This Dissertation

entitled

SEARCH FOR LEPTON FLAVOR VIOLATING DECAYS

OF HIGGS BOSONS

WITH THE CMS EXPERIMENT

typeset with NDdiss2$_\varepsilon$ v3.2017.2 (2017/05/09) on September 11, 2019 for

Nabarun Dev

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

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

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

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

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

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

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

A Dissertation

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

Doctor of Philosophy
in
Physics

by

Nabarun Dev

Colin Philip Jessop, Director

Graduate Program in Physics
Notre Dame, Indiana
September 2019

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

Abstract

by

Nabarun Dev

DEDICATED TO

To my family

CONTENTS

FIGURES

TABLES

ACKNOWLEDGMENTS

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

SYMBOLS

c	speed of light
h	Standard Model Higgs
H	Heavy Higgs
m	mass
e	elementary charge
E	energy

CHAPTER 1

INTRODUCTION

The standard model of particle physics is the most complete description of nature available today. The discovery of the Higgs Boson added another feather to the hat of the standard model...

...expand...

CHAPTER 2

THEORETICAL BASES

2.1 The Standard Model

2.2 Physics beyond the standard model

CHAPTER 3

EXPERIMENTAL SETUP

..introduce...

CHAPTER 4

OBJECT RECONSTRUCTION AND EVENT GENERATION

4.1 Introduction

This chapter is divided into two parts. In the first part, the procedure for the generation of simulated events is described. This is done in several distinct stages with the output of one stage serving as an input for the next. A suite of software packages, developed mostly by the particle and nuclear physics communities, is used to achieve this. This part concludes by detailing the simulated datasets used in the analyses described in this thesis. In the second part of this chapter, the reconstruction of physics objects is described in detail. It starts with a description of the particle-flow algorithm which a global event reconstruction scheme for the entire event. This is followed by descriptions of track, muon and electron reconstructions. Reconstruction of jets is described next followed by description of composite objects used in the analysis such as collinear mass and transverse mass.

4.2 Event Simulation

A pp collision at the LHC, like any hadronic collision, is more complex than the hard interaction of two participating partons. The proton being a composite object, the colliding partons from the hard interaction are accompanied by other quarks and gluons that interact and rearrange themseleves into colorless objects. A pp collision thus consists of: the Hard Scattering which represents the part of the collision where two partons in the initial state interact by exchanging high transverse momentum,

and the Underlying Event that represent the interaction of the everything else in the collision except the partons in hard scattering. In addition to the implementing the above, i.e. physics of a pp collision that produces a bunch of final state particles, the event simulation also has to include interactions of these particles with the CMS detector. Monte Carlo methods, that use generation of random numbers to simulate sampling from a given probability distribution, are used to model the above event simulations [1].

4.2.1 Monte Carlo method

Monte Carlo (MC) methods (named after a famous casino in the city state of Monaco) are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results [2]. In particle physics, these methods play a key role in generation of events and are used primarily for : generation of samples from specified probabilty distributions, and the calculation of integrals. Programs which implement the above method, called MC event generators, use generation of random numbers to make decisions about physics processes. These can range from selection of processes that are generated in the collision, to which decay channel a particle decays in, to making decisions on how the particle interacts with detector material. Usually, each such decision is the result of a draw from a distribution which depends only on the current state the process is in, and not on previous states. The MC generator is provided as input the distributions that represent the physics of the generated particles, their production, their decay modes and their couplings. A MC generator starts by using a pseudo-random number generator that usually outputs a random number between 0 and 1. Although, true random number generation can only be done by physical processess, modern pseudo-random number generators are known to generate numbers with a high degree of randommness. Starting from this distribution, the MC event generator uses one of the various methods such as the

inverse-transform method, or the rejection sampling method to convert this uniform distribution into a desired probablity distribution, $p(x)$. It is then possible to generate random numbers according to this distribution to simulate physical processes.

4.2.2 CMS simulation pipeline

The MC simulation of events in CMS consists of the following sequential steps. The first step is simulation of the Hard Scattering. As mentioned earlier, this represents the primary hard interaction in a collision where two partons in the initial state interact by exchanging high transverse momentum resulting in a final state with two or more partons. The parton density function (pdf) which parametrizes the distributions of the partons inside each hadron are used to model the momenta of incoming partons. It represents the probability of finding a parton of a certain flavour at a certain longitudinal momentum fraction, when the hadron, that contains it, is probed at a certain scale. The PDF are extracted from fits to the data, mainly from ep collisions, and various PDF sets are available for each parton flavour. Commonly used pdf sets include ones provided by the CTEQ, HERA (H1 and ZEUS) and NNPDF collaborations. The LHAPDF library provides a unified C++ interface to all major PDF sets. The matrix element formulation is used to model the hard scattering process to leading order in perturbative QCD, or to higher orders depending on the generator. The next step is simulation of the parton shower. The hadronization and radiation of quarks and gluons in the initial and final states cannot be feasibly encapsulated in the matrix element computation. Parton shower describes these missing parts. The matrix element calculations are combined with the parton shower by one of the different matching schemes which ensure that there is no double counting of terms present in both the matrix element and the partion shower expansion. The matching schemes that are most often used are MLM [3], CKKW [4] and FxFx [5]. The simulation of the Underlying Event comes next. Underlying event includes everything

in the collision that is not associated with the primary hard scattering process. This consists mostly of soft QCD interactions, and implemented using the MC event generators and interfaced with the matrix element simulation. The hadronization of the quarks and gluons is simulated next and it consists of recombination of individual partons into colorless hadrons. Lastly, the decay of short-lived particles is simulated.

An important part of the event generation chain is the simulation of pileup. The protons circulate inside the LHC not as a continuous beam but in discrete closely packed bunches. This leads to more than one proton-proton collision per bunch crossing, i.e. pileup both in-time and out-of-time (see chapter 3). Event generators add pile-up events to the hard scattering samples by randomly simulating soft inelastic collisions and overlapping them. The distribution of the number of pileup interactions in data is hard to predict. MC event generators usually produce events for a scenario with a higher number of pileup vertices, and with a flat disctribution of number of vertices . This is afterwards reweighted to match the observed distribution of pileup interactions in data.

Several MC generators have been developed. Some of these can produce all components of the above simulation pipeline while some calculate only the matrix element and need to be interfaced with other generators for the simulation of remaining parts. Pythia [6] and Herwig [7] can produce the entire chain while Powheg [8–13], aMC@NLO [14] and Madgraph [15] produce up to matrix element stage. Powheg and aMC@NLO can perform next-to-leading order calculations.

Finally, the Geant4 (GEometry ANd Tracking) [16] package is used to simulate the interaction of physical particles after the collision, produced by pipeline described above, with a sophisticated and complex simulation of the detector itself. This simulated detector response is used as input for the same physics reconstruction algorithms (desribed in the next section), that are used to reconstruct the data, thus enabling a direct comparison of the two. If differences are observed in the behavior of these

reconstruction algorithms for MC events in comparsion to observed data, the MC events are tuned to the behavior observed in data.

4.3 MC samples used for the analyses

The ggH and VBF Higgs boson samples are generated with POWHEG 2.0 while an extension of POWHEG 2.0 [17] is used for the WH and ZH simulated samples. For the H $\rightarrow \mu\tau_e$ analysis, only the gluon fusion (ggH) production mode has been considered. Samples are generated for a range of H masses from 200 to 900 GeV.

The Z + jets and W + jets processes are simulated using the MG5_aMC@NLO generator at leading order (LO) with the MLM jet matching and merging scheme. The same generator is also used for diboson production which is simulated at next-to-LO (NLO) with the FxFx jet matching and merging scheme. POWHEG 2.0 and 1.0 are used for top quark-antiquark ($t\bar{t}$) and single top quark production, respectively. The POWHEG and MADGRAPH generators are interfaced with PYTHIA 8 for parton showering, fragmentation, and decays.

As mentioned earlier in this chapter, additional pileup interactions are also a part of the MC generation pipeline. All simulated samples are reweighted to the pileup distribution observed in data. An event weight is applied based on the number of simulated pileup events and the instantaneous luminosity per bunch-crossing, averaged over the run period. Several other scale factors are used to reweight the events in order to get the MC simulation to match the data closely. These include scale factors based on trigger, lepton identification, lepton isolaton and b-jet tagging efficiencies.

4.4 Physics Object Reconstruction

This section begins with the description of the particle-flow algorithm followed by reconstruction of tracks and vertices, electrons, muons, jets and other physics objects.

4.4.1 Particle Flow

The overarching algorithm used by CMS to produce a unified global (synchronized for all sub-detectors) description of an event is the particle-flow (PF) algorithm [18]. The idea behind the PF algorithm is that if the basic building blocks or elements from the various sub-detector can be correlated in a well-defined way, then the description of the event and that of each particle in it can be refined by using the global information from the entire detector. The ALEPH experiment at the CERN LEP collider was the first experiment to use such a holistic approach towards event reconstruction. The CMS experiment, owing to its very granular layers of sub-detector, is the first hadron collider experiment to successfully use particle-flow. The first step of the PF algorithm is the linking of the several building-blocks or PF elements that a single particle can give rise to, across different sub-detector layers. The link algorithm tests pairs of neighbors in the $\eta - \phi$ plane and combines (links) them to form PF blocks. Reconstruction and identification algorithms are run according to a predefined sequence in each of these PF blocks. First, muon candidates are reconstructed and identified. If a muon candidate successfully passes PF quality criterion, the PF elements associated with it are removed from the block. Electron reconstruction proceeds next with electron candidates successfully becoming PF electrons if their tracks in the tracker, when extrapolated, have a corresponding energy deposit in the ECAL. The reconstruction procedure of muons, electrons and tracks are discussed in detail in the sections 4.4.3, 4.4.2, 4.4.4. The PF block now consists of photons and hadrons. To reduce fake track identification, tracks with momentum uncertainty larger than the resolution of the calorimeters are removed at this stage. The remaining tracks are then associated with charged hadrons. The remaining calorimeter energy deposits are then associated with photons (ECAL) and hadrons (HCAL). In this manner, PF finally produces a list of all electrons, photons, muons, charged hadrons and neutral hadrons in the event with optimally determined direction, charge and energy.

4.4.2 Track and primary vertex reconstruction

Tracks of charged particles, that traverse the CMS tracker (described in section **??**), are reconstructed [19] using hits from the pixel and strip detectors in the tracker. Hits are reconstructed by clustering signals above specified thesholds in the pixel and strip channels, and then estimating the cluster positions and uncertainties in a local orthogonal system plane of each sensor. During track reconstruction, a translation is made between the local coordinate system of these hits to the global coordinate system of the tracks. The software used to reconstruct tracks by CMS is called the Combinatorial Track Finder (CTF) and is adaptation of the Kalman filter [20]. Tracks are reconstructed using a iterative procedure with the basic idea being, tracks that are easiest to find (e.g., high p_T tracks, and tracks produced near the interaction region) are searched in the initial iterations, with subsequesnt iterations looking for more difficult sets of tracks (e.g., low p_T tracks , or tracks produced far from the interaction region). Hits unambiguously assigned to the track in the previous iterations are removed for the subsequent ones, thus reducing the combinatorial complexity. Each iteration can be divided into four sequential steps.

The first step is seed generation which provides initial track candidates that define the starting trajectory parameters and associated uncertainties of potential tracks. Charged particles follow helical paths in the quasi-uniform magnetic field of the tracker, requiring a total of five parameters to determine the trajectory. These five parameters are extracted using two or three hits in the inner region of the tracker. The seeds are constructed in the inner part (and then tracks constructed outwards, and not in the opposite manner) because the high granularity of pixel detectors (in contrast to outer strip layers) ensure that low fraction of channels are hit. Also, particles like pions and electrons interact inelastically with tracker material or lose energy due to bremsstrahlung radiation as they traverse through the tracker to its outer regions, making the idea of construcing seeds in the inner region a better choice.

The second step in track generation is track finding which is closely based on the Kalman filter. It extrapolates the seed trajectories along the expected path of a charged particle, beginning with an estimate of the track parameters provided by the trajectory seeds generated in the last step. It then uses the location and uncertainty of detected hits, and estimations of effects such as Coulomb scattering, at successive detector layers, to build track candidates, updating the parameters at each layer. First, using the parameters of the track candidate, evaluated at the current layer, an analytical extrapolation is done that determines which adjacent layers of the detector the trajectory can intersect. This takes into account the current uncertainty in that trajectory just like a Kalman filter. Secondly, a search is perfomed for silicon modules in these layers that are compatible with the extrapolated trajectory. All compatible modules in each layer are then grouped into mutually exclusive groups, such that no two modules in each group overlap. The collection of all hits from from one such module group forms a group of hits. Finally, new track candidates are formed by adding exactly one of the compatible hits from each group, to each original track candidate. The modules in a given group are mutually exclusive and a contribution of more than one hit from each group is not expected. The trajectory parameters of the new candidates are then updated by combining the information from the added hits with the extrapolated trajectory of the original track candidates. Fig 4.1 illustrates the reconstruction efficiency of tracks in case of isolated muons.

The third step in track generation track fitting. In this step the collection of hits from the last step are refitted using a Kalman filter and smoother, to provide a best possible estimate of parameters for each track trajectory. The procedure described above, in conditions as challenging as the LHC, can yields several fake tracks that are not associated with any charged particle passing through the tracker. The fourth and final step applies several quality requirements to the set of reconstructed tracks and substantially reduces the fake contribution. The requirements are based on criteria

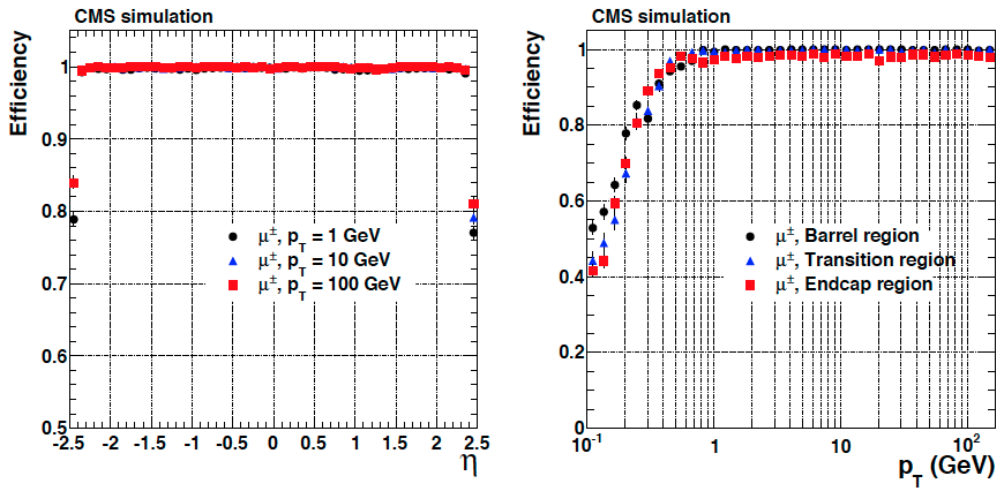


Figure 4.1: Track reconstruction efficencies for single isolated muons as a function of η and p_T [19].

such as the minimum number of layers the track has hits in, how compatible its origin is with a primary vertex, how good a fit it yields etc.

Proton-proton interaction vertices are reconstructed by selecting tracks that are produced promptly in the primary interaction region. The selected tracks are then clustered on the basis of their z-coordinates at their point of closest approach to the centre of the beam spot, which represents a 3-D profile of the region where the LHC beams collide inside the CMS detector. The exact positions of the vertices are then obtained from these clustered candidates, by using a fitting procedure, called the adaptive vertex fitter [21]. The vertex which has the largest sum of squared transverse momenta of tracks originating from it is considered the primary interaction vertex.

4.4.3 Muon Reconstruction

Hits in the muon system (described in section **??**) and tracks (muons being charged particles leave tracks in the tracker) from the tracker are used to reconstruct muons [22]. When muons traverse a muon subdetector (such as RPC, CSC or DT) in the muon system, they ionize the gas in the chambers. The electrical signals

produced on the wires and strips as a consequence of the ionization are read out by electronics systems that associate these "hits" with well-defined locations in the detector. Various algorithms depending on the subdetector technology are used to reconstruct these hits. Reconstruction of muon tracks using these hits first proceeds independently of track reconstruction in the tracker. These tracks, called *standalone-muon tracks*, are built using these reconstructed hits from the muon system using a Kalman filter. Muon tracks are also built inside-out by propagating tracker tracks (described in previous section) with transverse momentum above 0.5 GeV to the muon system and matching them to (straight-line) segments of hits in DT or CSC. If a match is found, the tracker track qualifies as a *tracker muon track*. Muon tracks are also built outside-in by matching standalone-muon tracks with tracker tracks, and combining information from both using a Kalman filter fit. These are called *global muon tracks*. The global muon reconstruction is especially efficient for muons leaving hits in several muon stations. The *tracker muon* reconstruction is more efficient for low p_T muon candidates but it can cause fake muon tracks due to hadronic particles which *punch-through* to the innermost muon stations. The *global muon* reconstruction has high efficiency for muons penetrating through more than one muon station, and reduces the muon misidentification rate compared to tracker muons. Combining both *tracker muon tracks* and *global muon tracks*, the efficiency for reconstructing a muon is as high as 99%. The particle-flow algorithm applies a set of requirements, based on various quality parameters from muon reconstruction as well as information from other sub-detectors, to reconstructed candidates. The PF muon candidates used in the analyses described in this thesis were required to satisfy the following set of criterion to be identified as a muon:

- Must be a global muon or a tracker muon.
- Must have at least one hit in the pixel subdetector of the tracker

- χ^2 of the compatibility between the position of the standalone and trackers tracks < 12
- Transverse impact parameter of the associated tracker track with respect to the primary vertex $d_{xy} < 2mm$
- Longitudinal distance of the (origin of)associated tracker track with respect to the primary vertex $d_z < 5mm$
- constraints on muon segment matching compatibility between tracker and muon system dependent on if it is a global muon

The efficiency of the above selection for muon identification is illustrated using a plot from a study performed by the CMS Muon Physics Object group in Fig. 4.2. As can be seen from the plots, there is a difference in the efficiencies in data and MC simulation. This is corrected using a set a of scale-factors applied as a function η and $\mathrm{p_T}$ to adjust the efficiency in simulation to get it to match the efficiency in data.

The momentum of muons is measured by CMS using one among different possible ways involving the tracker and muon system [23] and then using the PF algorithm to refine this measurement exploiting information from the full event.

