This Dissertation

entitled

SEARCH FOR LEPTON FLAVOR VIOLATING DECAYS

OF HIGGS BOSONS

WITH THE CMS EXPERIMENT

typeset with NDdiss2$_\varepsilon$ v3.2017.2 (2017/05/09) on September 4, 2019 for

Nabarun Dev

This LaTeX 2$_\varepsilon$ classfile conforms to the University of Notre Dame style guidelines as of Fall 2012. However it is still possible to generate a non-conformant document if the instructions in the class file documentation are not followed!

> Be sure to refer to the published Graduate School guidelines at `http://graduateschool.nd.edu` as well. Those guidelines override everything mentioned about formatting in the documentation for this NDdiss2$_\varepsilon$ class file.

This page can be disabled by specifying the "`noinfo`*" option to the class invocation.* (i.e.,`\documentclass[...,noinfo]{nddiss2e}`)

This page is *NOT* part of the dissertation/thesis. It should be disabled before making final, formal submission, but should be included in the version submitted for format check.

NDdiss2$_\varepsilon$ documentation can be found at these locations:

`http://graduateschool.nd.edu`
`https://ctan.org/pkg/nddiss`

SEARCH FOR LEPTON FLAVOR VIOLATING DECAYS OF HIGGS BOSONS WITH THE CMS EXPERIMENT

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in
Physics

by

Nabarun Dev

Colin Philip Jessop, Director

Graduate Program in Physics
Notre Dame, Indiana
September 2019

SEARCH FOR LEPTON FLAVOR VIOLATING DECAYS OF HIGGS BOSONS WITH THE CMS EXPERIMENT

Abstract

by

Nabarun Dev

DEDICATED TO

To my family

CONTENTS

FIGURES

TABLES

PREFACE

Long time ago in a galaxy far far away....(preface is optional)

ACKNOWLEDGMENTS

I would like to acknowledge the light side of the force, Master Kenobi and Grand Master Yoda.

SYMBOLS

c	speed of light
m	mass
e	elementary charge
E	energy

CHAPTER 1

EVENT SELECTION

1.1 Introduction

This chapter describes in detail the event selection criteria for the analyses, and how they were chosen. It starts by introducing the backgrounds that each of selection criterion is trying to reduce in order to get a higher ratio of number of signal events to background events, leading to a better sensitivity for the search. This is followed by the procedure for arriving at the best possible set of selection criterion. For the h $\rightarrow \mu\tau_e$, two methods of selection were developed. The first method developed involves placing requirements on several kinematic variables, and then using the resulting ditribution of M_{col} as discriminant for a binned likelihood fit (see section **??** for description of statistical procedures). We call this method M_{col} fit method. The second method developed involves using a Boosted Decision Trees (BDT) discriminator for classification of signal and background events. The output distribution of the BDT discriminator is then used to perform the fit. We call this method BDT method. The BDT method is found to have greater sensitivity, as discussed later in the chapter. However, the M_{col} fit method is also presented as a complementary method and acts like a cross-check for the BDT method. For H $\rightarrow \mu\tau_e$ analysis, only the M_{col} fit method is developed. This is in part due to the dificulties foreseen in training a BDT with much fewer events available in H $\rightarrow \mu\tau_e$ analysis, and in part since this is the very first time the H $\rightarrow \mu\tau_e$ search is being performed, a simpler analysis was felt to be adequate.

Both analyses were performed blinded [**?**] in the signal region. All selection criterion and methods described below were developed without the knowledge of the observed data in the range of variable spectra where the signal is expected to be present. This is considered an optimal way of eliminating the unintended biasing of a result in a particular direction and is a standard methodology in particle physics analyses.

1.2 h125: $h \to \mu\tau_e$ analysis

1.2.1 $h \to \mu\tau_e$: Final state signature and backgrounds

The signature of the $h \to \mu\tau_e$ analysis final state consists of a muon that comes promptly from the Higgs and has a hard p_T spectrum, along with a softer electron of opposite sign charge that comes from the tau lepton, and missing transverse momentum from the tau decay. It is interesting to note that the signature is similar to the $h \to \tau_\mu\tau_e$ decay that is allowed by the SM and since been observed [**?**], but with significant kinematic differences. In $h \to \mu\tau_e$ decay the μ comes directly from the Higgs resulting in its p_T spectrum peaking and spreading out to much higher values. Also there are fewer neutrinos in $h \to \mu\tau_e$, coming from the decay of the single τ. The decay products of this highly boosted tau are closely aligned, leading to a narrow separation between the e and the $\vec{p}_T^{\,miss}$ in the azimuthal plane. The same is not true in the $h \to \tau_\mu\tau_e$ decays. These differences are illustrated pictorially in Fig. 1.1.

The most dominant backgrounds consists of $Z \to \tau\tau$ events coming from Drell-Yan production and $t\bar{t}$ production. In $Z \to \tau\tau$ events, one τ can decay to an e and the other to a μ. This background peaks at lower values of M_{col} than the signal events but there is significant overlap with the signal spectrum. In $t\bar{t}$ production, each of the top quarks can decay into a bottom and a W with the W bosons then decaying to a e and μ. The other backgrounds are smaller and include (in no particular order)

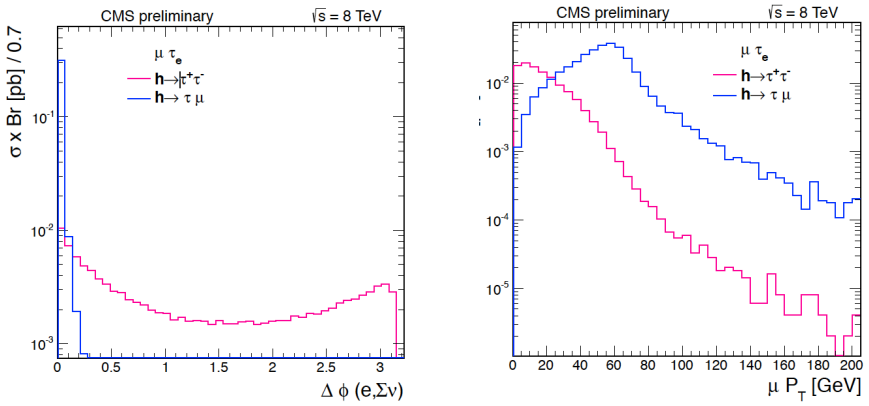


Figure 1.1: Illustration of the differences in p_T^{μ} and $\Delta\phi(e, \vec{p}_T^{miss})$ spectrums in $h \to \mu\tau_e$ and $h \to \tau_\mu\tau_e$ processes.

electroweak diboson production (WW, WZ and ZZ), h boson decays allowed by the SM ($H \to \tau\tau, WW$), $W\gamma^{(*)}$ + jets ,single top production, W + jets events, $Z \to \ell\ell$ ($\ell = e, \mu$) + jets and QCD multijet backgrounds. These backgrounds are described in more detail, along with there estimation and validation techniques in section **??**.

1.2.2 $h \to \mu\tau_e$:Baseline selection and categorization

A baseline selection is defined first in order to ensure that we have clean and well-defined events faithful to the final state signature of the signal process. An isolated and well-identified μ is thus required to be present along with a well-identified and isolated e of opposite sign charge. They are required to be separated by $\Delta R > 0.3$. The identification criterion applied for μ and e have been described in sections **??** and **??**. Isolation criterion, as measured by I_{rel} (described in **??**), are required to have values $I_{rel}^{e} < 0.15$ and $I_{rel}^{\mu} < 0.1$. The p_T of these candidates are required to be above minimal thresholds required by trigger, identification and isolation requirement. Both candidates are also required to be within the fiducial region of the detector. The μ is required to have $p_T^{\mu} > 26\,GeV$ and $|\eta^{\mu}| < 2.4$.The e is required to have $p_T^{e} > 10\,GeV$ and $|\eta^{e}| < 2.3$. Only events with two or fewer jets are considered. All jets

considered must have $p_T > 30\,\text{GeV}$, $|\eta| < 2.4$ and satisfy the loose identification criterion described in section **??**. Events with one or more jets arising from a b-quark (b-tagged jets) are vetoed. Cleaning events with b-tagged jets reduce some contribution from backgrounds which give rise to b-quarks such as $t\bar{t}$ and single top. Also, as described in **??**, any event with one or more jets within $\Delta R < 0.4$ of either lepton candidates is also rejected. Further, an event is rejected if it has additional μ or e, or any τ_{had} candidates. All the above baseline selection requirements have been summarized in Table 1.1. All the events were required to pass isolated muon triggers with a p_T threshold of 24 GeV. The trigger selection has been described in detail in section **??**. The distributions of the M_{col} and several other kinematic variables after the baseline selection just described, are shown in Figs. 1.2 and 1.3. These distributions act as the starting point for development of stricter kinematic selections looking at the different shapes of signal and backgrounds distributions for different variables.

Table 1.1: Baseline selection criteria for $h \to \mu\tau_e$ analysis.

