Search for heavy neutrinos in a 3-lepton final-state

Applications using supervised machine learning

by

William Hirst

THESIS

for the degree of

MASTER OF SCIENCE

Faculty of Mathematics and Natural Sciences
University of Oslo

Autumn 2022

Search for heavy neutrinos in a 3-lepton final-state

Applications using supervised machine learning

William Hirst

Search for heavy neutrinos in a 3-lepton final-state

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

Abstract

This will be the abstract.

Acknowledgments

Halla

Contents

Introduction

The Standard Model (SM) is perhaps one of the most successful scientific theories ever created. It accurately explains the interactions of leptons and quarks as well as the force carrying particles which mediate said interactions. In 2012 the SM achieved one of its crowning achievements when we discovered the Higgs boson. Much of the accolade was rightfully given to the theoretical work on the SM, but another aspect of the discovery was equally important. Data analysis was and is a crucial part of any new discovery in physics. One of the most important and exiting tools is Machine Learning (ML).

Outline of the Thesis

Chapter 1

The Standard model of elementary particles and beyond.

The SM is the most successful scientific theory ever created. It accurately explains the interactions of leptons and quarks as well as the force carrying particles which mediate said interactions. The model is a result of over a century of work demanding the contributions of grat minds like Paul Dirac, Erwin Schrodinger and Richard Feynman. In 2012 the SM achieved one of its crowning achievements when we discovered the Higgs boson.

1.1 The bulding blocks

As early as ancienct greece, humans pondered the nature of the most elementary building blocks of the universe. They imagined a rope of a given length, with a pair of scissors of adjustible size. Then one could ask, how many times can you cut the rope in half? If the answer is less than infinite, what are you left with?
In 1897, Joseph John Thomson discovered the first elementary particle using the Cathode Ray Tube. This particle we later named the electron. Prior to the time of discovery, we belived atoms to be the smalles building blocks. After the discovery of the electron, the discovery of the proton and neutron quickly followed. It was not until more than 50 years after the discovery of the proton (by Ernest Rutherford) that we discovered that also protons and neutrons could be further disected to smaller particles. We call these particles quarks. The "final-piece"[1] of the puzzle came in 1956 when we discovered the, at that time thought of as massless neutrino. Together with the electron, the neutrino is defined as a lepton. Together with the, quarks and leptons are called fermions.
Upon the evolution of the quantum mechanics and a physics as a whole, we started to divert our focus from the what and over to the how. How can we explain all the complex interactions between this relativly simple particles? Through the creation of SM and countless experiment, we discovered that fources are nothing but interactions between particles through what we call, force mediating particles. The SM decribes all forces as a fields which are mediated through a particles, we call bosons.
The four forces responsible for all the forces in the universe are electro-magnetism (Quantum Electro Dynamics (QED)), the weak-force, the strong-force(Quantum Chromo Dynamics (QCD)) and gravity. The boson most familiar to most is the photon. The photon is responsible for the mediation of QED and is responisble for all electro-magnetic effect, such as the ones allowing us to sea objects using our eyes. Similarly the W and Z bosons are responsible for the weak-force whcih allows for radioactive decay. And the gluon is responsible for QCD whcih holds protons and neutrinos together. Gravity is the only force not described in the SM, but would (if one day included) have its own force carrying particle, graviton.
The final building block in the universe introduced and described by SM is the Higgs boson. The Higgs boson was proposed by Peter Higgs in 1964 and discovered at CERN in 2012. The Higgs boson, also called the God particle is responsible for giving particles mass in a process called spontaneous symmetry breaking (more on this in later sections). Together the fermions and the bosons make up all the particles in the SM as it now stands.

[1] Given the nature of this thesis, the existence of further pieces is implied.

Generation	*Flavour*	Mass [MeV]	Charge [Elementary charge]
1*st*	e	0.511	-1
1*st*	ν_e	< 0.001	0
2*nd*	μ	105.66	-1
2*nd*	ν_μ	< 0.17	0
3*rd*	τ	1776.8	-1
3*rd*	ν_τ	< 18.2	0

Table 1.1: A list of all leptons along whith their generation, flavour, mass and charge.

Generation	*Flavour*	Mass [MeV]	Charge [Elementary charge]
1*st*	u	2.2	$-2/3$
1*st*	d	4.7	$+1/3$
2*nd*	c	1280	$-2/3$
2*nd*	s	96	$+1/3$
3*rd*	t	173100	$-2/3$
3*rd*	d	4180	$+1/3$

Table 1.2: A list of all quarks along whith their generation, flavour, mass and charge.

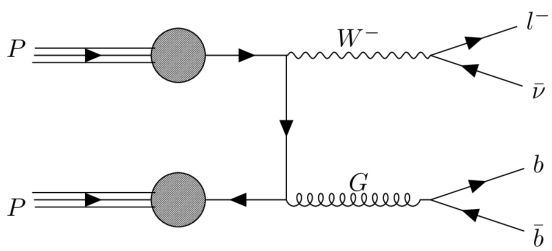
1.1.1 The leptons

The leptons are all elementary particle with half-integer spin, $\pm 1/2$. A lepton can either be charged or neutral. For reasons that are yet to be known, the leptons come in 3 genarations. Each genaration containing a pair of charged and neutral lepton. The first generation contains the electron, e^- and the electron-neutrino, ν_e. The second contains the muon, μ and the muon-neutrino, ν_μ. And the third generation contain the tau, τ^- and ν_τ. The generations are numbered by the mass of the charged lepton, where the first generation is the lightest. As is often the case in particle physics, the heavier a particle, the rarer. This is due to the heavier particles (higher generations) quickly decaying into lighter particles (lower generation), in a process we call particle decay. This explains why particle physicists neglect the τ when speaking about leptons, given that this is by far the heaviest and alos the rarest.
The charged leptons are all massive particles ranging from a fraction of 1eV to more than a thousand times that. The neutrinos were up until the turn of the millenia persumed to be massless. This was not only backed by experiments but also by the SM which seemed sheldom to be wrong. In 1998 it was discovered that neutrinos infact do have mass although being extremily light. Given the size of the masses we are yet to accuratly measure the mass of the neutrinos, but we have found them all to be less than 20 MeV. The fact that the neutrinos in fact do have mass is a problem which will be discussed further in later section. In table 1.2, a summary of all leptons are found, along with the respective mass and charge.

