This Dissertation

entitled

Search for lepton flavor violating decays of the Higgs boson

typeset with NDdiss2$_\varepsilon$ v3.0 (2005/07/27) on January 9, 2018 for

Fanbo Meng

This LaTeX 2$_\varepsilon$ classfile conforms to the University of Notre Dame style guidelines established in Spring 2004. 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.

It is YOUR responsibility to ensure that the Chapter titles and Table caption titles are put in CAPS LETTERS. This classfile does *NOT* do that!

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, but MUST be turned in to the proofreader(s) or the reviwer(s)!

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

`http://www.gsu.nd.edu`
`http://graduateschool.nd.edu`

Search for lepton flavor violating decays of the Higgs boson

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

by

Fanbo Meng,

Colin Jessop, Director

, Director

Graduate Program in Physics
Notre Dame, Indiana
January 2018

Search for lepton flavor violating decays of the Higgs boson

Abstract

by

Fanbo Meng

this going to put with the abstract, a summary of the whole analysis

NEW DEDICATION NAME

To be written

CONTENTS

FIGURES

TABLES

PREFACE

probably I will write some preface, but it depends $H \to \mu\tau_h$

ACKNOWLEDGMENTS

this part is going to be filled with acknowledge

CHAPTER 1

Introduction

1.1 Pre-LHC era on lepton Flavor violation(LFV) search and results

1.2 Overview of CMS Run1 and Run2 direct search on LFV

CHAPTER 2

Theory

2.1 Standard model

2.2 LFV in beyond standard model theories

CHAPTER 3

LHC and CMS experiment

3.1 LHC accelerator

3.2 CMS experiment

3.2.1 Tracker

3.2.2 ECAL

3.2.3 HCAL

3.2.4 Muon System

3.2.5 Trigger

probably I have to put something there

3.2.5.1 Level 1 Trigger

probably I have to put something there2

3.2.5.2 High Level Trigger

CHAPTER 4

Datasets

4.1 Datasets used in LFV analysis

4.2 Event reconstruction

4.2.1 Muon reconstruction and selection criteria

With the PF muon as the input, muons used in the analysis are further categorized into different identification, isolation categories. Muon ID suggested in CMS RunII, for the LHC data 2016, running period BCDEF, ICHEP medium muon ID is applied(table 4.1), for running period G and H, also the monte Carlo samples, standard medium muon ID(table 4.2)are applied to achieve the best performance for muon identification.

4.2.2 Tau lepton reconstruction

In Run I CMS experiment, tau lepton are constructed with hadrons plus strips(HPS) algorithm. In general, HPS starts with PF jets which are reconstructed with $anti - k_T$, as the initial seeds. π_0 components from the τ hadronic decays are first constructed and combined with the charge hadrons parts, to identify different τ decay modes and calculate τ four-momentum and other quantities [1].

TABLE 4.1

Muon ID used in the analysis, for the LHC data 2016, running period BCDEF.

ICHEP mediumID description	Technical description
Loose muon ID	PFLoose Muon
Fraction of valid tracker hits	> 0.49
1.Good Global muon	Global muon
	Normalized global-track $\chi^2 < 3$
	Tracker-Standalone position match < 12
	kick finder < 20
	Segment compatibility > 0.303
2. Tight segment compatibility	Segment compatibility > 0.451

TABLE 4.2

Muon ID used in the analysis, for the LHC data 2016, running period G and H, also the monte Carlo samples.

Standard mediumID description	Technical description
Loose muon ID	PFLoose Muon
Fraction of valid tracker hits	> 0.8
1.Good Global muon	Global muon
	Normalized global-track $\chi^2 < 3$
	Tracker-Standalone position match < 12
	kick finder < 20
	Segment compatibility > 0.303
2. Tight segment compatibility	Segment compatibility > 0.451

Photon conversions and the bremsstrahlung of electron/positron when traveling inside the CMS detector are well treated by the HPS algorithm. These phenomenons broaden the signature of the tau decay. With PF jets as input, the algorithm constructs strips out of electromagnetic particles and starts by taking the strip in which contains the most energetic electromagnetic particle as the center one. With the center strip, a window of the size $\Delta\eta = 0.05$ and $\Delta\phi = 0.2$ is taken. Within this window, if other charged particles are found, they are associated with the strip. The position of the strip is taken and four momentum of the strip is calculated. This procedure is repeated, until no strips can be constructed. The selected strips are required to have $P_T^{strip} > 1GeV$. The following decay topologies are taking into account by HPS:

- one charged particle without any strip, $h^{\pm}$ and the case when π^0 is not energetic enough to form a strip
- one charged particle plus one strip
- one charged particle plus two strips
- three charged partibles.

