Search For The Higgs Boson Decaying Into τ-Leptons In The Di-Electron Channel

Dissertation

zur Erlangung des Doktorgrades

des Department Physik

der Universität Hamburg

vorgelegt von

Jakob Maximilian Henry Salfeld-Nebgen

aus Melbourne

Hamburg

2014

Gutachter der Dissertation:	Prof. Dr. Peter Schleper Prof. Dr. Georg Weiglein
Gutachter der Disputation:	Prof. Dr. Peter Schleper Prof. Dr. Christian Sander
Datum der Disputation:	16.6.2014
Vorsitzender des Prüfungsausschusses:	Dr. Georg Steinbrück
Vorsitzender des Promotionsausschusses:	Prof. Dr. Peter Hauschildt
Leiterin des Fachbereichs Physik: Dekan der MIN-Fakultät:	Prof. Dr. Daniela Pfannkuche Prof. Dr. Heinrich Graener

Abstract

The first dedicated search for Higgs bosons decaying into tau pairs with two electrons and 4 neutrinos in the final state is presented. The search is performed with the CMS detector at the LHC based on an integrated luminosity of 4.9 fb^{-1} and 19.7 fb^{-1} at 7 TeV and 8 TeV, respectively. Events with two reconstructed electrons are selected before extensive multivariate analysis techniques are utilized to achieve an optimal background rejection. The ee-channel alone excludes $3.2 \times (\sigma^H_{SM} \times BR(H \to \tau\tau))$ for $m_H = 120$ GeV/c^2 and $3.7 \times (\sigma^H_{SM} \times BR(H \to \tau\tau))$ for $m_H = 125$ GeV/c^2 at 95% confidence level.

The ee-channel is combined into the official CMS $H \to \tau\tau$ analysis, to find, for the first time, direct evidence for couplings of the new Higgs-like boson to down-type fermions with 3.2σ [1] [2].

Kurzfassung

In dieser Arbeit wird zum ersten Mal eine Suche nach in 2 Tau-Leptonen zerfallende Higgs Bosonen mit 2 Elektronen und 4 Neutrinos im Endzustand vorgestellt. Die Suche basiert auf den Daten des CMS Experiments am LHC mit einer integrierten Luminosität von 4.9 fb^{-1} und 19.7 fb^{-1} bei einer Schwerpunktsenergie von 7 TeV and 8 TeV, welche in den Jahren 2011 und 2012 gesammelt wurden. Bevor eine komplexe multivariate Analysemethode angewendet wird, werden Ereignisse mit zwei rekonstruierten Elektronen selektiert. Der ee-Kanal alleine schliesst eine Higgs Signalstärke von $3.2 \times (\sigma^H_{SM} \times BR(H \to \tau\tau))$ für $m_H = 120$ GeV/c^2 und $3.7 \times (\sigma^H_{SM} \times BR(H \to \tau\tau))$ für $m_H = 125$ GeV/c^2 mit einem Konfidenzintervall von 95% aus.

Der ee-Kanal ist Teil der offziellen CMS $H \to \tau\tau$ Analyse, welche zum ersten Mal einen experimentellen Nachweis fuer die Kopplung des Higgs-ähnlichen Teilchens an down-type Fermionen mit einer Signifikanz von 3.2σ liefert [1] [2].

Dedicated to my Mother, Micaela Salfeld-Nebgen (†3.6.2006)

Contents

1. Introduction

The recent discovery of a new boson at the Large Hadron Collider (LHC) announced by the ATLAS and CMS collaborations on 4th of July 2012 [3] [4], was a milestone in the more than 50 year old quest for understanding the origin of electroweak symmetry breaking and thus mass of elementary particles. Subsequent measurements of both collaborations have indeed independently confirmed, that no significant deviations from the properties of the prognosticated scalar boson, introduced by the hypothesis of electroweak symmetry breaking, can be observed. Direct and indirect measurements of couplings of the new boson to Standard Model particles revealed compatibility, within uncertainties, with the predicted couplings of the Higgs boson, and spin and CP measurements favored the scalar CP-even hypothesis over CP-odd and non-zero integer spin hypotheses.

As a consequence of these experimental discoveries and observations made by the ATLAS and CMS collaborations, the Physics Nobel Prize 2013 was awarded to François Englert and Peter Higgs for the development of spontaneously broken local gauge theories and theoretical insights on the origin of mass [5] [6] [7]. Inspired by spontaneous gauge symmetry breaking in the BCS-theory of superconductivity [8], where photons acquire effective mass [9] when traveling through certain low-temperature materials due to the spontaneously broken $U(1)$ symmetry by the condensate of cooper-pairs in the ground-state, the Higgs (or BEH) mechanism introduces a scalar field with non-zero vacuum state to spontaneously break the local gauge symmetry and thereby gives mass to the corresponding vector bosons. The Higgs mechanism was incorporated into the framework of the $SU(2)_L \times U(1)_Y$ Standard Model electroweak gauge theory by Steve Weinberg [10] and Abdus Salam [11] and the renormalizability of spontaneously-broken gauge theories was subsequently proven by Geradus 't Hooft and Martin Veltmann [12] [13]. Fermions are predicted to acquire masses by so-called Yukawa couplings to the broken scalar field, preserving chiral and gauge symmetry [10].

As stated above, the theory predicts the existence of the Higgs boson, which is the only physical degree of freedom of the associated Higgs field introduced by the Higgs mechanism. Except for the mass of the Higgs boson itself, all couplings, the total width and quantum numbers of the Higgs boson can be predicted by the Standard Model.

The current status of Higgs property measurements can be found in [14] for the ATLAS collaboration, and in [15] for the CMS collaboration and shows good agreement with the Standard Model predictions for the Higgs boson. ATLAS and

CMS measure a combined signal strength of $\mu = 1.30 \pm 0.12(\text{stat})^{+0.14}_{-0.11}(\text{sys})$ and $\mu = 0.80 \pm 0.14$, respectively. The Higgs masses are measured by both collaborations in the $H \to \gamma\gamma$ and 4-lepton $H \to ZZ$ channels, and to be compatible with $m_H = 125.5 \pm 0.2(\text{stat})^{+0.5}_{-0.6}(\text{sys})$ GeV/c^2 for ATLAS and $m_H = 125.7 \pm 0.3(\text{stat}) \pm 0.3(\text{sys})$ GeV/c^2 for CMS.

Remarkably, both experiments find, for the first time, direct evidence for fermionic couplings of the new bosons [1] [16], and thereby establish strong evidence for the the existence of Yukawa couplings. Via gluon fusion Higgs production processes, indirect evidence for couplings to up-type quarks has already been established. In this respect, the results can be interpreted as first evidence for couplings of the new bosons to down-type fermions. The $H \to \tau\tau$ decay constitutes the most sensitive channel to study fermionic Higgs couplings. ATLAS observes a signal strength of $\mu = 1.4^{+0.5}_{-0.4}$ with an observed significance of 4.1 standard deviations, when 3.2σ are expected. CMS observes a signal strength of $\mu = 0.78 \pm 0.27$ with an observed significance of 3.2 standard deviations, when 3.7σ are expected.

The analysis presented in this thesis is part of the CMS combined $H \to \tau\tau$ search and covers one of the six inclusive channels, the ee-channel.

About the Analysis This thesis presents the first dedicated search for the Standard Model Higgs boson decaying into a pair of tau-leptons with 2 electrons and 4 neutrinos in the final state with the CMS detector. The analysis finally completes the full coverage of all final states of the inclusive search of Higgs bosons decaying into tau-leptons, performed by the CMS Collaboration.

At Tevatron, the ee-channel of the $H \to \tau\tau$ search has not been considered by the CDF and D0 collaborations [17] [18].

The ATLAS Collaboration uses the same analysis strategy for the ee, $\mu\mu$ and $e\mu$ ($\tau_{lep}\tau_{lep}$) channels [19] [16], only adjusting cuts on the di-lepton invariant mass, missing transverse energy and azimuthal angle between the two leptons, in order to suppress the Drell-Yan background present for the ee and $\mu\mu$ final states.

Early developments of the $\mu\mu$-channel [20], such as using tau decay length information (distance of closest approach significance) and a dedicated set of variables to discriminate Higgs signal against the overwhelming Drell-Yan background, have been studied and were found to be compatible with the ee-channel analysis. The strategy has gradually been re-designed to achieve a gain of about 30% in terms of 95% CL upper limits based on the same 2011 and 2012 CMS data sample. Remarkably, the combination of the ee and $\mu\mu$ channels yields a sensitivity of only about 10% worse than the $e\mu$-channel, which suggests that the large $Z \to ee/\mu\mu$ background is indeed successfully achieved to be separated from the $H \to \tau\tau$ signal.

The main features and achievements of this new analysis can be summarized as follows:

- the full 2011 and 2012 datasets are analyzed, corresponding to an integrated luminosity of 24.6 fb^{-1}, split into 4.9 fb^{-1} and 19.7 fb^{-1} at 7 TeV and 8 TeV, respectively;

- the analysis utilizes centrally provided jet energy and electron energy (adopted from $H \to ZZ$ analysis) corrections. Customized corrections are applied to trigger and electron selection efficiencies, as well as missing transverse energy resulting from imperfect Z-recoil modeling in simulated samples;

- the distance of closest approach between the two electron tracks, incorporating the decay-length information of the two tau leptons, is used at several stages of the analysis. This observable is in particular independent on the spatial resolution of the primary vertex measurement. Sophisticated correction methods are applied, to derive the shape in simulated Drell-Yan samples in a data-driven way (cf. Section 4.8);

- an elaborate method, using the distance of closest approach, is developed to estimate, in a data-driven way, the overwhelming $Z \to ee$ background. This is done in several phase space regions of various variables, used to discriminate Higgs signal against background, while maintaining correlations among the variables (cf. Section 5.4.5);

- a new multivariate method to discriminate $H \to \tau\tau$ against the two backgrounds, $Z \to ee$ and $Z \to \tau\tau$, has been developed (cf. Section 5.2). Two specifically trained classifiers (boosted decision trees) are constructed, and combined into one final discriminant (cf. Section 5.3), preserving powerful discrimination against both backgrounds and providing good transparency concerning all Standard Model background predictions and uncertainties, in the maximum likelihood fit for final signal extraction;

- contributions to official multivariate based CMS electron identification methods have been made and a new electron identification method was developed specifically suitable for the $H \to \tau\tau$ search, where electrons have on average non-zero impact parameter. Depending on the electron phase space and working point, an improvement of $5 - 13\%$ electron selection efficiency, with respect to the currently used identification, is achieved (cf. Section4.3).

The ee-channel is part of the combined CMS $H \to \tau\tau$ analysis, establishing first direct evidence for couplings of the new boson to down-type fermions. The result is published in [1] and [2], where in the latter the $H \to \tau\tau$ analysis has been combined with the $VH \to l + b\bar{b}$ analysis.

This thesis is organized as follows:

- Chapter 2 introduces the Higgs mechanism in the Standard Model, discusses the important Higgs production mechanisms at the LHC and the decay into tau leptons, to state the hypothesis to be tested in this thesis;
- Chapter 3 describes the experimental environment at the LHC and the CMS detector. General information is given on LHC and the CMS Experiment, and CMS sub-detector components are described in slightly more detail, to show the data acquisition and general experimental circumstances, under which the analysis is performed;
- Chapter 4 gives detailed information on the reconstruction of jets, missing transverse energy, electrons and additional global event observables. The electron reconstruction and identification is exhibited in greater detail, as contributions to the CMS electron identification techniques have been made;
- Chapter 5 is the main part of this thesis. The first dedicated $H \to \tau\tau \to 2e4\nu$ analysis is presented in detail. Advanced techniques are deployed and developed for optimal signal discrimination from Standard Model background processes and largely data-driven background estimation methods utilized. Various centrally provided or customized corrections to important observables are applied, and all necessary systematic uncertainties considered, before final background and signal predictions are passed to a profile likelihood ratio method for signal extraction;
- Chapter 6 highlights the results of the ee-channel analysis alone, as well as in combination with other CMS $H \to \tau\tau$ analysis. In particular, the same-flavor lepton channels, ee and $\mu\mu$, are combined, both utilizing the newly developed analysis strategy described in this thesis;
- Chapter 7 briefly summarizes the results and gives a very short outlook concerning possible improvements and experimental difficulties for the future.

2. Theoretical Aspects

In this chapter, the theoretical framework is outlined to describe the hypothesis of electroweak symmetry breaking to be tested in this thesis. A short summary of the Standard Model of particle physics is given in Section 2.1 with an emphasis on the electroweak theory and the Higgs mechanism [5] [6] [7] [21] [22] [23]. Section 2.2 summarizes the important production mechanisms of the Higgs boson at the LHC. Higgs boson decays into tau-leptons are finally discussed in Section 2.3.

2.1. Standard Model and Electroweak Symmetry Breaking

The Standard Model of particle physics (SM) [24] [10] [11] is the theoretical model presently used to describe the physical properties and dynamics of matter in terms of its fundamental constituents and their interactions. Based on global Poincaré space-time symmetry and local gauge invariance under transformations of the direct product of compact Lie groups $SU(3)_C \times SU(2)_L \times U(1)_Y$, as well as experimentally observed properties, the SM accomodates all known elementary particles of spin-1/2 fermions and spin-1 gauge bosons into irreducible representations of the gauge groups within the framework of a renormalizable Quantum Field Theory, as shown in Table 2.1. Matter particles, consisting of leptons and quarks, appear as right- and left-handed spinors living in irreducible representations of the Poincaré group. Gauge bosons are structured in adjoint representations of the corresponding gauge groups. The octet of gluons G^a_μ for the $SU(3)_C$ (C is color) gauge group establishes the interactions of Quantum Chromodynamics (QCD, strong force) by acting on $SU(3)_C$ triplets (fundamental representation) of quarks. In the electroweak sector, the gauge bosons are represented by the $SU(2)_L$ isotriplet W^i_μ together with the isosinglet B_μ of $U(1)_Y$. Left(right)-handed elementary fermion fields (quarks and leptons) are doublets(singlets) under $SU(2)_L$ and are charged under $U(1)_Y$, where Y is, together with the third component of the $SU(2)_L$ generator, isospin component I_3, related to the electromagnic charge via $Q = I_3 + \frac{Y}{2}$. I_3 is given by the dimension of the irreducible representation $SU(2)_L$: 0 for trivial states, $\pm\frac{1}{2}$ for doublets and $\pm$1,0 for the adjoint representation(isotriplet). All quarks and leptons appear in three generations. The left-handed e_L(electron), μ_L(muon), τ_L(tau-lepton) together with the associated neutrinos ν^i_L, form $SU(2)_L$ doublets, whereas right-handed neutrinos are not observed in nature, leading to right-handed

	field			spin	$SU(3)_C$	$SU(2)_L$	Y
quarks	$\begin{pmatrix} u \\ d \end{pmatrix}_L$	$\begin{pmatrix} c \\ s \end{pmatrix}_L$	$\begin{pmatrix} t \\ b \end{pmatrix}_L$	1/2	**3**	**2**	1/3
	u_R	c_R	t_R	1/2	**3**	**1**	4/3
	d_R	s_R	b_R	1/2	**3**	**1**	-2/3
leptons	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	1/2	**1**	**2**	-1
	e_R	μ_R	τ_R	1/2	**1**	**1**	-2
Higgs-doublet		$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}_L$		0	**1**	**2**	1
gauge bosons		G^a_μ		1	**8**	**1**	0
		W^i_μ		1	**1**	**3**	0
		B_μ		1	**1**	**1**	0

Table 2.1.: The field content of the Standard Model of particle physics. The Higgs boson, quarks and leptons are shown in their $SU(2)_L$ representations. Also shown, is the corresponding spin, hypercharge Y and representation under $SU(3)_C$, $SU(2)_L$ of each field.

$SU(2)_L$ singlets e_R, μ_R, τ_R. Accordingly, left-handed up- and down-type quarks, u(up) and d(down), s(strange) and c(charm) as well as t(top) and b(bottom) form $SU(2)_L$ doublets, and their right-handed counterparts form $SU(2)_L$ singlets.

The elementary particle content and fundamental interactions described by the SM, have been experimentally scrutinized up to the TeV scale with high accuracy, finding only very few deviations (e.g. muon anomalous magnetic moment of about 3σ) from the precise predictions of the SM.

The concept of gauge symmetry in general forbids mass terms for all fields, as these would directly spoil the gauge symmetry, which is in contradiction to the observed physically massive $W^\pm$ and Z bosons mediating the weak force and the massive leptons and quarks. This fact suggests, that the electroweak gauge symmetry is only a *hidden symmetry of Nature* or in other words *spontaneously broken*. The mechanism to generate masses of gauge bosons by spontaneaously breaking the corresponding gauge symmetries was found 1964 independently by Higgs and Brout, Englert. The accordingly named *Higgs mechanism* or *BEH mechanism*, states, that in contrast to the Goldstone Theorem applicable to global symmetries, the unphysical degrees of freedom for a broken scalar field due to an (partly)

unsymmetric vacuum state under a local gauge transformation get absorbed into longitudinal components of the corresponding gauge bosons, which thereby acquire masses. Fermions obtain masses indirectly, by *Yukawa couplings* to the introduced scalar field with degenerate vacuum state. It is in particular the aim of the analysis presented in this thesis to seek first experimental evidence for the existence of such Yukawa couplings.

The Higgs mechanism is a central part of the electroweak sector of the SM and shall be described in more detail in the following. In the electroweak sector of the SM gauge theory, the Lagrangian consists of three parts (QCD interactions are omitted):

$$\mathcal{L}_{\mathrm{SM}}^{\mathrm{EWK}} = \mathcal{L}_{\mathrm{gauge}} + \mathcal{L}_{\mathrm{fermion}} + \mathcal{L}_{\mathrm{Higgs}}. \tag{2.1}$$

The term $\mathcal{L}_{\mathrm{gauge}} \propto F^{\mu\nu}F_{\mu\nu}$ yields the usual curvature or field strength with tensors associated to the gauge fields which in turn take values in the Lie algebra of $SU(2)_L \times U(1)_Y$ with structure constants ϵ^{ijk} for the non-abelian part. $\mathcal{L}_{\mathrm{fermion}}$ is the kinematic term for all left- and right-handed fermions $f^i_{L,R}$ of the form ${f^i_{L,R}}^{\dagger}\bar{\gamma}^{\mu}D_{\mu}f^i_{L,R}$, where γ^{μ} are the gamma matrices. The covariant derivative for the $SU(2)_L \times U(1)_Y$ SM gauge theory can be written as

$$D_{\mu} = \partial_{\mu} + ig_2\mathbf{I}^i W^i_{\mu} + ig_1\frac{Y}{2}B_{\mu}\,, \tag{2.2}$$

with g_2 and g_1 being the coupling constants of the $SU(2)_L$ and $U(1)_Y$ interactions, respectively. The operators $\mathbf{I}^i$ and Y are the gauge group generators and a suitable representation can be written as $(Y, \mathbf{I}^i) = (\mathbf{1_2}, \sigma^{\mathbf{i}})$ (σ^i are pauli matrices). A gauge transformation, under which the electroweak lagrangian is invariant, is in general given by the following

$$\psi(x) \to \psi'(x) = e^{i\xi_a(x)\mathbf{I}^a + i\beta(x)Y}\psi(x)\,, \tag{2.3a}$$

$$W^a_{\mu} \to W'^a_{\mu} = W^a_{\mu} - \frac{1}{g_2}\partial_{\mu}\xi^a(x) - f^a_{bc}\xi^b W^c_{\mu}\,, \tag{2.3b}$$

$$B_{\mu} \to B'_{\mu} = B_{\mu} - \frac{1}{g_1}\partial_{\mu}\beta(x)\,. \tag{2.3c}$$

The Higgs mechanism manifests itself by introducing an additional scalar $SU(2)_L$ doublet

$$\Phi(x) = \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix} = \begin{pmatrix} \phi_1(x) + i\phi_2(x) \\ \phi_3(x) + i\phi_4(x) \end{pmatrix}. \tag{2.4}$$

with the lagrange density

$$\mathcal{L}^{\text{Higgs}} = (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi) + \mathcal{L}_{\text{Yukawa}} \tag{2.5}$$

$$= (D^\mu \Phi)^\dagger (D_\mu \Phi) + \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2 + \mathcal{L}_{\text{Yukawa}}. \tag{2.6}$$

The Higgs field has according to its $SU(2)_L$ doublet state and electromagnetic charge of the upper component, a hypercharge of $Y_\Phi = 1$. To spontaneously break the symmetry, the parameters in the Higgs potential $V(\Phi)$ are chosen, such that $\mu^2 > 0$ and $\lambda > 0$. This inevitably leads to a degenerate non-zero vacuum state of the Higgs field minimizing the potential, with $\Phi^\dagger \Phi = \sum_i \phi_i^2 = \frac{\mu^2}{2\lambda} = \frac{1}{2} v^2$. By choosing the *unitary gauge*, the vacuum expectation value (vev) of the Higgs field can be written as

$$\langle \Phi_0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{2.7}$$

The vacuum state carries weak- and hypercharge, and is therfore not invariant (non-vanishing) under general $SU(2)_L \times U(1)_Y$ gauge transformations. However, due to the correct choice of the Higgs hypercharge, the combination of generators $I_3 + Y/2$ annihilates the vacuum state, and thus the $SU(2)_L \times U(1)_Y$ gauge symmetry is spontaneously broken to the $U(1)_{em}$ gauge symmetry of electromagnetism. It is due to this, that the photon field stays massless and the three bosons of the weak force have mass. The choice of the unitary gauge is arbitrary, and it important to note, that generating masses with the Higgs mechanism is independent from the choice of gauge, and hence does not spoil renormalizability of the SM. In the unitary gauge, the unphysical degrees of freedom (Goldstone Bosons) of the Higgs field are absorbed into the gauge bosons, as shown in equation 2.3b and 2.3c. Still, the Goldstone Bosons interact with the electroweak generators via $v g^2 \xi^a(x) \partial_\mu W^{a,\mu}$ and can therefore be interpreted as *longitudinal components* in the direction of the momenta. The Higgs field is therefore transformed, such that

$$e^{i\xi_a(x)\mathbf{I}^a + i\beta(x)Y} \Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{2.8}$$

where all introduced fields have vanishing vev, and $H(x)$ describes "radial" excitations of the Higgs vev. To illustrate the physical particle content, the bosonic

fields can be rotated, without loss of generality, into the physical states:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \qquad \text{charged currents,} \tag{2.9}$$

$$Z_\mu^0 = \frac{g_2 W_\mu^3 - g_1 B_\mu}{\sqrt{g_1^2+g_2^2}} \qquad \text{neutral current,} \tag{2.10}$$

$$A_\mu = \frac{g_1 W_\mu^3 + g_2 B_\mu}{\sqrt{g_1^2+g_2^2}} \qquad \text{photon.} \tag{2.11}$$

The mixing of the third weak component and the $U(1)_Y$ generator is often expressed in terms of the weak mixing angle θ_W defined by $\sin(\theta_W) = g_1/\sqrt{g_1^2+g_2^2}$ and $\cos(\theta_W) = g_2/\sqrt{g_1^2+g_2^2}$. The Lagrange density for the Higgs field in the unitary gauge, is then of the form (constant term is omitted)

$$\begin{aligned}\mathcal{L}^{\text{Higgs}} = &\big(\frac{1}{2}\partial_\mu H \partial^\mu H - \lambda v^2 H^2\big) + \big(\frac{1}{4}v^2 g_2^2 W_\mu^+ W^{-\mu} + \frac{1}{8}(g_1^2+g_2^2)v^2 Z_\mu Z^\mu\big) \\ &+ \big(\frac{1}{2}g_2^2 v H W_\mu^+ W^{-\mu} + \frac{1}{4}(g_1^2+g_2^2) v H Z_\mu Z^\mu \\ &+ \frac{1}{4}g_2^2 H^2 W_\mu^+ W^{-\mu} + \frac{1}{8}(g_1^2+g_2^2) H^2 Z_\mu Z^\mu\big) - \lambda v H^3 - \frac{1}{4}\lambda H^4 \\ &+ \mathcal{L}_{\text{Yukawa}}\,.\end{aligned} \tag{2.12}$$

As can be seen, three mass terms appear for the $W_\mu^\pm$ and Z^0 bosons, with $m_{W^\pm} = vg_2/2$ and $m_Z = \frac{v}{2}\sqrt{g_1^2+g_2^2}$, and additionally the physical degree of freedom of the Higgs field gives rise to a particle with a mass of $m_H = v\sqrt{2\lambda}$, the *Higgs boson*. The Higgs field Lagrange density introduces two new parameters, λ and v (vev), of which v can be fixed by precision measurements of the Fermi Constant ($G_F = \frac{g_2^2\sqrt{2}}{8m_W^2}$) via muon decays, yielding $v = (\sqrt{2}G_F)^{-1/2} \approx 246$ GeV. With precise measurements of the mass of the Higgs-like boson found by Atlas and CMS, λ can be fixed, and thus the strength of the quartic Higgs coupling in equation 2.12 is fixed.

Masses of fermions, and especially τ-leptons, are generated via Yukawa couplings. The corresponding terms are $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge invariant and therefore necessary to be added to the renormalizable SM Lagrange density:

$$\mathcal{L}_{\text{Yukawa}} = -\sum_{i,j=1}^{3}\Big[y_{ij}^d (q_L^i)^\dagger \Phi d_R^j \;+ y_{ij}^u (q_L^i)^\dagger \Phi^c u_R^j \;+ y_{ij}^l (l_L^i)^\dagger \Phi l_R^j \;+ \text{h.c.}\Big]\,, \tag{2.13}$$

with $\Phi^c = i\sigma_2 \Phi^*$, and Yukawa terms for neutrino masses are omitted due to the absence of right-handed neutrinos. The complex 3×3 matrices y_{ij}^f with $f = (u, d, l)$ and $i, j = (1, 2, 3)$ (quark/lepton generation), are the Yukawa coupling matrices

and allow for general flavour mixing with respect to the physical appearence of quarks and leptons as mass-eigenstates. It turns out, that for electrons, muons and τ-leptons, the flavour- and mass-eigenstates coincide (neglecting neutrino masses), such that in the unitary gauge

$$\begin{aligned}\sum_{i,j=1}^{3} y_{ij}^{l}(l_L^i)^\dagger \Phi e_R^j \ + \text{h.c.} = y_e \ (e_L^\dagger e_R + e_R^\dagger e_L)\frac{v+H}{\sqrt{2}} \\ + y_\mu(\mu_L^\dagger \mu_R + \mu_R^\dagger \mu_L)\frac{v+H}{\sqrt{2}} \\ + y_\tau(\tau_L^\dagger \tau_R + \tau_R^\dagger \tau_L)\frac{v+H}{\sqrt{2}}\,,\end{aligned} \tag{2.14}$$

and therefore the masses of leptons are given by

$$m_l = y_l \frac{v}{\sqrt{2}} \tag{2.15}$$

and accordingly, a coupling of the Higgs boson to leptons appears, which is linearly proportional to the mass. The Yukawa couplings, and therefore lepton masses, are free parameters of the theory and can only be determined experimentally. For quarks, the mass-eigenstates do not coincide with the flavour-eigenstate, meaning that flavour transitions $d_i \to W^- u_j$ may occur. Instead, the Yukawa matrices in equation 2.13 can be diagonalized via $m^d = V_L^d y_{ij}^d V_R^{d\dagger}$ and $m^u = V_L^u y_{ij}^u V_R^{u\dagger}$, with $V_L^u V_L^{d\dagger} = V_{CKM}$ being the Cabibbo-Kobayashi-Maskawa (CKM) matrix [25] [26]. After this change of basis from the weak-eigenstates into the mass-eigenstates the Yukawa terms in the unitary gauge are of the form

$$y_i^{u,d}\ (q_{iL}^{\dagger u,d} q_{iR}^{u,d} + q_R^{\dagger u,d} q_{iL}^{u,d})\frac{v+H}{\sqrt{2}} = (m_i^{u,d} \ + \frac{m_i^{u,d}}{v}H)(q_{iL}^{\dagger u,d} q_{iR}^{u,d} + q_R^{\dagger u,d} q_{iL}^{u,d}). \tag{2.16}$$

In summary, by introducing the complex scalar $SU(2)_L$ doublet, the Higgs field, with the potential written in 2.5, mass terms for all necessary elementary particles are therefore generated. In addition, a new physical scalar Higgs boson is predicted by the SM, with couplings to massive fermions and vector bosons particles (including a vertex factor $n!$ for n particles of same type):

$$g_{Hff} = \frac{m_f}{v} \quad f = (u,d,c,s,t,b,e,\mu,\tau), \qquad g_{HVV} = 2\frac{m_V^2}{v} \quad V = (W,Z) \tag{2.17}$$

Therefore the thoery predicts a linear relation between masses and couplings: $\sqrt{g_{HVV}/(2v)}, g_{Hff} = 1/v \cdot m_{f,V}$. The measurements of couplings of the new boson by the CMS Collaboration confirms this prediction of the Higgs mechanism within uncertainties, as shown in Figure 2.1. For the fit of the couplings, the PDG

values [27] for the top quark and the running mass for the b-quark ($m_b(m_H = 125.7\ \text{GeV}) = 2.763\text{GeV}$) is used, and affect the loop induced couplings, e.g. gluon-Higgs, as predicted by the SM (more details in [28] [15]).

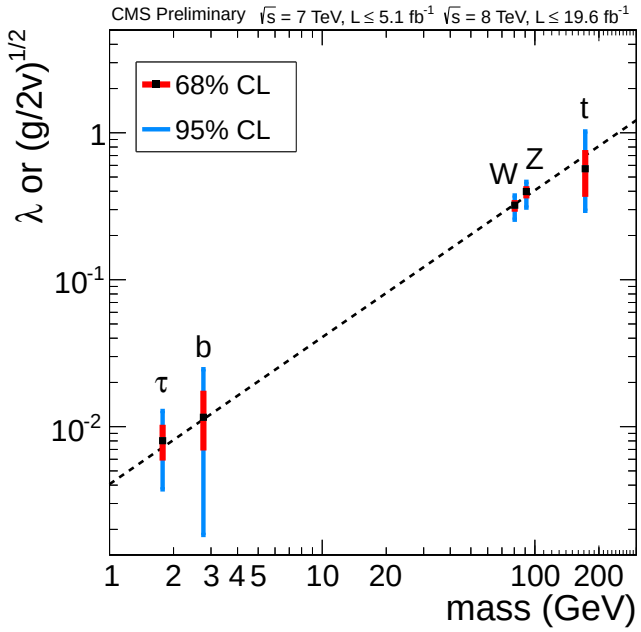

Figure 2.1.: CMS coupling fit results for the new boson to tau lepton, b-quarks, top-quarks and weak bosons. dotted line is SM prediction. [15]

2.2. Higgs Production at the LHC

The theoretical predictions for the properties of the Higgs boson are summarized in detail in [29] [30] [31], and citations therein.

The physics processes at the LHC are induced by the interaction of partons of the colliding protons. The densities of momentum fractions carried by individual partons are described by parton distribution functions, including CT10 [32], MSTW [33] and NNPDF [34]. Gluons and different types of quarks carry on average different fractions of the incoming protons, which are highly dependent on the energy scale (Q^2); as an example Figure 2.2 shows the the differental luminosity over the center of mass energy for partonic gluon-gluon, gluon-quark and quark-quark interactions for proton-proton collisions at a center of mass energy $\sqrt{s} = 14$ TeV [35]. Also shown in Figure 2.2, are comparisons of inclusive production cross-sections for SM physics processes of different type. As can be seen, the cross-section of Z-bosons exceeds the one for the Higgs boson by about five orders

of magnitude, leading to overwhelming partially irreducible background for Higgs searches.

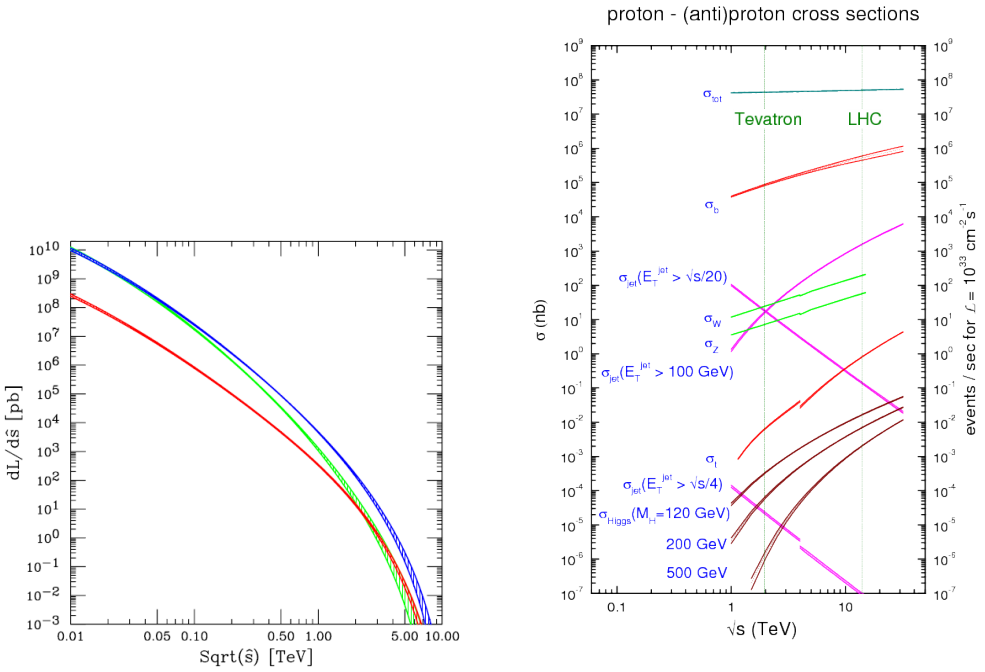

Figure 2.2.: Left: The differential parton-parton luminosity based on MSTW pdfs in picobarns over $\sqrt{\hat{s}} = \sqrt{x_1 \cdot x_2 \cdot s(14\ \text{TeV})}$. Green = gg, Blue = $(gq + g\bar{q} + qg + \bar{q}g)$, Red = $(q\bar{q} + \bar{q}q)$, where the sum runs over the five quark flavours d, u, s, c, b [35]. Right: Comparison of several inclusive hadronic production cross-sections, including Higgs, weak bosons and top- and bottom-quark production processes [36].

At the LHC, the SM predicts 4 major production mechanism to contribute to the overall cross-section of the Higgs boson: gluon-fusion, vector boson fusion, vector boson associated production via Bremsstrahlung and top-quark associated production. Examples of the corresponding Feynman diagrams are shown in Figure 2.4. The corresponding cross-sections as predicted by the SM as a function of the mass of the Higgs boson, as well as with fixed $m_H = 125$ GeV/c^2 and as a funtion of the center-of-mass energy, is shown in Figure 2.3. Higgs boson production rates via gluon fusion exceed the other production mechanisms by at least one order of magnitude. Table 2.2 summarizes the Higgs boson production cross-sections for the gluon fusion, vector boson fusion and W/Z and top-quark associated production mechanisms at the LHC for Higgs masses from 90-140 GeV/c^2.

Gluon Fusion The Higgs boson production via gluon-gluon fusion is mediated by triangular loops of heavy quarks. Due to the enhanced partonic gluon luminosity at high energies (see Figure 2.2) and the large couplings to especially top-quarks, the production mechanism has the largest contribution to the overall Higgs production cross-section. The cross section is highly dependent on higher order QCD corrections. Up to next-to-leading order (NLO) in α_s, the cross section prediction has been calculated maintaining the full dependency on the Higgs and top- and bottom-quark masses. The NLO corrections increase the leading order prediction by about 80%. Additionally, the NNLO contributions have been calculated in the approximation of the large top-mass limit ($m_t \to \infty$), increasing the cross section by about 20%. NLO electroweak contributions have been computed to further enhance the cross section by about 5%. Matching of the NNLO hard scattering matrix element to the soft-gluon contributions at next-to-next-to-leading logarithm (NNLL) accuracy have been achieved.

Vector Boson Fusion (VBF) The VBF production of the Higgs boson has the second largest contribution to the overall cross-section of the Higgs boson at the LHC, as shown in Figure 2.3. The process is induced by $t-$ and $u-$channel (Mandelstam variables, s-channel can be negelcted) Z or W boson exchange interaction of two (anti-)quarks of the incoming protons, with the Higgs boson being radiated off the weak bosons, linking the two quarks. VBF production via W bosons exceeds the one with Z bosons by a factor of about 3, due to the larger W coupling to fermions. The VBF production mechanism leads to a distinct final state topology with two jets in the final state. Due two the absence of color exchange at leading order between the two incoming protons (weak bosons are color singlets), additional gluon radiation in the central rapidity region between the two final-state quarks is strongly surpressed, and the quarks color connect with the respective proton remnant. This leads to two high energetic jets at, on average, large pseudo-rapidities, with large pseudo-rapidity gaps and without additional high energetic jet activity in between, which is a distinguished feature in comparison to background processes and the other Higgs boson production mechanisms. Further, Higgs boson production via VBF is dependent on the weak vector boson masses, and thus enhanced with respect to the weak Z boson production via VBF, which is suppressed by $(g_2)^4$. These features of the VBF Higgs production can be exploited independent of the Higgs decay mode, and provide complementary possibilities to study the bosonic couplings of the Higgs boson, especially in the $H \to \tau\tau$ channel. VBF cross section predictions have been calculated including NNLO QCD and NLO electroweak contributions of about 5-10%, respectively.

Z/W+H and tt+H The Higgs boson production via Bremstrahlung off weak vector bosons and heavy quark associated Higgs production processes, are addi-

tional important LHC Higgs production mechanism. The cross sections at the LHC, shown in Figure 2.3, are somewhat low compared to the gluon-fusion and VBF production cross sections. In the $H \rightarrow \tau\tau \rightarrow 2e4\nu$ analysis, these processes are inclusively taken into account, however due to the small corresponding production cross sections no dedicated strategy is pursued to exploit the distinct production signatures. But in general, the vector boson associated production offers the most sensetive possibilty to measure Higgs boson decays into bottom quarks, where final state leptons from vector boson decays can be used to trigger events and surpress the otherwise overwhelming QCD background. Also for the $H \rightarrow \tau\tau$ channel, 12 dedicated, but overall less sensitive, CMS analyses independent from the inclusive channels, aim to exploit the associated production mechanism. All theses analyses are combined in Chapter 6. Top-pair associated Higgs boson production, provides direct sensitivity to the top-quark Yukawa coupling of the Higgs boson, and the CMS and ATLAS collaborations pursued dedicated analysis [37] [38] [39] [40] to study this production mode for several Higgs boson decay modes. For Z/W associated Higgs production the cross section was calculated with NNLO QCD accuracy, increasing (mostly NLO corrections of Drell-Yan type) the LO cross section by about 30%. Electroweak NLO corrections have been calculated and enhance the cross-section by 5-10%. The cross section for ttH production is only known up to NLO QCD accuracy, increasing the LO cross section at the LHC by about 20%.

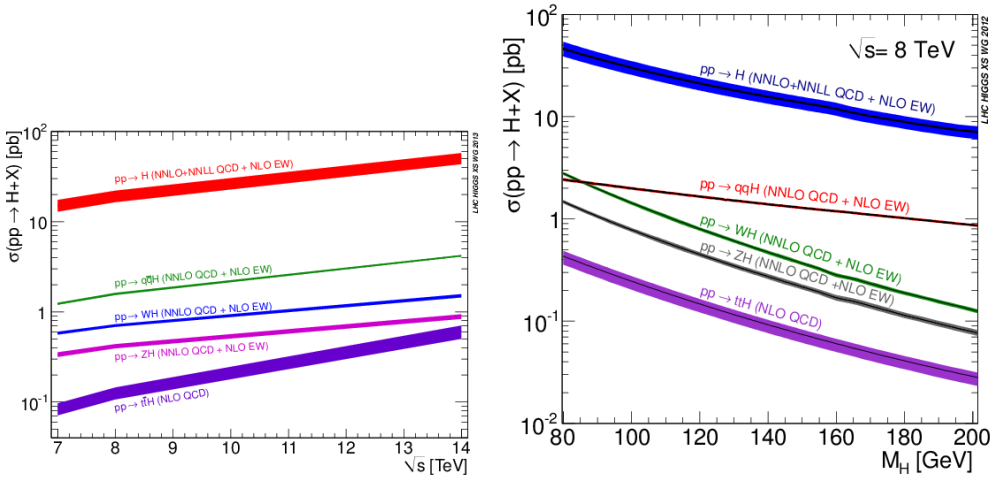

Figure 2.3.: Left: Higgs production cross sections at the LHC including NNLO QCD and NLO electroweak higher order corrections as a function of center of mass energy from 7-8 TeV for $m_H = 125$ GeV/c^2. Right: Higgs production cross sections at the LHC including NNLO QCD and NLO electroweak higher order corrections as a function of m_H for center of mass energy of 8 TeV.

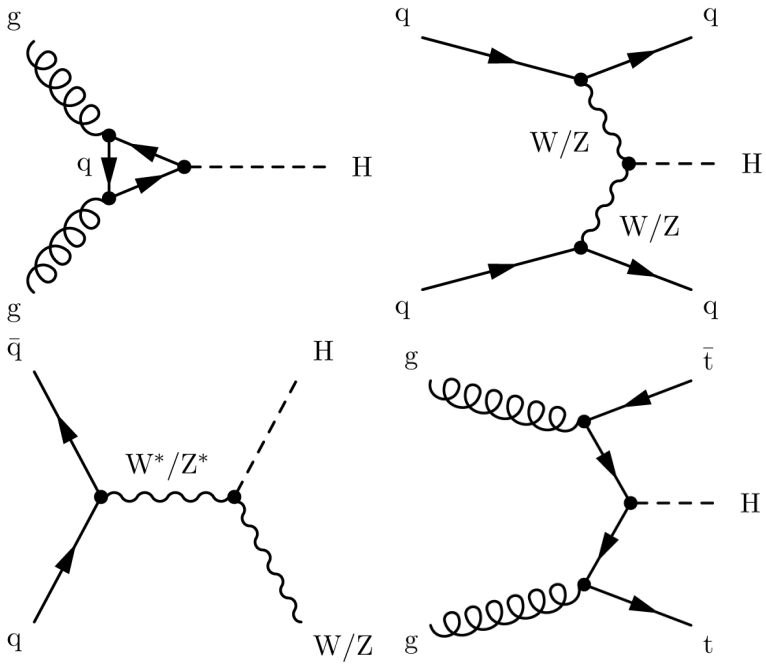

Figure 2.4.: Examples of Feynman diagrams for the LHC Higgs production mechanisms. Gluon-fusion (top-left), vector boson fusion (top-right), W/Z associated production (bottom-left) and top-quark associated production (bottom-right). [41]

m_H[GeV/c^2]	$\sigma(ggH)[pb]$	$\sigma(qqH)[pb]$	$\sigma(WH)[pb]$	$\sigma(ZH)[pb]$	$\sigma(t\bar{t}H)[pb]$
90	36.23	2.191	1.990	1.092	0.3202
95	32.69	2.084	1.695	0.9383	0.2786
100	29.68	1.988	1.447	0.8102	0.2433
105	27.01	1.897	1.242	0.7022	0.2133
110	24.70	1.809	1.071	0.6125	0.1871
115	22.66	1.729	0.9266	0.5358	0.1651
120	20.86	1.649	0.8052	0.4710	0.1459
125	19.27	1.578	0.7046	0.4153	0.1293
130	17.85	1.511	0.6169	0.3671	0.1149
135	16.57	1.448	0.5416	0.3259	0.1024
140	15.42	1.389	0.4768	0.2898	0.09150

Table 2.2.: Cross sections for the LHC Higgs production mechanisms at 8 TeV for m_H ranging from 90 to 140 GeV/c^2. Taken from [31].

2.3. Higgs Decay into Tau-Leptons

The partial decay widths are, once the Higgs boson mass is known, predicted by the SM, considering the couplings shown in equation 2.17. The couplings are proprtional to the mass, and therefore the Higgs boson has the tendency to primarly decay into the heavier bosons and fermions, dependent on the Higgs mass and phase space. In addition, decays into massless photons or gluons are possible mediated by loop diagrams, similiar to the production via gluon fusion. For the decay into fermions the partial decay width in Born approximation (leading order) is given by

$$\Gamma_{\text{Born}}(H \to f\bar{f}) = \frac{G_F N_c}{4\sqrt{2\pi}} m_H m_f^2 \beta_f^3 \,, \tag{2.18}$$

where $\beta = (1 - 4m_f^2/m_H^2)^{\frac{1}{2}}$ is the velocity of the final state fermions and N_c the color factor, equal to 1 for leptons and 3 for quarks. QCD correction up to third order (NNNLO) have been calculated and result in an uncertainty due to scale dependency of less than 0.2% for charm- and bottom-quarks and are negligable for decays into leptons, whereas electroweak corrections up to NLO result in an uncertainty of $1 - 2\%$. Partial decay widths into bosons are discussed in detail in references [29] [30] [31]. Figure 2.5 shows the branching ratios

$$BR(H \to P_i P_j) = \frac{\Gamma(H \to P_i P_j)}{\sum_{i,j} \Gamma(P_i P_j)}, \tag{2.19}$$

where P_j is any SM fermion or boson, as a function of the Higgs boson mass for all fermions and bosons with a considerable Higgs decay width fraction. For $m_H = 125$ GeV/c^2, the branching ratio into tau-leptons is the third largest, after the Higgs decay into W bosons and b-quarks. Towards lower masses, the decay into tau-leptons becomes the second most prominent after the Higgs decay into b-quarks. In particular, due to the large irreducible QCD background for $H \to b\bar{b}$, the decay into tau-leptons is the most sensitive decay mode to study and measure fermionic Higgs decays and couplings, compared to the decay into muons with a decreased branching ratio by 2 orders of magnitude.

The tau-lepton decays leptonically via weak interactions into muons or electrons or hadronically into pions and kaons. Reference [27] compiles the branching ratios of the tau-lepton in detail (measured at the BaBar and Belle experiments), summarized in Table 2.3. Hadronic tau-lepton decays make up about 64.76% of tau decays and are dominated by 1-prong decays $\tau^- \to \pi^-\nu_\tau$ and $\tau^- \to \pi^-\pi^0\nu_\tau$ and 3-prong decays $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$. Leptonic decays into muons and electrons have branching ratios of 17.83% and 17.41% respectively.

Consequently, the search for $H \to \tau\tau$ decays is split into 6 sub-channels with corresponding branch fractions (neutrinos omitted): fully leptonic channels $ee(BR = 3.18\%)$, $\mu\mu(BR = 3.03\%)$, $e\mu(BR = 6.21\%)$, and channels involving hadronically

tau decay mode	Br[%]
$\tau^- \to e^- \bar{\nu}_e \nu_\tau$	17.83
$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$	17.41
$\tau^- \to \nu_\tau + \text{hadrons}(\pi^{0,-,+}, K^{0,-,+})$	64.76

Table 2.3.: Branching ratios of leptonic or hadronic tau decays [27]. hadrons$(\pi^{0,-,+}, K^{0,-,+})$ is dominated by $\tau^- \to \pi^- \nu_\tau$, $\tau^- \to \pi^- \pi^0 \nu_\tau$, $\tau^- \to \pi^- \pi^0 \pi^0 \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$.

decaying tau-leptons $\tau_h\tau_h(BR = 41.94\%)$, $\mu\tau_h(BR = 22.55\%)$, $e\tau_h(BR = 23.09\%)$. The ee and $\mu\mu$ channels therefore have the smallest $H \to \tau\tau$ branching ratios.

In addition to $Z \to \tau\tau$ background, the same-flavor lepton channels have additional overwhelming $Z/\gamma* \to ee/\mu\mu$ background. This constitutes the challenging basis for the analysis presented in this thesis.

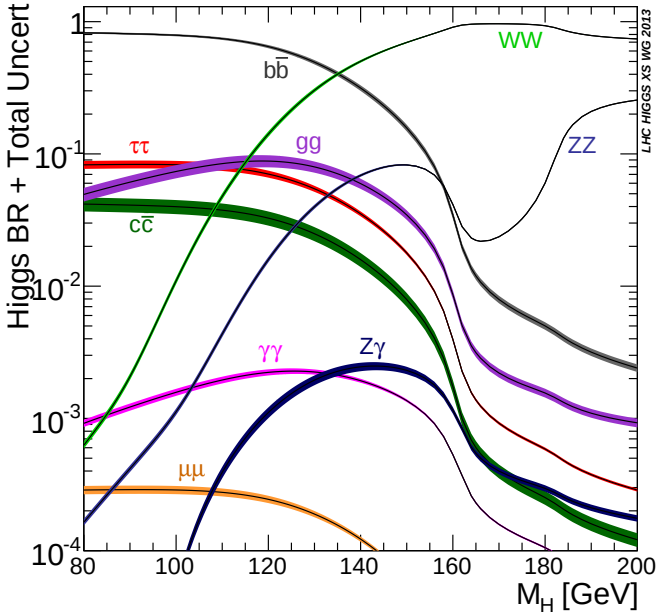

Figure 2.5.: Comparison of branching ratios and their uncertainties for all dominant Higgs boson decays as predicted by the SM as a function of m_H. [41]

m_H[GeV/c^2]	$H \to \tau\tau$ branching ratio
90	8.33×10^{-2}
95	8.32×10^{-2}
100	8.28×10^{-2}
105	8.17×10^{-2}
110	7.95×10^{-2}
115	7.89×10^{-2}
120	7.04×10^{-2}
125	6.32×10^{-2}
130	5.45×10^{-2}
135	4.49×10^{-2}
140	3.52×10^{-2}

Table 2.4.: $H \to \tau\tau$ branching ratios for m_H 90-140 GeV/c^2. Taken from [42].

3. Experimental Apparatus

This chapter describes the experimental set up to test the Higgs boson hypothesis at the Large Hadron Collider. Section 3.1 gives a short overview of the Large Hadron Collider and the proton beam configurations in 2011 and 2012 data-taking periods. The CMS Detector is described in detail in Section 3.2, and specific technical information for each CMS subdetector and general data acquisition is given in the subsections thereof.

3.1. The Large Hadron Collider

The Large Hadron Collider (LHC) is a circular proton-proton collider 26.7 km in circumference located at the Swiss-French border operated by the European Organization for Nuclear Research (CERN). Based on the synchrotron principle, it is designed to accelerate two beams of protons (ions) in opposite directions to be collided at a center-of-mass energy of up to $\sqrt{s} = 14$ TeV. Overall 1232 superconducting NbTi dipole magnets operated at 1.6 K and 8.3 Tesla are deployed around the two vacuum filled beam-pipes in the accelerator ring assembled into common cyrostatic modules filled with superfluid helium. Further superconducting multipoles are implemented around the ring to squeeze the beam of otherwise diverging protons. The two proton beams are accelerated in roughly 15 min to their peak energy via 16 niobium filmed copper radio frequency cavities cooled down to 4.5 K increasing the proton energy by about 16 MeV per turn. The two beam-pipes are separated except for smaller parts at the 4 interaction points along the ring where the two beams are brought to collisions to be studied by the 4 detectors ATLAS (A Toroidal LHC ApparatuS) [43], ALICE (A Large Ion Collider Experiment) [44], CMS (Compact Muon Solenoid) [45] [46] and LHCb [47].

Protons are first accelerated to 450 GeV/c in the pre-accelerator facilities at CERN including the Linac2 linear accelerator, Proton Synchrotron Booster (PSB) Proton Synchrotron (PS) and Super Proton Synchrotron (SPS) before being injected into the LHC in the form of discrete bunches of protons with equidistant spacing. By design, the bunch spacing is 25 ns for 2808 bunches a $\approx 10^{11}$ protons with a normalized transverse emittance of ϵ_n =3.75 μm and amplitude function β^* =0.55 m at the collision point.

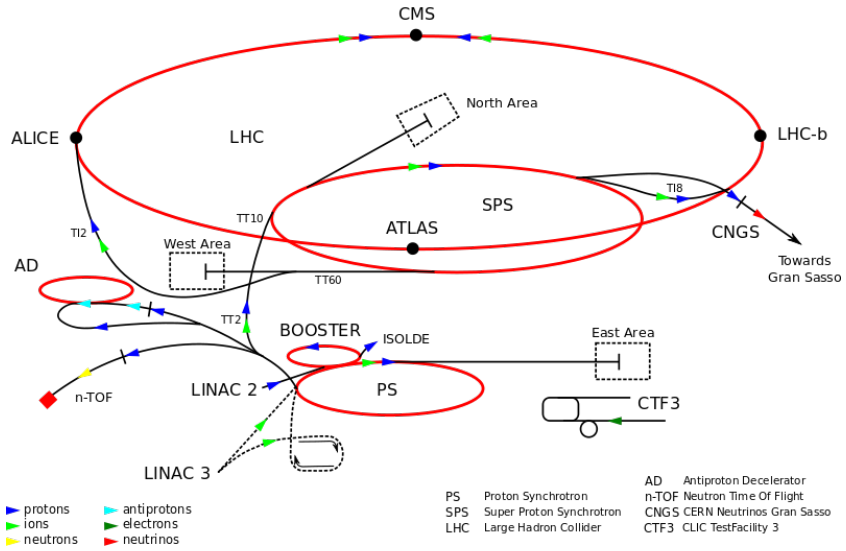

Figure 3.1.: Illustration of the CERN accelerator complex. The LHC is shown together with its pre-accelerator system consisting of Linac2, PSB, PS and SPS [48]

The instantaneous luminosity can be calculated via

$$\mathcal{L} = \frac{\gamma f k_B N_p^2}{4\pi\epsilon_n \beta*} F,$$

where $\gamma = E_{beam}/(m_{proton}c^2)$ is the relativistic gamma factor, k_B denotes the number of bunches per beam, f is the bunch-frequency, N_p is the number of protons per bunch and F is a factor accounting for the intersecting angle of the two beams. With the design beam-parameters of the LHC the instantaneous luminosity is $\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$ with on average 20 proton-proton interactions per bunch-crossing and a total interaction rate of 10^9 per second.

In 2011 and 2012 the LHC was running with a center-of-mass energy of 7 and 8 TeV respectively, and the bunch spacing was fixed to 50 ns with a maximum of 1380 bunches and 10^{11} protons per bunch. The beam parameters ϵ_n and β^* were gradually optimized to reach a peak luminosity of $\mathcal{L} = 3.7 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$ in 2011 and eventually $\mathcal{L} = 7.7 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$ in 2012. Figure 3.2 shows the cumulative distributions of integrated luminosity versus time for 2011 and 2012 delivered by the LHC accelerator and recorded by the CMS experiment.

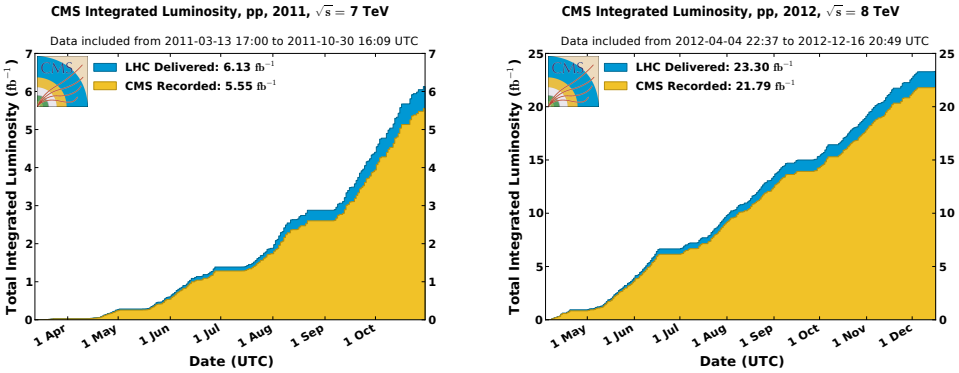

Figure 3.2.: Accumulative plots of integrated luminosity recorded by the CMS Experiment in 2011 (left) and 2012 (right) [49]

3.2. The CMS Detector

The CMS experiment [45] [46] is one of the two multipurpose detectors at the LHC and the instrument used to record proton-proton collisions for this analysis. It is designed to elucidate and scrutinize physics processes at the TeV scale and thus amongst others the mechanism of electroweak symmetry breaking. As a consequence, the design principles of the CMS detector are driven by the high luminosity environment provided by the LHC characterized by on average 20 proton-proton interactions at a center-of-mass energy of 14 TeV every 25 ns and the associated high radiation levels. Further, excellent momentum and energy resolution as well as highest particle identification efficiencies for photons, electrons, muons, taus, several jet-types and missing transverse energy up to the TeV scale are mandatory for the physics program to be realized.

Overall, the CMS detector is about 28.7 m long and has a diameter of 15.0 m with a weight of about 14000 tonnes. To keep the bias of measurements due to material budget at a minimum level, measurements and the associated subdetectors are organized radially around the beam axis from the inside out as a function of their sensitivity to biases induced by the material budget placed in front. For superb charged-particle momentum resolution and detection of secondary vertices and thus identification of taus and b-jets, the (pixel) tracking system is placed in the center of the detector. The tracker is surrounded by the high granular electromagnetic calorimeter to absorb and measure the energy of electrons and photons with high resolution and good π^0 rejection to be followed by the hadronic calorimeter with large hermetic coverage for precise measurements of missing transverse energy and jet energies. The solenoid is placed next to produce the magnetic field of 3.8 T without causing further “dead“ material in front of the calorimeters however with the caveat of only limited space to be occupied by the calorimeters. The

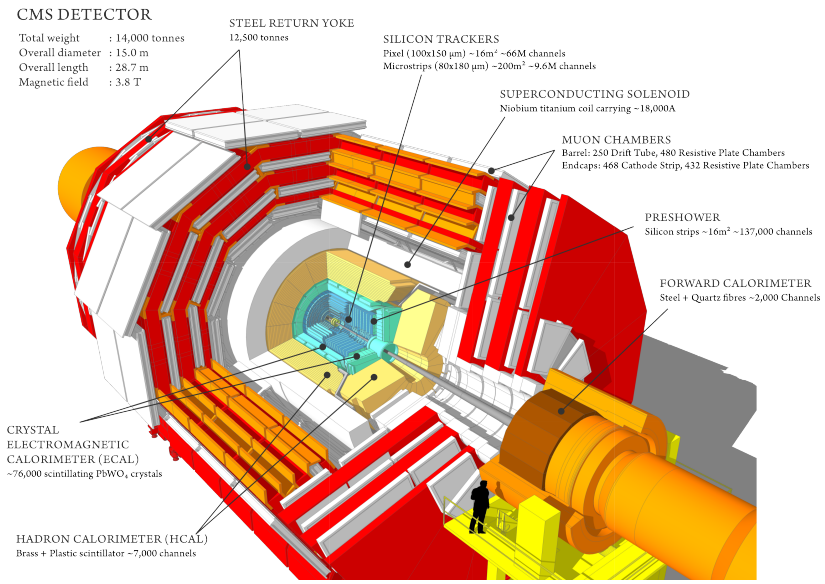

Figure 3.3.: Illustration of the CMS detector with labeled subsystems. [50]

outermost layer consists of the return-yoke for the magnetic field interweaved with the muon system for the identification of muons with long radiation length and to provide improved energy resolution for TeV muons complementing the muon track measurements. To cope with the high frequency of 40 MHz for the bunch-crossing rate, CMS deploys a distinct two-level selective event triggering system to reduce the number of events over time to about 100 Hz to be considered for further computing intense processing and reconstruction.