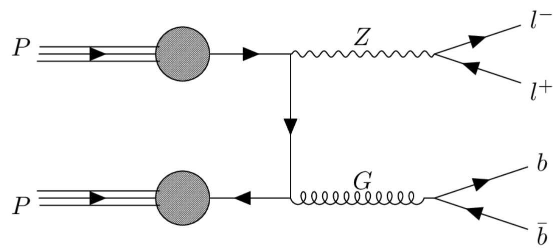
1.1.2 The quarks

> 'Three quarks for Muster Mark!
> Sure he hasn't got much of a bark.
> And sure any he has it's all beside the mark.' [1]

The poem above was written by James Joyce in 1939, and was the motivation for Gell-Mann coining the inner particles of hadrons, quarks. Quarks were introduced to explain some of the strong-force properties of hadrons. We categorize quarks as either up- or down quarks. All up-quarks have a negative charge equal to 2/3 that of the electron and all down quarks have a positive charge equal to 1/3 that of the electron. Similarly to leptons, all quarks have a spin equal to 1/2 and are divided in 3 genarations. Each generation of quarks are made of a pair of one up- and one down-quark. The first generation contains the up, u and the down, d quark, the second the charm, c and the strange, s quark and third the top, t and the bottom, b quark. Also similarly to leptons, the higher the generation and mass the rarer the quarks. Similarly to how difference in spin allows leptons to stay in an otherwise similar quantum state, the quarks have color. The colors of quarks are what connects them to the stron-force. The strong force is what allows quarks to change color and also explains a phenomenon known as



(a)



(b)

Figure 1.1: The Feynman diagram of both the W+jets 1.1a and the W+jets 1.1b.

color confinement. Briefly explained, color confinement results in quarks never existing in isolation but allways in pairs, mesons or in threes, baryons (like protons and neutrons). Given color confinement, quarks are never direclty observed in experiments, instead we detect the signature of quarks forming hadrons in a process called hadronisation. We call these signature jets.

1.2 Beyond the Standard Model

1.2.1 Why look beyond?

> 'There is nothing new to be discovered in physics now.
> All that remains is more and more precise measurement.' [2]

The quote above is rumored to have been spoken by William Thompson, better known as Lord Kelvin when adressing the British Association for the Advancement of Science in 1900. The statement followed a long period of advancements in the field of physics by the likes of Maxwell and Faraday. It would take less than half a decade before he would understand the magnitude of his misscalculation, when Einstein and Planck began the development of Quantum Mechanics. Just as Kelvin was wrong back then, would he be wrong today. For allthough SM explains a large range of phenomena, there are yet many mistories to explain in the universe and even problems rooted in SM. In this section I will explain some of the problems we hope to tackle in the future.

As mentioned in previous section, SM is yet to explain *gravity*. The hope has been to integrate gravity into SM through the discovery of a gravity-carrying particle, the gravtion. So far, no-such particle is found. *Dark matter* and *dark energy* are also not described by SM , even though the two make up more than 90% of the mass in the observable universe. Inflation is today the leading explaination to what happend in the early-stages (the first fraction of a second) of the universe, and explains a universe in which all space undergoes a rapid increase in rate of expantaion. None of the fields explained by SM are capable of causing any such expantion.

1.3 The background channels

The dominant SM backgrounds can be divided into two categories: (i) from leptonic τ decays and (ii) from fake leptons. In the first category, the dominant process is the pair production of WZ with W decaying leptonically and $Z \to \tau\tau$. The trilepton final states with no-OSSF pairs can arise from the subsequent leptonic decay of τ 's. We estimate this background process via Monte Carlo simulations.
The dominant processes of the second category are $\gamma^*/Z+$ jets and $t\bar{t}$, where two leptons come from $\gamma^*/Z \to \tau\tau$ or the prompt decay of t and $\bar{t}$, and a third lepton is faked from jets containing heavy-flavor mesons.

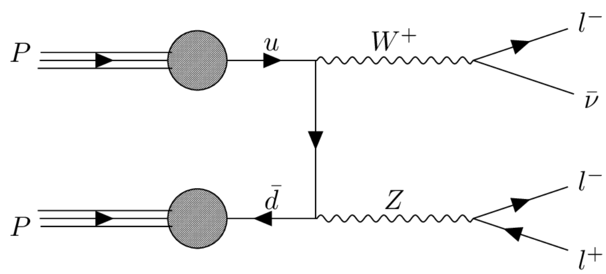


Figure 1.2: The Feynman diagram of the diboson WZ-channel.

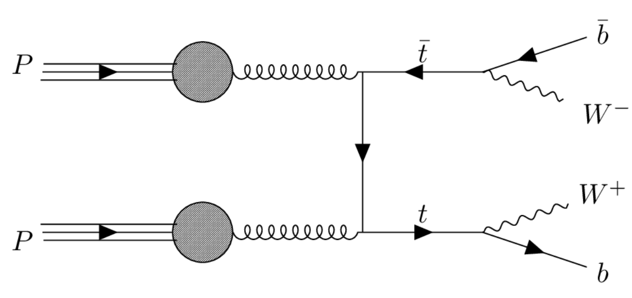


Figure 1.3: The Feynman diagram of the $t\bar{t}$-channel.

Chapter 2

Introduction to machine learning and statistical analysis

Machine learning is rapidly becomming an overwhelming presens in many different scientific fields. In areas ranging from cancer research to stock-trading, machine learning is being applied to probelms once thought as impossible to solve. Particle physics, like many other fields is no exeption. Jet flavor classification [3], separating jets from gluons [4] or using ML to create efficent Signal region (SR) are just some examples where ML is a vital tool. The traditional approach for ML in high-energy physics is through the use of supperivsed learning. Deep Neural Networks (DNN)

2.1 Machine Learning

2.1.1 Neural Networks in physics

2.1.2 Gradient Boosting and decision trees

In this rapport I will use the XGBoost-classifier which uses gradient-boosted trees. Gradient-boosting is a machine learning algorithm which uses a collective of "weak" classifiers in order to create one strong classifier. In the case of gradient-boosted trees the weak classifiers are a collective of shallow trees, which combine to form a classifiers that allows for deeper learning. As is the case for most gradient-boosting techniques, the collecting of weak classifiers is an iterative process.

