This Dissertation

entitled

SEARCH FOR LEPTON FLAVOR VIOLATING DECAYS OF

STANDARD MODEL HIGGS TO A MUON AND A TAU LEPTON

OR TO AN ELECTRON AND A TAU LEPTON

typeset with NDdiss2$_\varepsilon$ v3.2017.2 (2017/05/09) on September 1, 2020 for

Prasanna Kumar Siddireddy

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

STANDARD MODEL HIGGS TO A MUON AND A TAU LEPTON

OR TO AN ELECTRON AND A TAU LEPTON

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

Prasanna Kumar Siddireddy

_____

Colin Philip Jessop, Director

Graduate Program in Physics

Notre Dame, Indiana

September 2020

SEARCH FOR LEPTON FLAVOR VIOLATING DECAYS OF

STANDARD MODEL HIGGS TO A MUON AND A TAU LEPTON

OR TO AN ELECTRON AND A TAU LEPTON

Abstract

by

Prasanna Kumar Siddireddy

This dissertation presents searches for Lepton Flavor Violating decay of the Standard Model Higgs Boson (H) into a muon and a tau lepton or to an electron and a tau lepton. Data collected by the CMS detector in 2016-2018, in proton-proton collisions at the LHC, at a center-of-mass energy of $13\,\mathrm{TeV}$ was used to perform the search. Observed (expected) upper limits on the branching fraction of H decaying into a muon and a tau lepton, $\mathcal{B}(\mathrm{H} \to \mu\tau)$, was set at $XX\,(0.13)\,\%$ at 95% CL and branching fraction of H decaying into an electron and a tau lepton, $\mathcal{B}(\mathrm{H} \to \mathrm{e}\tau)$, was set at $XX\,(0.15)\,\%$ at 95% CL. These are the most stringent limits set on these processes till date.

CONTENTS

FIGURES

TABLES

CHAPTER 1

EVENT SELECTION

1.1 Introduction

This chapter summarizes the event selection criteria for the analysis. The signal topology consists of an isolated lepton, μ or e, along with an oppositely charged isolated tau lepton (τ_μ, τ_e, or τ_h). Jets misidentified as electrons or muons are suppressed by imposing isolation requirements. The events are first categorized into $\mu\tau$ and eτ and then further divided into leptonic and hadronic channels based on tau decay mode. Figure 1.1 shows the corresponding Feynman diagrams for the LFV $H \rightarrow \mu\tau$ and $H \rightarrow e\tau$ decays.

The final states of this analysis are similar to the $H \rightarrow \tau\tau$ decay allowed by the SM and since been observed [1]. However, there are some significant kinematic differences. The LFV $H \rightarrow \mu\tau_h$ and $H \rightarrow \mu\tau_e$ ($H \rightarrow e\tau_h$ and $H \rightarrow e\tau_\mu$) decays consist of a muon (an electron) that comes directly from the Higgs and has a hard p_T spectrum, along with a hadronically decaying tau or a softer electron (muon) that comes from the tau lepton of opposite sign charge, and missing transverse momentum from the tau decay. Also, there are fewer neutrinos in LFV decays, 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 visible decay products of the tau and the $\vec{p}_T^{\text{miss}}$ in the azimuthal plane. The same is not true in the $H \rightarrow \tau\tau$ decays. These differences are illustrated pictorially in Figures 1.2 and 1.3.

In each decay mode ($H \rightarrow e\tau_\mu$, $H \rightarrow e\tau_h$, $H \rightarrow \mu\tau_e$, $H \rightarrow \mu\tau_h$), a set of loose selection (preselection) for the respective signature is first defined. The jets in the
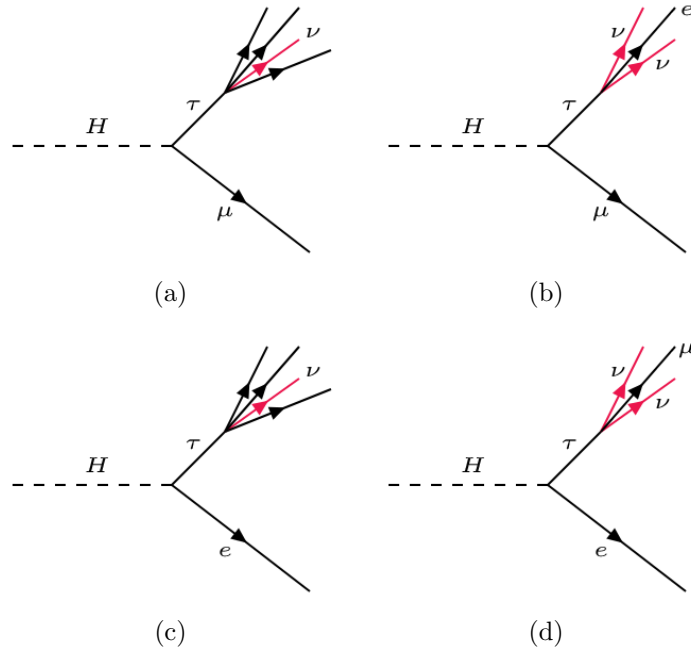
Figure 1.1. Feynman diagrams of lepton-flavor violating Higgs-boson decays. The first row shows diagrams for the Higgs boson coupling to $\mu\tau$ (a,b). Couplings to eτ (c,d) are shown in the second row. Taus can decay leptonically or hadronically. Feynman diagrams are shown for the leptonic decay of taus (b,d) and the hadronic decay of taus (a,c).
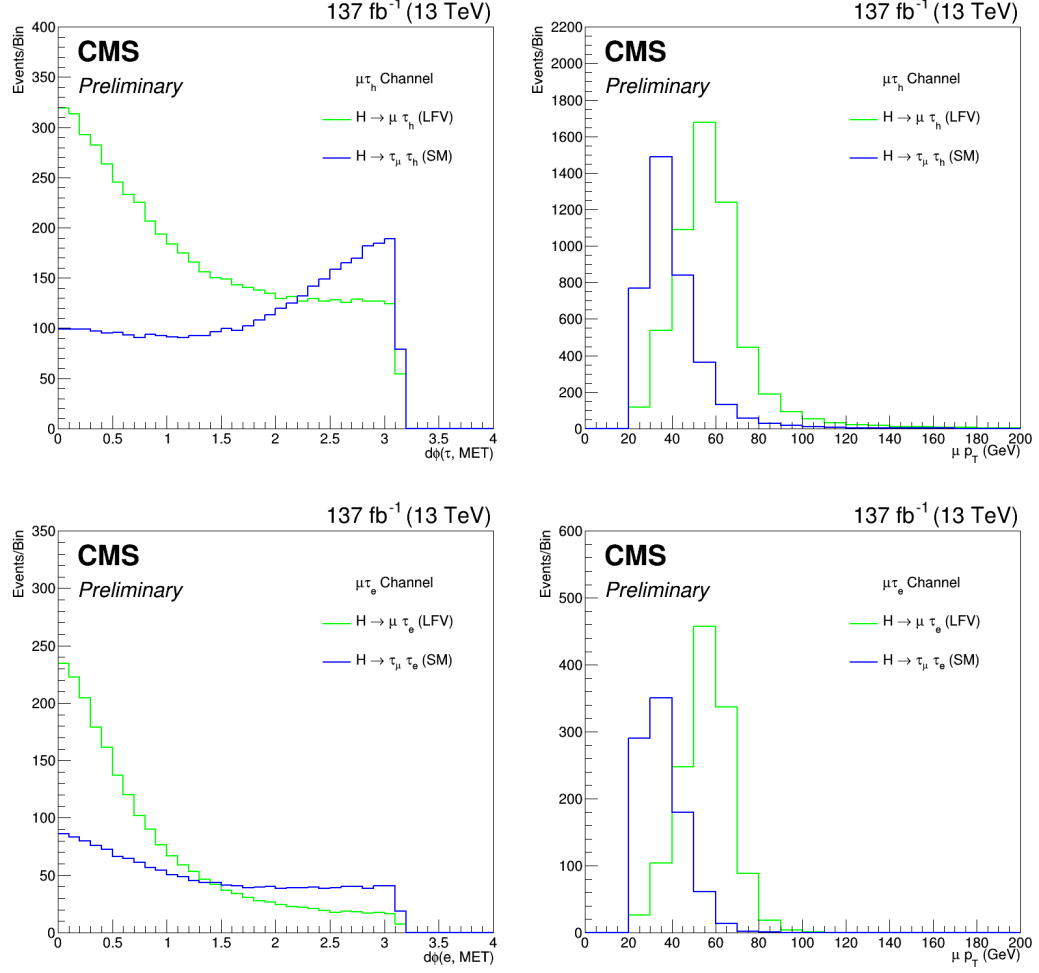
Figure 1.2. Illustration of the differences in $d\phi(\ell = \tau \ or \ e, MET)$ and p_T^{μ} spectrums in LFV and SM H $\rightarrow \tau\tau$ processes.
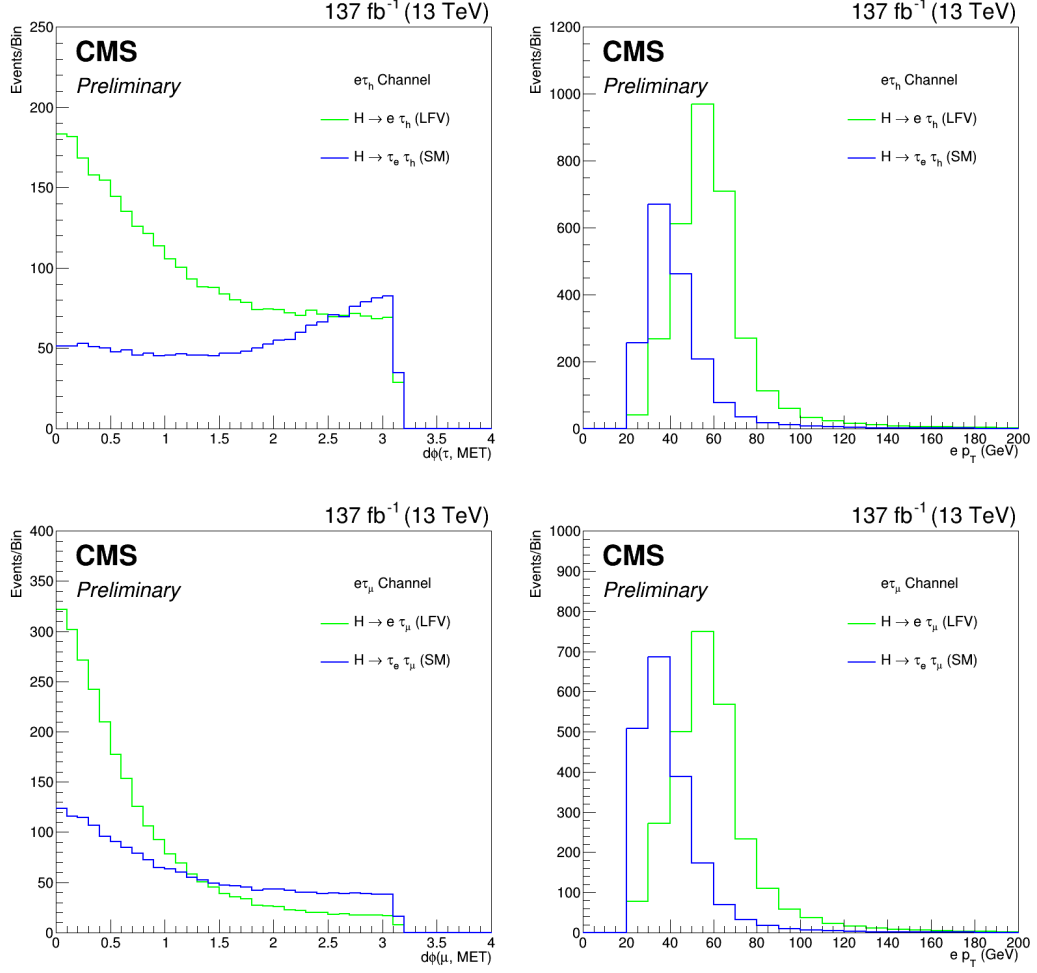
Figure 1.3. Illustration of the differences in $d\phi(\ell = \tau \ or \ \mu, MET)$ and $\mathrm{p_T^e}$ spectrums in LFV and SM H $\to \tau\tau$ processes.

event are required to have a $p_T > 30\,\mathrm{GeV}$ and $|\eta| < 4.7$. The event in each decay channel is divided into categories based on the number of jets in the event (0-jet, 1-jet, 2-jet) to enhance the contribution of different Higgs boson production mechanisms.

The 0-jet category enhances the Gluon Gluon Fusion (GGF) Higgs production contribution, while the 1-jet category enhances the GGF Higgs production with initial-state radiation. The 2-jet category is further broken into two based on the invariant mass of the two jets (M_{jj}). The threshold of 550 (500) GeV on M_{jj} for $\mu\tau$ (eτ) channels has been optimized to give the best-expected exclusion limits. Events with $M_{jj} < 550(500)\,\mathrm{GeV}$ enhances GGF Higgs production contribution while $M_{jj} \geq 550(500)\,\mathrm{GeV}$ enhances VBF Higgs production contribution.

To better discriminate between signal and background events, a Boosted Decision Trees (BDT) discriminator is trained using simulated events, using the TMVA tool of the ROOT analysis package [2]. After applying preselection, a binned likelihood is used to fit the distribution of a BDT discriminator for the signal and the background contributions, and we call this the BDT fit method. The collinear mass (M_{col}) and the transverse mass ($M_T(\ell)$) that are used as input variables to the BDT are defined in the following paragraphs. A brief description of the BDT is given in the next section.

The M_{col} provides an estimate of m_H using the observed decay products of the Higgs boson candidate. It is reconstructed using the collinear approximation based on the observation that, since $m_H \gg m_\tau$, the τ lepton decay products are highly Lorentz boosted in the direction of the τ candidate [3]. The momentum of the neutrino coming from τ decay can be approximated to have the same direction as the visible decay products of the $\tau(\vec{\tau}^{\mathrm{vis}})$. Figure 1.4 shows the corresponding Feynman diagram for the $\not{E}_T$ projected in the direction of the visible decay products of tau.

The component of the $\vec{p}_T^{\mathrm{miss}}$ in the direction of the visible τ lepton decay products, is used to estimate the transverse component of the neutrino momentum ($p_T^{\nu,\mathrm{est}}$). The

5
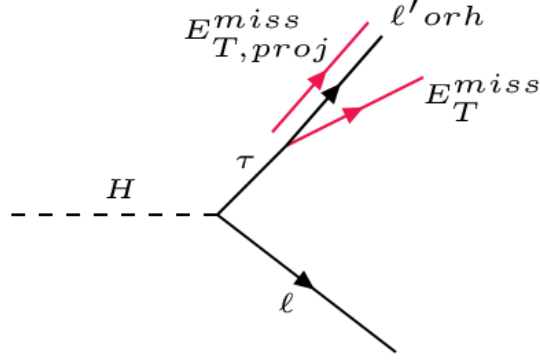
Figure 1.4. Estimation of the neutrino momentum $\not{E}_{T,proj}$ by using the component of the missing transverse energy $\not{E}_T$ which is collinear to the visible decay products of tau in the transverse plane.

collinear mass can then be derived from the visible mass of the $\mu - \tau$ or $e - \tau$ system (M_{vis}) as $M_{\mathrm{col}} = M_{\mathrm{vis}}/\sqrt{x_\tau^{\mathrm{vis}}}$, where x_τ^{vis} is the fraction of energy carried by the visible decay products of the τ ($x_\tau^{\mathrm{vis}} = p_{\mathrm{T}}^{\vec{\tau}^{\mathrm{vis}}}/(p_{\mathrm{T}}^{\vec{\tau}^{\mathrm{vis}}} + p_{\mathrm{T}}^{\nu,\mathrm{est}})$), and M_{vis} is the invariant mass of the visible decay products.

The $M_{\mathrm{T}}(\ell)$ is a variable constructed from the lepton momentum and the missing transverse momentum vectors: $M_{\mathrm{T}}(\ell) = \sqrt{2|\vec{p}_{\mathrm{T}}^{\ell}||\vec{p}_{\mathrm{T}}^{\mathrm{miss}}|(1 - \cos\Delta\phi_{\ell-\vec{p}_{\mathrm{T}}^{\mathrm{miss}}})}$, where $\Delta\phi_{\ell-\vec{p}_{\mathrm{T}}^{\mathrm{miss}}}$ is the angle in the transverse plane between the lepton and the missing transverse momentum, which is used to discriminate the Higgs boson signal candidates from the W + jets background.

An alternate analysis has been implemented to cross-check the results obtained from the BDT fit method. This approach involves placing requirements on several kinematic variables and then using the resulting distribution of M_{col} as a discriminant for a binned likelihood fit. Henceforth, we call this the M_{col} fit method. The BDT and M_{col} fit methods were performed blinded in the signal region [4]. The selection criterion described was developed without looking at the data in the region where the signal is expected. This approach is standard in particle physics analysis and

6

eliminates the experimenter's bias. We use a blinding criteria of $\frac{s}{\sqrt{s+b}} > 0.2$ for the plots that are shown in this chapter.

1.2 Boosted Decision Tree

A decision tree is a tree structure in which there is a condition on an attribute at each internal node. Each branch represents the outcome of this condition, and each leaf node represents a class label. The tree structure is built based on binary splits Figure 1.5. The starting point of the tree structure is called a root node containing all the events we want to classify. A sequence of binary splits is made using conditions on the input variables provided to the classifier. The variable which ensures the separation of the signal and the background is used for each split.
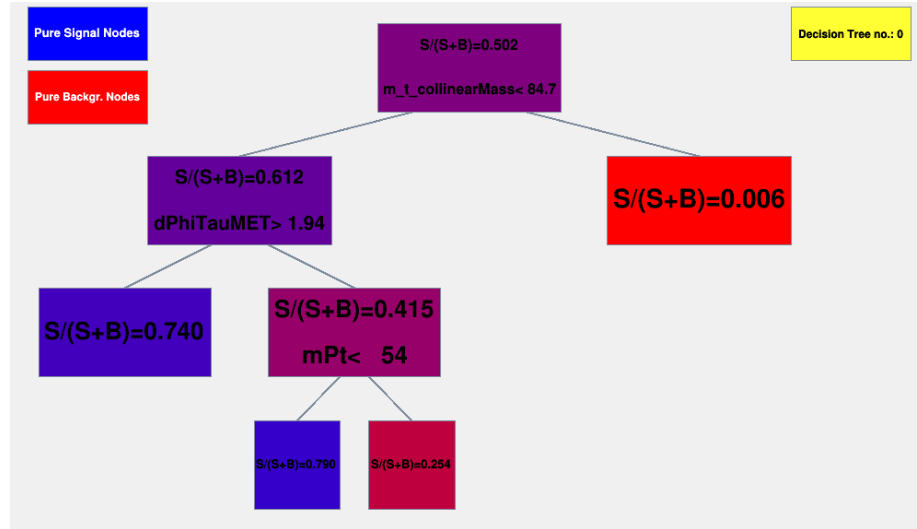
Figure 1.5. Illustration of decision tree.

Gini Index is used as the separation criterion, and it is defined by $p.(1-p)$, where

p is the purity of the node. The purity of a node is given by the ratio of signal events to all events in that node. Since the splitting criterion is always a cut on a single variable, the training procedure selects the variable and cut value that optimizes the increase in the Gini Index between the parent node and the sum of the two daughter nodes' indices, weighted by their relative fraction of events. The same variable can thus be used for splitting several nodes, and the splitting is stopped when a predefined depth of the tree, purity of leaf nodes, the minimum number of events in a leaf node is reached. An event which ends up in a leaf node with a majority of signal events is classified as a signal event and vice versa.

A single decision tree is a weak classifier. The performance of weak classifiers can be enhanced using the Boosting technique. This technique works by building classifiers using reweighted training data and then taking a weighted majority vote of the sequence of classifiers thus produced. AdaBoost (adaptive boosting) method was used for boosting. AdaBoost is adaptive in that subsequent classifiers are tweaked in favor of those instances misclassified by previous classifiers. The misclassified event weights depend on the training error of each decision tree. The training error is calculated as

$$\text{err}_m = \frac{\sum_{i=1}^{N} w_i I\left(y_i \neq DT_m\left(x_i\right)\right)}{\sum_{i=1}^{N} w_i} \tag{1.1}$$

in which the subscript m is the tree label and w is the event weight. The y_i is the true label for the event, 1 for signal and -1 for background. $DT_m(x_i)$ is the output of the decision tree. The variable $I(y_i \neq DT_m(x_i))$ equals 1 if $y_i \neq DT_m(x_i)$ or 0 otherwise. The weight for event i is updated using α_m which is calculated from the training error. β is the learning rate.

$$\alpha_m = \beta \times \ln\left((1 - \mathrm{err}_m) / \mathrm{err}_m\right) \tag{1.2}$$

$$w_i \to w_i \times e^{\alpha I(y_i \neq DT_m(x_i))} \tag{1.3}$$

By construction, the training error is $err_m \leq 0.5$ as the same training events used to classify the output nodes of the previous tree are used to calculate the training error. The learning rate parameter can be used to adjust the step size of each re-weighting. Event weights in each tree are renormalized to keep the summed weights constant. After the boosting and training processes, the final score of each event is $DT(x)$. A high score indicates a signal like event while a low score indicates a background like event.

$$DT(x) = \sum_{m=1}^{N_{tree}} \alpha_m DT_m(x) \tag{1.4}$$

This technique also helps in stabilizing the response of the classifiers for fluctuations in the training data. It utilizes a predefined depth of the tree instead of pruning it to avoid overfitting to the training data. All the BDT trainings were done with an ensemble of 850 decision trees, with each tree having a maximum depth of 3. The minimum node size is required to be 2.5%, and the learning rate is set to 0.5. A training to testing split of 50:50 was used.

