Measurement of the Higgs boson transverse momentum spectrum in the WW decay channel at 8 TeV and first results at 13 TeV

Lorenzo Viliani
of University of Florence

PhD Thesis

Abstract

The cross section for Higgs boson production in pp collisions is studied using the $H \to W^+W^-$ decay mode, followed by leptonic decays of the W bosons, leading to an oppositely charged electron-muon pair in the final state. The measurements are performed using data collected by the CMS experiment at the LHC with pp collisions at a centre-of-mass energy of 8TeV, corresponding to an integrated luminosity of 19.4fb^{-1}. The Higgs boson transverse momentum (p_{T}) is reconstructed using the lepton pair p_{T} and missing p_{T}. The differential cross section times branching fraction is measured as a function of the Higgs boson p_{T} in a fiducial phase space defined to match the experimental acceptance in terms of the lepton kinematics and event topology. The production cross section times branching fraction in the fiducial phase space is measured to be 39 ± 8 (stat) ± 9 (syst)fb. The measurements are compared to theoretical calculations based on the standard model to which they agree within experimental uncertainties.

Contents

Chapter 1

Electroweak and QCD physics at LHC

Chapter 2

The CMS experiment at the LHC

2.1 The Large Hadron Collider

2.2 The CMS experiment

2.3 The CMS trigger system

2.4 Objects definition and event reconstruction

Chapter 3

Higgs boson properties in the H $\rightarrow$ WW decay channel

3.1 Higgs boson measurements at LHC

The discovery of a new boson consistent with the standard model (SM) Higgs boson has been reported by ATLAS and CMS Collaborations in 2012. The discovery has been followed by a comprehensive set of studies of properties of this new boson in several production and decay channels and no evidence of deviation from the SM expectation has been found so far. The CMS studies in the H $\rightarrow$ WW $\rightarrow 2\ell 2\nu$ decay channel include the measurement of the Higgs properties, as well as constraints on the Higgs total decay width and gauge bosons anomalous couplings.

3.2 Higgs boson measurements in the H $\rightarrow$ WW decay channel

Chapter 4

Measurement of the Higgs boson transverse momentum at 8TeVusing $H \to WW \to 2\ell 2\nu$ decays

4.1 Introduction

In this chapter the measurement of the transverse momentum spectrum of the Higgs boson, produced in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 8$TeV, is reported. This measurement can be used to directly inspect the perturbative QCD theory in the Higgs sector. In particular the p_T^H variable is sensitive to the Higgs production mode and the differential distribution in this variable can be used to inspect the effects of the top quark mass in the gluon fusion top loop. Moreover, any observed deviation from the SM expectation, especially in the tail of the p_T^H distribution, could be a hint of physics beyond the SM.

Similar measurements have already been performed by CMS and ATLAS experiments in the ZZ and $\gamma\gamma$ Higgs decay channels. The measurement reported here is the first measurement of the Higgs p_{T} spectrum in the WW decay channel.

The cross section has been measured in a fiducial phase space defined using generator level variables in order to mimic the experimental acceptance and reduce the systematic uncertainties on the procedure of extrapolating the results in a larger phase space.

The Higgs transverse momentum has been reconstructed calculating the vector sum of the

dilepton system transverse momentum plus missing transverse energy

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{H}} = \vec{p}_{\mathrm{T}}^{\,\ell\ell} + \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} \tag{4.1}$$

The signal has been extracted subtracting all backgrounds by means of a binned Maximum Likelihood fit and has been then corrected for the efficiency of the analysis selections and for the detector resolution effects using an unfolding procedure.
The differential measurement has been performed in six bins of $p_{\mathrm{T}}^{\mathrm{H}}$ with variable widths, chosen to have approximately the same purity in each bin, as explained in section 4.3.

4.2 Datasets, Triggers and MC samples

This analysis is largely built on top of the already published H $\rightarrow$ WW measurments [1] in terms of code, selections and background estimates for both the gluon fusion (ggH) [2] and the vector boson fusion (VBF) [3] prodcution mechanisms.

4.2.1 Datasets and triggers

The datasets used for the analysis correspond to 19.4fb^{-1}at $\sqrt(s) = 8$ TeVof integrated luminosity composed of the following CMS data taking periods during 2012: 2012A (892 pb^{-1}), 2012B (4440 pb^{-1}), and 2012C (6898 pb^{-1}) and 2012D (7238 pb^{-1}). Data have been vchecked and validated and only data corresponding to good data taking quality are considered. The $\mathrm{e}^{\pm}\mu^{\mp}$ final state is considered in this analysis. The following five Primary Datasets have been used for the signal extraction: SingleElectron, SingleMu and MuEG (Muon-ElectronGamma).

For the data samples, the events are required to fire one of the unprescaled single-electron, single-muon or muon-electron triggers. A full description of these triggers in given in [4] for 8 TeVdata. Although identification and isolation criteria are also applied, a brief overview of the HLT transverse momentum (p_{T}) criteria on the leptons is given in Table 4.1. While the HLT lepton p_{T} thresholds of 17 and 8 GeVfor the double lepton triggers accommodate the offline lepton p_{T} selection of 20 and 10 GeV, the higher p_{T} thresholds in the single

lepton triggers help partially recovering double lepton trigger inefficiencies as a high p_{T} lepton is on average expected due to the kinematic of the Higgs decay.

Table 4.1: Highest transverse momentum thresholds applied in the lepton triggers at the HLT level. Double set of thresholds indicates the thresholds for each leg of the double lepton triggers.

Trigger Path	7 TeV	8 TeV
Single-Electron	$p_{\mathrm{T}} > 27$ GeV	$p_{\mathrm{T}} > 27$ GeV
Single-Muon	$p_{\mathrm{T}} > 15$ GeV	$p_{\mathrm{T}} > 24$ GeV
Muon-Electron	$p_{\mathrm{T}} > 17$ and 8 GeV	$p_{\mathrm{T}} > 17$ and 8 GeV
Electron-Muon	$p_{\mathrm{T}} > 17$ and 8 GeV	$p_{\mathrm{T}} > 17$ and 8 GeV

No trigger requirement is made on the simulated events but the combined trigger efficiency is estimated from data and applied to all simulated events. The detailed trigger efficiencies and the weighting procedure can be found in Appendix C of [2] [5]. The average trigger efficiency for signal events that pass the full event selection is measured to be about 96% in the $e\mu$ final state for a Higgs boson mass of about 125GeV.

4.2.2 Monte-Carlo samples

Several Monte Carlo event generators are used to simulate the signal and background processes:

- The POWHEG program [6] provides event samples for the $H \to WW$ signal for the Gluon Fusion (ggH) and VBF production mechanisms, as well as $\mathrm{t\bar{t}}$ and tW processes.
- The $q\bar{q} \to WW$, Drell-Yan, ZZ, WZ, $W\gamma$, $W\gamma^*$, tri-bosons and W+jets processes are generated using the MADGRAPH 5.1.3 [7] event generator.
- The VH process is simulated using PYTHIA 6.424 [8].

For leading-order generators samples, the CTEQ6L [9] set of parton distribution functions (PDF) is used, while CT10 [10] is used for next-to-leading order (NLO) ones. Cross section calculations [**LHCHiggsCrossSectionWorkingGroup:2011ti**] at next-to-next-to-leading order (NNLO) are used for the $H \to WW$ process (POWHEG NLO generator is tuned to reproduce NNLO accuracy on the on-shell Higgs p_T spectrum and scaled to NNLO

inclusive cross-section), while NLO calculations are used for background cross sections. For all processes, the detector response is simulated using a detailed description of the CMS detector, based on the GEANT4 package [11].

Minimum bias events are superimposed on the simulated events to emulate the additional pp interactions per bunch crossing (pile-up). The number of pile-up events simulated in the MC samples (in the same bunch crossing, in time, or in the previous or following one, out of time pile-up) have been generated poissonianly sampling from a distribution similar to what is expected from data. These samples are reweighted to represent the pile-up distribution as measured in the data. For a given range of analyzed runs, the mean number of pile-up interactions per bunch crossing is estimated per luminosity block using the instantaneous luminosity provided by the LHC, integrated over the entire run range and normalized. This distribution is then used to reweight the simulated pile-up distribution. The average number of pile-up events per beam crossing in the 2011 data is about 10, and in the 2012 data it is about 20.

We checked that the contribution of the ttH production mechanisms is negligible in each bin of p_T^H (below 1%) and has been neglected. In figure 4.1 is shown the relative fraction of the four different production modes in each bin of p_T^H.