We define an imperfect model $\mathcal{F}_m$, which is a collective of m number of weak classifiers, estimators. A prediction for the model on a given data-points, x_i is defined as $\mathcal{F}_m(x_i)$, and the observed value for the aforementioned data is defined as y_i. The goal of the iterative process is to minimize some cost-function $\mathcal{C}$ by introducing a new estimator h_m to compensate for any error, $\mathcal{C}(\mathcal{F}_m(x_i), y_i)$. In other words we define the new estimator as:

$$\tilde{\mathcal{C}}(\mathcal{F}_m(x_i), y_i) = h_m(x_i), \tag{2.1}$$

where we define $\tilde{\mathcal{C}}$ as some relation defined between the observed and predicted values such that when added to the initial prediction we minimize $\mathcal{C}$.

Using our new estimator h_m, we can now define a new model as

$$\mathcal{F}_{m+1}(x_i) = \mathcal{F}_m + h_m(x_i). \tag{2.2}$$

The XGBoost [**?**] framework used in this analysis enables a gradient-boosted algorithm, and was initially created for the Higgs ML challenge. Since the challenge, XGBoost has become a favorite for many in the ML community and has later won many other ML challenges. XGBoost often outperforms ordinary decision trees, but what is gains in results it looses in interpretability. A single tree can easily be analysed and dissected, but when the number of trees increases this becomes harder.

2.2 Statistical and multi-variabel analysis

The main idea of the search is to define a region where we will compare the Monte Carlo (MC)-background to the data and analyse any differences. Any analysis in the signal region hopes to either discovery or exclude certain Beyond Standard Model (BSM) physics. The standatd way of doing so is through statistics. In this rapport the statistical analysis will not be the focus and will therefore not be explained in heavy detail. Nonetheless a high-level understanding of the statistics is neccesarry when discussing the final results results. Therefore I will in this section discuss and define some basic statistical expressions and formulas.

Chapter 3

Implementation

3.1 Tools and Data

Every year technology for generating and measuring particle collisions is improved. As a consequence, the amount of data increases drastically. The ATLAS experiment is one of the largest particle detector experiments currently operating at the CERN laboratory near Geneva. ATLAS alone generates approximately 1 petabyte of raw data every second from proton proton collisions at the Large Hadron Collider (LHC). With amounts of data this large, datahandling and storing is a big challenge. Therefore, taking advantage of sofisticated numerical tools and data frameworks is pivotal if scientific development is to keep up with technological development.
In this section I will cover some of the tools and framewroks I have used to complete my analysis. Large amounts of details and explanations will not be covered. Instead this section will highlight which tools were used and some of the motivation for choosing them. Additionally I will cover some of the details regarding the data being used, both MC and real.

3.1.1 Monte Carlo Data

3.1.2 NTuples and RDataframe

3.1.3 Root

3.2 Features

The choice of which features to study and which to neglect are crucial in a search for new physics. This is particilarly true in the case of applying machine learning. The general motivation for including a given feature can be based on several factors. The first being its ability to provide a trend which we as researches can exploit when creating our regions. By this I mean that it is a variable were there is divirsity in distribution between the different channels. The second motivation is grounded in physics. Often we as physicsts tend to lean towards variables we know have some effect one the physics we are studying. For example the variable E_T^{miss}, can be directly used to either include or discard events were we do or do not expect final states with sufficent missing energy. The final motivation is grounded in the MC-simmulations ability to represent the variable. If there seems to be a clear deviation between the real and MC-data which does not stem from any new physics, we tend to discard them from the analysis.

3.2.1 Cuts and triggers

To allow for deep learning and a thorough analysis one must try and keep as much of the data as possible. At the same time, including large amounts of irrelevant data can be both redundant and destruvtive. Therefore simple cuts are necessary. The cuts applied in the analysis were grouped in two definitions, baseline and signal. The baseline requriemtns are written in table 3.1 and the signal requirments are written in table 3.2. Both sets of requirments were taken from tha ATLAS article from 2022 [5]. Given the definitions we demand that each event contains exactly three signal and three baseline leptons, thereby removing any event with more or less.

Requirement	Baseline electrons	Baseline muons
Identification		Loose
Overlap Removal	lepPassOR	lepPassOR
$\eta -$ cut	$\lvert\eta\rvert < 2.47$	$\lvert\eta\rvert < 2.7$
$\lvert z_0 \sin(\theta)\rvert$ cut	$\lvert z_0 \sin(\theta)\rvert < 0.5$ mm	$\lvert z_0 \sin(\theta)\rvert < 0.5$ mm

Table 3.1: Requirments for baseline electrons and muons.

Requirement	Signal electrons	Signal muons
Baseline	yes	yes
Identification	Tight	
Isolation	LooseVarRad	LooseVarRad
$\lvert d_0\rvert / \sigma_{d_0}$ cut	$\lvert d_0\rvert / \sigma_{d_0} < 5.0$	$\lvert d_0\rvert / \sigma_{d_0} < 3.0$

Table 3.2: Requirments for signal electrons and muons.

Leptons are identified in the detector by using a likelihood-based method combining information from different parts of the detector. The criteria of Loose or Tight identification are simply different thresholds in the discriminant, where Loose is defined as a lower threshold than Tight [6]. The overlap removal is used to solve any cases where the same lepton has been reconstructed as both a muon and an electron. The boolean of *lepPassOr* simpply applies a set of requirments to avoid any double counting. The cut for the longitudal track parameters, z_0 is applied to insure that the leptons originate for the primary vertex.
As for the requirements for the signal leptons, we require all baseline requirments are passed with the addition of a few more. We require Loose isolation for both electrons and muons. This means requiring criteria for a cone around the lepton and is used to supress QCD-background events. Similarly to the z_0-cut, the transverse track parameter is also used to ensure origin from primary vertex.

In addition to the simple cuts, we must insure a good comparison between MC- and real data. Often one finds large deviation betweem the two in the case of either very large or very small P_t. The latter case can often be caused by poor reconstruction or missidentification. These are issues we aim to solve by checking different triggers. Given our data set is composed of different data sets spread over many years, different triggers are used.