Variable	μ	e
p_T	$> 30\,\text{GeV}$	$> 10\,\text{GeV}$
$\|\eta\|$	< 2.4	< 2.3
I_{rel}	< 0.15	< 0.1
Cleaning requirements		
$\Delta R(\mu, \text{e}) > 0.3$		
No additional μ, e or τ_{had}		
No b-tagged jets with $p_T > 30\,\text{GeV}$		
No jets with $\Delta R(\mu, jet) < 0.4$ and $p_T > 30\,\text{GeV}$		
No jets with $\Delta R(\text{e}, jet) < 0.4$ and $p_T > 30\,\text{GeV}$		

At this point the events are divided into several buckets, called categories. This

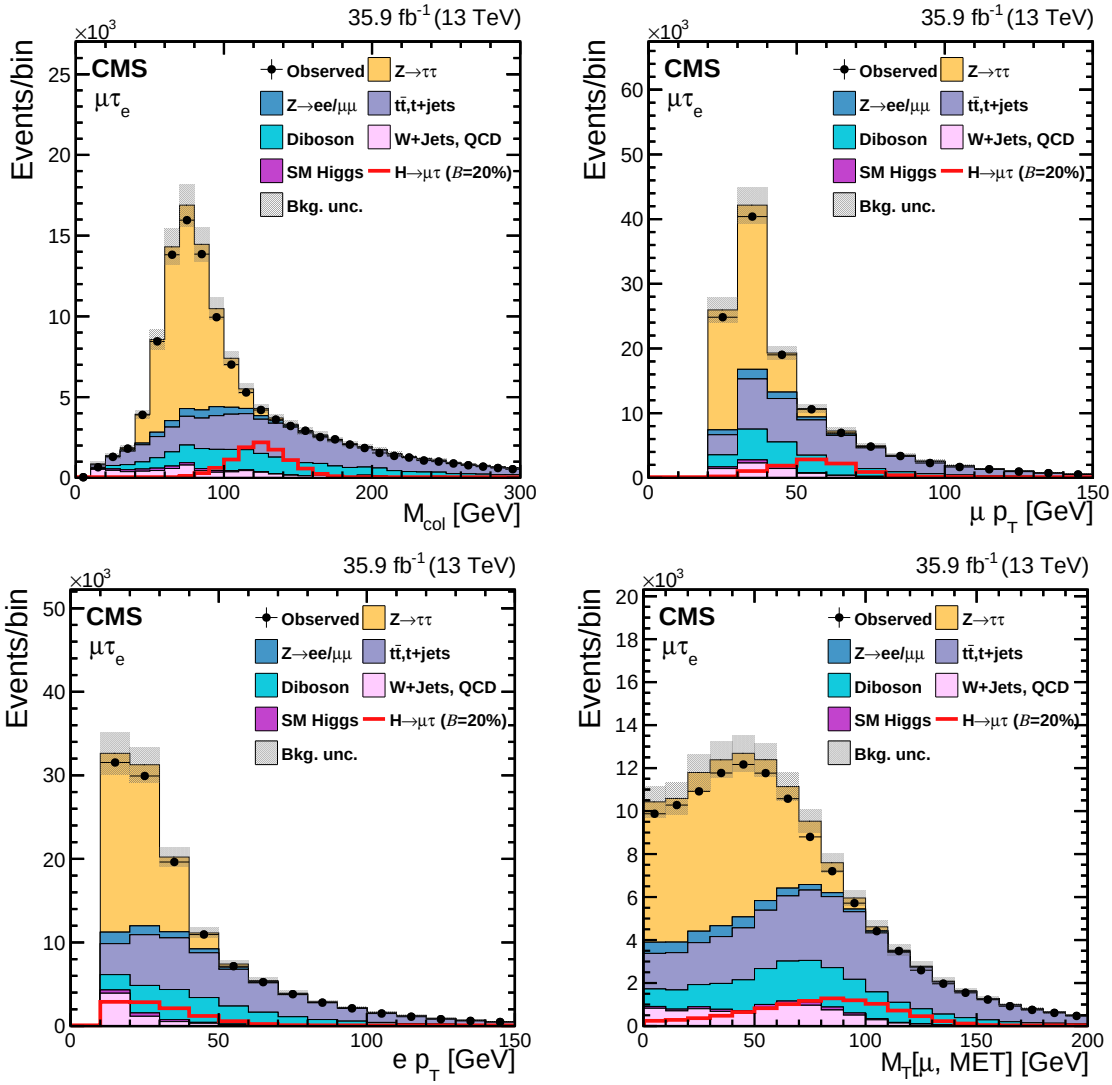


Figure 1.2: Distributions of kinematic variables after baseline selction for h $\rightarrow \mu\tau_e$ analysis (1).

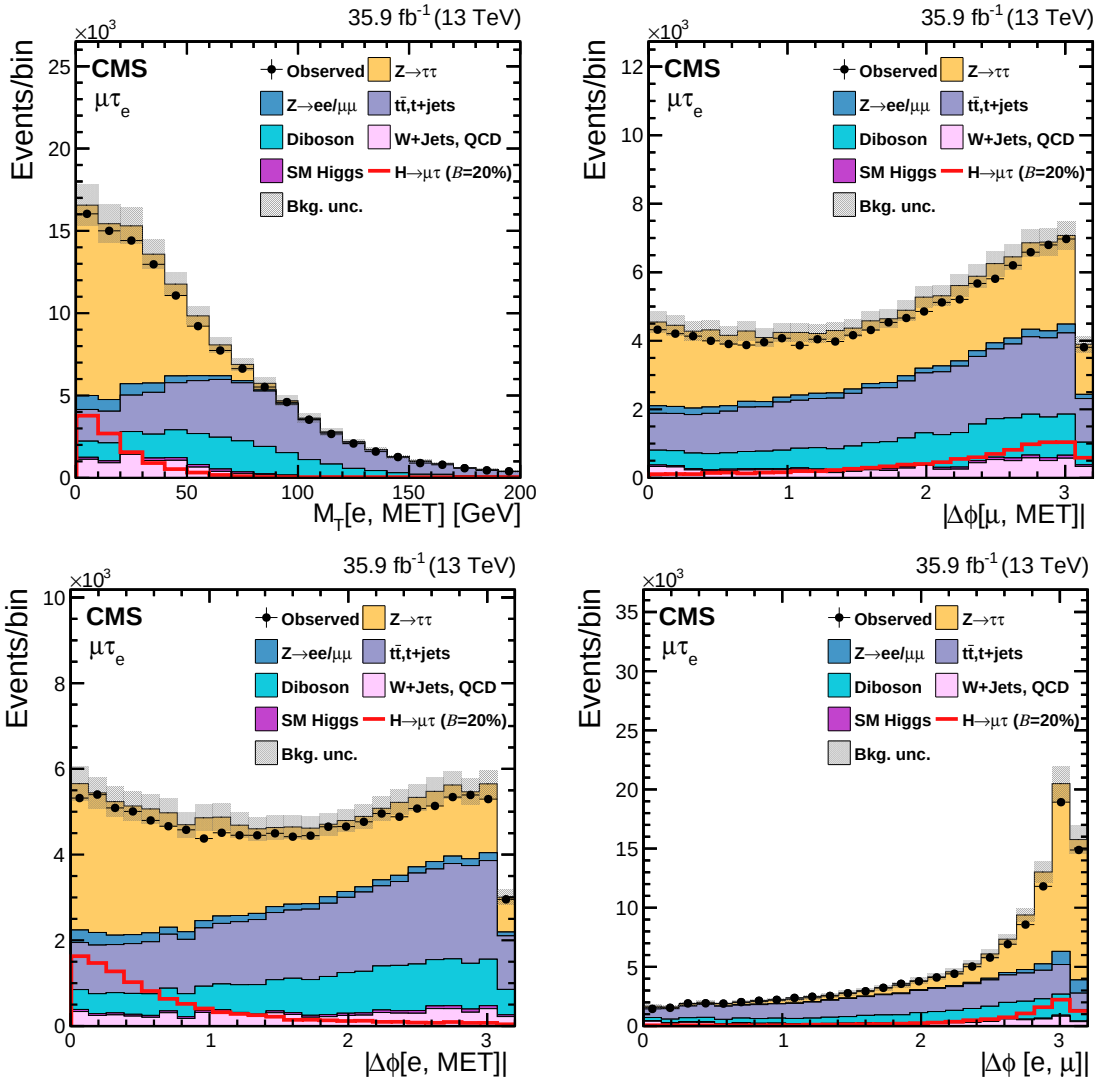


Figure 1.3: Distributions of kinematic variables after baseline selction for h $\rightarrow \mu\tau_e$ analysis (2).

is done on the basis of number of jets present in the event. In events with 2 jets the invariant mass of the di-jet system (M_{jj}) is also used for categorization. The topology of events containing different number of such jets can be different. For example, in events with one energetic jet the h produced can be boosted resulting in the azimuthal separation of the μ and e (that come from its decay) to be narrower than events with no jets. Each of this categories enchance the contribution of different h boson production mechanisms, and requiring different optimal selection criteria in each category helps increase the sensitivity of the search. The categories in order of decreasing number of signal events are:

- **0-jet category**: These are events that do not have any jet. This category enhances the gluon-gluon fusion (GGF) contribution.
- **1-jet category**: Events that have 1 jet are put in this category. This category enhances the GGF production with initial state radiation (ISR). Some VBF events where one jet has escaped detection can also enter this category.
- **2-jet GGF category**: This category contains events that have 2 jets with the additional requirement that $M_{jj} < 550\,\text{GeV}$. The dominant contribution comes from GGF production in association with two jets.
- **2-jet VBF category**: This category contains events that have 2 jets with the additional requirement that $M_{jj} \geq 550\,\text{GeV}$. The dominant contribution comes from VBF production which is characterized by presence of two jets with high dijet mass.