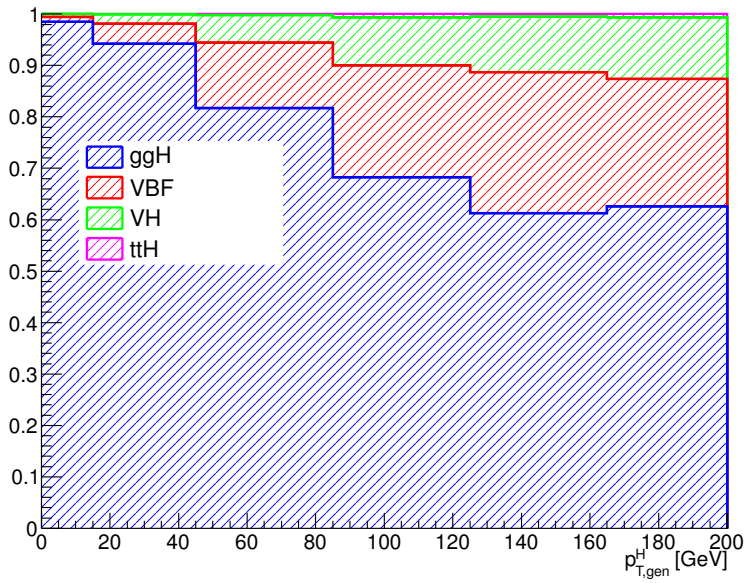


Figure 4.1: Relatve fraction of ggH, VBF, VH and ttH in each bin of the Higgs boson transverse momentum.

4.3 Analysis Strategy

The Higgs boson transverse momentum is measured in a fiducial phase space, which is defined at generator level requiring

- Exactly two status 3 leptons, an electron and a muon, originated from the H $\rightarrow$ WW $\rightarrow 2\ell 2\nu$ decays, with opposite charge, with $|\eta| < 2.5$ and $p_{\mathrm{T}} > 20$ GeV and $p_{\mathrm{T}} > 10$ GeV for the leading and subleading leptons respectively.
- Generator level invariant mass of the two leptons $m_{\ell\ell} > 12$ GeV.
- Vector sum of the two status 3 leptons $p_{\mathrm{T}}^{\ell\ell} > 30$ GeV.
- Generator level transverse mass $\sqrt{(p_{\mathrm{T}}^{\ell\ell} + p_{\mathrm{T}}^{\nu\nu})^2 - (\vec{p_{\mathrm{T}}^{\ell\ell}} + \vec{p_{\mathrm{T}}^{\nu}})^2} > 50$ GeV.

Experimentally, the Higgs boson transverse momentum is reconstructed as the vector sum of the lepton momenta in the transverse plane and MET.

$$\vec{p}_T^{\,H} = \vec{p}_T^{\,\ell\ell} + \vec{p}_T^{\,miss} \tag{4.2}$$

Compared to other differential analysis of the Higgs cross section, such as those in the ZZ and $\gamma\gamma$ decay channels, this analysis has to cope with the limited resolution due to the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ entering the transverse momentum measurement. The effect of the limited $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ resolution has two main implications on the analysis strategy:

- the choice of the binning in the transverse momentum spectrum needs to be reasonable when compared to the resolution. A detailed explanation of how the binning is defined is given in Sec. 4.5.
- Non negligible bin migration effects are present, and an unfolding procedure needs to be applied, not only to correct for selection efficiencies, as in ZZ and $\gamma\gamma$, but also to correct for bin migration effects. This is explained in Sec. **??**.

A detailed description of the fiducial region definition and about its optimization is given in appendix **??**.

The selection is essentially based on the one in the H $\rightarrow$ WW $\rightarrow 2\ell 2\nu$ published analysis [1] with one noticeable difference being the fact that in this analysis we do not make categories in the number of jets. The reason for this choice is that the number of jets

is strongly correlated with the transverse momentum, so making an inclusive analysis in the number of jets allows the dropping of most of the uncertainties related to the signal modeling of the number of jets produced in association with the Higgs boson. A detailed description of the selection is shown in Sec. 4.4.

The estimation of the backgrounds is different, to some extent, with respect to the one of the published $H \to WW \to 2\ell 2\nu$. This is mainly due to the absence of the jet binning. The techniques used to assess the backgrounds in each bin are discussed in Secs. 4.6.1, 4.6.2, 4.6.3.

Concerning the signal extraction, this analysis is again based on the already published $H \to WW \to 2\ell 2\nu$ analysis, although we fit the signal component in each of the transverse momentum bins, using two dimensional templates in the $m_{\ell\ell}$, m_{T} plane. The signal extraction is discussed in Sec. **??**.

Finally an unfolding procedure is needed to extract the differential distribution in a fiducial phase space. This is discussed in detail in Sec. **??**.

4.4 Event reconstruction and selections

4.4.1 Event reconstruction

The muons, electrons, jets and missing transverse energy ($\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$) reconstruction and criteria are described in in details in [12]. The following criteria are only a brief summary:

- **Muons:** *GlobalMuon* (with $\chi^2/ndof < 10$, at least one good muon hit and at least two muon segments in different muon stations) or *TrackerMuon* (provided it satisfies the "Tracker Muon Last Station Tight" selection). Several cut-based identification criteria are applied as well as the particle flow (PF) Isolation. In 2012, the PF Isolation is replaced by an MVA algorithm.
- **Electrons:** *GSF Electons.* A MVA identification criteria is applied as well as an MVA algorithm.

- **Jets:** *Anti-k_T PF jets* (with R=0.5 and applying L1, L2 and L3 jet energy corrections, including Pile-Up jet corrections from Fastjet method). Only jets and $|\eta| < 4.7$ are considered. A specific Pile-Up MVA-based rejection algorithm is applied.

- **$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$:** The $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is reconstructed the *PF Algorithm* or considering *only tracks* originating from the same vertex as the two leptons. In addition, the minimum of the projections of these two $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ to the closest lepton direction if they are in the same hemisphere, otherwise of their original values, is used in the analysis.

4.4.2 Event selection

Unlike the main H $\rightarrow$ WW $\rightarrow 2\ell 2\nu$ analysis, this analysis is inclusive in number of jets, so we do not have to define different jet multiplicity categories. The event selection consist of several steps. The first step is to select WW -like events applying a selection that is heavily based on the main analysis selection except for few different cuts explained below. The WW -like event preselection consists of the following set of cuts:

1. **Lepton preselection**:

   - at least two opposite-sign and opposite-flavour ($e\mu$) leptons reconstructed in the event;

   - $|\eta| < 2.5$ for electrons and $|\eta| < 2.4$ for muons;

   - $p_{\mathrm{T}} > 20$ GeV for the leading lepton. For the trailing lepton, the transverse momentum is required to be larger than 10 GeV.

2. **Extra lepton veto**: the event is required to have two and only two opposite-sign leptons passing the lepton selection.

3. **$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ preselection**: particle flow $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is required to be greater than 20GeV.

4. **Di-lepton mass cut**: $m_{\ell\ell} >$ 12GeVin order to reject low mass resonances and QCD backgrounds.

5. **Di-lepton p_T cut**: $p_T^{\ell\ell} > 30$GeV.

6. **projected $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ selection**: minimum projected $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ required to be larger than 20 GeV.

7. **Transverse mass**: $m_T^H > 60$GeVto reject Drell-Yan to $\tau\tau$ events.

In addition to the WW -like preselection other cuts are applied in order to reduce the top background ($t\bar{t}$ ans single-top), which is one of the main backgrounds in this final state. We operate two different selections depending on the number of jets with $p_T > 30$ GeVin the event. This is done to suppress the top background both in the low p_T^H region, where 0-jets events have the biggest contribution, and for higher values where also larger jet multiplicity events are important. The selection for 0-jets events relies on a soft muon veto, which rejects events with non-isolated soft muons (likely belonging to b-jets), and on a soft jets (with $p_T < 30$ GeV) anti b-tagging requirement. The latter requirement exploits the Track Counting High Efficiency tagger (TCHE) to reject soft jets that are likely to come from b quarks hadronization. These are exactly the same requirements applied in the 0-jets bin of the main analysis.

For events with a jet multiplicity greater or equal than one, we apply a different selection with respect to the main analysis. In this case we exploit the good b-tagging performances of the *JetBProbability* tagger to reject all the jets with $p_T > 30$ GeVthat are likely to come from a b quark. This jet veto relies on a cut on the *JetBProbability* tagger discriminant as has been also done in the VH ($H \to WW \to 2\ell 2\nu$) analysis [13]. Any jet with a discriminant value below 1.4 is identified as a non b-jet. The analysis selection requires no b-tagged jets with $p_T > 30$ GeV.