1.3   H $\to \mu\tau_{\mathrm{h}}$ channel

The first step is to require the events to pass an isolated muon trigger. For 2016 data, this trigger has a muon p_{T} threshold of $24\,\mathrm{GeV}$. However, for the 2017 and 2018 data, the trigger with a $24\,\mathrm{GeV}$ threshold is prescaled. Prescaling corresponds to collecting one out of every n events to reduce the event rate. We use this trigger in conjunction with the isolated muon trigger with a muon p_{T} threshold of $27\,\mathrm{GeV}$.

In addition to the event passing the trigger, the reconstructed leptons corresponding to the trigger have to match the HLT objects within $\Delta R < 0.5$.

Next, the preselection begins by requiring an isolated μ and an isolated τ_{h} candidates of opposite electric charge and separated by $\Delta \mathrm{R} > 0.5$. The muon candidate is required to have $\mathrm{p}_{\mathrm{T}}^{\mu} > 26\,\mathrm{GeV}$, $|\eta| < 2.1$ and isolation $I_{\mathrm{rel}}^{\mu} < 0.15$. The hadronic tau candidate is required to have $\mathrm{p}_{\mathrm{T}}^{\tau_{\mathrm{h}}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.3$. Events containing additional electrons, muons, or τ_{h} candidates are vetoed. Events with at least one b jet tagged by DeepCSV algorithm are rejected in order to suppress the $t\bar{t}$ background.

A BDT is trained after applying preselection criteria. The signal training sample considered is a mixture of simulated GGF and VBF events, weighted according to their respective SM production cross-sections. The misidentified lepton background and $\mathrm{Z} \rightarrow \tau\tau$ background are the dominant backgrounds in this channel. The background used for training the BDT is a data sample of misidentified lepton events with the same charge assignment for both leptons along with the Drell-Yan MC sample with signal selections. The input variables to the BDT are $\mathrm{p}_{\mathrm{T}}^{\mu}, \mathrm{p}_{\mathrm{T}}^{\tau_{\mathrm{h}}}$, M_{col}, $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, M_{T} $(\tau_{\mathrm{h}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$, $\Delta\eta(\mu, \tau_{\mathrm{h}})$, $\Delta\phi(\mu, \tau_{\mathrm{h}})$, and $\Delta\phi(\tau_{\mathrm{h}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$. The distribution of the input variables to the BDT can be seen in Figure 1.6.

The selection on $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is motivated by the presence of neutrinos in the τ lepton decays. The neutrino is expected to be collinear with τ_{h}, which leads to selection on the $\Delta\phi(\tau_{\mathrm{h}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$ variable. The two leptons are usually in the opposite direction in the azimuthal plane, which leads to selection on the $\Delta\phi(\mu, \tau_{\mathrm{h}})$ variable. The BDT discriminator distributions of simulated signal, data, and backgrounds for each category in H $\rightarrow \mu\tau_{\mathrm{h}}$ channel, are shown in results chapter.

In the M_{col} fit method, additional selection criteria require $M_{\mathrm{T}}(\tau_{\mathrm{h}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}) < 105\,\mathrm{GeV}$ in the $0-, 1-$, and $2-$jet GGF categories and $M_{\mathrm{T}}(\tau_{\mathrm{h}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}) < 85\,\mathrm{GeV}$ in the $2-$jet VBF category. The M_{col} distributions of simulated signal, data, and backgrounds for each category in H $\rightarrow \mu\tau_{\mathrm{h}}$ channel, are shown in results chapter.
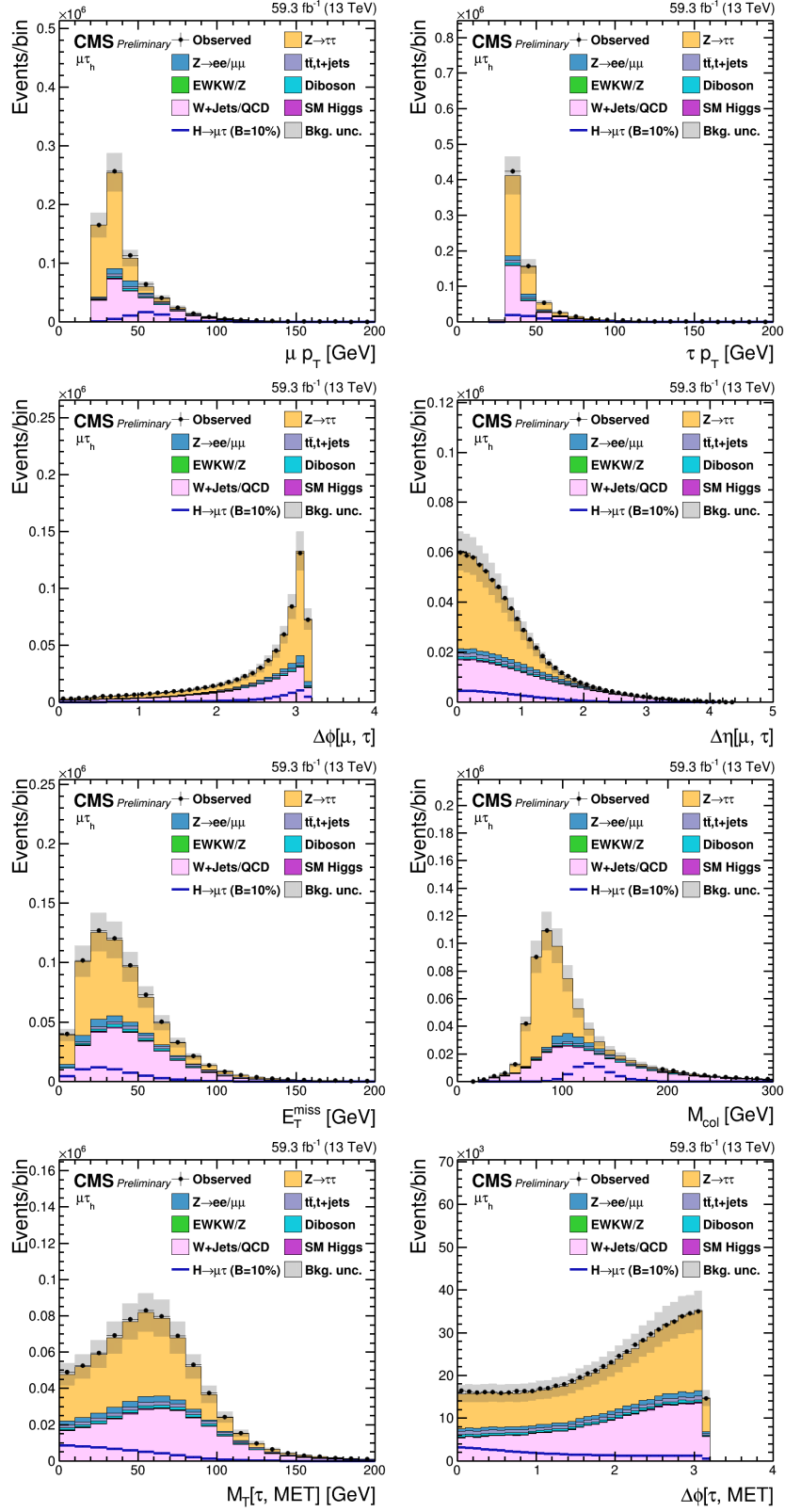
Figure 1.6. Distribution of the input variables to the BDT for the H $\to \mu\tau_{\mathrm{h}}$ process.

11

TABLE 1.1

EVENT SELECTION CRITERIA FOR THE KINEMATIC VARIABLES

FOR THE $H \to \mu\tau$ CHANNELS

Variable	$H \to \mu\tau_{\mathrm{h}}$	$H \to \mu\tau_e$
p_T^{e}	—	> 13
p_T^{μ}	> 26	> 24
$p_T^{\tau_{\mathrm{h}}}$	> 30	—
$\lvert\eta\rvert^{\mathrm{e}}$	—	< 2.5
$\lvert\eta\rvert^{\mu}$	< 2.1	< 2.4
$\lvert\eta\rvert^{\tau_{\mathrm{h}}}$	< 2.3	—
$I_{\mathrm{rel}}^{\mathrm{e}}$	—	< 0.1
I_{rel}^{μ}	< 0.15	< 0.15
$I_{\mathrm{rel}}^{\tau_{\mathrm{h}}}$	DNN τ_{h} ID	—
Trigger	$\mu(24)$ (all years)	$e(12)$ and $\mu(23)$ (all years)

	M_{col} fit selection							
	0-jet	1-jet	2-jet		0-jet	1-jet	2-jet	
			GGF	VBF			GGF	VBF
M_{jj}	—	—	< 550	≥ 550	—	—	< 550	≥ 550
p_T^{μ}		—			> 30	> 26	> 26	> 26
$M_{\mathrm{T}}(\mu)$		—			> 60	> 40	> 15	> 15
$M_{\mathrm{T}}(\tau_{\mathrm{h}})$	< 105	< 105	< 105	< 85		—		
$\Delta\phi(\mathrm{e}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$		—			< 0.7	< 0.7	< 0.5	< 0.3
$\Delta\phi(\mathrm{e}, \mu)$		—			> 2.5	> 1.0	—	—

12
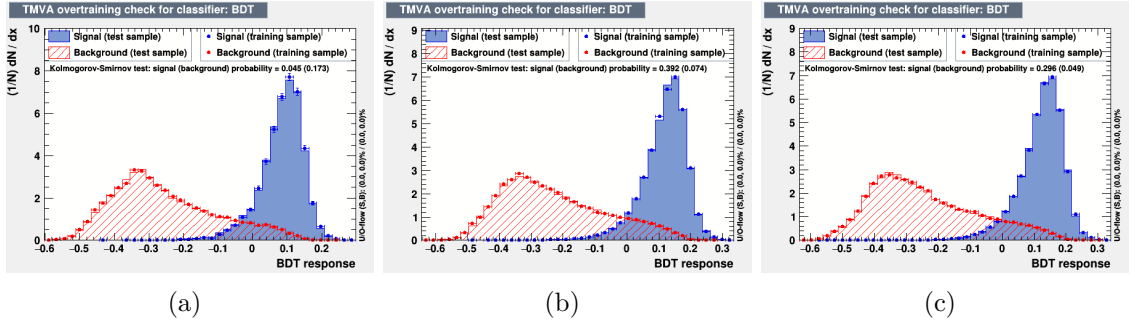
Figure 1.7. Overtraining check as performed in TMVA for the trained BDT in H → $\mu\tau_{\rm h}$ channel for 2016 (a), 2017 (b), and 2018 (c).

1.4 H → $\mu\tau_e$ channel

The events are required to pass the cross-trigger with $p_{\rm T}$ thresholds on the muon and the electron. The $p_{\rm T}$ threshold on the muon is 23 GeV, and on the electron is 12 GeV. The cross-trigger also places a constraint on the two leptons' longitudinal impact parameter to the primary vertex. However, this constraint is not present in the initial 2016 data samples and 2016 MC samples. In addition to the event passing the trigger, the reconstructed leptons corresponding to the trigger have to match the HLT objects within $\Delta R < 0.5$.

The preselection criteria for H → $\mu\tau_e$ channel requires an isolated muon and an isolated electron candidates of opposite charge and separated by $\Delta R > 0.3$. The muon candidate is required to have $p_{\rm T}^{\mu} > 24$ GeV, $|\eta| < 2.4$ and isolation $I_{\rm rel}^{\mu} < 0.15$. The electron candidate is required to have $p_{\rm T}^{\rm e} > 13$ GeV, $|\eta| < 2.5$ and isolation $I_{\rm rel}^{\rm e} < 0.1$. The $p_{\rm T}$ thresold of the electron and the muon are dictated by the cross-trigger we use for selecting the events of this channel. Events containing additional electrons, muons, $\tau_{\rm h}$ candidates or at least one b jet tagged by DeepCSV algorithm are removed.

Similar to the H → $\mu\tau_{\rm h}$ channel, a BDT is trained after applying preselection

13

criteria. The signal training is done in a similar way while for background training dominant contributors, $t\bar{t}$ and $Z \to \ell\ell$ ($\ell = e, \mu, \tau$) events are mixed and weighted by their respective production cross-sections. The $t\bar{t}$ process contributes dominantly for the $2-$jet category, with significant contribution to $1-$jet category. $Z \to \ell\ell$ background processes dominantly contribute the $0-$ and $1-$jet categories. The QCD multijet background has the third-largest contribution, so we use the same sign control region in data as additional background for training. The input variables to the BDT are: p_T^μ, p_T^e, M_{col}, M_T (μ, $\vec{p}_T^{\,\text{miss}}$), M_T (e, $\vec{p}_T^{\,\text{miss}}$), $\Delta\phi(e, \mu)$, $\Delta\phi(\mu, \vec{p}_T^{\,\text{miss}})$, and $\Delta\phi(e, \vec{p}_T^{\,\text{miss}})$. The distribution of the input variables to the BDT can be seen in Figure 1.8. The BDT discriminator distributions of simulated signal, data, and backgrounds for each category in $H \to \mu\tau_e$ channel, are shown in results chapter.

In the M_{col} fit method, additional selection criteria require a stringent selection on muons, $p_T^\mu > 30\,\text{GeV}$ for $0-$jet category and $p_T^\mu > 26\,\text{GeV}$ in rest of the categories. The M_T (μ, $\vec{p}_T^{\,\text{miss}}$) is required to be greater than 60, 40, 15 and 15 GeV for $0-, 1-, 2-$jet GGF and VBF categories, respectively, while azimuthal separation between the electron and $\vec{p}_T^{\,\text{miss}}$ is required to be less than 0.7, 0.7, 0.5 and 0.3 for $0-, 1-, 2-$jet GGF and VBF categories, respectively. For the $0-$ and $1-$jet categories $\Delta\phi(e, \mu) > 2.5$ and 1.0, respectively. The preselections and the selections for $H \to \mu\tau_h$ and $H \to \mu\tau_e$ channels in all categories are summarized in Table 1.1. The M_{col} distributions of simulated signal, data, and backgrounds for each category in $H \to \mu\tau_e$ channel, are shown in results chapter.

1.5 $H \to e\tau_h$ channel

The first step is to require the events to pass a single electron trigger. For 2016 data, this trigger has an electron p_T threshold of 25 GeV. However, for the 2017 and 2018 data, the trigger with the 25 GeV threshold is prescaled. Single-electron triggers with an electron p_T threshold of 27 GeV, 32 GeV, and 35 GeV are used in conjunction

Figure 1.8. Distribution of the input variables to the BDT for the H $\rightarrow \mu\tau_e$ process.

Figure 1.9. Overtraining check as performed in TMVA for the trained BDT in H $\to \mu\tau_e$ channel for 2016 (a), 2017 (b), and 2018 (c).

with the cross-trigger with an electron p_T threshold of 24 GeV and tau p_T threshold of 30 GeV. In addition to the event passing the trigger, the reconstructed leptons corresponding to the trigger have to match the HLT objects within $\Delta R < 0.5$.

The preselection in this channel requires an isolated electron and an isolated τ_h candidates of opposite charge and separated by $\Delta R > 0.5$. The electron candidate is required to have $p_T^e > 27$ GeV, $|\eta| < 2.1$ and isolation $I_{rel}^{\mu} < 0.15$. The hadronic tau candidate is required to have $p_T^{\tau_h} > 30$ GeV and $|\eta| < 2.3$. Events containing additional electrons, muons, or τ_h candidates or at least one b jet tagged by DeepCSV algorithm are removed.

A BDT is trained after applying preselection criteria. The same training samples, as used in the H $\to \mu\tau_h$ channel, are considered. The list of input variables to BDT training stays the same, except for the addition of the visible mass, M_{vis} variable, and removal of $\vec{p}_T^{\,miss}$. The M_{vis} variable is more useful as the relative composition of the two channels' backgrounds is different. In particular, Z $\to$ ee+jets background contributes more with respect to Z $\to \mu\mu$+jets background. The distribution of the input variables to the BDT can be seen in Figure 1.10. The BDT discriminator distributions of simulated signal, data, and backgrounds for each category in H $\to$ eτ_h channel, are shown in results chapter.

16

Figure 1.10. Distribution of the input variables to the BDT for the $H \to e\tau_h$ process.

In the M_{col} fit method, additional selection criteria require $M_{\text{T}}(\tau_{\text{h}}, \vec{p}_{\text{T}}^{\text{miss}}) < 60\,\text{GeV}$ in all the categories. The M_{col} distributions of simulated signal, data, and backgrounds for each category in $\text{H} \to \text{e}\tau_{\text{h}}$ channel, are shown in results chapter.
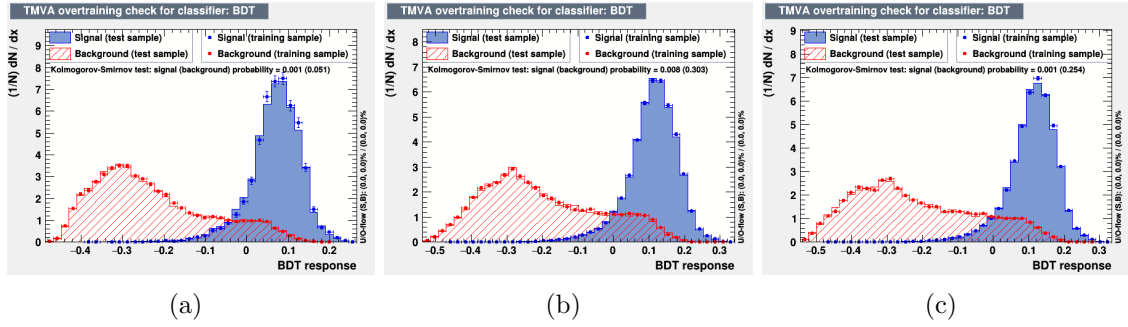
Figure 1.11. Overtraining check as performed in TMVA for the trained BDT in $\text{H} \to \text{e}\tau_{\text{h}}$ channel for 2016 (a), 2017 (b), and 2018 (c).

1.6 $\quad$ H $\to$ eτ_μ channel

The events are required to pass the cross-trigger with p_{T} thresholds on the electron and the muon. The p_{T} threshold on the electron is $23\,\text{GeV}$, and on the muon is $8\,\text{GeV}$. The cross-trigger also places a constraint on the longitudinal impact parameter of the two leptons to the primary vertex. However, this constraint is not present in the initial 2016 data samples and 2016 MC samples. In addition to the event passing the trigger, the reconstructed leptons corresponding to the trigger have to match the HLT objects within $\Delta R < 0.5$.

The preselection criteria for this channel requires an isolated electron and an isolated muon candidates of opposite charge and separated by $\Delta R > 0.4$. The electron

TABLE 1.2

EVENT SELECTION CRITERIA FOR THE KINEMATIC VARIABLES

FOR THE H $\to$ eτ CHANNELS

Variable	H $\to$ eτ_h	H $\to$ eτ_μ
p_T^e	> 27	> 24
p_T^μ	—	> 10
$p_T^{\tau_\mathrm{h}}$	> 30	—
$\|\eta\|^\mathrm{e}$	< 2.1	< 2.5
$\|\eta\|^\mu$	—	< 2.4
$\|\eta\|^{\tau_\mathrm{h}}$	< 2.3	—
$I_\mathrm{rel}^\mathrm{e}$	< 0.15	< 0.1
I_rel^μ	—	< 0.15
$I_\mathrm{rel}^{\tau_\mathrm{h}}$	DNN τ_h ID	—
Trigger	e(25) (2016)	e(23) and μ(8) (all years)
	e(27) (2017) / e(32) (2018)	
	e(24) and τ_h (30) (2017, 2018)	

	M_col fit selection							
	0-jet	1-jet	2-jet		0-jet	1-jet	2-jet	
			ggH	VBF			ggH	VBF
M_{jj}	—	—	< 500	≥ 500	—	—	< 500	≥ 500
p_T^e		—			> 30	> 26	> 26	> 26
$M_T(\mathrm{e})$		—			> 60	> 40	> 15	> 15
$M_T(\tau_\mathrm{h})$	< 60	< 60	< 60	< 60		—		
$\Delta\phi(\mu, \vec{p}_T^\mathrm{miss})$		—			< 0.7	< 0.7	< 0.5	< 0.3
$\Delta\phi(\mathrm{e}, \mu)$		—			> 2.5	> 1.0	—	—

19

candidate is required to have $p_T^e > 24\,\text{GeV}$, $|\eta| < 2.4$ and isolation $I_{rel}^e < 0.1$. The muon candidate is required to have $p_T^\mu > 10\,\text{GeV}$, $|\eta| < 2.5$ and isolation $I_{rel}^\mu < 0.15$. The p_T thresold of the electron and the muon are dictated by the trigger we use for selecting the events of this channel. Events containing additional electrons, muons, τ_h candidates or at least one b jet tagged by DeepCSV algorithm are removed. The selections for both $H \to e\tau_h$ and $H \to e\tau_\mu$ channels in all categories are also summarized in Table 1.2.

The BDT training is done after applying preselection criteria. The same training samples as used in the $H \to \mu\tau_e$ channel are considered. The list of input variables to BDT training also stays the same, except for the addition of the visible mass, M_{vis} variable, and removal of M_T (e, $\vec{p}_T^{\,miss}$). The distribution of the input variables to the BDT can be seen in Figure 1.12. The BDT discriminator distributions of simulated signal, data, and backgrounds for each category in $H \to e\tau_\mu$ channel, are shown in results chapter.

