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 $\mathrm{H} \to \mathrm{W}^+\mathrm{W}^-$ 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 8 TeV, corresponding to an integrated luminosity of $19.4\,\mathrm{fb}^{-1}$. 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

2.5. The CMS framework

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 8 TeV using $\mathbf{H \to WW \to 2\ell 2\nu}$ decays

4.1. Introduction

The Higgs boson production at hadron colliders is characterized by $p_{\mathrm{T}}^{\mathrm{H}}$ and η. The η distribution is essentially driven by the PDF of the partons in the colliding hadrons, and it is only mildly sensitive to radiative corrections. The $p_{\mathrm{T}}^{\mathrm{H}}$ distribution is instead sensitive to QCD radiative corrections. Considering the ggH production mode, at LO in perturbation theory, $\mathcal{O}(\alpha_s^2)$, the Higgs boson is always produced with $p_{\mathrm{T}}^{\mathrm{H}}$ equal to zero. Indeed in order to have p_{T} different from zero, the Higgs boson has to recoil at least against one parton. Higher order corrections to the ggH process are numerically large and are known at NLO including full top quark mass dependence [1, 2], and at NNLO using the so-called large-m_{t} approximation [3, 4, 5], in which the top quark mass is assumed to be very large and the fermionic loop is replaced by an effective vertex of interaction. Starting from the NLO, the Higgs boson can be produced recoiling against other final state partons, resulting in a finite $p_{\mathrm{T}}^{\mathrm{H}}$. For this reason the LO process for Higgs production at $p_{\mathrm{T}} \neq 0$ is at $\mathcal{O}(\alpha_s^3)$, and the counting of perturbative orders differs between inclusive Higgs boson production and $p_{\mathrm{T}}^{\mathrm{H}}$ distribution. Also, NNLO QCD corrections in the $p_{\mathrm{T}}^{\mathrm{H}}$ observable have recently been shown [6].

When $p_{\mathrm{T}}^{\mathrm{H}} \sim m_{\mathrm{H}}$ the QCD radiative corrections to $p_{\mathrm{T}}^{\mathrm{H}}$ differential cross section are theoretically evaluated using fixed-order calculations. When $p_{\mathrm{T}}^{\mathrm{H}} \ll m_{\mathrm{H}}$ the perturbative expansion does not converge due to the presence of large logarithmic terms of the form $\alpha_s^n \ln^{2n} m_{\mathrm{H}}^2/p_{\mathrm{T}}^2$, leading to a divergence of $d\sigma/dp_{\mathrm{T}}$ in the limit of $p_{\mathrm{T}} \to 0$. For computing the $p_{\mathrm{T}}^{\mathrm{H}}$ spectrum in this region soft-gluon resummation techniques are used, and matched to the fixed-order calculation in the $p_{\mathrm{T}}^{\mathrm{H}} \sim m_{\mathrm{H}}$ region. For the $p_{\mathrm{T}}^{\mathrm{H}}$ differential cross section the large-m_{t} calculation is a crude approximation, since it is known that the top quark mass has a non-negligible effect on the shape of the spectrum. Moreover the inclusion of the bottom quark contribution in the fermionic loop can significantly modify the p_T^H shape [7], as shown in Fig. 4.1. Hence, a precise experimental measurement of the p_T^H spectrum is important to test the existing SM calculations.

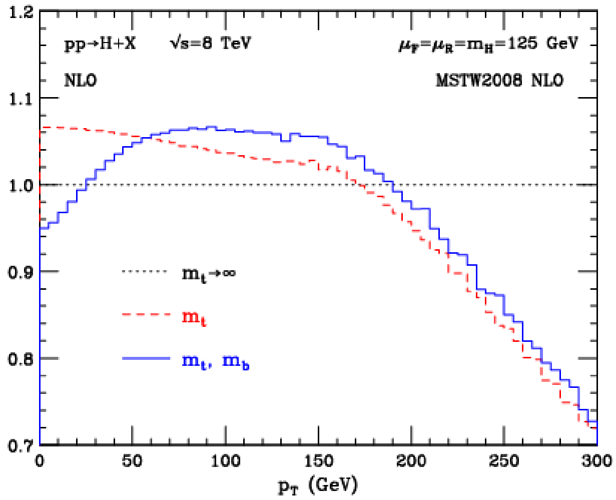


Figure 4.1.: p_T^H distribution computed at NLO (α_s^3) and normalized to the calculation obtained in the large-m_t approximation. The red dashed line corresponds to the calculation including the top quark mass while the blue line refers to the calculation including also the bottom quark effects.

Possible extensions of the SM predict a modification of the Higgs boson couplings to gluons and to the top quark. Many of these models actually predict the existence of new states that interact with the SM Higgs boson, but are beyond the direct production reach at the actual LHC energies. The effect of these new states could however show up as a deviation of the Higgs boson couplings with respect to the SM expectation. The modification of the couplings, as shown in Refs. [8, 9], can change the kinematics of the Higgs boson production and the effect can be particularly sizeable in the tail of the p_T^H distribution. Other models, such as Composite Higgs [10], predict the existence of top-partners, which are heavy resonances with the same quantum numbers as the top quark, that can interact with the Higgs boson in the ggH fermionic loop, changing the p_T^H shape with respect to what the SM predicts [11]. The measurement of the p_T^H spectrum is thus a useful tool for indirect searches of new particles predicted by theories beyond the SM.

Measurements of the fiducial cross sections and of several differential distributions, using the $\sqrt{s} = 8$ TeV LHC data, have been reported by ATLAS [12, 13, 14] and CMS [15, 16] for the $H \to ZZ \to 4\ell$ ($\ell = e, \mu$) and $H \to \gamma\gamma$ decay channels. In this chapter a measurement of the fiducial cross section times branching fraction ($\sigma \times \mathcal{B}$) and p_T spectrum for Higgs boson production in $H \to WW \to e^{\pm}\mu^{\mp}\nu\nu$ decays, based on $\sqrt{s} = 8$ TeV LHC data, is reported.

The analysis is performed looking at different flavour leptons in the final state in order to suppress the sizeable contribution of backgrounds containing a same-flavour lepton pair originating from Z boson decay.

Although the H $\rightarrow$ WW $\rightarrow 2\ell 2\nu$ channel has lower resolution in the $p_{\mathrm{T}}^{\mathrm{H}}$ measurement compared to the H $\rightarrow \gamma\gamma$ and H $\rightarrow$ ZZ $\rightarrow 4\ell$ channels because of neutrinos in the final state, the channel has a significantly larger $\sigma \times \mathcal{B}$, exceeding those for H $\rightarrow \gamma\gamma$ by a factor of 10 and H $\rightarrow$ ZZ $\rightarrow 4\ell$ by a factor of 85 for a Higgs boson mass of 125 GeV [17], and is characterized by good signal sensitivity. Such sensitivity allowed the observation of a Higgs boson at the level of 4.3 (5.8 expected) standard deviations for a mass hypothesis of 125.6 GeV using the full LHC data set at 7 and 8 TeV [18].

The measurement is performed in a fiducial phase space defined by kinematic requirements on the leptons that closely match the experimental event selection.

The effect of the limited detector resolution, as well as the selection efficiency with respect to the fiducial phase space are corrected to particle level with an unfolding procedure [19], as explained in Sec. 4.7.

4.2. Data sets, triggers and MC samples

This analysis relies on the published H $\rightarrow$ WW measurements [18] in terms of code, selections and background estimates for both the ggH and VBF prodocution mechanisms.

4.2.1. Data sets and triggers

The data sets used for the analysis correspond to 19.4 fb^{-1} at $\sqrt{s} = 8$ TeV of 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 checked 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.

For the data samples, the events are required to fire one of the unprescaled single-electron, single-muon or muon-electron triggers. Due the rather high LHC instantaneous luminosity the single-lepton triggers must have high HLT p_{T} thresholds, otherwise the rate of these triggers would be too large to be sustained. The double-lepton triggers allow to lower down the p_{T} thresholds while keeping a sustainable trigger rate, thus maintaining a good sensitivity to the Higgs boson signal, for which the lepton p_{T} can be rather small. A brief overview of the HLT p_{T} criteria on the leptons is given in Table 4.1. While the HLT lepton p_{T} thresholds of 17 and 8 GeV for 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	8 TeV
Single-Electron	$p_{\mathrm{T}} > 27$ GeV
Single-Muon	$p_{\mathrm{T}} > 24$ GeV
Muon-Electron	$p_{\mathrm{T}} > 17$ and 8 GeV
Electron-Muon	$p_{\mathrm{T}} > 17$ and 8 GeV

The trigger is not simulated in MC samples but the combined trigger efficiency is estimated from data and applied as a weight to all simulated events. The trigger efficiency for single and double lepton triggers is calculated using a Tag and Probe technique separately for muons and electrons, in bins of η and p_{T}. The Tag and Probe method uses a known mass resonance (e.g. J/Ψ, Z) to select particles of the desired type, and probe the efficiency of a particular selection criterion on these particles. In general the "tag" is an object that passes a set of very tight selection criteria designed to isolate the required particle type. Tags are often referred to as a golden electrons or muons and the fake rate for passing tag selection criteria should be very small. A generic set of the desired particle type (i.e. with potentially very loose selection criteria) known as "probes" is selected by pairing these objects with tags such that the invariant mass of the combination is consistent with the mass of the resonance. Combinatoric backgrounds may be eliminated through any of a variety of background subtraction methods such as fitting, or sideband subtraction. The definition of the probe objects depend on the specifics of the selection criterion being examined. The simple expression to get the efficiency ϵ as a function of p_{T} and η is given below:

$$\epsilon(p_{\mathrm{T}}, \eta) = \frac{N_{\mathrm{pass}}^{\mathrm{probe}}}{N_{\mathrm{pass}}^{\mathrm{probe}} + N_{\mathrm{fail}}^{\mathrm{probe}}} \tag{4.1}$$

For double lepton triggers the efficiency is calculated separately for each leg of the trigger and then combined together. In the calculation the efficiencies of the two trigger legs are considered as independent, given that the correlations are very small. The combined efficiency is then used as a kinematics-dependent weight to be applied on top of simulated events.

The event efficiency ϵ_{ev} for an event with two leptons to pass the single lepton trigger is given by the following formula:

$$\epsilon_{\rm ev} = 1 - (1 - \epsilon_{S,\ell 1}) \cdot (1 - \epsilon_{S,\ell 2}) \quad , \tag{4.2}$$

where $\epsilon_{S,\ell 1}$ and $\epsilon_{S,\ell 2}$ are the efficiencies for the leading and subleading lepton to pass the single lepton trigger. In other words, the dilepton event passes the single lepton trigger if either one of the two leptons passes the single lepton trigger, excluding the cases for which both leptons pass the trigger. For double lepton triggers, the event efficiency can be written as:

$$\epsilon_{\rm ev} = \epsilon^{\rm lead}_{D,\ell 1} \cdot \epsilon^{\rm trail}_{D,\ell 2} + (1 - \epsilon^{\rm lead}_{D,\ell 1} \cdot \epsilon^{\rm trail}_{D,\ell 2}) \cdot \epsilon^{\rm trail}_{D,\ell 1} \cdot \epsilon^{\rm lead}_{D,\ell 2} \quad , \tag{4.3}$$

where $\epsilon^{{\rm lead}(trail)}_{D,\ell 1}$ is the efficiency of the first lepton to pass the leading (trailing) leg of the double lepton trigger, and $\epsilon^{{\rm lead}(trail)}_{D,\ell 2}$ is the efficiency of the second lepton to pass the leading (trailing) leg of the double lepton trigger. The final event efficiency applied to reweight the events in simulation is given by the boolean OR of the event efficiencies corresponding to the single and double lepton triggers, which, using Eqs. (4.2) and (4.3), can be written as:

$$\begin{aligned} \epsilon_{\rm ev} = {} & 1 - (1 - \epsilon_{S,\ell 1}) \cdot (1 - \epsilon_{S,\ell 2}) + \\ & + (1 - \epsilon_{S,\ell 1}) \cdot (1 - \epsilon_{S,\ell 2}) \cdot \\ & \cdot [\epsilon^{\rm lead}_{D,\ell 1} \cdot \epsilon^{\rm trail}_{D,\ell 2} + (1 - \epsilon^{\rm lead}_{D,\ell 1} \cdot \epsilon^{\rm trail}_{D,\ell 2}) \cdot \epsilon^{\rm trail}_{D,\ell 1} \cdot \epsilon^{\rm lead}_{D,\ell 2}] \quad . \end{aligned} \tag{4.4}$$

The term that multiplies the double lepton trigger event efficiency is needed to ensure that the events passing the double lepton trigger do not pass also the single lepton trigger.

4.2.2. Monte-Carlo samples

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

- The first version of the POWHEG program [20, 21, 22, 23, 24] (POWHEG V1) 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 [25], with NLO accuracy.

- The qq $\to$ $\mathrm{W^+W^-}$, Drell-Yan, ZZ, WZ, Wγ, Wγ^*, tri-bosons and W+jets processes are generated using the MADGRAPH 5.1.3 [26] event generator.

- The gg$\to W^+W^-$ process is generated using the GG2WW 3.1 generator [27] and its cross section is scaled to the approximate NLO prediction [28, 29].
- The VH process is simulated using PYTHIA 6.426 [30].

For leading-order generators samples, the CTEQ6L [31] set of parton distribution functions (PDF) is used, while CT10 [32] is used for next-to-leading order (NLO) ones. Cross section calculations [33] at next-to-next-to-leading order (NNLO) are used for the $H \to WW$ process, while NLO calculations are used for background cross sections. The $H \to WW$ process simulation is reweighted so that the p_T^H spectrum and inclusive production cross section closely match the SM calculations that have NNLO+NNLL pQCD accuracy in the description of the Higgs boson inclusive production, in accordance with the LHC Higgs Cross Section Working Group recommendations [17]. The reweighting of the p_T^H spectrum is achieved by tuning the POWHEG generator, as described in detail in Ref. [34]. Cross sections computed with NLO pQCD accuracy [17] are used for the background processes. The contribution of the $t\bar{t}H$ production mechanisms is checked to be negligible in each bin of p_T^H (below 1%) and is not included among the different production mechanisms. In Fig. 4.2 the relative fraction of the four production mechanisms is shown for each p_T^H bin.

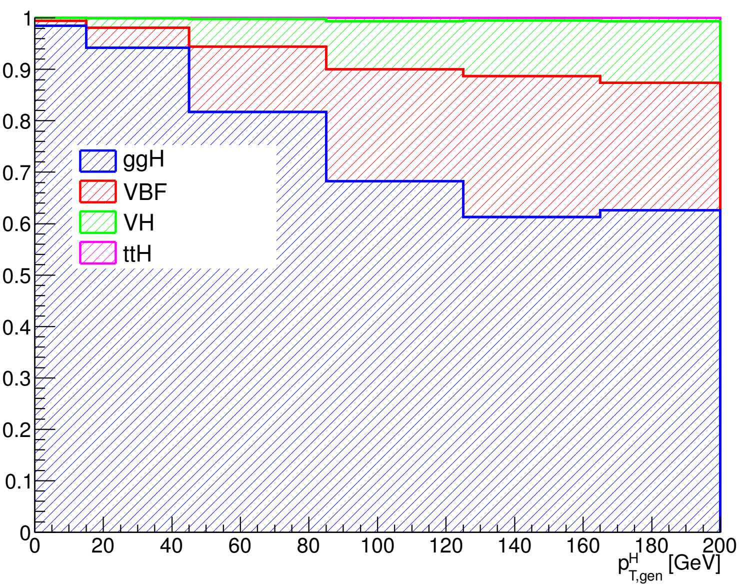


Figure 4.2.: Relatve fraction of ggH, VBF, VH and ttH in each bin of the Higgs boson transverse momentum.

For all processes, the detector response is simulated using a detailed description of the CMS detector, based on the GEANT4 package [35].

Minimum bias events are superimposed on the simulated events to emulate the additional pp interactions per bunch crossing (pileup). 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 pileup) have been generated poissonianly sampling from a distribution similar to what is expected from data. For a given range of analyzed runs, the mean number of pileup 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. The average number of pileup events per beam crossing in the 2011 data is about 10, and in the 2012 data it is about 20.

The simulated events are reweighted to correct for observed differences between data and simulation in the number of pileup events, as shown in Fig. 4.3, trigger efficiency, and lepton reconstruction and identification efficiencies [18].

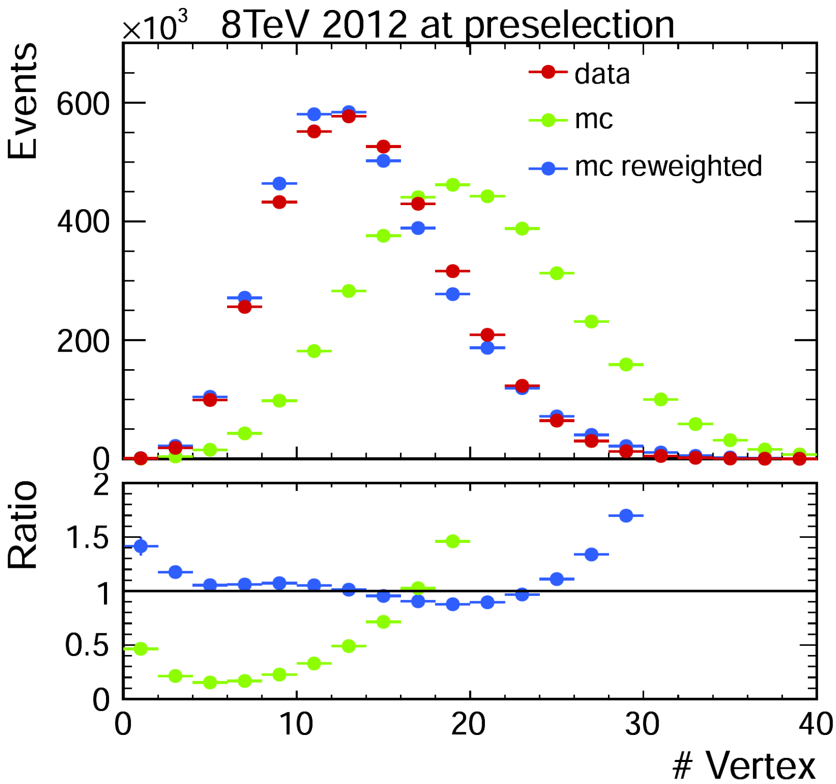


Figure 4.3.: Distribution of the number of vertices in data and in simulation, before and after applying the pile-up reweighting.

For the comparison of the measured unfolded spectrum with the theoretical predictions, two additional MC generators are used for simulating the SM Higgs boson production in the ggH process: HRes 2.3 [36, 7] and the second version of the Powheg generator (Powheg V2) [37]. HRes is a partonic level MC generator that computes the SM Higgs boson cross section at NNLO accuracy in pQCD and performs the NNLL resummation of soft-gluon effects at small p_{T}. The central predictions of HRes are obtained including the exact top and bottom quark mass contribution to the gluon fusion loop, fixing the renormalization and factorization scale central values at a Higgs boson mass of 125 GeV. The cross section normalization is scaled, to take into account electroweak corrections, by a factor of 1.05 and the effects of threshold resummation by a factor of 1.06 [38, 39]. The upper and lower bounds of the uncertainties are obtained by scaling up and down both the renormalization and the factorization scales by a factor of two. The POWHEG V2 generator is a matrix element based generator that provides a NLO description of the ggH process in association with zero jets, taking into account the finite mass of the bottom and top quarks. The POWHEG prediction is tuned using the POWHEG damping factor *hdump* of 104.17 GeV, in order to match the p_T^H spectrum predicted by HRES in the full phase space. This factor reduces the emission of additional jets in the high p_T regime, and enhances the contribution from the Sudakov form factor in the limit of low p_T. The POWHEG generator is interfaced to the JHUGEN generator version 5.2.5 [40, 41, 42] for the decay of the Higgs boson to a W boson pair and interfaced with PYTHIA 8 [43] for the simulation of parton shower and hadronization effects.

4.3. Analysis Strategy

The analysis presented here is based on that used in the previously published $H \to WW \to 2\ell 2\nu$ measurements by CMS [18], modified to be inclusive in the number of jets. This modification significantly reduces the uncertainties related to the modelling of the number of jets produced in association with the Higgs boson.

4.3.1. Event reconstruction and selections

The electron selection is based on two multivariate discriminants, one specialised in identifying the electron object and the other for isolation. The cut value for each discriminant is optimised to provide a good fake electron rejection and to improve the signal acceptance.

Muons are reconstructed using the standard CMS selection and are required to be identified both in the tracker (*Tracker Muon*) and in the muon chambers (*Global Muon*). Additionally quality criteria on the muon track are required, such as to have at least 10 hits in the tracker (at least one of which in the pixel detector) and to have $\chi^2/ndf < 10$. Muon isolation is based on the Particle-Flow algorithm. An MVA approach is considered, based on the radial distributions of the Particle-Flow candidates inside a cone of radius 0.5 around the muon direction.

The efficiencies for the identification and isolation of the electrons and muons are measured in data and in simulation selecting a pure sample of leptons coming from the $Z \to \ell\ell$ decay, and using a Tag and Probe technique very similar to the one described in Sec. 4.2.1 for the trigger efficiency. In this case, the probe lepton is defined by loose isolation and identification requirements and the efficiency to pass the tight analysis selections is measured performing a simultaneous fit of signal plus background in two categories, corresponding to events in which the probe lepton pass or fail the analysis requirements. For the electrons, the resonant signal contribution in the fit is modelled as the convolution of a Breit-Wigner and a Crystal-Ball function. A polynomial function is added to take into account the tail in the low mass region. For muons the signal is fitted using the sum of two Voigtian functions. For both electrons and muons the background contribution is modelled as a third order Bernstein polynomial function. The efficiencies for data and simulation are extracted as parameters of the fit and are used as scale factors to correct the MC simulation to precisely model the data.

This part would probably end up in the Object Reconstruction chapter

Jets in this analysis are reconstructed by combining the energy measured in the calorimeters and tracks from charged particles on basis of the standard CMS particle flow algorithm and using the anti-k_T clustering algorithm with R = 0.5. Events will be classsified into zero jet, one jet and VBF topologies by counting jets within $|\eta| < 4.7$ and for $p_{\mathrm{T}} > 30\,\mathrm{GeV}$.

In addition to the standard CMS PF $E_{\mathrm{T}}^{\mathrm{miss}}$, in this analysis a *projected* $E_{\mathrm{T}}^{\mathrm{miss}}$ variable is also used. The *projected* $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the component of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ transverse to the nearest lepton if the lepton is situated within the azimuthal angular window of $\pm\pi/2$ from the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ direction, or the $E_{\mathrm{T}}^{\mathrm{miss}}$ itself otherwise. Since the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution is degraded by pileup, the minimum of two projected $E_{\mathrm{T}}^{\mathrm{miss}}$ variables is used: one constructed from all identified particles (full projected $E_{\mathrm{T}}^{\mathrm{miss}}$), and another constructed from the charged particles only (track projected $E_{\mathrm{T}}^{\mathrm{miss}}$).

Background events from $\mathrm{t\bar{t}}$ and tW production are rejected applying a soft-muon veto and b-tagging veto. The former selection requires that in the event there are no muons from b-decays passing the following cuts:

- the muon is reconstructed as TrackerMuon;
- the number of hits of the muon in the Silicon Tracker is greater than 10;
- the transverse impact parameter of the muon is less than 0.2 cm;
- if $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ then the muon is required to be non-isolated with $ISO/p_{\mathrm{T}} > 0.1$.

The latter veto rejects events that contain jets tagged as b-jets using two different algorithms for high and low p_{T} jets. For jets with p_{T} between 10 and 30 GeV, the Track-Counting-High-Efficiency (TCHE) algorithm, with a cut at 2.1 on the discriminating variable, is applied. For jets above 30 GeV, a better performing algorithm, Jet-Probability (JP), is used. Jets are identified as b-jets by the JP algorithm if the discriminating variable has a value above 1.4. In the following a b-tagged jet is defined as a jet, within $|\eta| < 2.4$ (b-tagging requires the tracker information), with a value of the discriminating variable above the mentioned thresholds for the two algorithms.

The event selection consists of several steps. The first step is to select WW-like events applying a selection that consists of the following set of cuts:

1. **Lepton preselection**:
   - at least two opposite-charge and opposite-flavour (eμ) isolated leptons reconstructed in the event;
   - $|\eta| < 2.5$ for electrons and $|\eta| < 2.4$ for muons;
   - $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ 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. **$E_{\mathrm{T}}^{\mathrm{miss}}$ preselection**: particle flow $E_{\mathrm{T}}^{\mathrm{miss}}$ is required to be greater than 20 GeV.
4. **projected $E_{\mathrm{T}}^{\mathrm{miss}}$ selection**: minimum projected $E_{\mathrm{T}}^{\mathrm{miss}}$ required to be larger than 20 GeV.
5. **Di-lepton mass cut**: $m_{\ell\ell} > 12\,\mathrm{GeV}$ in order to reject low mass resonances and QCD backgrounds.
6. **Di-lepton p_{T} cut**: $p_{\mathrm{T}}^{\ell\ell} > 30\,\mathrm{GeV}$.
7. **Transverse mass**: $m_{\mathrm{T}} > 60\,\mathrm{GeV}$ to reject Drell-Yan to $\tau\tau$ events.

In addition to the WW-like preselection other cuts are applied in order to reduce the top quark background (both $\mathrm{t\bar{t}}$ and tW), which is one of the main backgrounds in this final state. Two different selections are used depending on the number of jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ in the event. This is done to suppress the top quark background both in the low $p_{\mathrm{T}}^{\mathrm{H}}$ region, where 0-jets events have the largest contribution, and for higher $p_{\mathrm{T}}^{\mathrm{H}}$ values where also larger jet multiplicity events are important. The selection for 0-jets events relies on the soft-muon veto and on a soft jets (with $p_{\mathrm{T}} < 30\,\mathrm{GeV}$) anti b-tagging requirement. The latter requirement exploits the TCHE algorithm to reject soft jets that are likely to come from b quarks hadronization.

For events with a jet multiplicity greater or equal than one, a different selection is applied. In this case we exploit the good b-tagging performances of the JP tagger to reject all the jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ that are likely to come from a b quark hadronization. The analysis selection requires to have no events containing b-tagged jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$.

A cut-flow plot is reported in Fig. 4.4, showing the effect of each selection using signal and background simulations. In the first bin, labelled as “No cut”, no selectio is applied and the bin content corresponds to the total expected number of events with a luminosity of 19.4 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 all the other selections are progressively reported, showing the effect of each cut on the background and signal yields. For each selection the expected signal over background ratio is also shown, which, after the full selection requirements, reach a maximum value of about 3%.

4.3.2. Fiducial phase space

The Higgs boson transverse momentum is measured in a fiducial phase space, whose requirements are chosen in order to minimize the dependence of the measurements on the underlying model of the Higgs boson properties and its production mechanism.

The exact requirements are determined by considering the two following correlated quantities: the reconstruction efficiency for signal events originating from within the fiducial phase space (fiducial signal efficiency ϵ_{fid}), and the ratio of the number of reconstructed signal events that are from outside the fiducial phase space ("out-of-fiducial" signal events) to the number from within the fiducial phase space. The requirement of having a small fraction of out-of-fiducial signal events, while at the same time preserving a high value of the fiducial signal efficiency ϵ_{fid}, leads to a loosening of the requirements on the low-resolution variables, $E_{\text{T}}^{\text{miss}}$ and m_{T}, with respect to the analysis selection.

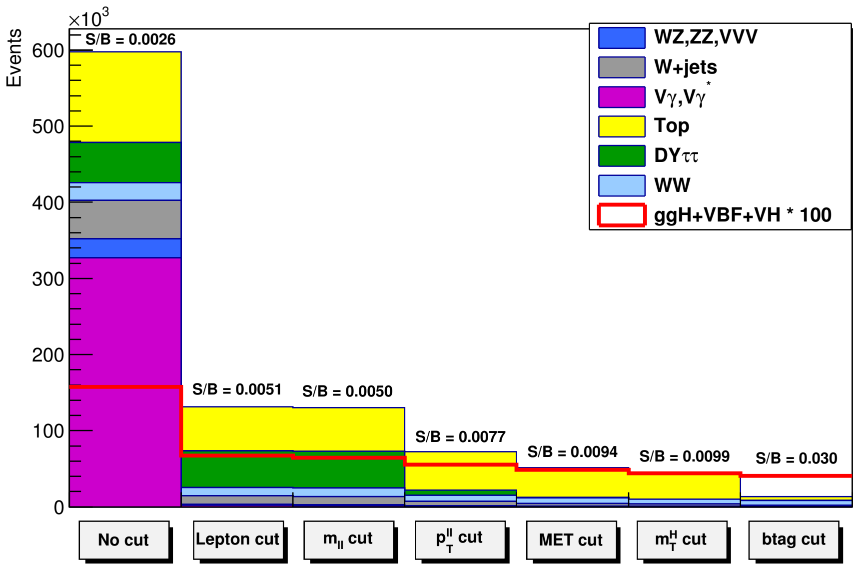


Figure 4.4.: Effect of single selections on MC samples. The signal (red line) is multiplied by 100 and superimposed on stacked backgrounds. In each bin, corresponding to a different selection, is reported the expected number of events in MC at a luminosity of 19.46 fb^{-1}.

The fiducial phase space used for the cross section measurements is defined at the particle level by the requirements given in Table 4.2. The leptons are defined as Born-level leptons, i.e. before the emission of final-state radiation (FSR), and are required not to originate from leptonic τ decays. The effect of including FSR is evaluated to be of the order of 5% in each $p_{\mathrm{T}}^{\mathrm{H}}$ bin. For the VH signal process, the two leptons are required to originate from the $\mathrm{H} \to \mathrm{WW} \to 2\ell 2\nu$ decays in order to avoid including leptons coming from the associated W or Z boson.

Table 4.2.: Summary of requirements used in the definition of the fiducial phase space.

Physics quantity	**Requirement**
Leading lepton p_{T}	$p_{\mathrm{T}} > 20\,\mathrm{GeV}$
Subleading lepton p_{T}	$p_{\mathrm{T}} > 10\,\mathrm{GeV}$
Pseudorapidity of electrons and muons	$\lvert\eta\rvert < 2.5$
Invariant mass of the two charged leptons	$m_{\ell\ell} > 12\,\mathrm{GeV}$
Charged lepton pair p_{T}	$p_{\mathrm{T}}^{\ell\ell} > 30\,\mathrm{GeV}$
Invariant mass of the leptonic system in the transverse plane	$m_{\mathrm{T}}^{\ell\ell\nu\nu} > 50\,\mathrm{GeV}$
$E_{\mathrm{T}}^{\mathrm{miss}}$	$E_{\mathrm{T}}^{\mathrm{miss}} > 0$

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

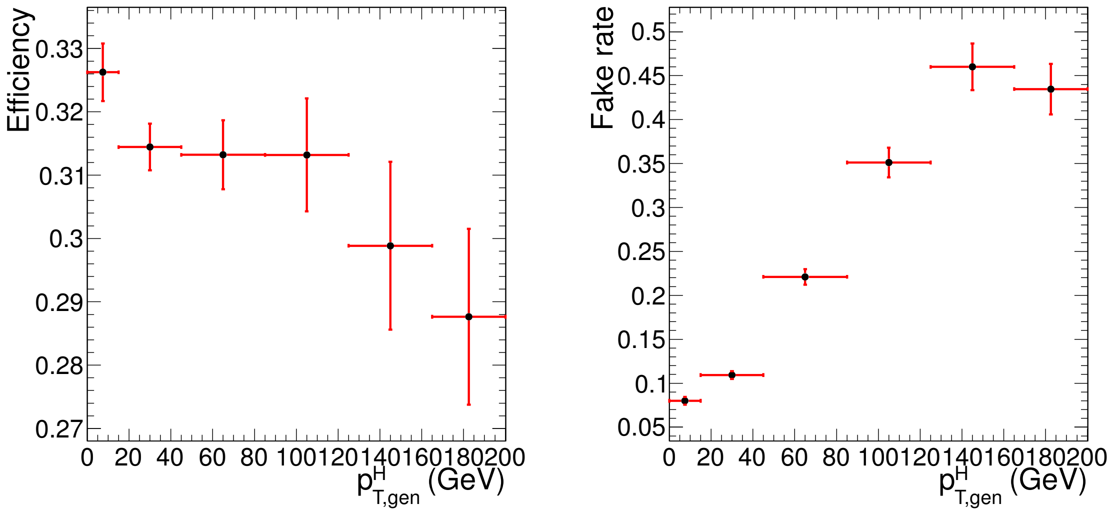
4.3.3. Binning of the $p_{\mathrm{T}}^{\mathrm{H}}$ distribution

Experimentally, the Higgs boson transverse momentum is reconstructed as the vector sum of the lepton momenta in the transverse plane and $E_{\mathrm{T}}^{\mathrm{miss}}$.

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{H}} = \vec{p}_{\mathrm{T}}^{\,\ell\ell} + \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} \tag{4.5}$$

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 $E_{\mathrm{T}}^{\mathrm{miss}}$ entering the transverse momentum measurement. The effect of the limited $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution has two main implications on the analysis strategy. The first one is that the choice of the binning in the $p_{\mathrm{T}}^{\mathrm{H}}$ spectrum needs to take into account the detector resolution. The second implication is that migrations of events across bins are significant and an unfolding procedure needs to be applied to correct for selection efficiencies and bin migration effects.

Given these aspects the criterion that was used to define the $p_{\mathrm{T}}^{\mathrm{H}}$ bin size 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



(a) (b)

Figure 4.5.: Efficiency of the full selection (a) and fake rate (b) as a function of $p_{\mathrm{T}}^{\mathrm{H}}$.

also reconstructed in that bin, $N_i^{\mathrm{GEN|RECO}}$, divided by the number of events reconstructed there, N_i^{RECO}:

$$P_i = \frac{N_i^{\mathrm{GEN|RECO}}}{N_i^{\mathrm{RECO}}} \quad . \tag{4.6}$$

The bin width is chosen in such a way as to make the smallest bins able to ensure a purity of about 60% on a ggH signal sample. Following this prescription we have divided the whole $p_{\mathrm{T}}^{\mathrm{H}}$ range in the following six bins: [0-15 GeV], [15-45 GeV], [45-85 GeV], [85-125 GeV], [125-165 GeV], [165-∞ GeV].

The efficiency of the analysis selection with respect to the fiducial phase space is reported in Fig. 4.5 (a) for each $p_{\mathrm{T}}^{\mathrm{H}}$ bin. The efficiency denominator is the number of events that are inside the fiducial phase space, while the numerator is the number of events that pass both the analysis and the fiducial phase space selections. The fake rate, defined by the ratio of signal events that pass the analysis selection but are not within the fiducial phase space, divided by the total number of events passing both the analysis and the fiducial phase space selections is shown in Fig. 4.5 (b). For both the selection efficiency and the fake rate, all the signal production mechanisms are included. The overall efficiency and fake rate are $\epsilon = 0.362 \pm 0.005$ and $fake\ rate = 0.126 \pm 0.004$ respectively, where only statistical uncertainties are taken into account.

If a 4π acceptance is defined, requiring just that the Higgs decays to WW and then to $2\ell 2\nu$, the efficiency becomes $\epsilon = 0.0396 \pm 0.0003$ and the fake rate is zero.

4.4. Background estimation

4.4.1. Top quark background

In this analysis the top quark background is divided into two different categories depending on the number of jets in the event. In the two categories different selections are applied, especially concerning the b-tagging requirements.

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. The orthogonal phase space is defined inverting the b-veto requirement of the signal region, in such a way to have a control region enriched in top quark events. Then, using this efficiency, the number of events with the associated uncertainty is propagated from the control region to the signal region. The number of surviving top events in the signal region would then be:

$$N_{bveto}^{\mathrm{signal}} = N_{btag}^{\mathrm{control}} \cdot \frac{1 - \epsilon_{\mathrm{top}}}{\epsilon_{\mathrm{top}}} \tag{4.7}$$

where $N_{\mathrm{btag}}^{\mathrm{control}}$ is the number of events in the control region and ϵ_{top} is the efficiency as measured in data.

The methods to estimate the top background contribution in the two jet categories are different and are explained below.

0-jets category

Most of the top background, composed of $\mathrm{t\bar{t}}$ and tW processes, is rejected in the 0-jet bin by the jet veto. The top-tagging efficiency in the zero jet bin, $\epsilon_{\mathrm{tag}}^{0-jet}$, is the probability for a top event to fail one of either the b-tagging veto or the soft muon veto, and is defined as:

$$\epsilon_{\mathrm{tag}} = \frac{N_{\mathrm{tag}}^{\mathrm{control}}}{N^{\mathrm{control}}} \quad , \tag{4.8}$$

where N^{control} is the number of events in the top control phase space defined requiring one b-tagged jet with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, and $N_{\mathrm{tag}}^{\mathrm{control}}$ is the subset of those events that pass either the soft muon tagging or the low-p_{T} b jet tagging. The purity of this control sample, as estimated from simulation, is about 97%. The remaining 3% background contribution is estimated from simulation and subtracted from the numerator and denominator of Eq. (4.9). The efficiency $\epsilon_{\text{top}}^{0-jet}$ can then be estimated using the following formula:

$$\epsilon_{\text{top}}^{0-jet} = f_{\text{t}\bar{\text{t}}} \cdot \epsilon_{2b} + f_{\text{tW}} \cdot (x \cdot \epsilon_{2b} + (1-x) \cdot \epsilon_{\text{tag}}) \quad , \tag{4.9}$$

$$\epsilon_{2b} = 1 - (1 - \epsilon_{\text{tag}})^2 \quad , \tag{4.10}$$

where $f_{\text{t}\bar{\text{t}}}$ and f_{tW} are the $\text{t}\bar{\text{t}}$ and tW fractions respectively, x is the fraction of tW events containing 2 b jets, and ϵ_{2b} is the efficiency for a top event with 0 counted jets, i.e. two soft b jets, to pass the top veto. For the ratio of $\text{t}\bar{\text{t}}$ and tW cross-sections an uncertainty of 17% is assumed. The fraction $f_{\text{t}\bar{\text{t}}}$ is estimated using MC simulation of the $\text{t}\bar{\text{t}}$ and tW processes at NLO accuracy.

Using this procedure a data/simulation scale factor of 0.98 ± 0.17 is found, and is applied to correct the MC simulation in order to match the data.

Category with more than 0 jets

The strategy for the estimation of the top background in events with at least one jet with p_{T} greater than 30 GeV is the following. First of all the efficiency for tagging a b jet is measured both in data and simulation and the values are used to correct the simulation for different b-tagging efficiencies in data and simulation. This evaluation is performed in a control region, called CtrlTP, containing at least two jets, using a Tag&Probe technique. The procedure to extract these scale factors is presented in Sec. 4.4.1. Then a larger statistics control region, CtrlDD, is defined 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. 4.4.1.

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 $\text{t}\bar{\text{t}}$ events.

The CrtlTP control region is defined selecting the events which pass the lepton preselection cuts listed in Sec. 4.3.1, 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. From 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 a variable that discriminates between true b-jets and other jets in a $t\bar{t}$ sample is chosen. 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:

$$N_{TPP} = N_s \epsilon_s + N_b \epsilon_b \tag{4.11}$$

$$N_{TFP} = N_s(1 - \epsilon_s) + N_b(1 - \epsilon_b) \tag{4.12}$$

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_{b} 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.