1.2.3 $h \to \mu\tau_e$: M_{col} fit selection

In the M_{col} fit method, the selection is performed by placing kinematic cuts on several variables to enhance the signal-to-background ratio. There are several variables considered for this and they include: the azimuthal separation ($\Delta\phi$) between μ

and e, between e and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, between μ and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, denoted respectively by $\Delta\phi(\mathrm{e},\mu)$, $\Delta\phi(\mathrm{e},\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$, $\Delta\phi(\mu,\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$, and the transverse mass between μ and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, between e and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, devnoted respectively by $M_T(\mu)$ and $M_T(\mathrm{e})$. The $\mathrm{h} \rightarrow \mu\tau_e$ decay being a 2-body decay, the μ and e are expected to be well separated in the azimuthal plane. Therefore, selecting events with a $\Delta\phi(\mathrm{e},\mu)$ larger than a threshold can help reject background events while keeping the signal that is peaked at high $\Delta\phi(\mathrm{e},\mu)$ values. This can be seen from Fig 1.3 (bottom right). Both neutrinos in the signal process come from the decay of the same τ. These neutrinos form the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. As mentioned earlier, the τ being much lighter than the h, it is highly boosted and its decay products i.e. e and the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ are expected to be close to each other in the azimuthal direction. Thus $\Delta\phi(\mathrm{e},\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$ is expected to peak at values close to zero for signal events, as seen in Fig 1.3 (bottom left). Given that all backgrounds have relatively flat shape for this variable throughout the $\Delta\phi$ range, requiring $\Delta\phi(\mathrm{e},\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$ to be lower than a threshold works as a strong rejection criterion against the backgrounds. Following a similar line of reasoning, the μ is expected to be well separated from the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ resulting in $\Delta\phi(\mu,\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$ for signal events to peak at high values, as seen in Fig 1.3 (top right). Further, as the $M_T(\ell)$ (defined in section **??**) contains negative of the cosine of $\Delta\phi(\ell,\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$ term, it is expected to be peak at values similar to $\Delta\phi(\ell,\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$. This can be seen from Fig 1.3 (top left) and Fig 1.3 (bottom right) which show signal events for $M_T(\mu)$ and $M_T(\mathrm{e})$ peak at relatively higher and lower values than most backgrounds respectively. In particular, requiring $M_T(\mu)$ to be larger than a threshold can help reject a lot of $\mathrm{Z} \rightarrow \tau\tau$ events which is the most dominant background in the 0-jet category. All the above variables have some amount of correlation with one another (see the correlation matrix shown in Fig. 1.6. The optimization procedure used to arrive at the most optimal set of kinematic thresholds for these variables is described in detail in the next paragrhaph. The thresholds on the $\mathrm{p_T}$ of the μ and e have not been made stricter to avoid biasing the selection

toward energetic leptons that sculpt the background M_{col} distribution to mimic the signal peak. This effect could potentially reduce the shape discrimination power of the signal extraction procedure. Only in the 0-jet category category the requirement on $\mathrm{p_T}$ of the μ is made marginally stricter by requiring $\mathrm{p_T^{\mu}} > 30\,\mathrm{GeV}$. All other lepton $\mathrm{p_T}$ requirements are allowed to remain the same as baseline selection and are not included in the optimization procedure.

The aim of the optimization procedure is to maximize the sensitivity of the analysis. In other words, we want to select a set of thresholds which increases a quantity such as the $\frac{S}{\sqrt{S+B}}$ ratio where S and B are the number of estimated signal and background events respectively. It is also necessary to ensure alongwith, that the entire spectrum of distribution of the discriminant variable (that is used int the final max-likelihood fit to extract results) is well-populated, escpecially in the region where the signal is expected to appear. A bad fit can potentially degrade the sensitivity of the analysis. Taking both of the above points into consideration, the thresholds have been optimized to obtain the most stringent (lowest) possible expected limits. The definition and procedure of extacting the expected limit is given in section **??**). To do the optimization of the kinematic thresholds, we start by requiring the baseline selection. Then for a variable in consideration,e.g.- $\Delta\phi(\mathrm{e}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$, we look at the expected limit while making the threshold progressively stricter until we reach a point where making the threshold any stricter degrades (increases) the expected limit. We repeat this procedure for all variables and note the stringent expected limit for each (by tightening thresholds of only that variable). This concludes one round of the optimization. For the next round we start by requiring the baseline selection. In addition we require that the variable that achieved the best possible expected limit among all variables in the last round satisfy its corresponding threshold. Lets call this variable variable1. We now repeat the same procedure as the last round for all but variable1. Say the variable that gave us the best possible expected limit this

round is variable2. For the start of the following round variable2 is is required to satisfy its corresponding threshold. Then all the other variables (including variables that were had chosen thresholds in earlier rounds such as variable1 here) are made to go through the same procedure. This is done because the optimum value of threshold for variables chosen earlier might shift as new variables are chosen. This process is continued until the expected limit becomes no further stringent in sucessive rounds. This optimization was done separately for each of the four categories. The final set of thresholds arrived at in this way for the h $\rightarrow \mu\tau_e$ $M_{\rm col}$ fit analysis are listed in Table. 1.2. This method of choosing the optimal set of thresholds is sometimes called the n-1 procedure, and the idea is conceptually similar to forward/backward selection methods used in statistical learning to build optimal models.

TABLE 1.2

FINAL SELECTION CRITERIA FOR h $\rightarrow \mu\tau_e$ $M_{\rm col}$ FIT ANALYSIS.

Category	0-jet	1-jet	2-jet GGF	2-jet
$\mathrm{p}_{\mathrm{T}}^{\mu}$	$> 30\,\mathrm{GeV}$	–	–	–
$M_T(\mu)$	$> 60\,\mathrm{GeV}$	$> 40\,\mathrm{GeV}$	$> 15\,\mathrm{GeV}$	$> 15\,\mathrm{GeV}$
$\Delta\phi(\mathrm{e}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$	< 0.7	< 0.5	< 0.3	< 0.3
$\Delta\phi(\mathrm{e}, \mu)$	> 2.5	> 2.0	–	–

1.2.4 h $\rightarrow \mu\tau_e$: BDT method selction

In the BDT method, a boosted decision trees (BDT) classifier is used to discriminate signal events from background events. A decision tree is a classifier which works by building a tree structure based on binary splits (as shown in Fig. 1.4). Starting from the root node of the tree (which contains all the events which we want to classify), a sequence of binary splits is made using input variables provided to the classifier. At each split, the variable which provides best purity of split or equivalently, in our case the best separation of signal and background events, is used. The same variable can thus be used for splitting several nodes and the splitting is continued until a desired some stopping criterion such as depth of the tree, purity of leaf nodes , minimum number of events in a leaf node etc. is reached. All events end up in one of the leaf nodes. If an event ends up in a leaf node in which signal events form the majority fraction, it is classified as a signal event. Otherwise, it is classified as a background event. Boosting is a class of ensemble machine learning techniques which help in enhancing performance of weak classifiers by sequentially building classifiers using reweighted (boosted) versions of the training data and then taking a weighted majority vote of the sequence of classifiers thus produced. Boosting also stabilizes the response of the classifiers with respect to fluctuations in the training data. In other words it helps avoid overfitting to the training data. When the boosting technique is a applied to produce an ensemble of decision trees, the resulting ensemble of classifiers is called a Boosted Decision Trees classifier. A detailed overview of how decision trees and boosting works, and the chosen value of parameters used in training the BDTs for this analysis is given in appendix A.

The BDT is trained using events that satisfy the baseline selection criteria. Simulated GGF and VBF events weighted by their cross-section are used as signal events for training. For background, a mixture of $t\bar{t}$ and Drell-Yan events are used, also weighted by their respective cross-sections. The $t\bar{t}$ and Drell-Yan backgrounds are

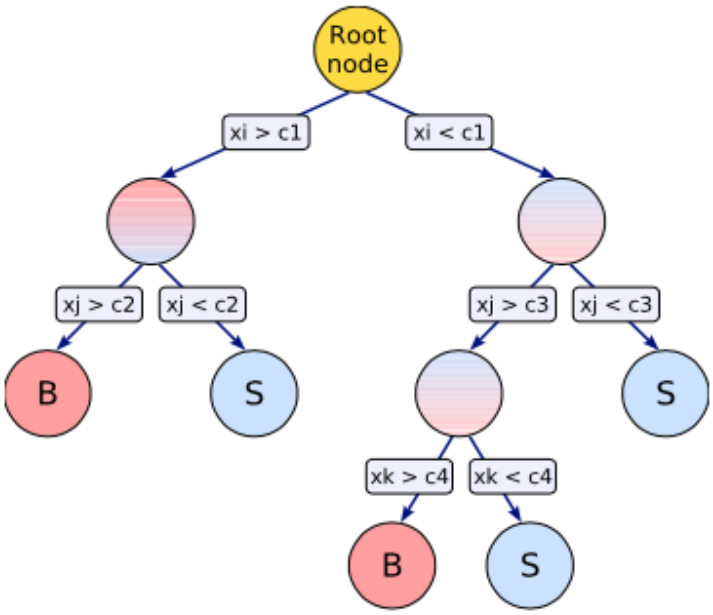


Figure 1.4: Illustration of decision tree. [**?**]

the most dominant backgrounds. The Drell-Yan background is the most dominant background in 0-jet and 1-jet category, while the $t\bar{t}$ background is the most dominant in both 2-jet categories. It also has many kinematic characteristics in common with diboson and single-top bankgrounds. A suite of input variables is used in training of the BDT. They are as follows:

- Transverse mass between the μ and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$: $M_T(\mu)$.
- Transverse mass between the e and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$: $M_T(\mathrm{e})$.
- Azimuthal angle between the e and μ: $\Delta\phi(\mathrm{e},\mu)$.
- Azimuthal angle between the e and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$: $\Delta\phi(\mathrm{e},\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$.
- Azimuthal angle between the μ and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$: $\Delta\phi(\mu,\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$.
- Collinear mass: M_{col}.
- Muon $\mathrm{p_T}$: $\mathrm{p}_{\mathrm{T}}^{\mu}$.