In the M_{col} fit method, additional selection criteria require a stringent selection on electrons, $p_T^e > 30\,\text{GeV}$ for $0-$jet category and $p_T^e > 26\,\text{GeV}$ in rest of the categories. The M_T (e, $\vec{p}_T^{\,miss}$) is required to be greater than 60, 40, 15 and 15 GeV for $0-, 1-, 2-$jet GGF and VBF categories, respectively, while azimuthal separation between the muon and $\vec{p}_T^{\,miss}$ is required to be less than 0.7, 0.7, 0.5 and 0.3 for $0-, 1-, 2-$jet GGF and VBF categories, respectively. For the $0-$ and $1-$jet categories $\Delta\phi(e, \mu) > 2.5$ and 1.0, respectively. The selections for both $H \to e\tau_h$ and $H \to e\tau_\mu$ channels in all categories are also summarized in Table 1.2. The M_{col} distributions of simulated signal, data, and backgrounds for each category in $H \to e\tau_\mu$ channel, are shown in results chapter.
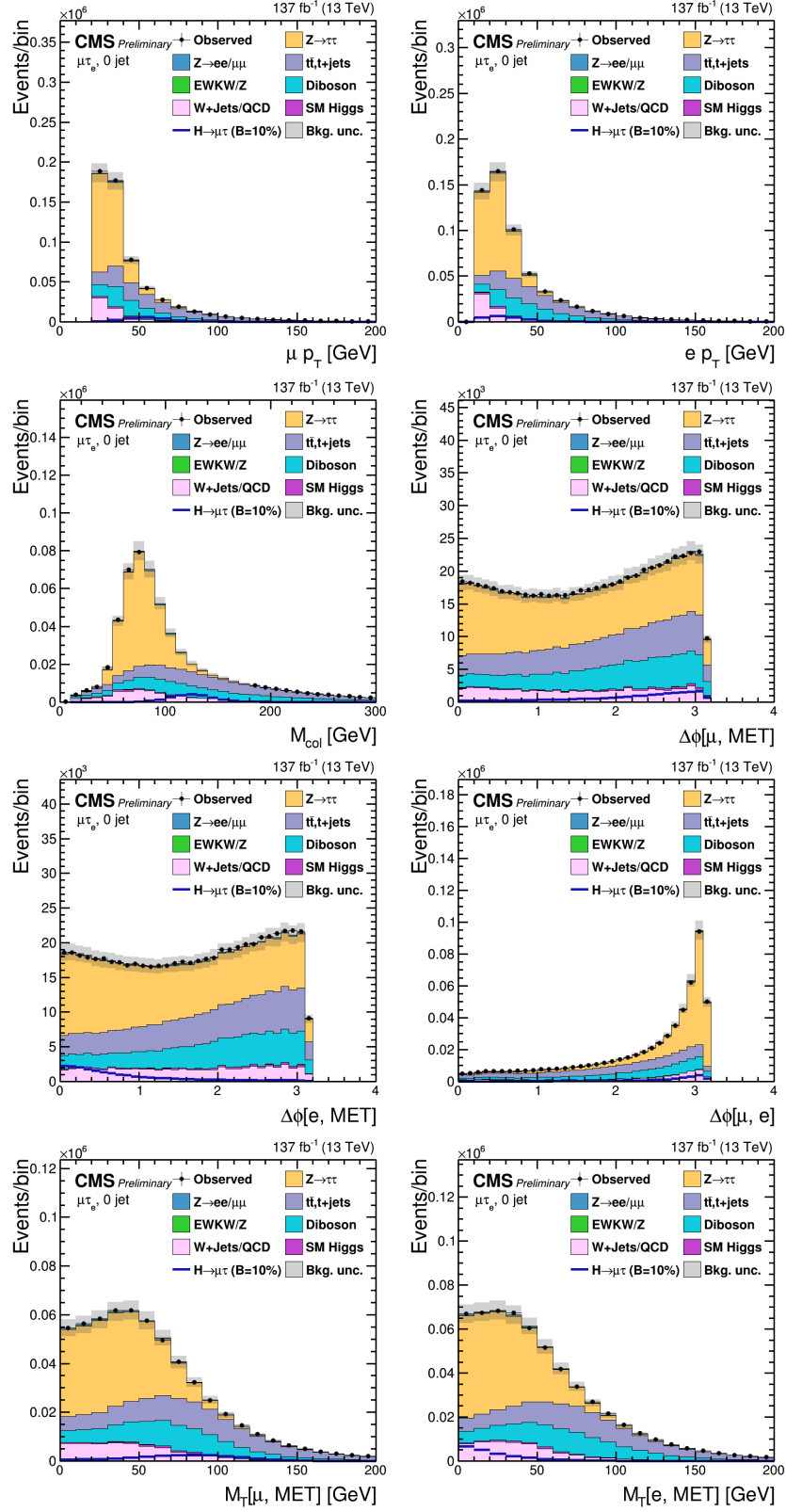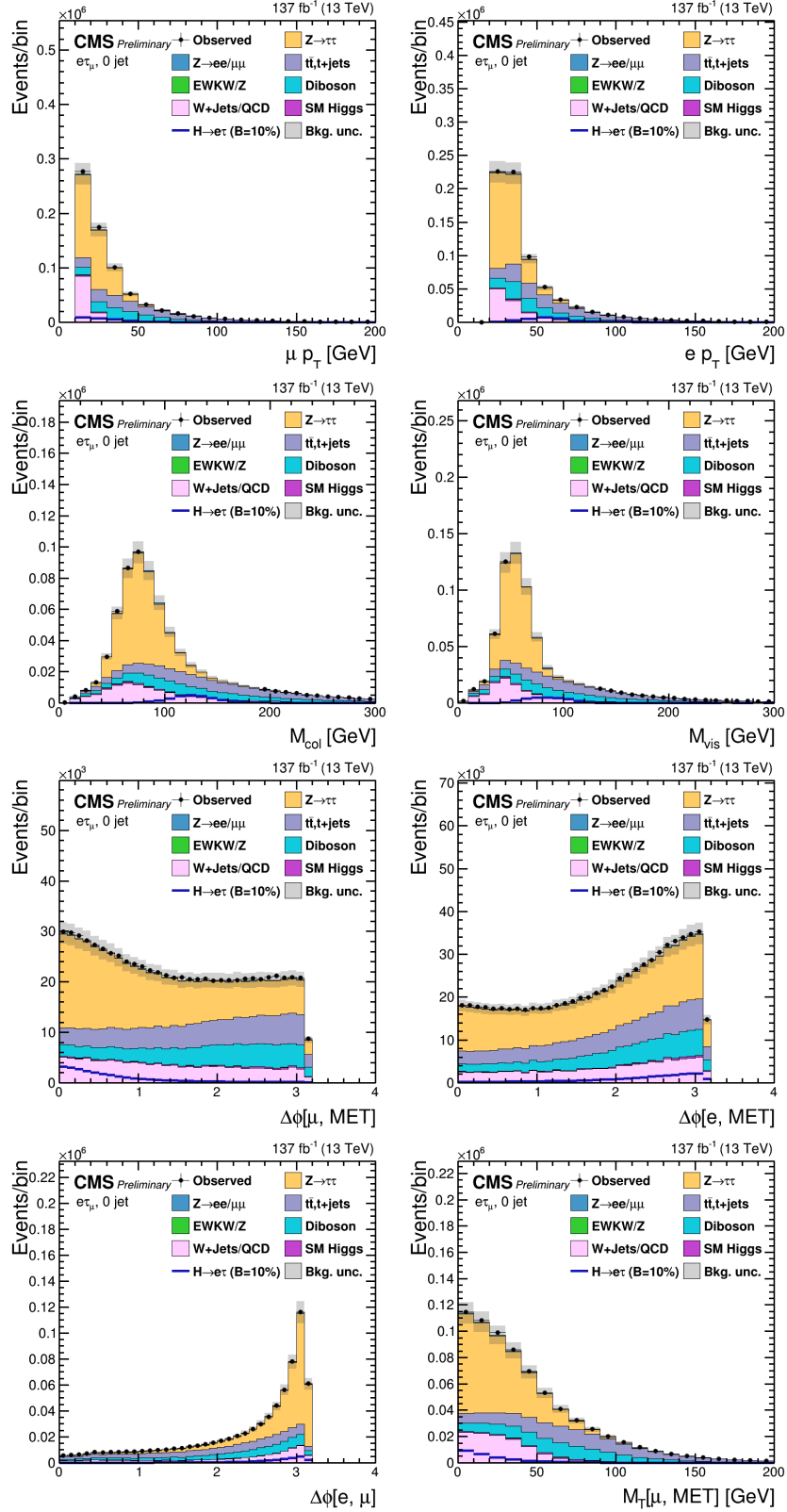
Figure 1.12. Distribution of the input variables to the BDT for the
$H \to e\tau_\mu$ process.
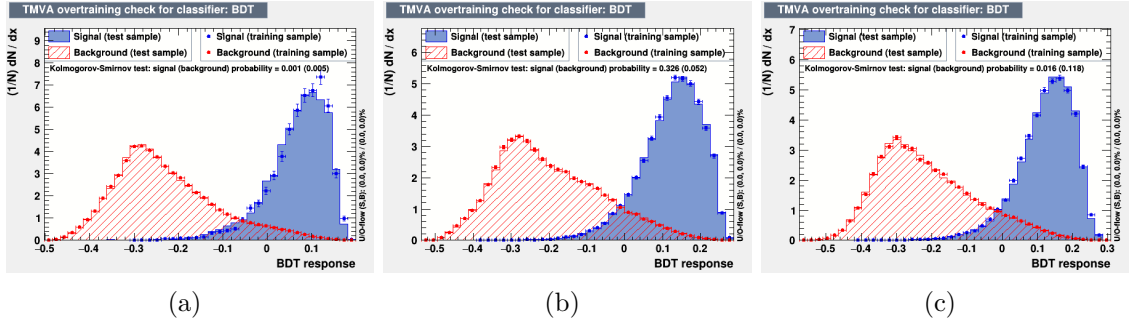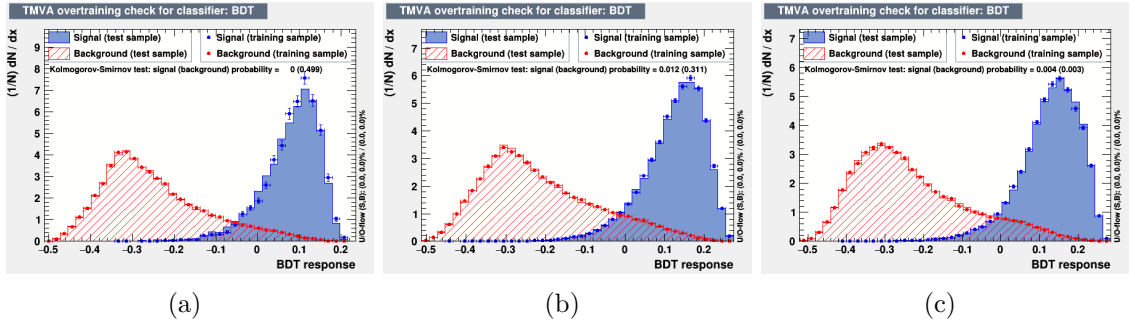
Figure 1.13. Overtraining check as performed in TMVA for the trained BDT in H → eτ_μ channel for 2016 (a), 2017 (b), and 2018 (c).

CHAPTER 2

BACKGROUND ESTIMATION

2.1   Introduction

The signal is a pair of oppositely charged leptons with different flavors, an isolated lepton, e or μ, accompanied by an isolated $\tau(\tau_\mu, \tau_e, \text{ or } \tau_h)$ lepton. The dominant contribution for such a signature comes from $Z \to \tau\tau$ process, in which the μ or e arises from a τ decay. The other dominant contribution comes from $W + \text{jets}$ and QCD multijets processes, where one or more of the jets are misidentified as leptons. In the leptonic channels, the $t\bar{t}$ process also has a dominant contribution.

Other contributions come from the processes in which a lepton pair is produced from the weak decays of quarks and vector bosons. These processes include Higgs boson production ($H \to \tau\tau$, WW), WW, WZ, and ZZ. There are non-negligible contributions from processes like $W\gamma^{(*)} + \text{jets}$, single top quark production, and $Z \to \ell\ell$ ($\ell = e, \mu$). Feynman diagrams of background processes to LFV Higgs boson decays are shown in Figure 2.1.

The dominant contributors, $Z \to \tau\tau$ and misidentified lepton backgrounds, are estimated from data using either a fully data-driven or semi data-driven approach. All the other backgrounds are estimated from simulated samples. The background estimates are validated in different orthogonal control regions constructed to have enhanced contributions from the dominant backgrounds.
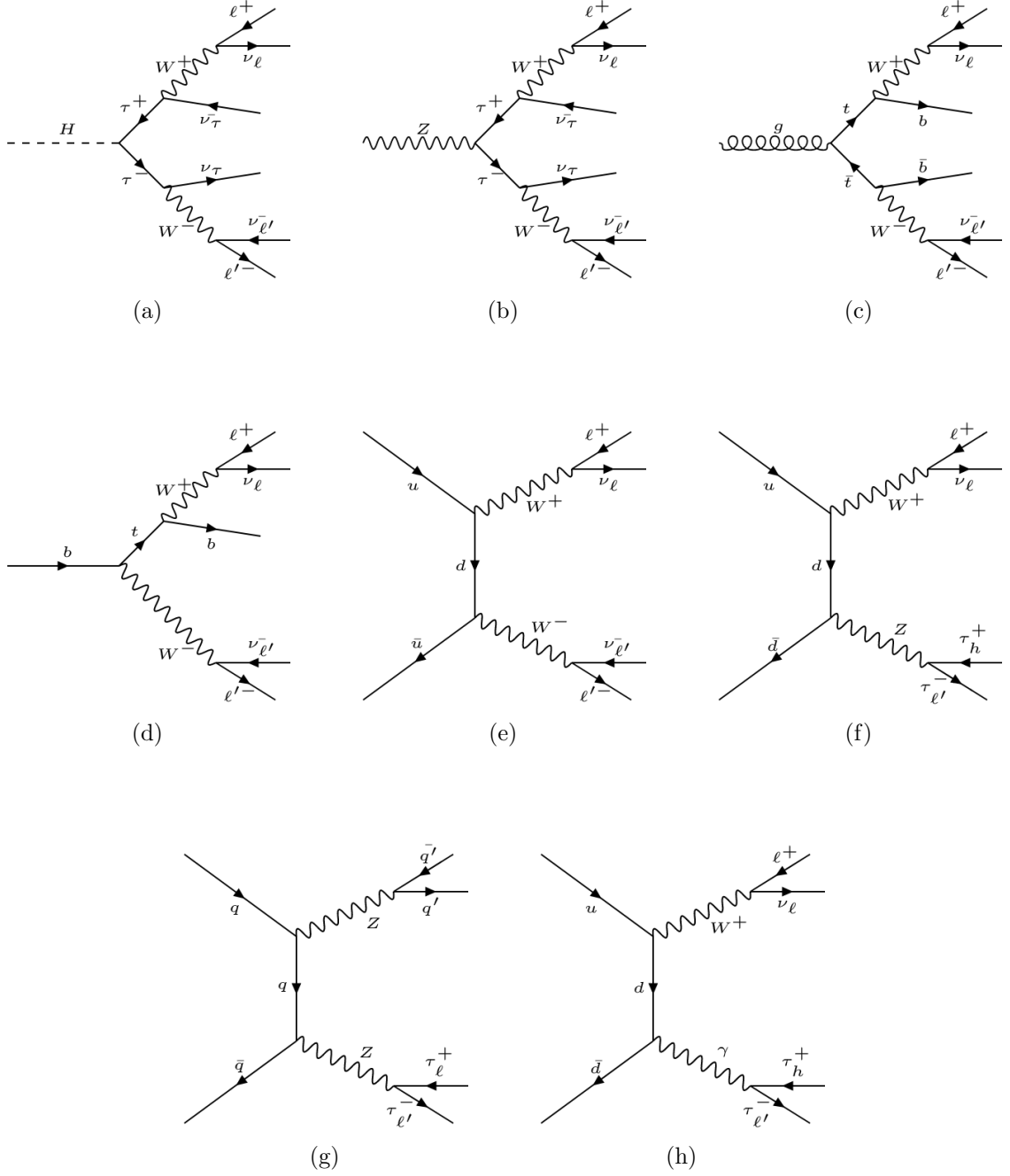
Figure 2.1. Feynman diagrams of background processes to LFV Higgs boson decays: (a) H → ττ, (b) Z → ττ, (c) $t\bar{t}$, (d) Single Top, (e) WW, (f) WZ, (g) ZZ, and (h) Wγ$^{(*)}$.

2.2   Embedding technique

The $Z \to \tau\tau$ background is estimated from data using the embedding technique [5]. The embedding technique allows for an estimation of the genuine $\tau\tau$ standard model backgrounds from data that minimizes uncertainties arising from a poor event description, with minimal simulation input. Events with a pair of oppositely charged muons are selected in data so that $Z \to \mu\mu$ events largely dominate it. These data events are chosen independently of the event selection criteria that are described in Chapter 1.

The muons are removed from the selected events and replaced with simulated τ leptons with the same kinematic properties as that of the replaced muon. In that way, a set of hybrid events is obtained that relies on simulation only for the decay of the tau leptons. The description of the underlying event or the production of associated jets is taken entirely from data, and there is no reliance on the simulation. This technique results in a more accurate description of the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$, jet related variables, and an overall reduction in the systematic uncertainties that arise due to the usage of simulated samples.

Embedded samples cover all backgrounds with two real τ leptons decaying semi-hadronically or leptonically. This includes a small fraction of $t\bar{t}$, Diboson, and electroweak W/Z events. The events from the $t\bar{t}$, Diboson, and electroweak W/Z MC samples where both tau candidates match genuine taus at the generator level are removed to avoid any double counting. A schematic for the embedding technique can be seen in Figure 2.2.

The $Z \to \tau\tau$ background is validated by looking at the agreement between observed data and estimated background in a region enriched with $Z \to \tau\tau$ events. In $H \to \mu\tau_{\mathrm{h}}$ channel, this region is constructed by requiring, in addition to the preselection, $M_{\mathrm{T}}(\mu) < 40\,\mathrm{GeV}$, $40\,\mathrm{GeV} < M_{\mathrm{vis}}(\mu, \tau) < 80\,\mathrm{GeV}$, and $P_\zeta(\mu, \tau) > -25$. P_ζ is the difference of the projections of $\mathrm{p}_{\mathrm{T}}^\ell$ plus MET, and the $\mathrm{p}_{\mathrm{T}}^\ell$ on the axis bisecting the
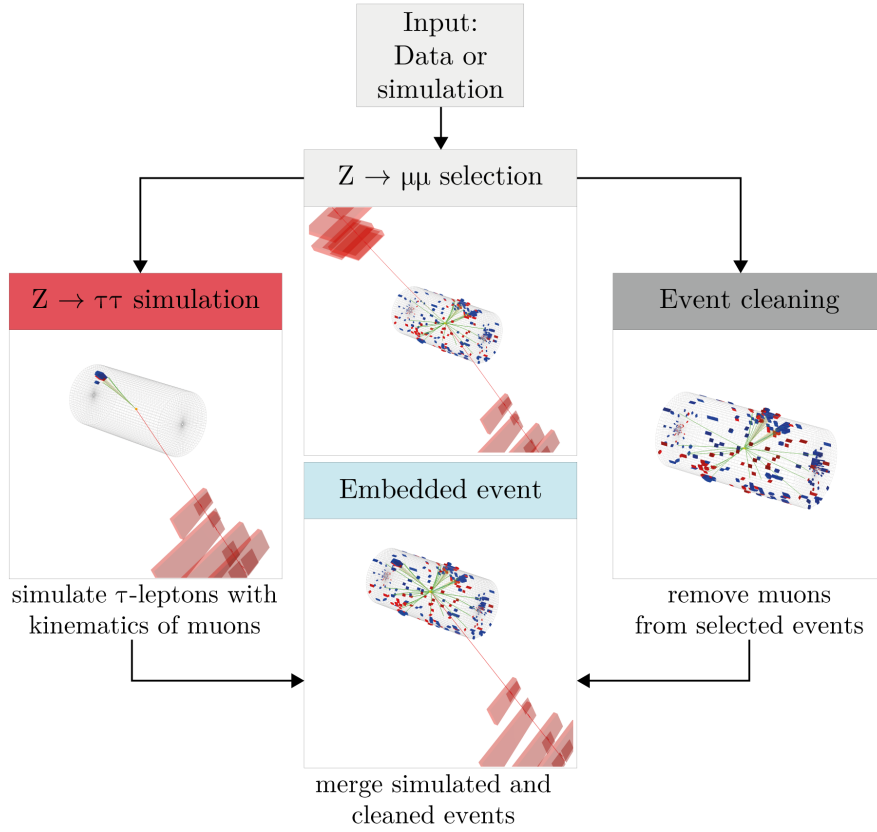
Figure 2.2. Schematic of Embedding Technique

two leptons. In H $\to$ eτ_h channel, the same requirements are placed with the muon variables replaced by corresponding electron variables.

In H $\to$ $\mu\tau_e$ channel, this region is constructed by requiring, in addition to the preselection, $M_\mathrm{T}(\mu) < 60\,\mathrm{GeV}$, $30\,\mathrm{GeV} < M_\mathrm{vis}(\mu, \mathrm{e}) < 70\,\mathrm{GeV}$, and $\mathrm{p}_\mathrm{T}^\mu < 40\,\mathrm{GeV}$. In H $\to$ eτ_μ channel, the same requirements are placed with the muon variables replaced by corresponding electron variables and vice versa. Figures 2.3 and 2.4 show the comparison of data with background estimates in the Z $\to \tau\tau$ control regions for the H $\to \mu\tau$ and H $\to$ eτ channels.

