

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

9

Abstract

10 The cross section for Higgs boson production in pp collisions is studied
11 using the $H \rightarrow W^+W^-$ decay mode, followed by leptonic decays of the
12 W bosons, leading to an oppositely charged electron-muon pair in the
13 final state. The measurements are performed using data collected by the
14 CMS experiment at the LHC with pp collisions at a centre-of-mass energy
15 of 8 TeV, corresponding to an integrated luminosity of 19.4 fb^{-1} . The
16 Higgs boson transverse momentum (p_T) is reconstructed using the lepton
17 pair p_T and missing p_T . The differential cross section times branching
18 fraction is measured as a function of the Higgs boson p_T in a fiducial
19 phase space defined to match the experimental acceptance in terms of
20 the lepton kinematics and event topology. The production cross section
21 times branching fraction in the fiducial phase space is measured to be
22 $39 \pm 8\text{ (stat)} \pm 9\text{ (syst)}\text{fb}$. The measurements are compared to theoretical
23 calculations based on the standard model to which they agree within
24 experimental uncertainties.

²⁵ **Contents**

Chapter 1

²⁷ Electroweak and QCD physics at
²⁸ LHC

Chapter 2

²⁹ The CMS experiment at the LHC

³⁰ In this chapter, the main characteristics of the Large Hadron Collider (LHC particle
³¹ accelerator and Compact Muon Solenoid (CMS) experiment are described.

³² 2.1 The Large Hadron Collider

³³ The LHC [1–4] at CERN, officially inaugurated on 21st October 2008, is the largest and
³⁴ most powerful hadron collider ever built. Installed in the underground tunnel which hosted
³⁵ the Large Electron Positron Collider (LEP) [5–7], the leptonic accelerator in operation until
³⁶ 2nd November 2000, the LHC accelerator has the shape of a circle with a length of about
³⁷ 27 km and is located underground at a depth varying between 50 m to 175 m, straddling the
³⁸ Franco-Swiss border near Geneva. It is designed to collide two 7 TeV counter-circulating
³⁹ beams of protons resulting in a center-of-mass energy of 14 TeV, or two beams of heavy
⁴⁰ ions, in particular lead nuclei at an energy of 2.76 TeV/nucleon in the center-of-mass frame.

⁴¹ The transition from a leptonic collider to a hadronic collider entailed the following
⁴² advantages: first, it has been possible to build a machine that having the same size of the
⁴³ previous one (and therefore accommodated in the same LEP tunnel, substantially reducing
⁴⁴ the cost and time of construction), could reach a higher energy in the center-of-mass
⁴⁵ frame. This is due to the much lower amount of energy loss through synchrotron radiation
⁴⁶ emitted by the accelerated particles, that is proportional to the fourth power of the ratio
⁴⁷ E/m between their energy and their mass. Secondly, the composite structure of protons
⁴⁸ compared to the elementary structure of electrons allows LHC to be able to simultaneously
⁴⁹ access a wider energy spectrum, despite the production of many low energies particles in a
⁵⁰ complex environment. This is a particularly important feature for a machine dedicated to
⁵¹ the discovery of “new” physics.

⁵² In Fig. 2.1 a schematic description of the accelerator complex installed at CERN is shown.
⁵³ The acceleration is performed in several stages [4]. The protons source is a *Duoplasmatron*:
⁵⁴ the protons are obtained by removing electrons from a source of hydrogen gas and then
⁵⁵ sent to the LINAC2, a 36 m long linear accelerator which generates a pulsed beam with

an energy of 50 MeV using Radio Frequency Quadrupoles (RFQ) and focusing quadrupole magnets. The beam is subsequently sent to the Proton Synchrotron Booster (PSB), a circular accelerator consisting of four superimposed synchrotron rings with a circumference of about 160 m, which increases the proton energy up to 1.4 GeV. Then, protons are injected into the Proton Synchrotron (PS), a single synchrotron ring with a circumference of about 600 m where the energy is increased to 25 GeV. The sequential combination of these two synchrotrons also allows to create a series of protons bunches interspersed by 25 ns as required for the correct operation of LHC. The final proton injection stage is the Super Proton Synchrotron (SPS), a synchrotron with a circumference of approximately 7 km where protons reach an energy value of 450 GeV. Subsequently, protons are extracted and injected into the LHC ring via two transmission lines, to generate two beams running in opposite directions in two parallel pipes and which are accelerated up to the energy of interest. In the two pipes an ultrahigh vacuum condition is maintained (about 10^{-10} Torr) to avoid the spurious proton interactions with the gas remnants. At full intensity, each proton beam consists of 2808 bunches and each bunch contains around 10^{11} protons. The beams are squeezed and collide for a length of about 130 m at four interaction points where the four main experiments (ALICE, ATLAS, CMS and LHCb) are placed:

- CMS (Compact Muon Solenoid) [8] and ATLAS (A Toroidal LHC ApparatuS) [9] are two general-purpose detectors designed to investigate the largest possible spectrum of physics. In particular, they have been devoted to the detection of particles produced by a Higgs boson decay and to look for any evidence of possible new physics. The use of two detectors chasing the same objectives but designed independently is crucial for a cross-check of any possible new discovery;
- LHCb (LHC beauty) [10] is an experiment primarily designed to study CP (combined Charge conjugation and Parity symmetry) violation in electroweak interactions and to study asymmetries between matter and antimatter through the analysis of rare decays of hadrons containing b quarks. The detector is also able to perform measurements in the forward region, at small polar angles with respect to the beam line;
- ALICE (A Large Ion Collider Experiment) [11] is an experiment studying heavy ions collisions, through the production of a new state of matter called quark-gluon plasma.

Two other smaller experiments are located along the circumference of the LHC accelerator, TOTEM and LHCf, which focus on particles emitted in the forward direction. TOTEM (TOTal Elastic and diffractive cross section Measurement) [12] measures the proton-proton interaction cross section and accurately monitors the luminosity of the LHC using detectors positioned on either side of the CMS interaction point. LHCf (LHC forward) [13] is made up of two detectors which sit along the LHC beamline, at 140 m either side of the ATLAS collision point. It makes use of neutral particles thrown forward by LHC collisions as a source to simulate the interaction with the atmosphere of very high energy cosmic rays (between 10^{17} TeV and 10^{20} TeV) in laboratory conditions.

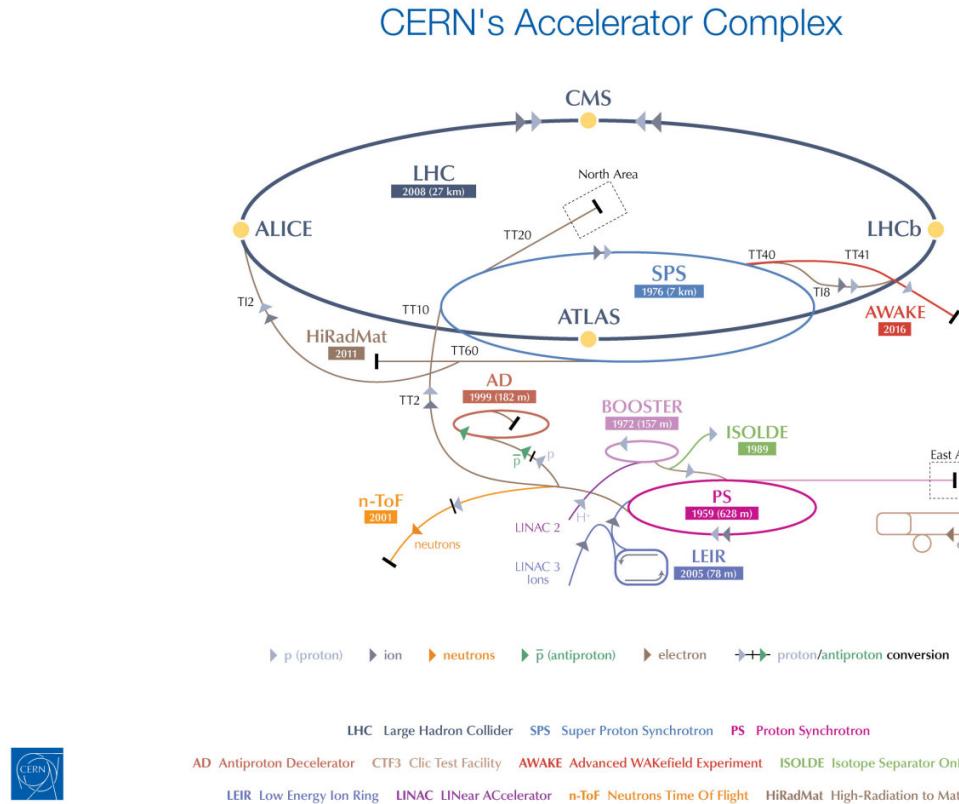


Figure 2.1: Schematic description of the accelerator complex installed at CERN.

95 A series of about 1200 magnetic dipoles bend the beams along the accelerator ring.
 96 They are located along the “arc” structures of the circumference. The ring, in fact, can
 97 be subdivided into octants, with eight curve regions (the “arcs”) separated by rectilinear
 98 regions. In these straight regions, instead, almost 400 focusing and defocusing quadrupoles
 99 are located, which maintain the beam stable along the orbit, and some other small multipolar
 100 magnets (sextupoles and octupoles) are used to make additional minor corrections to the
 101 beam direction. A radio frequency acceleration system, consisting of 16 superconducting
 102 radio-frequency resonant cavities, is used to increase the proton energy by 0.5 MeV with
 103 each beam revolution. The 7 TeV per-beam-energy limit on the LHC is not determined by
 104 the electric field generated by the radiofrequency cavity but by the magnetic field necessary
 105 to maintain the protons in orbit, given the current technology for the superconducting
 106 magnets, which is about 5.4 T on average.

107 One of the most important parameters of an accelerator is the instantaneous luminosity
 108 \mathcal{L} , which gives a measure of the rate of events one can expect given the process cross section.
 109 In fact, for a given physics process with cross section σ , producing N events for unit of

time, the instantaneous luminosity is defined by the following equation:

$$N = \sigma \mathcal{L} . \quad (2.1)$$

The LHC design luminosity is $\mathcal{L} = 10^{34} \text{cm}^{-2}\text{s}^{-1}$, leading to around 1 billion proton interactions per second.

The instantaneous luminosity is a parameter which depends on the construction characteristics of the accelerator, and can be expressed by the following approximated formula:

$$\mathcal{L} = f \frac{n_1 n_2}{4\pi\sigma_x\sigma_y} , \quad (2.2)$$

where n_1 and n_2 are the number of particles contained in the two bunches colliding at a frequency f , and σ_x and σ_y are the beam sizes in the transverse plane. At LHC, the bunches collide with $f = 40 \text{MHz}$ and the transverse size of the beam can be squeezed down to around $15 \mu\text{m}$. Then, the integrated luminosity L is defined as the time integral of the instantaneous luminosity:

$$L = \int \mathcal{L} dt . \quad (2.3)$$

The main parameters of the LHC machine are listed in Table 2.1.

The LHC started to be operative in September 2008 but, due to a faulty interconnection between two magnets which caused a helium leakage in the tunnel, the operation was stopped and restarted in March 2010. During 2010 and 2011 LHC ran successfully and provided proton proton collisions at a center-of-mass energy of 7TeV , delivering a total integrated luminosity of about 6.1fb^{-1} . The encouraging results in the Higgs boson search provided by the ATLAS and CMS Collaborations led to the decision of extending the data taking period to the end of 2012, and to increase the center-of-mass energy up to 8TeV . During 2012, LHC delivered to the experiments an integrated luminosity of 23.3fb^{-1} . After the first long shutdown (LS1), a two years period started in the early 2013 where the LHC operation stopped for maintenance and upgrade, the LHC started again delivering proton proton collisions on 3rd June 2015, at the new record center-of-mass energy of 13TeV . During the 2015 the LHC delivered an integrated luminosity of 4.2fb^{-1} . Nowadays, LHC is still colliding bunches of protons at $\sqrt{s} = 13 \text{TeV}$, reaching unprecedented instantaneous luminosities and delivering a total integrated luminosity of 31fb^{-1} . The cumulative delivered luminosity versus time for the different LHC data taking periods is shown in Fig. 2.2.

As the instantaneous luminosity increases, the probability of multiple proton proton interactions to occur in a single bunch crossing grows higher as well. In this instance, the main goal is the identification and reconstruction of a single primary collision where the physics event of interest occurs among the background of the additional proton proton interactions. Such backgrounds are due to processes occurring with very high probability,

Table 2.1: LHC technical parameters for proton proton collisions.

Parameter	Value
Maximum dipole magnetic field	8.33 T
Dipole operating temperature	1.9 K
Beam energy at injection	450 GeV
Beam energy at collision (nominal)	7 TeV
Beam energy at collision (2012)	4 TeV
Beam energy at collision (2015–2016)	6.5 TeV
Maximum instantaneous luminosity (nominal)	$10^{34} \text{ cm}^{-2}\text{s}^{-1}$
Maximum instantaneous luminosity (2012)	$7.7 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$
Maximum instantaneous luminosity (2015–2016)	$1.2 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$
Number of bunches per proton beam (nominal)	2808
Number of bunches per proton beam (2012)	1380
Number of bunches per proton beam (2015–2016)	2220
Maximum number of protons per bunch	$1.69 \cdot 10^{11}$
Bunch separation in time (nominal)	25 ns
Bunch separation in time (2012)	50 ns
Bunch separation in time (2015–2016)	25 ns
Collision frequency (nominal)	40 MHz
Collision frequency (2012)	20 MHz
Collision frequency (2015–2016)	40 MHz
Energy loss per turn at 14 TeV	7 keV

¹⁴¹ like the production of low- p_T jets. These additional collisions are known as pile up (PU).¹⁴² During the LHC current run the average number of pile up events is 23, with some event exhibiting over 45 pile up collisions.

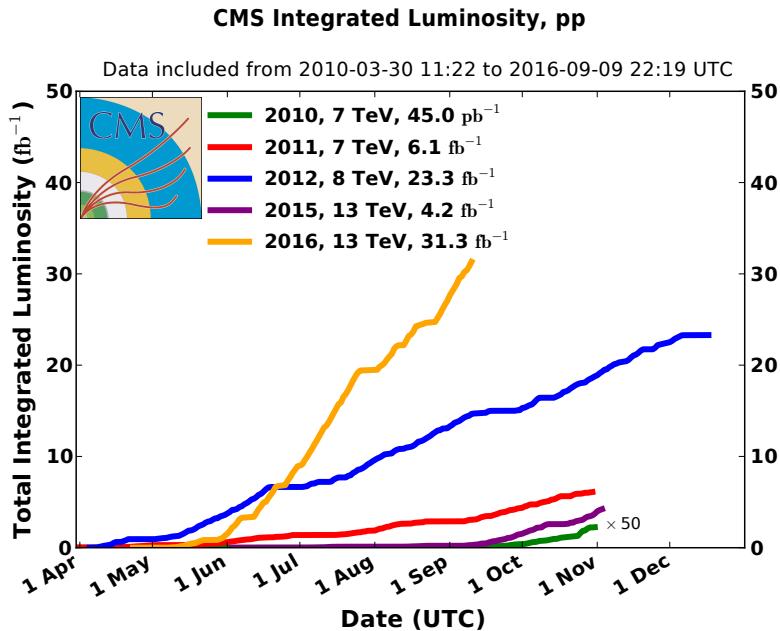


Figure 2.2: Cumulative luminosity versus day delivered to CMS during proton proton collisions.

¹⁴⁴ 2.2 The *Compact Muon Solenoid* experiment

The CMS apparatus is a general purpose detector situated in one of the four LHC interaction points¹. The detector is designed to investigate a wide range of physics, from the search of the Higgs boson, to SM measurements and BSM physics searches. To achieve this goal, the detector is able to identify and reconstruct all the physics objects that may be produced in the proton proton collisions: electrons, muons, photons and jets. The main feature of the CMS detector is a superconducting solenoidal magnet which is capable to produce a 3.8 T magnetic field. Such a strong magnetic field is the key aspect which permits to have a compact design of the detector. The detector has a cylindrical structure, which is typical of general purpose detectors, which consists of several cylindrical detecting layers, coaxial with the beam direction (*barrel* region), closed at both ends with detecting disks (*endcap* region), in such a way to ensure the hermetic closure of the apparatus.

The coordinate system used by CMS is a right-handed Cartesian system, with the origin at the center of the detector, in the nominal beam collision point. The x -axis is chose to point radially towards the center of the LHC circumference and the y -axis is directed upwards along the vertical. The z -axis is oriented along the beam direction, according to the anticlockwise direction of the LHC ring if seen from above. The CMS cylindrical symmetry and the Lorentz invariant description of the proton proton collisions, suggest the

¹The CMS detector is placed in a cavern 100 m underground in the area called Point 5, near the village of Cessy, in France.

¹⁶² use of a pseudo-angular reference frame, described by the triplet of coordinates (r, ϕ, η) ,
¹⁶³ where r is the distance from the z -axis, ϕ is the azimuthal angle, measured starting from
¹⁶⁴ the x -axis positive direction, and η is the pseudorapidity, defined by the following equation:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) , \quad (2.4)$$

¹⁶⁵ where θ is the polar angle. The used of pseudorapidity is preferred over the polar angle
¹⁶⁶ because differences in pseudorapidity are Lorentz invariant under boosts along the z -axis.
¹⁶⁷ In the limit of ultrarelativistic particles the pseudorapidity coincides with the rapidity y :

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) , \quad (2.5)$$

¹⁶⁸ where E is the particle energy and p_z is the momentum projection along the z -axis.

¹⁶⁹ The schematic view of the CMS detector, which has a length of 21.5 m, a diameter of
¹⁷⁰ 15 m and a weight of about 14000 tons, is shown in Fig. 2.3. From the inner region to the
¹⁷¹ outer one, the various CMS sub-detectors are:

- **Silicon tracker:** it occupies the region $r < 1.2$ m and $|\eta| < 2.5$. It is composed of an inner silicon pixel vertex detector and a surrounding silicon microstrip detector, with a total active area of about 215 m². It is used to reconstruct charged particle tracks and vertices;
- **Electromagnetic calorimeter (ECAL):** placed in the region $1.2 \text{ m} < r < 1.8$ m and $|\eta| < 3$, it consists of many scintillating crystals of lead tungstate (PbWO_4). It is used for the measurement of the trajectory and the energy released by electrons and photons;
- **Hadronic calorimeter (HCAL):** it is placed in the region $1.8 \text{ m} < r < 2.9$ m and $|\eta| < 5$. It is made up of brass layers alternated with plastic scintillators and it is used to measure the direction and energy deposited by the hadrons produced in the interactions;
- **Superconducting solenoidal magnet:** it occupies the region $2.9 \text{ m} < r < 3.8$ m and $|\eta| < 1.5$ and generates an internal uniform magnetic field with an intensity of 3.8 T, pointing along the direction of the beams. The magnetic field is necessary to bend the trajectories of charged particles, in order to allow the measurement of their momentum through the curvature observed in the tracking system. The magnetic field lines are closed by an external 21 m long iron yoke, that has a diameter of 14 m. Outside the return yoke, a residual 1.8 T magnetic field is present, pointing in the opposite direction with respect to the internal field;
- **Muon system:** the outermost system, which is placed in the region $4 \text{ m} < r < 7.4$ m and $|\eta| < 2.4$, has the purpose of reconstructing the tracks of muons passing through

194
195
196
197

it. It consists of Drift Tubes (DT) in the barrel region and Cathode Strip Chambers (CSC) in the endcaps. A complementary system of Resistive Plate Chambers (RPC)
is used both in the barrel and endcaps. The muon chambers are housed inside the
iron structure of the return yoke.

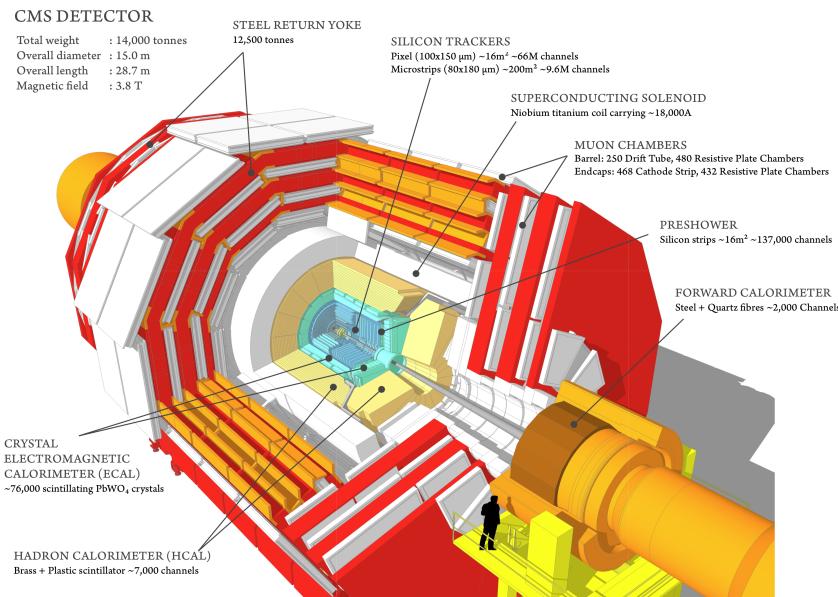


Figure 2.3: Schematic view of the CMS detector showing its sub-detectors.

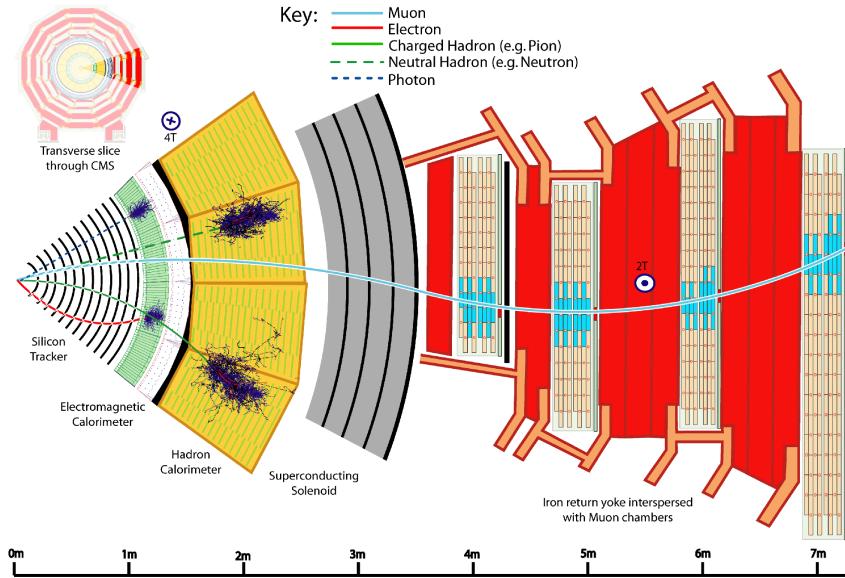


Figure 2.4: Schematic view of a slice of the CMS detector, showing the sub-detectors response to the passage of different types of particles.

198 In Fig. 2.4 the response of the various CMS sub-detectors to the passage of different types
199 of particles is sketched. In the following sections a brief description of each sub-detector is
200 given.

201 2.2.1 The solenoid

202 The CMS magnet [14], which contains the tracker, the electromagnetic and the hadronic
203 calorimeters, is the biggest superconducting solenoid ever built. The solenoid can generate
204 a magnetic field of 3.8 T in the internal bore, which has a diameter of 6 m and a length of
205 12.5 m. The energy stored in the magnet is about 2.7 GJ at full current. The superconductor
206 is made of four Niobium-Titanium layers and it is cooled down to about 4 K through a
207 liquid Helium cooling plant. In case of a quench, when the magnet loses its superconducting
208 property, the energy is dumped to resistors within 200 ms. The magnet return yoke of
209 the barrel is composed with three sections along the z -axis; each one is split into 4 layers
210 (holding the muon chambers in the gaps). Most of the iron volume is saturated or nearly
211 saturated, and the field in the yoke is about the half (1.8 T) of the field in the central
212 volume.

213 2.2.2 The tracker

214 The silicon tracker is the detector closest to the beam collision point. Its goal is the
215 high resolution reconstruction of the trajectories of charged particles originating from the
216 interaction point and the identification of the position of secondary vertices produced by
217 particles with a short mean life time (in particular hadrons containing the b quark, that
218 decay after few hundred of μm). The events produced in the proton proton collisions can
219 be very complex and track reconstruction is an entangled pattern recognition problem.
220 Indeed, at the nominal instantaneous luminosity of operation, an average of about 20 pile up
221 events overlapping to the event of interest are expected, leading to about 1000 tracks to be
222 reconstructed every 25 ns. In order to make the pattern recognition easier, two requirements
223 are fundamental:

- 224 • a low occupancy detector;
- 225 • a large redundancy of the measured points (*hits*) per track.

226 The first requirement is achieved building a detector with high granularity². The redundancy
227 of the hits is instead achieved having several detecting layers, and is necessary to reduce the
228 ambiguity on the assignment of the hits to a given track. Nevertheless, the amount of tracker
229 material has to be as low as possible, in order to avoid compromising the measurement

2The granularity of a detector is defined as the angular range ($\Delta\eta \times \Delta\phi$) that each individual element is able to resolve.

of the particle trajectory. An excessive amount of material would indeed deteriorate the measurement, mainly because of the increased probability of particle multiple scattering. The outer detectors such as ECAL would be influenced by the material as well, for example because of the increased probability for a photon to convert to an electron-positron pair in the tracker material. For this reasons, the tracker layers are limited in number and thickness.

The tracker comprises a large silicon strip detector with a small silicon pixel detector inside it. In the central η region, the pixel tracker consists of three co-axial barrel layers at radii between 4.4 cm and 10.2 cm and the strip tracker consists of ten co-axial barrel layers extending outwards to a radius of 110 cm. Both sub-detectors are completed by endcaps on either side of the barrel, each consisting of two disks in the pixel tracker, and three small plus nine large disks in the strip tracker. The endcaps extend the acceptance of the tracker up to $|\eta| < 2.5$. A three-dimensional schematic view of the tracker is shown in Fig. 2.5, while in Fig. 2.6 a pictorial representation of a slice of the tracker is displayed, showing the various layers of the sub-detectors.

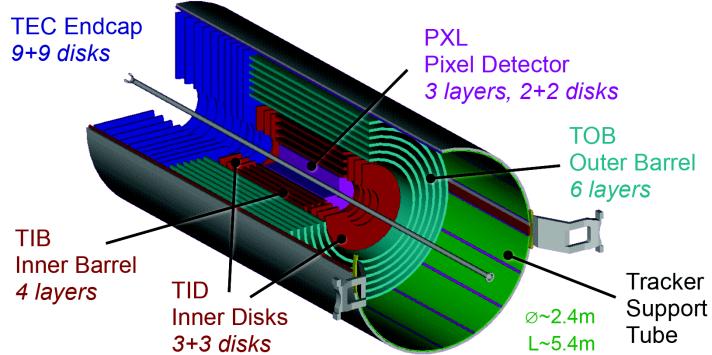


Figure 2.5: Three-dimensional schematic view of the CMS silicon tracker.

The whole tracker has a cylindrical shape with a length of 5.8 m and a diameter of 2.5 m, with the axis aligned to the beams direction. The average number of hits per track is 12-14, in order to have a high reconstruction efficiency and a low rate of fake tracks.

The material budget of the tracker as obtained from a simulation of the detector is shown in Fig. 2.7, reported both in units of radiation length t/X_0 and in units of nuclear interaction length t/λ_I , as a function of η . The region $1 < |\eta| < 2$ exhibits a larger material budget due to the presence of cables and services.