A χ^2 simultaneous fit of the *probe* p_{T} spectrum for *tag-pass-probe* and *tag-fail-probe* pairs is performed, deriving the shapes for true b-jets and non-b-jets from the simulation, and extracting N_{s}, N_{b}, ϵ_{s} and ϵ_{b} from the fit. The result of the fit on simulation is shown in Fig. 4.6. The relevant efficiencies are:

$$\epsilon_s^{MC} = 0.7663 \pm 0.0072 \tag{4.13}$$

$$\epsilon_b^{MC} = 0.208 \pm 0.015 \tag{4.14}$$

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 simulation statistic accuracy.

In order to assess the robustness of the fit, 5000 toy MC samples have been generated with a statistics equivalent to the one expected in data and the same fit is performed. All the 5000 fit succeeded, and the pull distributions for ϵ_{s} and ϵ_{b} parameters are shown in Fig. 4.7. The plots show the pull of the efficiencies measured in the fit, where the pull variable for each toy i is defined as:

$$pull(\epsilon_{\mathrm{s(b)}}) = \frac{\epsilon^{\mathrm{true}}_{\mathrm{s(b)}} - \epsilon^{i}_{\mathrm{s(b)}}}{\sigma(\epsilon^{i}_{\mathrm{s(b)}})} \tag{4.15}$$

The pulls are centered on 0 and have σ close to 1, as expected.

An example fit for one of the toys is shown in Fig. 4.8

Before running the fit on data, the shapes used in the fit have been validated. To do so, a purer top enriched phase space has been defined by requiring exactly two jets with *JetBProbability* score higher than 1.5 and no additional b-tagged jets, rejecting also jets with 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. 4.9 and shows good agreement.

Finally the fit has been performed on data, as shown in Fig. 4.10, providing the following efficiencies:

$$\epsilon_s^{Data} = 0.769 \pm 0.022 \tag{4.16}$$

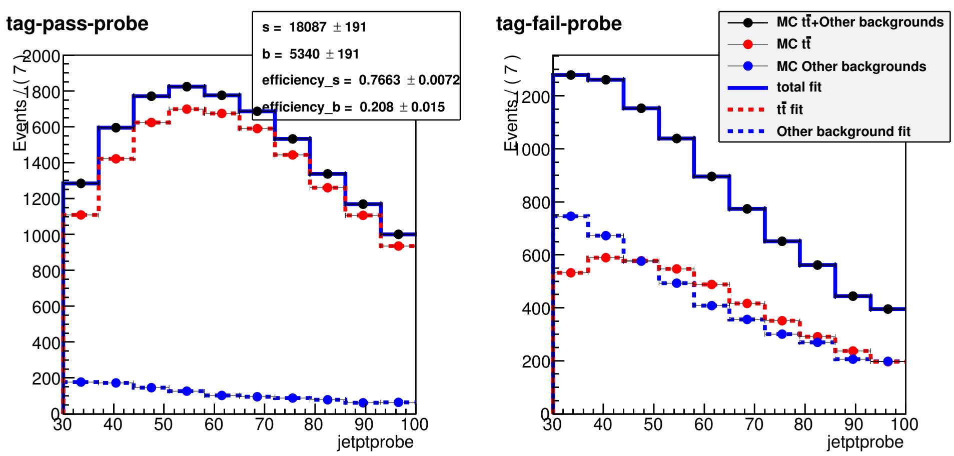


Figure 4.6.: Simultaneous fit of the *tag-pass-probe* and *tag-fail-probe* pairs in the MC.

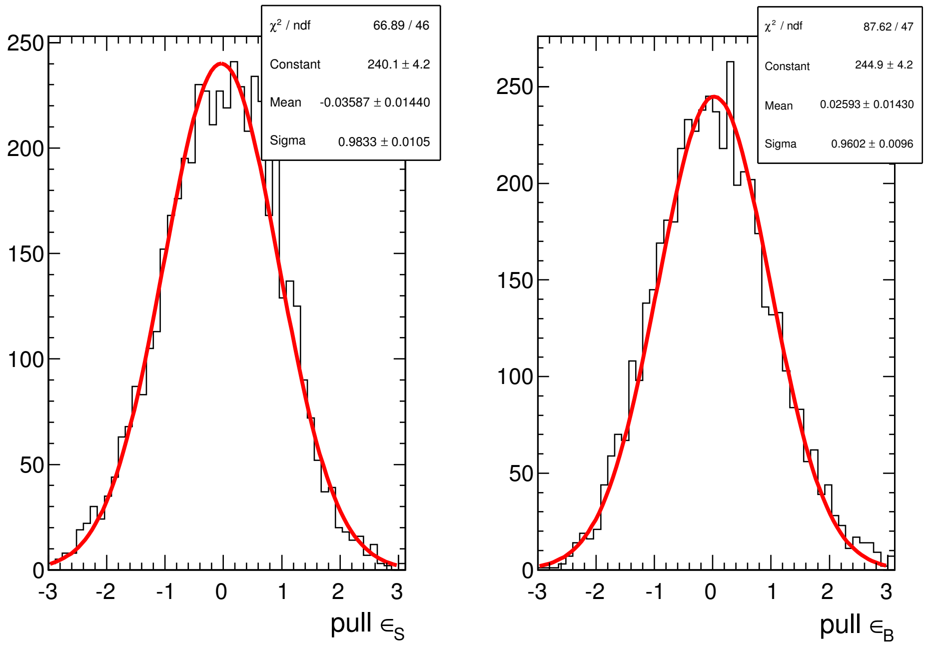


Figure 4.7.: Pulls of the ϵ_s and ϵ_b parameters in 5000 toy MC.

$$\epsilon_b^{Data} = 0.121 \pm 0.054 \tag{4.17}$$

Further studies have been performed to assess the effect of the relative uncertainty on the $\mathrm{t\bar{t}}$ and tW event fractions. The same procedure described above has been applied to different simulation templates obtained varying the $\mathrm{t\bar{t}}$ and tW fractions within theoretical uncertainties, and the effect on the parameters extracted with the fit procedure is found to be well below the fit uncertainties.

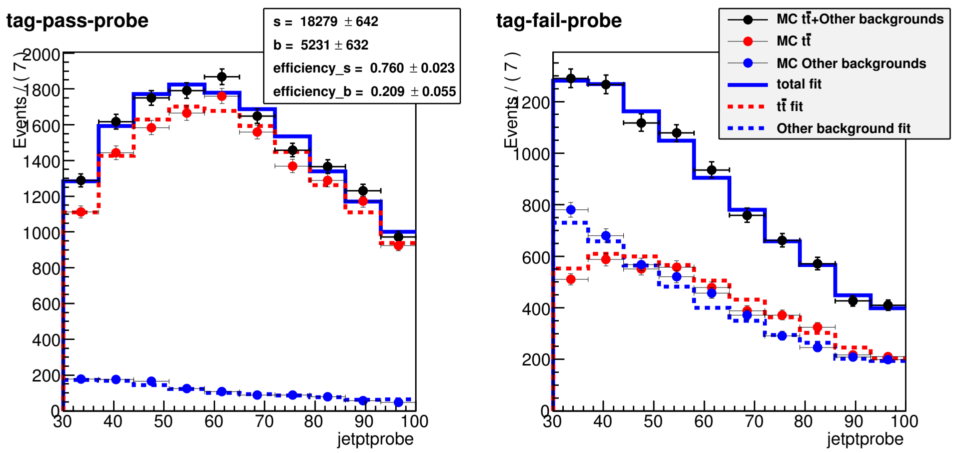


Figure 4.8.: Fit of a toy MC sample.

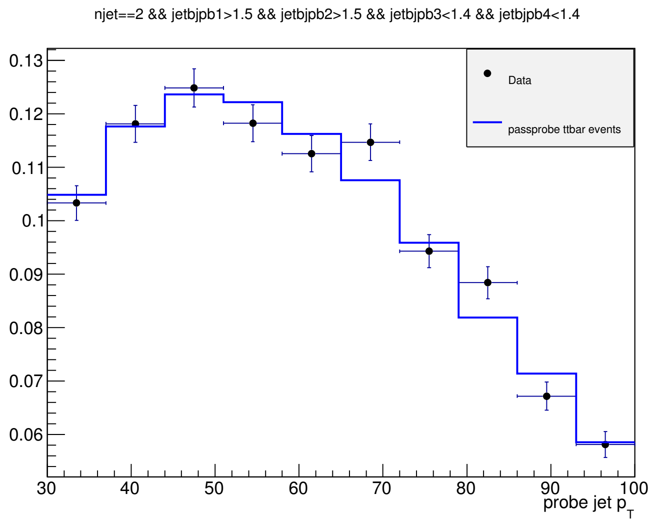


Figure 4.9.: Shape comparison for the *probe* p_{T} spectrum in data and in MC in a very pure $t\bar{t}$ sample.

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 is called CtrlDD. CtrlDD is defined according to the lepton preselection cuts defined in Sec. 4.3.1, and requiring in addition at least one jet with *JetBProbability* score higher than 1.4.

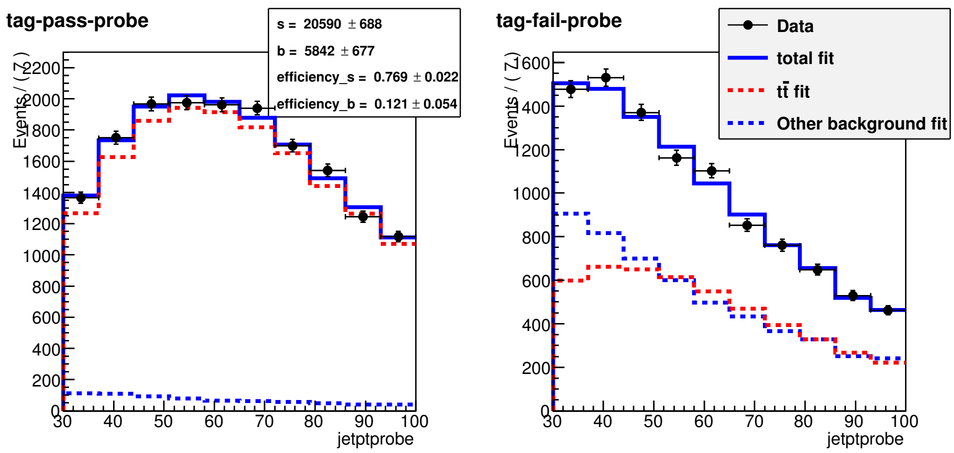


Figure 4.10.: Simultaneous fit of the *tag-pass-probe* and *tag-fail-probe* pairs in data.

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

$$\alpha = \frac{N^{SIG}_{\mathrm{t\bar{t}}\ MC}}{N^{CtrlDD}_{\mathrm{t\bar{t}}\ MC}} \quad . \tag{4.18}$$

The number of events in the CtrlDD region in data is counted, subtracting the expected number of events from non-$\mathrm{t\bar{t}}$ backgrounds, and obtaining $N^{CtrlDD}_{\mathrm{t\bar{t}}\ Data}$. Finally the number of expected $\mathrm{t\bar{t}}$ events in the signal region ($N^{SIG}_{\mathrm{t\bar{t}}\ Data}$) is obtained as:

$$N^{SIG}_{\mathrm{t\bar{t}}\ Data} = \alpha N^{CtrlDD}_{\mathrm{t\bar{t}}\ Data} . \tag{4.19}$$

In evaluating α and its error the b-tagging efficiencies determined in Sec. 4.4.1 are used. For each event an efficiency scale factor and a mistag rate scale factor are derived, depending on whether the event falls in the signal or CtrlDD region.

$$SF_{SIG} = \left(\frac{1-\epsilon_s^{Data}}{1-\epsilon_s^{MC}}\right)^{min(2,n_{b-jets})} \left(\frac{1-\epsilon_b^{Data}}{1-\epsilon_b^{MC}}\right)^{n_{non-b-jets}} \tag{4.20}$$

$$SF_{CtrlDD} = \left(\frac{\epsilon_s^{Data}}{\epsilon_s^{MC}}\right)^{(jet1==b-jet)} \left(\frac{\epsilon_b^{Data}}{\epsilon_b^{MC}}\right)^{(jet1==non-b-jets)} \tag{4.21}$$

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 have been measured on data are close to the one in the simulation, it was 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. 4.20 and Eq. 4.21 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 simulation. This results in an up and a down variation of the scale factors in the signal and CtrlDD regions, that is used to derive an error on α.

p_T^H [GeV]	N_{CTRL}^{DATA}	N_{CTRL}^{TOP}	N_{SIG}^{TOP}	α	$\Delta\alpha$
[0–15]	406.71	358.78	117.83	0.328	0.075
[15–45]	2930.14	2703.44	859.08	0.318	0.071
[45–85]	5481.02	5207.48	1506.05	0.289	0.065
[85–125]	4126.35	4032.56	861.22	0.214	0.052
[125–165]	1612.64	1654.27	304.69	0.184	0.055
[165–∞]	647.50	760.37	201.70	0.265	0.147

Table 4.3.: Data driven scale factors related to the top quark background estimation.

A data driven estimation of the top quark background with the method described above is performed in each of the p_T^H bins independently. The reason to make this estimation in p_T^H bins, rather than inclusively is explained in Fig. 4.11, where the p_T^H distribution is shown in the CtrlDD region normalized to the cross section measured by a specific CMS analysis [44]. As shown in the ratio plot, an overall normalization factor would not be able to accommodate for the variations of the data/simulation ratio from bin to bin.

The α factors for each bin and the number of events in signal, CtrlDD regions in MC as well as in data are listed in Tab. 4.3.

A comparison of the $m_{\ell\ell}$ distribution in the six p_T^H bins used in the analysis in CtrlDD after the data driven correction is shown in Fig. 4.12

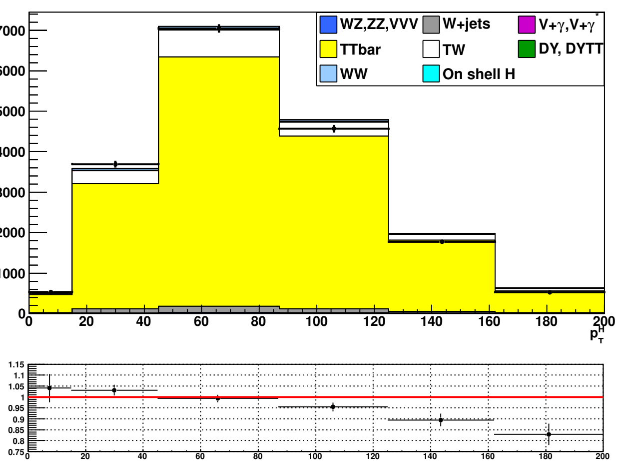
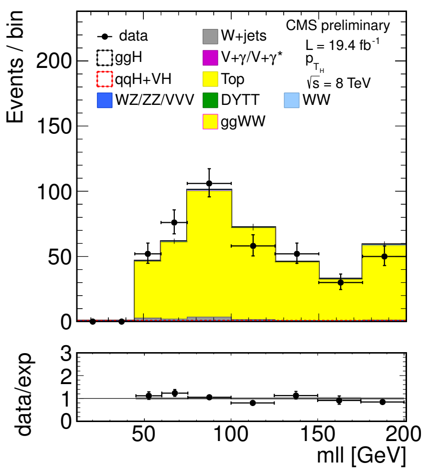
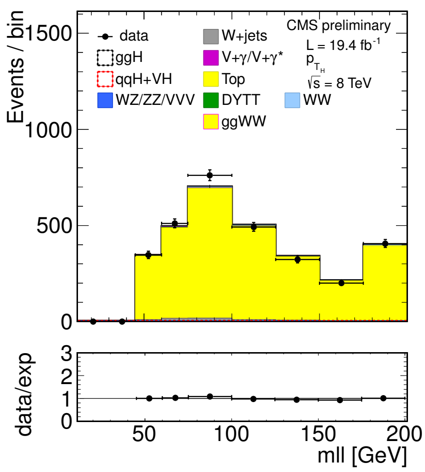


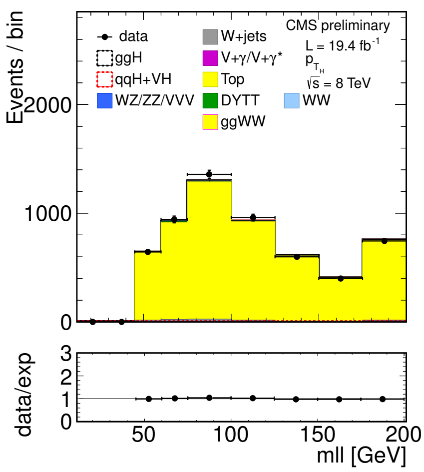
Figure 4.11.: p_T^H distribution in the CtrlDD control region.



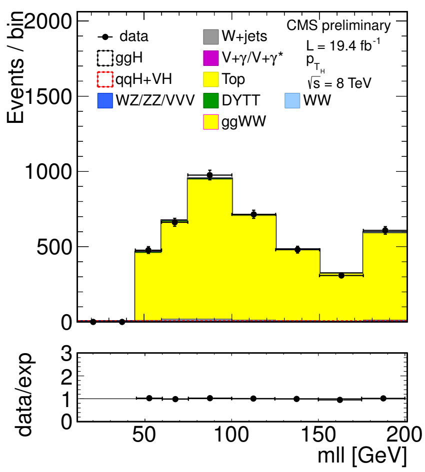
(a) $p_T^H < 15\,\text{GeV}$



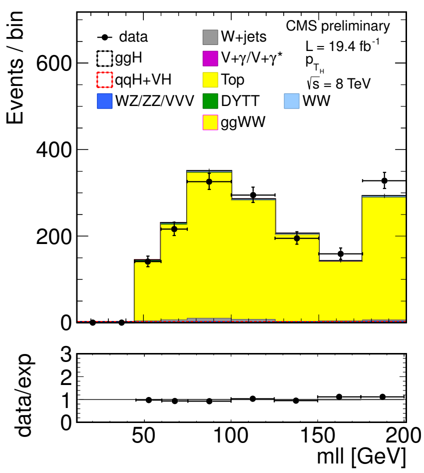
(b) $15\,\text{GeV} < p_T^H < 45\,\text{GeV}$



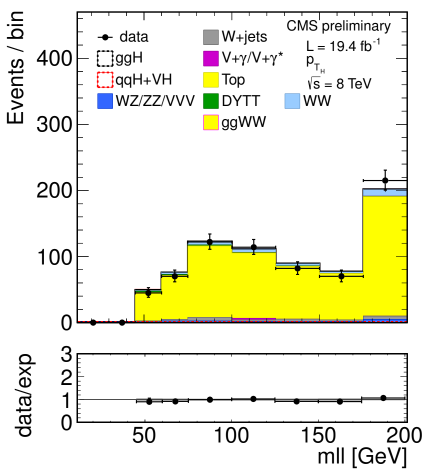
(c) $45\,\text{GeV} < p_T^H < 85\,\text{GeV}$



(d) $85\,\text{GeV} < p_T^H < 125\,\text{GeV}$



(e) $125\,\text{GeV} < p_T^H < 165\,\text{GeV}$



(f) $p_T^H > 165\,\text{GeV}$

Figure 4.12.: $m_{\ell\ell}$ distributions in the CtrlDD region for the different p_T^H bins.

4.4.2. WW background

For what the $qq \to W^+W^-$ background shape is concerned, the prediction from the simulation is used. This background is divided into six different parts, corresponding to the six p_T^H bins defined in the analysis. The normalization of the $qq \to W^+W^-$ background is left free to float in each bin, in such a way to adjust it in order to match the data during the fit procedure. In this way the shape difference between the p_T^{WW} theory prediction and the distribution provided by the simulation, which is obtained with the MADGRAPH generator, is minimized.
In figure 4.13 a comparison is shown between the p_T^{WW} spectra of two different $qq \to W^+W^-$ samples: one obtained with the MADGRAPH generator and the other after applying to the same distribution a reweighting 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_T^H. In order to assess the effect of this discrepancy on the shapes of the variables used for the signal extraction, $m_{\ell\ell}$ and m_T, the shapes have been checked in all p_T^H bins, comparing different MC samples. The MADGRAPH sample used for the nominal shape is compared to the MADGRAPH sample with NLO+NNLL reweighting, a POWHEG sample with NLO accuracy and an aMC@NLO sample. The results of this comparison are shown in figures 4.14 and 4.15. The shape discrepancy among the different models is included as an additional systematic uncertainty.

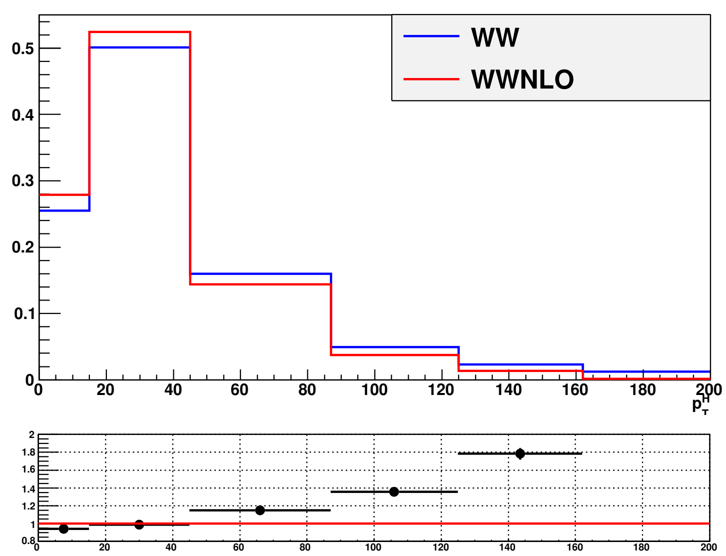
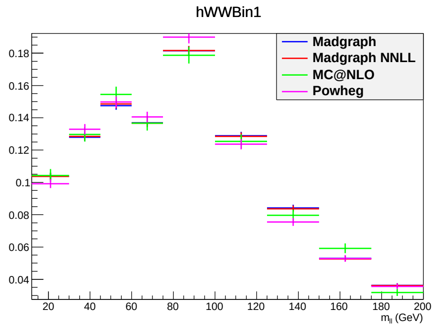
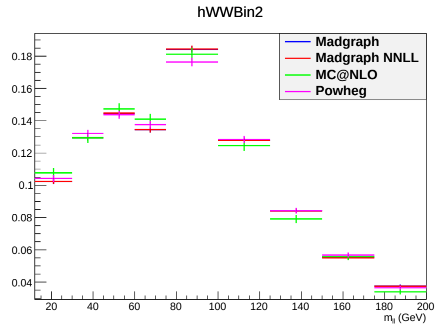


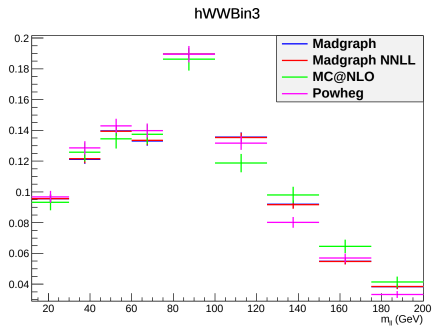
Figure 4.13.: Comparison between the p_T^{WW} distributions obtained with two different MC generators: the blue line corresponds to the MADGRAPH generator and the red line refers to he same sample in which a reweighting has been applied in order to match the theoretical prediction at NLO+NNLL precision.



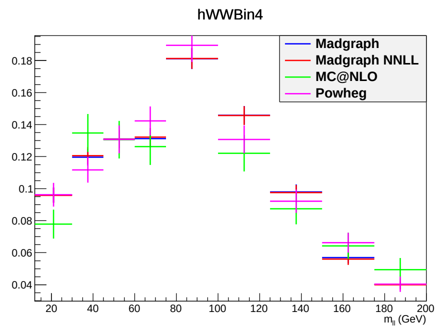
(a) $0 < p_T^H < 15$ GeV



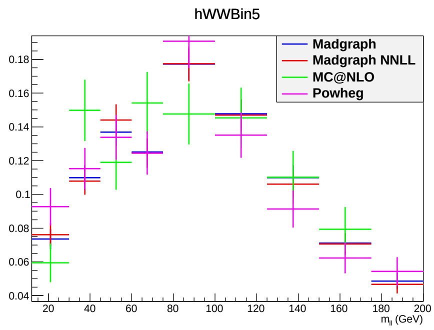
(b) $15 < p_T^H < 45$ GeV



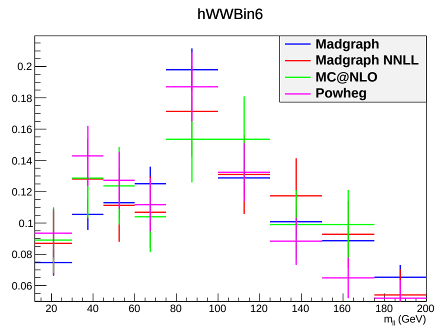
(c) $45 < p_T^H < 85$ GeV



(d) $85 < p_T^H < 125$ GeV



(e) $125 < p_T^H < 165$ GeV

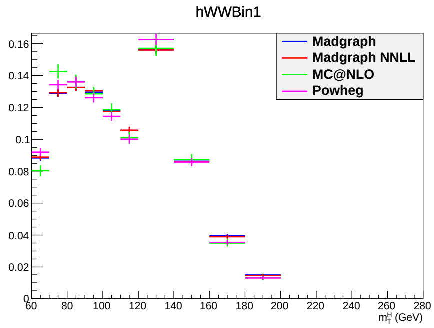


(f) $p_T^H > 165$ GeV

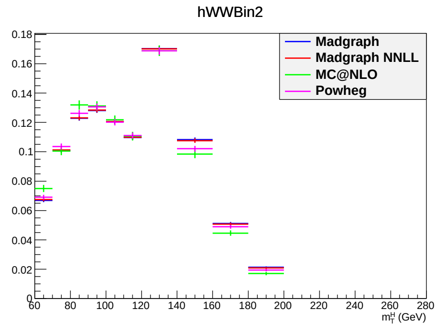
Figure 4.14.: Comparison between the default WW background sample and other theoretical models for the $m_{\ell\ell}$ distributions in every p_T^H bin.

The gluon-induced WW process, i.e. gg→ W^+W^-, has a sub-dominant contribution with respect to the quark-induced process, being the cross section ratio between the two of about 5%. The $m_{\ell\ell}$ and m_T shapes for this background are taken from simulation while the cross section is scaled to the approximate NLO calculation [28, 29].

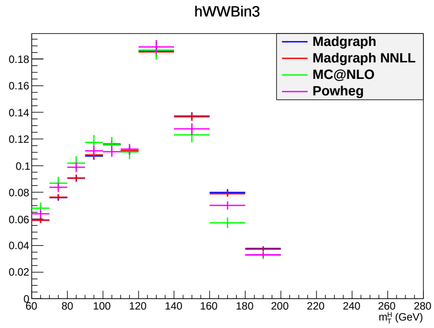
The agreement of the $m_{\ell\ell}$ and m_T shapes between simulation and data for this background was checked in a signal-free control, defined selecting events with values of $m_{\ell\ell}$



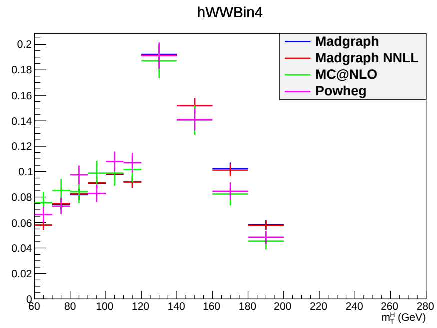
(a) $0 < p_T^H < 15\,\mathrm{GeV}$



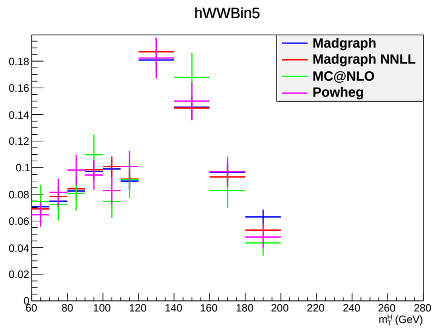
(b) $15 < p_T^H < 45\,\mathrm{GeV}$



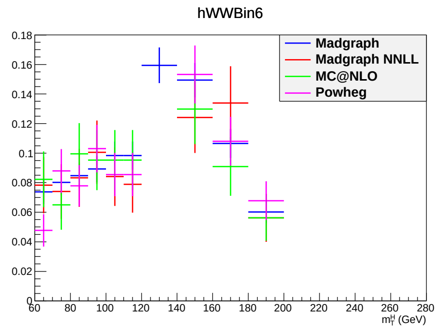
(c) $45 < p_T^H < 85\,\mathrm{GeV}$



(d) $85 < p_T^H < 125\,\mathrm{GeV}$



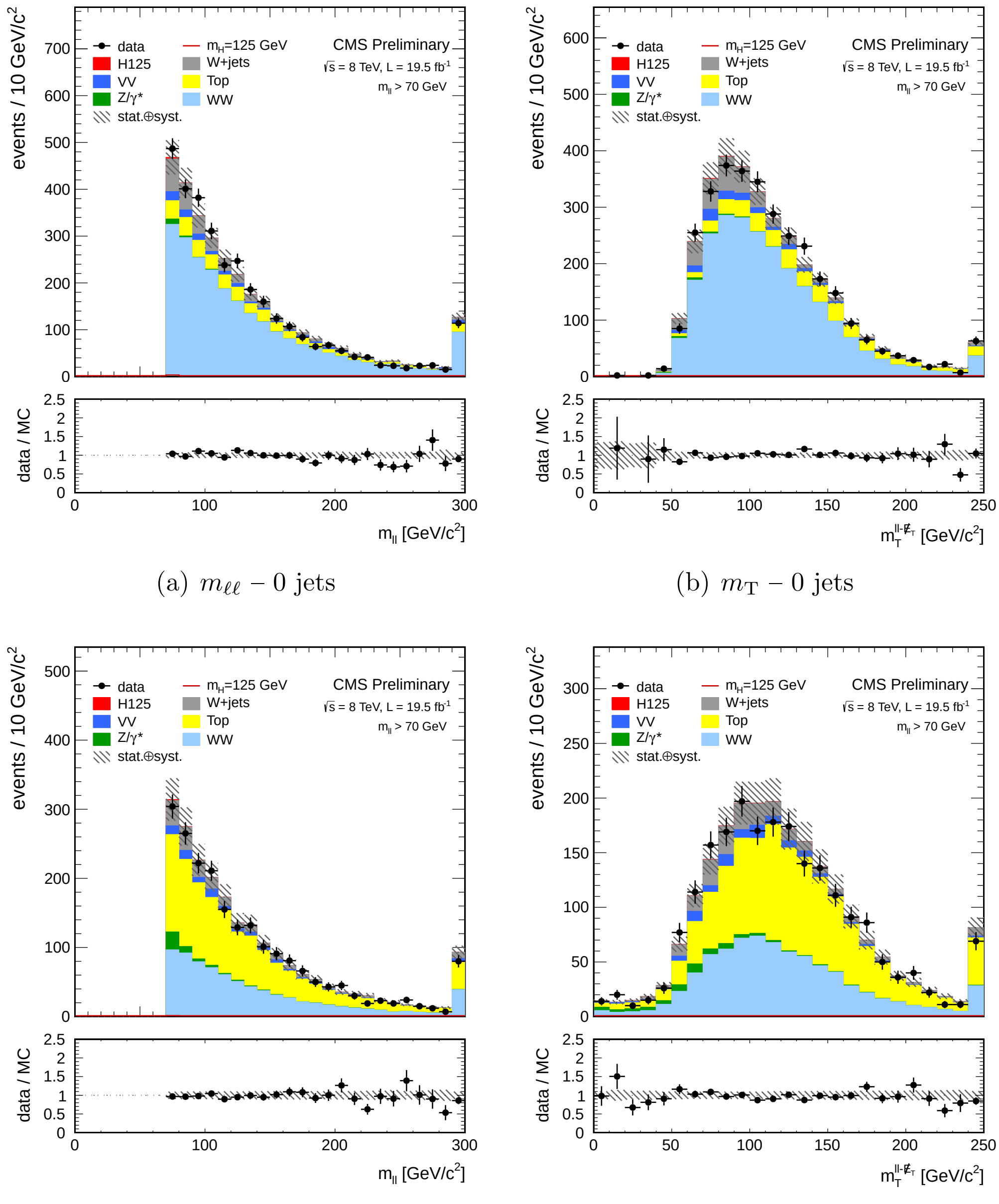
(e) $125 < p_T^H < 165\,\mathrm{GeV}$



(f) $p_T^H > 165\,\mathrm{GeV}$

Figure 4.15.: Comparison between the default WW background sample and other theoretical models for the m_T distributions in every p_T^H bin.

greater than 70 GeV. A comparison of the $m_{\ell\ell}$ and m_T shapes in data and simulation is shown in Fig. 4.16 for events containing zero and one jets, inclusive in p_T^H.



(a) $m_{\ell\ell}$ – 0 jets (b) m_{T} – 0 jets

(c) $m_{\ell\ell}$ – 1 jet (d) m_{T} – 1 jet

Figure 4.16.: Comparison of the $m_{\ell\ell}$ and m_{T} shapes in data and simulation for events with zero and 1 jets, inclusive in $p_{\mathrm{T}}^{\mathrm{H}}$. The events are required to pass the analysis requirements and, in order to define a signal-free control region, to have $m_{\ell\ell} > 70\,\mathrm{GeV}$.

4.4.3. Other backgrounds

W+jets background

Events in which W bosons are produced in association with jets, as well as multi-jet events, constitute a background for this analysis, because one or more jets can be misidentified as leptons. The rate at which jets are misidentified as leptons may be not accurately described in simulation, hence a data driven method is used to estimate this background.

The idea is to estimate the background containing one or two fake leptons selecting events with relaxed lepton quality criteria, i.e. looser with respect to the selections used at the analysis level, and computing the efficiencies for real and fake leptons to pass the tight lepton quality requirements of the analysis. A data-driven approach 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 data set of events dominated by dijet production. To measure the fake rate we count how many fakeable objects pass the full lepton selection of the analysis, parameterized 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:

$$Fake\ Rate = \frac{\#of\ fully\ reconstructed\ leptons}{\#of\ fakeable\ objects} \tag{4.22}$$

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

The definition of the denominator is of large impact in the systematic uncertainties related to this method. For the 2012 data taking period a summary of the selections used for the numerator and the denominator of Eq. (4.22) is shown below for electrons and muons respectively. For electrons the denominator is defined by the following requirements:

- $\sigma_{\mathrm{i}\eta\mathrm{i}\eta} < 0.01(0.03)$ for barrel (endcap);
- $|\Delta\phi_{\mathrm{in}}| < 0.15(0.10)$ for barrel (endcap);
- $|\Delta\eta_{\mathrm{in}}| < 0.007(0.009)$ for barrel (endcap);
- $H/E < 0.12(0.10)$ for barrel (endcap);
- electron conversion rejection;
- $|d_0| < 0.02\,\mathrm{cm}$;

- $\frac{\sum_{\mathrm{trk}} E_{\mathrm{T}}}{p_{\mathrm{T}}^{\mathrm{ele}}} < 0.2$;
- $\frac{\sum_{\mathrm{ECAL}} E_{\mathrm{T}}}{p_{\mathrm{T}}^{\mathrm{ele}}} < 0.2$;
- $\frac{\sum_{\mathrm{HCAL}} E_{\mathrm{T}}}{p_{\mathrm{T}}^{\mathrm{ele}}} < 0.2$.

For muons the selection are loosened with respect to the tight analysis selection requiring that:

- $|d_0| < 0.02\,\mathrm{cm}$;
- MVA isolation output > -0.6.

The dijet enriched data set used for the fake rate measurement, which is selected using single lepton triggers with low p_{T} thresholds, it is not a pure sample containing just fake leptons, but may still contain prompt leptons coming from the W and Z boson decays. To reject muons from the W decay, the events are required to have $E_{\mathrm{T}}^{\mathrm{miss}} < 20\,\mathrm{GeV}$ and a W transverse mass below 20 GeV as well. Muons from the Z decay are instead remove requiring $m_{\mu\mu} > 20\,\mathrm{GeV}$ and $m_{\mu\mu} \notin [76, 106]\,\mathrm{GeV}$. For electrons the Z mass peak veto is enlarged to $m_{\mathrm{ee}} \notin [60, 120]\,\mathrm{GeV}$. Finally both electrons and muons are required to be isolated from the leading jet in the event, i.e. $\Delta\phi(\ell, j) > 1$. The residual prompt lepton contamination from EW processes such as W/Z+jets production, which can bias the fake rate measurement, is estimated using simulation and subtracted from both the numerator and denominator. The contamination from EW processes is different for the numerator and denominator and is particularly important for relatively high lepton p_{T} values.

In addition to the fake rate, also a prompt lepton rate is evaluated, defined as the probability of a prompt lepton passing the loose requirements to also pass the tight analysis selections. The prompt rate is also measured in data, defining a control region enriched in $\mathrm{Z} \to \ell\ell$ events, selecting dilepton events with an invariant mass of the two leptons in the Z peak mass region.

Both the fake and prompt rate are used to reweight the data samples used in the analysis in order to obtain directly from data the contribution of the fake lepton background. The method to apply those rates is explained below in the simple case of just one lepton in the data sample, i.e. data selected by single lepton triggers, but can be straightforwardly generalized to situations with more than one lepton. Suppose that the total number of leptons passing the loose requirements, N_ℓ, is made up of N_p prompt and N_f fake leptons. N_p and N_f cannot be directly measured but one can measure the number of events where no leptons, N_{t0}, or one lepton, N_{t1}, pass the tight analysis requirement. These numbers are related by the following equations:

$$
\begin{aligned}
N_\ell = N_p + N_f &= N_{t0} + N_{t1} \\
N_{t0} &= (1-p)N_p + (1-f)N_f \\
N_{t1} &= pN_p + fN_f
\end{aligned}
\tag{4.23}
$$

where p and f are the prompt and fake rates respectively. Equation (4.23) can be inverted to obtain the number of prompt and fake leptons:

$$
\begin{aligned}
N_p &= \frac{1}{p-f}\left[(1-f)N_{t1} - fN_{t0}\right] \\
N_f &= \frac{1}{p-f}\left[pN_{t0} - (1-p)N_{t1}\right]
\end{aligned}
\tag{4.24}
$$

The number of fake events passing the tight analysis requirement is $N_{\text{fake}} = fN_f$. The fake background contribution is estimated directly from data, applying the kinematics-dependent weights (f and p are estimated in bins of p_{T} and η) defined in Eq.(4.24).

The prompt and fake rate estimations after the removal of the EW contribution are shown in Tables 4.4 and 4.5 separately for electrons and muons.

The region obtained by reversing the opposite sign lepton requirement in the analysis selection is enriched with W+jets events where one of the jets is misidentified as a lepton. The fake rate procedure can be applied to this same-sign control region to perform a closure test of the method. The results of the closure test on same-sign events gives good agreement with the expectations.

The systematic uncertainty on the prompt and fake rate estimation is evaluated by varying the jet thresholds in the dijet control sample, and an uncertainty on the background normalization is added according to the agreement with data in the same-sign control region. The systematic uncertainty amounts to about 36% of the fake background yield.