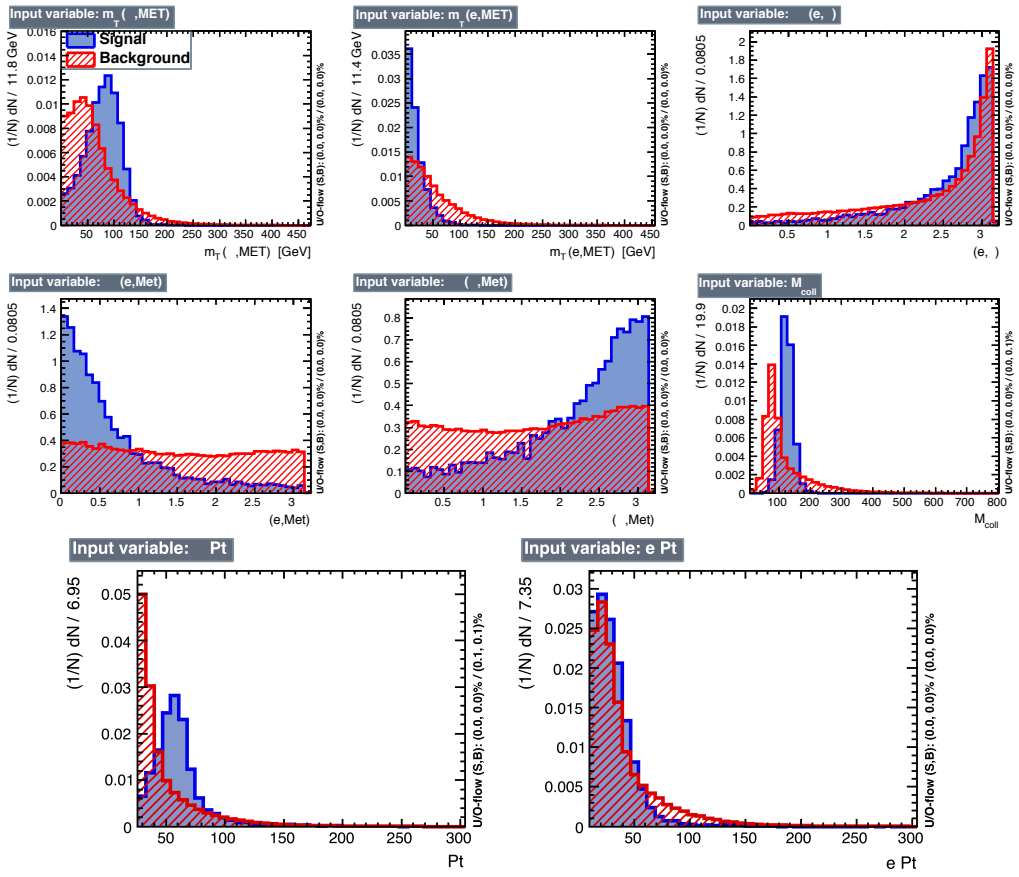


Figure 1.5: Normalized distributions of the input variables for BDT method. The signal (blue) is composed of a weighted mixture of GGF and VBF events, whereas the background (red) is made of $t\bar{t}$ and Drell-Yan events. All events were required to satisfy the baseline selection criteria.

- Electron p_T: p_T^e.

The distributions of these variables normalized to the total number of events in the input sample to the BDT is shown in Fig. 1.5. The correlations between these variables in signal and background events are shown in Fig. 1.6.

The training was done with a 800 decision tree ensemble, each tree having a maximum depth of 4. The gini-index criterion was used for splitting the data at each node. Further, AdaBoost (adaptive bossting) method was used for boosting (see appendix A for details of these techniques). A training to testing split of 70:30 split

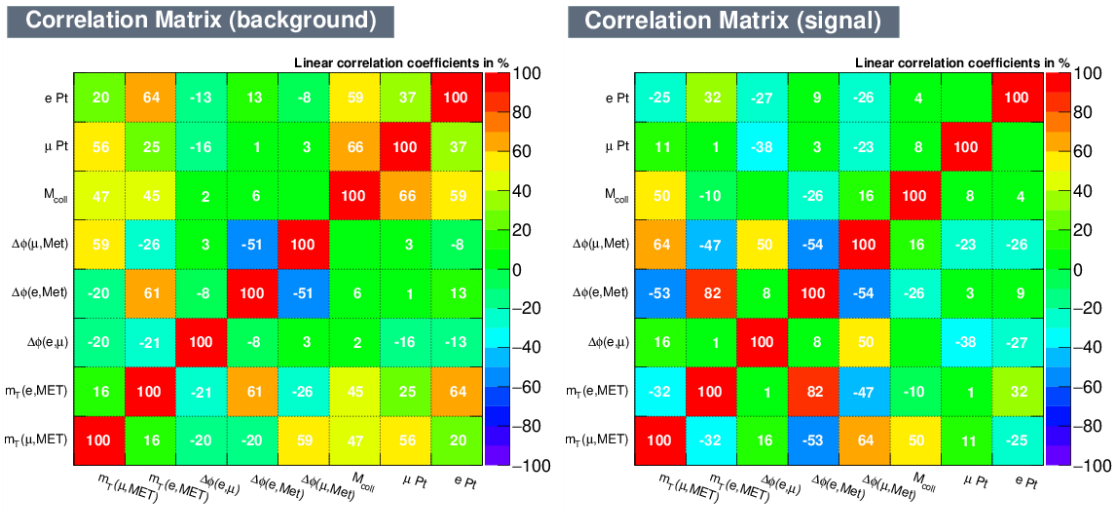


Figure 1.6: Correlations between input variables for signal events (right) and background events (left).

was used. Fig. 1.7 shows the distribution of the BDT response for training and testing samples. The training and testing distributions for both signal and background events match well, suggesting that there is no overtraining. The distribution of BDT response is used in max-likelihood fit to extract results, as discussed in section 2.3.3.

1.3 Heavy higgs:H $\to \mu\tau_e$ analysis

1.3.1 H $\to \mu\tau_e$: Final state signature and backgrounds

The signature of the H $\to \mu\tau_e$ analysis final state is very similar to that of h $\to \mu\tau_e$. It also consists of a muon that comes promptly from the Higgs and has a hard p_T spectrum, along with a softer electron that comes from the tau lepton, and missing transverse momentum from the tau decay. The p_T^{μ} spectrum is expected to be harder for higher H boson masses. The topologies being similar, the kinematic properties discussed in section 1.2.1 for h $\to \mu\tau_e$ analysis also apply to the H $\to \mu\tau_e$ analysis. The H boson mass peaks for all the simulated samples illustrated in Fig 1.8.

The most dominant backgrounds for H $\to \mu\tau_e$ consists of events from $t\bar{t}$ and

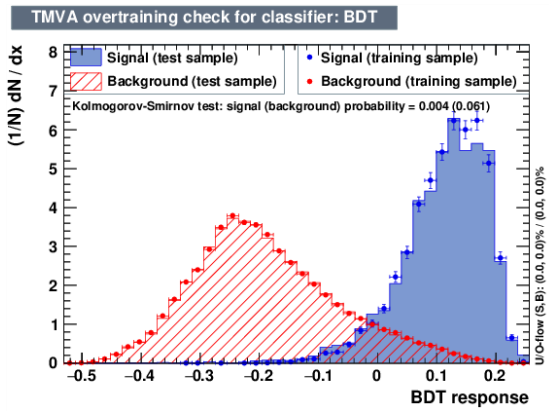


Figure 1.7: Distribution of BDT response for training (dots) and test(fill) distributions for both signal(blue) and background(red) events.

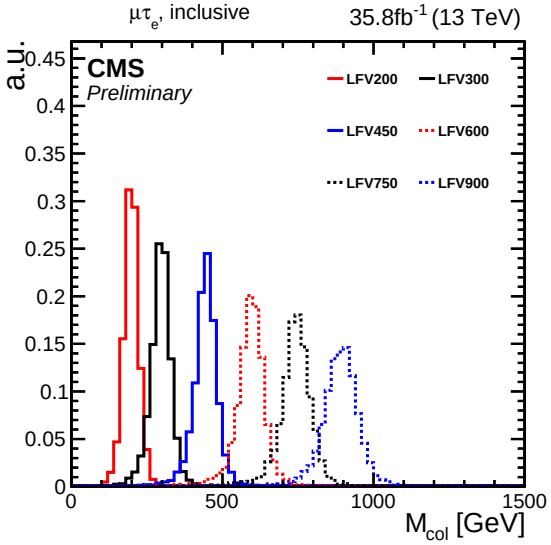


Figure 1.8: Illustration of simulated signal mass peaks for H $\rightarrow \mu\tau_e$ analysis for different H boson masses.

electroweak diboson production. Unlike h $\rightarrow \mu\tau_e$ analysis, Z $\rightarrow \tau\tau$ events from Drell-Yan production form a very small background as the Z $\rightarrow \tau\tau$ spectrum peaks at much lower values (around Z boson mass) of collinear mass that the signal events coming from heavy H boson decays. The other backgrounds come from h boson decays (H $\rightarrow \tau\tau$, WW), W$\gamma^{(*)}$ + jets ,single top production, W + jets events, $Z \rightarrow \ell\ell$ (ℓ = e, μ) + jets and QCD multijet backgrounds. These backgrounds are described in more detail, along with there estimation and validation techniques in section **??**.

1.3.2 H $\rightarrow \mu\tau_e$: Baseline selection and categorization

The baseline selection for H $\rightarrow \mu\tau_e$ is similar to that of h $\rightarrow \mu\tau_e$ with the exception of higher p_T thresholds. Just like h $\rightarrow \mu\tau_e$, an isolated and well-identified μ is thus required to be present along with an well-identified and isolated e of opposite sign charge. They are required to be separated by $\Delta R > 0.3$. The identification and isolation criteria have been described in sections **??**, **??** and **??**. All events are required to pass a single muon trigger with the threshold of 50 GeV. The trigger selection has been described in detail in section **??**. The μ is required to have $p_T^{\mu} >$ 53 GeV and $|\eta^{\mu}| < 2.4$. The e is required to have $p_T^{e} > 10$ GeV and $|\eta^{e}| < 2.3$. Only events with zero or one jet are considered. Jets must have $p_T > 30$ GeV, $|\eta| < 2.4$ and satisfy the loose identification criterion described in section **??** to be considered. As only GGF production mode is considered for the H $\rightarrow \mu\tau_e$ analysis, events with more than one jet make negligent contribution and are rejected. All other other criteria are same as the h $\rightarrow \mu\tau_e$ analysis. The entire set of baseline selection criteria for H $\rightarrow \mu\tau_e$ has been summarized in table 1.3.

The events are then divided into categories, with motivations similar to the h $\rightarrow \mu\tau_e$ analysis (see section 1.2.2), on the basis of number of jets present in the event. The two categories for H $\rightarrow \mu\tau_e$ are:

- **0-jet category**: These are events that do not have any jet. This category

Table 1.3: Baseline selection criteria for $H \to \mu\tau_e$ analysis.

Variable	μ	e
p_T	$> 53\,\text{GeV}$	$> 10\,\text{GeV}$
$\lvert\eta\rvert$	< 2.4	< 2.3
I_{rel}	< 0.15	< 0.1
Cleaning requirements		
$\Delta R(\mu, \text{e}) > 0.3$		
No additional μ, e or τ_{had}		
No b-tagged jets with $p_T > 30\,\text{GeV}$		
No jets with $\Delta R(\mu, jet) < 0.4$ and $p_T > 30\,\text{GeV}$		
No jets with $\Delta R(\text{e}, jet) < 0.4$ and $p_T > 30\,\text{GeV}$		

enhances the gluon-gluon fusion (GGF) contribution.