All of the charged hadrons and strips are required to be contained in the $\Delta R = 2.8/P_T^{\tau_h}$ core, where the $P_T^{\tau_h}$ is the reconstructed τ_h transverse momentum and ΔR is defined as $\Delta R = \sqrt{(}\Delta\phi^2 + \Delta\eta^2)$. The τ_h candidate is also required to match the direction of the seed PF jet within $\Delta R = 0.1$. Assuming all of the charged hadrons to be pions and taking in the associated strips, the HPS algorithm requires that different decay topologies meet the intermediate meson mass as listed in Table. 4.3. Tau lepton isolation in the cut based criterion is determined by if there are extra charged hadrons or photons besides the ones from τ_h in the cone

$\Delta R = 0.5$ in τ_h direction. Isolation cuts differ in which energy of the photons and hadrons are considered, ending up in loose, medium and tight level of cuts. The exact of the energy selection is suggested by the study of QCD dijet events. The loose cut brings in approximate 1% of fake τ from jets.

TABLE 4.3

Dominant hadronic τ lepton decays branching fractions and the associated intermediate resonance. The h stands for both π and K. The table is symmetric under charge conjugation.

Decay mode	Resonance	Mass (MeV/c^2)	Branching fraction(%)
$\tau^- \to h^- v_\tau$			11.6%
$\tau^- \to h^- \pi^0 v_\tau$	ρ^-	770	26.0%
$\tau^- \to h^- \pi^0 \pi^0 v_\tau$	α_1^-	1200	9.5%
$\tau^- \to h^- h^+ h^- v_\tau$	α_1^-	1200	9.8%
$\tau^- \to h^- h^+ h^- \pi^0 v_\tau$			4.8%

In CMS RunII, Tau reconstruction algorithm HPS has been improved [2]. The major improvement lies in Dynamic strip instead of fix size strip. Tau decay products can also affect the isolation. Charged pions in tau decay products experience nuclear interaction with tracker materials, which can results in low P_T

secondary particles. Photons from the neutral pion decay can also go through pair production into e^+e-, which further spread because of bremsstrahlung and the magnetic field. Broadening the strip is need in these cases in order to better cover the tau decay production. On the other hand, if the tau is boosted, high P_T decay products tends to be more concentrate and smaller strip size will be better. Similar to RunI tau reconstruction, the algorithm starts with hightest P_T charged particle as seeds for the strip. Starting from the seed strip, a window in η and ϕ direction is set.

$$\delta\eta = f(P_T^{\gamma}) + f(P_T^{strip}) \qquad f(P_T) = 0.2 \cdot P_T^{-0.66}$$
$$\delta\phi = g(P_T^{\gamma}) + g(P_T^{strip}) \qquad g(P_T) = 0.35 \cdot P_T^{-0.71}$$

The window is determined from single τ gun MC simulation. 95% of the decay product will be covered in that range. The upward and downward limits for η is 0.15 and 0.05, for ϕ the range is 0.3 and 0.05. The position of strip is set as P_T weighted average against all of the objects.

$$\eta_{strip} = \frac{1}{P_T^{strip}} \cdot \sum P_T^{\gamma} \cdot \eta_{\gamma}$$
$$\phi_{strip} = \frac{1}{P_T^{strip}} \cdot \sum P_T^{\gamma} \cdot \phi_{\gamma}$$

Construct the strip until no seed strip can be found. After the construction

of the τ lepton, for different decay mode, m_τ is required to lie in different mass windows [3]. The conditions of different hadronic decay mode mass window are listed in the Table. 4.4. With respect to RunI, the difference in mass window is δm, which originates from dynamic clustering. δm is calculated as:

$$\delta m = \sqrt{\Big(\frac{\partial m_\tau}{\partial \eta_{strip}} \cdot f(P_T^{strip})\Big)^2 + \Big(\frac{\partial m_\tau}{\partial \phi_{strip}} \cdot g(P_T^{strip})\Big)^2}$$

with:

$$\frac{\partial m_\tau}{\partial \eta_{strip}} = \frac{P_z^{strip} \cdot E_\tau - E_{strip} \cdot P_z^\tau}{m_\tau}$$
$$\frac{\partial m_\tau}{\partial \phi_{strip}} = \frac{-(P_y^\tau - P_y^{strip}) \cdot P_x^{strip} + (P_x^\tau - P_x^{strip}) \cdot P_y^{strip}}{m_\tau}$$

In current algorithm, $\tau^- \rightarrow h^- h^+ h^- v_\tau$ is not included, because of the jets contamination. This hadronic τ decay mode composed of 4.8% of total branching fraction. The $h^-\pi^0$ and $h^-\pi^0\pi^0$ are analyzed together, which is referred as $h^-\pi^0$.

The analysis with 2016 datasets, MVA based τ isolation criteria is used, which keeps high identification efficiency while maintains relatively low fake rate compared with cut based criteria.