Drell-Yan to $\tau\tau$ background

The low $E_{\text{T}}^{\text{miss}}$ threshold in the eμ final state requires the consideration of the contribution from $Z/\gamma^* \rightarrow \tau^+\tau^-$, that estimated from data. This is accomplished by selecting $Z/\gamma^* \rightarrow \mu^+\mu^-$events in data and replacing both muons with a simulated $\tau \rightarrow \ell\nu_\tau\bar{\nu_\ell}$ decay [18], thus obtaining a “hybrid” event. The Z boson four-momentum is reconstructed in data from the four-momenta of the daughter muons. Then a simulation step allows the replacement

Table 4.4.: Measured prompt rate for electrons and muons in bins of η, p_{T}. Only the statistical uncertainties are shown.

Electron prompt rate			
p_{T} range [GeV]	$0 < \eta \leq 1.4442$	$1.4442 < \eta \leq 1.566$	$1.566 < \eta$
$10 < p_{\mathrm{T}} \leq 15$	0.5738 ± 0.0045	0.5366 ± 0.0204	0.2947 ± 0.0047
$15 < p_{\mathrm{T}} \leq 20$	0.7091 ± 0.0020	0.5484 ± 0.0185	0.4477 ± 0.0034
$20 < p_{\mathrm{T}} \leq 25$	0.7175 ± 0.0013	0.6297 ± 0.0067	0.6200 ± 0.0001
$25 < p_{\mathrm{T}} \leq 50$	0.9219 ± 0.0002	0.8404 ± 0.0007	0.8509 ± 0.0001
$p_{\mathrm{T}} > 50$	0.9693 ± 0.0002	0.9398 ± 0.0021	0.9385 ± 0.0005

Muon prompt rate		
p_{T} range [GeV]	$0 < \eta \leq 1.5$	$1.5 < \eta \leq 2.5$
$10 < p_{\mathrm{T}} \leq 15$	0.7119 ± 0.0003	0.7582 ± 0.0006
$15 < p_{\mathrm{T}} \leq 20$	0.8049 ± 0.0018	0.8495 ± 0.0001
$20 < p_{\mathrm{T}} \leq 25$	0.9027 ± 0.0008	0.8948 ± 0.0012
$25 < p_{\mathrm{T}} \leq 50$	0.9741 ± 0.0001	0.9627 ± 0.0002
$p_{\mathrm{T}} > 50$	0.9900 ± 0.0001	0.9875 ± 0.0003

of the muon objects with τ leptons, in such a way to preserve the Z boson momentum direction is preserved in its rest frame. The $Z/\gamma^* \rightarrow \tau^+\tau^-$decay is simulated with the TAUOLA package [45] to correctly describe the τ-polarization effects.

After replacing muons from $Z/\gamma^* \rightarrow \mu^+\mu^-$decays with simulated τ decays, the set of pseudo-$Z/\gamma^* \rightarrow \tau^+\tau^-$events undergoes the reconstruction step. Good agreement in kinematic distributions for this sample and a MC based $Z/\gamma^* \rightarrow \tau^+\tau^-$sample is found. The global normalization of pseudo-$Z/\gamma^* \rightarrow \tau^+\tau^-$events is checked in the low m_{T} spectrum where a rather pure $Z/\gamma^* \rightarrow \tau^+\tau^-$sample is expected.

This method allows to avoid the simulation of very large MC samples that would be needed for an accurate description of this process.

ZZ, WZ and Wγ 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.

Table 4.5.: Measured electrons and muons fake rates in bins of η and p_{T}, after the EWK correction. Only statistical uncertainties are shown.

electron fake rate				
p_{T} range [GeV]	$0 < \eta \leq 1$	$1 < \eta \leq 1.479$	$1.479 < \eta \leq 2$	$2 < \eta \leq 2.5$
$10 < p_{\mathrm{T}} \leq 15$	0.045 ± 0.005	0.033 ± 0.004	0.008 ± 0.002	0.021 ± 0.005
$15 < p_{\mathrm{T}} \leq 20$	0.044 ± 0.003	0.049 ± 0.003	0.017 ± 0.001	0.017 ± 0.002
$20 < p_{\mathrm{T}} \leq 25$	0.041 ± 0.002	0.064 ± 0.003	0.025 ± 0.002	0.025 ± 0.002
$25 < p_{\mathrm{T}} \leq 30$	0.059 ± 0.003	0.101 ± 0.005	0.041 ± 0.003	0.043 ± 0.003
$30 < p_{\mathrm{T}} \leq 35$	0.084 ± 0.006	0.111 ± 0.009	0.058 ± 0.006	0.066 ± 0.005
muon fake rate				
p_{T} range [GeV]	$0 < \eta \leq 1$	$1 < \eta \leq 1.479$	$1.479 < \eta \leq 2$	$2 < \eta \leq 2.5$
$10 < p_{\mathrm{T}} \leq 15$	0.131 ± 0.002	0.154 ± 0.004	0.194 ± 0.005	0.241 ± 0.009
$15 < p_{\mathrm{T}} \leq 20$	0.143 ± 0.007	0.191 ± 0.012	0.235 ± 0.016	0.308 ± 0.027
$20 < p_{\mathrm{T}} \leq 25$	0.198 ± 0.005	0.239 ± 0.009	0.221 ± 0.011	0.271 ± 0.021
$25 < p_{\mathrm{T}} \leq 30$	0.182 ± 0.011	0.228 ± 0.018	0.195 ± 0.022	0.287 ± 0.045
$30 < p_{\mathrm{T}} \leq 35$	0.170 ± 0.021	0.244 ± 0.036	0.195 ± 0.041	0.289 ± 0.111

The $\mathrm{W}\gamma^{(*)}$ background, where the photon decays to an electron-positron pair, is expected to be very small, thanks to the stringent photon conversion requirements. This background also includes events where a real photon is produced in association with the W boson. These events constitute a background for this analysis because the photon can interact with the tracker material converting to an electron-positron pair.

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 $\mathrm{W}\gamma^*$ of 1.5 ± 0.5 based on a dedicated measurement of tri-lepton decays, $\mathrm{W}\gamma^* \to e\mu\mu$ and $\mathrm{W}\gamma^* \to \mu\mu\mu$, is applied [18]. The contribution of $\mathrm{W}\gamma^{(*)}$ 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.

4.5. Systematic uncertainties

Systematic uncertainties play an important role in this analysis where no strong mass peak is expected due to the presence of undetected neutrinos in the final state. One of the most important sources of systematic uncertainty is the normalization of the backgrounds that are estimated on data control samples whenever is possible.

A summary of the main sources of systematic uncertainty and the corresponding estimate is reported in Table 4.6. A detailed description of each source of systematic uncertainty is discussed in the following sections.

Table 4.6.: Main sources of systematic uncertainties and their estimate. The first category reports the uncertainties in the normalization of background contributions. The experimental and theoretical uncertainties refer to the effect on signal yields. A range is specified if the uncertainty varies across the p_T^H bins.

Uncertainties in backgrounds contributions	
Source	Uncertainty
$t\bar{t}$, tW	20–50%
W+ jets	40%
WZ, ZZ	4%
$W\gamma^{(*)}$	30%
Effect of the experimental uncertainties on the signal and background yields	
Source	Uncertainty
Integrated luminosity	2.6%
Trigger efficiency	1–2%
Lepton reconstruction and identification	3–4%
Lepton energy scale	2–4%
E_T^{miss} modelling	2%
Jet energy scale	10%
Pileup multiplicity	2%
b mistag modelling	3%
Effect of the theoretical uncertainties on signal yield	
Source	Uncertainty
b jet veto scale factor	1–2%
PDF	1%
WW background shape	1%

4.5.1. Background normalization uncertainties

The signal extraction is performed subtracting the estimated backgrounds to the event counts in data. This uncertainty depends on the background:

- **$t\bar{t}$ and tW backgrounds:** The efficiency on jets b-tagging is estimated using the Tag&Probe technique in data and simulation control regions, as explained in 4.4.1. A per-jet scale factor, which takes into account the possibly different efficiency of the anti b-tagging selection in data and simulation, is computed by means of the efficiency measured with the Tag&Probe method. The Tag&Probe method has been used also to measure the mistag rates in data and simulation, which are the probability to b-tag a jet that is not produced by the hadronization of a b quark. These factors are used to reweigh the Top MC samples as explained in 4.4.1. The uncertainties provided by the Tag&Probe fit are then propagated to the factor α that is used in the top data driven estimation 4.4.1. These uncertainties are embedded in a systematic error that affects the shape of the Top background in each $p_{\mathrm{T}}^{\mathrm{H}}$ bin.

  Provided that the simulated samples include both $t\bar{t}$ and tW processes, a systematic uncertainty related to the tW/$t\bar{t}$ fraction has been included. In fact, a relative variation of the contribution of these two processes could modify the shape of the MC sample, and is thus included as a shape uncertainty affecting the top quark background shape in each $p_{\mathrm{T}}^{\mathrm{H}}$ bin in a correlated way.

- **W+jets background:** It is estimated with data control sample as described in Sec.4.4.3. With 19.4 fb^{-1} at 8 TeV, the uncertainty receives similar contributions from statistics and systematic error (mainly jet composition differences between the fake rate estimation sample and the application sample), the total error being about 40%, dominated by the closure test of the method on a same-sign control region.

- **WZ,ZZ,W$\gamma^{(*)}$ backgrounds:** those backgrounds, which are expected to give a small contribution, are estimated from simulation. Uncertainties on the cross sections reported in [46, 47] are 4% for WZ and 2.5% for ZZ. A 30% uncertainty is assigned to the $W\gamma$ [48] yield and another 30% on $W\gamma^{(*)}$ contribution according to the uncertainty on the normalization study (see Sec. 4.4.3).

4.5.2. Experimental uncertainties

The following experimental systematic sources have been taken into account:

- **Luminosity:** Using the online luminosity monitoring CMS reached an uncertainty on the luminosity of 2.6% at 8 TeV.

- **Trigger efficiency.** The uncertainties for both electrons and muons are at 1-2% level, which is added together to the lepton efficiency uncertainty.

- **Lepton reconstruction and identification efficiency:** The lepton reconstruction and identification efficiencies are measured with the Tag&Probe method in data. To correct for the difference in the lepton identification efficiencies between data and MC, a scale factor is applied to MC. The uncertainties resulting from this procedure on the lepton efficiencies are 4% for electrons and 3% for muons.

- **Muon momentum and electron energy scale:** The momentum scale of leptons have relatively large uncertainties due to different detector effects. For electrons a scale uncertainty of 2% for the barrel, and 4% for the endcaps respectively, is assigned. For muons, a momentum scale uncertainty of 1.5%, independent of its pseudorapidity, is assigned.

- **$E_{\mathrm{T}}^{\mathrm{miss}}$ modeling:** The $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement is affected by the possible mis-measurement of individual particles addressed above, as well as the additional contributions from the pile-up interactions. The effect of the missing transverse momentum resolution on the event selection is studied by applying a Gaussian smearing of 10% on the x- and y-components of the missing transverse momentum. All correlated variables, like the transverse mass, are recalculated.

- **Jet energy scale (JES) uncertainties:** It affects both the jet multiplicity and the jet kinematic variables, such as m_{jj}. We estimate this uncertainty applying variations of the official jet uncertainties on the JES (which depend on η and p_{T} of the jet [49]) and compute the variation of the selection efficiency.

- **b jets mistag modeling:** A fraction of signal events is rejected because erroneously identified as b jets by the b-tagging algorithms. The mistag rate, as measured with the Tag&Probe technique described in Sec. 4.4.1, comes with an uncertainty due to different modeling of the b-tagging performance in data and simulation.

- **Pileup multiplicity:** Some of the variables used in the analysis are affected by the average number of pileup interactions. The simulated events have been reweighted according the instantaneous luminosity measured on data. The error in the average number of pileup interactions measured in data and the simulation of the modeling and physics aspects of the pileup simulation gives an uncertainty of 5% on the distribution used in the reweighting procedure. This uncertainty is propagated through all the analysis, and the estimated uncertainty on the efficiency is 2%.

4.5.3. Theoretical uncertainties

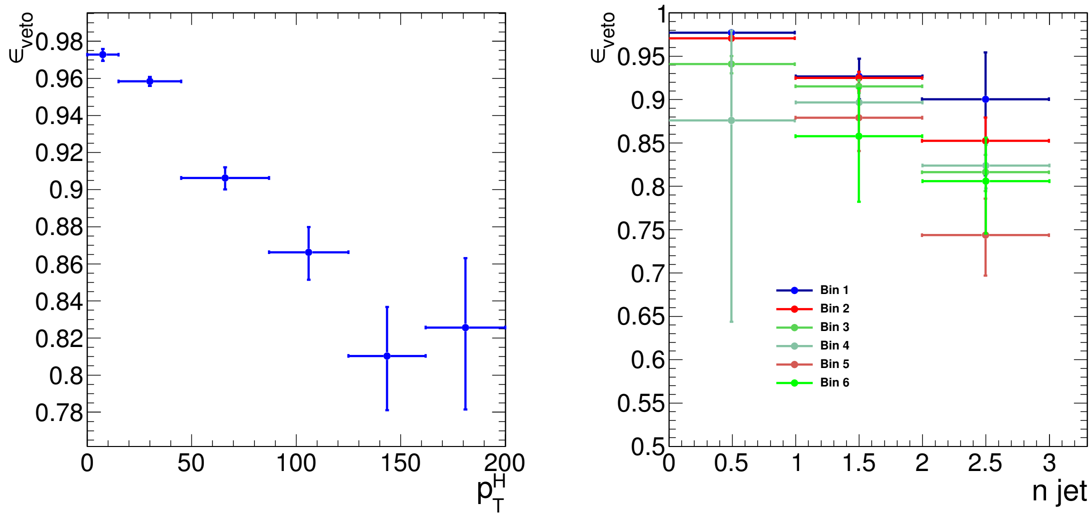
- **QCD scale uncertainties:** The uncertainties on the total cross sections due to the choice of the renormalization and factorization scale are assigned to MC-driven backgrounds. For the signal processes these uncertainties are separated in two categories: those affecting the selection efficiency and those affecting the jet bin fractions. The effect of renormalization and factorization scale on the selection efficiency is of the order of 2% for all processes. Although this analysis is inclusive in number of jets, the effect of the QCD scale variation on the jet bin migrations has to be taken into account because of the b-tagging veto efficiency. The efficiency of this selection depends on jet multiplicity and the effect of the QCD scale variation has been evaluated using the Stewart-Tackman method, as explained in 4.5.3.

- **PDFs uncertainties:** The utilization of different PDF sets can affect both the normalization and the shapes of the signal contributions. The uncertainty related due to the variations in the choice of PDFs is considered following the PDF4LHC [50, 51] prescription, using CT10, NNPDF2.1 [52] and MSTW2008 [53] PDF sets.
- **WW:** Due to the fact that the WW shape is entirely taken from simulation, the analysis is strongly relying on theoretical models and can thus be strongly affected by their uncertainties. Especially higher order QCD radiative effects have an influence on the generated WW shape. To study this impact, the shapes of the distributions produced with the MadGraph generator (which is the generator for the MC simulation used in the analysis) are compared to the ones produced with mc@nlo. The comparison is performed separately in each bin of $p_{\mathrm{T}}^{\mathrm{H}}$ and the uncertainty includes shape differences originating from the renormalization and factorization scale choice. A comparison of the $m_{\ell\ell}$ and m_{T} shapes for the WW background using different MC generators is reported in section 4.4.2.

Jet multiplicity uncertainty

The jet bin uncertainty on the ggH production mode has been evaluated using the Stewart-Tackman method, following the recipe proposed in Refs. [54, 17]. Three independent nuisance parameters have to be associated with the inclusive ggH production cross sections $\sigma_{\geq 0}$, $\sigma_{\geq 1}$ and $\sigma_{\geq 2}$, which corresponds to the cross sections with ≥ 0 jets, ≥ 1 jet and ≥ 2 jets respectively. According to the agreement on the treatment of uncertainties in the combination of ATLAS and CMS results [55], these nuisance parameters are labelled as *QCDscale_ggH*, *QCDscale_ggH1in* and *QCDscale_ggH2in*. However, in case the analysis is split in exclusive jet multiplicity bins, the jet bin uncertainties can be evaluated taking into account the correct correlations among the three nuisances following the Stewart-Tackman prescription. Even though this analysis is inclusive in number of jets, the jet binning uncertainties must be included due to the presence of the b-jet veto, that introduces a dependency of the selection efficiency on the number of jets in the event. The veto efficiency has been evaluated in all the $p_{\mathrm{T}}^{\mathrm{H}}$ bins defined in the analysis and as a function of jets multiplicity. The results are shown in Figs. 4.17(a) and 4.17(b). The drop of the veto efficiency at high values of $p_{\mathrm{T}}^{\mathrm{H}}$ is due to the correlation with jets multiplicity.

The first step of this procedure is to take the inclusive ggH cross section, σ_{ggH}, and to convert the relative QCD up/down scale uncertainties, ϵ_+ and ϵ_-, to a log-normal uncertainty, i.e. $\kappa = \sqrt{\exp(\epsilon_+) \cdot \exp(\epsilon_-)}$. The exclusive cross sections, σ_0, σ_1 and σ_2, can



(a) (b)

Figure 4.17.: (a) Efficiency of the b-tagging veto in different bins of p_T^H. (b) Efficiency of the b-tagging veto in different bins of p_T^H, as a function of number of jets.

be calculated starting from σ_{ggH} and using the selection efficiencies for the three jet bins. For every exclusive cross section the corresponding relative uncertainty is computed varying the renormalization (μ_R) and factorization (μ_F) scales independently of a factor 2 and 1/2, and taking the cross section value corresponding to half of the maximum variation. The inclusive cross sections are then obtained summing the exclusive cross sections and propagating the uncertainties, i.e. $\sigma_{\geq 0} = \sigma_0 + \sigma_1 + \sigma_2$, $\sigma_{\geq 1} = \sigma_1 + \sigma_2$, $\sigma_{\geq 2} = \sigma_2$.

The three nuisance parameters, including all the proper correlations among the jet bins, are defined according to Table 4.7, where the f_n constants represent the exclusive theoretical n jet bin fractions, i.e. $f_0 = \sigma_0/\sigma_{\geq 0}$, $f_1 = \sigma_1/\sigma_{\geq 0}$, $f_2 = \sigma_2/\sigma_{\geq 0}$.

The nuisance parameters reported in table 4.7 have then been calculated for each p_T^H bin embedding the b-jet veto efficiency and using the following formulas:

$$QCDscale\_ggH = \frac{\Delta^0_{\geq 0} \cdot f_0 \cdot \varepsilon_0 + \Delta^0_{\geq 1} \cdot f_1 \cdot \varepsilon_1}{\Delta^0_{\geq 0} \cdot f_0 \cdot \varepsilon_0 + \Delta^0_{\geq 1} \cdot f_1 \cdot \varepsilon_0} \quad , \tag{4.25}$$

$$QCDscale\_ggH1in = \frac{\Delta^1_{\geq 1} \cdot f_1 \cdot \varepsilon_1 + \Delta^1_{\geq 2} \cdot f_2 \cdot \varepsilon_2}{\Delta^1_{\geq 1} \cdot f_1 \cdot \varepsilon_1 + \Delta^1_{\geq 2} \cdot f_2 \cdot \varepsilon_1} \quad , \tag{4.26}$$

$$QCDscale\_ggH2in = 1 \quad , \tag{4.27}$$

Table 4.7.: Numerical calculation for the systematic uncertainties of jet binning.

Nuisance parameter	0-jet bin	1-jet bin	2-jet bin
QCDscale_ggH	$\Delta^{0}_{\geq 0} = (\kappa_{\geq 0})^{\frac{1}{f_0}}$		
QCDscale_ggH1in	$\Delta^{0}_{\geq 1} = (\kappa_{\geq 1})^{-\frac{f_1+f_2}{f_0}}$	$\Delta^{1}_{\geq 1} = (\kappa_{\geq 1})^{\frac{f_1+f_2}{f_1}}$	
QCDscale_ggH2in		$\Delta^{1}_{\geq 2} = (\kappa_{\geq 2})^{-\frac{f_2}{f_1}}$	$\Delta^{2}_{\geq 2} = (\kappa_{\geq 2})$

where ε_0, ε_1 and ε_2 are the selection efficiencies for the three jet categories. These nuisance parameters are expected to be equal to one in case the efficiency is independent on the number of jets, i.e if $\varepsilon_0 = \varepsilon_1 = \varepsilon_2$.
The numerical values obtained following this procedure are reported in Table 4.8 for each $p_{\mathrm{T}}^{\mathrm{H}}$ bin.

Table 4.8.: Values of the jet binning nuisance parameters for different $p_{\mathrm{T}}^{\mathrm{H}}$ bins.

Nuisance parameter	$p_{\mathrm{T}}^{\mathrm{H}}$ bin [GeV]					
	[0-15]	[15-45]	[45-85]	[85-125]	[125-165]	[165-∞]
QCDscale_ggH	0.998	0.993	0.989	1.000	1.000	1.000
QCDscale_ggH1in	0.997	0.993	0.984	0.975	0.946	0.974

4.5.4. Statistics uncertainty of the simulated samples

Due to the large range of weights used to correct the simulated distributions in order to match those in data, the effective size of the MC samples are sometimes smaller than the actual number of events in the sample. The statistical uncertainties of the event yields estimated from MC samples are included as nuisance parameters in the fit and have a small impact on the final result.

4.5.5. Treatment of systematic uncertainties in the shape analysis

One can distinguish between normalization uncertainties, where a systematic effect is changing the normalization of a given process assuming the shape is not affected, and shape uncertainties where the actual change in the shape of the distribution is taken into account. The normalization uncertainties enter the shape analysis as a constant normalization factor, whereas for shape uncertainties the nominal and the $+1\sigma$ and -1σ shapes enter the analysis in form of three histograms with the same normalization.

For the W+jets background, the shape differences for different jet p_T thresholds in the di-jet control sample are considered separately for electron and muon fakes, while the other sources of systematics are taken as normalization uncertainties as in the cut-based analysis.

Effects from experimental uncertainties are studied by applying a scaling and smearing of certain variables of the physics objects, followed by a subsequent recalculation of all the correlated variables. This is done for simulation, to account for possible systematic mis-measurements of the data. All experimental sources from Section 4.5.2 but luminosity are treated both as normalization and shape uncertainties. For background with a data-driven normalization estimation, only the shape uncertainty is considered.

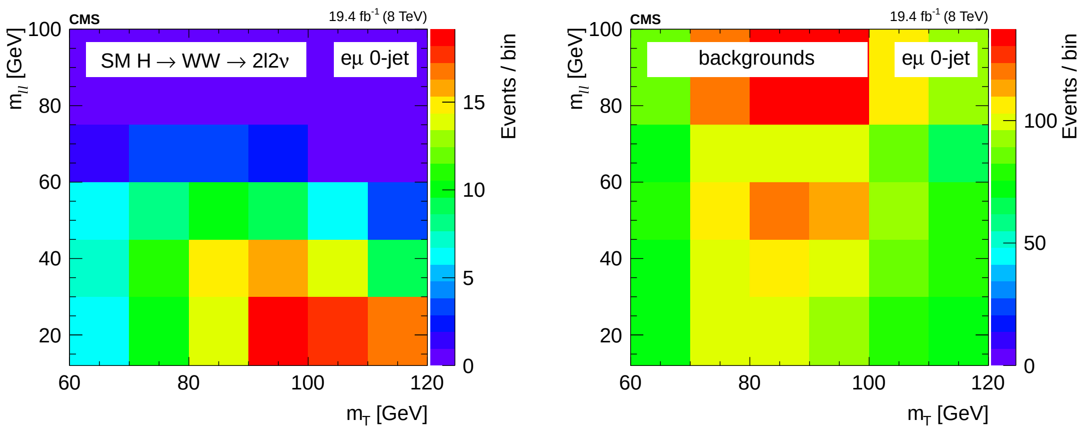
To account for statistical uncertainties, for each distribution going into the shape analysis, the $+1\sigma$ and -1σ shapes were obtained by adding/subtracting the statistical error in each bin and renormalizing it to the nominal distribution. In addition to this procedure a constant normalization uncertainty due to the finite statistics of the MC sample used to extract the shape is assigned.

4.6. Signal extraction

According to the "blinding" policy of the CMS Collaboration, the strategy of the analysis has been scrutinized and approved by a selected committee of internal reviewers before looking at the data in the signal region. This approach prevents the analysts from being biased by the data in the developing phase of the analysis. Below are shown the results after having looked at the data.

4.6.1. Fitting procedure

The signal, including ggH, VBF, and VH production mechanisms, is extracted in each bin of p_T^H by performing a binned maximum likelihood fit simultaneously in all p_T^H bins to a two-dimensional template for signals and backgrounds in the $m_{\ell\ell}$–m_T plane. The variables used for the two-dimensional template are chosen for their power to discriminate signal and background contributions. This is shown in Fig. 4.18, where the two-dimensional MC distributions are shown for the signal and background processes in the 0-jets category.



(a) (b)

Figure 4.18.: Two-dimensional $m_{\ell\ell}$–m_{T} distribution for signal (a) and background (b) processes in the 0-jets category.

Six different signal strength parameters are extracted from the fit, one for each $p_{\mathrm{T}}^{\mathrm{H}}$ bin. The relative contributions of the different Higgs production mechanisms in the signal template are taken to be the same as in the SM. The systematic uncertainty sources are considered as nuisance parameters in the fit.

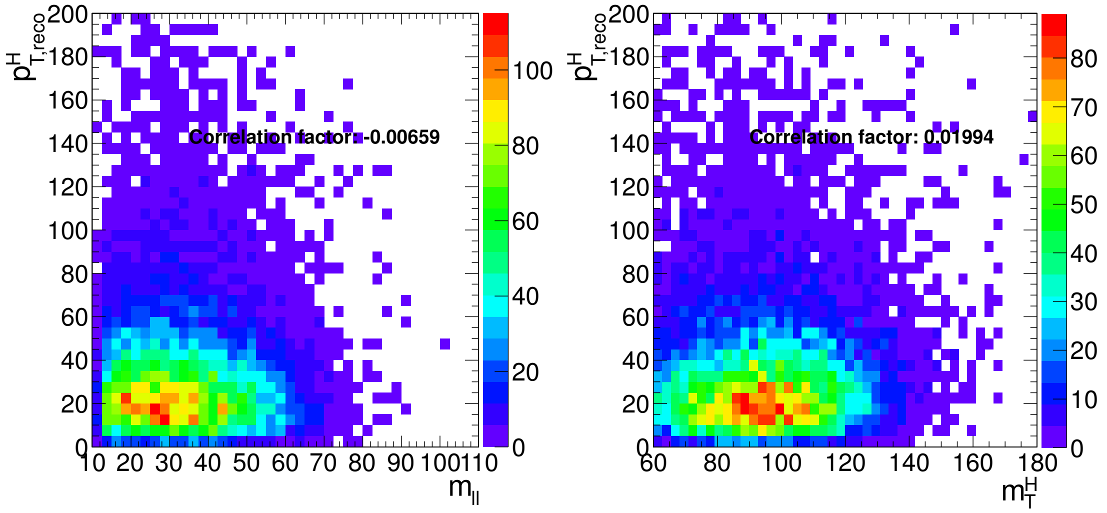
The binning of the $m_{\ell\ell}$ and m_{T} templates is chosen to be:

- $m_{\ell\ell}$: $[12, 30, 45, 60, 75, 100, 125, 150, 175, 200]$
- m_{T}: $[60, 70, 80, 90, 100, 110, 120, 140, 160, 180, 200, 220, 240, 280]$

To avoid a dependence of the results on the variables used for the template fit, $m_{\ell\ell}$ and m_{T} need to be uncorrelated with respect to $p_{\mathrm{T}}^{\mathrm{H}}$. This as been verified and the correlation between the discriminating variables and $p_{\mathrm{T}}^{\mathrm{H}}$ is shown in Fig. 4.19 and Fig. 4.20 for ggH and VBF production modes respectively.

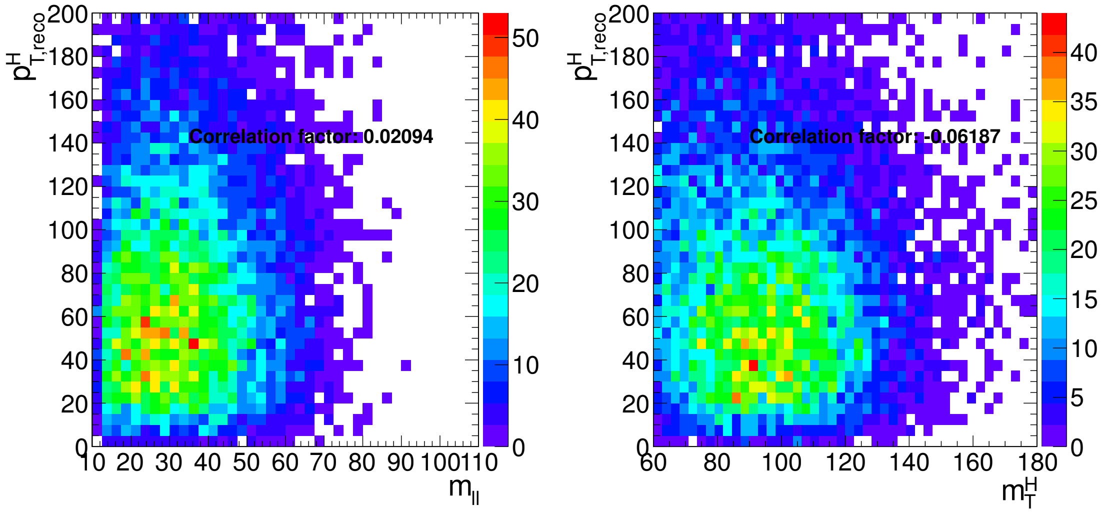
The relative contribution for different production mechanisms in the input signal template is taken to be the same as the SM. The signal strength μ in each bin, i.e. the ratio between the measured cross section and the SM one, $\mu = \sigma/\sigma_{\mathrm{SM}}$, is allowed to float between -10 and +10, thus allowing negative values. This is mainly intended to allow the error bars to float below 0.

Because of detector resolution effects, some of the reconstructed $\mathrm{H} \to \mathrm{WW}$ signal events might originate from outside the fiducial phase space. These out-of-fiducial signal events cannot be precisely handled by the unfolding procedure and must be subtracted from the measured spectrum. The $p_{\mathrm{T}}^{\mathrm{H}}$ distribution of the out-of-fiducial signal events is taken from



(a) (b)

Figure 4.19.: Correlation between p_T^H and $m_{\ell\ell}$ (a) and between p_T^H and m_T (b) after the full selection for the ggH production mode.



(a) (b)

Figure 4.20.: Correlation between p_T^H and $m_{\ell\ell}$ (a) and between p_T^H and m_T (b) after the full selection for the VBF production mode.

simulation, and each bin is multiplied by the corresponding measured signal strength before performing the subtraction.

At the end, the number of events in each bin i of the measured spectrum is:

$$N_i = \mu_i(s_i - f_i) \quad , \tag{4.28}$$

where s_i and f_i are respectively the number of signal and fake events expected from simulation and μ_i is the measured signal strength.

The fit makes use of the binned maximum likelihood approach.The likelihood function, $\mathcal{L}$, restricted to the $p_{\mathrm{T}}^{\mathrm{H}}$ bin j, can be written as: CHECK!!

$$\mathcal{L}(data|\mu_j, \theta) = \prod_{i=0}^{N_{\mathrm{bins}}} \frac{(\mu_j s_i(\theta) + b_i(\theta))^{n_i}}{n_i!} e^{-\mu_j s_i(\theta) - b_i(\theta)} \cdot p(\tilde{\theta}|\theta) \quad , \tag{4.29}$$

where *data* corresponds to the experimental observation and μ_j is the signal strength in the bin j, i.e. the parameter of interest of the fit, which multiplies the signal yield. The index i runs over the bins of the $m_{\ell\ell}$-m_{T} two-dimensional histogram corresponding the $p_{\mathrm{T}}^{\mathrm{H}}$ bin j, s_i and b_i are the expected number of signal and background events respectively in bin i, and n_i is the total number of observed events in bin i. The set of parameters θ represents the full suite of nuisance parameters used to incorporate the systematic uncertainties. Each nuisance parameter is constrained in the fit including the prior distributions functions $p(\tilde{\theta}|\theta)$ in the likelihood, where $\tilde{\theta}$ is the set of default values for the θ parameters. For the major part of the nuisance parameters a log-normal prior distribution is used, with a standard deviation corresponding to the given systematic uncertainty. For some nuisance parameters, as the ones related to the statistical uncertainty coming from the background measurement in data control regions, a Gamma distribution is instead recommended. A log-uniform distribution is used for the uncertainty related to the normalization of background contributions that are left unconstrained in the fit, such as the WW background process. Finally, some of the experimental uncertainties, related to the shape of signal and background processes, are modelled by means of additional histograms as explained in Sec. 4.5.5. The nuisance parameters correlations across different $p_{\mathrm{T}}^{\mathrm{H}}$ bins are taken into account. Moreover the nuisance parameters can also be correlated (or anti-correlated) between signal and different background processes. As an example, the uncertainty related to the integrated luminosity measurement is fully correlated for all the signal and background processes.

Before running the fit on the data, the same procedure has been applied to the so called *Asimov data set*[1], which provides a simple method to estimate the signal sensitivity before looking at the data [56].

[1] In a parallel reality imagined by the science fiction writer I. Asimov, politics was run in a peculiar way: instead of mobilizing millions of people to cast their vote to deliberate on something, an algorithm was used to select an individual "average" person, and then this person was asked to take the decision on that matter.

4.6.2. Signal and background yields

A comparison of data and background prediction is shown in Fig. 4.21, where the $m_{\ell\ell}$ distribution is shown for the six $p_{\mathrm{T}}^{\mathrm{H}}$ bins. Distributions correspond to the m_{T} window of [60, 110] GeV, in order to emphasize the signal contribution [18]. The m_{T} distributions are shown in Fig. 4.22 and correspond to the $m_{\ell\ell}$ window of [12, 75] GeV.

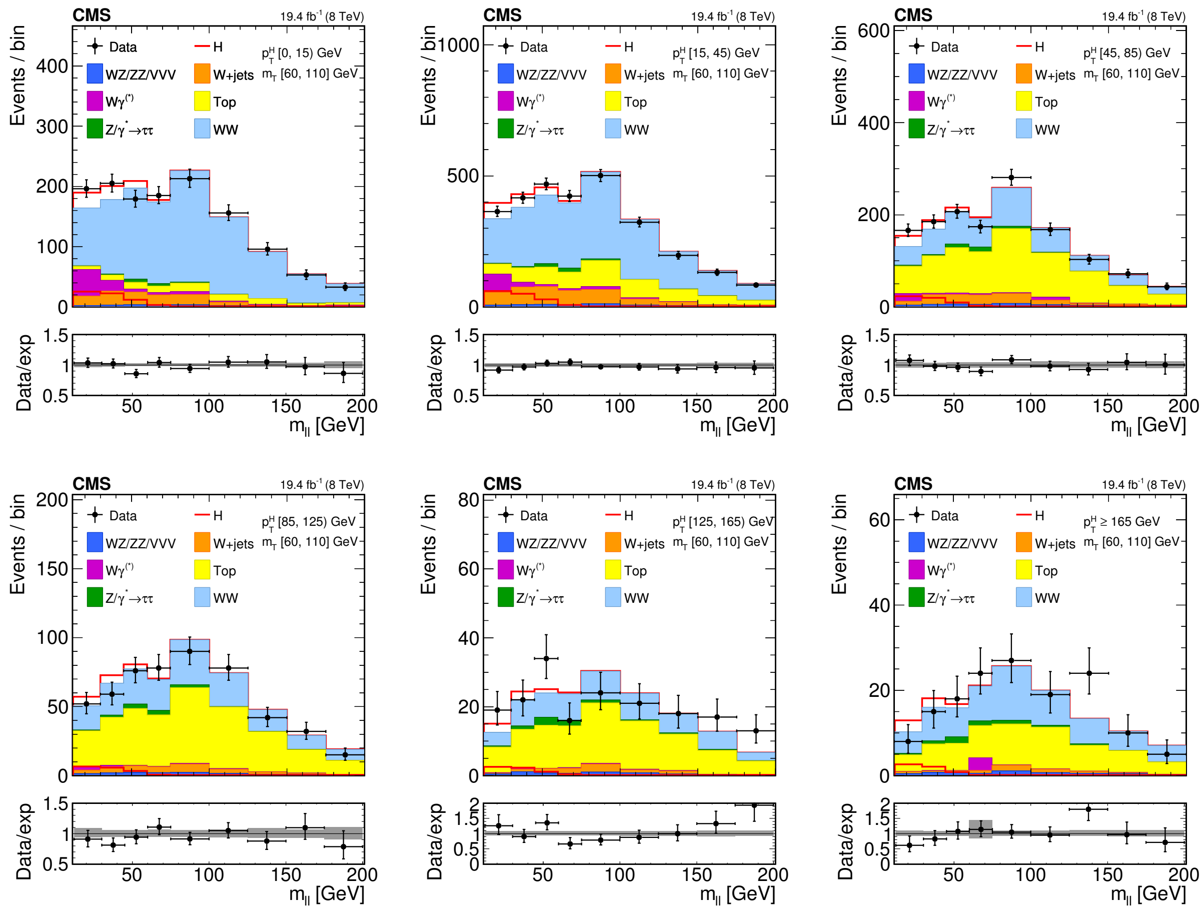


Figure 4.21.: Distributions of the $m_{\ell\ell}$ variable in each of the six $p_{\mathrm{T}}^{\mathrm{H}}$ bins. Background normalizations correspond to the values obtained from the fit. Signal normalization is fixed to the SM expectation. The distributions are shown in an m_{T} window of [60,110] GeV in order to emphasize the Higgs boson (H) signal. The signal contribution is shown both stacked on top of the background and superimposed to it. Ratios of the expected and observed event yields in individual bins are shown in the panels below the plots. The uncertainty band shown in the ratio plot corresponds to the envelope of systematic uncertainties after performing the fit to the data.

The signal and background yields after the analysis selection are reported in Table 4.9.