A cut-flow plot is reported in figure 4.2 showing the effect of each selection on top of Monte Carlo samples. In the first bin, labelled as *No cut*, no selection has been applied and the bin content correspond to the total expected number of events with a luminosity of 19.46 fb^{-1}. All the events in this bin have at least two leptons with a loose transverse momentum cut of 8 GeV. In the following bin the lepton cuts are applied, including the requirement to have two opposite-sign and opposite-flavour leptons and the extra lepton veto. Then are progressively reported all the other selections, showing the effect of each cut on backgrounds and signal. For each selection is also reported the expected signal over background ratio which after the full selection reach a maximum value around 3%.

The selection efficiency is shown in Fig. 4.3 (a). The efficiency denominator is the number of events that pass the acceptance, while the numerator is the number of events that pass both the selection and the acceptance, in each $p_{\mathrm{T}}^{\mathrm{H}}$ bin. The fake rate, defined by the ratio of signal events that pass the selection but are not within the acceptance, divided

by the total number of events passing both the selection and the acceptance is shown in Fig. 4.3 (b). For both the selection efficiency and the fake rate the signal samples included correspond to the *ggH*, *VBF* and *VH* production mechanisms. The overall efficiency and fake rate are: $\epsilon = 0.362 \pm 0.005$ and $fake\ rate = 0.126 \pm 0.004$, where the errors are only statistical.

If we define a 4π acceptance, requiring just that the Higgs decays to WW and then to $2l2\nu$, the efficiency is $\epsilon = 0.03960 \pm 0.00033$.

4.5 Binning of the $p_{\mathrm{T}}^{\mathrm{H}}$ spectrum

Given the limited resolution on $p_{\mathrm{T}}^{\mathrm{H}}$, a criterion is needed to establish bi size. The criterion that we have chose is devised to keep under control the bin migrations due to the finite resolution. For any given bin i we can define the purity P_i on a signal sample as the number events that are generated and also reconstructed in that bin, $N_i^{GEN|RECO}$, divided by the

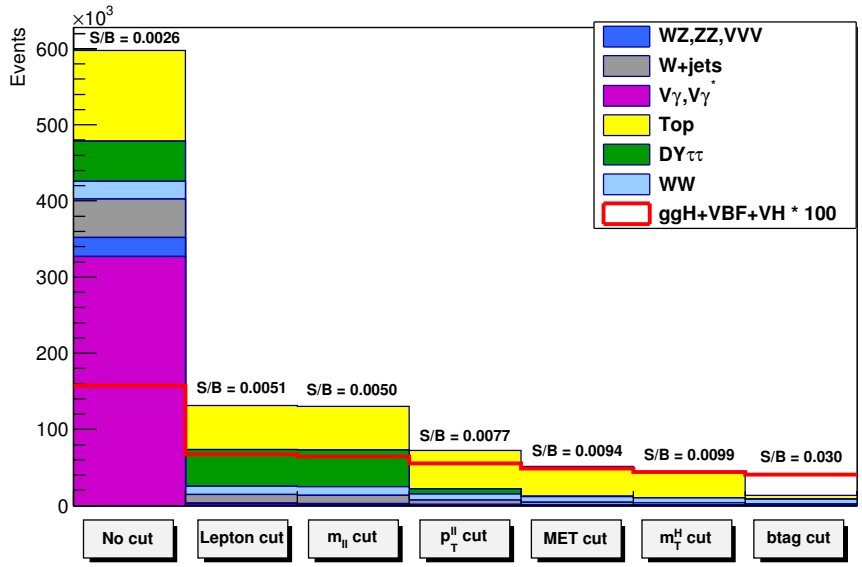
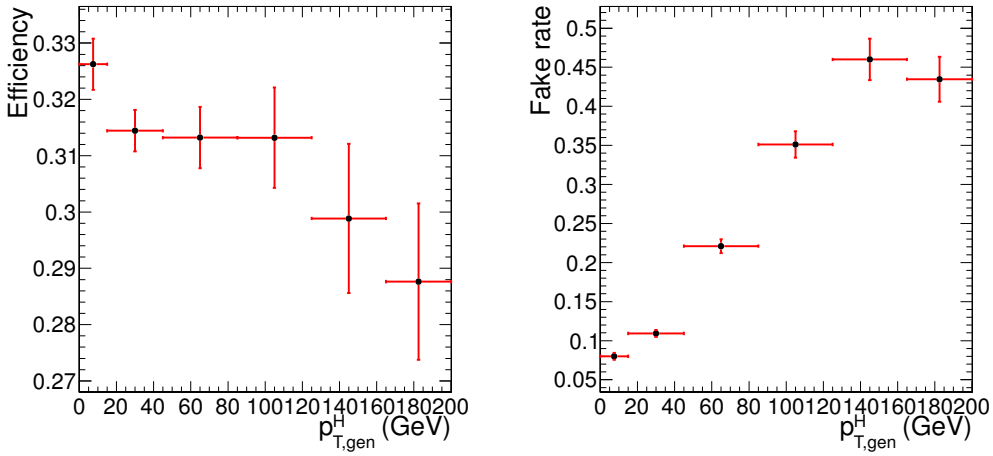


Figure 4.2: Effect of single selections on MC samples. The signal (red line) is multiplied by 100 and superimposed on stacked backgrounds. In each bin, corresponding to a different selection, is reported the expected number of events in MC at a luminosity of 19.46 fb^{-1}.



(a) (b)

Figure 4.3: Efficiency of the full selection (a) and fake rate (b) as a function of p_T^H.

number of events reconstructed there N_i^{RECO}:

$$P_i = \frac{N_i^{GEN|RECO}}{N_i^{RECO}} \quad . \tag{4.3}$$

Where $N_i^{GEN|RECO}$ is the number of events that are both generated and reconstructed in a $p_{\mathrm{T}}^{\mathrm{H}}$ bin i, while N_i^{RECO} is the number of events that are reconstructed in bin i. We have chosen the bin width in such a way as to make the smallest bins able to ensure a purity of about 60% on a gluon fusion sample. Following this prescription we have divided the whole $p_{\mathrm{T}}^{\mathrm{H}}$ range in six different bins: [0-15 GeV], [15-45 GeV], [45-85 GeV], [85-125 GeV], [125-165 GeV], [165-∞ GeV]. A two-dimensional histogram has been made putting the GEN level $p_{\mathrm{T}}^{\mathrm{H}}$ on the x-axis (calculated using the WW system transverse momentum) and the RECO one on the y-axis. Each row is then normalized to one in order to directly have the purity in the diagonal bins. Also the effect of bin migration due to finite detector resolution effects can be assessed from this plot.
This two dimensional plot is shown in Fig. 4.4 (a) and (b) for gluon fusion and VBF signals for $m_{\mathrm{H}} = 125$GeV.

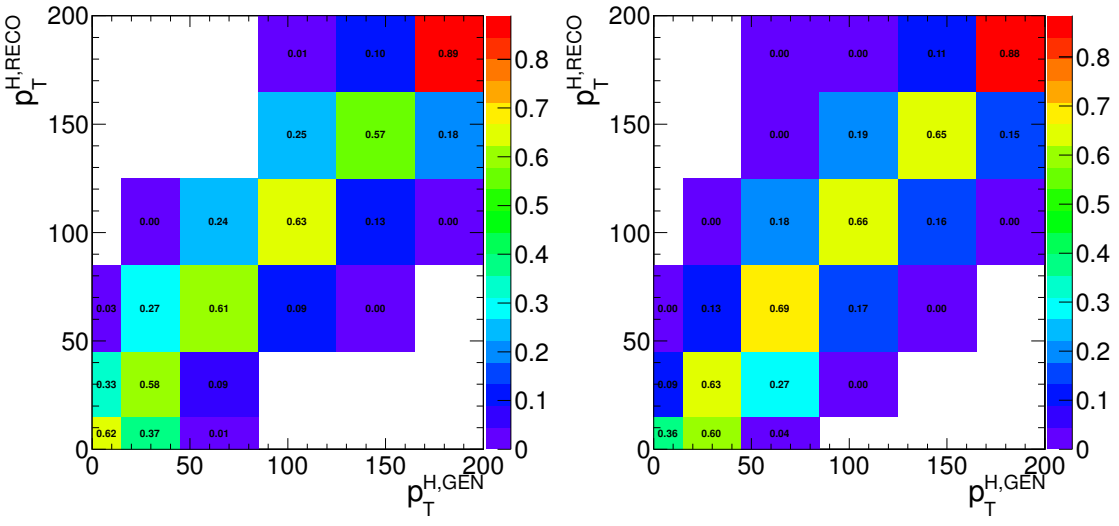
4.6 Background estimation

4.6.1 $t\bar{t}$ background

In this analysis the top background is divided into two different categories depending on the number of jets in the event. For 0-jets events we use the main analysis selections, thus also the top background estimation is the same. For events with more than 0 jets different cuts with respect to the main analysis are applied and the estimation of the top background is performed exploiting a Tag&Probe technique, explained in details later.

