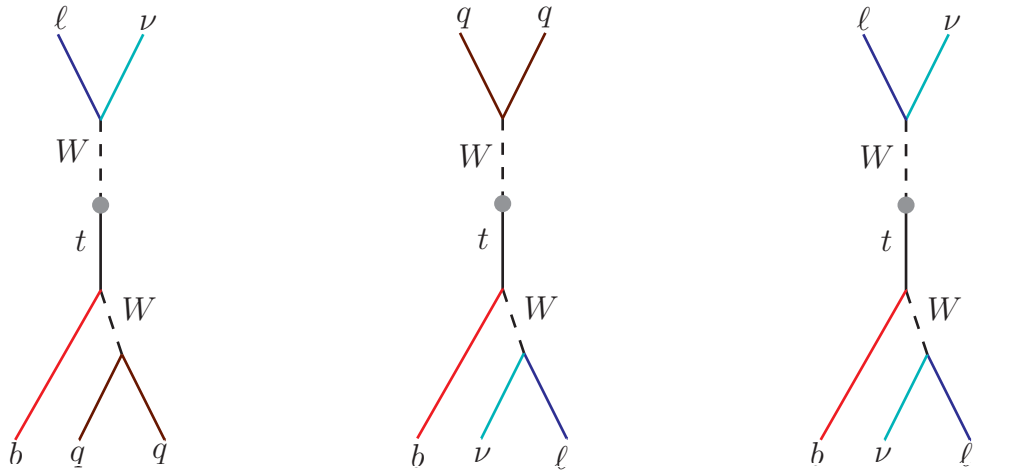
Event selection

"*The unity of all science consists alone in its method, not in its material.*" -- Karl Pearson

In thus chapter we will discuss the event selection used in our analysis. As discussed earlier we are looking into the decay channel where b^*/B' decays into a W boson and a top-quark. In our analysis we consider two channels, namely single-lepton and dilepton.

1. **Single-lepton Channel**: In this channel we have a final state that can be written as $b^* \to Wt \to WWb \to \ell + \nu_\ell + q\bar{q}'b$ or $B' \to Wt \to WWb \to \ell + \nu_\ell + q\bar{q}'b$ with an extra forward jet as shown in Figure 2.4. We consider the charge-conjugate decay modes in this thesis. The prompt W boson or the W boson from the top-quark decay decays leptonically and the other W boson decays hadronically. When the prompt W boson decays hadronically we call that final state hadronic W and when the W boson from the top-quark decays hadronically we call that final state hadronic top. The respective regions are illustrated in Figures 4.1a and 4.1b

2. **Dilepton Channel** : In this channel we have a final state that can be written as $b^* \to Wt \to e\mu + \nu\nu + b$ or $B' \to Wt \to e\mu + \nu\nu + b$ with the similar requirement of a forward jet. In this case both the prompt W boson and the W boson decaying from the top-quark decay leptonically. This region is illustrated in Figure 4.1c



(a) Hadronic top final state (b) Hadronic W final state (c) Dilepton final state

Figure 4.1: Different final states for $b^*/B' \to Wt$

4.1 Object definitions

Electrons Events which have a single top-quark in their final state are triggered in ATLAS by requiring a high p_{T}. The inner detector tracks are matched to electron candidates reconstructed from calorimeter clusters. Electrons used in this event selection are so-called "`tight++`" electrons [42]. Selected electrons have $E_{\mathrm{cluster}}/\cosh(\eta_{\mathrm{track}}) > 25$ GeV and $|\eta_{\mathrm{cluster}}| < 2.47$, where $E_{\mathrm{cluster}}/\cosh(\eta_{\mathrm{track}})$ is the transverse energy. We also place a veto on electrons in the calorimeter barrel-endcap region with the condition $1.37 < |\eta_{\mathrm{cluster}}| < 1.52$. In order to suppress background from different processes for high p_{T} isolated electrons, we have to ensure that there is as low calorimeter activity around the electron candidates. We use cuts on two isolation variables for this purpose, the first being the energy deposited in the calorimeter around the electron with $\Delta R = 0.2$ and the second being the p_{T} sum of tracks around the electron with $\Delta R = 0.3$.

Muons Muons are reconstructed by matching hits from the muon spectrometer to inner detector tracks. We select muons that have $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$. In addition to this, the muon candidates must be classified as tight and have as author `MuidCombined` [43]. We use certain isolation criteria to reduce background contamination from events producing muons from heavy flavour quarks. We require $I_l < 0.05$, where I_l is the ratio of the sum of p_{T} of tracks in a cone of variable size to the p_{T} of the muon. Selected muon candidates should not overlap with reconstructed jets with $\Delta R < 0.4$.

Jets The anti-k_t algorithm [44] is used to reconstruct the jets used for our analysis. Locally calibrated topological clusters is used as an input for finding the jets. Jet energies are calibrated using energy and η dependent correction factors derived from simulation. Residual corrections from in-situ measurements are added to these correction factors [45]. Jets with two different radii are used. Small-R jets (j_4) have $\Delta R = 0.4$, $p_{\mathrm{T}} > 25$ GeV, and $|\eta| < 4.5$. For high masses of b^*/B' the decay particles have high p_{T}. In such cases, the top-quarks become boosted i.e. their decay products merge at the detector level. So we require large-R jets to properly identify the final decay products. Large-R jets (j_{10}) have a radius of 1.0 [46]. These jets are trimmed to minimise the impact of energy depositions from pile-up. The trimming process is designed to get rid of any initial radiation, multiple interaction or pileup contamination of these jets. Jets which contribute less than 5% of the large-R jet p_{T} are removed. The large-R jets used in this analysis have $p_{\mathrm{T}} > 200$ GeV and $|\eta| < 2.0$.

b-quark jet In final states having a top-quark as a decay product we need to distinguish between jets originating from b quarks and those from other sources. Since the weakly decaying b hadrons have a relatively long lifetime (typically 1.5 ps), they decay at a secondary vertex which can be distinguished from the primary one. We use a neural network based algorithm called the MV1 for identifying the b-quarks [47]. A top-quark pair MC sample is used to calculate the tagging weights used to tag a jet as b-jet. The working point used has a 70% efficiency of selecting b-jets in the MC sample.

Missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$, missing transverse energy, is a measure of the magnitude of the missing momentum of the escaping neutrinos. It also gives us an estimate of energy losses due to detector resolution and inefficiencies. It is calculated as the vector sum over all the topological clusters (3-D objects built from calorimeter cells) in an event. Subsequently, object level corrections are applied to refine its definition.

4.2 Single-lepton channel

As shown in Figures 4.1a and 4.1b, the single-lepton channel final state can be written as $Wt \rightarrow \ell + E_{\mathrm{T}}^{\mathrm{miss}}$ + jets. For B' we require the presence of one extra forward jet. So the selected signal events should have the following criteria

- The events should have exactly one isolated lepton (electron or muon).
- They should have two or three small-R central jets with $|\eta| < 2.5$. Exactly one of the jets should be b-tagged. For B' event candidates in the signal region, at least one additional forward small-R jet with $2.5 < |\eta| < 4.5$ should be present.
- To reduce the fraction of mis-identified lepton background we require $E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV and $m_T(\ell.E_{\mathrm{T}}^{\mathrm{miss}}) + E_{\mathrm{T}}^{\mathrm{miss}} > 60$ GeV, where $m_T(\ell.E_{\mathrm{T}}^{\mathrm{miss}})$, the transverse mass of the lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined as
$$m_T(\ell, E_{\mathrm{T}}^{\mathrm{miss}}) = \sqrt{2p_{\mathrm{T}}(\ell)E_{\mathrm{T}}^{\mathrm{miss}}(1 - \cos\Delta\phi(p_{\mathrm{T}}(\ell), E_{\mathrm{T}}^{\mathrm{miss}}))} \tag{4.1}$$
where $p_{\mathrm{T}}(\ell)$ is the value of the lepton transverse momentum and $\Delta\phi(p_{\mathrm{T}}(\ell), E_{\mathrm{T}}^{\mathrm{miss}})$ is the difference of the azimuthal angle of lepton and missing transverse momenta.
- The selected events must have at least one large-R jet with $p_{\mathrm{T}} > 200$ GeV and $m_{j10} > 50$ GeV. If more than one large-R jets is present, we consider the one with highest p_{T}

We use cuts on the difference of azimuthal angle ($\Delta\phi$) between the lepton and the large-R jet and the minimal distance between the small-R and large-R jets to define our signal region. The large-R jet cut of $m_{j10} > 50$ GeV and a topology cut $\Delta\phi(\ell, j_{10}) > 1.5$ are applied to all the categories. As shown in Figures 4.1a and 4.1b, we define two signal regions, **hadronic top** and **hadronic W**. The exact definitions of the regions are given in Table 4.1.

Category reference	number of b-tags	forward jet	topology cuts
b^*, hadronic top	1	no cut	$\min\Delta R(\ell, j_4) > 1.5$ $\max\Delta R(j_4, j_{10}) < 2.0$
b^*, hadronic W	1	no cut	$\min\Delta R(\ell, j_4) < 1.5$ $\max\Delta R(j_4, j_{10}) > 2.0$
B', hadronic top	1	≥ 1	$\min\Delta R(\ell, j_4) > 1.5$ $\max\Delta R(j_4, j_{10}) < 2.0$
B', hadronic W	1	≥ 1	$\min\Delta R(\ell, j_4) < 1.5$ $\max\Delta R(j_4, j_{10}) > 2.0$

Table 4.1: Event categorisation of the b^* or B' decays in the single-lepton channel.

The invariant mass of the final state, i.e. the vector sum of the four momenta of the lepton, the small-R jets, and the missing transverse momentum ($\sum(\mathbf{p_l} + \mathbf{p_{small-R\ jet}} + \mathbf{p_T^{miss}})$ is used as our discriminant for separating signal and background events in the statistical analysis.

In the single-lepton channel, the two major background contributions come from the top-pair and W+ jets production. Other backgrounds with smaller contributions are single top-quark production, Z+jets and other diboson processes. We use four control regions, two regions corresponding to top-pair enriched regions and the other two corresponding to W+jets enriched regions. Both the top-pair

and W+jets control region are subdivided into the hadronic top and the hadronic W sub-categories. The definition of both the regions is similar to that of the signal regions except we have exactly zero b-tagged jets for the W+jets region and at least two b-tagged jets for the top-pair region.

4.3 Dilepton channel

As shown in Figure 4.1c, for the dilepton channel we can write the final state as $Wt \rightarrow e\mu + b + E_{\mathrm{T}}^{\mathrm{miss}}$. The requirement of an extra forward jet for B' is also needed in the dilepton case. The criteria used for selecting signal events are

- The selected events should have exactly one isolated electron and exactly one isolated muon with opposite charges.
- For the jets, we require $p_{\mathrm{T}} > 25$ GeV. Exactly one of the central jets should be b-tagged, as they are beyond the acceptance of the tracking detectors. Jets having $2.5 < |\eta| < 4.5$ are also taken into account. These jets cannot be b-tagged. For B' the existence of a small-R jet with $1.5 < |\eta| < 4.5$ is required.
- The minimum azimuthal angle difference between the leptons and the leading small-R jet j_0 should satisfy the requirement $\min\Delta\phi(\ell, j_0) < 0.9$.

We use two signal regions for our analysis, **1-jet-1tagged** and **2-jet-1tagged**. The transverse mass of the final state ($\sqrt{(E_{\mathrm{T}}^{\mathrm{miss}})^2 - (p_{\mathrm{T}}^{\mathrm{miss}})^2}$) is used as the discriminant to separate signal and background events. The strict cut on the number of b-jets suppresses the top-pair background. The second untagged jet for the second signal region is allowed to have looser p_{T} cut of $p_{\mathrm{T}} > 30$ GeV. The main background contribution comes from the top-pair. Other backgrounds are single top-quark production, Z+jets and other diboson processes. Since our final state is an $e\mu$ one, the Z boson can decay to a tau pair and $e\mu$ only for the background processes. We define a control region **2-jet-2tagged** for a better modelling of the top-pair.

CHAPTER 5

Statistical methods

"*I believe that we do not know anything for certain, but everything probably.*" -- Christiaan Huygens

In all experimental sciences, statistical methods play a crucial role in reaching the correct scientific inferences. In high energy physics, we use these methods to test the validity of different theoretical models with respect to the collision data obtained in our colliders. In this chapter, a detailed description of the statistical methods used in our analysis will be presented. We use both Bayesian and frequentist methods for setting limits on the parameter of interest and frequentist methods for performing hypothesis tests.

5.1 Basic concepts

In this section, some ideas about statistical inference are presented. There are two main schools of thought regarding the definition of probability namely frequentist and Bayesian. To quote [19] "In frequentist statistics, probability is interpreted as the frequency of the outcome of a repeatable experiment". As defined in [48] the interpretation of probability in Bayesian statistics can be stated as "a measure of the degree of belief that an event will occur".

5.1.1 Statistical inference

As mentioned earlier, statistical inference is the process of comparing our data to different theoretical models. Let us consider an experiment where we obtain a set of measured physical quantities $\vec{D}$. We would like to compare these values to the predictions of a model M having parameters $\vec{\lambda}$. Let $\vec{x}$ be the measured values of the physical quantities, λ_i, predicted by the given model M. Now we compare the predicted and observed values using different statistical methods. In most cases, we either use a frequentist or a Bayesian approach. In experimental sciences, we reach a concrete result after performing the same experiment numerous times. In this context frequentist methods are used. But while performing a given experiment, we have a set of prior information about our experiment. Incorporating these information is one of the main motivations behind using Bayesian approach.

Frequentist approach In the frequentist approach we define the probability as $P(\vec{x} = \vec{D}|\vec{\lambda}, M)$ i.e the probability that $\vec{x} = \vec{D}$ given the model M represented by the parameter vector $\vec{\lambda}$. Using this definition of probability we define a probability distribution function (p.d.f.) $f(x_i|\vec{\lambda})$ for each of the measured quantities x_i. We then define the likelihood $L(\vec{\lambda})$ as

$$L(\vec{\lambda}) = \prod_i f(x_i|\ \vec{\lambda}) \tag{5.1}$$

We use this likelihood to estimate the best approximate values of the different parameters in the parameter vector $\vec{\lambda}$. The uncertainties on these values are also computed to give the measurement a proper meaning in the frequentist framework. If the measured value of a parameter λ_i is x_i and if we assume that the error distribution of that parameter is Gaussian with a given σ then we say that the observed value of the given parameter is $x \pm \sigma$. In the frequentist framework this statement can be rephrased as "The statement that $(x - \sigma) < \lambda_i < (x + \sigma)$ has a 68% probability of being true" [49].

Bayesian approach In Bayesian statistics we have a probability distribution $P_0(\vec{\lambda}|M)$, called the *prior*, describing our prior belief in the knowledge of $\vec{\lambda}$ according to our model M. After our experiment, Bayes' theorem gives us the posterior probability distribution $P(\vec{\lambda}, M|\vec{D})$. In our notation we can formulate this as following,

$$P(\vec{\lambda}, M|\vec{D}) = \frac{P(\vec{x} = \vec{D}|\vec{\lambda}, M)P_0(\vec{\lambda}|M)}{P(\vec{D})} = \frac{P(\vec{x} = \vec{D}|\vec{\lambda}, M)P_0(\vec{\lambda}|M)}{\int P(\vec{x} = \vec{D}|\vec{\lambda}, M)P_0(\vec{\lambda}|M)d\vec{\lambda}} \tag{5.2}$$

This posterior distribution is used to compute different properties of the parameter. We can compute the mean, standard deviation of the posterior distribution to get the observed value and the respective uncertainties. In Bayesian statistics, we should ensure that our measurement is robust i.e. show that any arbitrary prior doesn't change our measurement to a large extent. We will discuss about the different choice of priors in detail in Section 5.2.3.

5.2 Parameter estimation

Parameter estimation is the method of estimating the "best possible value" of a given parameter and the uncertainties on those estimated values. We use various methods, both frequentist and Bayesian to compute these values. There are two types of frequentist methods which can be used, maximum likelihood method and least squares method. Here we will elaborate only the first method. The Bayesian method used is called marginalisation.

5.2.1 Method of maximum likelihood

In the frequentist method we use a consistent, unbiased, and robust estimator $\hat{\lambda}_i$ [19] to estimate the observed value of λ_i. We use the definition of likelihood defined in (5.1). The product is over the N measured quantities. We maximise the likelihood with respect to the each of the n values of the parameter vector $\vec{\lambda}$, λ_i, to get the maximum likelihood estimators. the solution of the following equations gives us the value of the estimators.

$$\frac{\partial \ln L}{\partial \lambda_i} = 0 \quad i = 1, 2.....n \tag{5.3}$$

The log-likelihood is used because of the useful properties of the logarithm function and the possibility of the likelihood taking very large values. In most algorithms used in high energy physics the negative log likelihood is minimised. In most cases the probability distribution $f(x_i)$ is Gaussian. In that case, it can be shown that for repeated independent experiments the estimated mean $\hat{\lambda}_i$ is distributed according to a Gaussian distribution with a mean of $\hat{\lambda}_i$ [50].

In high energy experiments, we have large samples of data so it is convenient to use binned data. If our data is binned over B bins and the content of each bin is N_i the likelihood is multinomial and is expressed by [50]

$$L(\vec{\lambda}) = N_{\text{tot}}! \prod_{i=1}^{B} \frac{P_i(\vec{\lambda})^{N_i}}{N_i!} \tag{5.4}$$

where $N_{\text{tot}} = \sum N_i$ is the sample size of the data and $P_i(\vec{\lambda})$ is the expected probability that a given event will appear in the ith bin. It can be expressed in terms of the p.d.f as

$$P_i(\vec{\lambda}) = \int_{x_i^{\text{low}}}^{x_i^{\text{up}}} f(x|\,\vec{\lambda})dx \tag{5.5}$$

x_i^{up} and x_i^{low} being the upper and lower bin edges of the ith bin respectively. The corresponding negative log likelihood can be written as

$$-\ln L = -\sum_{i=i}^{B} N_i \ln P_i(\vec{\lambda}) + \text{const.} \tag{5.6}$$

In our analysis, we will use a binned log likelihood fit to find the estimators $\hat{\lambda}_i$ of the parameters of interest. The details of that procedure will be explained in later sections.

5.2.2 Marginalisation

In the Bayesian approach, we use the posterior p.d.f., as defined in (5.2), to find the different properties of our parameters of interest. The posterior p.d.f. gives the distribution for the complete parameter vector. But in most cases, we are interested in the posterior p.d.f.s of individual parameters. We use a procedure called marginalisation to compute these individual posterior p.d.f.s. The marginalised posterior p.d.f. of a given parameter λ_i is computed by integrating the full posterior p.d.f. over the parameter space of all the parameters $\lambda_{j\neq i}$. The marginalised posterior p.d.f. is written as [51]

$$P(\lambda_i, M|\vec{D}) = \int P(\vec{\lambda}, M|\vec{D})d\vec{\lambda}_{j\neq i} \tag{5.7}$$

With the help of the following formulae, the marginalised posterior p.d.f.s gives us the desired properties like mean, median, standard deviation of different parameters in the parameter vector.

Mean

$$\bar{\lambda}_i = \int P(\lambda_i, M|\vec{D})\lambda_i d\lambda_i$$

Median

The median $\lambda_i^{\text{median}}$ is defined by

$$0.5 = \int_{\lambda_i^{\text{min}}}^{\lambda_i^{\text{median}}} P(\lambda_i, M|\vec{D})d\lambda_i$$

Standard deviation

$$\sigma_i = \sqrt{\bar{\lambda}_i^2 - (\bar{\lambda}_i)^2}$$

5.2.3 Priors

In a Bayesian method, the *prior* $P_0(\vec{\lambda}|M)$, as defined in (5.2), describes our prior "degree of belief" about the possible values $\vec{\lambda}$ can take. The choice of our prior shouldn't affect the measurement, it should be an uninformative one. The flat prior is the simplest case of an uninformative prior. The Jeffreys Prior [52] is a commonly used prior in Bayesian analyses. It is defined as

$$\pi(\vec{\lambda}) = \sqrt{\det I(\lambda)_{ij}} \tag{5.8}$$

where $I(\lambda)_{ij}$ is the Fisher Information Matrix defined as

$$I(\lambda)_{ij} = -E\left(\frac{\partial^2 \ln L}{\partial\, \lambda_i \lambda_j}\right) \tag{5.9}$$

Jeffreys prior is invariant under a re-parametrisation of the parameters $\lambda_i \rightarrow \lambda_i'$, thus making it an objective prior. If the likelihood L assumes a Gaussian distribution then it can be shown that the Jeffreys prior is uniform.

5.2.4 Template fitting of Monte Carlo samples

In high energy physics experiments, we obtain distributions of different physical observables from our collision data. These distributions have contributions from different processes. In real data, we don't have information about the individual contributions from different processes. We use a method called *Template fitting* [53] to find these individual contributions. We use Monte Carlo simulations of the distributions for each process and fit it to our observed data. We use binned histograms, which are called templates, to do the fit. This method compares the shapes of the individual templates to the shapes of the data templates. The individual normalisation factors can be calculated using the methods mentioned in the previous sub-sections.

Suppose we have P processes and each distribution is binned over B bins. Let E_k^j and N_k^j be the respective expected and observed events for the jth process in the kth bin. The events in a given bin k are assumed to be independent and follow a Poisson distribution with a mean E_k . In that case can write the likelihood L as

$$L = \prod_{k=1}^{B} \frac{E_k^{N_k}}{N_k!} \cdot e^{-E_k} \tag{5.10}$$

The same expression can be obtained if we assume the total number of observed events, N, to follow a Poisson distribution with a mean E [50]. The total number of expected number of events E can be written as the sum of the individual contributions from each process. The expected number of events in bin k can be thus written as [19]

$$E_k = \sum_{j=1}^{P} p_j\, A_{jk} \tag{5.11}$$

deleting where p_j is the scale factor a given process j and A_{jk} are the expected number of events for process j in bin k. The data events a_{jk} for a given process j in bin k are produced via a Poisson distribution.

Frequentist method For the frequentist approach we use the method as described in Section 5.2.1. Using (5.10) and (5.11), we can write the log likelihood as

$$\ln L = \sum_{k=1}^{B} N_k \ln E_k - N_k + \sum_{j=1}^{P} \sum_{k=1}^{B} a_{jk} \ln A_{jk} - A_{jk} \tag{5.12}$$

We maximise the log likelihood to get the estimates of p_j by solving the following $P \times (B + 1)$ equations [54]

$$\sum_{k=1}^{B} \frac{N_k A_{jk}}{E_k} - A_{jk} = 0 \; \forall j \tag{5.13}$$

$$\frac{N_k p_j}{E_k} - p_j + \frac{a_{jk}}{A_{jk}} - 1 = 0 \; \forall j, k \tag{5.14}$$

Bayesian method For the Bayesian approach, we use the method as described in Section 5.2.2. We use the definitions of likelihood and marginalised likelihood as mentioned in (5.10), (5.11), and (5.7) to get the marginalised posterior distribution for a given process.

$$L(p_j|N) = \int L(\vec{p}|N) d\vec{p}_{i \neq j} \tag{5.15}$$

Consequently, we can estimate p_j by finding the mean or median of these posterior distributions. We can subsequently compute different quantiles to compute the errors on p_j.

5.3 Nuisance Parameters

In any particle physics experiment, dealing with uncertainties is one of the most complex and difficult steps. As any physics experiment, the two main sources of uncertainties are systematic and statistical uncertainties. In collider experiments statistical and systematic uncertainties are not necessarily independent, thus making their estimation a complex process. The quantities used to estimate the uncertainties are called *nuisance parameters*.

Systematic Uncertainties In particle physics experiments, we define systematic uncertainties in the following manner [55].

"Systematic uncertainties are measurement errors which are not due to statistical fluctuations in real or simulated data samples."

The common sources of systematic uncertainties in collider experiments are

- Poor knowledge of trigger efficiencies, detector resolutions, and detector acceptances.
- Poor reconstruction of the detector environment.
- Uncertainties on input parameters like luminosity.
- Uncertainties coming from the generation of Monte Carlo simulations.
- Theoretical uncertainties.

Statistical Uncertainties Statistical uncertainties in any analyses come from fluctuations in data and Monte Carlo simulations. The statistical uncertainties from the data are generally larger. We consider the uncertainties on the scale factor, p_j for each process j, and the statistical fluctuations in different bins as our main sources of statistical uncertainties.

5.3.1 Incorporating uncertainties

The likelihood contains all the information about our statistical model. To estimate the effects of different uncertainties, we have to include them into the likelihood in a consistent manner. The easiest way of doing that is multiplicating one factor for each uncertainty as a nuisance parameter into the Likelihood. We assume a Gaussian contribution for each parameter. then the Poisson Likelihood in (5.10) can be re-written as [56]

$$L = \prod_{k=1}^{B} \frac{E_k^{N_k}}{N_k!} \cdot e^{-E_k} \prod_i G(\theta_i) \tag{5.16}$$

These Gaussian functions can be interpreted as priors in the Bayesian approach and constraints in the frequentist approach. We can represent both systematic and statistical uncertainties in this manner. In some cases, we take a log-Gaussian prior to make sure that the yields do not become negative. The Gaussians are generally taken as $G(0, 1)$, where the mean zero represents no variation. The effects of these uncertainties can be computed using both the frequentist and Bayesian methods.