2.3   Misidentified lepton background

Misidentified lepton background corresponds to processes where jets are misidentified as leptons. They mostly arise from two sources, W + jets, and QCD multijet events. In W + jets background events, one of the lepton candidates is from the W boson decay while the other is a jet misidentified as a lepton. In QCD multijet events, both the lepton candidates are misidentified jets.

In two channels of this analysis ($\mu\tau_\mathrm{h}$ and eτ_h), the contributions from misidentified lepton backgrounds have been estimated using a fully data-driven approach. In the leptonic channels ($\mu\tau_\mathrm{e}$ and eτ_μ), a semi data-driven approach is adopted. The results from the semi-data driven approach are found consistent with the fully data-driven method and are undertaken due to limited statistics in the leptonic channel.

2.3.1   Fully data-driven approach

The misidentified lepton background is estimated from collision data by defining a control region with the same selection as the signal region, but loosening the isolation requirements on one of the leptons, to enrich the contribution from W + jets and QCD multijets. The misidentification rates are evaluated using events with a Z boson candidate, and at least one jet that can be misidentified as a lepton and then
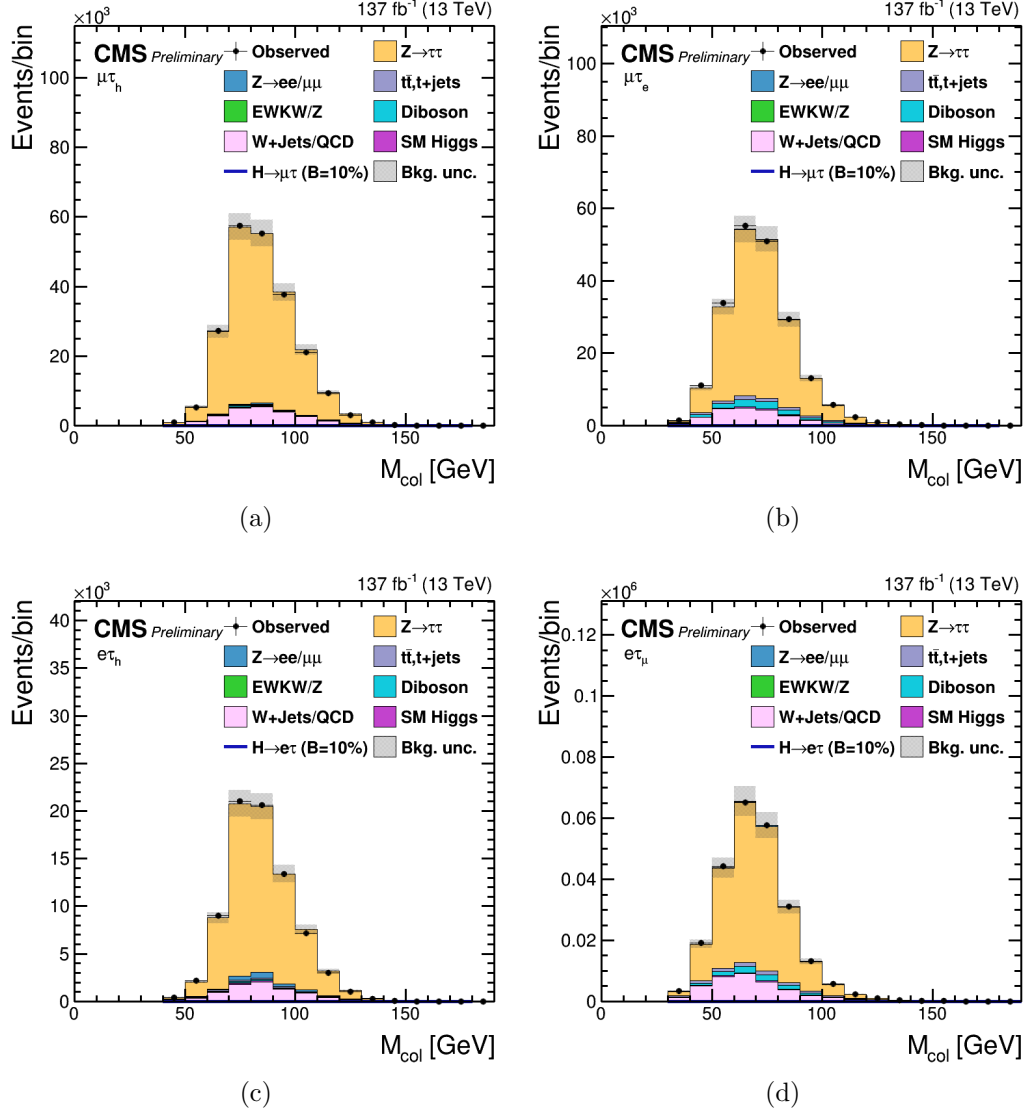
Figure 2.3. Distributions of M_{col} discriminator in the $Z \to \tau\tau$ control regions for the (a) $H \to \mu\tau_{\text{h}}$, (b) $H \to \mu\tau_e$, (c) $H \to e\tau_{\text{h}}$, and (d) $H \to e\tau_\mu$ channels.
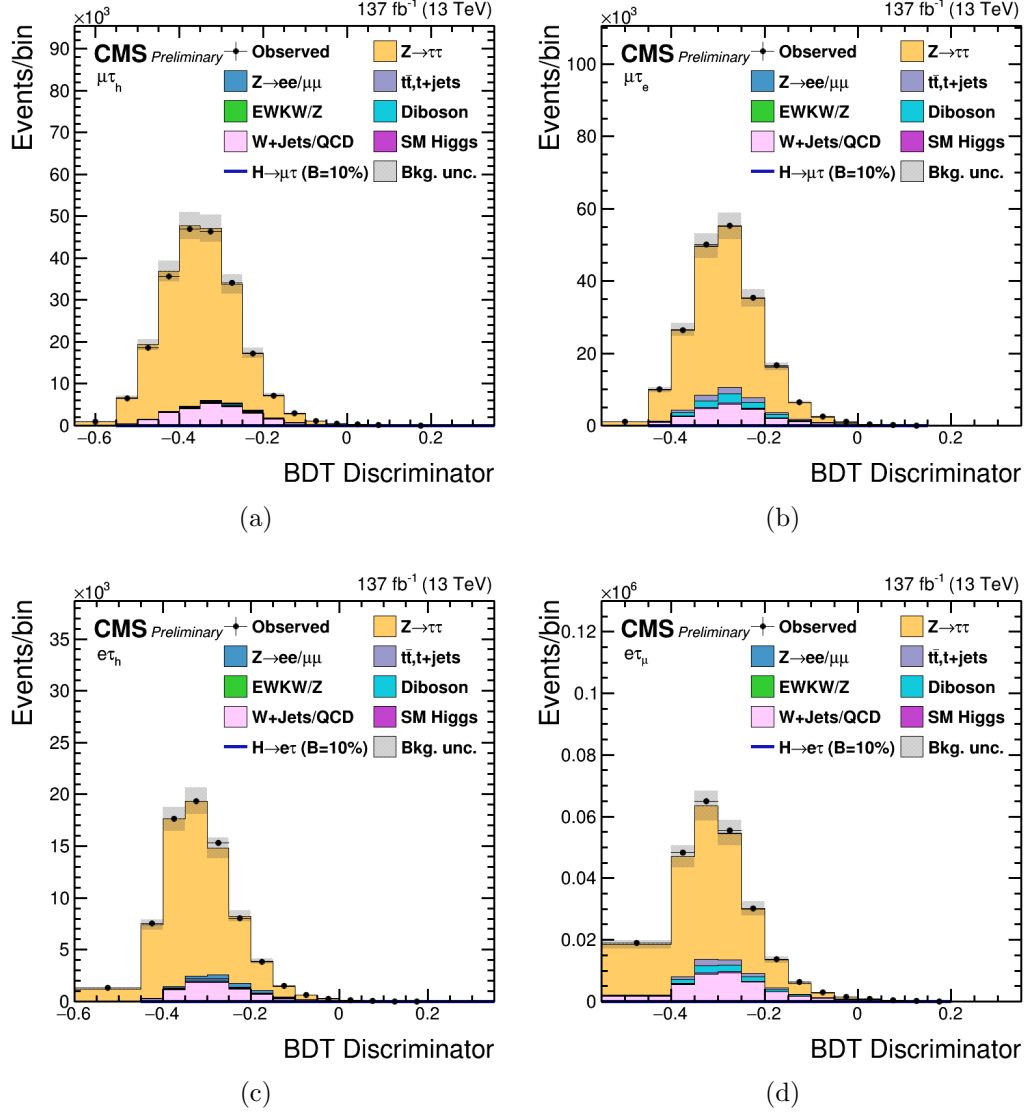
Figure 2.4. Distributions of BDT discriminator in the $Z \to \tau\tau$ control regions for the (a) $H \to \mu\tau_h$, (b) $H \to \mu\tau_e$, (c) $H \to e\tau_h$, and (d) $H \to e\tau_\mu$ channels.

applied to the control region, to estimate the misidentified background of the signal region. The signal region contrasted with the control regions used for determining the misidentified background can be seen in Figure 2.5.
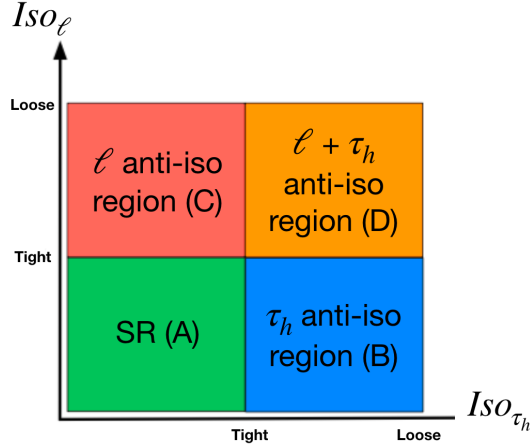
Figure 2.5. Signal region (green) contrasted with the control regions used
for estimating the misidentified background

The probabilities with which jets are misidentified as an electron, muon, or hadronically decaying tau are labeled as f_e, f_μ, and f_{τ_h}, respectively. The Z boson candidate is formed using two muons with $p_T^\mu > 26\,\text{GeV}$ and $|\eta| < 2.4$ and $I_{rel}^\ell < 0.15$ for measuring the jet $\rightarrow \tau_h, \mu, e$ misidentification rate. The muons are required to be oppositely charged and have their invariant mass $(M_{\mu\mu})$ between 70 and 110 GeV.

The contribution from diboson events, where the jet candidate corresponds to a real lepton, is subtracted using simulation. The jet is required to pass the same lepton identification criteria as used in the signal region. A "signal-like" sample is defined if the jet passes the tight lepton isolation, else a "background-like" sample is

defined if it only passes the looser lepton isolation. These two samples are used to estimate f_{e}, f_{μ}, and $f_{\tau_{\text{h}}}$ using the following:

$$f_i = \frac{N_i(\text{signal-like})}{N_i(\text{background-like}) + N_i(\text{signal-like})}$$

where $N_i(\text{signal-like})$ is the number of events with a third lepton candidate that passes the tight lepton isolation, while $N_i(\text{background-like})$ is the number of events that pass only the looser lepton isolation and index $i = \text{e}, \mu$, or τ. The lepton selection criteria is summarized in Tables 1.1 and 1.2.

To estimate the misidentified μ and e contribution in the background-like category, lepton isolation is required to be $0.15 < I_{\text{rel}}^{\mu} < 0.25$ and $0.15 < I_{\text{rel}}^{\text{e}} < 0.5$, respectively. The misidentification rate is computed as a function of the lepton p_{T}. To estimate the τ_{h} misidentified contribution, τ_{h} candidates are required to pass the loose Working Point (WP) of Deep Neural Network (DNN) discrimination against jets but fail the tight WP used for the signal selection. The τ_{h} misidentification rate shows a p_{T} dependence that varies with the τ decay mode and $|\eta|$ and are thus evaluated as a function of $\text{p}_{\text{T}}^{\tau}$ for the different decay modes and two $|\eta|$ regions ($|\eta| < 1.5$ or $|\eta| > 1.5$).

In the H $\rightarrow$ eτ_{h} channel, the τ_{h} misidentification rate is evaluated using events with a Z boson candidate that is formed using two electrons with $\text{p}_{\text{T}}^{\text{e}} > 27\,\text{GeV}$ and $|\eta| < 2.5$ and $I_{\text{rel}}^{\ell} < 0.15$. The electrons are required to be oppositely charged and have their invariant mass (M_{ee}) between 70 and 110 GeV. The reason for using Z $\rightarrow$ ee events for evaluating the τ_{h} misidentification rate in H $\rightarrow$ eτ_{h} channel is because the DNN WPs used for discriminating τ_{h} against electrons and muons is different in this channel compared to the H $\rightarrow$ $\mu\tau_{\text{h}}$ channel. The misidentification rates that are evaluated using this control region are compatible with the measurement in Z $\rightarrow$ $\mu\mu$ events.

(a)

(b)
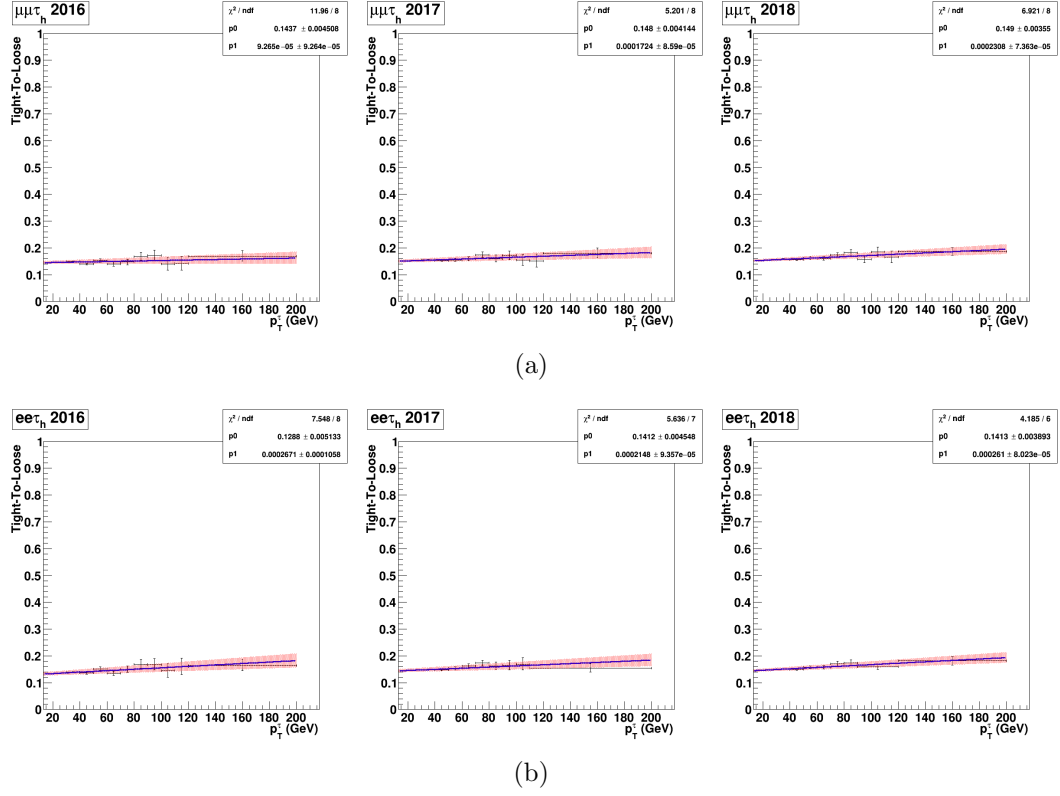
Figure 2.6. Fit performed to τ_h misidentification rates for $H \rightarrow \mu\tau_h$ (a) and $H \rightarrow e\tau_h$ (b) channel as a function of τ_h p_T for the different years. The misidentification rates used are further parametrized based on τ_h Decay Mode along with the pseudorapidity of τ_h. However, here only the inclusive misidentification rates are shown. The misidentification rates are labeled as "tight-to-loose" to clarify that they are calculated as a ratio of the number of events passing the tight WP to the loose WP of DNN discrimination against jets.

(a)

(b)
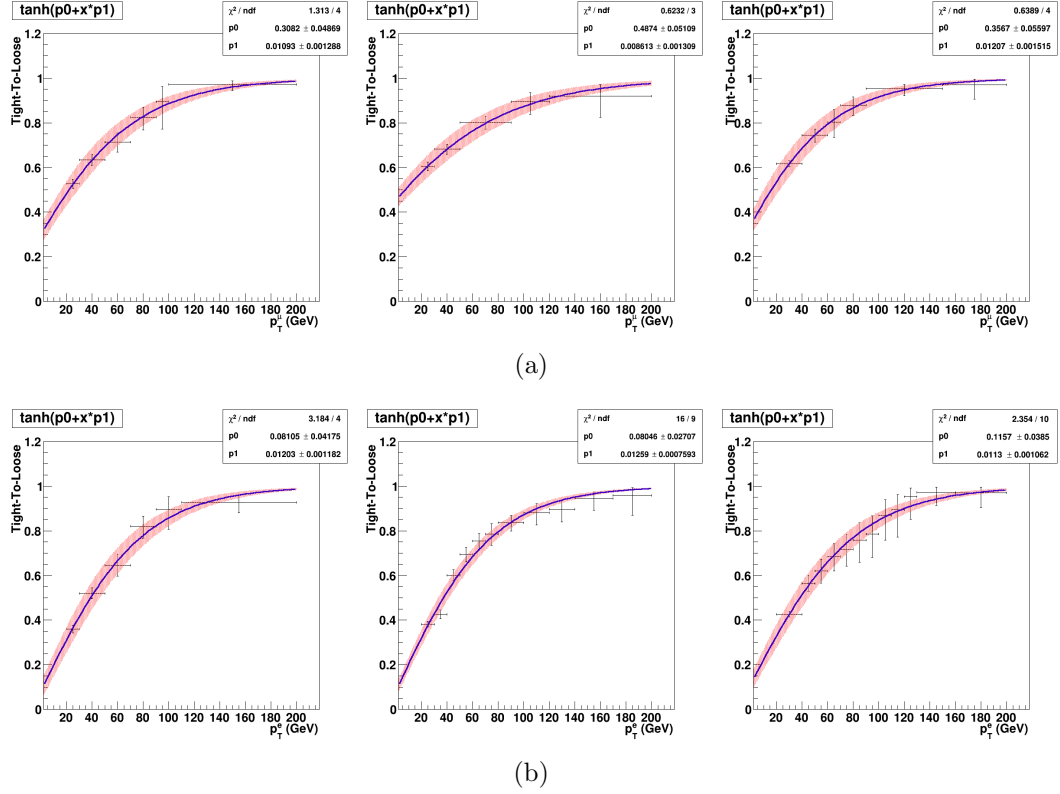
Figure 2.7. Fit performed to the μ (a) and e (b) misidentification rates as a function of their p_T for 2016 (Left), 2017 (Center), and 2018 (Right). The misidentification rates are labeled as "tight-to-loose" to clarify that they are calculated as a ratio of the number of events passing the tight isolation to the loose isolation. The hyperbolic tangent function is used for performing the fit.

Each event in the control region defined using the collision data with the same selection as the signal region, but loosening the isolation requirements on one of the leptons is then weighted by a factor $f_i/(1 - f_i)$ depending on the lepton p_T for electrons and muons or p_T, η and decay mode for the τ lepton candidates. Both background yields and shape distributions are thus estimated. Events with the possibility of double-counting due to two misidentified leptons are subtracted using a weight. For example, events with a misidentified $\mu(e)$ and a misidentified τ_h are subtracted in the $H \to \mu\tau_h$ ($H \to e\tau_h$) channel using a weight, $f_\tau f_\ell/[(1 - f_\tau)(1 - f_\ell)]$, where $\ell = \mu$ or e.

The estimation of the background is validated in events where the two leptons have the same electric charge. The misidentification rate f_i is applied to events passing preselection and by inverting the lepton pair's charge requirement. The same-sign selection enhances the misidentified lepton background. The background estimation is also validated in a W boson enriched control sample. This control region is obtained by applying the signal-like requirement and $M_T(\ell, \vec{p}_T^{\,miss}) > 60\,\text{GeV}$ ($\ell = $ e or μ) and $M_T(\tau_h, \vec{p}_T^{\,miss}) > 80\,\text{GeV}$. The same strategy is applied in the $H \to e\tau_h$ channel, which results in a similar agreement. Figures 2.8 and 2.9 show the comparison of data with background estimates in the same-sign and W boson enriched control regions for the $H \to \mu\tau_h$ and $H \to e\tau_h$ channels.