4.3 Event simulation

TABLE 4.4

τ hadronic decay mode hypothesis signatures compatibility tests. m_τ is required to be in the mass window

Decay mode	Mass window
$\tau^- \to h^-\pi^0 v_\tau$	$0.3 - \delta m_\tau < m_\tau < 1.3 \cdot \sqrt{P_T/100} + \delta m_\tau$
$\tau^- \to h^-\pi^0\pi^0 v_\tau$	$0.4 - \delta m_\tau < m_\tau < 1.2 \cdot \sqrt{P_T/100} + \delta m_\tau$
$\tau^- \to h^-h^+h^- v_\tau$	$0.8 - \delta m_\tau < m_\tau < 1.5 + \delta m_\tau$

CHAPTER 5

LFV analysis background estimation

CHAPTER 6

LFV event selection

For both 8TeV analysis $H \to e\tau_h$ channel and 13TeV analysis $H \to \mu\tau_h$, events are selected in several steps. The loose selection on the different IDs, energy, geometry parameters of the analysis related objects are applied. In both $H \to e\tau_h$ and $H \to \mu\tau_h$ analysis, cut-based analyses are applied. In $H \to \mu\tau_h$, a multivariate analysis with Boosted decision tree (BDT) is exploited to provide more sensitive results.

6.1 $H \to \mu\tau_h$

6.1.1 Loose selection

In $H \to \mu\tau_h$ events, tau leptons from signal events decay hadronically. SM higgs is much heavier than the LFV decay products μ and τ, so μ and τ are expected to have high P_T. Since the decay products are boosted, a cut on the $\Delta R > 0.3$ is applied. ΔR is defined as $\Delta R = \sqrt{(}\Delta phi^2 + \Delta eta^2)$. Higgs has no charge, so μ and τ candidates are required to have opposite sign of charges. Further, the events with additional μ and τ that pass a loose selection, the events with jets that are identified by the combined secondary vertex(CSVv2) b-tagging algorithm [4] as a b quark jets will be vetoed. Muons from signal events are boosted and isolated, as mentioned in previous chapter, the trigger HLT_IsoMu24

or HLT_IsoTkMu24 is used. The trigger select isolated muons that have energy higher than 24GeV. An further P_T cut on reconstructed μ, $P_T > 26$GeV and $|\eta| < 2.4$ are applied. Muons are required to pass the recommended Medium muon ID(chapter 4.2). For the LHC data 2016, running period BCDEF, ICHEP medium muon ID is applied(table 4.1), for data running period G and H, also the monte Carlo samples, standard medium muon ID(table 4.2)are applied to achieve the best performance for muon identification.

Hadronic taus are required to have $P_T > 30$GeV,$|\eta| < 2.3$, passing old tau decay mode finding(chapter 4.2), an MVA based tight tau isolation ID(Chapter 4.2) and tau discriminators against electrons and muons. These discriminators are very loose MVA based rejection against electrons and cut-based tight rejection against muons.

Events in the analysis are divided into four categories based on the number of jets in an event. In 2-jets category, it is furthered divided into 2 categories, 2-jet gluon gluon fusion higgs production(ggH) category and 2-jet vector boson fusion(VBF) category based on the value of 2 jets invariant mass(M_{jj}) . In 0-jet category, the signal mainly comes from ggH. In 1-jet category, the dominant signal production mode is also ggH, but with a boosted jet associated with the production, some of the VBF higgs signal also shows up in this category. In the 2-jet ggH category, signal evens mainly come from ggH and in 2-jet VBF, VBF production dominants the production mode. The following is a more detailed list of the selection condition in each categories.

0-jet: No events have jets pass the loose PF ID and with jet $P_T > 30$ GeV, $|\eta| < 4.7$.

1-jet: Events with one jet passes losse PF ID and jet $P_T > 30$ GeV, $|\eta| < 4.7$.

2-jets ggH: Events have two jets passing loose PF ID, $P_T > 30$ GeV, $|\eta| < 4.7$ and a

requirement on the invariant mass of the two jets, $M_{jj} < 550GeV$.

2 jets VBF: Events with two jets pass loose PF ID. Jets $P_T > 30$ GeV, $|\eta| < 4.7$ and $M_{jj} > 550GeV$ are required.

The threshold on M_{jj} has been optimized to give the best expected exclusion limits.

6.1.2 Cut-based analysis

With the loose selection, including the categorization, a further cut-based selection strategy is applied. Variables that can help distinguish signal from background used in this analysis are P_T^{μ}, P_T^{τ} and $M_T(\tau_h)$. The lepton P_T variables are very powerful background discriminant variables, but it will also cause the problem that signal picks under the background. Leptons from signal process incline to have higher P_T values, by cutting tighter on the lepton P_T, more background events can be removed. However this will also reshape some of the backgrounds, making them peak closer under the signal so that signal processes will be affected more by the background statistics fluctuation. In the $H \to e\tau_h$ analysis, the effect of cutting hard on lepton P_T will be shown. So in $H \to \mu\tau_h$ search, lepton P_T variables are kept at loose values and tune on other variables to achieve better signal significance.