252 The pixel detector

The pixel detector, shown in Fig. 2.8, is mainly used as starting point in the CMS track reconstruction and is of fundamental importance for the reconstruction of primary and

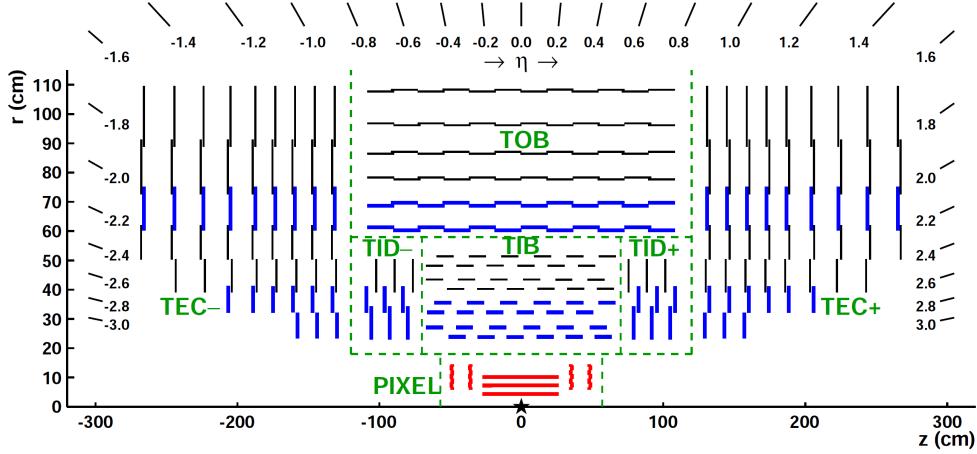


Figure 2.6: Pictorial view of a tracker slice in the r - z plane. Pixel modules are shown in red, single-sided strip modules are depicted as black thin lines and strip stereo modules are shown as blue thick lines.

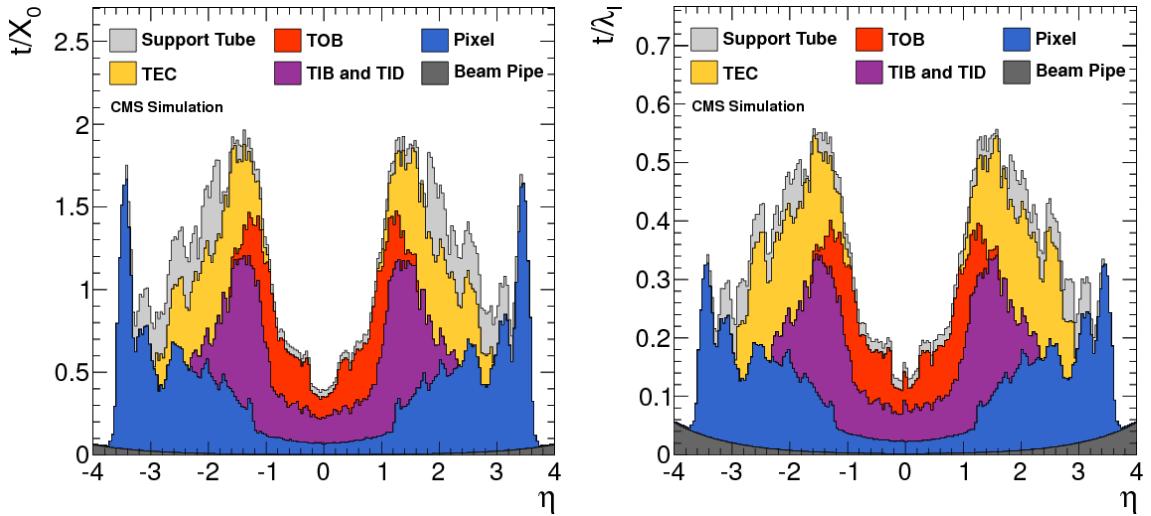


Figure 2.7: Total thickness t of the tracker material expressed in units of X_0 (left) and λ_I (right), as a function of η . The contribution to the total material budget of each part of the detector is shown.

secondary vertices. The pixel detector is placed in the closest position to the collision point, where the amount of radiation is larger. It is placed in the region $|\eta| < 2.5$ and consists of three cylindrical layers 53 cm long in the barrel region, located at $r = 4.4, 7.3$ and 10.2 cm, and two pairs of endcap disks with radii between 6 and 15 cm at $z = \pm 34.5$ and ± 46.5 cm, covering a total area of about 1 m^2 . The detector is composed of many modules, for a total of 768 in the barrel and 672 in the endcaps. Each endcap is composed of 24 segments, each one tilted with respect to the adjacent ones and containing 7 modules. Each module

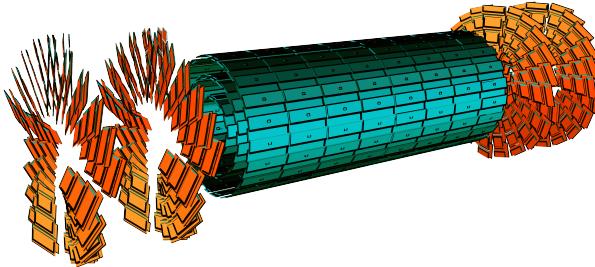


Figure 2.8: Schematic view of the CMS pixel detector.

262 consists of several units which contain a highly segmented silicon sensor with a thickness of
 263 $250\text{ }\mu\text{m}$. In order to achieve an optimal vertex position resolution in both the (r, ϕ) and
 264 z -coordinates, a design with a rectangular pixel shape with an area of $150 \times 100\text{ }\mu\text{m}^2$ was
 265 adopted, with the $100\text{ }\mu\text{m}$ size oriented along the (r, ϕ) direction in the barrel region, and
 266 along the z -direction in the endcap region. The achievable hit reconstruction resolution is
 267 about $10 - 15\text{ }\mu\text{m}$ in the barrel and $15\text{ }\mu\text{m}$ in the endcaps.

268 The microstrip detector

269 In this region of the detector the radiation flow is low enough to allow the use of a
 270 less segmented device, such as the silicon microstrip detector. The microstrip tracker is
 271 composed of 15148 silicon modules, covering a total area of about 193 m^2 with a total of
 272 9.3 million strips. Two types of modules are installed: single sided modules consist of one
 273 sensor stucked onto a carbon fiber support together with the readout electronics, with the
 274 silicon strips laying along the z direction in the barrel and along the (r, ϕ) direction in the
 275 endcaps. The other type of module, referred to as stereo-module, consists of two sensors
 276 stucked together back to back and tilted of a relative angle of 100 mrad . This combination
 277 allows a three-dimensional measurement of the particle interaction point, providing the
 278 information along the z -direction. The whole microstrip tracker is 5.4 m long and extends
 279 up to $r = 1.1\text{ m}$. As the pixel detector, the microstrip detector consists of a barrel and an
 280 endcap region and is divided into four distinct parts, as shown in Fig. 2.6. The barrel is
 281 made up of the following parts:

- 282 • TIB (*Tracker Inner Barrel*): it consists of four cylindrical coaxial layers, covering the
 283 region up to $|z| < 65\text{ cm}$. In this region the detectors have a thickness of $300\text{ }\mu\text{m}$ and
 284 the strips are separated by a variable pitch between 80 and $120\text{ }\mu\text{m}$. The first two
 285 layers are composed of stereo modules while the other layers have single-sided modules.
 286 Since the strips are oriented along the z axis, the position resolution is more precise
 287 in the (r, ϕ) direction, about $23 - 34\text{ }\mu\text{m}$, with respect to the z direction, where a
 288 resolution of about $230\text{ }\mu\text{m}$ is obtained thanks to the stereo modules.

- TOB (*Tracker Outer Barrel*): it consists of six cylindrical coaxial layers, placed in the region $55 \text{ cm} < r < 65 \text{ cm}$ and $|z| < 110 \text{ cm}$. Stereo modules are mounted on the two inner layers. Since the density of particles passing through this region is lower with respect to the TIB, the pitch between the strips is larger ($120 - 180 \mu\text{m}$) and the strips are longer (190 mm). The spatial resolution varies in the range $25 - 52 \mu\text{m}$ in the (r, ϕ) direction, and is about $530 \mu\text{m}$ in the z coordinate in the stereo modules.

The endcaps are also made up of two parts:

- TID (*Tracker Inner Disk*): it consists of six disks, three per side, placed orthogonally with respect to the beam axis, between the TIB and the TOB. The modules are positioned in a ring shape, with the strips oriented in the radial direction, and they are alternately placed on the internal and on the external side of the disk. The two innermost rings of the TID are equipped with stereo modules. The thickness of the silicon is 300 μm .
- TEC (*Tracker EndCap*): each one of the two TEC is made of nine disks which extend to the region $120 \text{ cm} < |z| < 280 \text{ cm}$. Each disk is divided into 8 slices in each of which a number ranging from 4 to 7 modules are mounted in a ring shape, depending on the position along z . Also in this case the modules are alternately mounted on the internal and on the external side of the disk, with the strips radially oriented. On the two innermost rings and on the fifth one the stereo modules are installed to measure the z coordinate. The thickness of the sensors range between 300 and 500 μm depending on the disk.

The tracker is operated at low temperature in order to reduce those radiation damage induced effects that have a temperature dependence, such as the increase of the leakage current and the long-term increase of the depletion voltage (also called reverse annealing)³.

The alignment of the tracker modules is very important to obtain a high spatial resolution. Deviations are caused by assembly inaccuracy, deformations due to cooling and stress from the magnetic field. Therefore, three methods are used for the tracker alignment. The geometry was determined during the assembly to an accuracy of 80 to 150 μm . An infrared laser system is used for continuous monitoring of the position of selected tracker modules. The final alignment is done with tracks from well known physics processes, e.g. cosmic muons or muon pairs from the J/Ψ , Υ or Z decays.

³The tracker in Run 1 was operated at a temperature of +4°C, but during the Long Shutdown 1 a new cooling dry gas plant has been installed and the tracker is now operating at the lower temperature of -15°C.

320 2.2.3 The electromagnetic calorimeter (ECAL)

321 The main function of an electromagnetic calorimeter is to identify electrons and photons
 322 and to measure accurately their energy. The CMS electromagnetic calorimeter (ECAL) [15,
 323 16], shown in Fig. ??, is a hermetic homogeneous calorimeter with cylindrical geometry,
 324 composed of many scintillating crystals of lead tungstate (PbWO_4) with a truncated
 325 pyramidal shape. As the other detectors it consists of two parts, the ECAL barrel (EB),
 326 which contains 61200 crystals, and two endcaps (EE) containing 7324 crystals each one.

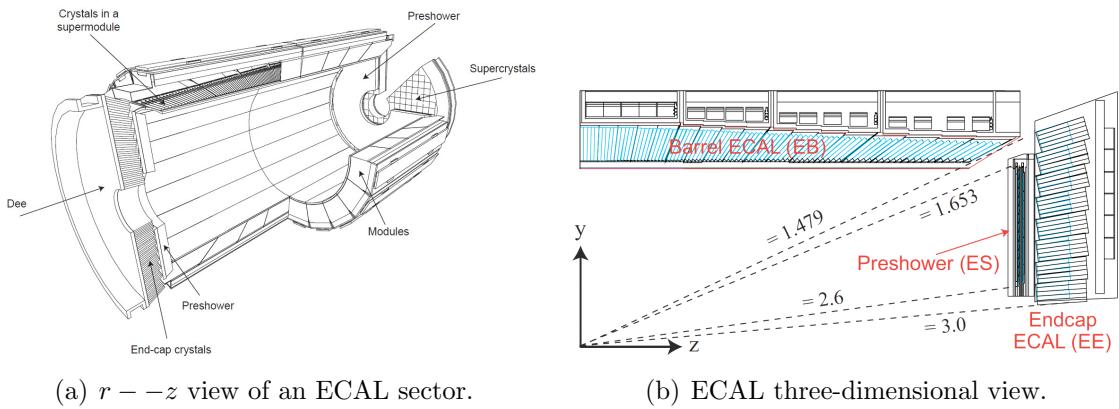


Figure 2.9: Schematic representation of the CMS electromagnetic calorimeter.

327 The characteristics of the PbWO_4 crystals make them an appropriate choice for operation
 328 at LHC. The high density ($\rho = 8.3 \text{ g/cm}^3$), short radiation length ($X_0 = 0.89 \text{ cm}$) and
 329 small Molière radius⁴ (2.2 cm) allow to build a compact and high granularity calorimeter.
 330 Another advantage of this material is the radiation hardness and the fast scintillation
 331 decay time ($\tau = 10 \text{ ns}$), that permits to collect about 80% of the produced light within
 332 the 25 ns interval between two consecutive bunch crossings. The main drawbacks of this
 333 material are the low light yield ($\sim 10 \text{ photoelectrons/MeV}$) and the strong dependence
 334 on the operating temperature, that makes it necessary to keep the crystals at a stabilized
 335 temperature (18°C).

336 The crystals are grouped into 5×5 matrices called *towers*. The barrel has an inner
 337 radius of 129 cm, a length of 630 cm and extends in the region $|\eta| < 1.479$. The crystals
 338 in the barrel have the following dimensions: $22 \times 22 \text{ mm}^2$ at the front face, $26 \times 26 \text{ cm}^2$
 339 at the rear face, and a length of 23 cm, corresponding to $25.8X_0$, and are mounted in a
 340 quasi-projective geometry, in order to have the long side tilted by 3° with respect to the
 341 direction pointing to the interaction point, both in the η and ϕ coordinates. This is done to
 342 avoid the empty spaces between adjacent crystals to be aligned with the direction pointing

⁴The Molière radius R_M characterizes the transverse development of an electromagnetic shower in a calorimeter. On average 90% of the energy deposited by a shower is contained inside a cylinder with radius R_M .

³⁴³ to the interaction point. The granularity of the EB is about 1° . Avalanche photodiodes
³⁴⁴ (APDs) are used as photodetectors connected with the crystals in the barrel region.

³⁴⁵ Each endcap covers the region $1.479 < |\eta| < 3$ and is formed by two semicircular
³⁴⁶ aluminium halves called *dees*. Crystals in endcaps have a length of 22 cm, a frontal area
³⁴⁷ equal to $28.6 \times 28.6 \text{ mm}^2$ and a rear surface of $30 \times 30 \text{ mm}^2$. In the endcaps the crystals
³⁴⁸ are arranged in a $\eta - \phi$ symmetry. The photodetectors used to collect the light produced
³⁴⁹ in the endcap crystals are single stage vacuum phototriodes (VPTs), because this region
³⁵⁰ experiences a rather high particle flux and VPTs are more robust against radiation damages
³⁵¹ with respect to APDs. A preshower system is installed in front of the ECAL endcaps in
³⁵² order to separate the showers produced by a primary γ from those produced by forward
³⁵³ emitted π^0 . This detector, which covers the region $1.653 < |\eta| < 2.6$, is a sampling
³⁵⁴ calorimeter consisting of two lead disks ($2X_0$ and $1X_0$ thick respectively) that initiate the
³⁵⁵ electromagnetic shower from incoming photons or electrons, with silicon strip sensors after
³⁵⁶ each disk, which measure the deposited energy as well as the shower transverse profile.

³⁵⁷ The energy resolution of a homogeneous calorimeter can be expressed by the sum in
³⁵⁸ quadrature of three terms, as shown in the following formula:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2 \quad (2.6)$$

³⁵⁹ The stochastic term a dominates at low energies: it includes the contribution of statistical
³⁶⁰ fluctuations in the number of generated and collected photoelectrons. This term takes
³⁶¹ into account the crystal light emission, the light collection efficiency and the photodetector
³⁶² quantum efficiency⁵. The noise term b includes the contributions of pile up events and
³⁶³ electronic noise, both due to the photodetector and preamplifier. These contributions
³⁶⁴ depend on η and on the LHC operational luminosity. The constant term c , dominant at
³⁶⁵ high energies, takes into account several contributions. The most relevant are the non-
³⁶⁶ uniformity of the longitudinal light collection, the intercalibration errors and the leakage
³⁶⁷ of energy from the rear side of the crystal. The ECAL barrel resolution for electrons was
³⁶⁸ measured using test beams to be:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\% \text{ GeV}^{1/2}}{\sqrt{E}}\right)^2 + \left(\frac{12\% \text{ GeV}^{1/2}}{E}\right)^2 + (0.3\%)^2 , \quad (2.7)$$

³⁶⁹ where E is the energy measured in GeV.

⁵The quantum efficiency is the ratio between the number of collected electron-hole pairs (or photoelectrons)
and the number of photons incident on the photodetector.

370 2.2.4 The hadron calorimeter (HCAL)

371 The hadron calorimeter (HCAL) [hcal] is used together with ECAL to make a complete
372 calorimetric system for the jet energy and direction measurement. Moreover, thanks
373 to its hermetic structure, it can measure the energy imbalance in the transverse plane,
374 E_T^{miss} , a typical signature of non interacting particles, such as neutrinos. The HCAL is
375 a sampling calorimeter covering the region $|\eta| < 5$. As shown in Fig. ??, it is divided
376 in four sub-detectors: HB (*Barrel Hadronic Calorimeter*), located in the barrel region
377 inside the solenoid, extending up to $|\eta| < 1.4$; HE (*Endcap Hadronic Calorimeter*), placed
378 in the endcaps region inside the magnet, covering the region $1.3 < |\eta| < 3$ and partially
379 overlapping with the HB coverage; HO (*Outer Hadronic Calorimeter*), also known as
380 *tail-catcher*, placed along the inner wall of the magnetic field return yoke, just outside of
381 the magnet; HF (*Forward Hadronic Calorimeter*), a sampling calorimeter consisting of
382 quartz fibers sandwiched between iron absorbers, consisting of two units placed in the very
383 forward region ($3 < |\eta| < 5$) outside the magnetic coil. The quartz fibers emit Cherenkov
384 light with the passage of charged particles and this light is detected by radiation resistant
photomultipliers.



Figure 2.10: Longitudinal view of the CMS detector showing the HCAL sub-detectors.

³⁸⁶ **2.2.5 The muon system**

³⁸⁷ **2.3 The CMS trigger system**

³⁸⁸ **2.3.1 The Level-1 trigger**

³⁸⁹ **2.3.2 The high level trigger (HLT)**

³⁹⁰ **2.4 Objects definition and event reconstruction**

³⁹¹ **2.4.1 The Particle Flow technique**

³⁹² **2.4.2 Leptons reconstruction and identification**

³⁹³ **Electrons**

³⁹⁴ **Muons**

³⁹⁵ **2.4.3 Jets reconstruction and identification**

³⁹⁶ **Jet b tagging**

³⁹⁷ **2.5 The CMS framework**

Chapter 3

³⁹⁸ Higgs boson properties in the ³⁹⁹ $H \rightarrow WW$ decay channel

⁴⁰⁰ 3.1 Higgs boson phenomenology at LHC

⁴⁰¹ The discovery of a new boson consistent with the standard model (SM) Higgs boson has
⁴⁰² been reported by ATLAS and CMS Collaborations in 2012. The discovery has been followed
⁴⁰³ by a comprehensive set of studies of properties of this new boson in several production and
⁴⁰⁴ decay channels and no evidence of deviation from the SM expectation has been found so
⁴⁰⁵ far.

⁴⁰⁶ 3.1.1 Higgs boson production mechanisms

⁴⁰⁷ 3.1.2 Higgs boson decay channels

⁴⁰⁸ 3.2 The $H \rightarrow WW$ decay channel

⁴⁰⁹ The CMS studies in the $H \rightarrow WW \rightarrow 2\ell 2\nu$ decay channel include the measurement of the
⁴¹⁰ Higgs properties, as well as constraints on the Higgs total decay width and gauge bosons
⁴¹¹ anomalous couplings.

⁴¹² **3.2.1 The H → WW → 2ℓ2ν signature**

⁴¹³ **3.2.2 Physics objects definition**

⁴¹⁴ **3.2.3 Background processes**

⁴¹⁵ **3.3 Experimental results of the Higgs boson measurements**

⁴¹⁶

Chapter 4

Measurement of the Higgs boson transverse momentum at 8 TeV using $H \rightarrow WW \rightarrow 2\ell 2\nu$ decays

4.1 Introduction

The Higgs boson production at hadron colliders is characterized by p_T^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_T^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_T^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 [17, 18], and at NNLO using the so-called large- m_t approximation [19–21], 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_T^H . For this reason the LO process for Higgs production at $p_T \neq 0$ is at $\mathcal{O}(\alpha_s^3)$, and the counting of perturbative orders differs between inclusive Higgs boson production and p_T^H distribution. Also, NNLO QCD corrections in the p_T^H observable have recently been shown [22].

When $p_T^H \sim m_H$ the QCD radiative corrections to p_T^H differential cross section are theoretically evaluated using fixed-order calculations. When $p_T^H \ll m_H$ the perturbative expansion does not converge due to the presence of large logarithmic terms of the form $\alpha_s^n \ln^{2n} m_H^2/p_T^2$, leading to a divergence of $d\sigma/dp_T$ in the limit of $p_T \rightarrow 0$. For computing the p_T^H spectrum in this region soft-gluon resummation techniques are used, and matched to the fixed-order calculation in the $p_T^H \sim m_H$ region. For the p_T^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

⁴⁴⁴ bottom quark contribution in the fermionic loop can significantly modify the p_T^H shape [23],
⁴⁴⁵ as shown in Fig. ???. Hence, a precise experimental measurement of the p_T^H spectrum is
⁴⁴⁶ important to test the existing SM calculations.

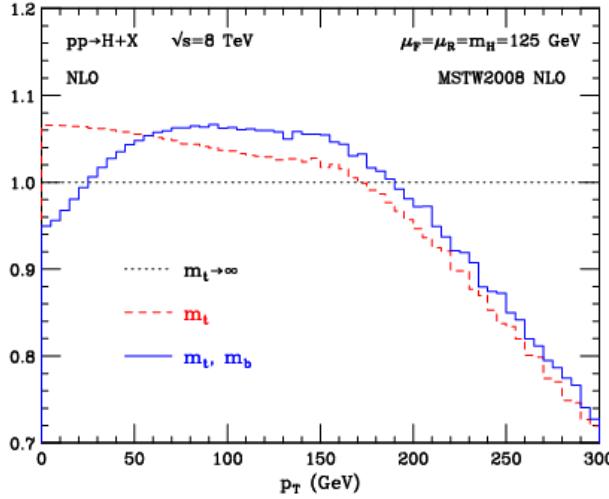


Figure 4.1: p_T^H distribution computed at NLO (α_s^3) and normalized to the calculation obtained in the large- m_t approximation. The red dashed line corresponds to the calculation including the top quark mass while the blue line refers to the calculation including also the bottom quark effects.

⁴⁴⁷ Possible extensions of the SM predict a modification of the Higgs boson couplings to
⁴⁴⁸ gluons and to the top quark. Many of these models actually predict the existence of new
⁴⁴⁹ states that interact with the SM Higgs boson, but are beyond the direct production reach at
⁴⁵⁰ the actual LHC energies. The effect of these new states could however show up as a deviation
⁴⁵¹ of the Higgs boson couplings with respect to the SM expectation. The modification of
⁴⁵² the couplings, as shown in Refs. [24, 25], can change the kinematics of the Higgs boson
⁴⁵³ production and the effect can be particularly sizeable in the tail of the p_T^H distribution.
⁴⁵⁴ Other models, such as Composite Higgs [26], predict the existence of top-partners, which
⁴⁵⁵ are heavy resonances with the same quantum numbers as the top quark, that can interact
⁴⁵⁶ with the Higgs boson in the ggH fermionic loop, changing the p_T^H shape with respect to
⁴⁵⁷ what the SM predicts [27]. The measurement of the p_T^H spectrum is thus a useful tool for
⁴⁵⁸ indirect searches of new particles predicted by theories beyond the SM.

⁴⁵⁹ Measurements of the fiducial cross sections and of several differential distributions, using
⁴⁶⁰ the $\sqrt{s} = 8$ TeV LHC data, have been reported by ATLAS [28–30] and CMS [31, 32] for
⁴⁶¹ the $H \rightarrow ZZ \rightarrow 4\ell$ ($\ell = e, \mu$) and $H \rightarrow \gamma\gamma$ decay channels. In this chapter a measurement of
⁴⁶² the fiducial cross section times branching fraction ($\sigma \times \mathcal{B}$) and p_T spectrum for Higgs boson
⁴⁶³ production in $H \rightarrow WW \rightarrow e^\pm \mu^\mp \nu\nu$ decays, based on $\sqrt{s} = 8$ TeV LHC data, is reported.

⁴⁶⁴ The analysis is performed looking at different flavour leptons in the final state in order
⁴⁶⁵ to suppress the sizeable contribution of backgrounds containing a same-flavour lepton pair
⁴⁶⁶ originating from Z boson decay.

467 Although the $H \rightarrow WW \rightarrow 2\ell 2\nu$ channel has lower resolution in the p_T^H measurement
 468 compared to the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ \rightarrow 4\ell$ channels because of neutrinos in the final
 469 state, the channel has a significantly larger $\sigma \times \mathcal{B}$, exceeding those for $H \rightarrow \gamma\gamma$ by a factor
 470 of 10 and $H \rightarrow ZZ \rightarrow 4\ell$ by a factor of 85 for a Higgs boson mass of 125 GeV [33], and is
 471 characterized by good signal sensitivity. Such sensitivity allowed the observation of a Higgs
 472 boson at the level of 4.3 (5.8 expected) standard deviations for a mass hypothesis of 125.6
 473 GeV using the full LHC data set at 7 and 8 TeV [34].

474 The measurement is performed in a fiducial phase space defined by kinematic require-
 475 ments on the leptons that closely match the experimental event selection.

476 The effect of the limited detector resolution, as well as the selection efficiency with respect
 477 to the fiducial phase space are corrected to particle level with an unfolding procedure [35],
 478 as explained in Sec. ??.

479 4.2 Data sets, triggers and MC samples

480 This analysis relies on the published $H \rightarrow WW$ measurements [34] in terms of code, selections
 481 and background estimates for both the ggH and VBF production mechanisms.

482 4.2.1 Data sets and triggers

483 The data sets used for the analysis correspond to 19.4 fb^{-1} at $\sqrt{s} = 8 \text{ TeV}$ of integrated
 484 luminosity composed of the following CMS data taking periods during 2012: 2012A
 485 (892 pb^{-1}), 2012B (4440 pb^{-1}), and 2012C (6898 pb^{-1}) and 2012D (7238 pb^{-1}). Data have
 486 been checked and validated and only data corresponding to good data taking quality are
 487 considered. The $e^\pm \mu^\mp$ final state is considered in this analysis.

488 For the data samples, the events are required to fire one of the unprescaled single-electron,
 489 single-muon or muon-electron triggers. Due to the rather high LHC instantaneous luminosity
 490 the single-lepton triggers must have high HLT p_T thresholds, otherwise the rate of these
 491 triggers would be too large to be sustained. The double-lepton triggers allow to lower
 492 down the p_T thresholds while keeping a sustainable trigger rate, thus maintaining a good
 493 sensitivity to the Higgs boson signal, for which the lepton p_T can be rather small. A brief
 494 overview of the HLT p_T criteria on the leptons is given in Table ???. While the HLT lepton
 495 p_T thresholds of 17 and 8 GeV for the double lepton triggers accommodate the offline
 496 lepton p_T selection of 20 and 10 GeV, the higher p_T thresholds in the single lepton triggers
 497 help partially recovering double lepton trigger inefficiencies as a high p_T lepton is on average
 498 expected due to the kinematics 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	$p_T > 27$ GeV
Single-Muon	$p_T > 24$ GeV
Muon-Electron	$p_T > 17$ and 8 GeV
Electron-Muon	$p_T > 17$ and 8 GeV

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

$$\epsilon(p_T, \eta) = \frac{N_{\text{pass}}^{\text{probe}}}{N_{\text{pass}}^{\text{probe}} + N_{\text{fail}}^{\text{probe}}} \quad (4.1)$$

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

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

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

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

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

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

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

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

534 4.2.2 Monte Carlo samples

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

- 537 • The first version of the POWHEG program [36–40] (POWHEG V1) provides event
 538 samples for the $H \rightarrow WW$ signal for the gluon fusion (ggH) and VBF production
 539 mechanisms, as well as $t\bar{t}$ and tW processes [41], with NLO accuracy.
- 540 • The $qq \rightarrow W^+W^-$, Drell-Yan, ZZ, WZ, $W\gamma$, $W\gamma^*$, tri-bosons and $W+jets$ processes
 541 are generated using the MADGRAPH 5.1.3 [42] event generator.

- The $gg \rightarrow W^+W^-$ process is generated using the GG2WW 3.1 generator [43] and its cross section is scaled to the approximate NLO prediction [44, 45].
- The VH process is simulated using PYTHIA 6.426 [46].

For leading-order generators samples, the CTEQ6L [47] set of parton distribution functions (PDF) is used, while CT10 [48] is used for next-to-leading order (NLO) ones. Cross section calculations [49] at next-to-next-to-leading order (NNLO) are used for the $H \rightarrow WW$ process, while NLO calculations are used for background cross sections. The $H \rightarrow WW$ 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 [33]. The reweighting of the p_T^H spectrum is achieved by tuning the POWHEG generator, as described in detail in Ref. [50]. Cross sections computed with NLO pQCD accuracy [33] 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. ?? the relative fraction of the four production mechanisms is shown for each p_T^H bin.