The CMS detector is located at point 5 diametral to the CERN site in a 100 m deep cavern. The coordinate system for the detector has its origin at the nominal vertex with the x-axis in the horizontal plane radially inwards with respect to the LHC ring, the y-axis pointing vertically upwards and the z-axis pointing along the beamline in the direction of the Jura Range. In polar coordinates the azimuthal angle ϕ in the x-y plane goes counter-clockwise with respect to the positive x-axis and θ also counter-clockwise with respect to the positive z-axis. Conveniently instead of using the coordinate θ, the pseudorapidity defined as:

$$\eta = -\log\left(\tan\left(\frac{\theta}{2}\right)\right)$$

is used with the property of being Lorentz invariant and that the particle flux at hadron colliders is approximately constant as a function of η.

3.2.1. Tracking Detectors

The CMS tracking system [51] [52] [45] is the most central part of the detector and is entirely based on silicon tracking technology. Its purpose is to measure the trajectory of charged particles crucial for precise momentum reconstruction and track extrapolations to the calorimeter or primary vertices towards the beamline. Close to the interaction vertex with high occupancy the tracker consists of three silicon pixel layers providing precise position measurements in three spatial dimensions. The three layers of the pixel detector are located at radii 4.4 cm, 7.3 cm and 10.2 cm extending to $z = \pm 26.5$ cm parallel to the beam-axis. Additionally, two pixel layers perpendicular to the beam-axis are deployed in both forward detector regions at about $z = \pm 34.5$ and ± 46.5 cm extending to radii of about 6 cm and 15 cm , respectively. The pixel detector consists of 66 million pixel elements with the size of $100 \times 150 \mu\text{m}^2$ across 1440 modules and attains a spatial resolution of 10 μm in the r-ϕ plane and 20 μm in z direction. Improved spatial resolution with respect to the size of the pixels is achieved by taking into account secondary adjacent pixel hits due to charge sharing induced by lorentz drift.

The outer parts of the detector, shown in Figure 3.4 with less occupancy consist of silicon strip detectors which provide precise spatial measurements only in the r-ϕ plane and slightly less precise resolution in z direction. In the barrel, the silicon strip detector consists of two parts. The Tracker Inner Barrel (TIB) with four layers placed at radii between 25.5 and 49.8 cm from the beam line extending to $|z| < 70$ cm and the Tracker Outer Barrel (TOB) with additional 6 layers placed at radii ranging from 60.8 to 108.0 cm with a total length of 2.4 m contributing to the track measurement up to $|\eta| < 1.3$. The first two layers of the TIB and TOB are stereo layers to overlap in the z-ϕ plane with an angle of 110 mrad with the other associated layers to improve the resolution in z direction. In the r-ϕ plane a spatial resolution of 16-28 μm in the TIB and 25-41 μm in the TOB is achieved. Due to the strip design the resolution in z direction is 230 μm and 530 μm for the TIB and TOB respectively.

In the endcaps, the outer tracking system is divided into the Tracker Inner Disk (TID) and Tracker End Cap (TEC). The TEC consists of 9 layers on each side perpendicular to the beamline positioned between 124 and 280 cm away from the nominal vertex. Each disk (layer) is segmented into at most 7 rings of different radii and 8 sectors in ϕ. The first two rings and in addition the fifth ring contain stereo modules for improved spatial resolution Furthermore, the Tracker Inner Disk (TID) is installed to fill the gap between the TIB and TEC consisting of three layers on each side positioned between 80 and 90 cm away from the nominal vertex. Overall the silicon strip detector consists of 15148 modules with about 9.3 million strips of semi-conductor p-n junctions with a strip pitch width between 80 μm and 120 μm in the TIB and between 120 μm and 180 μm in the rest of the detectors.

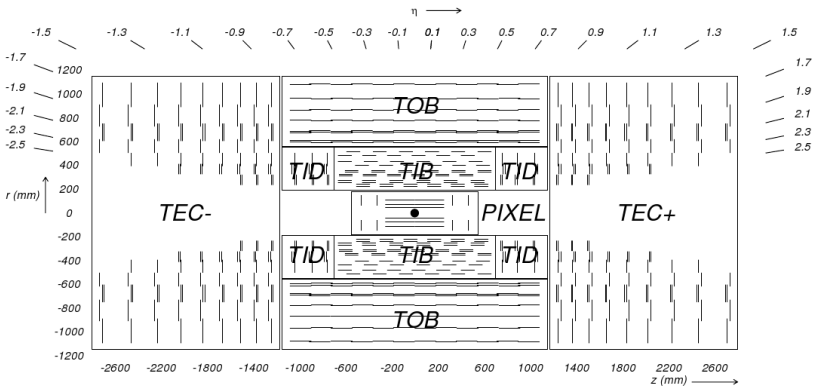

Figure 3.4.: Illustration of the CMS tracker with the individual subdetectors labeled. Stereo modules are shown as double lines. [53]

The CMS tracker covers a region of up to $|\eta| < 2.5$ and the thickness of the silicon sensors vary from 285 μm in the pixel detector up to 500 μm in the TEC and TOB. It is cooled to operate at $-10\,^{\circ}\mathrm{C}$ and together with the sophisticated readout electronics, cabling and cooling system the tracker has a considerable amount of material budget, shown in Figure 3.5, varying from 0.4 radiation lengths to a maximum of 1.8 radiation lengths at the barrel-endcap transition region $|\eta| \approx 1.5$, which leads to significant production of bremsstrahlung from electrons and photon conversions.

The electronic signal generated by electron-hole pairs produced by particles interacting with the semiconductor p-n junctions operated in reverse bias are amplified and considered for further signal processing and reconstruction.

3.2.2. Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) [54] is in combination with the tracking detector the key subdetector to identify electrons and photons and facilitates high resolution energy measurements. The CMS ECAL is a highly homogeneous and hermetic calorimeter composed out of 75848 radiation resistant lead tungstate ($PbWO_4$) crystals of truncated pyramidal shape. The high density of the crystals and small Moliere radius allow for a compact design of the ECAL with about 25 interaction lengths and enables a fast calorimeter response with 80% of light being collected in 25 ns.

The ECAL barrel (EB) covers the pseudorapidity range $|\eta| < 1.479$ and accomodates 61200 crystals to achieve a granularity of about $(\Delta\phi, \Delta\eta) \approx (0.0174, 0.0174)$ (i.e. 360-fold in ϕ and 190-fold in η). With a crystal length of 23 cm, this corresponds to a cross-section of 22 $\times$ 22 mm^2 at the front face and 26 $\times$ 26 mm^2

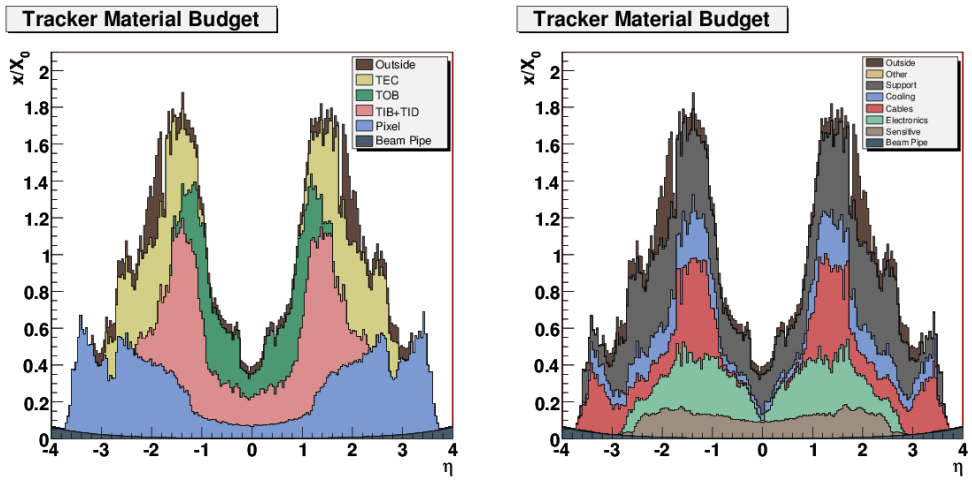

Figure 3.5.: Material budget of the CMS tracking system as function of η separated into tracker sub-detector (left) and material (right). [53]

at the rear face of each crystal. The crystals are mounted in a quasi-projective geometry with respect to the vector from the nominal vertex with a distance of 1.29 m from the beamline to the center of the face of each crystal. Grouped into 5×2 matrices, 40 to 50 matrices are collected into one module, and 4 modules are then assembled into 1 of the overall 36 supermodules in the barrel, as illustrated in Figure 3.6. Scintillation photons emitted in electromagnetic showers are collected in magnetic-field-resistant silicon avalanche photodiodes (APDs). The light yield of the lead tungstate is rather low, about 4.5 photo electrons per MeV, and both the light yield of the crystal as well as the APD amplification are sensitive to temperature variations, estimated to be about -2.1%°C^{-1} at nominal operating temperature of (18 ± 0.05)°C.

The ECAL endcaps (EE) cover the range $1.479 < |\eta| < 3.0$ and consist of 2×3662 crystals per endcap, mounted in two disk halves (Dees). The crystals have a length of 22 cm with a front face cross section of 28.6 × 28.6 mm^2 and a rear face cross section of 30 × 30 mm^2. Photoelectrons are collected via vacuum phototriodes, suitable for the high radiation environment and magnetic field orientation in the endcap regions. For improved π^0/photon discrimination, a preshower detector (PS) covering the pseudo-rapidity range $1.65 < |\eta| < 2.61$, is placed in front of the EE, adding 3 radiation lengths. The PS is a 20 cm thick sampling calorimeter, composed out two alternating layers of lead and silicon strip sensors of orthogonal orientation. The pitch width of one strip is 1.9 mm and thus gives significantly enhanced spatial resolution.

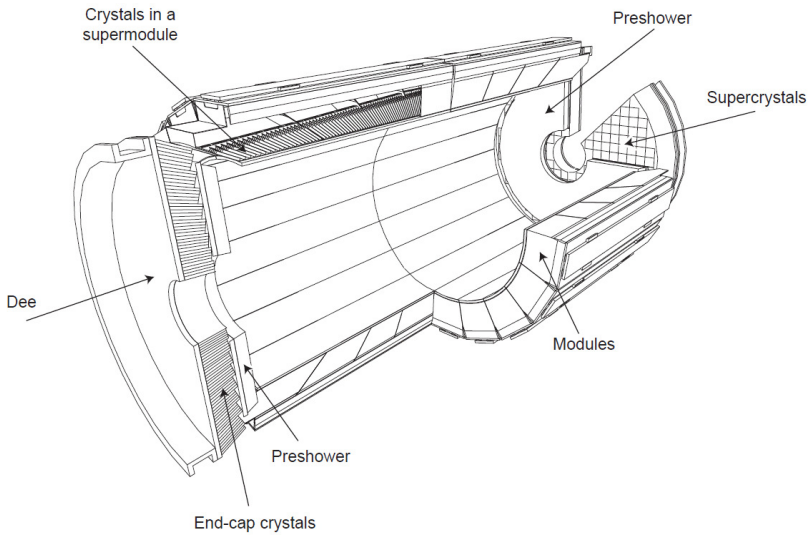

Figure 3.6.: Illustration of the CMS tracker with the individual subdetectors labeled. [53]

3.2.3. Hadronic Calorimeter

The Hadronic Calorimeter (HCAL) [55] of the CMS detector is used to measure the energy of charged and neutral hadrons, crucial for the reconstruction of jets and missing transverse energy. Especially for the measurement of vector boson fusion processes with two forward jets in the final state, a large hermetic coverage is necessary. The HCAL consists of three sub-detectors, the HCAL barrel (HB) covering the central region up to $|\eta| < 1.3$, the HCAL endcaps (HE) covering $1.3 < |\eta| < 3.0$, and the HCAL forward detector (HF) covering the the region $3.0 < |\eta| < 5.2$. Due to the radial constraints imposed by the solenoid, an additional HCAL Outer (HO) subdetector element is placed just behind the magnet to complement the energy measurement for hadron showers extending beyond the the HB. The effective thickness of the HCAL ranges between roughly 6 hadron interaction lengths to roughly 11 interaction lengths for $|\eta| > 1.3$ and the magnet adds about 1 interaction length in front of the HO.

Based on a sampling calorimeter design, the HB and HE consist of brass hadron absorber layers interleaved with thin plastic active scintillator layers with a thickness of a few millimeters. The HB has a segmentation of $(\Delta\phi, \Delta\eta) = (0.087, 0.087)$ and 16 layers of brass with a thickness ranging from about 51 cm to 57 cm radially stacked from 1.77 m to 2.95 m with respect to the beamline. In the endcap the HE has a segmentation of $(\Delta\phi, \Delta\eta) = (0.087, 0.087)$ for $|\eta| < 1.6$ and $(\Delta\phi, \Delta\eta) \approx (0.17, 0.17)$ for $|\eta| \geq 1.6$ with about 12-17 longitudinal brass layers of

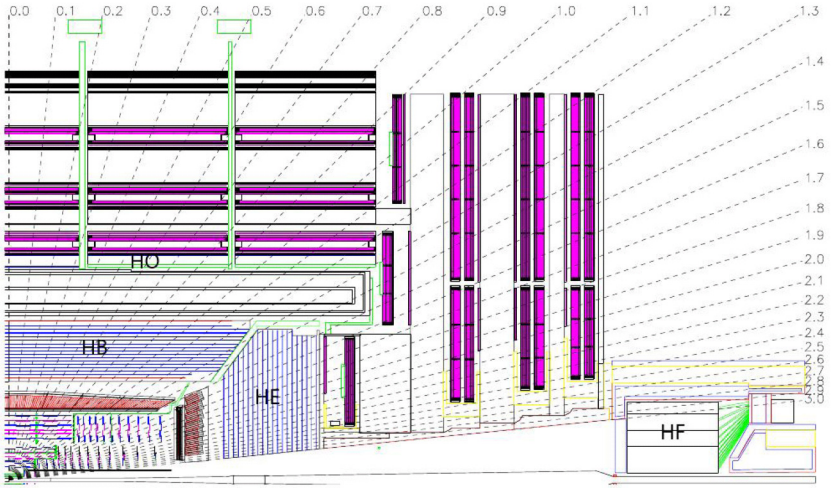

Figure 3.7.: Illustration of the CMS detector quadrant in the r-z plane with labeled HCAL subdetectors and η coverage. [53]

about 8 cm thickness. Light from hadronic showers is collected in each scintillator layer and is guided through optical fibres to hybrid photodiodes (HPDs) where light is proportionally transduced into electronic signal. Due to low signal-to-noise performance of HPDs, in depth read-out of each calorimeter tower is highly limited. In the HB the signal in each tower is added up and not segmented in depth. In the HE the longitudinal segmentation ranges from 2 to 3 segments.

Due to the harsh radiation environment of about 10 MGy in the forward regions of the detector, the HF uses more robust quartz fibres as active material and steel as absorber material. The HF is placed at $|z| = 11.2$ m with a diameter of 2.6 m and is functionally longitudinally subdivided with half of the fibres running over the full length of the detector and the other half only starting at a depth of 22 cm from the front of the detector. This establishes partial separation of electrons and photons predominantly absorbed in the front of the detector from pure hadronic showers mostly extending over both calorimeter segments. The Cerenkov light emitted in the quartz fibres is then channeled into photomultiplier tubes, bundled with a segmentation of $(\Delta\phi, \Delta\eta) \approx (0.17, 0.17)$.

3.2.4. Solenoid Magnet

The CMS Solenoid [56] is a core part of the detector design and facilitates the measurement of momenta of charged particles, especially muons. The superconducting magnet surrounds the tracker and calorimetric system of CMS and is designed to operate at 4 T within a bore of 12.5 m length and 6 m diameter. Four layers of coiled superconducting NbTi are cooled to 4.6 K using liquid helium to conduct a

nominal current of 19.41 kA.

The solenoid is complemented with a 10000 t iron return yoke consisting of 5 longitudinal wheels in the barrel and 2 endcaps, each composed out of 3 disks. In the barrel each wheel consists of 3 layers of 30, 63 and 63 cm thickness and the 3 disks in the endcap have a thickness of 25, 60 and 60 cm. The return yoke incorporates the muon system and guides the magnetic field outside the bore of the magnet, to enhance the resolution of muon momenta measurements.

3.2.5. Muon Detectors

The detection of muons is not mandatory for the analysis developed in this thesis, however constitutes an important part of the global event description. The CMS muon system [57] illustrated in Figure 3.8 covers the region up to $|\eta| < 2.4$ and can be divided into three subsystems. In the barrel region $|\eta| < 1.2$ where background radiation is small and the magnetic field is mainly uniform and in the return-yoke, Drift Tube chambers (DT) are installed consisting of 4 stations of different radii inside and around the return-yoke of the magnet across 5 rings longitudinally around the inner parts of the detector. The DTs are complemented with Cathode Strip Chambers (CSC) in the endcap covering the region $0.9 < |\eta| < 2.4$ where the muon and background(neutron) rate is higher and the magnetic field is non-uniform. The DT and CSC are used for precise spatial and thus p_T measurements with a resolution of 200 μm and 1 mrad but have the caveat of imprecise timing reconstruction with respect to the beam-crossing time. Therefore, in the region $|\eta| < 1.6$ additional fast responding Resistive Plate Chamber (RPC) systems are installed with enhanced timing resolution and improved capabilities for triggering for design luminosity.

3.2.6. Data Acquisition and Trigger System

The CMS Data Acquisition and Trigger System (TriDAQ) [53] is designed and used to analyze and read-out the detector information at a LHC bunch-crossing rate of 40 MHz (25 ns bunch-spacing) with 20 collisions per bunch-crossing, for the design luminosity of $\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$. At design luminosity, the initial data flow is larger than 1TByte/s, and therefore a dedicated Trigger System is deployed to first reduce the rate for event processing to $O(100\text{kHz})$ with the Level-1 Trigger (L1), and an additional reduction by a factor of 1000 is realized with the subsequent High Level Trigger (HLT). The data flow for offline event reconstruction and final analysis is then of the order of a few 100 Hz, and well managable. The L1 is installed within a network of custome build hardware and firmware such as Field Progammable Gate Arrays (FPGAs) and Application Specific Integrated Circuits (ASICs) located in the sub-detectors front-end electronics nodes of specific trigger electronic devices. The complete electronic detector output of 55 million channels

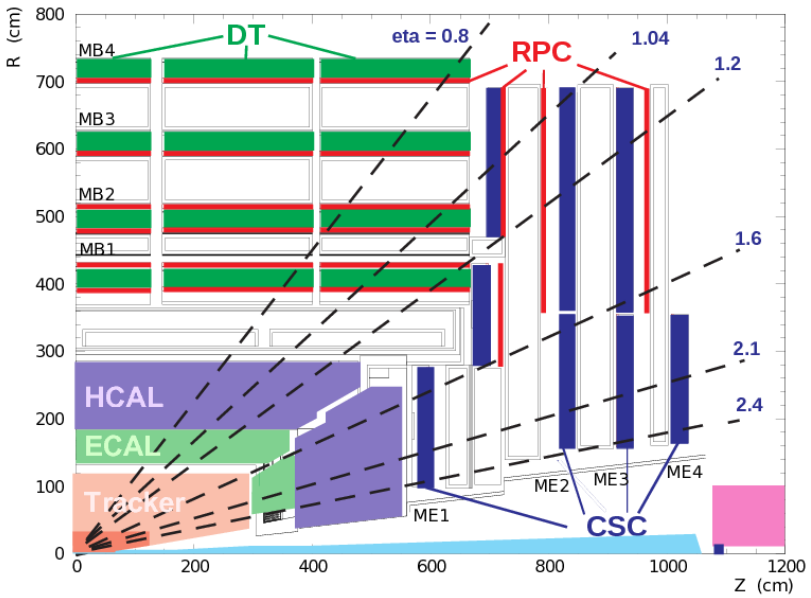

Figure 3.8.: Illustration of the CMS muon system quadrant in the r-z plane with the individual subdetectors labeled. [53]

at 40 mHz is pipelined and feedbacked within the L1 system and read-out buffers with a latency of 3.5 μs, to decide if the event can be discarded or kept for further processing.

The decision of the L1 trigger is based on the existence of specific physics object primitives build from fast responding ECAL and HCAL sub-detector outputs and the Muon system response. Global detector signatures such as higher energetic electromagnetic showers, hits in the muon chambers, missing transverse momentum or overall transverse hadronic energy serve as building block of trigger primitives. For electron/photon candidates, ECAL trigger towers consist of 25 crystals, and two adjacent ECAL trigger towers are summed to test L1 transverse energy thresholds. Additionally, a fine-grained veto bit is installed to measure the ratio of the highest energetic 2×5 $\eta - \phi$ strip and the total energy of the 2 ECAL trigger towers and further, the ratio of the ECAL tower and the geometrically attached HCAL tower energies is taken as a L1 quality requirement.

After the positive L1 decision, the event is forwarded to the software algorithmic based online HLT computing farm running on about 5000 CPU-cores. The data fragments of each sub-detector pass the event building network, to be rebuild as a single event in several stages. More sophistacted event reconstruction algorithms are installed, to sequentially reconstruct the event in more detail, and to reject an event as early as possible based on intermediate HLT trigger requirements, such as track quality or track to calorimeter cluster matching cuts.

If the HLT decision is positiv, the event is forwarded to the CERN central Tier-0 site for full offline event reconstruction and data archival of several reconstruction steps distributed over the world-wide LHC Grid network.

The CMS TriDAQ system therefore consists of the CMS detector front-end read-out electronics, L1 and HLT trigger system, online event building and data and detector system quality monitoring.

4. Event and Object Reconstruction

The event and object related reconstruction techniques with the CMS detector relevant for this analysis are discussed in this chapter. Particular attention is given to the reconstruction and identification of electrons as the central objects of the analysis and where significant improvements have been achieved. In particular, early contributions and commissioning studies to converge to a multivariate based CMS electron identification for the $H \to WW/ZZ$ analysis in [4], have been made, and a new official MVA based electron identification version has been developed, suitable for the $H \to \tau\tau$ search (cf. Section 4.3.2).

In addition, the reconstruction of missing transverse energy and jets are discussed to provide a solid basis of the most important event characteristics of the $H \to \tau\tau$ search. Involved algorithms for the reconstruction of the invariant mass of the di-tau system, and measurements of the distance of closest approach are described, to establish the basis for the newly developed search strategy for the di-electron channel.

4.1. Datasets and Triggers

The analysis is performed on the data collected in LHC run periods of 2011 and 2012 corresponding to an integrated luminosity of 24.6 fb^{-1} split into 4.9 fb^{-1} and 19.7 fb^{-1} at 7 TeV and 8 TeV, respectively. The precision to which the luminosity is known is 2.2% for 2011 and 2.6% at 2012 data taking periods (cf. Section 5.5).

During CMS data-taking, events are recorded by the CMS triggering system and stored according to their detector response signatures. Simple event selection procedures are applied online to collect the events of interest for the analysis to be performed. In the $H \to 2\tau \to 2e4\nu$ analysis an asymmetric double electron trigger is used and requires two L1 trigger electron primitives with p_T thresholds of 8 GeV and 17 GeV satisfying further geometrical calorimeter cluster and track requirements at HLT level. The transverse momentum is calculated from 3×3 clusters of ECAL crystals and HCAL cells, which are computed from the sum of the central tower and the highest energetic tower of the adjacent ones. Further online trigger event selection criteria are then based on the superclusters of calorimeter cells in the neighborhood of the trigger seed to account for possible Bremsstrahlung resulting from interactions of electrons with the tracker material and the geometrically attached track hypothesis seeded by the central ECAL seed and matched to the

hits in pixel detector and tracker layers. More specifically cuts on the following variables are applied:

- H/E: ration of the energy deposits in HCAL and ECAL
- $\sigma_{i\eta i\eta}$: the width of the calorimeter cluster in η direction
- $\Delta\eta_{in}$ and $\Delta\phi_{in}$: distance in η and ϕ coordinates respectively between supercluster position and track extrapolated from the primary vertex
- isolation computed from the transverse momentum of the tracks and calorimeter deposits around the center of the supercluster within the geometric cone $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.3$

Table 4.1.: Overview of the high level triggers used in the analysis of the 2011 and 2012 data-taking. The runs above 190000 belong to the 2012 data taking period.

Trigger	Run range	Int. Lumi. $[fb^{-1}]$
HLT Ele17 CaloIdL CaloIsoVL Ele8 CaloIdL CaloIsoVL	160404 - 170053	
HLT Ele17 CaloIdT CaloIsoVL TrkIdVL TrkIsoVL		4.9
Ele8 CaloIdT CaloIsoVL TrkIdVL TrkIsoVL	170053 - 180252	
HLT Ele17 CaloIdT CaloIsoVL TrkIdVL TrkIsoVL		
Ele8 CaloIdT CaloIsoVL TrkIdVL TrkIsoVL	190456 - 208686	19.7

Events firing this trigger online are then considered for further event selections and analysis steps described in Chapter 5.

Table 4.2.: Overview of the datasets used in this analysis and the corresponding integrated luminosities. The primary data stream is used according to the trigger selection

Dataset	Int. Luminosity fb^{-1}
/Run2011A-May10ReReco-v1/AOD	0.21
/Run2011A-PromptReco-v4/AOD	0.95
/Run2011A-Aug05ReReco-v1/AOD	0.39
/Run2011A-Oct03ReReco-v1/AOD	0.71
/Run2011B-PromptReco-v1/AOD	2.69
Total 7 TeV	4.95
/Run2012A-22Jan2013-v1/AOD	0.87
/Run2012B-22Jan2013-v1/AOD	4.41
/Run2012C-22Jan2013-v1/AOD	7.05
/Run2012D-22Jan2013-v1/AOD	7.36
Total 8 TeV	19.69

4.2. Primary Vertex Reconstruction

Vertices are reconstructed using the so-called Deterministic Annealing (DA) [58] clustering of tracks to be deployed for Adaptive Vertex Fitting [59]. Vertex candidates are required to be longitudinally within $24cm$ of the nominal detector center and radially within $2cm$ from the beamspot. Reconstructed vertices are required to have more than 4 degrees of freedom, which are sums of weighted tracks according to the χ^2 of their distance with respect to the associated best-fit vertex position (see [59] for details). The hard interaction vertex is selected to be the primary vertex with maximum summed up p_T of tracks associated to the vertex.

With increasing instantaneous luminosity of the LHC machine, the data-taking conditions have been constantly changing. In particular the mean number of proton-proton interactions per bunch-crossing varied, coursing altering effects over time due to pile-up (Section 3.1 and Figure 4.2). As a consequence, differences in the distribution of the number of reconstructed primary vertices for simulated samples of physics processes and observed data shown in Figure 4.1 (left) become significant. Corrections to simulated events are thus applied by re-weighting each event as a function of the number of simulated pile-up vertices to match the experimental conditions in the observed data. After these corrections the agreement between the distributions of the number of reconstructed primary vertices for simulated and observed events shown in Figure 4.1 is well improved.

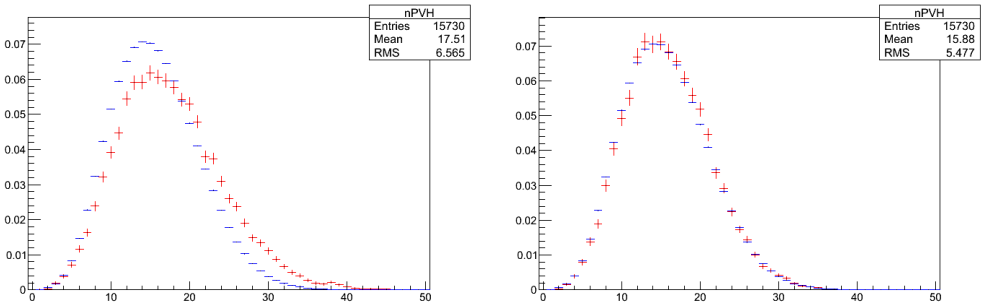

Figure 4.1.: Distribution of reconstructed primary vertices per bunch-crossing with hard interaction for 2012 8 TeV data taking period (blue) and simulated ggH signal sample (red) before (left) and after correction.

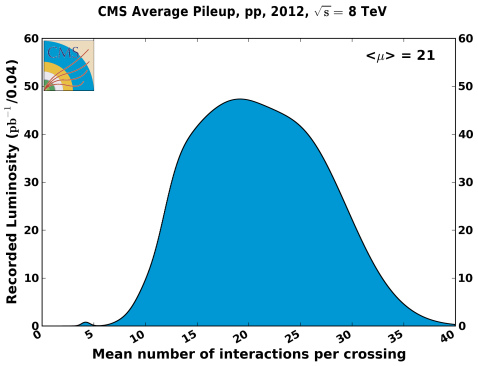

Figure 4.2.: Mean number of proton-proton interactions per bunch crossing for 2012 8 TeV data taking period. For constant instantaneous luminosity the number of proton-proton interactions per bunch-crossing follows a Poisson distribution. Due to changing LHC beam parameter configurations over time the observed distribution is an accumulation of several Poisson distributions. Taken from [49]

4.3. Electron Reconstruction and Identification

The final-state of the Higgs decay mode analyzed in this thesis is characterized by the presence of two electrons. In this section, the generic reconstruction of electrons within the CMS detector [60] as well as the more refined selection criteria to identify electrons is described. The latter is a compromise of high selection efficiency for real electrons and a low fake rate for processes faking electrons and defines the final collection of electron candidates to be considered for further analysis. In the analysis performed in this thesis real electrons come from $H/Z \to \tau\tau \to 2e4\nu$ and the direct $Z \to ee$ decays, which represents the largest background. However also physics processes faking electron signatures in the detector play a non-negligible effect and lead to a background contribution from QCD multijet and W+Jets events. In general, the following processes can fake real electron signatures:

- photon conversion: photons produced via hard bremsstrahlung of leptons interacting with the tracker material, π^0 di-photon decays or direct proton-proton interaction, undergo asymmetric conversion in the tracker material producing one electron carrying the majority of the photon energy.
- semileptonic decays of heavy flavor hadrons: partons fragmenting into leading B or D hadrons which decay semileptonically into an electron carrying the majority of the momentum of the initial parton and having a small impact parameter.
- inelastic charge exchange of charged pions and kaons: $K^{\pm}$ or $\pi^{\pm}$ interact with protons or neutrons in the nuclei of the electromagnetic calorimeter producing a π^0 which then decays into photons. The track from the pions or kaons can be matched to the cluster in the electromagnetic calorimeter which contains almost all of the energy such that E/p is close to one.

4.3.1. Electron Reconstruction

The electron reconstruction with the CMS detector is thoroughly described in [60]. Information from the pixel detector, silicon strip tracker and the ECAL is used for track finding and appropriate calorimeter clustering to build electron candidates.

Electrons interacting with the tracker material can radiate considerable fractions of their energy via bremsstrahlung emission. More specifically, about 35% of electrons radiate more than 70% of their energy before reaching the ECAL (Figure 4.3).

Due to the bending of the electron track in the strong magnetic field the spray of emitted photons off the electron track results in ECAL cluster energy deposits spread in ϕ direction in addition to the energy cluster of the electron itself. To reconstruct the initial energy of the electron, superclusters, i.e. clusters of clusters,

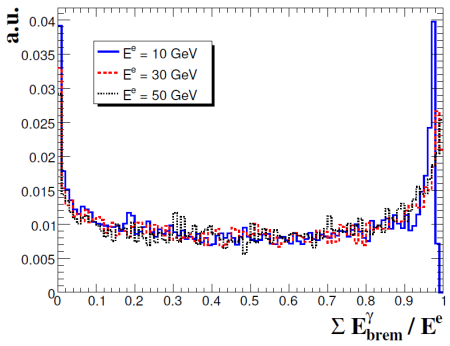

Figure 4.3.: Distribution of fractions of energy emitted via bremsstrahlung (ΣE^{γ}_{brem}) of the generated electron energy for electrons of 10,30 and 50 GeV. [60]

are build starting from the highest energetic seed cluster using the so called hybrid and island clustering algorithms [60] extending to a maximum of 0.3rad in ϕ direction.

The energy weighted mean position of the supercluster in the η-ϕ-plane is then propagated towards the pixel detector for both charge hypotheses, where a first compatible hit is being searched for within a relaxed $\Delta\phi$ and Δz window. The track seeded by the supercluster is then reconstructed using the pattern recognition technique called Gaussian sum filter (Gsf) [62] in combination with the Bethe-Heitler model allowing for proper modeling of energy loss due to bremsstrahlung. Different to the Kalman Filter (KF) technique the Gsf algorithm uses a weighted sum of Gaussians to represent each trajectory state. The trajectory with the largest weight is then taken as the reconstructed electron track.

The cluster based seeding procedure is complemented with a tracker based seeding for low pT electrons where loose track-to-cluster matching and track quality criteria are imposed on the track collection to which the Gsf algorithm is applied to reduce the number of possibilities for computation.

The electron track reconstruction efficiency measured as the fraction of generated electrons which have a reconstructed track is shown in Figure 4.5.

For the charge of the electron the majority decision of 3 measurements, i.e. the charge from the Gsf track, the matched KF track and the relative position of the innermost track hit to the supercluster is used.

A refined procedure is used to reconstruct the electron momentum for the analysis to achieve optimal separation of $Z \rightarrow \tau\tau \rightarrow 2e4\nu$, $H \rightarrow \tau\tau \rightarrow 2e4\nu$ and $Z \rightarrow ee$ in the di-electron and reconstructed SVFit di-τ mass spectra used in the final BDT signal vs. background discriminant. The energy reconstruction can be divided into three steps: ECAL cluster energy corrections as well as ECAL cluster

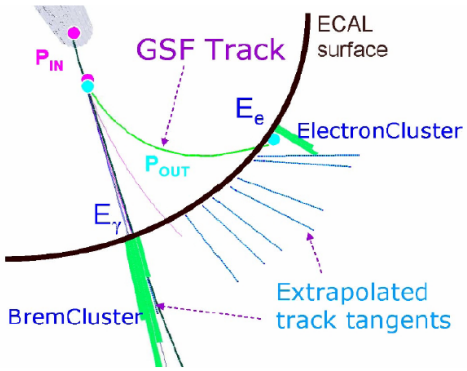

Figure 4.4.: Graphic illustrating energy loss of an electron due to bremsstrahlung within the detector. Taken from [61].

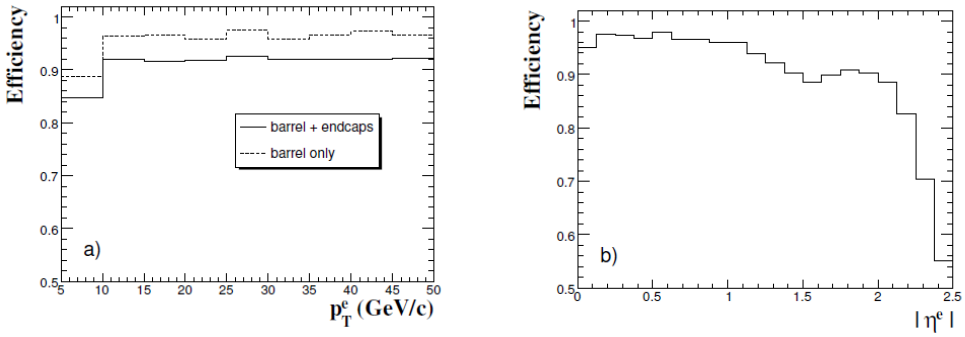

Figure 4.5.: Electron track reconstruction efficiency as a function of p_T averaged over ECAL barrel only and ECAL barrel and endcaps together (left) and as a function of η (right) [60].

energy and track momentum combination, electron energy scale calibration and smearing for MC [63] [64] [65].

First, the electron momentum is estimated from a combined measurement of the ECAL supercluster energy and the Gsf track momentum [63]. The ECAL resolution increases with the electron momentum whereas the track momentum resolution decreases for higher p_T electrons, shown in Figure 4.6. At first a multivariate [66] regression based on 60 variables is applied comparing the ECAL_{raw} energy to the generated electron energy to correct imperfect clustering of the electron supercluster which accounts for effects like missing electron-showers in the reconstructed cluster, energy leakage into the HCAL, additional ECAL energy deposits from pile-up overlapping with the electron shower and energy leakage in gaps between crystals and modules or EB to EE transitions. Then an improvement of momentum resolution of ≈40% for electrons with $p_T = 15$ GeV/c is achieved by the linear combination of the corrected ECAL supercluster energy and the Gsf track momentum measured at the distance of closest approach to the beamspot in the transverse plane again using a multivariate regression technique, as shown in Figure 4.6. The following measurements are used as input variables for the regression:

- The corrected ECAL energy and its relative error;
- The track momentum and its relative error;
- The ratio of the ECAL energy and the track momentum, and the error on this ratio;
- The ratio of the two relative errors;
- The electron category based on shower shape, η and p_T;
- Two flags for ECAL driven and tracker driven electrons
- A flag for electrons in the barrel or endcaps.

Secondly, time-dependent [65] absolute electron energy scale corrections are applied to electrons from data by categorizing $Z \rightarrow ee$ events in bins of η and $R_9 = E_{3x3}/E_{SC}$ of both electrons for different run ranges [64]. By fitting the mass distributions with a Breit-Wigner fixed to PDG values convoluted with a Crystal Ball in each category for data and MC and extracting the relative differences of offsets of the two measured peak positions from the fitted Crystal Ball mean Δm_{MC} and Δm_{data} to $m_Z^{PDG} = 91.188 GeV$,

$$\Delta P = \frac{\Delta m_{Data} - \Delta m_{MC}}{m_Z^{PDG}}.$$

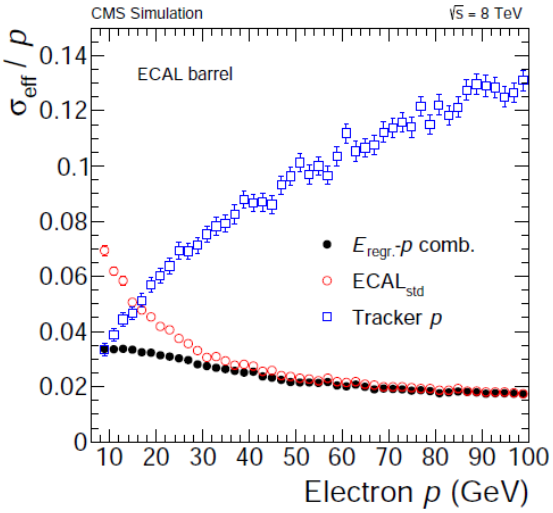

Figure 4.6.: Expected effective momentum resolution σ_{eff}/p for electrons in the EB as a function of the momentum for the corrected ECAL-only (circles), the tracker-only (squares), and the combined estimates (dots). [64]

the correction factor applied to data is determined to be $(1 - \Delta P)$.

Finally, for electron energies in simulated events a Gaussian random multiplicative factor centered at $1 + \Delta P$ with a resolution of $\Delta\sigma$ is applied to correct the energy in MC. In categories of η and the shower shape variable R_9 the resolution $\Delta\sigma$ and scale correction ΔP, being effectively nominal after the corrections applied to electrons from data, are estimated from a maximum likelihood method for pairs of $(\Delta P, \Delta\sigma)$.

The effect on the di-electron invariant mass spectrum of using the energy regression, electron energy scale correction and MC smearing explained above can be seen in Figure 4.7. Additional di-electron phase space dependent corrections to MC are applied in this analysis as discussed in Section 5.4.5 which further improve the data to predicted background agreement, as for example shown in Figure 5.14

4.3.2. Electron Identification

To further clean the reconstructed electron collection from fake processes described in the beginning of this section, a multivariate Boosted Decision Tree method is used to identify isolated leptons [67] [64].

In addition to the electron identification (ID) used in [64] (*NonTrig*), which is used as the default in this analysis, another electron ID specified to the analysis presented in this thesis and the $H \to \tau\tau$ in general (*TrigNoIP*) is developed. In this

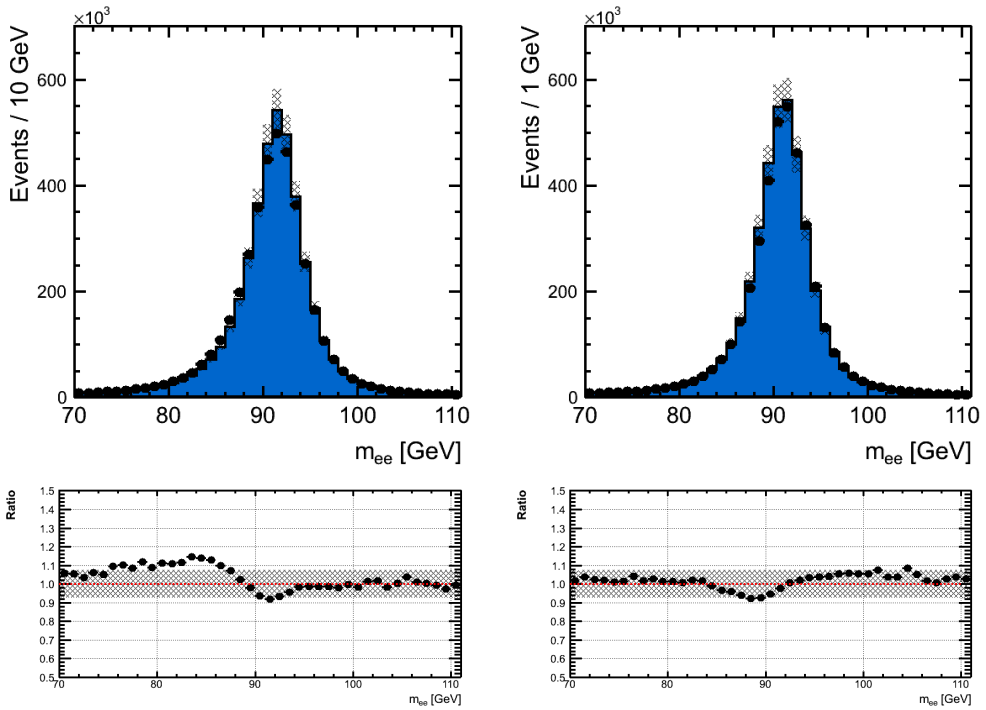

Figure 4.7.: Left: Di-electron invariant mass spectrum after electron selection without energy regression, energy scale corrections and MC smearing. Right: Di-electron invariant mass spectrum after electron selection with energy regression, energy scale corrections and MC smearing

paragraph, both approaches on the electron identification level are described and compared. Additionally, in Appendix D a preliminary comparison on analysis-level is shown. Further studies are needed for the $e\tau$, $e\mu$ and ee channels to re-optimize the corresponding analyses with respect to the TrigNoIP Electron ID and establish the best possible gain in sensetivity for RunII.

4.3.2.1. NonTrig Electron ID

The *NonTrig* ID training of the BDT is based on information such as shower or cluster shape, track information and track to cluster matching variables for electron candidates, which can be significantly different for real electrons and faking processes, as shown in Figures 4.9 - 4.11 as an illustration.

More specifically the following track-ECAL shower matching variables are used:

- E_{SC}/p_{in}, where E_{SC} is the Super-Cluster energy of the electron and p_{in} is the track momentum measurement at the innermost track position

- E_e/p_{out}, where E_e is the energy of the cluster closest to the track extrapolation to the ECAL and p_{out} is the track momentum at the outer-most track position
- $1/E_{SC} - 1/p_{gsf-mean}$, where $p_{gsf-mean}$ is track momentum taken from the weighted mean of all gsf trajectory states
- $|\Delta\eta_{in}|$, η-difference of the electron super-cluster and the track extrapolated from the innermost track state to the ECAL
- $|\Delta\phi_{in}|$, same as above for $|\Delta\eta_{in}|$ but for the ϕ-coordinate
- $|\Delta\eta_{out}|$, η-difference of the cluster closest to the track extrapolation to the ECAL and the track extrapolated from the outermost track state to the ECAL
- E_{HCAL}/E_{SC} and E_{Pre-Sh}/E_{SC}, where E_{HCAL} is the energy deposit in the HCAL behind the electron super-cluster and E_{Pre-Sh} is the ECAL Pre-Shower energy deposit towards the endcaps

In addition, the following shower shape variables are used for training:

- $\sigma_{i\eta i\eta}$, η-width of the 5x5-Block ECAL cluster centered at the highest energetic crystals, exploiting the pattern of the shower in the seed cluster
- $\sigma_{i\phi i\phi}$, ϕ-width of the 5x5-Block ECAL cluster centered at the highest energetic crystals, exploiting the pattern of the shower in the seed cluster
- η-width, η-width of the electron super-cluster, exploiting the pattern of the electron super-cluster
- ϕ-width, ϕ-width of the electron super-cluster, exploiting the pattern of the electron super-cluster
- $R_9 = E_{3x3}/E_{SC}$, where E_{3x3} is the energy in the 3x3-Block ECAL cluster centered at the highest energetic crystal
- $(E_{5x5} - E_{1x5})/E_{5x5}$, where E_{1x5} is the 1x5 ϕ-strip around the cluster seed and E_{5x5} is the energy in the 5x5-Block ECAL cluster centered at the highest energetic crystal

The variable $\sigma_{i\eta i\eta}$ and likewise $\sigma_{i\phi i\phi}$ is defined as follows:

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_{i\in 5\times 5} w_i \cdot (\eta_i - \langle\eta\rangle_{5\times 5})^2}{\sum_{i\in 5\times 5} w_i}}, w_i = \max\Big(0, 4.7 + ln\frac{E_i}{E_{5\times 5}}\Big) \tag{4.1}$$

where η_i and E_i are the η-coordinate and Energy of the i'th crystal in the 5x5-Block around the highest energetic crystal and $\langle\eta\rangle_{5\times5}$ is the centroid of the 5x5-Block of crystals. Accordingly, the variables η-width and ϕ-width of the supercluster are defined via the sum over all clusters in the supercluster.

The following track related variables complete the set of input parameters:

- $f_{brem} = (p_{in} - p_{out})/p_{in}$, where p_{in} and p_{out} are the track momenta measured at the innermost and outermost track state respectively. This is the electron energy fraction emitted by Bremsstrahlung.
- χ^2_{Gsf}, chi-squared test for the Gsf track reconstruction
- χ^2_{KF}, chi-squared test for the Kalman-Filter track reconstruction
- $Nhits_{KF}$, number of hits in the tracker associated to the KF track reconstruction

The training for the *NonTrig* ID is optimized for the $H \to ZZ \to 4l$ analysis where low p_T signal electrons are expected from the virtual Z-Boson decay. Hence the signal training sample is chosen to be a mixture of 2011 7 TeV simulated $H \to ZZ \to 4e$ events for Higgs mass hypotheses of 115-120-130-140 GeV/c^2.

Z+jets processes constitute the most dominant background resulting from false lepton identification in the $H \to ZZ \to 4l$ analysis. Therefore, to be as close to the analysis use case, however still unbiased the background training sample, fake electrons accompanying W bosons, i.e. W+1fake, are directly extracted from 2011 7 TeV data, using various cuts to reduce signal contamination from Z boson and leptonic top quark decays. No pre-selection quality requirements for signal or fake training electrons are imposed. This mirrors best the analysis use case where 2 electrons from the virtual Z of $H \to ZZ \to 4e$ decays do not necessarily pass trigger requirements and can in particular be well below the trigger p_T trigger thresholds of 10 GeV/c.

To achieve optimal separation between electrons and electron faking processes over the full range of the p_T and η spectra, where the input variables are depend on the tracker and ECAL architecture, the training is performed separately in six distinct bins: $|\eta| < 0.8$, $0.8 < |\eta| < 1.479$ and $1.479 < |\eta| < 2.5$ in η and 5 GeV/c $< p_T \leq 10$ GeV/c and $p_T > 10$ GeV/c in p_T.

The performance in terms of *receiver operating characteristic* (ROC) curves in comparison to the *TrigNoIP* ID is shown in Figures 4.12 - 4.14.

4.3.2.2. TrigNoIP Electron ID

For the *TrigNoIP* electron ID, the following modifications to the BDT training configuration of the *NonTrig* ID are applied:

- the signal electrons in all sub-channels of the $H \to \tau\tau$ analysis have to pass the p_T trigger threshold of at least 10 GeV/c. Hence the separation of the training in bins of η and p_T are adjusted to be: $|\eta| < 0.8$, $0.8 < |\eta| < 1.479$ and $1.479 < |\eta| < 2.5$ in η and 10 GeV/c $< p_T \leq 20$ GeV/c and $p_T > 20$ GeV/c in p_T. The low p_T bins of the *NonTrig* ID are redundant for triggering electrons where thresholds are above 10 GeV/c. This is adopted from the MVA based electron ID [67] used in the CMS $H \to WW$ analysis.

- the signal electrons in all sub-channels of the $H \to \tau\tau$ analysis pass the quality requirements imposed by the electron-leg of the cross trigger used to collect data. Thus a set of pre-selection requirements is imposed on the electron candidates entering the training, allowing the BDT to focus on the optimal separation in the variable phase-space of interest.

- the *NonTrig* ID is Pile-Up dependent, see Figures 4.15 - 4.17. The variable $\rho = \langle dE^{PU+UE}/d\eta d\phi \rangle$ is thus added to the set of input of variables for the *TrigNoIP* ID, yielding less Pile-Up dependency in the selection efficiency.

- the initial training samples for the *TrigNoIP* ID were changed to Z+jets 2012 MC at 8 TeV center of mass energy for signal electrons from Z Boson decays and multijet events extracted directly from 8 TeV data collected in 2012 for the fake training sample.

Some parts of these configurations coincide with the MVA based electron ID [67] used in the CMS $H \to WW$ analysis. The *TrigNoIP* ID however excludes explicitly impact parameter information, such that no inefficiency for electrons from tau decays is introduced. This is one of the important differences to the electron ID used in the $H \to WW$ analysis.

Except for the ρ variable, the input observables for the *TrigNoIP* electron ID are the same as for the *NonTrig* ID and shown in Figures 4.9- 4.11. The 2012 Z+jets MC was Pile-Up re-weighted as explained in Section 4.2. An additional re-weighting to the distinct electron p_T spectrum of $H \to \tau\tau \to 2e + 4\nu$ decays was studied but did not reveal significant improvement, and is thus not applied. The gradient boost method with 2000 trees and a shrinkage parameter of 0.10 was used. Depending on the η and p_T bin, the number of input events for the training varies but is always above 20000 both for background and signal ensuring that no significant bias from the training sample is introduced. An example for the BDT output is shown in Figure 4.8.

The QCD multijet training sample for fake electrons extracted from 2012 data was collected using dedicated single electron triggers. To achieve high purity of electron faking processes the events are required to have exactly one reconstructed electron candidate with a p_T of less than 35 GeV/c and missing transverse energy

less than 20 GeV/c to suppress real electrons from W decays. After these requirements the selected sample of multijet events with one quark or gluon misidentified as a promptly produced electron has a purity of above 97%.

As discussed above, the electron candidates are required to pass pre-selection criteria mimicking the quality criteria imposed by electron HLTs on electron candidates before entering the BDT training. These are as follows:

- $\sigma_{i\eta i\eta} < 0.01, (0.03)$ for EB (EE)
- $|\Delta\eta_{in}| < 0.15, (0.10)$
- $|\Delta\phi_{in}| < 0.007, (0.009)$
- $E_{HCAL}/E_{SC} < 0.12, (0.10)$
- relative track/ECAL/HCAL isolation with $\Delta R = 0.3$ to be ≤ 0.2
- $|d_z| < 0.1$ cm
- $|d_0| < 0.02$ cm
- photon conversion rejection as described in Section 4.3.3

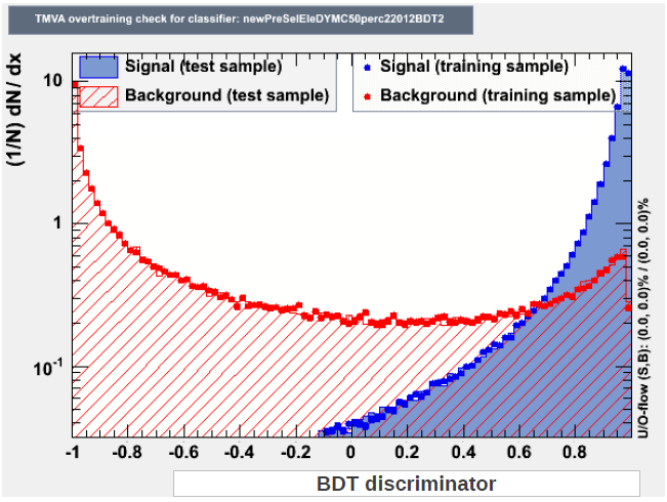

Figure 4.8.: Electron discriminator output for the *TrigNoIP* ID training configuration for signal (blue) and background (red). Overlaid are the outputs of the training samples themselves (dots) and the testing sample (filled) histograms. No large differences are observed, showing that no bias from the training sample is introduced.

Table 4.3.: Cut values (>) for *TrigNoIP* BDT based electron identification in the $e\mu$, ee analysis

	$\|\eta\| < 0.8$	$0.8 \geq \|\eta\| \leq 1.479$	$\|\eta\| \geq 1.479$
$p_T \leq 20$ GeV/c	-0.5375	-0.375	-0.025
$p_T > 20$ GeV/c	0. 325	0. 775	0.775

Table 4.4.: Cut values (>) for *TrigNoIP* BDT based electron identification in the $e\tau_h$ analysis

	$\|\eta\| < 0.8$	$0.8 \geq \|\eta\| \leq 1.479$	$\|\eta\| \geq 1.479$
$p_T > 20$ GeV/c	0.55	0.9	0.925

Comparisons of the *TrigNoIP* and *NonTrig* ID are shown in Figures 4.12- 4.14. A significant gain in electron identification performance is observed for the *TrigNoIP* ID over the η and p_T range of interest.

Working points for the *TrigNoIP* ID in each η and p_T bin are chosen to yield the same fake rate as the *NonTrig* ID for the working points used in the $e\mu$, ee (given in Table 4.5) and $e\tau_h$ final state analysis and shown in Table 4.3 and 4.4. In Figures 4.15 - 4.17, the efficiency over the number of primary vertices for the extracted working points are shown for both electron IDs. Depending on the working point the gain in signal efficiency for the same fake rate is 5-13% per electron.

4.3.3. Electron Selection

In this section, the final selection of electron candidates for this analysis are described. Electrons candidates considered in this analysis are required to be within $|\eta| < 2.3$ of the detector and are required to pass cuts applied on the *NonTrig* BDT outputs dependent on the transverse momentum and pseudorapidity, shown in Table 4.5.

Table 4.5.: Cut values (>) for BDT based electron identification *NonTrig* used in the ee and $e\mu$ channels.

	$\|\eta\| < 0.8$	$0.8 \geq \|\eta\| \leq 1.479$	$\|\eta\| \geq 1.479$
$p_T \leq 20$ GeV/c	0.925	0.915	0.965
$p_T > 20$ GeV/c	0.905	0.955	0.975

The impact parameter in the transverse plane with respect to the primary vertex is required to be $|d_0| < 0.02$cm. The longitudinal impact parameter with respect

to the primary vertex is required to be $|d_z| < 0.2$cm. Also required is that there are no missing hits in the inner layers of the tracker, no hits before the associated vertex and a probability $P > 10^{-6}$ for the vertex fit to reduce the selection of electrons from photon conversions.

Furthermore, electrons are required to be isolated for further discrimination against electron candidates within jets. The analysis exploits the particle-flow based relative isolation variable computed as:

$$\mathrm{Iso}_e^{\mathrm{PF}} = \frac{\sum (p_T\mathrm{charged} + p_T\gamma + p_T\mathrm{neutral})}{p_T^e}, \tag{4.2}$$

where the sum runs over the transverse momenta of all charged particles emerging from the hard interaction vertex, and photons and neutral hadrons within a cone $\Delta R < 0.4$ around the lepton momentum. To remove the electron footprint within the isolation cone, veto cones for the constituents are defined in Table 4.6. Due to missing tracks for neutral hadrons, the assignment to the primary vertex is not possible. The $\Delta\beta$-correction is applied to the isolation variable to subtract pile-up contributions from neutral hadrons and is estimated to be 50% of the charged hadrons from pile up interactions.

$$\mathrm{Iso}_e^{\mathrm{PF}-\Delta\beta} = \frac{\sum_{\mathrm{charged}} p_T + \max\left(0, \sum_{\gamma p_T} + \sum_{\mathrm{neutral}} p_T - \Delta\beta\right)}{p_T^e}, \tag{4.3}$$

with

$$\Delta\beta = 0.5 \times \sum_{PU-charged} p_T. \tag{4.4}$$

Table 4.6.: Electron isolation veto cones (ΔR in $\eta - \phi$ space)

	Charged cand.	Photon cand.	Neutral cand.	PU Charged cand.
e(Barrel)	0.01	0.08	none	none
e(Endcap)	0.015	0.08	none	none

The $\Delta\beta$-corrected relative isolation variable $\mathrm{Iso}_e^{\mathrm{PF}-\Delta\beta}$, is required to be less than 0.1 (0.15) for electrons in Barrel (Endcap) regions.

The kinematic distributions of the selected electrons are displayed in Figure 5.2.