Frequentist method In the frequentist approach, the negative log likelihood gets an additional contribution from each of the nuisance parameters. Using (5.16) we get

$$-\ln L = -\sum_{k=1}^{B} [N_k \ln E_k - N_k] + \sum_i s_i^2 \tag{5.17}$$

where s_i^2 is the contribution from θ_i. This can be minimised for individual θ_i to give us an estimator of the parameters of interest for the respective uncertainties. In that way, we can understand the effect of each uncertainty. We can repeat the same process for any set of given uncertainties. There is also a method called profiling where we maximise the following profiled likelihood [19].

$$L(\vec{\lambda}) = L(\vec{\lambda}, \hat{\hat{\vec{\theta}}}(\vec{\lambda})) \tag{5.18}$$

where $\hat{\hat{\vec{\theta}}}$ are the values of the nuisance parameters which maximise the likelihood for given $\vec{\lambda}$.

Bayesian method In the Bayesian approach, we use the method of marginalisation to integrate out the nuisance parameters. After integrating out all the nuisance parameters we get our posterior p.d.f. for the parameter vector. We can also integrate out $\theta_{j\neq i}$ to find the effect of θ_i on the posterior p.d.f. Using (5.7), the marginalised posterior likelihood for any parameter λ_i can be re-written as

$$L(\lambda_i, \vec{\theta}, M|\vec{D}) = \int \int L(\vec{\lambda}, \vec{\theta}, M|\vec{D}) d\vec{\lambda}_{j\neq i} d\vec{\theta} \tag{5.19}$$

We can also write the marginalised posterior for a given nuisance parameter θ_i as

$$L(\theta_i, \vec{\lambda}, M|\vec{D}) = \int\int L(\vec{\lambda}, \vec{\theta}, M|\vec{D}) d\vec{\theta}_{j\neq i} d\vec{\lambda} \tag{5.20}$$

As defined in (5.16), the priors on the nuisance parameters are Gaussian so the posterior likelihoods will also be Gaussian. These post-fit posteriors of the different nuisance parameters give us an estimate of their effect on our parameter vector.

5.3.2 Morphing

Some uncertainties affect both the shape and rate of our distributions. We generally have distributions for the nominal and $\pm 1\sigma$ variations for a given uncertainty. If any uncertainty affects the shape of the nominal distributions, we would like to get an estimate of its effect for any arbitrary variation. In such cases, we introduce a parameter f as a continuous measure of the variation. We call this parameter the *morphing parameter*. It is defined such that $f = 0, \pm 1$ are the nominal and $\pm\,\sigma$ variations respectively.

Suppose the normalisation of the efficiencies for the jth process in the ith bin is represented as ϵ_{ji}. If ϵ^0_{ji}, ϵ^+_{ji}, and ϵ^-_{ji} are the respective efficiencies for the nominal and $\pm\ 1\sigma$ variations, then in the simplest case, we write the efficiency ϵ_{ji} [56]

$$\epsilon_{ji} = \epsilon^0_{ji} + f\frac{\epsilon^+_{ji} - \epsilon^-_{ji}}{2} \tag{5.21}$$

More sophisticated interpolation and extrapolation algorithms are generally used for modelling the spectra for any arbitrary variation. Some of these algorithms are explained in the next chapter.

5.4 Hypothesis tests

One of the main goals of high energy physics experiments is to look for new processes and particles. In a collider, all types of processes, hopefully including the ones we are searching for, occur. The method of testing different hypotheses against each other is called *hypothesis testing*.

For search experiments, we test the background-only hypothesis. If there is presence of a significant signal, we can reject the background-only or the null hypothesis with a higher degree of confidence. We can also check the null hypothesis against the signal-plus-background hypothesis to search for the presence of a signal. In reality, our data either represents the null hypothesis or the signal plus background hypothesis. When we make a statement about the validity of the null hypothesis (H_0) we can make two types of errors namely.

1. Type I error : The null hypothesis has been rejected although it is true. This error can happen with a probability α.

2. Type II error : The null hypothesis has not been rejected although it is false. This error can happen with a probability β.

The probabilities of committing or not committing the aforementioned errors can be shown in Table 5.1 [57].

We generally use a test statistic Q that helps us to quantify the difference between different hypotheses and the null hypothesis. A test statistic is chosen in such a manner that our test is unbiased. We should make sure that $1-\beta \not\leq \alpha$. If $1-\beta \leq \alpha$, the probability to choose the null hypothesis when it is false, $(1-\beta)$, is greater than the probability to choose the null hypothesis when it is true, (α), our test is

	H_0 is true	H_0 is false
H_0 rejected	Type I error (Probability α)	Right decision (Probability $(1-\beta)$)
H_0 not rejected	Right decision (Probability $(1-\alpha)$)	Type II error (Probability β)

Table 5.1: Probabilities of occurrence of Type I and Type II errors

biased. From the Neyman-Pearson lemma we know that the likelihood ratio is an optimal test statistic. We define Q as

$$Q(x) = \frac{L(x|H_i)}{L(x|H_0)} \; \forall H_i \tag{5.22}$$

5.4.1 p-Value and Significance

A deviation from the null hypothesis doesn't always suggest a presence of a signal. While quoting discovery or evidence, we have to make sure that Type I error is not happening. We define a quantity called $p-value$ which gives us the measure how much different hypotheses differ from the null hypothesis. In most cases, we measure the deviation of the signal-plus-background hypothesis from the null hypothesis. We have distributions for the Q-values of both null and signal plus background hypothesis. The following figure shows such a comparison.

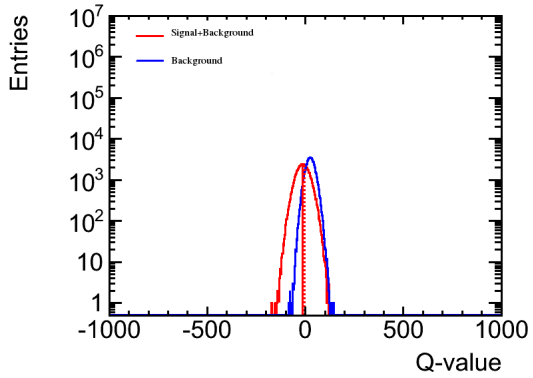


Figure 5.1: Q-value plot for hypothesis test

In the Figure 5.1 the graph on the right is the Q-value distribution of the null hypothesis and the graph on the left is the Q-value distribution of the signal plus background hypothesis. The vertical line indicates the median of the Q-value distribution of the signal-plus-background hypothesis. We define p-value as the area of the Q-value distribution of the null hypothesis enclosed by the Q-value distribution of the signal-plus-background hypothesis. So it gives us a measure of overlap or difference between the two hypotheses. It can be shown that p-value can also be defined as the probability of making a Type I error

The significance Z is the p-value described in terms of σ of a unit Gaussian. So we define significance as following

$$\int_Z^\infty e^{-\frac{t^2}{2}} dt = p \tag{5.23}$$

In particle physics, we consider 3σ as evidence and 5σ as the discovery of the signal.

5.4.2 Hypertest and trials factor

In particle physics, we generally have large samples of data binned across numerous bins. Suppose we have 1000 bins and we get a discrepancy in one bin for a given hypothesis test. On the basis of such a test we cannot reject the null hypothesis. If we had devised our test in some other region we could have got different results. The choice of such tests is completely *ad hoc*. So we cannot use one such test to come to any conclusion. We should "look elsewhere", i.e. we should use numerous hypothesis tests across the whole spectrum to reach a concrete result. So we perform a set of hypothesis tests called *hypertest*.

Suppose we devise N independent hypothesis tests and define $p = \alpha$ as the p-value for which we reject H_0. The probability that for one of these tests $p_i \leq \alpha$ is [58]

$$P(\text{one test has } p \leq \alpha) = 1 - (1 - \alpha)^N \tag{5.24}$$

N is called the "trials factor". If we have $\tilde{N}$ independent tests among the N hypothesis tests then we can say that we can reject H_0 with a Type I error of $1 - (1 - \alpha)^{\tilde{N}}$ where we call $\tilde{N}$ as the "effective trials factor". To account for all the tests we define a global p-value using a new test-statistic Q_{new} [58]

$$Q_{new} = -\log(\min_i p_i) \tag{5.25}$$

We use this global p-value to quote evidence or discovery. Hypertests are constructed using as many possible sets of different hypothesis tests. In this way we get rid of the "look-elsewhere effect".

5.5 Limit Setting

Experiments designed to search for new particles and processes do not always have the desired sensitivity. In cases where the number of background events are comparable to the number of observed events we use a method called *limit setting* to set physical boundaries on the possible values for the parameter of interest. In most cases, the signal strength is used as the parameter of interest. The signal strength is hypothesised using the theoretical models which are to be tested in a given analysis. In case of non-discovery, we generally quote a 95% upper limit on the signal strength. There are both frequentist and Bayesian methods of interval estimation which can be used setting limits, but in particle physics a method called CL_s method is often used to do limit setting.

Frequentist method In frequentist methods, confidence intervals are used for setting limits. The simplest method of setting upper confidence levels is the Neyman construction [59]. Using this method, we can set a $(1 - \alpha)$ confidence level upper limit, λ_{up} if

$$P(x < \lambda_{\text{up}}|\hat{\lambda}) = 1 - \alpha \tag{5.26}$$

where $\hat{\lambda}$ is the true value of λ. For non-discovery, we can say that λ can be excluded for values $\lambda < \lambda_{\text{up}}$ at $(1 - \alpha)$ C.L..

Bayesian method In Bayesian approaches a method called credible intervals are used to set limits. We can set a $(1 - \alpha)$ confidence level upper limit, λ_{up} if

$$\int_{-\infty}^{\lambda_{\text{up}}} P(\lambda|D)d\lambda = 1 - \alpha \tag{5.27}$$

where $P(\lambda|D)$ is the posterior p.d.f.

5.5.1 CL_s Method

The CL_s method [60] is a way to set exclusion limits on the signal strength. In particle physics, we compare the expected number of background events against our data to set limits on our signal strength. In cases where the background expectation is more than the observed data, a negative value of the fitted signal strength fits better from a purely statistical point of view. But this is clearly an unphysical result. If we use classical frequentist methods in such cases, even zero signal strength is excluded. To get rid of such situations the CL_s method was introduced. The CL_s confidence levels are neither similar to frequentist confidence intervals nor identical to Bayesian credible intervals.

We define a test statistic Q such that it increases with increasing sensitivity. We use the following definitions to construct the *CLs* exclusion limit.

$$CL_{s+b} = P_{s+b}(Q \leq Q_{obs}) = \int_{-\infty}^{Q_{obs}} f_{s+b}\, dQ \tag{5.28}$$

$$CL_b = P_b(Q \leq Q_{obs}) = \int_{-\infty}^{Q_{obs}} f_b\, dQ \tag{5.29}$$

where f_{s+b} and f_b are the respective p.d.f.s for the signal plus background and background only hypothesis.

We define CL_s as [60]

$$CL_s = \frac{CL_{s+b}}{CL_b} \tag{5.30}$$

So CL_s is not a confidence level in the strict statistical sense, but it is a ratio of confidence levels. We exclude a signal at $(1 - \alpha)$ confidence if

$$CL_s \geq \alpha \tag{5.31}$$

$$\text{or, } \frac{p_s}{1 - p_b} < \alpha \tag{5.32}$$

where p_s and p_b are the respective p-values of signal and background hypothesis. For the CL_s method, we use likelihood ratio as our test statistic.

Merits of CL_s method

- This method gives us higher coverage than classical methods for excluding a given signal hypothesis i.e. the probability of the signal strength being in the excluded interval decreases when we use the CL_s method.

- The test statistic increases monotonically with increasing sensitivity.
- This method helps us to avoid stronger exclusion limits if the background expectation is less than data.
- For low values of the signal strength, CL_s has a better sensitivity than CL_{s+b} has.

5.6 Methods used

In this thesis we conduct a search for the b^*/B' signal. We use the following statistical methods in this thesis to do this

- In the first step we test our background-only (Standard Model) hypothesis versus signal-plus-background (b^*/B'-plus-Standard Model) hypothesis. We do this to reject or accept the background-only (Standard Model) hypothesis. We use a type of hypertest to perform this step.
- In the second step we set limits on the signal strength for various signal hypotheses. We set our limits by using both Bayesian and CL_s methods. These methods use Bayesian and frequentist methods of parameter estimation respectively. In this thesis we try to compare the results obtained by these two methods.

CHAPTER 6

Statistical tools

"*Statistics are somewhat like old medical journals, or like revolvers in newly opened mining districts. Most men rarely use them, and find it troublesome to preserve them so as to have them easy of access; but when they do want them, they want them badly.*" -- John Shaw Billings

In the previous chapter we have discussed the different statistical methods used in particle physics. We use different tools in our analyses which implement these methods. In this chapter, the details and working of the statistical tools used in this analysis are presented. We use four tools, `BumpHunter`. `Histfactory`, Bayesian Analysis Toolkit (`BAT`), and Binned Log Likelihood (`BILL`) fitter,

6.1 BumpHunter

`BumpHunter` [58] is a statistical tool that checks the validity of the null (background-only) hypothesis against the signal-plus-background hypothesis. It looks for bumps or excess of data over the background. It employs a hypertest, as described in Section 5.4.2, to get rid of the "look-elsewhere" effect. The hypothesis tests which constitute the hypertest calculates the difference between the data and the background for different regions of the parameter space. The minimum width of the window considered for any hypothesis test is two bins. This ensures that we don't misinterpret a single bin fluctuation as an excess. For each of these hypothesis tests, `BumpHunter` calculates a local p-value. It makes sure that the whole parameter space is scanned by the hypertest. The global p-value of the hypertest is calculated using (5.25). The test statistic t of each of the hypothesis tests is defined as

$$t = \begin{cases} 0 \text{ if } d_i \leq b_i \\ f(d_i - b_i) \text{ otherwise} \end{cases} \tag{6.1}$$

where $f(d_i - b_i)$ is a positive, monotonically increasing function. This ensures that with an increasing excess of data the test-statistic t increases. We use the Monte Carlo simulations of the background to describe our null hypothesis, but `BumpHunter` also allows fitting a functional form of the background to the observed data. For most practical purposes it is difficult to predict the correct function. `BumpHunter` generates thousands of pseudoexperiments where the null-hypothesis is a distribution chosen at random between the error bands of the original background histogram. These pseudoexperiments give us a distribution of p-values which we use to calculate the final p-value. The final p-value used for checking the validity of the null hypothesis is given by

$$\text{p-value} = \frac{\#\text{Pseudoexperiments with } p \geq p_0}{\#\text{Pseudoexperiments generated}} \tag{6.2}$$

where p_0 is the global p-value for the hypertest comparing the original MC distribution and data.

Incorporating systematic uncertainties

Let us assume a factor λ that gives us the measure of all systematic uncertainties. The variations change the background hypothesis. Our goal is to minimise this effect while looking for an excess of data. We re-parametrise our local p-value in the following manner to incorporate the effect of all the systematic uncertainties.

$$p_i^{\text{syst}} = \exp(-\lambda^2/2) \cdot p_i \tag{6.3}$$

For each hypothesis test, we choose that value of λ which minimises the effect of the uncertainties. We achieve this by scanning over a large range of λ values. Thus the local p-value is defined as

$$p_i^{\text{syst}} = \max(\exp(-\lambda^2/2) \cdot p_i) \tag{6.4}$$

After this, we follow the method as described above to get the final p-value.

6.2 HistFactory

`Histfactory` [61] is a statistical tool to build different likelihood functions in the `RooStats` framework. This tool can incorporate the effects of various systematic and statistical uncertainties. Different morphing algorithms are also a part of the framework to model the uncertainties in a precise manner. `Histfactory` uses binned ROOT histograms to build the likelihood functions. The likelihood function is generally defined as

$$\mathcal{L}(N_{csb}, a_p|\phi_p, \alpha_p, \gamma_b) = \prod_{c\in channels} \prod_{s\in samples} \prod_{b\in bins} \text{Pois}(N_{csb}|E_{csb}) \cdot G(L_0|\lambda, \Delta_L) \cdot \prod_{\in \mathbb{S}+\Gamma} f_p(a_p|\alpha_p) \tag{6.5}$$

$$E_{csb} = \lambda_{cs}\, \gamma_{cb}\, \phi_{cs}\, \eta_{cs}\, \sigma_{csb} \tag{6.6}$$

- N_{csb} and E_{csb} are the observed and expected events for channel c in sample s in bin b.
- $G(L_0|\lambda, \Delta_L)$ is the Gaussian constraint term on the luminosity.
- $f_p(a_p|\alpha_p)$ are the constraints or priors on the different nuisance parameters α_p. a_p is an auxiliary measurement constraining α_p The most commonly used constraint is the Gaussian constraint. For normalisation uncertainties, we use a log-Gaussian constraint to make sure that the yields are non-negative.
- λ_{cs} is the luminosity parameter for channel c in sample s. This parameter is constant in most cases but this can also be varied according to the experiment.
- γ_{cb} is the bin-by-bin scale factor. The bin-by-bin scale factor is a function of the nuisance parameter vector $\boldsymbol{\alpha}$.
- $\mathbb{S} = \{\gamma_{csb}\}$ and $\Gamma = \{\alpha_p\}$ are sets containing all the nuisance parameters and bin-by-bin uncertainties respectively.
- ϕ_{cs} is the product of all unconstrained normalisation uncertainties. This can include the different properties of the parameter of interest.
- η_{cs} are the different rate uncertainties used. In most cases, these include the normalisation uncertainties of the given channels.

- σ_{csb} are the binned histograms which contain the shape and rate information for the different uncertainties in $\boldsymbol{\alpha}$.

Monte Carlo uncertainties

The histograms used as inputs for calculating the likelihood are Monte Carlo simulations of different processes. A statistical uncertainty arises when estimated events in the histograms are quite low. If we use the Barlow-Beeston method [54] we have to consider a rate uncertainty for each bin of the given samples. For most analyses, this would add hundreds of extra nuisance parameters, thus making the calculation of the likelihood computationally expensive. Instead of this, `Histfactory` uses an uncertainty for each bin b. We can rewrite the Poisson distribution to include the Monte Carlo (MC) uncertainty.

$$\text{Pois}(N_{csb}|E_{csb}(\boldsymbol{\alpha}) + \gamma_b E_b^{MC}(\boldsymbol{\alpha}))\ \text{Pois}(M_b|\gamma_b\tau_b) \tag{6.7}$$

where $E_{csb}(\boldsymbol{\alpha})$ is the expected number of events in bin b without considering the MC uncertainty and $E_b^{MC}(\boldsymbol{\alpha})$ are the events by considering the MC uncertainty. γ_b is the nuisance parameter reflecting the rate difference between the expected estimates and the MC estimates. $\text{Pois}(M_b|\gamma_b\tau_b)$ is a Poisson distribution of M_b around $\gamma_b\tau_b$, τ_b being a reparametrised variable containing the information about the original MC sample.

Morphing

As mentioned in Section 5.3.2, we need to incorporate different morphing algorithms to precisely model our uncertainties. `Histfactory` employs certain extrapolation and interpolation algorithms to do so. The algorithms used are combinations of linear, quadratic and exponential algorithms. Here we will only describe the algorithms used in our analysis. We use an algorithm which uses polynomial interpolation and exponential extrapolation. We can write the rate uncertainties $\eta_s(\boldsymbol{\alpha})$ and shape uncertainties σ_{sb} as

$$\eta_s(\boldsymbol{\alpha}) = \prod_{NP} I_{\text{poly|exp}}(\alpha_p; 1, \eta_{sp}^+, \eta_{sp}^-) \tag{6.8}$$

$$\sigma_{sb}(\boldsymbol{\alpha}) = \sigma_{sb}^0 \prod_{NP} I_{\text{poly|exp}}(\alpha_p; \sigma_{sbp}^0, \sigma_{sp}^+, \sigma_{sbp}^-) \tag{6.9}$$

where the product is over all the nuisance parameters (NP). The interpolation or extrapolation is done around 1 for the rate uncertainties and σ_{sb}^0 for the shape uncertainties for a given sample s in bin b. The function $I_{\text{poly|exp}}$ can be written as

$$I_{\text{poly|exp}}(\alpha; I^0, I^+, I^-, \alpha_0) = \begin{cases} (I^+/I_0)^\alpha & \alpha \geq \alpha_0 \\ 1 + \sum_{i=1}^6 a_i \alpha & |\alpha| < \alpha_0 \\ (I^+/I_0)^{-\alpha} & \alpha \leq -\alpha_0 \end{cases} \tag{6.10}$$

Here we use a six-degree polynomial. This method helps us to avoid discontinuous derivatives at $\alpha = 0$. It also makes sure that the extrapolated and interpolated values of the various nuisance parameters are non-negative.

6.3 Bayesian Analysis Toolkit (BAT)

The Bayesian Analysis Toolkit (BAT) [51] is a statistical tool that uses Bayesian methods to perform data analysis to check the validity of various theoretical models. It uses methods described in Section 5.2.2 and Section 5.2.3 to find the various properties of the parameters of interest. BAT uses Markov chain Monte Carlo (MCMC) methods to perform the numerical integration involved in the process of marginalisation as shown in (5.7). It takes binned histograms of the different processes as input.

The numerical implementation of marginalisation involves integrating over the multidimensional space of all the nuisance parameters. This integration can be thought as an approximation to a multidimensional random walk problem. Markov chain Monte Carlo (MCMC) is a commonly used method in statistics to model a multidimensional random walk. MCMC samples random numbers based on the nuisance parameter prior distributions to reach a stable distribution.

6.3.1 Markov chain Monte Carlo

A Markov chain is a sequence of random numbers $X_1, X_2, ..., X_n$ where the conditional probability distribution of X_{n+1} depends on X_n only. These chains are stochastic processes having a well-defined stationary distribution. A stochastic process is called stationary if

- $\forall\ k \in \mathbb{Z}^+$ the k-tuple $(X_{n+1}....X_{n+k})$ is independent of n.

A Markov chain is stationary if and only if X_n is independent of n. So it can easily be inferred that the chain can be completely defined using stationary distributions denoted by the *transition probability matrix* P_{ij}.

$$P(X_{n+1} = x_j | X_n = x_i) = p_{ij}\ i, j = 1....n \tag{6.11}$$

If the chain fulfils the following conditions, the chain is ergodic i.e. the probability to be in a given state is independent of the initial state.

- **Recurrence** The various states of the Markov chains are recurrent i.e.the probability of coming back to X_i after starting at X_i is 1 for infinite (long enough) run time.
- **Irreducibility** This property ensures that the transition from $X_i \rightarrow X_j\ \forall i, j$ can be achieved in finite time.
- **Aperiodicity** This condition makes sure that the there is no fixed period for coming back to a given state X_i. The chain can return to the state at irregular times.

BAT uses multiple Markov chains to compute the stationary distribution $\pi(x_i)$. The posterior p.d.f. $P(\lambda_i, \vec{v}, M|\vec{D})$ as defined in (5.2) and (5.19) is the stationary distribution computed by BAT. This is illustrated in 6.1. An algorithm called the Metropolis-Hastings algorithm [62, 63] is used to compute the posterior p.d.f.. The algorithm is a two-step algorithm

1. A state $X_i = \vec{x}$ described by the distribution $\pi(\vec{x})$ is chosen at random. A new state $\vec{y}$ is proposed according to a symmetric proposal function $f(\vec{y}, \vec{x})$
2. We define a quantity $r = \frac{\pi(\vec{y})}{\pi(\vec{x})}$ to decide the transition of the Markov chain. Subsequently, r is compared with a random number, U, sampled from a uniform distribution within $[0, 1]$. If $U < r$, $X_{i+1} = \vec{y}$ otherwise $X_{i+1} = \vec{x}$.

For some reasonable proposal functions f, this algorithm fulfils the conditions of an ergodic Markov chain with a stationary distribution $\pi(\vec{x})$. The algorithm is implemented in BAT using a pre-run and a main run. In the pre-run various Markov chains are run in parallel to ensure convergence of different

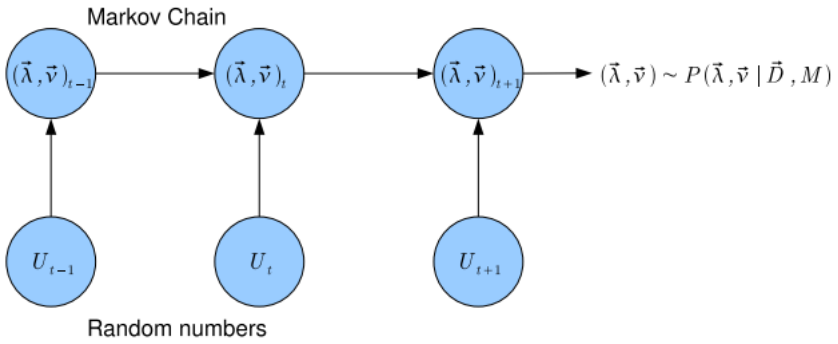


Figure 6.1: An illustration of MCMC [51]

run parameters. The convergence criterion is set to be $R \approx 1$ where R is defined as the square root of the estimated variance of the desired posterior and the mean of the variance of all chains.

$$R = \sqrt{\frac{\hat{\sigma}_\pi}{\bar{\sigma}_{chains}}} \tag{6.12}$$

$$\bar{\sigma}_{chains} = \frac{1}{m(n-1)} \sum_{j=1}^{m} \sum_{1=1}^{n} (x_i - \bar{x}_j)^2 \tag{6.13}$$

$$\hat{\sigma}_\pi = \frac{n-1}{n} \cdot \bar{\sigma}_{chains} + \frac{1}{m-1} \sum_{j=1}^{m} (x_j - \bar{x})^2 \tag{6.14}$$

Here m is the number of chains and n are the number of elements or iterations in the Markov chain. The main run is performed after this to compute the probability distribution of interest.