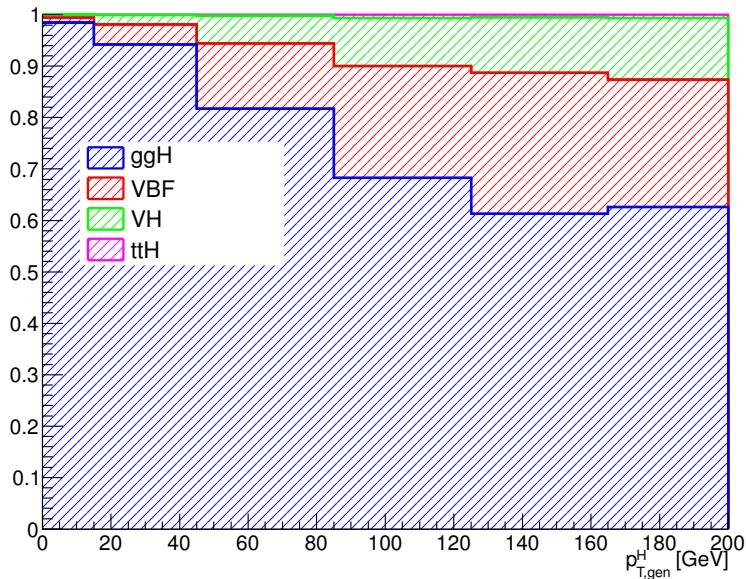


Figure 4.2: Relative fraction of ggH, VBF, VH and ttH in each bin of the Higgs boson transverse momentum.

For all processes, the detector response is simulated using a detailed description of the CMS detector, based on the GEANT4 package [51].

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. ??, trigger efficiency, and lepton reconstruction and identification efficiencies [34].

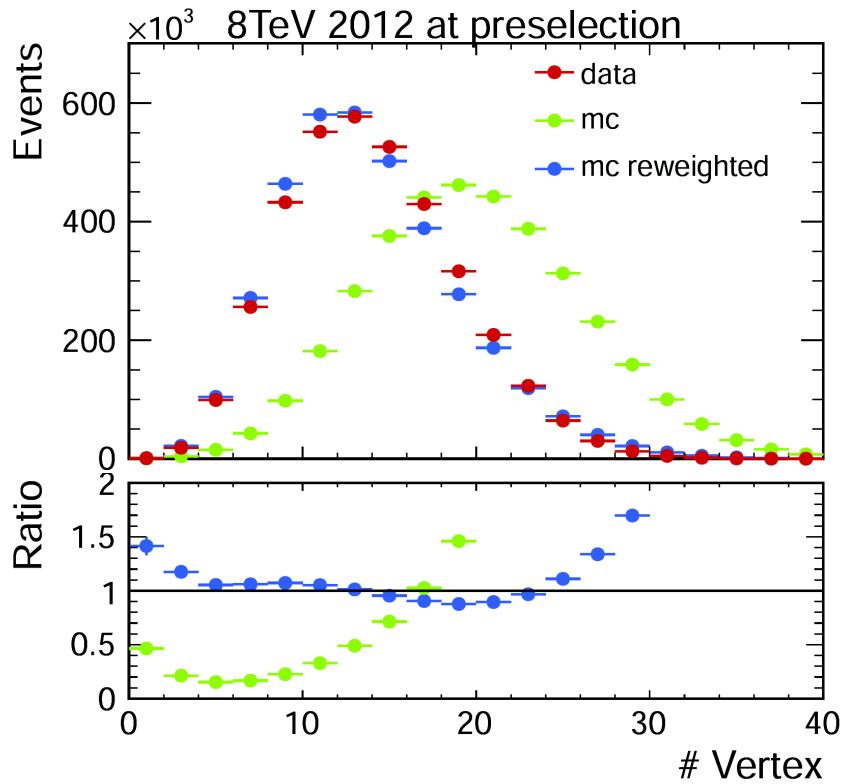


Figure 4.3: Distribution of the number of vertices in data and in simulation, before and after applying the pile-up reweighting.

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 [23, 52] and the second version of the POWHEG generator (POWHEG V2) [53]. 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

579 renormalization and factorization scale central values at a Higgs boson mass of 125 GeV.
 580 The cross section normalization is scaled, to take into account electroweak corrections, by a
 581 factor of 1.05 and the effects of threshold resummation by a factor of 1.06 [54, 55]. The
 582 upper and lower bounds of the uncertainties are obtained by scaling up and down both the
 583 renormalization and the factorization scales by a factor of two. The POWHEG V2 generator
 584 is a matrix element based generator that provides a NLO description of the ggH process
 585 in association with zero jets, taking into account the finite mass of the bottom and top
 586 quarks. The POWHEG prediction is tuned using the POWHEG damping factor $hdump$ of
 587 104.17 GeV, in order to match the p_T^H spectrum predicted by HRES in the full phase space.
 588 This factor reduces the emission of additional jets in the high p_T regime, and enhances the
 589 contribution from the Sudakov form factor in the limit of low p_T . The POWHEG generator
 590 is interfaced to the JHUGEN generator version 5.2.5 [56–58] for the decay of the Higgs
 591 boson to a W boson pair and interfaced with PYTHIA 8 [59] for the simulation of parton
 592 shower and hadronization effects.

593 4.3 Analysis Strategy

594 The analysis presented here is based on that used in the previously published $H \rightarrow WW \rightarrow 2\ell 2\nu$
 595 measurements by CMS [34], modified to be inclusive in the number of jets. This modifica-
 596 tion significantly reduces the uncertainties related to the modelling of the number of jets
 597 produced in association with the Higgs boson.

598 4.3.1 Event reconstruction and selections

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

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

609 The efficiencies for the identification and isolation of the electrons and muons are
 610 measured in data and in simulation selecting a pure sample of leptons coming from the
 611 $Z \rightarrow \ell\ell$ decay, and using a Tag and Probe technique very similar to the one described in
 612 Sec. ?? for the trigger efficiency. In this case, the probe lepton is defined by loose isolation
 613 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 classified into zero jet, one jet and VBF topologies by counting jets within $|\eta| < 4.7$ and for $p_T > 30 \text{ GeV}$.

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

Background events from $t\bar{t}$ 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 $p_T > 20 \text{ GeV}$ then the muon is required to be non-isolated with $ISO/p_T > 0.1$.

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.

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

652 1. **Lepton preselection:**

- 653 • at least two opposite-charge and opposite-flavour ($e\mu$) isolated leptons recon-
 654 structed in the event;
- 655 • $|\eta| < 2.5$ for electrons and $|\eta| < 2.4$ for muons;
- 656 • $p_T > 20$ GeV for the leading lepton. For the trailing lepton, the transverse
 657 momentum is required to be larger than 10 GeV.

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

660 3. **E_T^{miss} preselection:** particle flow E_T^{miss} is required to be greater than 20 GeV.

661 4. **projected E_T^{miss} selection:** minimum projected E_T^{miss} required to be larger than
 662 20 GeV.

663 5. **Di-lepton mass cut:** $m_{\ell\ell} > 12$ GeV in order to reject low mass resonances and QCD
 664 backgrounds.

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

666 7. **Transverse mass:** $m_T > 60$ GeV to reject Drell-Yan to $\tau\tau$ events.

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

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

680 A cut-flow plot is reported in Fig. ??, showing the effect of each selection using signal
 681 and background simulations. In the first bin, labelled as “No cut”, no selection is applied
 682 and the bin content corresponds to the total expected number of events with a luminosity
 683 of 19.4 fb^{-1} . All the events in this bin have at least two leptons with a loose transverse
 684 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_T^{miss} and m_T , with respect to the analysis selection.

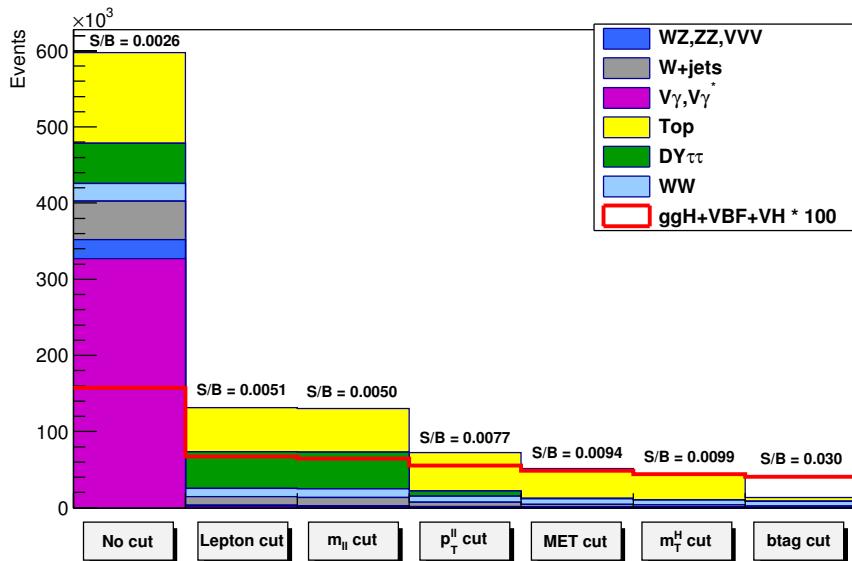


Figure 4.4: Effect of single selections on MC samples. The signal (red line) is multiplied by 100 and superimposed on stacked backgrounds. In each bin, corresponding to a different selection, is reported the expected number of events in MC at a luminosity of 19.46 fb^{-1} .

The fiducial phase space used for the cross section measurements is defined at the particle level by the requirements given in Table ???. 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_T^H bin. For the VH signal process, the two leptons are required to originate from the $H \rightarrow WW \rightarrow 2\ell 2\nu$ 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	$p_T > 20 \text{ GeV}$
Subleading lepton p_T	$p_T > 10 \text{ GeV}$
Pseudorapidity of electrons and muons	$ \eta < 2.5$
Invariant mass of the two charged leptons	$m_{\ell\ell} > 12 \text{ GeV}$
Charged lepton pair p_T	$p_T^{\ell\ell} > 30 \text{ GeV}$
Invariant mass of the leptonic system in the transverse plane	$m_T^{\ell\ell\nu\nu} > 50 \text{ GeV}$
E_T^{miss}	$E_T^{\text{miss}} > 0$

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

4.3.3 Binning of the p_T^H distribution

Experimentally, the Higgs boson transverse momentum is reconstructed as the vector sum of the lepton momenta in the transverse plane and E_T^{miss} .

$$\vec{p}_T^H = \vec{p}_T^{\ell\ell} + \vec{p}_T^{\text{miss}} \quad (4.5)$$

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_T^{miss} entering the transverse momentum measurement. The effect of the limited E_T^{miss} resolution has two main implications on the analysis strategy. The first one is that the choice of the binning in the p_T^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_T^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

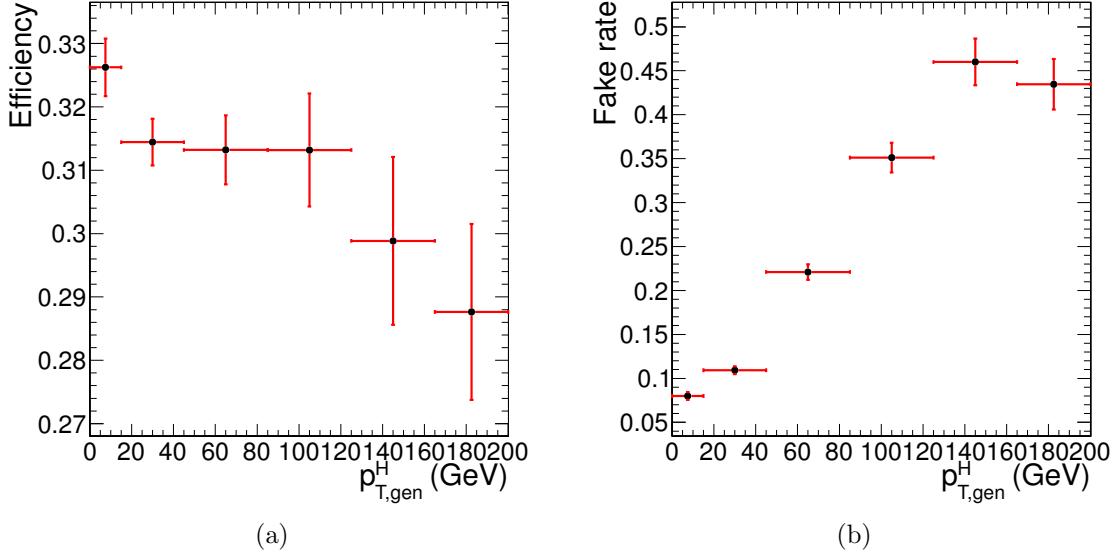


Figure 4.5: Efficiency of the full selection (a) and fake rate (b) as a function of p_T^H .

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

$$P_i = \frac{N_i^{\text{GEN|RECO}}}{N_i^{\text{RECO}}} \quad . \quad (4.6)$$

726 The bin width is chosen in such a way as to make the smallest bins able to ensure a purity
 727 of about 60% on a ggH signal sample. Following this prescription we have divided the whole
 728 p_T^H range in the following six bins: [0-15 GeV], [15-45 GeV], [45-85 GeV], [85-125 GeV],
 729 [125-165 GeV], [165-∞ GeV].

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

740 If a 4π acceptance is defined, requiring just that the Higgs decays to WW and then to
 741 $2\ell 2\nu$, the efficiency becomes $\epsilon = 0.0396 \pm 0.0003$ and the fake rate is zero.

742 4.4 Background estimation

743 4.4.1 Top quark background

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

747 The general strategy for determining the residual top events in the signal region is to
 748 first measure the top tagging efficiencies from an orthogonal region of phase space in data.
 749 The orthogonal phase space is defined inverting the b-veto requirement of the signal region,
 750 in such a way to have a control region enriched in top quark events. Then, using this
 751 efficiency, the number of events with the associated uncertainty is propagated from the
 752 control region to the signal region. The number of surviving top events in the signal region
 753 would then be:

$$N_{b\text{veto}}^{\text{signal}} = N_{b\text{tag}}^{\text{control}} \cdot \frac{1 - \epsilon_{\text{top}}}{\epsilon_{\text{top}}} \quad (4.7)$$

754 where $N_{b\text{tag}}^{\text{control}}$ is the number of events in the control region and ϵ_{top} is the efficiency as
 755 measured in data.

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

758 0-jets category

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

$$\epsilon_{\text{tag}} = \frac{N_{\text{tag}}^{\text{control}}}{N^{\text{control}}} \quad , \quad (4.8)$$

762 where N^{control} is the number of events in the top control phase space defined requiring
 763 one b-tagged jet with $p_T > 30 \text{ GeV}$, and $N_{\text{tag}}^{\text{control}}$ is the subset of those events that pass
 764 either the soft muon tagging or the low- p_T b jet tagging. The purity of this control sample,
 765 as estimated from simulation, is about 97%. The remaining 3% background contribution is

766 estimated from simulation and subtracted from the numerator and denominator of Eq. (??).
767 The efficiency $\epsilon_{\text{top}}^{0-jet}$ can then be estimated using the following formula:

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

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

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

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

775 Category with more than 0 jets

776 The strategy for the estimation of the top background in events with at least one jet with
777 p_T greater than 30 GeV is the following. First of all the efficiency for tagging a b jet is
778 measured both in data and simulation and the values are used to correct the simulation for
779 different b-tagging efficiencies in data and simulation. This evaluation is performed in a
780 control region, called CtrlTP, containing at least two jets, using a Tag&Probe technique.
781 The procedure to extract these scale factors is presented in Sec. ???. Then a larger statistics
782 control region, CtrlDD, is defined by requiring at least one b-tagged jet and we use the
783 simulation, corrected for the previously computed b-tagging efficiency scale factor, to derive
784 the factor that connects the number of events in CtrlDD to the number of events in the
785 signal region. This second step is explained in detail in Sec. ???.

786 Tag&Probe

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

The CrtlTP control region is defined selecting the events which pass the lepton preselection cuts listed in Sec. ??, 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:

$$N_{TPP} = N_s \epsilon_s + N_b \epsilon_b \quad (4.11)$$

812

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

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. ???. The relevant efficiencies are:

$$\epsilon_s^{MC} = 0.7663 \pm 0.0072 \quad (4.13)$$

821

$$\epsilon_b^{MC} = 0.208 \pm 0.015 \quad (4.14)$$

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 fraction 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. ???. The plots show the pull of the efficiencies measured in the fit, where the pull variable for each toy i is defined as:

$$pull(\epsilon_{s(b)}) = \frac{\epsilon_{s(b)}^{\text{true}} - \epsilon_{s(b)}^i}{\sigma(\epsilon_{s(b)}^i)} \quad (4.15)$$

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. ??

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. ?? and shows good agreement.

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

$$\epsilon_s^{Data} = 0.769 \pm 0.022 \quad (4.16)$$

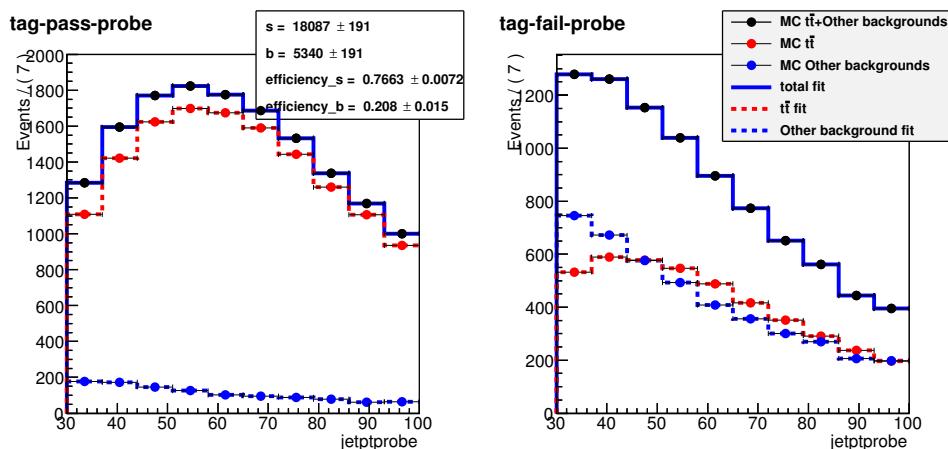


Figure 4.6: Simultaneous fit of the *tag-pass-probe* and *tag-fail-probe* pairs in the MC.

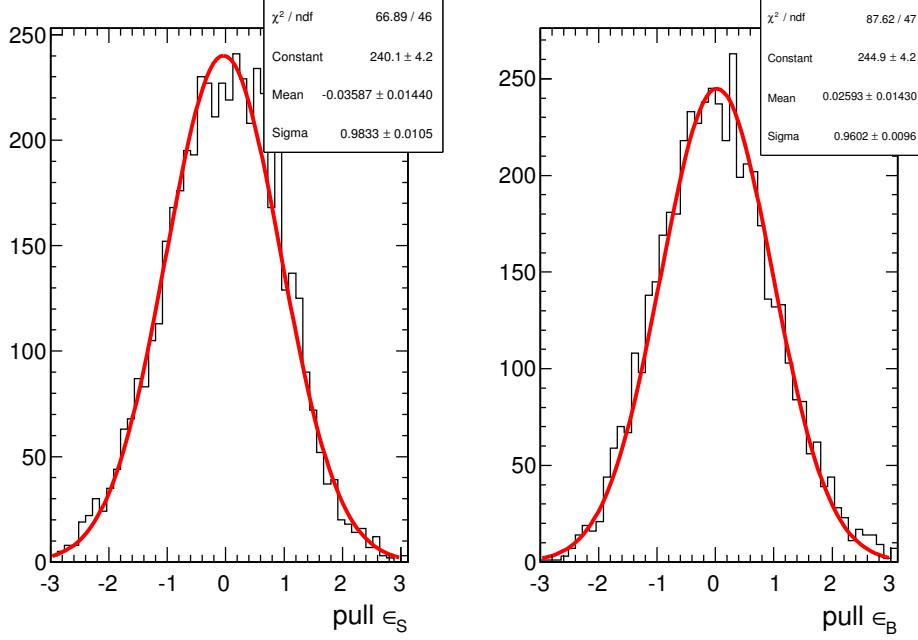


Figure 4.7: Pulls of the ϵ_s and ϵ_b parameters in 5000 toy MC.

$$\epsilon_b^{Data} = 0.121 \pm 0.054 \quad (4.17)$$

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

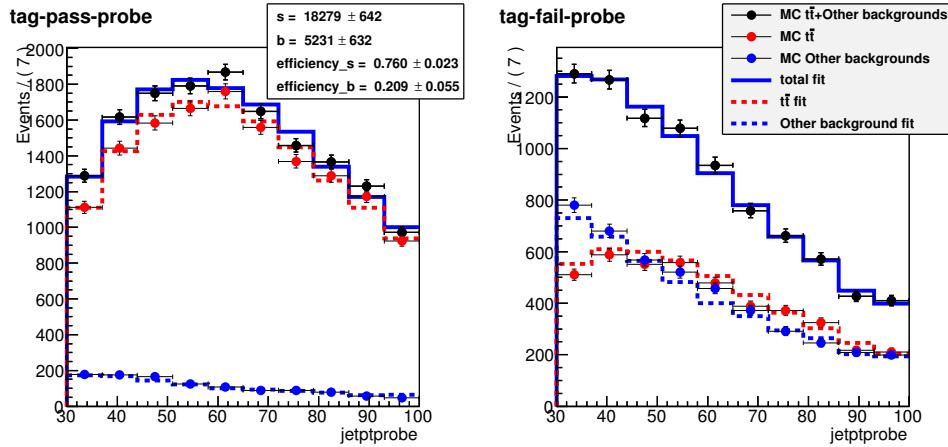


Figure 4.8: Fit of a toy MC sample.

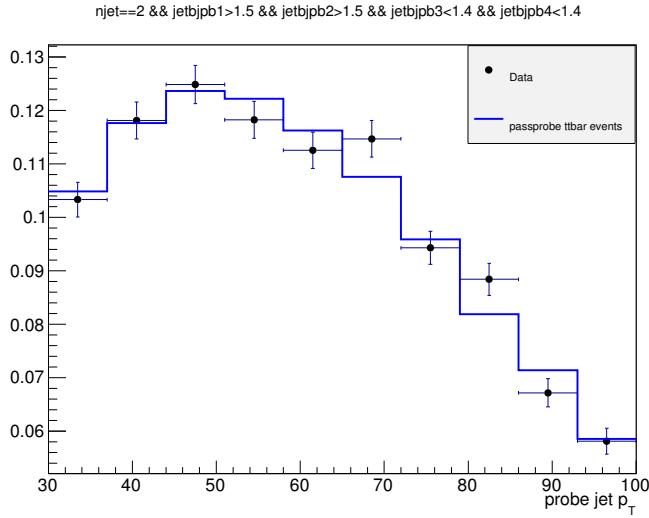


Figure 4.9: Shape comparison for the *probe p_T* spectrum in data and in MC in a very pure $t\bar{t}$ sample.

847 Data driven estimation

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

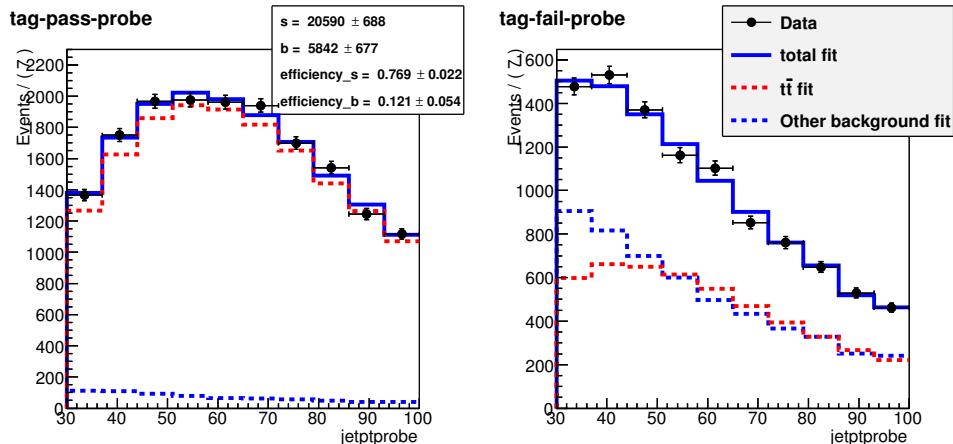


Figure 4.10: Simultaneous fit of the *tag-pass-probe* and *tag-fail-probe* pairs in data.

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

$$\alpha = \frac{N_{tt\ MC}^{SIG}}{N_{tt\ MC}^{CtrlDD}} . \quad (4.18)$$

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

$$N_{tt\ Data}^{SIG} = \alpha N_{tt\ Data}^{CtrlDD} . \quad (4.19)$$

In evaluating α and its error the b-tagging efficiencies determined in Sec. ?? 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.

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

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

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. ?? and Eq. ?? 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. ??, where the p_T^H distribution is shown in the CtrlDD region normalized to the cross section measured by a specific CMS analysis [60]. 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. ??.

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. ??

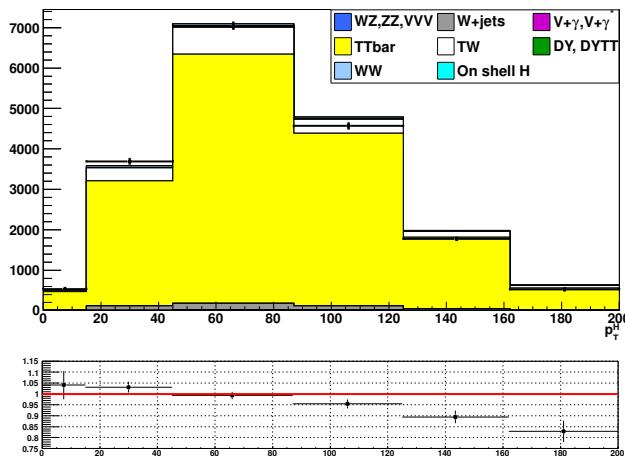


Figure 4.11: p_T^H distribution in the CtrlDD control region.

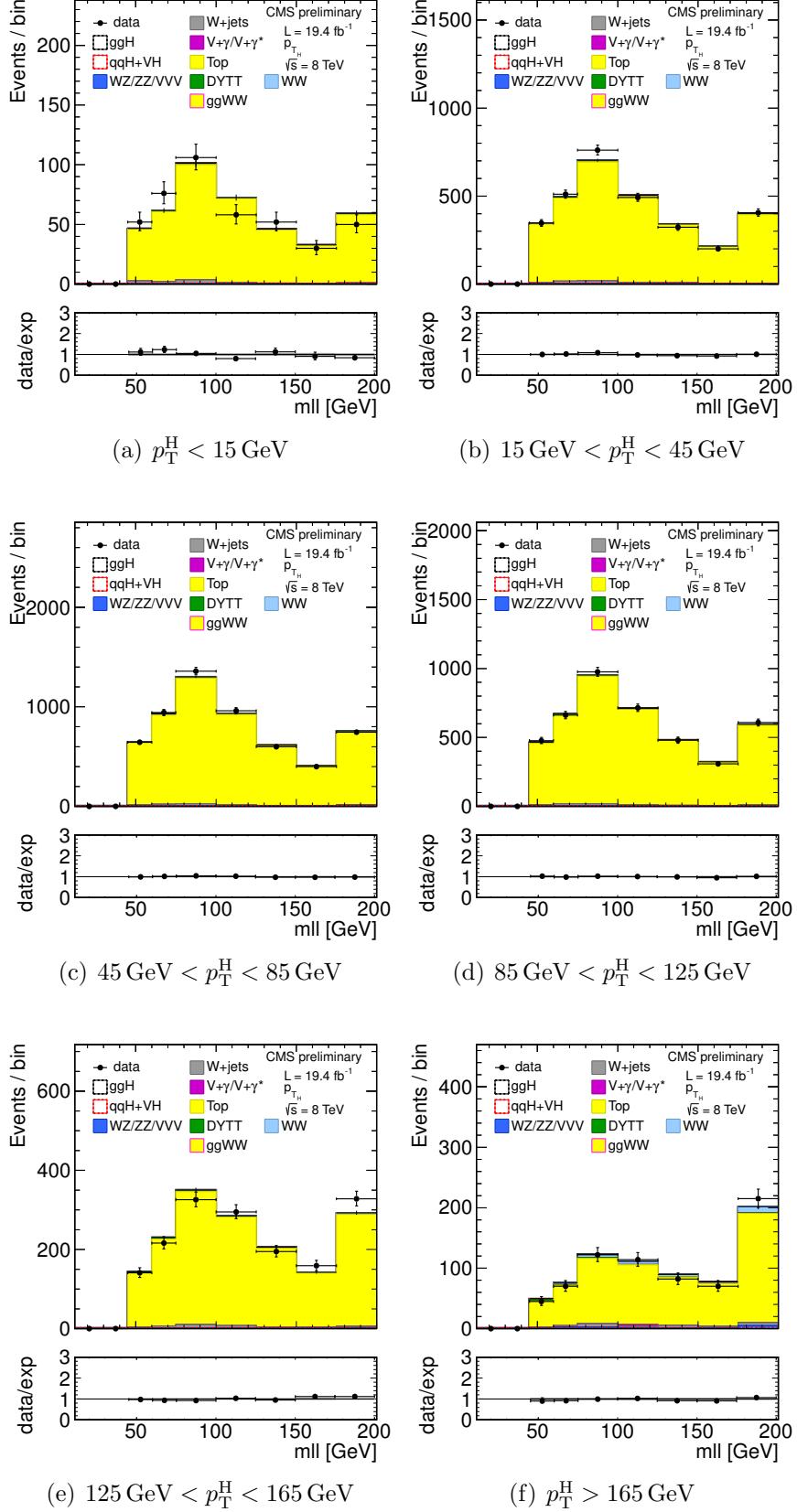


Figure 4.12: $m_{\ell\ell}$ distributions in the CtrlDD region for the different p_T^H bins.