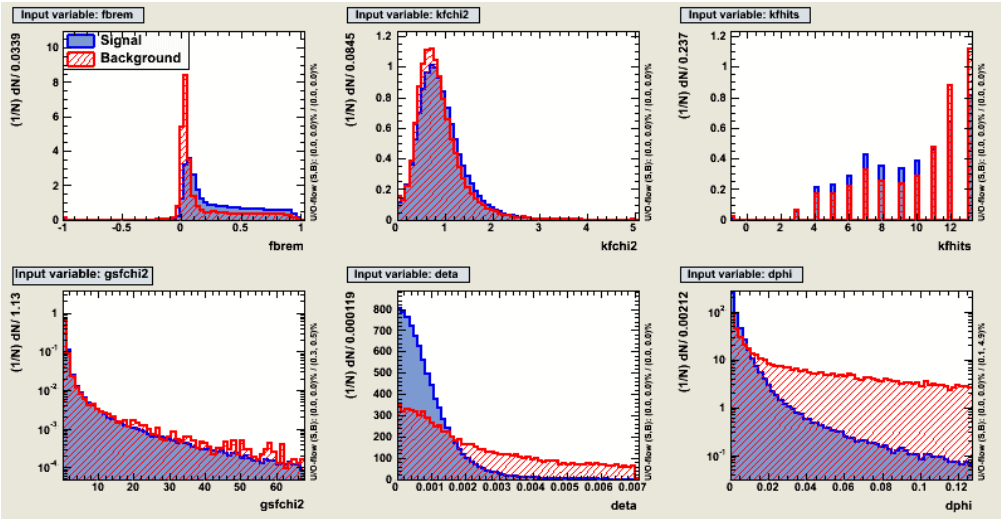

Figure 4.9.: Discriminating input variables for the electron IDs for signal (blue) and fake (red) electrons in the bin $|\eta| < 0.8$ and $p_T > 20$ GeV/c.
Top l.t.r.: f_{brem}, χ^2_{KF}, $Nhits_{KF}$
Bottom l.t.r.: χ^2_{Gsf}, $|\Delta\eta_{in}|$, $|\Delta\phi_{in}|$

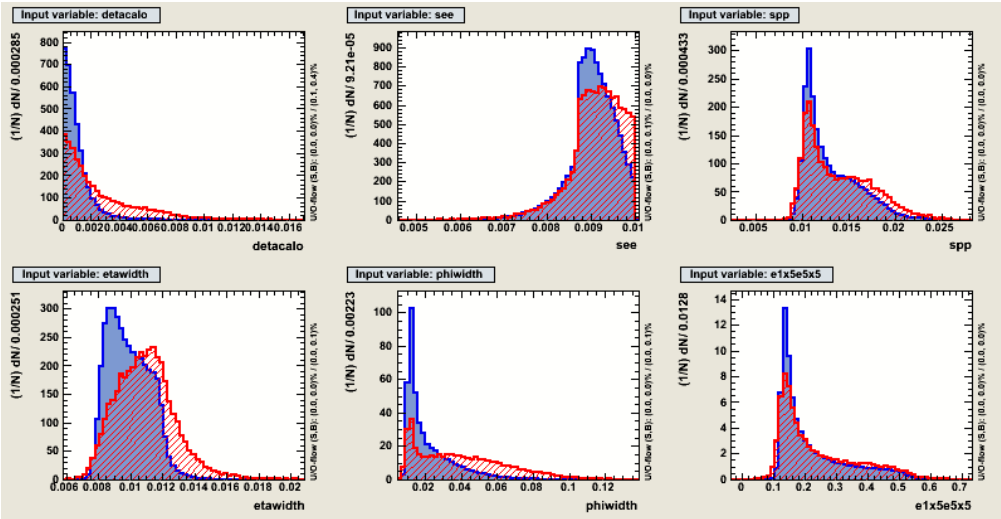

Figure 4.10.: Discriminating input variables for the electron IDs for signal (blue) and fake (red) electrons in the bin $|\eta| < 0.8$ and $p_T > 20$ GeV/c.
Top l.t.r.: $|\Delta\eta_{out}|$, $\sigma_{i\eta i\eta}$, $\sigma_{i\phi i\phi}$
Bottom l.t.r.: η-width, η-width, $(E_{5x5} - E_{1x5})/E_{5x5}$

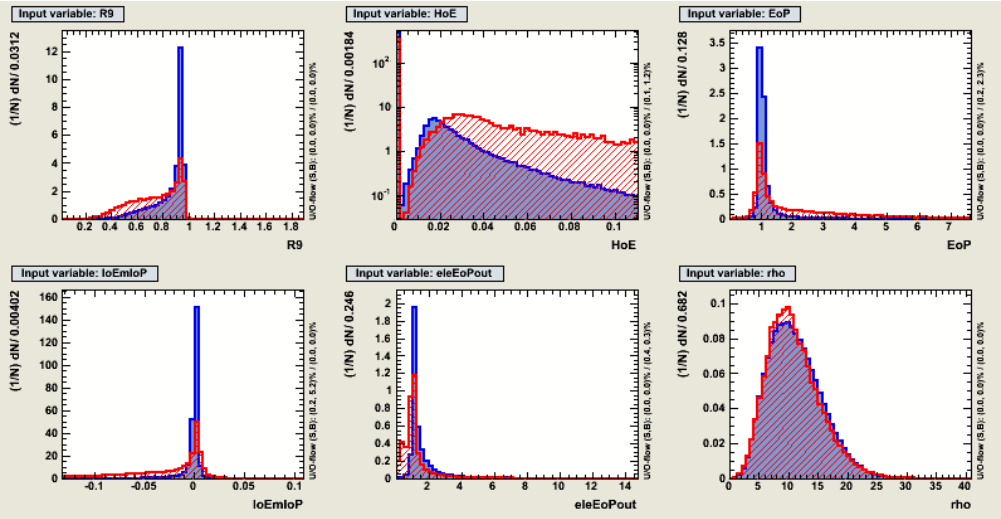

Figure 4.11.: Discriminating input variables for the electron IDs for signal (blue) and fake (red) electrons in the bin $|\eta| < 0.8$ and $p_T > 20$ GeV/c.
Top l.t.r.: R_9, E_{HCAL}/E_{SC}, E_{SC}/p_{in}
Bottom l.t.r.: $1/E_{SC} - 1/p_{gsf-mean}$, E_e/p_{out}, ρ

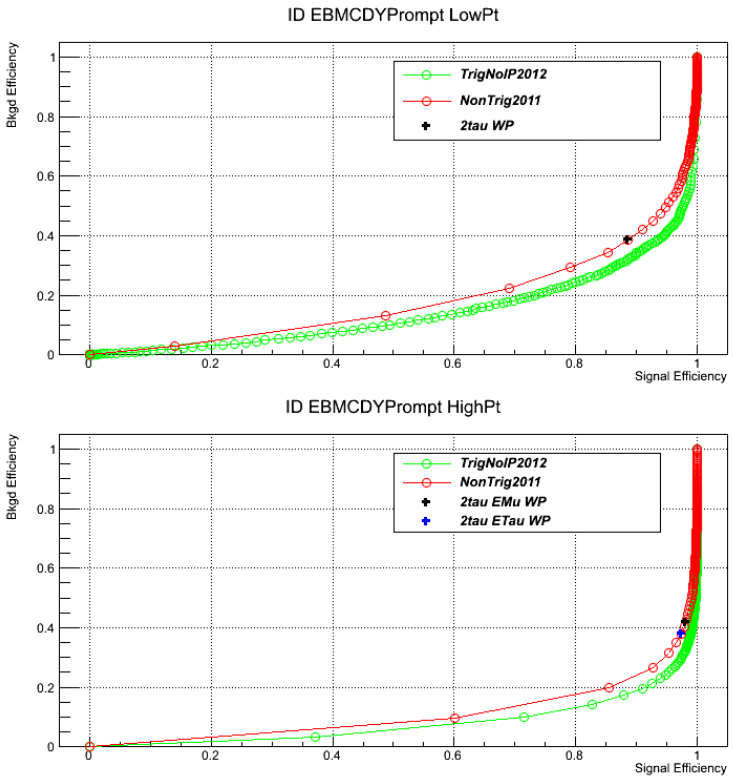

Figure 4.12.: Top: ROC curves for the *NonTrig* (red) and *TrigNoIP* (green) electron IDs on 2012 Z+jets MC in the bin $|\eta| < 0.8$ and 10 GeV/c $< p_T \leq 20$ GeV/c. The black cross depicts the *NonTrig* ID working point used in the $e\mu$ and ee final state analyses. The denominator is defined as the set of electron probes passing the trigger quality requirements.
Bottom: ROC curves for the *NonTrig* (red) and *TrigNoIP* (green) electron IDs on 2012 Z+jets MC in the bin $|\eta| < 0.8$ and $p_T > 20$ GeV/c. The black (blue) cross depicts the *NonTrig* ID working point used in the $e\mu$ and ee ($e\tau_h$) final state analyses. The denominator is defined as the set of electron probes passing the trigger quality requirements.

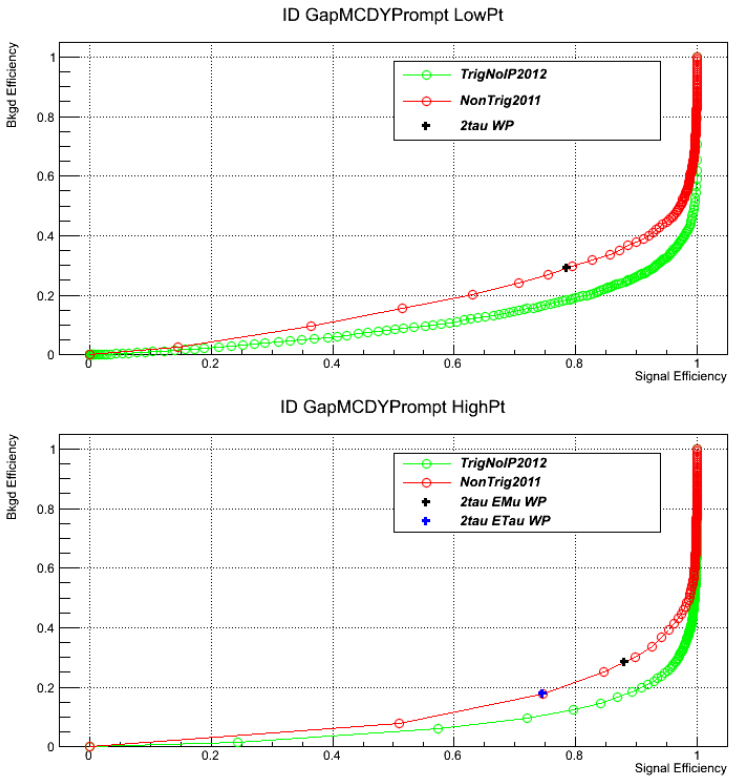

Figure 4.13.: Top: ROC curves for the *NonTrig* (red) and *TrigNoIP* (green) electron IDs on 2012 Z+jets MC in the bin $0.8 < |\eta| < 1.479$ and 10 GeV/c $< p_T \leq 20$ GeV/c. The black cross depicts the *NonTrig* ID working point used in the $e\mu$ and ee final state analyses. The denominator is defined as the set of electron probes passing the trigger quality requirements.
Bottom: ROC curves for the *NonTrig* (red) and *TrigNoIP* (green) electron IDs on 2012 Z+jets MC in the bin $0.8 < |\eta| < 1.479$ and $p_T > 20$ GeV/c. The black (blue) cross depicts the *NonTrig* ID working point used in the $e\mu$ and ee ($e\tau_h$) final state analyses. The denominator is defined as the set of electron probes passing the trigger quality requirements

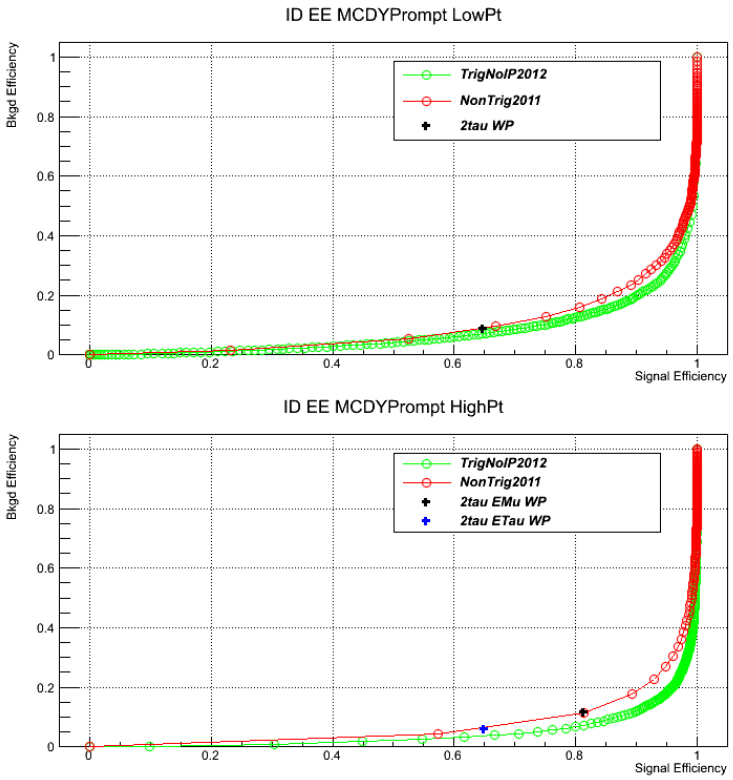

Figure 4.14.: Top: ROC curves for the *NonTrig* (red) and *TrigNoIP* (green) electron IDs on 2012 Z+jets MC in the bin $1.479 < |\eta| < 2.5$ and 10 GeV/c $< p_T \leq 20$ GeV/c. The black cross depicts the *NonTrig* ID working point used in the $e\mu$ and ee final state analyses. The denominator is defined as the set of electron probes passing the trigger quality requirements.
Bottom: ROC curves for the *NonTrig* (red) and *TrigNoIP* (green) electron IDs on 2012 Z+jets MC in the bin $1.479 < |\eta| < 2.5$ and $p_T > 20$ GeV/c. The black (blue) cross depicts the *NonTrig* ID working point used in the $e\mu$ and ee ($e\tau_h$) final state analyses. The denominator is defined as the set of electron probes passing the trigger quality requirements

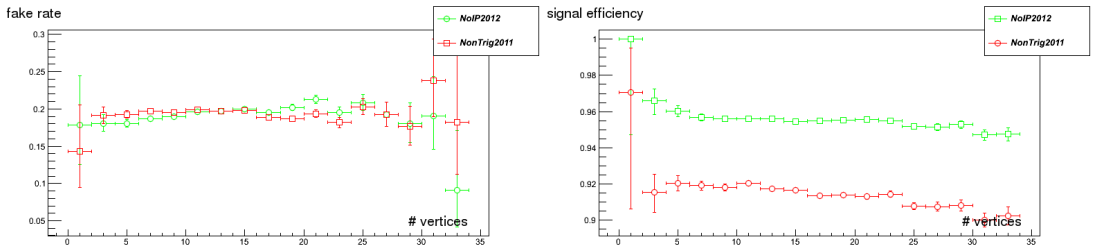

Figure 4.15.: Left: Fake rate over the number of primary vertices for the working point used in the $e\mu$ and ee analyses for $p_T < 20$ GeV/c.
Right: Signal efficiency over the number of primary vertices for the working point used in the $e\mu$ and ee analyses for $p_T < 20$ GeV/c.

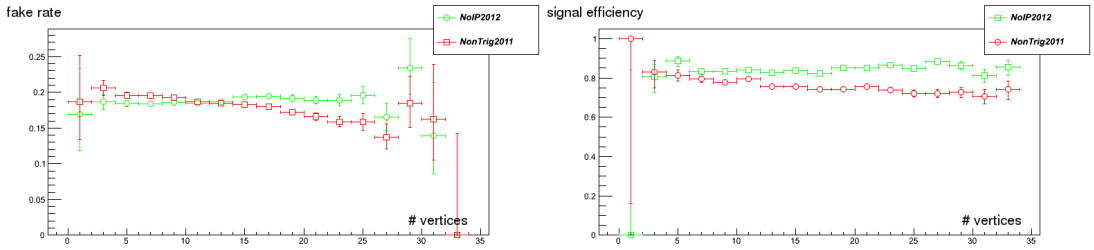

Figure 4.16.: Left: Fake rate over the number of primary vertices for the working point used in the $e\mu$ and ee analyses for $p_T > 20$.GeV/c.
Right: Signal efficiency over the number of primary vertices for the working point used in the $e\mu$ and ee analyses for $p_T > 20$ GeVc.

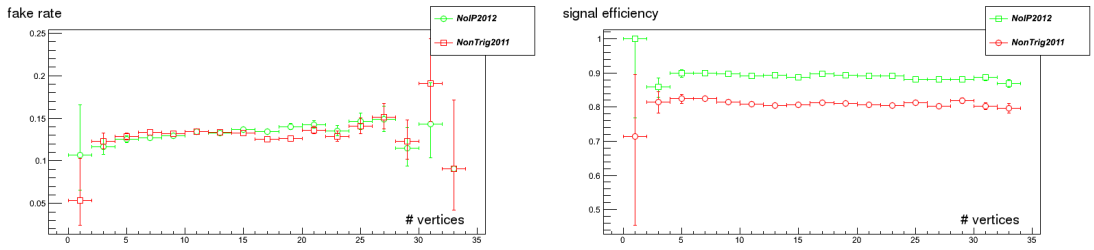

Figure 4.17.: Left: Fake rate over the number of primary vertices for the working point used in the $e\tau_h$ analyses for $p_T > 20$ GeV/c.
Right: Signal efficiency over the number of primary vertices for the working point used in the $e\tau_h$ analyses for $p_T > 20$ GeV/c.

4.4. Particle Flow Algorithm

The Particle Flow Algorithm [68] refers to a method of directly reconstructing events in terms of all produced particles, namely charged hadrons, neutral hadrons, muons, electrons and photons, by combining all CMS sub-detector responses and in particular tracker and calorimeter information. The combination of tracker and calorimeter information on reconstruction level is in contrast to the reconstruction of for example jets solely based on their calorimeter clusters, where directional biases for charged hadrons due to the magnetic field are not accounted for and the jet energy resolution is largely affected by the energy resolution of the HCAL ($\sim 90\% - 120\%/\sqrt{E}$). On average jet energy fractions of charged particles, photons and neutral hadrons are about 65%, 25% and 10% respectively, and hence ∼90% of a jets energy measurement can in principle be more accurately measured based on the tracking of the charged particles and ECAL clusters from jet associated photons.

Due to the high granularity of the CMS tracking device, ECAL and slightly coarser HCAL, the Particle Flow Algorithm is suitable for the CMS event reconstruction procedure and results in an improved jet energy resolution by about 40% as well as an improved directional jet angular resolution by a factor of two.

Tracks are iteratively reconstructed with the Kalman Filter technique with loosened seeding criteria in each iteration. Hits in the pixel and silicon detector which were associated to tracks of the previous iteration are removed. The tracks are then extrapolated from the last hit in the tracker towards the calorimeters and *linked* to the ECAl and HCAL clusters if the extrapolated track is within the boundaries of a cluster. Based on the distance in η-ϕ-coordinates of the extrapolated track and the cluster seeds, the algorithm assigned track to cluster matching quality measures to each link to be taken into account for resolving ambiguities. Additionally, ECAL clusters are linked to tangents of the tracks at each layer of the tracking device to account for Bremsstrahlung in the identification of electrons.

Iteratively, detector responses of unambiguous links such as for muons are grouped together and removed for the next iteration. Electrons are reconstructed similar to as explained in Section 4.3 with additional quality requirements on reconstruction level reassembled in a multivariate discriminator, to reduce the fake rate and thus release the corresponding tracks and calorimeter clusters for next iterations in case the electron candidate fails the quality requirements. Dedicated cleaning and cluster calibration methods are then used to identify and disentangle charged hadrons from neutral particles (hadrons or photons) by comparing the momentum and energy of the linked tracks and clusters. Neutrals are then identified as significant energy excesses of calorimeter clusters over the measured momentum of the linked tracks.

For this analysis particle flow particles are in particular used to reconstruct jets (Section 4.5) and missing transverse energy (Section 4.6), both of which are playing

central roles in the Higgs search strategy.

4.5. Jet Reconstruction and B-Tagging

The $H \to \tau\tau \to 2e4\nu$ analysis, heavily relies on jet multiplicities as well as their kinematics in each selected event:

- jets constitute the recoil to boosted Higgs production processes and thus enhance the resolution of the svFit di-tau mass reconstruction (Section 4.7)
- the SM Higgs production via vector boson fusion has a distinct 2-jet topology, and therefore jet-related variables are used as input variables to the multivariate analysis in the VBF category described in Section 5.2;
- the event categorization of the analysis discussed in Section 5.2 is entirely based on the multiplicity of high energetic jets. Higgs productions via gluon fusion and vector boson fusion have on average higher final-state jet multiplicities compared to the Z boson background produced via quark anti-quark annihilation. The jet multiplicity is therefore suitable to separate signal from background.

4.5.1. Jet Reconstruction

In this analysis, jets are clustered from particles reconstructed and identified by the particle flow algorithm. For clustering the anti-kT algorithm [69] with a distance of $\Delta R = 0.5$ is used.

Several steps to correct jet energies are implemented. First, energy contributions from pile-up, underlying event and detector noise are subtracted using the jet area method [70] [71], where the event-by-event median, possibly dependent on the pseudo-rapidity, of hadronic activity over the detector area not due to the hard interaction, ρ, is measured and subtracted after multiplying ρ with the area of the jet. Further, non-linear p_T and η biases of the jet energy response due to the different geometrical coverages of the sub-detectors (tracker and calorimeters) are corrected on simulated events such that the reconstructed jet energy is on average calibrated to the energy of the initial MC particle fragmenting and hadronizing into the jet. Finally, p_T and η dependent data-to-MC differences are corrected using di-jet and $Z/\gamma^* + jets$ events. The jet energy resolution after these corrections varies from 4% for jets at 1 TeV to 15% for low energetic ($\approx$ 20 GeV) jets.

Typically, Higgs bosons produced via vector boson fusion are accompanied by two jets in the forward regions of the detector. In these regions of the detector, where the tracking detector is not available and thus tracks of charged particles cannot directly be associated to the primary vertex, it is difficult to distinguish

jets of the hard interaction from jets of pile-up interactions. Therefore, a pile-up jet identification algorithm [72] is deployed to reduce the pile-up contamination of VBF-like events. The Jet ID is based on a multivariate boosted decision tree and uses two sets of variables describing the jet shape and vertex compatibility. The most important vertex related variables, which are primarily relevant for the region covered by the tracker $|\eta|$ <2.5, are for each jet defined as follows:

$$\beta = \frac{\sum_{\mathrm{HS-charged}} p_T}{\sum_{\mathrm{all-charged}} p_T}, \qquad \beta^* = \frac{\sum_{\mathrm{PU-charged}} p_T}{\sum_{\mathrm{all-charged}} p_T} \tag{4.5}$$

where "HS-charged" are the charged PF particles in the jet associated to the hard-scattering primary vertex, "PU-charged" are the PF charged particles associated to pile-up interaction vertices. Variables encoding the shape of a jet become more relevant for $|\eta|$ >2.5, where the discrimination is mainly based on the calorimeter response, and the two most important are defined as follows:

$$\langle \Delta R^2 \rangle = \frac{\sum_{\mathrm{allparticles}} p_T^2 \Delta R^2}{\sum_{\mathrm{allparticles}} p_T^2} \tag{4.6}$$

$$p_T^{\Delta R}(X) = \frac{1}{p_T^{jet}} \sum_{\mathrm{allparticleswith} X \leq \Delta R < X+0.1} p_T. \tag{4.7}$$

The sums run over all PF particles clustered in jet, p_T^{jet} denotes the transverse momentum of the jet being identified, ΔR is the distance of a particle with respect to the jet axis in $\eta - \phi$ space and $p_T^{\Delta R}(X)$ is calculated for 5 values of X, i.e. 0.0 - 0.4 in steps of 0.1. In addition to these variables the total number of charged as well as neutral PF particles clustered into the jet are taken as an input to the BDT discriminant.

After the cut on the BDT discriminator for jets $p_T > 30$ GeV/c a signal efficiency of approximately 99% is achieved with a pile-up rejection of about 90% in the central region of the detector, where tracking is available. In the tracker-endcap transition region the signal efficiency is about 95% with a background rejection of about 60% and in the endcap region ($0.5 > |\eta| > 0.3$) the background rejection is lowered to roughly 40% at a signal efficiency of about 80%.

4.5.2. B-tagging

In this analysis events with b-tagged jets with $p_T > 20$ GeV/c and $|\eta| < 2.4$ are rejected to reduce the $\bar{t}t$ background as well as the $t\bar{t}$ contamination of the $Z \to \tau\tau$ embedded sample (Section 5.4.4) to a negligible level.

To tag b-jets the *Combined Secondary Vertex* (CSV) algorithm [73] is used. The algorithm combines track/life-time and secondary vertex related variables into two Likelihood ratio to discriminate b from c quarks and b from light quarks and gluons to be linearly combined afterwards. According to the quality of the secondary vertex fit, jets are classified into "real", "pseudo vertex" and "no vertex" and the training is then done on the following discriminating variables:

- the 2D flight distance significance in the transverse plane;
- the vertex mass;
- the number of tracks at the vertex;
- the ratio of the energy carried by tracks at the vertex with respect to all tracks in the jet;
- the pseudorapidities of the tracks at the vertex with respect to the jet axis;
- the 2D IP significance of the first track that raises the invariant mass above the charm threshold of 1.5 GeV/c^2 (tracks are ordered by decreasing IP significance and the mass of the system is recalculated after adding each track);
- the number of tracks in the jet;
- the 3D IP significances for each track in the jet.

A jet is considered as *b*-tagged if it has a discriminator value of the *Combined Secondary Vertex* (CSV) tagger $d_{\mathrm{CSV}}^{\mathrm{BTag}} > 0.679$. As mentioned above, for *b*-tagged jets, the momentum threshold is lowered to $p_T > 20$ GeV/c, while the allowed pseudorapidity range is restricted to $|\eta| < 2.4$. The b-tagging efficiency ranges from approximately 70% in the inner tracking system ($|\eta| < 1.5$) to about 60% for $|\eta| > 1.5$ at a mis-tag rate of about 1.5% [74]. The b-tagging efficiency and mis-tag rate measured on data is propagated to the number of final state b-quarks and light partons on parton level in simulated MC events to match the exact b-tagging efficiencies.

Figure 5.3 shows the total jet multiplicity and *b*-jet multiplicity distributions for data and simulated events after the preselection of di-electron events as described in Section 5.1.1. The leading and subleading jet kinematic variables are shown in Figure 5.4. A good agreement between data and simulation can be observed.

4.6. Missing Transverse Energy

Due to the 4 neutrinos intrinsic to the final-state analyzed in this thesis, the measurement of the missing transverse energy (E_T^{miss} or MET) in each event is a crucial ingredient for this analysis:

- the missing transverse momentum is utilized by the Secondary Vertex Fit (SVfit) algorithm employed for the reconstruction of the invariant mass of tau-lepton pairs (Section 4.7),
- the E_T^{miss} is used in the multi-variate analysis exploiting the VBF production of the SM Higgs boson. (Section 5.2.1).

The analysis uses the E_T^{miss} reconstruction determined with a multivariate regression technique which is described in detail in reference [75]. In principle, the missing transverse energy for each event can be computed as the negative vectorial sum of all reconstructed particle flow objects in the event (PFMET). However it is found that the PFMET resolution is primarily affected by the relative energy resolution of 10-15% [70] of the hadronic components in the event, which leads to significant dependencies of the E_T^{miss} resolution on additional pile-up interactions with hadronic deposits in the detector.

The idea of the E_T^{miss} regression (MVAMET) [75] is to separate charged and neutral hadrons associated to the hard scattering process from those associated to pile-up interactions, and then based on these two sets to reduce the contribution to the vectorial sum of each part which is likely not to be the cause of the E_T^{miss} in the event. In Z boson production processes with small boost and hence low energetic hadronic recoil, the E_T^{miss} resolution is predominantly deteriorated by additional hadronic particles from pile-up interactions and therefore only considering particles from the hard scattering process yields a more accurate E_T^{miss} measurement. Also for events with genuine sources of E_T^{miss}, such as leptonically decaying $t\bar{t}$ and $Z \to \tau\tau$ events, taking into account hadronic particles associated to pile-up interactions have almost no effect on the nominal E_T^{miss} value but worsen the resolution considerably. However for highly boosted Z boson productions, and therefore high energetic hadronic recoil, high energetic neutral hadrons from the hard interaction may not be correctly associated to the hard scattering vertex and thus only by adding the additional hadrons from all interactions in the bunch crossing will give an accurate E_T^{miss} measurement.

More specifically, in addition to the two highest energetic jets in the event and number of reconstructed primary vertices, the scalar sum of all PF particles, magnitude and azimuthal angle of the transverse hadronic recoil of the following E_T^{miss} variables enter the regression:

- the negative vectorial sum of all PF particles in the transverse plane (i.e. PFMET),

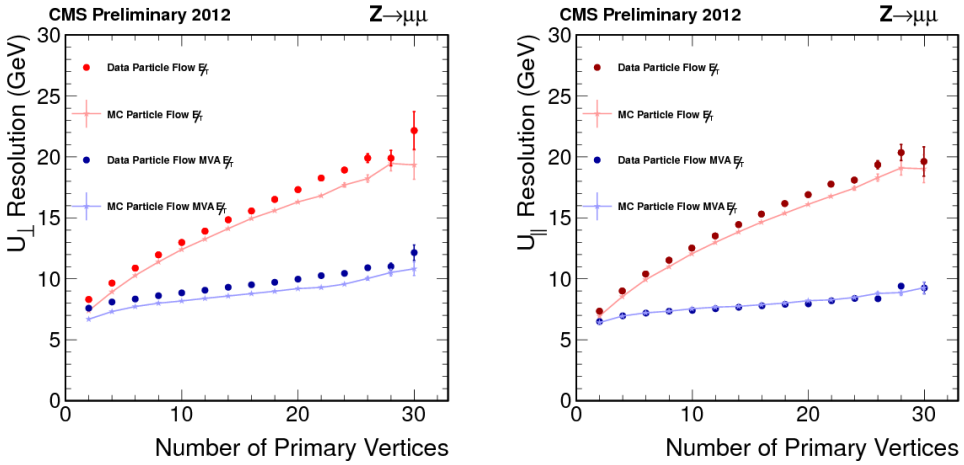

Figure 4.18.: Resolutions of the perpendicular (left) parallel (right) components of the hadronic recoil, for the mva based MVAMET regression (blue) and PFMET (red). Taken from [76].

- the negative vectorial sum of all charged PF particles that have been associated to the selected hard-scatter vertex,
- the negative vectorial sum of all charged PF particles that have been associated to the hard-scatter vertex and all neutral PF particles within jets that have passed the MVA pileup jet ID,
- the negative vectorial sum of all charged PF particles that have not been associated to the hard-scatter vertex and all neutral PF particles within jets that have failed the MVA pileup jet ID,
- the negative vectorial sum of all charged PF particles that have been associated to the hard-scatter vertex and all neutral PF particles (also those that have not been clustered into jets) plus the positive vectorial sum of all neutral PF particles within jets that have failed the MVA pileup jet ID.

The resolution for PFMET and the MVAMET measurement is compared in Figure 4.18. An improved resolution of 40% for bunch-crossings with 15 reconstructed primary vertices can be observed for MVAMET with respect to PFMET.

Due to the imperfect modeling of the simulation of underlying events and (out-of-time) pile-up, and therefore an improper modeling of Z boson recoil, the E_T^{miss} distribution for simulated events is expected to be different compared to data. In Figure 4.20 (left) a comparison of the E_T^{miss} distributions for simulated events and data is shown and significant differences can be observed in the region $E_T^{miss} < 100$

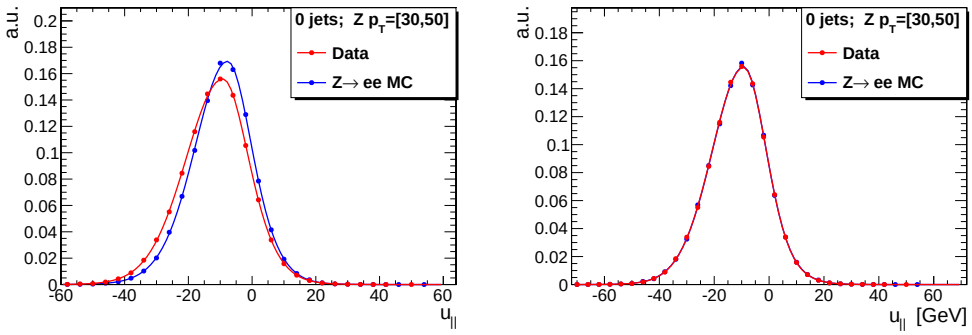

Figure 4.19.: Distributions of the MVAMET after preselection of di-electron events before (left) and after (right) applying the recoil corrections via isomorphic mapping in the bin q_T^Z =[30-50] and for events with no jets above 30GeV/c.

GeV. Therefore, in order to reproduce the correct E_T^{miss} measurements observed in data, recoil corrections to simulated physics processes have to be applied.

For the recoil corrections to the dominant background of $Z + jets \rightarrow ee + jets$ events a customized method called "isomorphic mapping" is applied. Events with 60 GeV/c^2 < M_{ee} <120 GeV/c^2 in both MC and data are selected and categorized into bins of the transverse momentum of the Z boson and jet multiplicity, q_T^Z =[0-10], [10-20], [20-30], [30-50], [50-1000] GeV/c for events with 0, 1 or ≥2 jets of p_T >30 GeV/c. Then in each bin the balance of the transverse momentum of the hadronic recoil and the Z boson parallel and perpendicular to the Z boost, i.e. $u_{||} = \hat{u}_{||} - q_T^Z$ and $u_\perp = \hat{u}_\perp$, $\hat{u}_{||(\perp)}$ is the hadronic recoil parallel(perpendicular) to the Z boost, is fitted both for MC and data where additional backgrounds have been subtracted. The fit function is chosen to be a two sided double Gaussian which is found to suitably describe the shapes of both $u_{||}$ and $u_\perp$, where especially the component parallel to the Z boost shows asymmetric behavior. Then the isomorphic mapping technique (Appendix A) reproduces precisely the E_T^{miss} components $u_{||}$ and $u_\perp$ measured in data for simulated $Z + jets$ events and preserves all correlations of E_T^{miss} with other observables in the event.

As an illustration, Figure 4.19 shows the $u_{||}$ component for events with no jets above 30 GeV/c and a Z boson boost of 30-50 GeV/c. After the isomorphic mapping a perfect agreement can be observed. All other fits are shown in Appendix A.

Recoil corrections for less dominant electroweak backgrounds and as well to the $H \rightarrow \tau\tau$ signal with genuine E_T^{miss} are applied via the "type-II recoil corrections" method explained in [75], where recoil responses and resolutions are fitted and rescaled as a function of the boson(s) p_T. Due to the data-driven modeling of the second largest background of $Z \rightarrow \tau\tau$ events, no E_T^{miss} corrections have to be

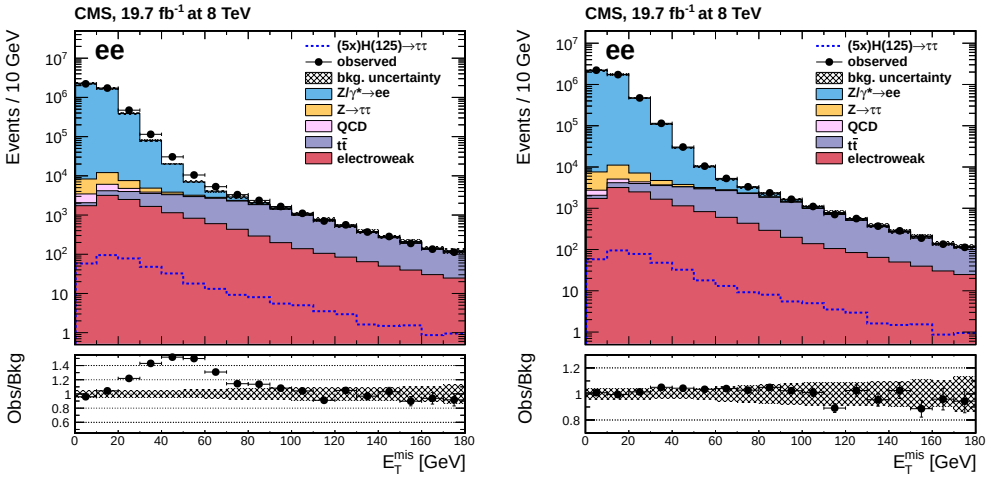

Figure 4.20.: Distributions of the MVAMET after preselection of di-electron events before (left) and after (right) applying the recoil corrections via isomorphic mapping. In the small panels below, the data/MC ratio is displayed.

applied for this sample as it is directly extracted from data and thus incorporates the correct E_T^{miss} response and resolution.

In Figure 4.20 a data-to-MC comparison for the E_T^{miss} distributions is shown. A significant improvement with respect to the uncorrected E_T^{miss} distribution, left plot, for the modeling of E_T^{miss} can be seen in the right plot where the recoil corrections explained above are applied.

4.7. Di-Tau Mass Reconstruction

The reconstructed invariant di-tau mass is a key variable in this analysis to separate $Z \to \tau\tau$ from $H \to \tau\tau$ processes. Due to the distinct undetectable neutrinos in the final states of the leptonic decays of di-tau resonances, the di-tau system is underconstrained in terms of measurable observables and can only be hypothetically reconstructed with additional theoretical assumptions on tau decays and refined algorithms.

Good approaches of di-tau mass reconstruction have a high efficiency for valid physical solutions, result in a good mass resolution (<20%) and are well applicable in terms of CPU-intensity (O(1 event/s)).

4.7.1. SVFit Di-Tau Mass Reconstruction Algorithm

In CMS the *svFit Algorithm* [1] is used for the invariant mass reconstruction for the $H \to \tau\tau$ analysis and will be described in the following.

The leptonic decay of taus is in principle fully described by the measurable four-momentum of the visible decay product p^{vis} and three additional experimentally unconstrained variables chosen to be:

- x, the fraction of the τ-lepton energy in the laboratory frame carried by the visible decay products;
- ϕ, the azimuthal angle of the τ-lepton direction in the laboratory frame;
- $m_{\nu\nu}$, the invariant mass of the two-neutrino system in leptonic τ decays;

From these variables the four-momentum of the initial tau lepton can in principle be fully reconstructed. Additional measurable constraints come from the event by event based measurement of the missing transverse energy $\vec{E}_T^{\text{miss}}$, i.e. the two components E_x^{miss} and E_y^{miss}. Due to the fact that the three variables x, ϕ and $m_{\nu\nu}$ cannot be measured directly the svFit Algorithm uses a maximum Likelihood method to estimate the mass of the di-tau system of $Z/H \to \tau\tau$ decays. Different $M_{\tau\tau}^i$ hypotheses can then be tested by integration of the likelihood kernel over the non fixed phase phase space

$$P(M_{\tau\tau}^i) = \int \delta\left(M_{\tau\tau}^i - M_{\tau\tau}(\vec{y}, \vec{a_1}, \vec{a_2})\right) f(\vec{z}, \vec{y}, \vec{a_1}, \vec{a_2})\, \mathrm{d}\vec{a_1}\, \mathrm{d}\vec{a_2} \tag{4.8}$$

where $\vec{a_1} = (x_1, \phi_1, m_{\nu\nu,1})$ and $\vec{a_2} = (x_2, \phi_2, m_{\nu\nu,2})$ are the unconstrained parameters for the two τ-leptons, and $\vec{z} = (E_x^{\text{miss}}, E_y^{\text{miss}})$ and $\vec{y} = (p_1^{\text{vis}}, p_2^{\text{vis}})$ assemble the constraining parameters of the measured $\vec{E}_T^{\text{miss}}$ and the momenta of the visible decay products. Boundaries of $m_{\nu\nu}$ are taken to be within the physically allowed phase space $0 \leq x \leq 1$ and $0 \leq m_{\nu\nu} \leq m_\tau\sqrt{1-x}$. The test is done for a discrete set of $M_{\tau\tau}^i \in [5, 3000]\text{GeV/c}^2$ in a series of 2.5% steps with respect to the preceding mass in the series. The reconstructed di-tau mass $\hat{M}_{\tau\tau}$ is then taken to be the value of $M_{\tau\tau}^i$ which maximizes the probability 4.8.

The likelihood kernel $f(\vec{z}, \vec{y}, \vec{a_1}, \vec{a_2})$ factorizes into two parts. The first part describes the matrix elements of the two leptonic tau decays [77]:

$$\mathcal{L}_{\tau,\text{l}} = \frac{\mathrm{d}\Gamma}{\mathrm{d}x\, \mathrm{d}m_{\nu\nu}\, \mathrm{d}\phi} \propto \frac{m_{\nu\nu}}{4m_\tau^2}[(m_\tau^2 + 2m_{\nu\nu}^2)(m_\tau^2 - m_{\nu\nu}^2)], \tag{4.9}$$

where the dependence on tau polarization has been neglected.

The second part incorporates the experimentally estimated event by event $\vec{E}_T^{\text{miss}}$ resolution based on the significance measurement of the missing transverse energy [78] [79], combined with the neutrino kinematics:

$$\mathcal{L}_\nu(E_x^{\text{miss}}, E_y^{\text{miss}}) = \frac{1}{2\pi\sqrt{|V|}} \times \exp\left[-\frac{1}{2}\begin{pmatrix} E_x^{\text{miss}} - \sum p_x^\nu \\ E_y^{\text{miss}} - \sum p_y^\nu \end{pmatrix}^T V^{-1} \begin{pmatrix} E_x^{\text{miss}} - \sum p_x^\nu \\ E_y^{\text{miss}} - \sum p_y^\nu \end{pmatrix}\right], \tag{4.10}$$

where the only source of missing transverse energy is assumed to be due to neutrinos from tau decays. V declares the $\vec{E}_T^{\text{miss}}$ covariance matrix [79] and encodes its resolution, $|V|$ denotes its determinant.

The resolution of reconstructed di-tau masses where both taus decay leptonically is 20% but is enhanced for events where the di-tau system is boosted and thus the $\vec{E}_T^{\text{miss}}$ resolution is better, i.e. in the VBF or Boost-high categories defined for this analysis (Section 5.2).

In Figure 4.21 comparisons of the distributions of the di-electron invariant mass spectrum (left), the reconstructed di-tau mass spectrum with the svFit Algorithm (middle) and the reconstructed di-tau mass spectrum for boosted di-tau systems in the VBF category (right) (Section 5.2) for $Z \to \tau\tau \to 2e4\nu$ events and $H \to \tau\tau \to 2e4\nu$ processes with $m_H = 125\text{GeV/c}^2$ are shown. A better separation of the Higgs signal from Drell-Yan di-tau decays can be observed for the svFit Mass spectrum compared to the invariant mass of the di-electron system. As can be seen in the right plot, for events selected in the VBF category the separation power of the svFit reconstructed di-tau mass is further enhanced compared to the inclusively selected sample.

Direct comparisons of the separation power in terms of ROC curves, i.e. fraction yields of signal and background for different cut values on the different mass spectra, are illustrated in Figure 4.22. The svFit mass for boosted Z or Higgs boson productions in the VBF category performs best, followed by the svFit mass in the inclusive sample and the invariant di-electron mass spectrum performs worst.

4.7.2. Collinear approximation

In contrast to the svFit Mass reconstruction algorithm there exist less evolved methods to reconstruct the invariant mass of a di-tau system, where more constraining theoretical assumptions enter the calculation. The *Collinear Approximation* is built on the assumption, that the neutrinos of each of the two tau decays are collinear with the associated visible decay products. This implies that $\theta_{1(2)}^{\text{vis}} = \theta_{1(2)}^{\text{mis}}$ and $\phi_{1(2)}^{\text{vis}} = \phi_{1(2)}^{\text{mis}}$ where $\phi_{1(2)}^{\text{vis}}$ and $\theta_{1(2)}^{\text{vis}}$ are the angles associated to the momentum in polar coordinates of the two visible decay products (electrons) $\vec{p}_{1(2)}^{\text{vis}} = (p_{1(2)}^{vis}, \phi_{1(2)}^{\text{vis}}, \theta_{1(2)}^{\text{vis}})$ of the two tau decays and $\phi_{1(2)}^{\text{mis}}$ and $\theta_{1(2)}^{\text{mis}}$ consolidate the angles of the two neutrinos associated to each of the leptonic tau decays $\tau \to e\nu\nu$.

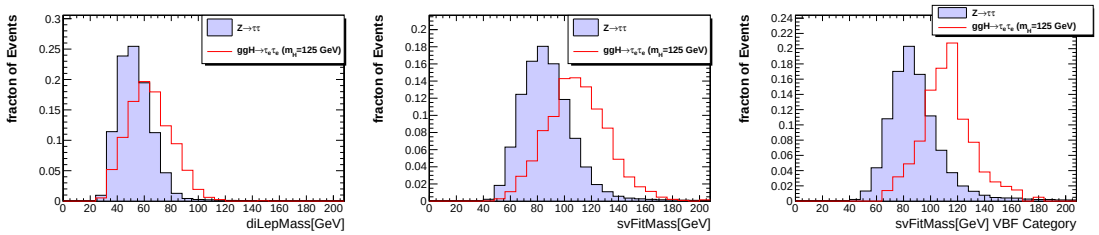

Figure 4.21.: Comparisons of the distribution of invariant di-electron mass (left), svFit reconstructed di-tau mass for non-boosted di-tau systems (middle) and boosted (VBF) di-tau systems (right), for $Z \to \tau\tau \to 2e4\nu$ and $ggH \to \tau\tau \to 2e4\nu$ events with $m_H = 125$ GeV/c^2.

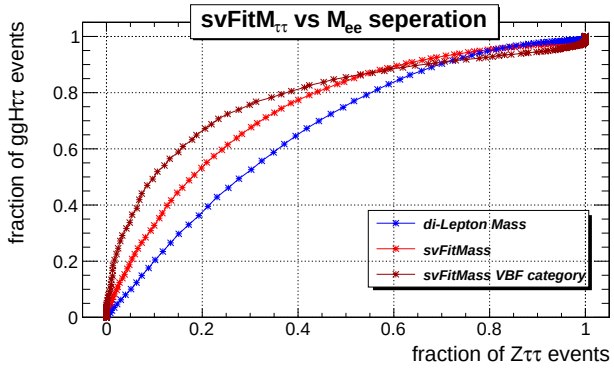

Figure 4.22.: Illustration of discrimination power between $Z \to \tau\tau$ and $H \to \tau\tau$ events, for invariant di-electron mass, svFit reconstructed di-tau mass for boosted (VBF) and non-boosted events.

The absolute values of the momenta of the invisible decay products of each tau,assuming that the only sources of missing transverse energy $\vec{E}_T^{\text{miss}}$ are due to the neutrinos, can then be calculated with the following set of linear equations:

$$\begin{pmatrix} E_x^{\text{miss}} \\ E_y^{\text{miss}} \end{pmatrix} = \begin{pmatrix} \sin\theta_1^{\text{vis}}\cos\phi_1^{\text{vis}} & \sin\theta_2^{\text{vis}}\cos\phi_2^{\text{vis}} \\ \sin\phi_1^{\text{vis}}\cos\phi_{\text{vis}}^1 & \sin\phi_2^{\text{vis}}\cos\phi_2^{\text{vis}} \end{pmatrix} \begin{pmatrix} p_1^{\text{mis}} \\ p_2^{\text{mis}} \end{pmatrix}. \tag{4.11}$$

Solutions for invisible momentum components of the two taus p_1^{mis} and p_2^{mis} can then be written as follows:

$$p_1^{\text{mis}} = \frac{\sin\phi_2^{\text{vis}} E_x^{\text{miss}} - \cos\phi_2^{\text{vis}} E_y^{\text{miss}}}{\sin\theta_1^{\text{vis}}\sin(\phi_2^{\text{vis}} - \phi_1^{\text{vis}})} \tag{4.12}$$

$$p_2^{\text{mis}} = \frac{\cos\phi_1^{\text{vis}} E_y^{\text{miss}} - \sin\phi_1^{\text{vis}} E_x^{\text{miss}}}{\sin\theta_2^{\text{vis}}\sin(\phi_2^{\text{vis}} - \phi_1^{\text{vis}})} \tag{4.13}$$

The invariant di-tau mass calculated with the assumptions of the Collinear Approximation is then $M_{\tau\tau}^{CA} \approx M_{ee}/\sqrt{x_1 x_2}$, where M_{ee} is the invariant mass of the visible decay products, i.e. the two electrons, and $x_{1(2)} = \frac{p_{1(2)}^{vis}}{p_{1(2)}^{vis}+p_{1(2)}^{mis}}$ are the momentum fractions of the taus carried by the corresponding electrons.

As can be seen from the solutions in equation 4.12, the Collinear Approximation exhibits significant ill-defined shortcomings. For $\phi_2^{\text{vis}} - \phi_1^{\text{vis}} \to n\pi$ with $n \in \mathbb{N}$ the solutions for the tau decay associated neutrino momenta $p_{1(2)}^{\text{mis}}$ diverge towards unphysically high values leading to long unphysical tails for the reconstructed di-tau mass $M_{\tau\tau}^{CA}$ with $x_{1(2)} \to 0$. Hence the Collinear Approximation does not provide valid solutions for non-boosted di-tau resonances. Also for events where $\vec{E}_T^{\text{miss}}$ is not in the plane spanned by the momenta of the visible decay products the Collinear Approximation solutions for either p_1^{mis} or p_2^{mis} can end up negative which again is physically non valid.

Whether or not the Collinear Approximation yields positive solutions for both p_1^{mis} and p_2^{mis} can be used to discriminate $Z \to 2e$ and di-tau like events, especially for boosted production modes, and is taken as boolean input variable for the multivariate analysis method used in this analysis described in Section 5.2.

As discussed in Section 4.6, the missing transverse energy due to mis-measurements of hadronic components of the Z boson recoils is centrally shifted and the resolution worse towards the opposite direction of the Z boson boost. Therefore, in addition to the absence of neutrinos, the properties of missing transverse energy due to mis-measured Z recoils leads to effectively negative $p_{1(2)}^{\text{mis}}$ and thus enhances the separation power of the validity of the Collinear Approximation.

4.8. Distance of Closest Approach

The 3-dimensional distance of closest approach significance (DCA) between the two electron tracks is a central variable for this analysis. It serves as a discriminating variable between $Z/\gamma* \rightarrow ee$ and $H/Z \rightarrow \tau\tau \rightarrow 2e4\nu$ processes. Due to the significant mean life-time of taus of $2.91\times10^{-13}s$ or decay-length of about 900 μm for a boost of 20 GeV, the tracks of the two electrons from tau decays appear to be displaced with respect to each other. In particular, the DCA is not affected by the limited resolution of the primary vertex reconstruction, of about 100-200 μm (depending on the number of tracks) in parallel and azimuthal direction from the beam-axis.

The DCA significance is used as an input variable for the boosted decision trees described in Section 5.2, to discriminate prompt Drell-Yan decays into electrons from di-tau events. Furthermore, the DCA significance serves as the basis for the data-driven background estimation of the most dominant background component, namely $Z/\gamma* \rightarrow ee$ processes, as explained in Section 5.4.5.

For Drell-Yan decays into electrons the DCA significance is not well modeled in MC. Corrections are applied using the *isomorphic mapping technique* similar to the the E_T^{miss} recoil corrections. In bins of the opening angle between the two electron tracks and p_T difference as well as sum of the two electrons, the DCA significance of simulated $Z/\gamma* \rightarrow ee$ events is mapped onto the observed DCA significance in data after subtracting all other backgrounds and fitting a two-sided double-gaussian function to both distributions. The variables for the binning are chosen such that corrections to more precise DCA significance measurements corresponding to opening angles closer to π and leptons with higher transverse momentum are disentangled from less precise kinematical phase-spaces. Figure 4.23 shows the impact of the corrections to di-electron events. With the corrections applied, a good description of the observed DCA significance with simulated event processes is obtained.

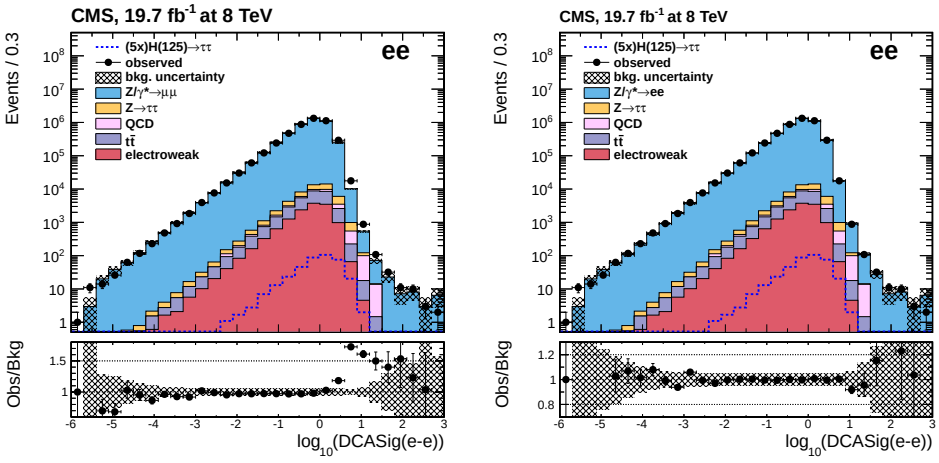

Figure 4.23.: Distributions of the DCA significance after preselection of di-electron events before (left) and after (right) applying isomorphic mapping corrections.Experimental data (8 TeV) are shown as circles, predicted background samples are represented by filled histograms. Also displayed as dashed histograms are the signal expected for the SM Higgs boson with $m_H = 125$ GeV/c^2 and scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show quadratic sum of statistical and systematic uncertainties.

5. Data Analysis and Search Strategy

The purpose of this chapter is to give a detailed overview of the new search strategy for $H \to \tau\tau \to 2e4\nu$ decays. The general physics object reconstruction and selection techniques have been established in Chapter 4.

A sketch of the novel $H \to \tau\tau \to 2e4\nu$ analysis strategy is shown in Figure 5.1, and is divided into the following steps:

- A simple event selection, based on trigger and electron selection requirements, is imposed, to select di-electron events. This is described in Section 5.1;

- To enhance the sensitivity, kinematic features of Higgs production mechanisms are exploited. Mutually exclusive event categories are defined, based on the jet multiplicity, to consider the distinct VBF two-jet event topology and several kinematic features of boosted Higgs decays. Details are discussed in Section 5.2;

- The $H \to \tau\tau \to 2e4\nu$ search suffers from two large backgrounds: $Z \to ee$ and $Z \to \tau\tau \to 2e4\nu$. The two backgrounds have in general different event topologies and the $Z \to ee$ background is by about two order of magnitude larger than the $Z \to \tau\tau$ background. In total, 4 boosted decision trees (BDTs) are constructed: two specifically trained for events in the VBF category, and two for the Boost and 0Jet categories. In both cases, one BDT separates $Z \to ee$ from $Z/H \to \tau\tau$ events, and the other one separates $Z \to \tau\tau$ from $H \to \tau\tau$ events. Detailed information is given in Section 5.2.1 and 5.2.2;

- In each of the 5 categories, a one-dimensional final discriminant is constructed, based on the two corresponding BDTs. The final discriminants combine the two corresponding BDTs, maintaining the separation power against both backgrounds, $Z \to ee$ and $Z \to \tau\tau$. By using one-dimensional final discriminants, good transparency and control regarding effects of systematic uncertainties and various background predictions is established. (The more aggressive approach of using the unfolded two-dimensional correlation of the two BDTs directly for signal extraction, may be studied and used in the future.) The final discriminant is discussed in Section 5.3;

- Data-driven background estimation techniques (except for di-boson and W+Jets processes) are then deployed to accurately predict the shape and normalization for each SM background process. The specifically developed background estimation method for $Z \to ee$ processes, uses template fits of the the DCA significance observable, previously corrected as described in Section 4.8. $Z \to \tau\tau$ contributions are estimated, utilizing the embedding technique, commonly used across all channels of the CMS $H \to \tau\tau$ analysis. Section 5.4 details the methods for relevant backgrounds;

- Theoretical and experimental systematic uncertainties are addressed, and final discriminator shape-altering or solely normalization related uncertainties identified, Section 5.5;

- Finally, a maximum profile likelihood model is deployed to test the Higgs boson hypothesis with the modified frequentist approach. The final result of the ee-channel alone is achieved, after combining all categories for the 7 and 8 TeV data-taking periods, and deriving expected and observed 95% confidence level upper limits on the Higgs boson cross-section as a function of the mass. The ee-channel is further combined with the all CMS $H \to \tau\tau$ searches. The profile likelihood model and the modified frequentist approach is described in Section 5.6.

Throughout this chapter, plots are shown for the 8 TeV data-taking period only, to facilitate an acceptable and readable structure of the document. Control plots for the analysis of 7 TeV data are shown in Appendix E. In all plots, the SM background include all relevant corrections discussed so far, and are derived, using the background estimation methods discussed in Section 5.4.

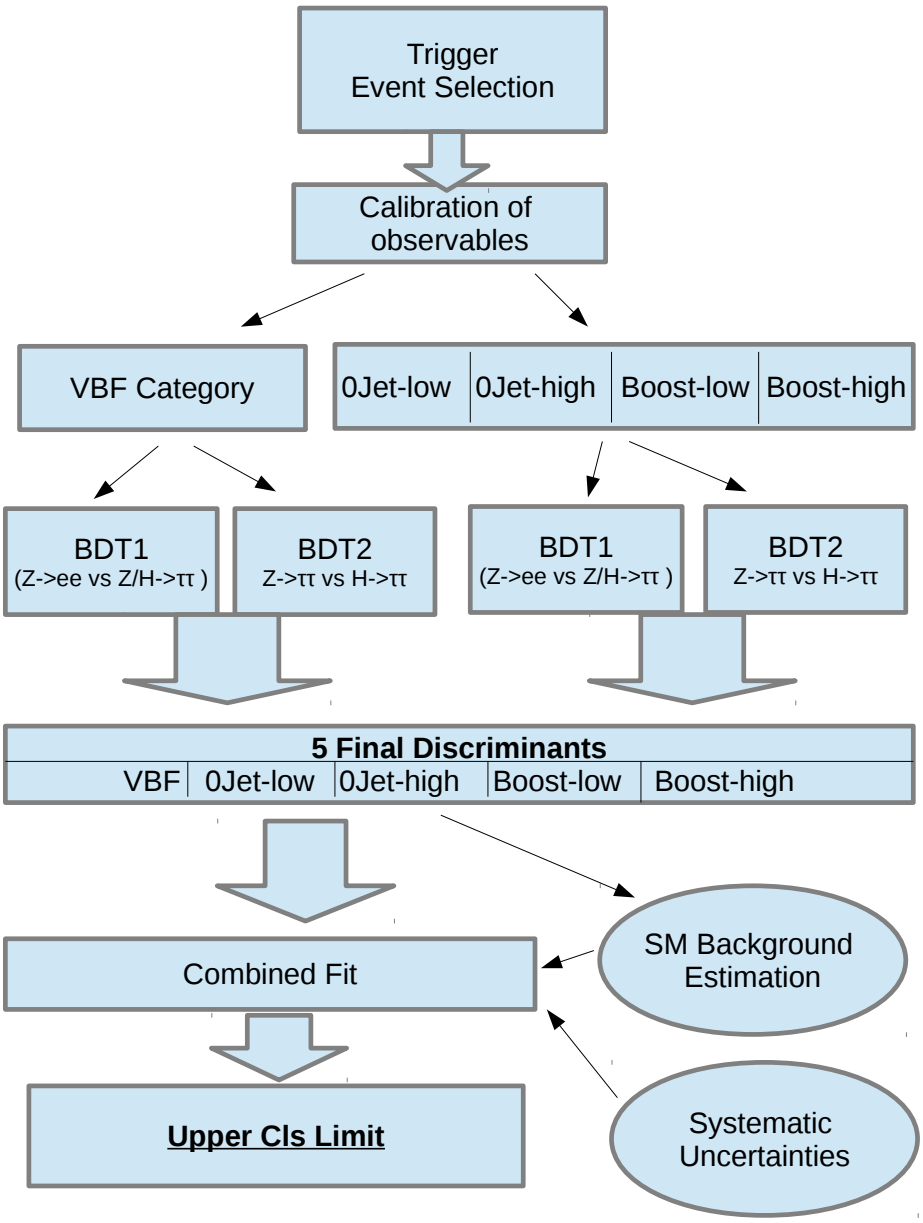

Figure 5.1.: Overview of the design and analysis flow of the CMS $H \to \tau\tau \to 2e4\nu$ analysis.