0-jets bin

The general strategy for determining the residual top events in the signal region is to first measure the top tagging efficiencies from an orthogonal region of phase space in data. Then, using this efficiency, propagate from the control region defined as the inversion of one of the top rejection cuts. The number of surviving top events would then be:



(a) ggH (b) qqH

Figure 4.4: Reconstructed versus generated $p_{\mathrm{T}}^{\mathrm{H}}$ for gluon fusion (a) and VBF (b). Plots are normalized by rows, so that the bin purity is shown on the diagonal.

$$N_{bveto}^{\text{signal}} = N_{btag}^{\text{control}} \cdot \frac{1 - \epsilon_{\text{top}}}{\epsilon_{\text{top}}} \tag{4.4}$$

where $N_{btag}^{\text{control}}$ is the number of events in the inverted control region and ϵ_{top} is the efficiency as measured in data. A full description of the method can be found in [2].

The amount of tW and $t\bar{t}$ backgrounds contaminating the signal phase space is evaluated in a region obtained inverting the b-veto requirement on jets, and then extrapolated to the signal region.

Greater than 0-jets bin

The strategy for the estimation of the top background in events with at least one jet with p_{T} greater than 30 GeVis the following. We first estimate a per jet scale factor for the b-tagging efficiency. This evaluation is performed in a control region, that we will call CtrlTP, containing at least two jets, using a Tag&Probe technique. The procedure to extract these scale factors is presented in Sec. 4.6.1. Then we define a larger statistics control region, CtrlDD, by requiring at least one b-tagged jet and we use the simulation, corrected for the previously computed b-tagging efficiency scale factor, to derive the factor that connects the number of events in CtrlDD to the number of events in the signal region. This second step is explained in detail in Sec. 4.6.1.

Tag&Probe

The Tag&Probe technique is a method to estimate the efficiency of a selection on data. In can be applied whenever one has two objects in one event, by using one of the two, the *tag*, to identify the process of interest, and using the second, the *probe*, to actually measure the efficiency of the selection being studied. In our case we want to measure the b-tagging efficiency, so what we need is a sample with two b-jets per event. The easiest way to construct such a sample is to select $t\bar{t}$ events.

We define a control region, called CrtlTP, which contains events passing the lepton preselection cuts listed in Sec. 4.4.2, and have at least two jets with p_T greater than 30 GeV. One of the two leading jets is required to have a *JetBProbability* score higher than

0.5. Form events in this control region we built *tag-probe* pairs as follows. For each event the two leading jets are considered. If the leading jet passes the *JetBProbability* cut of 0.5, that is considered a *tag*, and the sub-leading jet is the *probe*. In order to avoid any bias that could arise from the probe being always the second jet, the pair is tested also in reverse order, meaning that the sub-leading jet is tested against the *tag* selection, and in case it passes, then the leading jet is used as *probe* in an independent *tag-probe* pair. This means that from each event passing the CrtlTP cuts one can build up to two *tag-probe* pairs.

If the *tag* selection were sufficient to suppress any non top events, one could estimate the efficiency by dividing the number of *tag-probe* pairs in which the *probe* passes the analysis cut *JetBProbability* >1.4 (*tag-pass-probe*) by the total number of *tag-probe* pairs. However this is not the case. In order to estimate the efficiency in the presence of background we have chosen a variable that discriminates between true b-jets and other jets in a $t\bar{t}$ sample. The variable is the p_T of the *probe* jet. For real b-jets this variable has a peak around 60 GeV, while it does not peak for other jets. The idea is to fit simultaneously the p_t spectrum for *probe* jets in *tag-pass-probe* and *tag-fail-probe* pairs, linking together the normalizations of the two samples as follows:

$$N_{TPP} = N_s\epsilon_s + N_b\epsilon_b \tag{4.5}$$

$$N_{TFP} = N_s(1-\epsilon_s) + N_b(1-\epsilon_b) \tag{4.6}$$

where N_{TPP} is the number of *tag-pass-probe* pairs, N_{TFP} is the number of *tag-fail-probe* pairs, N_s is the number of *tag-probe* pairs in which the probe is a b-jet, N_s is the number of *tag-probe* pairs in which the probe is a not b-jet, ϵ_s is the b-tagging efficiency, ϵ_b is the probability of identifying as b-jet a non-b-jets, i.e. the mistag rate.

We have performed a χ^2 simultaneous fit of the *probe* p_T spectrum for *tag-pass-probe* and *tag-fail-probe* pairs, deriving the shapes for true b-jets and non-b-jets from the simulation, and extracting from the fir N_s, N_b, ϵ_s and ϵ_b. The result of the fit on MC simulation is shown in Fig. 4.5. The relevant efficiencies are:

$$\epsilon_s^{MC} = 0.7663 \pm 0.0072 \tag{4.7}$$

$$\epsilon_b^{MC} = 0.208 \pm 0.015 \tag{4.8}$$

We have checked that these values are consistent with the true value for the b-tagging efficiency. The true value is computed by selecting jets that are matched within a cone of $\Delta R < 0.5$ with a generator level b-quark, and counting the faction of those that pass the *JetBProbability* cut of 1.4. This means that the *tag-probe* method does not introduce biases within the MC statistic accuracy.

In order to assess the robustness of the fit we have generated 5000 toy MC samples and fitted them. The toy MC were generated with a statistics equivalent to the one expected in data. All the 5000 fit succeeded, and the pull distributions for ϵ_s and ϵ_b parameters are shown in Fig. 4.6. The pulls are well centered on 0 and have σ close to 1, as expected. An example fit for one of the toys is shown in Fig. 4.7

Before running the fit on data we have tried to validate the shapes used in the fit with data. To do so, we have made a much more pure $t\bar{t}$ selection, by requiring exactly two jets with *JetBProbability* score higher than 1.5 and no additional b-tagged jets, even if they have p_{T} smaller than 30 GeV. On this purer sample we have compared data against the shape used to fit the true b-jets in the *tag-pass-probe* distribution. The result is shown in Fig. 4.8 and shows good agreement.

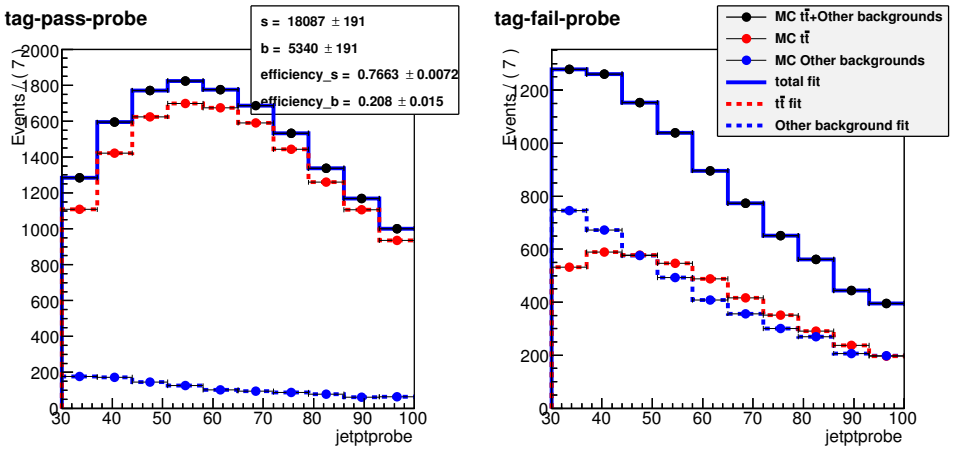


Figure 4.5: Simultaneous fit of the *tag-pass-probe* and *tag-fail-probe* pairs in the MC.

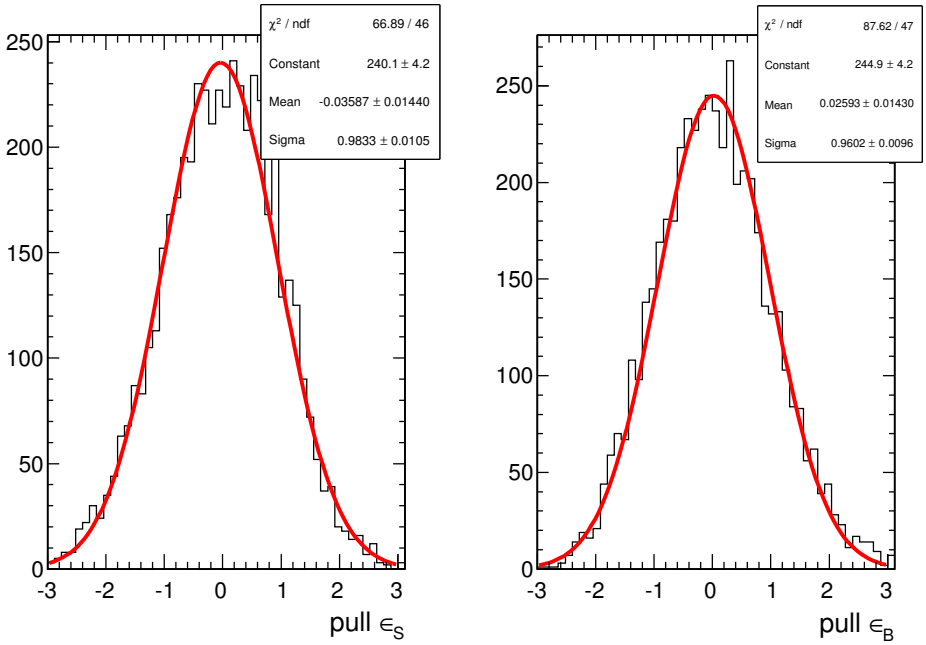


Figure 4.6: Pulls of the ϵ_s and ϵ_b parameters in 5000 toy MC.