4.4.4 Electron Reconstruction

Besides muons, electron form the other primary part of the final state of the decay we are searching for in this thesis. Electrons, in the CMS, are reconstructed using clusters of energy formed in the ECAL (described in section **??**) and associating them with tracks from the tracker [24]. The reconstruction of electrons is made complicated by the fact that they can radiate a significant amount of energy before reaching the ECAL. This happens due to the radiation of bremsstrahlung photons caused by the ineraction of electrons with atoms as they pass through the tracker. This loss can

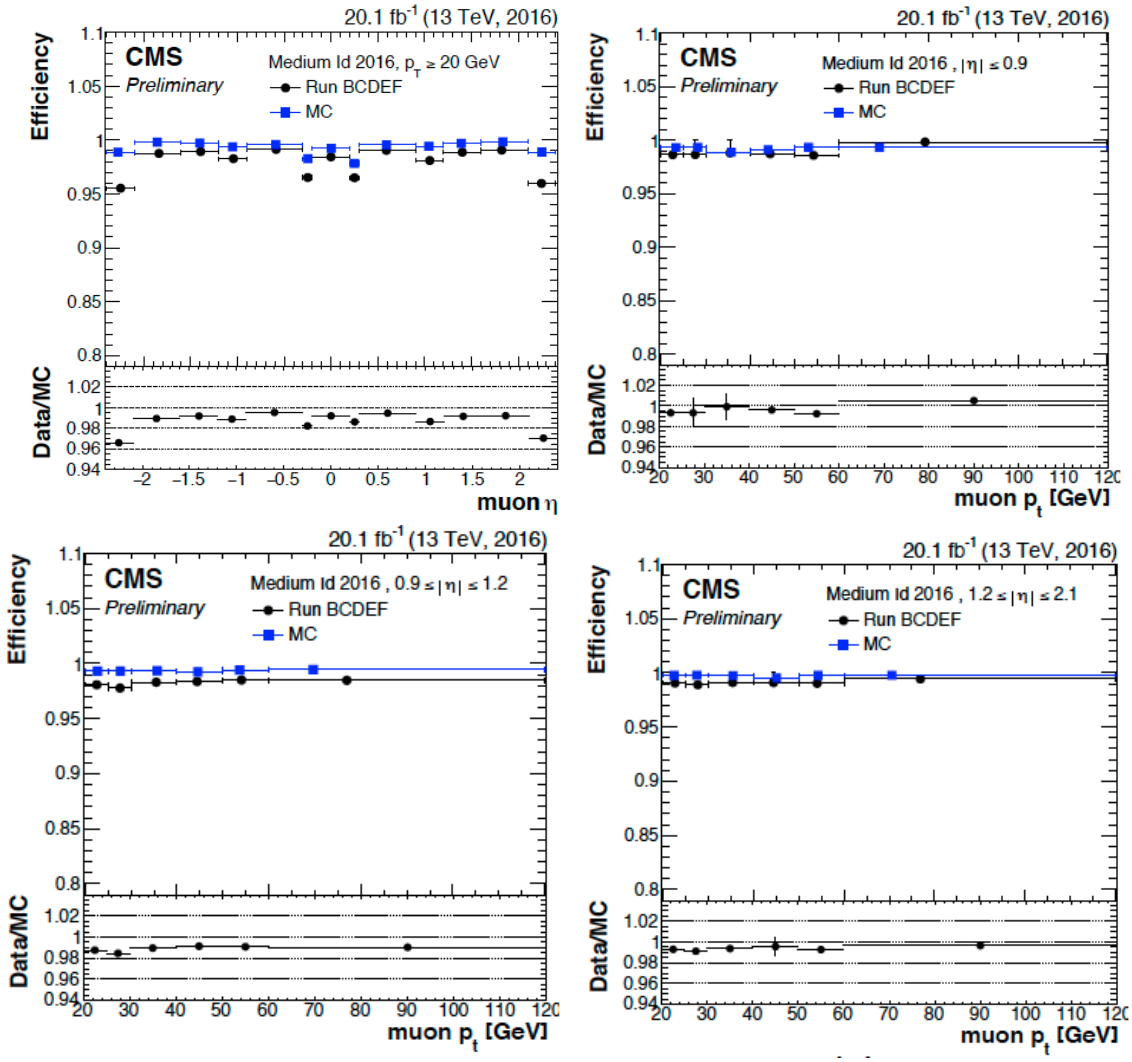


Figure 4.2: Efficiency of muon identification as a function of η and p_T, for data (black) and simulation (blue)

range from 33% to as high as 86% depending on η (as a consequence of the fact that the amount of detector material the electron has to cross is *eta* dependent). In order to measure the an electron's energy, clustering algorithms thus need to take into account the energy from these bremsstrahlung photon showers, together with the deposit made in the ECAL by the electron. The energy from these radiated photons spreads primarily in the ϕ direction, owing to bend in electron trajectory in the magnetic field of CMS. The spread in the η direction is relatively small. These facts are used by the clustering algorithms.

The algorithm used to cluster the electron energy deposit in the ECAL barrel is called the *hybrid* algorithm. It exploits the above property of the electron shower shape, and uses the geometry of the ECAL to form clusters that are narrow in η direction but wide in ϕ direction. Starting with a (seed) crystal containing the largest amount of energy deposited in a considered region above a certain threshold (1 GeV), it adds 5x1 arrays of crystals in $\eta \times \phi$ around the seed crystals in both directions of ϕ if the energy contained in the arrays is above another predefined threshold (0.1 GeV). Contiguous arrays are merged into clusters, and finally a electron supercluster is formed from all such strip clusters which have at least one seed strip with energy above another predefined threshold (0.35 GeV). The position of the supercluster is computed as the energy-weighted mean of the cluster positions, whereas its energy is simply taken as the sum of the energy of all its constituent clusters. In the ECAL endcap a different clustering algorithm is used owing to different geometrical arrangement of the crystals. This algorithm called the '5x5' algorithm starts similarly with a seed crystal with maximum energy in a local region, and satisfying the minimum energy requirement of 0.18 GeV. Clusters of 5x5 crystals are progressively grouped around the seed crystal, making a supercluster, if the total cluster energy exceeds 1 GeV and they are withing ± 0.7 and ± 0.3 respectively in η and ϕ around the seed crystal. The position and energy of the supercluster is calculated in the same manner as the barrel.

The energy from the preshower is also added into the supercluster, using it's most energetic cluster and it's maximum distance in ϕ to other clusters, and extrapolating it to the preshower plane to define the spread in the preshower. The thresholds used in the above algorithms were optimized using simulation and adjusted durting data taking periods.

The standard track reconstruction (section 4.4.2) is not efficient for electrons. This is because the standard approach is compromised of the large radiative losses in the tracker leading to a poor estimation of track parameters [19]. Therefore, a dedicated tracking procedure is used for electron candidates that used information not only from the tracker but also the ECAL. Just like the standard track reconstruction procedure, the first step in electron track reconstruction is seeding. This is done in two ways and the results are then combined. In the first method, superclusters from ECAL are used. As mentioned earlier, owing to strong magnetic field, the bremsstrahlung photons emitted by the electrons deposit energy in the ECAL at η values similar to that of the electron, but at diffrent ϕ leading to a spread. The ECAL supercluster algorithms described above recover this energy. The position and energy of these reconstructed superclusters along with the assumption that the electrons originated close to the center of the beam spot can be used to constrain the trajectory of the electron through the tracker. Hits in the first layers of the trackers compatible with these trajectories are deemed electron seeds. In the second method of seeding, the "opposite" is done. Tracks constructed by the regular tracking algorithm are extrapolated to the ECAL and matched with a supercluster. The seeds corresponding to such matching tracks are retained as electron seeds. The seed collections from these two methods are merged leading to a increase in overall efficiency of the seeding procedure. These seeds are then used to intiate electron track finding phase and fitting phases. This track finding procedure is similar to that used in standard tracking except for small adjustments. The χ^2 fit thresholds used

by the Kalman filter to decide whether a hit is compatible with a trajectory (see section 4.4.2) is weakened to accommodate tracks that deviate from their expected trajectory because of bremsstrahlung. Similar adjustments are made to the penalties assigned to track candidates for passing through a tracker layer without being assigned a hit. The final track fit uses a modified version of the Kalman filter, called the Gaussian Sum Filter (GSF), to account for the fact that the energy loss of an electron traversing the tracker material is non-Gaussian. This makes it unsuitable to use a conventional Kalman filter algorithm which assumes gaussian distribution. The GSF technique deals with this by approximating this non-Gaussian energy-loss distribution as the sum of several Gaussian functions, and is found to perform much better than the regular fitting procedure.

Finally, electron candidates are constructed by associating a electron track (called GSF track) produced by the above procedure with a supercluster in the ECAL. For ECAL-seeded candidates this association is made by a geometrical matching in $\eta - \phi$, while for tracker-seeded candidates a multivariate (MVA) technique that combines information from supercluster and GSF track is used. The electron charge is estimated using a combination of three procedures involving the use of the GSF track curvature, use of ECAL supercluster position and its relative postion in ϕ to that of the first hit in the GSF track, and also by using KF tracks that have common hits with the GSF tracks. The combination of a best vote of three methods reduces the charge misidentification probability to 1.5% compared to 10% when using just the GSF track curvature method. Like other variables, the momentum of electrons is also estimated using a combination of tracker and ECAL measurements.

Further, several quality requirements are used on reconstructed electron candidates to identify (real/signal) electrons to supress fake sources such as photon conversions, jets misidentified as electrons etc. These requirements are based on variables that fall into three broad categories: variables that compare measurements from

ECAL and the tracker, variables that come only from ECAL (such as transverese shape of electromagnetic showers, ratio of energy fractions deposited in the HCAL to the ECAL) and purely tracking based variables (such as information from GSF track, difference between the information from GSF and KF-fitted tracks). These variables can be used in two ways: a cut-based method that uses the variables above directly to apply threshold requirements, or a multivariate (MVA) technique that uses all these variables as an input to a Boosted Decision Trees classifier to obtain a combined discriminator variable on which a threshold is applied. The BDT based method has much better performance as is illustrated in Fig 4.3Two separate BDTs are trained depending on whether electron is required to pass a HLT triggering requirement or it is not. The trigger selection used in the analyses described in this thesis uses trigger based on muons. The BDT based identification criterion for non-triggering electrons is thus used in this analysis. The threshold corresponding to the working point used in this analysis has an efficiency of approximately 80%. The difference in efficiencies of electron identification based on the above criteria in data and MC simulation is corrected using a set a of scaled factors, applied as a function η and p_T. This adjusts the efficiency in simulation to get it to match the efficiency in data.

4.4.5 Hadronic tau leptons

Tau leptons decay hadronically in several ways or decay modes. The primary decay modes consist of: one charged hadron and up to two neutral pions, or three charged hadrons. The algorithm that is used to reconstruct hadronically decaying tau leptons in CMS is called the hadrons-plus-strips (HPS) algorithm [25, 26]. This algorithm proceeds in two steps. In the first step, the topology of the candidate is checked to match the topology of one of the decay modes. The second step consists of a MVA-based discriminator (built using variables such as lifetime information, decay mode etc.), that is used to reject electrons, muons, quarks or gluon jets wrongly

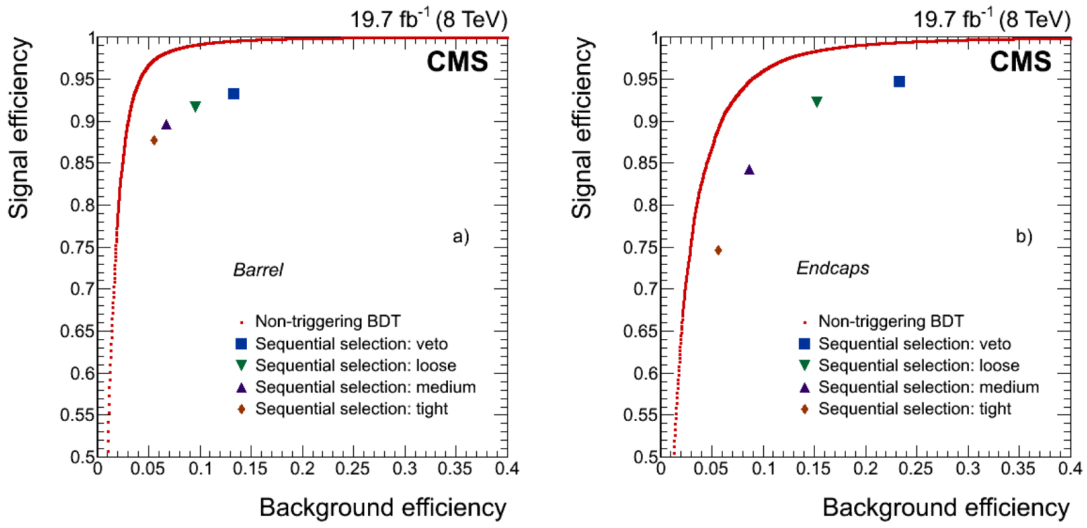


Figure 4.3: Performance of the BDT-based electron identification algorithm (red dots) compared with results from several working points of cut-based selection for electron candidates in the ECAL barrel (left), and endcaps (right).

identified as hadronic taus. The reconstruction efficiency is also improved, in the case of converted photons from a neutral-pion decay, by considering PF photons and electrons from a strip along the ϕ direction. The final states in the analyses presented here do not contain hadronically decaying taus. Hence, all events which contain hadronically decaying taus are rejected.

4.4.6 Jet Reconstruction

Jets are clusters of particles that are experimental signatures of quark and gluons which hadronize (due to color confinement) producing a narrow spray or "jet" of particles [27]. Jets are reconstructed in CMS by clustering PF objects. In order to group together objects into a jet, CMS uses the anti-k_T clustering algorithm. This belongs to a broader class of clustering algorithms called sequential clustering algorithms which cluster objects into jet in a sequential order following a predefined set of rules. The general form of a sequential clustering algorithm is based on the quantities d_{ij}, which represents the distance between two entities, and d_{iB} which

represents the distance of the i-th entity from the beam axis. These distances are defined as:

$$d_{ij} = min(k_{ti}^{2p}, k_{tj}^{2p})\frac{\Delta_{ij}^2}{R^2}$$

$$d_{iB} = k_{ti}^{2p}$$

where $\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$, k_{ti} is the transverse momentum of the i-th entity and R is the radius parameter which is set as 0.4. The parameter p governs the relative power of energy versus geometrical scales and the particular value of $p = -1$ defines the anti-k_T clustering algorithm. The algorithm first computes distance d_{ij} between all entity pairs present at that stage. If the minimum of those distances (say between entity i and j) is smaller than the minimum distance d_{iB} of any entity from the beam axis, those entities i and j are combined into a single entity. Otherwise, the entity closest to the beam axis is considered a jet and removed from the list of entities to be further clustered. This continues until all entities are clustered. The anti-k_T is dominated by high p_T particles which it clusters first, subsequently including softer and softer constituents. Soft particles tend to cluster with hard particles before they tend to cluster among themselves. A hard particle that has no hard neighbours within a distance 2R accumulates all the soft particles within a circle of radius R. It tries to produce jets with fairly conical shapes that are centered around the hardest particles of the event, and with boundaries that resilient to the effect of soft radiation.

Jets being complex objects suffer from several effects that cause there energy, reconstructed as described above, to differ from their true values. Multiplicative correction factors are applied to calibrate their p_T and to ensure a uniform response in η [28, 29]. A total of four multiplicative corrections are applied. Firstly, the energy coming from pileup that has been clustered into the jet needs to be corrected for. A correction is applied based on the *hybrid jet area* method which is a combination of two methods viz., the average offset method and the jet area method. The average

offset method uses zero bias events to measure the average amount of energy added to the event due to pileup. The assumption is that averaging over zero bias events makes this measurement insenstive to high p_T objects and primarily represents soft pileup contributions. The average offset is measured in bins of η and number of pileup vertices (N_{PV}) averaged over ϕ. The correction is then given by $1 - \frac{<Offset(N_{PV},\eta)>}{p_T^{RAW}}$, where p_T^{RAW} is the uncorrected jet p_T. The drawback of this method is that it assumes that every jet contains the same amount of pileup contribution. The jet area method, on the other hand , calculates corrections on a jet-by-jet basis. It calculates an energy density per event by clustering jets using the k_T algorithm (this has a value of parameter p=1 and favours clustering soft jets as opposed to hard ones) and dividing the p_T by jet area, which is defined as the region in $\eta - \phi$ occupied by soft particles clustered in the jet. The median of this distribution (ρ) for an event is expected to be insensitive to hard particles and thus ρA_j is a good approximation of pileup contribution to the i-th jet. The drawback of this second approach, however is that it doesn't take into account the fact that the detector response is η dependent. The *hybrid jet area* method combines these two methods to calculate a jet-by-jet correction depending on η and N_{PV}. Secondly, a MC calibration factor, which corrects the energy of reconstructed jets to match the generated MC particle jet energy on average, is applied. This factor is based on simulated events. Finally, two other factors are used that each calibrate the energy response of reconstructed jets to be uniform with respect to η and p_T. These are also measured using simulated events. A QCD dijet sample is used to uniformize the dependence in η. Conservation of momentum in the transverse plane tells us that the sum of momentum in the transverse plane should be zero. Using jets that are approximately back-to-back in the azimuthal direction but at different η regions of the detector, the difference in response between these two η regions can be ascertained and corrected/uniformized. Using the same method of measuring residual response in the transverse direction in

$\gamma + jets$ or $Z + jets$ events, the absolute jet energy scale as a function of p_T can be uniformized.

4.4.7 Missing Transverse Energy: $\vec{p}_T^{miss}$

The CMS detector is unable to detect neutrinos (and other hypothetical particles) that are weakly interacting. However, the momentum balance (or imbalance) in the plane transverse to beam direction can be used to infer their presence. This "missing" transverse momentum vector is referred to as the $\vec{p}_T^{miss}$ and its magnitude is referred to as p_T^{miss}. It is defined as the negative vector sum of the p_T of entire list of objects reconstructed in the event by the above reconstruction algorithms and refined by particle-flow (PF objects):

$$\vec{p}_T^{miss} = -\Sigma \vec{p}_T \tag{4.1}$$

The $\vec{p}_T^{miss}$ plays an important role in this analysis as it helps gauge the momentum of the netrinos from the decaying tau lepton. The $\vec{p}_T^{miss}$ reconstrucion is directly dependent on the reconstruction of all the other objects in the event, from jets to muons to electrons. Consequently, it is sensitive to all the effects that influence the precise reconstruction and calibration of these objects. The largest effects come from biases in jet reconstruction and pileup (which are interconnected). Jet energy corrections described in previous section and pileup mitigation techniques discussed earlier can help significantly reduce the bias in $\vec{p}_T^{miss}$ reconstruction [30].

4.4.8 Relative isolation

The isolation of an object is the measure of the absence of other objects in its viscinity. In other words, it is the measure of how "isolated" an object is. It is calculated by summing up the p_T of all objects in a cone with predefined radius $\Delta R = 0.4$ around the lepton. The relative isolation which is obtained by dividing

the isolation by the p_T of the lepton is then given by:

$$I^{\ell}_{\text{rel}} = \left(\sum p_T^{\text{charged}} + \max \left[0, \sum p_T^{\text{neutral}} + \sum p_T^{\gamma} - p_T^{\text{PU}}(\ell) \right] \right) / p_T^{\ell}, \qquad (4.2)$$

where p_T^{charged}, p_T^{neutral}, and p_T^{γ} indicate the p_T of a charged particle, a neutral particle, and a photon within the cone, respectively. The contribution to isolation sum from charged particles in the cone coming from pileup is excluded by requiring the track origins be consistent with the primary vertex associated with the hard interaction. However, the same procedure cannot be adopted for neutral particles in the cone coming from pileup. This contribution , $p_T^{\text{PU}}(\ell)$, is estimated using a jet area method for electrons [31, 32], and simply as half of the scalar p_T sum of charged particles from pileup inside the cone for muons. The definition ensures that the total contribution to the isolation cone sum for neutral particles is always greater than or equal to 0.

The relative isolation is an useful variable in this analysis as prompt objects are usually isolated. More importantly, for an analyses with electrons and muons in the final state such as this, a tightened isolation requirement helps to reject those events where a jet is misidentied as either one of these leptons. Strict isolation requirements are thus used in the event selection in this analysis as described in sections 5.2.2 and 5.3.2.

4.4.9 Collinear Mass: M_{col} and Transverse Mass: M_T

An important variable of interest in this analyis is the collinear mass, M_{col}. As mentioned in later chapters, M_{col} is used as the signal variable in the $H \to \mu\tau_e$ analysis and in the M_{col} fit method of the $h \to \mu\tau_e$ analysis. The visible mass of the muon-electron system, M_{vis} is not a very good estimator of the Higgs boson mass. This is because the neutrinos from the tau decay do not interact with the detector

and the energy they carry is 'lost'. M_{col} provides a better estimate of the Higgs mass by approximating the neutrino momenta using the collinear approximation [33]. The primary idea is the following. The mass of the Higgs being much larger than that of the tau lepton causes the tau lepton to become highly boosted. Consequently, the decay products of the tau lepton, i.e. the electron, the tau neutrino and the electron neutrino, are produced in a highly collimated region around the direction of the tau lepton momentum. The momenta of the neutrinos ($\mathrm{p}_{\mathrm{T}}^{\nu,\,\mathrm{est}}$) can thus be approximated from the projection of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ in the direction of momenta of the visible tau decay product (electron), i.e. $\vec{\mathrm{p}}_{\mathrm{T}}^{\,\mathrm{e}}$. The visible fraction of the tau lepton momentum is then given as $x_{\tau}^{\mathrm{vis}} = \mathrm{p}_{\mathrm{T}}^{\vec{\tau}^{\mathrm{vis}}}/(\mathrm{p}_{\mathrm{T}}^{\vec{\tau}^{\mathrm{vis}}} + \mathrm{p}_{\mathrm{T}}^{\nu,\,\mathrm{est}})$. Finally, M_{col} is given as $M_{\text{col}} = M_{\text{vis}}/\sqrt{x_{\tau}^{\mathrm{vis}}}$. A very simple illustration in Figure 4.4 shows the superimposition the M_{col} and M_{vis} spectrums for a 300 GeV Higgs boson. Evidently, M_{col} is a better estimator of the mass and also has a sharper peak.