- **1-jet category**: Events that have 1 jet are put in this category. This category enhances the GGF production with initial state radiation (ISR).

The distributions of several kinematic variables after the baseline selection and categorization are shown in Figs. 1.9 and 1.10.

1.3.3 $H \to \mu\tau_e$: mcol fit selection

Just like the M_{col} fit method in $h \to \mu\tau_e$, the selection is performed by placing kinematic cuts on several variables to enhance the signal-to-background ratio. The variables considered are: $\Delta\phi(\text{e}, \mu)$, $\Delta\phi(\text{e}, \vec{p}_T^{\text{miss}})$, $\Delta\phi(\mu, \vec{p}_T^{\text{miss}})$, $M_T(\mu)$ and $M_T(\text{e})$. In addition, the p_T of the μ and e are also considered. Since we are looking for a decay in an extended mass range (200-900 GeV) in $H \to \mu\tau_e$, and not in a particular region like the $h \to \mu\tau_e$ analysis, the potential effect of background mimicking the signal, in particular due to higher p_T thresholds of the leptons, is not apparent. The motivations for using these variables remain much the same like the $h \to \mu\tau_e$ analysis owing to similarities in topology. They are motivated by the facts that the only source of MET is the τ, and the τ being lighter than the H, its visible products are closely aligned, and the p_T spectrum of the prompt lepton (μ) is hard.

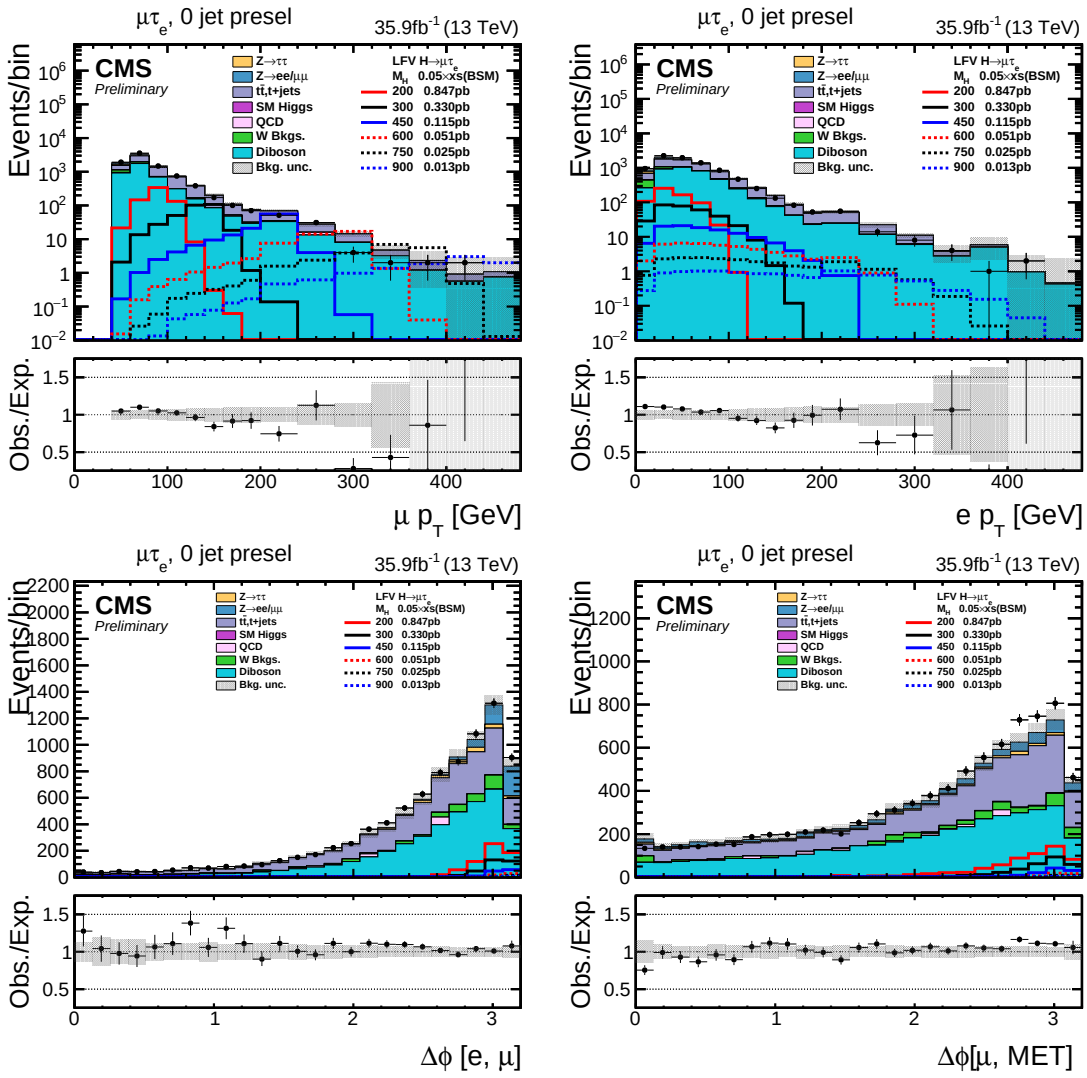


Figure 1.9: Distributions of kinematic variables after baseline selction for 0-jet category of H $\rightarrow \mu\tau_e$ analysis.

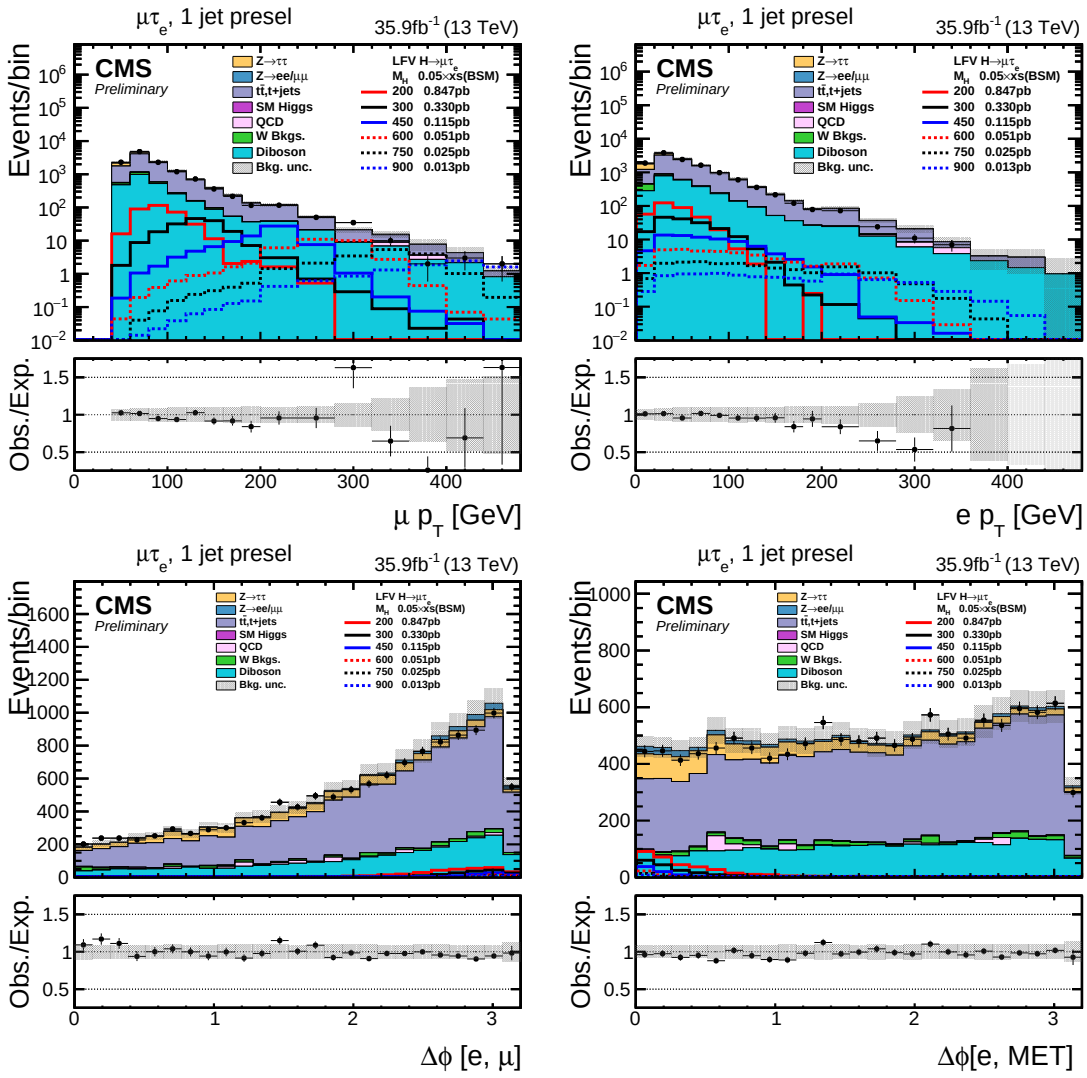


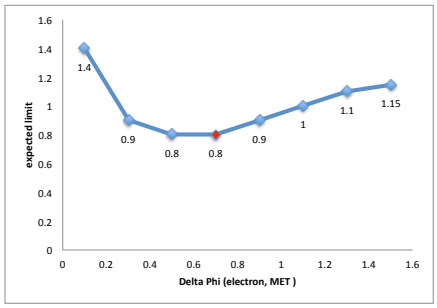
Figure 1.10: Distributions of kinematic variables after baseline selction for 1-jet category of H $\rightarrow \mu\tau_e$ analysis.