3.2.2 Lepton variables selection

Now we will have a look at what variables from the leptons that were included in the analysis. All low level information on the momentum of the leptons were added into the dataset: i.e the transverse momentum P_t, the pseudo rapidity η and the azimuthal angle ϕ. All momentum features were represented individually for each lepton. For example P_t was added as three columns, $P_t(l_1)$, $P_t(l_2)$ and $P_t(l_3)$, where the ordering of the leptons were based on the momentum from highest (l_1) to lowest (l_3). Similarly I added information regarding the charge ($\pm$) and flavor (electron, muon) of each lepton. Based on the momentum variables the transverse mass m_t of each lepoton was calculated and included along with the energy E_t^{miss} and azimuthal angle ϕ^{miss} of the missing transverse momentum.
The variables described in the section above are often considered as low-level features. These are very useful in many (if not all) searches and contribute little to no bias to your analysis. But, in the case of final-state spesific searches such as mine, one can allow one self of adding physics motivated higher-level physics. The higher level features calculated in this thesis were inspired by [5] (ATLAS 2022).
Firstly I added different mass variables, namely m_{lll} and $m_{ll}(OSSF)$. The first being the trilepton invariant mass

2015	2016	2017 + 2018
HLT_2e12_lhloose_L12EM10VH	HLT_2e17_lhvloose_nod0	HLT_2e17_lhvloose_nod0_L12EM15VHI
HLT_e17_lhloose_mu14	HLT_e17_lhloose_nod0_mu14	HLT_e17_lhloose_nod0_mu14
HLT_mu18_mu8noL1	HLT_mu22_mu8noL1	HLT_mu22_mu8noL1
		HLT_2e24_lhvloose_nod0

Table 3.3: Trigger requirments for events produced in their respective years.

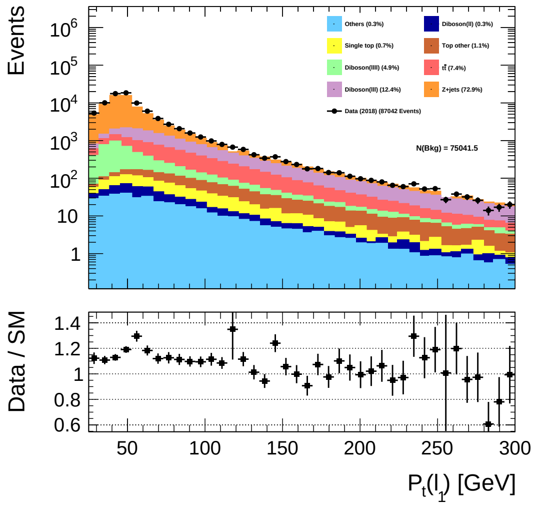
and the latter being the dilepton invariant mass of the pair with Opposite Sign Same Flavour (OSSF). In the case of more than one possible OSSF-pair, the pair with the highest invariant mass was chosen. Secondly I added variables composed of the sum of different set of momentum. These variables are the sum of all three leptons $H_t(lll)$, of the pair with Same Sign (SS) $H_t(SS)$ and the sum of the momentum for all three leptons added with the missing transverse energy $H_t(lll) + E_t^{miss}$. Finally I added the significance of the E_t^{miss}, $S(E_t^{miss})$.

3.2.3 Jet variables selection

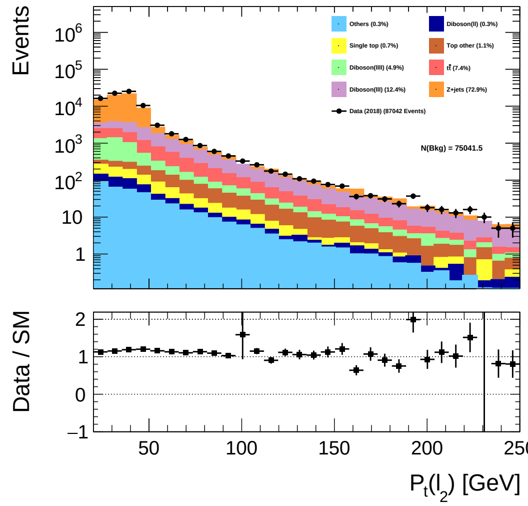
Now we can have a look at the jet-features. Given the final-state of interest should be independant of jets, there are not many features added with jet information. But, given the risk of missidentification and errors in reconstruction, some features were added. The first features were the number of jets, both all signal jets and number of b-jets. The latter information was divided in two columns based on the efficency of a multivarant analysis used to sepereata jet-flavors. The effeciencies used are 77% and %85. The last information added for the jets were the mass of the leading pair (based on p_t) di-jet mass.

3.2.4 Validation

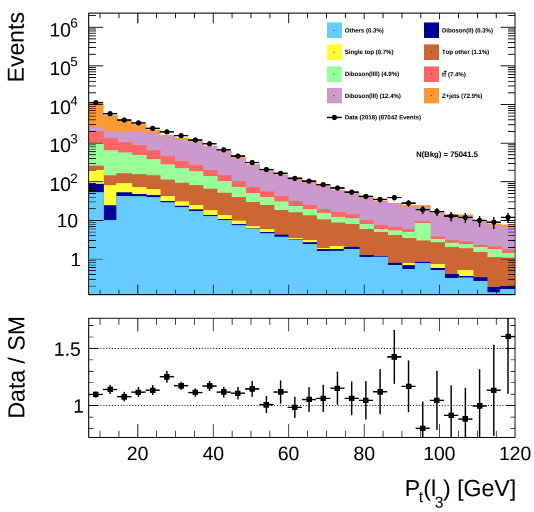
As mentioned in previous sections, the comparison between ML- and real data is a crucial part of the analysis, and we must therefore insure an adequate reconstruction of the real data. This is not only true for the low-level features taken directly from the ML simulation, but also for the the higher-level features. Therefore we will in this section compare both sets of data for all features included in the anlysis.