6.4 Binned Log Likelihood (BILL) fitter

The BILL fitter [64] uses the frequentist method of binned maximum log-likelihood fit as described in Section 5.2.1 to evaluate the estimators of the parameters of interest. It uses a product of Poisson likelihood for each bin and Gaussian constraint terms for various systematic uncertainties. The likelihood for P processes in B bins can be written as

$$\mathcal{L}(\beta_1, ..., \beta_P) = \prod_{k=1}^{B} \frac{e^{-E_k} \cdot E_k^{N_k}}{N_k!} \prod_{j=2}^{P} G(\beta_j, \Delta_j) \tag{6.15}$$

- β_1 is the relative signal cross-section normalised to the theory prediction.
- $\beta_2 ... \beta_P$ are the relative background cross-sections (rates) normalised to the theory predictions.
- Δ_j is the relative uncertainty on the predicted rate of the process j.
- E_k and N_k are the expected and observed events in bin k respectively.

Binned histograms are used as inputs by BILL to do the fit. It uses MINUIT to do the negative log likelihood fit. The fit is done by creating thousands of pseudoexperiments. For a process j, we create a Poisson distribution around the expected number of events E_j. Subsequently, we draw a value at random

from the distribution and fit to the value. Each of these experiments gives us an estimate of $\hat{\beta}_j$. After performing all these pseudoexperiments we get a distribution for $\hat{\beta}_j$, the median of which gives us the fitted value of $\hat{\beta}_j$.

Incorporating systematic uncertainties

BILL incorporates both rate and shape uncertainties by integrating over the whole parameter space of uncertainties. It assumes a Gaussian constraint on all the systematics. For each uncertainty, BILL introduces a nuisance parameter δ_i which gives a measure of the uncertainty. In each pseudoexperiment a δ_i is chosen at random from a Gaussian distribution. We use histograms corresponding to $\pm 1\sigma$ variation of the systematic variations. The choice of δ_i for a given pseudoexperiment is decided by these variations. The pseudoexperiments calculate the fitted β_j by varying the systematics both individually and simultaneously. These values give us an estimate of the uncertainties on the post-fitted scale factors.

CHAPTER 7

Statistical analysis

"*Life's most important questions are, for the most part, nothing but probability problems.*" -- Pierre-Simon Laplace

In this chapter we describe the statistical analysis used in the search for $b^*/B' \to Wt$. The statistical methods and tools used for this analysis have been explained in detail in Chapters 5 and 6. We first test the validity of the signal-plus-background (b^*/B'-plus-Standard Model) hypothesis against the background-only (Standard Model) hypothesis. Subsequently, we set limits on the cross-section times branching section for the process $pp \to b^*/B' \to Wt$. This analysis is done using ATLAS data taken in 2012 at $\sqrt{s} = 8$ TeV with an integrated luminosity of 20.3 fb^{-1}.

7.1 Hypothesis test

For testing the validity of signal-plus-background hypothesis, we use the statistical tool `BumpHunter`. The details of this method and the tool have been described in Sections 5.4 and 6.1. Rather than using the original Monte Carlo simulations of the background, we fit the backgrounds to the data in the control regions and propagate their scale factors to the signal regions. This method ensures that we get use the correct normalisation scale factors of the backgrounds. The post-fit errors also give us a better estimate of the effect of all the systematic uncertainties. We compare these backgrounds against the data in the signal regions to look for any possible excess. This test gives us a global p-value as defined in (6.2) which is used to find the significance of this test.

7.2 Limit setting

In this analysis we also set limits on the signal strength μ, defined as the product of the cross-section for the process $pp \to b^*/B'$ times the branching fraction $\mathcal{BR}(b^*/B' \to Wt)$. We set limits on the strength of signal hypotheses for varying masses of b^*/B'. The theoretical cross-sections are hypothesised according to the models described in Sections 2.3.1 and 2.3.2. The cross-sections for the different mass hypotheses for b^* and B' are given in Table 7.2 and Table 7.3 respectively. For B' we consider the theoretical cross-section corresponding to $\lambda = 2, 3, 4$, where λ is a coupling parameter. The limits are set for different mass hypotheses using the CL_s method and a Bayesian method. We use frequentist and Bayesian methods of likelihood template fitting, as described in Section 5.2.4, to fit the various binned signal and background templates to extract the respective limits. The likelihood is a product over the individual likelihoods of all the signal and control regions. The different signal and control regions used in this analysis is shown in Table 7.1. The distributions for the discriminant of the different backgrounds and signals of varying masses in all these regions are shown in Figures 7.1-7.3.

Single lepton						Dilepton		
SR		$t\bar{t}$ CR		W+ jets CR		SR		CR
hadT	hadW	hadT	hadW	hadT	hadW	1-jet 1-btag	2-jet 1-btag	2-jet 2-btag

Table 7.1: Different signal and control regions. SR and CR denote signal region and control region respectively. hadT and hadW denotes hadronic top and hadronic W decay modes respectively.

M_{b^*} [GeV]	$\sigma(b^* \to Wt)$ [pb]	Scale [pb]	
		Down	Up
600	34.74	39.31	30.71
800	7.52	8.69	6.54
1000	1.99	2.34	1.71
1100	1.09	1.28	0.93
1200	0.61	0.73	0.52
1300	0.35	0.42	0.30
1400	0.21	0.25	0.17
1600	0.08	0.09	0.06

Table 7.2: Total cross-section for $b^* \to Wt$ in the mass range under investigation for couplings $f_g = f_L = f_R = 1$, and cross-section when the renormalisation and factorisation scale is varied down and up by a factor two. These variations are done separately for the up and down case.

$M_{B'}$ [GeV]	$\sigma(B' \to Wt)$ [fb]		
	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$
500	438	—	—
600	226	249	—
700	104	169	9.5
800	52	97	95
900	28	54	69
1000	15	30	43
1200	4.8	10.2	16

Table 7.3: Cross-section for the process $B' \to Wt$ production for several different B' masses and coupling values λ. The branching fraction is taken into account here.

7.2.1 Bayesian approach

In the Bayesian approach used here, we use Gaussian priors, $f_{\mathcal{N}}(\theta_i; 0, 1)$, on the various nuisance parameters θ_i and a flat prior, $\pi(\mu)$, on the signal strength. Using (5.2), (5.7), and (5.16) the marginalised posterior p.d.f. of the signal strength can be written as

$$L(\mu|\vec{D}) \propto \int_{\theta} \mathcal{L}(\mu, \theta)\, \pi(\mu) \prod_i f_{\mathcal{N}}(\theta_i; 0, 1) \tag{7.1}$$

Here the Poisson likelihood is a function of the signal strength and different nuisance parameters. As mentioned in earlier chapters, the nuisance parameters contain information about all systematic and statistical uncertainties. The scale factors of the background processes are also treated as nuisance

parameters in the Bayesian approach.

$$\mathcal{L}(\mu, \theta) = \prod_{k}^{\text{bins}} \frac{{E_k}^{N_k} \mathrm{e}^{-E_k}}{N_k!} \tag{7.2}$$

$$E_k \equiv E_k(\mu, \theta) = \mu E_k^{\text{signal}}(\theta) + \sum_{j}^{\text{bkgs.}} E_k^j(\theta) \tag{7.3}$$

We use binned distributions of the discriminant to calculate the likelihood. These distributions are shown in Figures 7.1-7.3. The value of the likelihood is calculated using `Histfactory` and the marginalised posterior PDF is computed using `BAT`. $E_k^j(\theta)$ is estimated for arbitrary θ_i using morphing algorithms of `Histfactory` as illustrated in Section 6.2. We use one Markov chain with $O(10^6)$ iterations for computing the marginalised posterior. Subsequently, we compute the one-sided Bayesian 95% C.L. upper limit, $\mu_{\text{up}}^{\text{obs}}$, using the equation

$$\int_0^{\mu_{\text{up}}^{\text{obs}}} L(\mu)\, \mathrm{d}\mu = 0.95 \tag{7.4}$$

For extracting the expected limits, we fit to Asimov data [65]. The Asimov dataset is defined in such a manner that if we fit to it, the estimated values are the true expected values of the parameters of interest. This procedure helps us to reduce the computation time in a drastic manner. The errors on the expected limits are computed using

$$\mu_{\text{up}\pm N} = \mu_{\text{up}}^{\text{med}} \left(\Phi^{-1}(1 - 0.05\Phi(\pm N)) \pm N \right) \tag{7.5}$$

where $N = 1, 2$ corresponds to $1, 2\ \sigma$ error bands respectively [66]. Here $\Phi(x)$ denotes the Gaussian cumulative distribution function (c.d.f.) and $\Phi^{-1}(y)$ is the Gaussian quantile and it can be written as the inverse of the Gaussian c.d.f..

7.2.2 Frequentist approach

We use `BILL` fitter to compute the CL_s limit as defined in Section 5.5.1. We generate 10000 pseudoexperiments to find the upper limits. The negative log likelihood fit method is used for fitting the signal and background templates, as described in Section 6.4. The fit incorporates the systematic uncertainties, but they are not profiled. The templates used for the fit are same for both the Bayesian and frequentist approach.

7.3 Systematic uncertainties

In this section, we would describe the different systematic uncertainties that have been considered in our analysis. The uncertainties are chosen according to the top group recommendations [67]. Most of the systematic uncertainties were independent of the signal or background process, but uncertainties like generator systematic and normalisation uncertainties are process dependent.

- **Luminosity uncertainty** The uncertainty on the integrated luminosity is 2.8%. We use a rate uncertainty to account for this variation.
- **Lepton energy scale and resolution**: For estimating the scale uncertainty, the lepton p_{T} is scaled by $\pm 1\sigma$ and the event selection is re-applied for the respective variations. For the resolution uncertainty, we smear the lepton p_{T} and re-apply the event selection. The variation in the acceptance gives us a measure of the uncertainty.

- **Jet energy resolution (JER)** This uncertainty is calculated by smearing the jet p_{T}. It gives us a measure of the variation of the selection efficiency of jets.

- **Jet vertex fraction (JVF)** This is the uncertainty associated with the jet vertex fraction cut in a given analysis. Jets having $p_{\mathrm{T}} < 50$ GeV and $|\eta| < 2.4$ contribute to this uncertainty.

- **Jet energy scale (JES)** This is one of the major sources of uncertainty for many top-quark physics analyses. It depends on the η and p_{T} of the reconstructed jet. Multiple factors contribute to this uncertainty, so we need different sub-components to describe this uncertainty accurately. The different uncertainties used are
  - Statistical uncertainty. (JesEffectiveStat)
  - Modelling uncertainty. (JesEffectiveModel)
  - Uncertainty from the detector components. (JesEffectiveDet)
  - Uncertainty from detector sources and modelling (JesEffectiveMix)
  - η inter-calibration uncertainty. (modelling and statistical)
  - Uncertainties arising from pileup.
  - Single particle uncertainty.
  - Uncertainty from response of b-jet. (BJesUnc)

- **Tagging uncertainties** These uncertainties arise from the variation in efficiency of tagging the different jets. We have three kinds of uncorrelated tagging uncertainties, namely b-tag, c-tag and mistag. They represent the tagging uncertainties corresponding to b-jets, c-jets and other jets. They are computed by measuring the change in tagging efficiencies while we vary the η, p_{T}, and flavour-dependent scale factors of the jets.

- **Large-R jet energy and mass scale** Since our analysis uses large-R jets, we consider the scale uncertainties arising from the energy and mass of these jets separately.

- **Large-R jet energy and mass resolution** We also consider the resolution uncertainties for the mass and energy of these jets separately.

- **Parton distribution function** This uncertainty comes from the variation in acceptance due to different choice of parton distribution functions. We have modelled the uncertainty by re-weighting the top pair, Wt, and the signal MC samples according to the momentum distributions of the colliding partons, as predicted by three different sets of parton distribution functions.

- **Initial and final state radiation (ISR/FSR)** These uncertainties reflect the effect of varying radiation in the initial or final state. As it is done in other top group analyses, these uncertainties are calculated using certain `Alpgen` and `Pythia` samples.

- **Generator dependence** These uncertainties account for the variation in acceptance due to a different choice of Monte Carlo generators.

- **Parton shower modelling** We consider the uncertainties coming from modelling uncertainties of the parton shower.

- **Normalisation uncertainties** Various normalisation uncertainties of the backgrounds are taken into account. We assign rate uncertainties for cross-section uncertainties. These uncertainties are log-Gaussian i.e. they are Gaussian distributions of the log of the parameter. This ensures that the post-fit yields of the backgrounds are not negative. In addition to these uncertainties, we use an extra rate uncertainty associated with extra jets, called Berends scaling. For the dilepton channel, we assign a 24% uncertainty in the one and two jet bins. In the single-lepton channel we use a 34% uncertainty for the three jet bin and a 24% uncertainty for the two jet bin. The cross-section uncertainties used for the different backgrounds are shown in the following table [67].

Background	Cross-section uncertainty
$t\bar{t}$	$+5.7\%/-5.3\%$
W+jets	$\pm 4\%$
Wt	$\pm 6.8\%$
Z+jets	$\pm 4\%$
Diboson	$\pm 5\%$
Single top	$\pm 3.9\%$

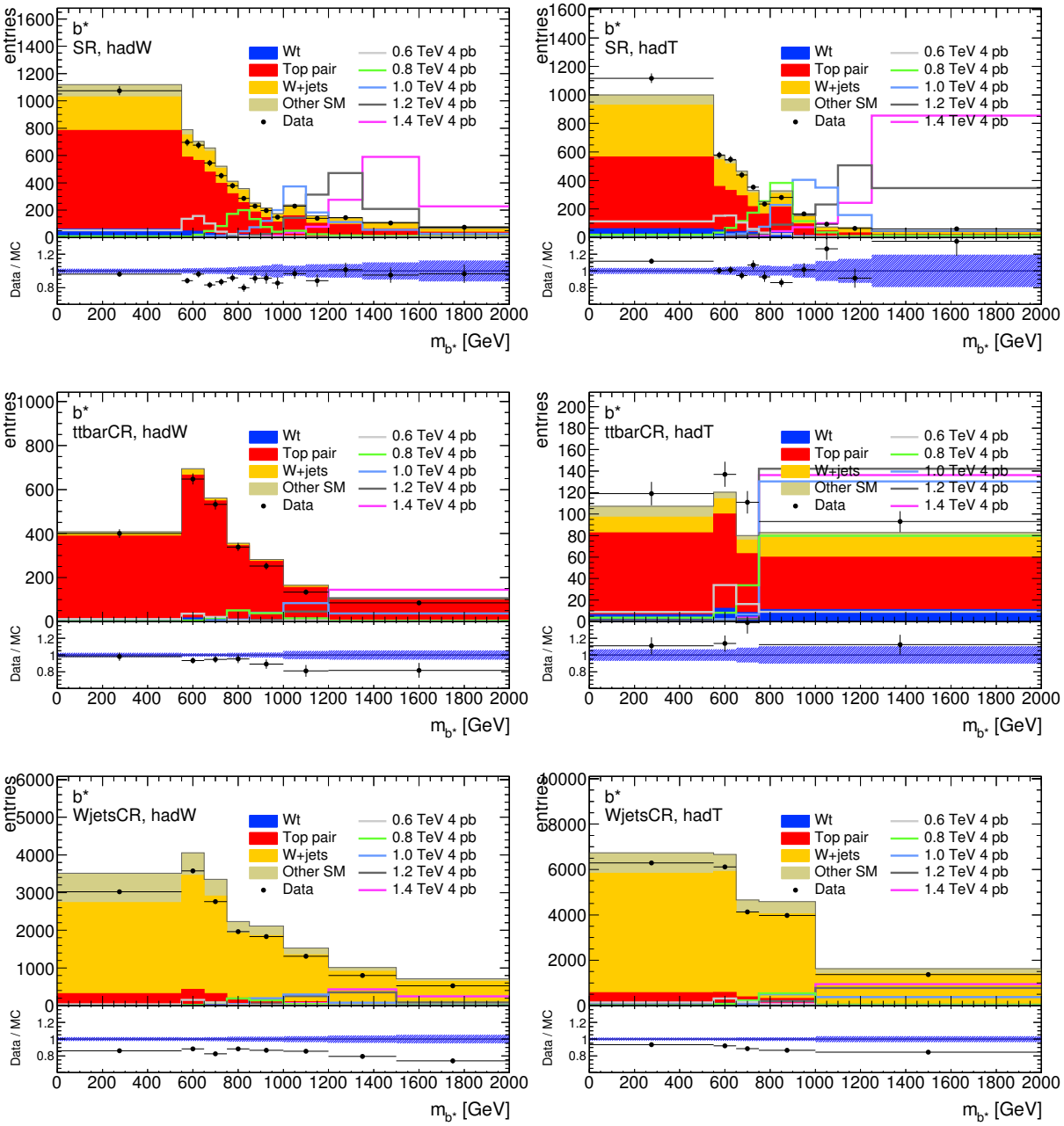


Figure 7.1: Distributions of the discriminant for b^* in the single-lepton channel. The signal templates have been scaled up, but they are normalised to 1 pb in the fit. The ratio plot shows the comparison between data and Monte Carlo. The uncertainty band denotes the statistical uncertainty. Small backgrounds like the diboson, other single top-quark backgrounds, and fakes are merged in the figure, but treated separately in the fit.

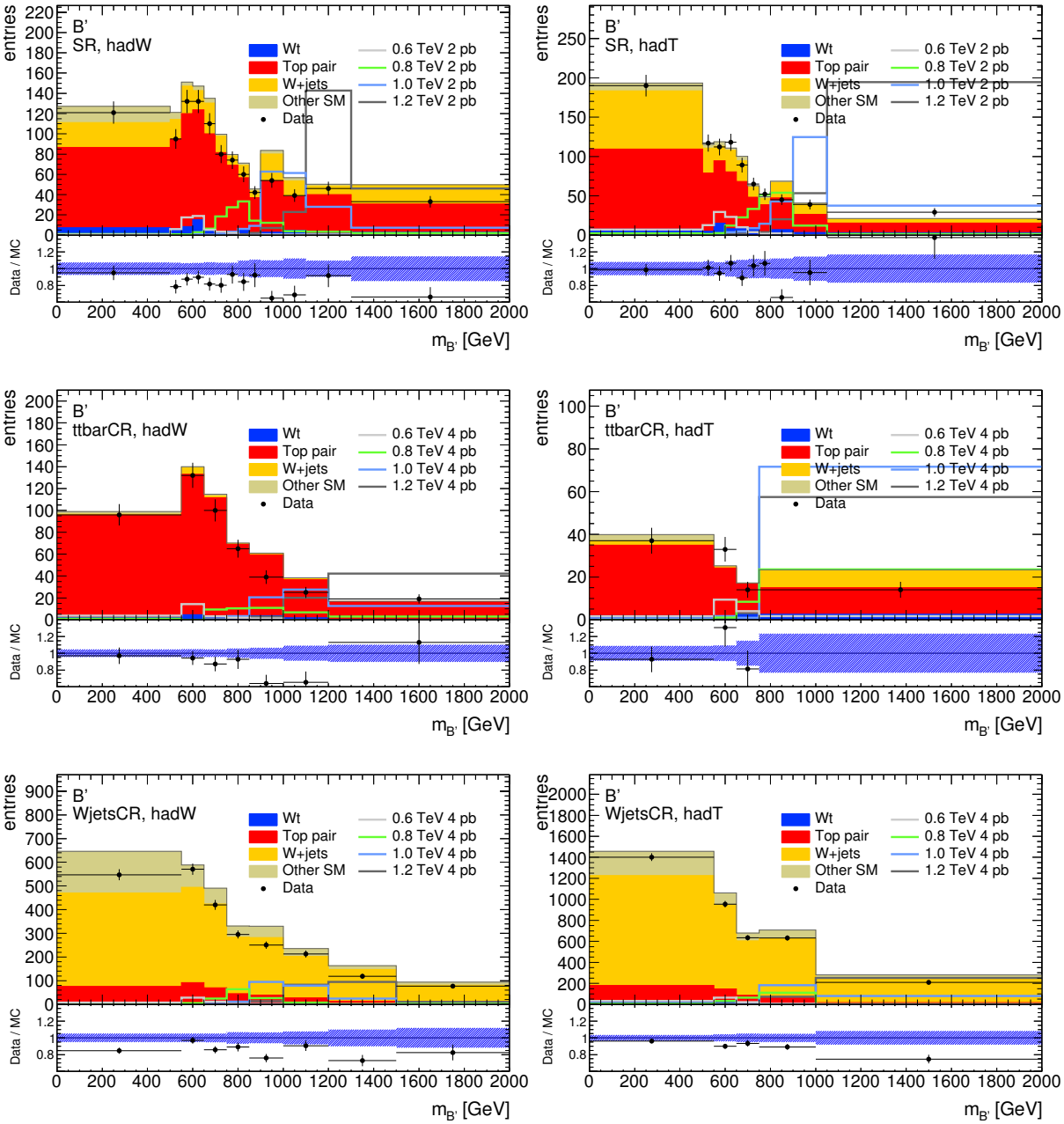


Figure 7.2: Distributions of the discriminant for B' in the single-lepton channel. The signal templates have been scaled up, but they are normalised to 1 pb in the fit. The ratio plot shows the comparison between data and Monte Carlo. The uncertainty band denotes the statistical uncertainty. Small backgrounds like the diboson, other single top-quark backgrounds, and fakes are merged in the figure, but treated separately in the fit.

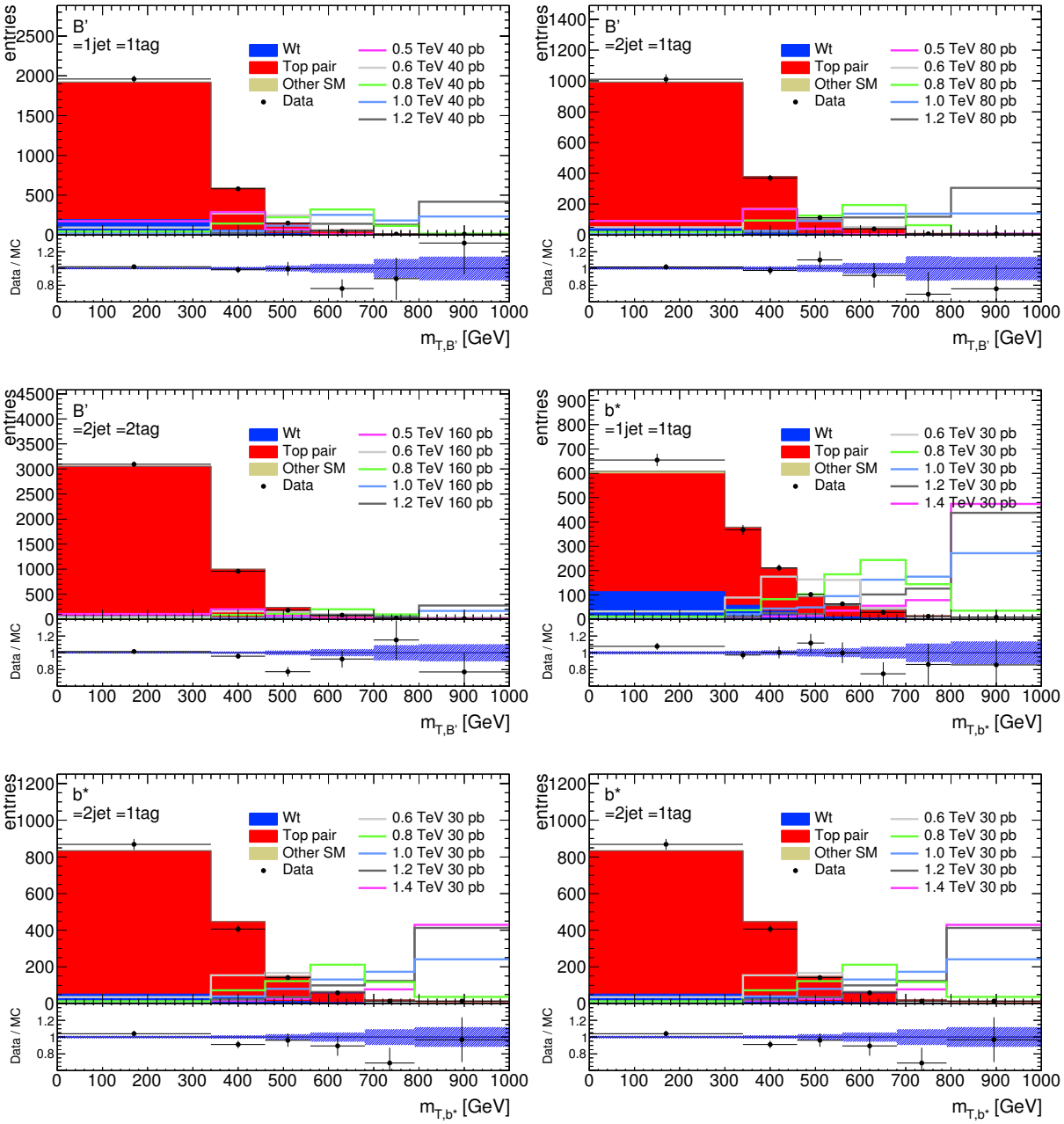


Figure 7.3: Distributions of the discriminant for B' and b^* in the dilepton channel. The signal templates have been scaled up, but they are normalised to 1 pb in the fit. The ratio plot shows the comparison between data and Monte Carlo. The uncertainty band denotes the statistical uncertainty. Small backgrounds like the diboson, other single top-quark backgrounds, and fakes are merged in the figure, but treated separately in the fit.

CHAPTER 8

Results

"*It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong.*" -- Richard P. Feynman

In this chapter the results obtained in our analysis will be presented. The statistical analysis was performed for both the single-lepton and dilepton channels as defined in Chapter 4. We first test our background-only hypothesis against our signal-plus-background hypothesis. Since this procedure cannot rule out the background hypothesis, we subsequently set limits on the signal strength. The signal strength and its likelihood has been defined in Chapter 6. The hypothesis testing was done for the signal regions. The limits are calculated by fitting in all signal and control regions for a given channel. The signal and control regions used are shown in Table 7.1.