We have finally performed the fit on data, as shown in Fig. 4.9, which results in in the following efficiencies:

$$\epsilon_s^{Data} = 0.769 \pm 0.022 \tag{4.9}$$

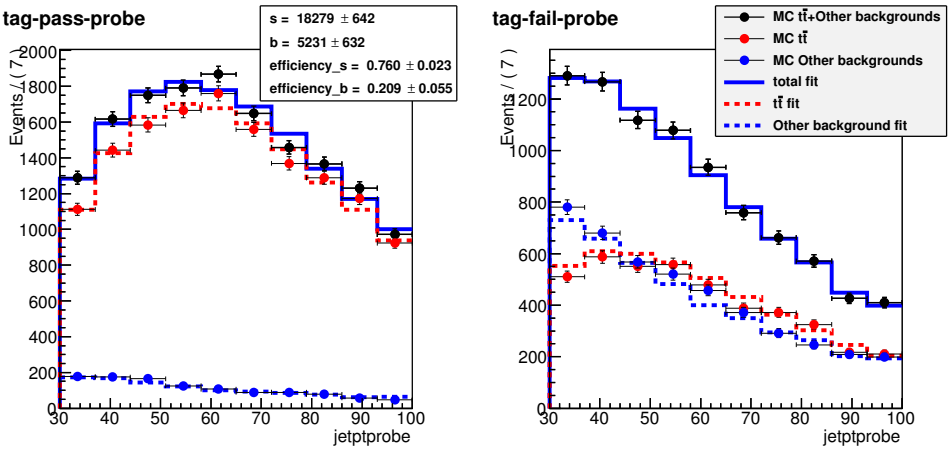


Figure 4.7: Fit of a toy MC sample.

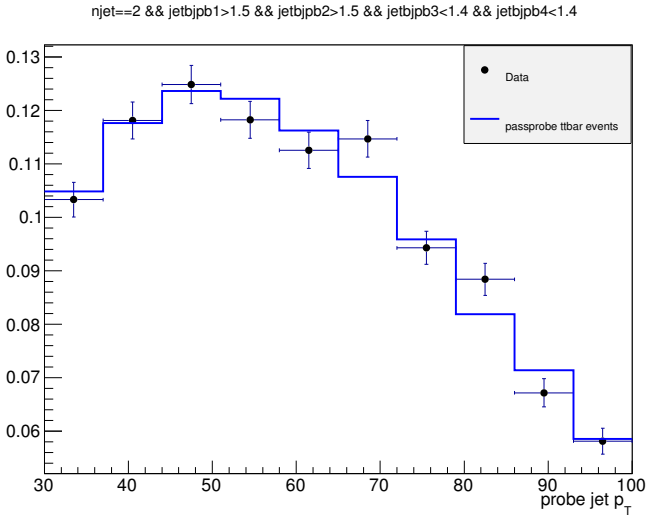


Figure 4.8: Shape comparison for the *probe* p_{T} spectrum in data and in MC in a very pure $t\bar{t}$ sample.

$$\epsilon_b^{Data} = 0.121 \pm 0.054 \tag{4.10}$$

Further checks on the Tag&Probe efficiencies are shown in Appendix **??**, which concern the uncertainty related to the not perfect knowledge of the $tW/t\bar{t}$ ratio in the MC.

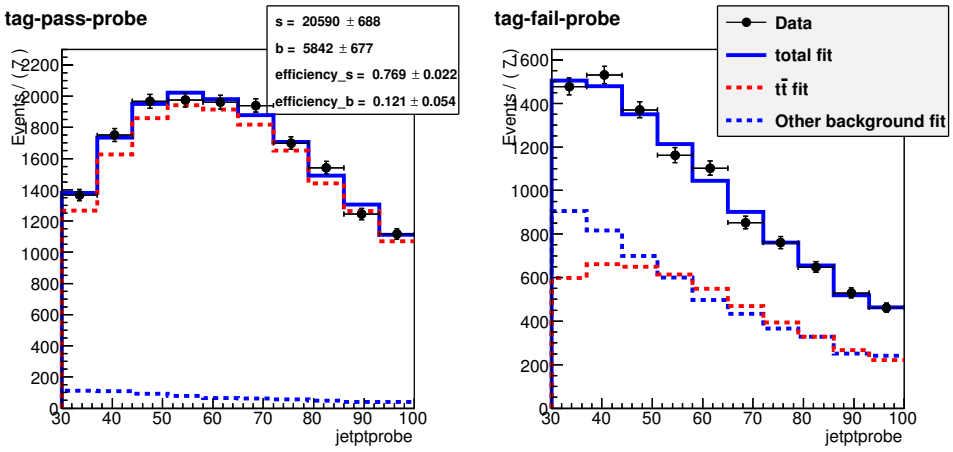


Figure 4.9: Simultaneous fit of the *tag-pass-probe* and *tag-fail-probe* pairs in data.

Data driven estimation

In addition to the b-tagging efficiency, the other ingredient to estimate the $t\bar{t}$ background is the process cross section. The idea is to measure the cross section in a $t\bar{t}$ enriched control region, that we call CtrlDD. CtrlDD is defined according to the lepton preselection cuts defined in Sec. 4.4.2, and requiring in addition at least one jet with *JetBProbability* score higher than 1.4.

From the simulation we derive the factor α that connects CrtlDD to the signal region, from the ratio of $t\bar{t}$ events in the two regions.

$$\alpha = \frac{N^{SIG}_{t\bar{t}\ MC}}{N^{CtrlDD}_{t\bar{t}\ MC}}. \tag{4.11}$$

We then count events CtrlDD in data, we subtract the expected number of events from non-$t\bar{t}$ backgrounds, and we obtain $N^{CtrlDD}_{t\bar{t}\ Data}$. We finally obtain the number of expected $t\bar{t}$ events in the signal region ($N^{SIG}_{t\bar{t}\ Data}$) as:

$$N^{SIG}_{t\bar{t}\ Data} = \alpha N^{CtrlDD}_{t\bar{t}\ Data}. \tag{4.12}$$

In evaluating α and its error we made use of the b-tagging efficiencies determined in Sec. 4.6.1. For each event we derive an efficiency scale factor and a mistag rate scale factor, depending on whether the event is in the signal or CtrlDD regions.

$$SF_{SIG} = \left(\frac{1-\epsilon_s^{Data}}{1-\epsilon_s^{MC}}\right)^{min(2,n_{b-jets})} \left(\frac{1-\epsilon_b^{Data}}{1-\epsilon_b^{MC}}\right)^{n_{non-b-jets}} \tag{4.13}$$

$$SF_{CtrlDD} = \left(\frac{\epsilon_s^{Data}}{\epsilon_s^{MC}}\right)^{(jet1==b-jet)} \left(\frac{\epsilon_b^{Data}}{\epsilon_b^{MC}}\right)^{(jet1==non-b-jets)} \tag{4.14}$$

where n_{b-jets}) is the number of true b-jets in the event and $n_{non-b-jets}$ is the number of non-b-jets in the event. The writing $jet1 == b-jet$ ($jet1 == non-b-jets$) is a boolean flag that is true when the leading jet, the one used for the CtrlDD selection, is (not) a true b-jet.

Since the efficiency and mistag rate that we have measured on data are close to the one in the MC we have decided to assume a scale factor of 1 for both b-tagging efficiency and mis-tag rate. This means that the central values of the scale factors defined in Eq. 4.13 and Eq. 4.14 is 1, but these numbers have an error that is derived assuming an uncertainty on ϵ_s^{Data} and ϵ_b^{Data} that covers both the statistical error from the fit of the two quantities and the difference with respect to the MC. This results in an up variation and a down variation of the scale factors in the signal region and CtrlDD regions, that is used to derive an error on α.