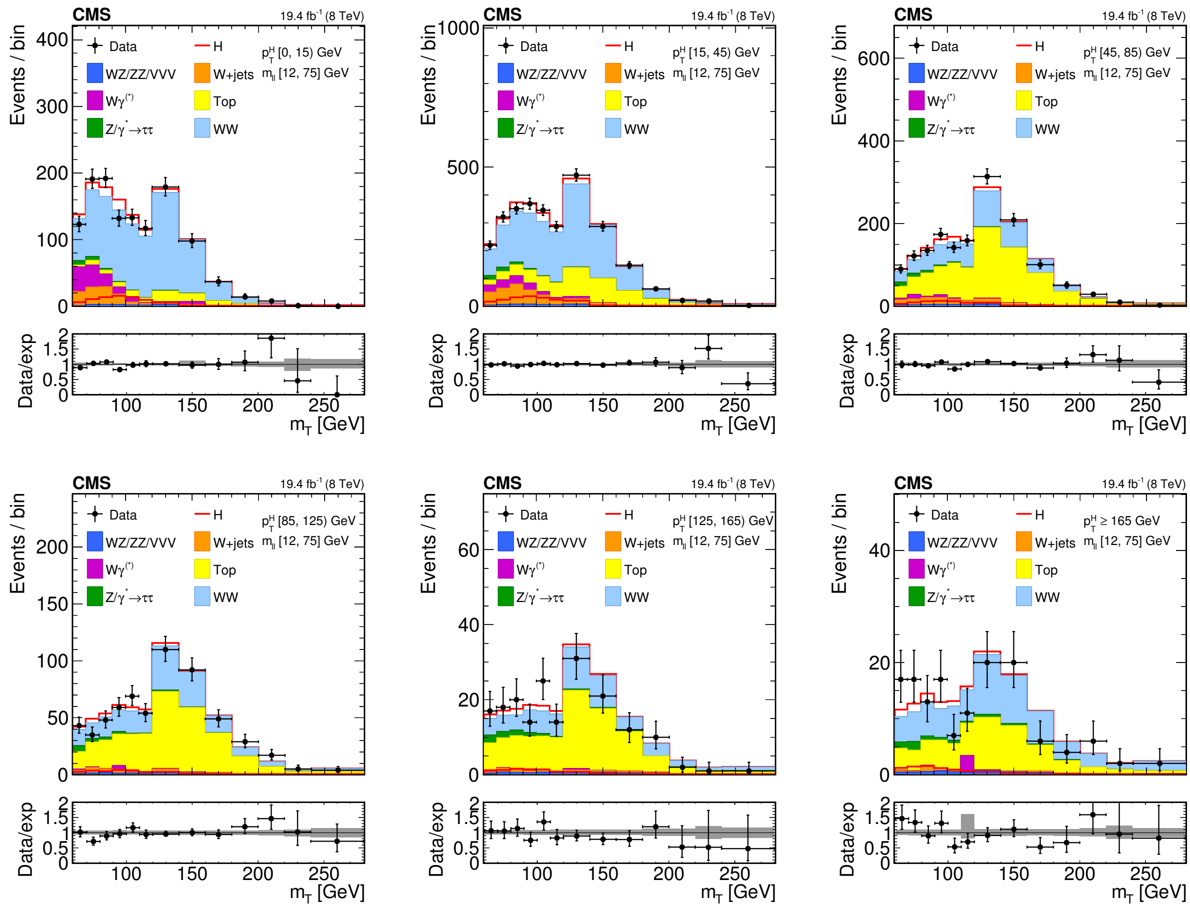


Figure 4.22.: Distributions of the m_{T} variable in each of the six $p_{\mathrm{T}}^{\mathrm{H}}$ bins. Background normalizations correspond to the values obtained from the fit. Signal normalization is fixed to the SM expectation. The distributions are shown in an $m_{\ell\ell}$ window of [12,75] GeV in order to emphasize the Higgs boson (H) signal. The signal contribution is shown both stacked on top of the background and superimposed to it. Ratios of the expected and observed event yields in individual bins are shown in the panels below the plots. The uncertainty band shown in the ratio plot corresponds to the envelope of systematic uncertainties after performing the fit to the data.

Table 4.9.: Signal prediction, background estimates and observed number of events in data are shown in each p_T^H bin for the signal after applying the analysis selection requirements. The total uncertainty on the number of events is reported. For signal processes, the yield related to the ggH are shown, separated with respect to the contribution of the other production mechanisms (XH=VBF+VH). The WW process includes both quark and gluon induced contribution, while the Top process takes into account both $t\bar{t}$ and tW.

$\mathbf{p_T^H}$ [GeV]	**0-15**	**15-45**	**45-85**	**85-125**	**125-165**	**165-∞**
ggH	73 ± 3	175 ± 5	59 ± 3	15 ± 2	5.1 ± 1.5	4.9 ± 1.4
XH=VBF+VH	4 ± 2	15 ± 4	16 ± 4	8 ± 2	3.8 ± 1.1	3.0 ± 0.8
Out-of-fiducial	9.2 ± 0.5	19.9 ± 0.7	11.4 ± 0.6	4.4 ± 0.3	1.6 ± 0.2	2.4 ± 0.2
Data	2182	5305	3042	1263	431	343
Total background	2124 ± 128	5170 ± 321	2947 ± 293	1266 ± 175	420 ± 80	336 ± 74
WW	1616 ± 107	3172 ± 249	865 ± 217	421 ± 120	125 ± 60	161 ± 54
Top	184 ± 38	1199 ± 165	1741 ± 192	735 ± 125	243 ± 51	139 ± 49
W+jets	134 ± 5	455 ± 10	174 ± 6	48 ± 4	14 ± 3	9 ± 3
WZ+ZZ+VVV	34 ± 4	107 ± 10	71 ± 7	29 ± 5	14 ± 3	13 ± 4
$Z/\gamma^* \to \tau^+\tau^-$	23 ± 3	67 ± 5	47 ± 4	22 ± 3	12 ± 2	10 ± 2
W$\gamma^{(*)}$	132 ± 49	170 ± 58	48 ± 30	12 ± 9	3 ± 3	5 ± 10

The spectrum shown in Fig. 4.23 is obtained after having performed the fit and after the subtraction of the out-of-fiducial signal events, but before undergoing the unfolding procedure. The theoretical distribution after the detector simulation and event reconstruction is also shown for comparison.

In order to assess the robustness of the fit, several toy MC samples have been produced, with a statistical accuracy corresponding to the one expected in data. The distribution of the signal strengths extracted in each bin using the toy MC samples and the their pull distributions are shown in Fig. 4.24.

4.7. Unfolding

To facilitate comparisons with theoretical predictions or other experimental results, the signal extracted performing the fit has to be corrected for detector resolution and efficiency effects and for the efficiency of the selection defined in the analysis. An unfolding procedure is used relying on the RooUnfold package [57], which provides the tools to run various unfolding algorithms.

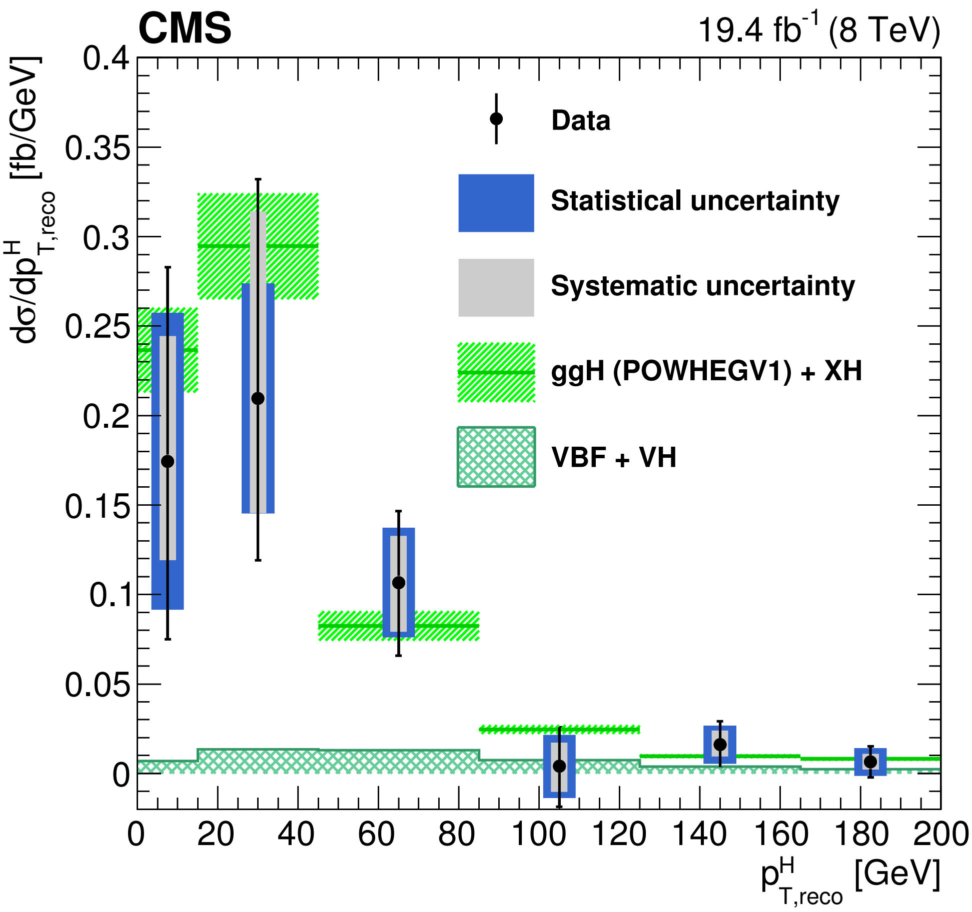
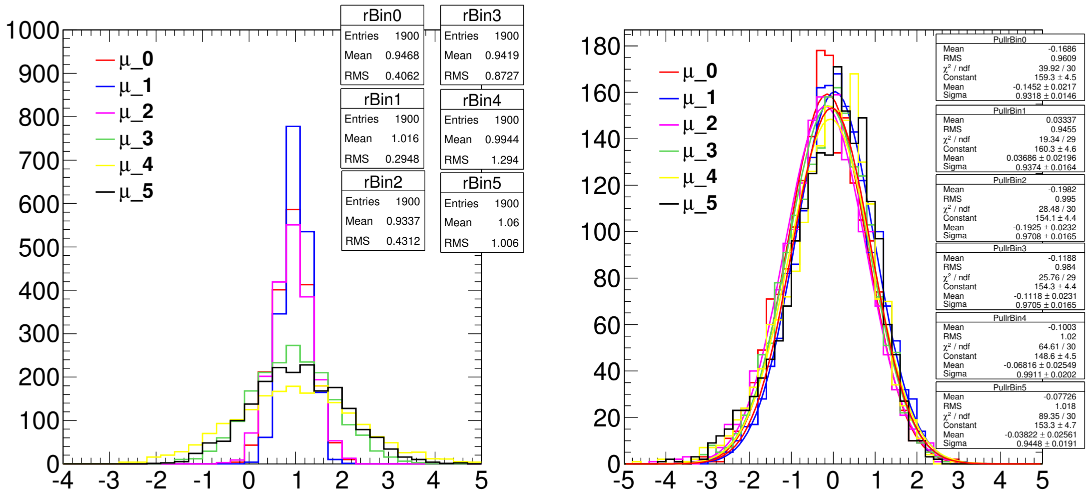


Figure 4.23.: Differential Higgs boson production cross section as a function of the reconstructed $p_{\mathrm{T}}^{\mathrm{H}}$, before applying the unfolding procedure. Data values after the background subtraction are shown together with the statistical and the systematic uncertainties, determined propagating the sources of uncertainty through the fit procedure. The line and dashed area represent the SM theoretical estimates in which the acceptance of the dominant ggH contribution is modelled by Powheg V1. The sub-dominant component of the signal is denoted as XH=VBF+VH, and is shown with the cross filled area separately.

The basic principle behind the unfolding procedure in this analysis is to use MC signal samples to make the “true” distribution of the variable of interest, which is obtained using simulated events before particle interaction with the detector, and the same distribution obtained using events reconstructed after the full Geant4 simulation of the CMS detector and event reconstruction. These two distributions are used to calculate the detector response matrix M:



(a) (b)

Figure 4.24.: Signal strength distribution as extracted from the fit of toy MC samples (a). Distribution of the pull of the signal strength parameters (b).

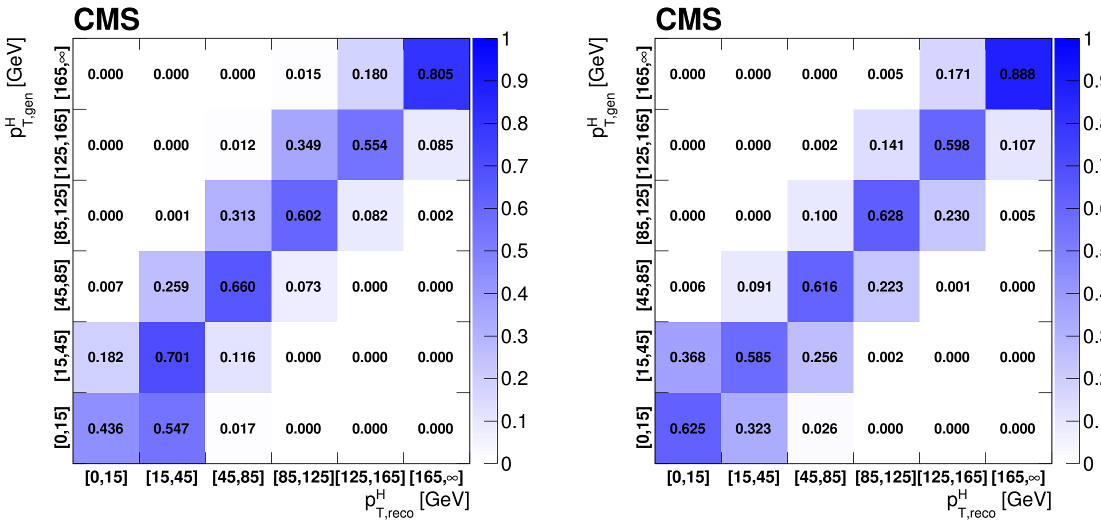
$$R_i^{\mathrm{MC}} = \sum_{j=1}^{n} M_{ij} T_j^{\mathrm{MC}} \quad , \tag{4.30}$$

where R^{MC} and T^{MC} are two n-dimensional vectors representing the distribution before and after event processing through CMS simulation and reconstruction. The dimension n of the two vectors corresponds to the number of bins in the distributions, equal to six in this analysis. The response matrix M includes all the effects related to the detector and analysis selection that affect the R^{MC} distribution. The goal of the unfolding procedure is to obtain the T^{truth} distribution starting from the measured R^{observed} distribution by inverting the matrix M. To avoid the large variance and strong negative correlation between the neighbouring bins [19], the unfolding procedure in this analysis relies on the singular value decomposition [58] method based on the Tikhonov regularization function. Since the response matrix is in general limited by the statistical uncertainties of simulated samples and given the finite data statistical accuracy, a simple inversion could lead to large fluctuations between bins in the unfolded result. In particular, if the off-diagonal elements of the response matrix are sizeable, the unfolded distribution has large variance and strong negative correlations between the neighbouring bins [19]. Several unfolding methods with regularization are available in literature, such as a method based on the Bayes' theorem, which overcome the unfolding instability using an iterative procedure [59]. One possible solution is the utilization of regularization methods. Such methods introduce a regularization function that controls the smoothness of the distribution and depends generally on one regularization parameter, which can be controlled to achieve the desired degree of smoothness. The choice of the regularization parameter is particularly critical, and it should represent an optimal trade-off between taming the fluctuations in the unfolded result, and biasing the unfolded distribution towards the one used to build the response matrix. The main feature of this method is the use of the singular value decomposition of the response matrix, including an additional term to suppress the oscillatory component of the solution, i.e. the regularization term, which represents some *a priori* knowledge of the final solution. The regularization parameter is chosen to obtain results that are robust against numerical instabilities and statistical fluctuations, following the prescription described in Ref. [58]. Maybe I should add an appendix describing the SVD method in details

The response matrix is built as a two-dimensional histogram, with the generator-level $p_{\mathrm{T}}^{\mathrm{H}}$ on the y axis and the same variable after the reconstruction on the x axis, using the same binning for both distributions. The resulting detector response matrix, including all signal sources and normalized by row, is shown in Fig. 4.25(a). The value of the diagonal bins corresponds to the stability S. The same matrix, normalized by column, is shown in Fig. 4.25(b). In this case the diagonal bins correspond to the purity P. The S and P parameters, defined in Sec. 4.3, provide an estimate of the $p_{\mathrm{T}}^{\mathrm{H}}$ resolution and migration effects. The main source of bin migrations effects in the response matrix is the limited resolution in the measurement of $E_{\mathrm{T}}^{\mathrm{miss}}$.

The resulting detector response matrix, which includes the effects of all signal sources and is represented by normalizing each row to unity is shown in Fig. 4.25(a). This representation shows the stability S in the diagonal bins, where S is defined as the ratio of the number of events generated and reconstructed in a given bin, and the number of events generated in that bin. In addition, a deconvolution matrix is constructed by normalizing each column to unity and is shown in Fig. 4.25(b). This latter representation shows the purity P in the diagonal bins, where P is defined as the ratio of the number of events generated and reconstructed in a given bin, and the number of events reconstructed in that bin. The S and P parameters provide an estimate of the $p_{\mathrm{T}}^{\mathrm{H}}$ resolution and of migration effects. The response matrix built including all signal sources is shown in Fig. 4.25. In order to point out either the purity or the stability in diagonal bins, each column or row of the matrix was respectively normalized to unity. The matrix obtained in the first case is what is actually called detector response matrix, while in the other case the matrix is usually referred to as detector deconvolution matrix.

Several closure tests are performed in order to validate the unfolding procedure. To estimate the uncertainty in the unfolding procedure due to the particular model adopted for building the response matrix, two independent gluon fusion samples are used, corresponding to two different generators: POWHEG V1 and JHUGEN generators, both interfaced to PYTHIA 6.4. The JHUGEN generator sample is used to build the response matrix while the POWHEG V1 sample is used for the measured and the MC distributions at generator level.



(a) Response matrix (b) Deconvolution matrix

Figure 4.25.: Response matrix (a) and deconvolution matrix (b) including all signal processes. The matrices are normalized either by row (a) or by column (b) in order to show the purity or stability respectively in diagonal bins.

The result of this test shows good agreement between the unfolded and the distribution from MC simulation.

In order to further prove the choice of the regularization parameter, a large number of simulated pseudo-experiments has been generated to verify that the coverage of the unfolded uncertainties obtained with this procedure is as expected. From each pseudo-experiment the reconstructed p_T^H spectrum is obtained and then unfolded using the procedure described above, including only the statistical uncertainties. The coverage is calculated for each p_T^H bin, counting the number of pseudo-experiments for which the statistical uncertainty covers the true value. The confidence intervals are calculated using the Clopper-Pearson approach, and the results are shown in Table 4.10 for different values of the regularization parameter: starting from $k_{reg} = 2$ (stronger regularization) up to $k_{reg} = 5$ (weaker regularization). The criterion for choosing the best k_{reg} value is to increase the regularization as much as possible without introducing a bias, i.e. until a 68% coverage is fulfilled. This criterion leads to the same result as the prescription described in Ref. [58], strengthening the choice of $k_{reg} = 3$.

4.7.1. Treatment of systematic uncertainties

An important aspect of this analysis is the treatment of the systematic uncertainties and the error propagation through the unfolding procedure. The sources of uncertainty are

Table 4.10.: Coverage interval for each bin and for different values of the regularization parameter, obtained using pseudo-experiments.

$p_{\mathrm{T}}^{\mathrm{H}}$ bin [GeV]	Coverage			
	$k_{\mathrm{reg}} = 2$	$k_{\mathrm{reg}} = 3$	$k_{\mathrm{reg}} = 4$	$k_{\mathrm{reg}} = 5$
0–15	$0.654^{+0.015}_{-0.016}$	$0.704^{+0.015}_{-0.015}$	$0.727^{+0.014}_{-0.015}$	$0.755^{+0.014}_{-0.014}$
15–45	$0.701^{+0.015}_{-0.015}$	$0.665^{+0.015}_{-0.016}$	$0.683^{+0.015}_{-0.015}$	$0.733^{+0.014}_{-0.015}$
45-85	$0.717^{+0.014}_{-0.015}$	$0.706^{+0.015}_{-0.015}$	$0.709^{+0.015}_{-0.015}$	$0.716^{+0.014}_{-0.015}$
85–125	$0.634^{+0.016}_{-0.016}$	$0.681^{+0.015}_{-0.015}$	$0.714^{+0.015}_{-0.015}$	$0.739^{+0.014}_{-0.015}$
125–165	$0.599^{+0.015}_{-0.016}$	$0.650^{+0.015}_{-0.016}$	$0.700^{+0.015}_{-0.015}$	$0.751^{+0.014}_{-0.014}$
165–∞	$0.632^{+0.016}_{-0.016}$	$0.674^{+0.015}_{-0.015}$	$0.701^{+0.015}_{-0.015}$	$0.722^{+0.014}_{-0.015}$

divided into three categories, depending on whether the uncertainty affects only the signal yield (type A), both the signal yield and the response matrix (type B), or only the response matrix (type C). These three classes propagate differently through the unfolding procedure.

Type A uncertainties are extracted directly from the fit in the form of a covariance matrix, which is passed to the unfolding tool as the covariance matrix of the measured distribution. The nuisance parameters belonging to this category are the background shape and normalization uncertainties. To extract the effect of type A uncertainties a dedicated fit is performed, fixing to constant all the nuisance parameters in the model, but type A nuisance parameters. The correlation matrix among the six signal strengths corresponding to the six $p_{\mathrm{T}}^{\mathrm{H}}$ bins, including all type A uncertainties, is shown in Fig. 4.26. The correlation $\mathrm{cor}(i,j)$ of bins i and j is defined as:

$$\mathrm{cor}(i,j) = \frac{\mathrm{cov}(i,j)}{s_i s_j} \quad , \tag{4.31}$$

where $\mathrm{cov}(i,j)$ is the covariance of bins i and j, and (s_i, s_j) are the standard deviations of bins i and j, respectively.

The nuisance parameters falling in the type B class are:

- the b veto scale factor. It affects the signal and background templates by varying the number of events with jets that enter the selection. It also affects the response matrix because the reconstructed spectrum is harder or softer depending on the number of jets, which in turn depends on the veto.

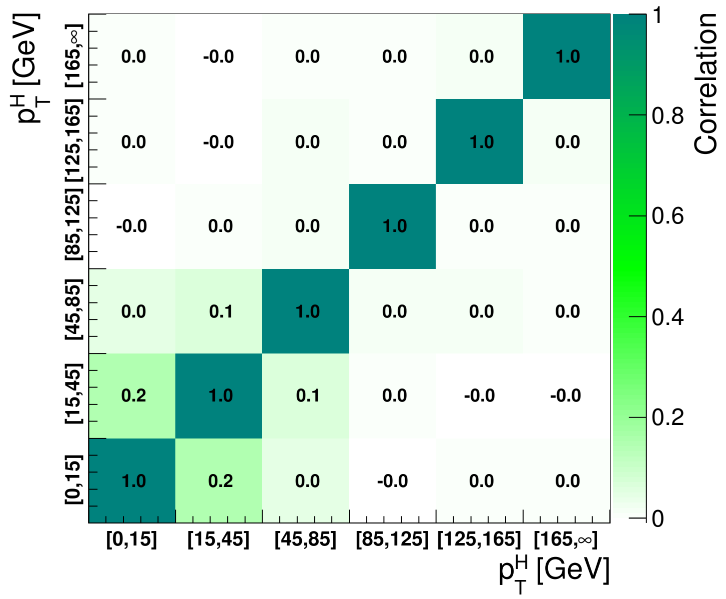


Figure 4.26.: Correlations among the signal strengths corresponding to the six p_T^H bins including all type A uncertainties.

- the lepton efficiency scale factor. It affects the signal and background template shape and normalization. It affects the response matrix by varying the reconstructed spectrum;
- the E_T^{miss} scale and resolution, which have an effect similar to the above;
- lepton scale and resolution. The effect is similar to the above;
- jet energy scale. It affects the signal and background template shape and normalization. It also affects the response matrix because, by varying the fraction of events with jets, the b veto can reject more or fewer events, thus making the reconstructed spectrum harder or softer.

The effect of each type B uncertainty is evaluated separately, since each one changes the response matrix in a different way. In order to evaluate their effect on the signal strengths parameters, two additional fits are performed, each time fixing the nuisance parameter value to ± 1 standard deviation with respect to its nominal value. The results of the fits are then compared to the results of the full fit obtained by floating all the nuisance parameters, thus determining the relative uncertainty on the signal strengths due to each nuisance parameter, as shown in Tab. 4.11. Using these uncertainties, the measured spectra for each type B source are built. The effects are propagated through the unfolding by building the corresponding variations of the response matrix and unfolding the measured spectra with the appropriate matrix.

Table 4.11.: Effect of all the Type B uncertainties on the signal strengths of each bin. In the table are reported the signal strength variations corresponding to an up or down scaling of each nuisance.

Type B uncertainty	Effect on signal strength ($+1\sigma/-1\sigma$ [%])					
	[0–15]	[15–45]	[45–85]	[85–125]	[125–165]	[165–∞]
b veto	-10.1/-8.8	7.3/12.2	-6.3/3.1	-14.4/-4.8	-5.4/14.5	-7.9/17.8
lepton efficiency	-14.7/-3.9	4.5/15.1	-5.7/2.5	-13.2/-5.3	-0.2/7.6	-0.1/6.8
$E_{\mathrm{T}}^{\mathrm{miss}}$ resolution	-12.5/0.0	15.4/-0.0	-12.8/-0.0	8.7/0.0	-20.9/-0.0	10.5/0.0
$E_{\mathrm{T}}^{\mathrm{miss}}$ scale	-14.4/-6.8	-0.0/17.7	-6.1/-7.1	9.6/-20.9	2.3/32.4	2.5/2.6
lepton resolution	-12.5/-0.0	11.2/0.0	-2.4/0.0	-13.4/-0.0	9.9/0.0	-4.6/-0.0
electron momentum scale	-2.7/-13.1	15.9/9.9	10.8/-16.8	16.2/-33.1	30.9/-14.4	12.6/-10.9
muon momentum scale	-7.0/-10.7	11.8/8.9	1.1/-8.7	-0.7/-14.4	14.5/-4.6	8.0/-1.6
jet energy scale	-10.9/-10.1	9.0/9.0	-3.0/-2.9	-10.3/-8.9	0.3/3.4	5.2/3.1

Type C uncertainties are related to the underlying assumption on the Higgs boson production mechanism used to extract the fiducial cross sections. These are evaluated using alternative response matrices that are obtained by varying the relative fraction of the VBF and ggH components within the experimental uncertainty, as given by the CMS combined measurement [60]. Three different response matrices are built, corresponding to the nominal, scaled up, and scaled down VBF/ggH ratio. The nominal matrix assumes the SM VBF/ggH ratio, while up- and down-scaled matrices are constructed by varying the SM signal strengths within the experimental constraints for VBF and ggH in such a way as to obtain the maximal variation of the VBF/ggH ratio allowed by the experimental constraints. These three matrices are used to unfold the reconstructed spectrum with the nominal VBF/ggH fraction, and obtain an uncertainty on the unfolded spectrum.

4.8. Results

In order to unfold the spectrum, the procedure described in section 4.7 has been pursued. The statistical plus type A systematic uncertainties are propagated by the unfolding procedure into the final spectrum, taking into account the signal strengths covariance matrix. The type B systematic uncertainty has been propagated using the following procedure: for each $p_{\mathrm{T}}^{\mathrm{H}}$ bin, we compute the upper bound of the systematic band computing the square sum of all the signal strength variations that deviate in the up direction with respect to the bin central value, whether or not this variation corresponds to the up or down shift of the systematic uncertainty. The same is done for the lower bound of the systematic band. If both the up and down shifts of a given nuisance parameter lead to a same direction variation of the signal strength, only the larger variation is considered.

The unfolded $p_{\mathrm{T}}^{\mathrm{H}}$ spectrum is shown in Fig. 4.27. Statistical, systematic, and theoretical uncertainties are shown as separate error bands in the plot. The unfolded spectrum is

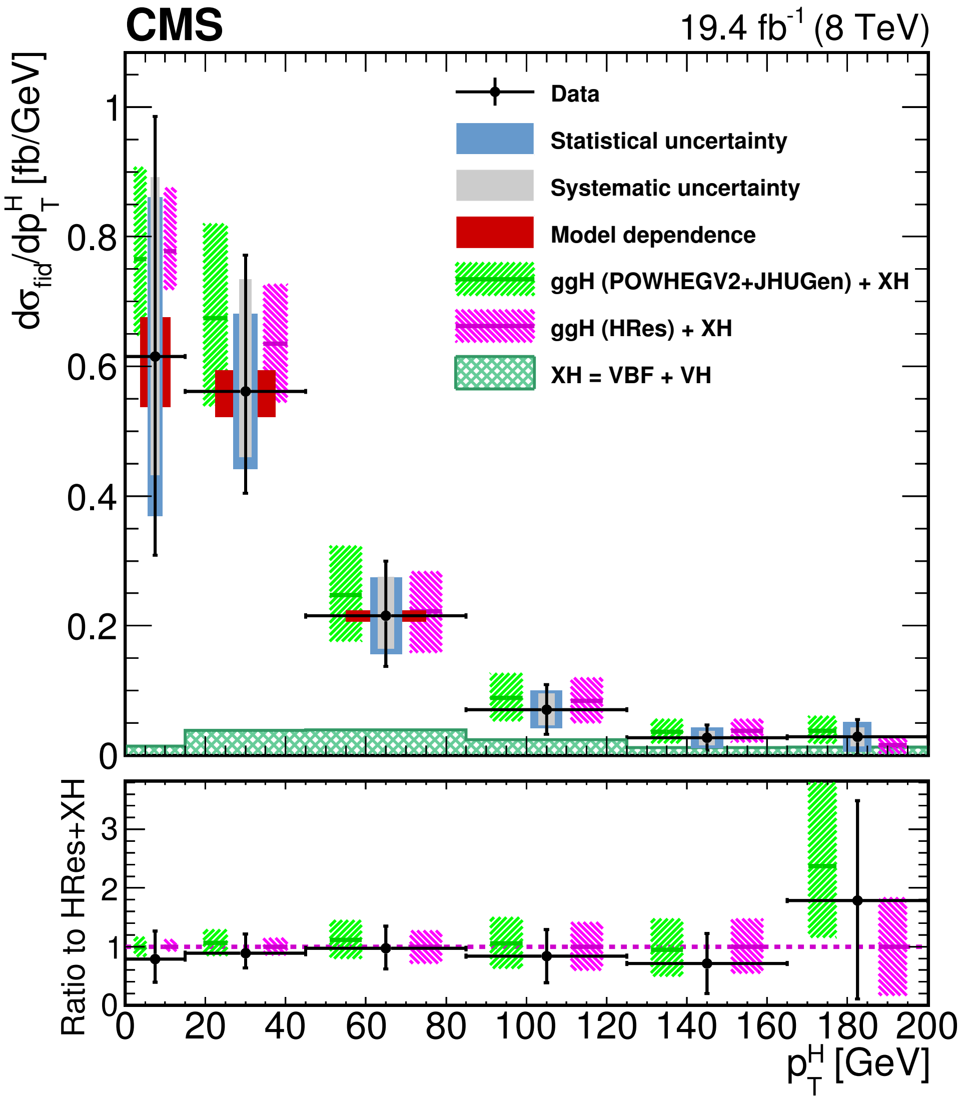


Figure 4.27.: Higgs boson production cross section as a function of $p_{\mathrm{T}}^{\mathrm{H}}$, after applying the unfolding procedure. Data points are shown, together with statistical and systematic uncertainties. The vertical bars on the data points correspond to the sum in quadrature of the statistical and systematic uncertainties. The model dependence uncertainty is also shown. The pink (and back-slashed filling) and green (and slashed filling) lines and areas represent the SM theoretical estimates in which the acceptance of the dominant ggH contribution is modelled by HRes and Powheg V2, respectively. The subdominant component of the signal is denoted as XH=VBF+VH and it is shown with the cross filled area separately. The bottom panel shows the ratio of data and Powheg V2 theoretical estimate to the HRes theoretical prediction.

compared with the SM-based theoretical predictions where the ggH contribution is modelled

Table 4.12.: Differential cross section in each p_T^H bin, together with the total uncertainty and the separate components of the various sources of uncertainty.

p_T^H [GeV]	$d\sigma/dp_T^H$ [fb/GeV]	Total uncertainty [fb/GeV]	Statistical uncertainty [fb/GeV]	Type A uncertainty [fb/GeV]	Type B uncertainty [fb/GeV]	Type C uncertainty [fb/GeV]
0-15	0.615	+0.370/-0.307	±0.246	±0.179	+0.211/-0.038	+0.0782/-0.0608
15-45	0.561	+0.210/-0.157	±0.120	±0.093	+0.146/-0.041	+0.0395/-0.0327
45-85	0.215	+0.084/-0.078	±0.059	±0.037	+0.047/-0.034	+0.0089/-0.0084
85-125	0.071	+0.038/-0.038	±0.029	±0.017	+0.018/-0.017	+0.0018/-0.0022
125-165	0.027	+0.020/-0.019	±0.016	±0.009	+0.007/-0.007	+0.0003/-0.0006
165-∞	0.028	+0.027/-0.027	±0.023	±0.012	+0.008/-0.007	+0.0002/-0.0006

using the HRES and POWHEG V2 programs. The comparison shows good agreement between data and theoretical predictions within the uncertainties. The measured values for the differential cross section in each bin of p_T^H are reported together with the total uncertainty in Table 4.12.

Figure 4.28 shows the correlation matrix for the six bins of the differential spectrum. The correlation of of bins is defined as in Eq. (4.31).

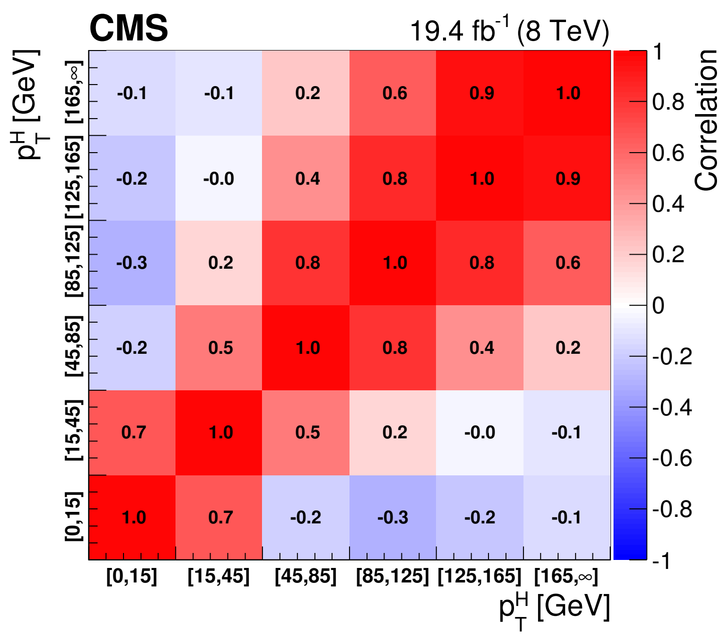


Figure 4.28.: Correlation matrix among the p_T^H bins of the differential spectrum.

To measure the inclusive cross section in the fiducial phase space, the differential measured spectrum is integrated over $p_{\mathrm{T}}^{\mathrm{H}}$. In order to compute the contributions of the bin uncertainties of the differential spectrum to the inclusive uncertainty, error propagation is performed taking into account the covariance matrix of the six signal strengths. For the extrapolation of this result to the fiducial phase space, the unfolding procedure is not needed, and the inclusive measurement has only to be corrected for the fiducial phase space selection efficiency ϵ_{fid}. Dividing the measured number of events by the integrated luminosity and correcting for the overall selection efficiency, which is estimated in simulation to be $\epsilon_{\mathrm{fid}} = 36.2\%$, the inclusive fiducial $\sigma \times \mathcal{B}$, σ_{fid}, is computed to be:

$$\sigma_{\mathrm{fid}} = 39 \pm 8\ (\mathrm{stat}) \pm 9\ (\mathrm{syst})\ \mathrm{fb} \quad , \tag{4.32}$$

in agreement within the uncertainties with the theoretical estimate of 48 ± 8 fb, computed integrating the spectrum obtained with the POWHEG V2 program for the ggH process and including the XH contribution.

Chapter 5.

Search for the SM Higgs boson in the H $\rightarrow$ WW channel with the first 13 TeV LHC data

5.1. Introduction

In this chapter, the first search for the SM Higgs boson decaying to a W boson pair at 13 TeV is presented, using a total integrated luminosity of 2.3 fb^{-1}, collected during the 2015 proton proton data taking period of the LHC.

Final states in which the two W bosons decay leptonically are studied. Therefore, events with a pair of oppositely-charged leptons, exactly one electron and one muon, a substantial amount of missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, due to the presence of neutrinos in the final state, and either zero or one jet are selected. This signature is common to other processes, which enter the analysis as backgrounds. The main background comes from WW production, irreducible background that shares the same final states and can only be separated by the use of certain kinematic properties. Another important background is W+jets, where a jet can mimick a leptonic signature. Background coming from top quark events, i.e. $\mathrm{t\bar{t}}$ and single top production, is also important, followed by other processes such as Drell-Yan, WZ, and other EWK production. The analysis strategy follows the one used during Run 1 in the same channel, described in Chapter 4, with a few different aspects that are described in the next sections.

With respect to 8 TeV, the ggH production cross section at 13 TeV is expected to increase of a factor of 2, thus raising the number of expected signal events. In addition, the cross section for the background processes is increasing as well. The WW production cross section increases of a factor of 1.8 and and the $\mathrm{t\bar{t}}$ cross section of a factor of 3.5, due to the enhancement of the gluon PDFs at higher center of mass energies.

5.2. Data and simulated samples

Data recorded in proton proton collisions at 13 TeV during 2015 was used in the analysis, with a total integrated luminosity of 2.3 fb^{-1}. Single and double lepton triggers are used similarly to the same analysis at 8 TeV. The HLT paths and descriptions of the triggers used in this analysis are described in Tables 5.1 and 5.2 for electrons and muons respectively.

Table 5.1.: HLT paths related to Electrons

HLT Path	Description
HLT_Ele23_WPLoose_Gsf_v*	Single Electron trigger. Best trigger to be used for 2015 data. In HWW, we are using "Trigger safe" Id. Turn on is at around Ele p_T = 30 GeV
HLT_Ele17_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*	Double Electron Trigger. Best trigger to cover the turn on region from single electron trigger. "DZ" filter is also present. Its efficiency is also calculated separately.
HLT_Ele12_CaloIdL_TrackIdL_IsoVL_v*	This electron leg of HLT_Mu17_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v* same as Ele12 leg of double electron trigger.
HLT_Ele17_CaloIdL_TrackIdL_IsoVL_v*	This electron leg of HLT_Mu8_TrkIsoVVL_Ele17_CaloIdL_TrackIdL_IsoVL_v* same as Ele17 leg of double electron trigger.

The trigger efficiencies are measured in data and applied on simulated events as described in Sec. 4.2.1.

Concerning the simulated samples, several different Monte Carlo (MC) generators were used. In the simulation, 'lepton' includes also τ. Higgs signal samples have been simulated in all channels with POWHEG V2 [61, 21, 34], designed to describe the full NLO properties of these processes. In particular, for Higgs produced via gluon fusion [23], and vector-boson-fusion (VBF) [24], the decay of the Higgs boson into two W boson and subsequently into leptons was done using JHUGEN v5.2.5 [62]. For associated production with a vector boson (W^+H, W^-H, ZH) [63], including gluon fusion produced ZH (ggZH), the Higgs decay was done via PYTHIA 8.1 [43]. Alternative signal samples were produced with aMC@NLO [26], or with POWHEG V2 but decayed via PYTHIA 8.1 for gluon fusion and VBF assuming a Higgs boson mass of 125 GeV. In the following, the mass of the SM Higgs boson is assumed to be 125 GeV.