8.1 `BumpHunter` results

8.1.1 Single-lepton channel

We test our background-only-hypothesis with `BumpHunter` in each of the signal regions. The test gives us a global p-value, as defined in (6.1), and we quote the corresponding significance σ using (5.23). In our analysis, we use 20000 pseudoexperiments while running `BumpHunter`. We present the distribution of the control region fitted Monte Carlo (the measure of our background-only hypothesis) versus the data in the signal region in Figures 8.1a and 8.2a. We also show a distribution of the bin-by-bin significance in these plots. The corresponding distributions of the `BumpHunter` test statistic, as defined in (6.1), which are used for the calculation of the global p-value is shown in Figures 8.1b and 8.2b. The global p-value and significance which was obtained for the hadronic top and hadronic W signal regions for both b^* and B' is quoted in Table 8.1. The significance obtained in the different signal regions is quite low, thus indicating the absence of any signal.

b^* SR	p-value	σ
Hadronic top	0.19	0.87
Hadronic W	0.44	0.14

B' SR	p-value	σ
Hadronic top	0.09	1.34
Hadronic W	0.49	0.01

Table 8.1: p-value and significance σ in the signal regions of the single-lepton channel for b^* and B'

8.1.2 Dilepton channel

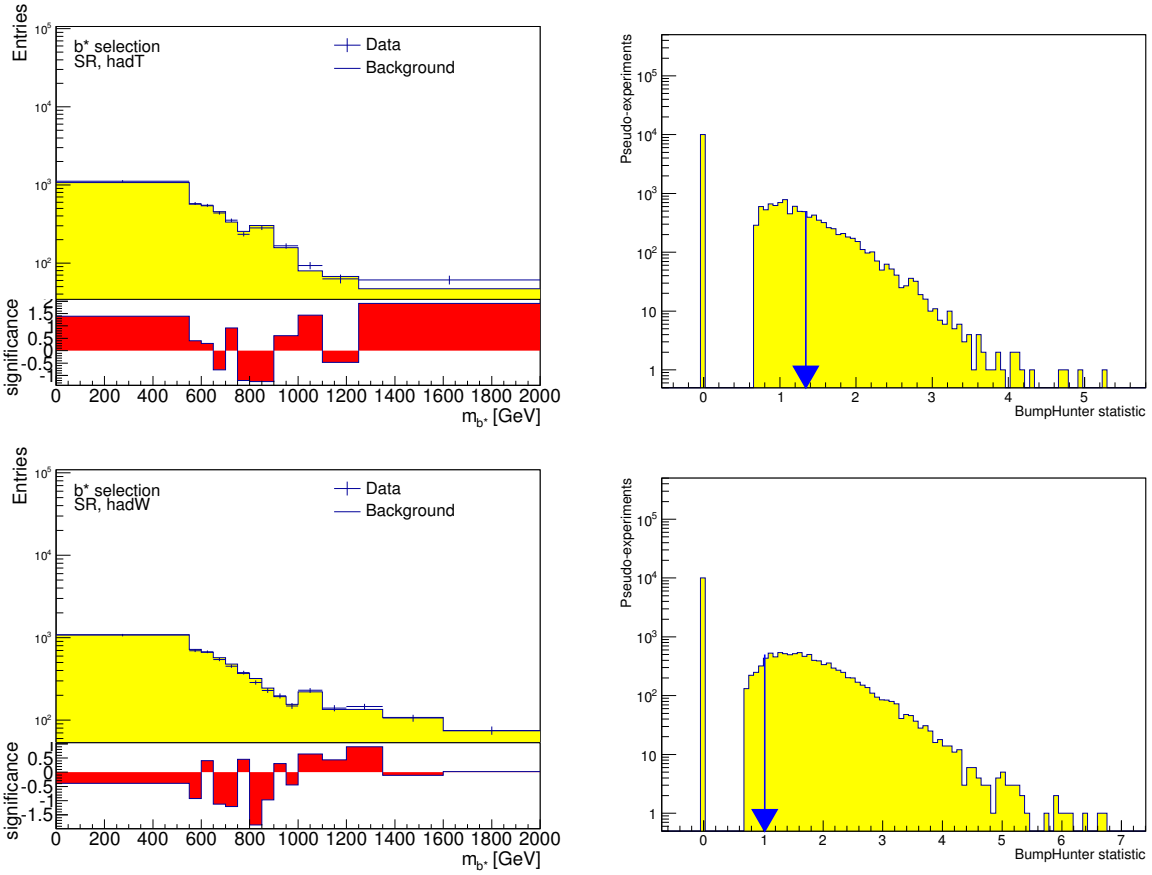
The aforementioned analysis is also performed for both signal regions of the dilepton channel, the 1-jet-1tag region and the 2-jet-1tag region. Figures 8.3a and 8.4a shows the control region fitted background

b^* SR	p-value	σ
1-jet 1-tag	0.40	0.27
2-jet 2-tag	0.50	0.00

B' SR	p-value	σ
1-jet 1-tag	0.35	0.38
2-jet 2-tag	0.24	0.70

Table 8.2: p-value and significance σ in the signal regions of the dilepton channel for b^* and B'

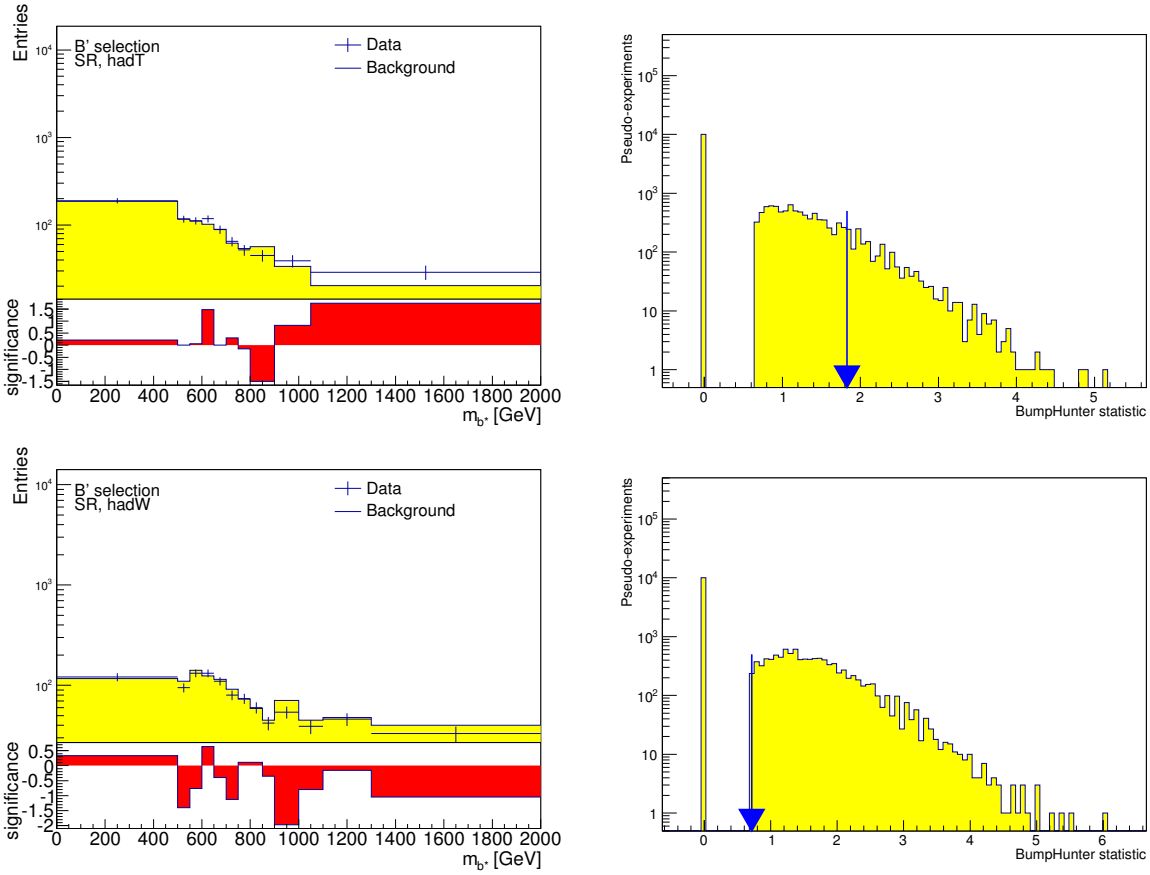
versus the data including a distribution of bin-by-bin significance. Figures 8.3b and 8.4b shows the corresponding distributions of the BumpHunter statistic. The global p-value and significance for the respective signal regions is quoted in Table 8.2. The significance obtained in the different signal regions of the dilepton channel is also small, thus indicating the absence of any signal.



(a) Data versus CR fitted MC and bin-by-bin significance

(b) BumpHunter test-statistic distribution

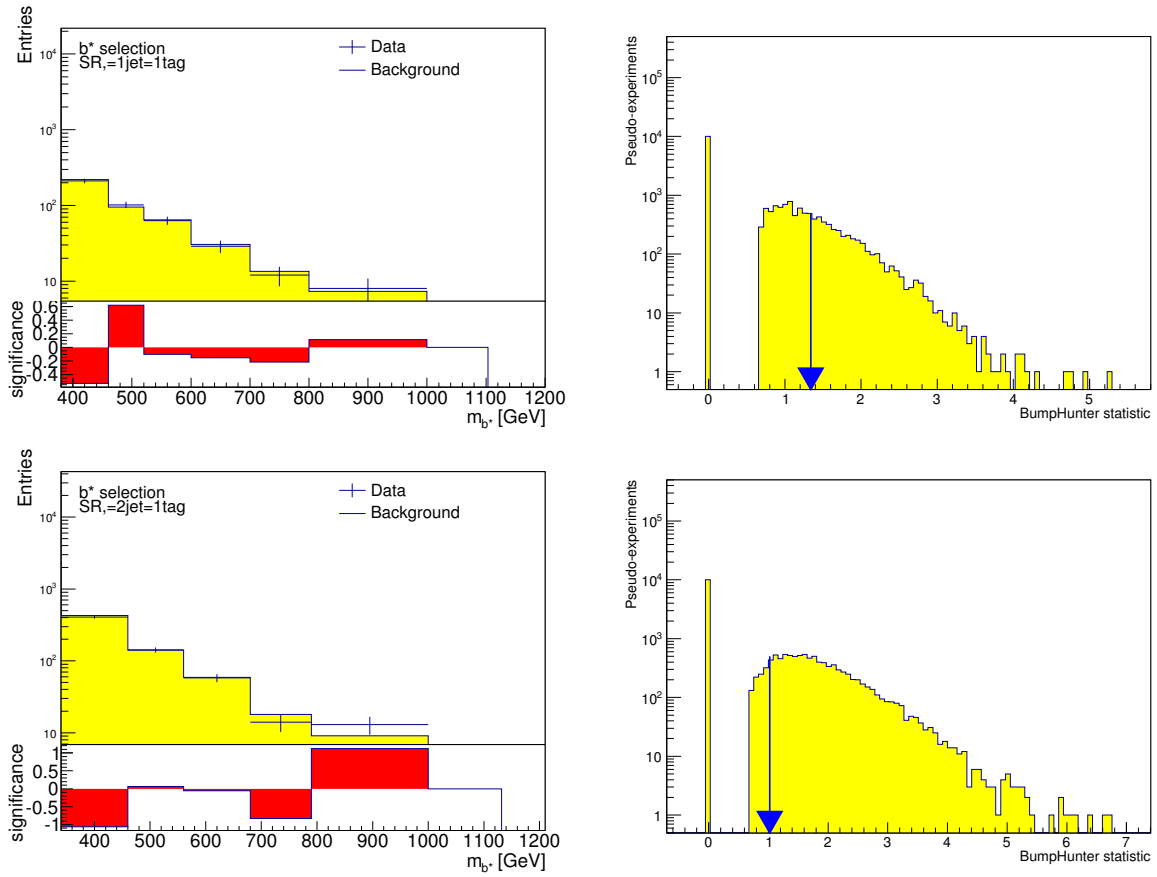
Figure 8.1: Figures 8.1a show the data versus CR fitted background for the hadronic top and hadronic W signal regions for B'. The significance plots illustrate the bin-by-bin significance. Figures 8.1b show a distribution of the BumpHunter test statistic for the pseudoexperiments performed by varying the background within its error range. The blue arrow denotes the value of the test statistic for the control region fitted background compared with the data.



(a) Data versus CR fitted MC and bin-by-bin significance

(b) `BumpHunter` test-statistic distribution

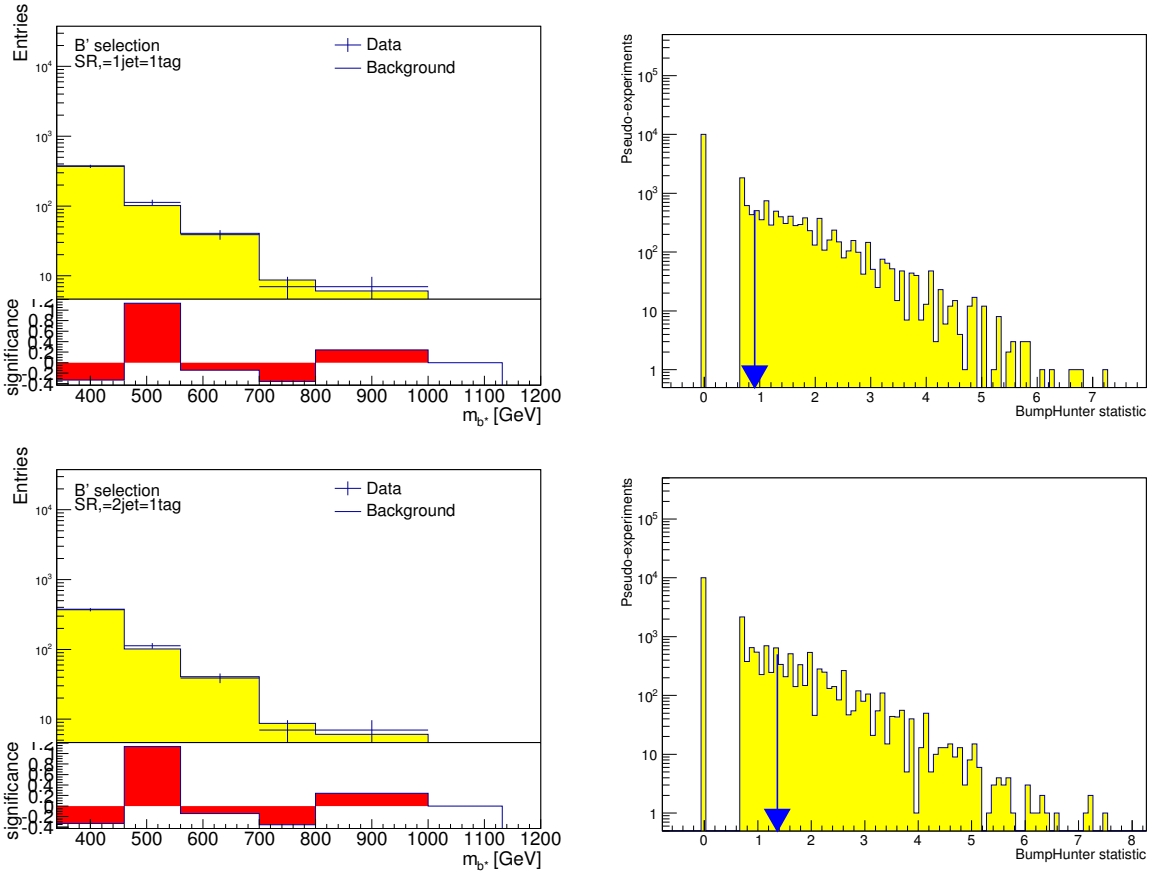
Figure 8.2: Figures 8.2a show the data versus CR fitted background for the hadronic top and hadronic W signal regions for B'. The significance plots illustrate the bin-by-bin significance. Figures 8.2b show a distribution of the `BumpHunter` test statistic for the pseudoexperiments performed by varying the background within its error range. The blue arrow denotes the value of the test statistic for the control region fitted background compared with the data.



(a) Data versus CR fitted MC and bin-by-bin significance

(b) `BumpHunter` test-statistic distribution

Figure 8.3: Figures 8.3a show the data versus CR fitted background for the hadronic top and hadronic W signal regions for B'. The significance plots illustrate the bin-by-bin significance. Figures 8.3b show a distribution of the `BumpHunter` test statistic for the pseudoexperiments performed by varying the background within its error range. The blue arrow denotes the value of the test statistic for the control region fitted background compared with the data.



(a) Data versus CR fitted MC and bin-by-bin significance

(b) `BumpHunter` test-statistic distribution

Figure 8.4: Figures 8.2a show the data versus CR fitted background for the hadronic top and hadronic W signal regions for B'. The significance plots illustrate the bin-by-bin significance. Figures 8.2b show a distribution of the `BumpHunter` test statistic for the pseudoexperiments performed by varying the background within its error range. The blue arrow denotes the value of the test statistic for the control region fitted background compared with the data.

8.2 Nuisance parameters

After performing our likelihood fit, we estimate the effect the different nuisance parameters (NPs). In our analysis, both systematic uncertainties and normalisation uncertainties are treated as nuisance parameters. We assume a Gaussian prior for each of the nuisance parameter, θ_i. So we expect the posterior distributions of the NPs to have similar shapes. We study the posterior distributions of the nuisance parameters to find their effect on our fit. We perform a background-only fit to Asimov data and real data to find the expected and observed constraints, respectively, of the nuisance parameters. The expected effect of the nuisance parameters on the fit for b^* in the combined channel is shown in Figure 8.5. The summary of the observed nuisance parameters for b^* in the single-lepton, dilepton and the combined channels is shown in Figures 8.6-8.8. The observed effect of the nuisance parameters for B' in the combined channel is illustrated in Figure 8.9. Similar plots for other channels can be found in Appendix A.

In the left part of these plots, we have plotted the median of the posterior distribution of θ_i and taken the $\pm 1\sigma$ quantiles of these distribution as the respective errors. The y-axis denotes the uncertainties that are considered in our analysis. They have been described in Section 7.3. The NPs having a suffix xsec denote the cross-section normalisation uncertainties. The suffix berends denote the Berends scaling normalisation uncertainty. NPs having the prefix Fat denote uncertainties arising from large-R jets. NPs containing the phrase Jes represent systematic uncertainties arising from the jet energy scale. The uncertainties containing the phrase tag depict tagging uncertainties. The generator systematics are denoted by the NPs having the prefix Gen. The PDF and PS NPs are the parton distribution function and parton shower uncertainties respectively. NPs having suffix el or mu denote uncertainties arising from electrons or muons. The wjets_flavor uncertainty is a 10% rate uncertainty assigned to account for the flavour composition of the W+jets. The green band is the $\pm 1\sigma$ variation of a unit Gaussian. When we fit to Asimov data, we set a prefit value of $\theta_i = 0$, so we expect most nuisance parameters to be unit Gaussian after the fit, as observed in Figure 8.5.

The NPs which are constrained have posteriors with width less than unity. Another implication of this constraint is that their posterior distributions are not centred i.e. they have a median away from zero. For example, in Figure 8.5 the wjets_berends4 NP (Berends scaling uncertainty) has a median away from zero and a width less than unity implying that it is constrained and thus has an effect on the fit. Similarly, if we look at the btag0 NP (b-tagging systematic) we see that it resembles a unit Gaussian, thus implying the fact that it has no effect on the fit. For the fit to observed data, we expect a larger variation than in the expected case because there are no prefit assumptions in the fit to data, as it can be seen in Figures 8.6-8.9. If we compare the expected and observed nuisance parameter distributions we observe a reasonable agreement between the nature of the constraints of the NPs. For example, we observe that the wjets_berends4 uncertainty is constrained in both cases, as seen in Figures 8.5 and 8.8. When we compare the NPs for B' and b^* we observe similar trends in the constraints of the NPs. Figures 8.6-8.8 show the observed nuisance parameter distributions for the single-lepton, dilepton and combined channels. We clearly observe the fact that the NPs in the combined channel is more sensitive to the single-lepton channel than the dilepton channel.

The plot on the right side of Figures 8.5-8.9 illustrates the relative impact of the respective nuisance parameters on the signal strength of different signal hypotheses of varying masses. We define the relative impact of a nuisance parameter θ_i on the signal strength μ as posterior $\mathrm{cov}(\theta_i, \mu)/\sigma(\mu)$, where posterior $\mathrm{cov}(\theta_i, \mu)$ is the covariance of the posterior distributions of θ_i and μ and σ_μ is the standard deviation of the posterior distribution of the signal strength μ. The values of the relative impact for all the NPs were calculated by using the posterior distributions, after fitting our likelihood to data for different masses of b^*/B'.

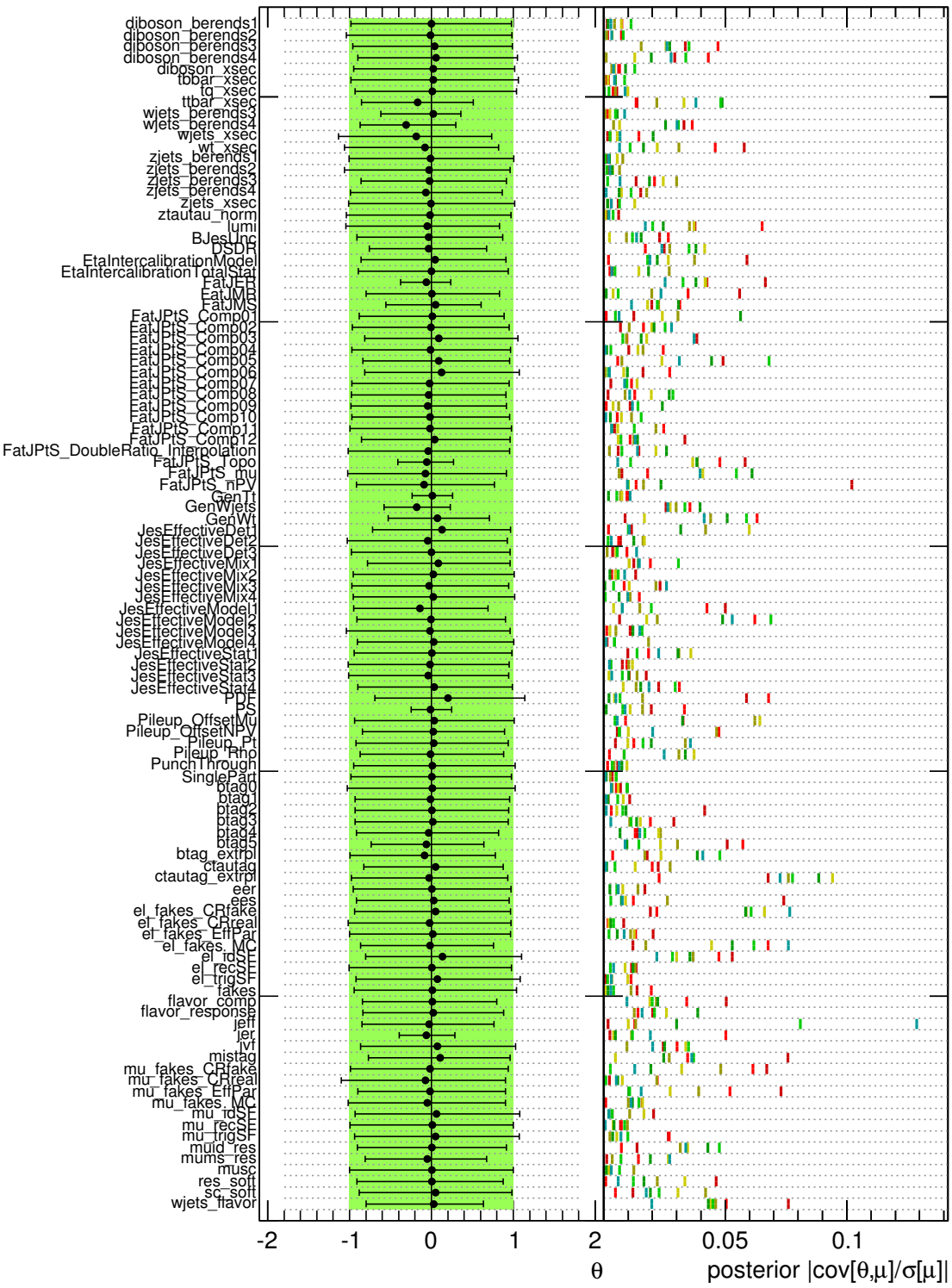


Figure 8.5: Expected nuisance parameter distributions for b^* in the combined fit of the single-lepton and dilepton channels. The left part visualises the median and $\pm 1\sigma$ quantiles of the posterior p.d.f.s on each nuisance parameter before and after the fit of the background-only hypothesis. The right part visualizes the relative impact (as defined in the text) on each of the signal hypotheses. Signals with lower mass are denoted by reddish whereas signals with heavier masses by bluish colours.