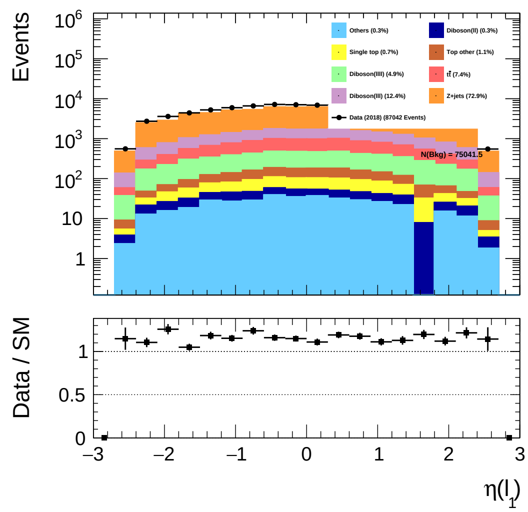
(a)



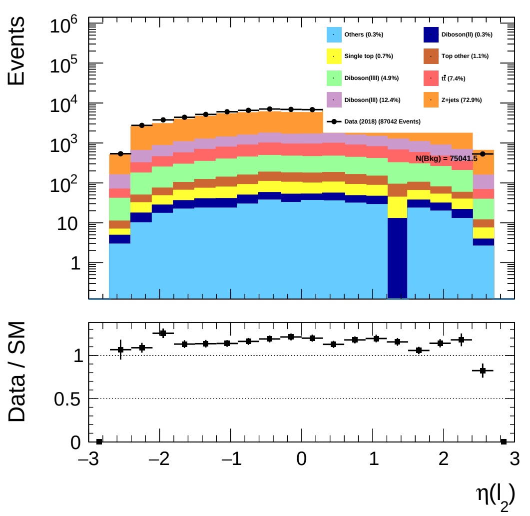
(b)



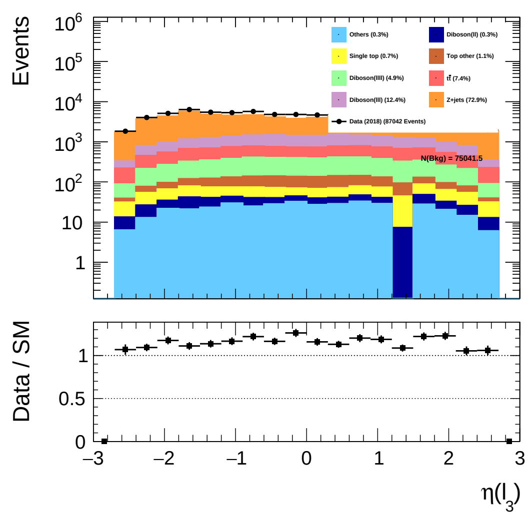
(c)



(d)

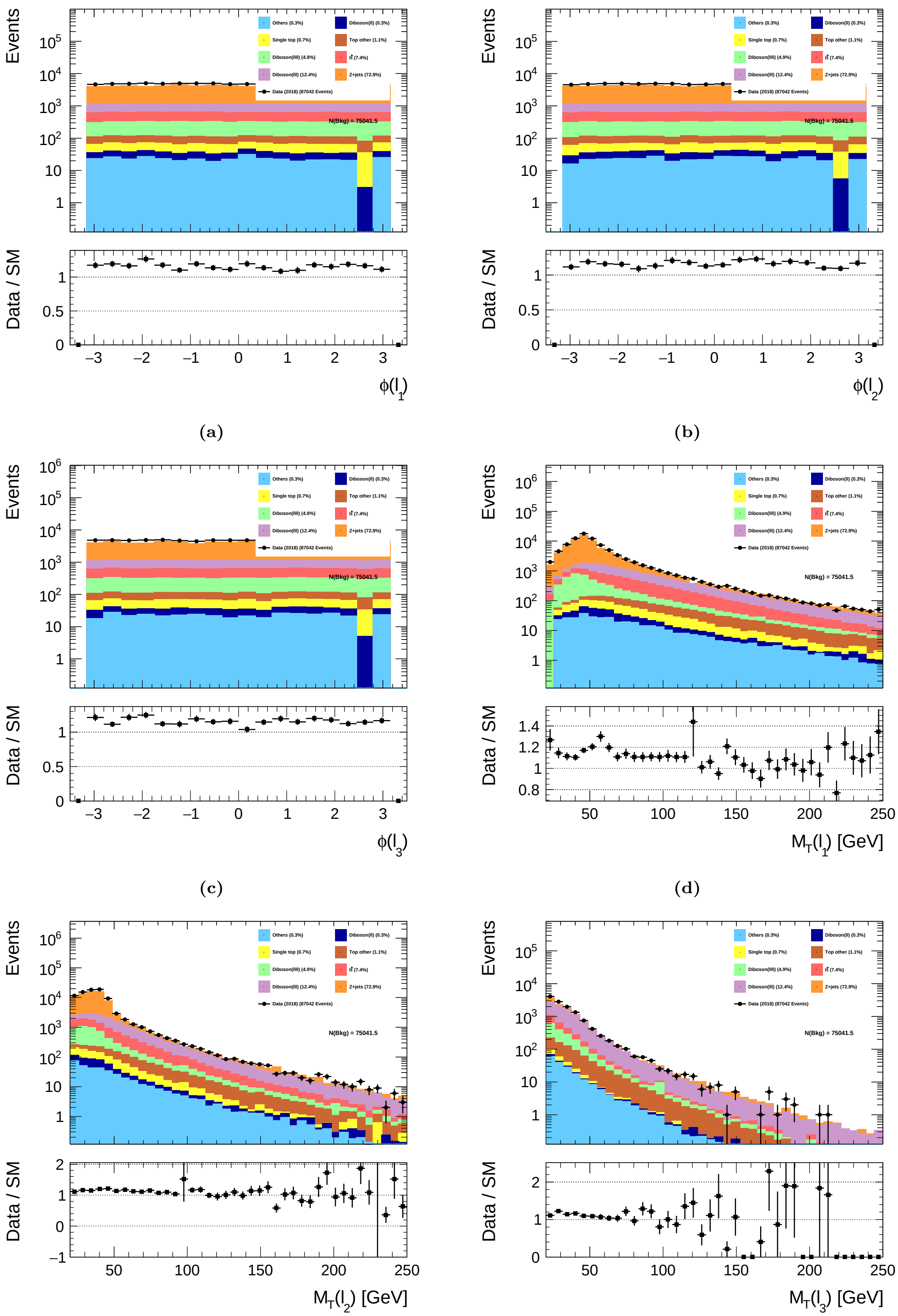


(e)



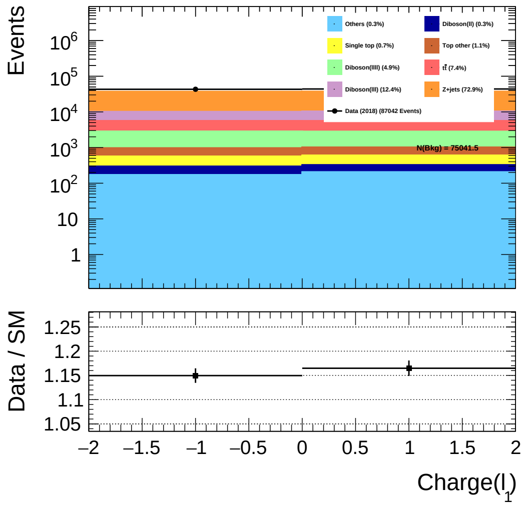
(f)