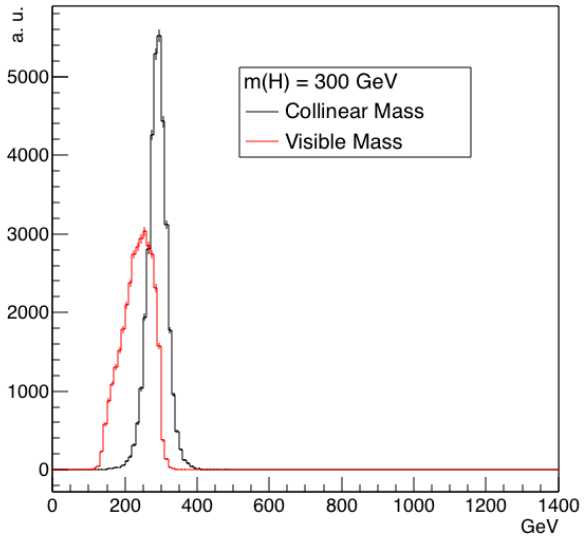


Figure 4.4: M_{col} and M_{vis} distributions for Higgs mass of 300 GeV.

Another variable that is used in this analyses and is useful to discriminate signal from background (see chapter 5) is the transverse mass, $M_{\mathrm{T}}(\ell)$ ($\ell = \mu, \mathrm{e}$). It is defined as follows: $M_{\mathrm{T}}(\ell) = \sqrt{2|\vec{p}_{\mathrm{T}}^{\,\ell}||\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}|(1 - \cos\Delta\phi_{\ell-\mathrm{p}_{\mathrm{T}}^{\mathrm{miss}}})}$, where $\Delta\phi_{\ell-\mathrm{p}_{\mathrm{T}}^{\mathrm{miss}}}$ is the angle between the lepton transverse momentum and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$.

CHAPTER 5

EVENT SELECTION

5.1 Introduction

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

Both analyses were performed blinded [34] in the signal region. All selection criterion and methods described below were developed without the knowledge of the observed data in the range of variable spectra where the signal is expected to be present. This is considered an optimal way of eliminating the unintended biasing of a result in a particular direction and is a standard methodology in particle physics analyses.

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

5.2.1 $h \to \mu\tau_e$: Final state signature and backgrounds

The signature of the $h \to \mu\tau_e$ analysis final state consists of a muon that comes promptly from the Higgs and has a hard p_T spectrum, along with a softer electron of opposite sign charge that comes from the tau lepton, and missing transverse momentum from the tau decay. It is interesting to note that the signature is similar to the $h \to \tau_\mu\tau_e$ decay that is allowed by the SM and since been observed [35], but with significant kinematic differences. In $h \to \mu\tau_e$ decay the μ comes directly from the Higgs resulting in its p_T spectrum peaking and spreading out to much higher values. Also there are fewer neutrinos in $h \to \mu\tau_e$, coming from the decay of the single τ. The decay products of this highly boosted tau are closely aligned, leading to a narrow separation between the e and the $\vec{p}_T^{\,miss}$ in the azimuthal plane. The same is not true in the $h \to \tau_\mu\tau_e$ decays. These differences are illustrated pictorially in Fig. 5.1.

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

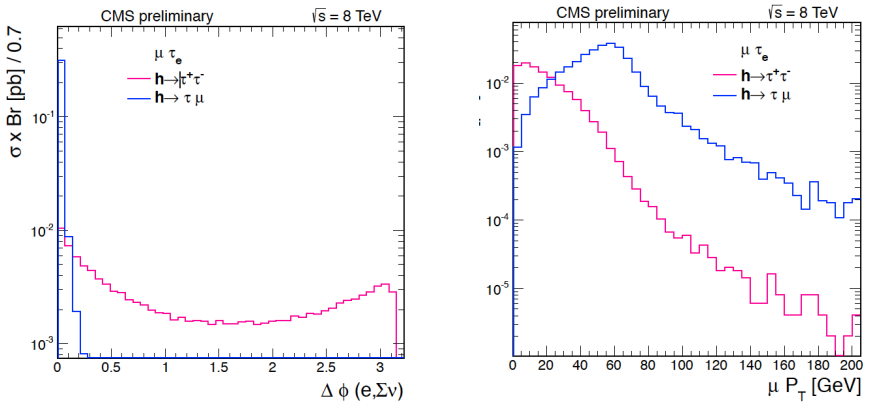


Figure 5.1: Illustration of the differences in p_T^{μ} and $\Delta\phi(e, \vec{p}_T^{miss})$ spectrums in $h \to \mu\tau_e$ and $h \to \tau_\mu\tau_e$ processes.

electroweak diboson production (WW, WZ and ZZ), h boson decays allowed by the SM ($H \to \tau\tau, WW$), $W\gamma^{(*)}$ + jets ,single top production, W + jets events, $Z \to \ell\ell$ ($\ell = e, \mu$) + jets and QCD multijet backgrounds. These backgrounds are described in more detail, along with there estimation and validation techniques in section 6.

5.2.2 $h \to \mu\tau_e$:Baseline selection and categorization

A baseline selection is defined first in order to ensure that we have clean and well-defined events faithful to the final state signature of the signal process. An isolated and well-identified μ is thus required to be present along with a well-identified and isolated e of opposite sign charge. They are required to be separated by $\Delta R > 0.3$. The identification criterion applied for μ and e have been described in sections 4.4.3 and 4.4.4. Isolation criterion, as measured by I_{rel} (described in 4.4.5), are required to have values $I_{rel}^{e} < 0.15$ and $I_{rel}^{\mu} < 0.1$. The p_T of these candidates are required to be above minimal thresholds required by trigger, identification and isolation requirement. Both candidates are also required to be within the fiducial region of the detector. The μ is required to have $p_T^{\mu} > 26\,GeV$ and $|\eta^{\mu}| < 2.4$.The e is required to have $p_T^{e} > 10\,GeV$ and $|\eta^{e}| < 2.3$. Only events with two or fewer jets are considered.

All jets considered must have $p_T > 30\,\text{GeV}$, $|\eta| < 2.4$ and satisfy the loose identification criterion described in section 4.4.6. Events with one or more jets arising from a b-quark (b-tagged jets) are vetoed. Cleaning events with b-tagged jets reduce some contribution from backgrounds which give rise to b-quarks such as $t\bar{t}$ and single top. Also, as described in 4.4.6, any event with one or more jets within $\Delta R < 0.4$ of either lepton candidates is also rejected. Further, an event is rejected if it has additional μ or e, or any τ_{had} candidates. All the above baseline selection requirements have been summarized in Table 5.1. All the events were required to pass isolated muon triggers with a p_T threshold of 24 GeV. The trigger selection has been described in detail in section **??**. The distributions of the M_{col} and several other kinematic variables after the baseline selection just described, are shown in Figs. 5.2 and 5.3. These distributions act as the starting point for development of stricter kinematic selections looking at the different shapes of signal and backgrounds distributions for different variables.

Table 5.1: Baseline selection criteria for $h \to \mu\tau_e$ analysis.

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

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

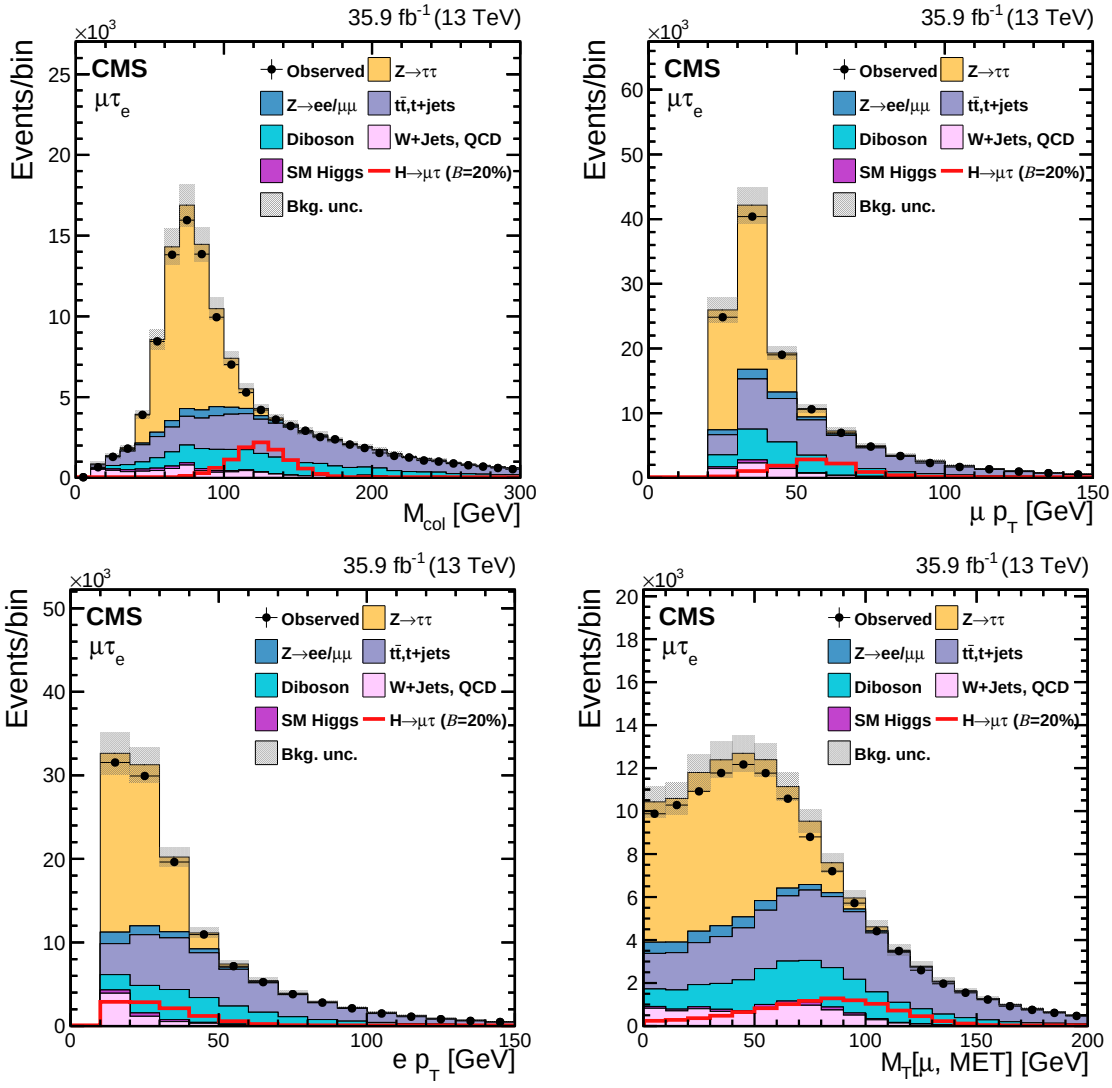


Figure 5.2: Distributions of kinematic variables after baseline selction for h $\rightarrow \mu\tau_e$ analysis (1).

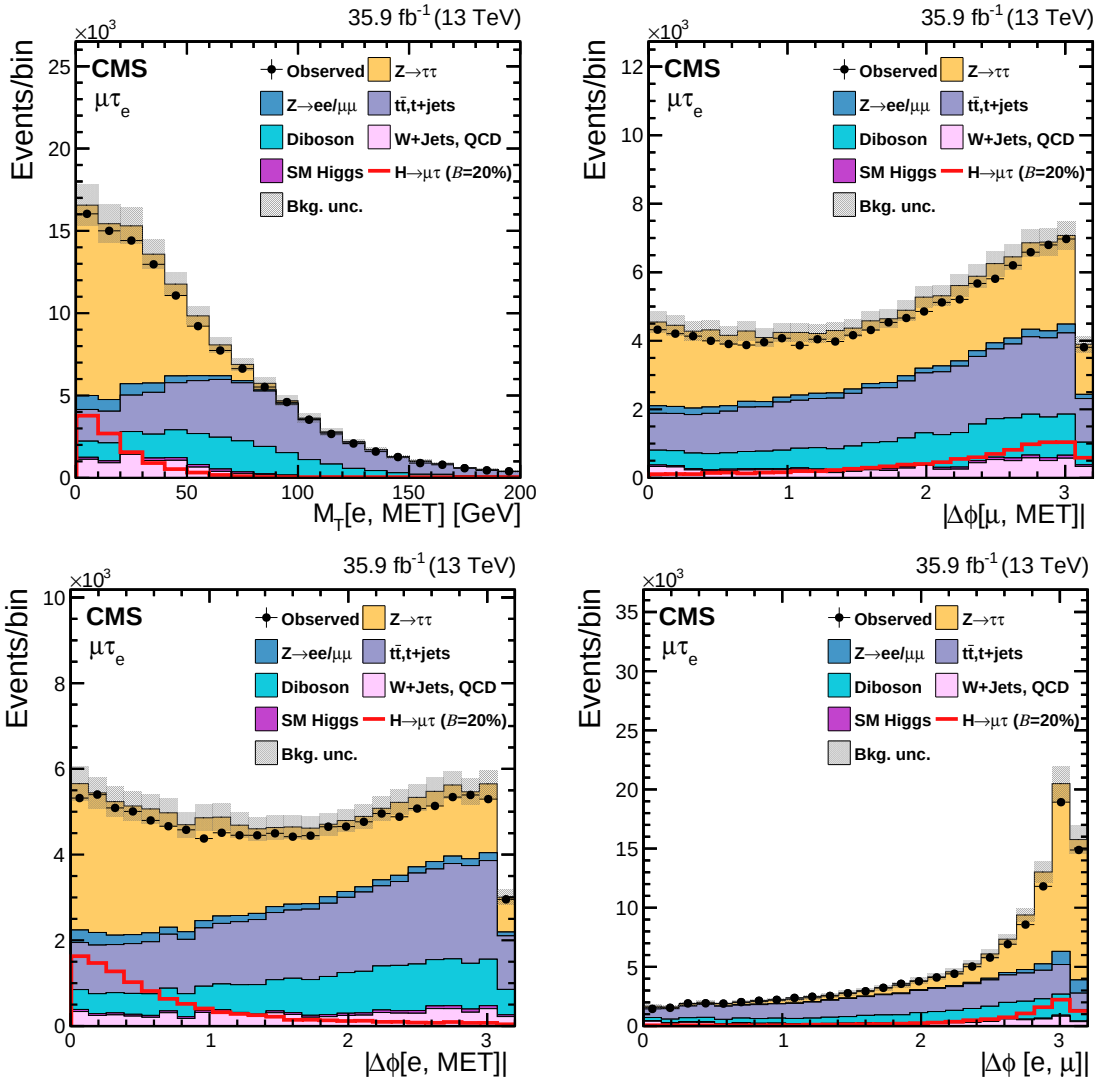


Figure 5.3: Distributions of kinematic variables after baseline selction for h $\rightarrow \mu\tau_e$ analysis (2).

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

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

5.2.3 $h \to \mu\tau_e$: M_{col} fit selection

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

and e, between e and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, between μ and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, denoted respectively by $\Delta\phi(\mathrm{e},\mu)$, $\Delta\phi(\mathrm{e},\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$, $\Delta\phi(\mu,\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$, and the transverse mass between μ and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, between e and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, devnoted respectively by $M_T(\mu)$ and $M_T(\mathrm{e})$. The $\mathrm{h} \to \mu\tau_e$ decay being a 2-body decay, the μ and e are expected to be well separated in the azimuthal plane. Therefore, selecting events with a $\Delta\phi(\mathrm{e},\mu)$ larger than a threshold can help reject background events while keeping the signal that is peaked at high $\Delta\phi(\mathrm{e},\mu)$ values. This can be seen from Fig 5.3 (bottom right). Both neutrinos in the signal process come from the decay of the same τ. These neutrinos form the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. As mentioned earlier, the τ being much lighter than the h, it is highly boosted and its decay products i.e. e and the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ are expected to be close to each other in the azimuthal direction. Thus $\Delta\phi(\mathrm{e},\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$ is expected to peak at values close to zero for signal events, as seen in Fig 5.3 (bottom left). Given that all backgrounds have relatively flat shape for this variable throughout the $\Delta\phi$ range, requiring $\Delta\phi(\mathrm{e},\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$ to be lower than a threshold works as a strong rejection criterion against the backgrounds. Following a similar line of reasoning, the μ is expected to be well separated from the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ resulting in $\Delta\phi(\mu,\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$ for signal events to peak at high values, as seen in Fig 5.3 (top right). Further, as the $M_T(\ell)$ (defined in section **??**) contains negative of the cosine of $\Delta\phi(\ell,\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$ term, it is expected to be peak at values similar to $\Delta\phi(\ell,\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$. This can be seen from Fig 5.3 (top left) and Fig 5.3 (bottom right) which show signal events for $M_T(\mu)$ and $M_T(\mathrm{e})$ peak at relatively higher and lower values than most backgrounds respectively. In particular, requiring $M_T(\mu)$ to be larger than a threshold can help reject a lot of $\mathrm{Z} \to \tau\tau$ events which is the most dominant background in the 0-jet category. All the above variables have some amount of correlation with one another (see the correlation matrix shown in Fig. 5.6. The optimization procedure used to arrive at the most optimal set of kinematic thresholds for these variables is described in detail in the next paragrhaph. The thresholds on the $\mathrm{p_T}$ of the μ and e have not been made stricter to avoid biasing the selection

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

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

round is variable2. For the start of the following round variable2 is is required to satisfy its corresponding threshold. Then all the other variables (including variables that were had chosen thresholds in earlier rounds such as variable1 here) are made to go through the same procedure. This is done because the optimum value of threshold for variables chosen earlier might shift as new variables are chosen. This process is continued until the expected limit becomes no further stringent in sucessive rounds. This optimization was done separately for each of the four categories. The final set of thresholds arrived at in this way for the h $\rightarrow \mu\tau_e$ M_{col} fit analysis are listed in Table. 5.2. This method of choosing the optimal set of thresholds is sometimes called the n-1 procedure, and the idea is conceptually similar to forward/backward selection methods used in statistical learning to build optimal models.

TABLE 5.2

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

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

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

In the BDT method, a boosted decision trees (BDT) classifier is used to discriminate signal events from background events. A decision tree is a classifier which works by building a tree structure based on binary splits (as shown in Fig. 5.4). Starting from the root node of the tree (which contains all the events which we want to classify), a sequence of binary splits is made using input variables provided to the classifier. At each split, the variable which provides best purity of split or equivalently, in our case the best separation of signal and background events, is used. The same variable can thus be used for splitting several nodes and the splitting is continued until a desired some stopping criterion such as depth of the tree, purity of leaf nodes , minimum number of events in a leaf node etc. is reached. All events end up in one of the leaf nodes. If an event ends up in a leaf node in which signal events form the majority fraction, it is classified as a signal event. Otherwise, it is classified as a background event. Boosting is a class of ensemble machine learning techniques which help in enhancing performance of weak classifiers by sequentially building classifiers using reweighted (boosted) versions of the training data and then taking a weighted majority vote of the sequence of classifiers thus produced. Boosting also stabilizes the response of the classifiers with respect to fluctuations in the training data. In other words it helps avoid overfitting to the training data. When the boosting technique is a applied to produce an ensemble of decision trees, the resulting ensemble of classifiers is called a Boosted Decision Trees classifier. A detailed overview of how decision trees and boosting works, and the chosen value of parameters used in training the BDTs for this analysis is given in appendix A.