We have decided to make a data driven estimation of the $t\bar{t}$ background with the method described above in each of the $p_{\mathrm{T}}^{\mathrm{H}}$ bins independently. The reason why we have chosen to make this estimation in $p_{\mathrm{T}}^{\mathrm{H}}$ bins, rather than inclusively is explained in Fig. 4.10. In this plot the $t\bar{t}$ background is normalized to the cross section measured by CMS. The binning is the same chosen for the analysis. As shown in the ratio plot, an overall normalization factor would not be able to accommodate for the variations of the Data/MC ratio from bin to bin.

The α factors for each bin and the number of events in signal, CtrlDD region in MC as well as in data are listed in Tab. 4.2.

A comparison of the $m_{\ell\ell}$ distribution in the six $p_{\mathrm{T}}^{\mathrm{H}}$ bins used in the analysis in CtrlDD after the data driven correction is shown in Fig. 4.11

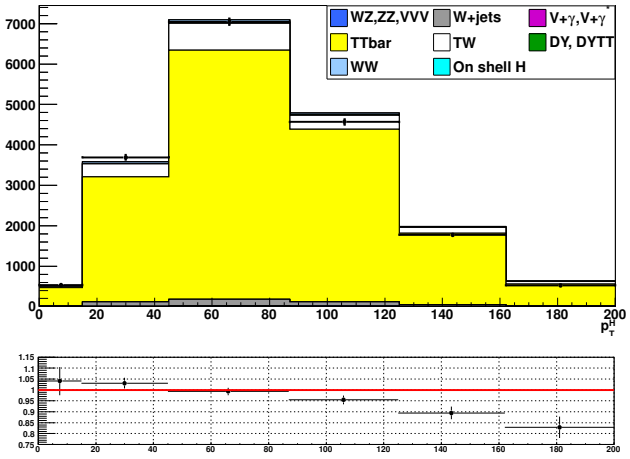


Figure 4.10: $p_{\mathrm{T}}^{\mathrm{H}}$ variable in the CtrlDD control region.

p_T^H bin	N_{CTRL}^{DATA}	N_{CTRL}^{TOP}	N_{SIG}^{TOP}	α	$\Delta\alpha$
1	406.71	358.78	117.83	0.328	0.075
2	2930.14	2703.44	859.08	0.318	0.071
3	5481.02	5207.48	1506.05	0.289	0.065
4	4126.35	4032.56	861.22	0.214	0.052
5	1612.64	1654.27	304.69	0.184	0.055
6	647.50	760.37	201.70	0.265	0.147

Table 4.2: Table of data driven scale factors.

4.6.2 WW background

For what the WW background shape is concerned the prediction from the Monte-Carlo simulation has been used. This background is divided into six different parts, corresponding to the six bins of $p_{\mathrm{T}}^{\mathrm{H}}$ considered. In each bin the normalization of the WW background is left free to float and is thus adjusted to match the data by the fit. In this way we minimize an effect that has been observed also in [14], that is a difference in shape between the p_T^{WW} theory prediction and the distribution provided by the MC simulation, in our case by MADGRAPH.

In figure 4.12 a comparison is shown between the p_T^{WW} spectra of two different qqWW samples: the blue line corresponds to the WW MADGRAPH samples that we use in this analysis and the red line refers to the same sample in which a reweighing has been applied in order to match the theoretical prediction at NLO+NNLL precision. A shape discrepancy can be clearly observed and the effect becomes larger at high values of $p_{\mathrm{T}}^{\mathrm{H}}$.

In order to assess if these discrepancy has a not negligible effect on the shapes of the variables that we use for the fit, $m_{\ell\ell}$ and m_{T}, we checked these distributions in every $p_{\mathrm{T}}^{\mathrm{H}}$ bin, comparing several samples. In particular we compared the MADGRAPH sample used for the nominal shape, the MADGRAPH sample with NLO+NNLL reweighting, a POWHEG NLO sample and an aMC@NLO sample. The results of this comparison are shown in figures 4.13 and 4.14. The discrepancy in shape among the different models is within the statistical accuracy of the MC samples.

4.6.3 Other backgrounds

W+jets background

Backgrounds containing one or two fake leptons are estimated from events selected with relaxed lepton quality criteria, using the efficiencies for real and fake leptons to pass the tight lepton quality cuts of the analysis.

A data-driven approach, described in detail in [15] and [16], is pursued to estimate this background. A set of loosely selected lepton-like objects, referred to as the 'fakeable object' or "denominator" from here on, is defined in a sample of events dominated by dijet production. The denominator object definition used in the full 2012 data is described in [17].

To measure the fake rate we count how many fakeable objects pass the full lepton selection of the analysis, parametrized as a function of the phase space of the fakeable lepton, therefore it is extracted in bins of η and p_T.

The ratio of the fully identified lepton, referred as "numerator", to the fakeable objects is taken as the probability for a fakeable object to fake a lepton:

$$\mathrm{F}ake\ Rate = \frac{\#of\ fully\ reconstructed\ leptons}{\#of\ fakeable\ objects} \tag{4.15}$$

It is then used to extrapolate from the loose leptons sample to a sample of leptons satisfying the full selection.

The details of the method implementation can be found in [2]. The systematic uncertainty is evaluated by varying the jet thresholds in the di-jet control sample, and by performing a closure test in the same-sign data sample (see [2]). In both cases it is about 36%.

Drell-Yan to $\tau\tau$ background

The low $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ threshold in eμ final state requires the consideration of the contribution from $Z/\gamma^* \to \tau^+\tau^-$ that is infact estimated from data.

This is accomplished by using $Z/\gamma^* \to \mu^+\mu^-$events and replacing muons with a simulated $\tau \to l\nu_\tau\bar{\nu}_e$ decay [18].

After replacing muons from $Z/\gamma^* \to \mu^+\mu^-$decays with simulated τ decays, the set of pseudo $Z/\gamma^* \to \tau^+\tau^-$events undergoes the reconstruction step.

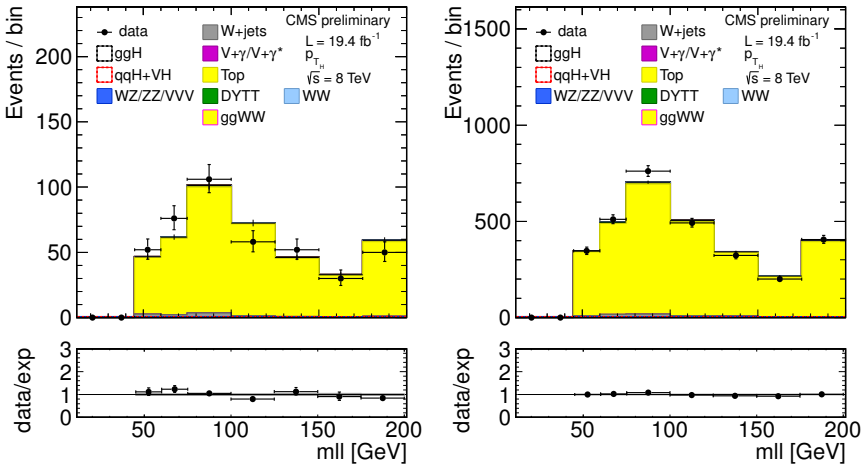
Good agreement in kinematic distributions for this sample and a Monte Carlo based $Z/\gamma^* \to \tau^+\tau^-$sample is found.

The global normalization of pseudo $Z/\gamma^* \to \tau^+\tau^-$events is checked in the low m_{T} spectrum where a rather pure $Z/\gamma^* \to \tau^+\tau^-$sample is expected.

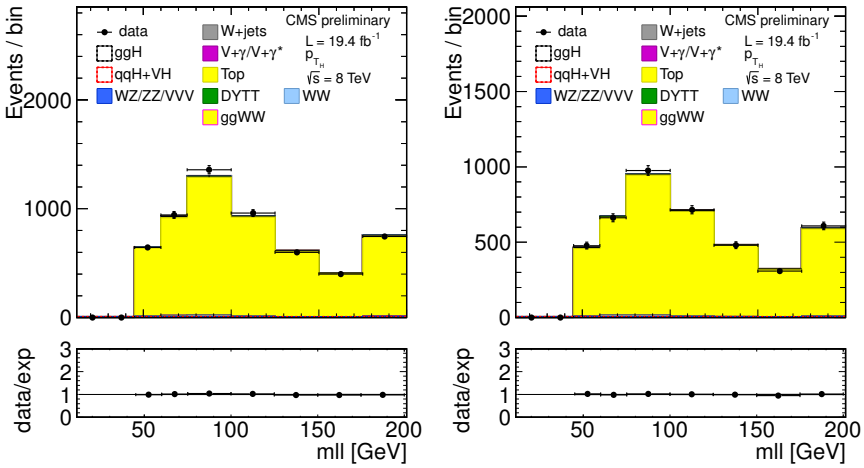
ZZ, WZ and W+γ backgrounds

The WZ and ZZ backgrounds are partially estimated from data when the two selected leptons come from the same Z boson. If the leptons come from different bosons the contribution is expected to be small. The WZ component is largely rejected by requiring only two high p_{T} isolated leptons in the event.