The WW production, irreducible background for the analysis, was simulated in different ways. POWHEG V2 [64] was used for $q\bar{q}$ produced WW in different decays. The cross section used for normalizing WW processes produced via $q\bar{q}$ was computed at next-to-next-to-leading order (NNLO) [65]. In order to control the top quark background processes, the analysis is performed with events that have no more than one high-p_T jet. The veto on

Table 5.2.: Muon trigger's elements description

HLT path	
HLT_IsoMu18_v*	single muon trigger
HLT_IsoTrMu20_v*	single muon trigger with tracker isolation
HLT_Mu17_TrkIsoVVL	leg for the HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*, HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v* and HLT_Mu17_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v* double lepton triggers
HLT_Mu8_TrkIsoVVL	leg for the HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v* and HLT_Mu8_TrkIsoVVL_Ele17_CaloIdL_TrackIdL_IsoVL_v* double lepton triggers
HLT_TkMu8_TrkIsoVVL	leg for the HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v* double muon trigger
$DZ_{\mu\mu}$	efficiency of DZ cut in the HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v* and HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v* double muon triggers, it is around 95%

high-p_{T} jets enhances the importance of logarithms of the jet p_{T}, spoiling the convergence of fixed-order calculations of the qq→WW process and requiring the use of dedicated resummation techniques for an accurate prediction of differential distributions [66, 67]. The p_{T} of the jets produced in association with the WW system is strongly correlated with its transverse momentum, $p_{\mathrm{T}}^{\mathrm{WW}}$, especially in the case where only one jet is produced. The simulated qq→WW events are reweighted to reproduce the $p_{\mathrm{T}}^{\mathrm{WW}}$ distribution from the p_{T}-resummed calculation.

Gluon fusion produced WW was generated, with and without Higgs diagrams, using MCFM v7.0 [68]. A $\mathrm{t\bar{t}}$ sample dilepton sample was also generated using POWHEG v2. The WW and $\mathrm{t\bar{t}}$ samples produced specifically for this analysis are presented in Table 5.3. Other background samples are used, a list of the most relevant ones is presented in Table 5.4.

All processes are generated using the NNPDF2.3 [69, 70] parton distribution functions (PDF) for NLO generators, while the LO version of the same PDF is used for LO generators. All the event generators are interfaced to PYTHIA 8.1 [43] for the showering of partons and hadronization, as well as including a simulation of the underlying event (UE) and multiple interaction (MPI) based on the CUET8PM1 tune [71]. To estimate the systematic uncertainties related to the choice of UE and MPI tune, the signal processes and the WW events are also generated with two alternative tunes which are representative of the errors on the tuning parameters. The showering and hadronization systematic uncertainty is

Table 5.3.: Simulated samples for $\mathrm{t\bar{t}}$ and WW production. The $gg \to$WW $\to 2\ell 2\nu$ (H diagr.) sample includes both ggH production, the ggWW component and the interference.

Process	$\sigma \times \mathcal{B}$ [pb]
$\mathrm{t\bar{t}} \to$WW $b\bar{b} \to 2\ell 2\nu b\bar{b}$	87.31
$\mathrm{q\bar{q}} \to$WW $\to 2\ell 2\nu$	12.178
$gg \to$WW $\to 2\ell 2\nu$	0.5905
$gg \to$WW $\to 2\ell 2\nu$ (H diagr.)	0.9544

Table 5.4.: Simulated samples for other backgrounds used in the analysis.

Process	$\sigma \times \mathcal{B}$ [pb]
Single top	71.7
Drell-Yan (10 GeV $< m_{\ell\ell} <$ 50 GeV)	20471.0
Drell-Yan ($m_{\ell\ell} >$ 50 GeV)	6025.26
WZ$\to 2\ell$2q	5.5950
ZZ$\to 2\ell$2q	3.2210
WWZ	0.1651
WZZ	0.05565
ZZZ	0.01398

estimated by interfacing the same MC samples with the HERWIG ++ 2.7 parton shower [72, 73]. For all processes, the detector response is simulated using a detailed description of the CMS detector, based on the GEANT4 package [35].

The simulated samples are generated with distributions for the number of pileup interactions that are meant to roughly cover, though not exactly match, the conditions expected for the different data-taking periods. In order to factorize these effects, the number of true pileup interactions from the simulation truth is reweighted to match the data. The re-weighting is propagated automatically to both the in-time pile up and the out-of-time one. In Fig. 5.1, the effect of this reweighting on a sample enriched in Drell-Yan events is shown. Before the reweighting the simulation is presented in the open red histogram; after the reweighting, it is represented by the solid green histogram that matched well the data. In order to select this sample, events with two leptons with $p_{\mathrm{T}} > 20$ GeV, opposite sign, and same flavour, are selected only if $|m_{\ell\ell} - m_{\mathrm{Z}}| < 15$ GeV.

The average number of pileup is approximately 11.5.

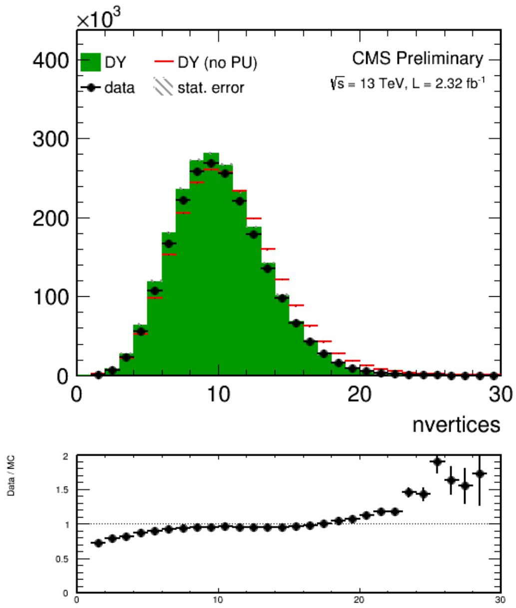


Figure 5.1.: Distributions of the number of vertices in a Drell-Yan enriched sample in data, together with the simulation before (red) and after (solid green) the pileup reweighting.

Different sources and calculations are used to obtain the cross sections for the different processes at 13 TeV. For Higgs signal, the cross sections used are the ones reported by the LHC Higgs Cross Section Working Group [74], computed at NNLO and NNLL QCD and NLO EW for gluon fusion, and at NNLO QCD and NLO EW for the rest of the production modes. The branching fractions are the ones reported in Ref. [17].

The cross section used for normalizing $q\bar{q}$ produced WW processes was computed at next-to-next-to-leading order (NNLO) [65]. The leading-order (LO) cross section for ggWW is obtained directly from MCFM. For gluon fusion, the difference between LO and NNLO cross sections is significantly big. A scale factor of 1.4 is theoretically calculated [75]. For the LO simulation of the interference between gg$\rightarrow$WW and gluon fusion produced H$\rightarrow$WW a k-factor of 1.87 is applied. This k-factor is obtained as the average between LO to NNLO ggH scale factor and LO to NLO ggWW scale factor.

The cross sections of the different single top processes are estimated by the LHC Top Working group [76] at NLO. The $t\bar{t}$ cross section is also provided by the LHC Top Working group [77], and it is computed at NNLO, with NNLL soft gluon resummation.

Drell-Yan (DY) production of Z/γ^* is generated using aMC@NLO [26]. Other multi-boson processes, such as WZ,ZZ, and VVV (V=W/Z), are generated with aMC@NLO and normalized to the cross section obtained at NLO in generation.

All processes are generated using the NNPDF2.3 [69, 70] parton distribution functions (PDF) for NLO generators, while the LO version of the same PDF is used for LO generators. All the event generators are interfaced to PYTHIA 8.1 for the showering of partons and hadronization, as well as including a simulation of the underlying event (UE) and multiple interaction (MPI) based on the CUET8PM1 tune [71].

5.3. Analysis strategy

5.3.1. Event reconstruction

Regarding the electrons, muons, jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ definition and reconstruction, the standard CMS recommendations described in Chapter 2 are used. The specific selections used in this analysis are briefly summarised below.

Muons are identified according to the CMS recommendations for the medium working point, with the addition of some extra cuts, as defined by the following selections:

- identified by the standard medium muon selection described in Sec. 2.4; Not yet defined :)
- $p_{\mathrm{T}} > 10\,\mathrm{GeV}$;
- $|\eta < 2.4|$;
- $|d_{xy}| < 0.01\,\mathrm{cm}$ for $p_{\mathrm{T}} < 20\,\mathrm{GeV}$ and $|d_{xy}| < 0.02\,\mathrm{cm}$ for $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, d_{xy} being the transverse impact parameter with respect to the primary vertex;
- $|d_z| < 0.1\,\mathrm{cm}$, where d_z is the longitudinal distance of the muon track in the tracker extrapolated along the beam direction.

For the muon isolation, the CMS recommended particle flow isolation based on the tight working point is used, corresponding to a requirement on the isolation variable of $ISO_{\mathrm{tight}} < 0.15$. In addition a tracker relative isolation is also applied.

For the electron identification, the tight working point is used. In addition some additional cuts to make the selection “trigger-safe” are included. This is done because the electron triggers already include some identification and isolation requirements that are based on the raw detector information, while the offline selections make use of particle flow requirements. The “trigger-safe” selections are defined to make the the offline identification and isolation requirements tighter with respect to the online triggers.

The simulated events are corrected for the lepton trigger, identification and isolation efficiencies measured in data using the same techniques described in Sec. 4.3.1.

Jets are defined clustering the particle flow objects using the anti-k_t algorithm with a distance parameter of 0.4. The CHS pileup mitigation technique is used. The L1, L2, L3 and L2L3 jet energy correction described in Sec. 2.4 are applied. The reject jets coming from calorimeter or readout electronics noise, the loose working point for PF jet identification is used.

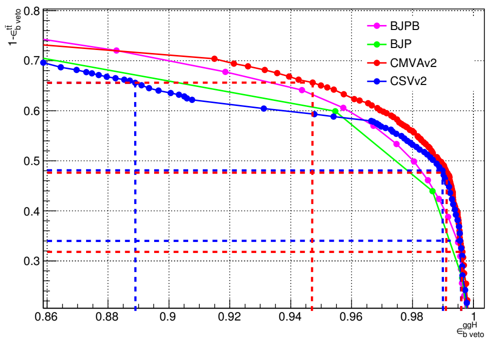
The b-tagging algorithm for this analysis is chosen comparing the performances of different algorithms using simulations for signal and background contributions in the phase space defined by the analysis kinematic requirements. More precisely, two MC samples are used, one corresponding the the $\mathrm{H} \to \mathrm{WW} \to 2\ell 2\nu$ signal produced via the ggH production mode and another corresponding to the $\mathrm{t\bar{t}}$ process. In fact, the first sample is enriched in light jets, i.e. originating by the hadronization of light quarks like u,d,c and s quarks, while the second sample is enriched in b jets, coming from the top quark decay. The b-veto efficiency, $\epsilon_{\mathrm{bveto}}$, is computed separately for the two samples and for the various b tagging algorithms. To compare the b tagging performance $\epsilon_{\mathrm{bveto}}$ is computed for different working points, i.e. different selections on the specific b tagging discriminator, and the results are reported in the form of a ROC curve. The ROC curves corresponding to events with 0, 1 and ≥ 2 jets are shown in Fig. 5.2. Events considered for this study are the ones passing the WW baseline selection.

The ROC curves show that the cMVAv2 algorithm has the best performance for the analysis phase space among the algorithms taken into account. For both the CSVv2 and cMVAv2 algorithms, three working points are defined corresponding to the mistag rates[1] of 10% for the loose, 1% for the medium and 0.1% for the tight working point. The distribution of the cMVAv2 discriminator associated to the leading jet both for the ggH and the $\mathrm{t\bar{t}}$ MC sample is shown in figure 5.3.

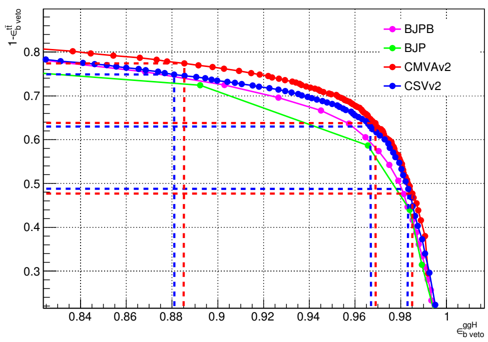
In order to determine the best working point for this analysis a preliminary significance assessment is performed, using a complete analysis procedure in which only statistical effects are taken into account (no systematics are included). The significance assessment was performed using a two dimensional discriminating variable consisting of the dilepton invariant mass versus the transverse mass. The assessment was performed with the following leptonic selection:

- two leptons, an electron and a muon with opposite charge, with leading lepton p_{T} greater than 20 GeV and sub-leading lepton p_{T} greater than 13 GeV;
- no other lepton (electron or muon) with p_{T} greated than 10 GeV;
- $m_{\ell\ell}$ greater than 12 GeV;
- PF type 1 corrected MET greater than 20 GeV;
- $p_{\mathrm{T}}^{\ell\ell}$ greater than 30 GeV.

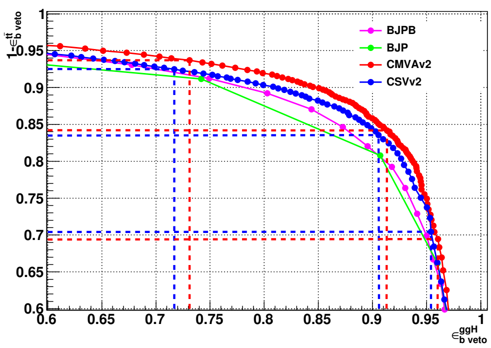
[1] The mistag rate is defined as the probability for a light jet to be identified as a b-jet by the b tagging algorithms.



(a) $N_{\mathrm{jets}} = 0$



(b) $N_{\mathrm{jets}} = 1$



(c) $N_{\mathrm{jets}} \geq 2$

Figure 5.2.: ROC curve for the b veto efficiency on signal and background events. The blue and red lines point out the signal efficiency and the background rejection corresponding to the three working points considered for the CSVv2 and the cMVAv2 algorithms respectively.

In addition to this global selection, two categories were identified:

- 0 jets: no jets above 30 GeV, jets between 20 GeV and 30 GeV are b-vetoed with the cMVAv2 WP under study;
- 1 jet: exactly 1 jet above 30 GeV, no b-tagged jets above 30 GeV according to the cMVAv2 WP under study.

The two categories were eventually combined together and the significance assessment was repeated for the three working points. With these selection we find the significance values listed in Table 5.5 for the three working points.

The working point providing the best significance in the combined $0 + 1$ jets category is found to be the loose one.

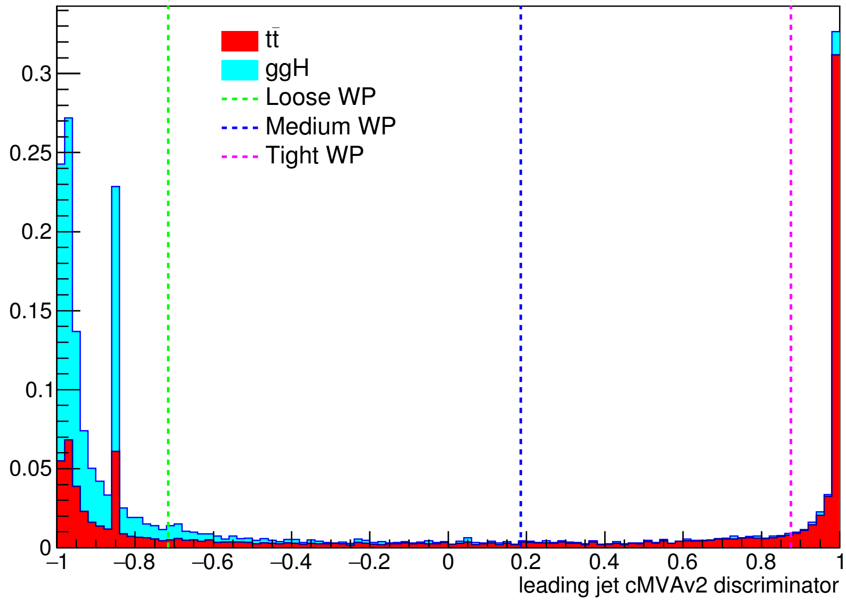


Figure 5.3.: cMVAv2 discriminator associated to the leading jet (with $p_{\mathrm{T}} > 30$ GeV) both for the ggH and the $t\bar{t}$ processes. The two processes are normalized to unity and stacked. The vertical dashed lines show the discriminator value corresponding to the three working points.

Table 5.5.: Significance corresponding to the three working points and for different jet categories using a shape analysis.

Jet category	Loose WP (-0.715)	Medium WP (0.185)	Tight WP (0.875)
0 jets	2.022	2.043	2.036
1 jet	1.439	1.404	1.305
$0 + 1$ jets	2.481	2.479	2.420

To correct for a possible different b tagging efficiency in data and simulation, the simulated events are reweighted using scale factors computed in bins of the jet η and p_{T}. These scale factors and the corresponding uncertainties are centrally calculated for each working point, in such a way to be employable by all the CMS analyses. The prescription to reweight the simulated events is the following. First of all one has to compute the b tagging efficiency using the MC samples, $\varepsilon_{\mathrm{MC}}(p_{\mathrm{T}}, \eta, f)$, for the chosen working point in bins of jet p_{T} and η. The efficiency has to be computed for different flavours f of the jets, b, c and light (u,d,s), using the jet matching information[2] which is available in all the MC samples. An MC-based event weight is then calculated computing the probability P_{MC} of a given b tagging configuration to occur, e.g.:

$$P_{\mathrm{MC}} = \prod_{i \,\in\, b-tagged-jets} \varepsilon_{\mathrm{MC}_i} \prod_{j \,\in\, non-b-tagged-jets} (1 - \varepsilon_{\mathrm{MC}_j}) \tag{5.1}$$

Afterwards, a similar probability is computed using data:

$$P_{\mathrm{DATA}} = \prod_{i \,\in\, b-tagged-jets} SF_i \varepsilon_{\mathrm{MC}_i} \prod_{j \,\in\, non-b-tagged-jets} (1 - SF_j \varepsilon_{\mathrm{MC}_j}) \quad , \tag{5.2}$$

where SF_i is the provided scale factor value for the relevant jet flavour, p_{T} and η. Products in Eqs. 5.1 and 5.2 run over all jets. The event weight is finally given by the ration $P_{\mathrm{DATA}}/P_{\mathrm{MC}}$.

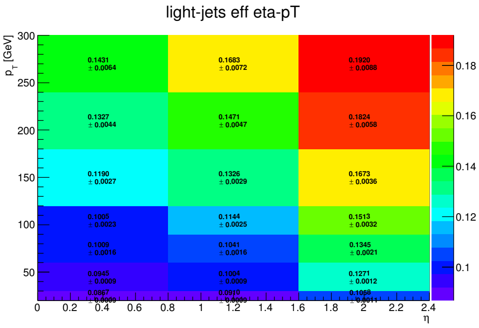
The b tagging efficiencies to be fed into Eq. 5.1 and Eq. 5.2 are derived using $\mathrm{t\bar{t}}$ simulated events and applying basic leptonic selections. These efficiencies are shown in Fig. 5.4 for light (a), c-jets (b) and b-jets (c), in bins of η and p_{T}. The uncertainties associated to the efficiencies are representative of the statistics of the simulated $\mathrm{t\bar{t}}$ sample, and are computed according to a binomial distribution.

The effect of the event reweighting is to correct the shape of the b tagging discriminator in simulation, moving events from the b tag region (discriminator greater than > -0.715) to the b veto region (discriminator < -0.715) and viceversa. A data/simulation comparison of the b tagging discriminator for the leading and subleading jets is performed to check the agreement after the application of the event weights. In order to evaluate the data/simulation agreement for b-jets, the data and simulation are compared in a top enriched control region, defined by the following requirements:

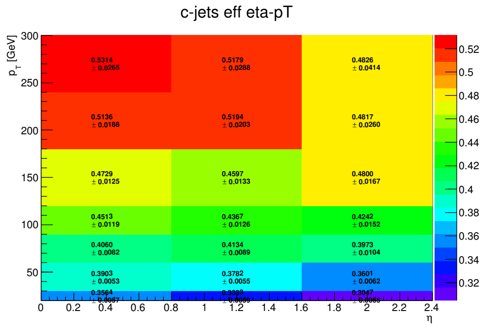
- two leptons, an electron and a muon with opposite charge, with leading lepton p_{T} greater than 20 GeV and sub-leading lepton p_{T} greater than 15 GeV;
- no other lepton (electron or muon) with p_{T} greater than 10 GeV;
- lepton invariant mass greater than 50 GeV;
- at least two jets with p_{T} greater than 30 GeV;
- at least one of the two leading jets with cMVAv2 btagging score greater than -0.715 (i.e. the loose working point).

In order to evaluate the agreement for light jets, a second control region is defined, populated by Z+light jet events, defined as follows:

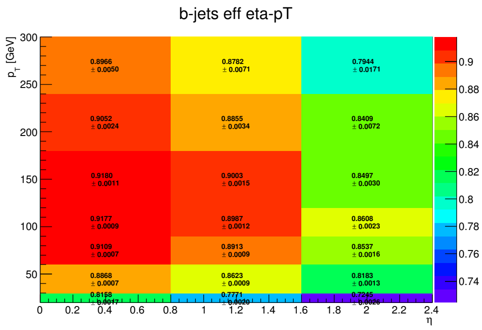
[2] There are a couple of techniques developed by the CMS Collaboration to assess the flavour of a reconstructed jet in simulation. The technique used here makes use of the flavour of the hadrons clustered into a jet.



(a)



(b)

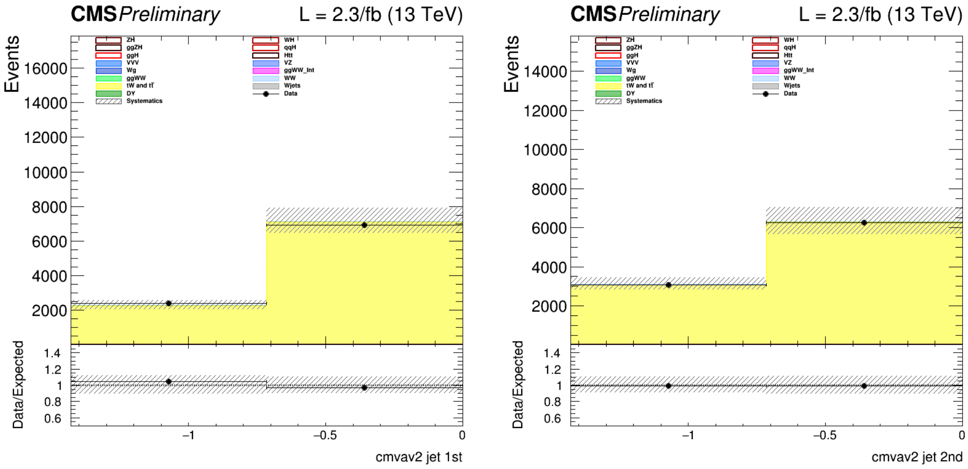


(c)

Figure 5.4.: B tagging efficiencies for light jets (a), c-jets (b) and b jets (c), as a function of η and p_T.

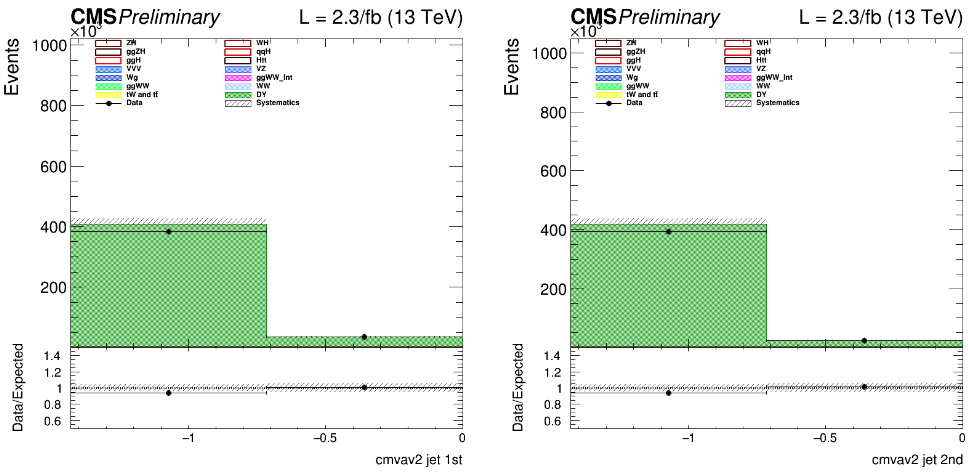
- two leptons, two electrons or two muons with opposite charge, with leading lepton p_T greater than 20 GeV and sub-leading lepton p_T greater than 15 GeV.
- no other lepton (electron or muon) with p_T greater than 10 GeV.
- lepton invariant mass greater between 80 GeV and 110 GeV.
- at least two jets with p_T greater than 30 GeV.
- at least one jet above 30 GeV.
- no jets above 20 GeV with a TCHE score above 2.1.

Although a Z+jets sample is dominated by light flavor jets, a b-veto on an alternative algorithm (TCHE) is applied to reduce the contamination from b-jets, especially above the cMVAv2 cut. This helps mitigating possible data/simulation discrepancies in the modeling of the heavy/light flavour ratio. The comparison between data and simulation after the event reweighting is shown in Figs. 5.5 and 5.6 for the b-jets and light jets enriched control regions, respectively.



(a) (b)

Figure 5.5.: B tagging cMVAv2 discriminator for the leading (a) and the subleading (b) jet in the b-jets enriched control region.



(a) (b)

Figure 5.6.: B tagging cMVAv2 discriminator for the leading (a) and the subleading (b) jet in the light jets enriched control region.

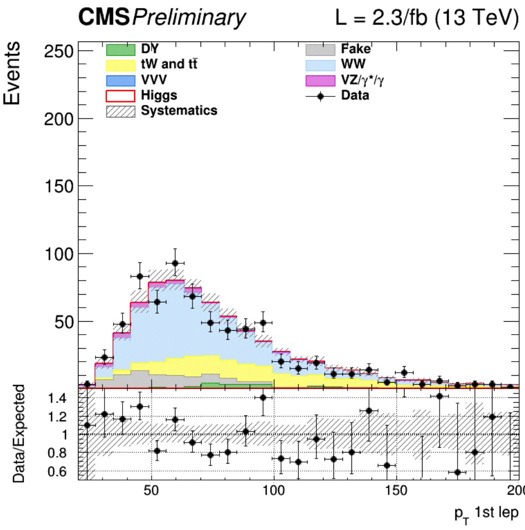
5.3.2. Event selection and background rejection

Since the ggH production mechanism, which is the main production mode for a Higgs mass of around 125 GeV, is characterized by the emission of few jets arising from initial or final state radiation, this analysis is limited to events with no jets or one jet. Due to the large DY background in di-electrons and di-muons events, only the $e\mu$ final state is studied in this early Run 2 data analysis, including the indirect contribution from τ leptons decaying to electron or muons. Exactly one electron and one muon are required to be reconstructed in the event with opposite charges and a minimum p_{T} of 10 (13) GeV for the muon (electron). One of the two leptons should also have a p_{T} greater than 20 GeV and both leptons are required to be well identified and isolated to reject fake leptons and leptons coming from QCD sources. To suppress background processes with three or more leptons in the final state, such as ZZ, WZ, Zγ, Wγ, or tri-boson production, no additional identified and isolated lepton with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ should be reconstructed. The low $m_{\ell\ell}$ region dominated by QCD production of leptons is not considered in the analysis and $m_{\ell\ell}$ is requested to be higher than 12 GeV. To suppress the background arising from DY events decaying to a τ lepton pair which subsequently decays to an $e\mu$ final state and suppress processes without genuine $E_{\mathrm{T}}^{\mathrm{miss}}$, a minimal $E_{\mathrm{T}}^{\mathrm{miss}}$ of 20 GeV is required. The DY background is further reduced by requesting $p_{\mathrm{T}}^{\ell\ell} > 30\,\mathrm{GeV}$. Finally the contribution from leptonic decays of single top and $\mathrm{t\bar{t}}$ production is reduced by requesting that no jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ are identified by the b tagging algorithm as originating from a b quark in the event.

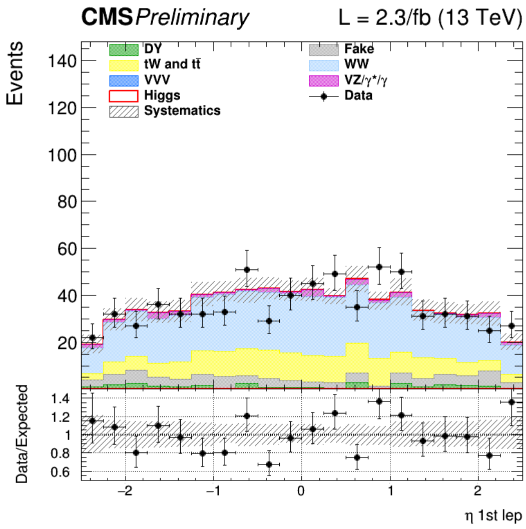
The requirements described above define the WW baseline selection. After those requirements the data sample is dominated by events arising from the non-resonant WW production and $\mathrm{t\bar{t}}$ production. To further reduce the effect of these backgrounds on the signal sensitivity, the events are categorized depending on the jet multiplicity, counting jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$. Events with zero associated jets mainly arise from the WW production, while WW and $\mathrm{t\bar{t}}$ productions have a similar contribution in the category with one jet. Higher jet multiplicity categories, which are sensitive to other Higgs production mechanisms, such as VBF, are not included in this analysis, given the very low expected yield for other production modes with the analysed integrated luminosity.

Distributions of some variables of interest for the 0 and 1 jet categories separately, but merging the eμ and μe final states together, are shown in Figs. 5.7, 5.8 and 5.9 after applying the WW baseline selections, with the addition of a cut on $m_{\ell\ell}$ to remove the Higgs signal contribution ($m_{\ell\ell} > 80\,\mathrm{GeV}$), and a cut on m_{T} to be orthogonal to the $\mathrm{Z}\gamma^* \to \tau\tau$ background control region ($m_{\mathrm{T}} > 60\,\mathrm{GeV}$).

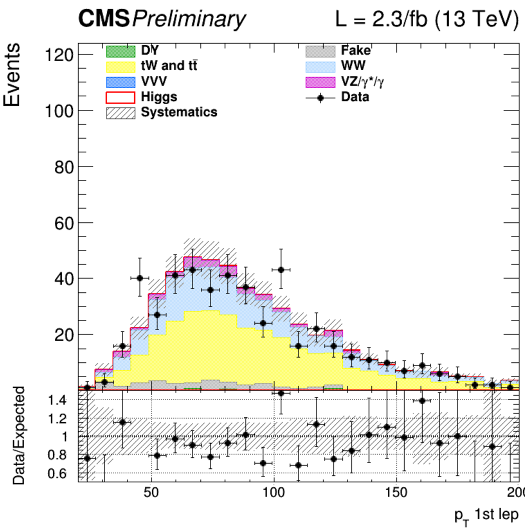
The W+jets background, where one jet can be misidentified as a lepton, is a sub-dominant background in the phase space defined by the analysis kinematic requirements. The 0 ad 1 jets categories are further split according to the lepton flavour to eμ and μe, where the first lepton refers to the leading one. In this way an improvement of about 10% in terms of the signal significance can be achieved, exploiting the different W+jets background contribution in the two categories. Indeed the probability for a jet to be misidentified as an electron or a muon is not the same.



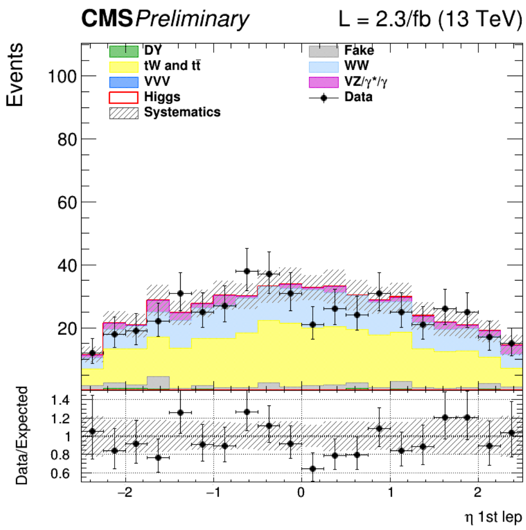
(a) 0 jets - $p_{\mathrm{T}}^{\ell,1}$



(b) 0 jets - $\eta^{\ell,1}$

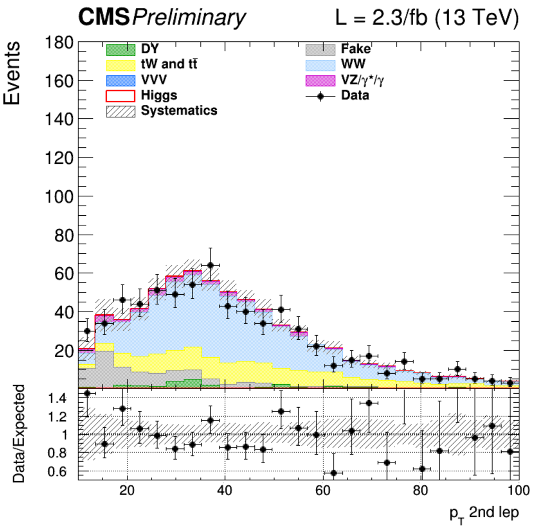


(c) 1 jet - $p_{\mathrm{T}}^{\ell,1}$

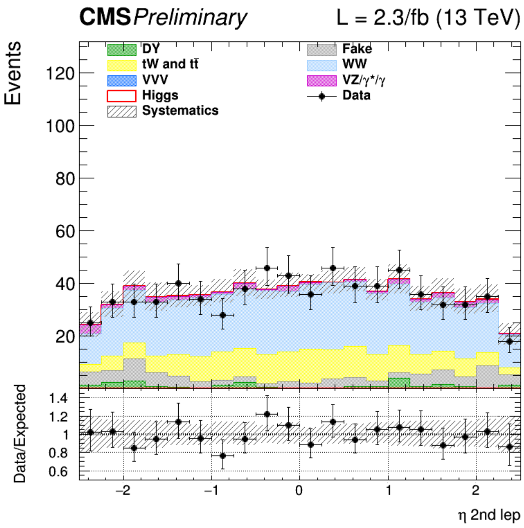


(d) 1 jet - $\eta^{\ell,1}$

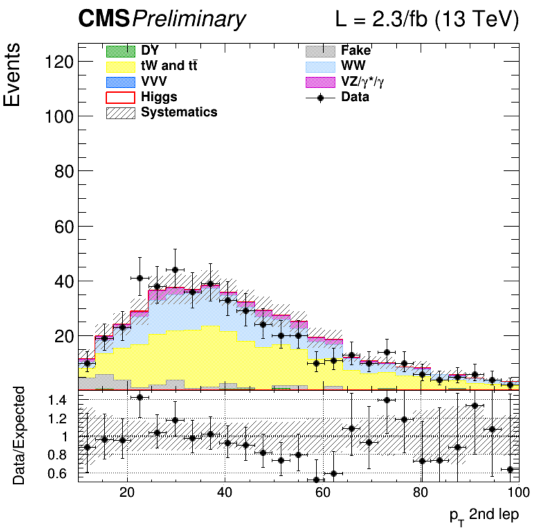
Figure 5.7.: Distributions of p_{T} (left) and η (right) of the leading lepton for events with 0 jet (upper row) and 1 jet (lower row), for the main backgrounds (stacked histograms), and for a SM Higgs boson signal with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ (superimposed and stacked red histogram) at the WW selection level. The last bin of the histograms includes overflows. The simulation of the WW background is normalized to data.



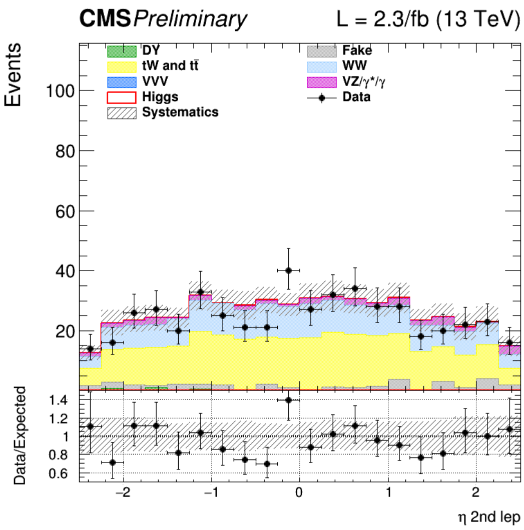
(a) 0 jets - $p_{\mathrm{T}}^{\ell,2}$



(b) 0 jets - $\eta^{\ell,2}$

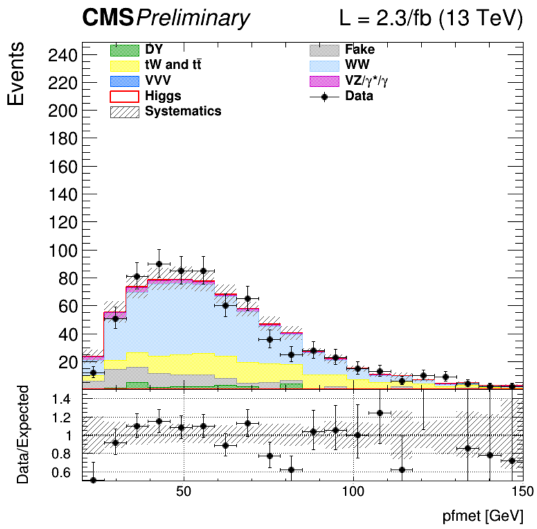


(c) 1 jet - $p_{\mathrm{T}}^{\ell,2}$

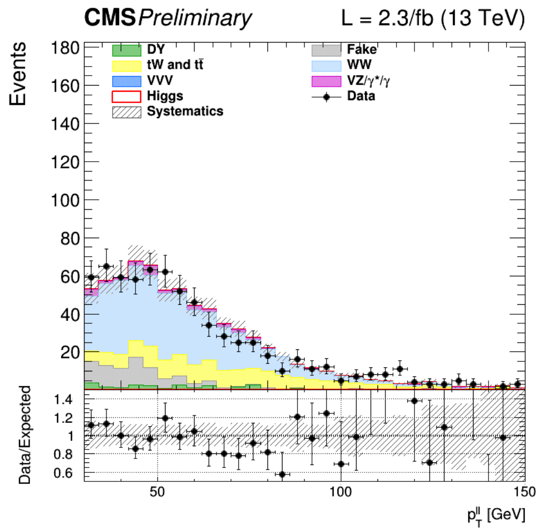


(d) 1 jet - $\eta^{\ell,2}$

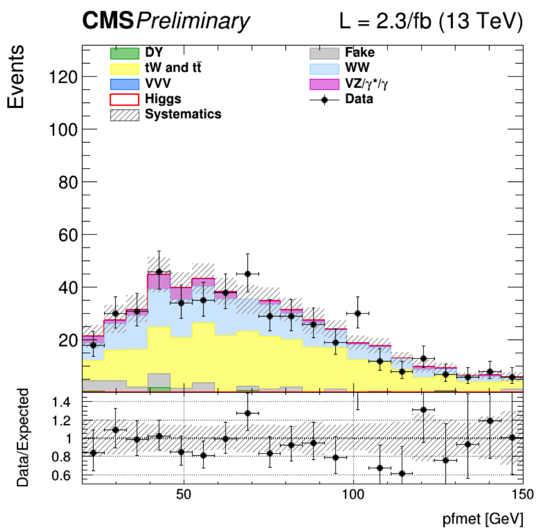
Figure 5.8.: Distributions of p_{T} (left) and η (right) of the subleading lepton for events with 0 jets (upper row) and 1 jet (lower row), for the main backgrounds (stacked histograms), and for a SM Higgs boson signal with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ (superimposed and stacked red histogram) at the WW selection level. The last bin of the histograms includes overflows. The simulation of the WW background is normalized to data.