5.1. Event Selection

Requirements for events to be considered for further analysis are discussed in this section. The object and event selection cuts used for this analysis are chosen to be fully consistent with the object selections across all CMS $H \to \tau\tau$ channels.

The central requirement is that events are required to have two electron candidates of opposite sign, passing the selection criteria described in Section 4.3.3. In each event, the highest p_T electrons are selected. Events are required to have fired the di-electron trigger examined in Section 4.1, and the selected electron candidates are required to be within a cone of 0.5 in $\eta - \phi$-space of the HLT electron trigger objects. The two electron candidates are required to be within $|\eta| < 2.3$ and to have transverse momenta above 20 GeV/c for the leading electron, and higher than 10 GeV/c for the trailing electron.

Jets are selected with the PU Jet ID discussed in Section 4.5.1 and are taken into account if the jet is within the pseudorapidity range of $|\eta| < 4.7$ and has a transverse momentum above 30 GeV/c.

Events are vetoed, if there is a b-tagged jet using the CSV medium working point explained in Section 4.5.1 with a p_T above 20 GeV/c within the pseudorapidity range of $|\eta| < 2.4$, suppressing the $t\bar{t}$ background by approximately 85%.

After the event categorization into mutually exclusive event categories, as explained in Section 5.2, further simple category dependent cuts on the azimuthal angle between the two electrons or di-electron invariant mass are imposed to reject QCD background.

To correct for possible mis-modeling of observables used for event selection, i.e. electron identification variables, electron efficiencies are measured (as discussed in Section 5.1.1), in both the observed data and simulated event samples, to extract scale factors to be applied event-by-event to the simulated samples to improve the precision of the background and signal predictions.

Figure 5.2 shows good data-to-MC agreement for p_T and η spectra of the leading and trailing electrons. In the p_T spectrum of the sub-leading electron, an edge is observed around 45 GeV, which is due to the combinatorics with the electron of higher momentum to have an invariant mass around the Z peak. The step in the η spectra around $\eta \approx 1.5$ is the effect of the suboptimal ECAL coverage in the barrel-to-endcap transition region. Figures 5.3 and 5.4 show good data-to-MC agreement for the number of jets, b-jets (before b-jet veto) as well as kinematical variables of the leading and trailing jets, after the event selection. A slight trend can be observed in the distribution of the jet multiplicities for jet multiplicities larger 5. This however is covered by uncertainties and has no significant effect on the analysis.

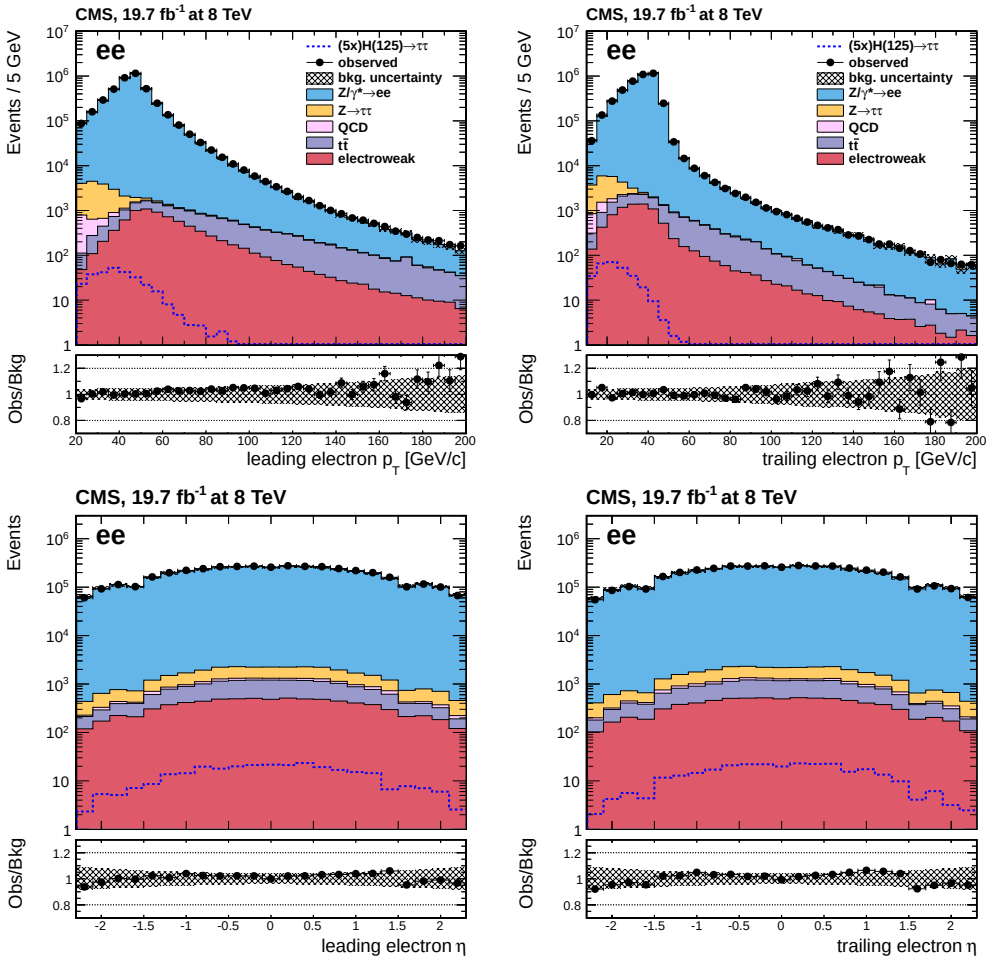

Figure 5.2.: Distributions of the leading (left) and trailing (right) electron transverse momentum (top) and pseudorapidity (bottom). Experimental data (8 TeV) are shown as circles, predicted background samples are represented by filled histograms. Also displayed as dashed histograms are the signal expected for the SM Higgs boson with $m_H = 125\ \mathrm{GeV/c^2}$ and scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

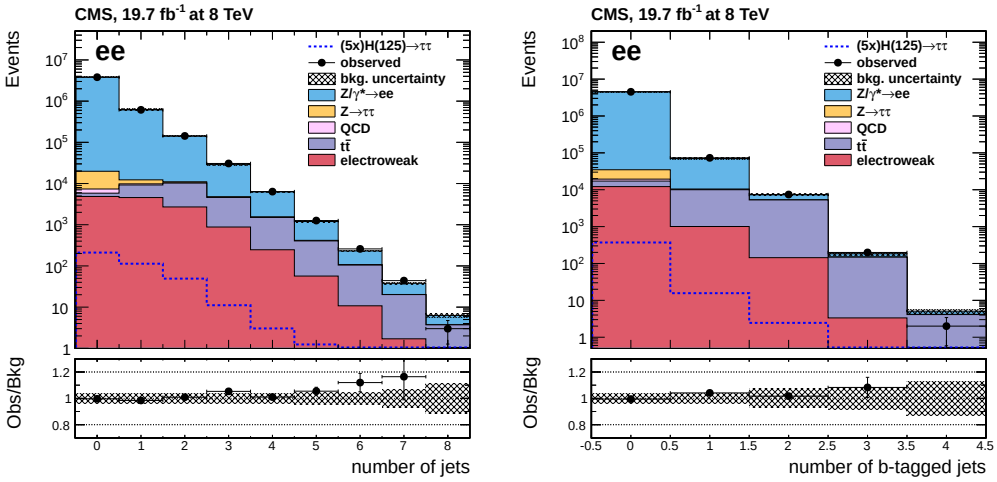

Figure 5.3.: Left plot: multiplicity of jets after preselection of the di-electron events; right plot: multiplicity of b-tagged jets.Experimental data (8 TeV) are shown as circles, predicted background samples are represented by filled histograms. Also displayed as dashed histograms are the signal expected for the SM Higgs boson with $m_H = 125$ GeV/c^2 and scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

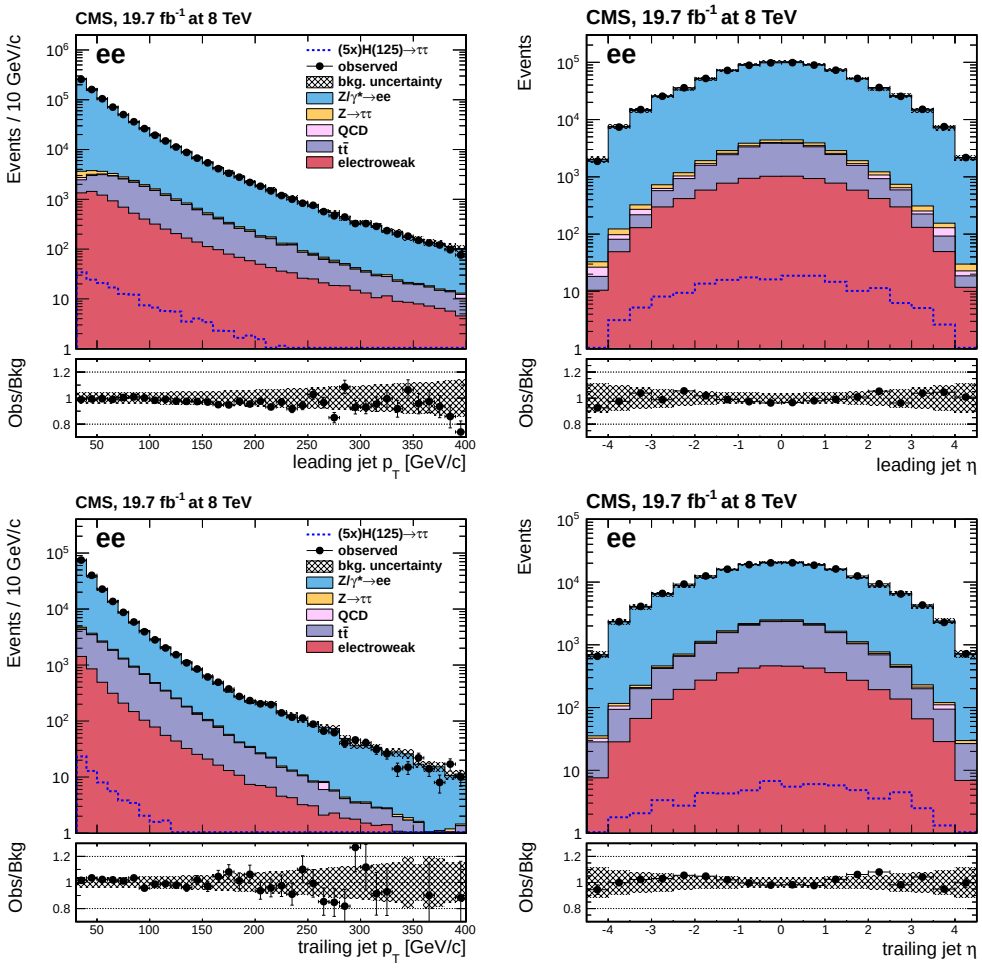

Figure 5.4.: Left plots: leading jet p_T (upper) and subleading jet p_T (lower) distributions after preselection of the di-electron events. Right plots: leading jet η (upper) and subleading jet η (lower) after preselection of the di-electron events. Experimental data (8 TeV) are shown as circles, predicted background samples are represented by filled histograms. Also displayed as dashed histograms are the signal expected for the SM Higgs boson with $m_H = 125$ GeV/c^2 and scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

5.1.1. Event Selection Efficiency

To achieve a precise prediction of the expected signal and background yields in this analysis, the event selection efficiency must be accurately measured and possible differences thereof for simulated Monte Carlo events have to be corrected. Although Monte Carlo simulations of the interaction of the particles with the detector material with the GEANT4 program [80] give excellent approximations of the real experimental situation in general, slight inaccuracy of the simulated detector model can appear for certain phase space regions where the tuning is imperfect. In addition small irregular time-dependent detector changing effects may not be incorporated in the simulation.

The event selection for this analysis is primarily based on the trigger and electron selection requirements. Therefore, a sophisticated method to determine the electron associated selection efficiencies is pursued. The total selection efficiency factorizes into the reconstruction, identification, isolation and trigger efficiency:

$$\epsilon_{tot} = \epsilon_{electron \to \mathrm{reco}} \times \epsilon_{\mathrm{reco} \to (\mathrm{ID+Iso})} \times \epsilon_{(\mathrm{ID+Iso}) \to \mathrm{trigger}}, \tag{5.1}$$

For $\epsilon_{electron \to \mathrm{reco}}$, the efficiency consists of two parts. First the ECAL supercluster reconstruction efficiency which is found to be 99.5% and second the electron associated track reconstruction efficiency shown in Figure 4.5 found to be $> 85\%$. No efficiency difference for data and Monte Carlo simulation is assumed.

Electron Selection Efficiency

To measure the electron selection and trigger efficiency a *Tag-And-Probe* method is used. Events with one electron passing tight selection criteria ("tag") are selected if there is another electron primitive ("probe") with no particular selection criteria imposed and if the invariant mass of the two is close to the mass of the Z boson. Using the dataset collected by single-electron-triggers fired by the tag electron ensures the absence of a priori biases for the probe electron. The tag electron suppresses background to the largely dominating $Z \to ee$ and the probe serves as electron candidate which can either pass or fail the selection criteria in question. By measuring the number of events for passing and failing probes, the efficiency of the selection step can be estimated. This is done by fitting the invariant mass spectrum from 60-120 GeV/c^2 of the tag and probe pair in bins of p_T and η of the probe assuming the fixed Breit-Wigner function convoluted with a Crystal-Ball function to model $Z \to ee$ signal and a product of an exponential and a linear function to model the invariant mass spectrum of background processes. The Breit-Wigner function is fixed to the mass and width of the Z boson (PDG [27]) and the parameters of the Crystal-Ball fit function are floating to allow modeling of energy resolution effects.

This procedure is done both for data and Monte Carlo and electron phase-space

2011 data		
p_T [GeV/c]	$\|\eta\|$	Data/MC Scale Factor
$10 - 15$	$0.0 - 1.2$	1.04
$10 - 15$	$1.2 - 2.3$	0.98
$15 - 20$	$0.0 - 1.2$	0.96
$15 - 20$	$1.2 - 2.3$	1.15
> 20	$0.0 - 1.2$	0.99
> 20	$1.2 - 2.3$	1.02

Table 5.1.: Electron Id and Isolation efficiency scale factors (2011 dataset), in the barrel and endcap regions for bins in electron p_T. Uncertainties are discussed in Section 5.5.2.

(p_T and η) dependent MC-to-data scale factors are extracted from the ratio of the relevant electron selection step efficiency measured in data and Monte Carlo. The scale factors are then applied to each Monte Carlo simulated sample to correct for possible biases explained in the beginning of this section. For illustration, fitted tag and probe invariant mass distributions for failing and passing probes are shown in Figures 5.5.

To ensure proper corrections to electron phase-space dependent effects, the binning for the electron probe for 2012 was chosen to be: in p_T 10-15 GeV/c, 15-20 GeV/c, 20-25 GeV/c, 25-30 GeV/c, 30-35 GeV/c, 35-50 GeV/c and $|\eta|$ <0.8, 0.8< $|\eta|$ <1.479, 1.479< $|\eta|$ <2.3. The estimated efficiencies and scale factors applied to correct simulated events are shown in Table 5.2 for 2012 8 TeV data and in Table 5.1 for 2011 7 TeV data.

Trigger Efficiency

Subsequent to the measurement of the efficiency of the electron identification and isolation $\epsilon_{\text{reco}\to\text{ID+Iso}}$, the efficiency for correctly identified and isolated electrons to fire the trigger $\epsilon_{\text{ID+Iso}\to\text{trigger}}$ is determined. The trigger requirement is omitted for simulated Monte Carlo events such that the trigger efficiency measurement solely performed on data directly translate into scale factors to be applied to simulated events. As a consequence of the electron quality requirements and p_T thresholds of the asymmetric di-electron trigger the efficiency is again p_T and η dependent for both electrons. In the following $\epsilon^{high}_{+(-)} = \epsilon^{high}_{+(-)}(p_{\text{T}}^{+(-)}, \eta^{+(-)})$ shall denote the efficiency of the positive (negative) electron to fire the high-p_T Ele17-leg of the di-electron trigger and $\epsilon^{low}_{+(-)} = \epsilon^{low}_{+(-)}(p_{\text{T}}^{+(-)}, \eta^{+(-)})$ the efficiency of the positive (negative) electron to fire the low-p_T Ele8-leg. Both efficiencies are measured as functions of p_T and η of the corresponding electron. In particular, a matching of offline selected electrons to online electron primitives passing the HLT trigger-path

2012 data				
p_T [GeV/c]	$\|\eta\|$	Data Efficiency	MC Efficiency	Data/MC Scale Factor
$10 - 15$	$0.0 - 0.8$	0.34	0.44	0.77
$10 - 15$	$0.8 - 1.5$	0.35	0.45	0.78
$10 - 15$	$1.5 - 2.3$	0.12	0.20	0.60
$15 - 20$	$0.0 - 0.8$	0.51	0.62	0.83
$15 - 20$	$0.8 - 1.5$	0.52	0.61	0.85
$15 - 20$	$1.5 - 2.3$	0.25	0.35	0.72
$20 - 25$	$0.0 - 0.8$	0.65	0.73	0.88
$20 - 25$	$0.8 - 1.5$	0.56	0.65	0.86
$20 - 25$	$1.5 - 2.3$	0.29	0.38	0.75
$25 - 30$	$0.0 - 0.8$	0.71	0.80	0.89
$25 - 30$	$0.8 - 1.5$	0.64	0.73	0.88
$25 - 30$	$1.5 - 2.3$	0.38	0.46	0.82
$30 - 35$	$0.0 - 0.8$	0.79	0.85	0.93
$30 - 35$	$0.8 - 1.5$	0.73	0.78	0.93
$30 - 35$	$1.5 - 2.3$	0.45	0.54	0.83
> 35	$0.0 - 0.8$	0.86	0.90	0.96
> 35	$0.8 - 1.5$	0.83	0.87	0.95
> 35	$1.5 - 2.3$	0.61	0.67	0.91

Table 5.2.: Electron Id and Isolation efficiency in data and MC and corresponding scale factors (2012 dataset), in the barrel and endcap regions for bins in electron p_T. Uncertainties are discussed in Section 5.5.2.

2011 Ele8 leg Efficiency		
p_T [GeV/c]	Barrel	Endcap
10−15	0.80	0.77
15−20	0.91	0.90
>20	0.97	0.97

Table 5.3.: 2011 dataset. HLT_Ele17_*_Ele8_*_v* trigger : efficiency of the Ele8 leg, in the barrel and endcap regions for bins in electron p_T. Uncertainties are discussed in Section 5.5.2.

2011 Ele17 leg Efficiency		
p_T [GeV/c]	Barrel	Endcap
10−15	0.26	0.41
15−20	0.81	0.83
>20	0.96	0.96

Table 5.4.: 2011 dataset. HLT_Ele17_*_Ele8_*_v* trigger : efficiency of the Ele17 leg, in the barrel and endcap regions for bins in electron p_T. Uncertainties are discussed in Section 5.5.2.

filter requirements of the corresponding double-trigger leg is performed.

The di-electron trigger efficiency can then be calculated from the efficiency of the two trigger-legs:

$$\begin{aligned}\epsilon_{\text{trig}}(p_{\text{T}}^{+},\eta^{+},p_{\text{T}}^{-},\eta^{-}) &= 1-(\bar{\epsilon}_{+}^{high}\cdot\bar{\epsilon}_{-}^{high}+\bar{\epsilon}_{+}^{low}\cdot\epsilon_{-}^{high}+\bar{\epsilon}_{-}^{low}\cdot\epsilon_{+}^{high}) \\ &= \epsilon_{+}^{high}\cdot\epsilon_{-}^{low}+\epsilon_{-}^{high}\cdot\epsilon_{+}^{low}-\epsilon_{+}^{high}\cdot\epsilon_{-}^{high}\end{aligned} \tag{5.2}$$

where, $\bar{\epsilon}=(1-\epsilon)$. The efficiency $\epsilon_{\text{trig}}(p_{\text{T}}^{+},\eta^{+},p_{\text{T}}^{-},\eta^{-})$ is then the scale factor applied to simulated events with two isolated electrons satisfying the identification criteria described in Section 4.3.3.

For 2011 data where a slightly looser di-electron trigger was used until run 170053 the trigger efficiency of both legs is measured as the weighted average of the corresponding data taking periods.

The di-electron trigger-leg efficiencies are presented in Tables 5.3-5.4 for 2011 data and 5.5-5.6 for 2012 data.

The effect of background contamination on the trigger efficiency after the electron identification requirements is found to be negligible. Therefore, the efficiency can simply be estimated from counting events for failing and passing probes.

2012 Ele8 leg Efficiency			
p_T [GeV/c]	$0.0 < \eta \leq 0.8$	$0.8 < \eta \leq 1.5$	$1.5 < \eta \leq 2.3$
10−15	0.78	0.78	0.77
15−20	0.89	0.91	0.90
20−25	0.94	0.95	0.96
25−30	0.96	0.97	0.97
>30	0.97	0.98	0.98

Table 5.5.: 2012 dataset. HLT_Ele17_*_Ele8_*_v* trigger : efficiency of the Ele8 leg, in the barrel and endcap regions for bins in electron p_T. Uncertainties are discussed in Section 5.5.2.

2012 Ele17 leg Efficiency			
p_T [GeV/c]	$0.0 < \eta \leq 0.8$	$0.8 < \eta \leq 1.5$	$1.5 < \eta \leq 2.3$
10−15	0.00	0.00	0.00
15−20	0.54	0.33	0.63
20−25	0.94	0.94	0.97
25−30	0.96	0.98	0.99
>30	0.97	0.98	0.99

Table 5.6.: 2012 dataset. HLT_Ele17_*_Ele8_*_v* trigger : efficiency of the Ele17 leg, in the barrel and endcap regions for bins in electron p_T. Uncertainties are discussed in Section 5.5.2.

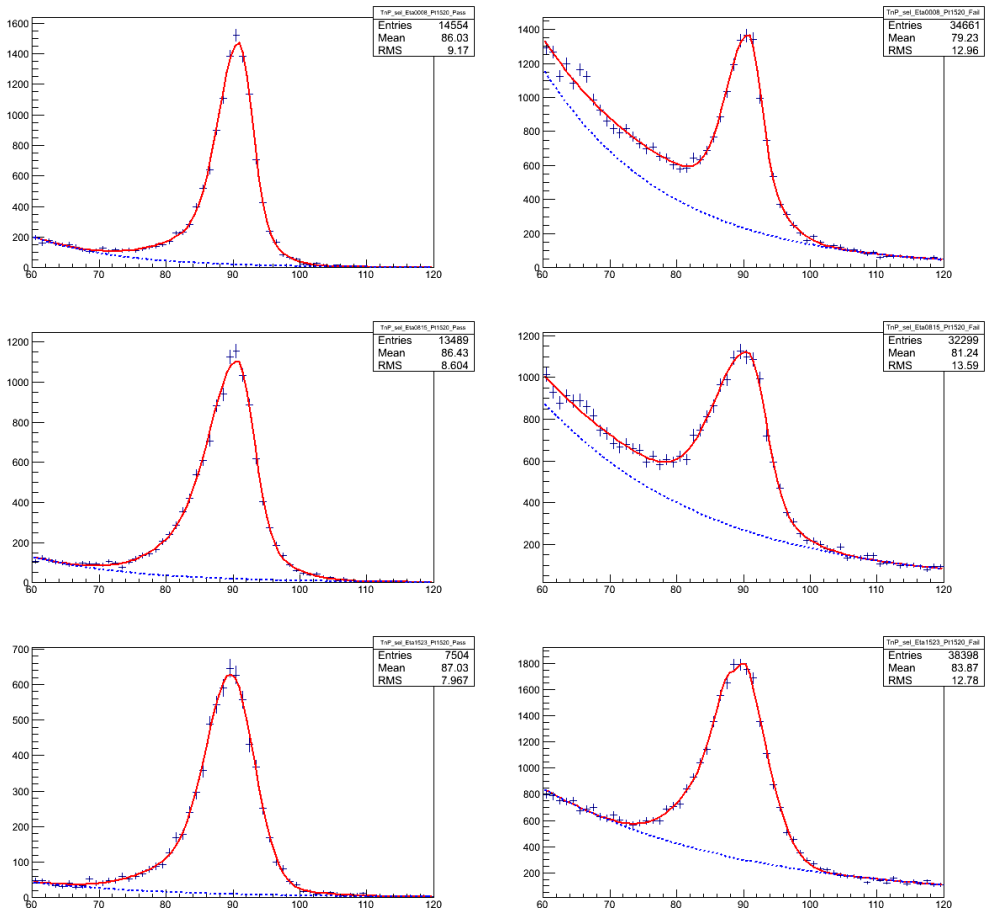

Figure 5.5.: Illustration of fits of the tag and probe invariant mass in bins of passing (left) and failing (right) probe p_T =15-20 GeV/c and $|\eta|$ <0.8, 0.8< $|\eta|$ <1.479, 1.479< $|\eta|$ <2.3 (top to bottom). The data spectra are fitted with the sum of a signal (solid line) and background (dashed line) model.

5.2. Event Categorization and Multivariate Analysis

After the event selection, an event categorization is imposed to enhance the sensitivity by exploiting the distinct event topologies of the dominant Higgs production mechanisms, with examples of Feynman diagrams shown in Figure 2.4. The key variable is the number of jets with a p_T of more than 30 GeV/c in the event, shown in Figure 5.3. The categories are defined such that the sensitive categories with enhanced SM Higgs signal are controlled via background dominated categories. For the Higgs production via vector boson fusion, two high energetic jets in the forward regions of the detector are expected and therefore, a dedicated two jet category is defined which represents the most sensitive category. In addition, Higgs production via gluon fusion is on average more likely to be accompanied with high energetic initial state radiation compared to $q\bar{q} \to Z$ processes. Furthermore, the presents of recoiling jets results in considerably boosted Z or Higgs boson production and hence a good resolution for the reconstruction of the di-tau invariant mass. These two facts are used to define the *boost* category.

- **VBF:** As explained above, the first event category is intended to exploit the production of Higgs bosons via vector boson fusion (VBF). This category is defined for events with 2 or more jets with $p_T > 30$ GeV/c within $|\eta| < 4.7$. Furthermore, no jet with $p_T > 30$ GeV/c in the rapidity gap between the two leading jets are required. The last condition is known as the central jet veto (CJV) and makes use of the suppressed central hadronic activity in vector boson fusion processes where no color-flow between the initial partons of the hard interaction (e.g. [81]). By requiring $m_{ee} > 30$ GeV/c^2, QCD background is suppressed, .

- **Boost:** This event category is intended to exploit the production of a high p_T Higgs boson, recoiling against high p_T jets. The higher p_T of the Higgs boson leads to the selection of an event with higher E_T^{miss} in the hard scattering process, which due to the better precision of the E_T^{miss} measurement improves the reconstruction of the invariant di-τ mass. In addition, it allows to distinguish the Higgs boson signal from the irreducible background from Drell-Yan production of tau pairs, which is expected to have a softer p_T spectrum. In this category, at least one jet with $p_T > 30$ GeV/c within $|\eta| < 4.7$ is required. The event is required not to be part of the *VBF* event category. To reject QCD background, where opposite sign electrons result from decay chains heavy flavor hadrons, only events with $m_{ee} > 30$ GeV/c^2 are selected.

- **0Jet:** All selected events that are not part of any other event category described above are collected in this event category. It only contains events without jets with $p_T > 30$ GeV/c within $|\eta| < 4.7$ and without b-tagged

jets with $p_T > 20$ GeV/c. QCD background is suppressed by requiring $\Delta\Phi(e^+, e^-) > 2.0$ to reject boosted low-mass hadron resonances.

Event overlap across categories is prevented, i.e. the categories are mutually exclusive. The *Boost*, and the *0Jet* event category are further split into two subcategories depending on the p_T of the leading lepton. Events are classified into a high-p_T lepton category if the p_T of the leading lepton exceeds 35 GeV/c. Otherwise, the event is classified into the low-p_T lepton category. In this subcategorization, the high p_T lepton class is introduced for a better discrimination against the irreducible background from Drell-Yan production of tau pairs. Thus in total five categories are defined: **VBF** , **Boost-high**, **Boost-low**, **0Jet-high** and **0Jet-low**.

In each of the five categories a dedicated signal to background discrimination via *boosted decision trees* is applied on which the final combined statistical analysis for signal extraction is based.

The event yields in each category for data as well as expected background and signal production processes are shown in Table 5.7. Predicted background yields per category sub-divided into SM processes are shown in Table 5.8. In both tables, the background event yields include the results of the background predictions described in Section 5.4, and subsequent corrections from the background-only maximum likelihood fit, discussed in Section 5.6 (pulls of the fit can be found in Appendix F).

Table 5.7.: Predicted and observed event yields in all event categories. The signal processes are normalized to the expected Standard Model Higgs production for $m_H = 125$ GeV/c^2 and the predicted background yields represent the result of the global-fit of the final discriminant (described in Section 5.6). Errors are extracted from the fit.

Event category	ggH	VBF	VH	Σsignal	Background	Data
0Jet-low p_T 7 TeV	3.6	0.0	0.1	3.7 ± 0.5	190990 ± 965	190890
0Jet-low p_T 8 TeV	14.1	0.2	0.3	14.5 ± 2.2	519423 ± 932	519376
0Jet-high p_T 7 TeV	3.9	0.0	0.5	4.4 ± 0.6	819965 ± 1781	820035
0Jet-high p_T 8 TeV	22.0	0.3	2.4	24.8 ± 3.4	3225021 ± 2061	3225144
Boost-low p_T 7 TeV	1.5	0.2	0.1	1.8 ± 0.2	10284 ± 101	10300
Boost-low p_T 8 TeV	4.6	0.6	0.3	5.5 ± 0.7	26557 ± 182	26604
Boost-high p_T 7 TeV	2.4	0.4	0.6	3.3 ± 0.4	144905 ± 732	144945
Boost-high p_T 8 TeV	11.6	1.9	3.2	16.7 ± 1.8	560110 ± 1948	560104
VBF 7 TeV	1.6	0.6	0.4	2.6 ± 0.4	35801 ± 282	35796
VBF 8 TeV	5.0	2.7	1.6	9.4 ± 1.1	140146 ± 1302	140070

Table 5.8.: Predicted background yields in all event categories, separately for each SM process for 7 and 8 TeV. Errors show the statistical uncertainties only. VV+(W+Jets) includes WW, WZ, ZZ and W+Jets processes, also denoted Di-boson in most of the plots.

Event category	$Z \to ee$	$Z \to \tau\tau$	$t\bar{t}$	(VV)+(W+jets)	QCD
0Jet-low p_T 8 TeV	509177 ±714	9201 ± 96	11 ± 3	248 ± 16	785 ± 28
0Jet-high p_T 8 TeV	3217880± 1794	3117 ± 56	324 ± 18	3423 ± 59	277 ± 17
Boost-low p_T 8 TeV	24923 ±158	1173 ± 34	109 ± 10	156 ± 12	197 ± 14
Boost-high p_T 8 TeV	552354± 743	1146 ± 34	1936 ± 44	4442 ± 67	232 ± 15
VBF 8 TeV	134454± 367	597 ± 24	2114 ± 46	2911 ± 54	70 ± 8
0Jet-low p_T 7 TeV	189283± 435	3426 ± 59	3 ± 2	74 ± 9	60 ± 8
0Jet-high p_T 7 TeV	818258± 905	922 ± 30	53 ± 7	723 ± 27	8 ± 3
Boost-low p_T 7 TeV	9791 ±99	409 ± 20	28 ± 5	41 ± 6	16 ± 4
Boost-high p_T 7 TeV	143537± 379	316 ± 18	336 ± 18	700 ± 26	16 ± 4
VBF 7 TeV	34773 ±187	179 ± 13	416 ± 20	427 ± 21	6 ± 2.5

5.2.1. Multivariate Analysis in the VBF Category.

After the event categorization, the VBF category is still largely dominated by the two major backgrounds $Z/\gamma^* \to ee$ and $Z \to \tau\tau \to 2e4\nu$. However, a set of variables with considerable discrimination power between Higgs signal and background processes has been identified (partly in [20]) and combined into several boosted decision trees for optimal discrimination.

The event topologies and kinematics are significantly different for both major background processes. Therefore, two separate BDTs were trained to first optimally discriminate the statistically overwhelming $Z/\gamma^* \to ee$ background from $Z/H \to \tau\tau \to 2e4\nu$ events ($\mathbf{BDT_1}$), and then secondly to discriminate $Z \to \tau\tau \to 2e4\nu$ from $H \to \tau\tau \to 2e4\nu$ processes ($\mathbf{BDT_2}$). The set of input variables for both BDTs is as follows where the invariant di-electron mass is only used in BDT_1 and the reconstructed di-tau mass is only used in BDT_2 as input variable

- m_{jj}, the invariant mass of the two leading jets;
- $\Delta\eta_{jj}$, the pseudorapidity difference between the two leading jets;
- E_T^{miss}, the missing transverse energy;
- DCASig($2e$), the electron distance of closest approach (DCA) significance;
- $\Delta\Phi(e^+, p_T^{miss})$, the azimuthal angle between direction of the positively charged lepton three-momentum and the missing transverse energy;
- the validity of collinear approximation, it is set to one if the collinear approximation for the reconstruction of tau lepton pair kinematics yields a physical solution for the neutrino energies, otherwise the variable takes the value 0. This variable is described in detail in Section 4.7.2;
- the invariant mass of the electron pair (used only in the training of the BDT_1);
- the di-τ invariant mass reconstructed with the SVFit algorithm, as described in Section 4.7 (only used in the training of the BDT_2).

Distributions of these variables for events selected in the *VBF* category are shown in Figures 5.8 and 5.9, where predicted background contributions are compared to 8 TeV data. 7 TeV control plots are shown in Appendix E. Overall good agreement between data and predicted SM processes can be observed. The correlation between the input observables can be monitored using Profile plots, i.e. mean and RMS of variable a in bins of variable b. Data to background prediction comparisons are shown in Appendix C.

m_{jj} and $\Delta\eta_{jj}$ encode the correlation of the two highest energetic jets which govern good separation power for the two jets associated with the VBF production and di-jets from higher order gluon initiated ISR [82]. For Higgs productions from gluon fusion processes leaking into the VBF category the theoretical uncertainties for the two tagged jets from higher order gluon radiations are considerably large and treated in Section 5.5.8.

E_T^{miss} inevitably separates signal from background well due to the 4 neutrino final state of the boosted Higgs decay compared to E_T^{miss} resulting from mis-measurements of the hadronic components in Z+jets final states. This variable introduces non-negligible systematic uncertainties but is used nevertheless due to its separation power in this category.

$\Delta\Phi(e^+, p_T^{miss})$ describes the azimuthal correlation of the electron momentum direction and the missing transverse energy direction. The boosted electronically decaying di-tau system leads to kinematically correlated electron momenta and missing transverse energy from the associated neutrinos and as a consequence to on average smaller azimuthal separation of E_T^{miss} and the positive (negative) electron compared Z+jets di-electron final states where missing transverse energy is only produced via mis-measurements of the hadronic recoil being asymmetric with respect to the Z boost direction.

Because of the boosted Higgs production in the VBF category the number of physical solutions of the collinear approximation for the di-tau system, where the missing transverse energy direction of the 4 neutrinos culminated with the electrons is collinear to the transverse momentum of the two electrons, is greatly enhanced. For $Z/\gamma^* + 2jets \to ee + 2jets$ processes, where there is no genuine missing transverse energy and the E_T^{miss} resolution is asymmetric as well as centrally shifted in the opposite direction with respect to p_T of the Z boson, the number of physical solutions is reduced on average by a factor of $\Delta\alpha/2\pi$ with respect to non-boosted Z boson production, with $\Delta\alpha$ being the average of the opening angle between the two electrons of $Z/\gamma^* + 2jets \to ee$ decays in the VBF category.

The invariant mass separates well $Z/\gamma^* \to ee$ from $Z/H \to \tau\tau \to 2e4\nu$ events where in the latter a sizable amount of energy is carried away by the 4 neutrinos. The di-electron invariant mass spectrum for di-tau events is broadened and distributed towards energies less than 80 GeV/c^2. On the other hand due to the neutrinos in the final state the separation of $Z \to \tau\tau$ from $H \to \tau\tau$ events is suboptimal which is why this variable is only used for BDT_1.

The reconstructed di-tau mass via the SVFit algorithm has good separation power between $Z \to \tau\tau$ and $H \to \tau\tau$, as discussed in Section 4.7, whereas between $H \to \tau\tau$ and $Z/\gamma^* \to ee$ the separation is less efficient. This is complementary to the case of the di-electron invariant mass and therefore in order to preserve good discrimination the reconstructed di-tau mass is only used in BDT_2.

Training samples for the two BDTs are kept complementary to the set of simulated events used for statistical inference. For the signal training sample a mixture

of qqH samples with masses 110 - 145 GeV/c^2 is used. Both for background and signal, only events satisfying the VBF category requirements are used for training. The adaptive boosting technique with 800 trees is used and the number of training events is above 16000.

For illustration the shape of both BDT discriminator outputs is shown in Figure 5.6, where qualitatively BDT_1 shows good discrimination between $Z/\gamma^* \to ee$ and $Z/H \to \tau\tau$ and BDT_2 between $Z \to \tau\tau$ and $H \to \tau\tau$. Data to predicted background comparisons of both BDTs for the 8 TeV dataset are shown in Figures 5.7.

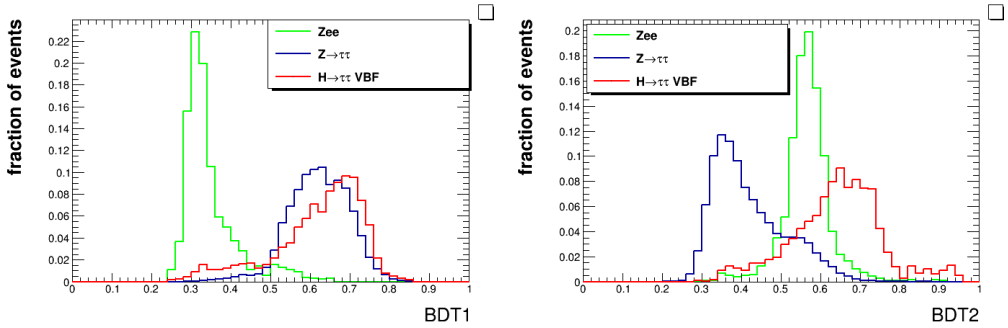

Figure 5.6.: Shapes for BDT_1 and BDT_2 of simulated samples for the two major backgrounds $Z/\gamma^* \to ee$ (green) and $Z \to \tau\tau$ (blue) and Higgs signal produced via vector boson fusion (red) with $m_H = 125\text{GeV/c}^2$.

Finally, the two BDTs are are combined into a global discriminant described in Section 5.3 for signal extraction and the statistical inference on the Higgs hypothesis.

5.2.2. Multivariate Analysis in the 0Jet and Boost Categories.

Similar to the VBF category two BDTs are trained inclusively for *Boost* and *0Jet* categories. BDT_1 is trained to $Z/\gamma^* \to ee$ background from $Z/H \to \tau\tau \to 2e4\nu$ events and BDT_2 is trained to discriminate between $Z \to \tau\tau$ and $H \to \tau\tau$ events. The set of variables, is as follows:

- The ratio of the transverse momentum of the di-electron system to the scalar sum of positive and negative electron momenta, $|\vec{p}_T^{\,1} + \vec{p}_T^{\,2}|/(\sum p_T(e))$;
- The electron distance of closest approach (DCA) significance, DCASig($2e$);
- The pseudorapidity of the di-electron system, $\eta(2e)$;
- The azimuthal angle between direction of the positively charged electron three-momentum and the missing transverse energy, $\Delta\Phi(\ell^+, p_T^{miss})$;

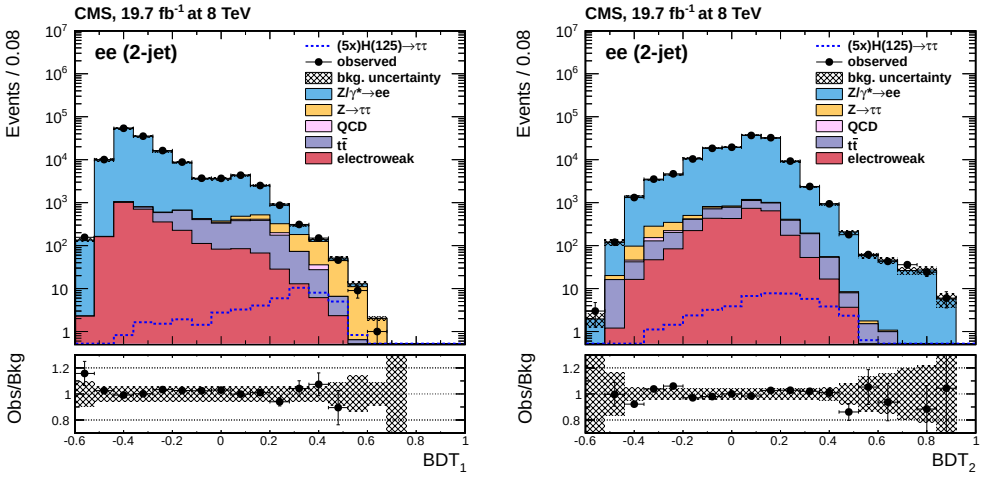

Figure 5.7.: Distributions for BDT_1 (left) and BDT_2 (right) for events in the VBF category. Experimental data (8 TeV) is shown as circles, predicted background samples are represented by filled histograms. Also displayed as dashed histograms is the signal expected for the SM Higgs boson with $m_H = 125$ GeV/c^2 and scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

- The decay angle θ^* of the positively charged electron in the rest frame of the di-electron system;
- The angle ω^* between three-momentum of the positively charged electron and production plane of the di-electron system. The Z boson production plane is defined as the plane spanned by the three-momentum vector of the di-electron system and the proton-beam axis;
- The discrete variable, reflecting validity of collinear approximation for the reconstruction of di-tau kinematics, as described in Section 4.7.2;
- The invariant mass of the electron pair.(used only in the training of the BDT_1);
- the di-τ invariant mass reconstructed with the SVFit algorithm (only used in the training of the BDT_2), described in Section 4.7.

Figures 5.12 to 5.14 show the distributions of these variables for events selected inclusively in the *0Jet-low*, *0Jet-high*, *Boost-low* and *Boost-high* categories, for both, predicted background and observed 8 TeV data. 7 TeV control plots are

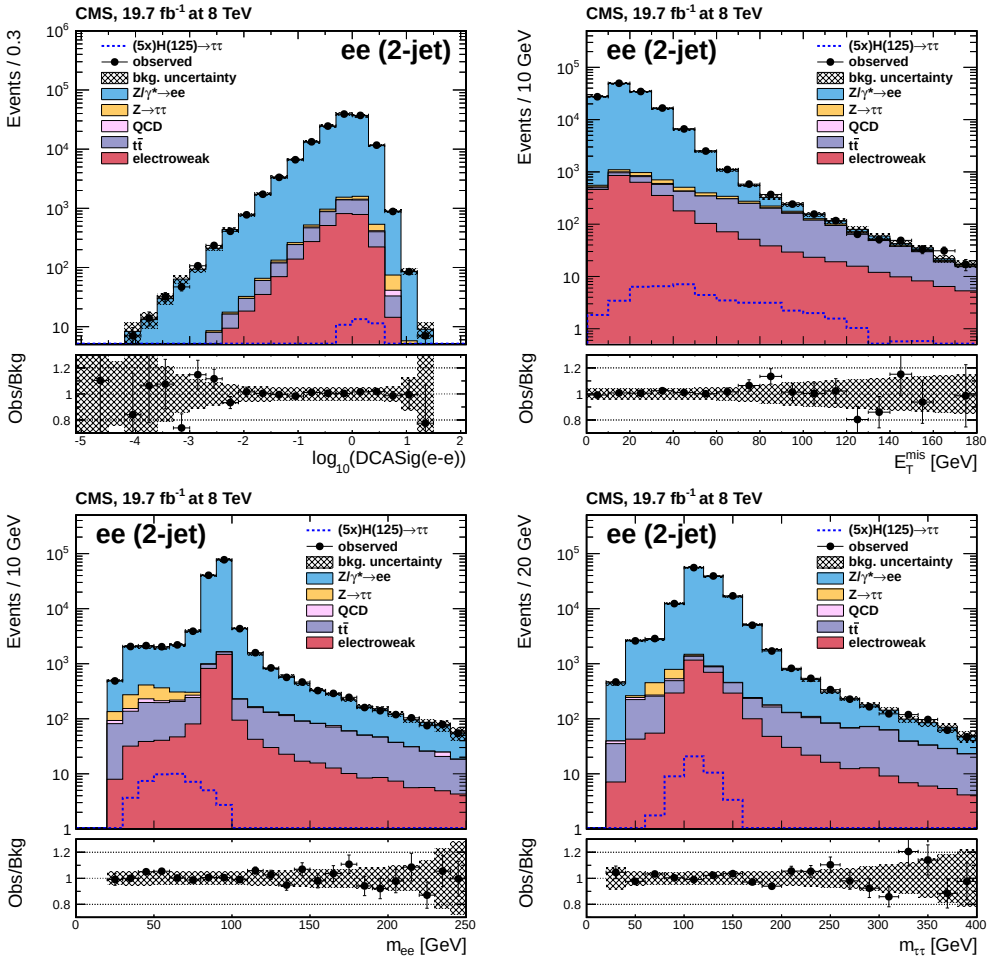

Figure 5.8.: Distributions of variables used in the multivariate analysis in the *VBF* event category in the di-electron channel. Experimental data (8 TeV) are shown as circles, predicted background samples are represented by filled histograms. Also displayed as dashed histograms are the signal expected for the SM Higgs boson with $m_H = 125$ GeV/c^2 and scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

shown in Appendix E. Overall good agreement between data and predicted backgrounds can be observed. The correlation between the input observables can be monitored using Profile plots, i.e. mean and RMS of variable a in bins of variable b. Data to background prediction comparisons are shown in Appendix C.

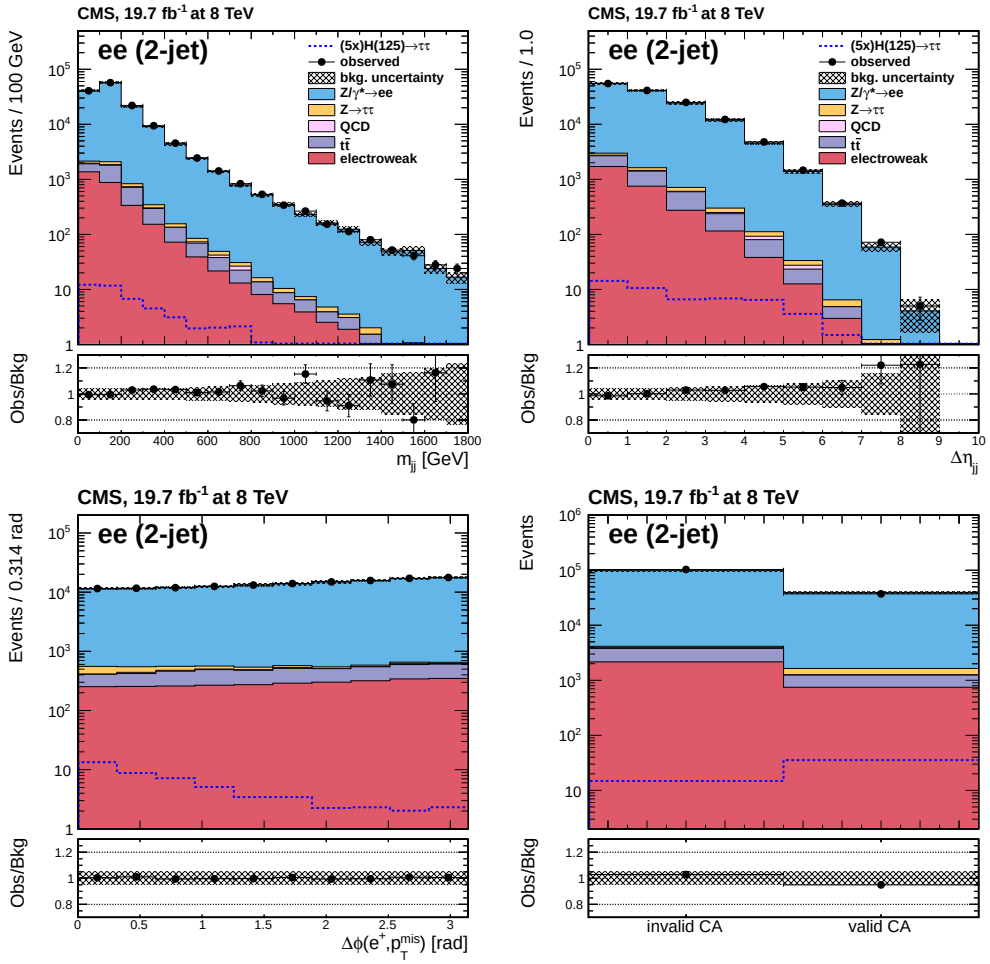

Figure 5.9.: Distributions of variables used in the multivariate analysis in the *VBF* event category. Experimental data (8 TeV) is shown as circles, predicted background samples are represented by filled histograms. Also displayed as dashed histograms are the signal expected for the SM Higgs boson with $m_H = 125$ GeV/c^2 and scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

Data to MC comparisons are shown in Figures 5.10 and 5.11 for BDT_1 and BDT_2 separately for each 0Jet and Boost low/high category.

The jet related variables used in the VBF category are dropped due to the dominant signal production via gluon fusion in these categories where the associated jets do not have distinct topological properties compared to the major backgrounds.

Instead of E_T^{miss} as in the VBF category the ratio of the transverse momentum of the di-electron system to the scalar sum of positive and negative electron momenta is used. Due to on average less boosted Higgs productions from gluon fusion processes, the E_T^{miss} component from energy mis-measurements contributes more substantially to possible genuine E_T^{miss} from the 4 neutrinos leading to a loss in separation. By omitting E_T^{miss} the associated systematic uncertainty additionally giving rise to a loss in sensitivity does not affect the Boost and 0Jet categories.

The pseudorapidity of the di-electron system, $\eta(2e)$, yields different distributions for the dominant Drell-Yan production mechanism $q\bar{q} \to Z/\gamma^*$ and major signal production via $gg \to H$. As a consequence of the parton density function of the proton, the momentum fraction of valence quarks is typically higher compared to the sea anti-quarks and thus the asymmetric initial state in the production of the Z boson in proton-proton collisions leads to momenta typically distributed towards higher pseudorapidities in the detector. This is different for the Higgs production via gluon fusion, where the constituents of the hard scattering process have symmetric initial states with the respect to their incoming momenta.

As explained for the VBF category, $\Delta\Phi(\ell^+, p_T^{miss})$ distributes towards lower values for the boosted Higgs di-tau decays where the genuinely produced E_T^{miss} due to the neutrinos is on average more collinear with the electrons in the final state as compare to Z+jets events where the mis-measured hadronic recoil leads to missing transverse energy on average more reverse to the Z boost.

The decay angle of the positive (negative) electron in the rest-frame of the Z boson for $Z/\gamma^* \to ee$ decays, is to leading order described by $\frac{1+\cos(\theta^*)^2}{2}$. However, the theoretically expected distribution is experimentally dubbed with the acceptance and event selection efficiency in this analysis. High (anti-)collinearity of the positive electron with the boost of the Z boson in the Z rest-frame is effectively the case for events where one of the two associated electrons has a low momentum when boosted back into the rest-frame of the detector leading to the distribution shown in Figure 5.14. For Higgs signal this variable is essentially flat due to the de-correlating effect of the leptonically decaying taus and the neutrinos involved.

The validity of the collinear approximation in these categories is slightly less discriminating as in the VBF category. Especially for events in the 0Jet categories unphysical solutions for di-tau events are likely due to the absence of recoiling jets and hence low boost of the Z/H. For events selected in the Boost categories physical solutions for di-tau events are again enhanced.

Similar to the VBF category the invariant mass of the di-electron pair and the reconstructed di-tau mass have complementary discrimination power with respect to the two major backgrounds $Z/\gamma^* \to ee$ and $Z \to \tau\tau$ where $Z/\gamma^* \to ee$ largely statistically dominates and would thus lead to suboptimal separation when both variables are used in one single BDT. Therefore, the invariant mass of the di-electron pair is only used for BDT_1 to discriminate $Z/\gamma^* \to ee$ from di-tau like

events and the reconstructed di-tau mass is only used in BDT_2 to discriminate $Z \to \tau\tau$ from $H \to \tau\tau$ processes. The SVFit reconstructed di-tau mass has higher resolution and hence better discrimination power for events in the Boost categories due to the improved E_T^{miss} resolution for boosted $Z/H \to \tau\tau$ decays.