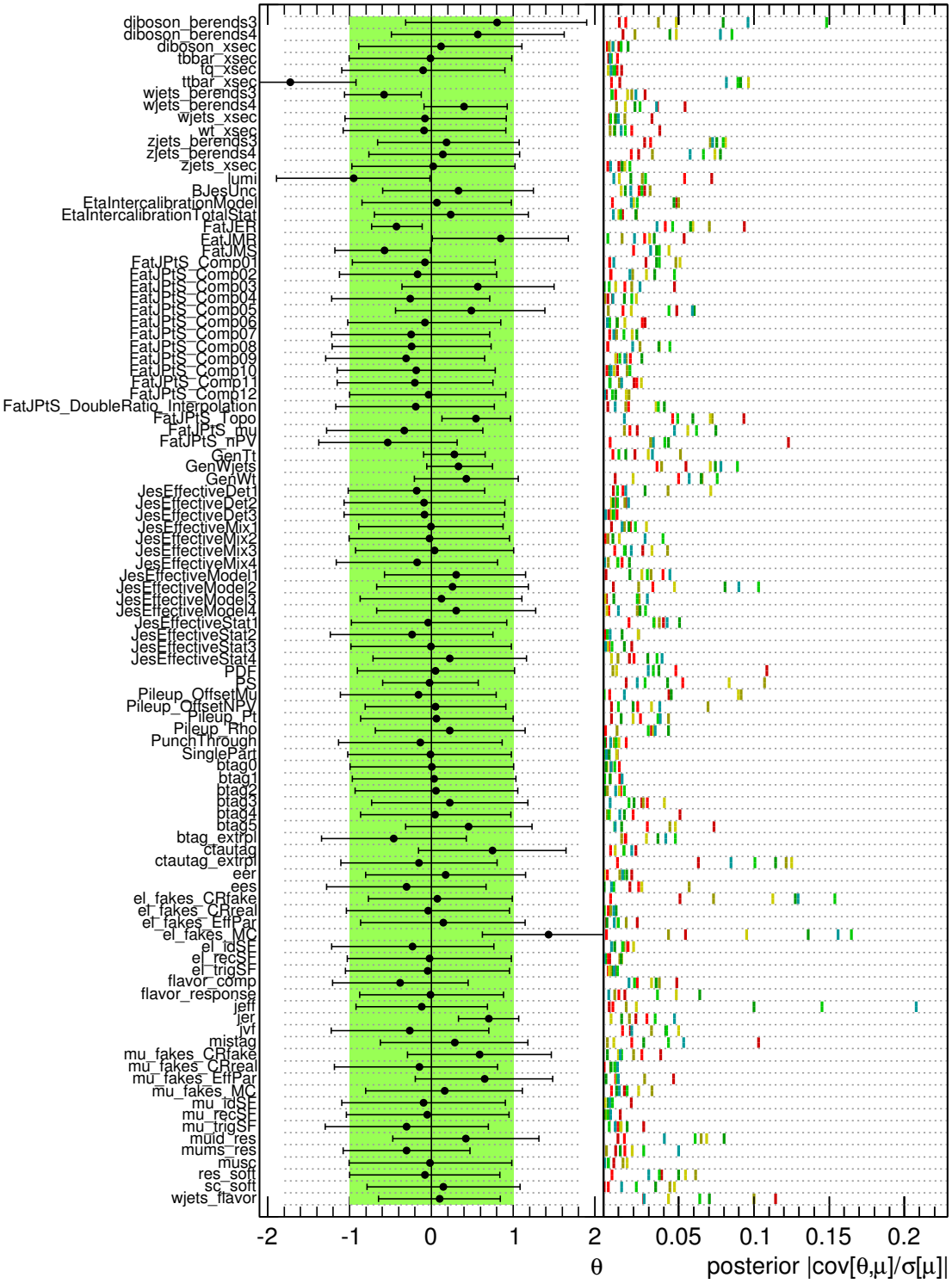


Figure 8.6: Observed nuisance parameter distributions for b^* in the single-lepton channel. The left part visualises the median and $\pm 1\sigma$ quantiles of the posterior p.d.f.s on each nuisance parameter before and after the fit of the background-only hypothesis. The right part visualizes the relative impact (as defined in the text) on each of the signal hypotheses. Signals with lower mass are denoted by reddish whereas signals with heavier masses by bluish colours.

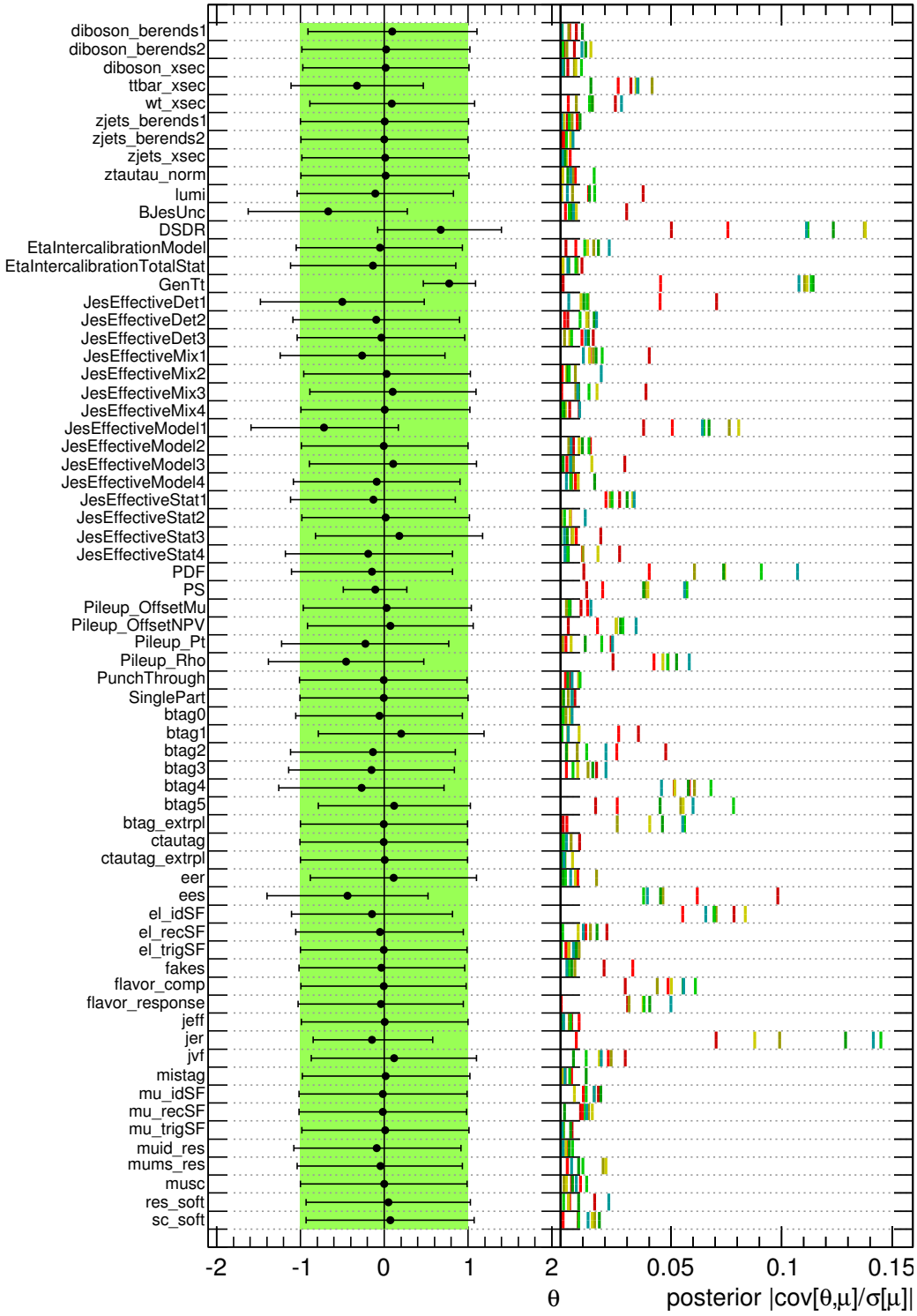


Figure 8.7: Observed nuisance parameter distributions for b^* in the dilepton channel. The left part visualises the median and $\pm 1\sigma$ quantiles of the posterior p.d.f.s on each nuisance parameter before and after the fit of the background-only hypothesis. The right part visualizes the relative impact (as defined in the text) on each of the signal hypotheses. Signals with lower mass are denoted by reddish whereas signals with heavier masses by bluish colours.

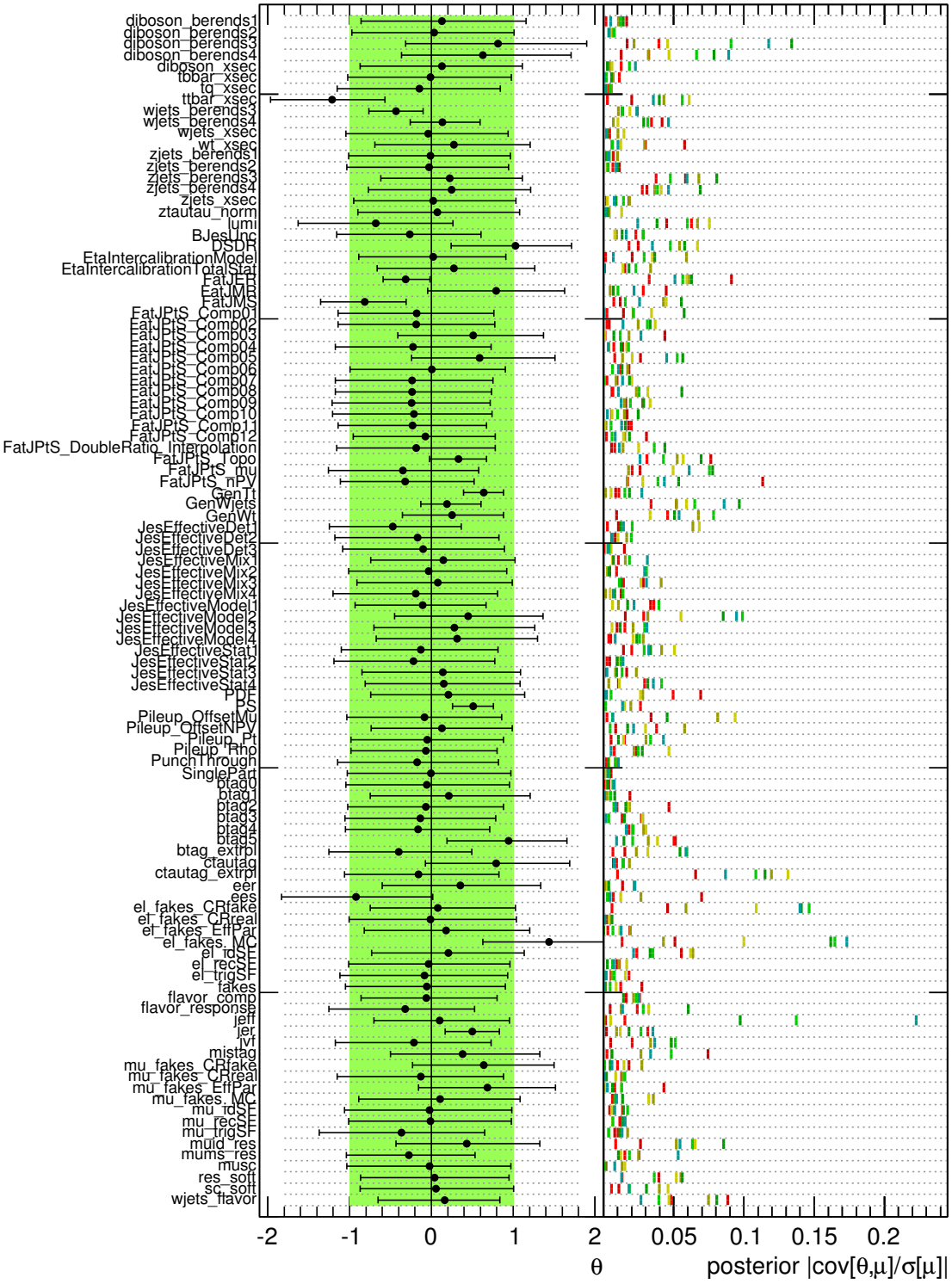


Figure 8.8: Observed nuisance parameter distributions for b^* in the combined fit of the single-lepton and dilepton channel. The left part visualises the median and $\pm 1\sigma$ quantiles of the posterior p.d.f.s on each nuisance parameter before and after the fit of the background-only hypothesis. The right part visualizes the relative impact (as defined in the text) on each of the signal hypotheses. Signals with lower mass are denoted by reddish whereas signals with heavier masses by bluish colours.

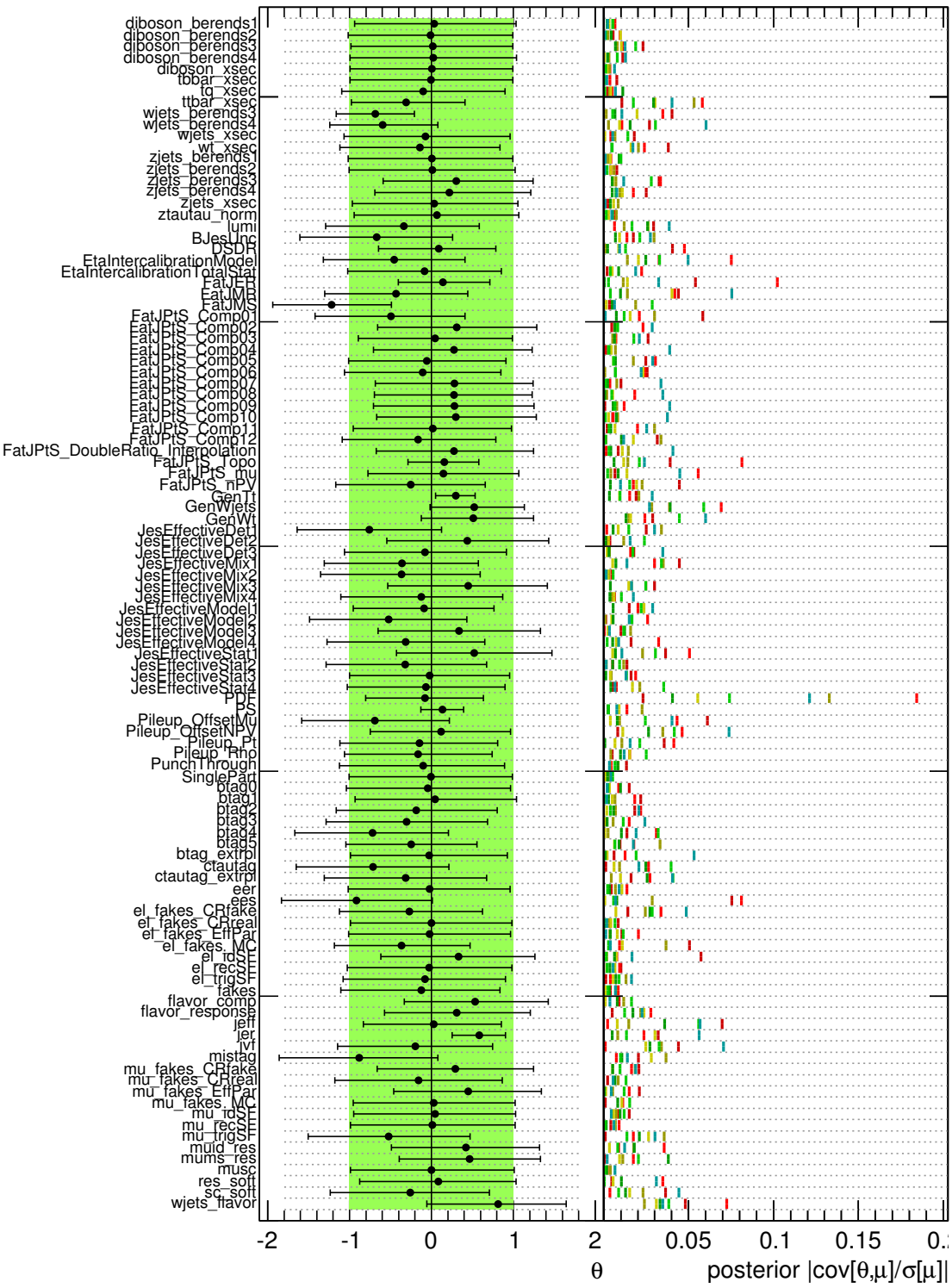


Figure 8.9: Observed nuisance parameter distributions for B' in the combined fit of the single-lepton and dilepton channel. The left part visualises the median and $\pm 1\sigma$ quantiles of the posterior p.d.f.s on each nuisance parameter before and after the fit of the background-only hypothesis. The right part visualizes the relative impact (as defined in the text) on each of the signal hypotheses. Signals with lower mass are denoted by reddish whereas signals with heavier masses by bluish colours.

8.3 Pre-and post-fit background distributions

We fit our background-only hypothesis to the data in both signal and control regions to get an estimate of the post-fitted backgrounds. We get the yields E_{ij} for process i in channel j by calculating the expectation value of the corresponding posterior distributions. The errors on these yields are taken to be standard deviation of the posterior distributions. These errors give us a measure of the total systematic and statistical uncertainties. We calculate a β factor, defined as the ratio of prefit and postfit yields for E_{ij}, for all the backgrounds in all the signal and control regions. These yields along with the total errors and β factors are shown in Tables 8.3-8.6. In our analysis the major backgrounds are the top-pair, Wt and the other single top-quark backgrounds (both of them are combined in the Table 8.3-8.6), diboson backgrounds and fakes. The β factors of the major backgrounds are close to one.

We plot these pre and post-fit yields, along with their respective errors in Figures 8.13-8.16. These distributions are for the discriminant used for fitting in the respective regions. Since the fit used to compute the yields is a background-only fit, we expect a reasonable agreement between data and Monte Carlo simulations. This can be mostly verified in Figures 8.13-8.16. The uncertainty bands in these plots include the total uncertainty i.e both statistical and systematic uncertainties. The post-fit plots illustrate how the fit helps us to constrain the total uncertainty. These plots show how the fit affects the composition of backgrounds having highly constrained normalisation uncertainties. This can be clearly seen in the case of W+jets in Figure 8.13. These post-fit distributions along with the nuisance parameter summary plots help us to verify the validity of our fit.

	Before fit	After fit	β
=1-jet =1tag (SR)			
$t\bar{t}$	1160	1165±36	1.00±0.03
Single t	236	253±26	1.07±0.11
WW,WZ,ZZ,Z+jets	13	13±4	1.03±0.32
Fakes	5	6±4	1.12±0.74
All backgrounds	1414	1436±33	
=2-jet =1tag (SR)			
$t\bar{t}$	1417	1409±33	0.99±0.02
Single t	106	97±10	0.91±0.09
WW,WZ,ZZ,Z+jets	2	3±2	1.26±0.69
Fakes	2	2±1	1.00±0.74
All backgrounds	1527	1511±34	
=2-jet =2tag (CR)			
$t\bar{t}$	4287	4253±65	0.99±0.02
Single t	125	106±20	0.85±0.16
WW,WZ,ZZ,Z+jets	1	1±0	1.19±0.68
Fakes	3	4±3	1.13±0.82
All backgrounds	4417	4364±64	

Table 8.3: Yields for b^* selection in the dilepton channel before and after fit of the background-only hypothesis to real data. The numbers for the major backgrounds are also shown separately. β is the ratio of the post- over pre-fit yields. The quoted error is the total uncertainty, calculated as explained in the text.

	Before fit	After fit	β
=1-jet =1tag (SR)			
$t\bar{t}$	2441	2485±44	1.02±0.02
Single t	286	276±23	0.96±0.08
WW,WZ,ZZ,Z+jets	16	15±5	0.90±0.33
Fakes	3	6±4	1.62±1.25
All backgrounds	2747	2780±39	
=2-jet =1tag (SR)			
$t\bar{t}$	1448	1457±34	1.01±0.02
Single t	82	81±9	0.98±0.11
WW,WZ,ZZ,Z+jets	6	5±4	0.80±0.62
Fakes	1	2±2	1.83±1.42
All backgrounds	1537	1545±35	
=2-jet =2tag (CR)			
$t\bar{t}$	4287	4229±67	0.99±0.02
Single t	125	119±19	0.95±0.15
WW,WZ,ZZ,Z+jets	1	1±0	1.16±0.62
Fakes	3	3±3	1.07±0.81
All backgrounds	4417	4352±65	

Table 8.4: Yields for B' selection in the dilepton channel before and after fit of the background-only hypothesis to real data. The numbers for the major backgrounds are also shown separately. β is the ratio of the post- over pre-fit yields. The quoted error is the total uncertainty, calculated as explained in the text.

	Before fit	After fit	β
hadronic top SR			
$t\bar{t}$	1973	1863±63	0.94±0.03
Single t	363	368±33	1.01±0.09
W+jets	1289	1358±145	1.05±0.11
WW,WZ,ZZ,Z+jets	154	232±97	1.51±0.63
Fakes	62	62±38	1.00±0.61
All backgrounds	3841	3883±72	
hadronic W SR			
$t\bar{t}$	3771	3387±115	0.90±0.03
Single t	521	520±40	1.00±0.08
W+jets	1370	1287±123	0.94±0.09
WW,WZ,ZZ,Z+jets	147	183±47	1.25±0.32
Fakes	118	102±42	0.87±0.35
All backgrounds	5929	5480±73	
hadronic top $t\bar{t}$ CR			
$t\bar{t}$	259	264±15	1.02±0.06
Single t	47	52±8	1.11±0.16
W+jets	60	74±20	1.24±0.33
WW,WZ,ZZ,Z+jets	18	28±13	1.60±0.76
Fakes	7	14±7	1.98±1.00
All backgrounds	390	433±19	
hadronic W $t\bar{t}$ CR			
$t\bar{t}$	2331	2155±49	0.92±0.02
Single t	142	139±12	0.98±0.08
W+jets	73	80±18	1.10±0.25
WW,WZ,ZZ,Z+jets	13	18±6	1.35±0.41
Fakes	12	13±10	1.16±0.83
All backgrounds	2573	2406±47	
hadronic top W+jets CR			
$t\bar{t}$	1586	1423±77	0.90±0.05
Single t	350	355±36	1.01±0.10
W+jets	19376	17558±817	0.91±0.04
WW,WZ,ZZ,Z+jets	1558	2142±686	1.38±0.44
Fakes	1395	442±427	0.32±0.31
All backgrounds	24265	21921±148	
hadronic W W+jets CR			
$t\bar{t}$	1420	1233±92	0.87±0.06
Single t	220	215±23	0.97±0.10
W+jets	14162	11839±625	0.84±0.04
WW,WZ,ZZ,Z+jets	1776	2054±580	1.16±0.33
Fakes	946	430±267	0.45±0.28
All backgrounds	18524	15771±132	

Table 8.5: Yields for b^* selection in the single-lepton channel before and after fit of the background-only hypothesis to real data. The numbers for the major backgrounds are also shown separately. β is the ratio of the post- over pre-fit yields. The quoted error is the total uncertainty, calculated as explained in the text.

	Before fit	After fit	β
hadronic t SR			
$t\bar{t}$	521	483±27	0.93±0.05
Single t	87	94±13	1.09±0.15
W+jets	237	217±39	0.92±0.17
WW,WZ,ZZ,Z+jets	19	22±7	1.15±0.38
Fakes	17	18±11	1.04±0.64
All backgrounds	881	835±34	
hadronic W SR			
$t\bar{t}$	810	746±34	0.92±0.04
Single t	123	125±13	1.02±0.11
W+jets	233	152±27	0.65±0.11
WW,WZ,ZZ,Z+jets	28	30±11	1.07±0.40
Fakes	24	27±14	1.13±0.56
All backgrounds	1218	1081±35	
hadronic t $t\bar{t}$ CR			
$t\bar{t}$	77	69±7	0.89±0.09
Single t	14	14±3	1.03±0.23
W+jets	11	11±6	1.01±0.53
WW,WZ,ZZ,Z+jets	2	2±1	0.98±0.42
Fakes	1	3±1	1.72±1.00
All backgrounds	106	99±8	
hadronic W $t\bar{t}$ CR			
$t\bar{t}$	479	428±18	0.89±0.04
Single t	43	38±5	0.89±0.11
W+jets	12	8±3	0.66±0.22
WW,WZ,ZZ,Z+jets	3	3±1	0.93±0.47
Fakes	3	6±4	1.91±1.30
All backgrounds	541	483±20	
hadronic t W+jets CR			
$t\bar{t}$	422	409±31	0.97±0.07
Single t	83	85±14	1.02±0.17
W+jets	3141	2836±172	0.90±0.05
WW,WZ,ZZ,Z+jets	253	284±79	1.12±0.31
Fakes	286	226±134	0.79±0.47
All backgrounds	4185	3840±76	
hadronic W W+jets CR			
$t\bar{t}$	328	323±27	0.99±0.08
Single t	42	48±11	1.14±0.25
W+jets	2017	1597±134	0.79±0.07
WW,WZ,ZZ,Z+jets	308	333±111	1.08±0.36
Fakes	183	173±68	0.95±0.37
All backgrounds	2879	2475±72	

Table 8.6: Yields for B' selection in the single-lepton channel before and after fit of the background-only hypothesis to real data. The numbers for the major backgrounds are also shown separately. β is the ratio of the post- over pre-fit yields. The quoted error is the total uncertainty, calculated as explained in the text.