The procedure for optimization of the thresholds of for these variables is exactly the same as described in section 1.2.3. Further to get better sensitivity in the entire mass range from 200 to 900 GeV, two separate sets of thresholds are optimized, for each category. One set is optimized to provide better sensitivity in the 200-450 GeV mass range. The simulated signal for the H mass of 200 GeV is used when calculating expected limits during the optimization procedure for this mass range. The other set is optimized to provide better sensitivity in 450-900 GeV mass range. The simulated signal for H mass of 450 GeV is used when calculating expected limits during the optimization procedure for this mass range. A few illustrations of the optimization procedure are shown in Fig. 1.11. The final set of thresholds arrived at in this manner, for both mass ranges and both categories of the $H \to \mu\tau_e$ M_{col} fit analysis, are listed in Table. 1.4. The M_{col} distributions after requiring these selections is used in a max-likelihood fit to extract results, as discussed in section 2.3.3.

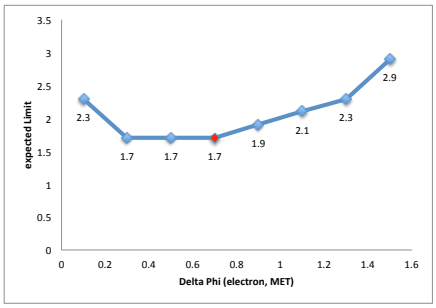
TABLE 1.4

FINAL SELECTION CRITERIA IN EACH CATEGORY OF THE $H \to \mu\tau_e$ ANALYSIS.

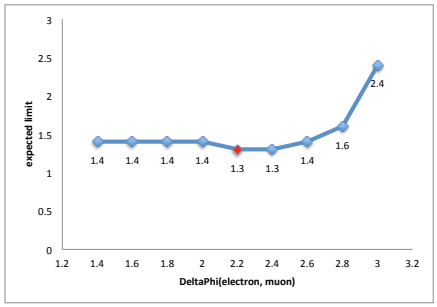
	Low mass range	High mass range
0-jet	$p_T^{\mu} > 60\,\text{GeV}$, $p_T^{e} > 10\,\text{GeV}$ $\Delta\phi(e, \vec{p}_T^{\,\text{miss}}) < 0.7$ $\Delta\phi(e, \mu) > 2.2$	$p_T^{\mu} > 150\,\text{GeV}$, $p_T^{e} > 10\,\text{GeV}$ $\Delta\phi(e, \vec{p}_T^{\,\text{miss}}) < 0.3$ $\Delta\phi(e, \mu) > 2.2$
1-jet	$p_T^{\mu} > 60\,\text{GeV}$, $p_T^{e} > 10\,\text{GeV}$ $\Delta\phi(e, \vec{p}_T^{\,\text{miss}}) < 0.7$ $\Delta\phi(e, \mu) > 2.2$	$p_T^{\mu} > 150\,\text{GeV}$, $p_T^{e} > 10\,\text{GeV}$ $\Delta\phi(e, \vec{p}_T^{\,\text{miss}}) < 0.3$ $\Delta\phi(e, \mu) > 2.2$



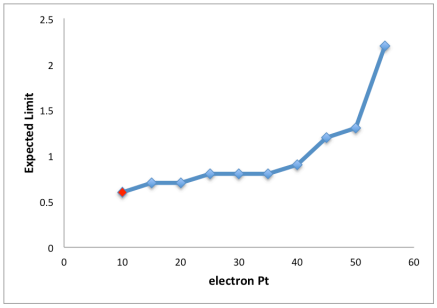
(a) Low mass range 0 jet $\Delta\phi(\mathrm{e}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$



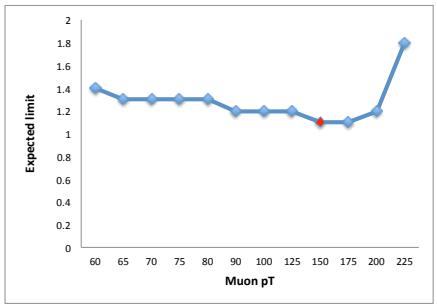
(b) Low mass range 1 jet $\Delta\phi(\mathrm{e}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$



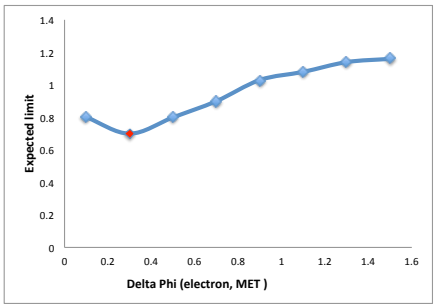
(c) Low mass range 0 jet $\Delta\phi(\mathrm{e}, \mu)$



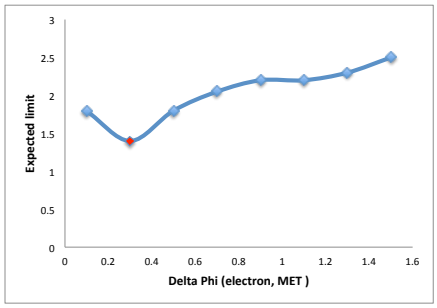
(d) Low mass range 0 jet $\mathrm{p}_{\mathrm{T}}^{\mathrm{e}}$



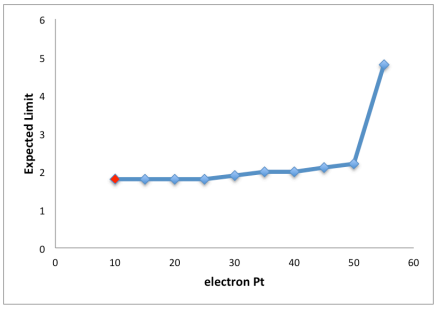
(e) High mass range 1 jet $\mathrm{p}_{\mathrm{T}}^{\mu}$



(f) High mass range 0 jet $\Delta\phi(\mathrm{e}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$



(g) High mass range 1 jet $\Delta\phi(\mathrm{e}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$



(h) High mass range 1 jet $\mathrm{p}_{\mathrm{T}}^{\mathrm{e}}$

Figure 1.11. Examples of cut optimisation for the $\mathrm{H} \to \mu\tau_e$ analysis

o

CHAPTER 2

SIGNAL EXTRACTION AND SYSTEMATIC UNCERTAINTIES

2.1 Introduction

The analysis is, in its essence, a sophisticated counting experiment. The presence of a signal is indicated by an excess of events over the predicted background, in the distribution of a signal variable. For our analyses the signal variables are collinear mass or BDT output, as described in Sections 1 and **??**. Given that there are several uncertainties, both experimental and theoretical and also due to the innate randomness in the process, it is possible that an excess is observed when there is no signal. So, when an excess is observed, a p-value which represents the probability that the excess is due to statistical fluctuations is computed. A very low p-value is taken to indicate that the excess corresponds to an observed signal and not merely a statistical fluctuation. Conversely, if no excess is observed (upper exclusion) limits are set on the product of branching fraction and production cross-section. A 95% CL (confidence level) is taken as a requirement for ruling out a signal at or above a certain value known i.e. upper exclusion limit. The first part of this chapter describes the statistical methods used, that very closely follow the procedure used for LHC Higgs boson search and described in [**?**].

Several sources of systematic uncertainties need to be considered when making the above measurement. The sources of these uncertainties can be theoretical, experimental or purely statistical in nature. Further, they can effect only the overall scale of the distributions (used to make the measurement), or effect there shape i.e.

change the scale differently in each bin of the distribution. All the uncertainties used in the analyses and their sources are described in the second part of this chapter.

2.2 Statistical methods for signal extraction

In the following section, the expected signal event yields are denoted by s, and backgrounds by b. The parameter μ that appears is the signal strength modifier, which changes the signal production cross-sections of all the production mechansims by exactly the same scale μ.

2.2.1 Likelihood function

The Poisson distribution is an appropriate model for n, the number of times an event occurs in an interval if the following assumptions are true [**?**].

- The occurrence of one event does not affect the probability that a second event will occur. That is, events occur independently.
- The rate at which events occur is constant. The rate cannot be higher in some intervals and lower in other intervals. This rate is the average number of events in the interval. λ.
- Two events cannot occur at exactly the same instant; instead, at each very small sub-interval exactly one event either occurs or does not occur.

The poisson probablity of distribution is then given by:

$$P(n_e vents) = \frac{e^{-\lambda}\lambda^n}{n!} \tag{2.1}$$

For a counting experiments such as ours, the above conditions approximately hold. The expected number of events is $\mu \cdot s + b$. The likelihood function $\mathcal{L}(data|\mu)$ is then

given by:

$$\mathcal{L}(data|\mu) = \prod_{i=1}^{bins} \frac{(\mu \cdot s_i + b_i)^{n_i}}{n_i!} e^{-\mu \cdot s_i - b_i} \tag{2.2}$$

,where n_i is the number of events observed in the bin i of the distribution, and s_i and b_i are expected number of signal and background events in that bin respectively.

2.2.2 Treatment of systematic uncertainties

All systematic uncertainties are handled by introducing them as nuisance parameters. Nuisance parameters are parameters that influence the model but are not of interest in our measurement, e.g., if we are interested in knowing only the mean of a population that is expected to be distributed as a gaussian, the standard deviation becomes a nuisance parameter for the model that we fit. In our experiment, the nuisance parameters are embedded into the likelihood function. In order for the likelihood function to have a clean factorised form [**?**], all sources of uncertainties considered are considered 100%-corrrelated or uncorrelated. If an uncertainty is partially correlated, it is either separated into 100%-corrrelated or uncorrelated components, or considered 100%-corrrelated or uncorrelated, depending on whichever is a more conservative estimate. The full suite of nuisance parameters is represented as θ. These effect the expected signal and backgeound yields which are now represented as $s(\theta)$ and $b(\theta)$. Each component of θ is associated with a default value $\tilde{\theta}$, reflecting our degree of belief on the real value of θ. The pdf (probablity distribution function) $\rho(\theta|\tilde{\theta})$ can then be interpreted as a posterior distribution from measurements of $\tilde{\theta}$. Using Bayes' theorem:

$$\rho(\theta|\tilde{\theta}) = \rho(\tilde{\theta}|\theta) \cdot \pi_\theta(\theta), \tag{2.3}$$

where the priors $\pi_\theta(\theta)$ are taken as flat distributions representing no prior knowledge of θ. This reformulation allows us to use the pdf of $\tilde{\theta}$ instead, i.e. $\rho(\tilde{\theta}|\theta)$ to directly constrain the likelihood of the measurement. The likelihood function after

the introduction of systematic uncertainties now becomes:

$$\mathcal{L}(data|\mu,\theta) = Poisson(data|\mu \cdot s(\theta) + b(\theta)) \cdot \rho(\tilde{\theta}|\theta) \tag{2.4}$$

Systematic unceraintites that effect only the overall scale of the distributions, correspond to a mutliplicative factor in the signal and/or background yields, and are described by log-normal pdfs. Log-normal pdfs are characterised by the width κ, and are well-suited for positively valued observables. The log-normal distribution looks like:

$$\rho(\theta|\tilde{\theta}) = \frac{1}{\sqrt{2\pi}\ \ln(\kappa)}\exp\ (\frac{\ln(\theta/\tilde{\theta})^2}{2(\ln\ \kappa)^2})\frac{1}{\theta} \tag{2.5}$$

Systematic uncertainties that effect the scale of the distribution differently in each been have the effect of altering its shape along with its scale. Such uncertainties are called shape uncertainties [**?**], and are modeled using a linear extrapolation method [**?**]. In practice, two alternate distributions obtained by varying the nuisance by ± 1 standard deviation are used, and a parameter is added to the likelihood that smoothly interpolates between these shapes.