2.3.2   Semi data-driven approach

In the $H \to e\tau_\mu$ and $H \to \mu\tau_e$ channels, QCD multijet background is estimated from data using events with an electron and a muon with the same electric charge. These events are selected by applying the preselection except for requiring both the leptons to have the same electric charge, and we call this the same sign (SS) control region. Contributions from other processes are estimated from simulation and subtracted from data in this SS control region. Extrapolation factors from the SS control
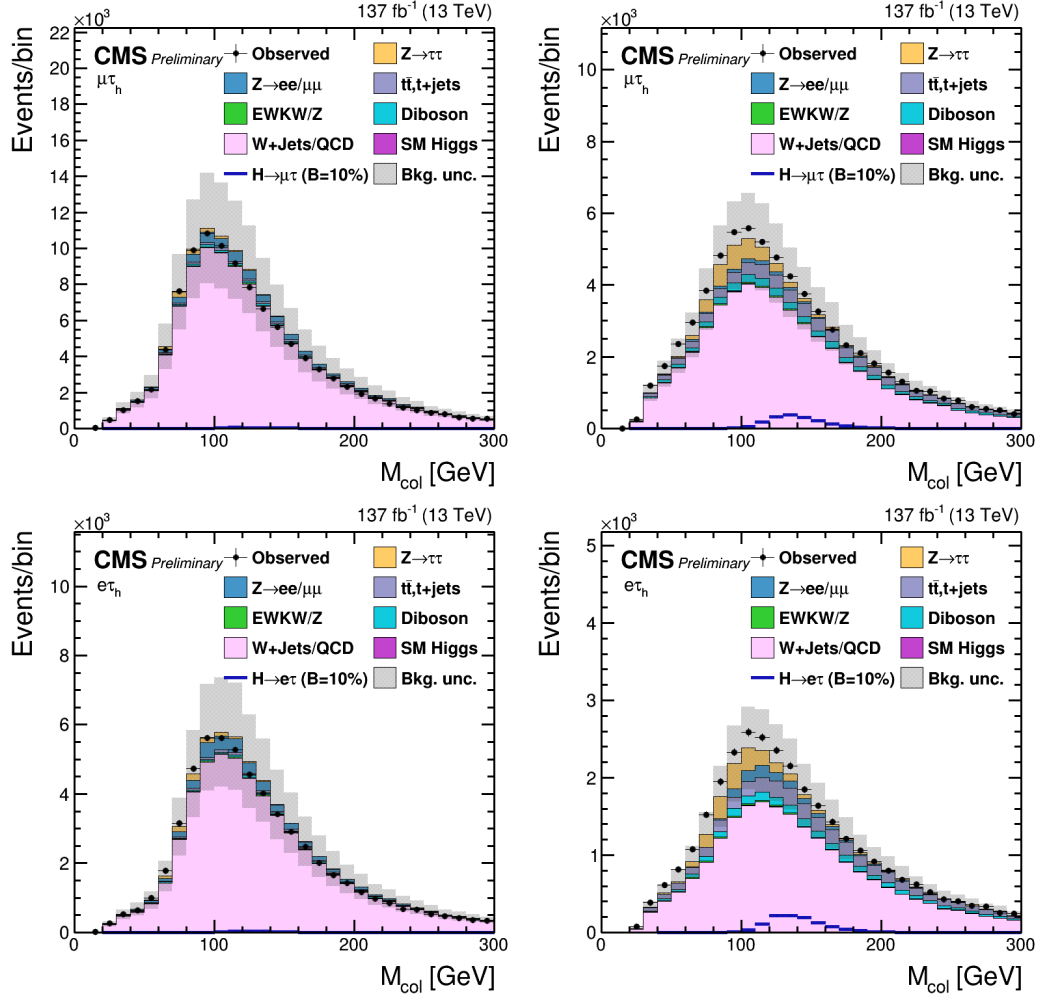
Figure 2.8. Distributions of M_{col} discriminator in the same-sign (Left) and W boson enriched (Right) control regions for the H $\rightarrow \mu\tau_{\text{h}}$ (top) and H $\rightarrow$ eτ_{h} (bottom) channels.
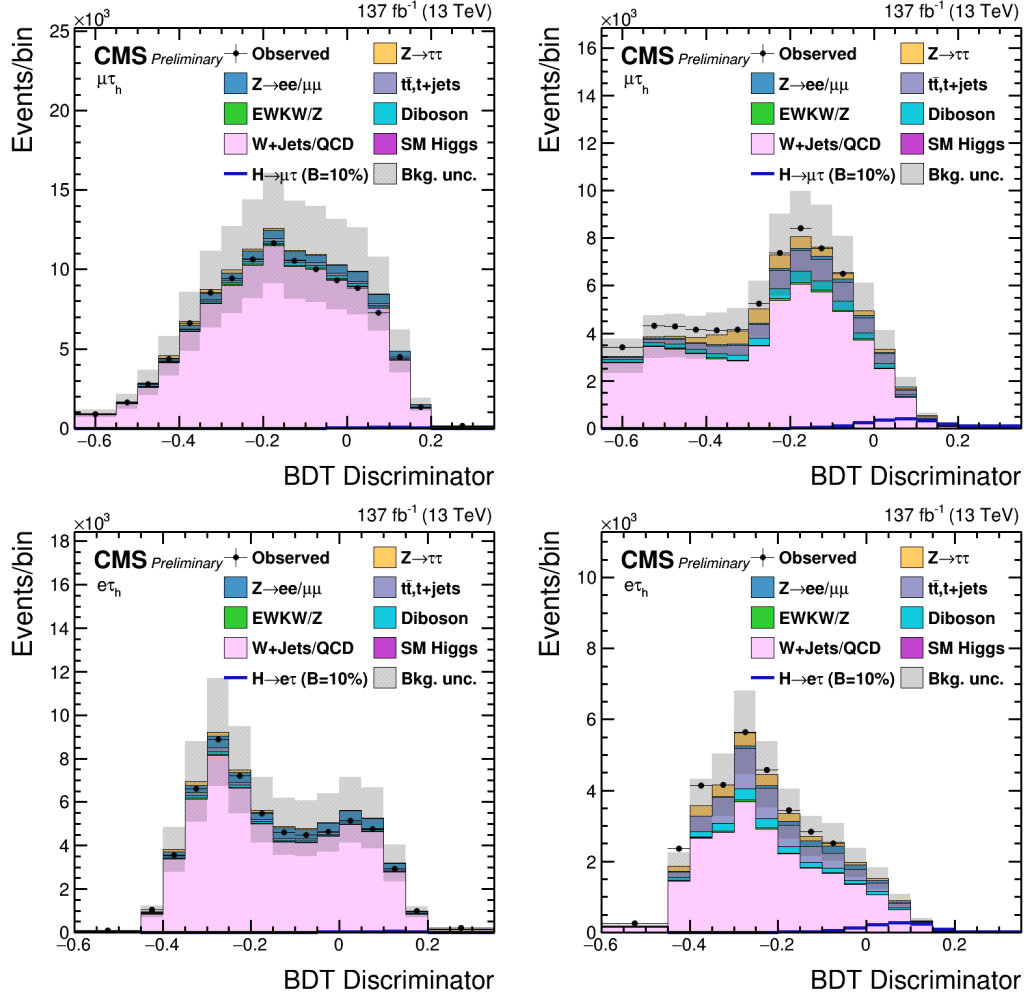
Figure 2.9. Distributions of BDT discriminator in the same-sign (Left) and W boson enriched (Right) control regions for the H → $\mu\tau_h$ (top) and H → eτ_h (bottom) channels.

region to the opposite sign (OS) signal region are measured in data as a function of the jet multiplicity and the ΔR separation between the electron and the muon. QCD OS/SS extrapolation factors that are measured can be seen in Figure 2.10.

The OS/SS extrapolation factor is estimated using events with an anti-isolated muon and an isolated electron. The contribution from $b\bar{b}$ events to the QCD multijet background gives rise to the ΔR dependency and is parameterized with a linear function. The OS/SS extrapolation factor is higher for events with low ΔR separation between the electron and the muon, decreasing as the ΔR separation increases. The OS/SS extrapolation factor also depends on the electron and muon p_T. This p_T dependence comes from the leptons arising from the semi-leptonic c quark decay. These leptons tend to be softer in p_T and less isolated resulting in a reduction in the number of such events passing the p_T and isolation requirements. Corrections of the QCD OS/SS extrapolation factors dependent on lepton p_T along with the correction to account for the mismodeling introduced by anti-isolating the muon to measure them can be seen in Figure 2.11.

As the OS/SS extrapolation factor is measured in a control region where the muon is anti-isolated, an additional correction is applied to cover for a potential mismodeling. This correction is calculated by measuring the OS/SS extrapolation factors in two different control regions. The first control region has events where the muon is isolated, and the electron is anti-isolated. The second control region has events where both the electron and the muon are anti-isolated. The ratio of the extrapolation factors measured in these two control regions is taken as the correction for accounting the potential mismodeling induced by anti-isolating the muon. Figure 2.12 shows the comparison of data with background estimates in the muon anti-isolated control regions for the H $\to \mu\tau_e$ channel.
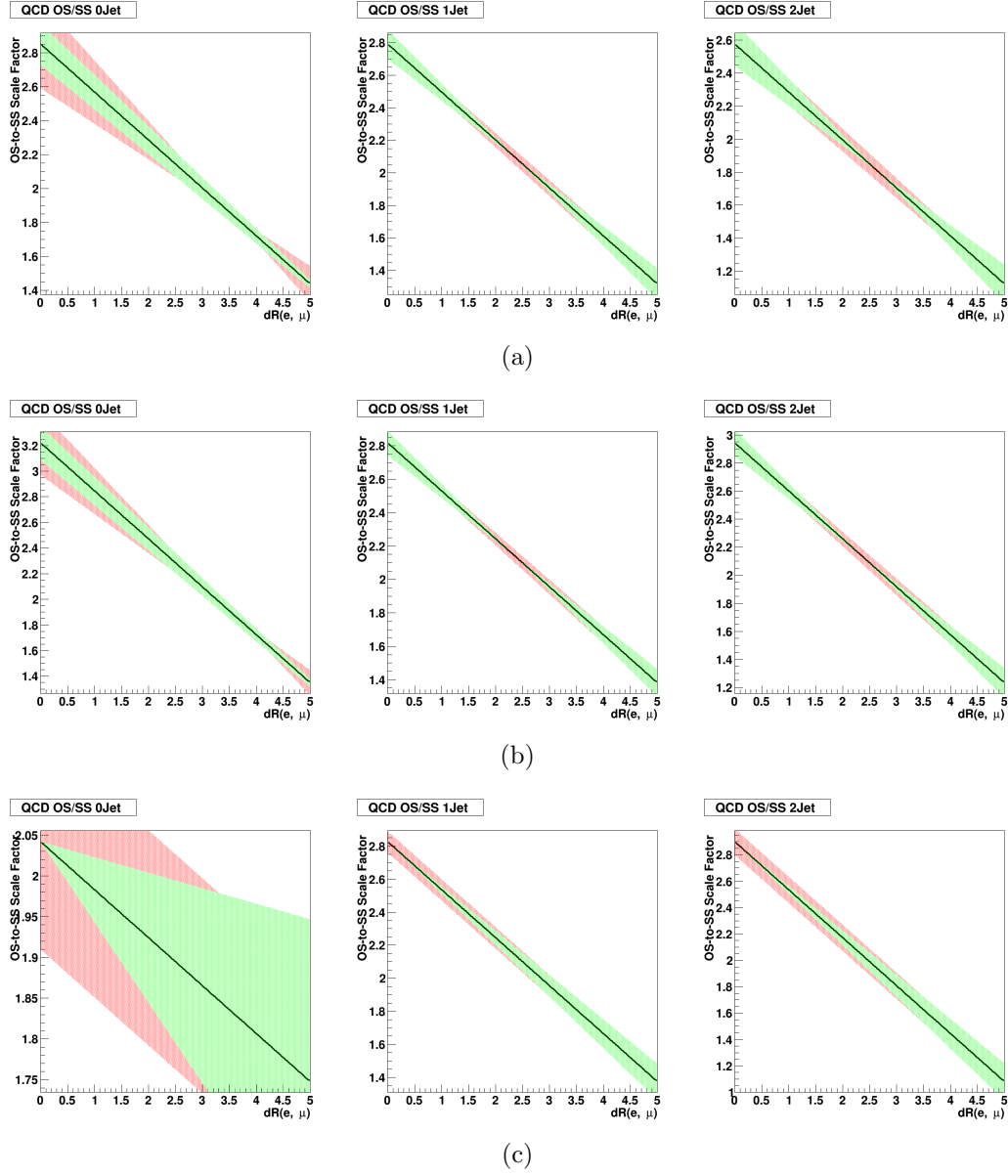
Figure 2.10. QCD OS/SS extrapolation factors in events with 0 Jets (Left), 1 Jet (Center), and 2 Jets (Right) for 2016 (a), 2017 (b), 2018 (c). The line is the best fit, and the shaded region corresponds to the shape uncertainties.

(a)

(b)
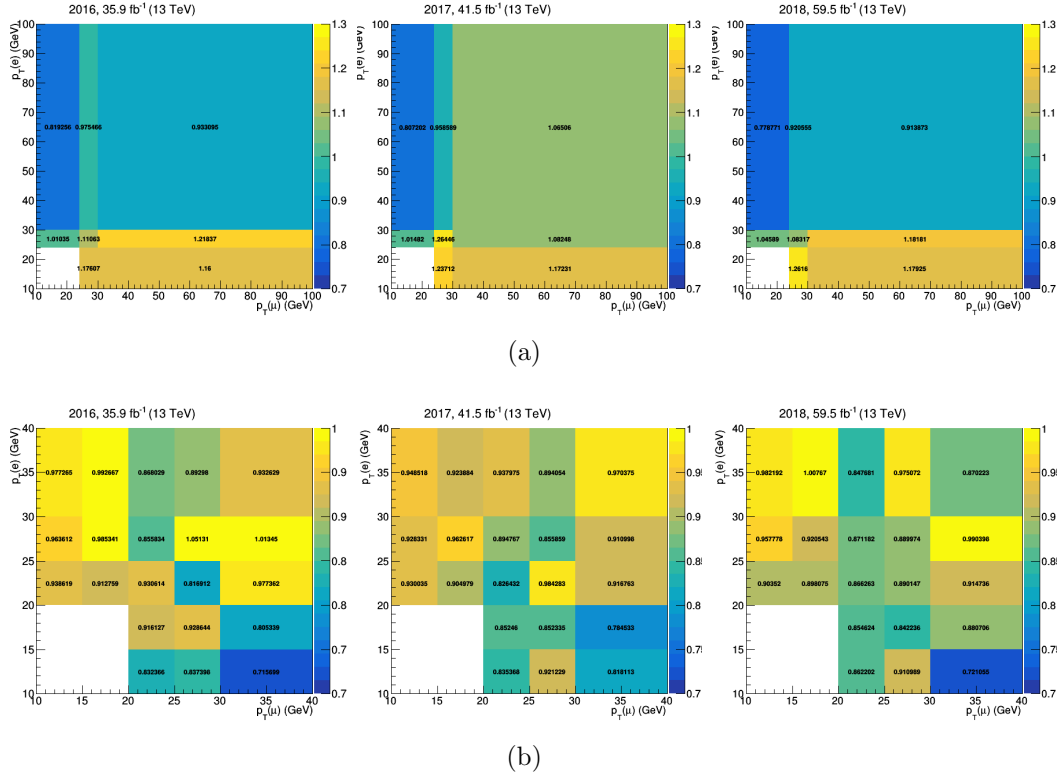
Figure 2.11. (a) Corrections of the QCD OS/SS extrapolation factors determined in the region with an anti-isolated muon as a function of the p_T of the electron and the muon, using data collected in 2016, 2017, and 2018. (b) Correction of the QCD OS/SS extrapolation factors to account for the mismodeling introduced by anti-isolating the muon to measure the SFs, using data collected in 2016, 2017, and 2018.

39
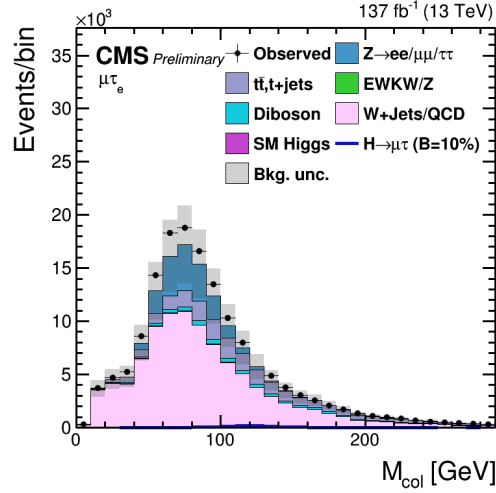
Figure 2.12. Distribution of M_{col} discriminator in the muon anti-isolated control regions for the H $\to \mu\tau_e$ channel.

2.4 MC Simulation

All the other backgrounds are estimated using MC simulation. In the leptonic channels, the $t\bar{t}$ process has a dominant contribution, and this background is validated in a dedicated control region defined by requiring the presence of at least one b jet tagged by the DeepCSV algorithm in the event. Figure 2.13 shows the background validation in this control region for the H $\to \mu\tau_e$ and H $\to$ eτ_μ channels.

The standard model, Higgs boson production, forms a small but non-negligible background. The contributions come mainly from H $\to \tau\tau$ and H $\to$ WW decays. The contribution from H $\to$ WW peaks at lower values than the signal in the distribution of the BDT discriminator due to the presence of additional neutrinos in the decay. The background is estimated from simulations using selection based on the BDT discriminator and kinematic variables, as described in Chapter 1.
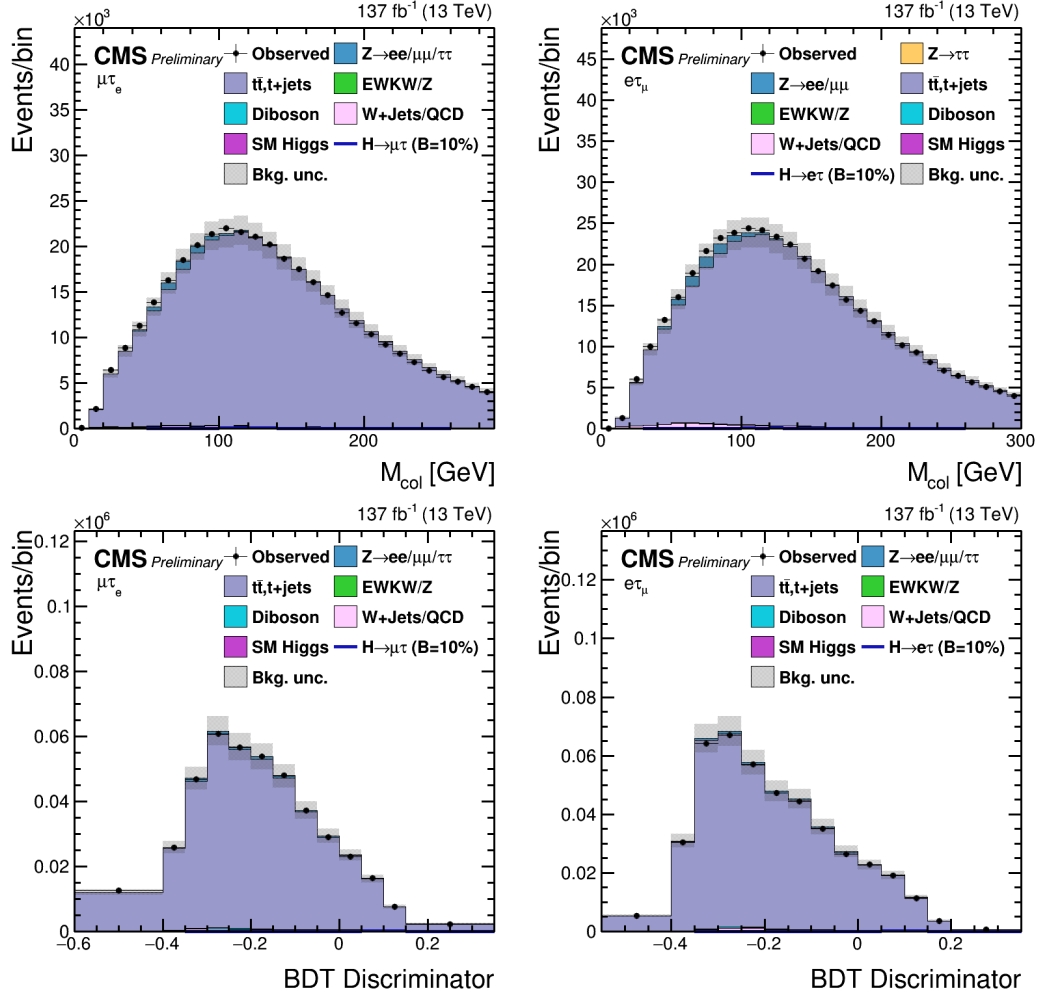
Figure 2.13. Distributions of M_{col} (BDT) discriminator in $t\bar{t}$ enriched control region for $H \to \mu\tau_e$ and $H \to e\tau_\mu$ channel.

CHAPTER 3

STATISTICAL METHODS AND SYSTEMATIC UNCERTAINTIES

3.1   Introduction

In the search for lepton flavor violating Higgs decays, a discovery will be the observation of events with the Higgs Boson decaying either into $\mu\tau$ or eτ. As the LFV is forbidden in SM, the SM can be taken as the background model, and discovery can be claimed if the observation is not compatible with this background model. The uncertainties resulting from theoretical, experimental, and statistical sources can give rise to an excess in the observation even when there is no signal at all. When an excess is observed, a p-value is computed. The p-value represents the probability that an observed excess is due to statistical fluctuations. The p-value has to be very low to indicate that the excess observed is due to a signal's presence and not merely a statistical fluctuation. However, if no excess is observed, upper exclusion limits can be set on the branching fraction. This chapter describes the statistical methods used for extracting the signal strength, followed by the systematic uncertainties involved in the analysis.