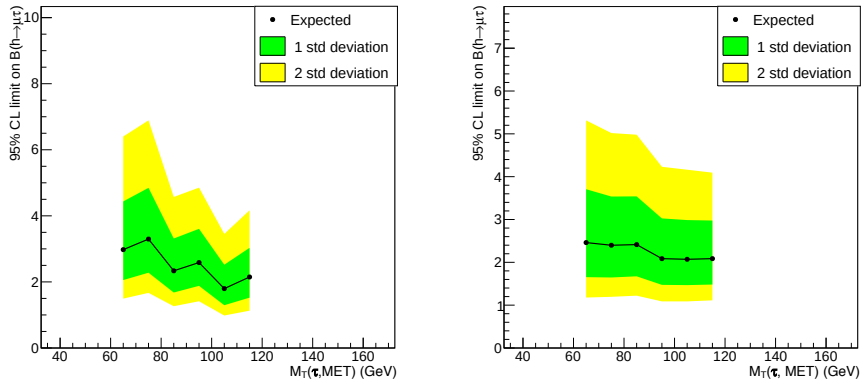
Tuning process is done only with Monte Carlo samples to not double use data. The final results are extract from data, so the double use of data may cause biases on the results. Misidentified lepton background estimated with full data driven method, is replaced with the semi data driven estimation in tuning. In $H \to \mu\tau_h$ channel, the variables tuned are M_{jj} and $M_T(\tau_h)$. Cuts have been optimized to have the most stringent expected limits with the Asimov dataset. If loose value

of the cut gives same expected limits as the tight one, then chose the loose cut value to have more statistics. Examples of obtaining the limits are shown in Fig. 6.1, $M_T(\tau_h)$ for the different categories and Fig. 6.2 for the optimization of M_{jj} in 2-jets categories. With this method and criteria, the cut optimized for $H \rightarrow \mu\tau_h$ is shown in Table. 6.1

6.1.3 Multivariate analysis

A Boost decision trees(BDT) method is chosen as the multivariate analysis method used in $H \rightarrow \mu\tau_h$ search. It provides better sensitivity compared with the cut-based analysis. In this analysis, the BDT is provided by the TMVA package [5]. BDT method takes in signal and background datasets with a selected set of input variables. Input variables are the ones that show distinguishing power between signal and background. The training output is a weight file, which contains a list of weights to indicate in percentage how likely an event is signal like with a give set of input variable values from that event. A more detail description of the BDT method is available in section 7.1. In this analysis, signal and background events are required to pass the loose selection criteria. All of the categories are combined. The signal events from gluon gluon fusion and vector boson fusion higgs production mode are mix by the weight with respect to their production cross section. The background sample used in the training is misidentified lepton background from the like sign region(Region II as in table **). The list of BDT input variables is following and the distribution of the variables are shown in Fig. 6.3.

- Transverse mass between the τ_h and E_T^{miss}, $M_T(\tau_h)$.



(a) 0 jet

(b) 1 jet

(c) 2 jets, gg-enriched

(d) 2 jets, VBF-enriched

Figure 6.1. Expected limits based on an Asimov dataset as a function of $M_T(\tau, MET)$ for the different categories.

TABLE 6.1

Selection criteria for each event category after cut optimization, for the $H \to \mu\tau_h$ channel

0-jet category
• $P_T^{\mu} > 26GeV$, $P_T^{\tau} > 30GeV$
• $M_T(\tau) < 105GeV$
• No jets with $P_T^{jet} > 30GeV$, $\|\eta\| < 4.7$, LooseID
1-jet category
• $P_T^{\mu} > 26GeV$, $P_T^{\tau} > 30GeV$
• $M_T(\tau) < 105GeV$
• One jet with $P_T^{jet} > 30GeV$, $\|\eta\| < 4.7$, LooseID
2-jet, gg-enriched category
• $P_T^{\mu} > 26GeV$, $P_T^{\tau} > 30GeV$
• $M_T(\tau) < 105GeV$
• $P_T^{jet1} > 30GeV$,$P_T^{jet2} > 30GeV$ $\|\eta_{jet1}\| < 4.7$,$\|\eta_{jet2}\| < 4.7$, LooseID
• $M_{jj} < 550GeV$
• Two jets with $P_T^{jet} > 30GeV$, $\|\eta\| < 4.7$, LooseID
2-jet, vbf-enriched category
• $P_T^{\mu} > 26GeV$, $P_T^{\tau} > 30GeV$
• $M_T(\tau) < 85GeV$
• $P_T^{jet1} > 30GeV$,$P_T^{jet2} > 30GeV$ $\|\eta_{jet1}\| < 4.7$,$\|\eta_{jet2}\| < 4.7$, LooseID
• $M_{jj} > 550GeV$
• Two jets with $P_T^{jet} > 30GeV$, $\|\eta\| < 4.7$, LooseID