2.2.3 Calculation of exclusion limits

The CL_s method [**?** **?** **?**] is used to set upper exclusion limits when no excess of data over background is observed. The test statistic used generally for hypothesis testing in searches at the LHC, uses profiling of nuisances as described above, and is based on the likelihood ratio [**?**], which by the Neyman-Pearson lemma is known as the most powerful discriminator. This is denoted by $\tilde{q_\mu}$, and is given by:

$$\tilde{q_\mu} = -2\ \ln\frac{\mathcal{L}(\text{data}|\mu,\hat{\theta_\mu})}{\mathcal{L}(\text{data}|\hat{\mu},\hat{\theta})},\ \text{with}\ 0 \leq \mu \leq \hat{\mu} \tag{2.6}$$

,where $\hat{\theta_\mu}$ refers to the conditional maximum likelihood estimators of θ, i.e. the

set of nuisances parameters that maximize the likelihood for a given signal strength μ, while $\hat{\mu}$ and $\hat{\theta}$ refer to the global maximum likelihood estimators for μ and θ. The lower constraint on $\hat{\mu}$ i.e., $\hat{\mu} \geq 0$ ensures that the signal rate cannot be negative, while the upper constraint that $\hat{\mu}$, which is the global maximum value, cannot be less than the value of μ under consideration is imposed to guarantee that upward fluctuations of data such that $\hat{\mu} \geq \mu$ are not considered as evidence against the signal hypothesis,i.e., a signal of strength μ.

Now, using equation 2.6, the observed value of the test statistic,$\tilde{q_\mu}^{obs}$, is calculated for the signal strength μ. Also, maximum likelihood estimators for the nuisance parameters, for the background-only($\mu = 0$) and signal-plus-background(current $\mu >$ 0 under consideration) hypotheses are calculated. They are denoted by $\hat{\theta_0}^{obs}$ and $\hat{\theta_\mu}^{obs}$ respectively, and are used to generate toy Monte carlo pseudo-datasets. These pseudo datasets are used to construct pdfs, using equation 2.6, of test statistics $f(\tilde{q_\mu}|0, \hat{\theta_0}^{obs})$ and $f(\tilde{q_\mu}|\mu, \hat{\theta_\mu}^{obs})$ by treating them as they were real data. Example of these distributions are shown in Fig. 2.1.

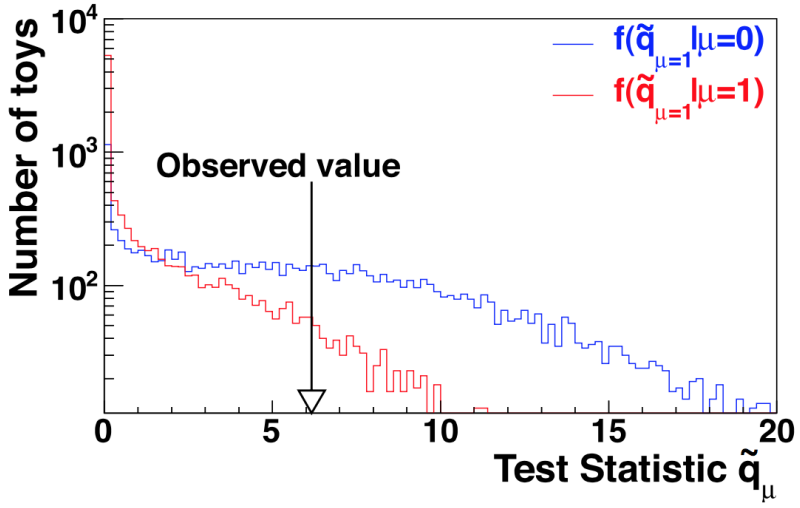


Figure 2.1: Test statistic distributions for ensembles of pseudo-data generated for signal-plus-background (red) and background-only (blue) hypotheses. [**?**]

Having constructed the above pdfs, it is now possible to calculate the probabilities of the observations under both hypotheses. The first quantity that we calculate is:

$$p_\mu = P(\tilde{q_\mu} \geq \tilde{q_\mu}^{obs} | \text{signal-plus-background}) = \int_{\tilde{q_\mu}^{obs}}^{\text{inf}} f(\tilde{q_\mu} | \mu, \hat{\theta_\mu}^{obs}) d\tilde{q_\mu} \tag{2.7}$$

The above quantity corresponds to $\text{CL}_{\text{s+b}}$ and measures the incompatibilty of data with signal-plus-background hypothesis. This quantity alone is not adequate for hypothesis testing in situations when the signal is so small that both hypotheses are compatible with the observation and a downward fluctuation of the background can lead to an inference of signal.

The second quantity we calculate is:

$$1 - p_b = P(\tilde{q_\mu} \geq \tilde{q_\mu}^{obs} | \text{background-only}) = \int_{\tilde{q_\mu}^{obs}}^{\text{inf}} f(\tilde{q_\mu} | 0, \hat{\theta_0}^{obs}) d\tilde{q_\mu} \tag{2.8}$$

This quantity corresponds to CL_{b} and measures the incompatibilty of data with the background. The incompatibilty of the data with background-only hypothesis alone doesn't tell us that it is indeed compatible with the signal, and so is not considered a good test of the signal hypothesis.

The ratio of the two quatities referred to as CL_{s} [**? ? ?**] helps deal with both situations above well, and is given by:

$$\text{CL}_{\text{s}} = \frac{p_\mu}{1 - p_b} \tag{2.9}$$

The 95% CL is then arrived at by iterating over μ until we have $\text{CL}_{\text{s}} = 0.05$. And the amount of signal or above, given by that μ, denoted as $\mu^{95\% CL}$, is said to be excluded at 95% CL.

2.2.4 Median expected Limits

Upper exclusion limits calculated using toy datasets of background-only expectation, are called expected limits. A large set of background-only pseudo-data is generated, and CL_s and $\mu^{95\%CL}$ is calculated for each of them. The median expected limit is calculated by integrating over this distribution until the 50% quantile is reached. The $\pm 1\sigma$ bands are calculated similary by integrating the distribution to the appropiate quantiles are reached. The calculation of median expected limits does not involve using the observed data and hence can be calculated when the analyses is blinded to prevent experimenter's bias (as mentioned in Section 1.1). This can be use to maximize the sensitivity of the search, as described in Sections 1.2.3 and 1.3.3. A more stringent(lower) median limit corresponds to a more sensitive search.

2.2.5 Quantifying an excess of events

In case an excess of data over background is observed, it is necessary to make sure beyond a reasonable doubt that the excess is not merely a fluctuation. This is quantified using the background-only p-value, which is the probability for the background to fluctuate and give an excess of events as large or larger than that observed. The same test statistic as equation 2.6 is used with the signal stength set to 0 to correspond to the background-only hypothesis:

$$\tilde{q_0} = -2 \ln \frac{\mathcal{L}(\text{data}|0, \hat{\theta}_0)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})}, \text{ with } 0 \leq \hat{\mu} \tag{2.10}$$

The constraint on $\hat{\mu}$ being greater than 0 is required so that a deficit of events in observed data is not interpreted in the same manner as we would an excess. In other words a departure from the background hypothesis in the form of deficit of events is not considered in favour of the signal hypothesis. Following the same procedure as calculation of observed limits 2.2.3 and generating pseudo-data, the distribution

$f(\tilde{q_0}|0,\hat{\theta_0}^{obs})$ is constructed. The p-value is then given by:

$$p_0 = P(\tilde{q_0} \geq \tilde{q_0}^{obs}) = \int_{\tilde{q_0}^{obs}}^{\inf} f(\tilde{q_0}|0,\hat{\theta_0}^{obs}) d\tilde{q_0} \tag{2.11}$$

The p-value can be converted to significance $\mathcal{Z}_0$, which is an equivalent way of quantifying an excess and is related to the p-value by the following:

$$p_0 = \int_{\mathcal{Z}_0}^{\inf} \frac{1}{\sqrt{2\pi}} exp(-x^2/2) dx \tag{2.12}$$

Broadly, the signficance corresponds to how far into the tail of the distribution (i.e., away from the most probable value), assuming background hypothesis, the test statistic value corresponding to the observed data lies. The farther it is, the less likely it is to have been a fluctuation. The conventional standard in high energy physics to be able to claim observation of a process is a significane of 5σ, which corresponds to a p-value of 2.8×10^{-7}.

2.2.6 Systematic uncertainties

It is important to consider all relevant sources of uncertainties when perfoming sophisticated counting experiments such as these. Uncertainties that introduced as a result of imprecise/inaccurate knowledge of the system or gaps in prior knowledge that is used in the measurement are called systematic uncertainties. They are a different class of uncertainties than those arising purely out randomness in statistical measurements, called statistical uncertainties. The sources of systematic uncertainties range from purely theoretical in nature to purely experimental. They can be categorized in the two following ways:

2.2.6.1 Normalization Uncertainties

The value of these uncertainties are independent of the signal/discriminant variable. To be more precise, these uncertainties are independent of the value of M_{col} or BDT response. Hence, they effect each bin of those distributions in exactly the same manner and thus change only the overall scale of the distribution without altering its shape.