882 4.4.2 WW background

883 For what the $q\bar{q} \rightarrow W^+W^-$ background shape is concerned, the prediction from the simulation
 884 is used. This background is divided into six different parts, corresponding to the six
 885 p_T^H bins defined in the analysis. The normalization of the $q\bar{q} \rightarrow W^+W^-$ background is left
 886 free to float in each bin, in such a way to adjust it in order to match the data during the fit
 887 procedure. In this way the shape difference between the p_T^{WW} theory prediction and the
 888 distribution provided by the simulation, which is obtained with the MADGRAPH generator,
 889 is minimized.

890 In figure ?? a comparison is shown between the p_T^{WW} spectra of two different $q\bar{q} \rightarrow W^+W^-$
 891 samples: one obtained with the MADGRAPH generator and the other after applying to the
 892 same distribution a reweighting in order to match the theoretical prediction at NLO+NNLL
 893 precision.

894 A shape discrepancy can be clearly observed and the effect becomes larger at high values
 895 of p_T^H . In order to assess the effect of this discrepancy on the shapes of the variables used for
 896 the signal extraction, $m_{\ell\ell}$ and m_T , the shapes have been checked in all p_T^H bins, comparing
 897 different MC samples. The MADGRAPH sample used for the nominal shape is compared
 898 to the MADGRAPH sample with NLO+NNLL reweighting, a POWHEG sample with NLO
 899 accuracy and an AMC@NLO sample. The results of this comparison are shown in figures
 900 ?? and ?? . The shape discrepancy among the different models is included as an additional
 901 systematic uncertainty.

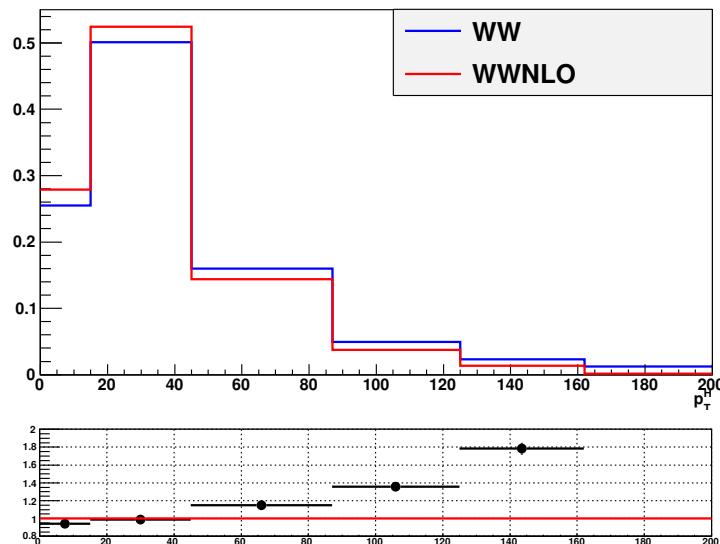


Figure 4.13: Comparison between the p_T^{WW} distributions obtained with two different MC generators: the blue line corresponds to the MADGRAPH generator and the red line refers to the same sample in which a reweighting has been applied in order to match the theoretical prediction at NLO+NNLL precision.

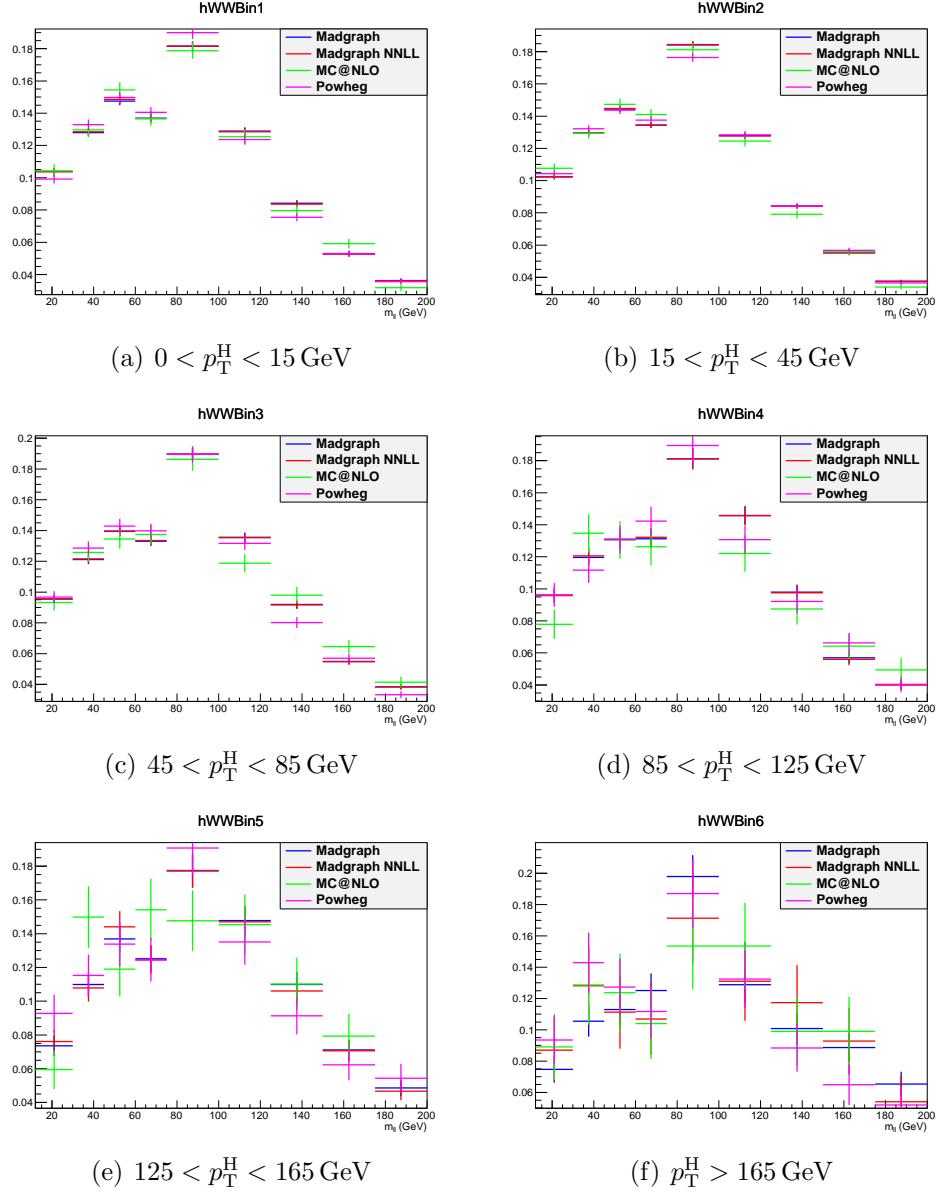


Figure 4.14: Comparison between the default WW background sample and other theoretical models for the $m_{\ell\ell}$ distributions in every p_T^H bin.

The gluon-induced WW process, i.e. $gg \rightarrow W^+W^-$, 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 [44, 45].

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 $m_{\ell\ell}$

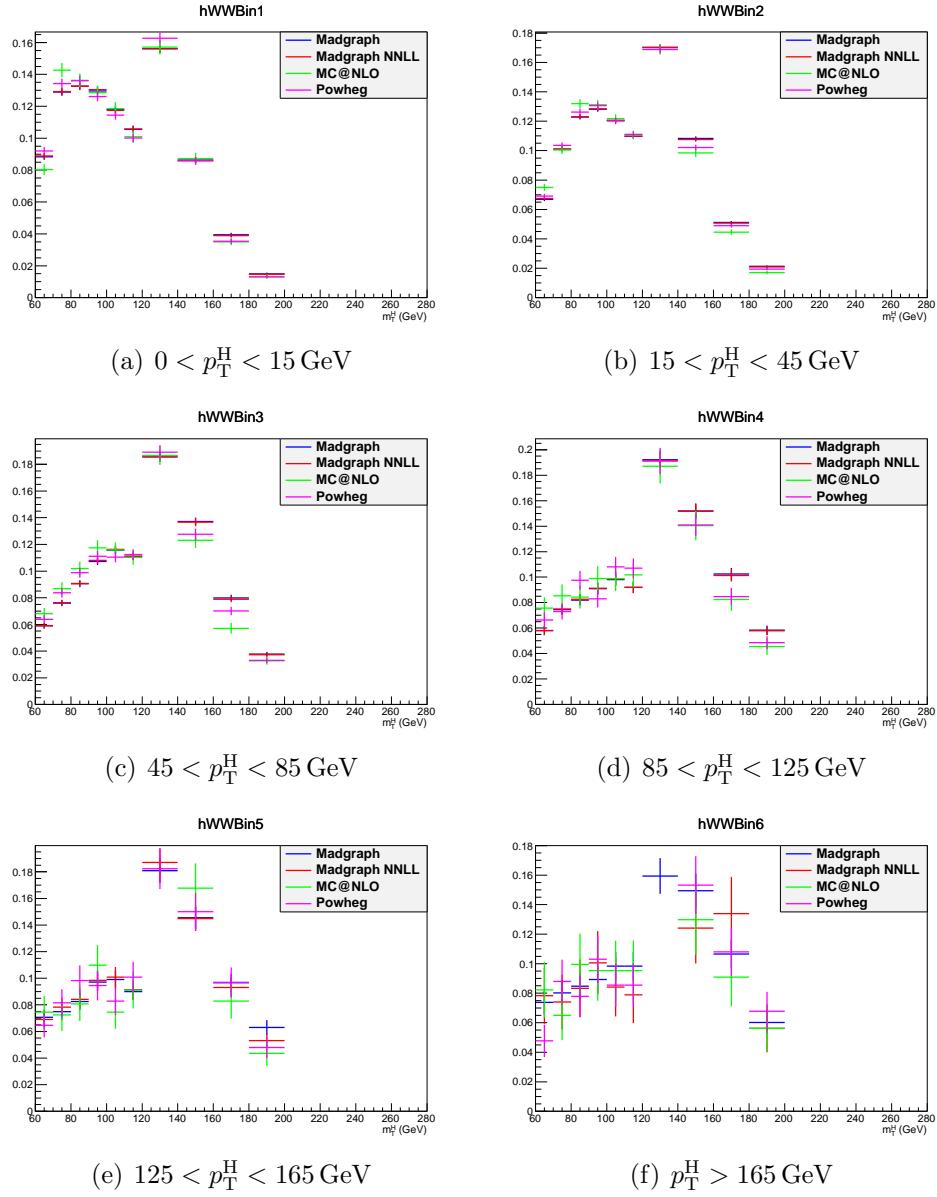


Figure 4.15: Comparison between the default WW background sample and other theoretical models for the m_T distributions in every p_T^H bin.

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

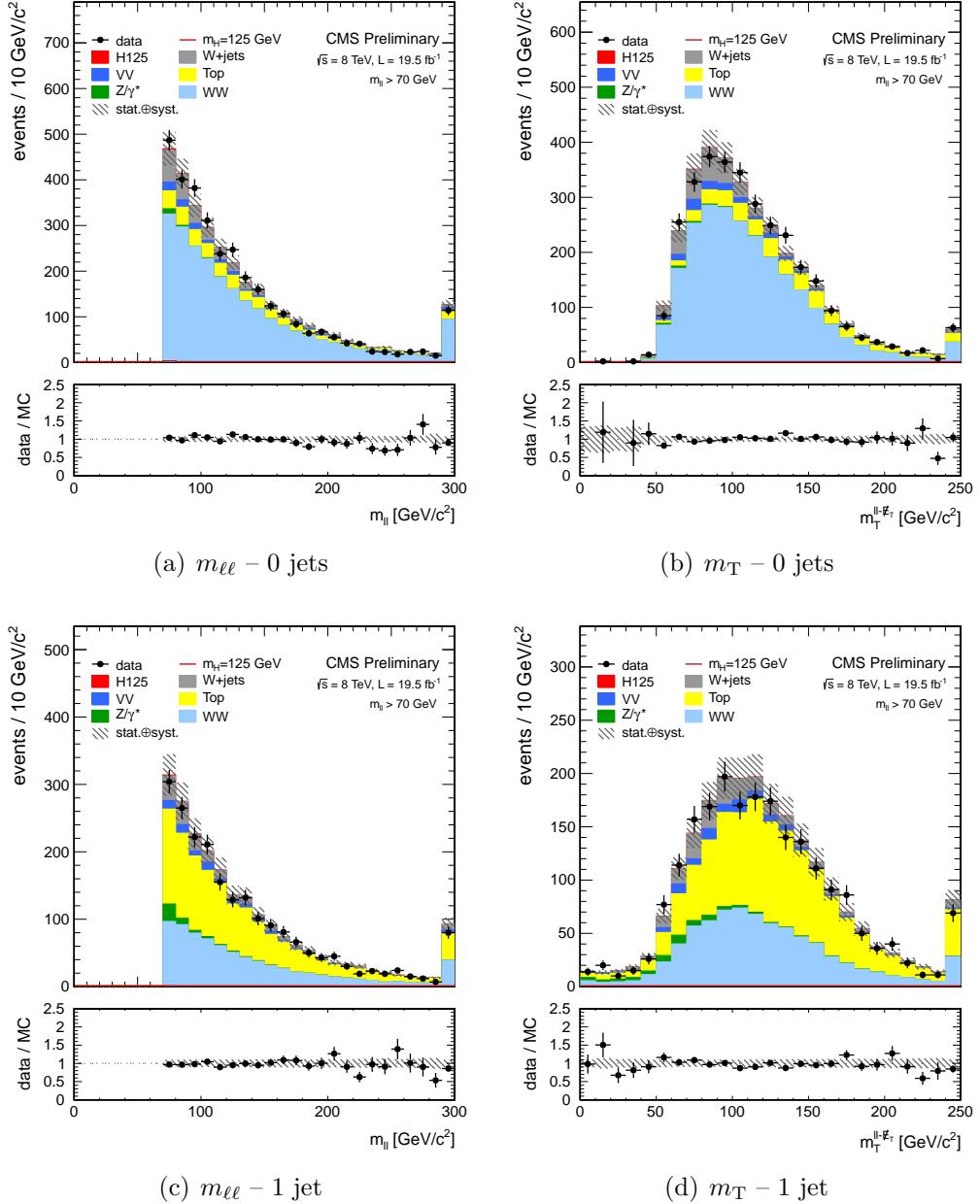


Figure 4.16: Comparison of the $m_{\ell\ell}$ and m_T shapes in data and simulation for events with zero and 1 jets, inclusive in p_T^H . The events are required to pass the analysis requirements and, in order to define a signal-free control region, to have $m_{\ell\ell} > 70$ GeV.

910 4.4.3 Other backgrounds

911 W+jets background

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

916 The idea is to estimate the background containing one or two fake leptons selecting
 917 events with relaxed lepton quality criteria, i.e. looser with respect to the selections used at
 918 the analysis level, and computing the efficiencies for real and fake leptons to pass the tight
 919 lepton quality requirements of the analysis. A data-driven approach is pursued to estimate
 920 this background. A set of loosely selected lepton-like objects, referred to as the “fakeable
 921 object” or “denominator” from here on, is defined in a data set of events dominated by
 922 dijet production. To measure the fake rate we count how many fakeable objects pass the
 923 full lepton selection of the analysis, parameterized as a function of the phase space of the
 924 fakeable lepton, therefore it is extracted in bins of η and p_T . The ratio of the fully identified
 925 lepton, referred as “numerator”, to the fakeable objects is taken as the probability for a
 926 fakeable object to fake a lepton:

$$927 \text{Fake Rate} = \frac{\text{\#of fully reconstructed leptons}}{\text{\#of fakeable objects}} \quad (4.22)$$

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

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

- 933 • $\sigma_{\text{inj}\eta} < 0.01(0.03)$ for barrel (endcap);
- 934 • $|\Delta\phi_{\text{in}}| < 0.15(0.10)$ for barrel (endcap);
- 935 • $|\Delta\eta_{\text{in}}| < 0.007(0.009)$ for barrel (endcap);
- 936 • $H/E < 0.12(0.10)$ for barrel (endcap);
- 937 • electron conversion rejection;
- 938 • $|d_0| < 0.02 \text{ cm};$

- 939 • $\frac{\sum_{\text{trk}} E_T}{p_T^{\text{ele}}} < 0.2$;
- 940 • $\frac{\sum_{\text{ECAL}} E_T}{p_T^{\text{ele}}} < 0.2$;
- 941 • $\frac{\sum_{\text{HCAL}} E_T}{p_T^{\text{ele}}} < 0.2$.

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

- 944 • $|d_0| < 0.02$ cm;
- 945 • MVA isolation output > -0.6 .

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

958 In addition to the fake rate, also a prompt lepton rate is evaluated, defined as the
 959 probability of a prompt lepton passing the loose requirements to also pass the tight analysis
 960 selections. The prompt rate is also measured in data, defining a control region enriched in
 961 $Z \rightarrow \ell\ell$ events, selecting dilepton events with an invariant mass of the two leptons in the Z
 962 peak mass region.

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

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

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.(??).

The prompt and fake rate estimations after the removal of the EW contribution are shown in Tables ?? and ?? 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.

988 Drell-Yan to $\tau\tau$ background

The low E_T^{miss} threshold in the $e\mu$ final state requires the consideration of the contribution from $Z/\gamma^* \rightarrow \tau^+\tau^-$, that is 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 [34], 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]	$0 < \eta \leq 1.4442$	$1.4442 < \eta \leq 1.566$	$1.566 < \eta$
$10 < p_T \leq 15$	0.5738 ± 0.0045	0.5366 ± 0.0204	0.2947 ± 0.0047
$15 < p_T \leq 20$	0.7091 ± 0.0020	0.5484 ± 0.0185	0.4477 ± 0.0034
$20 < p_T \leq 25$	0.7175 ± 0.0013	0.6297 ± 0.0067	0.6200 ± 0.0001
$25 < p_T \leq 50$	0.9219 ± 0.0002	0.8404 ± 0.0007	0.8509 ± 0.0001
$p_T > 50$	0.9693 ± 0.0002	0.9398 ± 0.0021	0.9385 ± 0.0005

Muon prompt rate		
p_T range [GeV]	$0 < \eta \leq 1.5$	$1.5 < \eta \leq 2.5$
$10 < p_T \leq 15$	0.7119 ± 0.0003	0.7582 ± 0.0006
$15 < p_T \leq 20$	0.8049 ± 0.0018	0.8495 ± 0.0001
$20 < p_T \leq 25$	0.9027 ± 0.0008	0.8948 ± 0.0012
$25 < p_T \leq 50$	0.9741 ± 0.0001	0.9627 ± 0.0002
$p_T > 50$	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 [61] 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 $W\gamma$ 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.

p_T range [GeV]	electron fake rate			
	$0 < \eta \leq 1$	$1 < \eta \leq 1.479$	$1.479 < \eta \leq 2$	$2 < \eta \leq 2.5$
$10 < p_T \leq 15$	0.045 ± 0.005	0.033 ± 0.004	0.008 ± 0.002	0.021 ± 0.005
$15 < p_T \leq 20$	0.044 ± 0.003	0.049 ± 0.003	0.017 ± 0.001	0.017 ± 0.002
$20 < p_T \leq 25$	0.041 ± 0.002	0.064 ± 0.003	0.025 ± 0.002	0.025 ± 0.002
$25 < p_T \leq 30$	0.059 ± 0.003	0.101 ± 0.005	0.041 ± 0.003	0.043 ± 0.003
$30 < p_T \leq 35$	0.084 ± 0.006	0.111 ± 0.009	0.058 ± 0.006	0.066 ± 0.005

p_T range [GeV]	muon fake rate			
	$0 < \eta \leq 1$	$1 < \eta \leq 1.479$	$1.479 < \eta \leq 2$	$2 < \eta \leq 2.5$
$10 < p_T \leq 15$	0.131 ± 0.002	0.154 ± 0.004	0.194 ± 0.005	0.241 ± 0.009
$15 < p_T \leq 20$	0.143 ± 0.007	0.191 ± 0.012	0.235 ± 0.016	0.308 ± 0.027
$20 < p_T \leq 25$	0.198 ± 0.005	0.239 ± 0.009	0.221 ± 0.011	0.271 ± 0.021
$25 < p_T \leq 30$	0.182 ± 0.011	0.228 ± 0.018	0.195 ± 0.022	0.287 ± 0.045
$30 < p_T \leq 35$	0.170 ± 0.021	0.244 ± 0.036	0.195 ± 0.041	0.289 ± 0.111

1009 The $W\gamma^{(*)}$ background, where the photon decays to an electron-positron pair, is expected
 1010 to be very small, thanks to the stringent photon conversion requirements. This background
 1011 also includes events where a real photon is produced in association with the W boson.
 1012 These events constitute a background for this analysis because the photon can interact with
 1013 the tracker material converting to an electron-positron pair.

1014 Since the WZ simulated sample has a generation level cut on the di-lepton invariant mass
 1015 ($m_{\ell\ell} > 12$ GeV) and the cross-section raises quickly with the lowering of this threshold, a
 1016 dedicated MADGRAPH sample has been produced with lower momentum cuts on two of
 1017 the three leptons ($p_T > 5$ GeV) and no cut on the third one. The surviving contribution
 1018 estimated with this sample is still very small, and since the uncertainty on the cross-section
 1019 for the covered phase space is large, a conservative 100% uncertainty has been given to it.
 1020 A k -factor for $W\gamma^*$ of 1.5 ± 0.5 based on a dedicated measurement of tri-lepton decays,
 1021 $W\gamma^* \rightarrow e\mu\mu$ and $W\gamma^* \rightarrow \mu\mu\mu$, is applied [34]. The contribution of $W\gamma^{(*)}$ is also constrained
 1022 by a closure test with same sign leptons on data, which reveals a good compatibility of the
 1023 data with the expected background.

1024 4.5 Systematic uncertainties

1025 Systematic uncertainties play an important role in this analysis where no strong mass peak
1026 is expected due to the presence of undetected neutrinos in the final state. One of the most
1027 important sources of systematic uncertainty is the normalization of the backgrounds that
1028 are estimated on data control samples whenever is possible.

1029 A summary of the main sources of systematic uncertainty and the corresponding estimate
1030 is reported in Table ???. A detailed description of each source of systematic uncertainty is
1031 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%
$W + \text{jets}$	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%

1032 4.5.1 Background normalization uncertainties

1033 The signal extraction is performed subtracting the estimated backgrounds to the event
1034 counts in data. This uncertainty depends on the background:

1035 • **tt** and tW backgrounds: The efficiency on jets b-tagging is estimated using the
1036 Tag&Probe technique in data and simulation control regions, as explained in ???. A
1037 per-jet scale factor, which takes into account the possibly different efficiency of the
1038 anti b-tagging selection in data and simulation, is computed by means of the efficiency
1039 measured with the Tag&Probe method. The Tag&Probe method has been used also
1040 to measure the mistag rates in data and simulation, which are the probability to b-tag
1041 a jet that is not produced by the hadronization of a b quark. These factors are used
1042 to reweigh the Top MC samples as explained in ???. The uncertainties provided by the
1043 Tag&Probe fit are then propagated to the factor α that is used in the top data driven
1044 estimation ???. These uncertainties are embedded in a systematic error that affects
1045 the shape of the Top background in each p_T^H bin.

1046 Provided that the simulated samples include both tt> and tW processes, a systematic
1047 uncertainty related to the tW/tt> fraction has been included. In fact, a relative variation
1048 of the contribution of these two processes could modify the shape of the MC sample,
1049 and is thus included as a shape uncertainty affecting the top quark background shape
1050 in each p_T^H bin in a correlated way.

- 1051 • **W+jets background:** It is estimated with data control sample as described in
1052 Sec.???. With 19.4 fb^{-1} at 8 TeV, the uncertainty receives similar contributions from
1053 statistics and systematic error (mainly jet composition differences between the fake
1054 rate estimation sample and the application sample), the total error being about 40%,
1055 dominated by the closure test of the method on a same-sign control region.
- 1056 • **WZ,ZZ,W $\gamma^{(*)}$ backgrounds:** those backgrounds, which are expected to give a
1057 small contribution, are estimated from simulation. Uncertainties on the cross sections
1058 reported in [62, 63] are 4% for WZ and 2.5% for ZZ. A 30% uncertainty is assigned to
1059 the $W\gamma$ [64] yield and another 30% on $W\gamma^{(*)}$ contribution according to the uncertainty
1060 on the normalization study (see Sec. ??).

1061 4.5.2 Experimental uncertainties

1062 The following experimental systematic sources have been taken into account:

- 1063 • **Luminosity:** Using the online luminosity monitoring CMS reached an uncertainty
1064 on the luminosity of 2.6% at 8 TeV.
- 1065 • **Trigger efficiency.** The uncertainties for both electrons and muons are at 1-2% level,
1066 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_T^{miss} modeling:** The E_T^{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 [65]) 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. ??, 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 ??.

- **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 [66, 67] prescription, using CT10, NNPDF2.1 [68] and MSTW2008 [69] 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_T^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 ??.

1123 Jet multiplicity uncertainty

1124 The jet bin uncertainty on the ggH production mode has been evaluated using the Stewart-
 1125 Tackman method, following the recipe proposed in Refs. [33, 70]. Three independent
 1126 nuisance parameters have to be associated with the inclusive ggH production cross sections
 1127 $\sigma_{\geq 0}$, $\sigma_{\geq 1}$ and $\sigma_{\geq 2}$, which corresponds to the cross sections with ≥ 0 jets, ≥ 1 jet and ≥ 2
 1128 jets respectively. According to the agreement on the treatment of uncertainties in the
 1129 combination of ATLAS and CMS results [71], these nuisance parameters are labelled as
 1130 *QCDscale_ggH*, *QCDscale_ggH1in* and *QCDscale_ggH2in*. However, in case the analysis is
 1131 split in exclusive jet multiplicity bins, the jet bin uncertainties can be evaluated taking into
 1132 account the correct correlations among the three nuisances following the Stewart-Tackman
 1133 prescription. Even though this analysis is inclusive in number of jets, the jet binning
 1134 uncertainties must be included due to the presence of the b-jet veto, that introduces a
 1135 dependency of the selection efficiency on the number of jets in the event. The veto efficiency
 1136 has been evaluated in all the p_T^H bins defined in the analysis and as a function of jets
 1137 multiplicity. The results are shown in Figs. ?? and ???. The drop of the veto efficiency at
 1138 high values of p_T^H is due to the correlation with jets multiplicity.