3.2   Statistical methods

The results on the branching fraction of the LFV Higgs-boson decays to $\mu\tau$, and eτ are estimated using a profile likelihood method. The SM Higgs boson production cross-sections are used for the signal model, while the branching fraction of the Higgs boson to $\mu\tau$ and eτ remain free parameters. The branching fraction is the

42

parameter of interest. In addition to it, the signal and background model contain in general nuisance parameters whose values must be derived from collision data. The profile likelihood method is implemented, assuming the asymptotic approximation. Distributions of the BDT discriminator and the collinear mass for signal and various background processes are fitted to collision data. Systematic uncertainties are represented as nuisance parameters, and they can affect the normalization or the shape of the distribution.

Poisson distribution can be used to model the expected number of events and the observed events for the situation at hand. The expected number of events is $\mu \cdot s + b$, where s is the expected signal event yields, and b is the expected background event yields. The parameter μ is the signal strength modifier, which changes the signal production cross-sections of all the production mechanisms by exactly the same scale of μ. The likelihood function measures the goodness of fit of a statistical model to a sample of data for given values of the unknown parameters. For our situation, the likelihood function $\mathcal{L}(data|\mu)$ is then given by:

$$\mathcal{L}(data|\mu) = \prod_{i=1}^{bins} \frac{(\mu \cdot s_i + b_i)^{n_i}}{n_i!} e^{-\mu \cdot s_i - b_i} \tag{3.1}$$

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.

The nuisance parameters which represent the systematic uncertainties are embedded into the likelihood function. The uncertainties considered are taken to be 100%-correlated or uncorrelated, thus ensuring that the likelihood function has a clean factorized form [6]. However, certain uncertainties have partial correlations across the years. There are also partial correlations among uncertainties between the embedded and MC samples. A partially correlated uncertainty is separated into 100%-correlated or uncorrelated components. The magnitude of the correlated com-

ponent will be ρ, and for the uncorrelated part, it will be $\sqrt{1 - \rho^2}$, where ρ is the magnitude of the correlation. For example, a 50% correlation will have a correlated component with 0.5 as magnitude and an uncorrelated part with 0.866 as magnitude.

The expected signal and background yields are effected by nuisance parameters and we parametrize them as $s(\theta)$ and $b(\theta)$. There is a default value $\tilde{\theta}$ for each component of θ. The default value reflects our degree of belief on the real value of θ. The probability distribution function (pdf) $\rho(\theta|\tilde{\theta})$ can then be interpreted as a posterior distribution from measurements of $\tilde{\theta}$. Using Bayes' theorem:

$$\rho(\theta|\tilde{\theta}) = \rho(\tilde{\theta}|\theta) \cdot \pi_\theta(\theta), \tag{3.2}$$

The priors $\pi_\theta(\theta)$ are taken as flat distributions representing no prior knowledge of θ. The likelihood of the measurement can be constrained by using the pdf of $\tilde{\theta}$. After incorporating the nuisance parameters, the likelihood function now becomes:

$$\mathcal{L}(data|\mu, \theta) = \prod_{i=1}^{bins} \frac{(\mu \cdot s_i(\theta) + b_i(\theta))^{n_i}}{n_i!} e^{-\mu \cdot s_i(\theta) - b_i(\theta)} \cdot \rho(\tilde{\theta}|\theta) \tag{3.3}$$

If no excess is observed then upper exclusion limits can be set on the branching fraction using the CL$_\text{s}$ method [7–9]. According to Neyman-Pearson lemma, the likelihood ratio is the most powerful discriminator. The likelihood ratio is used generally in the searches at the LHC for hypothesis testing, and it uses profiling of nuisances. The test statistic is denoted by $\tilde{q}_\mu$ and is given by:

$$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})}, \text{ with } 0 \leq \hat{\mu} \leq \mu \tag{3.4}$$

where $\hat{\theta}_\mu$ refers to the conditional maximum likelihood estimators of θ, while $\hat{\mu}$ and $\hat{\theta}$ refer to the global maximum likelihood estimators for μ and θ. The constraint $0 \leq \hat{\mu}$ ensures that the signal rate cannot be negative. In contrast, the upper constraint

$\hat{\mu} \leq \mu$ is imposed to guarantee that upward fluctuations of data such that $\hat{\mu} > \mu$ are not considered as evidence against the signal hypothesis.

The observed value of the test statistic, $\tilde{q}_{\mu}^{obs}$, is calculated for the signal strength μ, using equation 3.4. Maximum likelihood estimators for the nuisance parameters, under the background-only and signal-plus-background hypotheses, are calculated and are denoted by $\hat{\theta}_0^{\,obs}$ and $\hat{\theta}_{\mu}^{\,obs}$ respectively. They are used to generate toy Monte Carlo pseudo-datasets. These pseudo datasets are used to construct pdfs 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 if they were real data. An illustration of these distributions can be seen in Figure 3.1.
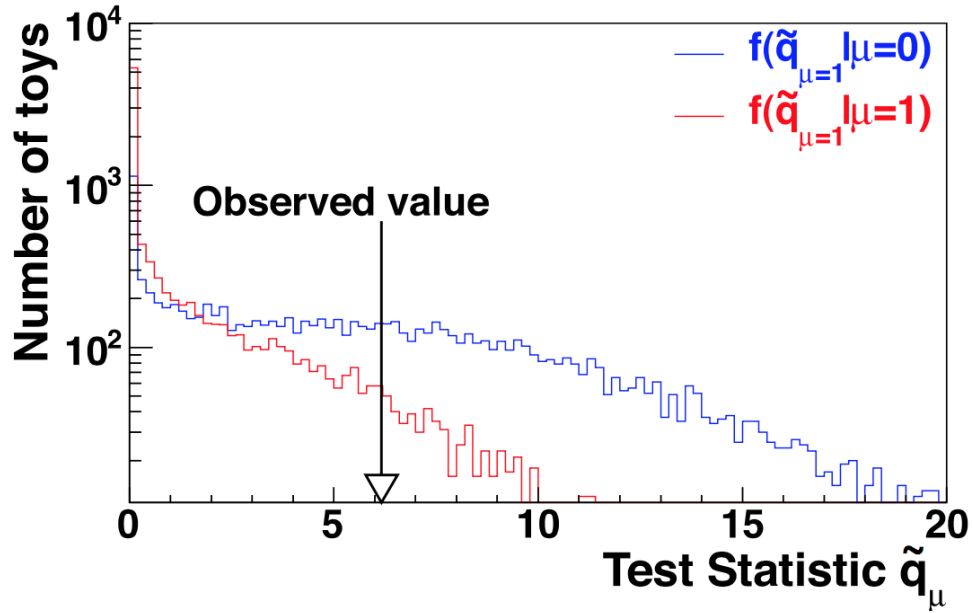
Figure 3.1: Test statistic distributions for ensembles of pseudo-data generated for background-only (blue) and signal-plus-background (red) hypotheses [6].

CL_{s+b} and CL_b correspond to the probabilities of the observations under both hy-

potheses. $\mathrm{CL_{s+b}}$ measures the incompatibility of data under the signal-plus-background hypothesis. $\mathrm{CL_b}$ measures the incompatibility of data under the background hypothesis. These quantities alone are not adequate for hypothesis testing. For example, 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. Also, the incompatibility of the data with the background-only hypothesis alone doesn't tell us that it is indeed compatible with the signal. The ratio of the two quantities referred to as $\mathrm{CL_s}$ helps deal with both the situations described above well. The 95% CL is then arrived at by iterating over μ until we have $\mathrm{CL_s} = 0.05$. The μ thus obtained is denoted as $\mu^{95\%CL}$ and is said to be excluded at 95% CL.

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu{}^{obs}|\text{signal-plus-background}) = \int_{\tilde{q}_\mu{}^{obs}}^{\inf} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu{}^{obs})d\tilde{q}_\mu \qquad (3.5)$$

$$1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu{}^{obs}|\text{background-only}) = \int_{\tilde{q}_\mu{}^{obs}}^{\inf} f(\tilde{q}_\mu|0, \hat{\theta}_0{}^{obs})d\tilde{q}_\mu \qquad (3.6)$$

$$\mathrm{CL_s} = \frac{p_\mu}{1 - p_b} \qquad (3.7)$$

The discussion until now pertains to calculating the observed limits when the data is unblinded. However, when the analysis is performed in a blinded manner, we first calculate the expected limits, which are upper exclusion limits calculated using toy datasets of background-only expectation. An extensive set of pseudo-data is generated using the background-only hypothesis, and $\mathrm{CL_s}$ and $\mu^{95\%CL}$ is calculated for each of them. A distribution is built as a function of the $\mu^{95\%CL}$ calculated for each of these pseudo-data. We then calculate the median expected limit from the 50% quantile of the cumulative distribution function. The $\pm 1\sigma$ and $\pm 2\sigma$ bands are calculated similarly by integrating the distribution until the appropriate quantiles are reached. The expected limits can be used to maximize the sensitivity of the search.

A more stringent median limit corresponds to a more sensitive search.
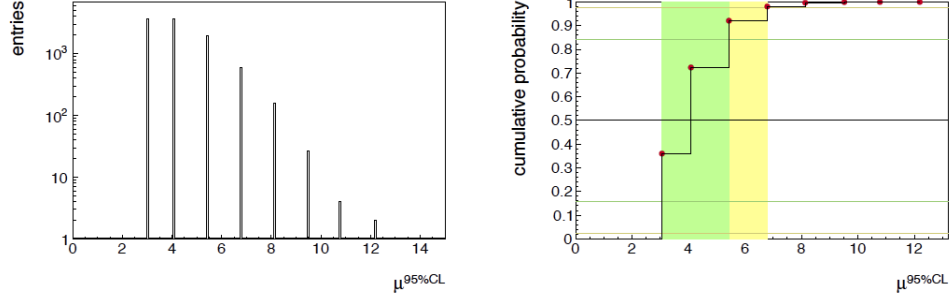
Figure 3.2: (Left) An example of differential distribution of possible limits on μ for the background-only hypothesis. (Right) The cumulative probability distribution of the plot on the left with 2.5%, 16%, 50%, 84%, and 97.5% quantiles defines the median expected limit as well as the 68% and 95% bands for the expected value of μ for the background-only hypothesis.

3.3 Systematic Uncertainties

Several sources of experimental and theoretical systematic uncertainties are taken into account as input to the maximum likelihood fit. These uncertainties affect both the normalization and the shape of the distributions of the different processes. The normalization effect is described by log-normal pdfs and corresponds to a multiplicative factor in the signal or background yields. These pdfs are characterized by the parameter κ and are well-suited for positively valued observables. The log-normal distribution looks like:

$$\rho(\theta|\tilde{\theta}) = \frac{1}{\sqrt{2\pi}\,\ln(\kappa)}\exp\left(\frac{\ln(\theta/\tilde{\theta})^2}{2(\ln\kappa)^2}\right)\frac{1}{\theta} \tag{3.8}$$

47

The distributions' shape is altered when the systematic uncertainties affect the scale of the distribution differently in each bin. Such uncertainties are called shape uncertainties [10] and are modeled using a linear interpolation method [11]. Two other distributions obtained by varying the nuisance parameter by ± 1 standard deviation are used for implementing the shape uncertainties. A parameter is added to the likelihood that smoothly interpolates between these two other distributions.

As the analysis is categorized into different final states, partial and complete correlations between the uncertainties in different categories are treated appropriately and are summarized in Tables 3.1 and 3.2.

The uncertainties to reconstruct a hadronically decaying τ and estimation of its identification efficiency for different p_T ranges are measured using a tag-and-probe method and found to be in the range of 2-3%. The uncertainties for different ranges of p_T are treated in an uncorrelated way. These uncertainties are also considered for the embedded $\tau\tau$ background, where they are treated 50% correlated with the simulation uncertainties. For the embedded samples, triggering on muons before being replaced by τ leptons have an uncertainty of about 4%. This uncertainty is treated as uncorrelated between the three years due to different triggering criteria. Uncertainty due to tracking measurement is also considered in correlated way between $h^{\pm}$ and $h^{\pm}h^{\mp}h^{\pm}$ decay modes and uncorrelated way between $h^{\pm}\pi^0$ and $h^{\pm}h^{\mp}h^{\pm}\pi^0$ decay modes.

Uncertainties due to electrons and muons misidentified as τ_h candidates correspond to 40% and between 10-70%, respectively, for different bins of p_T, η, and τ_h decay modes. The uncertainty due to τ_h energy scale uncertainty is treated as uncorrelated for different decay modes and 50% correlated between embedded and simulated backgrounds and ranges from 0.7 to 1.2%. The uncertainty due to the electron (muon) energy scale for misidentified leptons is independent of the τ_h energy scale and amount to 7% (1%). The effect of lepton energy resolution is found to be

TABLE 3.1

SYSTEMATIC UNCERTAINTIES IN THE EXPECTED EVENT

YIELDS FOR THE H $\to \mu\tau$ CHANNELS.

Systematic Uncertainty	H $\to \mu\tau_{\mathrm{h}}$	H $\to \mu\tau_e$
Muon ID/Isolation	2%	2%
Electron ID/Isolation	—	2%
Trigger	2%	2%
τ_{h} Identification	p_T dep. (2–3%)	—
$\mu \to \tau_{\mathrm{h}}$ ID	10–70%	—
e $\to \tau_{\mathrm{h}}$ ID	—	—
b-tagging veto	2.0–10%	2.0–10%
Prefiring	(0.2–1.3%)	(0.2–1.3%)
Jet energy scale and resolution	3–20%	3–20%
τ_{h} energy scale	0.7–1.2%	—
e $\to \tau_{\mathrm{h}}$ energy scale	1–7%	—
$\mu \to \tau_{\mathrm{h}}$ energy scale	1%	—
e energy scale	—	1–2.5%
μ energy scale	0.4–2.7%	0.4–2.7%
Unclustered energy scale	$\pm 1\sigma$	$\pm 1\sigma$

TABLE 3.2

SYSTEMATIC UNCERTAINTIES IN THE EXPECTED EVENT

YIELDS FOR THE H $\to$ eτ CHANNELS.

Systematic Uncertainty	H $\to$ eτ_{h}	H $\to$ eτ_{μ}
Muon ID/Isolation	—	2%
Electron ID/Isolation	2%	2%
Trigger	2% for SingleElectron, 5.4% for Cross-Trigger	2%
τ_{h} Identification	p$_{\text{T}}$dep. (2--3%)	—
$\mu \to \tau_{\text{h}}$ ID	—	—
e $\to \tau_{\text{h}}$ ID	40%	—
b-tagging veto	2.0–10%	2.0–10%
Prefiring	(0.2–1.3%)	(0.2–1.3%)
Jet energy scale and resolution	3–20%	3–20%
τ_{h} energy scale	0.7–1.2%	—
e $\to \tau_{\text{h}}$ energy scale	1–7%	—
$\mu \to \tau_{\text{h}}$ energy scale	1%	—
e energy scale	1–2.5%	1–2.5%
μ energy scale	—	0.4–2.7%
Unclustered energy scale	$\pm 1\sigma$	$\pm 1\sigma$

negligible for the study under consideration.

The jet energy scale (JES) is affected by several sources, and uncertainty due to them is evaluated as a function of p_T and pseudorapidity. JES's effect is propagated to the BDT discriminator by varying each source of uncertainty by one standard deviation and disseminating them to the fit template for each process and are of the order of 3-20%. The jet energy resolution uncertainties are also taken into account, and they mostly impact the M_{jj}-defined categories. In the particle flow candidate list, jets with $p_T < 10\,\text{GeV}$ are not considered. They fall under unclustered energy. The unclustered energy scale is considered independently for charged particles, neutral hadrons, photons, and very forward particles, affecting both shape and yield, and are treated as uncorrelated. The efficiency to classify a jet as b-tagged is different in data, and simulation and scale factors depend on jet p_T and η are used to correct simulation. The uncertainties in measuring these scale factors are treated as a systematic uncertainty.

The uncertainty to track leptons (e, μ), reconstruct them, and identify and isolate is measured using the tag-and-probe method in data in Z $\rightarrow$ ee and Z $\rightarrow$ $\mu\mu$ events and sums up to about 2% [12–15]. The uncertainty in the measurement of the muon momentum scale is in the range $0.4 - 2.7\%$ for different η_μ, while for the electron momentum scale, it is less than 1%. The selection of events using electron and muon based triggers results in additional 2% uncertainty in the yield of simulated processes.

The misidentification rates in the eτ_h and $\mu\tau_h$ final states are parameterized using a linear function on τ_h p_T. There are a couple of uncertainties that are ascribed per fit function. The normalization uncertainties in the data-driven estimations of misidentified-lepton backgrounds (jet $\rightarrow \tau_h, \mu, e$), is taken from the orthogonal control region as described in Chapter 2. Discriminators with different signal to background efficiency are used to single out τ_h against electrons and muons, which entails an additional 3% uncertainty for the eτ_h channel. The misidentified lepton background in

the eτ_μ and $\mu\tau_e$ final states is affected by different shape uncertainties. The statistical uncertainties arising from both fits of the OS/SS extrapolation factor as a function of the spatial separation between e and μ and lepton p_T are taken into account. The uncertainties from OS/SS extrapolation factor are estimated in the anti-isolated muon control region, which results in an additional uncertainty due to the inversion of muon isolation, with a combined effect of about 20% on the normalization. The predominant source of uncertainties in the simulated background processes, $Z \to ee$, $Z \to \mu\mu$, $Z \to \tau\tau$, WW, ZZ, Wγ, $t\bar{t}$, and single top production is the measurement of the cross-section for these processes.

The theoretical uncertainties affecting the measurement of Higgs boson production cross-section are the factorization and the renormalization scales, choice of the parton distribution functions (PDFs), and the strong coupling constant (α_S). These uncertainties affect the normalization of the signal shape and are taken from Ref. [16]. The variations of the QCD scales results in a 3.9%, 0.5%, 0.9%, and 0.8% uncertainty for the ggH, VBFH, ZH, WH cross-sections, respectively, while variations of PDF+α_s results in 3.2%, 2.1%, 1.3%, 1.9% uncertainty. The acceptance effects are also taken into account while varying the renormalization and factorization scales along with the PDF choice and α_S.

The uncertainty on the H $\to \tau\tau$ branching-fraction includes a 1.7% uncertainty due to missing higher-order corrections, a 0.99% parametric uncertainty on the quark masses, and a 0.62% parametric uncertainty on α_S. The uncertainty on the H $\to$ WW branching fraction includes a 0.99% uncertainty due to missing higher-order corrections, a 0.99% parametric uncertainty on the quark masses, and a 0.66% parametric uncertainty on α_S.

The bin-by-bin uncertainties account for the statistical uncertainties in each bin of the template distributions of every process. The Barlow-Beeston Lite [17, 18] approach is used, which assigns a single nuisance parameter to scale the sum of the

process yields in each bin, constrained by the total uncertainty, instead of requiring separate parameters, one per process. This is useful in order to minimize the number of parameters needed in the maximum-likelihood fit. They are treated uncorrelated between bins, processes, and categories.

The uncertainty of $2.3 - 2.5\%$ on the integrated luminosity affects all processes with the normalization taken directly from the simulation. Shape uncertainties related to the pileup have been considered by varying the weights applied to simulation. The weight variation is obtained by a 5% change of the total inelastic cross-section used to estimate the number of pileup events in data. Other minimum bias event modeling and simulation uncertainties are expected to be much smaller than those on the rate and are therefore neglected.

During the 2016 and 2017 data-taking, pre-firing has impacted the ECAL L1 trigger. A gradual shift in the timing of the inputs of the ECAL L1 trigger in the region at $|\eta| > 2.0$ caused a specific trigger inefficiency. For events containing an electron (a jet) with p_T larger than $\approx 50\,\mathrm{GeV}$ ($\approx 100\,\mathrm{GeV}$), in the region $2.5 < |\eta| < 3.0$ the efficiency loss is $\approx 10\text{-}20\%$, depending on p_T, η, and time. Correction factors were computed from data and applied to the acceptance evaluated by simulation. Uncertainty due to this correction factor is of the order of 1%.

CHAPTER 4

RESULTS

After applying the selection criteria, the expected limits are calculated from a final discrimination variable using the profile likelihood method with asymptotic approximation. Two different fits are considered, with differences in the discrimination variable employed and changes to the selection. The primary results are obtained from a fit to the BDT discriminator after the loose selection. The cross-check results are obtained from a fit to the M_{col} distribution after applying cut-based selection criteria. The fits are performed per channel and category and then combined to set 95% CL upper limits on the branching fraction of LFV Higgs decay in the H $\rightarrow$ $\mu\tau$ and H $\rightarrow$ eτ channels, $\mathcal{B}(\text{H} \rightarrow \mu\tau)$, and $\mathcal{B}(\text{H} \rightarrow \text{e}\tau)$, respectively. The BDT discriminator distributions of signal and background for each category for H $\rightarrow$ $\mu\tau$ and H $\rightarrow$ eτ can be seen in Figures 4.1, 4.2, 4.5, and 4.6 respectively. The M_{col} discriminator distributions of signal and background for each category for H $\rightarrow$ $\mu\tau$ and H $\rightarrow$ eτ can be seen in Figures 4.3, 4.4, 4.7, and 4.8 respectively.

The BDT analysis yields a sensitivity on the Higgs BR of XX(0.13) for H $\rightarrow$ $\mu\tau$ and XX(0.15) for H $\rightarrow$ eτ. The observed blinded and median expected 95% CL upper limits and the best fit branching fractions, for $\mathcal{B}(\text{H} \rightarrow \mu\tau)$ and $\mathcal{B}(\text{H} \rightarrow \text{e}\tau)$, assuming $m_{\text{H}} = 125\,\text{GeV}$, are reported in Tables 4.1 and 4.2. The limits are also summarized graphically in Figure 4.9 and in tabular form in Table 4.5. The median expected limits are significantly improved from previous results.