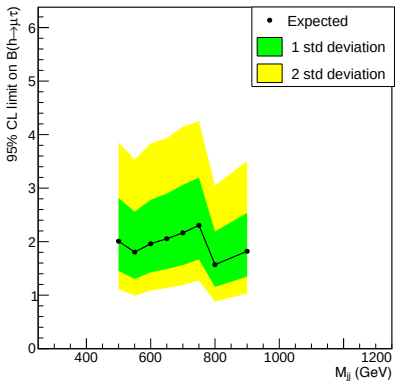


Figure 6.2. Expected limits based on an Asimov dataset as a function of M_{jj} for the 2 jet categories.

- Missing transverse energy, E_T^{miss}.
- Pseudorapidity difference between the μ and the τ_h candidate, $\Delta\eta(\mu, \tau_h)$.
- Azimutal angle between the μ and the τ_h, $\Delta\phi(\mu, \tau_h)$.
- Azimutal angle between the τ_h and the E_T^{miss}, $\Delta\phi(\tau_h, \mathrm{E}_T^{miss})$.
- Collinear mass, M_{col}.
- Muon P_T.
- τ_h P_T.

Fig. 6.4 shows BDT input variables in the signal and background training samples. The chosen input variables show low correlation in both samples. Fig. 6.5 shows the TMVA overtraining checks. Training samples are divided into two groups, one for training and one for testing. After the training, the algorithm applies the training output to the testing sample to check if they are in good matching.

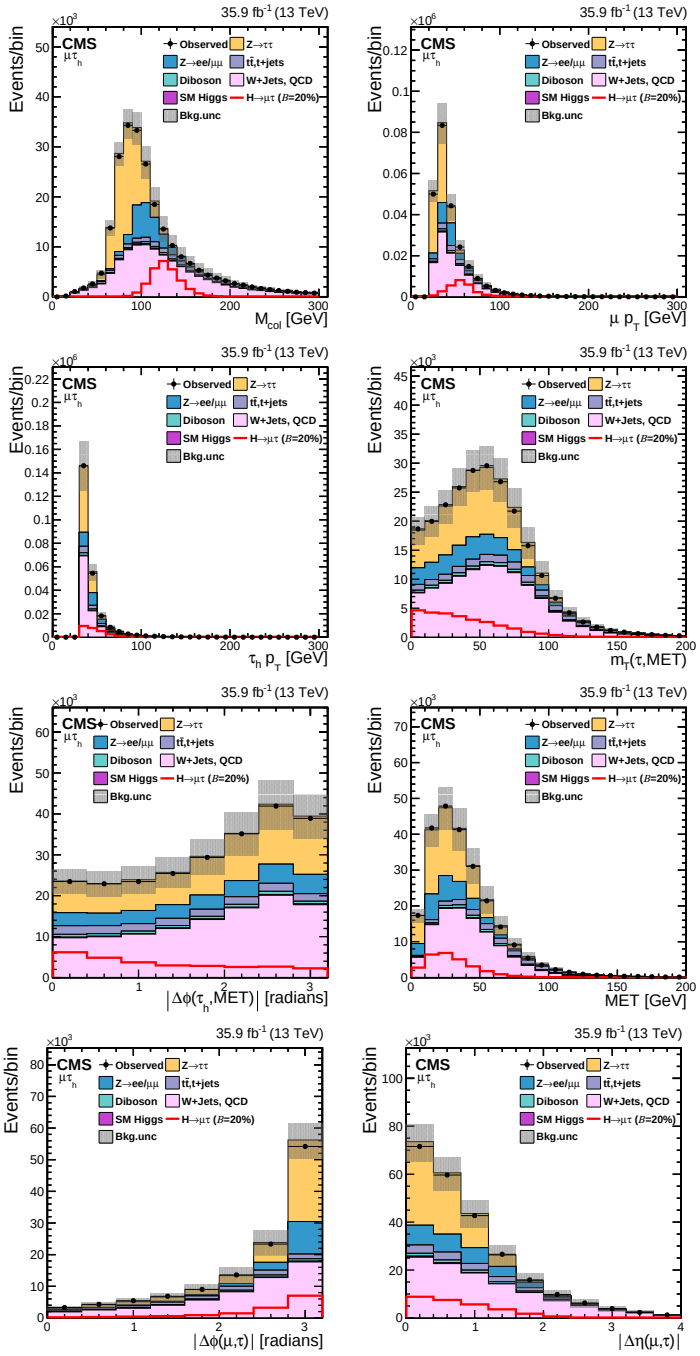
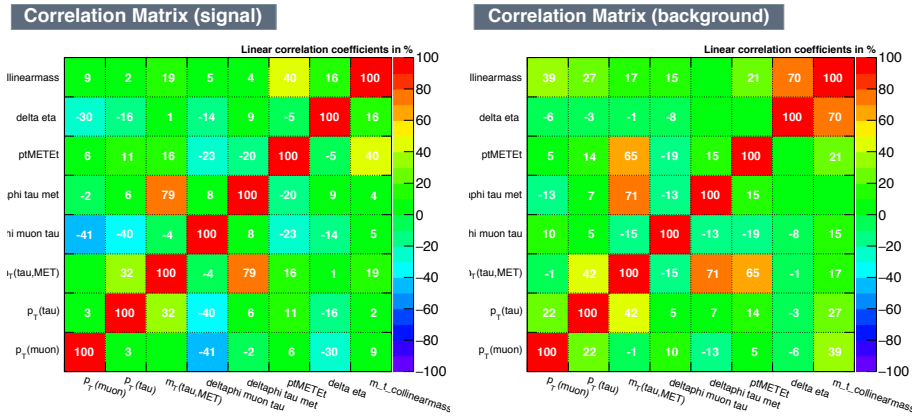