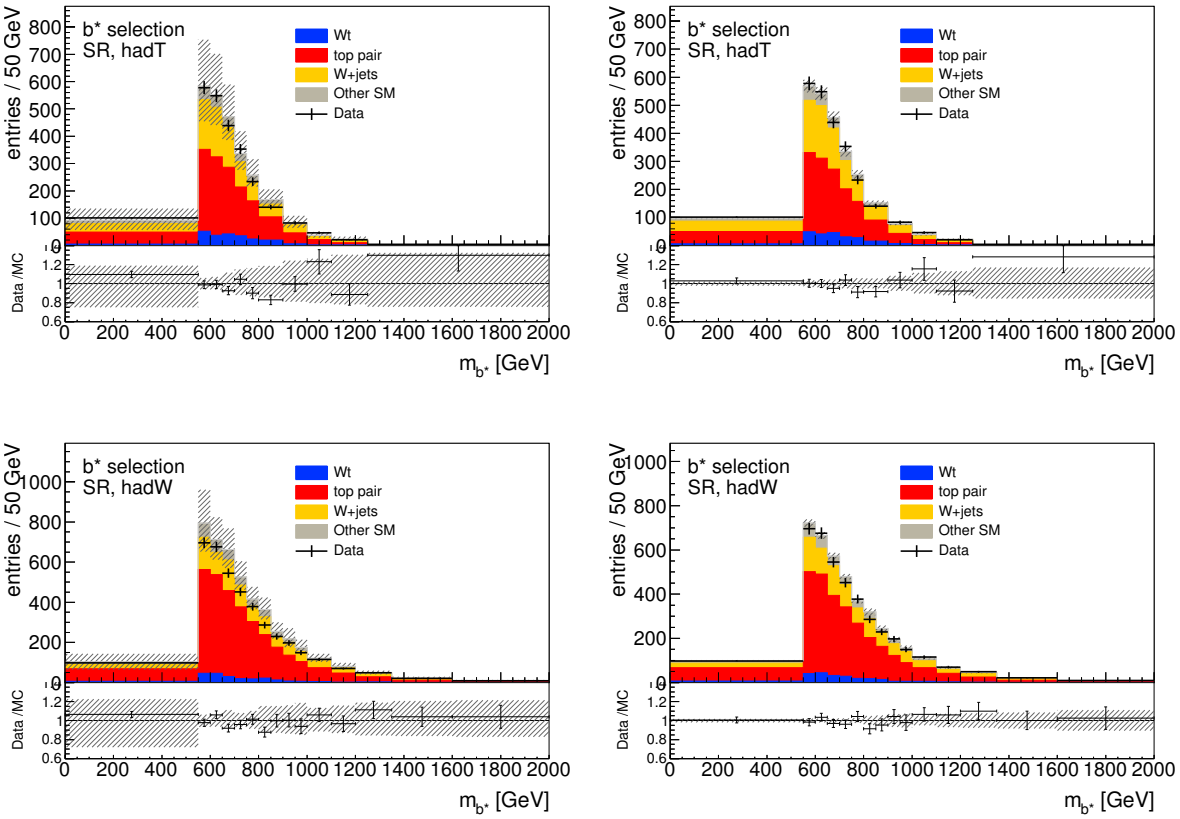


Figure 8.10: Pre-and post-fit plots for b^* in the signal regions of single-lepton channel. The plots on the left are pre-fit plots and the plots on the right are post-fit plots. The error bands include all uncertainties. Small backgrounds like the diboson, other single top-quark backgrounds, and fakes are merged in the figure, but treated separately in the fit.

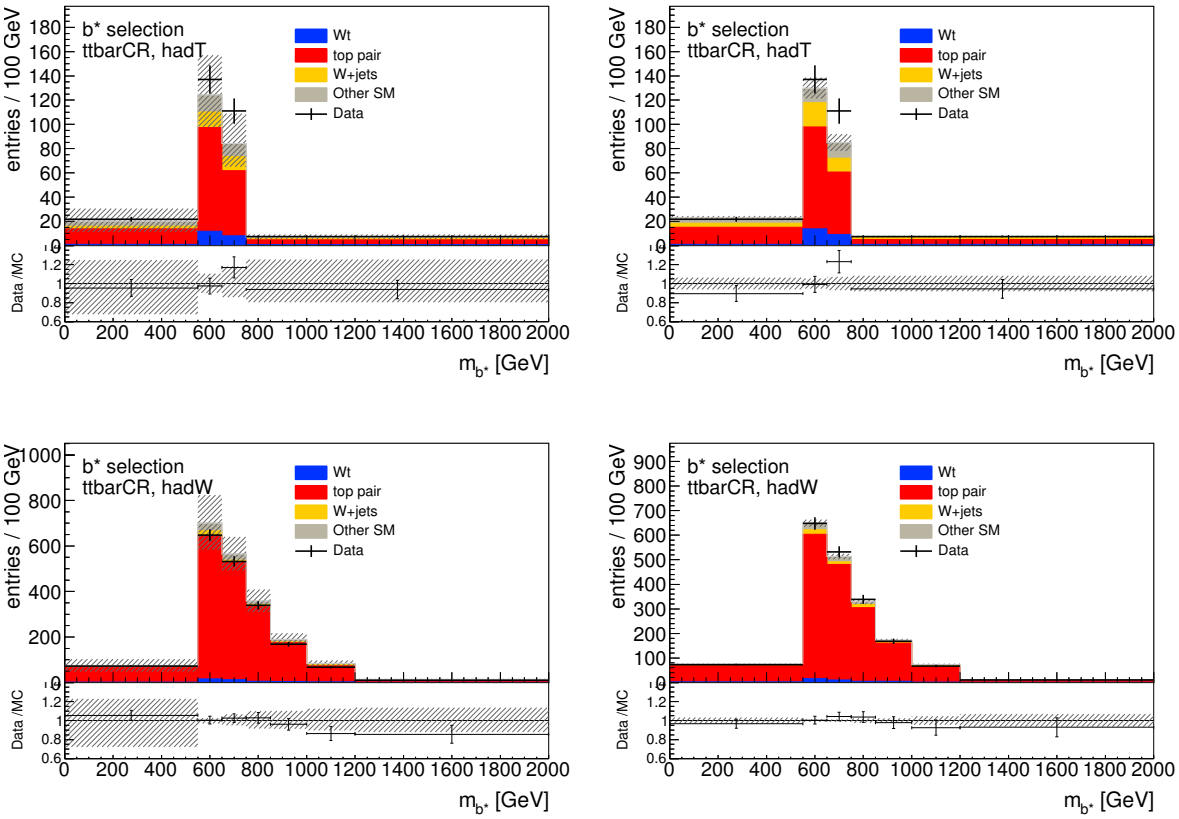


Figure 8.11: Pre-and post-fit plots for b^* in the $t\bar{t}$ control regions of single-lepton channel. The plots on the left are pre-fit plots and the plots on the right are post-fit plots. The error bands include all uncertainties. Small backgrounds like the diboson, other single top-quark backgrounds, and fakes are merged in the figure, but treated separately in the fit.

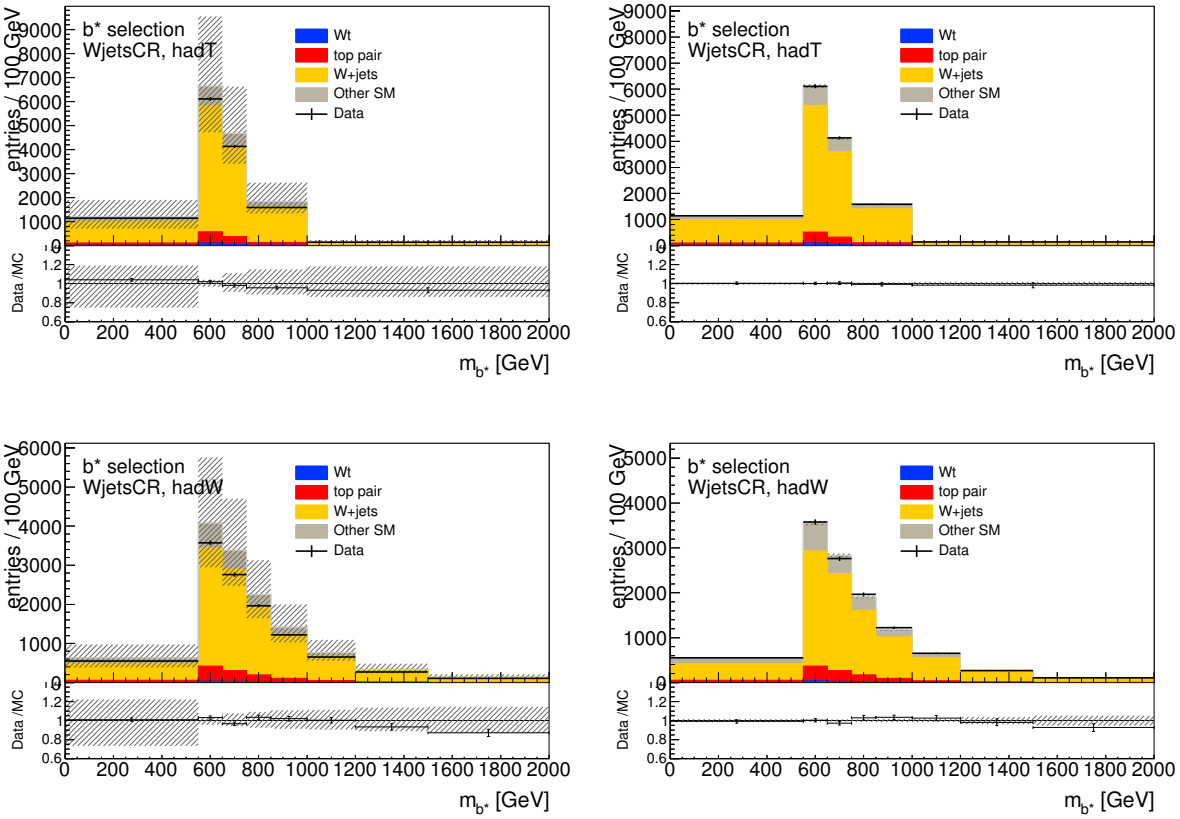


Figure 8.12: Pre- and post-fit plots for b^* in the W+jets control regions of single-lepton channel. The plots on the left are pre-fit plots and the plots on the right are post-fit plots. The error bands include all uncertainties. Small backgrounds like the diboson, other single top-quark backgrounds, and fakes are merged in the figure, but treated separately in the fit.

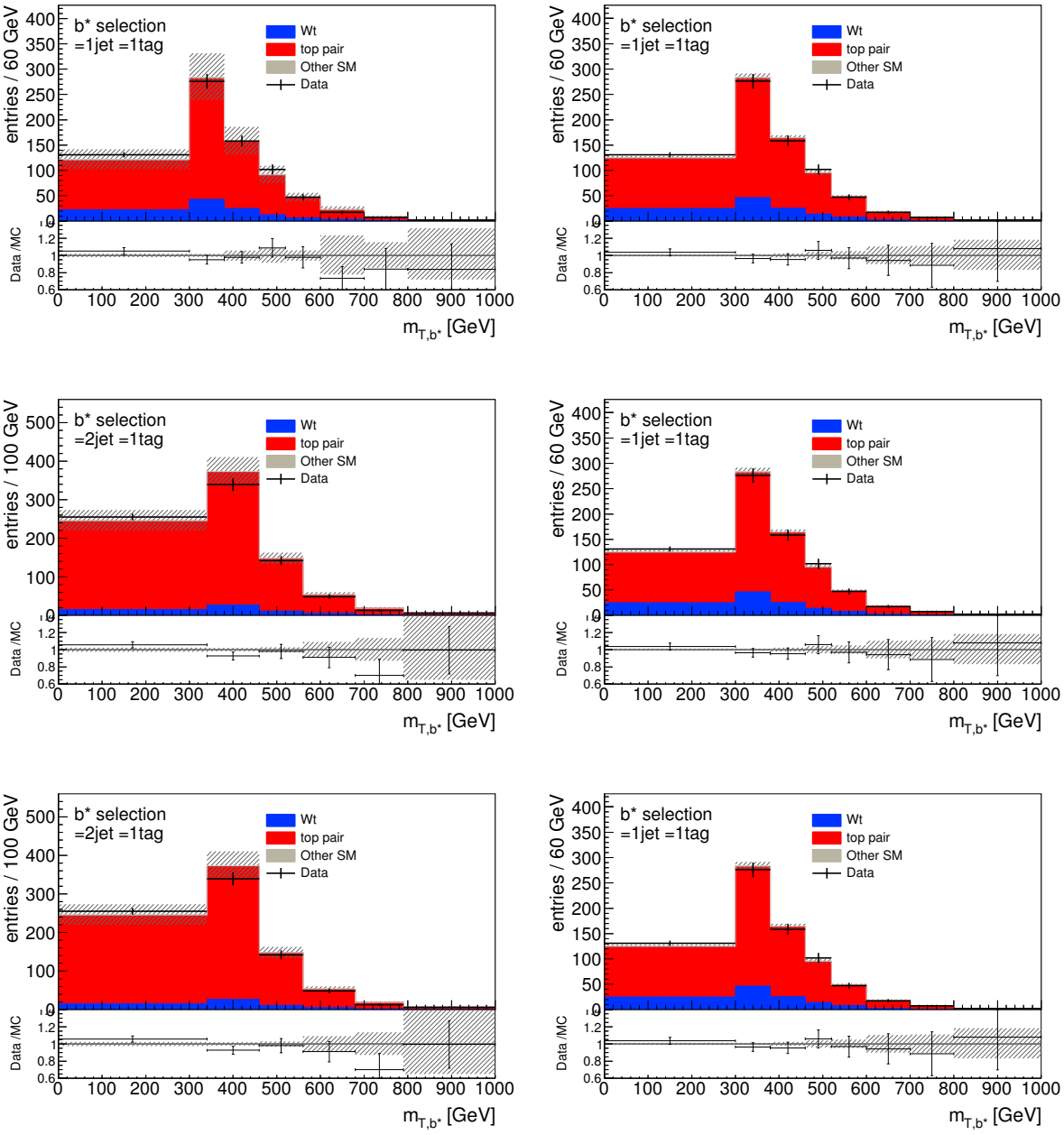


Figure 8.13: Pre-and post-fit plots for b^* in the dilepton channel. The plots on the left are pre-fit plots and the plots on the right are post-fit plots. The error bands include all uncertainties. Small backgrounds like the diboson, other single top-quark backgrounds, and fakes are merged in the figure, but treated separately in the fit.

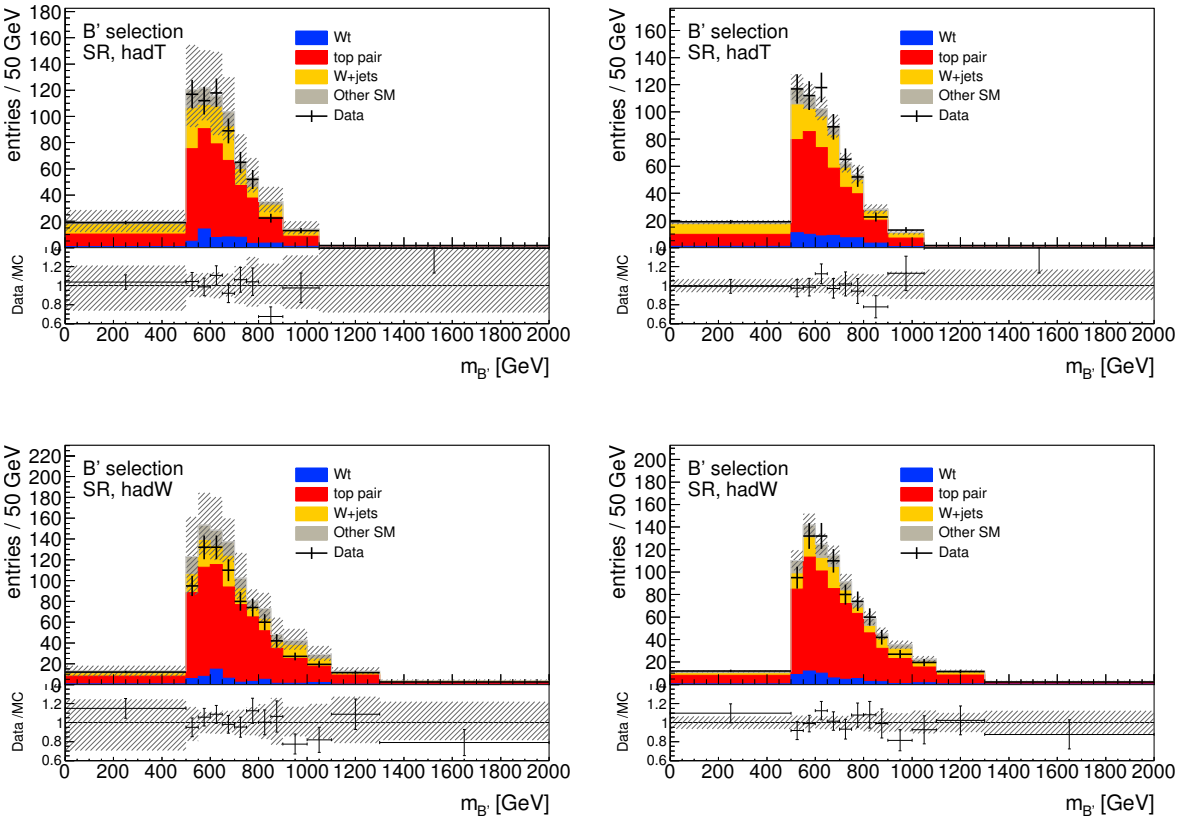


Figure 8.14: Pre-and post-fit plots for B' in the signal regions of single-lepton channel. The plots on the left are pre-fit plots and the plots on the right are post-fit plots. The error bands include all uncertainties. Small backgrounds like the diboson, other single top-quark backgrounds, and fakes are merged in the figure, but treated separately in the fit.

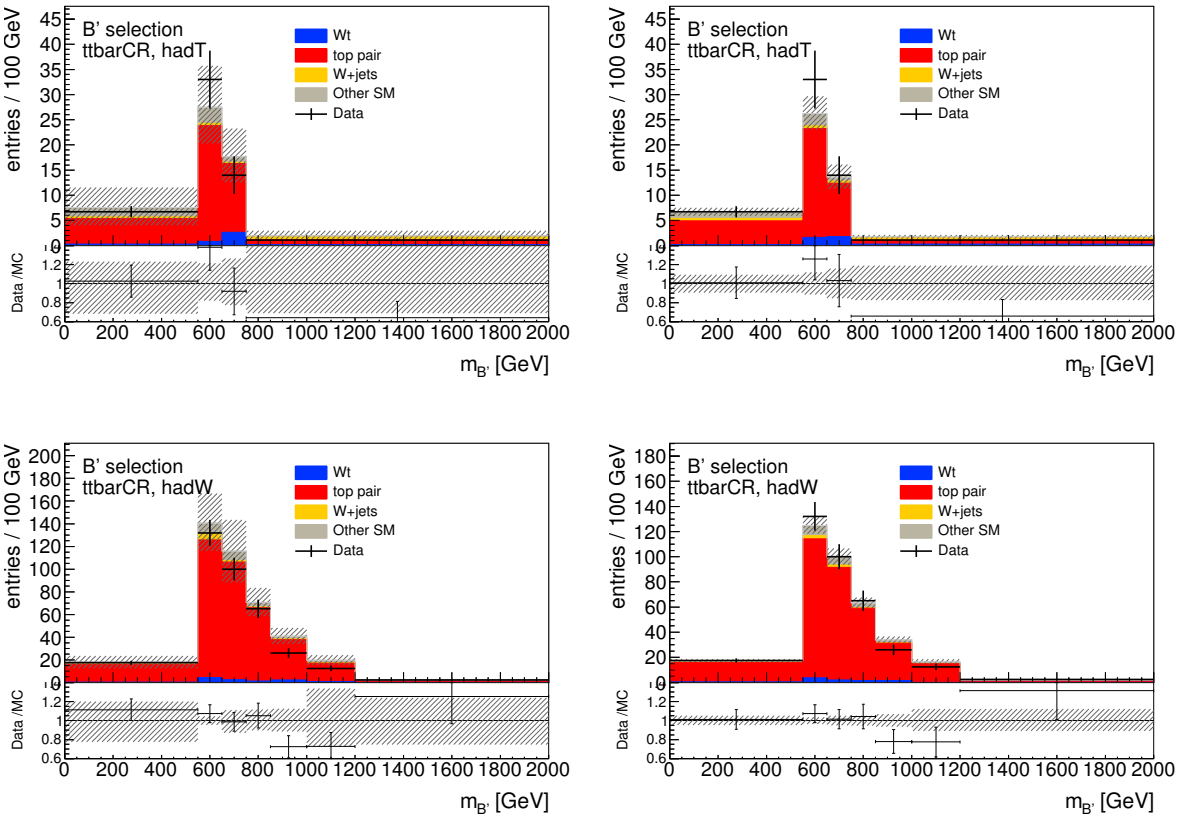


Figure 8.15: Pre-and post-fit plots for B' in the $t\bar{t}$ control regions of single-lepton channel. The plots on the left are pre-fit plots and the plots on the right are post-fit plots. The error bands include all uncertainties. Small backgrounds like the diboson, other single top-quark backgrounds, and fakes are merged in the figure, but treated separately in the fit.

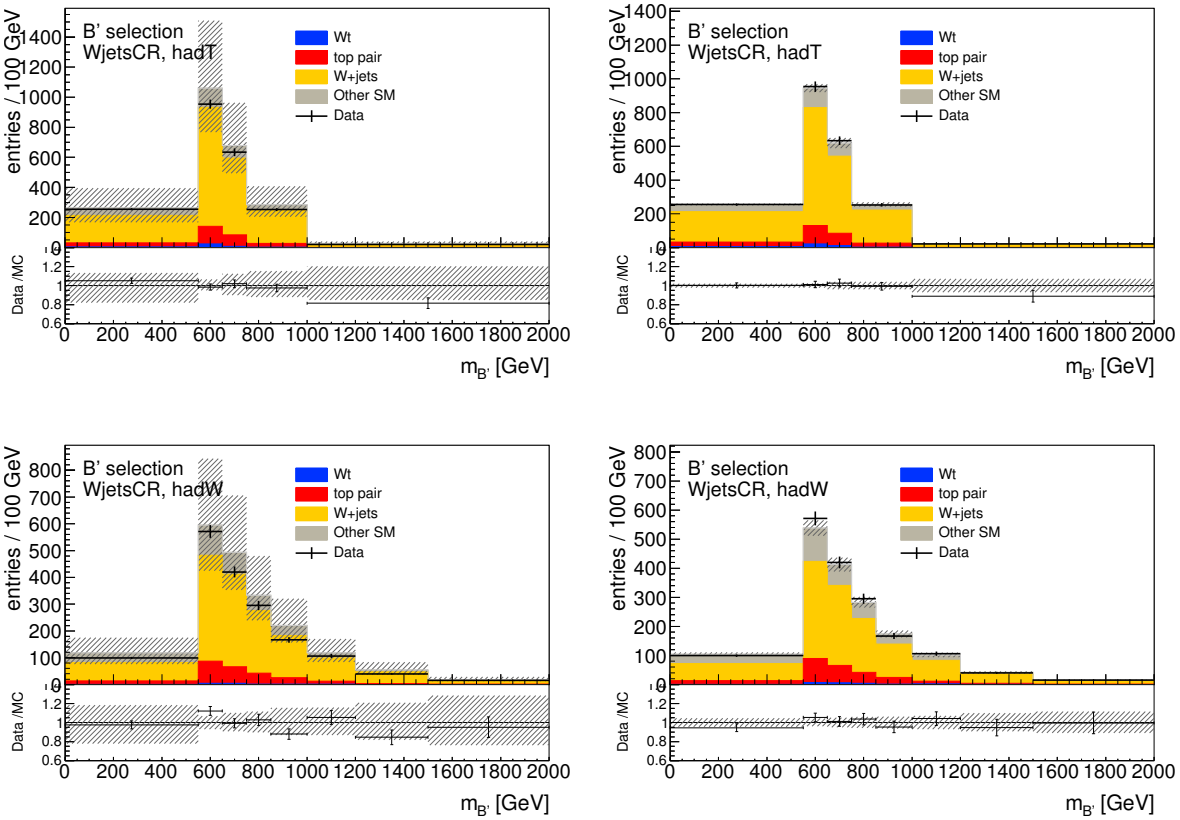


Figure 8.16: Pre-and post-fit plots for B' in the W+jets control regions of single-lepton channel. The plots on the left are pre-fit plots and the plots on the right are post-fit plots. The error bands include all uncertainties. Small backgrounds like the diboson, other single top-quark backgrounds, and fakes are merged in the figure, but treated separately in the fit.

8.4 Limits

The results obtained from hypothesis testing show that the background-only hypothesis cannot be rejected. So we set limits on the signal strength of b^*/B' for different mass hypotheses. For b^* we set limits by repeating the fitting procedure for m_{b^*} ranging from 600 to 1600 GeV. The mass points for which the limits were set are 600 GeV, 800 GeV, 1000 GeV, 1200 GeV, 1300 GeV, 1400 GeV, and 1600 GeV. For B' we used signals having a mass of 500 GeV, 600 GeV, 700 GeV, 800 GeV, 900 GeV, 1000 GeV, and 1200 GeV. The 95% C.L. limits on the production cross-section times branching ratio were set for the single-lepton channel, dilepton channel and the combined channels. Using these limits we extract a mass limit for the given particle. The mass limit for a hypothetical particle is defined as the mass for which the theory production cross-section equals the observed production cross-section and the observed cross-section is greater or lesser than the theoretical cross-section for decreasing or increasing masses respectively. The expected and observed 95% C.L. limit on the production cross-section times branching ratio as a function of the mass of b^*/B' is shown in Figures 8.17-8.19. The figures show the limits for the single-lepton, dilepton and combined channels. The green and yellow bands represent the $\pm 1, 2\sigma$ error band respectively. The observed limits are within $\pm 2\sigma$ error bands of the expected limit. For computing the expected mass limit for b^* in the single-lepton and the combined channel, we had to extrapolate our expected limits beyond 1600 GeV due to the unavailability of signal Monte Carlo simulations beyond 1600 GeV. We observe that the combined limits is more sensitive to the single-lepton limit than the dilepton one. The mass limits for b^* is shown in Table. 8.7. We could not set a mass limit on B' because of its extremely low theoretical cross-sections. We repeat the similar procedure with a frequentist method, described in Chapter 7, in the single-lepton and the dilepton channels. to compare with the results obtained by the Bayesian procedure. The mass limits for b^* is quoted in Table 8.8. The corresponding limits as a function of signal mass is shown in Figures 8.20 and 8.21. We see that the Bayesian and frequentist limits are comparable within an error range of 100-150 GeV. We can explain this difference by the difference in the treatment of the nuisance parameters by each of these methods. The Bayesian approach included profiling whereas the the frequentist approach did not. The exact values of the 95% C.L. limits on the production cross-section for different mass hypotheses of the signal are given in Appendix B.