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

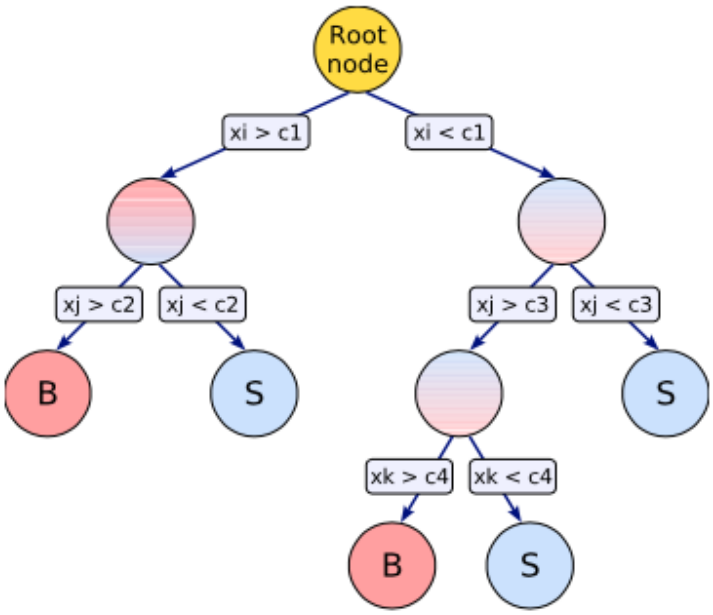


Figure 5.4: Illustration of decision tree. [36]

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

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

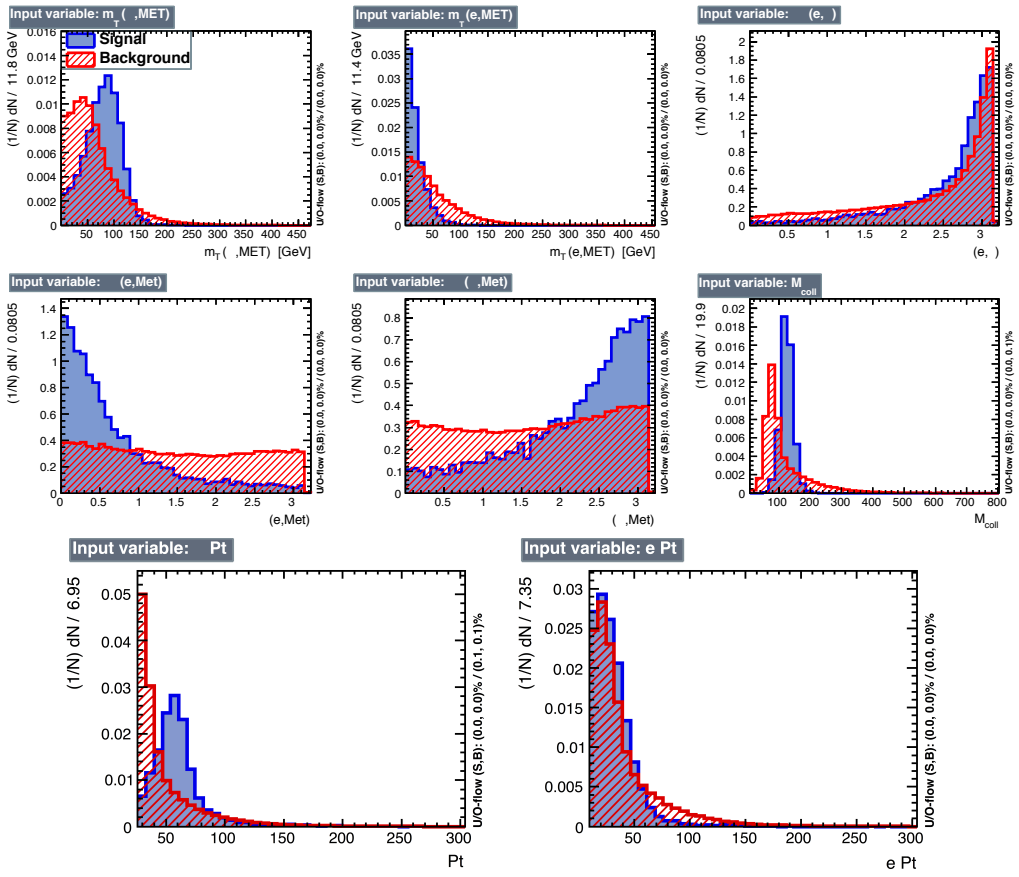


Figure 5.5: Normalized distributions of the input variables for BDT method. The signal (blue) is composed of a weighted mixture of GGF and VBF events, whereas the background (red) is made of $t\bar{t}$ and Drell-Yan events. All events were required to satisfy the baseline selection criteria.

- Electron p_T: p_T^e.

The distributions of these variables normalized to the total number of events in the input sample to the BDT is shown in Fig. 5.5. The correlations between these variables in signal and background events are shown in Fig. 5.6.

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

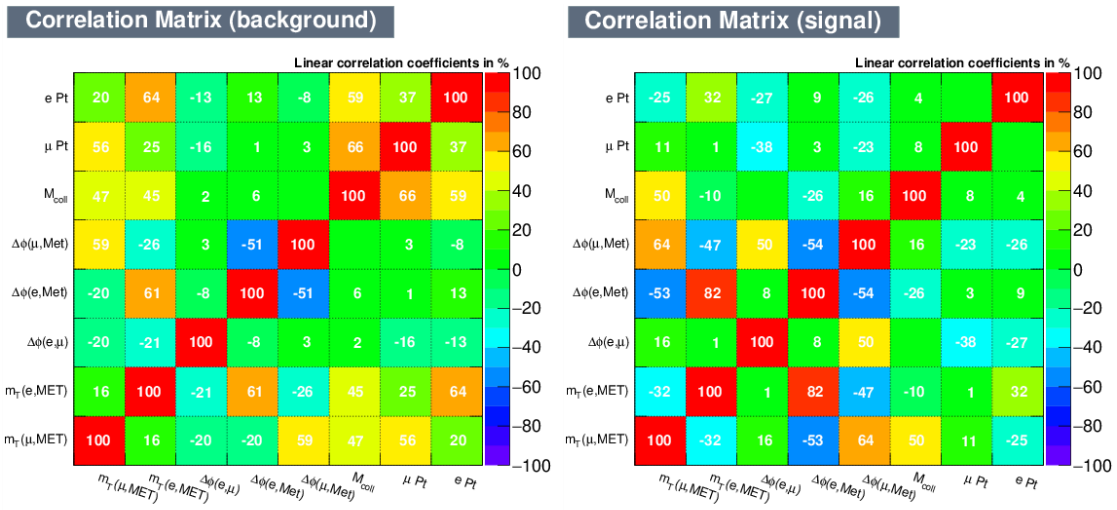


Figure 5.6: Correlations between input variables for signal events (right) and background events (left).

was used. Fig. 5.7 shows the distribution of the BDT response for training and testing samples. The training and testing distributions for both signal and background events match well, suggesting that there is no overtraining. The distribution of BDT response is used in max-likelihood fit to extract results, as discussed in section 7.

5.3 Heavy higgs:H $\rightarrow \mu\tau_e$ analysis

5.3.1 H $\rightarrow \mu\tau_e$: Final state signature and backgrounds

The signature of the H $\rightarrow \mu\tau_e$ analysis final state is very similar to that of h $\rightarrow \mu\tau_e$. It also consists of a muon that comes promptly from the Higgs and has a hard p_T spectrum, along with a softer electron that comes from the tau lepton, and missing transverse momentum from the tau decay. The p_T^{μ} spectrum is expected to be harder for higher H boson masses. The topologies being similar, the kinematic properties discussed in section 5.2.1 for h $\rightarrow \mu\tau_e$ analysis also apply to the H $\rightarrow \mu\tau_e$ analysis. The H boson mass peaks for all the simulated samples illustrated in Fig 5.8.

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

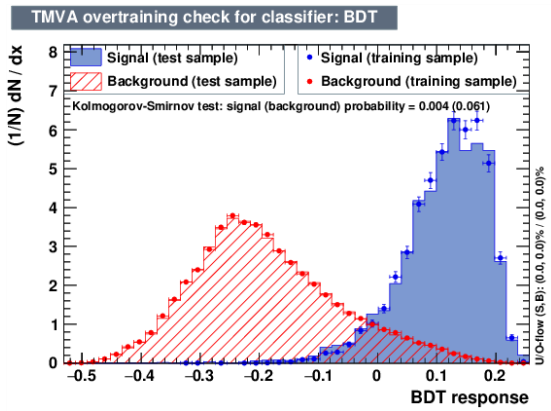


Figure 5.7: Distribution of BDT response for training (dots) and test(fill) distributions for both signal(blue) and background(red) events.

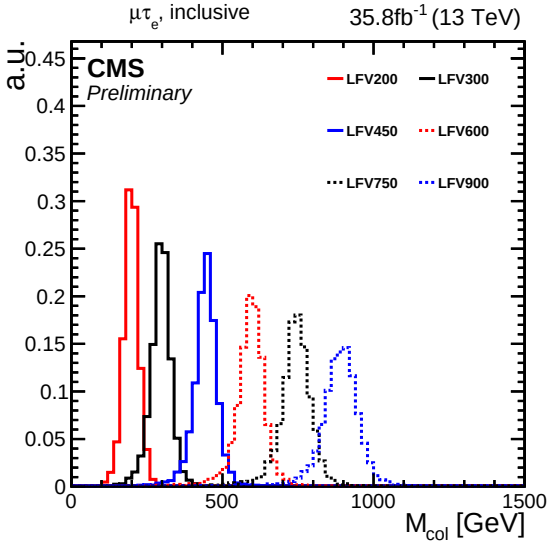


Figure 5.8: Illustration of simulated signal mass peaks for H $\rightarrow \mu\tau_e$ analysis for different H boson masses.

electroweak diboson production. Unlike h $\rightarrow \mu\tau_e$ analysis, Z $\rightarrow \tau\tau$ events from Drell-Yan production form a very small background as the Z $\rightarrow \tau\tau$ spectrum peaks at much lower values (around Z boson mass) of collinear mass that the signal events coming from heavy H boson decays. The other backgrounds come from h boson decays (H $\rightarrow \tau\tau$, WW), W$\gamma^{(*)}$ + jets ,single top production, W + jets events, $Z \rightarrow \ell\ell$ (ℓ = e, μ) + jets and QCD multijet backgrounds. These backgrounds are described in more detail, along with there estimation and validation techniques in section 6.

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

The baseline selection for H $\rightarrow \mu\tau_e$ is similar to that of h $\rightarrow \mu\tau_e$ with the exception of higher p_T thresholds. Just like h $\rightarrow \mu\tau_e$, an isolated and well-identified μ is thus required to be present along with an well-identified and isolated e of opposite sign charge. They are required to be separated by $\Delta R > 0.3$. The identification and isolation criteria have been described in sections 4.4.3, 4.4.4 and 4.4.5. All events are required to pass a single muon trigger with the threshold of 50 GeV. The trigger selection has been described in detail in section **??**. The μ is required to have $p_T^{\mu} > 53$ GeV and $|\eta^{\mu}| < 2.4$. The e is required to have $p_T^{e} > 10$ GeV and $|\eta^{e}| < 2.3$. Only events with zero or one jet are considered. Jets must have $p_T > 30$ GeV, $|\eta| < 2.4$ and satisfy the loose identification criterion described in section 4.4.6 to be considered. As only GGF production mode is considered for the H $\rightarrow \mu\tau_e$ analysis, events with more than one jet make negligent contribution and are rejected. All other other criteria are same as the h $\rightarrow \mu\tau_e$ analysis. The entire set of baseline selection criteria for H $\rightarrow \mu\tau_e$ has been summarized in table 5.3.

The events are then divided into categories, with motivations similar to the h $\rightarrow \mu\tau_e$ analysis (see section 5.2.2), on the basis of number of jets present in the event. The two categories for H $\rightarrow \mu\tau_e$ are:

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

Table 5.3: Baseline selection criteria for H $\to \mu\tau_e$ analysis.

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

enhances the gluon-gluon fusion (GGF) contribution.

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

The distributions of several kinematic variables after the baseline selection and categorization are shown in Figs. 5.9 and 5.10.

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

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

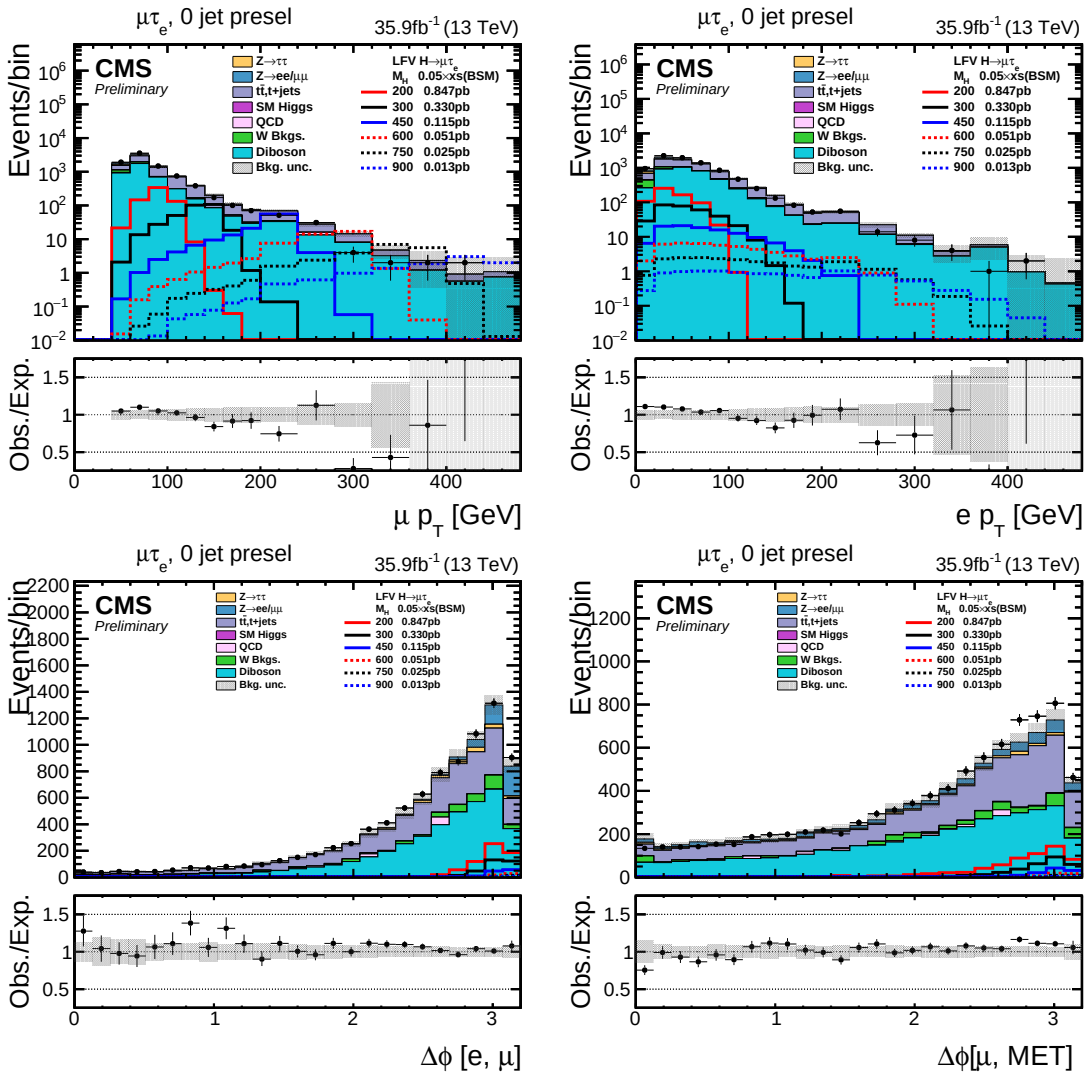


Figure 5.9: Distributions of kinematic variables after baseline selction for 0-jet category of H $\rightarrow \mu\tau_e$ analysis.

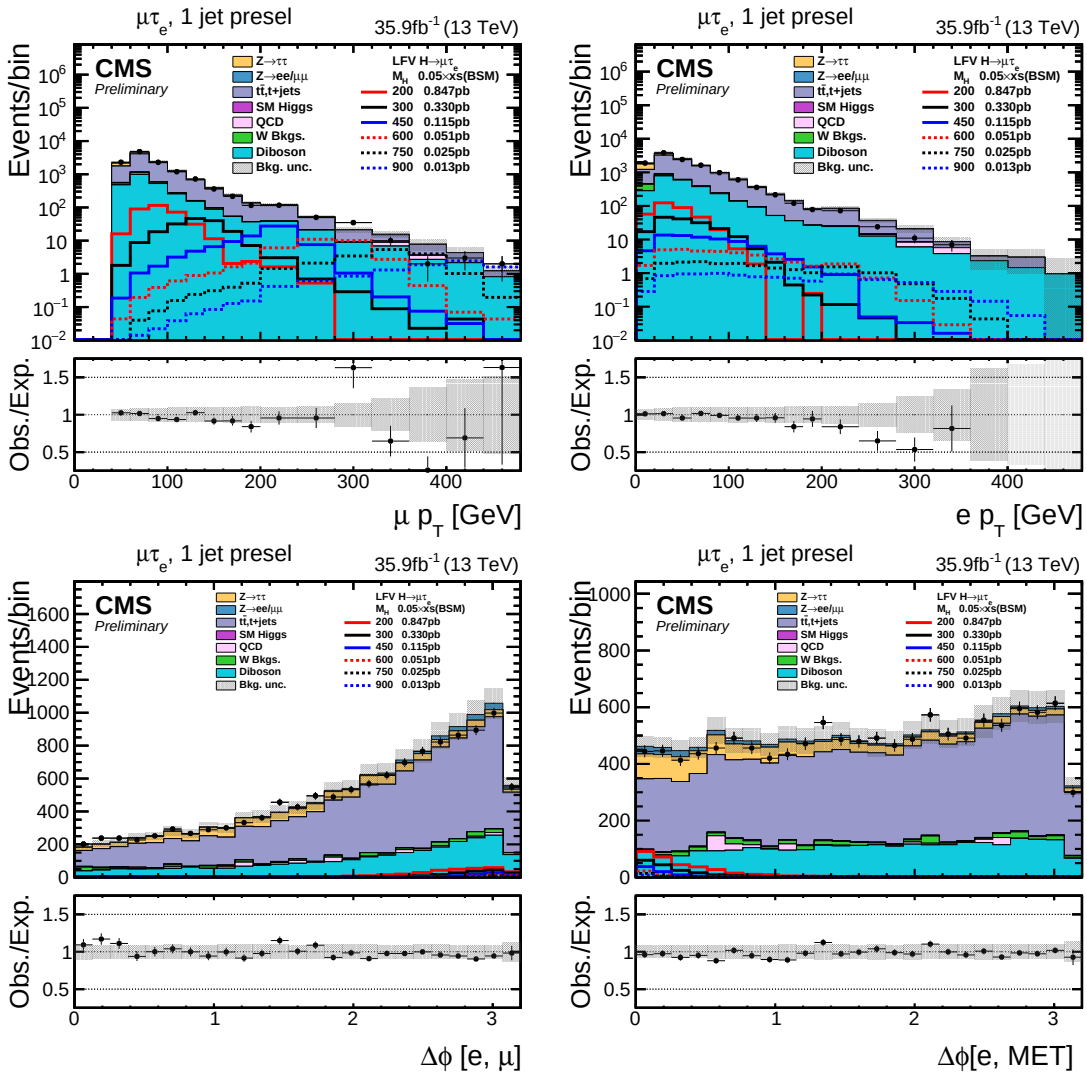


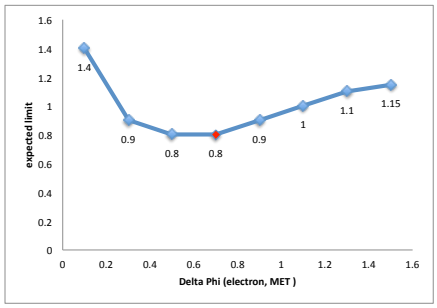
Figure 5.10: Distributions of kinematic variables after baseline selction for 1-jet category of H $\rightarrow \mu\tau_e$ analysis.