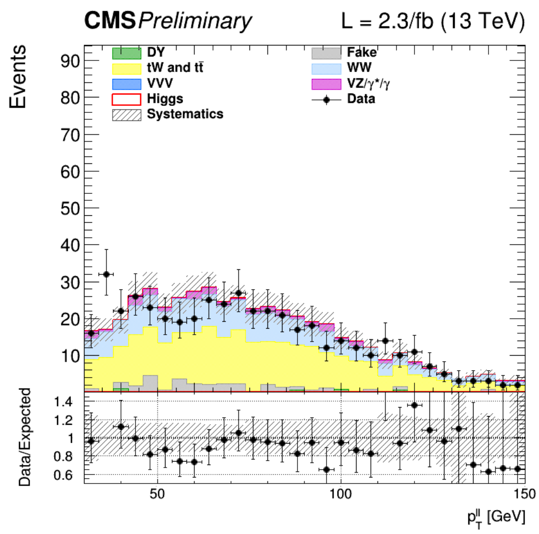
(a) 0 jets - $E_{\mathrm{T}}^{\mathrm{miss}}$



(b) 0 jets - $p_{\mathrm{T}}^{\ell\ell}$



(c) 1 jet - $E_{\mathrm{T}}^{\mathrm{miss}}$



(d) 1 jet - $p_{\mathrm{T}}^{\ell\ell}$

Figure 5.9.: Distributions of $E_{\mathrm{T}}^{\mathrm{miss}}$ (left) and $p_{\mathrm{T}}^{\ell\ell}$ (right) for events with 0 jets (upper row) and 1 jet (lower row), for the main backgrounds (stacked histograms), and for a SM Higgs boson signal with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ (superimposed and stacked red histogram) at the WW selection level. The last bin of the histograms includes overflows. The simulation of the WW background is normalized to data.

5.3.3. Signal extraction

To extract the Higgs boson signal contribution in the four previously mentioned categories, a similar approach to the one used in the Run 1 analysis [18] is pursued. The analysis is based on two-dimensional templates of $m_{\ell\ell}$ versus m_{T} to discriminate signal and background contributions. The $m_{\ell\ell}$ template is defined using 5 bins from $m_{\ell\ell} = 10\,\mathrm{GeV}$ up to $m_{\ell\ell} = 110\,\mathrm{GeV}$, while for the m_{T} template 7 bins are defined in the range $60\,\mathrm{GeV} < m_{\mathrm{T}} < 200\,\mathrm{GeV}$. The phase space with $m_{\mathrm{T}} < 60\,\mathrm{GeV}$ is used as an orthogonal control region to extract the normalization of the DY background. A binned maximum likelihood fit to the signal and background two-dimensional templates is performed to extract the signal strength in the four categories.

Distributions of the $m_{\ell\ell}$ and m_{T} variables after the WW level selection are shown in Fig. 5.10 for the 0 and 1 jet categories separately, but merging the eμ and μe final states together.

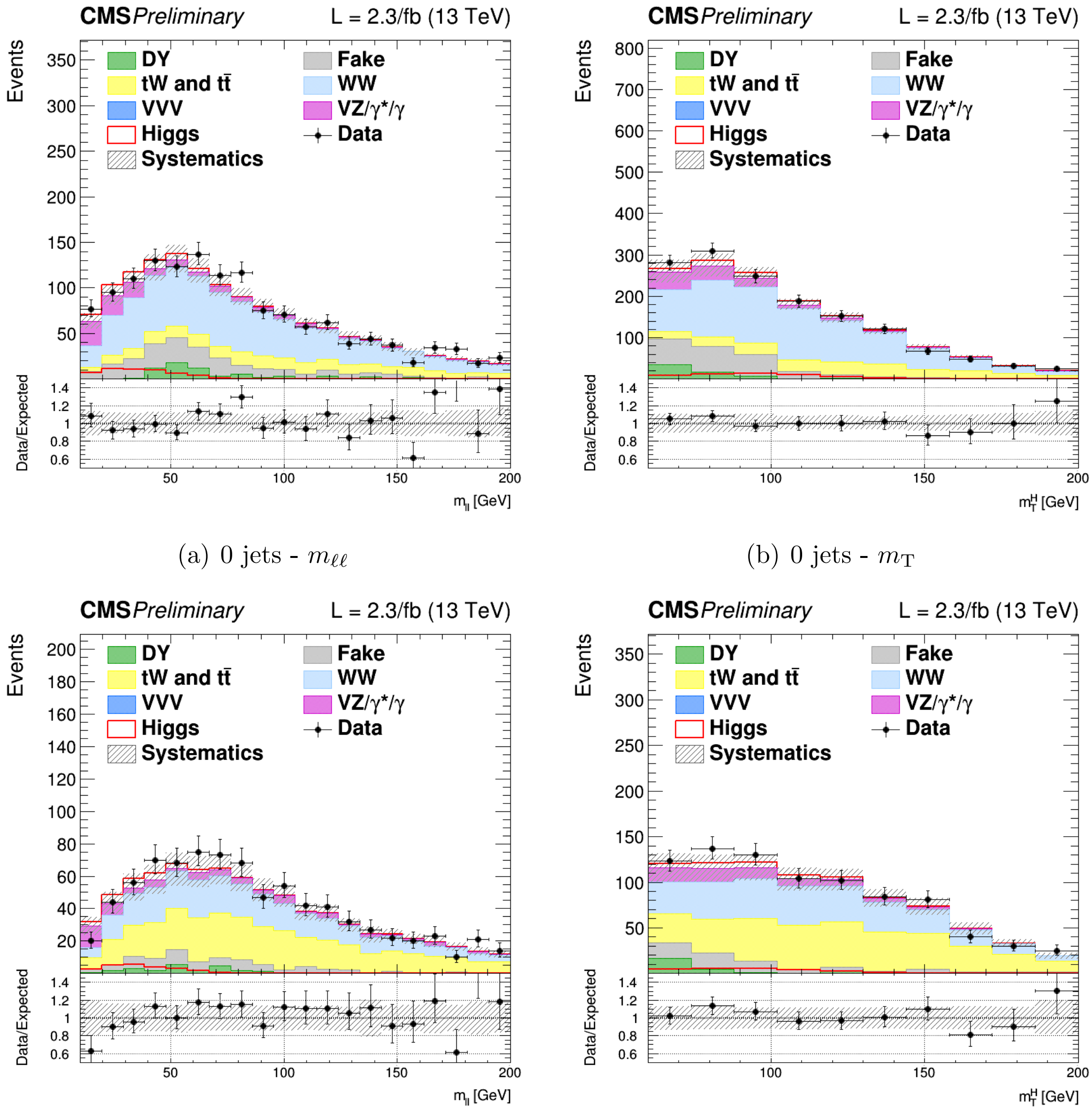
The statistical methodology used to interpret the data and to combine the results from the independent 0-jet and 1-jet categories in the eμ and μe final states has been developed by the ATLAS and CMS collaborations in the context of the LHC Higgs Combination Group [78, 60]. The number of events in each category and in each bin of two-dimensional template is modelled as a Poisson random variable, with a mean value given by the sum of the contributions from all the processes under consideration. Systematic uncertainties are represented by individual nuisance parameters with log-normal distributions. The uncertainties affect the overall normalization of the signal and backgrounds as well as the shape of the predictions across the distribution of the observables. Correlation between systematic uncertainties in different categories are taken into account.

5.4. Background estimation

The main background processes affecting the analysis signature, non-resonant WW production and top quark processes, are estimated using data. Backgrounds arising from an experimental misidentification of the objects, such as W+jets (also called "Fake"), are estimated using data as well. The other minor backgrounds are generally estimated directly from simulation as described in the following subsections.

5.4.1. WW background

The quark-induced WW background is simulated with NLO accuracy in perturbative QCD, and the transverse momentum of the diboson system is reweighted to match the NNLO+NNLL accuracy from theoretical calculations [66, 67]. However, given the large



(a) 0 jets - $m_{\ell\ell}$

(b) 0 jets - m_{T}

(c) 1 jet - $m_{\ell\ell}$

(d) 1 jet - m_{T}

Figure 5.10.: Distributions of $m_{\ell\ell}$ (left) and m_{T} (right) for events with 0 jets (upper row) and 1 jet (lower row), for the main backgrounds (stacked histograms), and for a SM Higgs boson signal with $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ (superimposed and stacked red histogram) at the WW selection level. The last bin of the histograms includes overflows. The simulation of the WW background is normalized to data.

uncertainties on the jet multiplicity distribution associated to this process, the normalization of this background is measured from data separately for the 0 and 1 jet categories. The normalization k-factors are extracted directly from the fit together with the signal strengths, leaving the WW normalization free to float separately in the two jet multiplicity categories. An orthogonal control region for the WW background normalization estimation is not needed in this case, owing to the different $m_{\ell\ell}$-m_{T} shape for signal and background.

The gluon-induced WW production is sub-dominant with respect to the quark-induced production, and its shape and normalization is fully taken from simulation, scaling the cross section to the theoretical prediction with NLO accuracy [75].

5.4.2. Top quark background

As explained in Sec. 5.3, the production of top quark pairs represents one of the dominant backgrounds in this analysis given its large cross section and a similar final state compared to the signal. A b-jet veto, based on the *cMVAv2* b tagging algorithm, is used to suppress this background and a reweighting procedure is applied on top of the simulated events to correct for different b tagging efficiency in data and simulation.

The top quark background normalization is measured using data, defining a b-jets enriched control region by inverting the b-jet veto. More precisely, the b-jets enriched control region for the 0-jet category is defined with the same WW baseline selection but requiring at least one jet with $20 < p_{\mathrm{T}} < 30\,\mathrm{GeV}$ to be identified as a b jet and no other jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$. For the 1-jet category, the b-jets enriched region is defined requiring exactly one jet with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ identified as a b-jet. To reduce other backgrounds in these two regions, the dilepton mass has to be greater than 50 GeV. Distributions of the $m_{\ell\ell}$ and m_{T} variables in the b-jets enriched control regions after applying the data driven estimation are shown in Figure 6.5, for the 0 and 1 jet categories separately.

The top quark background normalization is constrained during the fit procedure separately in the two jet categories, by means of the control regions defined above, which are treated in the fit as two additional categories.

5.4.3. Jet-induced (or Fake) background

One of the primary source belonging to this category arises from the misidentification of leptons in W+jets processes in the 0 jet category. Also, semileptonic $\mathrm{t\bar{t}}$ decays contribute especially for higher jet multiplicities. Multijet production and hadronic $\mathrm{t\bar{t}}$ decays are also taken into account, but have a much smaller contribution.

This background is fully estimated using data, with the technique described in Sec. 4.4.3. To check the agreement of the background estimated in this way with data, a control sample

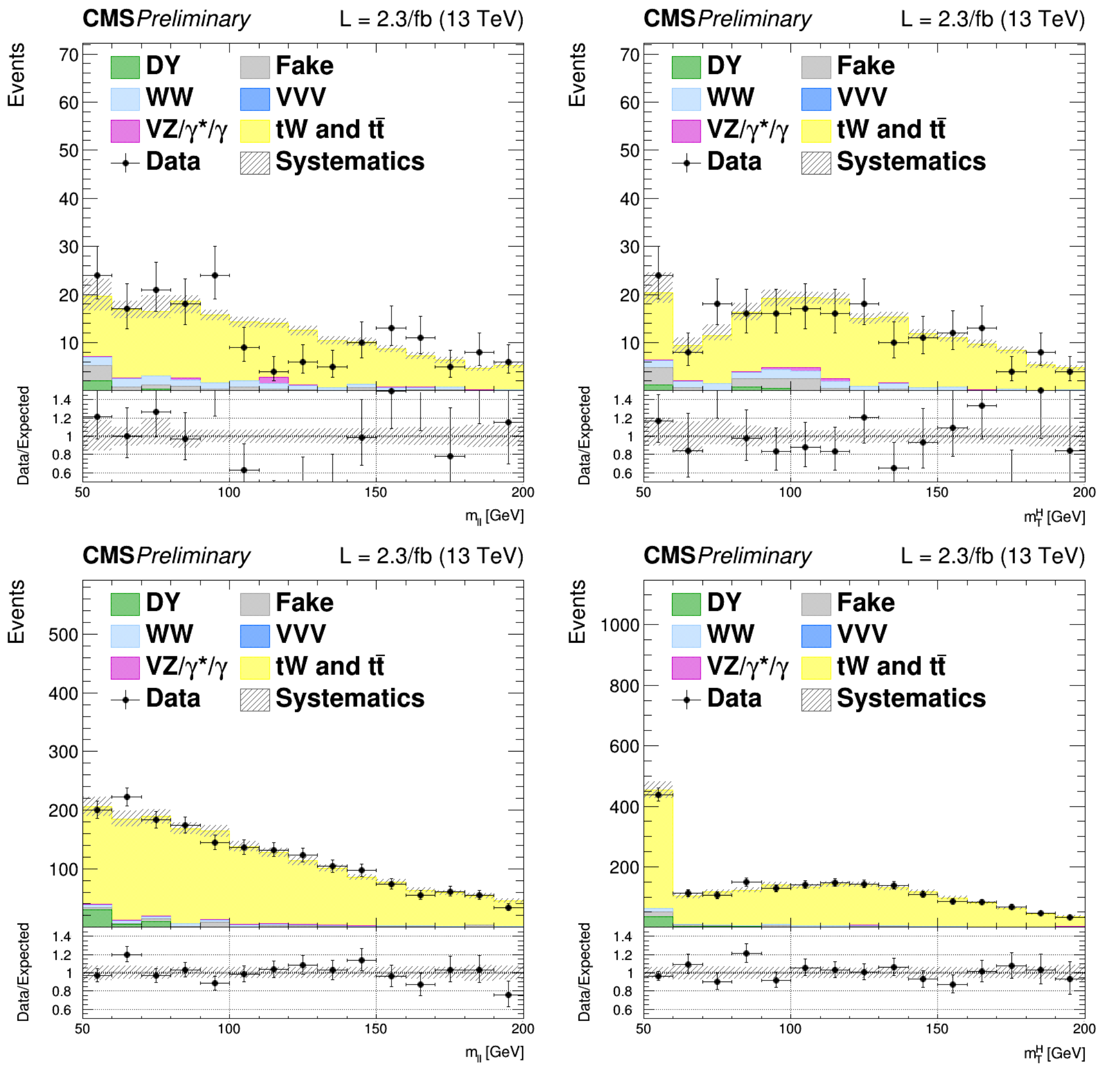
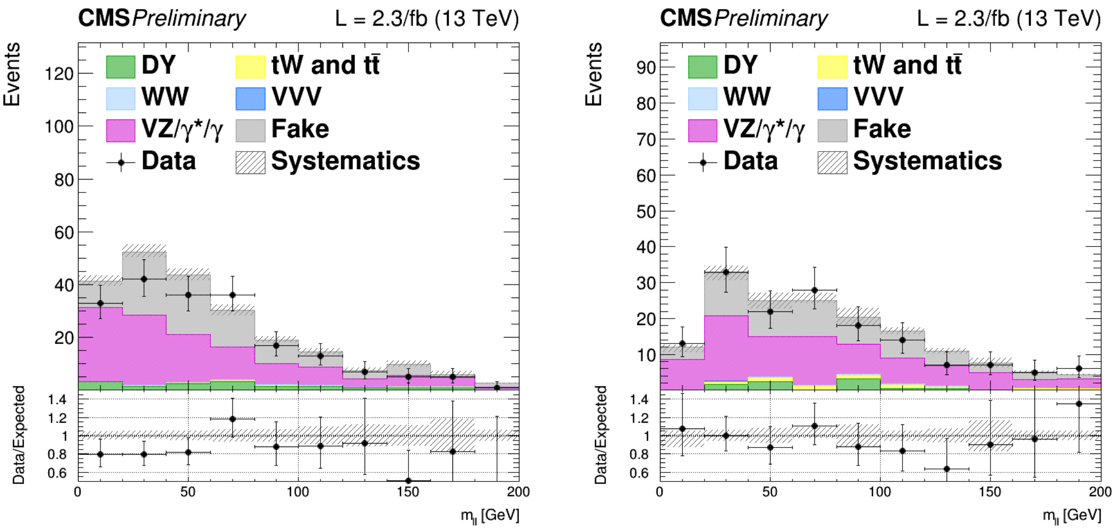


Figure 5.11.: Distributions of $m_{\ell\ell}$ (left) and m_{T} (right) for events with 0 jet (top) and 1 jet (bottom) in top enriched phase space. Scale factors estimated from data are applied. The first (last) bin includes underflows (overflows).

enriched in jet-induced events is defined. The events in the control sample are selected applying the WW baseline requirements but requesting an $e\mu$ pair with same charge, which significantly suppresses the WW and $t\bar{t}$ processes. The $m_{\ell\ell}$ distributions in this control region for the 0 and 1 jet categories are shown in Fig. 5.12. From the crosscheck in this control region, a global normalization factor of 0.8 is derived and applied to the jet-induced background.



(a) 0 jet (b) 1 jet

Figure 5.12.: Control plots for $m_{\ell\ell}$ in a fakes enriched phase space for events with 0 and 1 jet with $p_T > 30$ GeV, in $e\mu$ final state. Fake contribution has been scaled by 0.8 to match data.

5.4.4. DY background

This background contributes to the analysis phase space because of the Z/γ^* decays to a pair of τ leptons, which consequently decays to an $e\mu$ pair. This background process is predominant in the low m_T region, which is used as an orthogonal control region to determine the background normalization in the 0 and 1 jet categories separately. In particular this control region is defined by selecting events with $m_T < 60$ GeV and $30\,\text{GeV} < m_{\ell\ell} < 80$ GeV. The $m_{\ell\ell}$ distributions in these control regions for the 0 and 1 jet categories are shown in Fig. 5.13.

As for the top quark background, the normalization of this background in the 0 and 1 jet categories, is constrained directly in the fit by means of the control regions, which are treated as two additional categories.

The kinematics of this background is taken from simulation, after reweighting the Z boson p_T spectrum to match the observed distribution measured in data. In fact, this variable is not well reproduced by the MC generator used for simulating this process, especially in the bulk of the distribution, the discrepancy being ascribed to the missing contribution from resummed calculations.

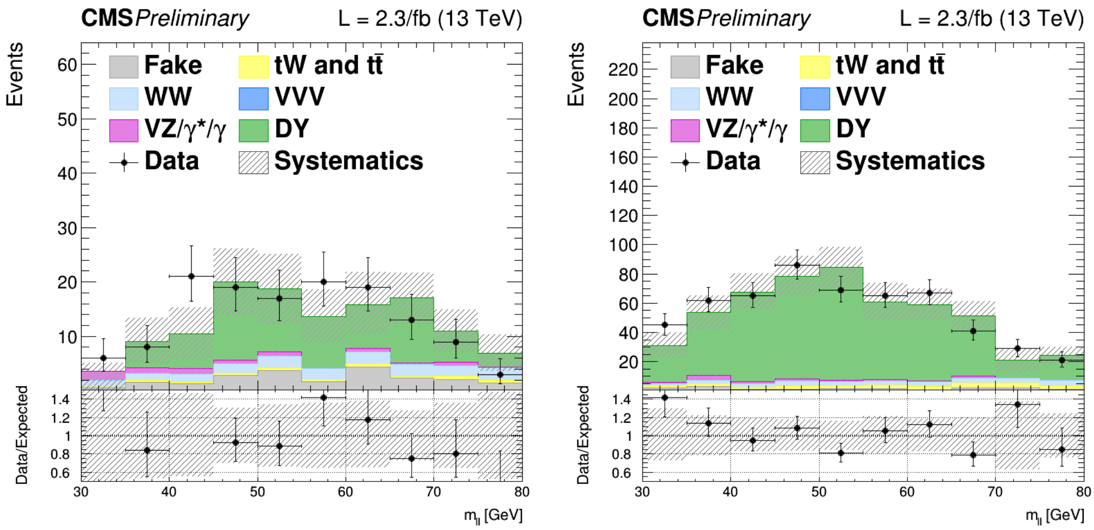


Figure 5.13.: Distributions of $m_{\ell\ell}$ for events with 0 jet (left) and 1 jet (right) in the DY→ $\tau\tau$ enriched control region. Scale factors estimated from data are applied.

5.4.5. Other backgrounds

The Wγ^* and the WZ electroweak processes can be gathered in the same physical process, although the final state kinematics is rather different. In particular, the invariant mass of the leptons arising from the γ^* decays is generally below 4 GeV, while the leptons from the Z boson decay are characterized by a larger invariant mass. Another background which can be experimentally identical to those is the Wγ production, where a real photon is produced in association with a W boson and consequently undergoes a photon conversion to leptons due to the interaction with the material constituting the first layers of the silicon tracker.

All these backgrounds may contribute to the signal phase space whenever one of the three leptons escape from the detector acceptance or is not identified. The shape and cross section of these backgrounds are taken from simulation. The only exception is the normalization of the Wγ^* background, being this process dominant in the low $m_{\ell\ell}$ region, which is scaled to data defining a proper control region. The control region is defined selecting events with three isolated muons, with $p_T > 10{,}5$ and 3 GeV for the first three leading muons respectively. The selection is further defined by $E_T^{miss} < 25$ GeV and E_T^{miss} projected to the leading muon < 45 GeV. The pair of muons with the smallest invariant mass is taken as coming from the γ^* decay. The k-factor measured in data for this background to be applied in the simulation is 1.98 ± 0.54.

All remaining backgrounds from di-boson and tri-boson production, which are of minor importance in the analysis phase space, are normalized according to their expected theoretical cross sections.

5.5. Systematic uncertainties

The systematic uncertainties affecting this measurement can be divided into three categories: the uncertainties on the background estimation, experimental uncertainties and theoretical uncertainties.

The first category includes the uncertainties related to the background normalization and shape. For the non-resonant WW production the shape is taken from simulation. The input normalization to the fit is set to the expected value from simulation, and an unconstrained nuisance parameter with a flat distribution is associated to this number. This is done separately for the two jet categories.

The top quark background shape is taken from simulation after correcting for the b tagging scale factors. An uncertainty due to these scale factors is included and affects both the normalization and the shape of the top quark background. The uncertainties on the normalization are treated similarly to the WW background case, but constraining the corresponding nuisances by means of the two control regions orthogonal to the signal phase space. A similar procedure is used for the DY background.

Effects due to experimental uncertainties are studied by applying a scaling and smearing of certain variables related to the physics objects, e.g. the p_{T} of the leptons, followed by a subsequent recalculation of all the correlated variables. This is done for simulation, to account for possible systematic mismodeling.

All experimental sources, except luminosity, are treated both as normalization and shape uncertainties, and are correlated among the signal and background processes and all the categories. The following experimental uncertainties are considered:

- the uncertainty determined by the CMS online luminosity monitoring, 2.7% for the first data collected at $\sqrt{s} = 13\,\mathrm{TeV}$;
- the acceptance uncertainty associated with the combination of single and double lepton triggers, which is 2%;
- the lepton reconstruction and identification efficiencies uncertainties, that are in the range 0.5-5% for electrons and 1-7% for muons depending on p_{T} and η;
- the muon momentum and electron energy scale and resolution uncertainties, that amount to 0.01-0.5% for electrons and 0.5-1.5% for muons depending on p_{T} and η;

- the jet energy scale uncertainties, that vary between 1-11% depending on the p_{T} and η of the jet;
- the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution uncertainty, that is taken into account by propagating the corresponding uncertainties on the leptons and jets;
- the scale factors correcting the b tagging efficiency and mistagging rate, that are varied within their uncertainties. This systematic uncertainty is anticorrelated between the top control regions and the other ones.

The uncertainties in the signal and background production rates due to theoretical uncertainties include several components, which are assumed to be independent: the PDFs and α_s, the underlying event and parton shower model, and the effect of missing higher-order corrections via variations of the renormalization and factorization scales.

The effects of the variation of PDFs, α_s and renormalization/factorization QCD scales, mainly affect the signal processes, being the most important backgrounds estimated using data driven techniques. However, the uncertainties on minor backgrounds that are estimated from simulation are taken into account. These uncertainties are split in the uncertainties on the cross section, which are computed by the LHC cross section working group [79], and on the selection efficiency [80]. The PDFs and α_s signal cross section normalization uncertainties are $^{+7.4\%}_{-7.9\%}$ and $^{+7.1\%}_{-6.0\%}$ for ggH and $\pm 0.7\%$ and $\pm 3.2\%$ for VBF Higgs production mechanism. The PDFs and α_s acceptance uncertainties are less than 1% for gluon- and quark-induced processes. The effect of the QCD scales variation on the selection efficiency is around 1-3% depending on the specific process. To estimate these uncertainties, the events are reweighted according to different QCD scales or different PDF sets and the selection efficiency is recomputed each time. For the QCD scale uncertainty the maximum variation with respect to the nominal value is taken as the uncertainty. For the case of PDF and α_s uncertainties, the distribution of the selection efficiency is built taking into account all the replicas in the NNPDF3.0 set and the uncertainty is estimated as the standard deviation of that distribution.

In addition, the categorization of events based on jet multiplicity introduces additional uncertainties on the ggH production mode related to missing higher order corrections. These uncertainties are evaluated following the prescription described in Sec. [**subsec:stewart-tackman**] and correspond to 5.6% for the 0-jet and 13% for the 1-jet bin categories.

The underlying event uncertainty is estimated by comparing two different PYTHIA 8 tunes, while parton shower modelling uncertainty is estimated by comparing samples interfaced with the PYTHIA 8 and HERWIG++ parton shower programs. The effect on the ggH (VBF) signal expected yield is about 5% (5%) for the PYTHIA 8 tune variation and about 7% (10%) for the parton shower description.

Other specific theoretical uncertainties are associated to some backgrounds. An uncertainty on the ratio of the $\mathrm{t\bar{t}}$ and tW cross sections is included. Indeed, these two processes are characterized by a different number of b-jets in the final state (2 b-jets for $t\bar{t}$ and 1 for tW) and the b-veto acts differently for the two. A variation of the relative ratio of the cross sections can thus cause a migration of events from the 0 to the 1 jet categories and viceversa. The corresponding uncertainty is of 8%, according to the theoretical cross section calculations [77, 76].

The gg $\to$ WW background LO cross section predicted by the MCFM generator is scaled to the NLO calculation, applying a k-factor of 1.4 with an uncertainty of 15% [75]. The interference term between the gg $\to$ WW and the ggH signal is also included and simulated with LO accuracy using MCFM. The k-factor to scale the interference term is 1.87, given by the geometrical average of the LO to NNLO gg$\to$H$\to$WW scale factor (2.5) and the LO to NLO gg $\to$ WW scale factor (1.4). The uncertainty on this value is estimated as the maximum variation with respect to the two scale factors mentioned above, and is found to be of 25%. Anyway, with the current amount of integrated luminosity, the interference contribution is found to be negligible.

For what the qq $\to$ WW background shape is concerned, an uncertainty related to the diboson p_{T} reweighting is evaluated varying the renormalization, factorization and resummation QCD scales.

Finally, the uncertainties due to the limited statistical accuracy of the MC simulations are also taken into account, including an independent uncertainty for each bin of the two-dimensional distribution, and for each category. The uncertainty for a certain bin and process is given by the standard deviation of the Poisson distribution with mean corresponding to the number of MC events in that bin.

5.6. Results

The expected and observed signal significance are shown in Table 5.6 for all the categories separately. Also, the observed signal strengths and the corresponding uncertainties are shown. The best fit signal strength obtained combining all the categories together is found to be $0.3^{+0.5}_{-0.5}$, corresponding to an observed significance of $0.7\,\sigma$, to be compared with the expected significance of $2.0\,\sigma$ for a Higgs boson mass of 125 GeV.

Maybe I should add the nuisances impact plots...

Table 5.6.: Observed and expected significance and signal strength the SM Higgs boson with a mass of 125 GeV for the 0-jet and 1-jet, μe and eμ, categories.

Category	Expected significance	Observed significance	σ/σ_{SM}
0-jet μe	1.1	1.3	$1.13\ ^{+0.9}_{-0.9}$
0-jet eμ	1.3	0.4	$0.33\ ^{+0.7}_{-0.7}$
1-jet μe	0.8	0	$-0.11^{+0.5}_{-1.7}$
1-jet eμ	0.9	0	$-0.54^{+1.4}_{-1.4}$
0-jet	1.6	1.3	$0.71^{+0.6}_{-0.5}$
1-jet	1.2	0	$-0.56^{+1.0}_{-1.0}$
Combination	2.0	0.7	$0.33^{+0.5}_{-0.5}$

Chapter 6.

Search for high mass resonances decaying to a W boson pair with first 13 TeV LHC data

6.1. Introduction

In this chapter, a search for a high mass spin-0 particle (from now on denoted as X) in the X$\rightarrow$WW$\rightarrow$ $\ell\nu\ell'\nu'$ decay channel is presented, where ℓ and ℓ' refer to an different flavour lepton pair, i.e. eμ. The search is based upon proton-proton collision data samples corresponding to an integrated luminosity of up to $2.3\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 13\,\mathrm{TeV}$, recorded by the CMS experiment at the LHC during 2015. This analysis represents a general extension of the SM Higgs boson search presented in 5 and is performed in a range of heavy scalar masses from $M_{\mathrm{X}} = 200\,\mathrm{GeV}$ up to 1 TeV, extending the range studied in a similar analysis performed using Run 1 LHC data [81], which provided upper limits on the production cross section of new scalar resonances up to 600 GeV.

Despite the discovery of a particle consistent with the SM Higgs boson in 2012, there is a possibility that this particle is only a part of a larger Higgs sector, and hence only partially responsible of the EW symmetry breaking. This can be achieved in different theoretical models that extends the SM, such as the two-Higgs-doublet models [82, 83, 84], or models in which the SM Higgs boson mixes with a heavy EW singlet, which predict the existence of an additional resonance at high mass, with couplings similar to those of the SM Higgs boson, as most recently described in [85, 86].

This analysis reports a generic search for a scalar particle with different resonance decay widths hypothesis, produced via the ggH and VBF production mechanisms. The results can then be interpreted in terms of different theoretical models. This analysis is heavily based on the SM Higgs search described in 5, in terms of physics objects, selections and background estimation. The differences and similarities are discussed in this chapter.

6.2. Data and simulated samples

The data sets, triggers, pile up reweighting, lepton identification and isolation used in this analysis are the same as the SM Higgs search and are described in Sec. 5.2.

Also, the same MC simulations are used for the background processes, the only exception being the DY background, for which the MG5_aMC@NLO generator is used with LO QCD accuracy, matching together events with up to four jets in addition to the vector boson with the MLM [87] matching scheme. Given that this analysis aims to probe regions of phase space where the DY contribution is very small, like in the high transverse mass region, the usage of a simulation of the inclusive DY process may lead to large uncertainties due to the limited simulation statistics in the sample. To partially overcome this issue, different DY samples are generated in restricted portions of the phase space defined by the H_{T} variable, i.e. the scalar sum of all the partons p_{T} in the event. For $H_{\mathrm{T}} < 100\,\mathrm{GeV}$ the inclusive simulation is used, while different samples are used for higher values of H_{T}. The samples are merged using the parton level information, and it has been verified that a smooth transition between different H_{T} regions is achieved, as shown in Fig. 6.1. The DY LO cross section obtained from the simulation is scaled using the LO to NNLO k-factor of 1.23.

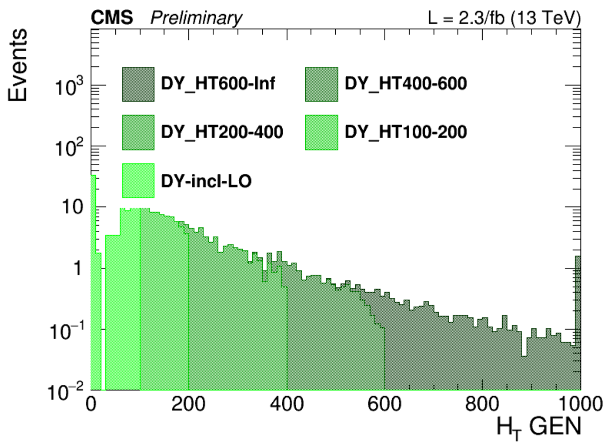


Figure 6.1.: Generator level H_{T} distribution for the merged DY sample.

In order to perform the resonance search in a large part of the mass spectrum, several signal samples for the gluon-gluon fusion and the vector boson fusion mechanisms have been generated corresponding to different Higgs boson masses in the range between 200 GeV and 1 TeV. The signal width for each mass point corresponds to the one expected for a SM Higgs boson at that mass. The samples are produced with a mass step of 50 GeV from 250 to 800 GeV and of 100 GeV from 800 to 1000 GeV. A finer stepping is used between 200 and 250 GeV. All the signal samples are generated with the POWHEG V2 generator, interfaced with the JHUGEN v6.2.8 generator, which handles the decay of the Higgs boson to $\mathrm{W^+W^-} \to 2\ell 2\nu$.

The interference effects among gg→X→WW, gg→WW and gg→H→WW are evaluated using the MCFM and JHUGen generators, as implemented in the MELA (Matrix Element Likelihood Approach) framework [62]. Details about the interference effects are given in Sec. 6.3.

6.3. Analysis strategy

The analysis strategy for the first results on the high mass search in the $W^+W^- \to 2\ell 2\nu$ decay channel closely follows the strategy presented in the 13 TeV SM Higgs search in the $H \to W^+W^- \to 2\ell 2\nu$ channel regarding the 0 and 1 jet categories. In addition a dedicated category to the VBF production mechanism is added, given that this production mode is particularly important in the high mass region. Indeed, assuming a SM Higgs boson, the ratio of cross sections $\sigma_{\mathrm{VBF}}/\sigma_{\mathrm{ggH}}$[1] increases with the Higgs boson mass, making the VBF production mechanism more and more important as the mass of the resonance approaches to high values.

This analysis is affected essentially by the same background processes as the SM Higgs boson search, with the difference that in this case the SM Higgs boson processes, including all production modes, are treated as backgrounds.

In addition to requiring the events to pass the single or double lepton triggers, exactly one electron and one muon are required to be reconstructed in the event with opposite charges and a minimum p_{T} of 20 GeV for both the muon and electron. Both leptons are required to be well identified and isolated to reject fake leptons and leptons coming from decays in flight. To suppress background processes with three or more leptons in the final state, such as diboson or triboson production, events with any additional identified and isolated lepton with $p_{\mathrm{T}} > 10$ GeV are rejected. To suppress the contribution of the SM production of the Higgs boson at 125 GeV, $m_{\ell\ell}$ is requested to be higher than 50 GeV. The other event requirements are identical to the 125 GeV Higgs boson search and are described in Sec. 5.3.2.

In addition to the 0 and 1 jet categories, a specific category sensitive to the VBF production mode is defined exploiting the characteristic signature of this process, where two energetic jets are emitted in the forward region of the detector and with large $\Delta\eta$ gap. Events belonging to the VBF-enriched category are selected by requiring at least two jets with $p_{\mathrm{T}} > 30$ GeV, an invariant mass $m_{\mathrm{jj}} > 500$ GeV and a gap in pseudorapidity $\Delta\eta_{\mathrm{jj}} > 3.5$.

In addition to the transverse mass variable m_{T}, which is used in the analysis selection to define the DY background control region, an additional variable is defined, that from

[1] The ggH notation is used for the gluon-gluon fusion production mode, even in the cases where a non-SM Higgs boson is created in the process.

now on will be labelled as "improved transverse mass" $m_{\mathrm{T}}^{\mathrm{i}}$. This variable is defined as the invariant mass of the four momentum resulting from the sum of the two leptons four momenta $(p_{\ell\ell}, \vec{p}_{\ell\ell})$ and four momentum $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}} = (E_{\mathrm{T}}^{\mathrm{miss}}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$, i.e.:

$$m_{\mathrm{T}}^{\mathrm{i}} = \sqrt{(p_{\ell\ell} + E_{\mathrm{T}}^{\mathrm{miss}})^2 - (\vec{p}_{\ell\ell} + \vec{p}_{\mathrm{T}}^{\mathrm{miss}})^2} \quad . \tag{6.1}$$

This variable allows having a better sensitivity to different resonance mass hypothesis as shown in Fig. 6.2, where the shape of the $m_{\mathrm{T}}^{\mathrm{i}}$ variable is shown for different SM Higgs mass hypothesis and it is compared to the standard m_{T} variable. The usage of this variable also provide a good discriminating power between signal and background, which depends on the particular signal mass hypothesis.

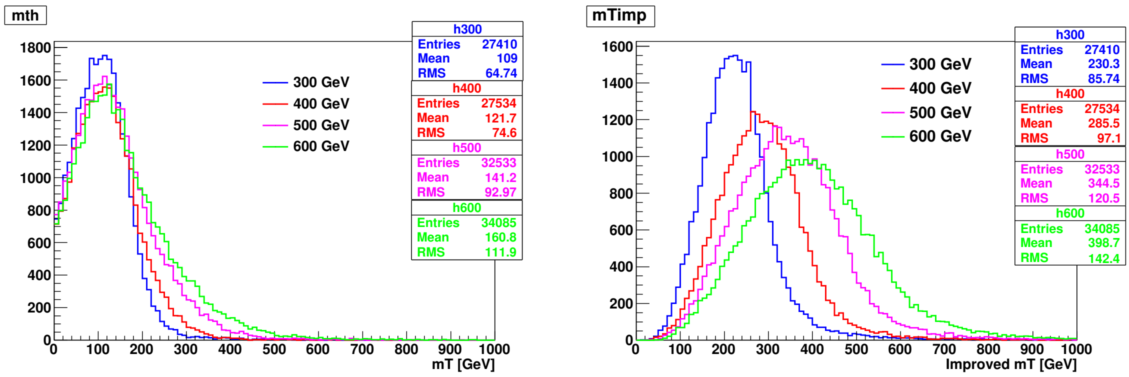


Figure 6.2.: Distribution of the m_{T} and $m_{\mathrm{T}}^{\mathrm{i}}$ variables at generator level for different resonance mass hypothesis.

The signal extraction is based on a binned maximum likelihood fit using the $m_{\mathrm{T}}^{\mathrm{i}}$ distribution for signal and background contributions as templates. The $m_{\mathrm{T}}^{\mathrm{i}}$ template is defined using the following bin boundaries:

- **0/1 jet:** [100,150,200,250,300,350,400,450,500,600,700,1000] ,
- **VBF:** [100,150,200,250,300,350,400,500,700,1000] ,

where the first number represents the lower edge of the first bin while the other numbers represent the upper edges. The last bin is an overflow bin.

In order to test different resonance decay widths hypotheses, the signal samples, which are generated with a decay width corresponding to the expected value for a SM Higgs boson at that mass (Γ_{SM}), are reweighted to obtain the desired width value (Γ'). In particular the following values are used: $\Gamma' = \Gamma_{\mathrm{SM}}$, $\Gamma' = 0.49 \times \Gamma_{\mathrm{SM}}$, $\Gamma' = 0.25 \times \Gamma_{\mathrm{SM}}$ and $\Gamma' = 0.09 \times \Gamma_{\mathrm{SM}}$. The reweighting is performed at generator level by computing the ratio of two relativistic

Breit Wigner distributions with different decay widths, $f(E, \Gamma', M_{\mathrm{X}})/f(E, \Gamma_{\mathrm{SM}}, M_{\mathrm{X}})$, where:

$$f(E) \propto \frac{1}{(E^2 - M^2)^2 + M^2\Gamma^2} \quad . \tag{6.2}$$

Here, $f(E, \Gamma_{\mathrm{SM}}, M_{\mathrm{X}})$ represents the distribution used for the simulation of the signal at a mass M_{X}, and $f(E, \Gamma', M_{\mathrm{X}})$ the distribution with the new decay width. Each event is multiplied by this ratio (which depends on the energy E of the event) to obtain the reweighted distribution.