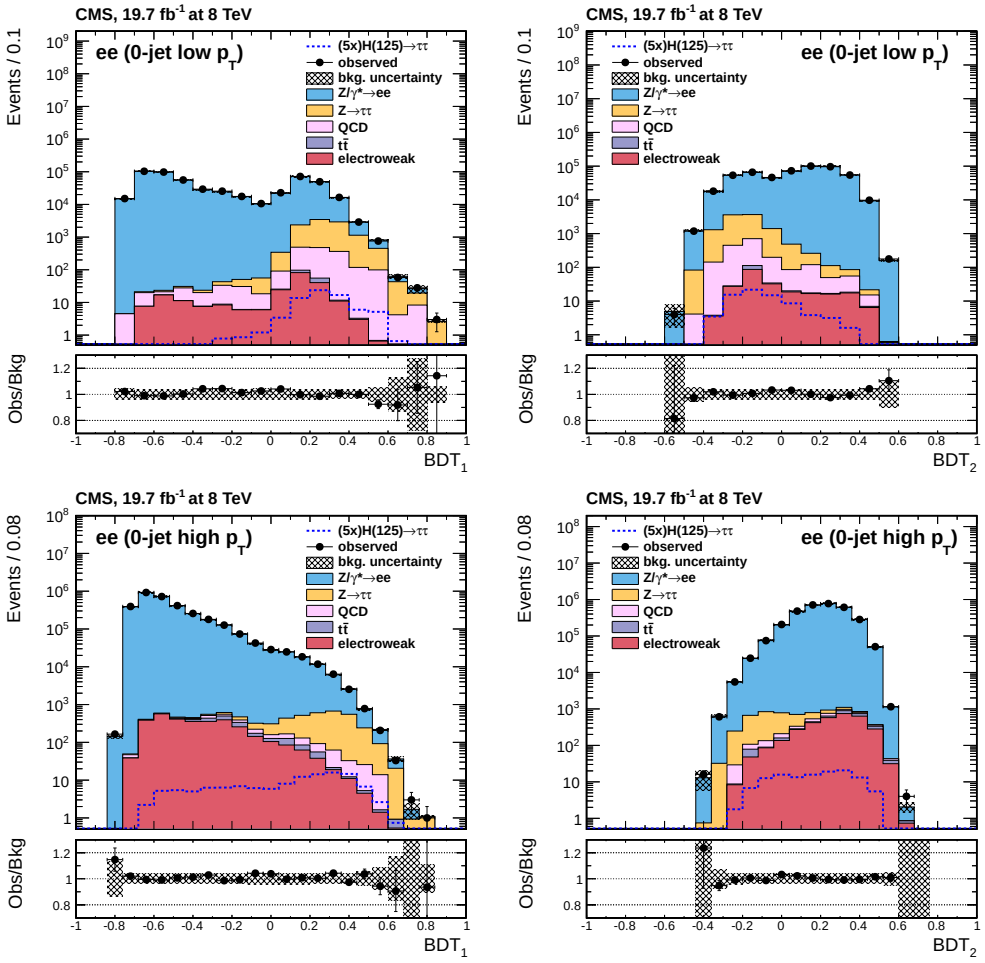

Figure 5.10.: Distributions for BDT_1 (left) and BDT_2 (right) for events selected in the 0Jet-low (top) and 0Jet-high (bottom) categories. Experimental data (8 TeV) are shown as circles, predicted background samples are represented by filled histograms. Also displayed as dashed histograms are the signal expected for the SM Higgs boson with $m_H = 125$ GeV/c^2 and scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

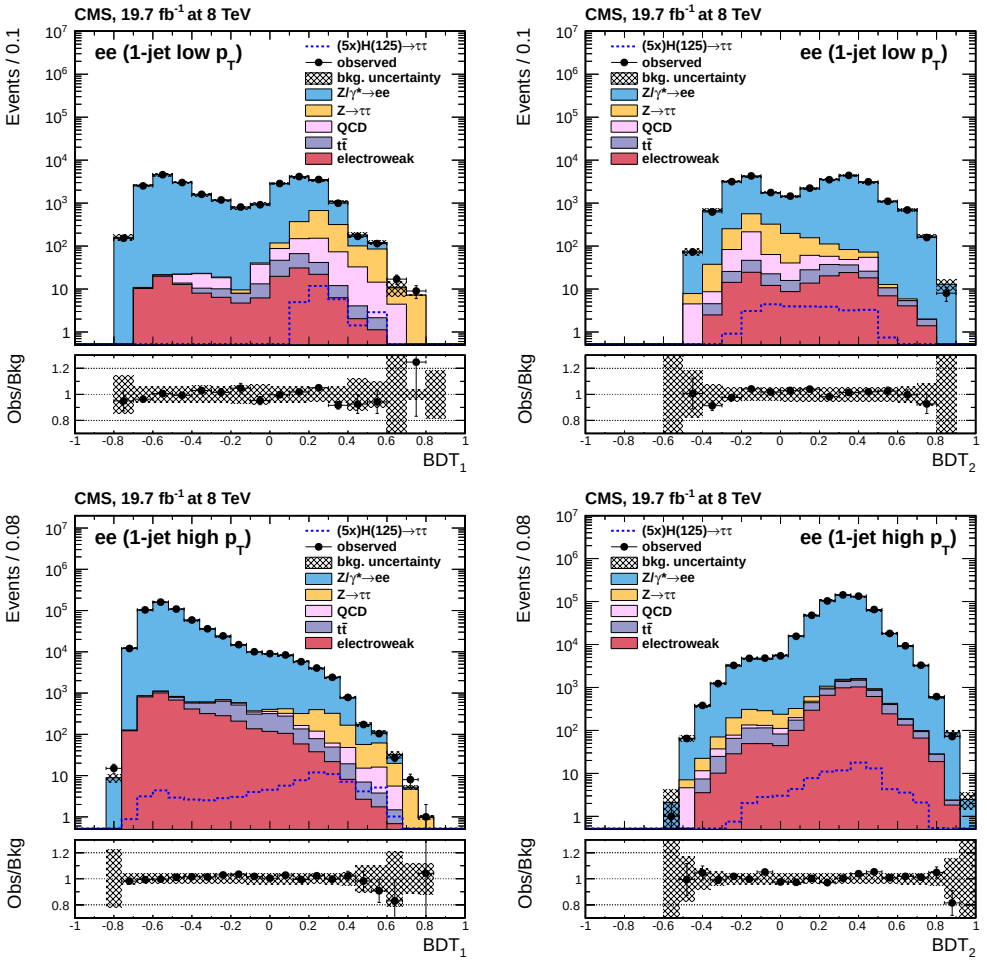

Figure 5.11.: Distributions for BDT_1 (left) and BDT_2 (right) for events selected in the Boost-low (top) and Boost-high (bottom) categories. Experimental data (8 TeV) is shown as circles, predicted background samples are represented by filled histograms. Also displayed as dashed histograms are the signal contributions expected for the SM Higgs boson with $m_H = 125$ GeV/c^2 and scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

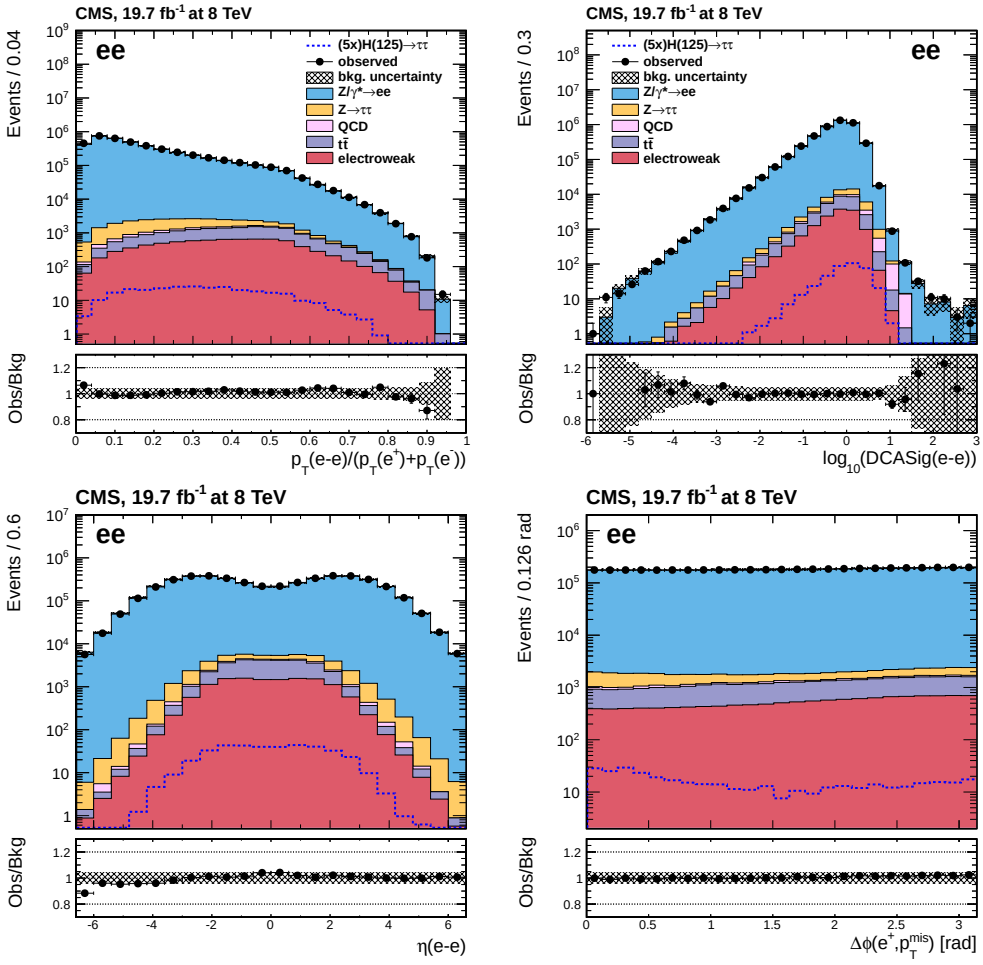

Figure 5.12.: Distributions of discriminating variables used to in the multivariate analysis in the *0Jet* and *Boost* categories. Data (circles) 8 TeV is compared with predicted background samples (filled histograms). Also shown is the SM Higgs boson signal with $m_H = 125 GeV/c^2$ scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

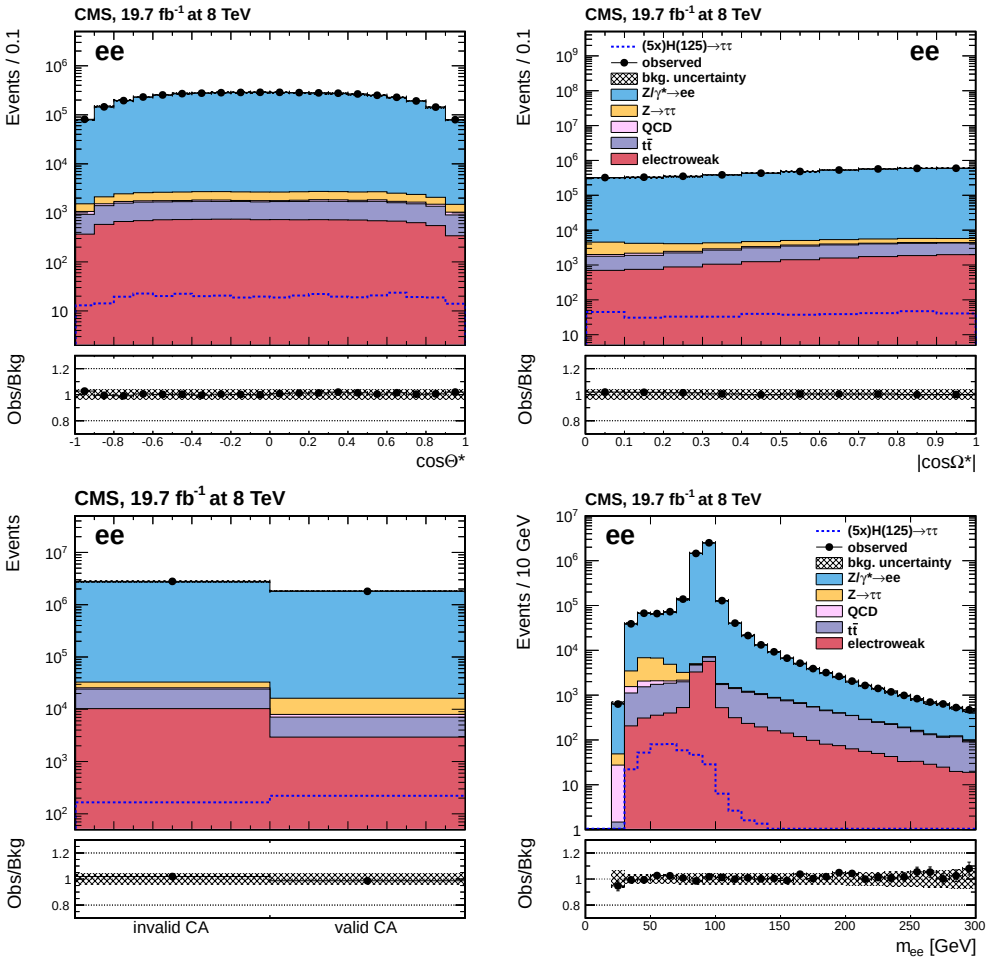

Figure 5.13.: Distributions of discriminating variables used to in the multivariate analysis in the *0Jet* and *Boost* categories. Data (circles) 8 TeV is compared with predicted background samples (filled histograms). Also shown is the SM Higgs boson signal with $m_H = 125 GeV/c^2$ scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

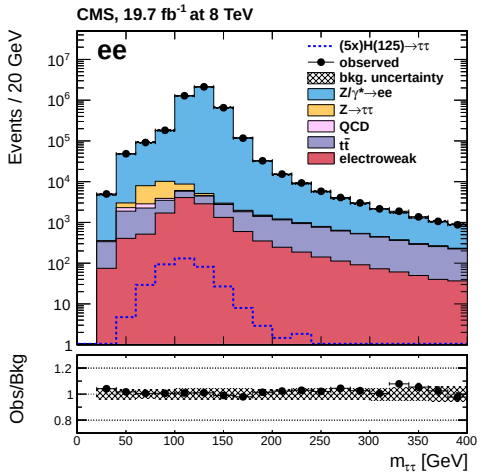

Figure 5.14.: Distribution of the reconstructed di-tau mass in the *0Jet* and *Boost* categories. Data (circles) is compared with predicted background samples (filled histograms). Also shown is the SM Higgs boson signal with $m_H = 125 GeV/c^2$ scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

5.3. Final Discriminant

With the two discriminators BDT_1 and BDT_2, both discriminating signal against the major backgrounds, a convenient method for statistical inference on the Higgs boson hypothesis would be to extract the signal in each category from the two-dimensional likelihood model for the corresponding two-dimensional BDT_1 vs BDT_2 distributions. However, in addition to the difficulty in controlling two-dimensional histograms, limited number of events for simulated signal and background samples result in poorly described shapes, where in turn a coarser choice of binning would lead to a degradation of the sensitivity of the analysis.

Therefore, separately in each of the five categories, BDT_1 and BDT_2 are combined into one global discriminant. This method then yields for each category a one-dimensional distribution for signal extraction which is well manageable and has well described shapes for each of the simulated samples.

The global discriminant is build from signal probability density functions F_{cat}^{signal} in the phase-space of BDT_1 and BDT_2. In the 0Jet and Boost low and high p_T categories, the signal PDFs are obtained from the two-dimensional BDT_1 vs. BDT_2 distributions of ggH samples with m_H = 110-145 GeV/c^2 combined for events classified within the corresponding category and normalized to one. For the VBF category the PDF is build from the two-dimensional distributions of qqH events selected selected in this category Higgs masses 110 to 145 GeV/c^2.

The global discriminant D_{cat} of an event i in category cat with BDT values BDT_1^i and BDT_2^i is then defined as the probability to be within the two-dimensional phase-space of $BDT_1 \leq BDT_1^i$ and $BDT_2 \leq BDT_2^i$ assuming the signal hypothesis.

$$D_{cat} = \int_{-1}^{BDT_1^i} \int_{-1}^{BDT_2^i} F_{cat}^{signal}(BDT_1, BDT_2) d(BDT_1) d(BDT_2).$$

D_{cat} for a certain event can be interpreted as the p-value assuming the signal hypothesis constructed as explained above.

For illustration in Figure 5.15 the two-dimensional BDT_1 vs BDT_2 distributions for events in the VBF category and Boost-high category is shown, separately for signal and background. As can be seen from these plots, signal distributes towards higher discriminator values in the upper right corner whereas background is distributed towards lower BDT_1 values for $Z/\gamma^* \to ee$ and lower BDT_2 values for $Z \to \tau\tau$ events. Shape comparisons for the two Drell-Yan backgrounds and signal are shown in Figure 5.16. Distinct shapes for the individual background and signal processes facilitate well the possibility to extract Higgs signal via template fits.

In Figures 5.17- 5.19 the final distributions in each category to be considered for signal extraction separately for 7 and 8 TeV data-taking periods are shown. A simultaneous template fit of the final distributions in all categories is performed to test the SM Higgs boson hypothesis for masses 110-145 GeV/c^2. The background

dominated categories 0Jet-low, 0Jet high and Boost-low give control over the inter-categorically correlated systematic uncertainties associated to each background in the signal sensitive categories, Boost-high and VBF.

Several other techniques to combine BDT_1 and BDT_2 were studied. The kNN method with adaptive phase space volume configuration was applied, for a varying number of 50-500 neighbors, to map local signal significances in the two-dimensional plane of each category into a corresponding one-dimensional discriminator. No improvement was found, and instead the performance degraded significantly in terms of expected exclusion limits on the Higgs signal strength parameter. Simple multiplication of the two discriminants resulted in unfavorable shape templates to be controlled by the profile likelihood fit and overall worse performance compared to the construction developed above. Additionally, the background-only hypothesis pdf, constructed similar to the signal hypothesis pdf, and also the ratio of the signal and background pdfs, have been found not to improve the resulting sensetivity with respect to the signal-only pdf.

These investigations provide good confidence in the choice of the construction of the final discriminant used for this analysis.

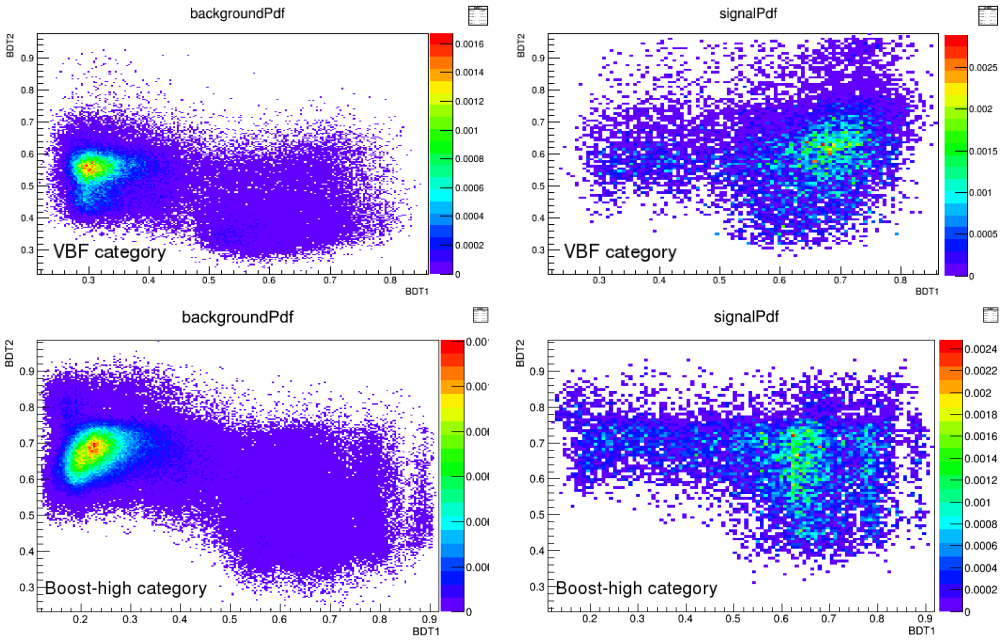

Figure 5.15.: Top: Distribution of background (left) and Higgs m_H =125 GeV/c^2 signal (right) events in the VBF category in the phase-space of BDT_1 and BDT_2. Bottom: Distribution of background (left) and Higgs m_H =125 GeV/c^2 signal (right) events in the Boost-high category in the phase-space of BDT_1 and BDT_2.

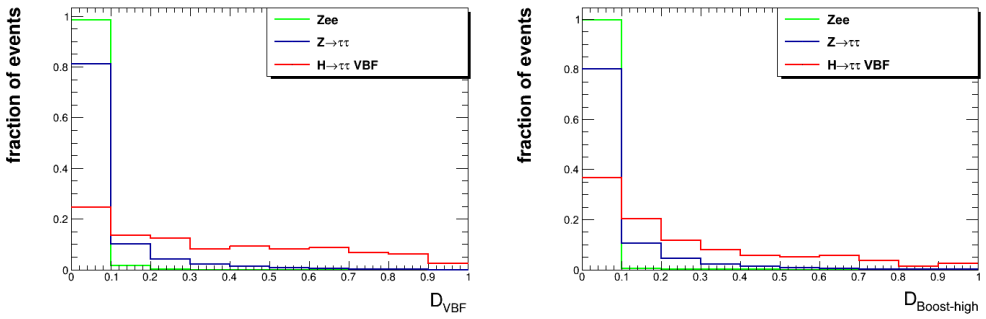

Figure 5.16.: Shape comparisons for the global discriminant D_{VBF} in the VBF category (left) and $D_{Boost-high}$ in the Boost-high category for signal and the two major backgrounds $Z/\gamma^* \to ee$ and $Z \to \tau\tau$.

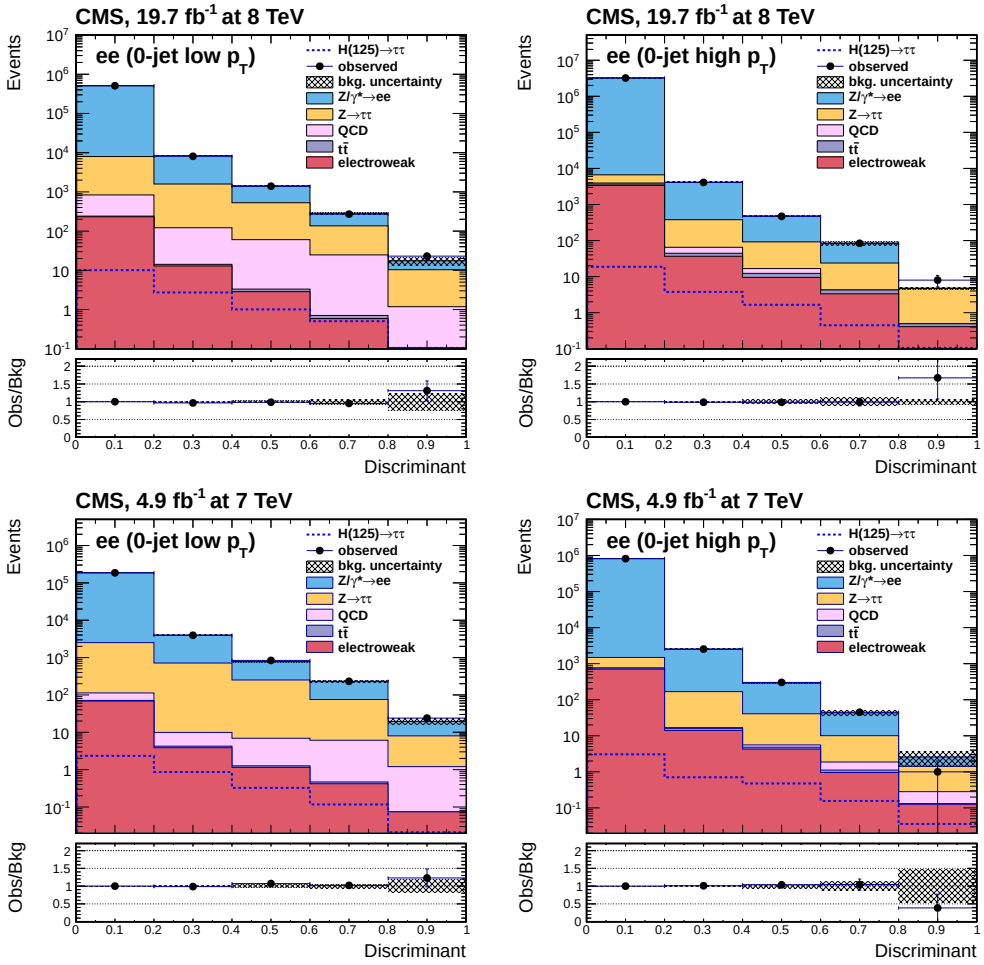

Figure 5.17.: Distribution of the global discriminant in the *0Jet* category in the di-electron channel at 7 and 8 TeV. Experimental data is shown as circles, predicted background samples are represented by filled histograms. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

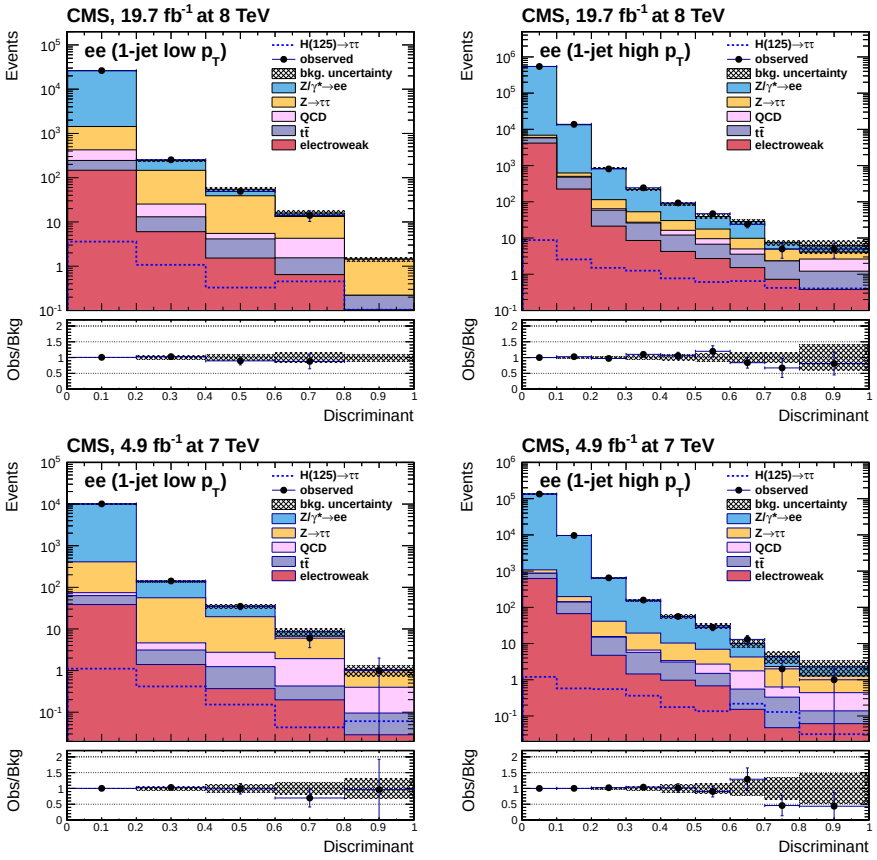

Figure 5.18.: Distribution of the global discriminant in the *1-Jet* category in the di-electron channel at 7 and 8 TeV. Experimental data is shown as circles, predicted background samples are represented by filled histograms. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

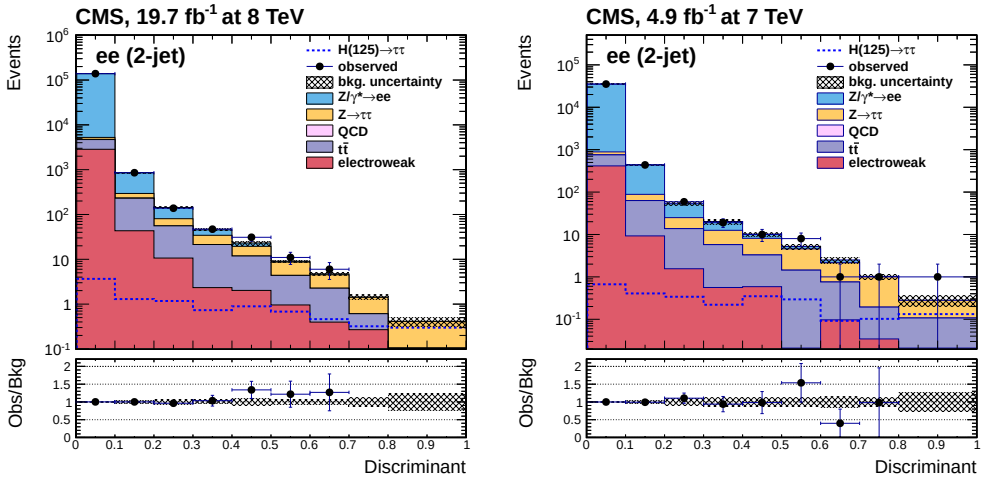

Figure 5.19.: Distribution of the global discriminant in the *VBF* category in the di-electron channel at 7 TeV (right) and 8 TeV (left). Experimental data is shown as circles, predicted background samples are represented by filled histograms. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

5.4. Background Modeling and Simulated Samples

The aim of the analysis is to probe the signal+background against the background-only hypothesis with respect to the observed data. Thus, precise knowledge about the proton-proton interactions and their Standard Model particle productions as well as the modeling of the detector is necessary. Depending on the particle production mechanism and final state, simulated event generators are used, or specifically controlled data subsamples with a high purity of the particle final state of interest are extracted. The latter method is called "data-driven" background estimation and for some background physics processes a combination of simulated event generators and data-driven techniques is used. In Table 5.9, the simulated event generators used in this analysis are listed. For the exclusive Z+NJets samples, the cross-section of the inclusive sample was re-weighted to the event yield with N jets on parton-level. The signal cross-section and branching ratios as a function of the Higgs mass follow the recommendations of the LHC Higgs cross-section group [31], shown in Table 2.4 and 2.2.

The ggH and VBF signal POWHEG [83] [84] samples and background MADGRAPH [85] samples are interfaced with PYTHIA [86] for parton-showering and fragmentation modeling. Associated Higgs production simulations are solely based on PYTHIA, and decays of taus in general, are simulated with the TAUOLA [87] program.

All generated simulated physics processes are then passed into to a detailed event reconstruction procedure using GEANT4 [80] to simulate the detector response and particle reconstruction within the CMS detector.

The analysis predicts all SM backgrounds in a data-driven way, except for the small contribution resulting from W+Jets, WW, WZ, and ZZ processes which are taken directly from simulation.

A sophisticated method for the prediction of $Z/\gamma* \to ee$ events is developed, to accurately predict the shape and normalization of these SM processes in the Final Discriminant in each event category. In bins of the phase-space of all input variables of the dedicated BDTs, described in Section 5.2.1 and 5.2.2, events are scaled by weights extracted from template fits of the DCA significance (cf. Section 4.8).

$Z \to \tau\tau$ process are predicted, using the embedding technique [88] [1], where a clean sample of $Z \to \mu\mu$ events is extracted from data, and muons are subsequently replaced by simulated tau decays.

$t\bar{t}$ and QCD contributions are predicted, via high purity control regions extrapolated into the signal region. Typically, high E_T^{miss} regions are used for $t\bar{t}$ processes and the side-band region for QCD is chosen by inverting the isolation cut on electrons and using di-electron events of same-sign.

Although the systematic uncertainties associated to the background predictions are summarized in Section 5.5.7, they are also mentioned here in context.

Table 5.9.: Monte Carlo event generator samples for both signal and background processes used in in the analysis. The cross sections for the background samples are already multiplied by the Monte Carlo filter efficiencies if applicable. All samples have been created within the official CMS production campaigns in fall 2011 (7 TeV) and summer 2012 (8 TeV).

Process	Generator	$\sigma \cdot \epsilon_{\mathrm{MC}}$ [pb]	
		7 TeV	8 TeV
SM $gg \to H$	POWHEG	–	–
SM qqH (VBF)	POWHEG	–	–
SM WH + ZH + ttH	PYTHIA	–	–
Z + Jets $m_{ll} < 50$	MADGRAPH	9530	11050
Z + Jets, $m_{ll} > 50$	MADGRAPH	3048	3504
Z + 1Jets $m_{ll} > 50$	MADGRAPH	not used	$3504 \times \frac{N_1}{N_{incl}}$
Z + 2Jets $m_{ll} > 50$	MADGRAPH	not used	$3504 \times \frac{N_2}{N_{incl}}$
Z + 3Jets $m_{ll} > 50$	MADGRAPH	not used	$3504 \times \frac{N_3}{N_{incl}}$
Z + 4Jets $m_{ll} > 50$	MADGRAPH	not used	$3504 \times \frac{N_4}{N_{incl}}$
$t\bar{t}$ (inclusive sample)	MADGRAPH	158	225
$t\bar{t}$ (dilepton decays of top pairs)	MADGRAPH	not used	25.0
W + Jets	MADGRAPH	31314	36257
$WW \to 2\ell 2\nu$	MADGRAPH	4.78	5.82
$WZ \to 3\ell\nu$	MADGRAPH	0.857	1.09
$WZ \to 2q2\ell$	MADGRAPH	1.79	2.27
$ZZ \to 2\ell 2\nu$	MADGRAPH	0.250	0.716
$ZZ \to 2q2\ell$	MADGRAPH	0.776	2.50
$ZZ \to 4\ell$	MADGRAPH	0.064	0.181

5.4.1. Di-boson and W + Jets backgrounds

Contributions from di-boson production and W + Jets backgrounds are solely estimated from Monte Carlo. The uncertainty of the W + Jets background is driven by limited Monte Carlo statistics of the selected W + Jets events and amounts to $25 - 50\%$ but has no visible effect on the analysis performance. An uncertainty of 30% is assigned on the normalization of the di-boson backgrounds.

5.4.2. QCD

For the multijet or QCD background with electrons from heavy flavor quark decays or mis-identified jets, a data-driven background technique is used. The contribution is estimated by exploiting the QCD-events-enriched control samples, selected by requiring same sign di-lepton pairs and applying inverted or relaxed electron

isolation criteria. In the following the estimation of the shape and normalization of the QCD background contribution is described.

For events preselected using the criteria outlined in Section 5.1, no indication of strong correlations between the electron isolation variables and the net charge of the selected electron pair is found. As discussed in Section 5.2, the preselection cut on the azimuthal angle between the momentum vectors of the positively and the negatively charged electrons, $\Delta\Phi(e^+, e^-) > 2.0$ rad in the *0Jet*, or the cut on the di-electron invariant mass, $m_{ee} > 30$ GeV/c^2 used in the *Boost* and *VBF* event categories, reduce opposite sign electron pairs originating either from the same quarkonia decays or from the same decay chain of heavy flavor mesons, $B \to \bar{D} + e^+\nu_e \to e^-\bar{\nu}_e + X$ (and charge conjugated mode) to an almost negligible level. These processes constitute the only possible source of correlations between the net charge of the electron pair and their isolation variables.

To model the shapes of the QCD background in each event category, a data sample of same-sign di-electron events selected with a medium isolation requirement is used:

- $0.25 < \mathrm{Iso}^{\mathrm{PF}}_{e,1} < 0.5$,
- $0.25 < \mathrm{Iso}^{\mathrm{PF}}_{e,2} < 0.5$,

This sample is characterized by a high purity of QCD events, namely $P_{\mathrm{QCD}} > 90$ %. The normalization of the QCD background in the signal region is then determined by means of two extrapolation factors.

The first extrapolation factor relates event yields of the opposite-sign to the same-sign QCD samples, and is determined with the sample of di-electron events selected by inverting the cut on the relative isolation variables of both electrons,

- $\mathrm{Iso}^{\mathrm{PF}}_{e,1} > 0.5$,
- $\mathrm{Iso}^{\mathrm{PF}}_{e,2} > 0.5$.

The OS/SS extrapolation factor is measured to be

$$F(\mathrm{OS/SS}) = 1.8 \pm 0.2.$$

The systematic error is evaluated by varying the cut on the relative isolation variable of both electrons between 0.3 and 0.7.

The second extrapolation factor relates the event yield in the sample of same-sign QCD events selected with the direct isolation requirement to the event yield in the sample of same-sign QCD events selected with medium isolation requirement. This extrapolation factor is determined using the relation,

$$F(\mathrm{Iso}) = \frac{N^{\mathrm{SS,dir}}_{\mathrm{QCD}}}{N^{\mathrm{SS,med}}_{\mathrm{QCD}}} = \frac{N^{\mathrm{SS,dir}}_{\mathrm{Data}} - f_{CF} \cdot N^{\mathrm{SS,dir}}_{\mathrm{non-QCD,MC}}}{N^{\mathrm{SS,med}}_{\mathrm{Data}} - f_{CF} \cdot N^{\mathrm{SS,med}}_{\mathrm{non-QCD,MC}}}, \tag{5.3}$$

where

- $N_{\text{Data}}^{\text{SS,med}}$ is the number of observed data events in the same-sign di-electron sample selected with the medium isolation requirement for both electrons;
- $N_{\text{Data}}^{\text{SS,dir}}$ is the number of observed data events in the same-sign di-electron sample selected with the direct isolation requirement for both electrons;
- $N_{\text{non-QCD,MC}}^{\text{SS,med}}$ is the number of non-QCD events in the same-sign di-electron sample selected with the medium isolation requirement for both electrons, this number is evaluated from Monte Carlo simulation;
- $N_{\text{non-QCD,MC}}^{\text{SS,dir}}$ is the number of non-QCD events in the same-sign di-electron sample selected with the medium isolation requirement for both electrons, this number is evaluated from Monte Carlo simulation;
- f_{CF} is the electron "charge-flip" scale factor, relating rate of the electron charge mis-measurements in the Monte Carlo simulation and data.

The scale factor f_{CF} is measured with the sample of the $Z \to ee$ events, where the charge of one of the two electrons is mis-measured. This sample is selected by requiring same sign di-electron pair, with both electrons passing the direct isolation requirement. The yields of the $Z \to ee$ events with electron "charge-flip" are determined separately in data and MC by fitting di-electron mass spectrum in the range from 60 to 120GeV/c^2. The fit is performed with the superposition of an exponential function, approximating the background contributions to the sample, and "Crystal-ball" function for the $Z \to ee$ contributions to the selected sample. The fits are illustrated in Figure 5.20. The scale factor f_{CF} is then computed as the ratio of the yields determined from fits in the data and Monte Carlo samples, $f_{CF} = 1.23 \pm 0.05$. The extrapolation factor from the medium to direct isolation region is measured to be

$$F(\text{Iso}) = 1.2 \pm 0.03. \tag{5.4}$$

The shapes of the QCD background in various control regions, defined by the net charge of the two electrons and the electron isolation requirements, are compared in Figure 5.21. Observed similarity of the shapes validates the usage of the same-sign di-electron sample selected with the medium isolation requirement for the modeling of the QCD background shapes in the signal region. The normalization of the QCD background in the signal region is obtained by applying the product of the two extrapolation factors, $F(\text{Iso}) \times F(\text{OS/SS})$, to the measured yield of QCD events in the SS di-electron sample selected with the medium isolation requirement for both electrons. Systematic uncertainties for this method are estimated from the statistics of the same-sign intermediate isolation sample which is added in quadrature to the uncertainty on the scale factors.

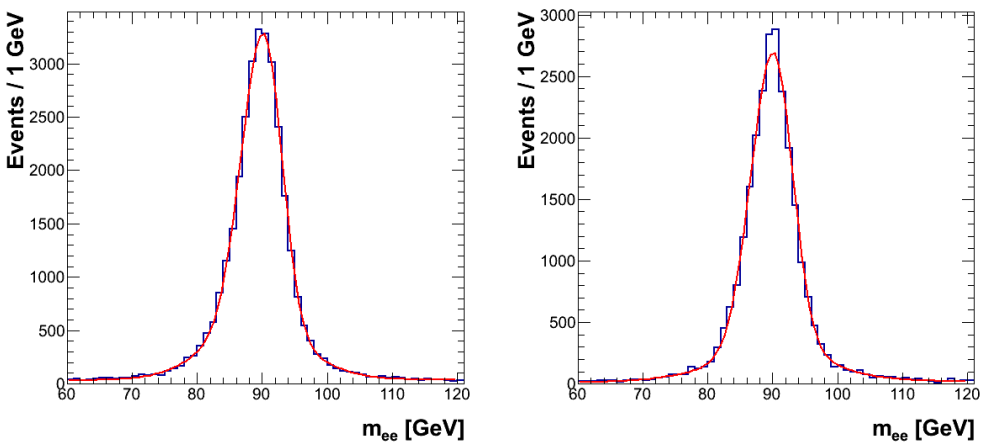

Figure 5.20.: The di-electron mass distribution in the selected sample of the $Z \rightarrow ee$ events with electron "charge-flip" in data (left plot) and Monte Carlo (right plot).

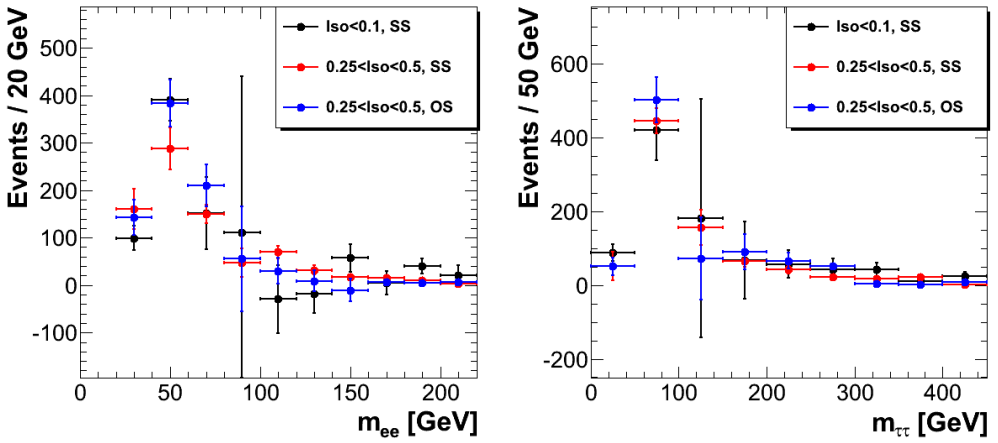

Figure 5.21.: The shapes of the QCD background in various control regions: the SS di-electron sample selected with the direct isolation requirement for both electrons (black points), the SS di-electron sample selected with the medium isolation requirement for both electrons (red points), the OS di-electron sample selected with the medium isolation requirement for both electrons (blue points). The left plot shows the distribution of the di-electron mass, the right plot shows the reconstructed di-tau mass.

5.4.3. Top-Pair Background

The $t\bar{t}$ background is evaluated with MC simulation accounting for the trigger efficiency, lepton reconstruction and selection efficiencies and acceptance. CMS measurements of the $t\bar{t}$ production cross-section are used to normalize the event yield [89]. These measurements are consistent with the approximate NNLO calculations [90] [91]. The theoretical predictions have an uncertainty of 8(10)% for the 7(8)TeV datasets, and are taken as a systematic uncertainty on the $t\bar{t}$ background normalization.

5.4.4. $Z \to \tau\tau$ Background

The $Z \to \tau\tau \to 2e4\nu$ decays of which the kinematics are very similar to the $H \to \tau\tau \to 2e4\nu$ decays, constitute the most signal-like background such that an accurate modeling is needed.

The normalization of the $Z \to \tau\tau$ background after preselection of di-electron events is estimated from simulation, accounting for the trigger efficiency, as well as for reconstruction and selection efficiencies. According to earlier studies by CMS [92], the inclusive Z boson production cross section is known to a precision of 2.5%. Additionally, there is an uncertainty associated with the non-linearity of the luminosity measurement of 2%, which is added in quadrature.

A data-driven method is used to derive the shape of this background. This is achieved by utilizing a hybrid sample of recorded CMS data and embedded simulated tau decays [88]. Separately for 7 TeV and 8 TeV data-taking periods, fully offline reconstructed $Z \to \mu\mu$ events are selected from data and the 4-vector of the reconstructed muons from the Z decay are replaced with generator τ-leptons, which subsequently decay into electrons and electron-neutrinos. Jet and E_T^{miss}related observables on which the event categorization of this analysis is based are thus unaffected from detector and physics process simulation discrepancies as they are directly extracted from data.

Possible biases from the initial selection of the di-muon pair in the $Z \to \mu\mu$ data events due to kinematically dependent trigger muon selection efficiency are unfolded and the the sample is re-weighted accordingly on an event-by-event basis.

Additional event category depended systematic normalization uncertainties are added, to account for a possible bias by the embedding technique by comparing event yields in each category of the true $Z \to \tau\tau$ Monte Carlo and $Z \to \tau\tau$ Monte Carlo embedded samples.

5.4.5. Di-Electron Drell-Yan Background

For the overwhelming $Z/\gamma* \to ee$ background estimation, which is data-driven, the shape difference of the distance of closest approach of the two electron tracks

for prompt $Z/\gamma* \rightarrow ee$ decays and signal-like $Z/H \rightarrow \tau\tau \rightarrow 2e4\nu$ decays is exploited. The distance of closest approach in the simulated $Z/\gamma* \rightarrow ee$ sample has been corrected in various kinematic phase space regions to reproduce the exact shape found in data (cf. Section 4.8). The $Z \rightarrow \tau\tau$ sample derived with the embedding technique 5.4.4 has been directly checked to predict the correct shape in the $e\mu$-channel, where $Z/\gamma* \rightarrow ee$ contributions are absent. Template fits in the DCA significance are performed in bins of dedicated *reduced Final Discriminants* ($FD_{reduced}$) and the invariant mass of the two selected electrons in each event category. The $FD_{reduced}$ (e.g. shown in Figure 5.22) is build similar to the Final Discriminant except that the BDT's do not include the following variables:

- DCA significance variable;
- di-electron invariant mass;
- reconstructed di-tau mass.

In each bin of $FD_{reduced}$ and m_{ee}, a kinematic phase-space dependent scale-factor is derived for the DY MADGRAPH sample in the corresponding bin from a template fit of the DCA significance distributions from $Z/\gamma* \rightarrow ee$ and $Z/H \rightarrow \tau\tau \rightarrow 2e4\nu$ to the one in data, where contributions from the $t\bar{t}$, $W + Jets$, QCD and di-boson backgrounds are subtracted. This method therefore predicts, data-driven, the $Z/\gamma* \rightarrow ee$ background contributions in various phase space regions of the BDT input variables, encoded in the Final Discriminant, in correlation with the di-electron invariant mass.

The binning for scale-factor extraction is chosen as follows:

- *0Jet-low*: m_{ee} = [0, 80], [80, 100], [100, 1000] (GeV/c^2) and $FD_{reduced}$ = [0.0, 0.5], [0.5, 1.0];
- *0Jet-high*: m_{ee} = [0, 80], [80, 100], [100, 140], [140, 1000] (GeV/c^2) and $FD_{reduced}$ = [0.0, 0.2], [0.2, 0.3], [0.3, 0.6], [0.6, 1.0];
- *Boost-low*: m_{ee} = [0, 80], [80, 100], [100, 1000] (GeV/c^2) and $FD_{reduced}$ = [0.0, 0.1], [0.1, 0.3], [0.3, 0.7], [0.7, 1.0];
- *Boost-high*: m_{ee} = [0, 80], [80, 100], [100, 1000] (GeV/c^2) and $FD_{reduced}$ = [0.0, 0.1], [0.1, 0.2], [0.2, 0.6], [0.6, 1.0];
- *VBF*: m_{ee} = [0, 80], [80, 90], [90, 100], [100, 1000] (GeV/c^2) and $FD_{reduced}$ = [0.0, 0.1], [0.1, 0.2], [0.2, 0.3], [0.3, 1.0]

The bin m_{ee} <80 GeV/c^2 has a sizable contribution from $Z \rightarrow \tau\tau \rightarrow ee$ and hypothetically signal. However in the template fit the $Z \rightarrow \tau\tau \rightarrow ee$ template normalization is left unconstrained such that the background estimation is unbiased

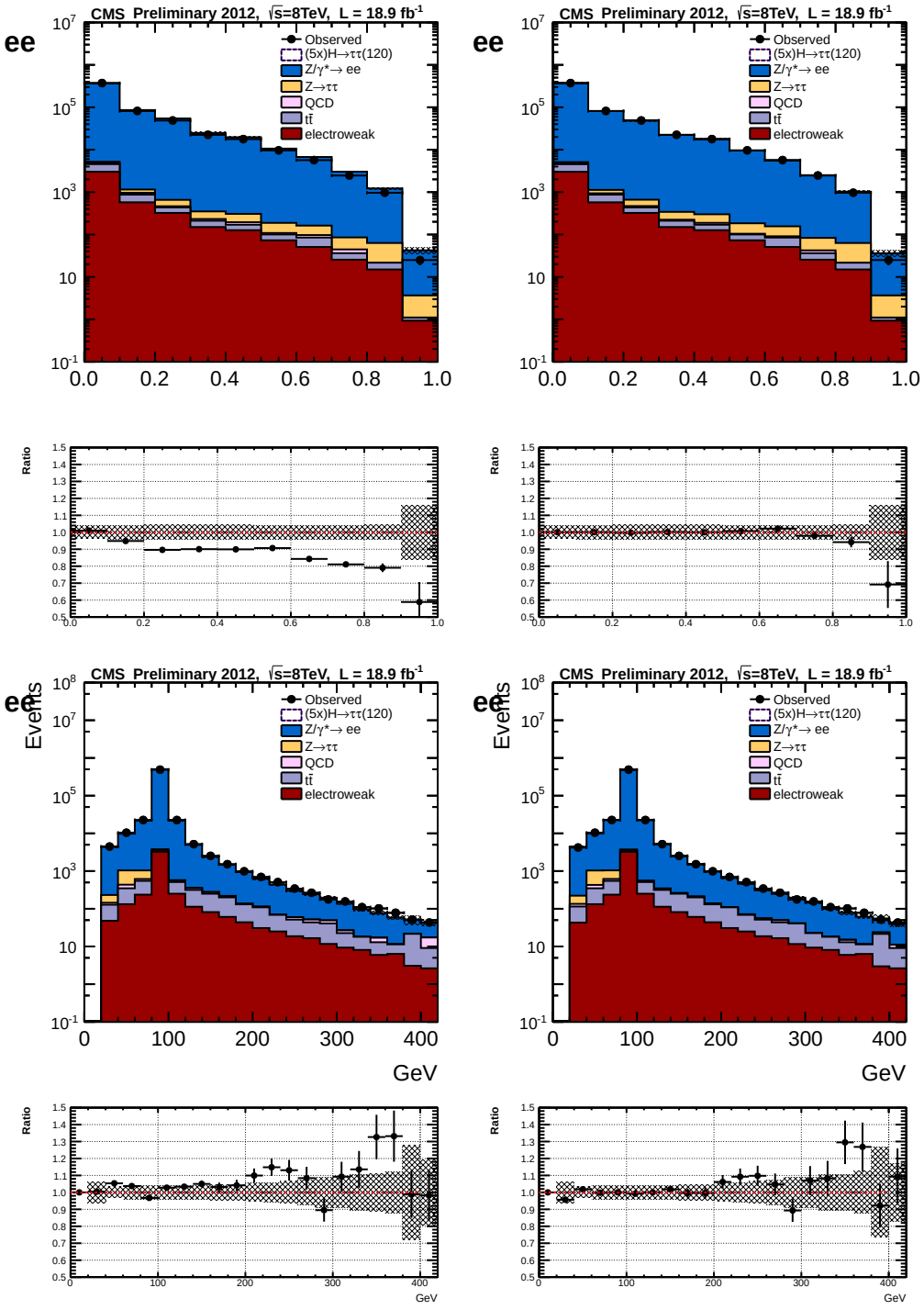

Figure 5.22.: Top: Distributions of the reduced Final Discriminant in the *Boost high* category. Before (left) DY background estimation and after (right) DY background estimation. Bottom: Distributions of the di-electron invariant mass in the *boost high* category. Before (left) DY background estimation and after (right) DY background estimation

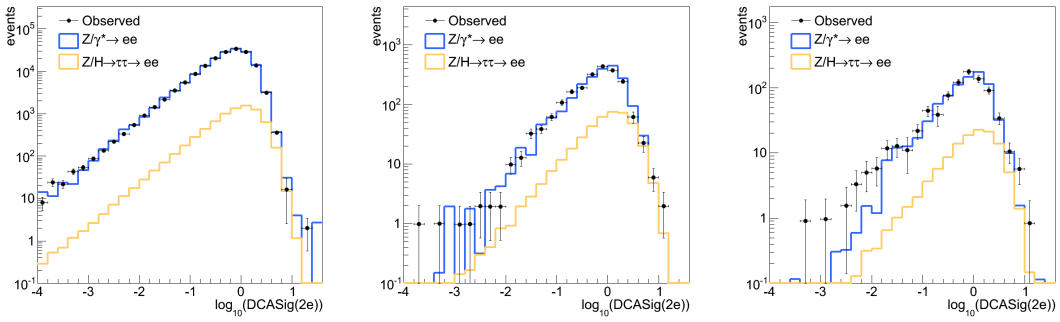

Figure 5.23.: The results of template fits are shown for events with the di-electron invariant mass in the region m_{ee} <80 GeV/c^2, and having the value of the 'reduced BDT' Final Discriminant, Left : *0Jet* category with 0.2 $> FD_{reduced} >$ 0.0. Center : *Boost* category with 0.6 $> FD_{reduced} >$ 0.2. Right : *VBF* category with 0.2 $> FD_{reduced} >$ 0.1.

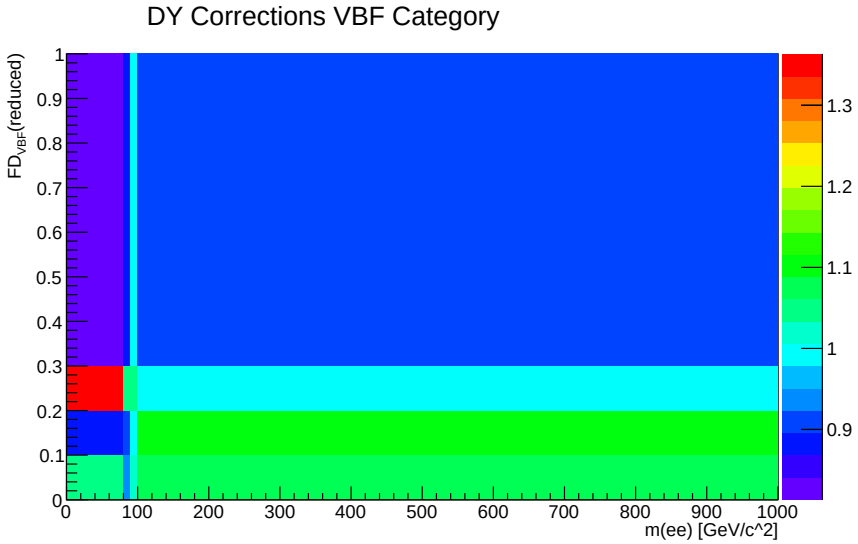

Figure 5.24.: Scale factors to be applied to $Z \rightarrow ee$ events in the VBF category, in bins of the reduced Final Discriminant $FD_{reduced}$ and invariant di-electron mass m_{ee}. The scale factors have been derived from template fits of the DCA significance.

from possible signal contamination having the same shape. Results of the template fits can be seen in Figure 5.23. For illustration, the scale factors to be applied to $Z \to ee$ in the VBF category are shown in Figure 5.24.

Two sources of systematic uncertainties of this background estimation are addressed:

- The error on the performed template fit in each bin;
- The normalization uncertainty in each category is estimated as the statistical uncertainty of the event yield in data with all other backgrounds subtracted.

From the uncertainties on the template fits 15 shape altering nuisance parameters are derived; 3 nuisance in each of the 5 category for the invariant mass regions:

- m_{ee} <80GeV/c^2;
- 80GeV/c^2 < m_{ee} <100GeV/c^2;
- m_{ee} >110GeV/c^2.

These bins maintain uncertainties corresponding to 3 different phase-space regions with high $Z \to \tau\tau$ contributions at low mass, the $Z \to ee$ dominated Z-peak region and the high mass region where mostly $t\bar{t}$ has been subtracted from the data-template.

5.5. Systematic Uncertainties

In this section the systematic uncertainties taken into account for the Higgs search are examined and discussed. There are several sources for uncertainties affecting the analysis. One source for uncertainties is related to the background estimates which can result in significant normalization uncertainties or uncertainties altering the shape of the final discriminant distribution for a particular background.

A second source is due to physics objects and the associated uncertainties on their energy scales, reconstruction and selection efficiencies.

Additionally, theoretical uncertainties addressing factorization and renormalization scale uncertainties, parton densities in the initial protons as well as the strong coupling α_s which affect parton shower and underlying event models are taken into account. Especially for Higgs production via gluon fusion processes selected in the VBF category, where on tree-level no jets are produced, the theoretical uncertainties have considerable impact on the shape of the final discriminant.

5.5.1. Luminosity

Energy deposits in the HF are used to measure the luminosity with the CMS detector online. Additionally, the luminosity is measured offline using the Pixel Cluster Counting Method where the number of hits in the CMS pixel detector is taken as a measure to be proportional to the luminosity. The absolute scale is calibrated on dedicated datasets using the Van Der Meer Scan technique with an uncertainty of 2.2% for 7 TeV and 2.6% for 8 TeV data [93] [94] [95] with contributions from the uncertainties of LHC proton beam parameters and the Van Der Meer Scan measurement.

Thus, to all signal and background contributions solely estimated from MC simulation, a normalization uncertainty of 2.2-2.6% is assigned to account for the uncertainty on the luminosity.

5.5.2. Electron Selection

The uncertainty on the electron selection efficiency, i.e. identification, isolation and trigger requirement as described in Section 5.1.1, is taken to be 3% per electron independent of the η and p_T phase space. This is estimated to cover both, the intrinsic uncertainty to the tag-and-probe method due to the choice of the signal and background model in each fit, and the statistical uncertainty associated to each scale factor.

This uncertainty therefore introduces an overall 6% normalization uncertainty for all signal samples and all background samples except the dominating $Z \to ee$ and the less dominating QCD backgrounds for which the normalization is estimated from data.

5.5.3. B-tag Efficiency

Due to the veto of events with b-tagged jets, uncertainties in the tagging efficiency for b-quark jets and in the mis-tagging efficiency for c-quark, light-flavor quarks and gluon jets result in normalization uncertainties varying from 1% to 3% [74] depending on the event categories.

5.5.4. Electron Energy Scale

Variables entering the boosted decision trees (described in Sections 5.2.1 and 5.2.2), used for discrimination in the final discriminant, are sensitive to the electron energy scale. As a consequence, and in addition to the effect on the normalization due to the selection cuts on the electron transverse momentum, the shape of the final discriminant depends on the electron energy scale and may change with respect to the uncertainty of the latter. The relative uncertainty on the electron energy

scale is estimated to be 1% in the barrel and 2.5% in the endcap, and propagated through the analysis to extract shape altered templates of the final discriminant to be used as a shape altering uncertainty in the final fit for signal extraction. Figure 5.25 shows the impact of 2σ electron energy scale up and down variations on the final discriminant in the VBF category for the $Z \to \tau\tau$ background component.

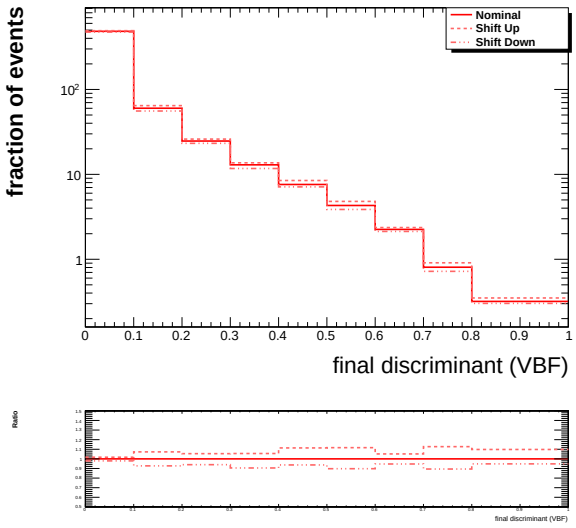

Figure 5.25.: $Z \to \tau\tau$ shape of the final discriminant in the VBF category in comparison with templates for 2σ up and down shift in the electron energy scale.

5.5.5. Jet Energy Scale

Uncertainties on the jet energy scale have influence on the shape of the final discriminant in the VBF category because of the usage of the discriminating variables m_{jj} and η_{jj} in the BDTs. Also, due to the event categorization based on jet multiplicities with p_T >30 GeV/c, inter-categorically anti-correlated normalizations uncertainties are introduced by the jet energy scale uncertainty. According to reference [70] the uncertainty on the jet energy scale varies between 3-7%, depending on the η and p_T of the jet.

In Figure 5.26 and 5.27, the effect of 2σ up and down variations of the jet energy scale in the VBF and 0Jet-low categories for top-pair and $Z \to ee$ processes are shown. For down variations, events migrate from event categories with higher jet multiplicities to categories with lower jet multiplicities, whereas for up variations events migrate towards categories with higher jet multiplicities

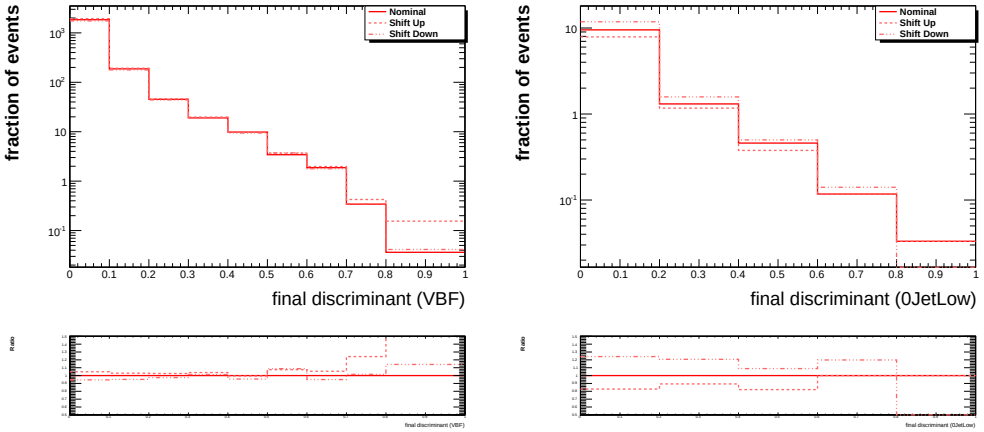

Figure 5.26.: $t\bar{t}$ shape of the final discriminant in the VBF(left) and 0Jet-Low (right) category in comparison with templates for 2σ up and down shifts in the jet energy scale.

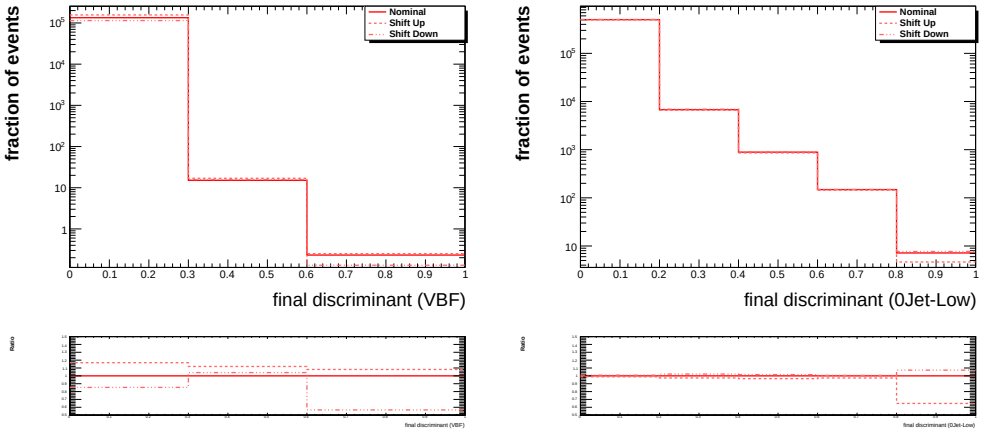

Figure 5.27.: $Z \rightarrow ee$ shape of the final discriminant in the VBF(left) and 0Jet-Low (right) category in comparison with templates for 2σ up and down shifts in the jet energy scale.

5.5.6. MET Scale

E_T^{miss} is directly used as input for the boosted decision trees in the VBF category and enters the svFit di-tau mass reconstruction which is used as BDT input variable in all event categories. The uncertainty on the E_T^{miss} scale gives therefore rise to shape altering uncertainties of the final discriminant.

The E_T^{miss} scale and its uncertainty is extracted from template fits of the E_T^{miss} spectrum. The fit is performed in 2 steps to determine the scale for $Z/\gamma^* \to ee$ events which are recoil corrected and $t\bar{t}$ events where no recoil corrections are applied.