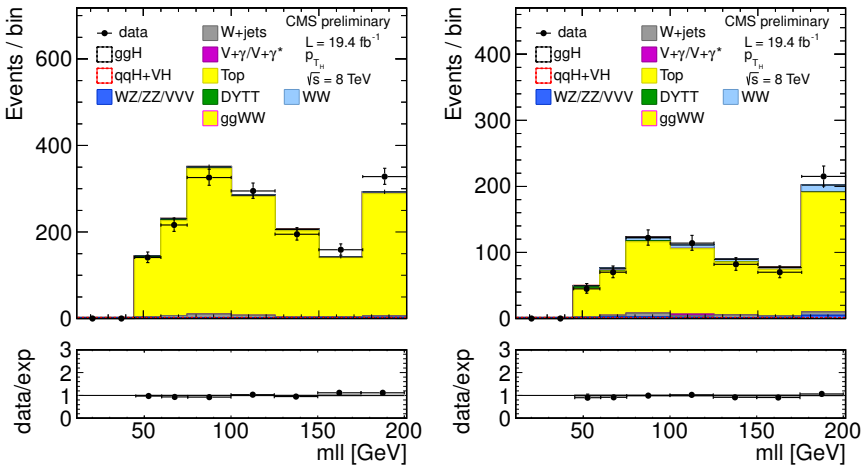
The W+$\gamma^{(*)}$ background, where the photon decays to an electron-positron pair, is expected to be very small, thanks to the stringent photon conversion requirements. Since the WZ simulated sample has a generation level cut on the di-lepton invariant mass ($m_{\ell\ell} >$ 12 GeV) and the cross-section raises quickly with the lowering of this threshold, a dedicated MADGRAPH sample has been produced with lower momentum cuts on two of the three leptons ($p_{\mathrm{T}} > 5$ GeV) and no cut on the third one. The surviving contribution estimated with this sample is still very small, and since the uncertainty on the cross-section for the covered phase space is large, a conservative 100% uncertainty has been given to it. A k-factor for W+γ^* of 1.5 ± 0.5 based on a dedicated measurement of tri-lepton decays, W+$\gamma^* \to e\mu\mu$ and W+$\gamma^* \to \mu\mu\mu$, is applied [19]. The contribution of W+$\gamma^{(*)}$ is also constrained by a closure test with same sign leptons on data, which reveals a good compatibility of the data with the expected background.



(a) $p_{\mathrm{T}}^{\mathrm{H}} < 15\mathrm{GeV}$ (b) $15\mathrm{GeV} < p_{\mathrm{T}}^{\mathrm{H}} < 45\mathrm{GeV}$



(c) $45\mathrm{GeV} < p_{\mathrm{T}}^{\mathrm{H}} < 85\mathrm{GeV}$ (d) $85\mathrm{GeV} < p_{\mathrm{T}}^{\mathrm{H}} < 125\mathrm{GeV}$



(e) $125\mathrm{GeV} < p_{\mathrm{T}}^{\mathrm{H}} < 165\mathrm{GeV}$ (f) $p_{\mathrm{T}}^{\mathrm{H}} > 165\mathrm{GeV}$

Figure 4.11: $m_{\ell\ell}$ distributions in the CtrlDD region for the different $p_{\mathrm{T}}^{\mathrm{H}}$ bins.

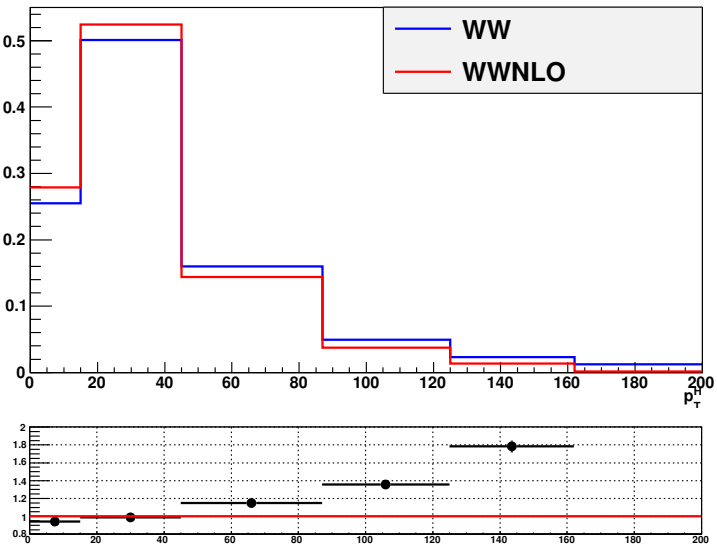
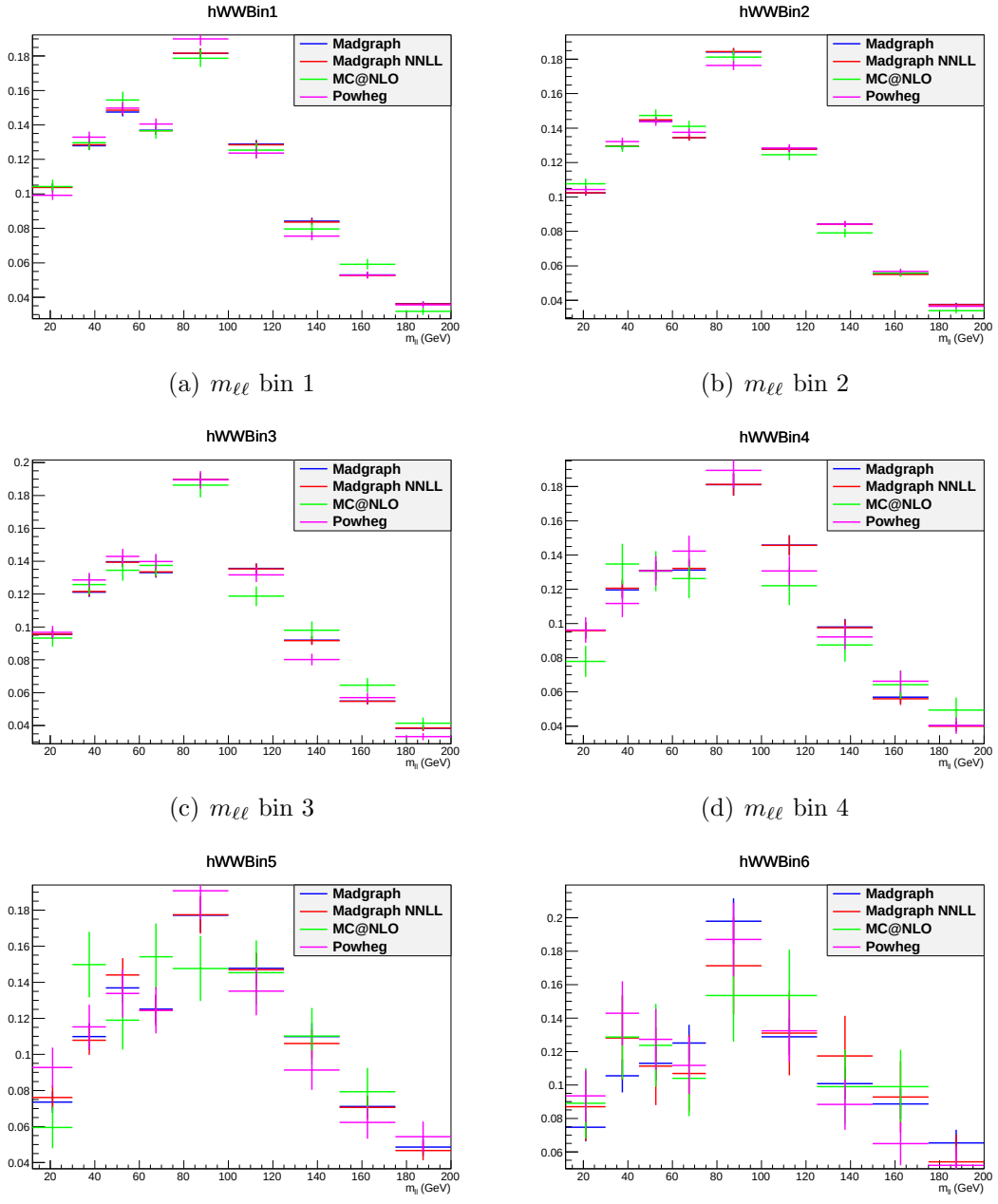