1139 The first step of this procedure is to take the inclusive ggH cross section, σ_{ggH} , and
 1140 to convert the relative QCD up/down scale uncertainties, ϵ_+ and ϵ_- , to a log-normal
 1141 uncertainty, i.e. $\kappa = \sqrt{\exp(\epsilon_+) \cdot \exp(\epsilon_-)}$. The exclusive cross sections, σ_0 , σ_1 and σ_2 , can

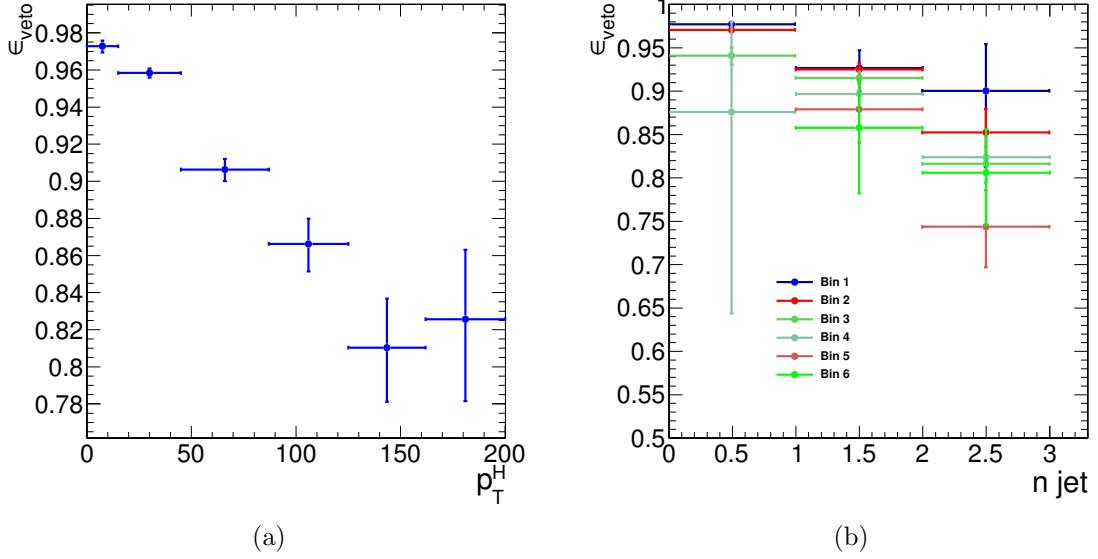


Figure 4.17: (a) Efficiency of the b-tagging veto in different bins of p_T^H . (b) Efficiency of the b-tagging veto in different bins of p_T^H , as a function of number of jets.

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 ??, 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 ?? have then been calculated for each p_T^H bin embedding the b-jet veto efficiency and using the following formulas:

$$QCDscale_{ggH} = \frac{\Delta_{\geq 0}^0 \cdot f_0 \cdot \epsilon_0 + \Delta_{\geq 1}^0 \cdot f_1 \cdot \epsilon_1}{\Delta_{\geq 0}^0 \cdot f_0 \cdot \epsilon_0 + \Delta_{\geq 1}^0 \cdot f_1 \cdot \epsilon_0} \quad , \quad (4.25)$$

$$QCDscale_{ggH1in} = \frac{\Delta_{\geq 1}^1 \cdot f_1 \cdot \epsilon_1 + \Delta_{\geq 2}^1 \cdot f_2 \cdot \epsilon_2}{\Delta_{\geq 1}^1 \cdot f_1 \cdot \epsilon_1 + \Delta_{\geq 2}^1 \cdot f_2 \cdot \epsilon_1} \quad , \quad (4.26)$$

$$QCDscale_{ggH2in} = 1 \quad , \quad (4.27)$$

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	$\Delta_{\geq 0}^0 = (\kappa_{\geq 0})^{\frac{1}{f_0}}$		
QCDscale_ggH1in	$\Delta_{\geq 1}^0 = (\kappa_{\geq 1})^{-\frac{f_1+f_2}{f_0}}$	$\Delta_{\geq 1}^1 = (\kappa_{\geq 1})^{\frac{f_1+f_2}{f_1}}$	
QCDscale_ggH2in		$\Delta_{\geq 2}^1 = (\kappa_{\geq 2})^{-\frac{f_2}{f_1}}$	$\Delta_{\geq 2}^2 = (\kappa_{\geq 2})$

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 ?? for each p_T^H bin.

Table 4.8: Values of the jet binning nuisance parameters for different p_T^H bins.

Nuisance parameter	p_T^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.

1166 4.5.5 Treatment of systematic uncertainties in the shape analysis

1167 One can distinguish between normalization uncertainties, where a systematic effect is
 1168 changing the normalization of a given process assuming the shape is not affected, and shape
 1169 uncertainties where the actual change in the shape of the distribution is taken into account.
 1170 The normalization uncertainties enter the shape analysis as a constant normalization factor,
 1171 whereas for shape uncertainties the nominal and the $+1\sigma$ and -1σ shapes enter the analysis
 1172 in form of three histograms with the same normalization.

1173 For the W+jets background, the shape differences for different jet p_T thresholds in the
 1174 di-jet control sample are considered separately for electron and muon fakes, while the other
 1175 sources of systematics are taken as normalization uncertainties as in the cut-based analysis.

1176 Effects from experimental uncertainties are studied by applying a scaling and smearing
 1177 of certain variables of the physics objects, followed by a subsequent recalculation of all
 1178 the correlated variables. This is done for simulation, to account for possible systematic
 1179 mis-measurements of the data. All experimental sources from Section ?? but luminosity are
 1180 treated both as normalization and shape uncertainties. For background with a data-driven
 1181 normalization estimation, only the shape uncertainty is considered.

1182 To account for statistical uncertainties, for each distribution going into the shape analysis,
 1183 the $+1\sigma$ and -1σ shapes were obtained by adding/subtracting the statistical error in each
 1184 bin and renormalizing it to the nominal distribution. In addition to this procedure a
 1185 constant normalization uncertainty due to the finite statistics of the MC sample used to
 1186 extract the shape is assigned.

1187 4.6 Signal extraction

1188 According to the “blinding” policy of the CMS Collaboration, the strategy of the analysis
 1189 has been scrutinized and approved by a selected committee of internal reviewers before
 1190 looking at the data in the signal region. This approach prevents the analysts from being
 1191 biased by the data in the developing phase of the analysis. Below are shown the results
 1192 after having looked at the data.

1193 4.6.1 Fitting procedure

1194 The signal, including ggH, VBF, and VH production mechanisms, is extracted in each bin
 1195 of p_T^H by performing a binned maximum likelihood fit simultaneously in all p_T^H bins to a
 1196 two-dimensional template for signals and backgrounds in the $m_{\ell\ell}$ – m_T plane. The variables
 1197 used for the two-dimensional template are chosen for their power to discriminate signal

and background contributions. This is shown in Fig. ??, where the two-dimensional MC distributions are shown for the signal and background processes in the 0-jets category.

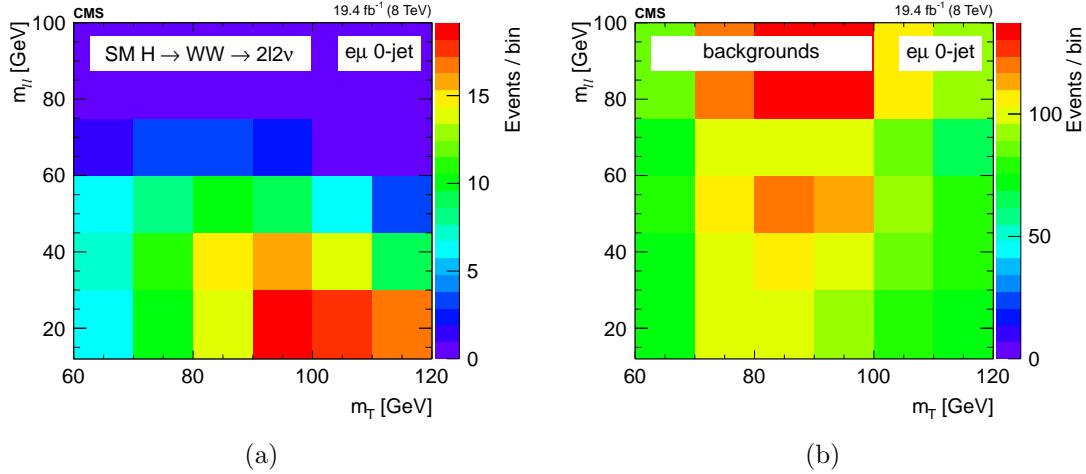


Figure 4.18: Two-dimensional $m_{\ell\ell}$ - m_T distribution for signal (a) and background (b) processes in the 0-jets category.

Six different signal strength parameters are extracted from the fit, one for each p_T^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_T^H . This has been verified and the correlation between the discriminating variables and p_T^H is shown in Fig. ?? and Fig. ?? 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, $\mu = \sigma/\sigma_{SM}$, 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 $H \rightarrow WW$ 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_T^H distribution of the out-of-fiducial signal events is taken from

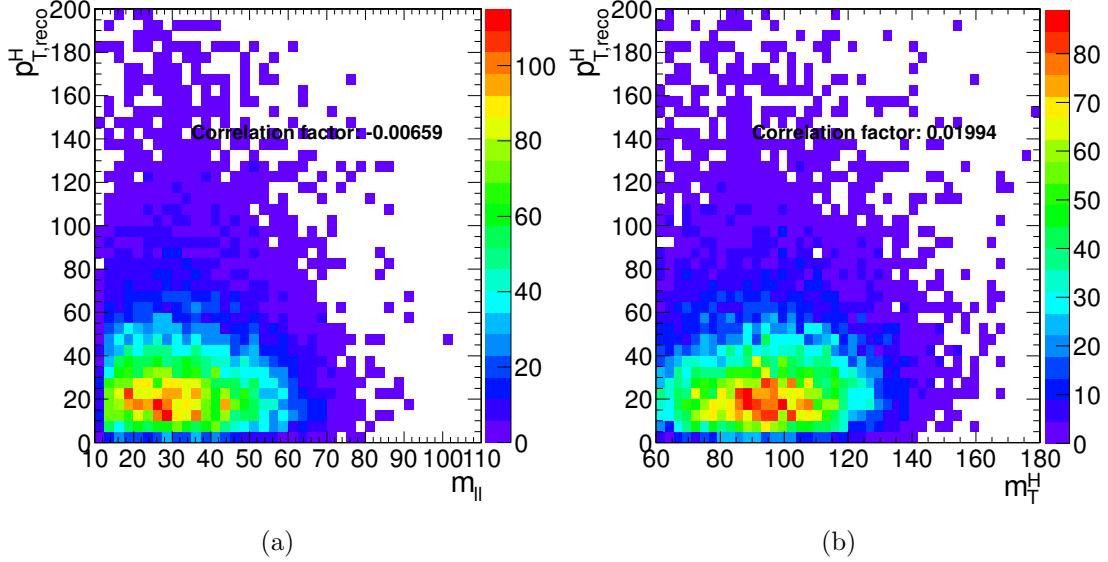


Figure 4.19: Correlation between p_T^H and $m_{\ell\ell}$ (a) and between p_T^H and m_T (b) after the full selection for the ggH production mode.

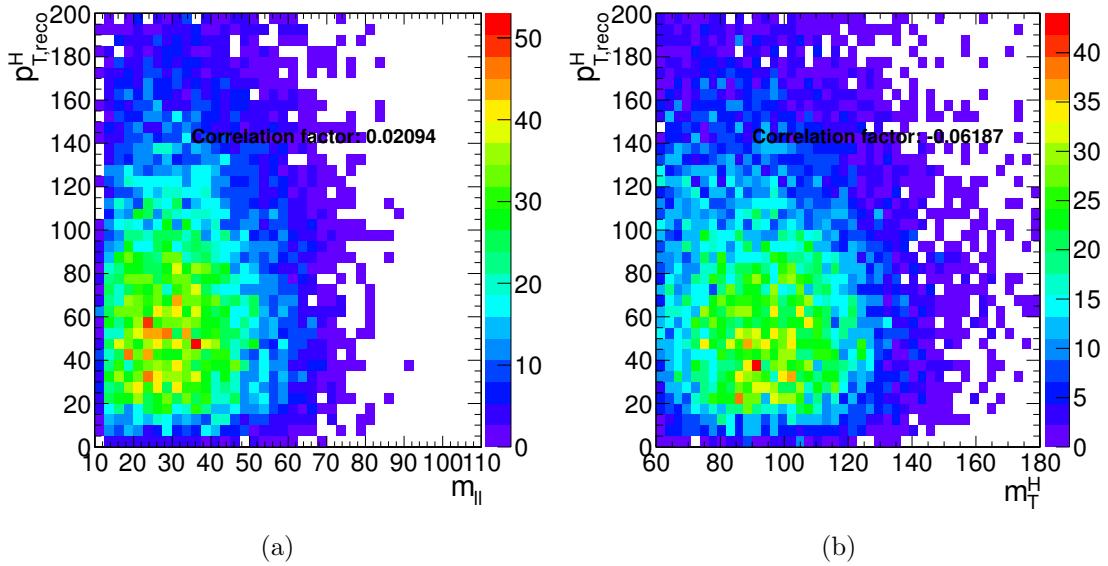


Figure 4.20: Correlation between p_T^H and $m_{\ell\ell}$ (a) and between p_T^H and m_T (b) after the full selection for the VBF production mode.

¹²²⁰ simulation, and each bin is multiplied by the corresponding measured signal strength before
¹²²¹ performing the subtraction.

1222 At the end, the number of events in each bin i of the measured spectrum is:

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

1223 where s_i and f_i are respectively the number of signal and fake events expected from
 1224 simulation and μ_i is the measured signal strength.

1225 The fit makes use of the binned maximum likelihood approach. The likelihood function,
 1226 \mathcal{L} , restricted to the p_T^H bin j , can be written as: **CHECK!!**

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

1227 where $data$ corresponds to the experimental observation and μ_j is the signal strength in
 1228 the bin j , i.e. the parameter of interest of the fit, which multiplies the signal yield. The
 1229 index i runs over the bins of the m_{ee} - m_T two-dimensional histogram corresponding the p_T^H
 1230 bin j , s_i and b_i are the expected number of signal and background events respectively in bin
 1231 i , and n_i is the total number of observed events in bin i . The set of parameters θ represents
 1232 the full suite of nuisance parameters used to incorporate the systematic uncertainties. Each
 1233 nuisance parameter is constrained in the fit including the prior distributions functions $p(\tilde{\theta}|\theta)$
 1234 in the likelihood, where $\tilde{\theta}$ is the set of default values for the θ parameters. For the major part
 1235 of the nuisance parameters a log-normal prior distribution is used, with a standard deviation
 1236 corresponding to the given systematic uncertainty. For some nuisance parameters, as the
 1237 ones related to the statistical uncertainty coming from the background measurement in data
 1238 control regions, a Gamma distribution is instead recommended. A log-uniform distribution
 1239 is used for the uncertainty related to the normalization of background contributions that
 1240 are left unconstrained in the fit, such as the WW background process. Finally, some of
 1241 the experimental uncertainties, related to the shape of signal and background processes,
 1242 are modelled by means of additional histograms as explained in Sec. ???. The nuisance
 1243 parameters correlations across different p_T^H bins are taken into account. Moreover the
 1244 nuisance parameters can also be correlated (or anti-correlated) between signal and different
 1245 background processes. As an example, the uncertainty related to the integrated luminosity
 1246 measurement is fully correlated for all the signal and background processes.

1247 Before running the fit on the data, the same procedure has been applied to the so called
 1248 *Asimov data set*¹, which provides a simple method to estimate the signal sensitivity before
 1249 looking at the data [72].

¹In 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.

1250 4.6.2 Signal and background yields

1251 A comparison of data and background prediction is shown in Fig. ??, where the $m_{\ell\ell}$
 1252 distribution is shown for the six p_T^H bins. Distributions correspond to the m_T window of
 1253 $[60, 110]$ GeV, in order to emphasize the signal contribution [34]. The m_T distributions are
 1254 shown in Fig. ?? and correspond to the $m_{\ell\ell}$ window of $[12, 75]$ GeV.

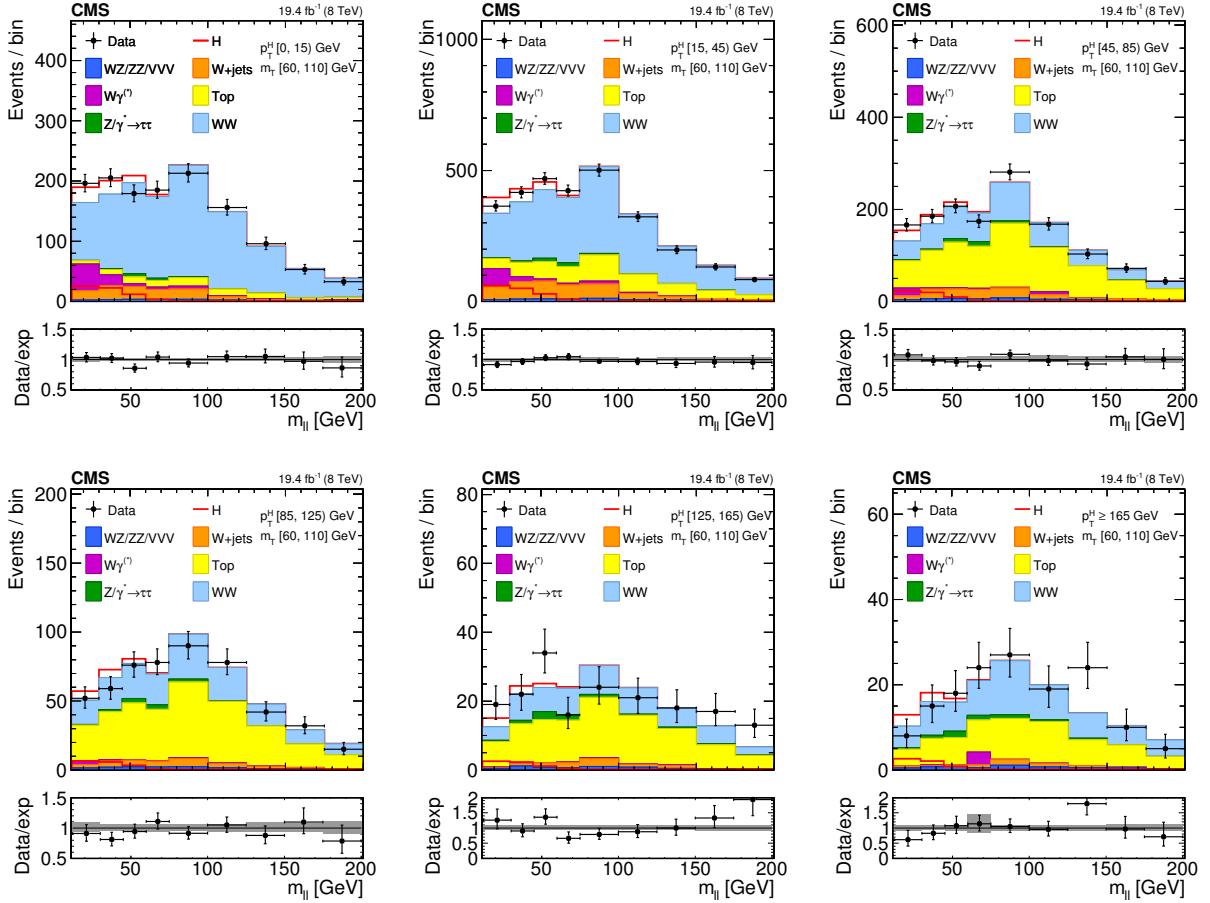


Figure 4.21: Distributions of the $m_{\ell\ell}$ variable in each of the six p_T^H bins. Background normalizations correspond to the values obtained from the fit. Signal normalization is fixed to the SM expectation. The distributions are shown in an m_T window of $[60, 110]$ GeV in order to emphasize 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.

1255 The signal and background yields after the analysis selection are reported in Table ??.

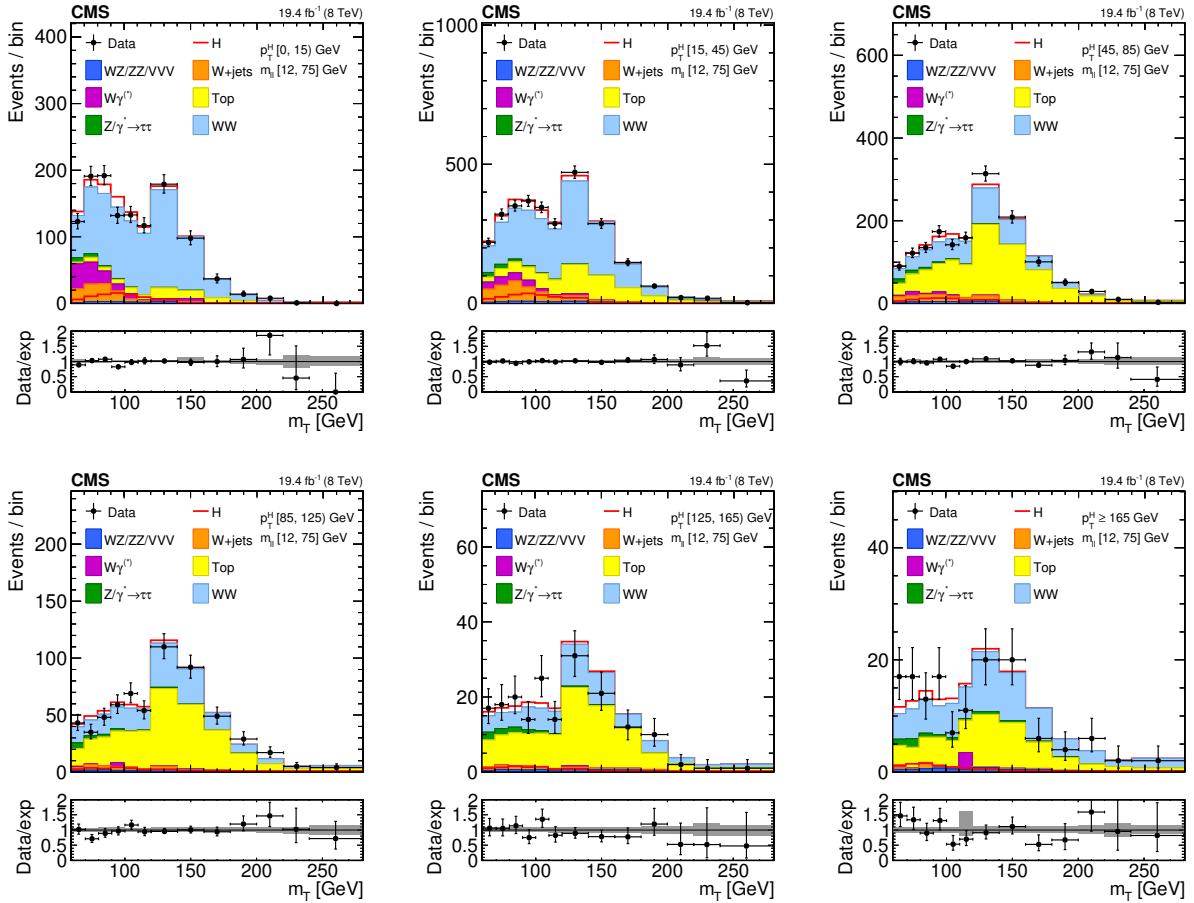


Figure 4.22: Distributions of the m_T variable in each of the six p_T^H bins. Background normalizations correspond to the values obtained from the fit. Signal normalization is fixed to the SM expectation. The distributions are shown in an $m_{\ell\ell}$ window of $[12, 75]$ GeV in order to emphasize 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 .

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
$Z/\gamma^* \rightarrow \tau^+\tau^-$	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

1256 The spectrum shown in Fig. ?? is obtained after having performed the fit and after the
 1257 subtraction of the out-of-fiducial signal events, but before undergoing the unfolding proce-
 1258 dure. The theoretical distribution after the detector simulation and event reconstruction is
 1259 also shown for comparison.

1260 In order to assess the robustness of the fit, several toy MC samples have been produced,
 1261 with a statistical accuracy corresponding to the one expected in data. The distribution
 1262 of the signal strengths extracted in each bin using the toy MC samples and their pull
 1263 distributions are shown in Fig. ??.

1264 4.7 Unfolding

1265 To facilitate comparisons with theoretical predictions or other experimental results, the
 1266 signal extracted performing the fit has to be corrected for detector resolution and efficiency
 1267 effects and for the efficiency of the selection defined in the analysis. An unfolding procedure
 1268 is used relying on the RooUnfold package [73], which provides the tools to run various
 1269 unfolding algorithms.

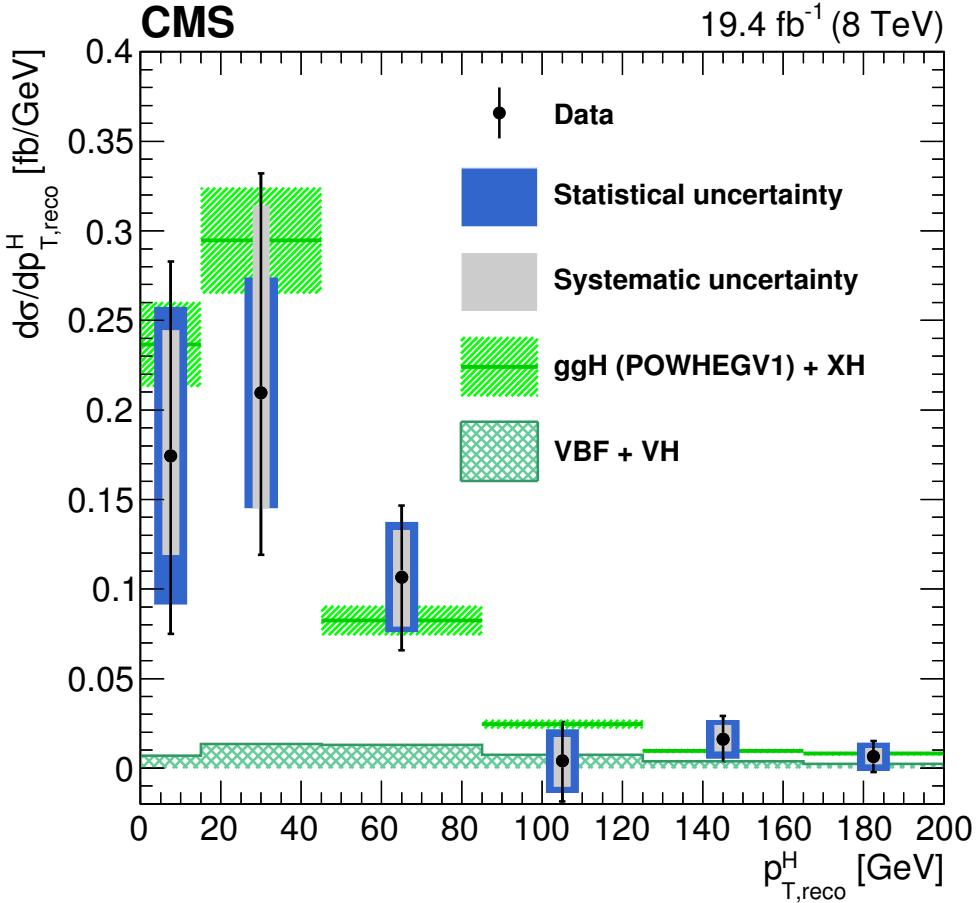


Figure 4.23: Differential Higgs boson production cross section as a function of the reconstructed p_{T}^H , before applying the unfolding procedure. Data values after the background subtraction are shown together with the statistical and the systematic uncertainties, determined propagating the sources 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.