When a resonance with a non negligible width is considered, it is important to take into account the interference effects both with the gg→WW background and the SM Higgs boson off-shell tail. A study of the interference effects for a resonance X produced through the gluon fusion mechanism is performed within the MCFM+JHUGen framework, and including NNLO corrections for cross section using HNNLO program [88] based on MCFM. The matrix element package MELA supports all of these processes and allows fast MC re-weighting and optimal discriminant calculation. The basic idea of this approach is to compute the matrix elements of the processes under study with the MCFM and JHUGen generators, including the interference terms, and using these matrix elements to compute an event weight used to reweight the simulated samples. Using this approach the simulated events can be reweighted according to different scenarios, for instance including some or all the interference terms, allowing a detailed study of the interference contribution. The effect of the various interference terms for the M_{X} variable at generator level is shown in Fig. 6.3, after having applied the WW baseline selections. As can be observed the contribution of the interference of the scalar resonance with the gg→WW background and with SM Higgs boson have opposite sign and partially cancel out. This cancellation effect is different for different resonance masses and depends on the event selection. In particular the interference term with the SM Higgs off-shell tail is positive for values below M_{X} while it turns negative above M_{X}. The contribution of the interference with the gg→WW background is instead characterized by an opposite sign lineshape, thus leading to a partial cancellation when considering the total interference.

The effect of the resulting interference contribution including all the different terms is shown in Fig. 6.4 for the $m_{\mathrm{T}}^{\mathrm{i}}$ signal templates, in the three categories separately and for different M_{X} hypotheses.

The interference contribution is thus not negligible, especially for large values of M_{X}, and is included in the analysis as part of the signal contribution. More specifically, during the fit procedure the signal yield is scaled by the signal strength parameter μ (which is the parameter of interest of the fit), while the interference yield is scaled by $\sqrt{\mu}$.

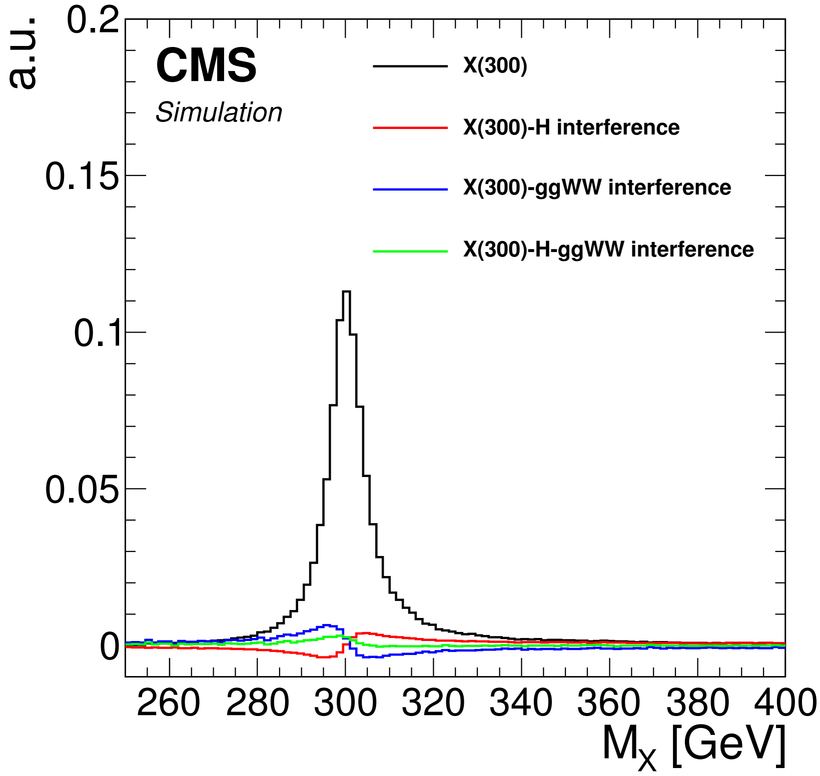


Figure 6.3.: Distribution of the M_X variable for a resonance mass of 300 GeV, showing the various interference terms after the WW baseline selections.

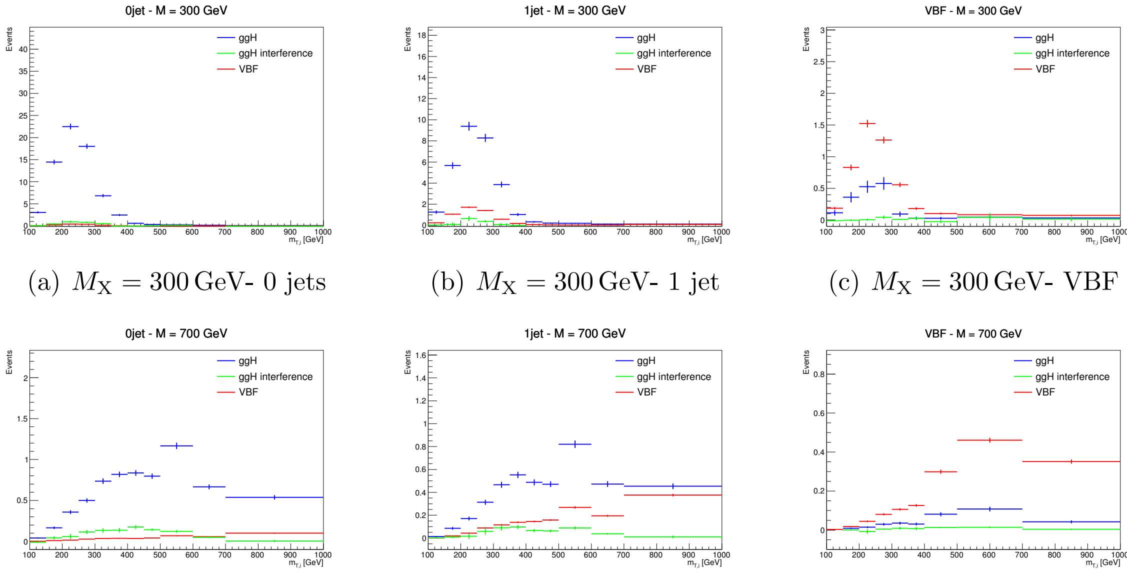
6.4. Background estimation

The background processes affecting the analysis phase space are the same as the ones contributing to the SM Higgs search described in Sec. 5.4. The techniques used for the background estimation are the same as well.

The most relevant difference is the addition of the 2 jets category. The WW, top quark and DY background normalizations are estimated in this category using data driven techniques, similarly to the other jet bins.

Given the slightly different WW baseline selection with respect to the SM Higgs search, also the control regions for the top quark and DY backgrounds estimation change, while the WW background normalization is estimated from data in the three signal regions separately, owing to the different m_T^i shapes for signal and background.

For the estimation of the top quark background, three control regions enriched in b-jets are defined by selecting events that pass the WW baseline selections and applying a b tagging requirement which depends on the jet category as follows:



(a) $M_X = 300\,\text{GeV}$- 0 jets (b) $M_X = 300\,\text{GeV}$- 1 jet (c) $M_X = 300\,\text{GeV}$- VBF

(d) $M_X = 700\,\text{GeV}$- 0 jets (e) $M_X = 700\,\text{GeV}$- 1 jet (f) $M_X = 700\,\text{GeV}$- VBF

Figure 6.4.: Distributions of the m_T^i variable for M_X =300 and 700 GeV, showing the signal (both the ggH and VBF processes) and the interference contributions in the three jet categories.

- 0 jets category: at least one b-tagged jet with $20 < p_T < 30\,\text{GeV}$ is required;
- 1 jet category: exactly one b-tagged jet with p_T above 30 GeV is required;
- 2 jets category: at least one b-tagged jet with p_T above 30 GeV is required.

Distributions of the m_T^i variable in the 0 jets, 1 jet and 2 jets top quark enriched control regions after applying the data driven estimation are shown in Fig. 6.5.

The jet induced background, also labelled as “non-prompt” background, so as to highlight that these events do not contain prompt leptons, is estimated using the same fake rate method described in 4.4.3. A crosscheck is performed selecting events passing the WW baseline selection but with an eμ pair with same charge. The m_T^i distributions for this phase space are shown in Fig. 6.6 for the three jet categories separately, showing agreement between data and simulation within the uncertainties.

Due to the cuts on the leptons p_T and on $m_{\ell\ell}$ in the WW baseline requirements, the contribution of the DY background decaying to a pair of τ leptons is very small in the signal regions, especially in the VBF phase space. The normalization of this background is estimated from a control region in data, defined in the same way as explained in 5.4.4, for the 0 and 1 jet categories. In the VBF category, given the very small number of expected events, the normalization of this background is taken from simulation.

Other minor background processes are estimated as described in 5.4.5.

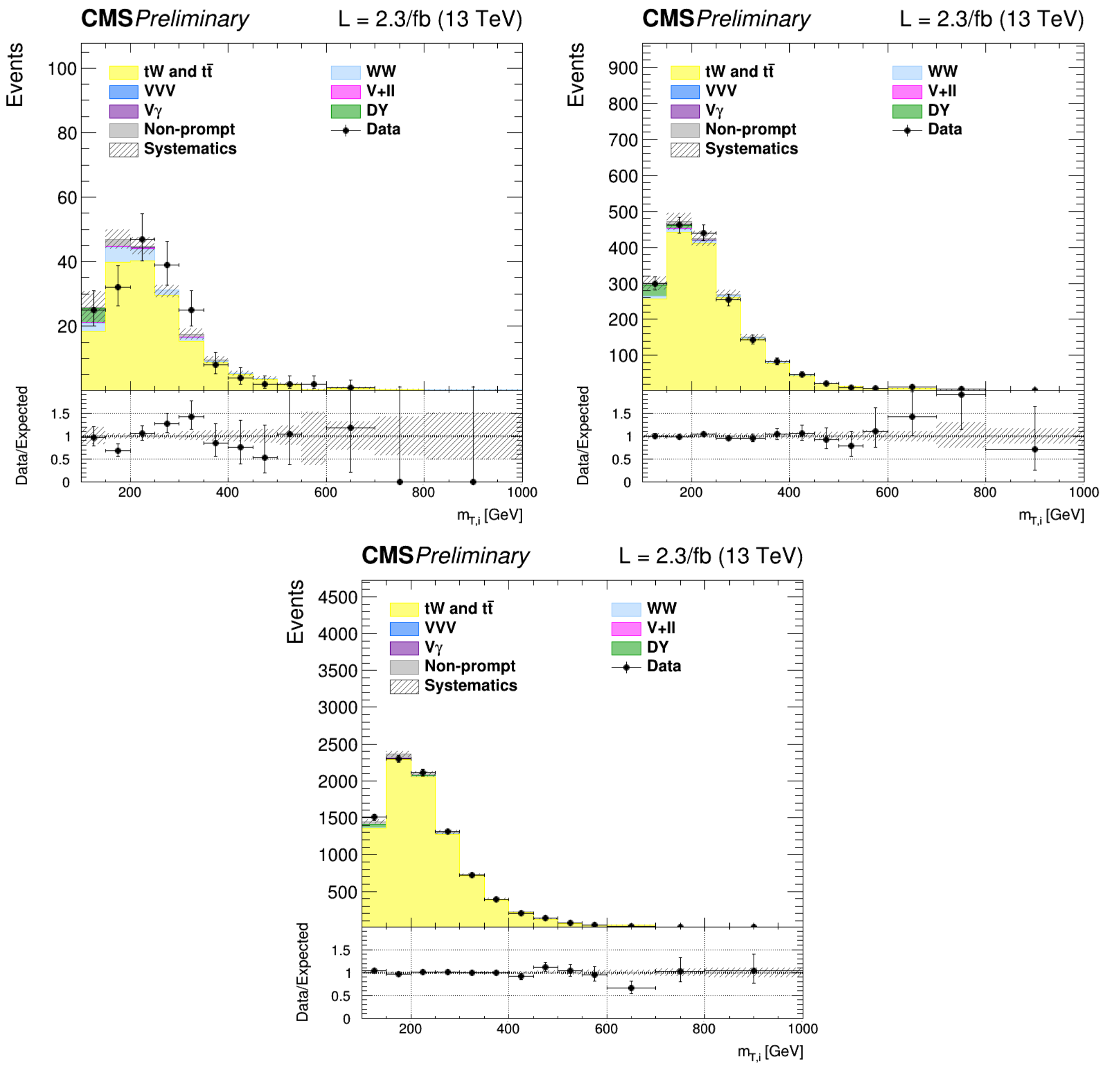


Figure 6.5.: Distributions of $m_{\mathrm{T}}^{\mathrm{i}}$ for events with 0 jet (top left), 1 jet (top right) and 2 jets (bottom) in top enriched control region. Scale factors estimated from data are not applied in the plots.

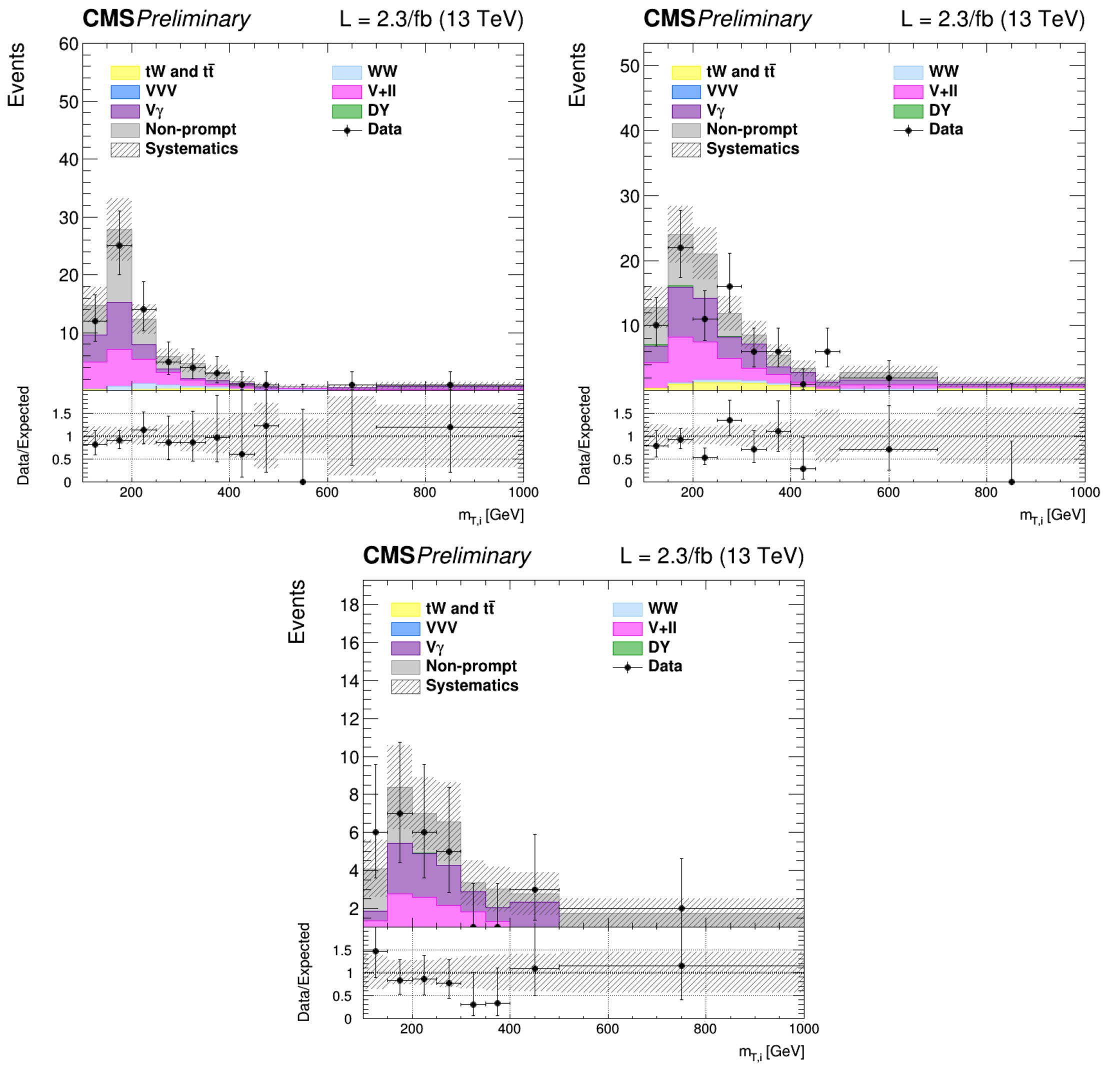


Figure 6.6.: Distributions of $m_{\mathrm{T}}^{\mathrm{i}}$ for events with 0 jet (top left), 1 jet (top right) and 2 jets (bottom) in the same-charge dilepton control region. The last bin of the histograms includes overflows.

6.5. Systematic uncertainties

The systematic uncertainties affecting this analasys are the same discussed in Sec. 5.5. The differences with respect to the SM Higgs boson search are described below.

The PDF and α_s uncertainties on the signal cross sections are taken from the computations performed by the LHC cross section working group [79], and are included for all the mass points. The value of these uncertainties depends on the resonance mass and vary from 3 and 5% for ggH and from 2 and 3% for VBF production modes. The PDFs and α_s uncertainties on the signal selection are evaluated for every resonance mass and are found to be less than 1% for both ggH and VBF.

The theoretical uncertainties in the signal yields due to the jet categorization are evaluated for all the ggH signals following the prescription described in Sec. 4.5.3.

An additional uncertainty on the modelling of the top pair background is derived from the observed discrepancy between data and Powheg V2 plus Pythia 8.1 simulation on the top quark p_{T} spectrum [89], which is particularly important in the tail of the $m_{\mathrm{T}}^{\mathrm{i}}$ distribution. Another uncertainty affecting the $m_{\mathrm{T}}^{\mathrm{i}}$ tail for the top quark background is the parton shower uncertainty. This is evaluated comparing the generator level $m_{\mathrm{T}}^{\mathrm{i}}$ distributions corresponding to two different simulations of the $\mathrm{t\bar{t}}$ process: one obtained using Pythia 8.1 for the showering and hadronization of the simulated events, and the other using Herwig++. The difference between the two is used to extract a shape uncertainty, which is less than 1% for low $m_{\mathrm{T}}^{\mathrm{i}}$ values and reaches about 6% in the $m_{\mathrm{T}}^{\mathrm{i}}$ tail.

6.6. Signal extraction and limit setting

The signal yield, including both the ggH and VBF production modes, is extracted performing a combined fit of the three categories to the $m_{\mathrm{T}}^{\mathrm{i}}$ simulation templates for backgrounds and signal, and is repeated for each resonance mass hypothesis. Moreover, fixed the mass of the resonance, the fit is performed again for the various hypotheses of the resonance decay width. A single signal strength μ is extracted from the fit, which multiplies both the ggH and VBF contributions. In other words it is assumed that the ratio of the two production mechanism stays the same as the one predicted by the SM[2].

The background yields expected from simulation corresponding to the three jet categories and after the analysis event selection are shown in Table 6.1. The signal yields corresponding to a selection of mass points and assuming $\Gamma' = \Gamma_{\mathrm{SM}}$ are shown in Table 6.2.

[2] This is an approximation which limits the amount of models that can be tested with the provided results. A future development od this analysis will also include the cases for which different ggH and VBF relative contributions are expected.

Table 6.1.: Expected yields estimated from simulation (except for the non-prompt contribution which is estimated using data) for each background process in the three analysis categories, after the analysis event selection. The uncertainties are shown for the processes estimated from simulation.

Background process	0 jets	1 jet	VBF
qq→WW	501.93 ± 0.00 (0%)	198.72 ± 0.00 (0%)	4.54 ± 0.00 (0%)
gg→WW	37.28 ± 5.77 (15%)	19.63 ± 3.04 (15%)	1.05 ± 0.16 (15%)
Top quark	188.75 ± 0.00 (0%)	330.05 ± 0.00 (0%)	25.06 ± 0.00 (0%)
DY	33.24 ± 0.00 (0%)	12.99 ± 0.00 (0%)	0.28 ± 0.00 (0%)
Non-prompt	64.21 ± 19.26 (30%)	31.69 ± 9.51 (30%)	2.10 ± 0.63 (30%)
Vγ	26.62 ± 0.72 (3%)	14.18 ± 0.38 (3%)	0.64 ± 0.02 (3%)
Vγ^*	4.44 ± 1.12 (25%)	3.39 ± 0.85 (25%)	0.14 ± 0.04 (25%)
VZ	13.51 ± 0.76 (6%)	11.67 ± 0.66 (6%)	0.28 ± 0.02 (6%)
VVV	0.01 ± 0.00 (3%)	0.02 ± 0.00 (3%)	0.00 ± 0.00 (3%)
SM H→WW	6.04 ± 0.40 (7%)	3.10 ± 0.11 (5%)	0.34 ± 0.02 (7%)
SM H→ $\tau\tau$	0.50 ± 0.05 (9%)	0.43 ± 0.04 (9%)	0.04 ± 0.00 (9%)
Total background	876.5	625.9	34.5

The strategy for computing the exclusion limits is based on the modified frequentist approach, also referred to as $\mathrm{CL_s}$, as described in [78]. The first step is to construct the likelihood function $\mathcal{L}(data|\mu,\theta)$:

$$\mathcal{L}(data|\mu,\theta) = Poisson(data|\mu\cdot s(\theta)+b(\theta))\cdot p(\tilde{\theta}|\theta) \quad , \tag{6.3}$$

where $data$ represents the experimental observation, s and b are the expected signal and background yields respectively and θ is the full set of nuisance parameters constrained by the prior distribution functions $p(\tilde{\theta}|\theta)$. The default values for the nuisance parameters are labelled as $\tilde{\theta}$.

For a binned shape analysis, $Poisson(data|\mu\cdot s+b)$ is the product of the Poisson probabilities to observe n_i events in bin i:

$$\prod_i \frac{(\mu\cdot s_i+b_i)^{n_i}}{n_i!}e^{-\mu\cdot s_i-b_i} \quad . \tag{6.4}$$

In order to test the compatibility of the data with the signal plus background (or the background only) hypothesis, the test statistic $\tilde{q}_\mu$ is constructed based on the profile

Table 6.2.: Expected signal yields for the ggH and VBF production modes estimated from simulation after the analysis event selection for different mass hypothesis assuming $\Gamma' = \Gamma_{\rm SM}$ in the three analysis categories. The errors correspond to the theoretical uncertainties in the signal estimation.

Mass (GeV)	0 jets	1 jet	VBF
	ggH signal yields		
200	90.21 ± 6.67 (7%)	37.47 ± 1.81 (5%)	1.25 ± 0.26 (21%)
400	66.35 ± 4.90 (7%)	32.65 ± 1.57 (5%)	2.04 ± 0.42 (21%)
600	13.86 ± 1.05 (8%)	8.56 ± 0.44 (5%)	0.68 ± 0.14 (21%)
800	3.20 ± 0.25 (8%)	2.32 ± 0.13 (6%)	0.22 ± 0.05 (21%)
1000	0.88 ± 0.07 (8%)	0.70 ± 0.04 (6%)	0.07 ± 0.02 (21%)
	VBF signal yields		
200	1.54 ± 0.06 (4%)	6.18 ± 0.25 (4%)	5.05 ± 0.20 (4%)
400	0.91 ± 0.04 (4%)	3.42 ± 0.14 (4%)	3.19 ± 0.13 (4%)
600	0.50 ± 0.02 (4%)	1.95 ± 0.08 (4%)	1.88 ± 0.08 (4%)
800	0.33 ± 0.01 (4%)	1.21 ± 0.05 (4%)	1.16 ± 0.05 (4%)
1000	0.22 ± 0.01 (4%)	0.79 ± 0.03 (4%)	0.69 ± 0.03 (4%)

likelihood reatio:

$$\tilde{q}_\mu = -2\ln\frac{\mathcal{L}(data|\mu,\hat{\theta}_\mu)}{\mathcal{L}(data|\hat{\mu},\hat{\theta})} \quad \text{with} \quad 0 \leq \hat{\mu} \leq \mu \quad , \tag{6.5}$$

where $\hat{\theta}_\mu$ refers to the conditional maximum likelihood estimators of θ, given the signal strength μ. The parameter estimators $\hat{\mu}$ and $\hat{\theta}$ correspond to the global maximum of the likelihood. The $0 \leq \hat{\mu}$ constraint is imposed to have a positive signal yield, e.g. background underfluctuations are forbidden, while $\hat{\mu} \leq \mu$ is imposed to have a one-sided confidence interval. The observed test statistic for the signal strength μ under test is referred to as $\tilde{q}_\mu^{\rm obs}$. The values of the nuisance parameters obtained maximising the likelihood function are labelled as $\hat{\theta}_0^{\rm obs}$ and $\hat{\theta}_\mu^{\rm obs}$ for the background only and signal plus background hypotheses, respectively. The pdf of the test statistic in constructed by generating toy MC pseudo-data for both the background only and signal plus background hypotheses, i.e. $f(\tilde{q}_\mu|\mu,\hat{\theta}_\mu^{\rm obs})$ and $f(\tilde{q}_\mu|0,\hat{\theta}_0^{\rm obs})$. These distributions can be used to define two p-values corresponding to the two hypotheses, p_μ and p_b:

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}}|\text{signal} + \text{background}) = \int_{\tilde{q}_\mu^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}})d\tilde{q}_\mu \quad , \tag{6.6}$$

$$1 - p_b = (\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}}|\text{background only}) = \int_{\tilde{q}_0^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}})d\tilde{q}_\mu \quad . \tag{6.7}$$

According to these definitions, p_μ and p_b can be identified with $\text{CL}_{\text{s+b}}$ and $1 - \text{CL}_{\text{b}}$, respectively. The $\text{CL}_{\text{s}}(\mu)$ is calculated using the following ratio:

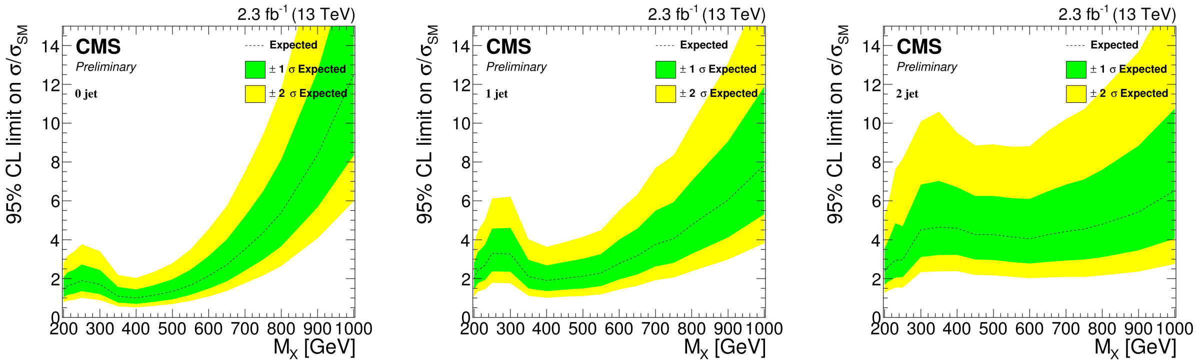
$$\text{CL}_{\text{s}}(\mu) = \frac{\text{CL}_{\text{s+b}}}{\text{CL}_{\text{b}}} = \frac{p_\mu}{1 - p_b} \quad . \tag{6.8}$$

If, for a given signal strength μ, $\text{CL}_{\text{s}} \leq \alpha$, then the hypothesis is excluded with a $(1 - \alpha)$ confidence level (CL). For instance, if one wants to quote the upper limit on μ with a 95% CL, the signal strength has to be adjusted until $\text{CL}_{\text{s}} = 0.05$.

The expected median upper limit, as well as the $\pm 1\sigma$ (68% CL) and $\pm 2\sigma$ (95% CL) bands, are determined generating a large amount of pseudo-data in the background only hypothesis and calculating CL_{s} and the 95% CL upper limit for each of them, as if they were real data. Then the cumulative distribution of the 95% CL upper limits is built and the median expected value is identified as the value at which the cumulative distribution crosses the 50% quantile. The $\pm 1\sigma$ ($\pm 2\sigma$) band is defined by the values at which the cumulative distribution crosses the 16% (2.5%) and 84% (97.5%) quantiles.

In order to assess the sensitivity of the analysis, the expected upper exclusion limits at 95% CL on the signal strength are shown in Fig. 6.7 for the three jet categories separately. For a given mass of the resonance, the limits are derived assuming a signal decay width $\Gamma' = \Gamma_{\text{SM}}$ and a cross section equal to the one expected from a SM Higgs boson at that mass. The other decay width hypothesis have also been tested, showing a very similar expected exclusion limit, suggesting that this analysis is not strongly sensitive to variations of the resonance decay width.

The 0 jets category is the most sensitive especially in the low mass region, while for very large masses of the resonance the 1 jet and VBF categories start being important. This is explained mainly by the fact that the VBF contribution increases, with respect to ggH, as the mass increases. The expected exclusion limit on the signal strength after the combination of the three categories is shown in Fig. 6.8. Comparing the limits in the single categories with the combination of the three it is evident how the higher jet multiplicity categories help in improving the results for large values of M_{X}.



(a) 0 jets (b) 1 jet (c) 2 jets

Figure 6.7.: Expected exclusion upper limits at 95% CL on the signal strength in the three categories, as a function of the resonance mass. The dashed line corresponds to median upper limit, while the green and yellow regions represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands, respectively. Limits are derived assuming a SM Higgs boson cross section and decay width for each mass point.

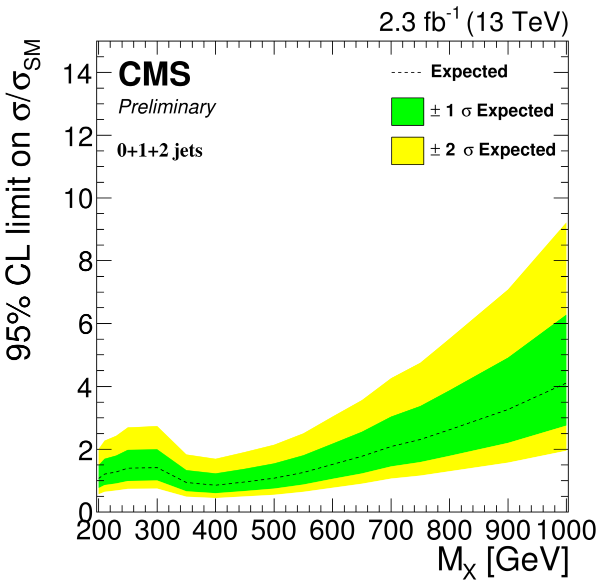


Figure 6.8.: Expected exclusion upper limit at 95% CL on the signal strength for the combination of the three categories, as a function of the resonance mass. The dashed line corresponds to median upper limit, while the green and yellow regions represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands, respectively. The limit is derived assuming a SM Higgs boson cross section and decay width for each mass point.

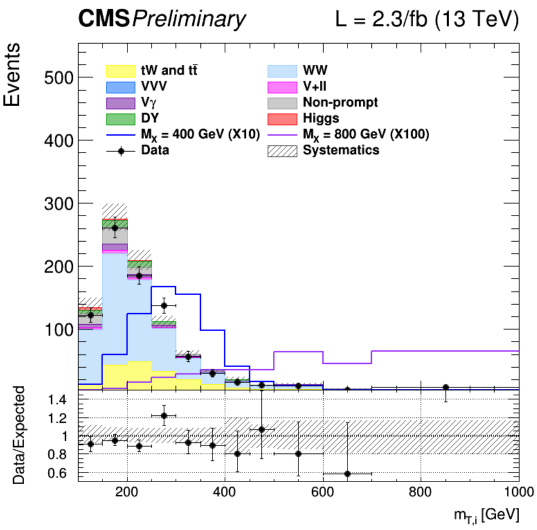
6.7. Results

The $m_{\mathrm{T}}^{\mathrm{i}}$ distributions for the signal region after the full analysis selection are shown in Fig. 6.9 for the three jet categories. Two different signal hypotheses corresponding to $M_{\mathrm{X}} = 400\,\mathrm{GeV}$ and $M_{\mathrm{X}} = 800\,\mathrm{GeV}$ are shown superimposed on the background for comparison.

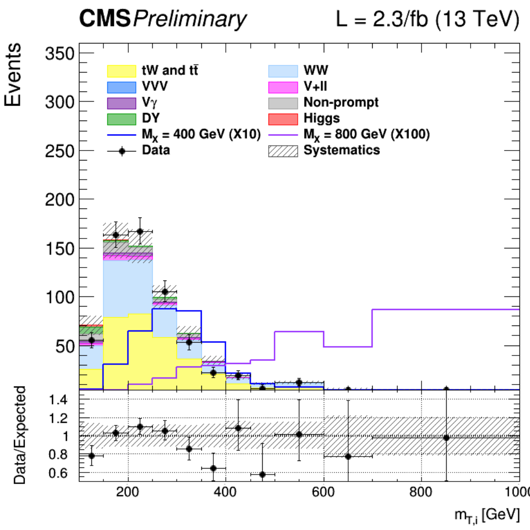
For every mass point from 200 GeV up to 1 TeV the observed p-value and the 95% CL upper exclusion limit are calculated for for five hypothesis of the signal width. The observed p-value as a function of the resonance mass for the combination of the three jet categories is shown in Table 6.3.

Table 6.3.: Observed p-value and corresponding significance (set to 0 in case of underfluctuations of the observed number of events) for the combination of the three jet categories for different resonance masses. Different values of the signal width are shown.

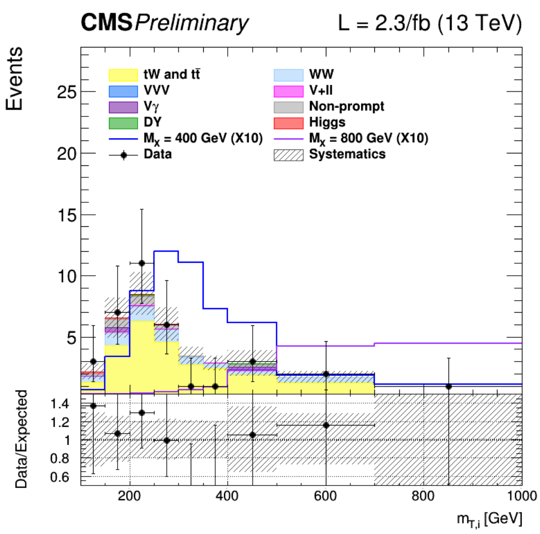
Mass [GeV]	$\Gamma = 0.09 \times \Gamma_{SM}$ p-value (signif.)	$\Gamma = 0.25 \times \Gamma_{SM}$ p-value (signif.)	$\Gamma = 0.49 \times \Gamma_{SM}$ p-value (signif.)	$\Gamma = \Gamma_{SM}$ p-value (signif.)
200	0.50 (0)	0.50 (0)	0.50 (0)	0.56 (0)
210	0.58 (0)	0.45 (0.1)	0.35 (0.4)	0.24 (0.7)
230	0.21 (0.8)	0.22 (0.8)	0.23 (0.7)	0.26 (0.6)
250	0.29 (0.5)	0.20 (0.8)	0.15 (1.0)	0.12 (1.2)
300	0.014 (2.2)	0.015 (2.2)	0.016 (2.1)	0.018 (2.1)
350	0.16 (1.0)	0.17 (1.0)	0.18 (0.9)	0.23 (0.7)
400	0.50 (0)	0.49 (0)	0.49 (0)	0.57 (0)
450	0.51 (0)	0.50 (0)	0.50 (0)	0.52 (0)
500	0.50 (0)	0.51 (0)	0.50 (0)	0.52 (0)
550	0.50 (0)	0.51 (0)	0.51 (0)	0.51 (0)
600	0.50 (0)	0.50 (0)	0.51 (0)	0.51 (0)
650	0.50 (0)	0.50 (0)	0.54 (0)	0.50 (0)
700	0.50 (0)	0.50 (0)	0.50 (0)	0.50 (0)
750	0.50 (0)	0.54 (0)	0.50 (0)	0.40 (0.3)
800	0.50 (0)	0.55 (0)	0.39 (0.3)	0.29 (0.6)
900	0.29 (0.6)	0.27 (0.6)	0.24 (0.7)	0.22 (0.8)
1000	0.18 (0.9)	0.18 (0.9)	0.18 (0.9)	0.18 (0.9)



(a) 0 jets



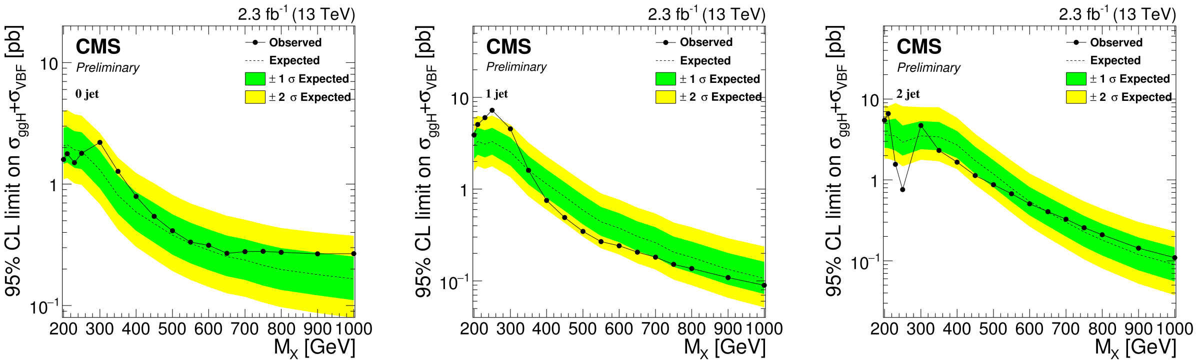
(b) 1 jet



(c) VBF

Figure 6.9.: Distributions of $m_{\mathrm{T}}^{\mathrm{i}}$ in the signal region for the 0 jets, 1 jet and VBF categories. Background normalisations correspond to the pre-fit value. Signal contributions for two mass hypotheses, $M_{\mathrm{X}} = 400\,\mathrm{GeV}$ and $M_{\mathrm{X}} = 800\,\mathrm{GeV}$, are shown superimposed on the background and scaled to facilitate the comparison.

In order to be independent on the particular model assumed for the signal cross section, the results are interpreted as exclusion limits on $\sigma \times \mathcal{B}$, where σ stands for the sum of the ggH and VBF cross sections, and $\mathcal{B}$ represents the X$\rightarrow$WW$\rightarrow 2\ell 2\nu$ branching ratio including all lepton flavours. The expected and observed upper exclusion limits on $\sigma \times \mathcal{B}$ for $\Gamma' = \Gamma_{\mathrm{SM}}$ are shown in Fig. 6.10.