Figure 3.1: The event distribution for for each channel over P_t for the first 3.1a , second 3.1b and third 3.1c lepton. Similarly the distribution over η for the 3.1d, second 3.1e and third 3.1f lepton

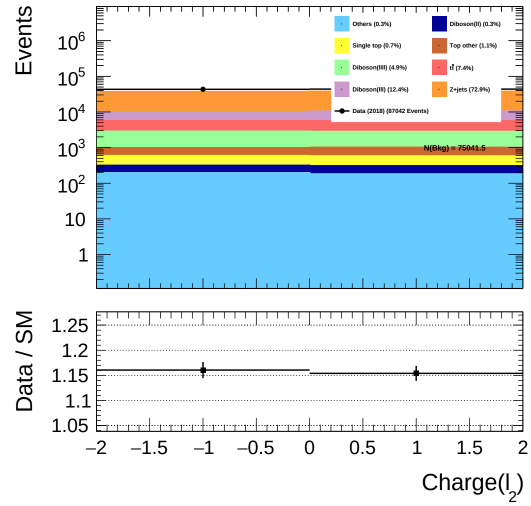


(a) (b) (c) (d) (e) (f)

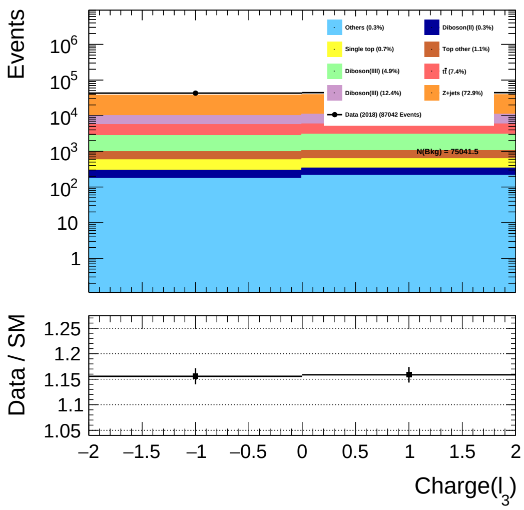
Figure 3.2: The event distribution for for each channel over ϕ for the first 3.2a , second 3.2b and third 3.2c lepton. Similarly the distribution over m_t for the first 3.2d, second 3.2e and third 3.2f lepton



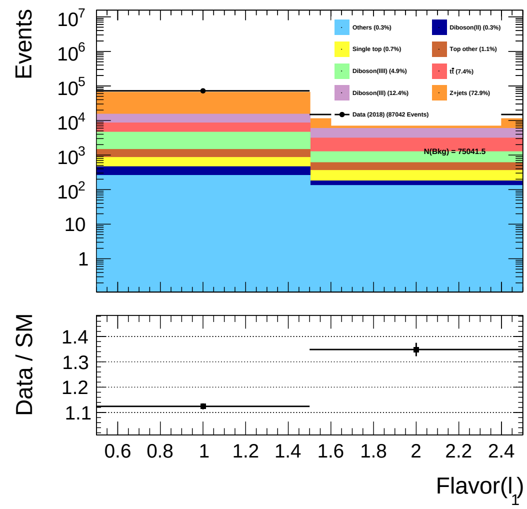
(a)



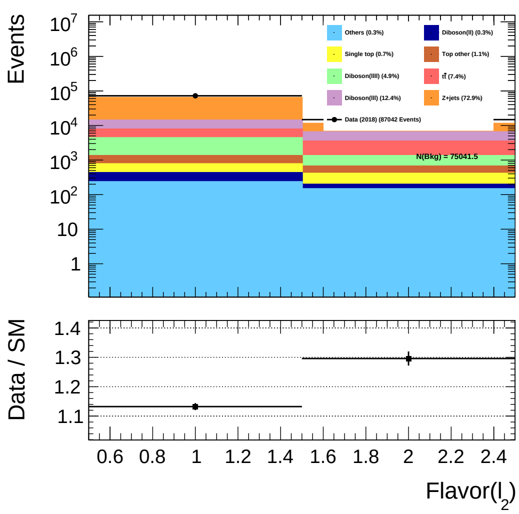
(b)



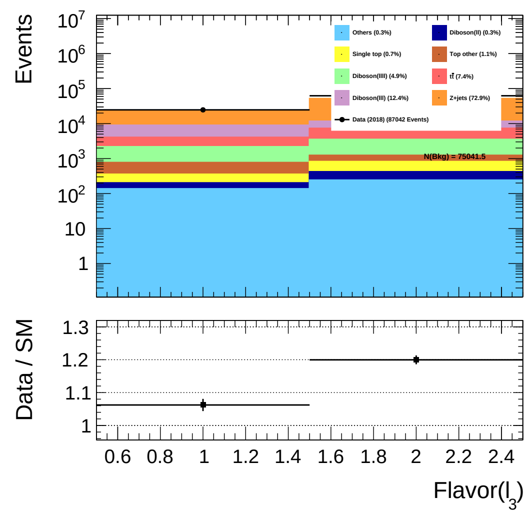
(c)



(d)

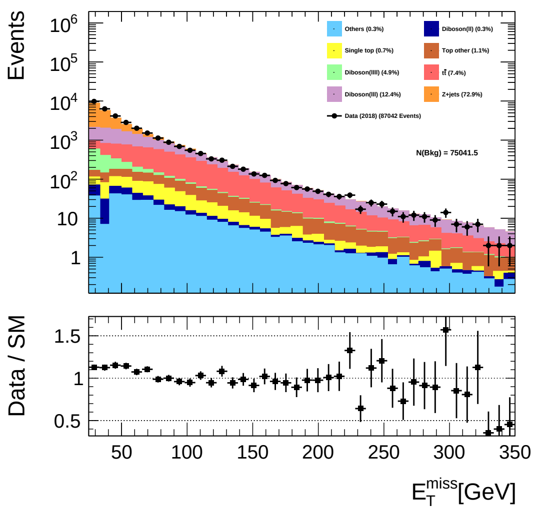


(e)