The procedure for optimization of the thresholds of for these variables is exactly the same as described in section 5.2.3. Further to get better sensitivity in the entire mass range from 200 to 900 GeV, two separate sets of thresholds are optimized, for each category. One set is optimized to provide better sensitivity in the 200-450 GeV mass range. The simulated signal for the H mass of 200 GeV is used when calculating expected limits during the optimization procedure for this mass range. The other set is optimized to provide better sensitivity in 450-900 GeV mass range. The simulated signal for H mass of 450 GeV is used when calculating expected limits during the optimization procedure for this mass range. A few illustrations of the optimization procedure are shown in Fig. 5.11. The final set of thresholds arrived at in this manner, for both mass ranges and both categories of the $H \to \mu\tau_e$ M_{col} fit analysis, are listed in Table. 5.4. The M_{col} distributions after requiring these selections is used in a max-likelihood fit to extract results, as discussed in section 7.

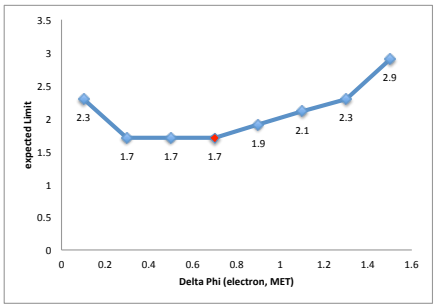
TABLE 5.4

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

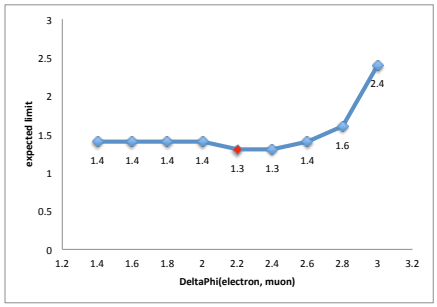
	Low mass range	High mass range
0-jet	$p_T^{\mu} > 60\,\text{GeV}$, $p_T^{e} > 10\,\text{GeV}$ $\Delta\phi(e, \vec{p}_T^{\,\text{miss}}) < 0.7$ $\Delta\phi(e, \mu) > 2.2$	$p_T^{\mu} > 150\,\text{GeV}$, $p_T^{e} > 10\,\text{GeV}$ $\Delta\phi(e, \vec{p}_T^{\,\text{miss}}) < 0.3$ $\Delta\phi(e, \mu) > 2.2$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



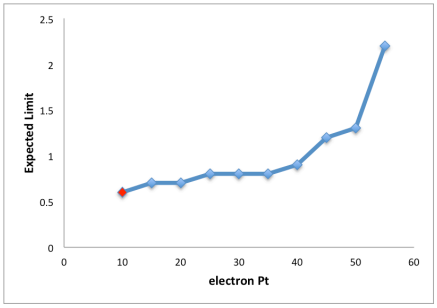
(a) Low mass range 0 jet $\Delta\phi(\mathrm{e}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$



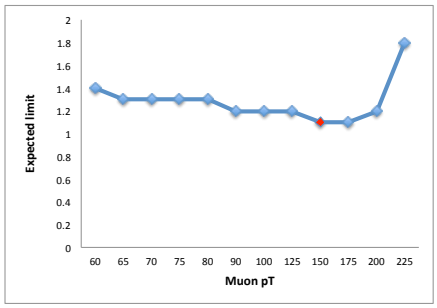
(b) Low mass range 1 jet $\Delta\phi(\mathrm{e}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$



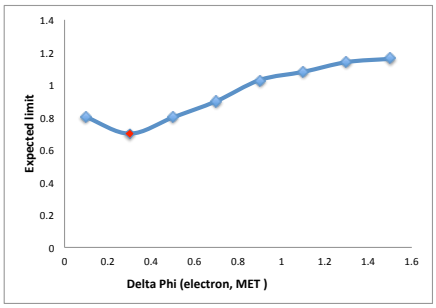
(c) Low mass range 0 jet $\Delta\phi(\mathrm{e}, \mu)$



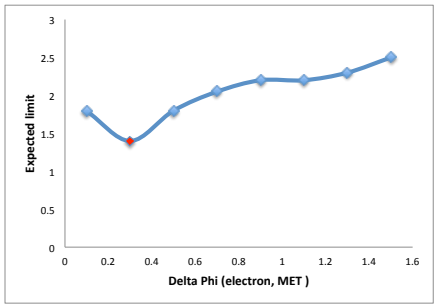
(d) Low mass range 0 jet $\mathrm{p}_{\mathrm{T}}^{\mathrm{e}}$



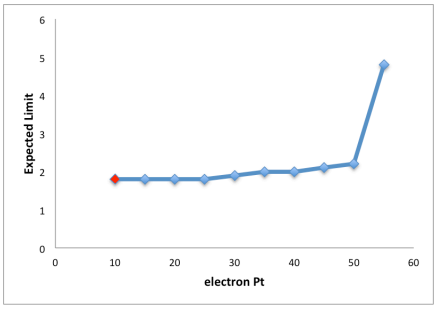
(e) High mass range 1 jet $\mathrm{p}_{\mathrm{T}}^{\mu}$



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



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



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

Figure 5.11. Examples of cut optimisation for the H $\rightarrow \mu\tau_e$ analysis

CHAPTER 6

BACKGROUND ESTIMATION AND VALIDATION

6.1 Introduction

This chapter describes the techniques used for estimation of the backgrounds in the analyses. Each background is estimated individually. For large backgrounds, the estimation is validated using regions enriched in those backgrounds.

6.2 h125: $h \to \mu\tau_e$ backgrounds

6.2.1 $Z \to \tau\tau$

The $Z \to \tau\tau$ background is the dominant background in 0-jet and 1-jet categories of the analysis. It is an irreducible background and arises when one τ coming from the Z boson decay further decays into a μ and the other decays into a e. This background is estimated from simulated monte-carlo events. In a $Z \to \ell\ell$ events from Drell-Yan production including $Z \to \tau\tau$, the $m_{\ell\ell}$ and Z_{p_T} distributions are found to be different in data and simulation. In order to correct for this, a set of reweighting factors is calculated using a dedicated control region enriched in $Z \to \mu\mu$ events. The set of reweighting factors are applied as a function of generator-level $m_{\ell\ell}$ and Z_{p_T} in the signal region of the analysis. A more detailed study of this effect and calculation of the reweighting factors can be found in the following references [35].

To validate this estimation, we look at agreement between observed data and simulation in a region enriched in $Z \to \tau\tau$ events. This region is constructed by requiring, in addition to the baseline selection, the p_T of the $\mu < 40$ GeV. The p_T in

$Z \to \tau\tau$ events is on softer side of the spectrum compared to other backgrounds which are more spread out, as seen in Fig. 5.2 (top right). The $M_T(\mu)$, as seen in Fig. 5.2 (bottom right), is required to be less than 60 GeV following similar reasoning. Further the invariant mass of the e and μ is required to be in between 30 GeV and 70 GeV in order to isolate the Z peak. The distributions of BDT response and M_{col} in this $Z \to \tau\tau$ enriched region are shown in Fig. 6.1, for the categories where this background is dominant. The plots show good agreement between data and background.

6.2.2 $t\bar{t}$

Tops decay into W bosons and a b-quark more than 90% of the time. The W boson can decay leptonically into a μ and e making it a background for the analysis. The b-tagging veto applied at the baseline selection level is able to somewhat supress this background. However it still forms a large fraction of the background for the analysis. In fact, it is the largest background in both 2-jet categories. It is also large in the 1-jet category. We estimate the $t\bar{t}$ background using simulation. The background estimation is validated in two separate control regions enriched in $t\bar{t}$. The first control region is formed requiring the baseline selection but with a inverted b-tagging veto. In other words, at least 1 b-tagged jet is required to be present in the event. The distributions of BDT response (top) and M_{col} (bottom) in this region are shown in Fig. 6.6 for categories where the $t\bar{t}$ background is large. The second control region is constructed using kinematic selection criteria. In particular, in addition to the baseline selection criteria with the b-tag veto removed, we require M_T(e) (see Fig. 5.3 top left) to be greater than 50 GeV. The distributions of BDT response (top) and M_{col} (bottom) in this second control region are shown in Fig. 6.3. Given that the uncertainty bands in these control region plots only contain uncertainties on normalization (and not shape-based uncertainties, as discussed in section 7, and included in the max likelihood fit used to extract results), the data over background

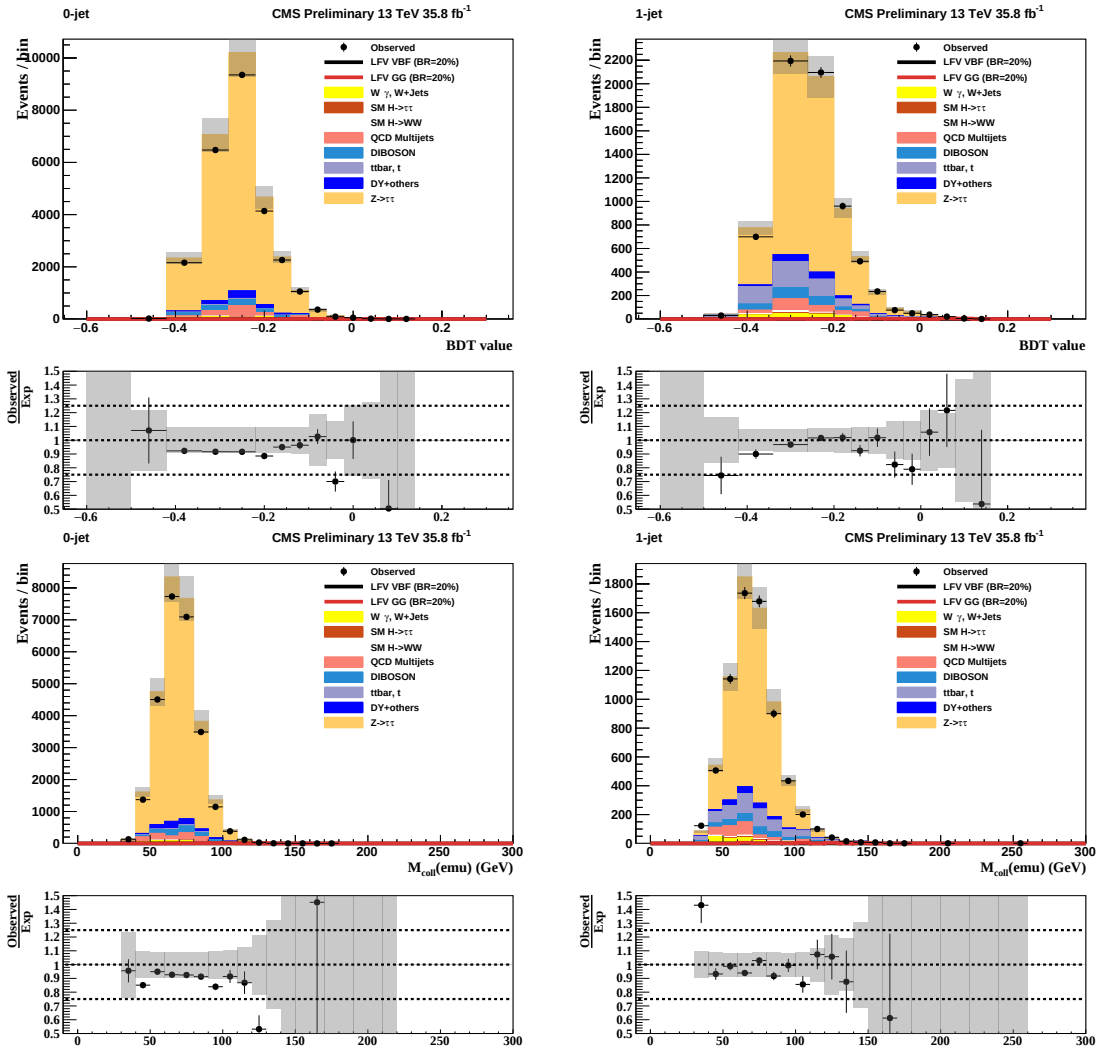


Figure 6.1: Distributions of BDT response (top) an M_{col} (bottom) in $\text{Z} \to \tau\tau$ enriched region for 0-jet (left) and 1-jet (right) categories.

estimation ratio is reasonable in these regions. Further, a normalization uncertainty of 10% is applied on the $t\bar{t}$ estimation in the signal region based on these control regions.

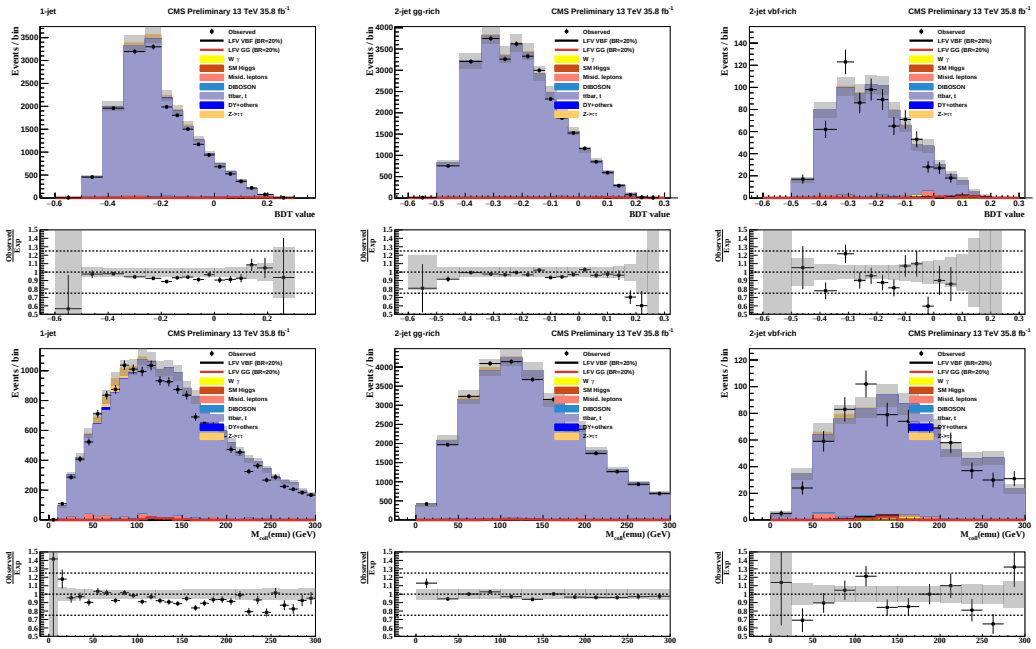


Figure 6.2. Distributions of BDT response (top) an M_{col} (bottom) in the first $t\bar{t}$ enriched region, as described in the text.

6.2.3 Misidentified lepton background

Another source of background which is relatively much smaller than $t\bar{t}$ or $\text{Z} \to \tau\tau$ arises from jets misidentified as leptons in W+jets or SM events comprised uniquely of jets produced through the strong interaction, referred to as quantum chromodynamics (QCD) multijet events. In W+jets events, one lepton candidate is a real lepton from the W boson decay while the other lepton is a misidentified jet. In QCD events,

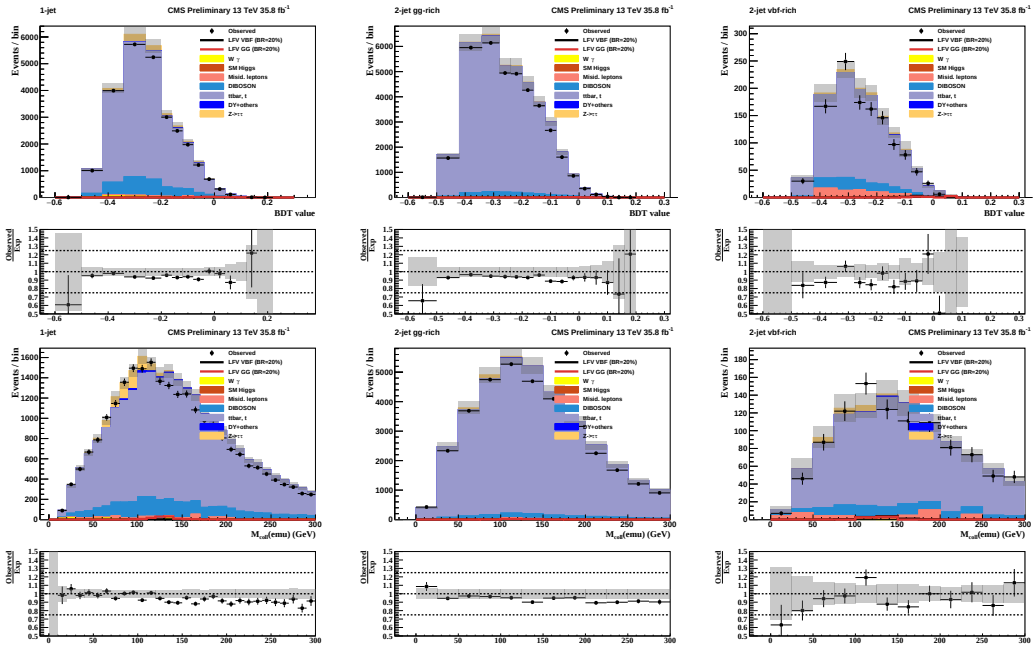


Figure 6.3. Distributions of BDT response (top) an M_{col} (bottom) in the second $t\bar{t}$ enriched region, as described in the text.

both leptons in the final state are misidentified jets. The baseline selection criteria requires the leptons to be well identified and isolated. This makes it difficult for a jet to masquerade as a lepton. In case of the μ, this is even more so since it is required to satisfy high p_T thresholds as well. Consequently, these events form a small part of the background. This is in contrast to a final state where the non-prompt lepton is a hadronically decaying τ instead of an electronically decaying one. This background would be much larger in such a case.

The W + jets background contribution to the misidentified-lepton background is estimated using simulation. The QCD multijet contribution is estimated from collision data events where the leptons have like-sign charge. The expected yield from non-QCD processes in this region is subtracted using simulation. The resulting sample is then rescaled to account for the differences between the composition in the like- and opposite-sign charge regions. The scaling factors are extracted from

samples enriched QCD multijet events, and the procedure is illustrated in Ref. []. This background is validated in a control region that is obtained by requiring the baseline selection but inverting the isolation criteria. In other words events with well-isolated μ and e are rejected. The particular isolation thresholds required for this region are:$0.1 < I^{\mathrm{e}}_{\mathrm{rel}} < 1$ or $0.15 < I^{\mu}_{\mathrm{rel}} < 0.25$. The distributions of BDT response and M_{col} in this qcd enriched region are shown in Fig. 6.4. The plots show good agreement between data and background.

6.2.4 Other backgrounds

The other backgrounds in the analysis make relatively much smaller contrbutions. Electroweak diboson production (WW, WZ and ZZ) contributes a similar number of events as the misidentified lepton background, and is estimated from simulation. WW events make the largest contribution, followed by WZ and ZZ events. This is because WZ and ZZ events have additional leptons in their final state which have to miss detection in order for the event to be a background. SM decays of the h boson also forms a small but non-negligible background. These come particularly from $\mathrm{h} \to \tau\tau$ and $\mathrm{h} \to \mathrm{WW}$ decays. Other backgrounds include $Z \to \ell\ell$ ($\ell = \mathrm{e}, \mu$) + jets, single-top quark production and $\mathrm{W}\gamma^{(*)}$ + jets. All of these are estimated using simulation.