The muons in the analysis are required to pass certain identification, isolation and triggering criteria (see chapter 1). The efficiencies for muon to pass these criteria are measured via tag-and-probe methods [22] using $Z \rightarrow \mu\mu$ events, and the scale factors are used to match the efficiency in MC to that in data. Briefly, the tag-and-probe method works in the following way. One of the muons (called the tag) is required to pass strict selection criterion, while the other (called the probe) is required to pass more relaxed criterion. Given the invariant mass of the $\mu - \mu$ system is required to within a very narrow window of the Z mass, the probe muon is also very likely to be a real muon. The percentage of probe muons that pass the criterion we are testing for (identification, isolation, trigger etc.) gives the efficiency. The efficiencies determined by this process, like any other quantity, are associated with systematic uncertainties. For the muons used in the analyses described here, a combined normalization uncertainty of 2% is associated with muon trigger, identification and isolation. Similar to muons, the efficiencies for electrons used in the analyses have also been measured via tag-and-probe methods [24] using $Z \rightarrow$ ee events. The uncertainties in efficiencies of electron identification and isolation criterion are also included as normalization uncertainty of 2% in the fit. Both the above uncertainties are applied to processes which are derived from MC simulation. As mentioned earlier, a b tagging veto is applied to the analysis in order supress backgrounds involving top quarks. The efficiency of b tagging procedure is different in MC simulation than data. A scaling procedure is applied to match these efficiencies, and the uncertainties associated with these factors

are found to not effect the shape of the M_{col} or BDT distributions. They are thus included in the fit as normalization uncertainties and range across categories from 2-4.5% and 2-2.5% for h $\to \mu\tau_e$ and H $\to \mu\tau_e$ analysis respectively.

Several backgrounds in the analyses are estimated using MC simulations (see chapter **??**). These include Z $\to \tau\tau$,$t\bar{t}$,W+jets, WW, WZ and ZZ, $Z \to \ell\ell$ (ℓ = e, μ)+jets, single-top quark production, W$\gamma^{(*)}$ + jets. The production cross-sections for these backgrounds determine the number of events each background would contribute. These cross-sections are measured experimentally and the uncertainty in those measurements are included in the fit. Given that a change in cross-section changes the overall number of events produced, it has no effect on the shape of distributions. Hence these uncertainties are included as normalization uncertainties. These uncertainties in general arise from: uncertainties on the parton distribution functions and strong coupling constant (called PDF+α_s); variations in renormalization and factorization scales. In the H $\to \mu\tau_e$ analysis a applied separate uncertainty is applied for PDF+α_s and renormalization/factorization scales for each of the backgrounds. In h $\to \mu\tau_e$ analysis, a combined uncertainty for each background is applied to cover both sources. All the above uncertainties are considered 100% correlated among categories. For each background, an 5% uncertainty, uncorrelated among all categories, is also applied to conservatively cover differences across categories. The QCD multi-jet background is estimated using a data-driven procedure. An uncertainty of 30% associated with this (corresponding to the uncertainty in the extrapolation factor from the same-sign to opposite-sign region) is included in the fit. All uncertainties are summarized in table 2.1.

Just like MC backgrounds described above, the signal in both h $\to \mu\tau_e$ and H $\to \mu\tau_e$ analysis comes from MC simulation. The signal process and the SM Higgs background are associated with uncertainties in the Higgs boson production cross sections. These come from variations in factorization/renormalization scales, as well

Table 2.1: The systematic uncertainties for the four channels. All uncertainties are treated as correlated between the categories, except those with more values separated by the $\oplus$ symbol. In the case of two values, the first value is the correlated uncertainty and the second value is the uncorrelated uncertainty for each individual category. In the case of three values, the first and second values correspond to the uncertainties arising from factorization and renormalization scales and PDF variations and are correlated between categories, while the third value is the uncorrelated uncertainty for each individual category. Two values separated by the "–" sign represent the range of the uncertainties from the different sources and/or in the different jet categories.

Systematic uncertainty	$h \to \mu\tau_e$	$H \to \mu\tau_e$
Muon trigger/ID/isolation	2%	2%
Electron trigger/ID/isolation	-	2%
b tagging veto	2.0–2.5%	2.0–2.5%
QCD multijet background	-	30%
$Z \to \tau\tau$ + jets background	0.1%⊕2%⊕5%	0.1%⊕2%⊕5%
$t\bar{t}$ background	10%⊕5%	10%⊕5%
W + jets background	-	
[] WW, ZZ, WZ background		
$W\gamma^{(*)}$ + jets background	-	10%⊕5%
Single top quark background	3%⊕5%⊕5%	–
$Z \to \ell\ell$ ($\ell = e, \mu$) + jets background	-	0.1
Jet energy scale	3–20%	3–20%
μ energy scale	0.2%	0.2%
e energy scale	-	0.1–0.5%
Unclustered energy scale	$\pm 1\sigma$	$\pm 1\sigma$
Integrated luminosity	2.5%	2.5%

as PDF+α_s, and result in changes only in normalization. These uncertainties are summarized for SM Higgs boson and heavier Higgs of different masses in table 2.1. They are taken from Handbook of LHC Higgs cross-sections found in Ref. [**?**].

Table 2.2: Theoretical uncertainties from [**?**] are applied to the Higgs boson production cross sections for the different masses. In the reference, the PDF and α_s uncertainties are computed following the recommendation of the PDF4LHC working group. The remaining Gaussian uncertainty accounts for additional intrinsic sources of theory uncertainty described in detail in the reference.

m_{H} (GeV)	Cross section (pb)	Theory, Gaussian (%)	PDF+α_s (%)
200	16.94	±1.8	±3.0
300	6.59	±1.8	±3.0
450	2.30	±2.0	±3.1
600	1	±2.1	±3.5
750	0.50	±2.1	±4.0
900	0.27	±2.2	±4.6

The estimation of a particular background, that is derived from simulation, needs to correspond to the number of events (of that background, having a particular cross-section) that would be produced in the amount proton-proton collision data that we are using for this search. In other words, the background estimations need to be normalized to the integrated luminosity of the data collected. Integrated luminosity (defined in chapter **??** is a measured quantity, and like all measured quantities, has an uncertainty associated with it. This amounts to 2.5% and, like other normalization unceratainties, only effects only the overall scale of distributions.

APPENDIX A

BOOSTED DECISION TREES

A.1 Introduction

BIBLIOGRAPHY

1. Vardan Khachatryan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9, 2014. doi: 10.1088/1748-0221/9/10/P10009.

2. Andy Buckley et al. General-purpose event generators for lhc physics. *Physics Reports*, 504:243, July 2011. doi: 10.1016/j.physrep.2011.03.005.

3. Wikipedia. Monte carlo method. Website. `https://en.wikipedia.org/wiki/Monte_Carlo_method`.

4. F. Krauss et al. J. Alwall, S. Hche. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *EPJC*, 53, 2008.

5. Leif Lonnblad. Correcting the color dipole cascade model with fixed order matrix elements. *JHEP*, 0205, 2002.

6. Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *JHEP*, 12:061, 2012. doi: 10.1007/JHEP12(2012)061.

7. Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852, 2007. doi: 10.1016/j.cpc.2008.01.036.

8. Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release notes. *EPJC*, 76, Apr 2016. doi: 10.1140/epjc/s10052-016-4018-8.

9. Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004. doi: 10.1088/1126-6708/2004/11/040.

10. Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 11:070, 2007. doi: 10.1088/1126-6708/2007/11/070.

11. Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010. doi: 10.1007/JHEP06(2010)043.

12. Simone Alioli, Keith Hamilton, Paolo Nason, Carlo Oleari, and Emanuele Re. Jet pair production in POWHEG. *JHEP*, 04:081, 2011. doi: 10.1007/JHEP04(2011)081.

13. Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO Higgs boson production via gluon fusion matched with shower in POWHEG. *JHEP*, 04:002, 2009. doi: 10.1088/1126-6708/2009/04/002.

14. E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini. Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM. *JHEP*, 02:088, 2012. doi: 10.1007/JHEP02(2012)088.

15. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: 10.1007/JHEP07(2014)079.

16. Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5: going beyond. *JHEP*, 06:128, 2011. doi: 10.1007/JHEP06(2011)128.

17. S. Agostinelli et al. GEANT4 — a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003. doi: 10.1016/S0168-9002(03)01368-8.

18. Gionata Luisoni, Paolo Nason, Carlo Oleari, and Francesco Tramontano. $HW^{\pm}/HZ$ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. *JHEP*, 10:083, 2013. doi: 10.1007/JHEP10(2013)083.

19. A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12:P10003, 2017. doi: 10.1088/1748-0221/12/10/P10003.

20. Rudolf Frhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth. A*, 262:444, 1987. doi: 10.1016/0168-9002(87)90887-4.

21. Rudolf Frhwirth Wolfgang Waltenberger and Pascal Vanlaer. Adaptive vertex fitting. *J. Phys. G*, 34, 2007.

22. A. M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13:P06015, 2018. doi: 10.1088/1748-0221/13/06/p06015.

23. Serguei Chatrchyan et al. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012. doi: 10.1088/1748-0221/7/10/P10002.

24. Vardan Khachatryan et al. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *JINST*, 10: P06005, 2015. doi: 10.1088/1748-0221/10/06/P06005.

25. Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008. doi: 10.1088/1126-6708/2008/04/063.

26. Serguei Chatrchyan et al. Determination of jet energy calibration and transverse momentum resolution in CMS. *JINST*, 6:11002, 2011. doi: 10.1088/1748-0221/6/11/P11002.

27. Vardan Khachatryan et al. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12:P02014, 2017. doi: 10.1088/1748-0221/12/02/P02014.

28. Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The catchment area of jets. *JHEP*, 04:005, 2008. doi: 10.1088/1126-6708/2008/04/005.

29. Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, 659:119, 2008. doi: 10.1016/j.physletb.2007.09.077.

This document was prepared & typeset with pdfLaTeX*, and formatted with* NDdiss2$_\varepsilon$ *classfile (v3.2017.2[2017/05/09])*