E_T^{miss} spectrum templates with scale shifts of $\pm 10\%$ in steps of 0.5% are produced and continuously differentiable interpolated with strictly positive bin entries. To determine the E_T^{miss} scale for $t\bar{t}$ events, the E_T^{miss} spectrum for E_T^{miss} >80 GeV in the inclusively selected events sample is fitted for fixed contributions from backgrounds other then the largely dominating $t\bar{t}$. As an example in Figure 5.28 the effect of a $\pm 4\%$ shift in MET scale for the top-pair background in the $t\bar{t}$ dominated region MET>80 GeV is shown. For $t\bar{t}$, the normalization as well as the E_T^{miss} scale is used as free parameter in the fit, yielding

$$\text{Norm.}(t\bar{t}) = 0.95 \pm 0.03,$$

for the top-pair background normalization parameter with respect to the MC prediction and the relative shift in the E_T^{miss} scale is estimated to be

$$\Delta(\alpha_{\text{MET}})(t\bar{t}) \approx (1 \pm 2)\%$$

The correlation matrix for the two fit parameters is

$$\begin{pmatrix} 1 & -0.15 \\ -0.15 & 1 \end{pmatrix}.$$

To measure the met scale uncertainty for $Z\gamma* \to ee$ events, the fit on the E_T^{miss} spectrum is done in the region MET<80 GeV separately in the inclusive 0Jet and Boost categories and in the VBF category with fixed normalization. The top-pair background is fixed to the fitted values above. Figure 5.28 illustrates the impact of a $\pm 1\%$ shift of the E_T^{miss} scale in the Drell-Yan dominated E_T^{miss} phase space. In all categories the E_T^{miss} scale for the $Z\gamma* \to ee$ background is found to be consistent with $\Delta(\alpha_{\text{MET}})(Z \to ee) \approx (0 \pm 0.5)\%$.

As an illustration the Δ-2Log(α_{MET}) dependence for the performed fits for $t\bar{t}$ (left) and $Z/\gamma* \to ee$ in the VBF category (right) is shown.

For Higgs signal and the electroweak backgrounds, W+Jets, WW, WZ and ZZ type II recoil corrections are applied yielding an uncertainty of 3% on the E_T^{miss}

scale.

The uncertainty is propagated through the analysis by varying the E_T^{miss} scale for each background according to the associated uncertainty and with respect to the nominal E_T^{miss} scale. This procedure yields shape templates for the final discriminants to be used as shape altering uncertainties for signal extraction.

In Figure 5.30 the affect of a 2σ shift in the E_T^{miss} scale on the $t\bar{t}$ background component of the VBF final discriminant is shown.

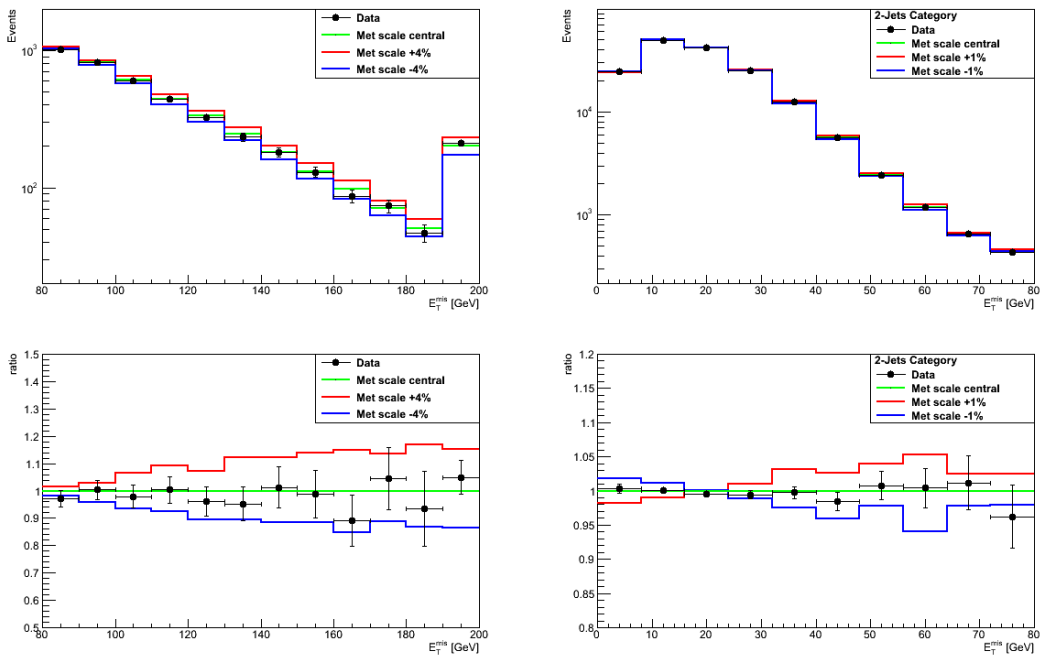

Figure 5.28.: The E_T^{miss} distribution in data and in MC for the inclusive sample of pre-selected di-electron events in the top-pair dominated region, MET$>$ 80 GeV, (left plot) and for the sample of events selected in the *VBF (2-Jets)* category in the $Z\gamma* \rightarrow ee$ dominated region (right plot). The colored lines indicate the effect of a $\pm 4\%$ shift in the E_T^{miss} scale in the left plot and a $\pm 1\%$ shift in the E_T^{miss} scale in the right plot.

5.5.7. Background Estimation Uncertainties

The systematic uncertainties associated to background estimations are also discussed in Section 5.4, and can be summarized as follows:

- **QCD:** The uncertainties of the scale factors are quadratically added to the statistical uncertainties associated to the event yields of the same sign

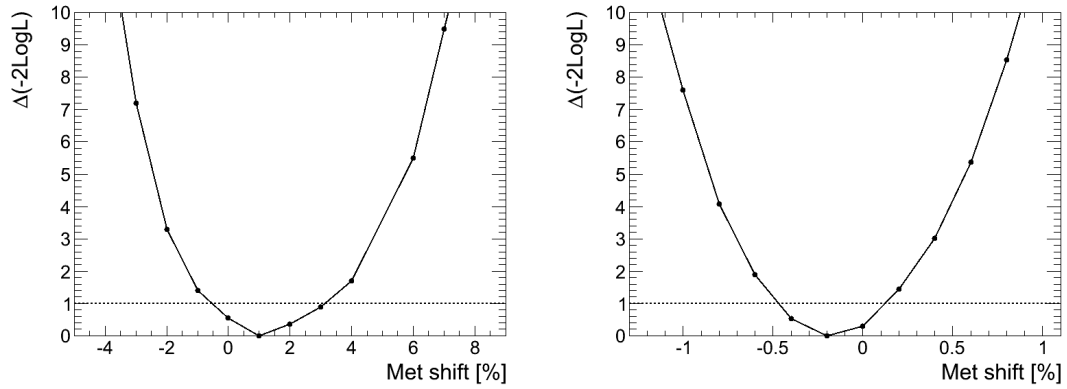

Figure 5.29.: The dependence of the $-2\log \mathrm{L}$ value on the E_T^{miss} scale for the $t\bar{t}$ events (left plot) and for the $Z\gamma* \rightarrow ee$ events selected in the *VBF (2-Jets)* category (right plot).

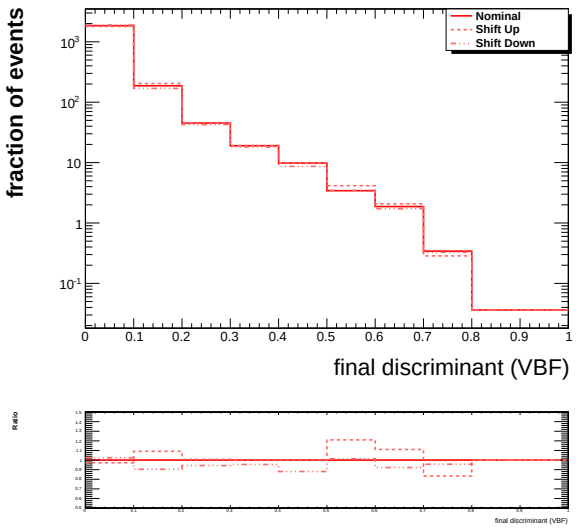

Figure 5.30.: $t\bar{t}$ shape of the final discriminant in the VBF category in comparison with templates for 2σ up and down shifts in the E_T^{miss} scale.

medium isolation sample in each category and taken into account as normalization uncertainty. It is estimated to vary from 20% in the 0 Jet categories to 60% in the VBF category.

- **Di-Boson and W+jets:** The normalization uncertainty is estimated from statistics of the simulated MC samples and results in 25-50% for W+jets and overall 30% for Di-Boson contributions.

- **Top-pair production:** The normalization uncertainty for $t\bar{t}$ is taken from the theoretical uncertainty on the NNLO cross section calculations and is 10%.

- $\boldsymbol{Z \rightarrow \tau\tau}$: Associated to the inclusive Z boson production cross-section a normalization of 3% is assigned. In addition event category dependent extrapolation uncertainties are addressed by comparing event yields of the pure MC simulated $Z \rightarrow \tau\tau$ sample and the simulated MC sample with embedded simulated tau decays. The extrapolation related event category dependent normalization uncertainties are estimated to vary between 7-9%.

- $\boldsymbol{Z/\gamma* \rightarrow ee}$: As explained in Section 5.4.5, the measurement of the $Z/\gamma* \rightarrow ee$ background results in independent shape and normalization altering uncertainties for low, medium and high di-electron invariant mass regions. In each category and each invariant mass bin, templates of 1σ shifts accounting for the statistical uncertainties of the fits in bins of the di-electron invariant mass are produced and treated as shape altering uncertainties in the final statistical inference. Shape comparisons for 1σ shifts in the low mass region for the final discriminator in the Boost-high and VBF category are shown in Figure 5.31. In addition, category dependent overall normalization uncertainties are estimated from the statistical uncertainty of the data yields, consisting of >95% $Z\gamma* \rightarrow ee$ events, with all backgrounds subtracted except $Z\gamma* \rightarrow ee$, resulting in uncertainties ranging from 0.1% to 1.5%.

5.5.8. Theoretical Uncertainties

Several theoretical uncertainties concerning underlying event modeling and parton showering (UEPS) or choices of factorization and renormalization scale (μ_f and μ_r), ambiguities of the proton parton density function (PDF) and biases due to missing higher order corrections are studied and taken into account.

The theoretical uncertainties on PDFs are studied as proposed in reference [96], with the set of PDFs CT10 [32], MSTW [33] and NNPDF [34]. The PDF affects the Higgs p_T spectrum and thus also influences the jet kinematics of the Higgs hadronic recoil leading to normalization uncertainties of 3.6% for qqH and 9.7%

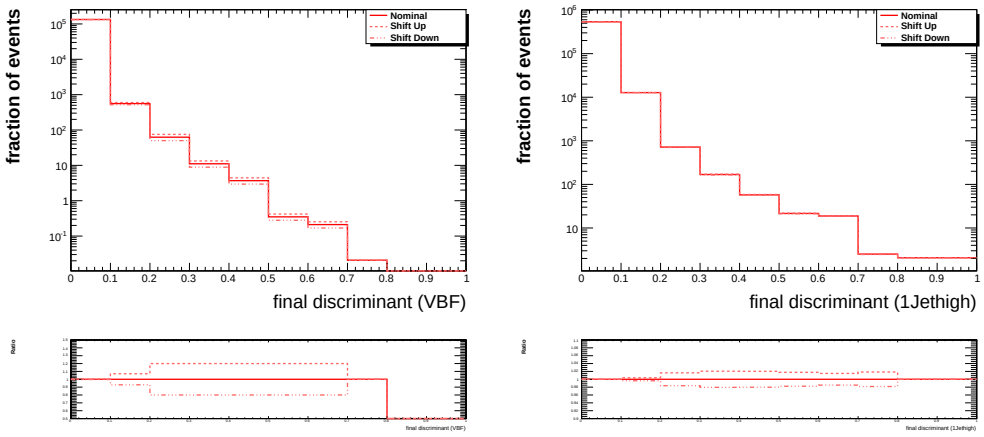

Figure 5.31.: Shape comparisons of the final discriminant in the Boost-high(right) and VBF category(left) for 1σ shifts of the DY background estimation in the low di-electron invariant mass region m_{ee} <80 GeV/c^2.

ggH Higgs production across all event categories. In addition, due to the usage of jet kinematics related variables in the VBF category the uncertainty on the PDF also introduces shape altering uncertainties for the final discriminant in this class of events. To estimate the shape altering uncertainties, events are re-weighted on an event-by-event basis dependent on the m_{jj}-η_{jj} phase-space. The weights are determined from the envelope of event yield ratios in bins of m_{jj} and η_{jj} after varying the PDFs according to their theoretical uncertainties. In Figure 5.32 the shape differences of the final discriminant in the VBF category for varying proton PDFs for the gluon fusion and VBF Higgs production mechanisms is shown.

For uncertainties on the UEPS, different Pyhtia tunes, the default ATLAS tune (AUET2) [97] and the CMS tune (Z2*) [98], are compared and normalization uncertainties are extracted. In Table 5.10 the associated uncertainty estimates in each category are shown.

The default renormalization and factorization scale for the simulated signal samples in chosen to be m_H. To account for the uncertainty or arbitrariness of choice, both scales are varied from $m_H/2$ to $2m_H$ and category dependent event yields are compared to extract the associated category dependent signal normalization uncertainties. The effect of scale variation on the shape of the final discriminant is also studied and found to be significant as can be seen in Figure 5.33 for VBF and gluon fusion signal processes in the VBF category.

The Powheg-Box sample used to model the ggH production has known inaccuracies regarding the modeling of the jet kinematics in the VBF category. Higgs events with two final state jets in this sample have one jet simulated by the NLO simulation and the second jet generated with the Parton Showering process. To

account for possible biases, shape comparisons of the VBF final discriminant are performed with improved, in particular hard scattering to soft-gluon matching, NLO ggH+nJets Monte Carlo simulations generated with Powheg-MiNLO [99] and aMC@NLO [100] and shown in Figure 5.34.

Table 5.10.: Normalization uncertainties for simulated signal samples associated to theoretical uncertainties for all event categories. The sign shows the effect of event migration between categories.

	qqH			ggH		
Category	PDF	UEPS	QCD scale	PDF	UEPS	QCD scale
0Jet-low	3.6%	8.9%	3.4%	9.7%	3.5%	10.3%
0Jet-high	3.6%	6.3%	2.8%	9.7%	4.2%	10.0%
Boost-low	3.6%	0.0%	0.8%	9.7%	-1.6%	10.9%
Boost-high	3.6%	0.4%	1.3%	9.7%	-2.2%	10.7%
Boost-high	3.6%	-1.4%	0.9%	9.7%	-7.4%	-18.2%

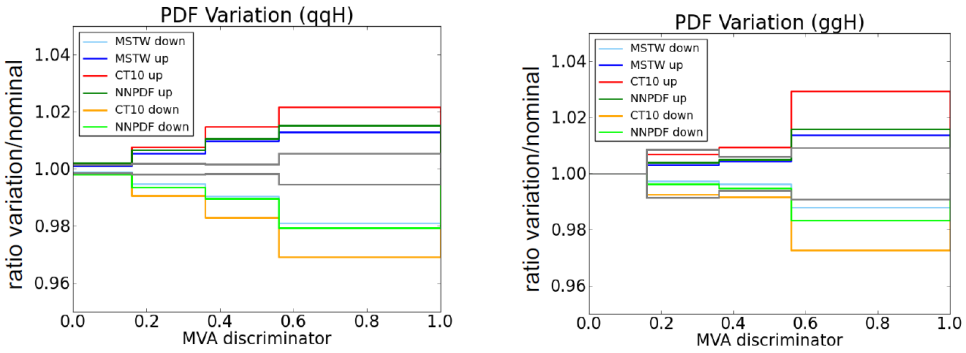

Figure 5.32.: Shape comparisons of the final discriminant in the VBF category for varying proton PDFs for VBF Higgs production (left) and gluon fusion Higgs production (right).

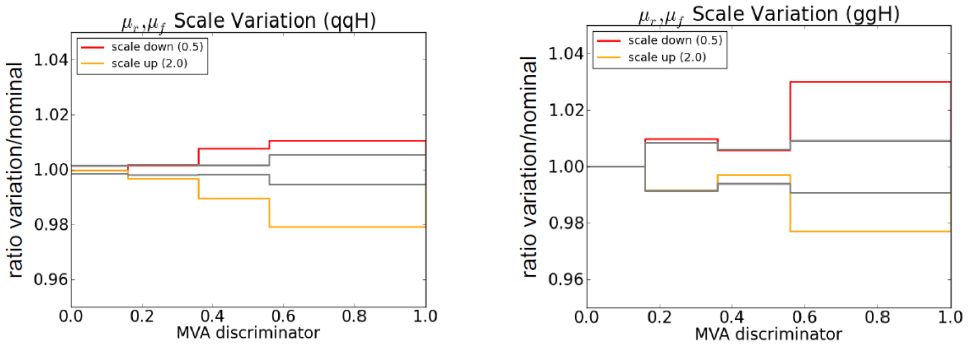

Figure 5.33.: Shape comparisons of the final discriminant in the VBF category for variation of μ_r and μ_f for VBF and gluon fusion Higgs production mechanisms.

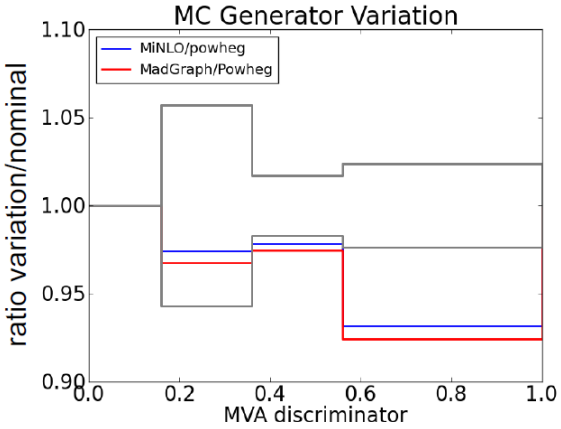

Figure 5.34.: Shape comparisons of the final discriminant in the VBF category Powheg-Box(nominal), Powheg-MiNLO and aMC@NLO for gluon fusion Higgs production.

5.6. Statistical Inference on Higgs Boson Hypothesis

To falsify the hypothesis of the existence of the Standard Model Higgs boson, a statistical hypothesis test is pursued and interpreted in terms of 95% confidence levels on the upper exclusion limits of the Standard Model Higgs production cross section, as well as p-values of the signal+background and background only hypothesis as a function of the Higgs boson mass.

The statistical procedure follows the recommendations of the LHC Higgs Combination Group [101] and is based on the modified frequentist method also referred to as the CL_s method [102] [103].

5.6.1. Profile Likelihood Model

A likelihood function is constructed to quantify the probability to observe the event yields in $2 \times (5 + 5 + 5 + 9 + 9) = 66$ bins of the final discriminator in each event category (0Jet-low, 0Jet-high, Boost-low, Boost-high, VBF) for 2011 7TeV and 2012 8 TeV data assuming a certain model hypothesis. The models to be tested in this analysis are the background-only hypothesis and the signal+background hypothesis, such that the predicted event yield in the i^{th} bin can be parametrized as:

$$\nu_i = \mu s_i(\vec{\theta}) + b_i(\vec{\theta}), \tag{5.5}$$

where s_i and b_i are the predicted event yields of Higgs signal and background in the i^{th} bin. The μ parameter represents the "signal strength modifier" and s_i is normalized such that for $\mu = 1$ the signal yield is equal to the expected yield of the Standard Model Higgs production. Both, signal and background yields, are possibly affected by systematic uncertainties which are represented by the set of nuisances $\vec{\theta}$. The set of nuisances assemble the systematic uncertainties described in Section 5.5 and are themselves constrained by posteriori probabilities $\rho(\vec{\theta}|\tilde{\vec{\theta}})$ to be interpreted as the pdf to describe the distribution of $\vec{\theta}$ given that the true value is $\tilde{\vec{\theta}}$, as measured in particular auxiliary measurements. After applying Bayes' Theorem and assuming a flat prior, ρ is converted to predictive measurement probabilities $p(\tilde{\vec{\theta}}|\vec{\theta})$ to describe the probability of the nuisances to yield $\tilde{\vec{\theta}}$ given that the true values are $\vec{\theta}$. The probability constrained on p is chosen to be of the form of a log-normal function for normalization uncertainties

$$p(x|\hat{x},\kappa) = \frac{1}{\sqrt{2\pi}\ln(\kappa)}\exp(-\frac{\ln(x/\hat{x})^2}{2\ln(\kappa)^2})\frac{1}{x} \tag{5.6}$$

with $\kappa = e^{\sigma}$ (σ is 1 standard deviation), and of the form of a normal distribution with mean 0 and width 1 for shape altering uncertainties for the parametrized

shifted templates discussed in Section 5.5.

The full likelihood is then constructed as a product of Poisson probabilities in each bin with constrained nuisances:

$$\mathcal{L}(\text{data}|\mu,\vec{\theta}) = \prod_i \frac{[\mu s_i(\vec{\theta}) + b_i(\vec{\theta})]^{n_i}}{n_i!} e^{\mu s_i(\vec{\theta})+b_i(\vec{\theta})} \times p(\tilde{\vec{\theta}}|\vec{\theta}), \tag{5.7}$$

where "data" represents the vector of observed data event yields n_i in each bin.

The test statistic is then defined as the profile likelihood ratio:

$$\tilde{q}_\mu = \begin{cases} -2\ln\frac{\mathcal{L}(\text{data}|\mu,\hat{\vec{\theta}}_\mu)}{\mathcal{L}(\text{data}|0,\hat{\vec{\theta}}_0)}) & \hat{\mu} < 0 \\ -2\ln\frac{\mathcal{L}(\text{data}|\mu,\hat{\vec{\theta}}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu},\hat{\vec{\theta}}}) & 0 \leq \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu, \end{cases} \tag{5.8}$$

where $\hat{\vec{\theta}}_\mu$ are the values of the nuisance parameters maximizing the likelihood given the observed data and fixed signal strength μ, and $\hat{\mu}$ and $\hat{\vec{\theta}}$ are the values which globally maximize the likelihood. The lower bound behavior for $\hat{\mu} < 0$ protects against unphysical upper limit results in case of background down fluctuations and the upper bound constraint for $\hat{\mu} > \mu$ imposes one-sided limit estimation and protects against interpreting signal excesses in the observed data as evidence against the signal hypothesis to be tested.

5.6.2. Observed Upper Limits

In order to extract observed upper limits on the signal strength for the Standard Model Higgs boson hypothesis, for each fixed value of μ two probability density functions of the test statistic 5.8 are built by generating toy Monte Carlo pseudo-datasets for the signal+background hypothesis for given signal strength and for the background-only hypothesis. Both sets of pseudo-data are generated around event counts following Poisson probabilities and the best fit values of the nuisances to the observed data, i.e. $\hat{\vec{\theta}}^{obs}_\mu$ for the signal+background pdf and $\hat{\vec{\theta}}^{obs}_0$ for the background-only pdf. The test statistic is then evaluated separately on both pseudo-datasets giving rise to the pdfs $f(\tilde{q}_\mu|\mu,\hat{\vec{\theta}}^{obs}_\mu)$ and $f(\tilde{q}_\mu|0,\hat{\vec{\theta}}^{obs}_0)$ to describe the pdf of the test statistic under the assumption of the signal+background and background-only hypothesis respectively.

Two p-values dependent on the test statistic evaluated on the observed data $\tilde{q}^{obs}_\mu$ are defined for the signal+background hypothesis and for the background-only hypothesis:

$$p_s(\mu) = P(\tilde{q}_\mu > \tilde{q}_\mu^{obs}|\mu, \hat{\vec{\theta}}_\mu^{obs}) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\vec{\theta}}_\mu^{obs})\mathrm{d}\tilde{q}_\mu \tag{5.9}$$

$$1 - p_b(\mu) = P(\tilde{q}_\mu > \tilde{q}_\mu^{obs}|0, \hat{\vec{\theta}}_0^{obs}) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|0, \hat{\vec{\theta}}_0^{obs})\mathrm{d}\tilde{q}_\mu. \tag{5.10}$$

The p-values define the probabilities to obtain at least as high values of the test statistic as the observed value under the assumption of the background-only hypothesis and signal+background hypothesis for the given signal strength to be tested.

The ratio of the two p-values is defined as

$$\mathrm{CL_s}(\mu) = \frac{p_s(\mu)}{1 - p_b(\mu)} \tag{5.11}$$

and a particular signal hypothesis with signal strength μ is then said to be excluded at 1-$\mathrm{CL}_s(\mu)$ confidence level (C.L.). For the exclusion of the signal hypotheses in this analysis the confidence level is chosen to be 95%, meaning that signal strengths corresponding to $\mathrm{CL_s}(\mu) \leq 0.05$ are excluded with the observed data. The signal strength $\mu^{95\%}$ is defined such that $1 - \mathrm{CL_s}(\mu^{95\%}) = 0.95$.

The p-value and significance of an excess in the observed data is then based on evaluating the test statistic for $\mu = 0$ on the toy Monte Carlo pseudo-datasets of the background-only hypothesis to construct the pdf $f(\tilde{q}_0|0, \hat{\vec{\theta}}_0^{obs})$. The p-value for an observed value $\tilde{q}_0^{obs}$ under the background-only hypothesis is then defined as

$$p_0 = P(\tilde{q}_0 > \tilde{q}_0^{obs}|0, \hat{\vec{\theta}}_0^{obs}) = \int_{\tilde{q}_0^{obs}}^{\infty} f(\tilde{q}_0|0, \hat{\vec{\theta}}_0^{obs})\mathrm{d}\tilde{q}_0.. \tag{5.12}$$

5.6.3. Expected Upper Limits

The expected upper limit can be derived by generating toy Monte Carlo data-sets based on the background-only hypothesis with the nominal values of the nuisances. On each toy the upper limit $\mu^{95\%}$ can be measured as described in the previous section yielding a distribution of upper limits. The 50% quantile of the distribution of $\mu^{95\%}$ upper limits obtained from background-only pseudo-datasets is then defined as the expected upper limit. Likewise, the $\pm 1\sigma$ uncertainty band is defined between the 16% and 84% quantiles and the $\pm 2\sigma$ band between the 2.5% and 97.5% quantiles of the $\mu^{95\%}$ distribution.

The described procedure is very computing intensive. A fast option to obtain the expected limit is to use the *asymptotic approximation* of the pdf of the test statistic for the regime of a large number of events sample [104]. The pdf of the test statistic $\tilde{q}_\mu$ is then of the analytic form

$$f(\tilde{q}_\mu|\mu') = \frac{1}{\sqrt{8\pi}}\frac{1}{\sqrt{\tilde{q}_\mu}}\exp\Big[-\frac{1}{2}(\sqrt{\tilde{q}_\mu}+\frac{\mu-\mu'}{\sigma})^2\Big] \tag{5.13}$$

$$+\begin{cases} \frac{1}{\sqrt{8\pi}}\frac{1}{\sqrt{\tilde{q}_\mu}}\exp\Big[-\frac{1}{2}(\sqrt{\tilde{q}_\mu}-\frac{\mu-\mu'}{\sigma})^2\Big] & \tilde{q}_\mu \leq \mu^2/\sigma^2 \\ \frac{1}{\sqrt{8\pi}(2\mu/\sigma)}\exp\Big[-\frac{1}{2}\frac{\tilde{q}_\mu-\frac{\mu^2-2\mu\mu'}{\sigma^2}}{\mu^2/\sigma^2}\Big] & \tilde{q}_\mu > \mu^2/\sigma^2, \end{cases} \tag{5.14}$$

with $\sigma = \mu/\tilde{q}_{\mu,A}$. The value $\tilde{q}_{\mu,A}$ is the test statistic evaluated on the "Asimov Dataset" [104], defined to be the dataset with the expected signal and background yields for nominal values of the nuisances. To evaluate the p-values 5.9 and 5.10 to determine the CL_s limits, μ' is set to $\mu' = \mu$ for the signal+background hypothesis and to $\mu' = 0$ for the background-only hypothesis. The median and the $\pm 1\sigma$ and $\pm 2\sigma$ bands can then directly be obtained from the cumulative function of the test statistic pdf in the asymptotic limit without generating toy Monte Carlo pseudo-data.

6. Search Results

In this chapter, the results of the search for the Higgs boson decaying into tau pairs are presented, based on data recorded with the CMS experiment corresponding to 4.9 fb^{-1} and 19.7 fb^{-1} at 7 TeV and 8 TeV, respectively. The results for the di-electron channel are specifically shown in detail in Section 6.1. The search strategy used for the ee-channel as described in Chapter 5, is identically also applied for the analysis of the $\mu\mu$-channel. Therefore, the combination of the same-flavor final states, i.e. the di-electron and di-muon channels together, are summarized in Section 6.2. Finally, in Section 6.3 the full combination of all $H \to \tau\tau$ final state searches: ee and $\mu\mu$, $\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$ and $e\mu$. In addition, analyses specifically exploiting di-tau final state W and Z boson associated Higgs production mechanisms are considered and combined with the di-tau Higgs inclusive searches.

In summary, the observed limit derived from the ee-channel alone is compatible with both the background-only and SM Higgs signal+background hypothesis within 1σ and excludes $3.7 \times \sigma^H_{SM}$ and $3.2 \times \sigma^H_{SM}$ for $m_H = 125$ GeV/c^2 and $m_H = 120$ GeV/c^2 at 95% confidence level, respectively.

The combined sensitivity of the same-flavor lepton channels, ee and $\mu\mu$ both using the analysis strategy outlined in this thesis, is compatible with the sensitivity of the $e\mu$-channel. This is remarkable, because although sharing the same branching ratio ($\mathcal{BR}(\tau\tau \to ee) + \mathcal{BR}(\tau\tau \to \mu\mu) \approx \mathcal{BR}(\tau\tau \to e\mu)$), the same-flavor lepton channels have overwhelming background contributions from $Z\gamma* \to ee/\mu\mu$ processes, not affecting the $e\mu$ final state analysis. This evidently demonstrates the power of the CMS search strategy for the same-flavor lepton channels developed in this thesis.

The combination of all $H \to \tau\tau$ final states reveals an observed excess of 3.2σ for $m_H = 125$ GeV/c^2 with respect to the background-only hypothesis and thus provides first direct evidence for a new resonance in the di-tau invariant mass spectrum consistent with the prediction of fermionic couplings of the Higgs boson.

6.1. Di-Electron Channel

The results of the statistical inference on the Higgs boson hypothesis as explained in Section 5.6 for the $H \to \tau\tau \to 2e4\nu$ analysis based on the final discriminants constructed in Section 5.3 are compiled in this section.

The expected significance of the Standard Model signal+background hypothesis

for the ee-channel alone is found to be below 1 standard deviations with respect to the background-only hypothesis. Therefore, the results are interpreted in terms of 95% C.L. exclusion limits on the signal strength parameter and no significant deviations of the observed exclusion limit with respect to the expected exclusion limit can be anticipated.

Figure 6.2 shows observed and expected 95% C.L. upper limits on the signal strength parameter μ as a function of m_H from 90 to 145 GeV/c^2 in steps of 5 GeV/c^2 for all event categories for 8 TeV data separately (7 TeV limits are shown Appendix E). Overall good agreement within 1 standard deviation can be seen between the expected and observed limits across all event categories. In Appendix F the pulls on all nuisances are listed and reveal good agreement with the nominal values ensuring that the various backgrounds are well under control and the systematics are appropriately taken into account.

In the 0Jet-low and Boost-low categories a sensitivity decrease towards higher m_H can be observed in contrast to the 0Jet-high and Boost-high categories where the sensitivity decreases towards lower m_H. This feature is due to the migration of high mass Higgs signal events into the high lepton p_T categories and the migration on low mass Higgs signal events into the lo lepton p_T categories. The VBF category is the most sensitive category followed by the Boost-high category.

The observed and expected exclusion limit of the combined event categories of both 7 and 8 TeV data-taking periods is displayed in Figure 6.3, with explicit numbers shown in Table 6.1. No significant deviation of the observed with respect to the expected exclusion limit is found above 1 standard deviation. The observed limit excludes $3.2 \times \sigma^H_{SM}$ for $m_H = 120$ GeV/c^2 and $3.7 \times \sigma^H_{SM}$ for $m_H = 125$ GeV/c^2 at 95% confidence level.

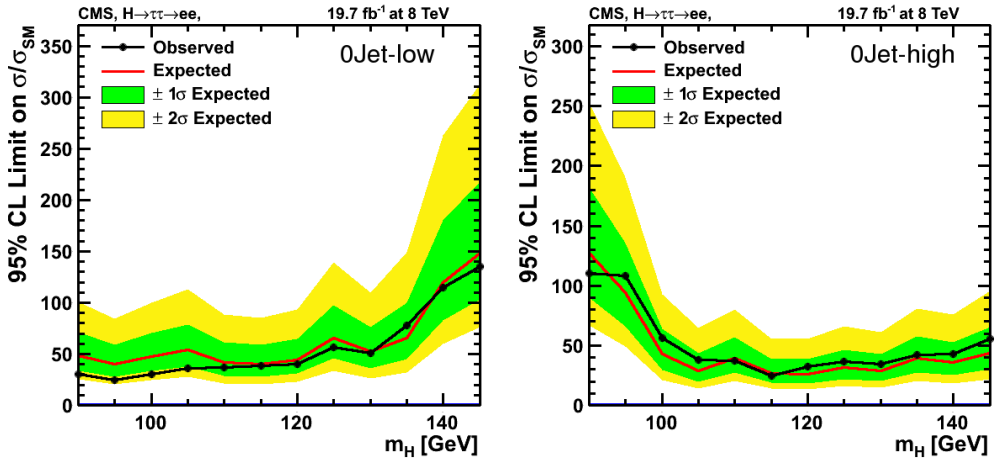

Figure 6.1.: Observed and expected 95% C.L. upper exclusion limits on the signal strength for the background dominated 0Jet-low (left) and 0Jet-high (right) categories of the ee-channel for 8 TeV data-taking period.

Table 6.1.: Explicit numbers of 95% C.L. upper exclusion limits (c.f. Figure 6.3) on the signal strength parameter μ for the search of $H \to \tau\tau$ in the ee-channel.

m_{H}	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	Obs. Limit
90 GeV	5.605	7.735	11.344	16.996	24.583	13.796
95 GeV	4.301	5.987	8.844	13.391	19.456	7.177
100 GeV	3.409	4.724	6.953	10.418	15.068	6.564
105 GeV	2.774	3.801	5.484	8.086	11.408	5.326
110 GeV	2.216	3.058	4.484	6.719	9.664	4.341
115 GeV	2.093	2.893	4.203	6.264	8.937	4.147
120 GeV	1.691	2.333	3.422	5.100	7.359	3.260
125 GeV	1.722	2.390	3.484	5.193	7.451	3.715
130 GeV	2.093	2.885	4.203	6.264	8.988	3.909
135 GeV	2.209	3.051	4.453	6.672	9.650	4.194
140 GeV	3.209	4.397	6.344	9.353	13.274	5.219
145 GeV	4.379	6.000	8.656	12.762	18.112	7.967

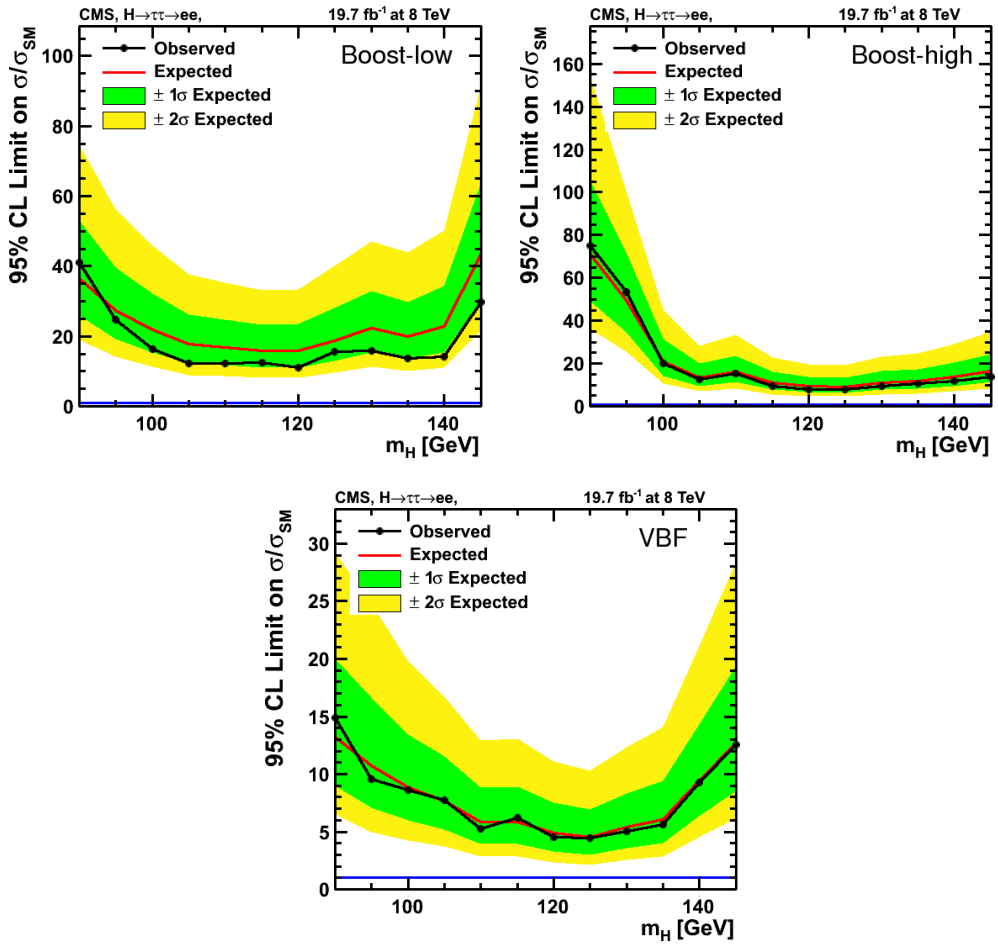

Figure 6.2.: Observed and expected 95% C.L. upper exclusion limits on the signal strength for the Boost-low (upper-left), Boost-high (upper right) and VBF (bottom) categories of the $H \to \tau\tau$ ee-channel for 8 TeV data-taking period.

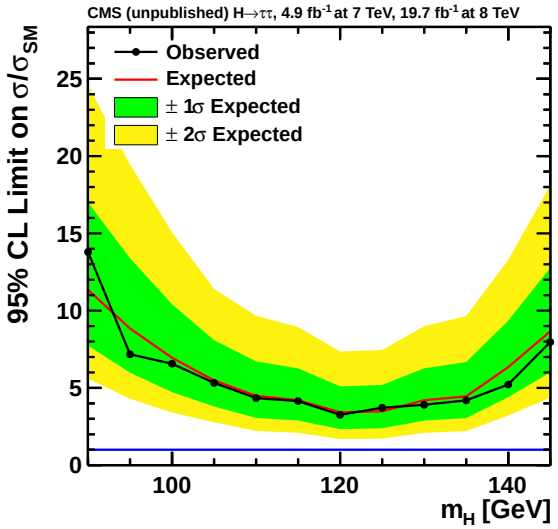

Figure 6.3.: Observed and expected 95% C.L. upper exclusion limits on the signal strength for $H \to \tau\tau$ ee-channel for full 2011 7 TeV and 2012 8 TeV data-taking periods and all categories combined.

6.2. Same-Flavor Lepton Channels

As stated before, the analysis strategy developed in this thesis was also applied to the $H \to \tau\tau \to 2\mu 4\nu$ final-state analysis. The only difference is that instead of selecting events with two electrons, events are required to have two appropriately identified muons [1]. Up to a per-mill level the branching ratio of the $\mu\mu$ final-state is the same as for the ee final-state and the background composition of statistically overwhelming Drell-Yan and additional Top-pair, Di-Boson and QCD processes coincide for the ee and $\mu\mu$ channels.

The final discriminants of both same-flavor lepton channels are then used to derive 95% upper limits on the Standard Model Higgs boson signal strength parameter. Figure 6.4 shows good agreement between the $ee + \mu\mu$ observed and expected 95% upper limits on the Standard Model Higgs production rate and about $2.1 \times \sigma^{H}_{SM}$ at 95% confidence level for $m_H = 125$ GeV/c^2 can be excluded. Explictit numbers are shown in Table 6.2.

Table 6.2.: Explicit numbers of 95% C.L. upper exclusion limits (c.f. Figure 6.4) on the signal strength parameter μ for the search of $H \to \tau\tau$ in the ee- and $\mu\mu$-channel.

m_{H}	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	Obs. Limit
90 GeV	4.367	5.921	8.469	12.317	17.198	8.232
95 GeV	3.266	4.456	6.406	9.394	13.216	4.379
100 GeV	2.440	3.315	4.750	6.890	9.634	4.110
105 GeV	1.758	2.388	3.422	4.990	7.000	2.206
110 GeV	1.411	1.933	2.789	4.112	5.836	2.361
115 GeV	1.409	1.914	2.742	3.977	5.562	2.299
120 GeV	1.215	1.657	2.383	3.475	4.904	2.044
125 GeV	1.128	1.532	2.195	3.202	4.491	2.085
130 GeV	1.352	1.838	2.633	3.840	5.418	2.229
135 GeV	1.525	2.072	2.969	4.330	6.073	2.405
140 GeV	1.934	2.628	3.766	5.462	7.684	2.805
145 GeV	2.871	3.885	5.547	8.001	11.153	4.304

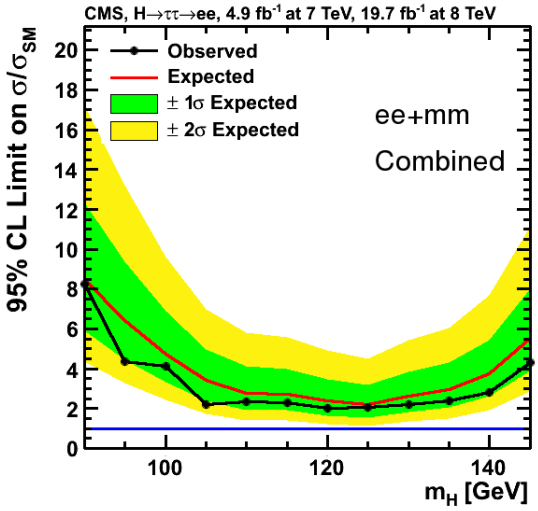

Figure 6.4.: Observed and expected 95% C.L. upper exclusion limits on the signal strength for the combination of the $H \to \tau\tau$ same-flavor lepton ($ee + \mu\mu$) channels.

6.3. Full $H \to \tau\tau$ Combination

To seek evidence for Higgs couplings to taus, the same-flavor lepton channels are combined with the more sensitive channels $\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$ and $e\mu$.

The analysis strategies for these channels are thoroughly described in reference [1]. Instead of the final discriminator, these channels use the svFit reconstructed invariant di-tau mass in each pre-defined event category directly for signal extraction. Also less sensitive analyses with 12 final states exploiting the vector boson associated Higgs production are combined into one "VH"-channel.

A comparison of the expected upper 95% C.L. limits separately for all final states is shown in Figure 6.5. As can be seen, the $\mu\tau_h$ final state is the most sensitive analysis and a comparison with Figure 6.4 shows that the same-flavor lepton channels are more sensitive for $m_H > 120$ GeV/c^2 than the 12 VH channels combined.

Figure 6.6 on the left shows the svfit reconstructed di-tau mass combined as the sum of each event category of the $\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$ and $e\mu$ channels weighted to the expected S/(S+B). As can be seen in the inset in the upper-left, an excess of events is observed in data with respect to the background-only hypothesis around $m_{\tau\tau} \approx 125$ GeV/c^2 compatible with the Standard Model Higgs boson hypothesis with $m_H = 125$ GeV/c^2. The excess is evaluated in terms of 95% C.L. upper exclusion limits on the Standard Model Higgs production rate as explained in Section 5.6, where the observed limit shows a significant deviation with respect to

the expected limit. Expected SM $H \to WW$ with $m_H = 125$ GeV/c^2 contributions are added to the background-only hypothesis. The local p-value and corresponding significance for the inclusive $H \to \tau\tau$ analyses, assuming the background-only hypothesis, as a function of m_H is shown in Figure 6.6 right. For $m_H = 125$ GeV/c^2 the observed significance of the excess is found to be 3.4 standard deviations with a corresponding best-fit signal strength value of $\hat{\mu} = 0.86 \pm 0.29$.

Observed and expected 95% upper limits for the fully combined inclusive and VH analyses are shown in Figure 6.7 (left) together with the corresponding expected and observed local p-values and significances (right). For $m_H = 125$ GeV/c^2 the observed significance of the excess is evaluated to be 3.2σ compared to the expected 3.7σ. Figure 6.8 shows the combined best-fit result $\mu = 0.78 \pm 0.27$ on the signal strength parameter for the SM Higgs boson with $m_H = 125$ GeV/c^2 and additionally the best-fit values for each final state analysis (left) and category (right) independently, including the combined $V + H$ channels [1].

All together the combined analysis of all $H \to \tau\tau$ final states provides experimental evidence for a new phenomenon in the di-tau invariant mass spectrum compatible with the Standard Model Higgs boson hypothesis prediction.

Additionally, the $H \to \tau\tau$ analysis is combined [2] with the CMS $VH \to V + b\bar{b}$ analysis [105]. The vector boson associated production mode with leptonically decaying vector bosons is most suitable to analyze $H \to b\bar{b}$ decays due to the lepton which can be used as a tag to reject the overwhelming QCD background. By combining the $VH \to V + b\bar{b}$ and $H \to \tau\tau$ analysis the expected significance is enhanced to 4.4σ. The best-fit value of the signal strength is found to be 0.83 ± 0.24 with an observed significance of 3.8σ for $m_H = 125$ GeV/c^2 with respect to the background-only hypothesis.

Figure 6.9 (left) shows the local p-value and associated significance in units of standard deviations for the background-only hypothesis to describe the observed data together with the Standard Model Higgs boson expectation separately for the inclusive $H \to \tau\tau$, $VH \to V + \tau\tau$ and $VH \to V + b\bar{b}$ channels and the combination of all three analysis. The right plot shows the profile-likelihood scan for the signal-strength value of the combined $H \to \tau\tau$ and $VH \to V + b\bar{b}$ channels with respect to the observed data. For $\mu = 0$ the significance is explicitly shown and overall the contributions from bosonic Higgs decays are added to the background.

The combination of CMS analyses exploring feasible fermionic Higgs decay channels supports the results of the $H \to \tau\tau$ analysis and provides further evidence for fermionic couplings of the new Higgs-like boson, increasing the observed significance to 3.8σ for $m_H = 125$ GeV/c^2 with respect to the background-only hypothesis. The results are compatible with the predictions of the Standard Model Higgs boson, and provide, for the first time, strong direct evidence for couplings of the Higgs boson to down-type fermions.

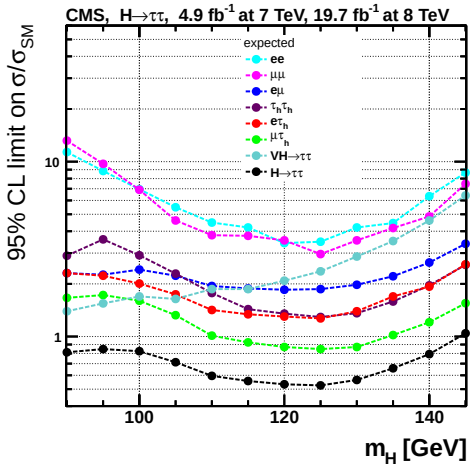

Figure 6.5.: Expected 95% C.L. upper limit on the signal strength parameter for the background only hypothesis separately shown for all inclusive $H \to \tau\tau$ final-state analysis and the combined VH channels.

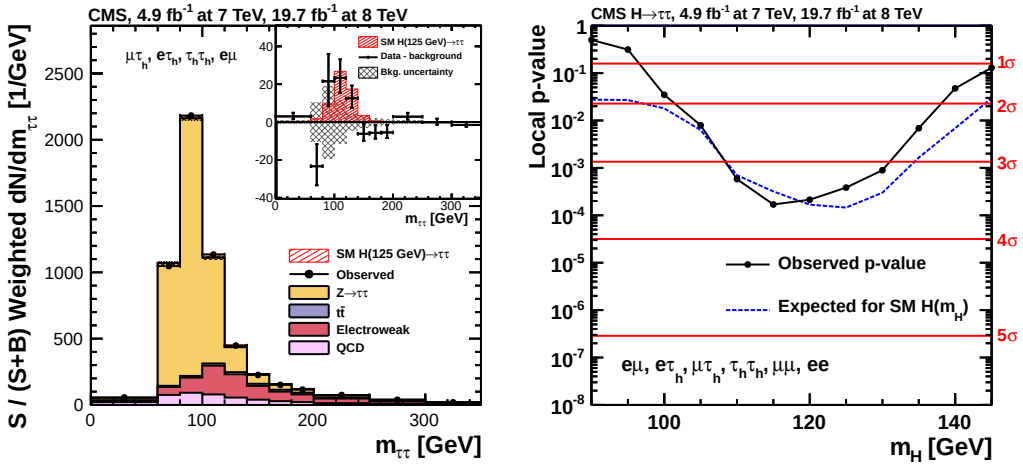

Figure 6.6.: Left: Combined $m_{\tau\tau}$ distribution of $\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$ and $e\mu$ states weighted by the ratio of the expected signal and signal+background yields in each event category. Right: Observed and expected local p-value and significance as a function of m_H with respect to the background-only hypothesis of the combined $H \to \tau\tau$ final states ee, $\mu\mu$, $\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$ and $e\mu$.

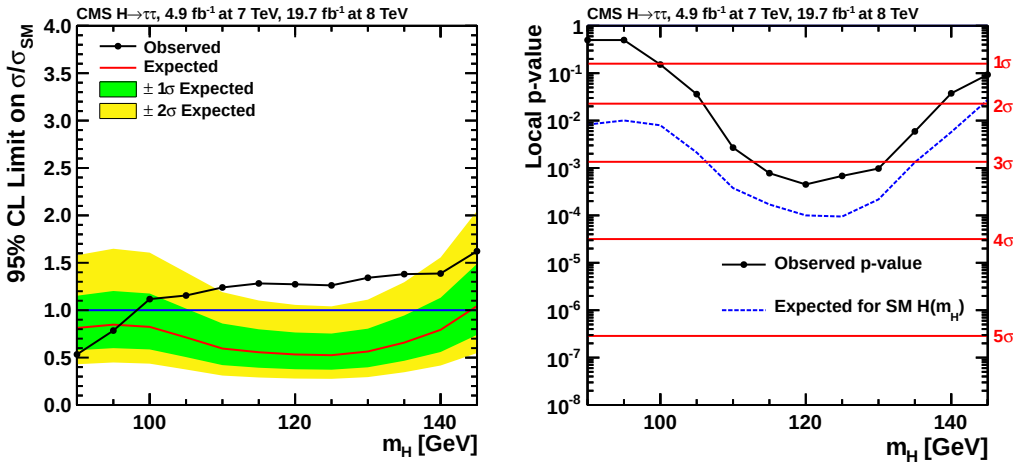

Figure 6.7.: Observed and expected 95% C.L. upper exclusion limits (left) and local p-values and significances in units of standard deviations (right) for the combined $H \to \tau\tau$ analysis as a function of m_H.

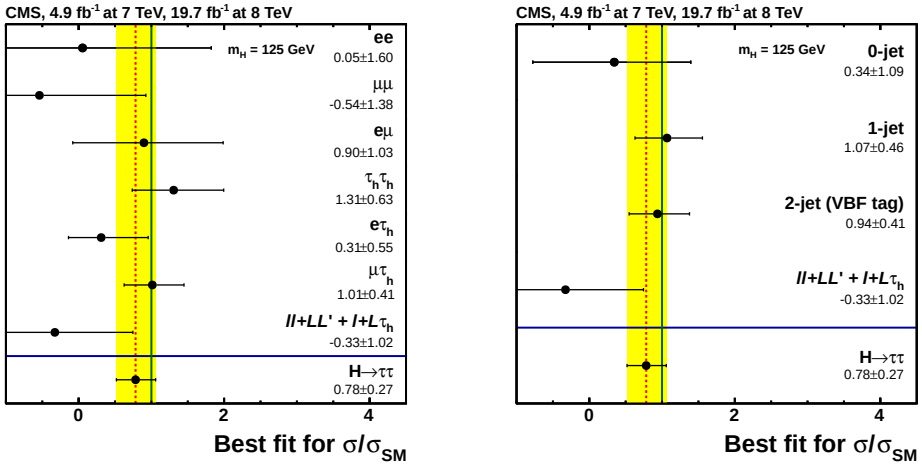

Figure 6.8.: Best-fit signal strength values, separately for each channel (left) and category (right), for $m_H = 125$ GeV/c^2. The combined value (dashed line) for the $H \to \tau\tau$ analysis in both plots corresponds to $\mu = 0.78 \pm 0.27$, obtained in the global fit combining all categories of all channels. The contribution from the $H(125GeV) \to WW$ process is treated as background normalized to the SM expectation.

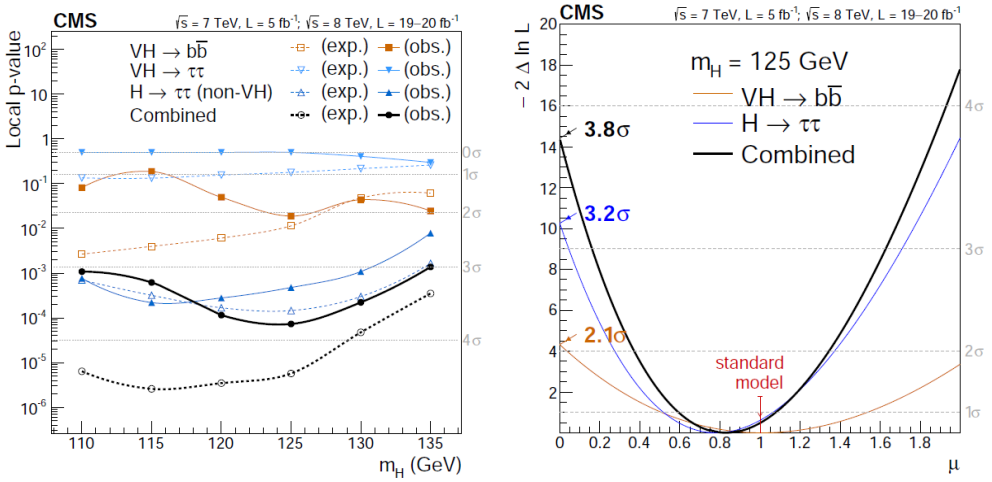

Figure 6.9.: Observed and expected local p-values (left) of the background-only prediction to describe the observed data and profile likelihood scan (right) for the signal strength parameter separately for the inclusive $H \to \tau\tau$, $VH \to \tau\tau$ and $VH \to V + b\bar{b}$ channels and the combination of all three analysis.

7. Conclusions and Outlook

The first dedicated search with the CMS Experiment for $H \to \tau\tau \to 2e4\nu$ decays has been performed and presented in this thesis. The search is performed using the full 2011 and 2012 dataset recorded with the CMS experiment, corresponding to an integrated luminosity of 4.9 fb^{-1} and 19.7 fb^{-1} at 7 TeV and 8 TeV, respectively.

The analysis strategy for the $H \to \tau\tau$ same-flavor ee and $\mu\mu$ lepton channels has been re-designed and a highly performant Drell-Yan background rejection method based on boosted decision trees was developed. A new approach to combine several boosted decision trees into one final discriminant obeying the discrimination against the two $Z \to ee$ and $Z \to \tau\tau$ backgrounds has been explored and applied to the analysis. Tau decay length information, the distance of closest approach between the two lepton tracks, was used to evaluate the Drell-Yan background in a data-driven way. In addition, a new electron identification algorithm based on boosted decision trees excluding impact parameter information has been developed.

The ee-channel alone can exclude $3.2 \times (\sigma^H_{SM} \times BR(H \to \tau\tau))$ for $m_H = 120$ GeV/c^2 and $3.7 \times (\sigma^H_{SM} \times BR(H \to \tau\tau))$ for $m_H = 125$ GeV/c^2 at 95% confidence level. Both ee and $\mu\mu$ channels together exclude $2.1 \times (\sigma^H_{SM} \times BR(H \to \tau\tau))$ for $m_H = 125$ GeV/c^2 at 95% confidence level.

The ee-channel is part of the CMS Collaboration wide $H \to \tau\tau$ search and combined with more sensitive final states, to observe, for the first time, direct experimental evidence for couplings of the Higgs boson to leptons. The observed significance was found to be 3.2σ with a best-fit signal strength of $\mu = 0.78 \pm 0.27$ for the Higgs boson hypothesis with $m_H = 125$ GeV/c^2.

Furthermore, the results of the $H \to \tau\tau$ search were combined with the CMS $VH \to b\bar{b}$ analysis. The combined result supports the findings of the $H \to \tau\tau$ channel alone, and results in strong evidence for couplings of the Higgs boson to down-type fermions with an observed significance of 3.8σ, where 4.4σ are expected.

In 2015 the LHC will run with a center-of-mass energy of 13 TeV, increasing the Higgs boson production cross-section by a factor of more than 2. The 5σ discovery in the $H \to \tau\tau$ channel can therefore be expected in about 2016, however demanding environmental conditions at higher instantaneous luminosities during RunII have to be managed. Trigger strategies and thresholds, especially for hadronic taus, have to be adjusted and increased pile-up effects have to be taken into account and studied.

With higher integrated luminosity (300-3000 fb^{-1}), measurements of the CP

properties of the Higgs boson may become feasible with the $H \to \tau\tau$ channel. Jet correlation variables may be used to study CP properties of Higgs coupling to bosons [106], but also fermionic Higgs couplings, which may be affected by mixed CP odd states at leading order, can be studied and measured using properties of 1-prong hadronic tau decays [107].

Performances of the ee- and $\mu\mu$-channel may be improved by considering the full 2D distributions of the two boosted decision trees in the maximum likelihood fit to derive CLs upper limits.

Further, the impact of using the TrigNoIP electron ID (cf. 4.3.2) in the $e\tau$ and ee channels should be studied to get an enhanced significance for the $H \to \tau\tau$ signal with electrons in the final state.

Appendices

A. MET Recoil Fits

Further information on the E_T^miss corrections, as described in section 4.6, applied to simulated samples, where hadronic recoils are mis-modeled, is given in this appendix. The method is as follows:

- The method exploits $Z \to ee$ events.
- Notations $u_{||}$ ($u_\perp$) are used for the balance $u_{||} = \hat{u}_{||} - q_T^Z$ ($u_\perp = \hat{u}_\perp$) recoil momentum projected on the axis parallel (perpendicular) to the transverse momentum of the Z-boson. $u_{||}$ and $u_\perp$ therefore the $\vec{E}_T^{miss}$ components parallel and perpendicular to the transverse Z momentum vector.
- The $u_{||}$, $u_\perp$ distributions are fitted in data and $Z \to ee$ MC sample in different Z p_T bins and jet multiplicity, N_{jets}, bins.
- The $u_{||}$ distributions are fitted with a double asymmetric Gaussian.
- The $u_\perp$ distributions are fitted with a double symmetric Gaussian.
- Cumulative functions $F_{\text{DATA}}(u_{||,\perp})$ and $F_{\text{MC}}^{Z\to ee}(u_{||,\perp})$ for given values of Z p_T and N_{jets}.
- Recoil corrections are performed via monotonic isomorphic mapping:
- $u_{||,\perp}^{\text{corr}} = F_{\text{DATA}}^{-1}(F_{\text{MC}}^{Z\to ee}(u_{||,\perp}^{\text{uncorr}}))$.