6.3 Heavy Higgs: $\mathrm{H} \to \mu\tau_e$ backgrounds

The background processes in the $\mathrm{H} \to \mu\tau_e$ analysis are similar to $\mathrm{h} \to \mu\tau_e$ but differ in relative contribution, and are overall much smaller. This is due to the fact that the $\mathrm{H} \to \mu\tau_e$ analyses searches for LFV decay in a higher mass, higher $\mathrm{p_T}$region. In particular, $\mathrm{Z} \to \tau\tau$ background which is the most dominant in $\mathrm{h} \to \mu\tau_e$ is now very small. The $\mathrm{Z} \to \tau\tau$ background peaks around the Z boson mass, and the high $\mathrm{p_T}$ cuts in this anlysis reject most of these events. The dominant backgrounds in $\mathrm{H} \to \mu\tau_e$ are $t\bar{t}$ production, followed by electroweak diboson production which have

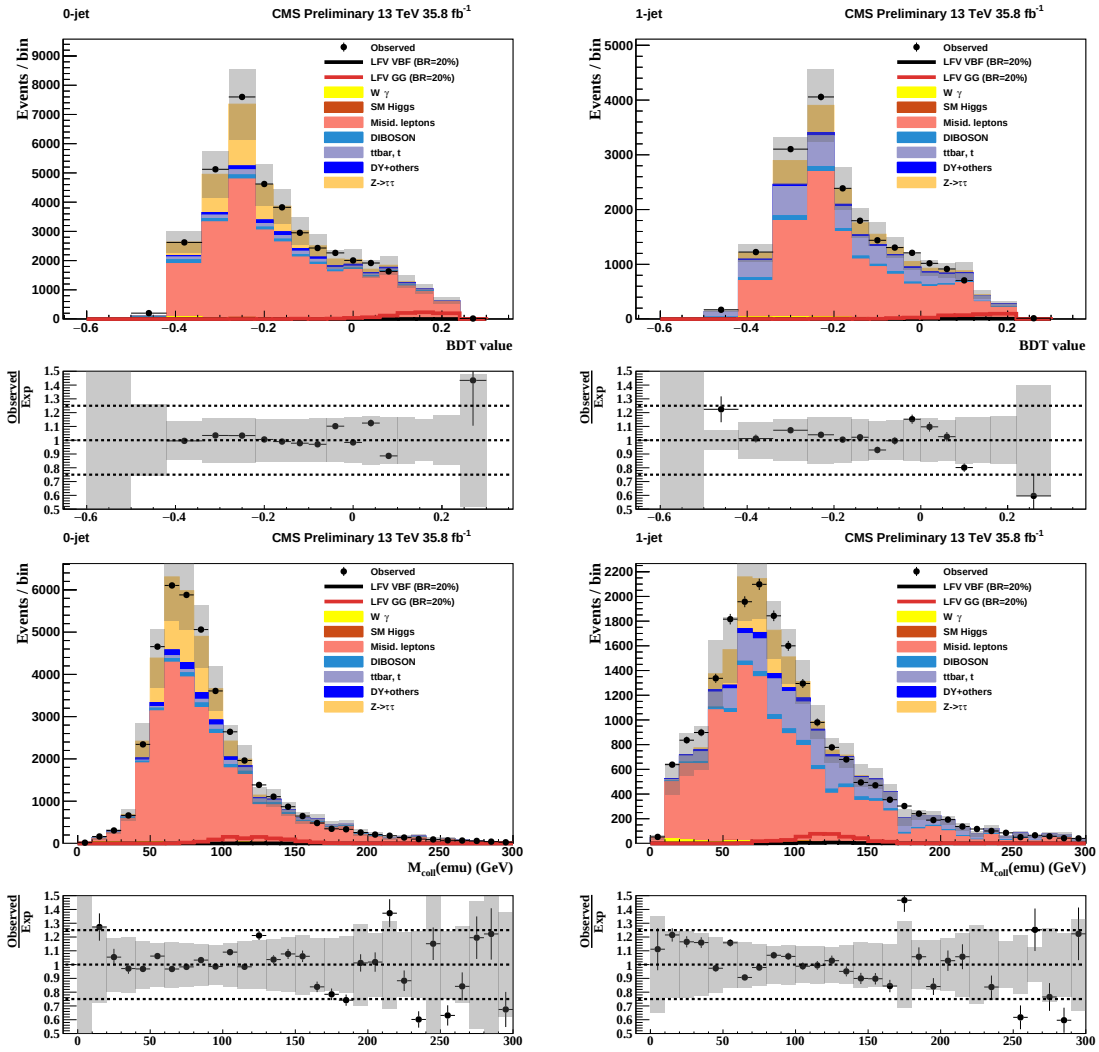


Figure 6.4: Distributions of BDT response (top) an M_{col} (bottom) in QCD enriched region for 0-jet (left) and 1-jet (right) categories.

a relatively flatter p_T distribution and survive the strict p_T requirements.

$t\bar{t}$ production is the largest background in the H $\to \mu\tau_e$ analysis. We estimate this background using simulation. A control region enriched in $t\bar{t}$ events is constructed by requiring the baseline selection with the b-tag veto removed, and with the additional requirement that at least 1 b-tagged jet be present. Fig. 6.5 (left) shows the M_{col} distribution of this sample. To take into account the residual data to background estimation difference, an overall normalization scale factor of 0.886 is extracted from this region, and is applied to the background estimation in the signal region. The same control region above is shown in Fig. 6.5 (right), after the background has been scaled by the above factor for illustration. Distributions of several other kinematic variables (after the above rescaling) in the $t\bar{t}$ control region are shown in Fig. 6.6. They show reasonable agreement between data and estimated background.

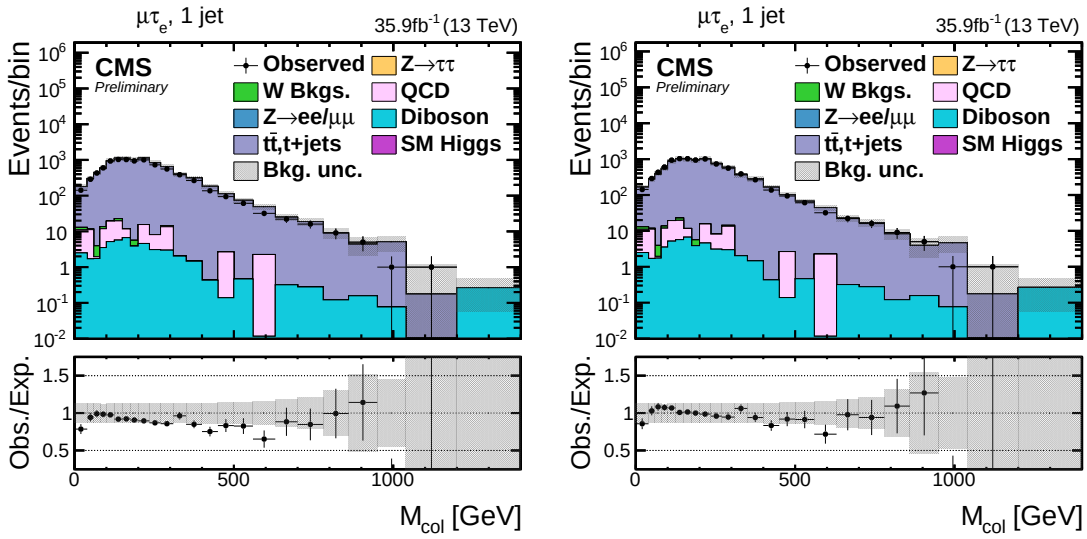


Figure 6.5: M_{col} distribution in $t\bar{t}$ enriched control region as defined in the text before the application of the scale factor (left) and after (right),for the H $\to \mu\tau_e$ analysis.

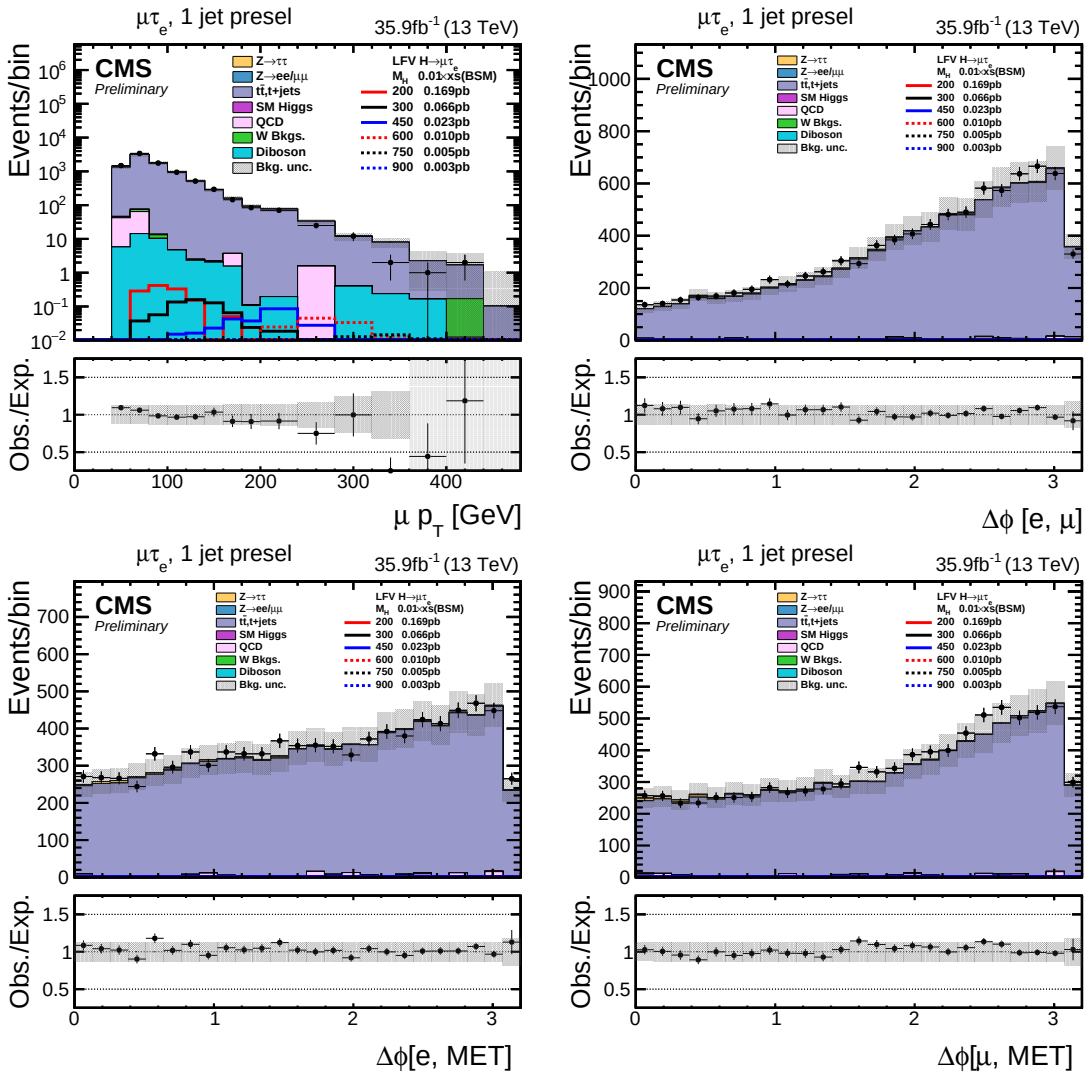


Figure 6.6: Distributions of several kinematic variables in the $t\bar{t}$ enriched control region for H $\rightarrow \mu\tau_e$ analysis.

Electroweak diboson production (WW, WZ and ZZ) forms the next largest background in $H \to \mu\tau_e$ analysis. It is estimated using simulation. All other backgrounds are much smaller. This can be seen from the distributions of kinematic variables after baseline selection, as can be seen from Figs. 5.9 and 5.10. The misidentified lepton background is even smaller here than $h \to \mu\tau_e$. The higher p_T requirement makes it even less likely for jets to be able to be misidentified as leptons. This background is estimated using the same technique as $h \to \mu\tau_e$, as described in section 6.2.3. The $Z \to \ell\ell$ ($\ell = e, \mu$) + jets and $Z \to \tau\tau$ backgrounds are estimated from simulation. Other backgrounds include SM h boson decays, $h \to WW$, $h \to \tau\tau$, single-top quark production and $W\gamma^{(*)}$ + jets, and are also estimated using simulation.

CHAPTER 7

SIGNAL EXTRACTION AND SYSTEMATIC UNCERTAINTIES

7.1 Introduction

The analysis is, in its essence, a sophisticated counting experiment. The presence of a signal is indicated by an excess of events over the predicted background, in the distribution of a signal variable. For our analyses the signal variables are collinear mass or BDT output, as described in chap 5 and 4. Given that there are several uncertainties, both experimental and theoretical and also due to the innate randomness in the process, it is possible that an excess is observed when there is no signal. So, when an excess is observed, a p-value which represents the probability that the excess is due to statistical fluctuations is computed. A very low p-value is taken to indicate that the excess corresponds to an observed signal and not merely a statistical fluctuation. Conversely, if no excess is observed (upper exclusion) limits are set on the product of branching fraction and production cross-section. A 95% CL (confidence level) is taken as a requirement for ruling out a signal at or above a certain value known i.e. upper exclusion limit. The first part of this chapter describes the statistical methods used, that very closely follow the procedure used for LHC Higgs boson search described in [37].

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

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

7.2 Statistical methods for signal extraction

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

7.2.1 Likelihood function

The Poisson distribution is an appropriate model for n, the number of times an event occurs in an interval if the following assumptions are true [38].

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

The poisson probablity of distribution is then given by:

$$P(n_{events}) = \frac{e^{-\lambda}\lambda^{n}}{n!} \tag{7.1}$$

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

given by:

$$\mathcal{L}(data|\mu) = \prod_{i=1}^{bins} \frac{(\mu \cdot s_i + b_i)^{n_i}}{n_i!} e^{-\mu \cdot s_i - b_i} \tag{7.2}$$

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

7.2.2 Treatment of systematic uncertainties

All systematic uncertainties are handled by introducing them as nuisance parameters. Nuisance parameters are parameters that influence the model but are not of interest in our measurement, e.g., if we are interested in knowing only the mean of a population that is expected to be distributed as a gaussian, the standard deviation becomes a nuisance parameter for the model that we fit. In our experiment, the nuisance parameters are embedded into the likelihood function. In order for the likelihood function to have a clean factorised form [37], all sources of uncertainties considered are taken to be 100%-corrrelated or uncorrelated. If an uncertainty is partially correlated, it is either separated into 100%-corrrelated or uncorrelated components, or considered 100%-corrrelated or uncorrelated, depending on whichever is a more conservative estimate. The full suite of nuisance parameters is represented as θ. These effect the expected signal and backgeound yields which are now represented as $s(\theta)$ and $b(\theta)$. Each component of θ is associated with a default value $\tilde{\theta}$, reflecting our degree of belief on the real value of θ. The pdf (probablity distribution function) $\rho(\theta|\tilde{\theta})$ can then be interpreted as a posterior distribution from measurements of $\tilde{\theta}$. Using Bayes' theorem:

$$\rho(\theta|\tilde{\theta}) = \rho(\tilde{\theta}|\theta) \cdot \pi_\theta(\theta), \tag{7.3}$$

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

the introduction of systematic uncertainties now becomes:

$$\mathcal{L}(data|\mu,\theta) = Poisson(data|\mu \cdot s(\theta) + b(\theta)) \cdot \rho(\tilde{\theta}|\theta) \tag{7.4}$$

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

$$\rho(\theta|\tilde{\theta}) = \frac{1}{\sqrt{2\pi}\ \ln(\kappa)}\exp\ (\frac{\ln(\theta/\tilde{\theta})^2}{2(\ln\ \kappa)^2})\frac{1}{\theta} \tag{7.5}$$

Systematic uncertainties that effect the scale of the distribution differently in each been have the effect of altering its shape along with its scale. Such uncertainties are called shape uncertainties [39], and are modeled using a linear extrapolation method [40]. In practice, two alternate distributions obtained by varying the nuisance by ± 1 standard deviation are used, and a parameter is added to the likelihood that smoothly interpolates between these shapes.

7.2.3 Calculation of exclusion limits

The CL_s method [41–43] is used to set upper exclusion limits when no excess of data over background is observed. The test statistic used generally for hypothesis testing in searches at the LHC, uses profiling of nuisances as described above, and is based on the likelihood ratio [44], which by the Neyman-Pearson lemma is known as the most powerful discriminator. This is denoted by $\tilde{q_\mu}$, and is given by:

$$\tilde{q_\mu} = -2\ \ln\frac{\mathcal{L}(\text{data}|\mu,\hat{\theta_\mu})}{\mathcal{L}(\text{data}|\hat{\mu},\hat{\theta})},\ \text{with}\ 0 \leq \mu \leq \hat{\mu} \tag{7.6}$$

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

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

Now, using equation 7.6, the observed value of the test statistic, $\tilde{q_\mu}^{obs}$, is calculated for the signal strength μ. Also, maximum likelihood estimators for the nuisance parameters, for the background-only($\mu = 0$) and signal-plus-background(current $\mu > 0$ under consideration) hypotheses are calculated. They are denoted by $\hat{\theta_0}^{obs}$ and $\hat{\theta_\mu}^{obs}$ respectively, and are used to generate toy Monte carlo pseudo-datasets. These pseudo datasets are used to construct pdfs, using equation 7.6, of test statistics $f(\tilde{q_\mu}|0, \hat{\theta_0}^{obs})$ and $f(\tilde{q_\mu}|\mu, \hat{\theta_\mu}^{obs})$ by treating them as they were real data. Example of these distributions are shown in Fig. 7.1.

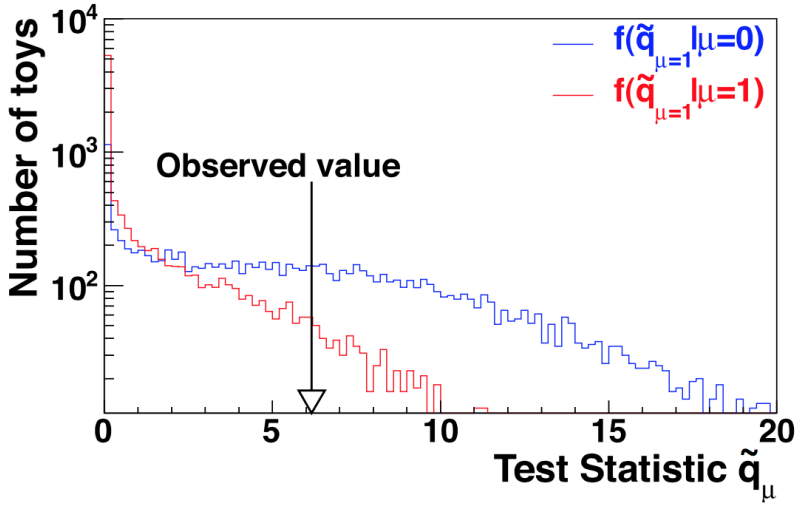


Figure 7.1: Test statistic distributions for ensembles of pseudo-data generated for signal-plus-background (red) and background-only (blue) hypotheses. [37]

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

$$p_\mu = P(\tilde{q_\mu} \geq \tilde{q_\mu}^{obs} | \text{signal-plus-background}) = \int_{\tilde{q_\mu}^{obs}}^{\text{inf}} f(\tilde{q_\mu} | \mu, \hat{\theta_\mu}^{obs}) d\tilde{q_\mu} \tag{7.7}$$

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

The second quantity we calculate is:

$$1 - p_b = P(\tilde{q_\mu} \geq \tilde{q_\mu}^{obs} | \text{background-only}) = \int_{\tilde{q_\mu}^{obs}}^{\text{inf}} f(\tilde{q_\mu} | 0, \hat{\theta_0}^{obs}) d\tilde{q_\mu} \tag{7.8}$$

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

The ratio of the two quatities referred to as CL_{s} [41–43] helps deal with both situations above well, and is given by:

$$\text{CL}_{\text{s}} = \frac{p_\mu}{1 - p_b} \tag{7.9}$$

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

7.2.4 Median expected Limits

Upper exclusion limits calculated using toy datasets of background-only expectation, are called expected limits. A large set of background-only pseudo-data is generated, and $\mathrm{CL_s}$ and $\mu^{95\%CL}$ is calculated for each of them. The median expected limit is calculated by integrating over this distribution until the 50% quantile is reached. The $\pm 1\sigma$ and $\pm 2\sigma$ bands are calculated similary by integrating the distribution to the appropiate quantiles are reached. The calculation of median expected limits does not involve using the observed data and hence can be calculated when the analyses is blinded to prevent experimenter's bias (as mentioned in Section 5.1). This can be use to maximize the sensitivity of the search, as described in Sections 5.2.3 and 5.3.3. A more stringent(lower) median limit corresponds to a more sensitive search.