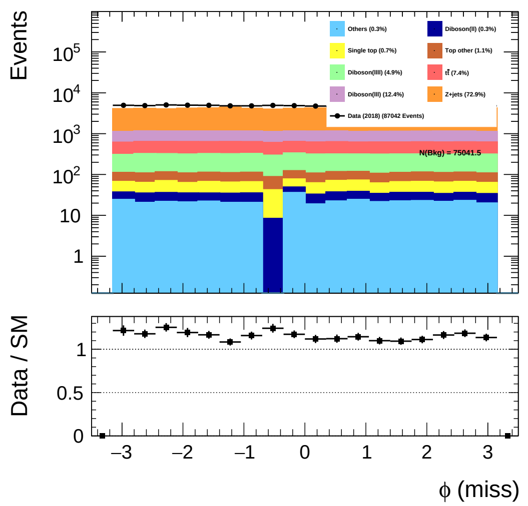


(f)

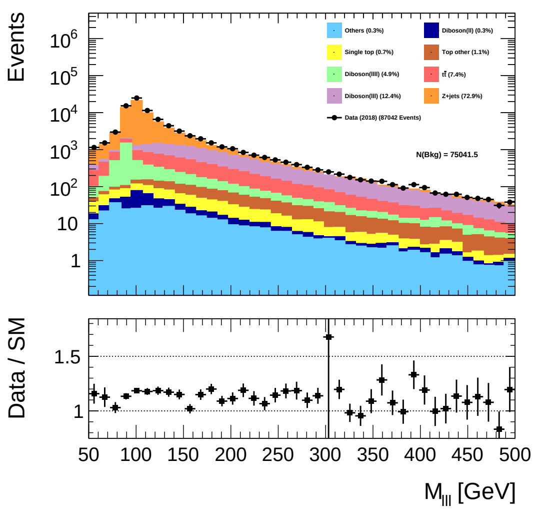
Figure 3.3: The event distribution for for each channel over the charge for the first 3.3a , second 3.3b and third 3.3c lepton. Similarly the distribution over the flavor for the first 3.3d, second 3.3e and third 3.3f lepton



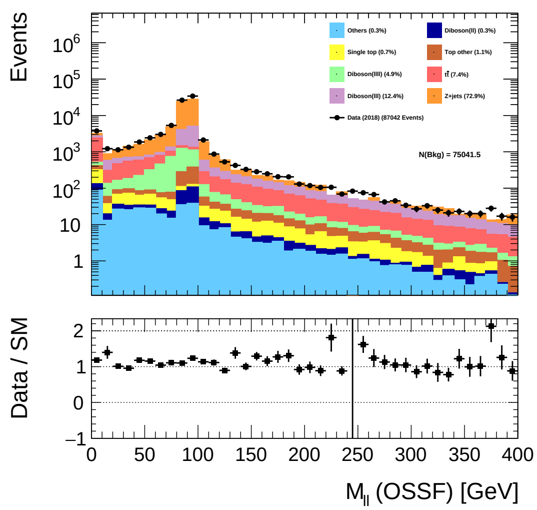
(a)



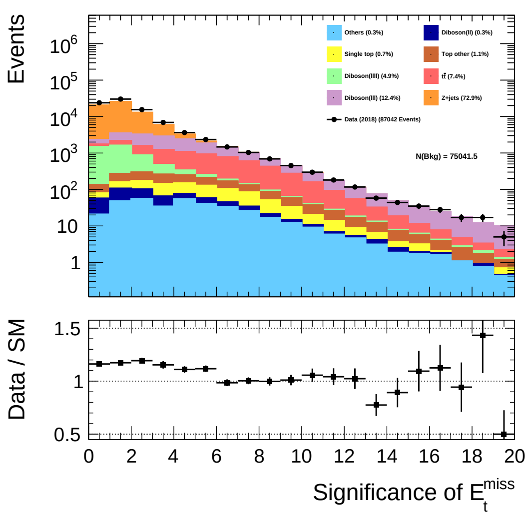
(b)



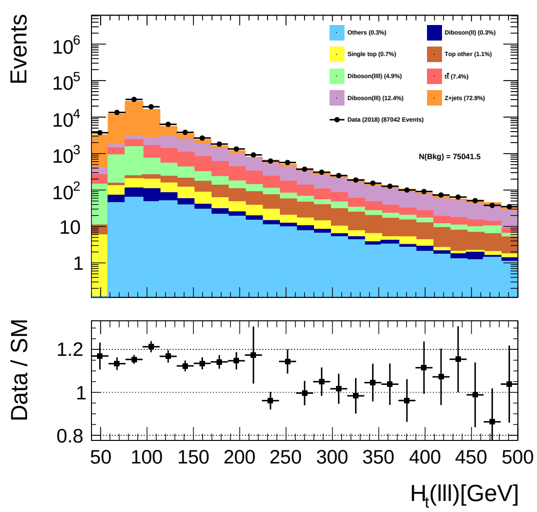
(c)



(d)



(e)



(f)

Figure 3.4: The event distribution for for each channel over the energy 3.4a and azimuthal angel 3.4b for the transverse momentum. The distribution of the invariant mass of the three leptons 3.4c and the OSSF pair 3.4d. The distribution over the significance of the missing transverse enegy 3.4e and the sum of P_t 3.4f.