1270 The basic principle behind the unfolding procedure in this analysis is to use MC signal
 1271 samples to make the “true” distribution of the variable of interest, which is obtained using
 1272 simulated events before particle interaction with the detector, and the same distribution
 1273 obtained using events reconstructed after the full GEANT4 simulation of the CMS detector
 1274 and event reconstruction. These two distributions are used to calculate the detector response
 1275 matrix M :

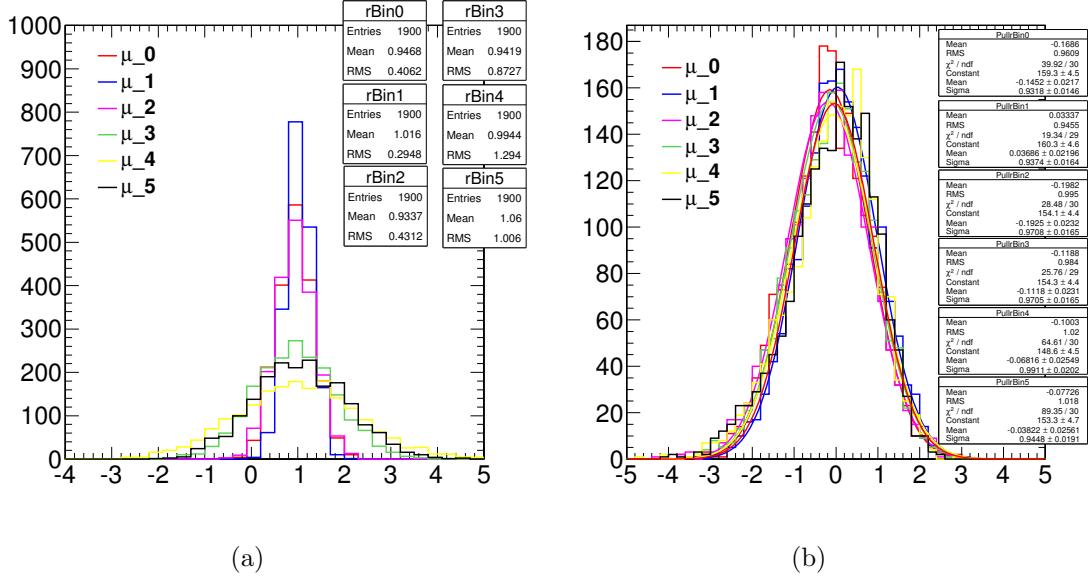


Figure 4.24: Signal strength distribution as extracted from the fit of toy MC samples (a). Distribution of the pull of the signal strength parameters (b).

$$R_i^{\text{MC}} = \sum_{j=1}^n M_{ij} T_j^{\text{MC}} \quad , \quad (4.30)$$

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 [35], the unfolding procedure in this analysis relies on the singular value decomposition [74] 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 [35]. 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 [75]. One possible solution is the utilization of regularization methods. Such methods introduce

1293 a regularization function that controls the smoothness of the distribution and depends
 1294 generally on one regularization parameter, which can be controlled to achieve the desired
 1295 degree of smoothness. The choice of the regularization parameter is particularly critical,
 1296 and it should represent an optimal trade-off between taming the fluctuations in the unfolded
 1297 result, and biasing the unfolded distribution towards the one used to build the response
 1298 matrix. The main feature of this method is the use of the singular value decomposition of
 1299 the response matrix, including an additional term to suppress the oscillatory component
 1300 of the solution, i.e. the regularization term, which represents some *a priori* knowledge
 1301 of the final solution. The regularization parameter is chosen to obtain results that are
 1302 robust against numerical instabilities and statistical fluctuations, following the prescription
 1303 described in Ref. [74]. **Maybe I should add an appendix describing the SVD method in**
 1304 **details**

1305 The response matrix is built as a two-dimensional histogram, with the generator-level
 1306 p_T^H on the y axis and the same variable after the reconstruction on the x axis, using the
 1307 same binning for both distributions. The resulting detector response matrix, including all
 1308 signal sources and normalized by row, is shown in Fig. ??(a). The value of the diagonal
 1309 bins corresponds to the stability S . The same matrix, normalized by column, is shown
 1310 in Fig. ??(b). In this case the diagonal bins correspond to the purity P . The S and P
 1311 parameters, defined in Sec. ??, provide an estimate of the p_T^H resolution and migration
 1312 effects. The main source of bin migrations effects in the response matrix is the limited
 1313 resolution in the measurement of E_T^{miss} .

1314 The resulting detector response matrix, which includes the effects of all signal sources and
 1315 is represented by normalizing each row to unity is shown in Fig. ??(a). This representation
 1316 shows the stability S in the diagonal bins, where S is defined as the ratio of the number of
 1317 events generated and reconstructed in a given bin, and the number of events generated in
 1318 that bin. In addition, a deconvolution matrix is constructed by normalizing each column
 1319 to unity and is shown in Fig. ??(b). This latter representation shows the purity P in
 1320 the diagonal bins, where P is defined as the ratio of the number of events generated and
 1321 reconstructed in a given bin, and the number of events reconstructed in that bin. The S
 1322 and P parameters provide an estimate of the p_T^H resolution and of migration effects. The
 1323 response matrix built including all signal sources is shown in Fig. ???. In order to point out
 1324 either the purity or the stability in diagonal bins, each column or row of the matrix was
 1325 respectively normalized to unity. The matrix obtained in the first case is what is actually
 1326 called detector response matrix, while in the other case the matrix is usually referred to as
 1327 detector deconvolution matrix.

1328 Several closure tests are performed in order to validate the unfolding procedure. To
 1329 estimate the uncertainty in the unfolding procedure due to the particular model adopted for
 1330 building the response matrix, two independent gluon fusion samples are used, corresponding
 1331 to two different generators: POWHEG V1 and JHUGEN generators, both interfaced to
 1332 PYTHIA 6.4. The JHUGEN generator sample is used to build the response matrix while the
 1333 POWHEG V1 sample is used for the measured and the MC distributions at generator level.

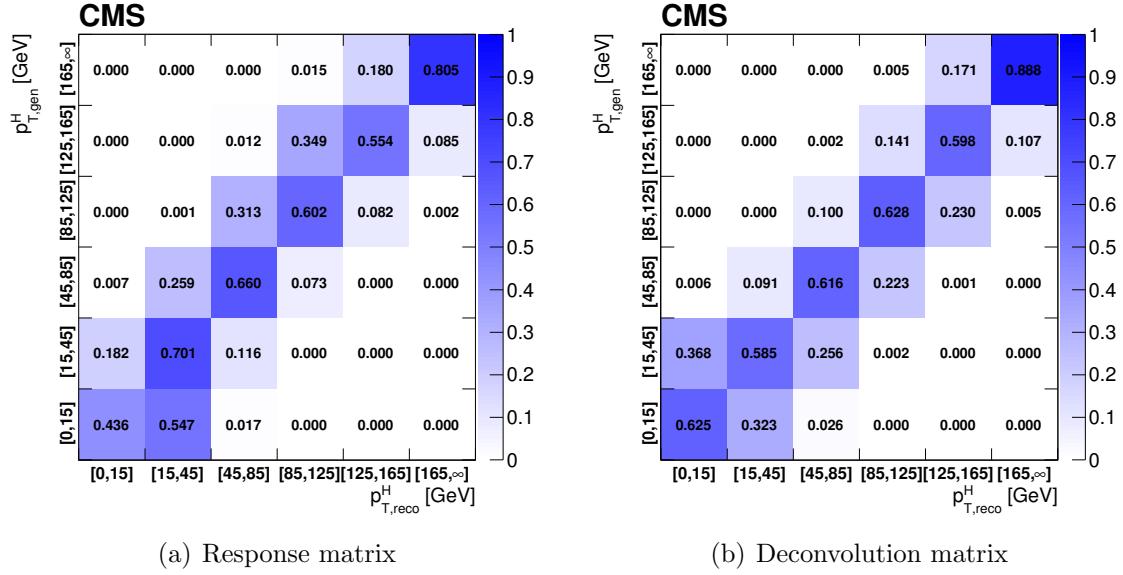


Figure 4.25: Response matrix (a) and deconvolution matrix (b) including all signal processes. The matrices are normalized either by row (a) or by column (b) in order to show the purity or stability respectively in diagonal bins.

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 ?? for different values of the regularization parameter: starting from $k_{\text{reg}} = 2$ (stronger regularization) up to $k_{\text{reg}} = 5$ (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. [74], strengthening the choice of $k_{\text{reg}} = 3$.

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_T^H bin [GeV]	Coverage			
	$k_{\text{reg}} = 2$	$k_{\text{reg}} = 3$	$k_{\text{reg}} = 4$	$k_{\text{reg}} = 5$
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_T^H bins, including all type A uncertainties, is shown in Fig. ???. The correlation $\text{cor}(i,j)$ of bins i and j is defined as:

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

where $\text{cov}(i,j)$ 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.

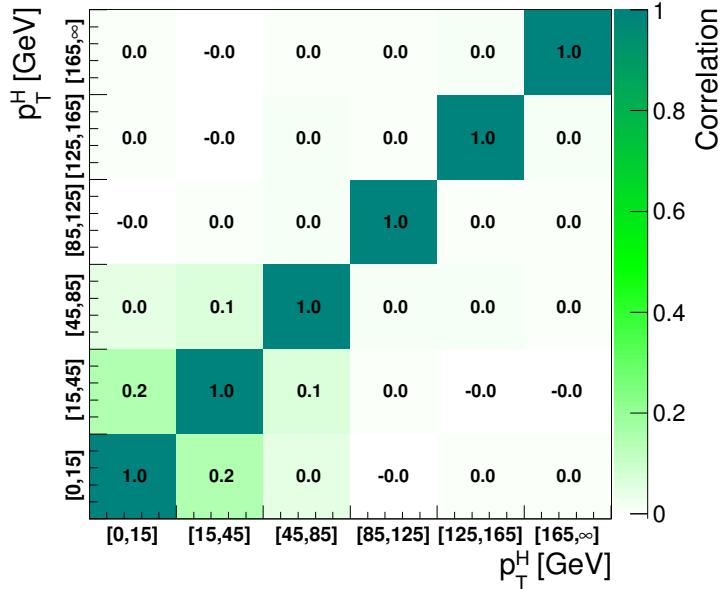


Figure 4.26: Correlations among the signal strengths corresponding to the six p_T^H bins including all type A uncertainties.

- 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. ???. 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_T^{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_T^{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

1388 Type C uncertainties are related to the underlying assumption on the Higgs boson
 1389 production mechanism used to extract the fiducial cross sections. These are evaluated
 1390 using alternative response matrices that are obtained by varying the relative fraction of
 1391 the VBF and ggH components within the experimental uncertainty, as given by the CMS
 1392 combined measurement [76]. Three different response matrices are built, corresponding to
 1393 the nominal, scaled up, and scaled down VBF/ggH ratio. The nominal matrix assumes
 1394 the SM VBF/ggH ratio, while up- and down-scaled matrices are constructed by varying
 1395 the SM signal strengths within the experimental constraints for VBF and ggH in such a
 1396 way as to obtain the maximal variation of the VBF/ggH ratio allowed by the experimental
 1397 constraints. These three matrices are used to unfold the reconstructed spectrum with the
 1398 nominal VBF/ggH fraction, and obtain an uncertainty on the unfolded spectrum.

1399 4.8 Results

1400 In order to unfold the spectrum, the procedure described in section ?? has been pursued. The
 1401 statistical plus type A systematic uncertainties are propagated by the unfolding procedure
 1402 into the final spectrum, taking into account the signal strengths covariance matrix. The
 1403 type B systematic uncertainty has been propagated using the following procedure: for each
 1404 p_T^H bin, we compute the upper bound of the systematic band computing the square sum
 1405 of all the signal strength variations that deviate in the up direction with respect to the
 1406 bin central value, whether or not this variation corresponds to the up or down shift of
 1407 the systematic uncertainty. The same is done for the lower bound of the systematic band.
 1408 If both the up and down shifts of a given nuisance parameter lead to a same direction
 1409 variation of the signal strength, only the larger variation is considered.

¹⁴¹⁰ The unfolded p_T^H spectrum is shown in Fig. ???. Statistical, systematic, and theoretical uncertainties are shown as separate error bands in the plot. The unfolded spectrum is

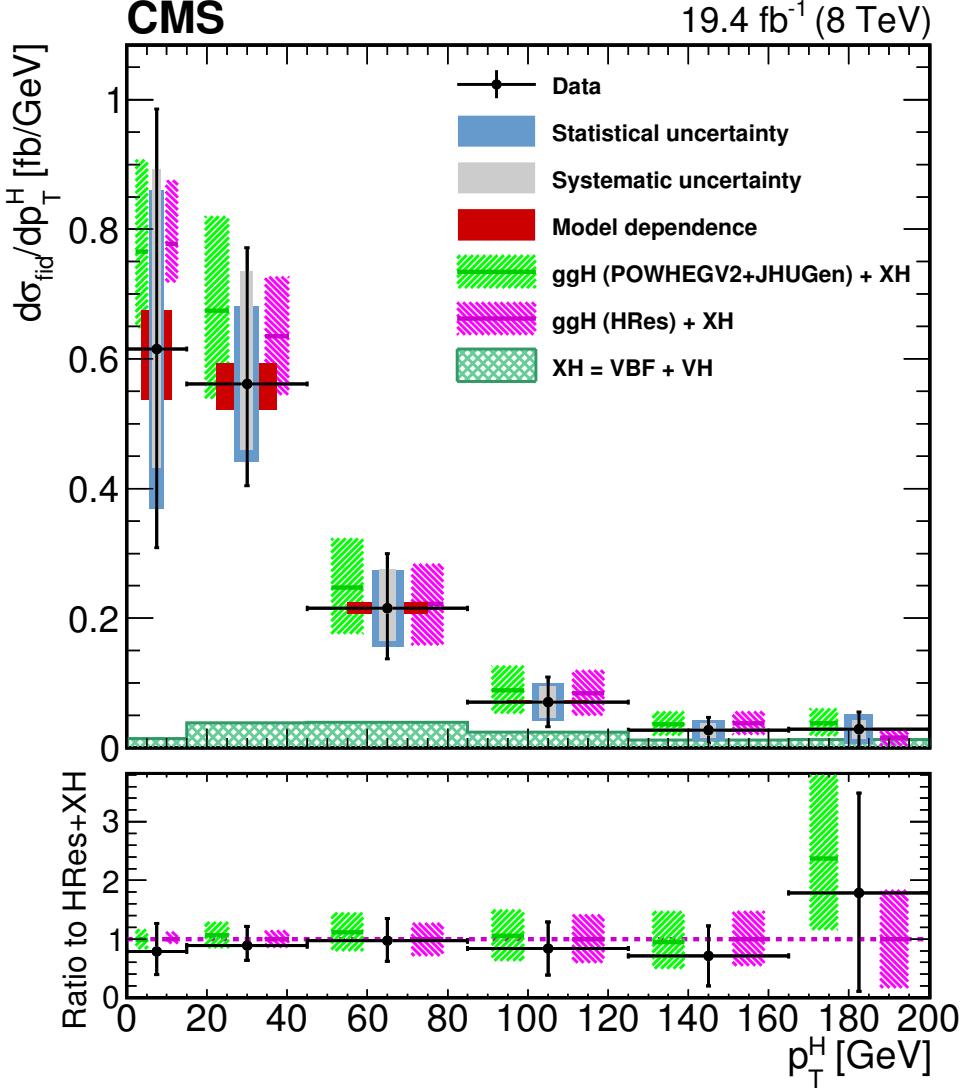


Figure 4.27: Higgs boson production cross section as a function of p_T^H , after applying the unfolding procedure. Data points are shown, together with statistical and systematic uncertainties. The vertical bars on the data points correspond to the sum in quadrature of the statistical and systematic uncertainties. The model dependence uncertainty is also shown. The pink (and back-sashed 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 $\text{XH}=\text{VBF}+\text{VH}$ 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.

¹⁴¹¹

¹⁴¹² 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

¹⁴¹³ using the HRES and POWHEG V2 programs. The comparison shows good agreement
¹⁴¹⁴ between data and theoretical predictions within the uncertainties. The measured values
¹⁴¹⁵ for the differential cross section in each bin of p_T^H are reported together with the total
¹⁴¹⁶ uncertainty in Table ??.

¹⁴¹⁷ Figure ?? shows the correlation matrix for the six bins of the differential spectrum. The
¹⁴¹⁸ correlation of bins is defined as in Eq. ??.

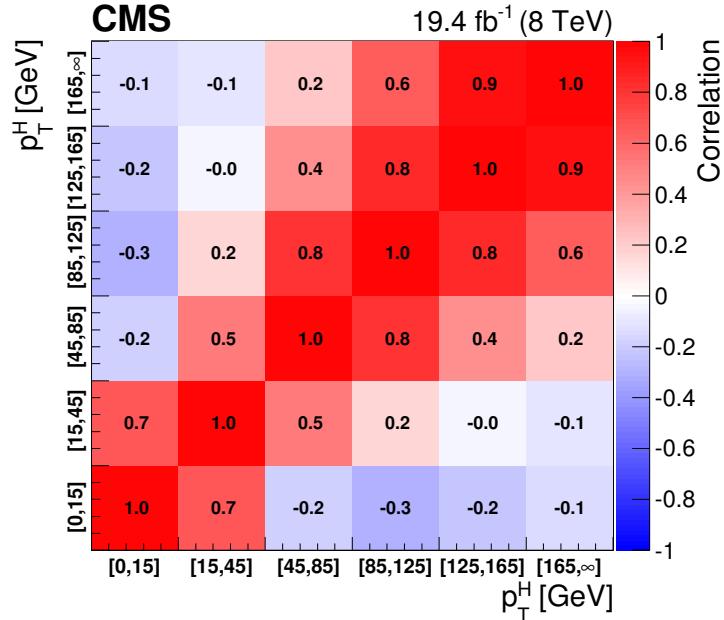


Figure 4.28: Correlation matrix among the p_T^H bins of the differential spectrum.

¹⁴¹⁹ To measure the inclusive cross section in the fiducial phase space, the differential
¹⁴²⁰ measured spectrum is integrated over p_T^H . In order to compute the contributions of the bin
¹⁴²¹ uncertainties of the differential spectrum to the inclusive uncertainty, error propagation
¹⁴²² is performed taking into account the covariance matrix of the six signal strengths. For
¹⁴²³ the extrapolation of this result to the fiducial phase space, the unfolding procedure is not
¹⁴²⁴ needed, and the inclusive measurement has only to be corrected for the fiducial phase
¹⁴²⁵ space selection efficiency ϵ_{fid} . Dividing the measured number of events by the integrated
¹⁴²⁶ luminosity and correcting for the overall selection efficiency, which is estimated in simulation
¹⁴²⁷ to be $\epsilon_{fid} = 36.2\%$, the inclusive fiducial $\sigma \times \mathcal{B}$, σ_{fid} , is computed to be:

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

¹⁴²⁸ in agreement within the uncertainties with the theoretical estimate of 48 ± 8 fb, computed
¹⁴²⁹ integrating the spectrum obtained with the POWHEG V2 program for the ggH process and
¹⁴³⁰ including the XH contribution.

Chapter 5

1431 Search for the SM Higgs boson in the 1432 $H \rightarrow WW$ channel with the first 1433 13 TeV LHC data

1434 5.1 Introduction

1435 In this chapter, the first search for the SM Higgs boson decaying to a W boson pair at
1436 13 TeV is presented, using a total integrated luminosity of 2.3 fb^{-1} , collected during the
1437 2015 proton proton data taking period of the LHC.

1438 Final states in which the two W bosons decay leptonically are studied. Therefore,
1439 events with a pair of oppositely-charged leptons, exactly one electron and one muon, a
1440 substantial amount of missing transverse energy, E_T^{miss} , due to the presence of neutrinos in
1441 the final state, and either zero or one jet are selected. This signature is common to other
1442 processes, which enter the analysis as backgrounds. The main background comes from
1443 WW production, irreducible background that shares the same final states and can only be
1444 separated by the use of certain kinematic properties. Another important background is
1445 W+jets, where a jet can mimick a leptonic signature. Background coming from top quark
1446 events, i.e. $t\bar{t}$ and single top production, is also important, followed by other processes such
1447 as Drell-Yan, WZ, and other EWK production. The analysis strategy follows the one used
1448 during Run 1 in the same channel, described in Chapter ??, with a few different aspects
1449 that are described in the next sections.

1450 With respect to 8 TeV, the ggH production cross section at 13 TeV is expected to
1451 increase of a factor of 2, thus raising the number of expected signal events. In addition, the
1452 cross section for the background processes is increasing as well. The WW production cross
1453 section increases of a factor of 1.8 and the $t\bar{t}$ cross section of a factor of 3.5, due to the
1454 enhancement of the gluon PDFs at higher center of mass energies.

1455 5.2 Data and simulated samples

1456 Data recorded in proton proton collisions at 13 TeV during 2015 was used in the analysis,
 1457 with a total integrated luminosity of 2.3 fb^{-1} . Single and double lepton triggers are used
 1458 similarly to the same analysis at 8 TeV. The HLT paths and descriptions of the triggers
 1459 used in this analysis are described in Tables ?? and ?? for electrons and muons respectively.

Table 5.1: HLT paths related to Electrons

HLT Path	Description
HLT_Ele23_WP Loose_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 $p_T = 30 \text{ GeV}$
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.

1460 The trigger efficiencies are measured in data and applied on simulated events as described
 1461 in Sec. ??.

1462 Concerning the simulated samples, several different Monte Carlo (MC) generators were
 1463 used. In the simulation, ‘lepton’ includes also τ . Higgs signal samples have been simulated
 1464 in all channels with POWHEG v2 [37, 50, 77], designed to describe the full NLO properties
 1465 of these processes. In particular, for Higgs produced via gluon fusion [39], and vector-boson-
 1466 fusion (VBF) [40], the decay of the Higgs boson into two W boson and subsequently into
 1467 leptons was done using JHUGEN v5.2.5 [78]. For associated production with a vector boson
 1468 (W^+H , W^-H , ZH) [79], including gluon fusion produced ZH ($ggZH$), the Higgs decay was
 1469 done via PYTHIA 8.1 [59]. Alternative signal samples were produced with AMC@NLO [42],
 1470 or with POWHEG v2 but decayed via PYTHIA 8.1 for gluon fusion and VBF assuming a
 1471 Higgs boson mass of 125 GeV. In the following, the mass of the SM Higgs boson is assumed
 1472 to be 125 GeV.

1473 The WW production, irreducible background for the analysis, was simulated in different
 1474 ways. POWHEG v2 [80] was used for $q\bar{q}$ produced WW in different decays. The cross
 1475 section used for normalizing WW processes produced via $q\bar{q}$ was computed at next-to-next-
 1476 to-leading order (NNLO) [81]. In order to control the top quark background processes, the
 1477 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%

1478 high- p_T jets enhances the importance of logarithms of the jet p_T , spoiling the convergence
 1479 of fixed-order calculations of the $q\bar{q} \rightarrow WW$ process and requiring the use of dedicated
 1480 resummation techniques for an accurate prediction of differential distributions [82, 83]. The
 1481 p_T of the jets produced in association with the WW system is strongly correlated with its
 1482 transverse momentum, p_T^{WW} , especially in the case where only one jet is produced. The
 1483 simulated $q\bar{q} \rightarrow WW$ events are reweighted to reproduce the p_T^{WW} distribution from the
 1484 p_T -resummed calculation.

1485 Gluon fusion produced WW was generated, with and without Higgs diagrams, using
 1486 MCFM v7.0 [84]. A $t\bar{t}$ sample dilepton sample was also generated using POWHEG v2. The
 1487 WW and $t\bar{t}$ samples produced specifically for this analysis are presented in Table ???. Other
 1488 background samples are used, a list of the most relevant ones is presented in Table ???.

1489 All processes are generated using the NNPDF2.3 [85, 86] parton distribution functions
 1490 (PDF) for NLO generators, while the LO version of the same PDF is used for LO generators.
 1491 All the event generators are interfaced to PYTHIA 8.1 [59] for the showering of partons
 1492 and hadronization, as well as including a simulation of the underlying event (UE) and
 1493 multiple interaction (MPI) based on the CUET8PM1 tune [87]. To estimate the systematic
 1494 uncertainties related to the choice of UE and MPI tune, the signal processes and the WW
 1495 events are also generated with two alternative tunes which are representative of the errors
 1496 on the tuning parameters. The showering and hadronization systematic uncertainty is

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

Process	$\sigma \times \mathcal{B}$ [pb]
$t\bar{t} \rightarrow WW b\bar{b} \rightarrow 2\ell 2\nu b\bar{b}$	87.31
$q\bar{q} \rightarrow WW \rightarrow 2\ell 2\nu$	12.178
$gg \rightarrow WW \rightarrow 2\ell 2\nu$	0.5905
$gg \rightarrow WW \rightarrow 2\ell 2\nu$ (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 ($10 \text{ GeV} < m_{\ell\ell} < 50 \text{ GeV}$)	20471.0
Drell-Yan ($m_{\ell\ell} > 50 \text{ GeV}$)	6025.26
$WZ \rightarrow 2\ell 2q$	5.5950
$ZZ \rightarrow 2\ell 2q$	3.2210
WWZ	0.1651
WZZ	0.05565
ZZZ	0.01398

1497 estimated by interfacing the same MC samples with the HERWIG ++ 2.7 parton shower [88,
1498 89]. For all processes, the detector response is simulated using a detailed description of the
1499 CMS detector, based on the GEANT4 package [51].

1500 The simulated samples are generated with distributions for the number of pileup
1501 interactions that are meant to roughly cover, though not exactly match, the conditions
1502 expected for the different data-taking periods. In order to factorize these effects, the number
1503 of true pileup interactions from the simulation truth is reweighted to match the data. The
1504 re-weighting is propagated automatically to both the in-time pile up and the out-of-time
1505 one. In Fig. ??, the effect of this reweighting on a sample enriched in Drell-Yan events is
1506 shown. Before the reweighting the simulation is presented in the open red histogram; after
1507 the reweighting, it is represented by the solid green histogram that matched well the data.
1508 In order to select this sample, events with two leptons with $p_T > 20 \text{ GeV}$, opposite sign,
1509 and same flavour, are selected only if $|m_{\ell\ell} - m_Z| < 15 \text{ GeV}$.

1510 The average number of pileup is approximately 11.5.

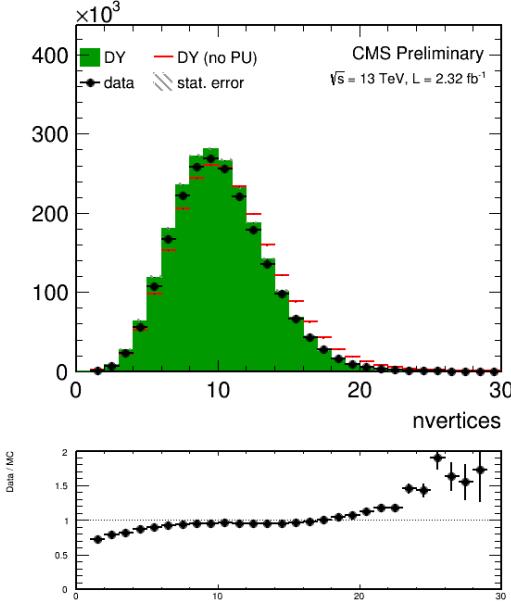


Figure 5.1: Distributions of the number of vertices in a Drell-Yan enriched sample in data, together with the simulation before (red) and after (solid green) the pileup reweighting.

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 [90], 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. [33].

The cross section used for normalizing $q\bar{q}$ produced WW processes was computed at next-to-next-to-leading order (NNLO) [81]. 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 [91]. 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 [92] at NLO. The $t\bar{t}$ cross section is also provided by the LHC Top Working group [93], and it is computed at NNLO, with NNLL soft gluon resummation.

Drell-Yan (DY) production of Z/γ^* is generated using AMC@NLO [42]. 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.

1529 All processes are generated using the NNPDF2.3 [85, 86] parton distribution functions
1530 (PDF) for NLO generators, while the LO version of the same PDF is used for LO generators.
1531 All the event generators are interfaced to PYTHIA 8.1 for the showering of partons and
1532 hadronization, as well as including a simulation of the underlying event (UE) and multiple
1533 interaction (MPI) based on the CUET8PM1 tune [87].

1534 5.3 Analysis strategy

1535 5.3.1 Event reconstruction

1536 Regarding the electrons, muons, jets and E_T^{miss} definition and reconstruction, the standard
1537 CMS recommendations described in Chapter 2 are used. The specific selections used in
1538 this analysis are briefly summarised below.

1539 Muons are identified according to the CMS recommendations for the medium working
1540 point, with the addition of some extra cuts, as defined by the following selections:

- 1541 • identified by the standard medium muon selection described in Sec. ??; Not yet defined
1542 :)
- 1543 • $p_T > 10 \text{ GeV}$;
- 1544 • $|\eta| < 2.4$;
- 1545 • $|d_{xy}| < 0.01 \text{ cm}$ for $p_T < 20 \text{ GeV}$ and $|d_{xy}| < 0.02 \text{ cm}$ for $p_T > 20 \text{ GeV}$, d_{xy} being the
1546 transverse impact parameter with respect to the primary vertex;
- 1547 • $|d_z| < 0.1 \text{ cm}$, where d_z is the longitudinal distance of the muon track in the tracker
1548 extrapolated along the beam direction.

1549 For the muon isolation, the CMS recommended particle flow isolation based on the
1550 tight working point is used, corresponding to a requirement on the isolation variable of
1551 $ISO_{\text{tight}} < 0.15$. In addition a tracker relative isolation is also applied.

1552 For the electron identification, the tight working point is used. In addition some
1553 additional cuts to make the selection “trigger-safe” are included. This is done because the
1554 electron triggers already include some identification and isolation requirements that are
1555 based on the raw detector information, while the offline selections make use of particle flow
1556 requirements. The “trigger-safe” selections are defined to make the the offline identification
1557 and isolation requirements tighter with respect to the online triggers.