WW
WWNLO
0.5
0.4
0.3
0.2
0.1
0
0
20
40
60
80
100
120
140
160
180
200
p_T^H
2
1.8
1.6
1.4
1.2
1
0.8

Figure 4.12



(a) $m_{\ell\ell}$ bin 1

(b) $m_{\ell\ell}$ bin 2

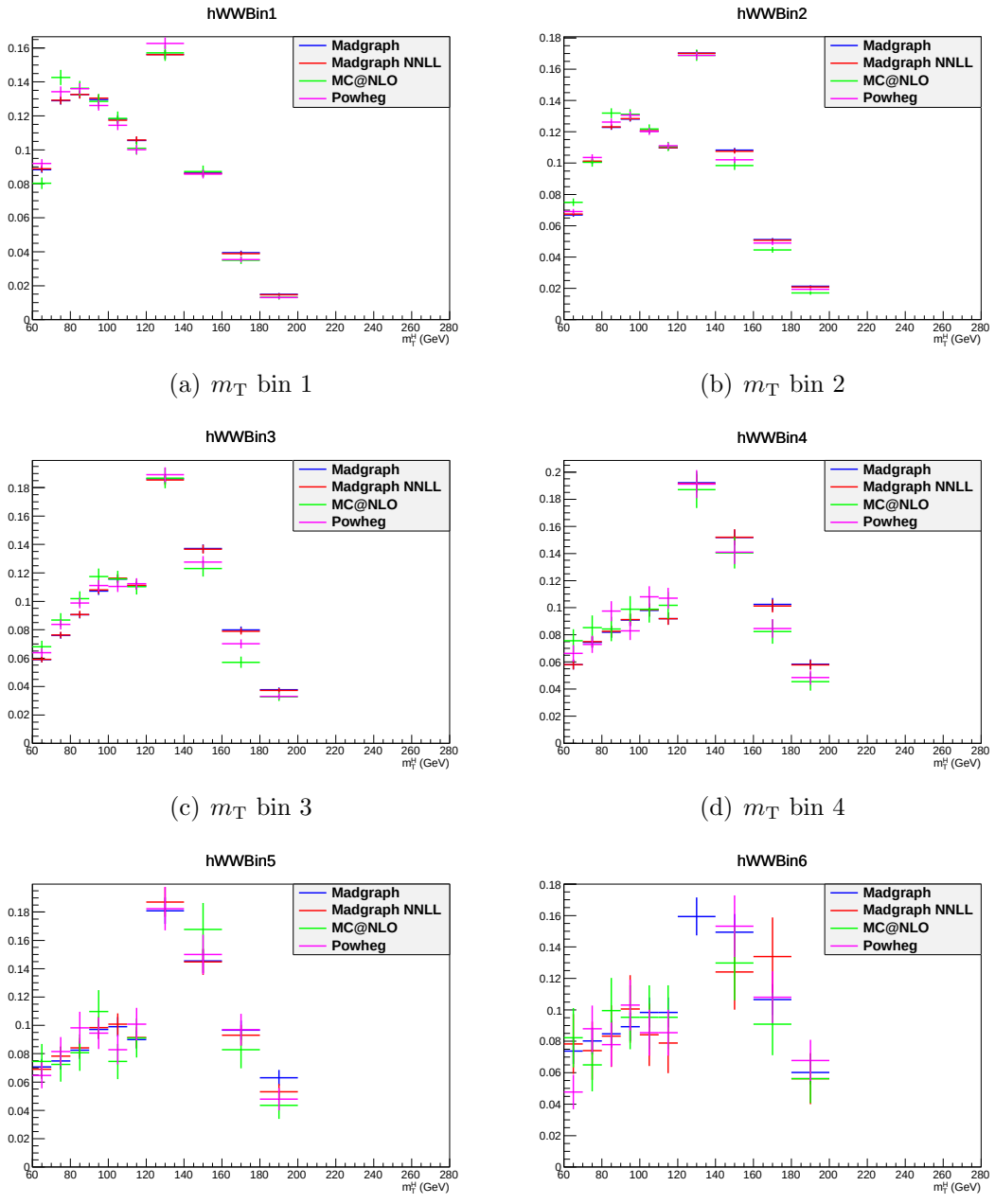
(c) $m_{\ell\ell}$ bin 3

(d) $m_{\ell\ell}$ bin 4

(e) $m_{\ell\ell}$ bin 5

(f) $m_{\ell\ell}$ bin 6

Figure 4.13: Comparison between the default WW background sample and other theoretical models for the $m_{\ell\ell}$ distributions in every $p_{\mathrm{T}}^{\mathrm{H}}$ bin.



(a) m_{T} bin 1

(b) m_{T} bin 2

(c) m_{T} bin 3

(d) m_{T} bin 4

(e) m_{T} bin 5

(f) m_{T} bin 6

Figure 4.14: Comparison between the default WW background sample and other theoretical models for the m_{T} distributions in every $p_{\mathrm{T}}^{\mathrm{H}}$ bin.

Chapter 5

First H → WW results at 13 TeV

5.1 Higgs boson search at 13TeV

5.2 Search for a high mass resonance in the WW decay channel at 13TeV

5.3 Conclusions

Bibliography

[1] Serguei Chatrchyan et al. "Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states". In: *JHEP* 01 (2014), p. 096. DOI: 10.1007/JHEP01(2014)096. arXiv: 1312.1129 [hep-ex].

[2] J. Brochero *et al.* "Higgs Boson Decaying to WW in the Leptonic Final State using 2011 and 2012 Data". In: *CMS Note* AN-2013/022 (2013).

[3] J. Brochero *et al.* "Search for a Standard Model Higgs boson produced via Vector Boson Fusion, in the decay channel $H \to WW \to l\nu l\nu$ using 2011 and 2012 data". In: *CMS Note* AN-2013/097 (2013).

[4] L. Bauerdick *et al.* "A Higgs Boson Search in the Fully Leptonic W+W- Final State (update for ICHEP2012 conference)". In: *CMS Note* AN-2012/228 (2012).

[5] L. Bauerdick *et al.* "A Higgs Boson Search in the Fully Leptonic W W Final State". In: *CMS Note* AN-2013/052 (2013).

[6] S. Frixione, P. Nason and C. Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method". In: *arXiv:0709.2092v1* (2007).

[7] J. Alwall *et al.* "Madgraph". In: *JHEP* 0709 (2007), p. 028.

[8] T. Sjostrand, S. Mrenna and P. Skands. "PYTHIA". In: *JHEP* 0605 (2006), p. 026.

[9] Hung-Liang Lai et al. "Uncertainty induced by QCD coupling in the CTEQ global analysis of parton distributions". In: *Phys. Rev. D* 82 (2010), p. 054021. DOI: 10.1103/PhysRevD.82.054021. arXiv: 1004.4624 [hep-ph].

[10] Huang-Liang Lai et al. "New parton distributions for collider physics". In: *Phys. Rev. D* 82 (2010), p. 074024. DOI: 10.1103/PhysRevD.82.074024. arXiv: 1007.2241 [hep-ph].

[11] S. Agostinelli et al. "GEANT4: A simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.

[12] J. Brochero *et al.* "Search for the Higgs Boson Decaying to WW in the Fully Leptonic Final State at 8 TeV". In: *CMS Note* AN-2012/194 (2012).

[13] CMS collaboration. *Search for associated Higgs boson production (VH) with $H \to W^+W^- \to \ell\nu\ell\nu$ and hadronic V decay in pp collisions at sqrt(s) = 7 TeV and 8 TeV.* CMS PAS 2013/084. 2013. URL: http://cds.cern.ch/record/1560844.

[14] CMS collaboration. *WW Cross Section Measurement at sqrt(s) = 8 TeV.* CMS AN 2014/056. 2014.

[15] H. Bakhshian et al. "Computing the contamination from fakes in leptonic final states". In: *CMS Note* AN-2010/261 (2010).

[16] H. Bakhshian et al. "Lepton fake rates in dilepton final states". In: *CMS Note* AN-2010/397 (2010).

[17] J. Brochero *et al.* "Higgs Boson Decaying to WW in the Fully Leptonic Final State". In: *CMS Note* AN-2012/378 (2012).

[18] M. Bluj, *et al.* "Modelling of tautau final states by embedding tau pairs in Z to mumu events". In: *CMS Note* (2011).

[19] CMS Collaboration. "Evidence for a particle decaying to W+W- in the fully leptonic final state in a standard model Higgs boson search in pp collisions at the LHC". In: *Hig12042 Twiki* (2012). URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/Hig12042TWiki#Study_on_W_g_background.