After the recoil correction via isomorphic mapping, the projected momenta of the simulated E_T^miss or recoil are observed to match excellently with the ones from experimental data.

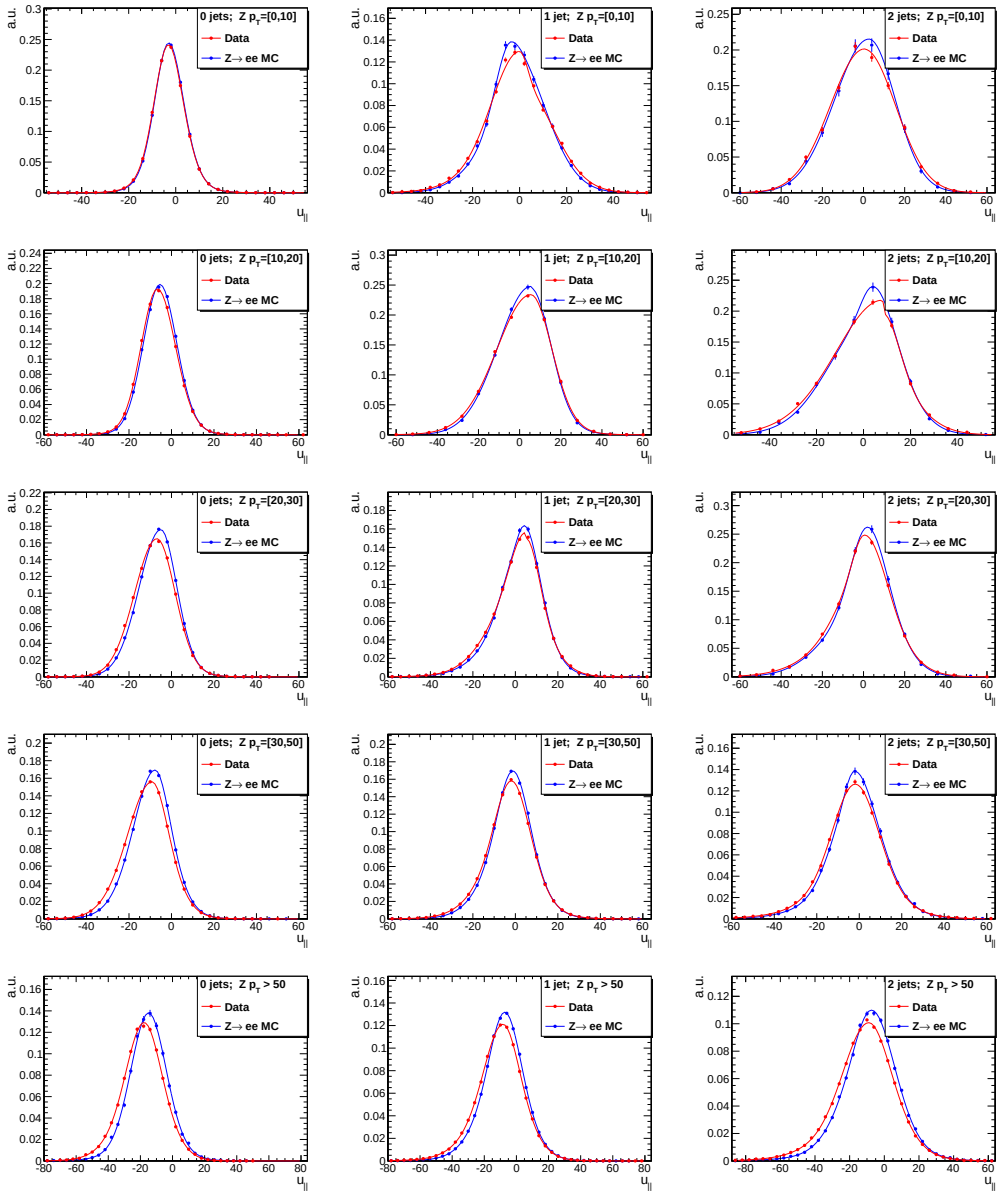

Figure A.1.: Uncorrected MET component perpendicular to the Z boost $u_{||}$ in 5×3 bins of the Z p_T and number of jets N_{Jets}.

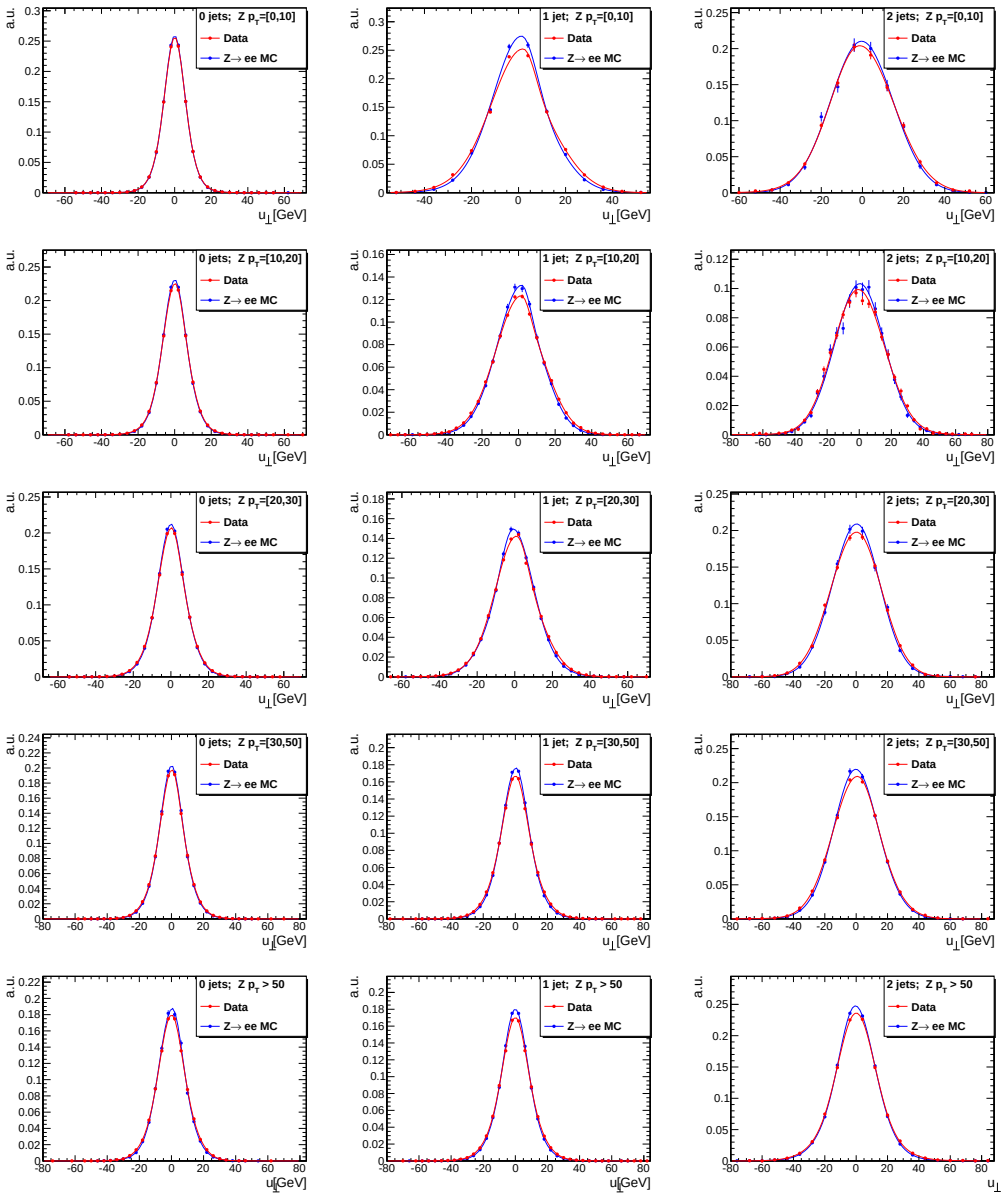

Figure A.2.: Uncorrected MET component perpendicular to the Z boost $u_\perp$ in 5×3 bins of the Z p_T and number of jets N_{Jets}.

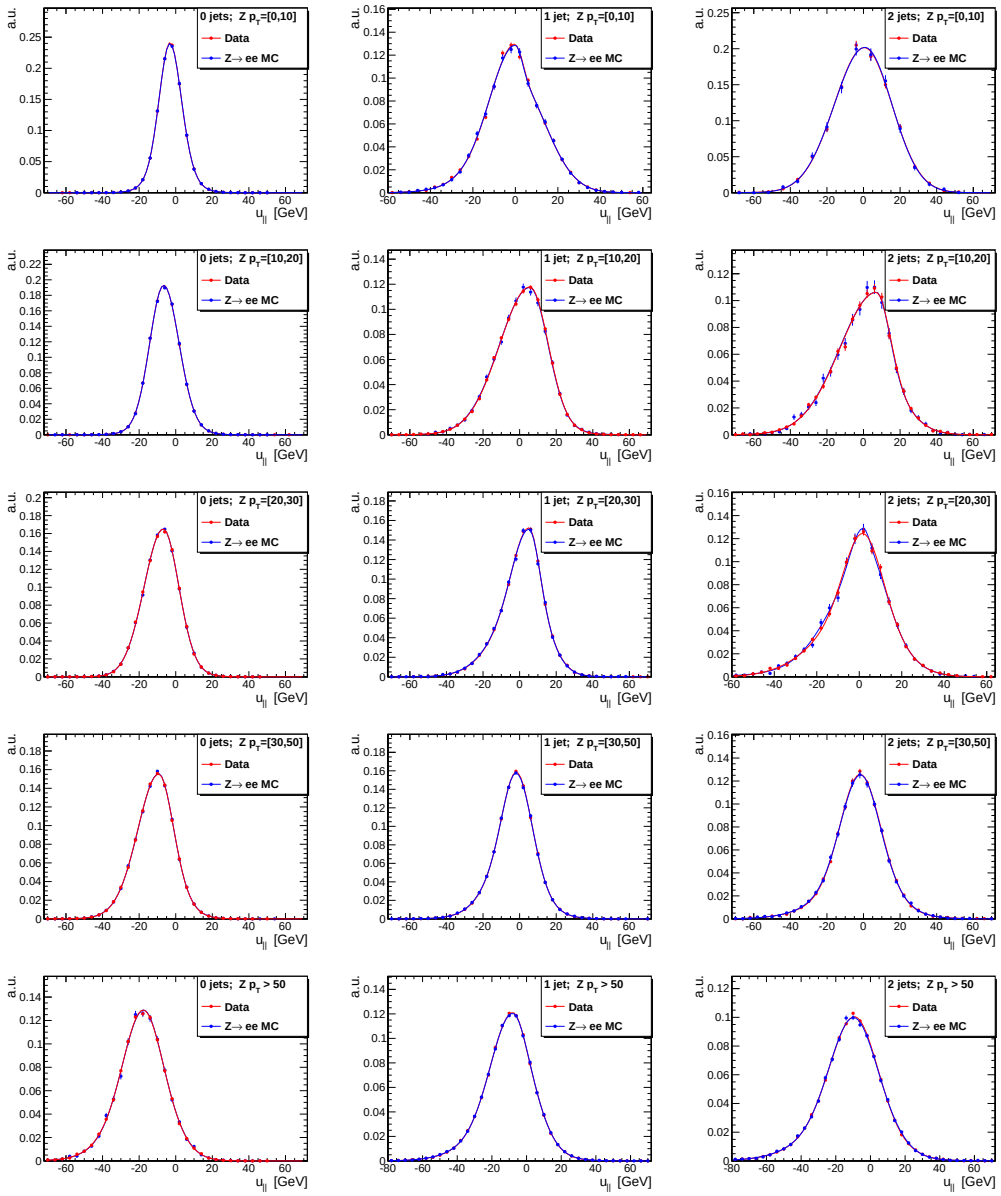

Figure A.3.: Isomorphic mapping corrected MET component parallel to the Z boost $u_{||}$ in 5×3 bins of the Z p_T and number of jets N_{Jets}.

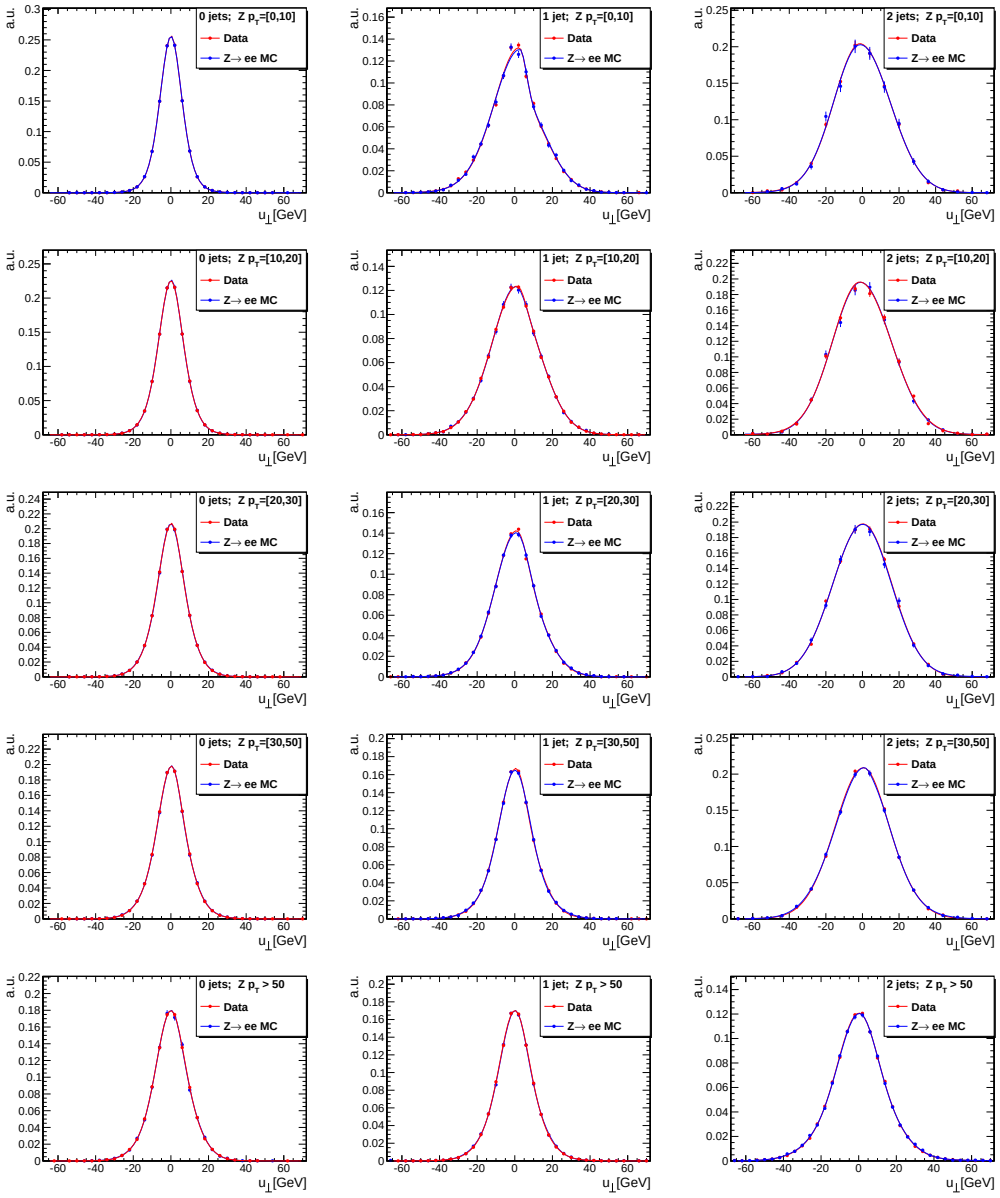

Figure A.4.: Isomorphic mapping corrected MET component perpendicular to the Z boost $u_\perp$ in 5×3 bins of the Z p_T and number of jets N_{Jets}.

B. Selection efficiency fits

This appendix collects the fits produced to determine the electron selection efficiency in data as, discussed in section 5.1.1.

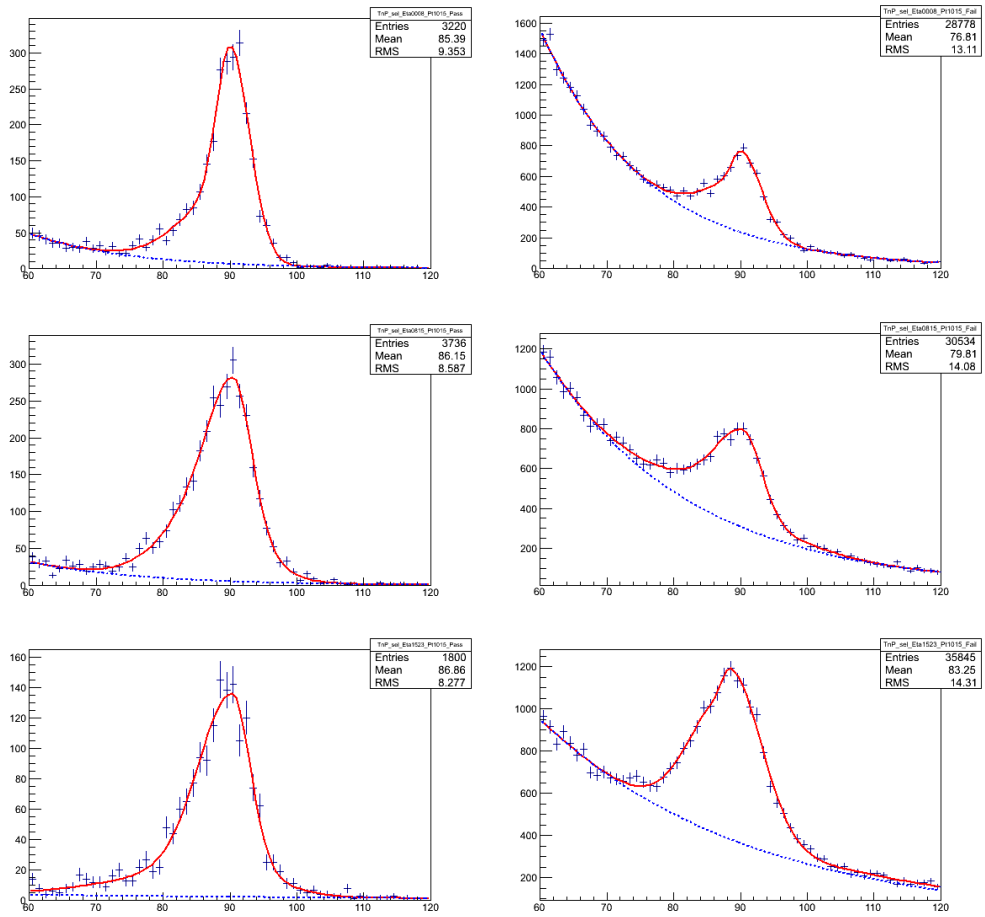

Figure B.1.: Illustration of fits of the tag and probe invariant mass in bins of probe p_T =10-15 GeV/c and $|\eta|$ <0.8, 0.8< $|\eta|$ <1.479, 1.479< $|\eta|$ <2.3 (top to bottom), for ID+Iso requirement passing (left) and failing (right) probes in data. The data spectra are fitted with the sum of a signal (solid line) and background (dashed line) model.

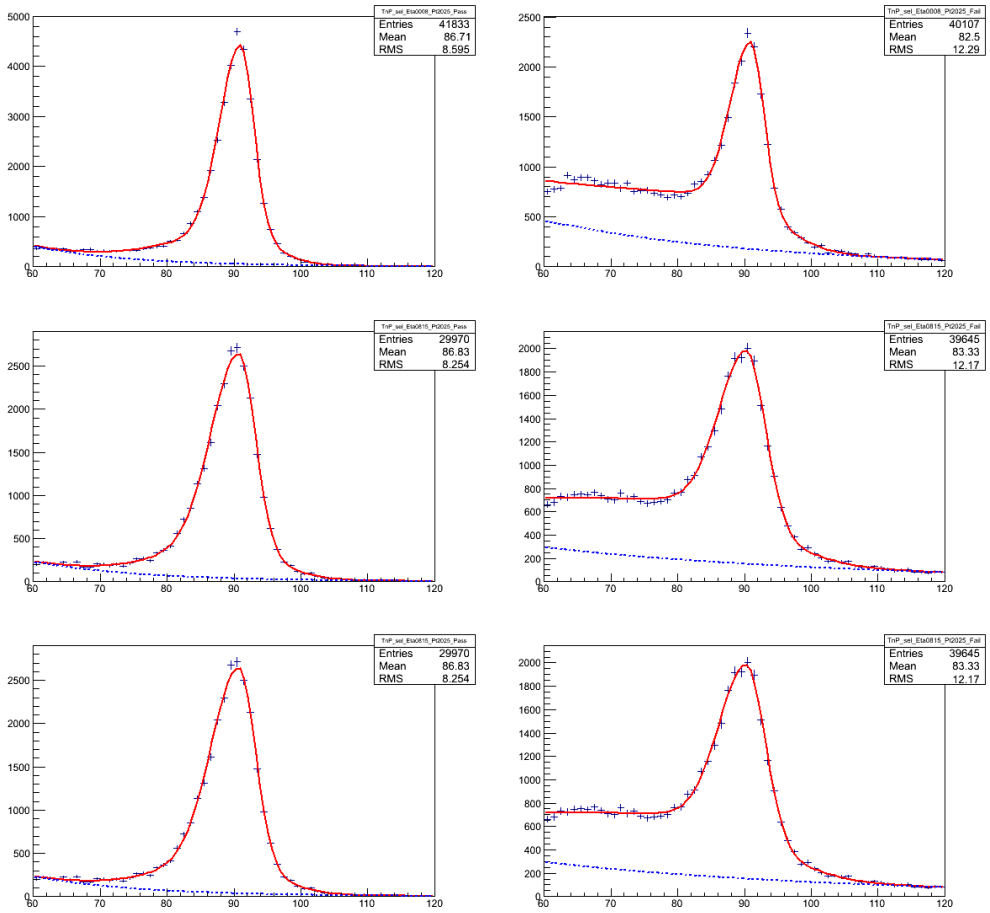

Figure B.2.: Illustration of fits of the tag and probe invariant mass in bins of passing (left) and failing (right) probe p_T =20-25 GeV/c and $|\eta|$ <0.8, 0.8< $|\eta|$ <1.479, 1.479< $|\eta|$ <2.3 (top to bottom), for ID+Iso requirement passing (left) and failing (right) probes in data. The data spectra are fitted with the sum of a signal (solid line) and background (dashed line) model.

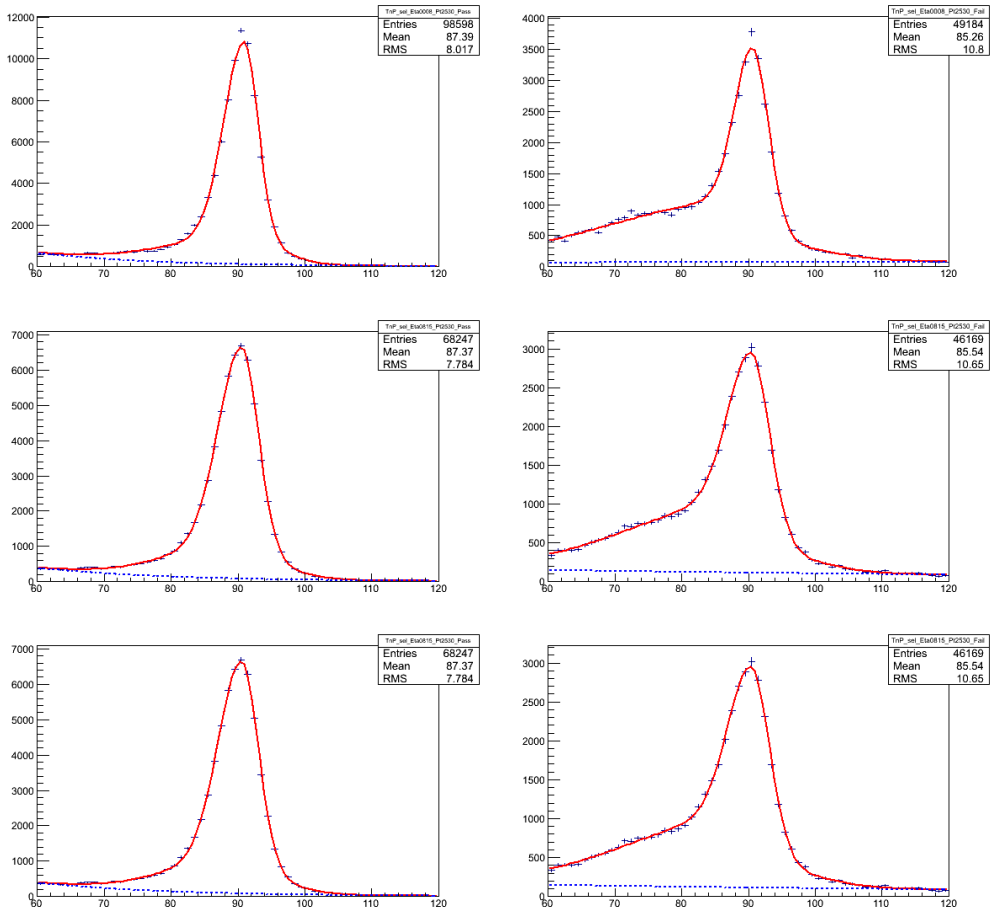

Figure B.3.: Illustration of fits of the tag and probe invariant mass in bins of passing (left) and failing (right) probe p_T =25-30 GeV/c and $|\eta|$ <0.8, 0.8< $|\eta|$ <1.479, 1.479< $|\eta|$ <2.3 (top to bottom), for ID+Iso requirement passing (left) and failing (right) probes in data. The data spectra are fitted with the sum of a signal (solid line) and background (dashed line) model.

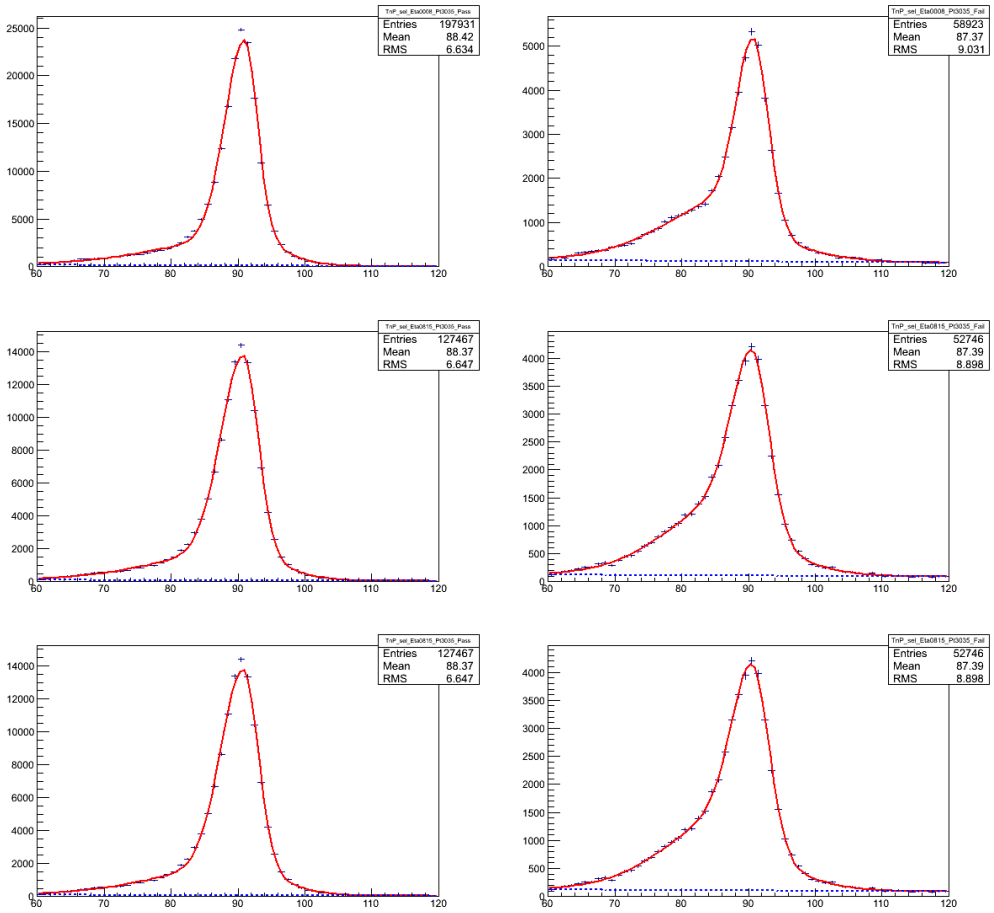

Figure B.4.: Illustration of fits of the tag and probe invariant mass in bins of passing (left) and failing (right) probe p_T =30-35 GeV/c and $|\eta|$ <0.8, 0.8< $|\eta|$ <1.479, 1.479< $|\eta|$ <2.3 (top to bottom), for ID+Iso requirement passing (left) and failing (right) probes in data. The data spectra are fitted with the sum of a signal (solid line) and background (dashed line) model.

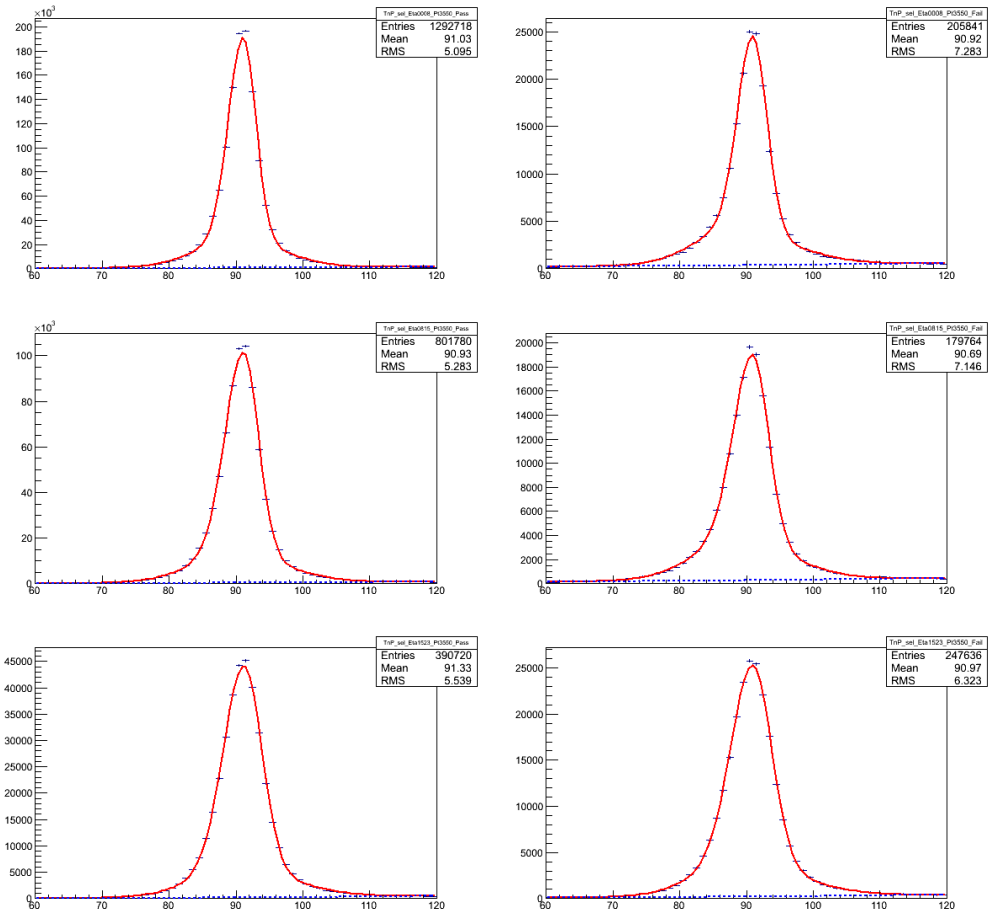

Figure B.5.: Illustration of fits of the tag and probe invariant mass in bins of passing (left) and failing (right) probe p_T =35-50 GeV/c and $|\eta|$ <0.8, 0.8< $|\eta|$ <1.479, 1.479< $|\eta|$ <2.3 (top to bottom), for ID+Iso requirement passing (left) and failing (right) probes in data. The data spectra are fitted with the sum of a signal (solid line) and background (dashed line) model.

C. Correlations of BDT input variables

This chapter compiles a representative collection of plots showing the correlation of BDT input variables discussed in Section 5.2.1 and 5.2.2. Displayed are Profile plots, showing the mean and RMS of one variable in bins of the other. In particular, the correlation in data is compared to the correlation in the predicted background. There is overall good agreement, showing that also the correlation of all input variables of the BDTs are well predicted in the analysis.

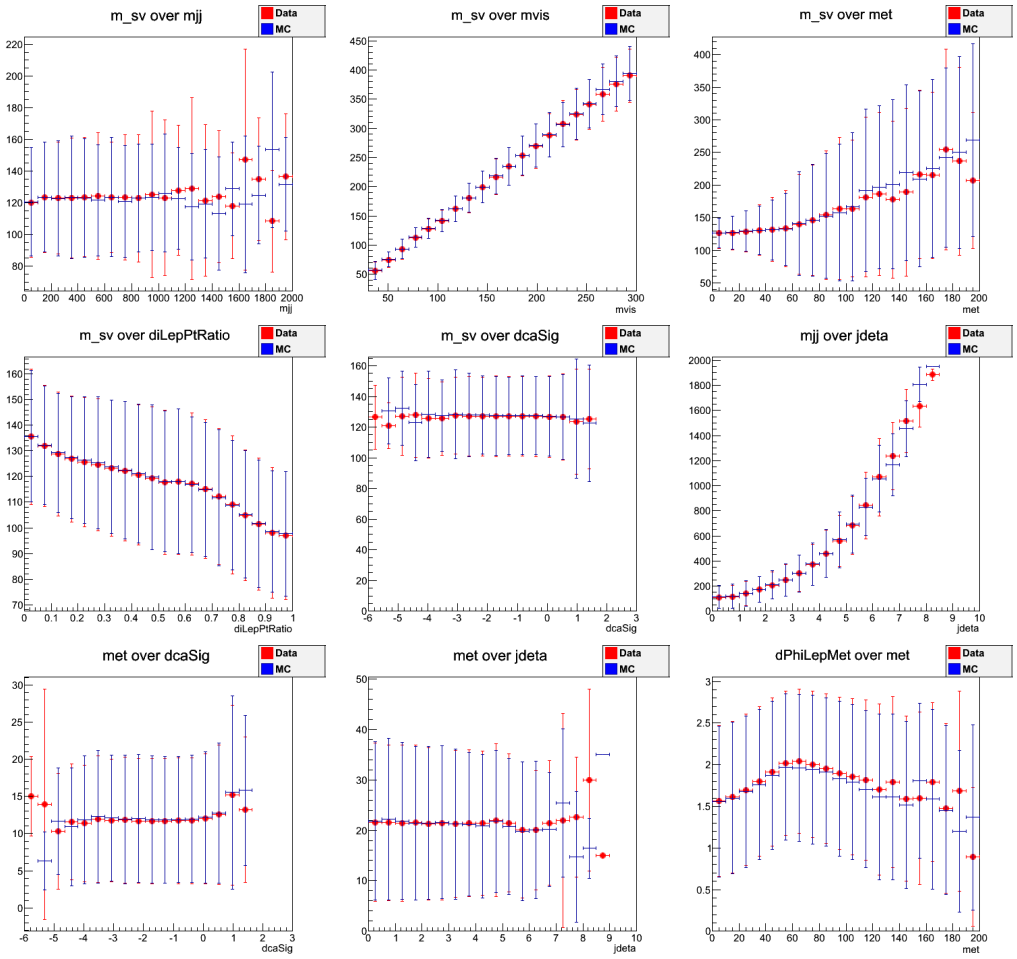

Figure C.1.: Profile plots for BDT input variables, described in Section 5.2.1 and 5.2.2, showing the mean and RMS of one variable in bins of the other.

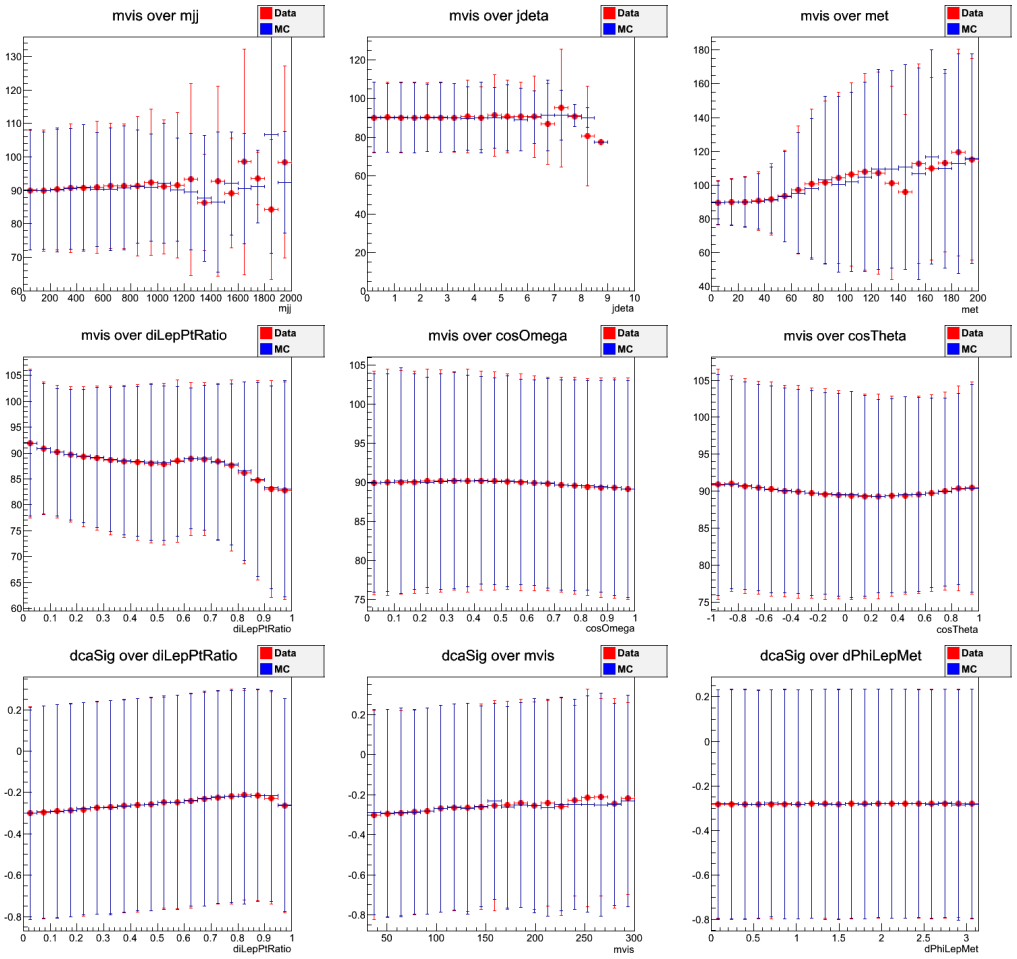

Figure C.2.: Profile plots for BDT input variables, described in Section 5.2.1 and 5.2.2, showing the mean and RMS of one variable in bins of the other.

D. Electron ID Comparison

A very preliminary study of the effect of the TrigNoIP Electron ID, discussed in Section 4.3.2, is shown in this chapter. The impact was estimated by exchanging the electron IDs only. As a first study, the same data-to-MC selection efficiency scale factors were kept and the background predictions, described in section 5.4, were assumed to scale trivially under the exchange of electron IDs. In addition, no new boosted decision tree training was done, neglecting the p_T dependent efficiency improvement and enhanced signal significance used for node splitting. Cuts on the BDT output were made using the working point shown in Table 4.3.

The expected exclusion limit is shown in Figure D.1. Improvements of roughly 10% in the low mass region $m_H < 120$ GeV/c^2 can be observed, however for $m_H \geq 120$ GeV/c^2, the improvement is only marginal.

Although, the electron ID has been shown to improve the performance of the ee-channel, more elaborate studies are needed to evaluate the possible gain by re-optimizing the analysis and reconsidering the full background prediction and efficiency measurements.

Also, the impact on the $e\tau$-channel has to be evaluated, where a tighter electron ID cut is used and therefore larger gain can be expected.

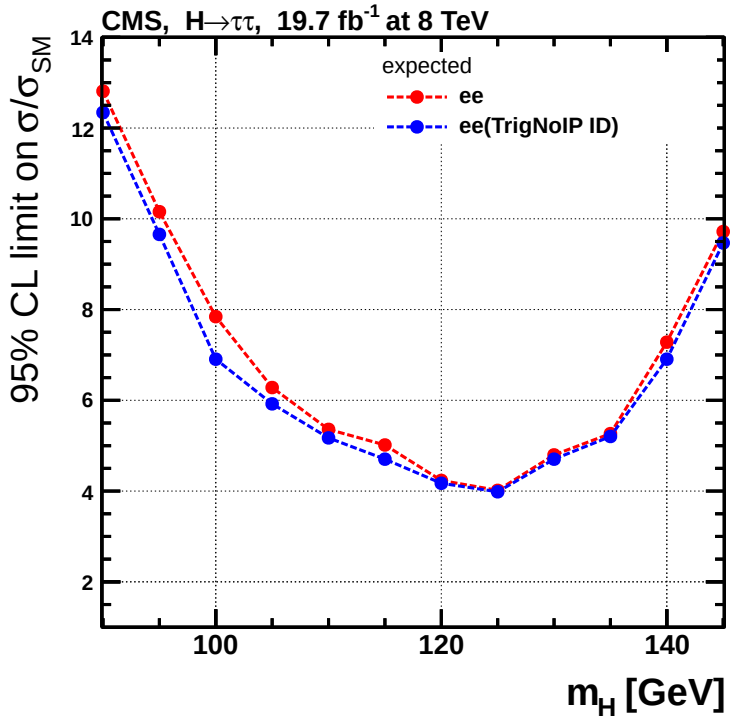

Figure D.1.: Comparison of expected 95% C.L. upper exclusion limits on the signal strength parameter for 8 TeV, for the TrigNoIP and NonTrig electron IDs, discussed in Section 4.3.2.

E. 7 TeV Analysis Control Plots and Limits

7 TeV control plots and observed and expected limits per category, are shown in this appendix. Electron transverse momentum and η spectrum is displayed in Figure E.1. Jet multiplicities are shown in Figure E.2. Figures E.4 and E.3 show variables used for the multivariate analysis strategy in the 0Jet and Boost categories. In Figures E.6 and E.5 BDT input variables in the VBF category are shown.

7 TeV observed and expected limits are shown in E.7.

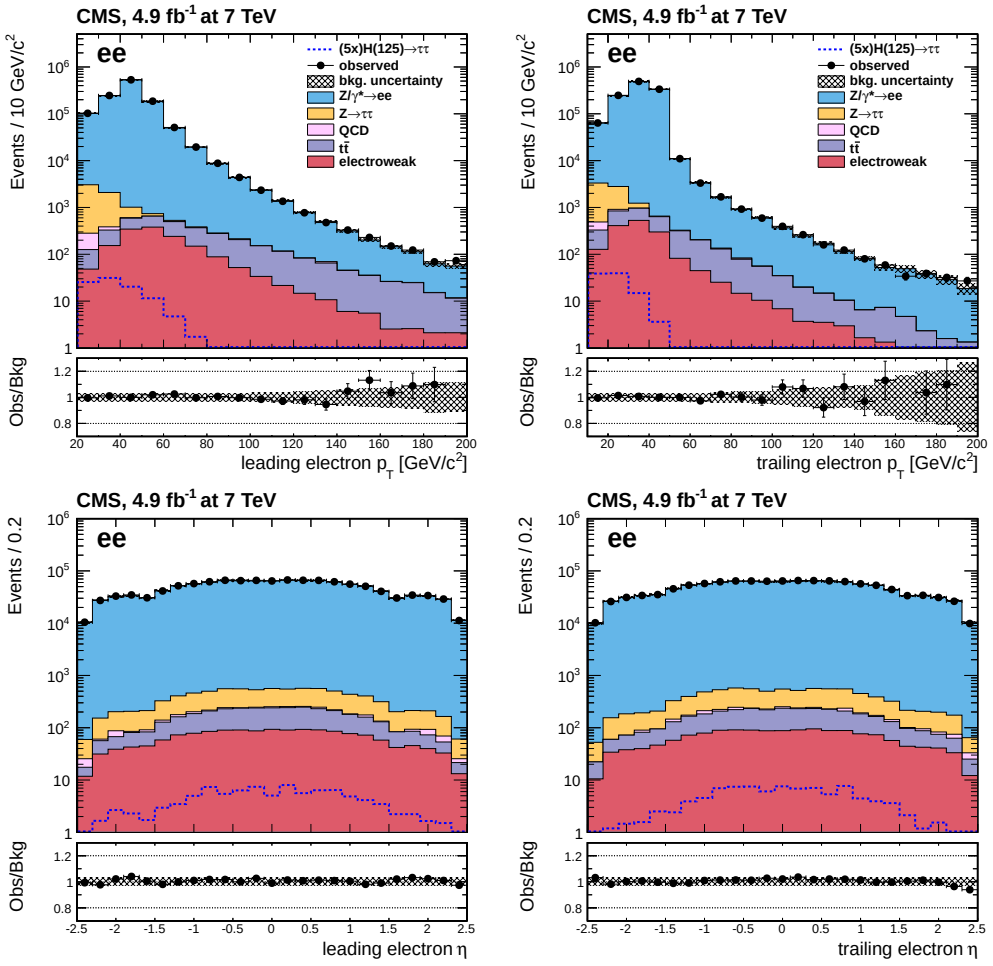

Figure E.1.: Distributions of the leading and trailing electron transverse momentum (top) and pseudorapidity (bottom). Experimental data (7 TeV) are shown as circles, predicted background samples are represented by filled histograms. Also displayed as dashed histograms are the signal expected for the SM Higgs boson with $m_H = 125$ GeV/c^2 and scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

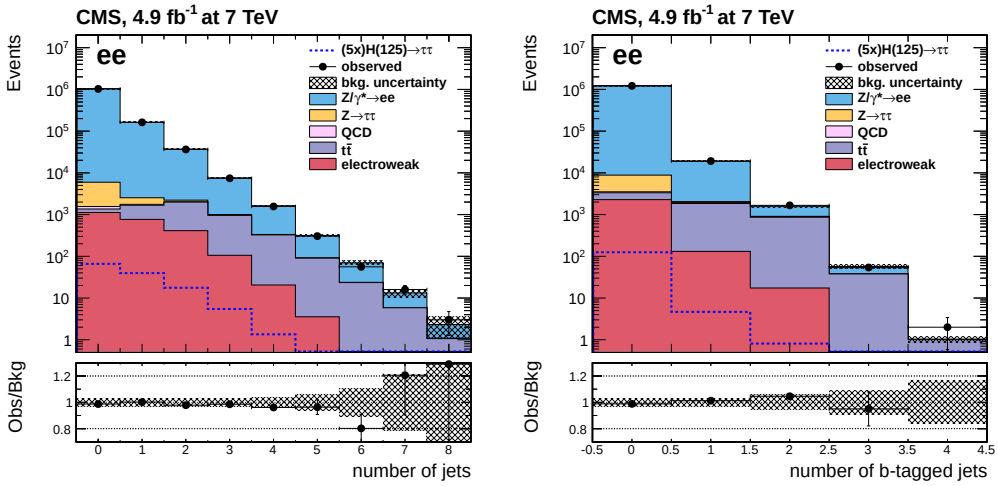

Figure E.2.: Left plot: multiplicity of jets after preselection of the di-electron events; right plot: multiplicity of b-tagged jets.Experimental data (7 TeV) are shown as circles, predicted background samples are represented by filled histograms. Also displayed as dashed histograms are the signal expected for the SM Higgs boson with $m_H = 125$ GeV/c^2 and scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

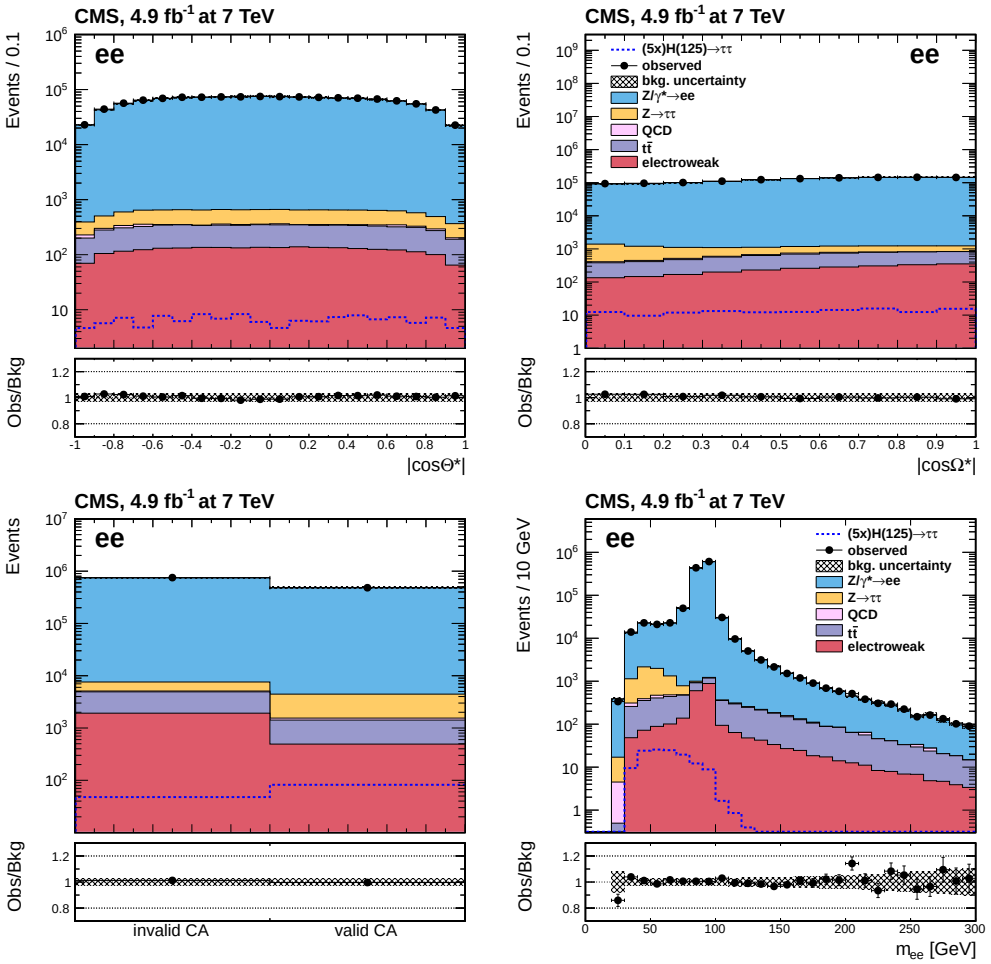

Figure E.3.: Distributions of discriminating variables used to in the multi-variate analysis in the *0-Jet* and *Boost* categories. Data (circles) 7 TeV is compared with predicted background samples (filled histograms). Also shown is the SM Higgs boson signal with $m_H = 125 GeV/c^2$ scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

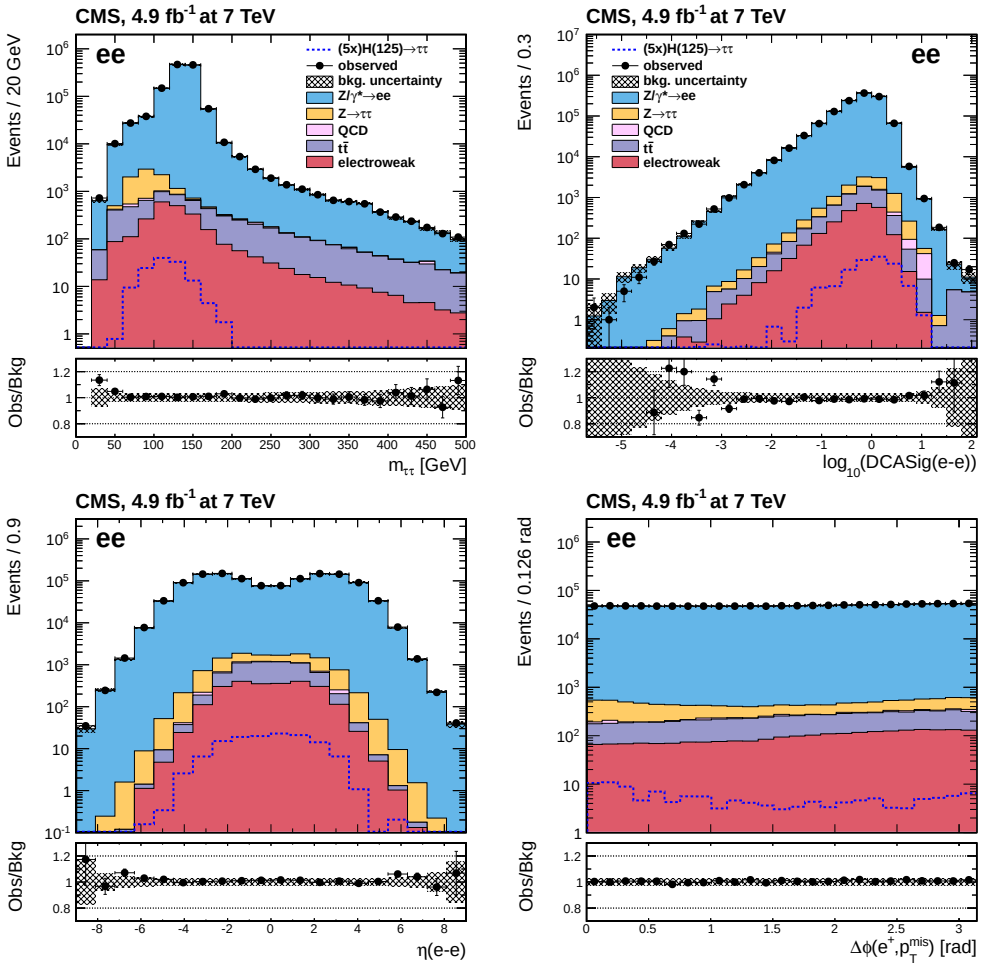

Figure E.4.: Distributions of discriminating variables used to in the multi-variate analysis in the *0-Jet* and *Boost* categories. Data (circles) 7 TeV is compared with predicted background samples (filled histograms). Also shown is the SM Higgs boson signal with $m_H = 125 GeV/c^2$ scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

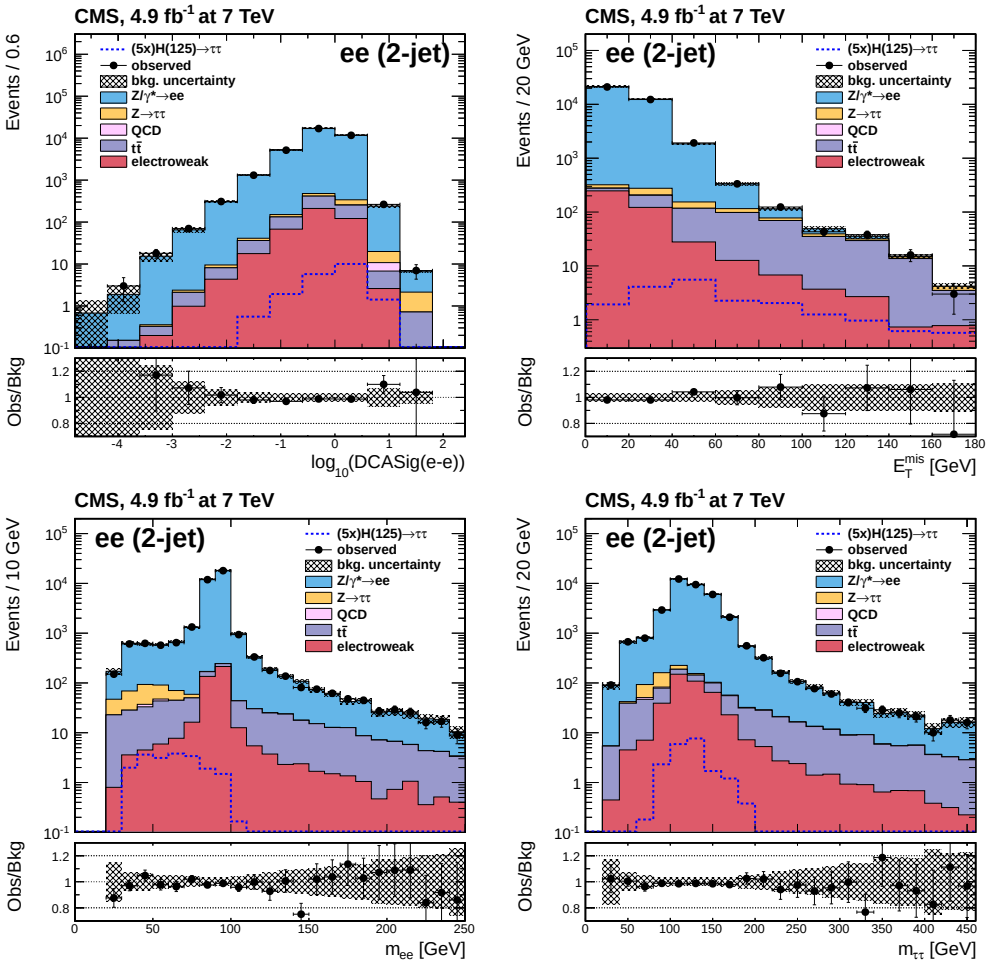

Figure E.5.: Distributions of variables used in the multi-variate analysis in the *VBF* event category in the di-electron channel. Experimental data (7 TeV) are shown as circles, predicted background samples are represented by filled histograms. Also displayed as dashed histograms are the signal expected for the SM Higgs boson with $m_H = 125$ GeV/c^2 and scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

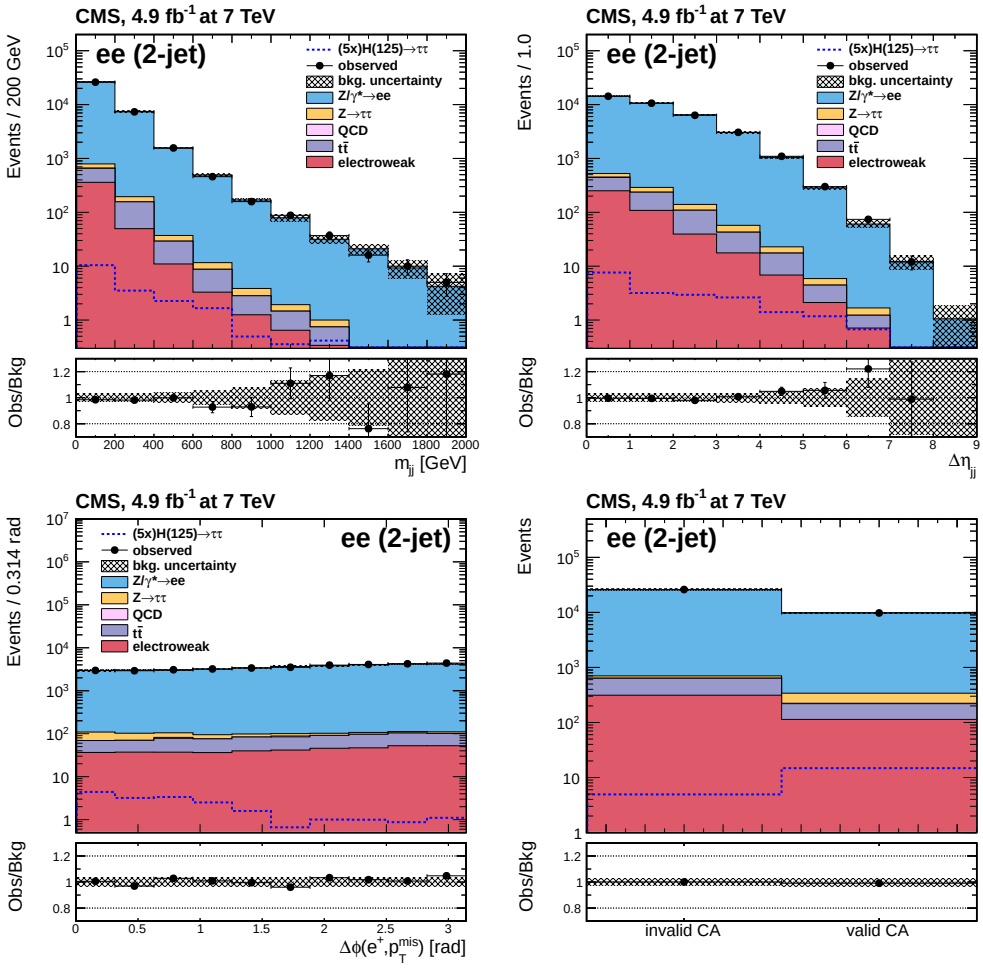

Figure E.6.: Distributions of variables used in the multi-variate analysis in the *VBF* event category. Experimental data (7 TeV) is shown as circles, predicted background samples are represented by filled histograms. Also displayed as dashed histograms are the signal expected for the SM Higgs boson with $m_H = 125$ GeV/c^2 and scaled to the SM cross section multiplied by a factor of 5. Error bars in ratio plots show the quadratic sum of statistical and systematic uncertainties.