Figure 6.3. Distributions of the input variables to the BDT for the $H \to \mu\tau_h$ channel.



(a) Signal sample variables correlation (b) Background sample variable correlation

Figure 6.4. Expected limits based on an Asimov dataset as a function of $M_T(\tau, MET)$ for the different categories.

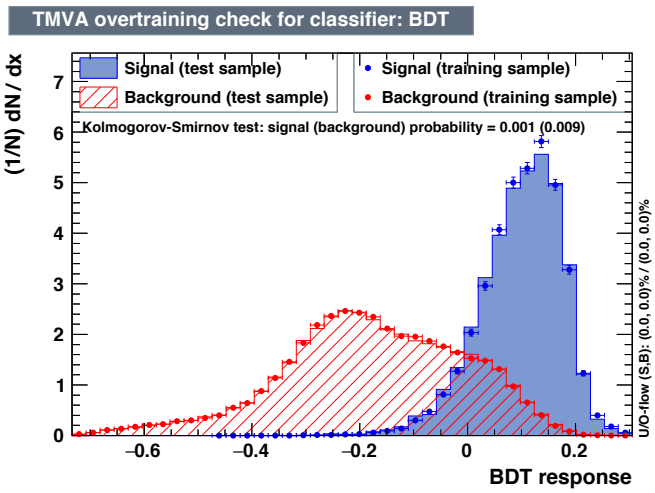


Figure 6.5. Overtraining checking for the BDT training in the TMVA package.

6.2 $H \to e\tau_h$

6.2.1 Loose selection

In H $\to e\tau_h$ channel, the trigger used is HLT_Ele27_WP80, which applies an electron P_T cut at 27GeV at HLT level. A further cut on electron $P_T > 30GeV$ is applied. Electrons are required to have $|\eta_e < 2.3|$ and $D_z < 0.2cm$. D_z is the longitudinal impact parameter that shows the displacement between primary vertex and track path. Electrons are also requirement to pass the MVA based tight ID and cut based PF tight isolation $PFiso < 0.1$. Tau candidates are required to have $P_T > 30GeV$, pseudorapidity in the range $|\eta^\tau < 2.3|$ and the longitudinal impact parameter $D_z < 0.2cm$. Tau isolation used is the cut based tight tau isolation. In addition, tau candidates pass the tau Decay mode finding and tau discriminator against electrons and muons. The analysis also requires no extra isolated electrons with $P_T > 10GeV$ and extra taus with $P_T > 20GeV$. Tau and electron candidates are required to have opposite sign of charges and separate with an $\Delta R > 0.4$ from any jets in the events with $P_T > 30GeV$. All of the requirements contribute to the selections of good qualities candidates. The datasets are binned into three categories according to the number of jets in the events:

0-jet: No events have jets pass the loose PF ID and with jet $P_T > 30GeV$, $|\eta| < 4.7$. This category enhances the gluon-gluon fusion contribution.

1-jet: Events with one jet passes loose PF ID and jet $P_T > 30GeV$, $|\eta| < 4.7$. This category enhances the gluon-gluon fusion production with initial state radiation.

2 jets: Events with two jets pass loose PF ID and with jet $P_T > 30$ GeV and

$|\eta| < 4.7$, This category contains both Higgs production mode and with an enhancement in VBF production mode.