7.2.5 Quantifying an excess of events

In case an excess of data over background is observed, it is necessary to make sure beyond a reasonable doubt that the excess is not merely a fluctuation. This is quantified using the background-only p-value, which is the probability for the background to fluctuate and give an excess of events as large or larger than that observed. The same test statistic as equation 7.6 is used with the signal stength set to 0 to correspond to the background-only hypothesis:

$$\tilde{q_0} = -2\ \ln\frac{\mathcal{L}(\mathrm{data}|0,\hat{\theta}_0)}{\mathcal{L}(\mathrm{data}|\hat{\mu},\hat{\theta})},\ \text{with}\ 0 \leq \hat{\mu} \tag{7.10}$$

The constraint on $\hat{\mu}$ being greater than 0 is required so that a deficit of events in observed data is not interpreted in the same manner as we would an excess. In other words a departure from the background hypothesis in the form of deficit of events is not considered in favour of the signal hypothesis. Following the same procedure as calculation of observed limits (as described in 7.2.3) and generating pseudo-data, the

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

$$p_0 = P(\tilde{q_0} \geq \tilde{q_0}^{obs}) = \int_{\tilde{q_0}^{obs}}^{\inf} f(\tilde{q_0}|0,\hat{\theta_0}^{obs}) d\tilde{q_0} \tag{7.11}$$

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

$$p_0 = \int_{\mathcal{Z}_0}^{\inf} \frac{1}{\sqrt{2\pi}} exp(-x^2/2) dx \tag{7.12}$$

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

7.2.6 Systematic uncertainties

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

7.2.6.1 Normalization Uncertainties

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

The muons in the analysis are required to pass certain identification, isolation and triggering criteria (see chapter 5). The efficiencies for muon to pass these criteria are measured via tag-and-probe methods [22] using $Z \to \mu\mu$ events, and the scale factors are used to match the efficiency in MC to that in data. Briefly, the tag-and-probe method works in the following way. One of the muons (called the tag) is required to pass strict selection criterion, while the other (called the probe) is required to pass more relaxed criterion. Given the invariant mass of the $\mu - \mu$ system is required to within a narrow window of the Z mass, the probe muon is also very likely to be a real muon. The percentage of probe muons that pass the criterion we are testing for (identification, isolation, trigger etc.) gives the efficiency. The efficiencies determined by this process, like any other quantity, are associated with systematic uncertainties. For the muons used in the analyses described here, a combined normalization uncertainty of 2% is associated with muon trigger, identification and isolation. Similar to muons, the efficiencies for electrons used in the analyses have also been measured via tag-and-probe methods [24] using $Z \to$ ee events. The uncertainties in efficiencies of electron identification and isolation criterion are also included as a normalization uncertainty of 2% in the fit. Both the above uncertainties are applied to processes which are derived from MC simulation. As mentioned earlier, a b tagging veto is applied in the analysis in order supress backgrounds involving top quarks. The efficiency of b tagging procedure is different in MC simulation than data. A scaling procedure is applied to match these efficiencies, and the uncertainties associated with

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

Several backgrounds in the analyses are estimated using MC simulations (see chapter 6). These include Z $\to \tau\tau$, $t\bar{t}$, W+jets, WW, WZ and ZZ, $Z \to \ell\ell$ ($\ell = \text{e}, \mu$)+jets, single-top quark production, W$\gamma^{(*)}$ + jets. The production cross-sections for these backgrounds determine the number of events each background would contribute. These cross-sections are measured experimentally and the uncertainty in those measurements are included in the fit. Given that a change in cross-section changes the overall number of events produced, it has no effect on the shape of distributions. Hence these uncertainties are included as normalization uncertainties. These uncertainties in general arise from: uncertainties on the parton distribution functions and strong coupling constant (called PDF+α_s); variations in renormalization and factorization scales. In the H $\to \mu\tau_e$ analysis a separate uncertainty is applied for PDF+α_s and renormalization/factorization scales for each of the backgrounds. In h $\to \mu\tau_e$ analysis, a combined uncertainty for each background is applied to cover both sources. All the above uncertainties are considered 100% correlated among categories. For each background, a 5% uncertainty, uncorrelated among all categories, is also applied to conservatively cover differences across categories. The QCD multi-jet background is estimated using a data-driven procedure. An uncertainty of 30% associated with this procedure (corresponding to the uncertainty in the extrapolation factor from the same-sign to opposite-sign region) is included in the fit. All uncertainties are summarized in table 7.1.

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

Table 7.1: The systematic uncertainties for the four channels. All uncertainties are treated as correlated between the categories, except those with more values separated by the $\oplus$ symbol. In the case of two values, the first value is the correlated uncertainty and the second value is the uncorrelated uncertainty for each individual category. In the case of three values, the first and second values correspond to the uncertainties arising from factorization and renormalization scales and PDF variations and are correlated between categories, while the third value is the uncorrelated uncertainty for each individual category. Two values separated by the "–" sign represent the range of the uncertainties from the different sources and/or in the different jet categories.

Systematic uncertainty	$h \to \mu\tau_e$	$H \to \mu\tau_e$
Muon trigger/ID/isolation	2%	2%
Electron trigger/ID/isolation	2%	2%
b tagging veto	2.0–4.5%	2.0–2.5%
QCD multijet background	30%	30%
$Z \to \tau\tau$ + jets background	10%⊕5%	0.1%⊕2%⊕5%
$t\bar{t}$ background	10%⊕5%	10%⊕5%
W + jets background	10%⊕5%	0.8%⊕3.8%⊕5%
WW, ZZ, WZ background	5%⊕5%	3.5%⊕5%⊕5%
$W\gamma^{(*)}$ background	10%⊕5%	10%⊕5%
Single top quark background	5%⊕5%	3%⊕5%⊕5%
$Z \to \mu\mu$/ee background	10%⊕5%	0.1%⊕2%⊕5%
[] Jet energy scale	3–20%	3–20%
μ energy scale	0.2%	0.2%
e energy scale	0.1–0.5%	0.1–0.5%
Unclustered energy scale	$\pm1\sigma$	$\pm1\sigma$
pileup	$\pm1\sigma$	$\pm1\sigma$
Integrated luminosity	2.5%	2.5%

as PDF+α_s, and result in changes only in normalization. These uncertainties are summarized for SM Higgs boson and heavier Higgs of different masses in table 7.1. They are taken from Handbook of LHC Higgs cross-sections found in Ref. [45].

Table 7.2: Theoretical uncertainties from [45] are applied to the Higgs boson production cross sections for the different masses. In the reference, the PDF and α_s uncertainties are computed following the recommendation of the PDF4LHC working group. The remaining Gaussian uncertainty accounts for additional intrinsic sources of theory uncertainty described in detail in the reference.

m_{H} (GeV)	Production mode	Cross section (pb)	Theory, Gaussian (%)	PDF+α_s (%)
125	GGF	48.58	±3.9	±3.2
125	VBF	3.782	±0.4	±2.1
200	GGF	16.94	±1.8	±3.0
300	GGF		6.59	±1.8
±3.0				
450	GGF	2.30	±2.0	±3.1
600	GGF	1	±2.1	±3.5
750	GGF	0.50	±2.1	±4.0
900	GGF	0.27	±2.2	±4.6

The estimation of a particular background, that is derived from simulation, needs to correspond to the number of events (of that background, having a particular cross-section) that would be produced in the amount proton-proton collision data that we are using for this search. In other words, the background estimations need to be normalized to (brought to the same scale as) the integrated luminosity of the data collected. This integrated luminosity (defined in chapter 3) is a measured quantity, and like all measured quantities, has an uncertainty associated with it. This amounts to 2.5% and, like other normalization unceratainties, only effects only the overall scale of distributions.

7.2.6.2 Shape Uncertainties

In this section, we describe the systematic uncertainties which not only alter the scale but also the shape of the distributions. We start with the description of uncertainties associated with jet energy corrections. As described in Section 4.4.6, the reconstruction of complex objects such as jets need to corrected using several correction factors. These factors have uncertainties associated with them and these are included in the fit as shape uncertainties. There are several different sources of these uncertainties and the effect of each is propagated to the fit by including alternate distributions for each process where each source of uncertainty has been moved by one standard deviation on either side. The effect of changing these sources is propagated through the jets in the analysis, and also other affected quantitites such as the p_T^{miss}. A total of 27 sources are considered, and these include effects of pileup, composition of jets, η dependence etc. They vary in the range from 3-20% and are considered uncorrelated.

Just like jets, there are shape uncertainties associated with the energy scale of electrons and muons. The effect of electron energy scale uncertainty is treated in a similar manner by propagating the effect of varying the scale to process distributions which are then included in the fit. The uncertainties are a result of the sum in quadrature of the following components: electron selection efficiency, pseudorapidity dependence, and shower-shape related categorization.The resolution systematics result to be negligible and are thus not considered in the fit. The value of this uncertainty ranges from 0.1-0.5%. The muon energy scale uncertainty amounts to 0.2% and is treated in the same manner as above.

Jets with $p_T < 15$ GeV or PF candidates which do not get clustered inside any jets are called unclustered energy. Their scale that effects the p_T^{miss} in particular, and is associated with a shape uncertainty that is treated the same way as others [30]. Four sources of unclustered energy scale uncertainty are considered, and estimated

independently for these four particle categories: charged particles, photons, neutral hadrons, and HF (very forward, high $|\eta|$) particles which are not contained in jets. The effect of these sources on the unclustered energy is propagated in the same manner as described above, and they are considered uncorrelated.

A set of weights is applied in order make the distribution of pileup in MC simulation match that of pp collision data. There is an uncertainty associated with this process. This is included by varying the weights by changing by 5%, in each direction, the total inelastic cross section used in the estimation of the pileup events in data [46]. These new set of weights are then applied event-by-event, producing alternate distributions that are included in the fit as shape uncertainties.

A shape uncertainty is also considered to deal with the fact that purely statistical fluctuations can change the shape of the distribution. These uncertainties are called bin-by-bin uncertainties as they account for statistical uncertainty in every bin of the distribution. Alternate distributions are created by varying the contents of each bin up and down, and these distributions are included in the fit as shape uncertainties **??**. Given that considering all bins of all processes will result in a very large number of such nuisances being considered, shape uncertainties for only those bins are considered in which there is more than 10% variation in the up and down shift.

All shape uncertainties are summarized in Table 7.1.

CHAPTER 8

RESULTS

In this chapter the results of both the searches are presented. The results for the $h \to \mu\tau_e$ search are first presented. Results for the $H \to \mu\tau_e$ search follow.

8.0.1 $h \to \mu\tau_e$ results

The resulting distributions of the signal variable (after applying all selection requirements as outlined in 5.2) are fit using a binned maximum likelihood fit. The entire procedure is described in detail in 7.2. All systematic uncertainties are included as nuisance parameters, and the fit is performed simultaneously across all categories. The BDT response distributions of signal and background are shown superimposed for each category in Fig 8.1. The distribution of M_{col} for the M_{col}-fit analysis are also shown in Fig 8.2. We do not observe an excess of signal over expected background. Hence, upper exclusion limits on $\mathcal{B}(h \to \mu\tau_e)$ are set, following the procedure described in 7.2.3. In table 8.0.1, the median expected limits, observed limits and the best fit branching fractions for $\mathcal{B}(h \to \mu\tau_e)$ are summarized. As noted earlier in this thesis, the tau lepton coming from the Higgs can also decay hadronically. This channel of the LFV Higgs decay, i.e. $h \to \mu\tau_h$ is studied in an analyses by different members of the same research team [47]. The limits on $\mathcal{B}(h \to \mu\tau_h)$ from that search are combined with limits on $\mathcal{B}(h \to \mu\tau_e)$, as calculated above. All limits are summarized graphically in Figure 8.3. The combined best fit branching fraction of $\mathcal{B}(h \to \mu\tau)$ is found to be 0.00 ± 0.12 for the BDT-fit analysis.

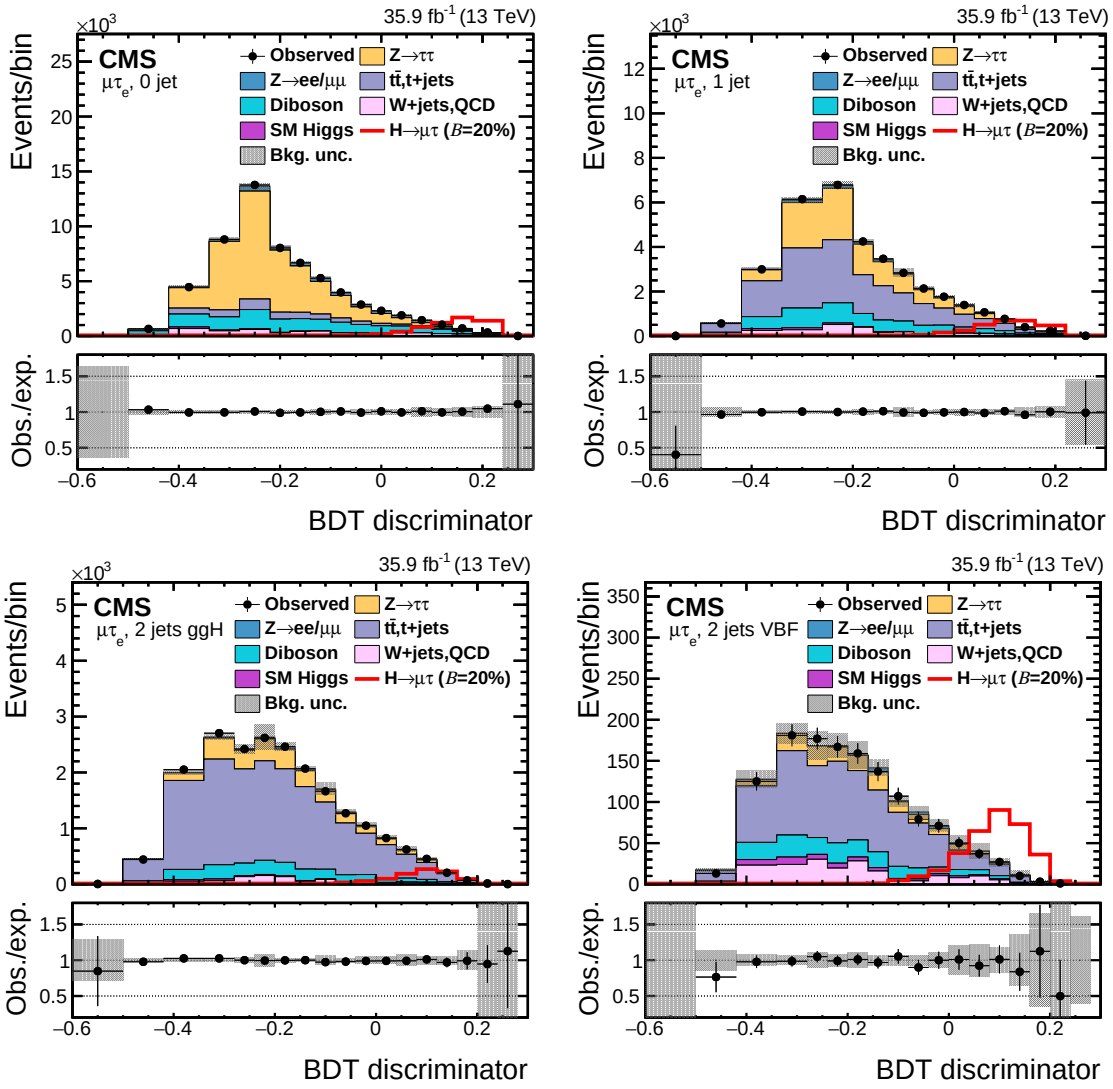


Figure 8.1: Distribution of BDT response in each category comparing signal and background estimations to observed collision data, for h $\rightarrow \mu\tau_e$ analysis. The bottom panel show the ratio of observed data and fitted background in each bin [47]

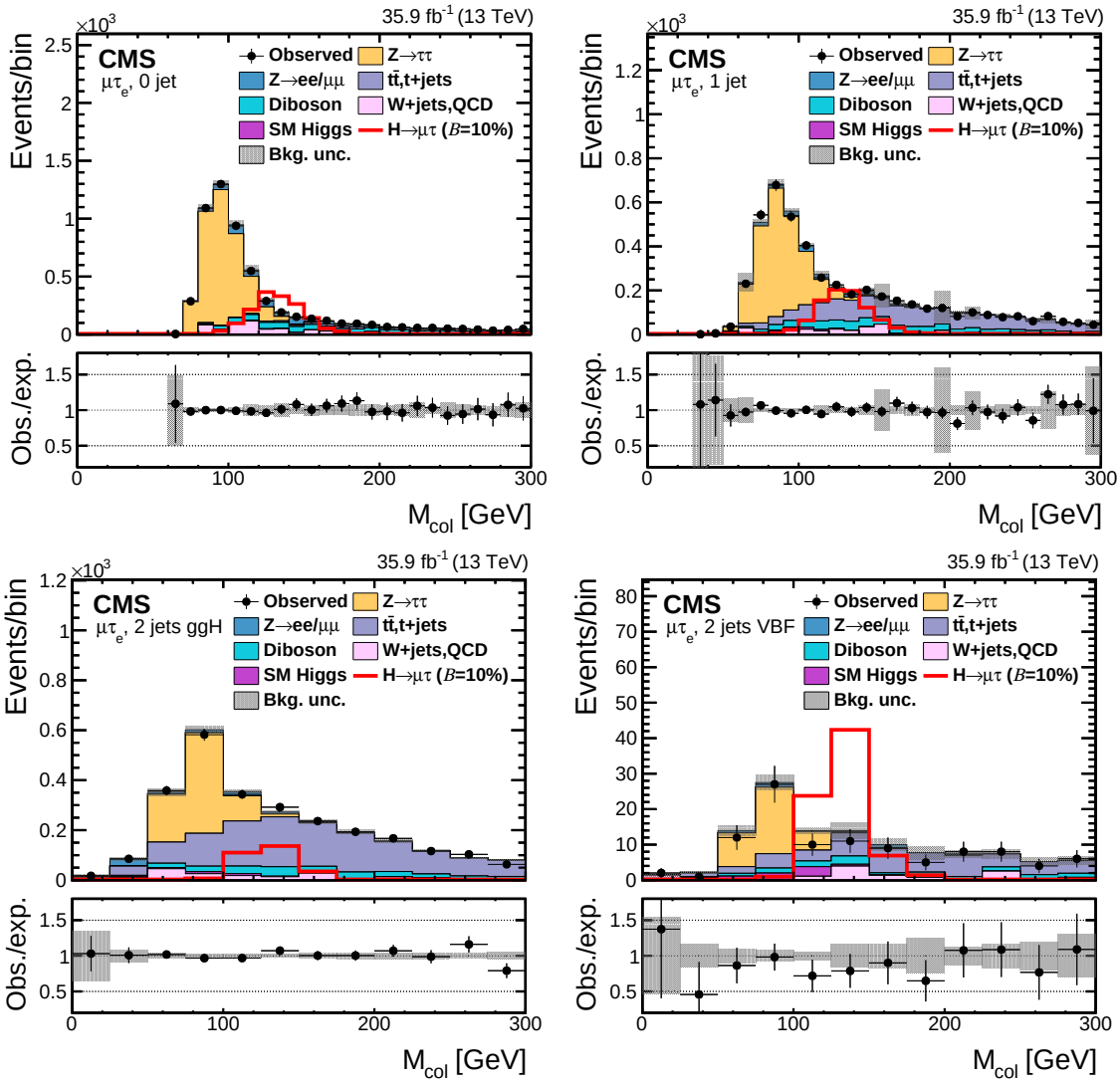


Figure 8.2: Distribution of M_{col} response in each category comparing signal and background estimations to observed collision data, for h $\rightarrow \mu\tau_e$ analysis. The bottom panel show the ratio of observed data and fitted background in each bin [47]

Table 8.1: Expected and observed upper limits at 95% CL, and best fit branching fractions in percent for each individual jet category, and combined,for the h $\to \mu\tau_e$ analysis.