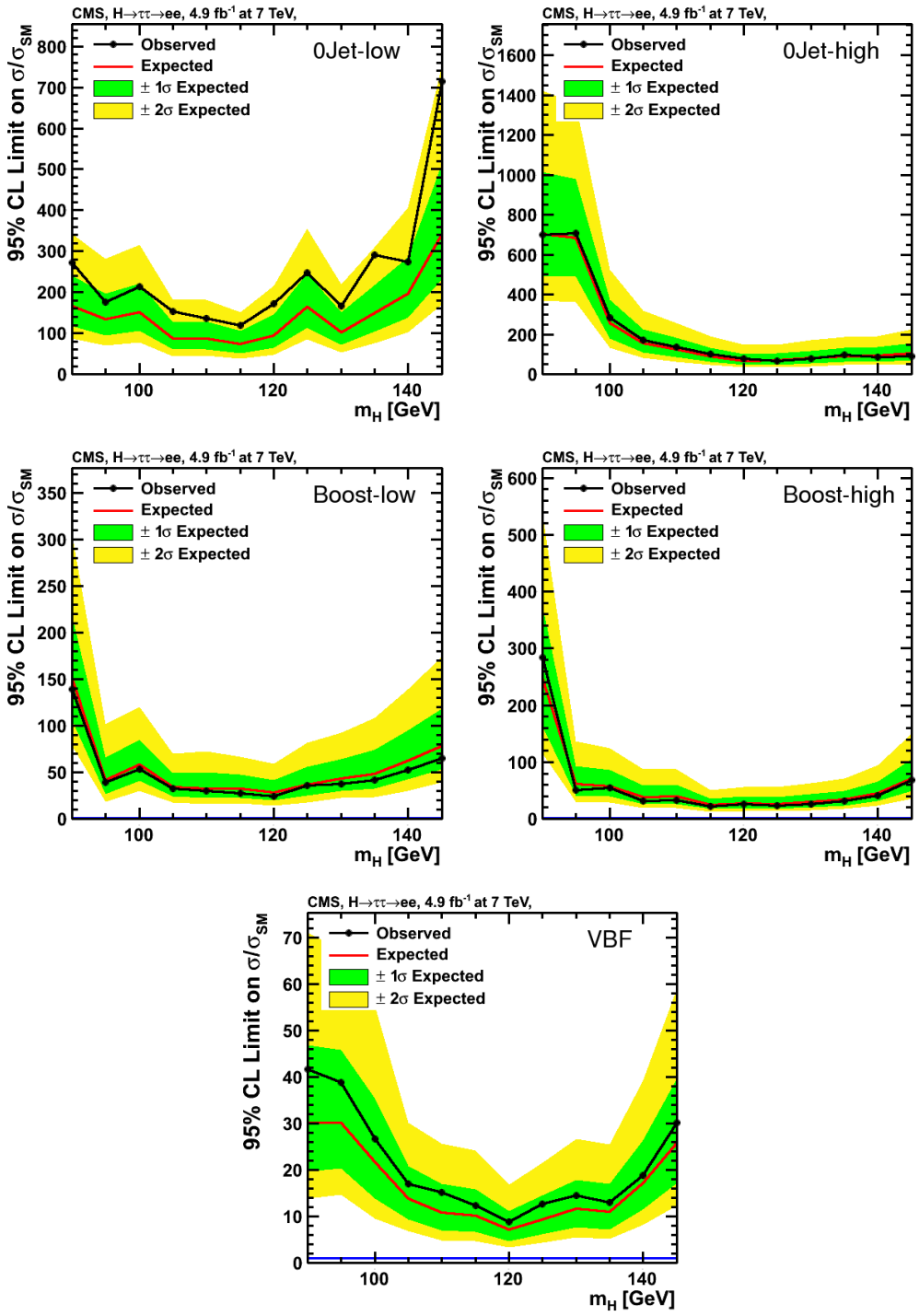

Figure E.7.: Observed and expected 95% C.L. upper exclusion limits on the signal strength for the 0Jet-low (upper-left), 0Jet-high (upper right), Boost-low (middle-left), Boost-high (middle right) and VBF (bottom) categories of the $H \to \tau\tau$ ee-channel for 7 TeV data-taking period.

F. Pulls ee-Channel

This chapter compiles the complete lists of pulls on the nuisance parameters, i.e. systematic uncertainties, discussed in Section 5.5. Added, are bin-by-bin uncertainties for each template, where uncertainties are merged in case the statistical uncertainty of a sample in one bin is below the threshold of 10% of the overall statistical uncertainties in that bin.

No large pulls above 2σ are observed and all pulls are well constrained around 1σ, suggesting that alll uncertainties are well taken into account. The largest pulls above 1σ are $CMS\_htt\_ee\_ee\_1jet\_high\_8TeV\_ZEE\_bin\_2$ (bin-by-bin uncertainty) with $+1.16 \pm 0.59$ and $CMS\_htt\_ee\_zeeShape\_0jet\_low\_mass0\_8TeV$ (uncertainty on the Drell-Yan background estimation (Section 5.4.5) at low mass in the 0Jet-low category) with -1.31 ± 0.81.

name	pre fit	b-only fit	$s+b$ fit	$\rho(\theta,\mu)$
CMS_eff_b_7TeV	0.00 ± 1.00	-0.02 ± 0.99 (-0.02σ, 0.99)	-0.02 ± 0.97 (-0.02σ, 0.98)	-0.00
CMS_eff_b_8TeV	0.00 ± 1.00	-0.06 ± 1.01 (-0.06σ, 1.01)	-0.06 ± 0.99 (-0.06σ, 1.00)	+0.01
CMS_eff_e	0.00 ± 1.00	-0.20 ± 0.67 (-0.20σ, 0.67)	-0.21 ± 0.69 (-0.21σ, 0.69)	-0.30
CMS_htt_ee_DiBosonNorm_7TeV	0.00 ± 1.00	-0.04 ± 0.96 (-0.04σ, 0.96)	-0.04 ± 0.97 (-0.04σ, 0.97)	-0.01
CMS_htt_ee_DiBosonNorm_8TeV	0.00 ± 1.00	$+0.30 \pm 0.89$ (+0.30σ, 0.90)	$+0.30 \pm 0.88$ (+0.30σ, 0.88)	-0.05
CMS_htt_ee_QCDNorm_0jet_high_7TeV	0.00 ± 1.00	-0.02 ± 0.99 (-0.02σ, 0.99)	-0.02 ± 0.99 (-0.02σ, 0.99)	-0.01
CMS_htt_ee_QCDNorm_0jet_high_8TeV	0.00 ± 1.00	-0.01 ± 0.97 (-0.01σ, 0.97)	-0.01 ± 1.01 (-0.01σ, 1.02)	+0.00
CMS_htt_ee_QCDNorm_0jet_low_7TeV	0.00 ± 1.00	$+0.12 \pm 1.00$ (+0.12σ, 1.00)	$+0.12 \pm 1.00$ (+0.12σ, 1.00)	-0.00
CMS_htt_ee_QCDNorm_0jet_low_8TeV	0.00 ± 1.00	-0.18 ± 0.97 (-0.18σ, 0.97)	-0.18 ± 0.97 (-0.18σ, 0.98)	+0.02
CMS_htt_ee_QCDNorm_1jet_high_7TeV	0.00 ± 1.00	-0.03 ± 0.98 (-0.03σ, 0.98)	-0.03 ± 0.98 (-0.03σ, 0.98)	-0.02
CMS_htt_ee_QCDNorm_1jet_high_8TeV	0.00 ± 1.00	$+0.03 \pm 0.99$ (+0.03σ, 1.00)	$+0.03 \pm 0.99$ (+0.03σ, 0.99)	-0.02
CMS_htt_ee_QCDNorm_1jet_low_7TeV	0.00 ± 1.00	-0.13 ± 0.96 (-0.13σ, 0.96)	-0.13 ± 0.96 (-0.13σ, 0.96)	+0.00
CMS_htt_ee_QCDNorm_1jet_low_8TeV	0.00 ± 1.00	$+0.05 \pm 0.98$ (+0.05σ, 0.99)	$+0.05 \pm 0.99$ (+0.05σ, 0.99)	-0.02
CMS_htt_ee_QCDNorm_vbf_7TeV	0.00 ± 1.00	-0.00 ± 0.99 (-0.00σ, 0.99)	-0.00 ± 0.99 (-0.00σ, 0.99)	-0.01
CMS_htt_ee_QCDNorm_vbf_8TeV	0.00 ± 1.00	-0.00 ± 0.99 (-0.00σ, 0.99)	-0.00 ± 0.99 (-0.00σ, 0.99)	+0.01
CMS_htt_ee_ee_0jet_high_7TeV_ZEE_bin_1	0.00 ± 0.99	-0.00 ± 0.99 (-0.00σ, 1.00)	-0.00 ± 0.98 (-0.00σ, 0.99)	-0.01
CMS_htt_ee_ee_0jet_high_7TeV_ZEE_bin_2	0.00 ± 0.99	$+0.03 \pm 0.88$ (+0.03σ, 0.89)	$+0.03 \pm 0.90$ (+0.03σ, 0.91)	+0.02
CMS_htt_ee_ee_0jet_high_7TeV_ZEE_bin_3	0.00 ± 0.99	$+0.17 \pm 0.76$ (+0.17σ, 0.77)	$+0.17 \pm 0.76$ (+0.17σ, 0.77)	-0.01
CMS_htt_ee_ee_0jet_high_7TeV_ZEE_bin_4	0.00 ± 0.99	$+0.11 \pm 0.76$ (+0.11σ, 0.77)	$+0.11 \pm 0.76$ (+0.11σ, 0.77)	-0.00
CMS_htt_ee_ee_0jet_high_7TeV_ZEE_bin_5	0.00 ± 0.99	-0.56 ± 0.85 (-0.56σ, 0.86)	-0.56 ± 0.85 (-0.56σ, 0.86)	-0.01
CMS_htt_ee_ee_0jet_high_8TeV_ZEE_bin_1	0.00 ± 0.99	$+0.01 \pm 1.00$ (+0.01σ, 1.01)	$+0.01 \pm 1.01$ (+0.01σ, 1.02)	-0.04
CMS_htt_ee_ee_0jet_high_8TeV_ZEE_bin_2	0.00 ± 0.99	-0.26 ± 0.78 (-0.26σ, 0.78)	-0.26 ± 0.78 (-0.26σ, 0.79)	+0.01
CMS_htt_ee_ee_0jet_high_8TeV_ZEE_bin_3	0.00 ± 0.99	-0.12 ± 0.68 (-0.12σ, 0.69)	-0.12 ± 0.68 (-0.12σ, 0.69)	-0.04
CMS_htt_ee_ee_0jet_high_8TeV_ZEE_bin_4	0.00 ± 0.99	-0.03 ± 0.68 (-0.03σ, 0.68)	-0.03 ± 0.68 (-0.04σ, 0.68)	-0.04
CMS_htt_ee_ee_0jet_high_8TeV_ZTT_bin_5	0.00 ± 0.99	$+0.17 \pm 0.95$ (+0.18σ, 0.96)	$+0.17 \pm 0.99$ (+0.18σ, 1.00)	-0.02
CMS_htt_ee_ee_1jet_high_7TeV_ZEE_bin_1	0.00 ± 0.99	-0.03 ± 0.97 (-0.03σ, 0.98)	-0.03 ± 0.97 (-0.03σ, 0.98)	-0.01
CMS_htt_ee_ee_1jet_high_7TeV_ZEE_bin_2	0.00 ± 0.99	$+0.06 \pm 0.79$ (+0.06σ, 0.80)	$+0.06 \pm 0.80$ (+0.06σ, 0.80)	-0.01
CMS_htt_ee_ee_1jet_high_7TeV_ZEE_bin_3	0.00 ± 0.99	$+0.17 \pm 0.77$ (+0.17σ, 0.78)	$+0.17 \pm 0.77$ (+0.17σ, 0.78)	-0.02
CMS_htt_ee_ee_1jet_high_7TeV_ZEE_bin_4	0.00 ± 0.99	$+0.18 \pm 0.75$ (+0.18σ, 0.76)	$+0.18 \pm 0.76$ (+0.18σ, 0.77)	-0.03
CMS_htt_ee_ee_1jet_high_7TeV_ZEE_bin_5	0.00 ± 0.99	$+0.03 \pm 0.74$ (+0.03σ, 0.75)	$+0.03 \pm 0.74$ (+0.03σ, 0.75)	-0.02
CMS_htt_ee_ee_1jet_high_7TeV_ZEE_bin_6	0.00 ± 0.99	-0.29 ± 0.76 (-0.30σ, 0.76)	-0.29 ± 0.76 (-0.30σ, 0.77)	-0.01
CMS_htt_ee_ee_1jet_high_7TeV_ZEE_bin_7	0.00 ± 0.99	$+0.41 \pm 0.79$ (+0.41σ, 0.80)	$+0.41 \pm 0.80$ (+0.41σ, 0.80)	-0.08

name	pre fit	b-only fit	$s+b$ fit	$\rho(\theta,\mu)$
CMS_htt_ee_ee_1jet_high_7TeV_ZEE_bin_8	0.00 ± 0.99	-0.63 ± 0.84 (-0.63σ, 0.85)	-0.63 ± 0.84 (-0.63σ, 0.85)	-0.05
CMS_htt_ee_ee_1jet_high_7TeV_ZEE_bin_9	0.00 ± 0.99	-0.46 ± 0.86 (-0.46σ, 0.87)	-0.46 ± 0.86 (-0.46σ, 0.87)	-0.01
CMS_htt_ee_ee_1jet_high_8TeV_ZEE_bin_1	0.00 ± 0.99	-0.10 ± 0.98 (-0.10σ, 0.99)	-0.10 ± 0.98 (-0.10σ, 0.99)	-0.01
CMS_htt_ee_ee_1jet_high_8TeV_ZEE_bin_2	0.00 ± 0.99	$+1.16 \pm 0.59$ (+1.17σ, 0.60)	$+1.16 \pm 0.60$ (+1.17σ, 0.60)	+0.07
CMS_htt_ee_ee_1jet_high_8TeV_ZEE_bin_3	0.00 ± 0.99	-0.51 ± 0.76 (-0.52σ, 0.77)	-0.51 ± 0.76 (-0.52σ, 0.77)	+0.01
CMS_htt_ee_ee_1jet_high_8TeV_ZEE_bin_4	0.00 ± 0.99	$+0.66 \pm 0.75$ (+0.67σ, 0.75)	$+0.66 \pm 0.75$ (+0.67σ, 0.76)	-0.04
CMS_htt_ee_ee_1jet_high_8TeV_ZEE_bin_5	0.00 ± 0.99	$+0.23 \pm 0.78$ (+0.23σ, 0.78)	$+0.23 \pm 0.78$ (+0.23σ, 0.79)	-0.05
CMS_htt_ee_ee_1jet_high_8TeV_ZEE_bin_6	0.00 ± 0.99	$+0.53 \pm 0.80$ (+0.54σ, 0.80)	$+0.53 \pm 0.80$ (+0.54σ, 0.81)	-0.07
CMS_htt_ee_ee_1jet_high_8TeV_ZEE_bin_7	0.00 ± 0.99	-0.47 ± 0.76 (-0.48σ, 0.77)	-0.47 ± 0.76 (-0.48σ, 0.77)	-0.10
CMS_htt_ee_ee_1jet_high_8TeV_ZEE_bin_8	0.00 ± 0.99	-0.34 ± 0.94 (-0.35σ, 0.94)	-0.34 ± 0.94 (-0.35σ, 0.95)	-0.05
CMS_htt_ee_ee_1jet_high_8TeV_ZEE_bin_9	0.00 ± 0.99	-0.25 ± 0.78 (-0.25σ, 0.78)	-0.25 ± 0.79 (-0.26σ, 0.80)	-0.16
CMS_htt_ee_ee_vbf_7TeV_ZEE_bin_1	0.00 ± 0.99	$+0.01 \pm 0.98$ (+0.01σ, 0.99)	$+0.01 \pm 0.98$ (+0.01σ, 0.99)	+0.01
CMS_htt_ee_ee_vbf_7TeV_ZEE_bin_2	0.00 ± 0.99	-0.19 ± 0.84 (-0.19σ, 0.85)	-0.18 ± 0.84 (-0.19σ, 0.84)	+0.03
CMS_htt_ee_ee_vbf_7TeV_ZEE_bin_3	0.00 ± 0.99	$+0.33 \pm 0.85$ (+0.33σ, 0.85)	$+0.33 \pm 0.84$ (+0.33σ, 0.85)	-0.02
CMS_htt_ee_ee_vbf_7TeV_ZEE_bin_4	0.00 ± 0.99	-0.14 ± 0.86 (-0.14σ, 0.87)	-0.14 ± 0.86 (-0.14σ, 0.87)	-0.02
CMS_htt_ee_ee_vbf_7TeV_ZEE_bin_5	0.00 ± 0.99	-0.03 ± 0.93 (-0.03σ, 0.94)	-0.03 ± 0.93 (-0.03σ, 0.94)	-0.03
CMS_htt_ee_ee_vbf_7TeV_ZEE_bin_6	0.00 ± 0.99	$+0.31 \pm 0.95$ (+0.32σ, 0.96)	$+0.31 \pm 0.95$ (+0.31σ, 0.96)	-0.07
CMS_htt_ee_ee_vbf_7TeV_ZEE_bin_7	0.00 ± 0.99	-0.20 ± 0.95 (-0.20σ, 0.96)	-0.20 ± 0.95 (-0.20σ, 0.96)	-0.01
CMS_htt_ee_ee_vbf_7TeV_ZTT_bin_8	0.00 ± 0.99	-0.00 ± 0.98 (-0.00σ, 0.99)	-0.00 ± 0.98 (-0.00σ, 0.98)	-0.01
CMS_htt_ee_ee_vbf_7TeV_ZTT_bin_9	0.00 ± 0.99	$+0.18 \pm 0.96$ (+0.18σ, 0.97)	$+0.18 \pm 0.97$ (+0.18σ, 0.98)	-0.18
CMS_htt_ee_ee_vbf_8TeV_TTJ_bin_6	0.00 ± 0.99	$+0.09 \pm 0.97$ (+0.09σ, 0.98)	$+0.08 \pm 0.97$ (+0.08σ, 0.98)	-0.06
CMS_htt_ee_ee_vbf_8TeV_TTJ_bin_7	0.00 ± 0.99	$+0.08 \pm 0.97$ (+0.08σ, 0.98)	$+0.08 \pm 0.97$ (+0.08σ, 0.98)	-0.06
CMS_htt_ee_ee_vbf_8TeV_TTJ_bin_8	0.00 ± 0.99	-0.16 ± 0.99 (-0.17σ, 1.00)	-0.16 ± 0.98 (-0.17σ, 0.99)	+0.01
CMS_htt_ee_ee_vbf_8TeV_ZEE_bin_1	0.00 ± 0.99	$+0.00 \pm 0.98$ (+0.00σ, 0.99)	$+0.00 \pm 0.99$ (+0.00σ, 1.00)	+0.01
CMS_htt_ee_ee_vbf_8TeV_ZEE_bin_2	0.00 ± 0.99	-0.04 ± 0.85 (-0.04σ, 0.85)	-0.04 ± 0.84 (-0.04σ, 0.85)	+0.08
CMS_htt_ee_ee_vbf_8TeV_ZEE_bin_3	0.00 ± 0.99	-0.22 ± 0.85 (-0.22σ, 0.86)	-0.22 ± 0.85 (-0.22σ, 0.86)	+0.00
CMS_htt_ee_ee_vbf_8TeV_ZEE_bin_4	0.00 ± 0.99	$+0.06 \pm 0.93$ (+0.06σ, 0.94)	$+0.06 \pm 0.93$ (+0.06σ, 0.94)	-0.02
CMS_htt_ee_ee_vbf_8TeV_ZEE_bin_5	0.00 ± 0.99	$+0.58 \pm 0.88$ (+0.58σ, 0.88)	$+0.58 \pm 0.88$ (+0.58σ, 0.89)	-0.11
CMS_htt_ee_ee_vbf_8TeV_ZTT_bin_9	0.00 ± 0.99	-0.08 ± 0.99 (-0.08σ, 1.00)	-0.08 ± 0.99 (-0.08σ, 1.00)	+0.01
CMS_htt_ee_zeeNorm_0jet_high_7TeV	0.00 ± 1.00	-0.11 ± 0.54 (-0.11σ, 0.54)	-0.11 ± 0.55 (-0.11σ, 0.55)	+0.01
CMS_htt_ee_zeeNorm_0jet_high_8TeV	0.00 ± 1.00	$+0.04 \pm 0.53$ (+0.04σ, 0.53)	$+0.04 \pm 0.53$ (+0.04σ, 0.54)	+0.05
CMS_htt_ee_zeeNorm_0jet_low_7TeV	0.00 ± 1.00	$+0.05 \pm 0.52$ (+0.06σ, 0.52)	$+0.05 \pm 0.53$ (+0.06σ, 0.53)	+0.01
CMS_htt_ee_zeeNorm_0jet_low_8TeV	0.00 ± 1.00	$+0.10 \pm 0.15$ (+0.10σ, 0.16)	$+0.10 \pm 0.16$ (+0.10σ, 0.16)	+0.11
CMS_htt_ee_zeeNorm_1jet_high_7TeV	0.00 ± 1.00	-0.08 ± 0.72 (-0.08σ, 0.72)	-0.08 ± 0.73 (-0.08σ, 0.73)	+0.01

name	pre fit	b-only fit	$s+b$ fit	$\rho(\theta,\mu)$
CMS_htt_ee_zeeNorm_1jet_high_8TeV	0.00 ± 1.00	-0.03 ± 0.33 (-0.03σ, 0.33)	-0.03 ± 0.34 (-0.03σ, 0.34)	+0.12
CMS_htt_ee_zeeNorm_1jet_low_7TeV	0.00 ± 1.00	-0.00 ± 0.96 (-0.00σ, 0.97)	-0.00 ± 0.97 (-0.00σ, 0.97)	+0.02
CMS_htt_ee_zeeNorm_1jet_low_8TeV	0.00 ± 1.00	$+0.53 \pm 0.59$ (+0.53σ, 0.59)	$+0.53 \pm 0.60$ (+0.53σ, 0.60)	+0.04
CMS_htt_ee_zeeNorm_vbf_7TeV	0.00 ± 1.00	-0.05 ± 0.35 (-0.05σ, 0.35)	-0.05 ± 0.36 (-0.05σ, 0.36)	+0.05
CMS_htt_ee_zeeNorm_vbf_8TeV	0.00 ± 1.00	-0.02 ± 0.45 (-0.02σ, 0.45)	-0.02 ± 0.43 (-0.02σ, 0.43)	+0.13
CMS_htt_ee_zeeShape_0jet_high_mass0_7TeV	0.00 ± 0.99	$+0.21 \pm 0.52$ (+0.21σ, 0.53)	$+0.21 \pm 0.53$ (+0.21σ, 0.54)	-0.01
CMS_htt_ee_zeeShape_0jet_high_mass0_8TeV	0.00 ± 0.99	-0.32 ± 0.77 (-0.32σ, 0.78)	-0.32 ± 0.77 (-0.32σ, 0.78)	-0.01
CMS_htt_ee_zeeShape_0jet_high_mass1_7TeV	0.00 ± 0.99	-0.07 ± 0.86 (-0.07σ, 0.87)	-0.07 ± 0.88 (-0.07σ, 0.89)	-0.00
CMS_htt_ee_zeeShape_0jet_high_mass1_8TeV	0.00 ± 0.99	-0.01 ± 0.98 (-0.01σ, 0.99)	-0.01 ± 0.98 (-0.01σ, 0.99)	-0.00
CMS_htt_ee_zeeShape_0jet_high_mass2_7TeV	0.00 ± 0.99	$+0.01 \pm 0.98$ (+0.01σ, 0.99)	$+0.01 \pm 0.99$ (+0.01σ, 1.00)	-0.01
CMS_htt_ee_zeeShape_0jet_high_mass2_8TeV	0.00 ± 0.99	-0.01 ± 0.98 (-0.01σ, 0.99)	-0.01 ± 1.00 (-0.01σ, 1.01)	-0.01
CMS_htt_ee_zeeShape_0jet_low_mass0_7TeV	0.00 ± 0.99	-0.12 ± 0.68 (-0.13σ, 0.69)	-0.12 ± 0.68 (-0.12σ, 0.69)	+0.05
CMS_htt_ee_zeeShape_0jet_low_mass0_8TeV	0.00 ± 0.99	-1.31 ± 0.81 (-1.32σ, 0.82)	-1.30 ± 0.82 (-1.32σ, 0.82)	+0.08
CMS_htt_ee_zeeShape_0jet_low_mass1_7TeV	0.00 ± 0.99	$+0.01 \pm 0.89$ (+0.01σ, 0.90)	$+0.01 \pm 0.91$ (+0.01σ, 0.92)	-0.02
CMS_htt_ee_zeeShape_0jet_low_mass1_8TeV	0.00 ± 0.99	-0.00 ± 0.98 (-0.00σ, 0.99)	-0.00 ± 1.01 (-0.00σ, 1.02)	-0.01
CMS_htt_ee_zeeShape_0jet_low_mass2_7TeV	0.00 ± 0.99	$+0.00 \pm 0.98$ (+0.00σ, 0.99)	$+0.00 \pm 0.99$ (+0.00σ, 1.00)	-0.00
CMS_htt_ee_zeeShape_0jet_low_mass2_8TeV	0.00 ± 0.99	-0.00 ± 0.99 (-0.00σ, 1.00)	-0.00 ± 1.00 (-0.00σ, 1.01)	-0.00
CMS_htt_ee_zeeShape_1jet_high_mass0_7TeV	0.00 ± 0.99	$+0.15 \pm 0.76$ (+0.15σ, 0.77)	$+0.15 \pm 0.78$ (+0.15σ, 0.78)	-0.01
CMS_htt_ee_zeeShape_1jet_high_mass0_8TeV	0.00 ± 0.99	$+0.27 \pm 0.92$ (+0.28σ, 0.92)	$+0.27 \pm 0.92$ (+0.28σ, 0.93)	-0.03
CMS_htt_ee_zeeShape_1jet_high_mass1_7TeV	0.00 ± 0.99	-0.06 ± 0.82 (-0.06σ, 0.83)	-0.06 ± 0.84 (-0.06σ, 0.85)	+0.00
CMS_htt_ee_zeeShape_1jet_high_mass1_8TeV	0.00 ± 0.99	$+0.18 \pm 0.97$ (+0.18σ, 0.98)	$+0.18 \pm 0.98$ (+0.18σ, 0.99)	+0.01
CMS_htt_ee_zeeShape_1jet_high_mass2_7TeV	0.00 ± 0.99	$+0.04 \pm 0.90$ (+0.04σ, 0.91)	$+0.04 \pm 0.90$ (+0.04σ, 0.91)	+0.01
CMS_htt_ee_zeeShape_1jet_high_mass2_8TeV	0.00 ± 0.99	$+0.27 \pm 0.97$ (+0.28σ, 0.97)	$+0.27 \pm 0.97$ (+0.28σ, 0.98)	-0.01
CMS_htt_ee_zeeShape_1jet_low_mass0_7TeV	0.00 ± 0.99	$+0.08 \pm 0.56$ (+0.08σ, 0.57)	$+0.08 \pm 0.58$ (+0.08σ, 0.58)	-0.05
CMS_htt_ee_zeeShape_1jet_low_mass0_8TeV	0.00 ± 0.99	$+0.07 \pm 0.95$ (+0.07σ, 0.96)	$+0.07 \pm 0.95$ (+0.07σ, 0.96)	-0.02
CMS_htt_ee_zeeShape_1jet_low_mass1_7TeV	0.00 ± 0.99	-0.04 ± 0.77 (-0.04σ, 0.78)	-0.04 ± 0.80 (-0.04σ, 0.81)	+0.04
CMS_htt_ee_zeeShape_1jet_low_mass1_8TeV	0.00 ± 0.99	-0.00 ± 0.98 (-0.00σ, 0.99)	-0.00 ± 1.00 (-0.00σ, 1.01)	-0.00
CMS_htt_ee_zeeShape_1jet_low_mass2_7TeV	0.00 ± 0.99	-0.03 ± 0.98 (-0.03σ, 0.99)	-0.03 ± 0.98 (-0.03σ, 0.99)	-0.00
CMS_htt_ee_zeeShape_1jet_low_mass2_8TeV	0.00 ± 0.99	-0.01 ± 0.98 (-0.01σ, 0.99)	-0.01 ± 0.99 (-0.01σ, 1.00)	-0.01
CMS_htt_ee_zeeShape_vbf_mass0_7TeV	0.00 ± 0.99	$+0.04 \pm 0.69$ (+0.04σ, 0.69)	$+0.04 \pm 0.68$ (+0.04σ, 0.69)	-0.02
CMS_htt_ee_zeeShape_vbf_mass0_8TeV	0.00 ± 0.99	-0.13 ± 0.69 (-0.13σ, 0.69)	-0.13 ± 0.68 (-0.13σ, 0.69)	+0.02
CMS_htt_ee_zeeShape_vbf_mass1_7TeV	0.00 ± 0.99	-0.05 ± 0.97 (-0.05σ, 0.98)	-0.05 ± 0.97 (-0.05σ, 0.98)	-0.00
CMS_htt_ee_zeeShape_vbf_mass1_8TeV	0.00 ± 0.99	-0.02 ± 0.97 (-0.02σ, 0.98)	-0.01 ± 0.97 (-0.01σ, 0.98)	+0.03
CMS_htt_ee_zeeShape_vbf_mass2_7TeV	0.00 ± 0.99	-0.01 ± 0.98 (-0.01σ, 0.99)	-0.01 ± 0.98 (-0.01σ, 0.99)	-0.01

name	pre fit	b-only fit	$s+b$ fit	$\rho(\theta,\mu)$
CMS_htt_ee_zeeShape_vbf_mass2_8TeV	0.00 ± 0.99	-0.01 ± 0.98 (-0.01σ, 0.99)	-0.01 ± 0.97 (-0.01σ, 0.98)	+0.02
CMS_htt_ee_ztt_extrap_0jet_high_7TeV	0.00 ± 1.00	$+0.00 \pm 0.98$ (+0.00σ, 0.99)	$+0.00 \pm 0.98$ (+0.00σ, 0.99)	+0.00
CMS_htt_ee_ztt_extrap_0jet_high_8TeV	0.00 ± 1.00	$+0.12 \pm 0.92$ (+0.12σ, 0.92)	$+0.12 \pm 0.94$ (+0.12σ, 0.94)	-0.02
CMS_htt_ee_ztt_extrap_1jet_high_7TeV	0.00 ± 1.00	-0.03 ± 0.99 (-0.03σ, 0.99)	-0.03 ± 0.99 (-0.03σ, 0.99)	+0.00
CMS_htt_ee_ztt_extrap_1jet_high_8TeV	0.00 ± 1.00	$+0.05 \pm 0.98$ (+0.05σ, 0.99)	$+0.05 \pm 0.98$ (+0.05σ, 0.99)	-0.03
CMS_htt_ee_ztt_extrap_1jet_low_7TeV	0.00 ± 1.00	-0.03 ± 0.94 (-0.03σ, 0.95)	-0.03 ± 0.95 (-0.03σ, 0.95)	-0.01
CMS_htt_ee_ztt_extrap_1jet_low_8TeV	0.00 ± 1.00	$+0.02 \pm 0.86$ (+0.02σ, 0.87)	$+0.02 \pm 0.87$ (+0.02σ, 0.88)	+0.01
CMS_htt_ee_ztt_extrap_vbf_7TeV	0.00 ± 1.00	$+0.09 \pm 0.97$ (+0.09σ, 0.97)	$+0.09 \pm 0.97$ (+0.09σ, 0.98)	-0.10
CMS_htt_ee_ztt_extrap_vbf_8TeV	0.00 ± 1.00	$+0.13 \pm 0.97$ (+0.13σ, 0.97)	$+0.13 \pm 0.98$ (+0.13σ, 0.98)	-0.10
CMS_htt_scale_met_7TeV	0.00 ± 0.99	$+0.00 \pm 1.01$ (+0.00σ, 1.02)	$+0.00 \pm 1.01$ (+0.00σ, 1.02)	-0.02
CMS_htt_scale_met_8TeV	0.00 ± 0.99	$+0.12 \pm 0.97$ (+0.12σ, 0.98)	$+0.12 \pm 0.96$ (+0.12σ, 0.97)	+0.01
CMS_htt_ttbarNorm_7TeV	0.00 ± 1.00	$+0.03 \pm 0.98$ (+0.03σ, 0.98)	$+0.03 \pm 0.98$ (+0.03σ, 0.98)	-0.04
CMS_htt_ttbarNorm_8TeV	0.00 ± 1.00	$+0.32 \pm 0.94$ (+0.32σ, 0.94)	$+0.31 \pm 0.94$ (+0.31σ, 0.95)	-0.14
CMS_htt_zttNorm_7TeV	0.00 ± 1.00	$+0.35 \pm 0.94$ (+0.35σ, 0.95)	$+0.35 \pm 0.95$ (+0.35σ, 0.95)	-0.00
CMS_htt_zttNorm_8TeV	0.00 ± 1.00	-0.48 ± 0.93 (-0.48σ, 0.93)	-0.48 ± 0.94 (-0.48σ, 0.94)	-0.01
CMS_scale_e_7TeV	0.00 ± 0.99	$+0.01 \pm 0.80$ (+0.01σ, 0.81)	$+0.01 \pm 0.80$ (+0.01σ, 0.81)	-0.07
CMS_scale_e_8TeV	0.00 ± 0.99	$+0.35 \pm 0.94$ (+0.36σ, 0.95)	$+0.35 \pm 0.98$ (+0.35σ, 0.99)	-0.09
CMS_scale_j_7TeV	0.00 ± 0.99	-0.06 ± 1.02 (-0.06σ, 1.03)	-0.06 ± 1.04 (-0.06σ, 1.05)	-0.02
CMS_scale_j_8TeV	0.00 ± 0.99	-0.04 ± 0.93 (-0.04σ, 0.94)	-0.04 ± 0.93 (-0.04σ, 0.94)	-0.03
QCDscale_VH	0.00 ± 1.00	$+0.00 \pm 0.99$ (+0.00σ, 0.99)	-0.00 ± 1.02 (-0.00σ, 1.02)	-0.00
QCDscale_ggH	0.00 ± 1.00	-0.00 ± 0.97 (-0.00σ, 0.98)	-0.00 ± 1.00 (-0.00σ, 1.00)	-0.01
QCDscale_ggH1in	0.00 ± 1.00	-0.00 ± 0.99 (-0.00σ, 0.99)	-0.00 ± 0.99 (-0.00σ, 1.00)	-0.03
QCDscale_ggH2in	0.00 ± 1.00	$+0.00 \pm 0.99$ (+0.00σ, 0.99)	$+0.01 \pm 0.99$ (+0.01σ, 1.00)	+0.07
QCDscale_ggH2in_Shape	0.00 ± 0.99	$+0.00 \pm 0.98$ (+0.00σ, 0.99)	-0.00 ± 0.98 (-0.00σ, 0.99)	-0.00
QCDscale_qqH	0.00 ± 1.00	$+0.00 \pm 0.99$ (+0.00σ, 0.99)	$+0.00 \pm 1.02$ (+0.00σ, 1.02)	-0.03
QCDscale_qqH_Shape	0.00 ± 0.99	$+0.00 \pm 0.98$ (+0.00σ, 0.99)	$+0.00 \pm 0.98$ (+0.00σ, 0.99)	+0.01
UEPS	0.00 ± 1.00	-0.00 ± 1.01 (-0.00σ, 1.01)	-0.00 ± 1.02 (-0.00σ, 1.02)	-0.03
lumi_7TeV	0.00 ± 1.00	-0.00 ± 0.99 (-0.00σ, 0.99)	$+0.00 \pm 0.99$ (+0.00σ, 0.99)	+0.02
lumi_8TeV	0.00 ± 1.00	$+0.03 \pm 0.99$ (+0.03σ, 0.99)	$+0.03 \pm 0.99$ (+0.03σ, 1.00)	-0.02
pdf_gg	0.00 ± 1.00	-0.00 ± 0.99 (-0.00σ, 0.99)	$+0.00 \pm 0.99$ (+0.00σ, 1.00)	+0.01
pdf_gg_Shape	0.00 ± 0.99	$+0.00 \pm 0.98$ (+0.00σ, 0.99)	$+0.00 \pm 0.98$ (+0.00σ, 0.99)	+0.01
pdf_qqbar	0.00 ± 1.00	$+0.00 \pm 1.01$ (+0.00σ, 1.01)	$+0.00 \pm 1.02$ (+0.00σ, 1.02)	+0.00
pdf_qqbar_Shape	0.00 ± 0.99	$+0.00 \pm 0.98$ (+0.00σ, 0.99)	-0.00 ± 0.99 (-0.00σ, 1.00)	+0.00
r	[-5.00, 5.00]	n/a	$+0.05 \pm 1.65$	+1.00

Bibliography

[1] S. Chatrchyan *et al.*, "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons," *JHEP*, vol. 1405, p. 104, 2014.

[2] S. Chatrchyan *et al.*, "Evidence for the direct decay of the 125 GeV Higgs boson to fermions," *Nature Phys.*, vol. 10, 2014.

[3] Atlas Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the {ATLAS} detector at the {LHC}," *Physics Letters B*, vol. 716, no. 1, pp. 1–29, 2012.

[4] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the {CMS} experiment at the {LHC}," *Physics Letters B*, vol. 716, no. 1, pp. 30–61, 2012.

[5] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys. Rev. Lett.*, vol. 13, pp. 321–323, Aug 1964.

[6] P. Higgs, "Broken symmetries, massless particles and gauge fields," *Physics Letters*, vol. 12, no. 2, pp. 132–133, 1964.

[7] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964.

[8] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, "Microscopic Theory of Superconductivity," *Phys. Rev.*, vol. 106, pp. 162–164, Apr 1957.

[9] P. W. Anderson, "Plasmons, Gauge Invariance, and Mass," *Phys. Rev.*, vol. 130, pp. 439–442, Apr 1963.

[10] S. Weinberg, "A Model of Leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967.

[11] A. Salam, "Weak and Electromagnetic Interactions," *Conf.Proc.*, vol. C680519, pp. 367–377, 1968.

[12] G. Hooft, "Renormalizable Lagrangians for massive Yang-Mills fields," *Nuclear Physics B*, vol. 35, no. 1, pp. 167–188, 1971.

[13] G. 't Hooft and M. Veltman, "Regularization and renormalization of gauge fields," *Nuclear Physics B*, vol. 44, no. 1, pp. 189–213, 1972.

[14] Atlas Collaboration, "Updated coupling measurements of the Higgs boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton collision data," Tech. Rep. ATLAS-CONF-2014-009, CERN, Geneva, Mar 2014.

[15] CMS Collaboration, "Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV," Tech. Rep. CMS-PAS-HIG-13-005, CERN, Geneva, 2013.

[16] ATLAS Collaboration, "Evidence for Higgs Boson Decays to the $\tau^+\tau^-$ Final State with the ATLAS Detector," Tech. Rep. ATLAS-CONF-2013-108, CERN, Geneva, Nov 2013.

[17] V. M. Abazov *et al.*, "Search for the standard model Higgs boson in tau lepton pair final states," *Phys.Lett.*, vol. B714, pp. 237–245, 2012.

[18] CDF Collaboration, "Search for a Low-Mass Standard Model Higgs Boson in the $\tau\tau$ Decay Channel in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96\ TeV$," *Phys. Rev. Lett.*, p. 181804, May 2012.

[19] ATLAS Collaboration, "Search for the Standard Model Higgs boson in H to tau tau decays in proton-proton collisions with the ATLAS detector," Tech. Rep. ATLAS-CONF-2012-160, CERN, Geneva, Nov 2012.

[20] CMS Collaboration, "Search for Neutral Higgs Bosons Decaying into Tau Leptons in the Dimuon Channel with CMS in pp Collisions at 7 TeV," Tech. Rep. CMS-PAS-HIG-12-007, CERN, Geneva, 2012.

[21] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles," *Phys. Rev. Lett.*, vol. 13, pp. 585–587, Nov 1964.

[22] T. W. B. Kibble, "Symmetry Breaking in Non-Abelian Gauge Theories," *Phys. Rev.*, vol. 155, pp. 1554–1561, Mar 1967.

[23] P. W. Higgs, "Spontaneous Symmetry Breakdown without Massless Bosons," *Phys. Rev.*, vol. 145, pp. 1156–1163, May 1966.

[24] S. L. Glashow, "Partial-symmetries of weak interactions," *Nuclear Physics*, vol. 22, no. 4, pp. 579–588, 1961.

[25] N. Cabibbo, "Unitary Symmetry and Leptonic Decays," *Phys. Rev. Lett.*, vol. 10, pp. 531–533, Jun 1963.

[26] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction," *Prog.Theor.Phys.*, vol. 49, pp. 652–657, 1973.

[27] Beringer, J. et al., "Review of Particle Physics," *Phys. Rev. D*, vol. 86, p. 010001, Jul 2012.

[28] A. David, A. Denner, M. Duehrssen, M. Grazzini, C. Grojean, G. Passarino, M. Schumacher, M. Spira, G. Weiglein, and M. Zanetti, "LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle," Tech. Rep. arXiv:1209.0040. LHCHXSWG-2012-001. CERN-PH-TH-2012-284, Sep 2012.

[29] Dittmaier, S. et al, *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*. Geneva: CERN, 2011. Working Group web page: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections.

[30] Dittmaier, S. et al, "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions," Tech. Rep. arXiv:1201.3084. CERN-2012-002, Geneva, 2012. Working Group web page: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections.

[31] Heinemeyer, S. et al, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group," Tech. Rep. arXiv:1307.1347. CERN-2013-004, Geneva, 2013. Working Group web page: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections.

[32] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, *et al.*, "Implications of CTEQ global analysis for collider observables," *Phys.Rev.*, vol. D78, p. 013004, 2008.

[33] A. Martin, W. Stirling, R. Thorne, and G. Watt, "Parton distributions for the LHC," *The European Physical Journal C*, vol. 63, no. 2, pp. 189–285, 2009.

[34] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, *et al.*, "A first unbiased global NLO determination of parton distributions and their uncertainties," *Nucl.Phys.*, vol. B838, pp. 136–206, 2010.

[35] J. M. Campbell, J. W. Huston, and W. J. Stirling, "Hard interactions of quarks and gluons: a primer for LHC physics," *Reports on Progress in Physics*, vol. 70, no. 1, p. 89, 2007.

[36] W. J. Stirling, "Progress in Parton Distribution Functions and implications for LHC," 2008.

[37] CMS Collaboration, "Search for the standard model Higgs boson produced in association with top quarks in multilepton final states," Tech. Rep. CMS-PAS-HIG-13-020, CERN, Geneva, 2013.

[38] CMS Collaboration, "Search for Higgs Boson Production in Association with a Top-Quark Pair and Decaying to Bottom Quarks or Tau Leptons," Tech. Rep. CMS-PAS-HIG-13-019, CERN, Geneva, 2013.

[39] CMS Collaboration, "Search for ttH production in events where H decays to photons at 8 TeV collisions," Tech. Rep. CMS-PAS-HIG-13-015, CERN, Geneva, 2013.

[40] Atlas Collaboration, "Search for the Standard Model Higgs boson produced in association with top quarks and decaying to $b\bar{b}$ in pp collisions at $\sqrt{s} =$ 8 TeV with the ATLAS detector at the LHC," Tech. Rep. ATLAS-CONF-2014-011, CERN, Geneva, Mar 2014.

[41] "LHCXSWG Webpage." https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSectionsFigures. Accessed: 2014-04-10.

[42] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira, "Standard model Higgs-boson branching ratios with uncertainties," *The European Physical Journal C*, vol. 71, no. 9, pp. 1–29, 2011.

[43] The ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider," *Journal of Instrumentation*, vol. 3, no. 08, p. S08003, 2008.

[44] The ALICE Collaboration, "The ALICE experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, no. 08, p. S08002, 2008.

[45] CMS Collaboration, "The CMS experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, no. 08, p. S08004, 2008.

[46] Della Negra, M. and others, "CMS: The Compact Muon Solenoid: Letter of intent for a general purpose detector at the LHC," 1992.

[47] The LHCb Collaboration, "The LHCb Detector at the LHC," *Journal of Instrumentation*, vol. 3, no. 08, p. S08005, 2008.

[48] C. Lefèvre, "The CERN accelerator complex. Complexe des accélérateurs du CERN." Dec 2008.

[49] "CMS Luminosity Public Results." https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Accessed: 2014-03-27.

[50] "SketchUp CMS." https://twiki.cern.ch/twiki/bin/view/CMSPublic/SketchUpCMS. Accessed: 2014-03-27.

[51] CMS Collaboration, "CMS, tracker technical design report," 1998.

[52] CMS Collaboration, *The CMS tracker: addendum to the Technical Design Report*. Technical Design Report CMS, Geneva: CERN, 2000.

[53] CMS Collaboration, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS, Geneva: CERN, 2006.

[54] CMS Collaboration, *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS, Geneva: CERN, 1997.

[55] CMS Collaboration, *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS, Geneva: CERN, 1997.

[56] CMS Collaboration, *The CMS magnet project: Technical Design Report*. Technical Design Report CMS, Geneva: CERN, 1997.

[57] CMS Collaboration, "CMS, the Compact Muon Solenoid. Muon technical design report," 1997.

[58] K. Rose, *Deterministic annealing, clustering, and optimization*. PhD thesis, California Institute of Technology, December 1991.

[59] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, "Adaptive Vertex Fitting," Tech. Rep. CMS-NOTE-2007-008, CERN, Geneva, Mar 2007.

[60] Baffioni, S. and Charlot, C. and Ferri, F. et al., "Electron reconstruction in CMS," *The European Physical Journal C*, vol. 49, no. 4, pp. 1099–1116, 2007.

[61] L. Bianchini and Y. Sirois, *Search for the Standard Model Higgs Boson decaying to tau leptons with the CMS experiment at LHC*. PhD thesis, Zurich, ETH, 2013. presented 18 Sep 2012.

[62] W. Adam, R. Früth, A. Strandlie, and T. Todor, "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC," 2005.

[63] N. Amapane et al., "Electron momentum determination using boosted regression trees," Tech. Rep. CMS-NOTE-2013-209, CERN, Geneva, Aug 2013.

[64] CMS Collaboration, "Measurement of the properties of a Higgs boson in the four-lepton final state," Tech. Rep. CERN-PH-EP-2013-220. CMS-HIG-13-002-003. arXiv:1312.5353, CERN, Geneva, Dec 2013. Comments: Submitted to Phys. Rev. D.

[65] S. Chatrchyan *et al.*, "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at âs = 7 TeV," *JINST*, vol. 8, p. P09009, 2013.

[66] A. Hocker, J. Stelzer, F. Tegenfeldt, H. Voss, K. Voss, *et al.*, "TMVA - Toolkit for Multivariate Data Analysis," *PoS*, vol. ACAT, p. 040, 2007.

[67] S. Xie, *Search for the Standard Model Higgs Boson Decaying to Two W Bosons at CMS*. PhD thesis, MIT, May 2012.

[68] CMS Collarboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET," Tech. Rep. CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr 2009.

[69] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *Journal of High Energy Physics*, vol. 2008, no. 04, p. 063, 2008.

[70] The CMS collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS," *Journal of Instrumentation*, vol. 6, no. 11, p. P11002, 2011.

[71] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas," *Physics Letters B*, vol. 659, no. 1–2, pp. 119–126, 2008.

[72] The CMS collaboration, "Pileup Jet Identification," Tech. Rep. CMS-PAS-JME-13-005, CERN, Geneva, 2013.

[73] CMS Collaboration, "Identification of b-quark jets with the CMS experiment," *J. Instrum.*, vol. 8, p. P04013. 67 p, Nov 2012. Comments: Submitted to the Journal of Instrumentation.

[74] CMS Collaboration, "Performance of b tagging at $\sqrt{(s)}$=8 TeV in multijet, ttbar and boosted topology events," Tech. Rep. CMS-PAS-BTV-13-001, CERN, Geneva, 2013.

[75] CMS collaboration, "Performance of Missing Transverse Momentum Reconstruction Algorithms in Proton-Proton Collisions at p s = 8 TeV with the CMS Detector," 2012.

[76] CMS Collaboration, "Search for the Standard-Model Higgs boson decaying to tau pairs in proton-proton collisions at $\sqrt(s) = 7$ and 8 TeV," Tech. Rep. CMS-PAS-HIG-13-004, CERN, Geneva, 2013.

[77] B. Bullock, K. Hagiwara, and A. Martin, "Tau polarization and its correlations as a probe of new physics," *Nuclear Physics B*, vol. 395, no. 3, pp. 499–533, 1993.

[78] H. Pi, P. Avery, D. Green, J. Rohlf, and C. Tully, "Measurement of missing transverse energy with the CMS detector at the LHC," *The European Physical Journal C - Particles and Fields*, vol. 46, no. 1, pp. 45–56, 2006.

[79] CMS collaboration, "Missing transverse energy performance of the CMS detector," *Journal of Instrumentation*, vol. 6, no. 09, p. P09001, 2011.

[80] S. Agostinelli et al., "Geant4—a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003.

[81] R. S. Fletcher and T. Stelzer, "Rapidity gap signals in Higgs-boson production at the SSC," *Phys. Rev. D*, vol. 48, pp. 5162–5167, Dec 1993.

[82] D. L. Rainwater, D. Zeppenfeld, and K. Hagiwara, "Searching for $H \to \tau^+\tau^-$ in weak boson fusion at the CERN LHC," *Phys.Rev.*, vol. D59, p. 014037, 1998.

[83] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method," *Journal of High Energy Physics*, vol. 2007, no. 11, p. 070, 2007.

[84] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO Higgs boson production via gluon fusion matched with shower in POWHEG," *Journal of High Energy Physics*, vol. 2009, no. 04, p. 002, 2009.

[85] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "MadGraph 5: going beyond," *Journal of High Energy Physics*, vol. 2011, no. 6, pp. 1–40, 2011.

[86] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 0605, p. 026, 2006.

[87] S. Jadach, J. H. Kuhn, and Z. Was, "TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons," *Comput.Phys.Commun.*, vol. 64, pp. 275–299, 1990.

[88] M. Zeise, "Study of Z Boson Decays into Pairs of Muon and Tau Leptons with the CMS Detector at the LHC,"

[89] CMS Collaboration, "Combination of top pair production cross section measurements," Tech. Rep. CMS-PAS-TOP-11-024, CERN, Geneva, 2011.

[90] R. Kleiss and W. Stirling, "Top quark production at hadron colliders: Some useful formulae," *Zeitschrift für Physik C Particles and Fields*, vol. 40, no. 3, pp. 419–423, 1988.

[91] U. Langenfeld, S. Moch, and P. Uwer, "Measuring the running top-quark mass," *Phys. Rev. D*, vol. 80, p. 054009, Sep 2009.

[92] CMS Collaboration, "Measurement of the W and Z inclusive production cross sections at sqrts=7 TeV with the CMS experiment at the LHC," Tech. Rep. CMS-PAS-EWK-10-005, CERN, Geneva, 2011.

[93] CMS Collaboration, "Absolute Calibration of Luminosity Measurement at CMS: Summer 2011 Update," Tech. Rep. CMS-PAS-EWK-11-001, CERN, Geneva, 2011.

[94] CMS Collaboration, "Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update," Tech. Rep. CMS-PAS-SMP-12-008, CERN, Geneva, 2012.

[95] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update," Tech. Rep. CMS-PAS-LUM-13-001, CERN, Geneva, 2013.

[96] S. Alekhin, S. Alioli, R. D. Ball, V. Bertone, J. Blumlein, *et al.*, "The PDF4LHC Working Group Interim Report," 2011.

[97] ATLAS Collaboration, "New ATLAS event generator tunes to 2010 data," Tech. Rep. ATL-PHYS-PUB-2011-008, CERN, Geneva, Apr 2011.

[98] CMS Collaboration, "Charged particle multiplicities in pp interactions at $\sqrt{s}$ = 0.9 , 2.36, and 7 TeV," *Journal of High Energy Physics*, vol. 2011, no. 1, 2011.

[99] K. Hamilton, P. Nason, and G. Zanderighi, "MINLO: multi-scale improved NLO," *Journal of High Energy Physics*, vol. 2012, no. 10, pp. 1–28, 2012.

[100] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations," *JHEP*, vol. 0206, p. 029, 2002.

[101] Atlas Collaboration, "Procedure for the LHC Higgs boson search combination in summer 2011," Tech. Rep. ATL-PHYS-PUB-2011-011, CERN, Geneva, Aug 2011.

[102] A. L. Read, "Presentation of search results: the CL s technique," *Journal of Physics G: Nuclear and Particle Physics*, vol. 28, no. 10, p. 2693, 2002.

[103] T. Junk, "Confidence level computation for combining searches with small statistics," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 434, no. 2-3, pp. 435–443, 1999.

[104] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *The European Physical Journal C*, vol. 71, no. 2, pp. 1–19, 2011.

[105] CMS Collaboration, "Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks," *Phys.Rev.*, vol. D89, p. 012003, 2014.

[106] T. Plehn, D. L. Rainwater, and D. Zeppenfeld, "Determining the structure of Higgs couplings at the LHC," *Phys.Rev.Lett.*, vol. 88, p. 051801, 2002.

[107] S. Berge, W. Bernreuther, and J. Ziethe, "Determining the CP parity of Higgs bosons at the LHC in their tau decay channels," *Phys.Rev.Lett.*, vol. 100, p. 171605, 2008.

Acknowledgements

It is my pleasure to devote this section to everyone who made my work possible.

I thank my supervisor Dr. Alexei Raspereza, for his dedication and guidance through the last 3 years. His deep interest in physics and his working-horse philosophy had great impact on me and it was a tremendous experience and my absolut pleasure to have worked with and for him.

I am very grateful to Prof. Dr. Peter Schleper for help and discussions in the beginning of my thesis work and for support during my examination. Also, I would like to thank Prof. Dr. Georg Weiglein and Prof. Dr. Christian Sander for discussions and for acting as examiners for my thesis.

I received a lot of support and help throughout my entire PhD phase from Dr. Roger Wolf. Not only did he teach me many technical aspects of physics, but also the social and political dynamics of big science collaborations. We had many enthusiastic and fun discussions and hatched many plans for scientific collaboration. Thanks a lot Roger!

I would like to thank the CMS KIT group of Prof. Dr. Günther Quast for a very fruitful collaboration on the same-flavour lepton channels. I had a great and memorable time working day and night with Thomas Müller, Raphael Friese, Joram Berger, Rene Caspart, Fabio Colombo, Manuel Zeise and Martin Niegel.

In 2010 Dr. Kerstin Borras gave me the opportunity to start my PhD work for the DESY CMS team, and I would like to thank her for putting me into this position and succesfully pursue the transition from mathematical physics to experimental particle physics at DESY and CERN at this very exciting time.

During my first year I worked under the supervision of Dr. Benjamin Lutz on the CMS HCAL SiPM Upgrade, and I would like to thank Benjamin for all the things about hardware he has taught me.

I had the great pleasure to work with Dr. Daniele Benedetti and Dr. Emanuiele Di Marco of the CMS EGamma POG who taught and helped me with my work on CMS electron identification during the time of the Higgs discovery.

I thank all my PhD colleagues, especially Niklas Pietsch and Hannes Schettler, for the pleasant time I shared with them at many places on earth. In Geneva, I had a lot of fun and unforgettable evenings with Ben Constance and Ilya Gorbunov. I would like to thank Dr. Matthias Schröder and also Dr. Johannes Hauk for suggestions to my thesis and thesis defence.

Outside of physics I always found social support from my friends Betti, Clemens, Flippe, Florian, Fritz, Daniel, Jakob and Raoul. I thank them for all the empathy

and support and all the good times we have spent together.

I would like to express my gratitude to my soul-mate Dora-Lisa, who initiated many of my insights in life and opened up a whole new world of experiences for me.

I am deeply indebted to thank my dear father and friend Carl Salfeld-Nebgen for his continuous help and love and all the opportunities he gave me throughout life. Thank you Dad, I could not have done it without you!