With the preselection and binning of the jets numbers, the M_{col} distribution of H $\rightarrow e\tau_h$ in different categories are shown in Fig. 6.6

6.2.2 Cut-based analysis

A set of kinematics variables are defined and set to further select signal events. In H $\rightarrow e\tau_h$ channel, similar to the H $\rightarrow \mu\tau_h$, muon and tau leptons from the signal events are highly boosted, so as P_T variables plays an important role in distinguishing from background events and the separation in ϕ direction is bigger between μ and τ for signal events. Missing energy from the τ neutrino in hadronic decay is one of the characters in signal events, so $M_T(\tau_h)$ is an useful variable. In two jets category, M_{jj} also used as a cut variable. The cuts have been optimized to have the most stringent expected limits with the Asimov dataset. The detailed cuts used for H $\rightarrow e\tau_h$ is shown in table. 6.2.

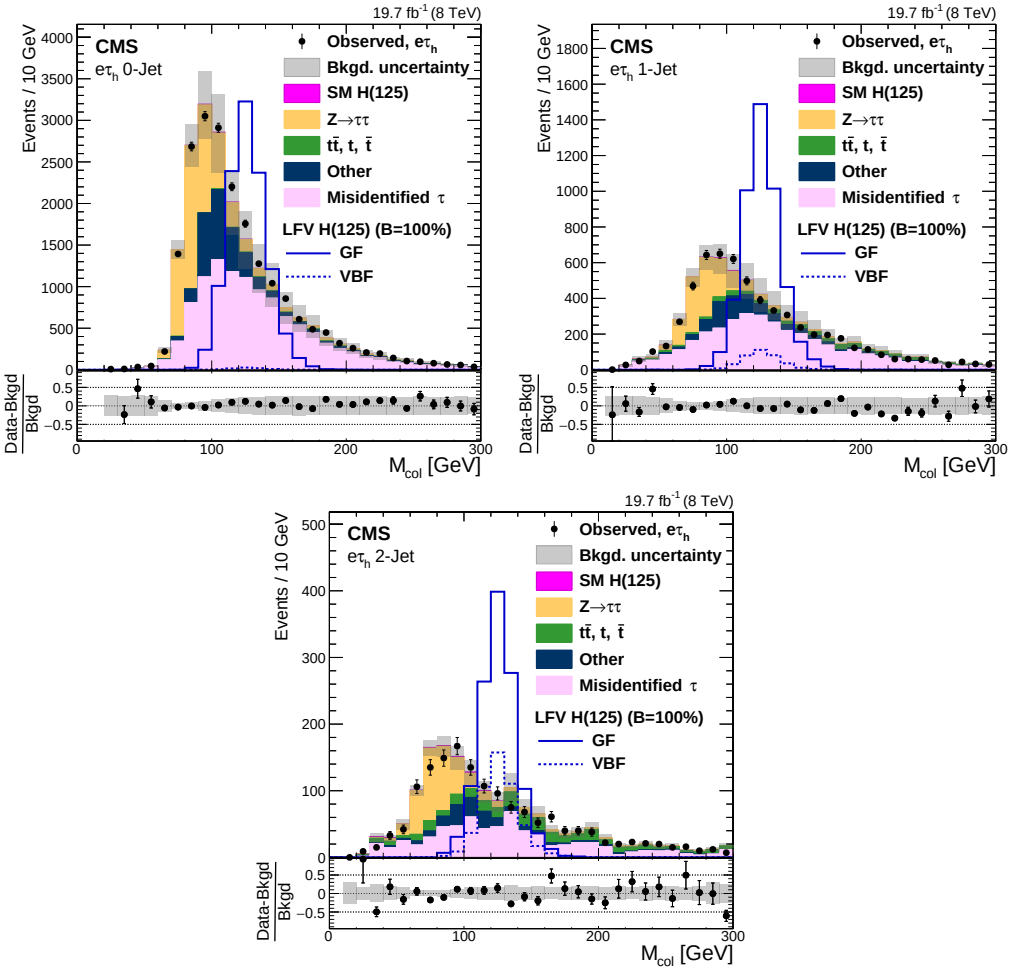


Figure 6.6. With loose selection conditions, the comparison of the observed collinear mass distributions with background from prediction. The shaded grey bands indicate the total background uncertainty. The open histograms correspond to the expected signal distributions for $\mathcal{B}(\mathrm{H} \to e\tau_h) = 100\%$ in the 0-jet, 1-jet and 2-jet categories, respectively.