In a cross-check analysis, a maximum-likelihood fit on M_{col} distribution after applying additional selections yields a sensitivity on the Higgs BR of XX(0.25) for

Figure 4.1. Distribution of the BDT discriminator for the $H \to \mu\tau_{\mathrm{h}}$ process in the BDT fit analysis, in the individual categories compared to the signal and background estimation. The simulated signal corresponds to $\mathcal{B}(H \to \mu\tau) = 5\%$. The bottom panel in each plot shows the fractional difference between the observed data and the background. Blinding criteria: $\frac{s}{\sqrt{s+b}} > 0.2$.

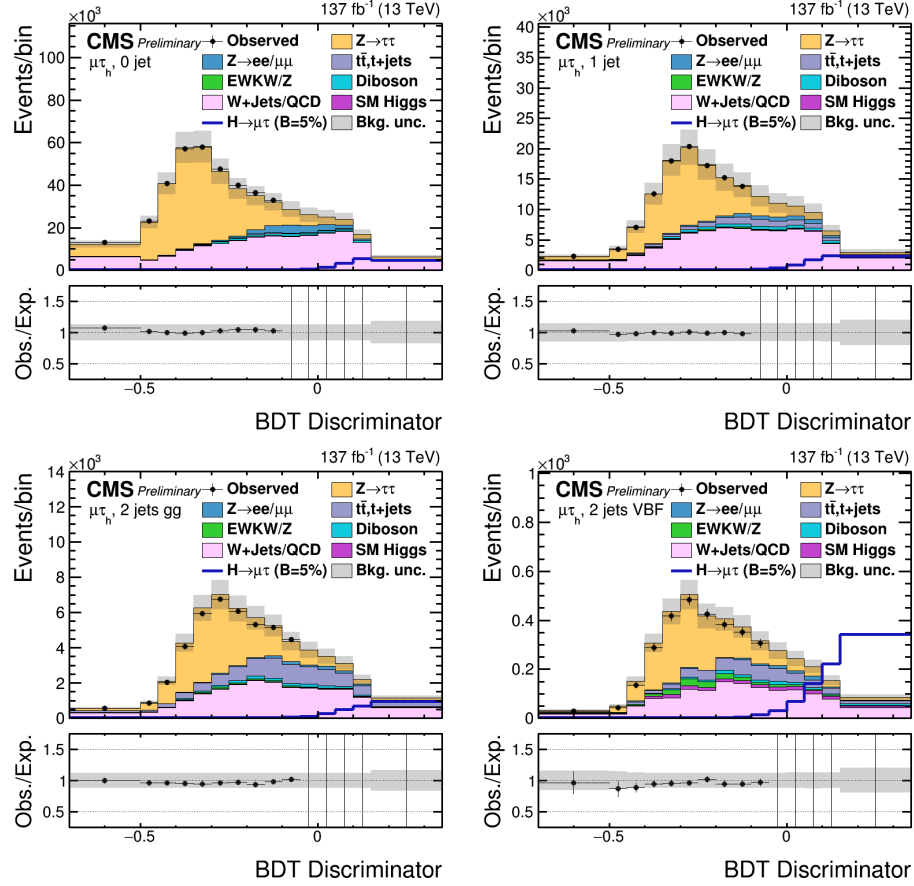Figure 4.2. Distribution of the BDT discriminator for the $H \to \mu\tau_e$ process in the BDT fit analysis, in the individual categories compared to the signal and background estimation. The simulated signal corresponds to $\mathcal{B}(H \to \mu\tau) = 10\%$. The bottom panel in each plot shows the fractional difference between the observed data and the background. Blinding criteria: $\frac{s}{\sqrt{s+b}} > 0.2$.

Figure 4.3. Distribution of the M_{col} discriminator for the H $\to \mu\tau_{\mathrm{h}}$ process in the M_{col} fit analysis, in the individual categories compared to the signal and background estimation. The simulated signal corresponds to $\mathcal{B}(\mathrm{H} \to \mu\tau) = 10\%$. The bottom panel in each plot shows the fractional difference between the observed data and the background. Blinding criteria: $\frac{s}{\sqrt{s+b}} > 0.2$.

Figure 4.4. Distribution of the M_{col} discriminator for the H $\to \mu\tau_e$ process in the M_{col} fit analysis, in the individual categories compared to the signal and background estimation. The simulated signal corresponds to $\mathcal{B}(\mathrm{H} \to \mu\tau) = 10\%$. The bottom panel in each plot shows the fractional difference between the observed data and the background. Blinding criteria: $\frac{s}{\sqrt{s+b}} > 0.2$.
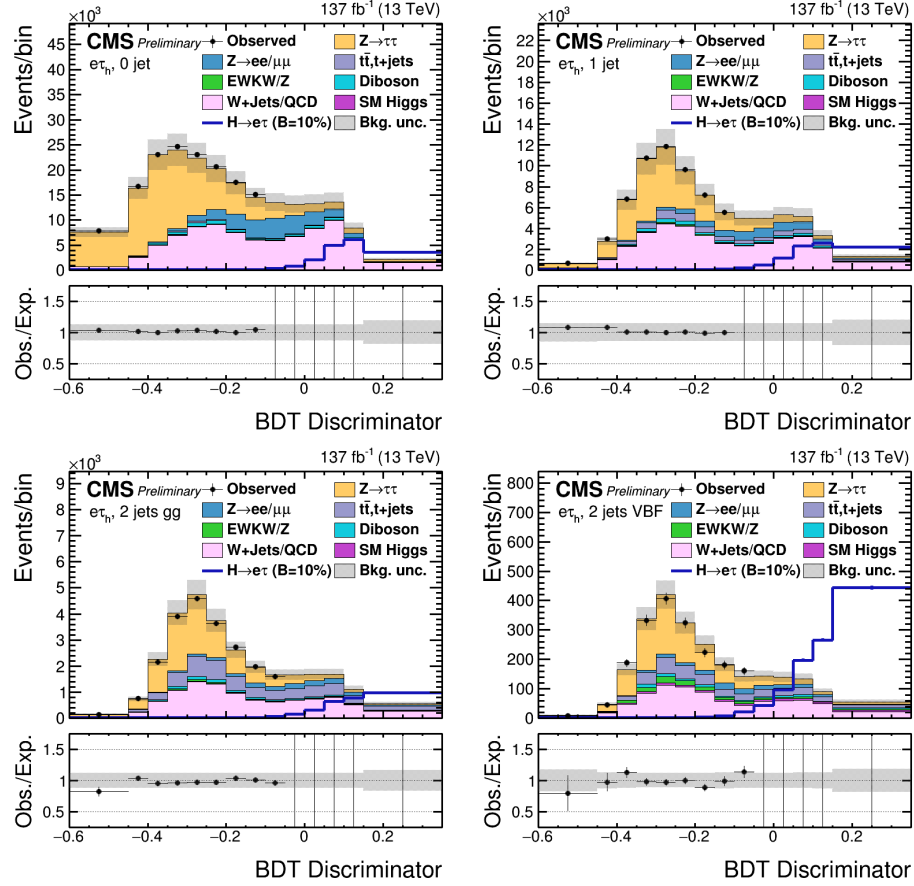
Figure 4.5. Distribution of the BDT discriminator for the $H \to e\tau_h$ process in the BDT fit analysis, in the individual categories compared to the signal and background estimation. The simulated signal corresponds to $\mathcal{B}(H \to e\tau) = 10\%$. The bottom panel in each plot shows the fractional difference between the observed data and the background. Blinding criteria: $\frac{s}{\sqrt{s+b}} > 0.2$.
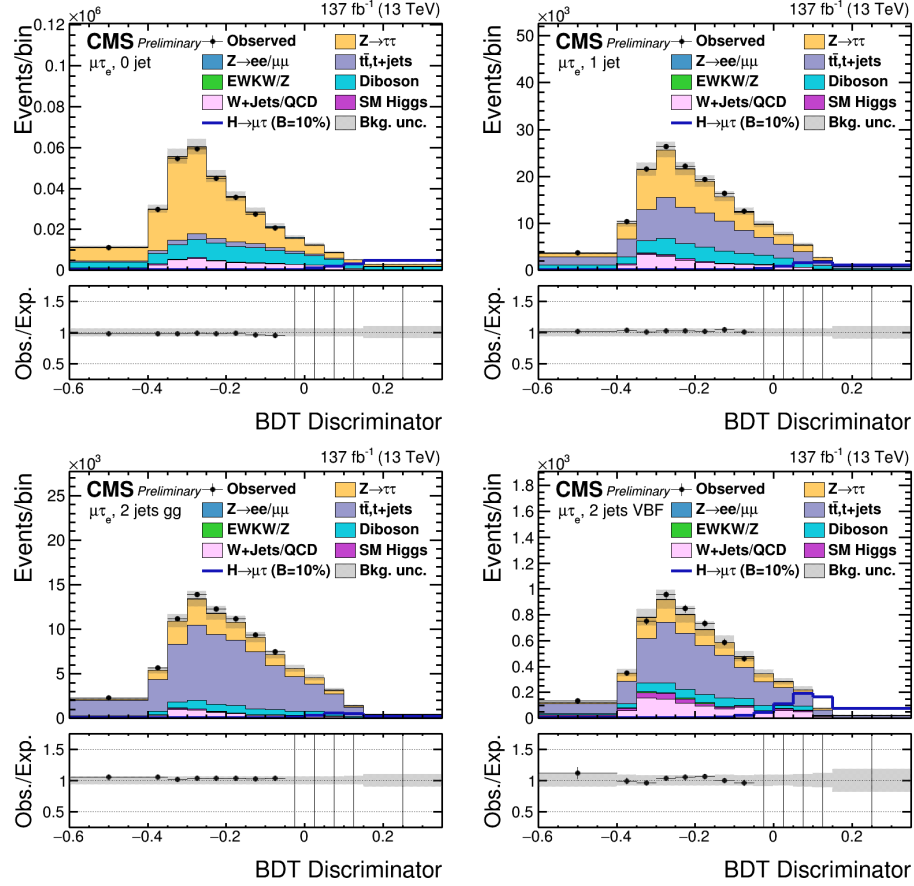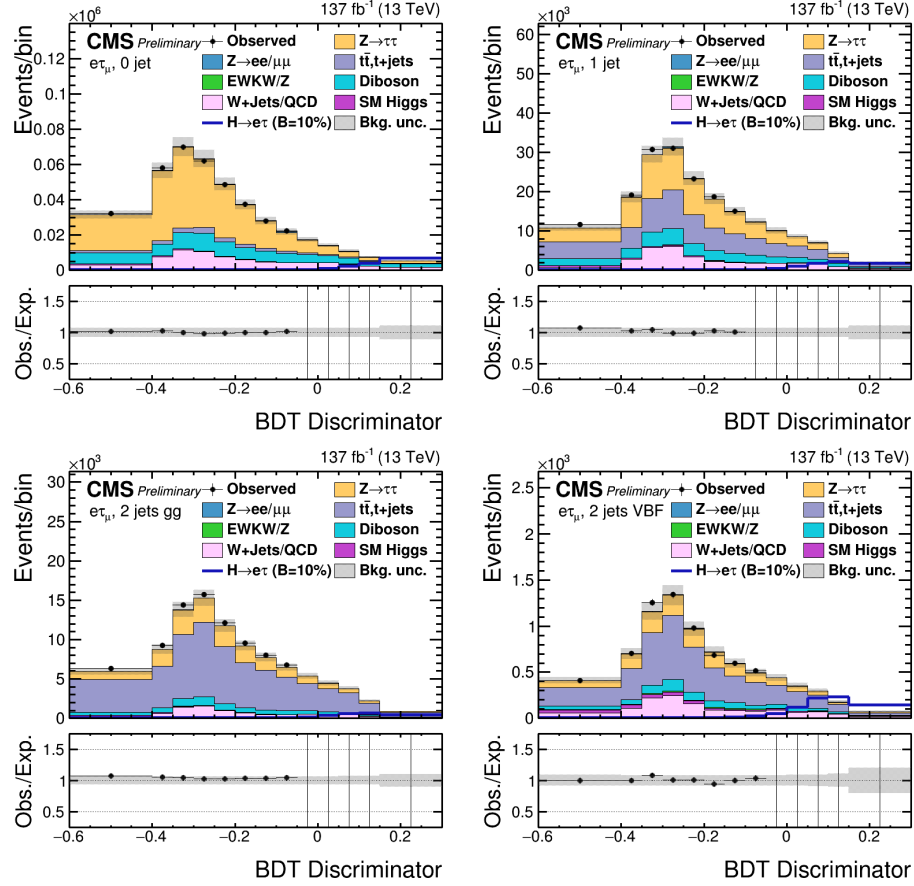
Figure 4.6. Distribution of the BDT discriminator for the $H \to e\tau_\mu$ process in the BDT fit analysis, in the individual categories compared to the signal and background estimation. The simulated signal corresponds to $\mathcal{B}(H \to e\tau) = 10\%$. The bottom panel in each plot shows the fractional difference between the observed data and the background. Blinding criteria: $\frac{s}{\sqrt{s+b}} > 0.2$.
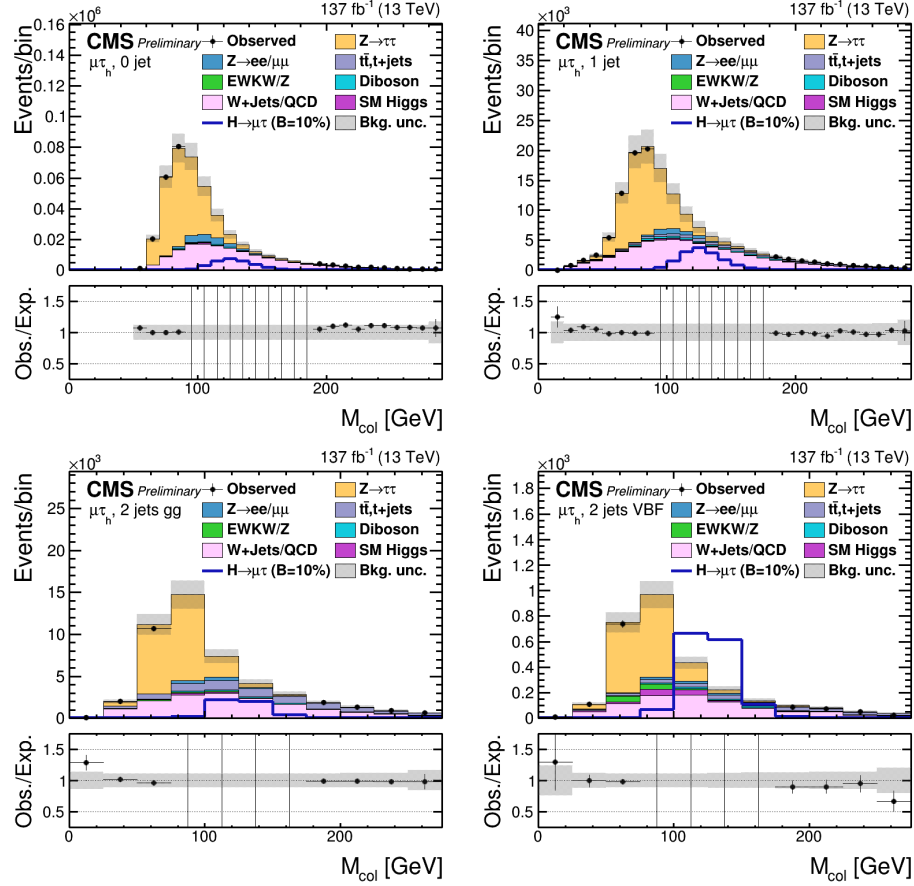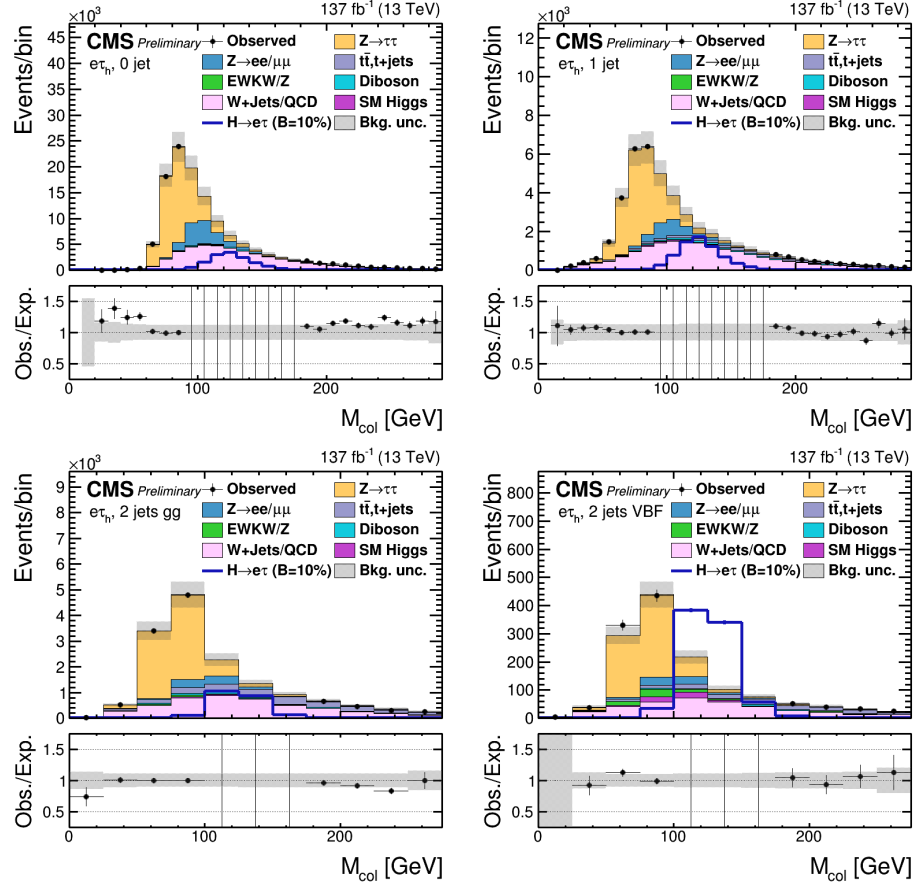
Figure 4.7. Distribution of the M_{col} discriminator for the H $\rightarrow$ eτ_{h} process in the M_{col} fit analysis, in the individual categories compared to the signal and background estimation. The simulated signal corresponds to $\mathcal{B}(\mathrm{H} \rightarrow \mathrm{e}\tau) = 10\%$. The bottom panel in each plot shows the fractional difference between the observed data and the background. Blinding criteria: $\frac{s}{\sqrt{s+b}} > 0.2$.
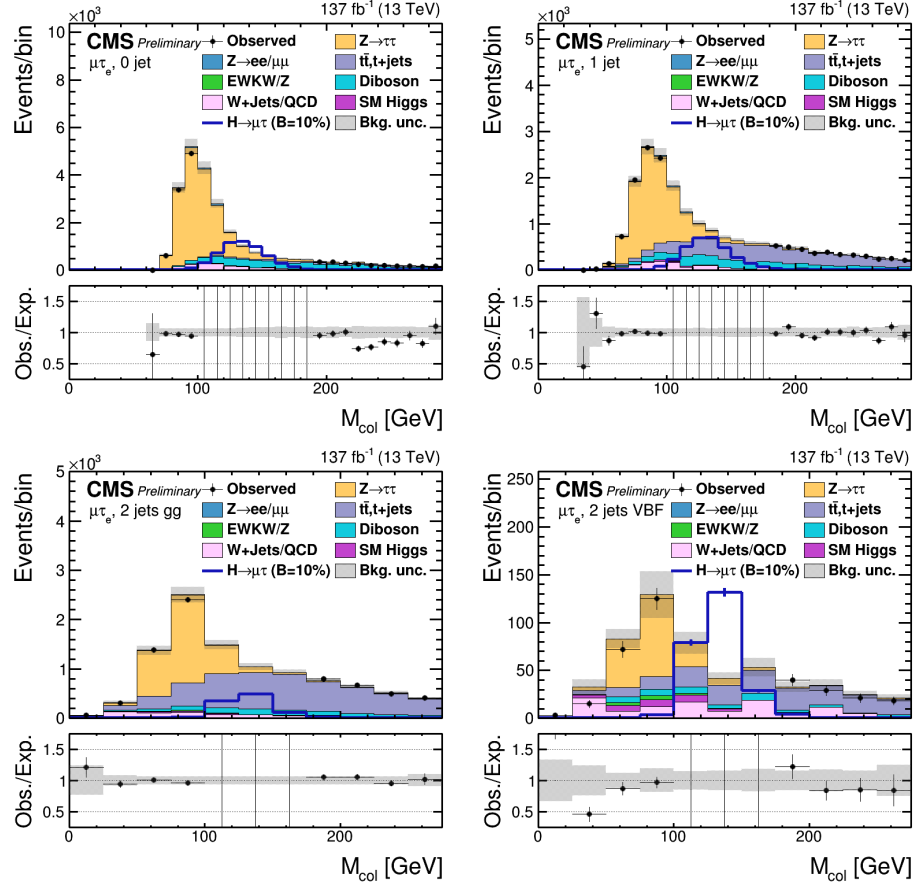
Figure 4.8. Distribution of the $M_{\rm col}$ discriminator for the $\mathrm{H} \to \mathrm{e}\tau_\mu$ process in the $M_{\rm col}$ fit analysis, in the individual categories compared to the signal and background estimation. The simulated signal corresponds to $\mathcal{B}(\mathrm{H} \to \mathrm{e}\tau) = 10\%$. The bottom panel in each plot shows the fractional difference between the observed data and the background. Blinding criteria: $\frac{s}{\sqrt{s+b}} > 0.2$.
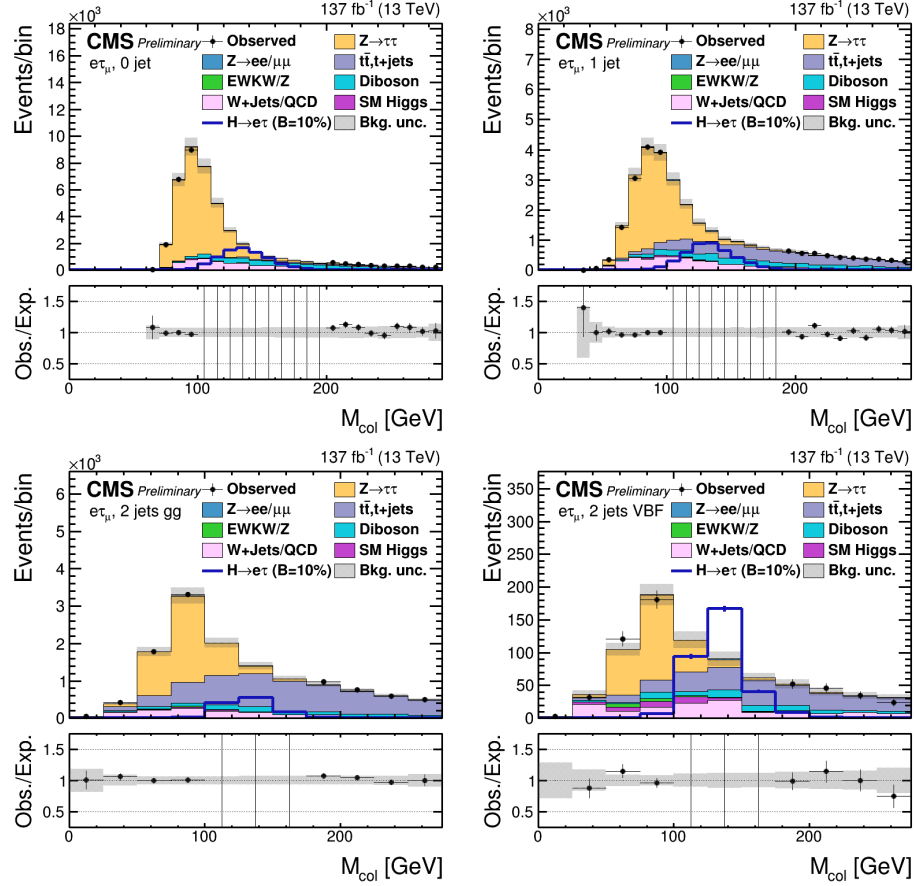
H $\to \mu\tau$ and $XX(0.27)$ for H $\to$ eτ. The observed blinded and median expected 95%
CL upper limits and the best fit branching fractions, for $\mathcal{B}(\text{H} \to \mu\tau)$ and $\mathcal{B}(\text{H} \to \text{e}\tau)$
resulting from the cross-check analysis are reported in Tables 4.3 and 4.4. The
limits are also summarized graphically in Figure 4.10. The BDT fit analysis is more
sensitive than the M_{col} fit analysis, and systematic uncertainties dominate results for
both cases.