1558 The simulated events are corrected for the lepton trigger, identification and isolation
1559 efficiencies measured in data using the same techniques described in Sec. ??.

1560 Jets are defined clustering the particle flow objects using the anti- k_t algorithm with a
 1561 distance parameter of 0.4. The CHS pileup mitigation technique is used. The L1, L2, L3
 1562 and L2L3 jet energy correction described in Sec. ?? are applied. The reject jets coming from
 1563 calorimeter or readout electronics noise, the loose working point for PF jet identification is
 1564 used.

1565 The b-tagging algorithm for this analysis is chosen comparing the performances of
 1566 different algorithms using simulations for signal and background contributions in the phase
 1567 space defined by the analysis kinematic requirements. More precisely, two MC samples are
 1568 used, one corresponding to the H \rightarrow WW \rightarrow 2 ℓ 2 ν signal produced via the ggH production
 1569 mode and another corresponding to the t \bar{t} process. In fact, the first sample is enriched
 1570 in light jets, i.e. originating by the hadronization of light quarks like u,d,c and s quarks,
 1571 while the second sample is enriched in b jets, coming from the top quark decay. The b-veto
 1572 efficiency, ϵ_{bveto} , is computed separately for the two samples and for the various b tagging
 1573 algorithms. To compare the b tagging performance ϵ_{bveto} is computed for different working
 1574 points, i.e. different selections on the specific b tagging discriminator, and the results are
 1575 reported in the form of a ROC curve. The ROC curves corresponding to events with 0, 1
 1576 and ≥ 2 jets are shown in Fig. ???. Events considered for this study are the ones passing
 1577 the WW baseline selection.

1578 The ROC curves show that the cMVAv2 algorithm has the best performance for the
 1579 analysis phase space among the algorithms taken into account. For both the CSVv2 and
 1580 cMVAv2 algorithms, three working points are defined corresponding to the mistag rates¹ of
 1581 10% for the loose, 1% for the medium and 0.1% for the tight working point. The distribution
 1582 of the cMVAv2 discriminator associated to the leading jet both for the ggH and the t \bar{t} MC
 1583 sample is shown in figure ??.

1584 In order to determine the best working point for this analysis a preliminary significance
 1585 assessment is performed, using a complete analysis procedure in which only statistical
 1586 effects are taken into account (no systematics are included). The significance assessment
 1587 was performed using a two dimensional discriminating variable consisting of the dilepton
 1588 invariant mass versus the transverse mass. The assessment was performed with the following
 1589 leptonic selection:

- 1590 • two leptons, an electron and a muon with opposite charge, with leading lepton p_T
 1591 greater than 20 GeV and sub-leading lepton p_T greater than 13 GeV;
- 1592 • no other lepton (electron or muon) with p_T greater than 10 GeV;
- 1593 • $m_{\ell\ell}$ greater than 12 GeV;
- 1594 • PF type 1 corrected MET greater than 20 GeV;
- 1595 • $p_T^{\ell\ell}$ greater than 30 GeV.

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

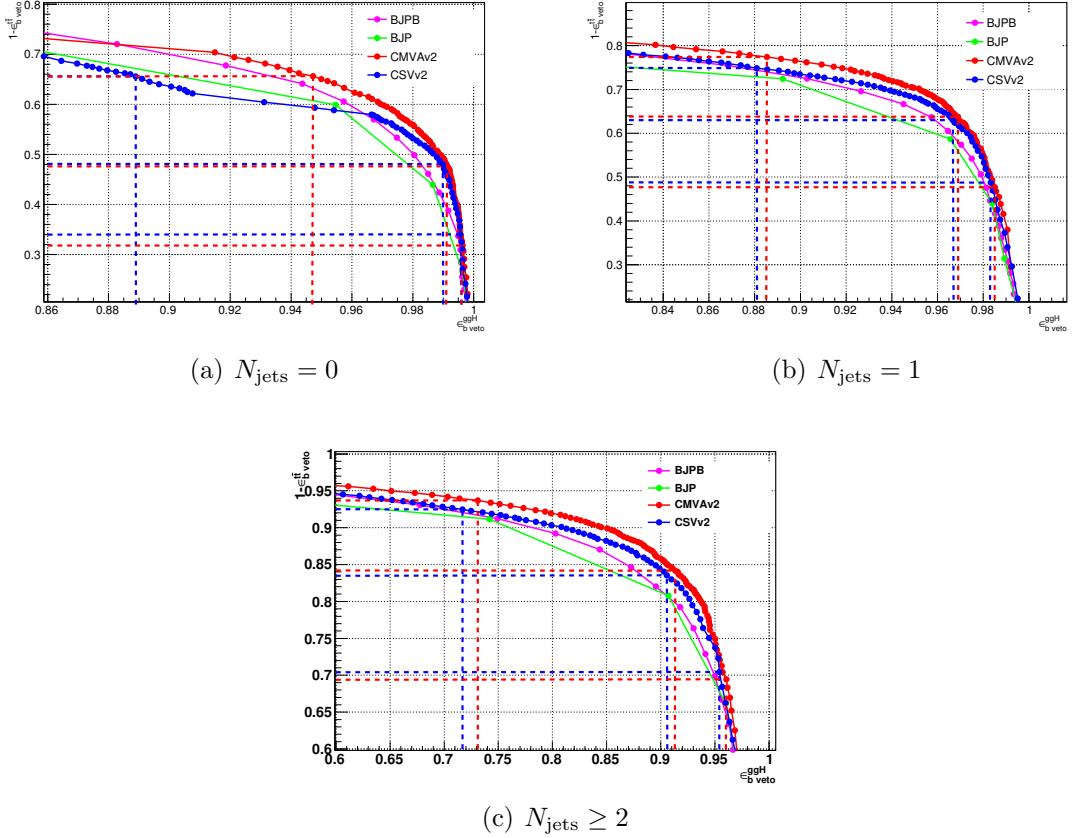


Figure 5.2: ROC curve for the b veto efficiency on signal and background events. The blue and red lines point out the signal efficiency and the background rejection corresponding to the three working points considered for the CSVv2 and the cMVAv2 algorithms respectively.

1596 In addition to this global selection, two categories were identified:

- 1597 • 0 jets: no jets above 30 GeV, jets between 20 GeV and 30 GeV are b-vetoed with the
1598 cMVAv2 WP under study;
- 1599 • 1 jet: exactly 1 jet above 30 GeV, no b-tagged jets above 30 GeV according to the
1600 cMVAv2 WP under study.

1601 The two categories were eventually combined together and the significance assessment was
1602 repeated for the three working points. With these selection we find the significance values
1603 listed in Table ?? for the three working points.

1604 The working point providing the best significance in the combined 0 + 1 jets category is
1605 found to be the loose one.

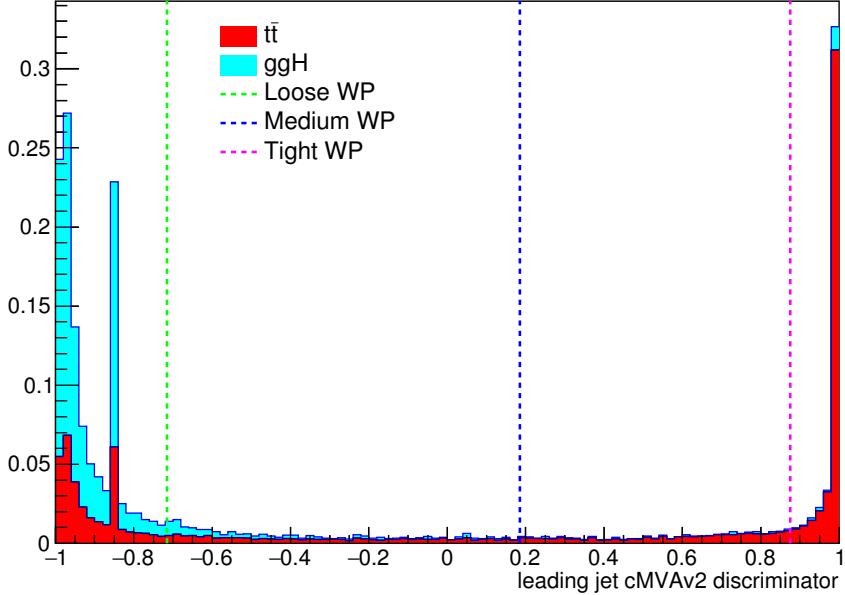


Figure 5.3: cMVAv2 discriminator associated to the leading jet (with $p_T > 30$ GeV) both for the ggH and the $t\bar{t}$ processes. The two processes are normalized to unity and stacked. The vertical dashed lines show the discriminator value corresponding to the three working points.

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

1606 To correct for a possible different b tagging efficiency in data and simulation, the
 1607 simulated events are reweighted using scale factors computed in bins of the jet η and p_T .
 1608 These scale factors and the corresponding uncertainties are centrally calculated for each
 1609 working point, in such a way to be employable by all the CMS analyses. The prescription
 1610 to reweight the simulated events is the following. First of all one has to compute the b
 1611 tagging efficiency using the MC samples, $\varepsilon_{MC}(p_T, \eta, f)$, for the chosen working point in bins
 1612 of jet p_T and η . The efficiency has to be computed for different flavours f of the jets, b,

1613 c and light (u,d,s), using the jet matching information² which is available in all the MC
1614 samples. An MC-based event weight is then calculated computing the probability P_{MC} of a
1615 given b tagging configuration to occur, e.g.:

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

1616 Afterwards, a similar probability is computed using data:

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

1617 where SF_i is the provided scale factor value for the relevant jet flavour, p_{T} and η . Products in
1618 Eqs. ?? and ?? run over all jets. The event weight is finally given by the ration $P_{\text{DATA}}/P_{\text{MC}}$.

1619 The b tagging efficiencies to be fed into Eq. ?? and Eq. ?? are derived using $t\bar{t}$ simulated
1620 events and applying basic leptonic selections. These efficiencies are shown in Fig. ?? for
1621 light (a), c-jets (b) and b-jets (c), in bins of η and p_{T} . The uncertainties associated to the
1622 efficiencies are representative of the statistics of the simulated $t\bar{t}$ sample, and are computed
1623 according to a binomial distribution.

1624 The effect of the event reweighting is to correct the shape of the b tagging discriminator
1625 in simulation, moving events from the b tag region (discriminator greater than > -0.715)
1626 to the b veto region (discriminator < -0.715) and viceversa. A data/simulation comparison
1627 of the b tagging discriminator for the leading and subleading jets is performed to check the
1628 agreement after the application of the event weights. In order to evaluate the data/simulation
1629 agreement for b-jets, the data and simulation are compared in a top enriched control region,
1630 defined by the following requirements:

- 1631 • two leptons, an electron and a muon with opposite charge, with leading lepton p_{T}
1632 greater than 20 GeV and sub-leading lepton p_{T} greater than 15 GeV;
- 1633 • no other lepton (electron or muon) with p_{T} greater than 10 GeV;
- 1634 • lepton invariant mass greater than 50 GeV;
- 1635 • at least two jets with p_{T} greater than 30 GeV;
- 1636 • at least one of the two leading jets with cMVAv2 btagging score greater than -0.715
1637 (i.e. the loose working point).

1638 In order to evaluate the agreement for light jets, a second control region is defined, populated
1639 by Z+light jet events, defined as follows:

²There 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.

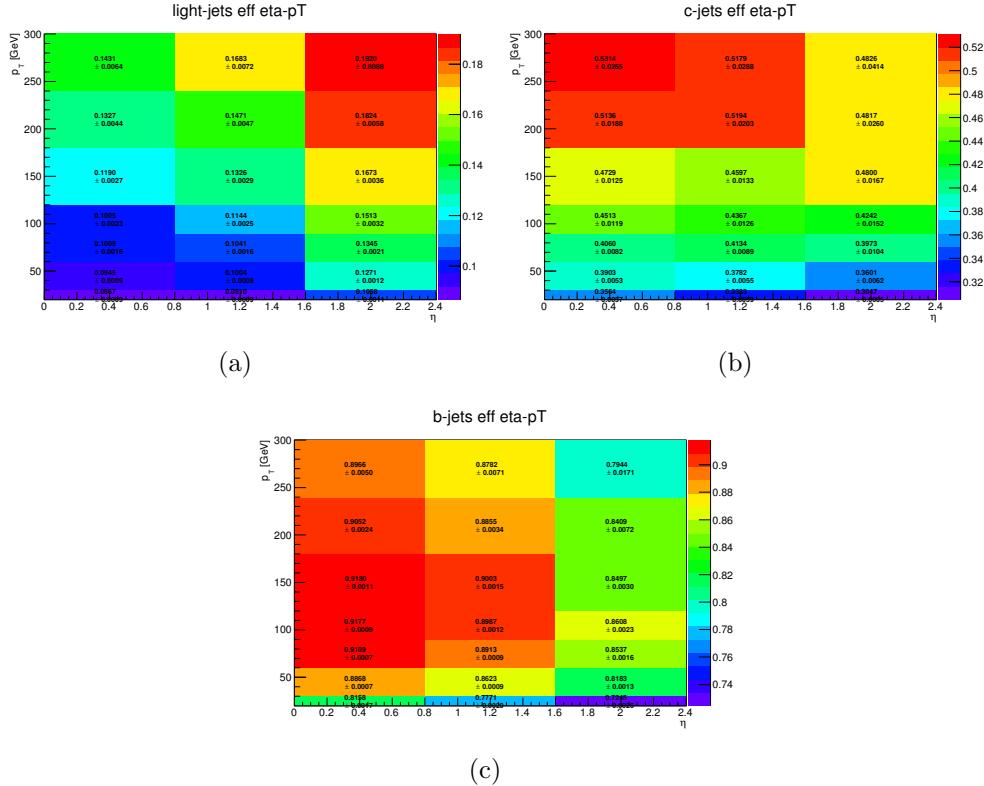


Figure 5.4: B tagging efficiencies for light jets (a), c-jets (b) and b jets (c), as a function of η and p_T .

- 1640 • two leptons, two electrons or two muons with opposite charge, with leading lepton p_T
1641 greater than 20 GeV and sub-leading lepton p_T greater than 15 GeV.
 - 1642 • no other lepton (electron or muon) with p_T greater than 10 GeV.
 - 1643 • lepton invariant mass greater between 80 GeV and 110 GeV.
 - 1644 • at least two jets with p_T greater than 30 GeV.
 - 1645 • at least one jet above 30 GeV.
 - 1646 • no jets above 20 GeV with a TCHE score above 2.1.
- 1647 Although a Z+jets sample is dominated by light flavor jets, a b-veto on an alternative
1648 algorithm (TCHE) is applied to reduce the contamination from b-jets, especially above the
1649 cMVAv2 cut. This helps mitigating possible data/simulation discrepancies in the modeling
1650 of the heavy/light flavour ratio. The comparison between data and simulation after the
1651 event reweighting is shown in Figs. ?? and ?? for the b-jets and light jets enriched control
1652 regions, respectively.

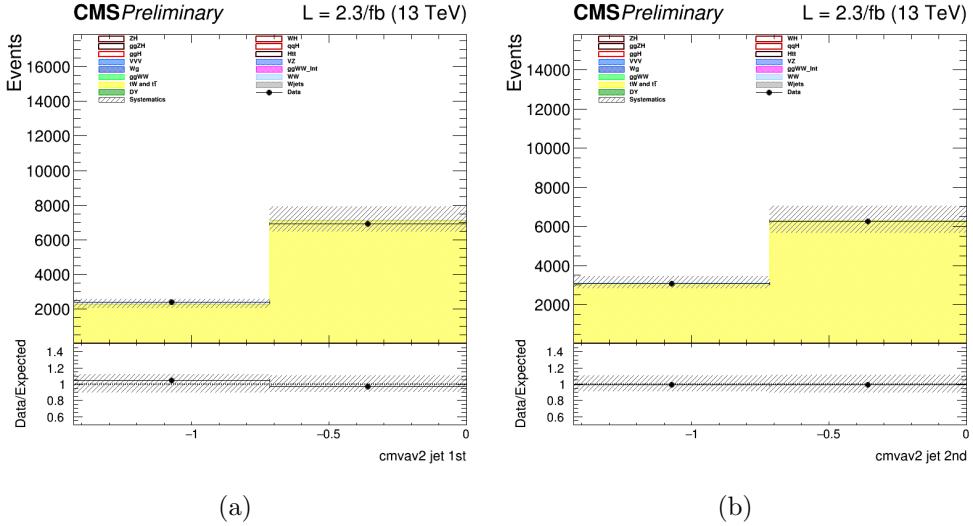


Figure 5.5: B tagging cMVAv2 discriminator for the leading (a) and the subleading (b) jet in the b-jets enriched control region.

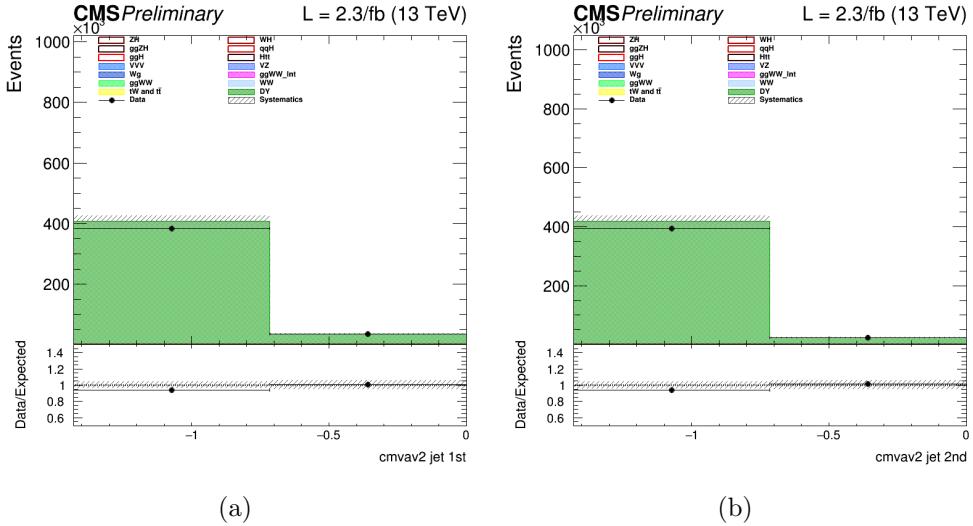


Figure 5.6: B tagging cMVAv2 discriminator for the leading (a) and the subleading (b) jet in the light jets enriched control region.

1653 5.3.2 Event selection and background rejection

1654 Since the ggH production mechanism, which is the main production mode for a Higgs
 1655 mass of around 125 GeV, is characterized by the emission of few jets arising from initial
 1656 or final state radiation, this analysis is limited to events with no jets or one jet. Due to
 1657 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, Z γ , W γ , or tri-boson production, no additional identified and isolated lepton with $p_T > 10$ 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_T^{miss} , a minimal E_T^{miss} of 20 GeV is required. The DY background is further reduced by requesting $p_T^{\ell\ell} > 30$ GeV. Finally the contribution from leptonic decays of single top and t \bar{t} production is reduced by requesting that no jets with $p_T > 20$ 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 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 $p_T > 30$ GeV. Events with zero associated jets mainly arise from the WW production, while WW and 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 $e\mu$ and μe final states together, are shown in Figs. ??, ?? and ?? after applying the WW baseline selections, with the addition of a cut on $m_{\ell\ell}$ to remove the Higgs signal contribution ($m_{\ell\ell} > 80$ GeV), and a cut on m_T to be orthogonal to the $Z\gamma^* \rightarrow \tau\tau$ background control region ($m_T > 60$ 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 $e\mu$ 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.

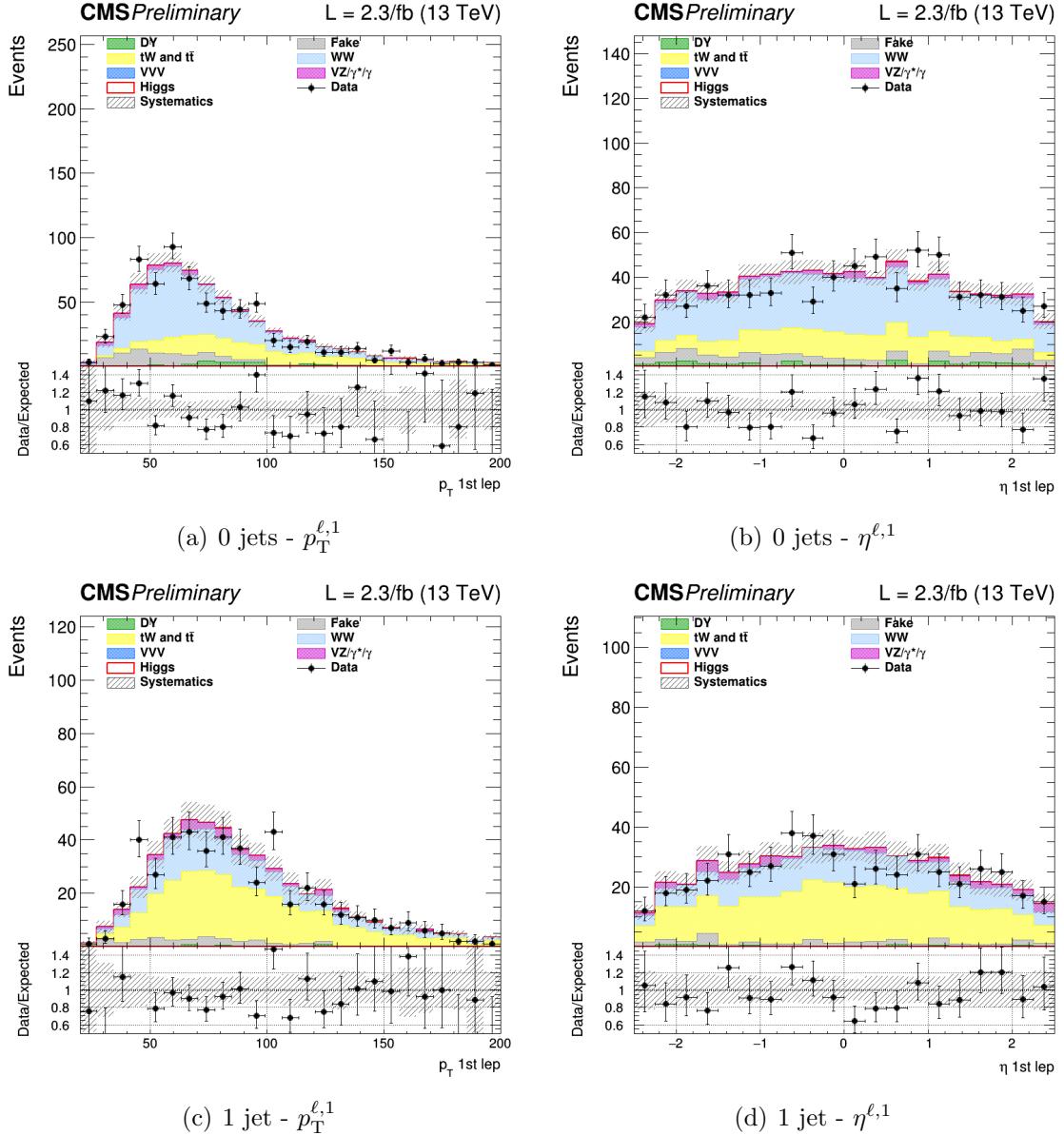


Figure 5.7: Distributions of p_T (left) and η (right) of the leading lepton for events with 0 jet (upper row) and 1 jet (lower row), for the main backgrounds (stacked histograms), and for a SM Higgs boson signal with $m_H = 125\text{ GeV}$ (superimposed and stacked red histogram) at the WW selection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

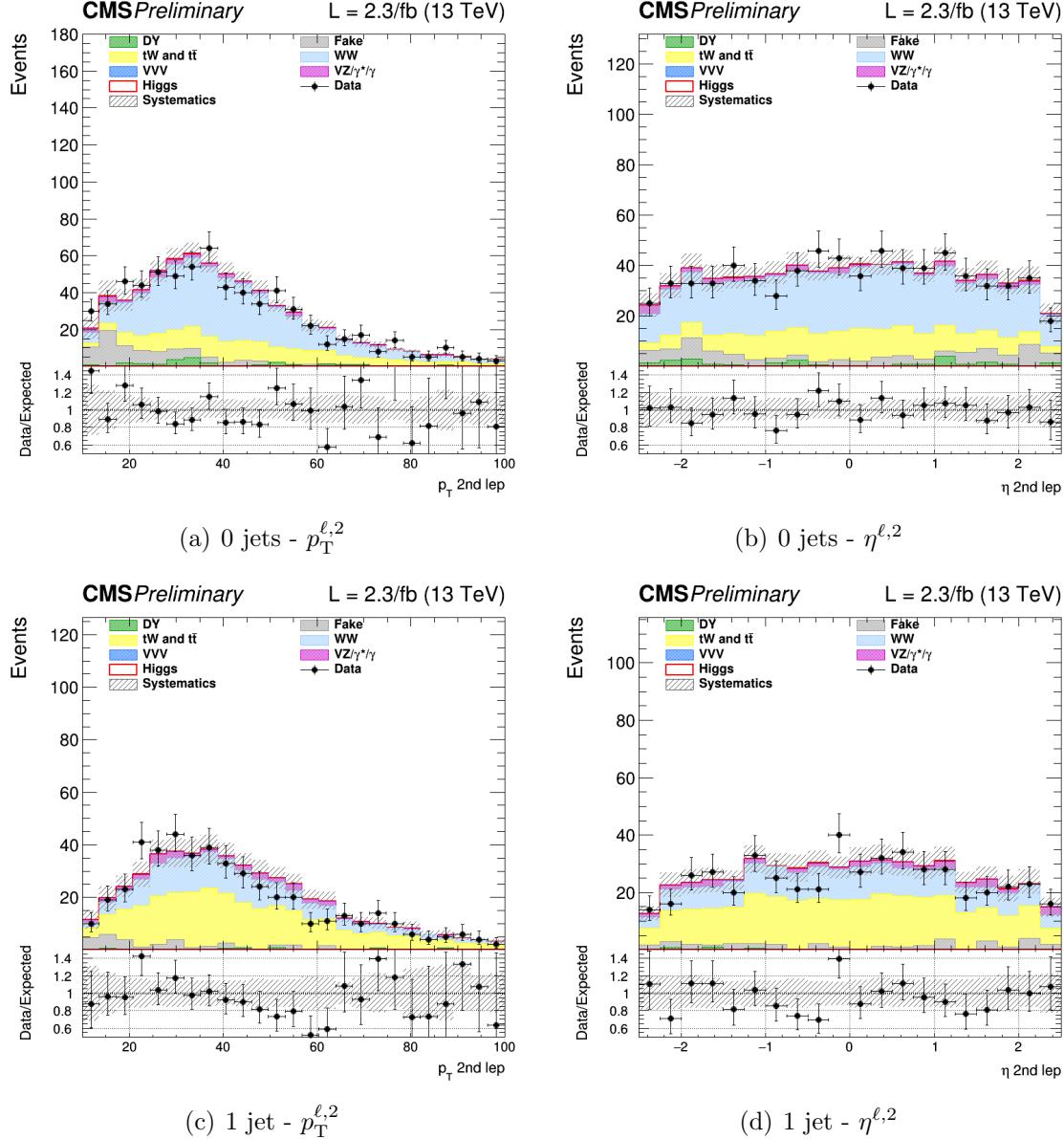


Figure 5.8: Distributions of p_T (left) and η (right) of the subleading lepton for events with 0 jets (upper row) and 1 jet (lower row), for the main backgrounds (stacked histograms), and for a SM Higgs boson signal with $m_H = 125 \text{ GeV}$ (superimposed and stacked red histogram) at the WW selection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