TABLE 6.2

Selection criteria for each event category after cut optimization, for the $H \to e\tau_h$ channel

0-jet category
• $P_T^e > 45GeV$, $P_T^\tau > 30GeV$, $\Delta\phi_{e\tau} > 2.3$
• $M_T(\tau) < 70GeV$
• No jets with $P_T^{jet} > 30GeV$, $\|\eta\| < 4.7$, LooseID
1-jet category
• $P_T^e > 35GeV$, $P_T^\tau > 40GeV$
• One jet with $P_T^{jet} > 30GeV$, $\|\eta\| < 4.7$, LooseID
2-jet category
• $P_T^e > 35GeV$, $P_T^\tau > 30GeV$
• $M_T(\tau) < 50GeV$
• $P_T^{jet1} > 30GeV$,$P_T^{jet2} > 30GeV$ $\|\eta_{jet1}\| < 4.7$,$\|\eta_{jet2}\| < 4.7$, LooseID
• $\Delta\eta(jet1, jet2) > 2.3$
• $M_{jj} > 400GeV$
• Two jets with $P_T^{jet} > 30GeV$, $\|\eta\| < 4.7$, LooseID

CHAPTER 7

Signal extraction and systematics

7.1 Boost decision trees method

7.2 statistical methods

7.3 Nuisance variables

CHAPTER 8

LFV Higgs decay searching results

CHAPTER 9

Conclusion

CHAPTER 10

Fake background

The misidentified lepton backgrounds are estimated with full data driven method from the collision data. In H $\rightarrow \mu\tau_h$ channel, the misidentification τ lepton rate is obtained from independent Z + jets data sets and then apply the rate to an control region that orthogonal to the signal region to estimate the misidentified lepton backgrounds.

Two set of Z + jets data sets are used, $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$, to estimate the misidentified lepton rate in order to increase the statistics in the estimation. In both cases, Z bosons are selected in the invariant mass window $70 < M_{ll} < 110GeV$. In $Z \rightarrow \mu\mu$, the trigger HLT_IsoMu24 or HLT_IsoTkMu24 is used. Both muons are required to have $P_T > 26GeV$, $|\eta| < 2.4$, cut based tight muon isolation($I^{\mu}_{rel} < 0.15$), passing the muon medium ID and in $Z \rightarrow ee$ case, the trigger HLT_singleE25eta2p1Tight is used. Both electrons are required to pass $P_T > 26GeV$, $|\eta| < 2.1$, cut based tight electron isolation($I^{e}_{rel} < 0.1$), passing MVANonTrigWP80 ID. In the Z+jets samples, with the selected Z boson in an event, the remaining jets are checked if they pass τ ID. The misidentified τ lepton ratio $\tau(f_\tau)$ is calculated as in equation . 10.1, together with one related ratio f_τ.

$$\tau(f_\tau) = \frac{f_\tau}{1 - f_\tau} \tag{10.1}$$

$$f_\tau = \frac{N_\tau(Z + jets\ tau\ tight\ Iso)}{N_\tau(Z + jets\ tau\ very\ loose\ Iso)} \tag{10.2}$$

f_τ is the ratio between number of jets pass tight τ MVA isolation ID and number of jets pass very loose τ MVA isolation ID. The jets that are identified as τ are also required to have $P_T > 30$GeV and $|\eta| < 2.3$. The control region which is orthogonal to the signal region is defined in the same criteria as the signal region, besides the requirement that τ lepton pass very loose isolation and not the tight isolation.

Exact method: For mutauh: estimate the tau fake rate and not the muon fake rate. The muon fake numbers can be commented. with Zmumu+jets and Zee+jets to increase the statistics. The tau fake rate depends on Tau decay mode and P_T.

For etauh: with the same tau fake rate as the mutauh channel, but also have extra electron fake rate. Electron fake rate is also estimated with both Zmumu+jets and Zee+jets samples to increase the statistics in high P_T regions. Control regions are checked to test the performance of this estimation. Same sign region and W+Jets enriched region. Z+jets, fake ratio, applied the weights to control region.

BIBLIOGRAPHY

1. C. Collaboration". Performance of τ-lepton reconstruction and identification in CMS. *JINST*, 7(arXiv:1109.6034. CMS-TAU-11-001. CERN-PH-EP-2011-137): P01001. 33 p, Sep 2011. URL `http://cds.cern.ch/record/1385560`.

2. C. Collaboration. Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2. 2016.

3. C. Collaboration. Reconstruction and identification of tau lepton decays to hadrons and tau neutrino at CMS. *JINST 11 (2016) P01019*, 11, 2016. doi: 10.1088/1748-0221/11/01/P01019.

4. C. Collaboration. Identification of b quark jets at the CMS Experiment in the LHC Run 2. Technical Report CMS-PAS-BTV-15-001, CERN, Geneva, 2016. URL `https://cds.cern.ch/record/2138504`.

5. A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, D. Dannheim, K. Danielowski, S. Henrot-Versille, M. Jachowski, K. Kraszewski, A. Krasznahorkay, Jr., M. Kruk, Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. Tegenfeldt, A. Voigt, K. Voss, M. Wolter, and A. Zemla. TMVA - Toolkit for Multivariate Data Analysis. *ArXiv Physics e-prints*, Mar. 2007.

This document was prepared & typeset with pdfLaTeX*, and formatted with* NDdiss2$_\varepsilon$ *classfile (v3.0[2005/07/27]) provided by Sameer Vijay.*