(a) 0 jets (b) 1 jet (c) 2 jets

Figure 6.10.: Expected and observed exclusion upper limits at 95% CL on $\sigma \times \mathcal{B}$ in the three categories, as a function of the resonance mass. The dashed line corresponds to median upper limit, while the green and yellow regions represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands, respectively. The dotted line represents the observed limit. Limits are derived assuming $\Gamma' = \Gamma_{\mathrm{SM}}$ for each mass point.

A mild excess is observed in the 0 jets category and, more evident, in the 1 jet category around 250-300 GeV. A deficit is instead in the VBF category around 250 GeV, which is mainly due to an underfluctuation of the background. This effect can be understood looking at the VBF shape in Fig. 6.9, where two adjacent data points, corresponding to the fifth and sixth bins, clearly underfluctuate with respect to the background prediction, causing the dip in the observed limit.

The exclusion limit resulting from the combination of the three categories is shown in Fig. 6.11, for the five Γ' hypotheses discussed before. From the combined exclusion limits no significant evidence of a deviation from the SM prediction is observed. The presence of a scalar resonance with $\sigma \times \mathcal{B}$ higher than the values reported in Fig. 6.11 is thus excluded with a 95% CL for masses ranging from 200 GeV up to 1 TeV.

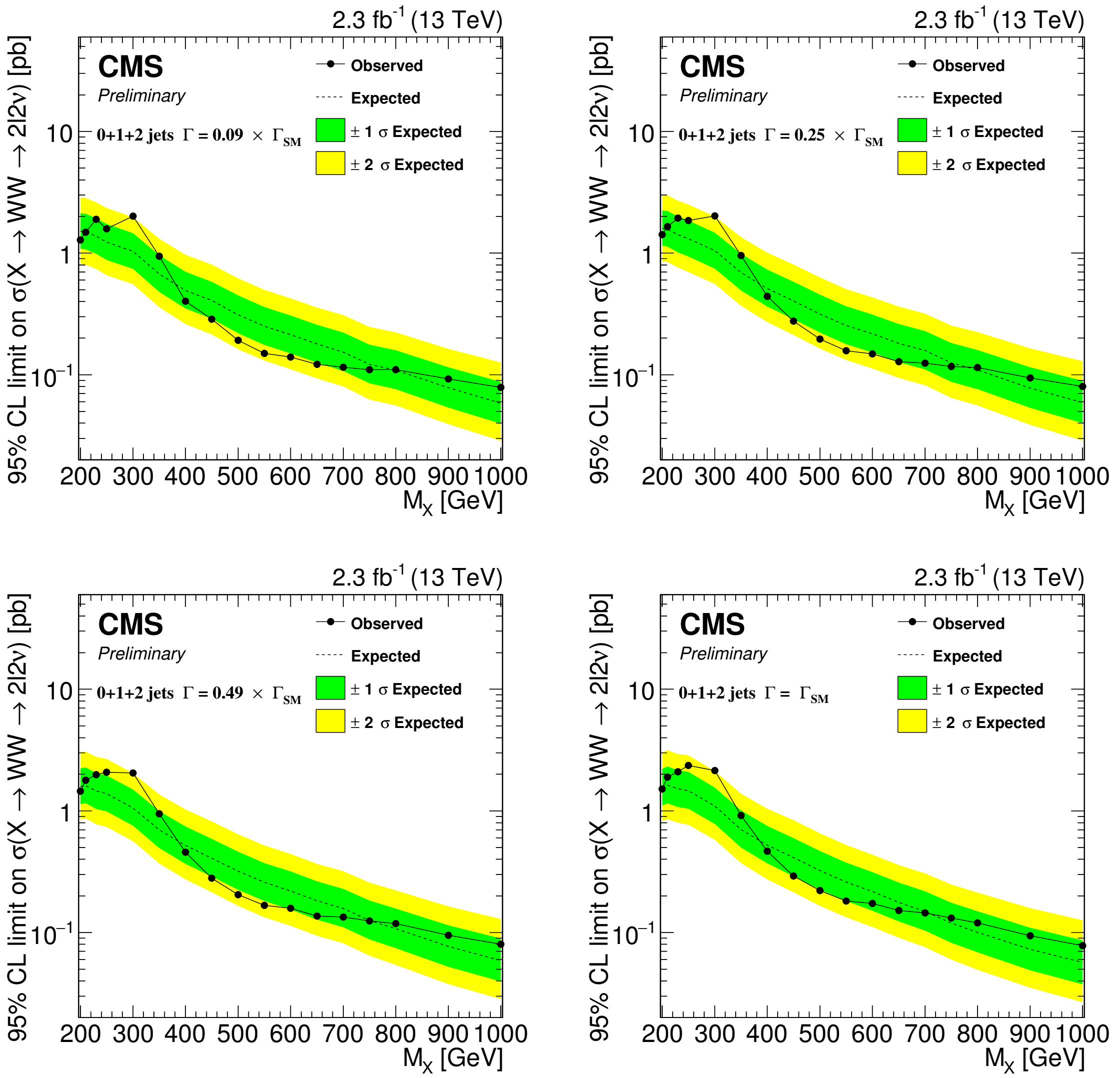


Figure 6.11.: Expected and observed exclusion limits at 95% CL on $\sigma \times \mathcal{B}$ for the combination of the three jet categories as a function of the resonance mass. The black dotted line corresponds to the observed value while the yellow and green bands represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties respectively. Limits are shown for four hypotheses of the resonance decay width.

Chapter 7.

Conclusions

Appendix A.

Fiducial region definition and optimization

The fiducial region must be chosen in such a way to be as close as possible to the selections applied in the analysis, in order to reduce the model dependece in the extrapolation step. That means that for optimizing the fiducial volume definition, the efficiency has to be maximized. Another parameter entering the game is the number of fake events, in other words the number of reconstructed events which do not belong to the fiducial phase space. This parameter should instead be as small as possible. Even if we have to observe the trend of these two quantities as a function of $p_{\mathrm{T}}^{\mathrm{H}}$, we can maximize the ratio between the overall efficiency and the overall fake rate as a proxy for establishing the "goodness" of the fiducial region.

Several different fiducial region definitions were tested and the results show that:

- **of cut:** The fiducial region definition must include only the opposite flavor combination including one electron and one muon. If we include also the combinations involving τ's the efficiency falls down.
- **Lepton cut:** Since the resolution on lepton transverse momentum is good, there is no need to loosen the cuts related these variables, i.e. we can use the same cuts defined in the analysis selection ($p_{\mathrm{T}}^{\ell,1} > 20\,\mathrm{GeV}$, $p_{\mathrm{T}}^{\ell,2} > 10\,\mathrm{GeV}$).
- **Di-lepton p_{T} cut:** As stated in the previous point, there is no need to loosen this cut, so we kept the same value as the analysis selection, i.e. $p_{\mathrm{T}}^{\ell\ell} > 30\,\mathrm{GeV}$.
- **Di-lepton mass cut:** $m_{\ell\ell} > 12\,\mathrm{GeV}$ as discussed before.
- **neutrino pair p_{T} cut:** Since the resolution on the measurement of the missing transverse energy is poor, the neutrino pair cut should not be included in the definition of the fiducial region, because it would increase the fake rate without increasing the efficiency, thus resulting in a lower ratio between overall efficiency and fake rate.
- **m_{T} cut:** Also the m_{T} cut that we have in the analysis selection, i.e. $m_{\mathrm{T}} > 60\,\mathrm{GeV}$, should be loosened or removed because it involves neutrinos and then increase the

fake rate. We decided eventually to keep this cut, loosening it to 50 GeV, because in addition to increase the number of fake events, it increases the efficiency as well.

The fake rate and the efficiency as a function of $p_{\mathrm{T}}^{\mathrm{H}}$ after the optimization discussed before are shown in figure A.1. To obtain these plots the fiducial region was modified adding in sequence the various cuts and computing the efficiency and the fake rate each time. In that way we can asses the composition of those distributions.

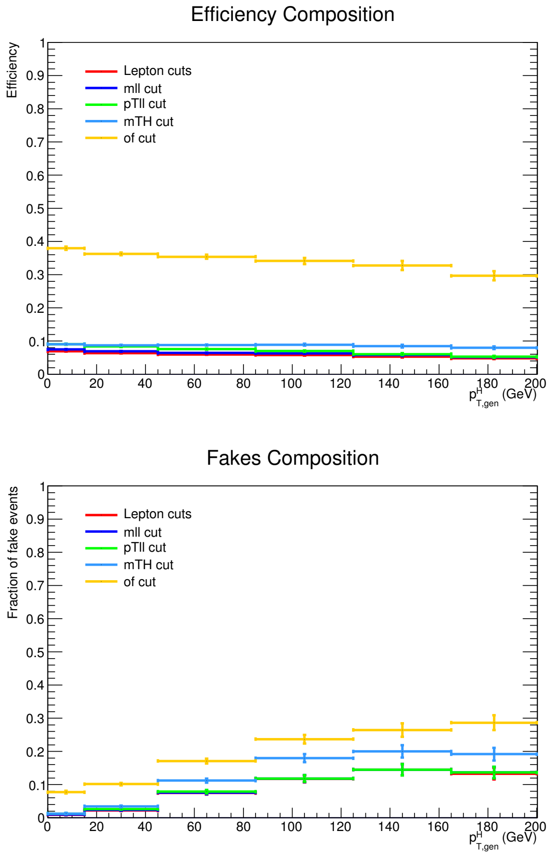


Figure A.1.: Efficiency and fake rate as a function on Higgs transverse momentum. The plots correspond to the optimized fiducial region definition and show the effect of adding each of the mentioned cuts in sequence.

The efficiency and fraction of fake events have been measured also as a function of the $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} cuts in the fiducial region. Since these two variables are correlated, the results are reported as two-dimensional histograms. In Fig. A.2 are reported the efficiency and fraction of fake events for these two variables.

The criterion adopted to define the fiducial region is a tradeoff between having a large efficiency and a small fraction of fake events. Especially when looking at the low resolution variables, such as $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T}, a suitable figure of merit has to be chosen for the estimation of the best cuts. Several different figures of merit have been checked, such as ϵ/f, $\epsilon - f$

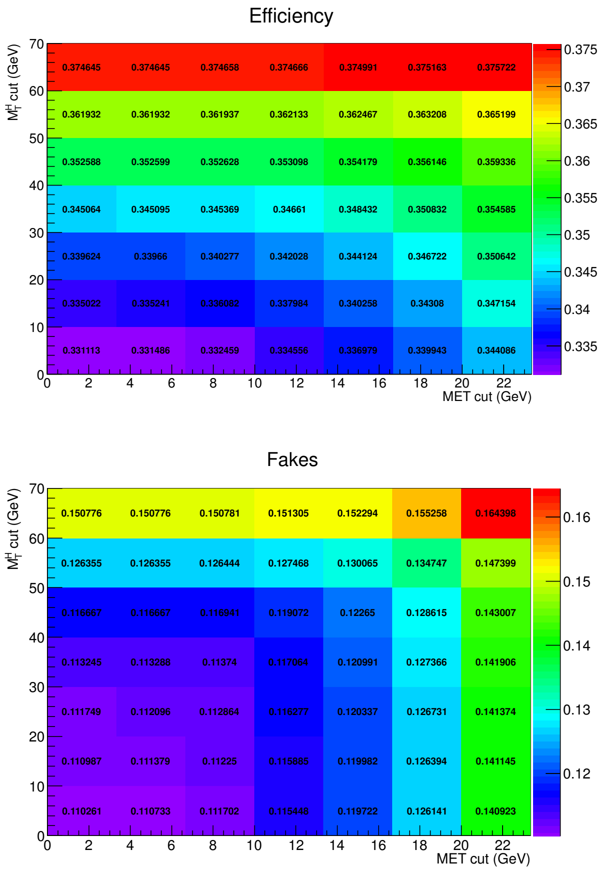


Figure A.2.: Efficiency and fake rate as a function of $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} cuts in the fiducial region.

and $(1-f)/\epsilon$. The results for these three different figures of merit are shown in Fig. A.3 as a function of the $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} cuts in the fiducial region.

Following the same criterion, similar plots as above have been obtained for an alternative model, given by varying up the ggH/VBF ratio within the experimental uncertainties. The results, shown in Fig. A.4 and Fig. A.5, show a similar trend with respect to the model with nominal ggH/VBF ratio.

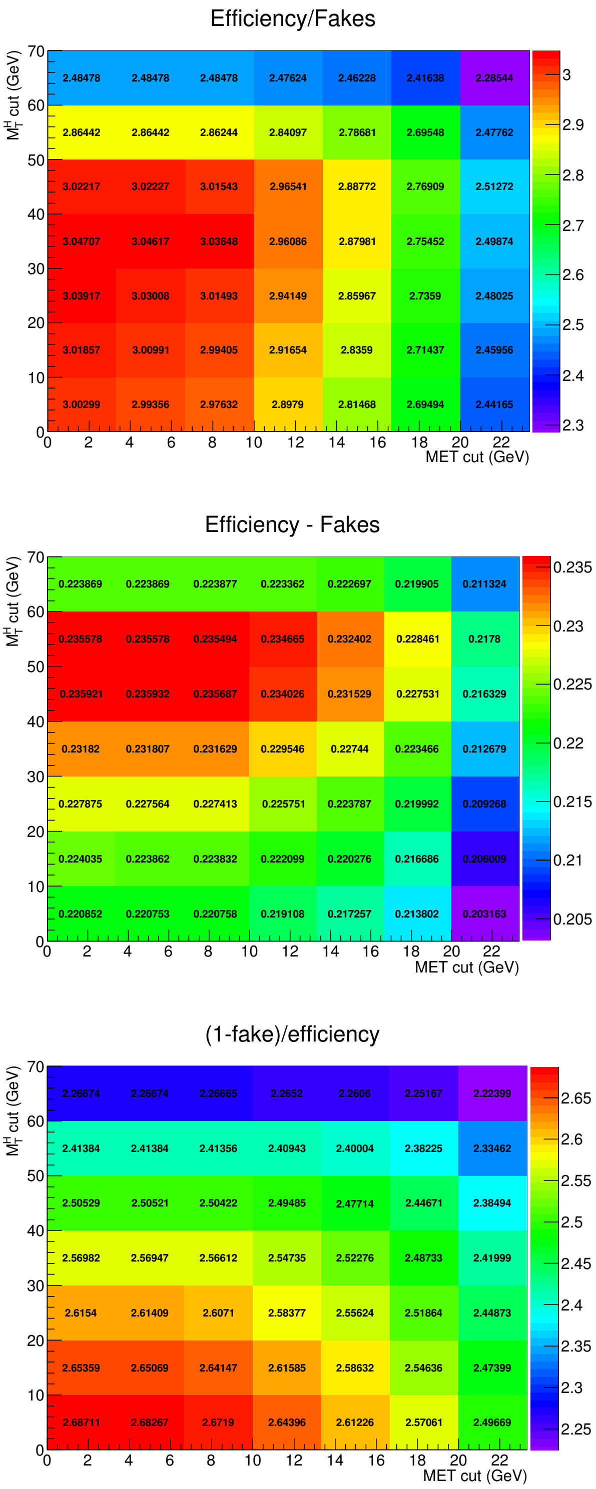


Figure A.3.: Different figures of merit as a function of $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} cuts in the fiducial region.

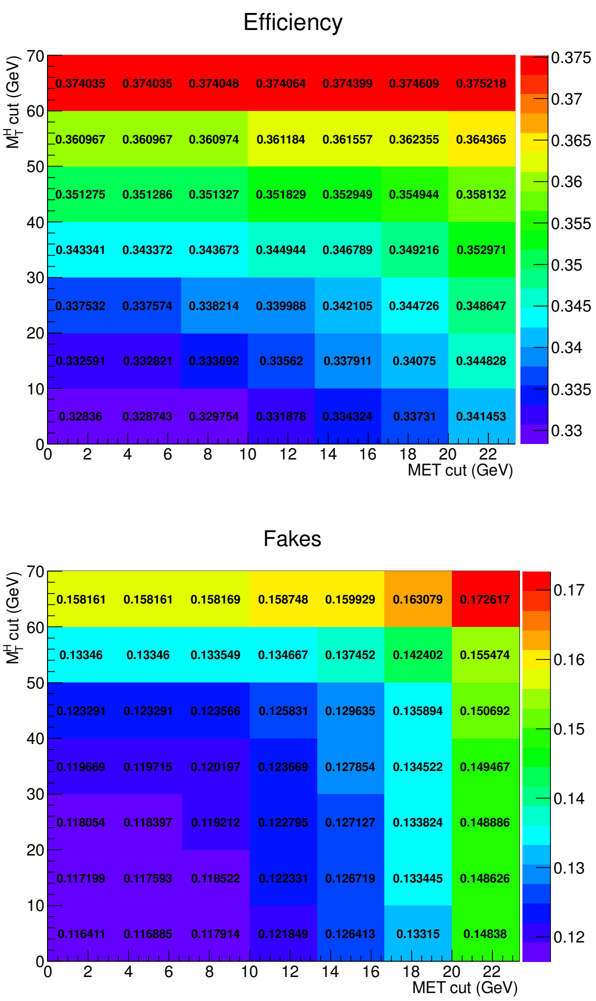


Figure A.4.: Efficiency and fake rate as a function of $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} cuts in the fiducial region, for the alternative model with an up variation of the ggH/VBF ratio.

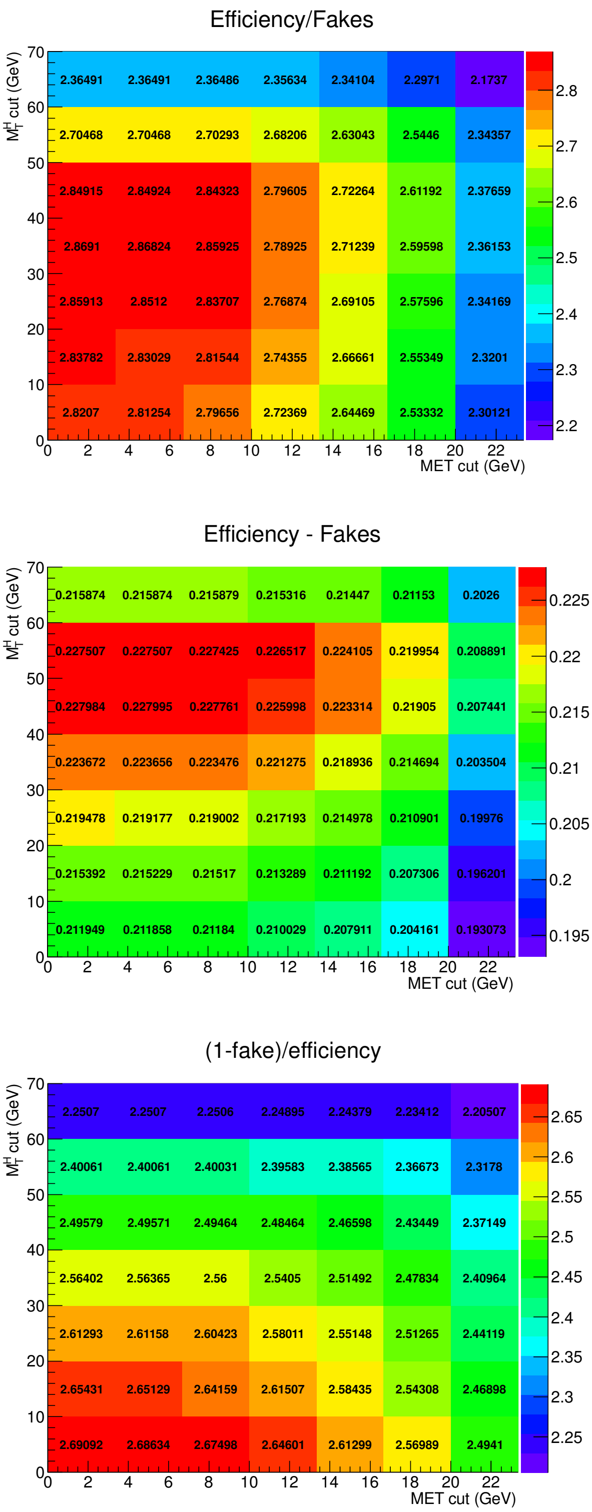


Figure A.5.: Different figures of merit as a function of $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} cuts in the fiducial region, for the alternative model with an up variation of the ggH/VBF ratio.

1923

Bibliography

[1] M. Spira et al. "Higgs boson production at the LHC". In: *Nucl. Phys.* B 453 (1995), pp. 17–82. DOI: 10.1016/0550-3213(95)00379-7. arXiv: hep-ph/9504378 [hep-ph].

[2] Robert Harlander and Philipp Kant. "Higgs production and decay: Analytic results at next-to-leading order QCD". In: *JHEP* 12 (2005), p. 015. DOI: 10.1088/1126-6708/2005/12/015. arXiv: hep-ph/0509189 [hep-ph].

[3] V. Ravindran, J. Smith, and W. L. van Neerven. "NNLO corrections to the total cross-section for Higgs boson production in hadron hadron collisions". In: *Nucl. Phys.* B 665 (2003), pp. 325–366. DOI: 10.1016/S0550-3213(03)00457-7. arXiv: hep-ph/0302135 [hep-ph].

[4] Stefano Catani and Massimiliano Grazzini. "An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC". In: *Phys. Rev. Lett.* 98 (2007), p. 222002. DOI: 10.1103/PhysRevLett.98.222002. arXiv: hep-ph/0703012 [hep-ph].

[5] Charalampos Anastasiou et al. "Higgs Boson Gluon-Fusion Production in QCD at Three Loops". In: *Phys. Rev. Lett.* 114 (2015), p. 212001. DOI: 10.1103/PhysRevLett.114.212001. arXiv: 1503.06056 [hep-ph].

[6] X. Chen et al. "NNLO QCD corrections to Higgs boson production at large transverse momentum". In: (2016). arXiv: 1607.08817 [hep-ph].

[7] Massimiliano Grazzini and Hayk Sargsyan. "Heavy-quark mass effects in Higgs boson production at the LHC". In: *JHEP* 09 (2013), p. 129. DOI: 10.1007/JHEP09(2013)129. arXiv: 1306.4581 [hep-ph].

[8] Aleksandr Azatov and Ayan Paul. "Probing Higgs couplings with high p_T Higgs production". In: *JHEP* 01 (2014), p. 014. DOI: 10.1007/JHEP01(2014)014. arXiv: 1309.5273 [hep-ph].

[9] Robert V. Harlander and Tobias Neumann. "Probing the nature of the Higgs-gluon coupling". In: *Phys. Rev. D* 88 (2013), p. 074015. DOI: 10.1103/PhysRevD.88.074015. arXiv: 1308.2225 [hep-ph].

[10] David Marzocca, Marco Serone, and Jing Shu. "General Composite Higgs Models". In: *JHEP* 08 (2012), p. 013. DOI: 10.1007/JHEP08(2012)013. arXiv: 1205.0770 [hep-ph].

[11] Andrea Banfi, Adam Martin, and Veronica Sanz. “Probing top-partners in Higgs+jets”. In: *JHEP* 08 (2014), p. 053. DOI: 10.1007/JHEP08(2014)053. arXiv: 1308.4771 [hep-ph].

[12] Georges Aad et al. “Fiducial and differential cross sections of Higgs boson production measured in the four-lepton decay channel in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ with the ATLAS detector”. In: *Phys. Lett. B* 738 (2014), p. 234. DOI: 10.1016/j.physletb.2014.09.054. arXiv: 1408.3226 [hep-ex].

[13] Georges Aad et al. “Measurements of fiducial and differential cross sections for Higgs boson production in the diphoton decay channel at $\sqrt{s} = 8\,\text{TeV}$ with ATLAS”. In: *JHEP* 09 (2014), p. 112. DOI: 10.1007/JHEP09(2014)112. arXiv: 1407.4222 [hep-ex].

[14] Georges Aad et al. “Measurements of the Total and Differential Higgs Boson Production Cross Sections Combining the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ Decay Channels at $\sqrt{s} = 8$ TeV with the ATLAS Detector”. In: *Phys. Rev. Lett.* 115 (2015), p. 091801. DOI: 10.1103/PhysRevLett.115.091801. arXiv: 1504.05833 [hep-ex].

[15] Vardan Khachatryan et al. “Measurement of differential cross sections for Higgs boson production in the diphoton decay channel in pp collisions at $\sqrt{s} = 8\,\text{TeV}$”. In: *Eur. Phys. J. C* 76 (2016), p. 13. DOI: 10.1140/epjc/s10052-015-3853-3. arXiv: 1508.07819 [hep-ex].

[16] Vardan Khachatryan et al. “Measurement of differential and integrated fiducial cross sections for Higgs boson production in the four-lepton decay channel in pp collisions at $\sqrt{s} = 7$ and 8 TeV”. In: *JHEP* 04 (2016), p. 005. DOI: 10.1007/JHEP04(2016)005. arXiv: 1512.08377 [hep-ex].

[17] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs cross sections: 3. Higgs properties.* CERN Report CERN-2013-004. 2013. DOI: 10.5170/CERN-2013-004. arXiv: 1307.1347 [hep-ph].

[18] 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].

[19] G. Cowan. “A survey of unfolding methods for particle physics”. In: *Conf. Proc.* C0203181 (2002), p. 248.

[20] Michael Krämer, Stephen Mrenna, and Davison E. Soper. “Next-to-leading order QCD jet production with parton showers and hadronization”. In: *Phys. Rev. D* 73 (2006), p. 014022. DOI: 10.1103/PhysRevD.73.014022. arXiv: hep-ph/0509127 [hep-ph].

[21] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[22] Nils Lavesson and Leif Lonnblad. "Extending CKKW-merging to one-loop matrix elements". In: *JHEP* 12 (2008), p. 070. DOI: `10.1088/1126-6708/2008/12/070`. arXiv: `0811.2912 [hep-ph]`.

[23] Simone Alioli et al. "NLO Higgs boson production via gluon fusion matched with shower in POWHEG". In: *JHEP* 04 (2009), p. 002. DOI: `10.1088/1126-6708/2009/04/002`. arXiv: `0812.0578 [hep-ph]`.

[24] Paolo Nason and Carlo Oleari. "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG". In: *JHEP* 02 (2010), p. 037. DOI: `10.1007/JHEP02(2010)037`. arXiv: `0911.5299 [hep-ph]`.

[25] Simone Alioli, Sven-Olaf Moch, and Peter Uwer. "Hadronic top-quark pair-production with one jet and parton showering". In: *JHEP* 01 (2012), p. 137. DOI: `10.1007/JHEP01(2012)137`. arXiv: `1110.5251 [hep-ph]`.

[26] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. DOI: `10.1007/JHEP07(2014)079`. arXiv: `1405.0301 [hep-ph]`.

[27] T. Binoth et al. "Gluon-induced W-boson pair production at the LHC". In: *JHEP* 12 (2006), p. 046. DOI: `10.1088/1126-6708/2006/12/046`. arXiv: `hep-ph/0611170 [hep-ph]`.

[28] Marco Bonvini et al. "Signal-background interference effects for $gg \to H \to W^+W^-$ beyond leading order". In: *Phys. Rev. D* 88 (2013), p. 034032. DOI: `10.1103/PhysRevD.88.034032`. arXiv: `1304.3053 [hep-ph]`.

[29] Giampiero Passarino. "Higgs CAT". In: *Eur. Phys. J. C* 74 (2014), p. 2866. DOI: `10.1140/epjc/s10052-014-2866-7`. arXiv: `1312.2397 [hep-ph]`.

[30] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. "PYTHIA 6.4 physics and manual". In: *JHEP* 05 (2006), p. 026. DOI: `10.1088/1126-6708/2006/05/026`. arXiv: `hep-ph/0603175 [hep-ph]`.

[31] 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]`.

[32] Hung-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]`.

[33] LHC Higgs Cross Section Working Group. "Handbook of LHC Higgs Cross Sections". In: *arXiv:1101.0593* (2011).

[34] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: `10.1007/JHEP06(2010)043`. arXiv: `1002.2581 [hep-ph]`.

[35] 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`.

[36] D. de Florian et al. “Higgs boson production at the LHC: transverse momentum resummation effects in the $H \to 2\gamma$, $H \to WW \to l\nu l\nu$ and $H \to ZZ \to 4l$ decay modes”. In: *JHEP* 06 (2012), p. 132. DOI: 10.1007/JHEP06(2012)132. arXiv: 1203.6321 [hep-ph].

[37] E. Bagnaschi et al. “Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM”. In: *JHEP* 02 (2012), p. 088. DOI: 10.1007/JHEP02(2012)088. arXiv: 1111.2854 [hep-ph].

[38] Stefano Actis et al. “NLO Electroweak Corrections to Higgs Boson Production at Hadron Colliders”. In: *Phys. Lett. B* 670 (2008), p. 12. DOI: 10.1016/j.physletb.2008.10.018. arXiv: 0809.1301 [hep-ph].

[39] Stefano Catani et al. “Soft gluon resummation for Higgs boson production at hadron colliders”. In: *JHEP* 07 (2003), p. 028. DOI: 10.1088/1126-6708/2003/07/028. arXiv: hep-ph/0306211 [hep-ph].

[40] Yanyan Gao et al. “Spin determination of single-produced resonances at hadron colliders”. In: *Phys. Rev. D* 81 (2010), p. 075022. DOI: 10.1103/PhysRevD.81.075022. arXiv: 1001.3396 [hep-ph].

[41] Sara Bolognesi et al. “Spin and parity of a single-produced resonance at the LHC”. In: *Phys. Rev. D* 86 (2012), p. 095031. DOI: 10.1103/PhysRevD.86.095031. arXiv: 1208.4018 [hep-ph].

[42] Ian Anderson et al. “Constraining anomalous HVV interactions at proton and lepton colliders”. In: *Phys. Rev. D* 89 (2014), p. 035007. DOI: 10.1103/PhysRevD.89.035007. arXiv: 1309.4819 [hep-ph].

[43] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. “A Brief Introduction to PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (2008), p. 852. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[44] Vardan Khachatryan et al. “Measurement of the $\mathrm{t\bar{t}}$ production cross section in the $\mathrm{e}\mu$ channel in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV”. In: (2016). arXiv: 1603.02303 [hep-ex].

[45] Stanislaw Jadach, Johann H. Kuhn, and Zbigniew Was. “TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons”. In: *Comput. Phys. Commun.* 64 (1990), p. 275. DOI: 10.1016/0010-4655(91)90038-M.

[46] CMS Collaboration. “Standard Model Cross Sections for CMS at 7 TeV”. In: *CMS Generator Group Twiki* (2010).

[47] John M. Campbell, R. Keith Ellis, and Ciaran Williams. “Vector boson pair production at the LHC”. In: *JHEP* 07 (2011), p. 018. DOI: 10.1007/JHEP07(2011)018. arXiv: 1105.0020 [hep-ph].

[48] J. Ohnemus. “Order alpha-s calculations of hadronic W+- gamma and Z gamma production”. In: *Phys. Rev.* D47 (1993), pp. 940–955. DOI: 10.1103/PhysRevD.47.940.

[49] JetMET group. "Jet energy uncertainties, https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources#2012_JEC". In: (). URL: https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources%5C#2012_JEC.

[50] Sergey Alekhin et al. "The PDF4LHC Working Group Interim Report". 2011.

[51] Michiel Botje et al. "The PDF4LHC Working Group Interim Recommendations". 2011.

[52] Richard D. Ball et al. "Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology". In: *Nucl. Phys. B* 849 (2011), p. 296. DOI: 10.1016/j.nuclphysb.2011.03.021. arXiv: 1101.1300 [hep-ph].

[53] A. D. Martin et al. "Parton distributions for the LHC". In: *Eur. Phys. J. C* 63 (2009), p. 189. DOI: 10.1140/epjc/s10052-009-1072-5. arXiv: 0901.0002 [hep-ph].

[54] Iain W. Stewart and Frank J. Tackmann. "Theory uncertainties for Higgs and other searches using jet bins". In: *Phys. Rev. D* 85 (2012), p. 034011. DOI: 10.1103/PhysRevD.85.034011. arXiv: 1107.2117 [hep-ph].

[55] "Procedure for the LHC Higgs boson search combination in summer 2011". In: (2011).

[56] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J.* C71 (2011). [Erratum: Eur. Phys. J.C73,2501(2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0,10.1140/epjc/s10052-013-2501-z. arXiv: 1007.1727 [physics.data-an].

[57] Tim Adye. "Unfolding algorithms and tests using RooUnfold". 2011.

[58] Andreas Hocker and Vakhtang Kartvelishvili. "SVD approach to data unfolding". In: *Nucl. Instrum. Meth. A* 372 (1996), p. 469. DOI: 10.1016/0168-9002(95)01478-0. arXiv: hep-ph/9509307 [hep-ph].

[59] G. D'Agostini. "A Multidimensional unfolding method based on Bayes' theorem". In: *Nucl. Instrum. Meth. A* 362 (1995), pp. 487–498. DOI: 10.1016/0168-9002(95)00274-X.

[60] Vardan Khachatryan et al. "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV". In: *Eur. Phys. J. C* 75 (2015), p. 212. DOI: 10.1140/epjc/s10052-015-3351-7. arXiv: 1412.8662 [hep-ex].

[61] Paolo Nason. "A New method for combining NLO QCD with shower Monte Carlo algorithms". In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146 [hep-ph].

[62] S. Bolognesi, Y. Gao and A.V. Gritsan *et al.* "JHUGen". In: *URL* http://www.pha.jhu.edu/spin/ (2011).

[63] Gionata Luisoni et al. "$HW^{\pm}$/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO". In: *JHEP* 10 (2013), p. 083. DOI: 10.1007/JHEP10(2013)083. arXiv: 1306.2542 [hep-ph].

[64] Tom Melia et al. "W+W-, WZ and ZZ production in the POWHEG BOX". In: *JHEP* 11 (2011), p. 078. DOI: 10.1007/JHEP11(2011)078. arXiv: 1107.5051 [hep-ph].

[65] T. Gehrmann et al. "W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD". In: *Phys. Rev. Lett.* 113.21 (2014), p. 212001. DOI: 10.1103/PhysRevLett.113.212001. arXiv: 1408.5243 [hep-ph].

[66] Patrick Meade, Harikrishnan Ramani, and Mao Zeng. "Transverse momentum resummation effects in W^+W^- measurements". In: *Phys. Rev.* D90.11 (2014), p. 114006. DOI: 10.1103/PhysRevD.90.114006. arXiv: 1407.4481 [hep-ph].

[67] Prerit Jaiswal and Takemichi Okui. "Explanation of the WW excess at the LHC by jet-veto resummation". In: *Phys. Rev.* D90.7 (2014), p. 073009. DOI: 10.1103/PhysRevD.90.073009. arXiv: 1407.4537 [hep-ph].

[68] John M. Campbell, R. Keith Ellis, and Ciaran Williams. "Bounding the Higgs width at the LHC: Complementary results from $H \to WW$". In: *Phys. Rev.* D89.5 (2014), p. 053011. DOI: 10.1103/PhysRevD.89.053011. arXiv: 1312.1628 [hep-ph].

[69] Richard D. Ball et al. "Parton distributions with QED corrections". In: *Nucl. Phys.* B877 (2013), pp. 290–320. DOI: 10.1016/j.nuclphysb.2013.10.010. arXiv: 1308.0598 [hep-ph].

[70] Richard D. Ball et al. "Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO". In: *Nucl. Phys.* B855 (2012), pp. 153–221. DOI: 10.1016/j.nuclphysb.2011.09.024. arXiv: 1107.2652 [hep-ph].

[71] Vardan Khachatryan et al. "Event generator tunes obtained from underlying event and multiparton scattering measurements". In: (2015). arXiv: 1512.00815 [hep-ex].

[72] Peter Richardson and Alexandra Wilcock. "Monte Carlo Simulation of Hard Radiation in Decays in Beyond the Standard Model Physics in Herwig++". In: *Eur. Phys. J.* C74 (2014), p. 2713. DOI: 10.1140/epjc/s10052-014-2713-x. arXiv: 1303.4563 [hep-ph].

[73] J. Bellm et al. "Herwig++ 2.7 Release Note". In: (2013). arXiv: 1310.6877 [hep-ph].

[74] *SM Higgs production cross sections at $\sqrt{s}$ = 13-14 TeV.* https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt1314TeV.

[75] Fabrizio Caola et al. "QCD corrections to W^+W^- production through gluon fusion". In: *Phys. Lett.* B754 (2016), pp. 275–280. DOI: 10.1016/j.physletb.2016.01.046. arXiv: 1511.08617 [hep-ph].

[76] *NLO single-top channel cross sections.* https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec.

[77] *NNLO+NNLL top-quark-pair cross sections.* https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO.

[78] *Procedure for the LHC Higgs boson search combination in Summer 2011.* Tech. rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. Geneva: CERN, Aug. 2011.

[79] *LHC Higgs Cross Section Working Group.* `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG`.

[80] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *J. Phys.* G43 (2016), p. 023001. DOI: `10.1088/0954-3899/43/2/023001`. arXiv: `1510.03865 [hep-ph]`.

[81] Vardan Khachatryan et al. "Search for a Higgs Boson in the Mass Range from 145 to 1000 GeV Decaying to a Pair of W or Z Bosons". In: *JHEP* 10 (2015), p. 144. DOI: `10.1007/JHEP10(2015)144`. arXiv: `1504.00936 [hep-ex]`.

[82] G. C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". In: *Phys. Rept.* 516 (2012), p. 1. DOI: `10.1016/j.physrep.2012.02.002`. arXiv: `1106.0034 [hep-ph]`.

[83] N. Craig and T. Scott. "Exclusive signals of an extended Higgs sector". In: *JHEP* 11 (2012), p. 083. DOI: `10.1007/JHEP11(2012)083`. arXiv: `1207.4835 [hep-ph]`.

[84] H. Haber and O. Stal. "New LHC benchmarks for the CP-conserving two-Higgs-doublet model". In: *Eur. Phys. J. C* 75 (2015), p. 491. DOI: `10.1140/epjc/s10052-015-3697-x`. arXiv: `1507.04281 [hep-ph]`.

[85] Suyong Choi, Sunghoon Jung, and P. Ko. "Implications of LHC data on 125 GeV Higgs-like boson for the Standard Model and its various extensions". In: *JHEP* 1310 (2013), p. 225. DOI: `10.1007/JHEP10(2013)225`. arXiv: `1307.3948`.

[86] Tania Robens and Tim Stefaniak. "Status of the Higgs singlet extension of the standard model after LHC run 1". In: *Eur. Phys. J. C* 75 (2015), p. 105. DOI: `10.1140/epjc/s10052-015-3323-y`. arXiv: `1501.02234 [hep-ph]`.

[87] Johan Alwall et al. "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions". In: *Eur. Phys. J. C* 53 (2008), pp. 473–500. DOI: `10.1140/epjc/s10052-007-0490-5`. arXiv: `0706.2569 [hep-ph]`.

[88] Massimiliano Grazzini. "NNLO predictions for the Higgs boson signal in the $\mathrm{H}\rightarrow\mathrm{WW}\rightarrow \ell\nu\ell\nu$ and $\mathrm{H}\rightarrow\mathrm{ZZ}\rightarrow 4\ell$ decay channels". In: *JHEP* 02 (2008), p. 043. DOI: `10.1088/1126-6708/2008/02/043`. arXiv: `0801.3232 [hep-ph]`.

[89] Vardan Khachatryan et al. "Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8\,\mathrm{TeV}$". In: *Eur. Phys. J. C* 75.11 (2015), p. 542. DOI: `10.1140/epjc/s10052-015-3709-x`. arXiv: `1505.04480 [hep-ex]`.