	Expected [GeV]	Observed [GeV]
Single-lepton	1660	1520
Dilepton	1265	1220
Combined	1660	1500

Table 8.7: Expected and observed mass limits for b^* computed using Bayesian method

	Expected [GeV]	Observed [GeV]
Single-lepton	1480	1410
Dilepton	1340	1320

Table 8.8: Expected and observed mass limits for b^* computed using frequentist method

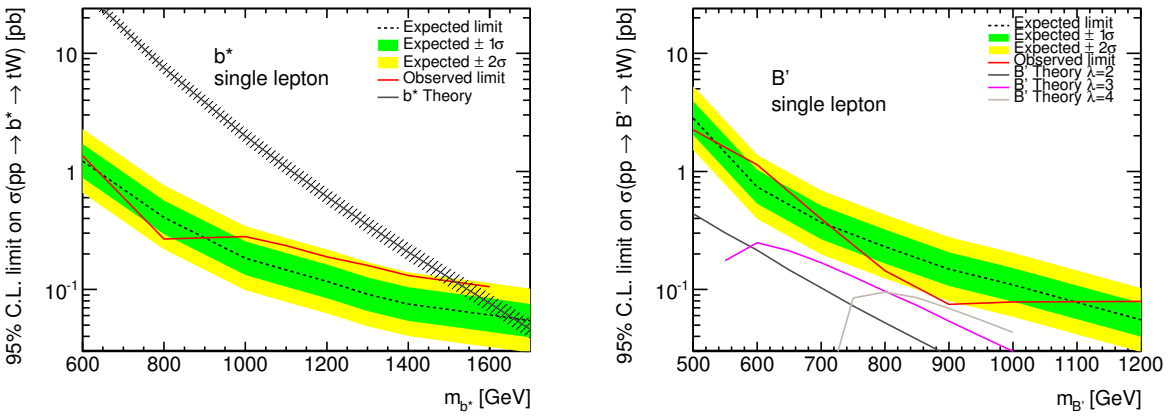


Figure 8.17: Observed and expected limits for b^* and B' in the single-lepton channel computed using Bayesian method.

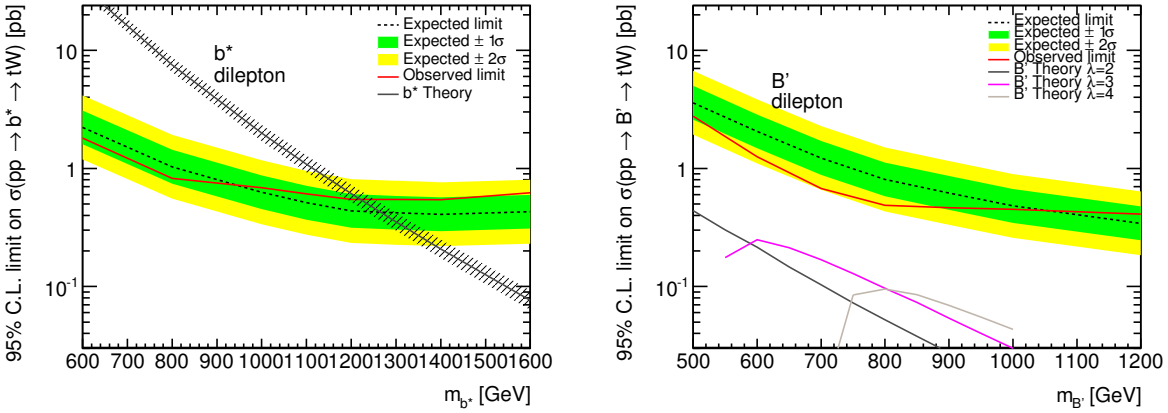


Figure 8.18: Observed and expected limits for b^* and B' in the dilepton channel computed using Bayesian method.

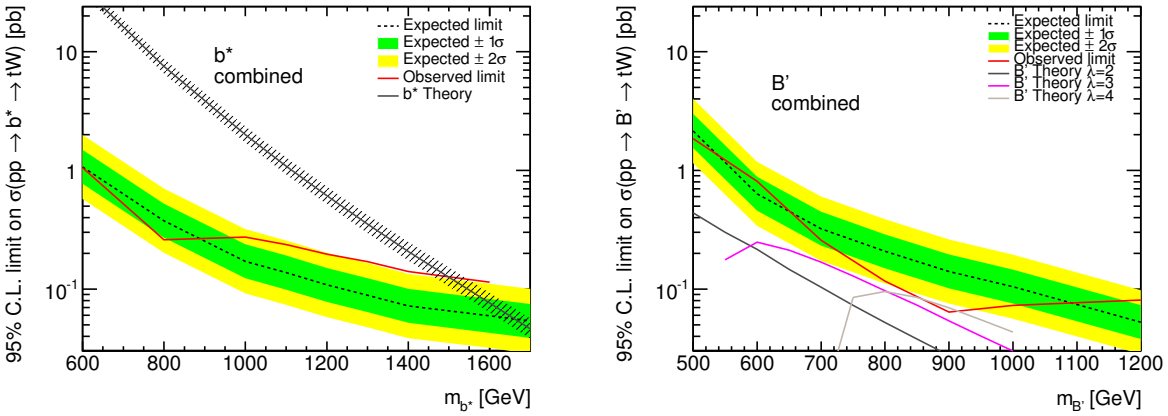


Figure 8.19: Observed and expected limits for b^* and B' in the combined channels computed using Bayesian method.

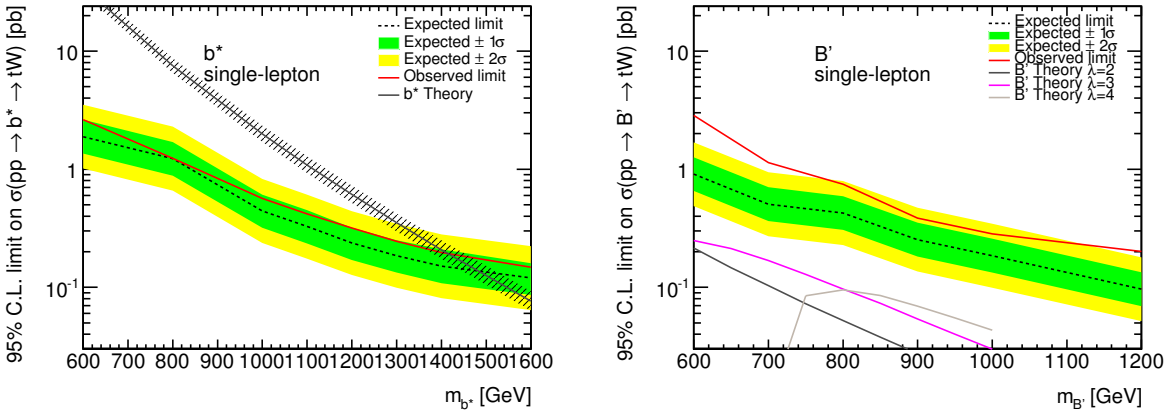


Figure 8.20: Observed and expected limits for b^* and B' in the single-lepton channel computed using frequentist method.

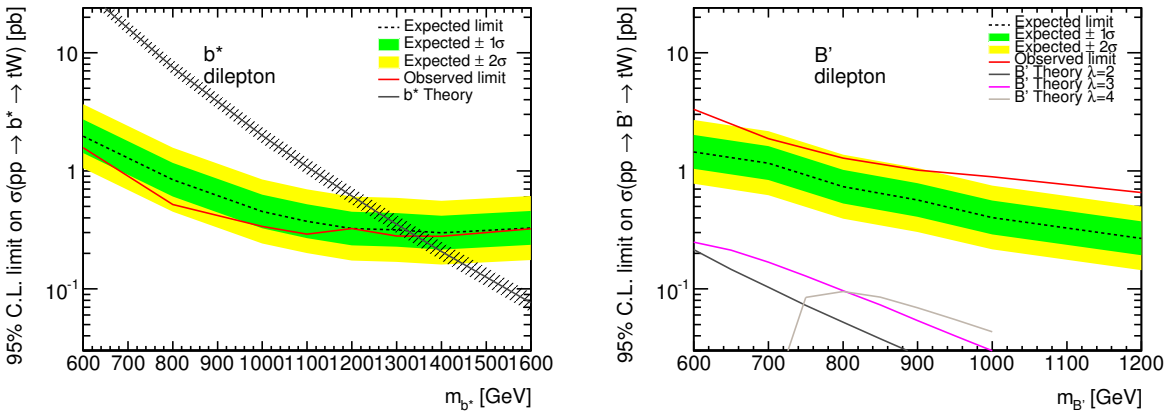


Figure 8.21: Observed and expected limits for b^* and B' in the dilepton channel computed using frequentist method.

CHAPTER 9

Conclusion

"*Nature is our kindest friend and best critic in experimental science if we only allow her intimations to fall unbiased on our minds.*" -- Michael Faraday

In this thesis, the statistical methods used in the search for two vector-like-quarks b^* and B' was presented. We considered the decay, $b^*/B' \rightarrow Wt$. Two decay modes of the Wt final state, namely the single-lepton decay mode and the dilepton decay mode was used for the search. This search was conducted using the $\sqrt{s}$ = 8 TeV data of proton-proton collisions taken at the ATLAS detector in 2012. The statistical analysis had two major steps. In the first step, we tried to look for any significant excess of data over background. After being unable to find any significant excess, we set 95% C.L. limits on the signal strength for varying mass hypotheses of the signal in the second step. Subsequently, we set an expected and observed mass limit on b^* and B'.

In the first part of the statistical analysis, we used a statistical package called `BumpHunter` to validate our background-only (Standard Model) hypothesis against our signal-plus-background hypothesis (Standard Model-plus-b^*/B'). In order to do this, we looked for any significant excess of data over background. We used control region fitted Monte Carlo simulations as a measure of our background. This package used a hypertest to scan the whole space of the parameter of interest. This ensured that we successfully avoided the "look-elsewhere" effect. In our analysis we scanned over the space of the discriminant. We used the mass of the signal in the single-lepton case and the transverse mass of the signal in the dilepton case as our discriminant. The results obtained by `BumpHunter` show us that we do not have any significant excess of data over background in the respective signal regions, thus confirming the background only (Standard Model) hypothesis.

The natural choice after the first part was to set limits on the signal strength. We use the cross-section times branching fraction of the process $pp \rightarrow b^*/B' \rightarrow Wt$ as our signal strength, μ. We used likelihood fits to do the limit setting. We used a Bayesian method to set limits on μ. We performed our fit in the single-lepton, dilepton, and the combined channel. The presence of four signal regions and five control regions made the analysis quite challenging. We used a tool called `Histfactory` to calculate the combined likelihood which included the information about all our uncertainties. We integrated over the whole parameter space of all the nuisance parameters to get the posterior probability distribution function (P.D.F.) of μ. This integration was done by the Bayesian Analysis Toolkit (`BAT`). `BAT` uses Markov Chain Monte Carlo to perform the integration. This P.D.F. was used to compute the limit on μ. We repeated this procedure for different mass hypotheses of the signal to extract a mass limit. A likelihood fit to the background only hypothesis was also performed to get an estimate of the constraints of the various nuisance parameters. This fit also illustrated how the fit constrains our uncertainties. Since we do not see any significant excess of data over background, we expect a reasonable agreement between data and Monte Carlo simulations of the background. The distributions obtained before and after the fitting procedure presented in the last chapter show this, thus illustrating the fact that our fit

works correctly. We also use a simple frequentist likelihood fit to compute CL_s limit on the signal strength in the single-lepton and dilepton channels. We use a tool called `Bill` fitter to implement this fit. We compare these limits with our Bayesian limits. For the combined limit we quote the Bayesian limit.

After computing the expected and observed limits, we were successful in excluding b^* below a certain mass, but due to the low theoretical cross-sections of B' we were unable to set mass limits for B'. We see that the Bayesian and frequentist limits were comparable within a range of 100-150 GeV. This can be explained by the different treatment of systematic uncertainties by the two approaches. The Bayesian approach included profiling whereas the frequentist approach did not. After setting limits we see that b^* can be excluded upto a mass of 1.5 TeV. Previously ATLAS had excluded b^* for masses up to 870 GeV [68]. So we see that this analysis was successful in improving the limits by a factor 1.7. The statistical methods developed during this thesis can be used in future searches for particles predicted by other Beyond the Standard Model theories during the second run of the Large Hadron Collider.

APPENDIX A

Nuisance parameters summary plots

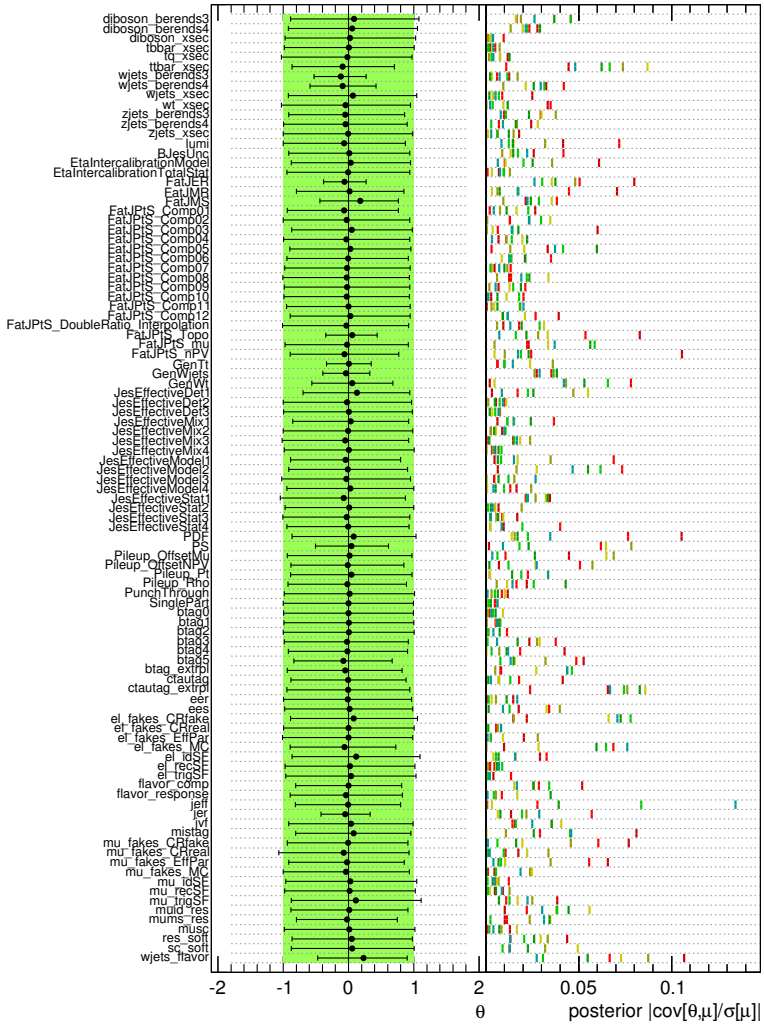


Figure A.1: Expected nuisance parameter distributions for b^* in the single-lepton channel. The left part visualises the median and $\pm 1\sigma$ quantiles of posterior p.d.f.s on each nuisance parameter before and after the fit of the background-only hypothesis. The right part visualizes the relative impact (as defined in the text) on each of the signal hypotheses. Signals with lower mass are denoted by reddish whereas signals with heavier masses by bluish colours.

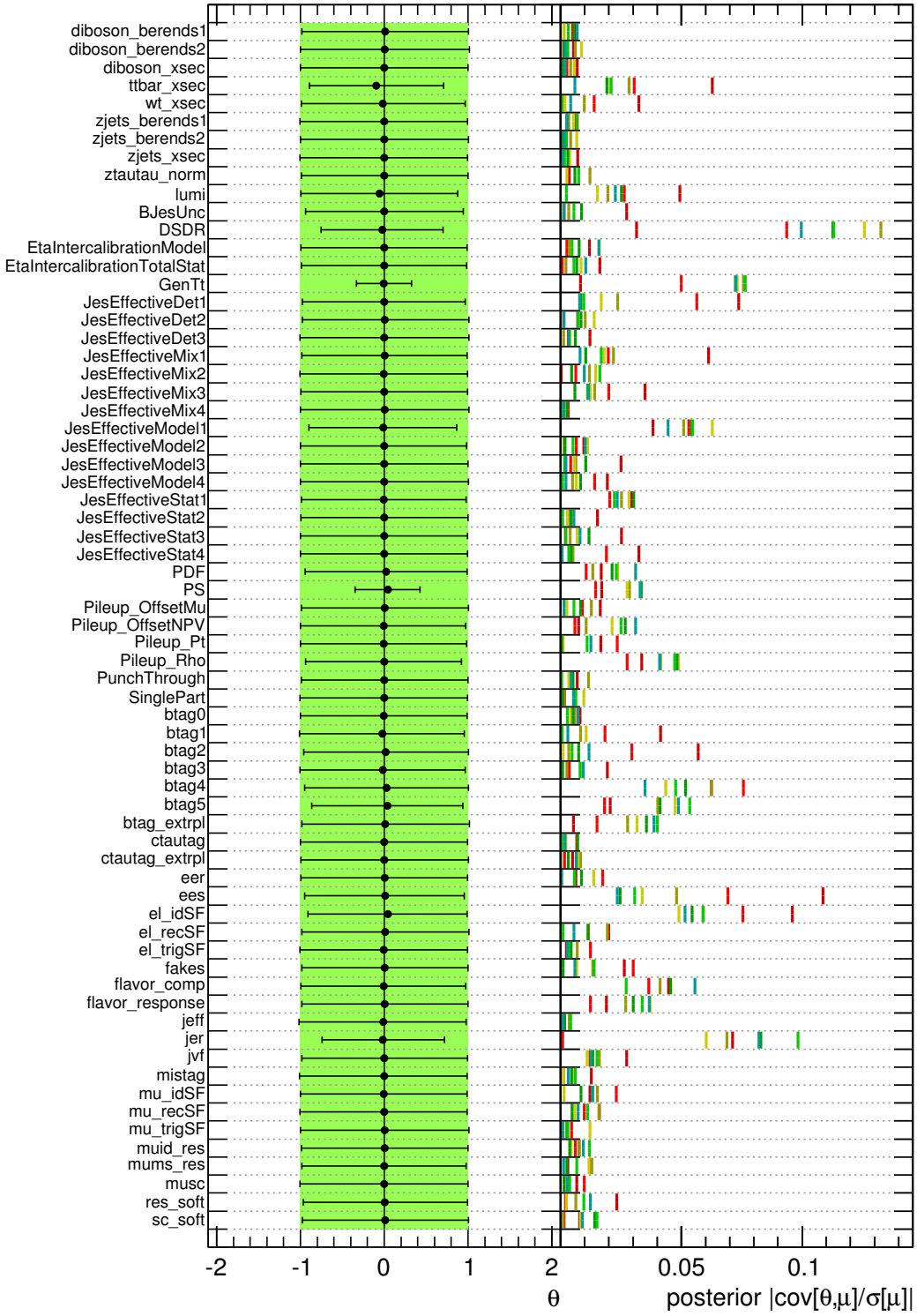


Figure A.2: Expected nuisance parameter distributions for b^* in the dilepton channel. The left part visualises the median and $\pm 1\sigma$ quantiles of the posterior p.d.f.s on each nuisance parameter before and after the fit of the background-only hypothesis. The right part visualizes the relative impact (as defined in the text) on each of the signal hypotheses. Signals with lower mass are denoted by reddish whereas signals with heavier masses by bluish colours.

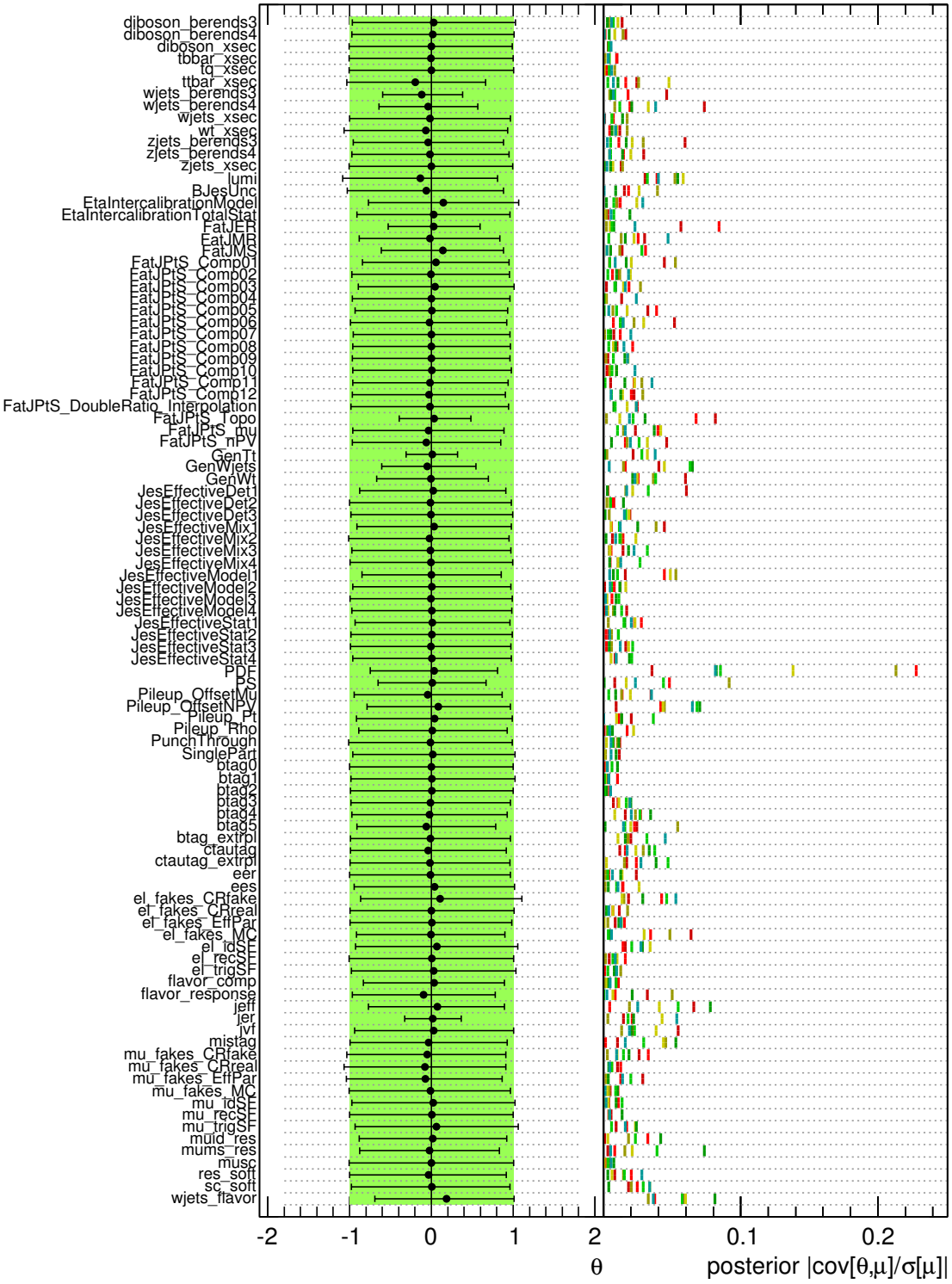


Figure A.3: Expected nuisance parameter distributions for B' in the single-lepton channel. The left part visualises the median and $\pm 1\sigma$ quantiles of the posterior p.d.f.s on each nuisance parameter before and after the fit of the background-only hypothesis. The right part visualizes the relative impact (as defined in the text) on each of the signal hypotheses. Signals with lower mass are denoted by reddish whereas signals with heavier masses by bluish colours.

Figure A.4: Expected nuisance parameter distributions for B' in the dilepton channel. The left part visualises the median and $\pm 1\sigma$ quantiles of the posterior p.d.f.s on each nuisance parameter before and after the fit of the background-only hypothesis. The right part visualizes the relative impact (as defined in the text) on each of the signal hypotheses. Signals with lower mass are denoted by reddish whereas signals with heavier masses by bluish colours.