The upper limits on $\mathcal{B}(\text{H} \to \mu\tau)$ and $\mathcal{B}(\text{H} \to \text{e}\tau)$ are subsequently used to put
constraints on LFV Yukawa couplings [19]. The LFV decays eτ and $\mu\tau$ arise at tree
level from the assumed flavor violating Yukawa interactions, $Y_{\ell^\alpha \ell^\beta}$, where ℓ^α, ℓ^β are
the leptons (e, μ, τ) of different flavors ($\alpha \neq \beta$). The decay width $\Gamma(\text{H} \to \ell^\alpha \ell^\beta)$ in
terms of the Yukawa couplings is given by:

$$\Gamma(\text{H} \to \ell^\alpha \ell^\beta) = \frac{m_{\text{H}}}{8\pi}(|Y_{\ell^\alpha \ell^\beta}|^2 + |Y_{\ell^\beta \ell^\alpha}|^2),$$

and the branching fractions is given by:

$$\mathcal{B}(\text{H} \to \ell^\alpha \ell^\beta) = \frac{\Gamma(\text{H} \to \ell^\alpha \ell^\beta)}{\Gamma(H \to \ell^\alpha \ell^\beta) + \Gamma_{\text{SM}}}.$$

The SM H decay width is assumed to be $\Gamma_{\text{SM}} = 4.1\,\text{MeV}$ [20] for $m_{\text{H}} = 125\,\text{GeV}$.
The 95% CL upper limit on the Yukawa couplings derived from the expression for the
branching fraction above is shown in Table 4.5. The limits on the Yukawa couplings
derived from the BDT fit analysis results are shown in Fig. 4.11.
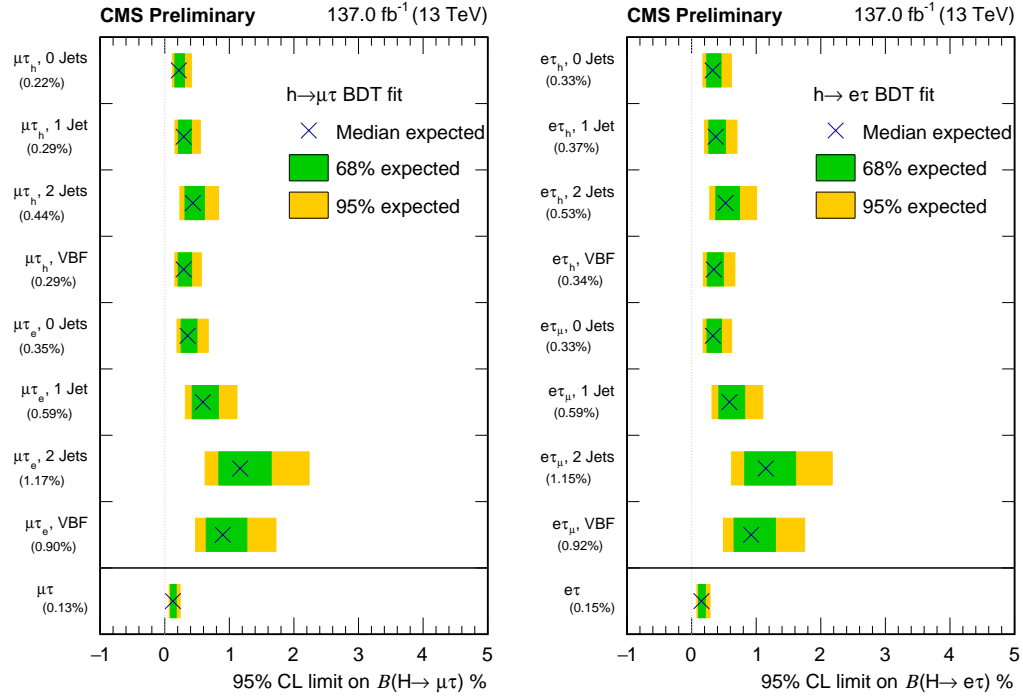
Figure 4.9. Observed (blinded) and expected 95% CL upper limits from the BDT fit analysis on the $\mathcal{B}(\mathrm{H} \to \mu\tau)$ (left) and $\mathcal{B}(\mathrm{H} \to \mathrm{e}\tau)$ (right) for each individual category and combined across all three years.
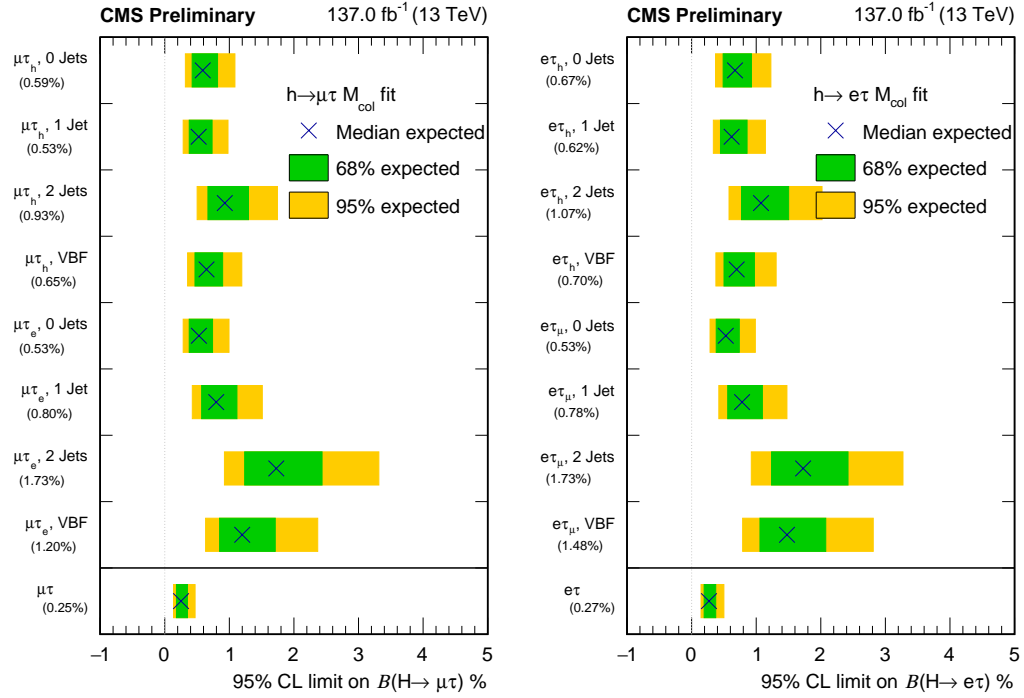
Figure 4.10. Observed (blinded) and expected 95% CL upper limits from the $M_{\rm col}$ fit analysis on the $\mathcal{B}(\mathrm{H} \to \mu\tau)$ (left) and $\mathcal{B}(\mathrm{H} \to \mathrm{e}\tau)$ (right) for each individual category and combined across all three years.
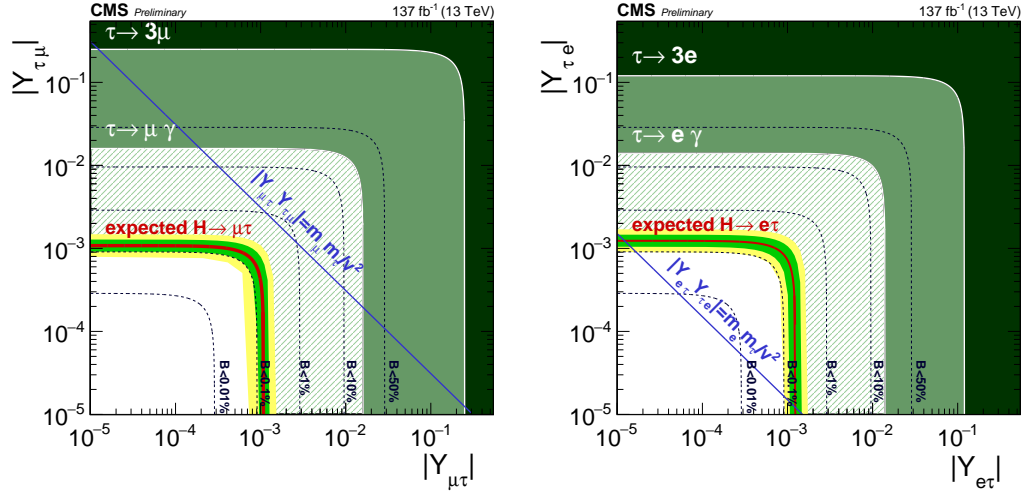
Figure 4.11. Constraints on the lepton flavor violating Yukawa couplings, $|Y_{\mu\tau}| - |Y_{\tau\mu}|$ (left), and $|Y_{e\tau}| - |Y_{\tau e}|$ (right). The expected (red line) and observed (black solid line) (blinded) limits are derived from the results shown in Figure 4.9. Their SM values approximate the flavor-diagonal Yukawa couplings. The green hashed region is derived by the CMS direct search presented in this paper. The green (yellow) band indicates the range that is expected to contain 68% (95%) of all observed limit variations from the expected limit. The shaded regions are derived constraints from null searches for $\tau \to 3\mu$ or $\tau \to 3e$ (dark green) [21] and $\tau \to \mu\gamma$ or $\tau \to e\gamma$ (lighter green) [22]. The blue diagonal line is the theoretical naturalness limit $|Y_{ij}Y_{ji}| \leq m_i m_j/v^2$ [22]

.

TABLE 4.1

OBSERVED AND EXPECTED UPPER LIMITS AT 95% CL AND BEST
FIT BRANCHING FRACTIONS IN PERCENT FOR EACH
INDIVIDUAL JET CATEGORY, AND COMBINED, IN THE H $\to \mu\tau$
PROCESS FROM BDT FIT ANALYSIS.

	Observed (expected) 95% CL limits (in%)				
	0-jet	1-jet	2-jets ggH	2-jets VBF	Combined
$\mu\tau_{\rm e}$	$\leq XX(0.35)$	$\leq XX(0.59)$	$\leq XX(1.17)$	$\leq XX(0.90)$	$\leq XX(0.28)$
$\mu\tau_{\rm h}$	$\leq XX(0.22)$	$\leq XX(0.29)$	$\leq XX(0.44)$	$\leq XX(0.29)$	$\leq XX(0.14)$
$\mu\tau$			$\leq XX(0.13)$		

	Best fit branching fractions (H $\to \mu\tau$) (in %)				
	0-jet(%)	1-jet(%)	2-jets ggH(%)	2-jets VBF(%)	Combined(%)
$\mu\tau_{\rm e}$	XX	XX	XX	XX	XX
$\mu\tau_{\rm h}$	XX	XX	XX	XX	XX
$\mu\tau$			XX		

TABLE 4.2

OBSERVED AND EXPECTED UPPER LIMITS AT 95% CL AND BEST
FIT BRANCHING FRACTIONS IN PERCENT FOR EACH
INDIVIDUAL JET CATEGORY, AND COMBINED, IN THE H $\to$ eτ
PROCESS FROM BDT FIT ANALYSIS.

	Observed (expected) 95% CL limits (in%)				
	0-jet	1-jet	2-jets ggH	2-jets VBF	Combined
eτ_μ	$\leq XX(0.33)$	$\leq XX(0.58)$	$\leq XX(1.13)$	$\leq XX(0.92)$	$\leq XX(0.27)$
eτ_h	$\leq XX(0.33)$	$\leq XX(0.37)$	$\leq XX(0.52)$	$\leq XX(0.34)$	$\leq XX(0.18)$
eτ			$\leq XX(0.15)$		

	Best fit branching fractions (H $\to$ eτ) (in %)				
	0-jet(%)	1-jet(%)	2-jets ggH(%)	2-jets VBF(%)	Combined(%)
eτ_μ	XX	XX	XX	XX	XX
eτ_h	XX	XX	XX	XX	XX
eτ			XX		

68

TABLE 4.3

OBSERVED AND EXPECTED UPPER LIMITS AT 95% CL AND BEST
FIT BRANCHING FRACTIONS IN PERCENT FOR EACH
INDIVIDUAL JET CATEGORY, AND COMBINED, IN THE H $\to \mu\tau$
PROCESS FROM M_{col} FIT ANALYSIS.

	Observed (expected) 95% CL limits (in%)				
	0-jet	1-jet	2-jets ggH	2-jets VBF	Combined
$\mu\tau_{\mathrm{e}}$	$\leq XX(0.53)$	$\leq XX(0.80)$	$\leq XX(1.73)$	$\leq XX(1.19)$	$\leq XX(0.40)$
$\mu\tau_{\mathrm{h}}$	$\leq XX(0.59)$	$\leq XX(0.52)$	$\leq XX(0.93)$	$\leq XX(0.65)$	$\leq XX(0.32)$
$\mu\tau$			$\leq XX(0.25)$		

	Best fit branching fractions (H $\to \mu\tau$) (in %)				
	0-jet(%)	1-jet(%)	2-jets ggH(%)	2-jets VBF(%)	Combined(%)
$\mu\tau_{\mathrm{e}}$	XX	XX	XX	XX	XX
$\mu\tau_{\mathrm{h}}$	XX	XX	XX	XX	XX
$\mu\tau$			XX		

TABLE 4.4

OBSERVED AND EXPECTED UPPER LIMITS AT 95% CL AND BEST
FIT BRANCHING FRACTIONS IN PERCENT FOR EACH
INDIVIDUAL JET CATEGORY, AND COMBINED, IN THE H $\to$ eτ
PROCESS FROM M_{col} FIT ANALYSIS.

	Observed (expected) 95% CL limits (in%)				
	0-jet	1-jet	2-jets ggH	2-jets VBF	Combined
eτ_μ	$\leq XX(0.53)$	$\leq XX(0.78)$	$\leq XX(1.72)$	$\leq XX(1.47)$	$\leq XX(0.42)$
eτ_{h}	$\leq XX(0.67)$	$\leq XX(0.61)$	$\leq XX(1.06)$	$\leq XX(0.70)$	$\leq XX(0.37)$
eτ			$\leq XX(0.27)$		

	Best fit branching fractions (H $\to$ eτ) (in %)				
	0-jet(%)	1-jet(%)	2-jets ggH(%)	2-jets VBF(%)	Combined(%)
eτ_μ	XX	XX	XX	XX	XX
eτ_{h}	XX	XX	XX	XX	XX
eτ			XX		

TABLE 4.5

SUMMARY OF OBSERVED AND EXPECTED UPPER LIMITS AT 95%
CL, BEST FIT BRANCHING FRACTIONS (BR) AND
CORRESPONDING CONSTAINTS ON YUKAWA COUPLINGS IN
PERCENT FOR H $\to \mu\tau$ AND H $\to$ eτ PROCESSES.

	Observed (expected) 95% upper limits (in %)		
	BR limits	Best Fit BR limits	Yukawa coupling constraints
H $\to \mu\tau$	$\leq XX(0.13)$	$\leq XX$	$\leq XX(1.02)\times10^{-3}$
H $\to$ eτ	$\leq XX(0.15)$	$\leq XX$	$\leq XX(1.12)\times10^{-3}$

BIBLIOGRAPHY

1. CMS Collaboration. Observation of the Higgs boson decay to a pair of τ leptons. *Phys. Lett. B*, 779:283, 2018. doi: 10.1016/j.physletb.2018.02.004.

2. A. Hoecker et al. *TMVA - Toolkit for Multivariate Data Analysis*, 2017. `http://tmva.sourceforge.net/`.

3. R. Keith Ellis, I. Hinchliffe, M. Soldate, and J. J. Van Der Bij. Higgs Decay to $\tau^+\tau^-$ A possible signature of intermediate mass Higgs bosons at high energy hadron colliders. *Nucl. Phys. B*, 297:221, 1988. doi: 10.1016/0550-3213(88)90019-3.

4. Aaron Roodman. Blind Analysis in Particle Physics . 2003. doi: arXiv:physics/0312102v1.

5. Albert M Sirunyan et al. An embedding technique to determine $\tau\tau$ backgrounds in proton-proton collision data. *JINST*, 14(06):P06032, 2019. doi: 10.1088/1748-0221/14/06/P06032.

6. CMS and ATLAS Collaborations. Procedure for the lhc higgs boson search combination in summer 2011. Technical report, August 2011.

7. A. L. Read. Presentation of search results: The CL_s technique. *Journal of Physics G*, 28, September 2002.

8. A. L. Read. Modified frequentist analysis of search results (The CL_s method). In *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 Jan 2000*, August 2000.

9. Thomas Junk. Confidence level computation for combining searches with small statistics. *Nuclear Instruments and Methods A*, 434, September 1999.

10. J. S. Conway. Incorporating nuisance parameters in likelihoods for multisource spectra. In *PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva*, January 2011.

11. A. L. Read. Linear interpolation of histograms. *Nucl. Instrum. Meth*, 425, April 1999.

12. 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.

13. 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.

14. Vardan Khachatryan et al. Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS. *JINST*, 11(01):P01019, 2016. doi: 10.1088/1748-0221/11/01/P01019.

15. Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2. Technical Report CMS-PAS-TAU-16-002, CERN, Geneva, 2016. URL https://cds.cern.ch/record/2196972.

16. D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2/2017, 10 2016. doi: 10.23731/CYRM-2017-002.

17. J. S. Conway. Incorporating nuisance parameters in likelihoods for multisource spectra, 2011.

18. Roger Barlow and Christine Beeston. Fitting using finite monte carlo samples. *Computer Physics Communications*, 77(2):219 – 228, 1993. ISSN 0010-4655. doi: https://doi.org/10.1016/0010-4655(93)90005-W. URL http://www.sciencedirect.com/science/article/pii/001046559390005W.

19. Roni Harnik, Joachim Kopp, and Jure Zupan. Flavor violating higgs decays. *JHEP*, 03:26, 2013. doi: 10.1007/JHEP03(2013)026.

20. A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira. Standard model Higgs-boson branching ratios with uncertainties. *Eur. Phys. J. C*, 71:1753, 2011. doi: 10.1140/epjc/s10052-011-1753-8.

21. K. Hayasaka et al. Search for Lepton Flavor Violating Tau Decays into Three Leptons with 719 Million Produced Tau+Tau- Pairs. *Phys. Lett. B*, 687:139–143, 2010. doi: 10.1016/j.physletb.2010.03.037.

22. Roni Harnik, Joachim Kopp, and Jure Zupan. Flavor violating higgs decays. *Journal of High Energy Physics*, 2013(3), Mar 2013. ISSN 1029-8479. doi: 10.1007/jhep03(2013)026. URL http://dx.doi.org/10.1007/JHEP03(2013)026.