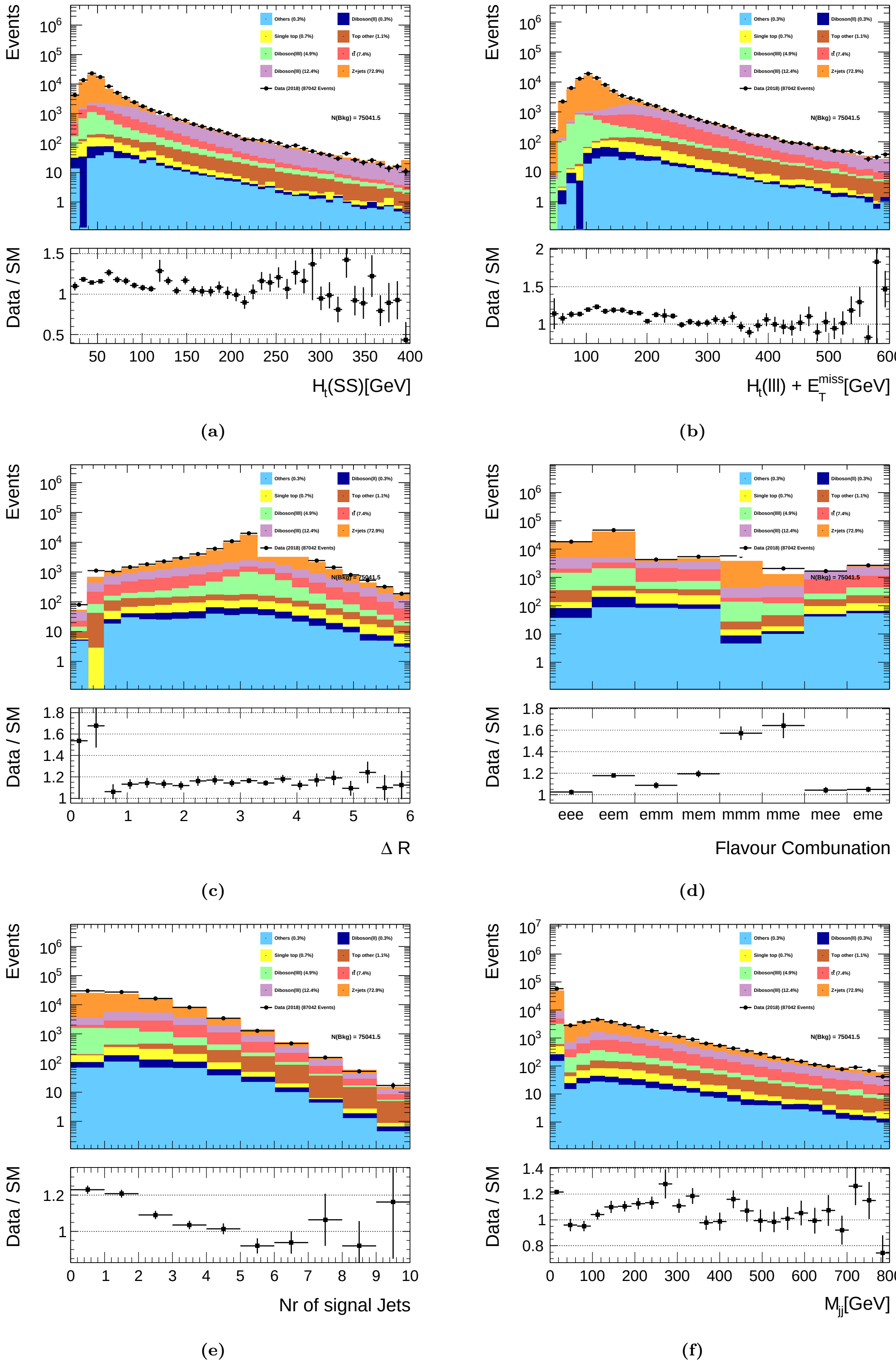
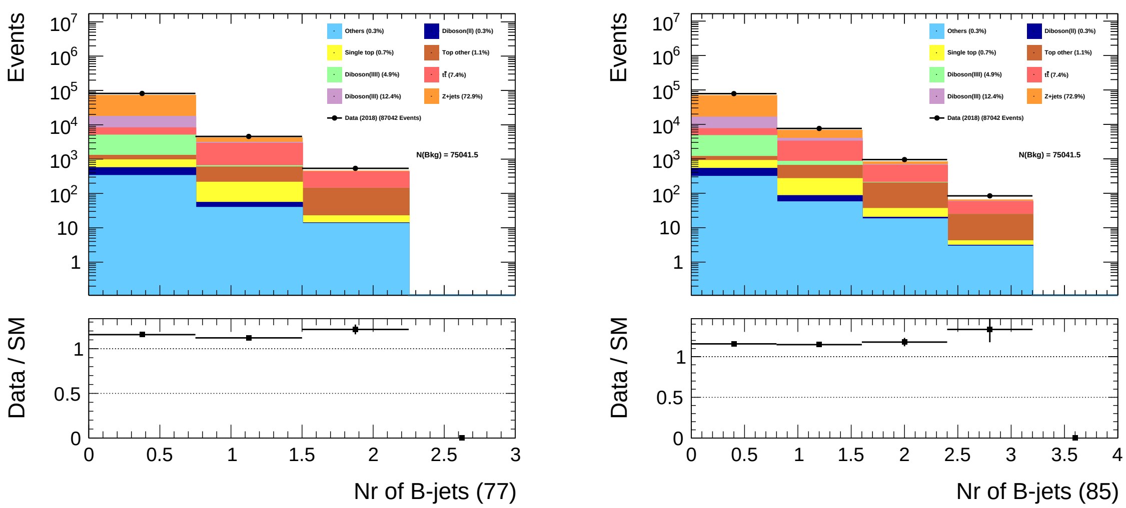


Figure 3.5: The event distribution for for each channel over the sum of P_t for the SS pair 3.5a and the sum over all three leptons added with E_t^{miss} 3.5b. The distribution over ΔR 3.5c and the flavor combination of the three leptons 3.5d. The distribution of number of jets 3.5e and the mass of the leading di-jet pair 3.5f.



(a) **(b)**

Figure 3.6: The event distribution for for each channel over the number of b-jets with 77% 3.6a and 85% 3.6b efficency.

Appendices

Appendix A

Acronyms

BSM Beyond Standard Model

DNN Deep Neural Networks

LHC Large Hadron Collider

MC Monte Carlo

ML Machine Learning

OSSF Opposite Sign Same Flavour

QCD Quantum Chromo Dynamics

QED Quantum Electro Dynamics

SM Standard Model

SR Signal region

SS Same Sign

Bibliography

[1] J. Joyce, *Finnegans Wake*. Penguin Books, New York, 1999.

[2] W. Thomson, *Lord kelvin addresed the british association for the advancement of science*, 1900.

[3] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban and D. Whiteson, *Jet flavor classification in high-energy physics with deep neural networks*, *Physical Review D* **94** (dec, 2016) .

[4] J. Pumplin, *How to tell quark jets from gluon jets*, *Phys. Rev. D* **44** (Oct, 1991) 2025–2032.

[5] M. Franchini, K. H. Mankinen, G. Carratta, F. Scutti, A. Gorisek, E. Lytken et al., *Search for type-III seesaw heavy leptons in dilepton final states in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, .

[6] M. Aaboud, , G. Aad, B. Abbott, D. C. Abbott, O. Abdinov et al., *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $$\sqrt{s} = 13$$ $$\text {TeV}$$*, *The European Physical Journal C* **79** (aug, 2019) .