Expected limits (%)

	0-jet	1-jet	2-jets	VBF	Combined
BDT fit analysis	<0.83	<1.19	<1.98	<1.62	<0.59
M_{col} fit analysis	<1.01	<1.47	<3.23	<1.73	<0.75

Observed limits (%)

	0-jet	1-jet	2-jets	VBF	Combined
BDT fit analysis	<1.30	<1.34	<2.27	<1.79	<0.86
M_{col} fit analysis	<1.08	<1.35	<3.33	<1.40	<0.71

Best fit branching fractions (%)

	0-jet	1-jet	2-jets	VBF	Combined
BDT fit analysis	0.61 ± 0.36	0.22 ± 0.46	0.39 ± 0.83	0.10 ± 1.37	0.35 ± 0.26
M_{col} fit analysis	0.13 ± 0.43	-0.22 ± 0.75	0.22 ± 1.39	-1.73 ± 1.05	-0.04 ± 0.33
combined $\mu\tau$ (BDT fit)	0.00 ± 0.12				

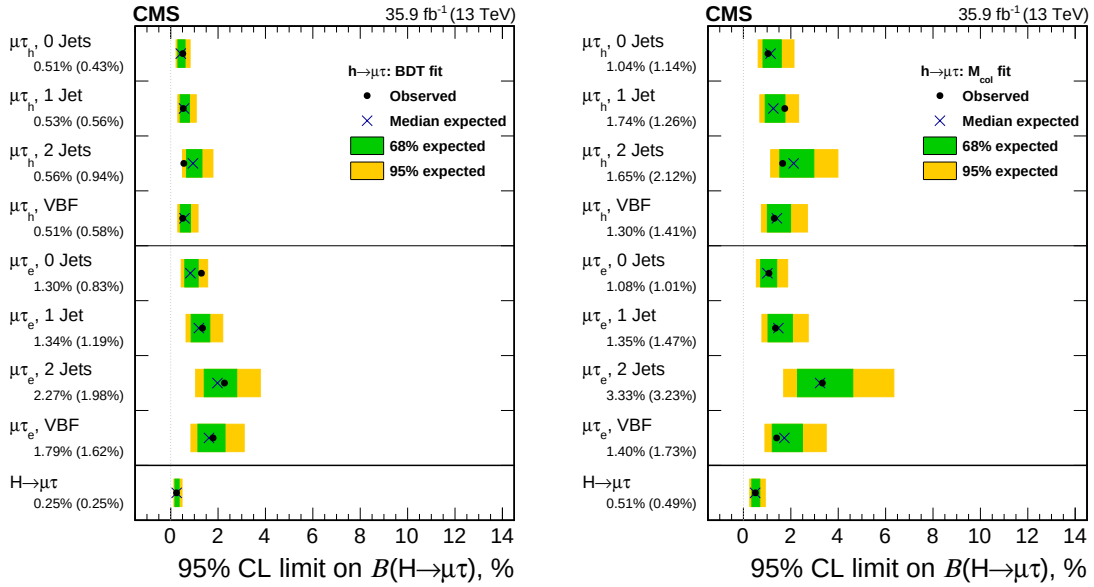


Figure 8.3: Observed and median expected upper exclusion limits for h $\to \mu\tau_e$, h $\to \mu\tau_{\text{h}}$ and combined h $\to \mu\tau$ channels, for the BDT fit (left) and M_{col} fit analysis (right). The $\pm 1\sigma$ and $\pm 2\sigma$ bands for expected limits are also shown in light green and yellow respectively [47].

The constraints on $\mathcal{B}(\mathrm{h} \to \mu\tau)$ can be transformed into constraints on Lepton Flavor Violating Yukawa Couplings $(Y_{\mu\tau}, Y_{\tau\mu})$. These couplings represent the strength of an interaction and are related to the decay width $\Gamma(\mathrm{h} \to \mu\tau)$ in the following way [48]:

$$\Gamma(\mathrm{h} \to \mu\tau) = \frac{m_h}{8\pi}(|Y_{\mu\tau}|^2 + |Y_{\tau\mu}|^2). \tag{8.1}$$

The decay width is also related to the branching fraction, $\mathcal{B}(\mathrm{h} \to \mu\tau)$ according to the following equation:

$$\mathcal{B}(\mathrm{h} \to \mu\tau) = \frac{\Gamma(\mathrm{h} \to \mu\tau)}{\Gamma(\mathrm{h} \to \mu\tau) + \Gamma_{SM}}. \tag{8.2}$$

,where the SM Higgs decay width is assumed to be $\Gamma_{SM} = 4.1$ MeV [49] for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. Using equations 8.1 and 8.2, we derive the constraints on Yukawa couplings at 95% CL. The limits for the Yukawa couplings are summarized in Table 8.2. Fig. 8.4 pictorially summarizes all existing limits on Yukawa couplings from different direct and indirect searches. It also shows the theoretical "naturalness" limit considering/expecting LFV couplings to be smaller than those of couplings for SM decays of the Higgs [48], which can be considered a benchmark for sensitivity of this search. The limits derived from this search are most stringent till date, and surpass the above benchmark.

Table 8.2: 95% CL observed upper limit on the Yukawa couplings, for the BDT fit and the M_{col} fit analysis.

	BDT fit	M_{col} fit
$\sqrt{\lvert Y_{\mu\tau}\rvert^2 + \lvert Y_{\tau\mu}\rvert^2}$	$< 1.43 \times 10^{-3}$	$< 2.05 \times 10^{-3}$

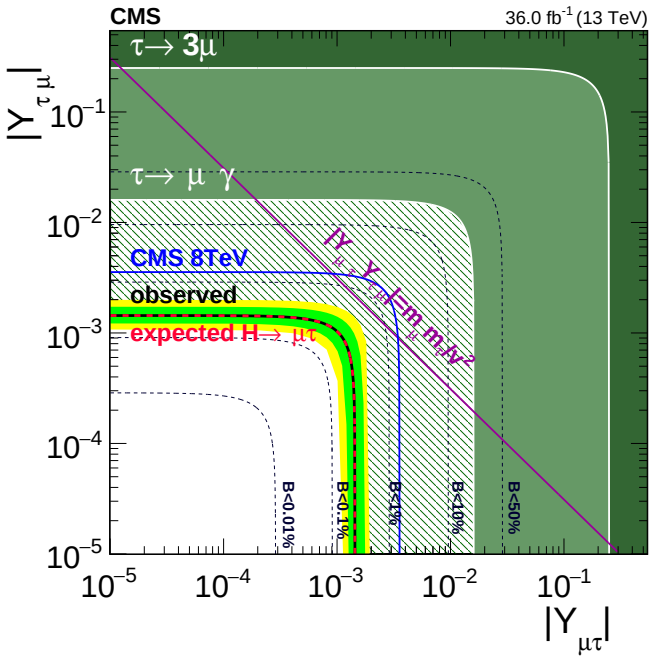


Figure 8.4: Observed (black solid)and median expected (red dashed) upper limits on h $\rightarrow \mu\tau$ Yukawa couplings from this analysis. The light green and yellow bands show the $\pm 1\sigma$ and $\pm 2\sigma$ spreads of the expected limit. Blue solid line shows the result from the previous CMS search with 8 TeV data [50]. The naturalness limit is shown as a purple straight line. [47]

8.0.2 $H \to \mu\tau_e$ results

The resulting M_{col} distributions for signal and background estimation (after applying all selection requirements as outlined in 5.3.1), after a binned maximum likelihood fit, are shown superimposed along with the observed data Fig 8.5. All systematic uncertainties are included as nuisance parameters, and the fit is performed simultaneously across all categories. We do not observe an excess over expected background in the entire range. Unlike the $h \to \mu\tau_e$ analysis described above where the production cross-section of the SM Higgs boson is known, here we are looking for LFV decay of a hypothetical heavy Higgs bosons of different masses. Hence, we set upper exclusion limits on production cross-section times branching fraction, $\sigma(gg \to H) \times \mathcal{B}(H \to \mu\tau_e)$. The procedure is the same as used above and described in 7.2.3. The observed and median expected upper limits at 95% CL on $\sigma(gg \to H) \times \mathcal{B}(H \to \mu\tau_e)$ are summarized in table 8.3 for different categories and Higgs masses. The limits are also summarized graphically in Figure 8.6. The observed (median expected) limits range from 159.4 (95.6) pb to 2.9 (4.9) pb for heavy Higgs masses in the range between 200 and 900 GeV. This search was combined with LFV heavy Higgs decay search with the tau lepton decaying hadronically, i.e. $H \to \mu\tau_h$ to produce constraints $H \to \mu\tau$. The combined observed (median expected) upper limits on $\sigma(gg \to H) \times \mathcal{B}(H \to \mu\tau)$ range from 51.9 (57.4) pb to 1.6 (2.1) pb. This is the first direct search till date to set limits on this decay.

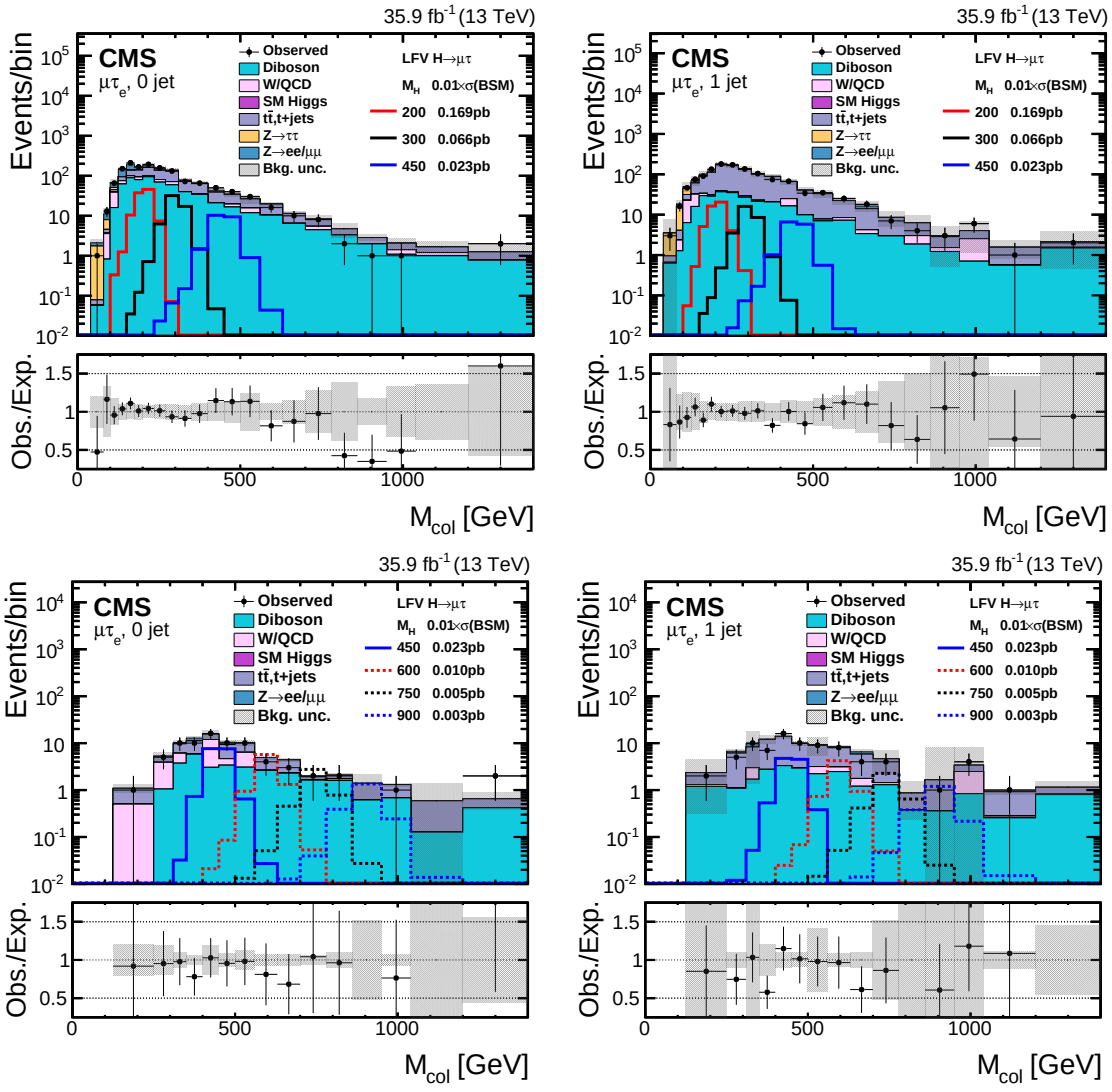


Figure 8.5: Distribution of M_{col} in 0-jet (left) and 1-jet (right) for lowmass (top) and highmass (range), comparing signal and background estimations to observed collision data, for $H \to \mu\tau_e$ analysis. The bottom panel show the ratio of observed data and fitted background in each bin [**?**]

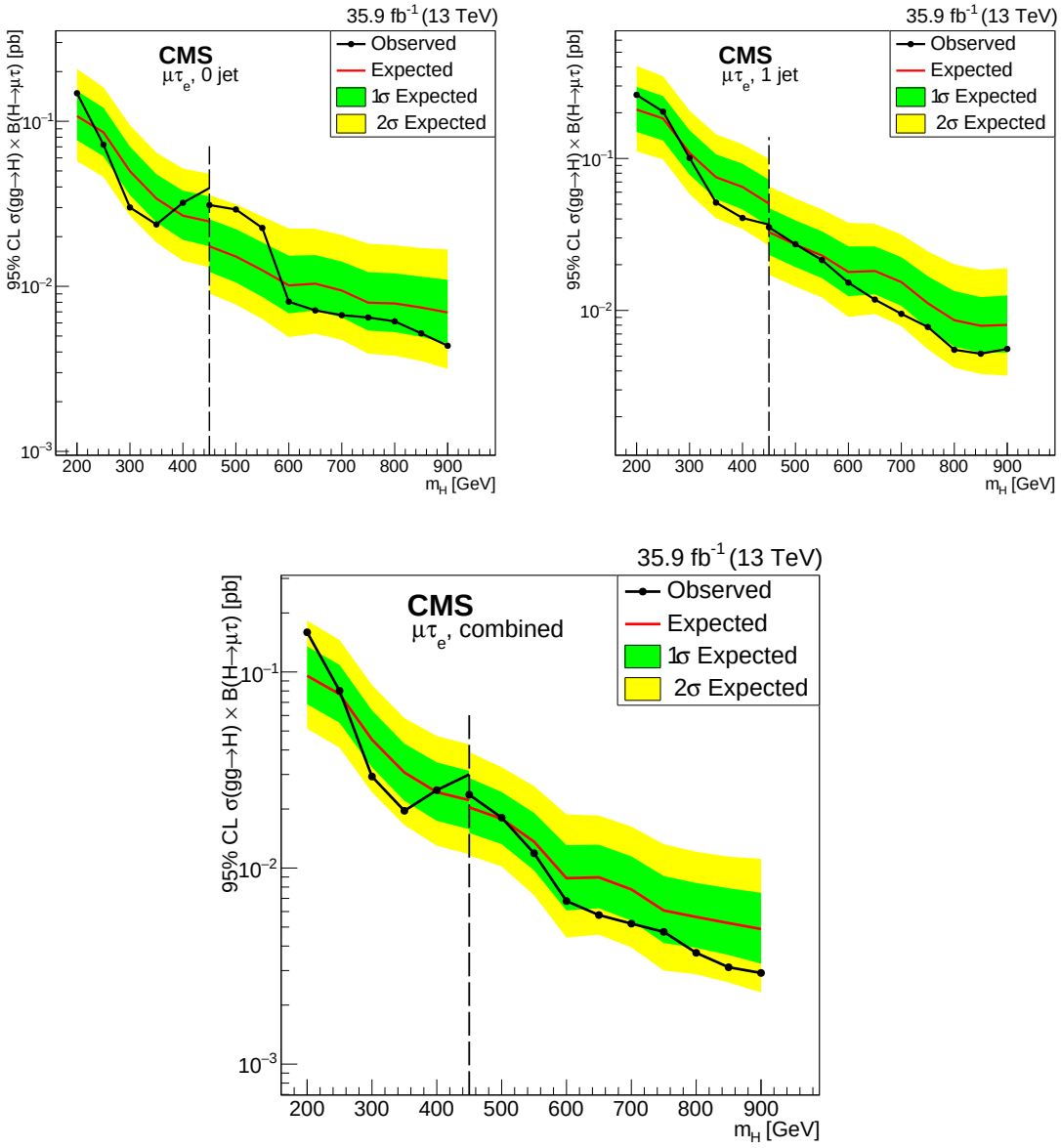


Figure 8.6: Observed and Median expected 95% upper exclusion limits for 0-jet (upper left), 1-jet (upper right) and combined (bottom),for the H $\rightarrow \mu\tau_e$ analysis. [**?**]

Table 8.3: The observed (median expected) 95% CL upper limits on $\sigma(\mathrm{gg} \to \mathrm{H}) \times \mathcal{B}(\mathrm{H} \to \mu\tau_e)$.

m_{H} (GeV)	0 jet	1 jet	comb
200	147.8 (107.5)	262.1 (209.8)	159.4 (95.6)
300	30.1 (49.8)	100.8 (108.6)	29.3 (45.2)
450	31.1 (17.5)	35.3 (32.8))	23.7 (20.4)
600	8.1 (10.4)	15.2 (17.9)	6.8 (8.9)
750	6.5 (8.0)	7.8 (18.2)	4.7 (6.1)
900	4.4 (6.9)	5.6 (15.4)	2.9 (4.9)

CHAPTER 9

CONCLUSION

APPENDIX A

BOOSTED DECISION TREES

A.1 Introduction

BIBLIOGRAPHY

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

26. A. M. Sirunyan et al. Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13: P10005, 2018. doi: 10.1088/1748-0221/13/10/p10005.

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

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

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

30. Albert M Sirunyan et al. Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector. *JINST*, 14:P07004, 2019. doi: 10.1088/1748-0221/14/07/P07004.

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

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

33. 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.

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

35. 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.

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

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

38. Wikipedia. Poisson distribution. Website, . `https://en.wikipedia.org/wiki/Poisson_distribution`.

39. 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.

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

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

42. 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.

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

44. Eilam Gross Glen Cowan, Kyle Cranmer and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *European Physics Journal C*, 71, February 2011.

45. LHC Higgs Cross Section Working Group. Handbook of LHC Higgs cross sections: 4. deciphering the nature of the Higgs sector. *CERN*, 2016. doi: 10.23731/CYRM-2017-002.

46. Albert M Sirunyan et al. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV. *JHEP*, 07:161, 2018. doi: 10.1007/JHEP07(2018)161.

47. CMS Collaboration. Search for lepton flavour violating decays of the Higgs boson to $\mu\tau$ and eτ in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 06:001, 2018. doi: 10.1007/JHEP06(2018)001.

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

49. 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.

50. Vardan Khachatryan et al. Search for lepton-flavour-violating decays of the Higgs boson. *Phys. Lett. B*, 749:337, 2015. doi: 10.1016/j.physletb.2015.07.053.

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