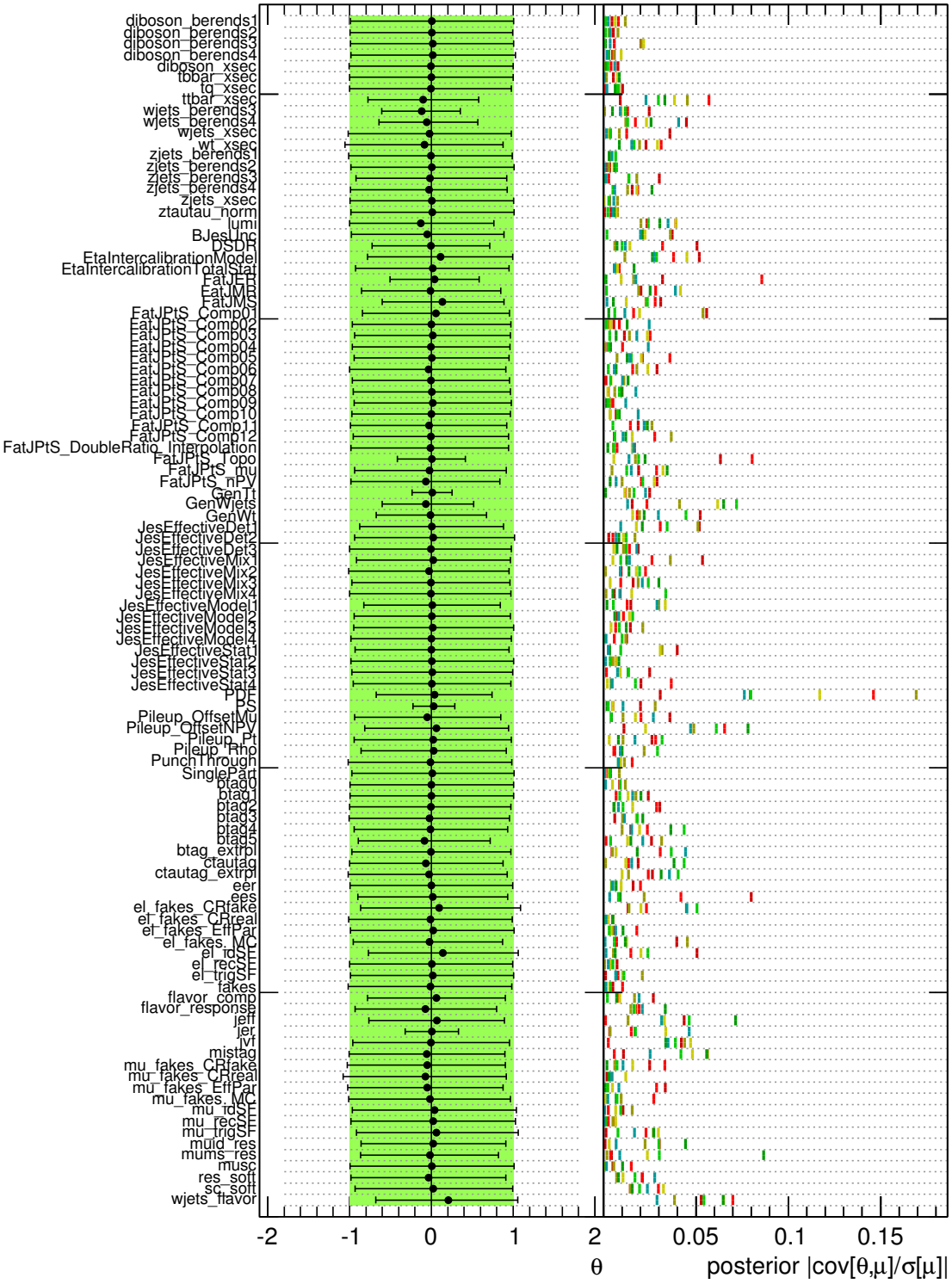


Figure A.5: Expected nuisance parameter distributions for B' in the combined fit of the single-lepton and dilepton channels. The left part visualises the median and $\pm 1\sigma$ quantiles of the posterior p.d.f.s on each nuisance parameter before and after the fit of the background-only hypothesis. The right part visualizes the relative impact (as defined in the text) on each of the signal hypotheses. Signals with lower mass are denoted by reddish whereas signals with heavier masses by bluish colours.

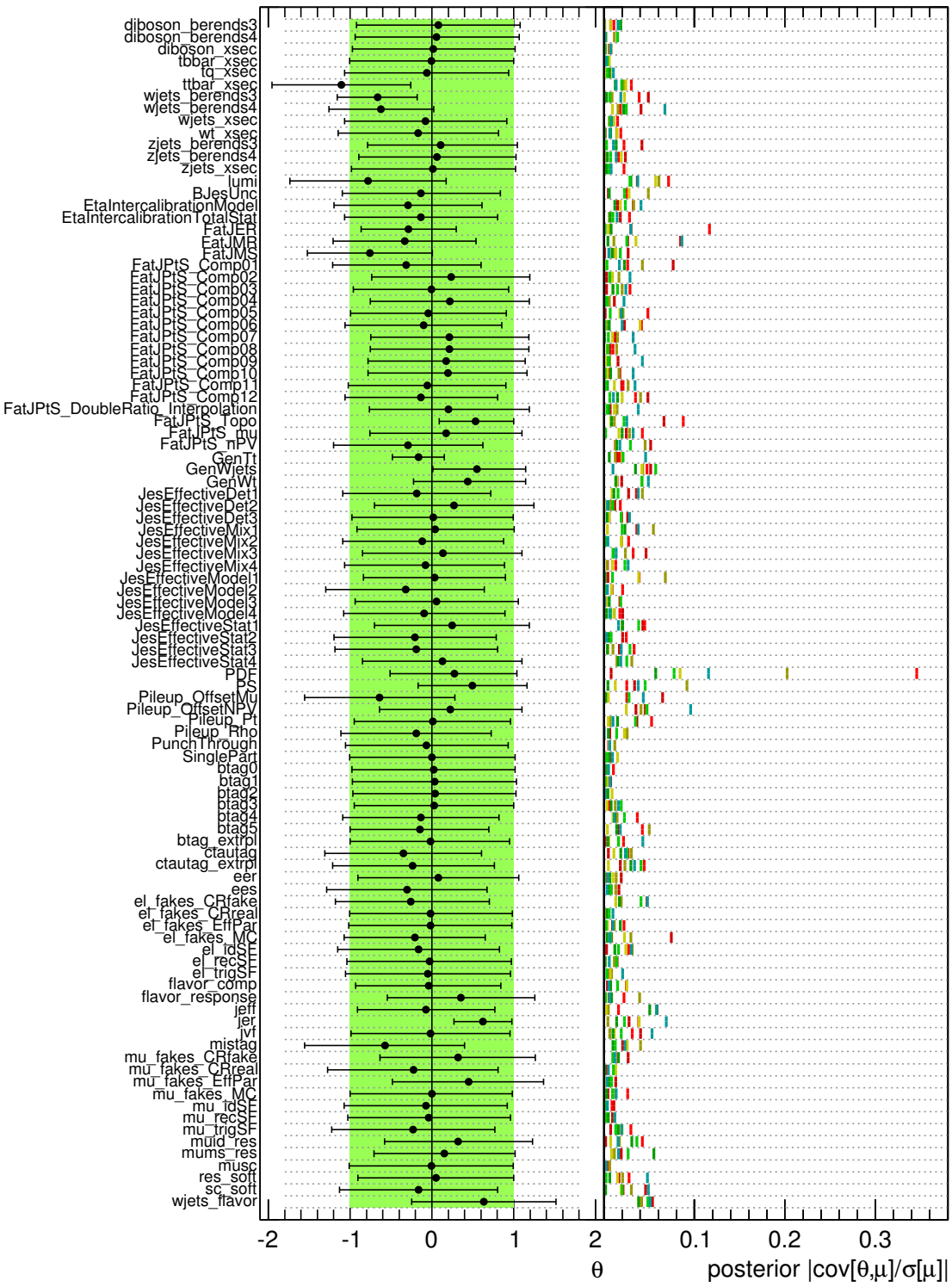


Figure A.6: Observed nuisance parameter distributions for B' in the single-lepton channel. The left part visualises the median and $\pm 1\sigma$ quantiles of the posterior p.d.f.s on each nuisance parameter before and after the fit of the background-only hypothesis. The right part visualizes the relative impact (as defined in the text) on each of the signal hypotheses. Signals with lower mass are denoted by reddish whereas signals with heavier masses by bluish colours.

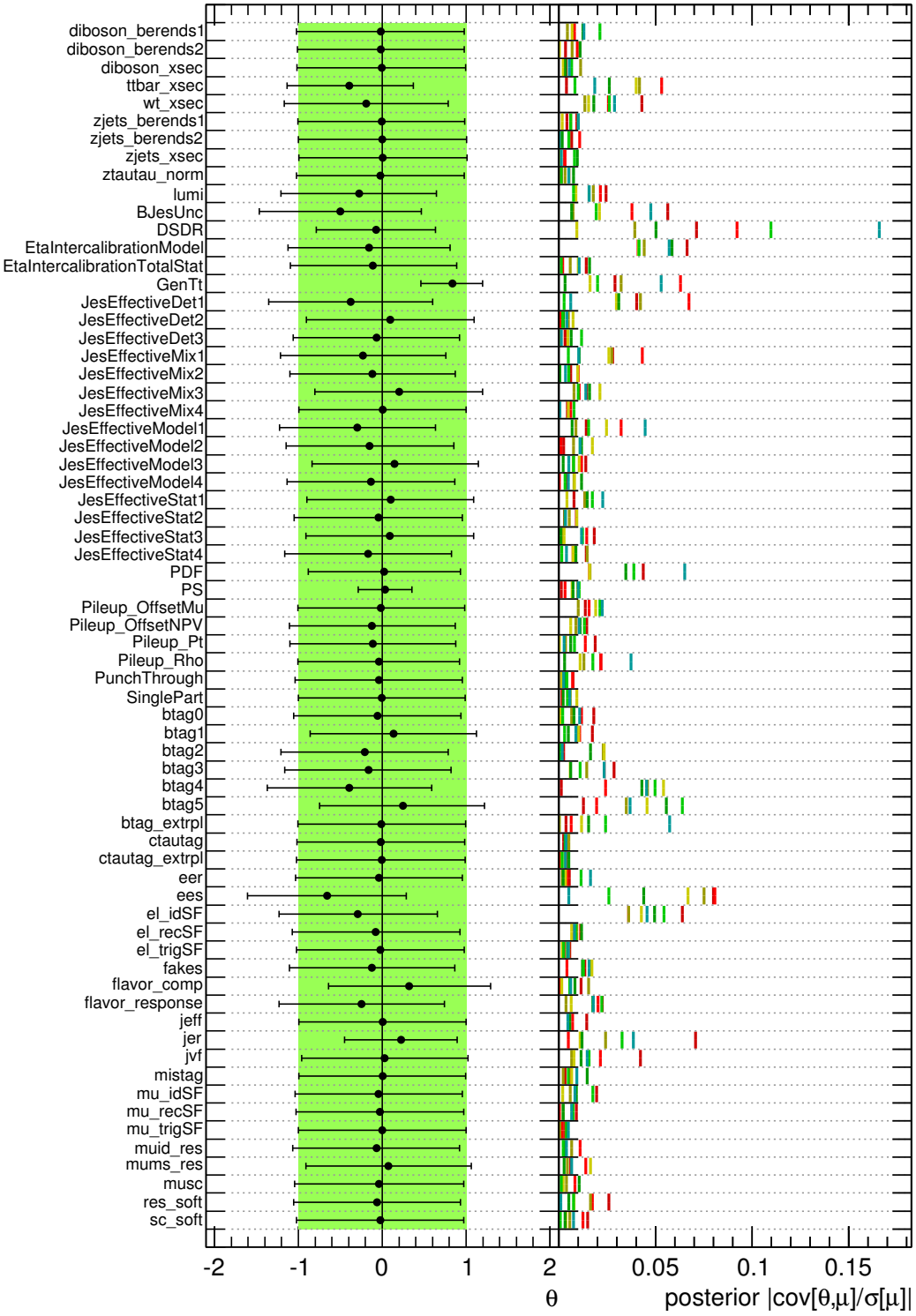


Figure A.7: Observed nuisance parameter distributions for B' in the dilepton channel. The left part visualises the median and $\pm 1\sigma$ quantiles of the posterior p.d.f.s on each nuisance parameter before and after the fit of the background-only hypothesis. The right part visualizes the relative impact (as defined in the text) on each of the signal hypotheses. Signals with lower mass are denoted by reddish whereas signals with heavier masses by bluish colours.

APPENDIX B

95% C.L. limits on production cross section of signal

mass [GeV]	600	800	1000	1200	1300	1400	1600
combined	1.07	0.38	0.17	0.11	0.09	0.07	0.06
single-lepton	1.23	0.41	0.18	0.12	0.09	0.08	0.06
dilepton	2.22	1.03	0.63	0.44	0.42	0.41	0.43

Table B.1: 95% C.L. expected upper limits for $\sigma(pp \to b^* \to tW)$ computed using Bayesian method in picobarns.

mass [GeV]	600	800	1000	1200	1300	1400	1600
combined	1.05	0.26	0.27	0.20	0.17	0.14	0.11
single-lepton	1.36	0.27	0.28	0.19	0.16	0.13	0.11
dilepton	1.81	0.82	0.69	0.55	0.55	0.54	0.62

Table B.2: 95% C.L. observed upper limits for $\sigma(pp \to b^* \to tW)$ computed using Bayesian method in picobarns.

mass [GeV]	500	600	700	800	900	1000	1200
combined	5.13	1.43	0.71	0.45	0.30	0.22	0.11
single-lepton	6.75	1.67	0.81	0.50	0.31	0.23	0.11
dilepton	8.60	4.64	2.68	1.73	1.31	1.01	0.71

Table B.3: 95% C.L. expected upper limits on $\sigma(pp \to B')$ computed using Bayesian method in picobarns.

mass [GeV]	500	600	700	800	900	1000	1200
combined	4.43	1.82	0.56	0.25	0.14	0.15	0.17
single-lepton	5.39	2.56	0.86	0.31	0.16	0.16	0.16
dilepton	6.60	2.84	1.48	1.04	0.98	0.94	0.85

Table B.4: 95% C.L. observed upper limits on $\sigma(pp \to B')$ computed using Bayesian method in picobarns.

mass [GeV]	600	800	1000	1200	1300	1400	1600
single-lepton	1.89	1.23	0.44	0.23	0.19	0.15	0.12
dilepton	1.96	0.84	0.45	0.32	0.32	0.30	0.33

Table B.5: 95% C.L. expected upper limits for $\sigma(pp \to b^* \to tW)$ computed using frequentist method in picobarns.

mass [GeV]	600	800	1000	1200	1300	1400	1600
single-lepton	2.63	1.23	0.57	0.32	0.25	0.20	0.15
dilepton	1.57	0.52	0.34	0.32	0.28	0.28	0.32

Table B.6: 95% C.L. observed upper limits for $\sigma(pp \to b^* \to tW)$ computed using frequentist method in picobarns.

mass [GeV]	600	700	800	900	1000	1200
single-lepton	2.05	1.10	0.91	0.53	0.39	0.20
dilepton	3.62	2.53	1.57	1.19	0.84	0.55

Table B.7: 95% C.L. expected upper limits on $\sigma(pp \to B')$ computed using frequentist method in picobarns.

mass [GeV]	600	700	800	900	1000	1200
single-lepton	2.85	1.14	0.75	0.39	0.28	0.20
dilepton	3.32	1.90	1.28	1.02	0.90	0.65

Table B.8: 95% C.L. observed upper limits on $\sigma(pp \to B')$ computed using frequentist method in picobarns.

APPENDIX B

Bibliography

[1] Planck Collaboration, P. Ade et al., *Planck 2015 results. XIII. Cosmological parameters* (2015), arXiv: 1502.01589 (cit. on p. 1).

[2] G. 't Hooft et al., "Naturalness, Chiral Symmetry, and Spontaneous Chiral Symmetry Breaking", *Recent Developments in Gauge Theories. Proceedings, Nato Advanced Study Institute, Cargese, France, August 26 - September 8, 1979*, ed. by G. 't Hooft, vol. 59, 1980 p. 135 (cit. on pp. 1, 9).

[3] M. Veltman, *The Infrared - Ultraviolet Connection*, Acta Phys.Polon. **B12** (1981) 437 (cit. on pp. 1, 9).

[4] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys.Lett. **B716** (2012) 1, arXiv: 1207.7214 [hep-ex] (cit. on pp. 1, 5, 8).

[5] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys.Lett. **B716** (2012) 30, arXiv: 1207.7235 [hep-ex] (cit. on pp. 1, 5, 8).

[6] R. P. Feynman, *Space-Time Approach to Quantum Electrodynamics*, Phys. Rev. **76** (6 Sept. 1949) 769 (cit. on p. 4).

[7] F. Halzen and A. D. Martin, *Quarks and Leptons*, Wiley and Sons, 1984 (cit. on p. 6).

[8] S. Glashow, *Partial Symmetries of Weak Interactions*, Nucl.Phys. **22** (1961) 579 (cit. on p. 6).

[9] S. Weinberg, *A Model of Leptons*, Phys.Rev.Lett. **19** (1967) 1264 (cit. on p. 6).

[10] A. Salam, *Weak and Electromagnetic Interactions*, Conf.Proc. **C680519** (1968) 367 (cit. on p. 6).

[11] J. Schwinger, *Quantum Electrodynamics. I. A Covariant Formulation*, Phys. Rev. **74** (10 Nov. 1948) 1439 (cit. on p. 6).

[12] R. P. Feynman, *Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction*, Phys. Rev. **80** (3 Nov. 1950) 440 (cit. on p. 6).

[13] J. Schwinger, *On Quantum-Electrodynamics and the Magnetic Moment of the Electron*, Phys. Rev. **73** (4 Feb. 1948) 416 (cit. on p. 6).

[14] M. Gell-Mann and F. Low, *Quantum electrodynamics at small distances*, Phys.Rev. **95** (1954) 1300 (cit. on p. 7).

[15] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum field theory*, Westview Press, 1995 (cit. on p. 7).

[16] D. Gross and F. Wilczek, *Asymptotically Free Gauge Theories. 1*, Phys.Rev. **D8** (1973) 3633 (cit. on p. 7).

[17] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys.Rev.Lett. **13** (1964) 508 (cit. on p. 8).

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

[19] Particle Data Group, *Review of Particle Physics*, Chin.Phys. **C38** (2014) 090001 (cit. on pp. 8, 21, 22, 24, 26).

[20] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49** (Feb. 1973) 652 (cit. on p. 8).

[21] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (12 June 1963) 531 (cit. on p. 8).

[22] S. P. Martin, *A Supersymmetry primer*, Adv.Ser.Direct.High Energy Phys. **21** (2010) 1, arXiv: hep-ph/9709356 [hep-ph] (cit. on p. 9).

[23] C. Burgess, *Introduction to Effective Field Theory*, Ann.Rev.Nucl.Part.Sci. **57** (2007) 329, arXiv: hep-th/0701053 [hep-th] (cit. on p. 10).

[24] A. K. Alok, A. Dighe and D. London, *Constraints on the four-generation quark mixing matrix from a fit to flavor-physics data*, Phys. Rev. D **83** (7 Apr. 2011) 073008 (cit. on p. 10).

[25] S. P. Martin, *Extra vector-like matter and the lightest Higgs scalar boson mass in low-energy supersymmetry*, Phys.Rev. **D81** (2010) 035004, arXiv: 0910.2732 [hep-ph] (cit. on p. 10).

[26] D. B. Kaplan, H. Georgi and S. Dimopoulos, *Composite Higgs Scalars*, Phys.Lett. **B136** (1984) 187 (cit. on p. 10).

[27] N. Vignaroli, *Discovering the composite Higgs through the decay of a heavy fermion*, JHEP **1207** (2012) 158, arXiv: 1204.0468 [hep-ph] (cit. on p. 10).

[28] M. Schmaltz and D. Tucker-Smith, *Little Higgs review*, Ann.Rev.Nucl.Part.Sci. **55** (2005) 229, arXiv: hep-ph/0502182 [hep-ph] (cit. on p. 10).

[29] J. Nutter et al., *Single Top Production as a Probe of B-prime Quarks*, Phys.Rev. **D86** (2012) 094006, arXiv: 1207.5179 [hep-ph] (cit. on p. 10).

[30] N. Vignaroli, Δ *F=1 constraints on composite Higgs models with LR parity*, Phys.Rev. **D86** (2012) 115011, arXiv: 1204.0478 [hep-ph] (cit. on p. 10).

[31] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001, ed. by L. Evans (cit. on p. 12).

[32] ATLAS Collaboration, *ATLAS Luminosity Results*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults (cit. on p. 12).

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

[34] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004 (cit. on p. 12).

[35] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005 (cit. on p. 12).

[36] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002 (cit. on p. 12).

[37] TOTEM Collaboration, *The TOTEM experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08007 (cit. on p. 12).

[38] J. Pinfold et al., *Technical Design Report of the MoEDAL Experiment*, tech. rep. CERN-LHCC-2009-006. MoEDAL-TDR-001, CERN, June 2009, URL: https://cds.cern.ch/record/1181486 (cit. on p. 12).

[39] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics* (2009), arXiv: 0901.0512 [hep-ex] (cit. on p. 13).

[40] ATLAS collaboration, *Readiness of the ATLAS liquid argon calorimeter for LHC collisions*, European Physical Journal C **70** (Dec. 2010) 723, arXiv: 0912.2642 [physics.ins-det] (cit. on p. 15).

[41] ATLAS Collaboration, *Readiness of the ATLAS Tile Calorimeter for LHC collisions*, European Physical Journal C **70** (Dec. 2010) 1193, arXiv: 1007.5423 [physics.ins-det] (cit. on p. 16).

[42] ATLAS Collaboration, *ATLAS EfficiencyMeasurements*, https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/EfficiencyMeasurements (cit. on p. 18).

[43] *Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector*, tech. rep. ATLAS-CONF-2011-063, CERN, Apr. 2011, URL: https://cds.cern.ch/record/1345743 (cit. on p. 18).

[44] G. P. Salam and G. Soyez, *A Practical Seedless Infrared-Safe Cone jet algorithm*, JHEP **0705** (2007) 086, arXiv: 0704.0292 [hep-ph] (cit. on p. 18).

[45] B. Acharya et al., *Object selection and calibration, background estimations and MC samples for top quark analyses using the full 2012 data set*, tech. rep. ATL-COM-PHYS-2013-1016, CERN, July 2013, URL: https://cds.cern.ch/record/1563201 (cit. on p. 18).

[46] *Performance of large-R jets and jet substructure reconstruction with the ATLAS detector*, tech. rep. ATLAS-CONF-2012-065, CERN, July 2012, URL: https://cds.cern.ch/record/1459530 (cit. on p. 18).

[47] *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*, tech. rep. ATLAS-CONF-2011-102, CERN, July 2011, URL: https://cds.cern.ch/record/1369219 (cit. on p. 18).

[48] G. D'Agostini, *Probability and measurement uncertainty in physics: A Bayesian primer* (1995), arXiv: hep-ph/9512295 [hep-ph] (cit. on p. 21).

[49] R. Barlow, "Fundamental Concepts", *Data Analysis in High Energy Physics*, Wiley, 2013 (cit. on p. 22).

[50] O. Behnke and L. Moneta, "Parameter Estimation", *Data Analysis in High Energy Physics*, ed. by O. Behnke et al., Wiley, 2013 (cit. on pp. 22–24).

[51] F. Beaujean et al., *BAT: The Bayesian analysis toolkit*, J.Phys.Conf.Ser. **331** (2011) 072040 (cit. on pp. 23, 35, 36).

[52] H. Jeffreys, *An Invariant Form for the Prior Probability in Estimation Problems*, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences **186**.1007 (1946) (cit. on p. 24).

[53] C. Hensel and K. Kröninger, "Statistical Methods Commonly Used in High Energy Physics", *Data Analysis in High Energy Physics*, ed. by O. Behnke et al., Wiley, 2013 (cit. on p. 24).

[54] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput.Phys.Commun. **77** (1993) 219 (cit. on pp. 25, 34).

[55] R. Wanke, "How to Deal with Systematic Uncertainties", *Data Analysis in High Energy Physics*, ed. by O. Behnke et al., Wiley, 2013 (cit. on p. 25).

[56] J. Conway, *Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra* (2011), arXiv: 1103.0354 [physics.data-an] (cit. on pp. 26, 27).

[57] G. Schott, "Hypothesis Testing", *Data Analysis in High Energy Physics*, ed. by O. Behnke et al., Wiley, 2013 (cit. on p. 27).

[58] G. Choudalakis, *On hypothesis testing, trials factor, hypertests and the BumpHunter* (2011), arXiv: 1101.0390 [physics.data-an] (cit. on pp. 29, 32).

[59] J. Neyman, *Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability*, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **236**.767 (1937) 333, ISSN: 0080-4614 (cit. on p. 29).

[60] A. L. Read, *Modified frequentist analysis of search results (the CL_s method)*, CERN-OPEN-2000-205 (2000), URL: http://cds.cern.ch/record/451614 (cit. on p. 30).

[61] K. Cranmer et al., *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, tech. rep. CERN-OPEN-2012-016, New York U., Jan. 2012 (cit. on p. 33).

[62] N. Metropolis et al., *Equation of State Calculations by Fast Computing Machines*, The Journal of Chemical Physics **21**.6 (1953) 1087 (cit. on p. 35).

[63] W. K. Hastings, *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*, Biometrika **57**.1 (1970) (cit. on p. 35).

[64] ATLAS Top Group, *Bill Tool*, https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/BillTool (cit. on p. 36).

[65] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, The European Physical Journal C-Particles and Fields **71**.2 (2011) 1 (cit. on p. 40).

[66] "Frequentist Limit Recommendation, Draft 1.1", *ATLAS Statistics Forum*, 2011 (cit. on p. 40).

[67] ATLAS Top Group, *'TopSystematicUncertainties for 8 TeV 2012 analyses*, https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TopSystematicUncertainties (cit. on pp. 40, 42).

[68] ATLAS Collaboration, *Search for single b^*-quark production with the ATLAS detector at $\sqrt{s} = 7$ TeV*, Phys.Lett. **B721** (2013) 171, arXiv: 1301.1583 [hep-ex] (cit. on p. 72).

List of Figures

List of Tables