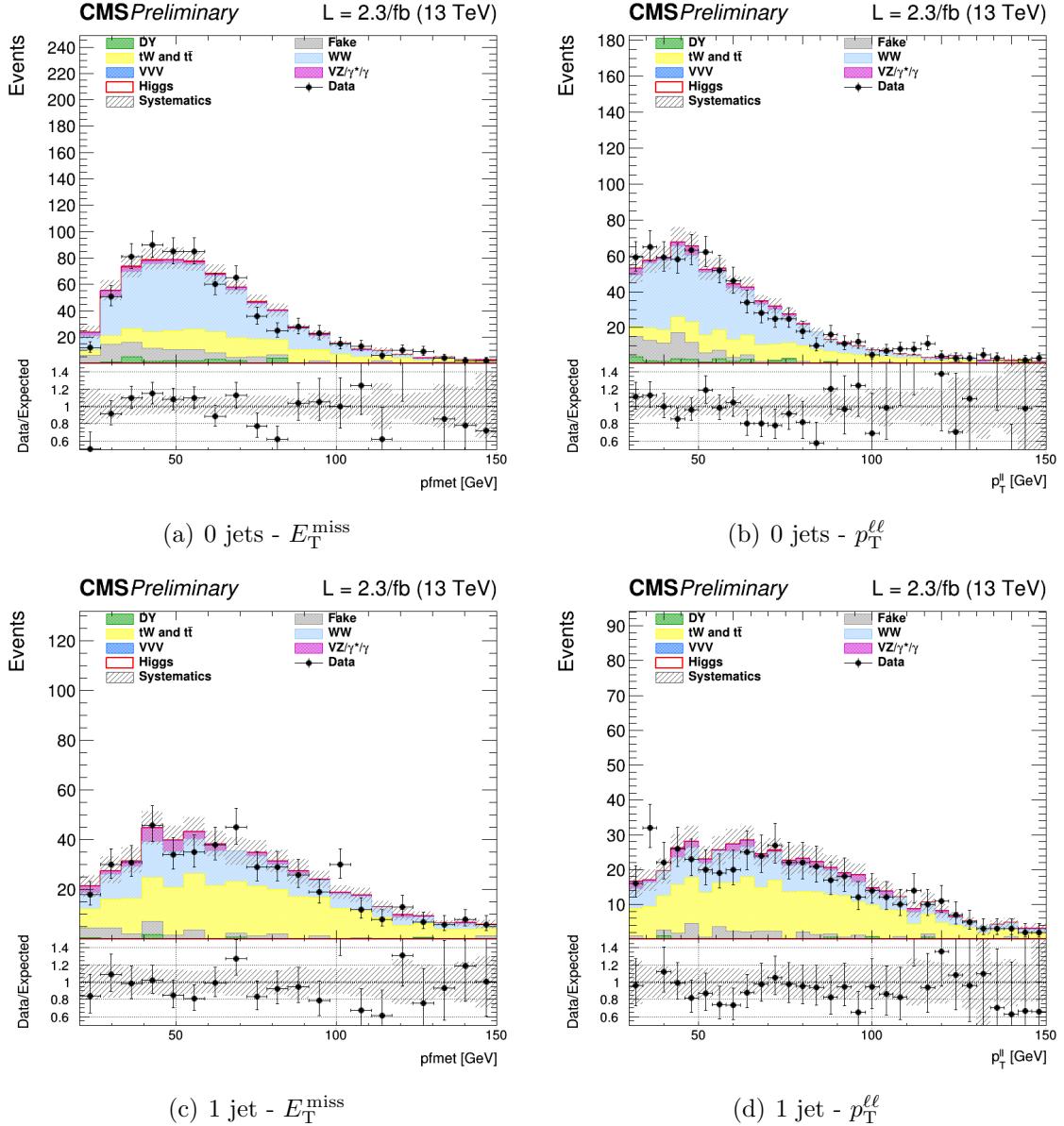


Figure 5.9: Distributions of E_T^{miss} (left) and $p_T^{\ell\ell}$ (right) for events with 0 jets (upper row) and 1 jet (lower row), for the main backgrounds (stacked histograms), and for a SM Higgs boson signal with $m_H = 125$ GeV (superimposed and stacked red histogram) at the WW selection level.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

¹⁶⁹⁵ **5.3.3 Signal extraction**

¹⁶⁹⁶ To extract the Higgs boson signal contribution in the four previously mentioned categories,
¹⁶⁹⁷ a similar approach to the one used in the Run 1 analysis [34] is pursued. The analysis is
¹⁶⁹⁸ based on two-dimensional templates of $m_{\ell\ell}$ versus m_T to discriminate signal and background
¹⁶⁹⁹ contributions. The $m_{\ell\ell}$ template is defined using 5 bins from $m_{\ell\ell} = 10 \text{ GeV}$ up to $m_{\ell\ell} =$
¹⁷⁰⁰ 110 GeV , while for the m_T template 7 bins are defined in the range $60 \text{ GeV} < m_T < 200 \text{ GeV}$.
¹⁷⁰¹ The phase space with $m_T < 60 \text{ GeV}$ is used as an orthogonal control region to extract the
¹⁷⁰² normalization of the DY background. A binned maximum likelihood fit to the signal and
¹⁷⁰³ background two-dimensional templates is performed to extract the signal strength in the
¹⁷⁰⁴ four categories.

¹⁷⁰⁵ Distributions of the $m_{\ell\ell}$ and m_T variables after the WW level selection are shown in
¹⁷⁰⁶ Fig. ?? for the 0 and 1 jet categories separately, but merging the $e\mu$ and μe final states
¹⁷⁰⁷ together.

¹⁷⁰⁸ The statistical methodology used to interpret the data and to combine the results from
¹⁷⁰⁹ the independent 0-jet and 1-jet categories in the $e\mu$ and μe final states has been developed
¹⁷¹⁰ by the ATLAS and CMS collaborations in the context of the LHC Higgs Combination
¹⁷¹¹ Group [76, 94]. The number of events in each category and in each bin of two-dimensional
¹⁷¹² template is modelled as a Poisson random variable, with a mean value given by the sum
¹⁷¹³ of the contributions from all the processes under consideration. Systematic uncertainties
¹⁷¹⁴ are represented by individual nuisance parameters with log-normal distributions. The
¹⁷¹⁵ uncertainties affect the overall normalization of the signal and backgrounds as well as the
¹⁷¹⁶ shape of the predictions across the distribution of the observables. Correlation between
¹⁷¹⁷ systematic uncertainties in different categories are taken into account.

¹⁷¹⁸ **5.4 Background estimation**

¹⁷¹⁹ The main background processes affecting the analysis signature, non-resonant WW pro-
¹⁷²⁰ duction and top quark processes, are estimated using data. Backgrounds arising from an
¹⁷²¹ experimental misidentification of the objects, such as W+jets (also called “Fake”), are
¹⁷²² estimated using data as well. The other minor backgrounds are generally estimated directly
¹⁷²³ from simulation as described in the following subsections.

¹⁷²⁴ **5.4.1 WW background**

¹⁷²⁵ The quark-induced WW background is simulated with NLO accuracy in perturbative
¹⁷²⁶ QCD, and the transverse momentum of the diboson system is reweighted to match the
¹⁷²⁷ NNLO+NNLL accuracy from theoretical calculations [82, 83]. However, given the large

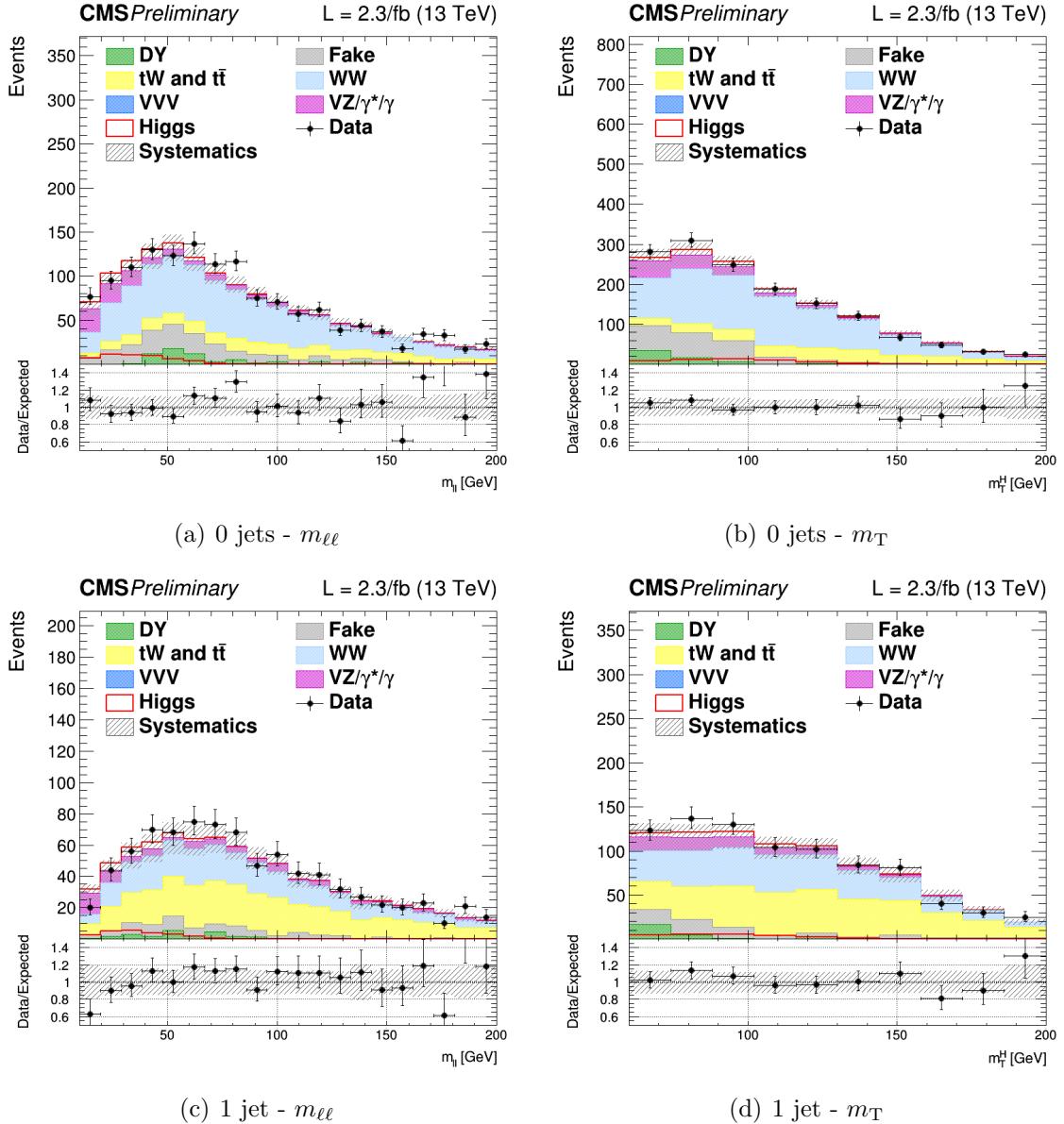


Figure 5.10: Distributions of $m_{\ell\ell}$ (left) and m_T (right) for events with 0 jets (upper row) and 1 jet (lower row), for the main backgrounds (stacked histograms), and for a SM Higgs boson signal with $m_H = 125$ GeV (superimposed and stacked red histogram) at the WW selection level. The last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

1728 uncertainties on the jet multiplicity distribution associated to this process, the normalization
 1729 of this background is measured from data separately for the 0 and 1 jet categories. The
 1730 normalization k-factors are extracted directly from the fit together with the signal strengths,
 1731 leaving the WW normalization free to float separately in the two jet multiplicity categories.
 1732 An orthogonal control region for the WW background normalization estimation is not
 1733 needed in this case, owing to the different $m_{\ell\ell}$ - m_T shape for signal and background.

1734 The gluon-induced WW production is sub-dominant with respect to the quark-induced
 1735 production, and its shape and normalization is fully taken from simulation, scaling the
 1736 cross section to the theoretical prediction with NLO accuracy [91].

1737 5.4.2 Top quark background

1738 As explained in Sec. ??, the production of top quark pairs represents one of the dominant
 1739 backgrounds in this analysis given its large cross section and a similar final state compared
 1740 to the signal. A b-jet veto, based on the *cMVAv2* b tagging algorithm, is used to suppress
 1741 this background and a reweighting procedure is applied on top of the simulated events to
 1742 correct for different b tagging efficiency in data and simulation.

1743 The top quark background normalization is measured using data, defining a b-jets
 1744 enriched control region by inverting the b-jet veto. More precisely, the b-jets enriched
 1745 control region for the 0-jet category is defined with the same WW baseline selection but
 1746 requiring at least one jet with $20 < p_T < 30 \text{ GeV}$ to be identified as a b jet and no other
 1747 jets with $p_T > 30 \text{ GeV}$. For the 1-jet category, the b-jets enriched region is defined requiring
 1748 exactly one jet with $p_T > 30 \text{ GeV}$ identified as a b-jet. To reduce other backgrounds in
 1749 these two regions, the dilepton mass has to be greater than 50 GeV . Distributions of the
 1750 $m_{\ell\ell}$ and m_T variables in the b-jets enriched control regions after applying the data driven
 1751 estimation are shown in Figure ??, for the 0 and 1 jet categories separately.

1752 The top quark background normalization is constrained during the fit procedure sepa-
 1753 rately in the two jet categories, by means of the control regions defined above, which are
 1754 treated in the fit as two additional categories.

1755 5.4.3 Jet-induced (or Fake) background

1756 One of the primary source belonging to this category arises from the misidentification of
 1757 leptons in W+jets processes in the 0 jet category. Also, semileptonic $t\bar{t}$ decays contribute
 1758 especially for higher jet multiplicities. Multijet production and hadronic $t\bar{t}$ decays are also
 1759 taken into account, but have a much smaller contribution.

1760 This background is fully estimated using data, with the technique described in Sec. ??.
 1761 To check the agreement of the background estimated in this way with data, a control sample

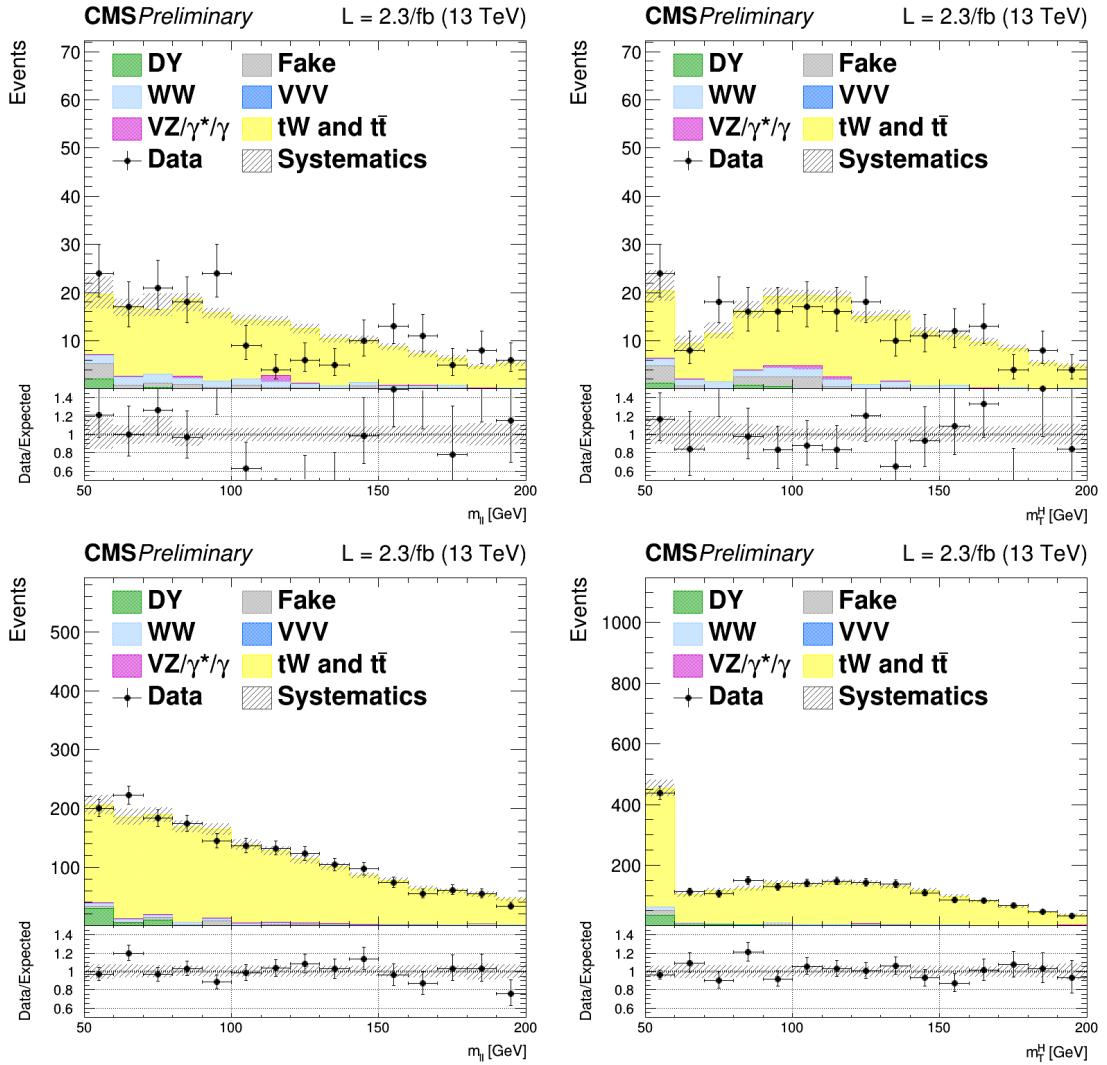


Figure 5.11: Distributions of $m_{\ell\ell}$ (left) and m_T (right) for events with 0 jet (top) and 1 jet (bottom) in top enriched phase space. Scale factors estimated from data are applied. The first (last) bin includes underflows (overflows).

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. ???. From the crosscheck in this control region, a global normalization factor of 0.8 is derived and applied to the jet-induced background.

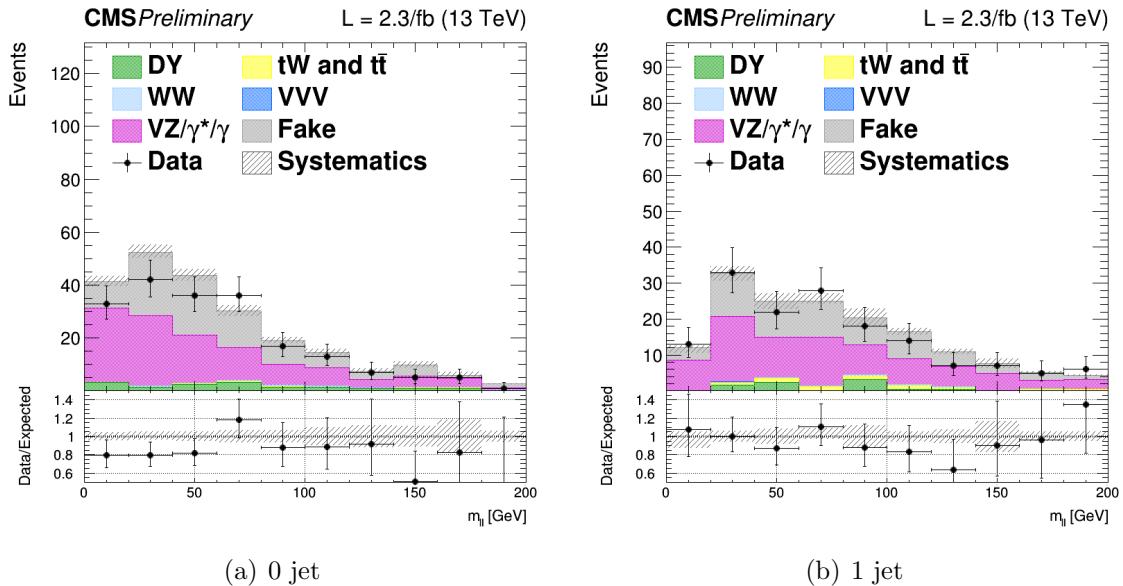


Figure 5.12: Control plots for $m_{\ell\ell}$ in a fakes enriched phase space for events with 0 and 1 jet with $p_T > 30$ GeV, in $e\mu$ final state. Fake contribution has been scaled by 0.8 to match data.

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 $30 \text{ GeV} < m_{\ell\ell} < 80 \text{ GeV}$. The $m_{\ell\ell}$ distributions in these control regions for the 0 and 1 jet categories are shown in Fig. ???.

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.

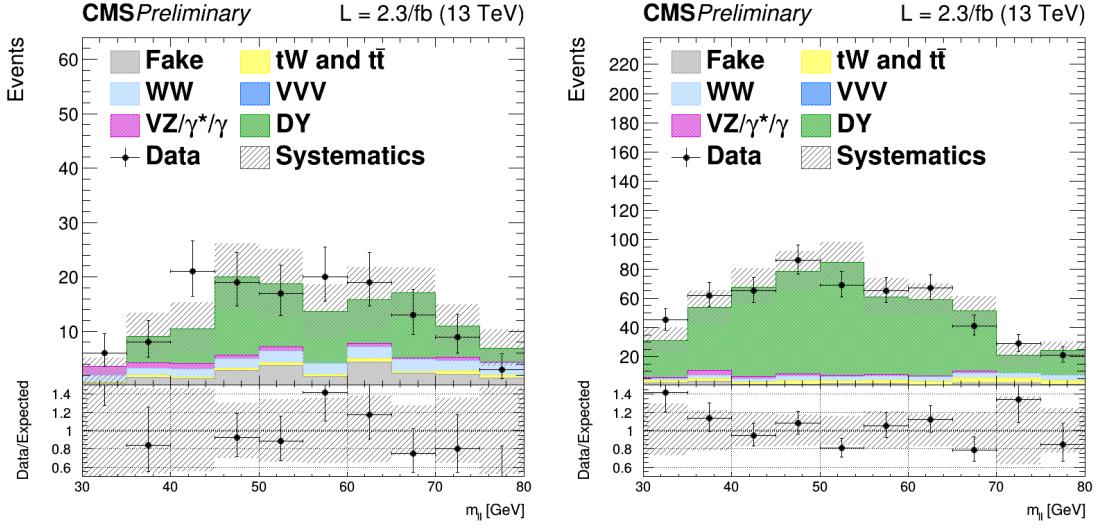


Figure 5.13: Distributions of $m_{\ell\ell}$ for events with 0 jet (left) and 1 jet (right) in the $DY \rightarrow \tau\tau$ enriched control region. Scale factors estimated from data are applied.

5.4.5 Other backgrounds

The $W\gamma^*$ 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 $W\gamma$ 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 $W\gamma^*$ 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 $E_T^{\text{miss}} < 25$ 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 .

1802 All remaining backgrounds from di-boson and tri-boson production, which are of mi-
1803 nor importance in the analysis phase space, are normalized according to their expected
1804 theoretical cross sections.

1805 5.5 Systematic uncertainties

1806 The systematic uncertainties affecting this measurement can be divided into three categories:
1807 the uncertainties on the background estimation, experimental uncertainties and theoretical
1808 uncertainties.

1809 The first category includes the uncertainties related to the background normalization
1810 and shape. For the non-resonant WW production the shape is taken from simulation.
1811 The input normalization to the fit is set to the expected value from simulation, and an
1812 unconstrained nuisance parameter with a flat distribution is associated to this number.
1813 This is done separately for the two jet categories.

1814 The top quark background shape is taken from simulation after correcting for the b
1815 tagging scale factors. An uncertainty due to these scale factors is included and affects
1816 both the normalization and the shape of the top quark background. The uncertainties on
1817 the normalization are treated similarly to the WW background case, but constraining the
1818 corresponding nuisances by means of the two control regions orthogonal to the signal phase
1819 space. A similar procedure is used for the DY background.

1820 Effects due to experimental uncertainties are studied by applying a scaling and smearing
1821 of certain variables related to the physics objects, e.g. the p_T of the leptons, followed by
1822 a subsequent recalculation of all the correlated variables. This is done for simulation, to
1823 account for possible systematic mismodeling.

1824 All experimental sources, except luminosity, are treated both as normalization and shape
1825 uncertainties, and are correlated among the signal and background processes and all the
1826 categories. The following experimental uncertainties are considered:

- 1827 • the uncertainty determined by the CMS online luminosity monitoring, 2.7% for the
1828 first data collected at $\sqrt{s} = 13 \text{ TeV}$;
- 1829 • the acceptance uncertainty associated with the combination of single and double lepton
1830 triggers, which is 2%;
- 1831 • the lepton reconstruction and identification efficiencies uncertainties, that are in the
1832 range 0.5-5% for electrons and 1-7% for muons depending on p_T and η ;
- 1833 • the muon momentum and electron energy scale and resolution uncertainties, that
1834 amount to 0.01-0.5% for electrons and 0.5-1.5% for muons depending on p_T and η ;

- 1835 • the jet energy scale uncertainties, that vary between 1-11% depending on the p_T and
1836 η of the jet;
- 1837 • the E_T^{miss} resolution uncertainty, that is taken into account by propagating the corre-
1838 sponding uncertainties on the leptons and jets;
- 1839 • the scale factors correcting the b tagging efficiency and mistagging rate, that are varied
1840 within their uncertainties. This systematic uncertainty is anticorrelated between the
1841 top control regions and the other ones.

1842 The uncertainties in the signal and background production rates due to theoretical
1843 uncertainties include several components, which are assumed to be independent: the PDFs
1844 and α_s , the underlying event and parton shower model, and the effect of missing higher-order
1845 corrections via variations of the renormalization and factorization scales.

1846 The effects of the variation of PDFs, α_s and renormalization/factorization QCD scales,
1847 mainly affect the signal processes, being the most important backgrounds estimated using
1848 data driven techniques. However, the uncertainties on minor backgrounds that are estimated
1849 from simulation are taken into account. These uncertainties are split in the uncertainties
1850 on the cross section, which are computed by the LHC cross section working group [95],
1851 and on the selection efficiency [96]. The PDFs and α_s signal cross section normalization
1852 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
1853 mechanism. The PDFs and α_s acceptance uncertainties are less than 1% for gluon- and
1854 quark-induced processes. The effect of the QCD scales variation on the selection efficiency is
1855 around 1-3% depending on the specific process. To estimate these uncertainties, the events
1856 are reweighted according to different QCD scales or different PDF sets and the selection
1857 efficiency is recomputed each time. For the QCD scale uncertainty the maximum variation
1858 with respect to the nominal value is taken as the uncertainty. For the case of PDF and α_s
1859 uncertainties, the distribution of the selection efficiency is built taking into account all the
1860 replicas in the NNPDF3.0 set and the uncertainty is estimated as the standard deviation of
1861 that distribution.

1862 In addition, the categorization of events based on jet multiplicity introduces additional un-
1863 certainties on the ggH production mode related to missing higher order corrections. These un-
1864 certainties are evaluated following the prescription described in Sec. [subsec:stewart-tackman
1865] and correspond to 5.6% for the 0-jet and 13% for the 1-jet bin categories.

1866 The underlying event uncertainty is estimated by comparing two different PYTHIA
1867 8 tunes, while parton shower modelling uncertainty is estimated by comparing samples
1868 interfaced with the PYTHIA 8 and HERWIG++ parton shower programs. The effect on the
1869 ggH (VBF) signal expected yield is about 5% (5%) for the PYTHIA 8 tune variation and
1870 about 7% (10%) for the parton shower description.

1871 Other specific theoretical uncertainties are associated to some backgrounds. An uncer-
1872 tainty on the ratio of the $t\bar{t}$ and tW cross sections is included. Indeed, these two processes

1873 are characterized by a different number of b-jets in the final state (2 b-jets for $t\bar{t}$ and 1
1874 for tW) and the b-veto acts differently for the two. A variation of the relative ratio of the
1875 cross sections can thus cause a migration of events from the 0 to the 1 jet categories and
1876 viceversa. The corresponding uncertainty is of 8%, according to the theoretical cross section
1877 calculations [92, 93].

1878 The $gg \rightarrow WW$ background LO cross section predicted by the MCFM generator is scaled
1879 to the NLO calculation, applying a k-factor of 1.4 with an uncertainty of 15% [91]. The
1880 interference term between the $gg \rightarrow WW$ and the ggH signal is also included and simulated
1881 with LO accuracy using MCFM. The k-factor to scale the interference term is 1.87, given by
1882 the geometrical average of the LO to NNLO $gg \rightarrow H \rightarrow WW$ scale factor (2.5) and the LO
1883 to NLO $gg \rightarrow WW$ scale factor (1.4). The uncertainty on this value is estimated as the
1884 maximum variation with respect to the two scale factors mentioned above, and is found
1885 to be of 25%. Anyway, with the current amount of integrated luminosity, the interference
1886 contribution is found to be negligible.

1887 For what the $qq \rightarrow WW$ background shape is concerned, an uncertainty related to
1888 the diboson p_T reweighting is evaluated varying the renormalization, factorization and
1889 resummation QCD scales.

1890 Finally, the uncertainties due to the limited statistical accuracy of the MC simulations
1891 are also taken into account, including an independent uncertainty for each bin of the
1892 two-dimensional distribution, and for each category. The uncertainty for a certain bin
1893 and process is given by the standard deviation of the Poisson distribution with mean
1894 corresponding to the number of MC events in that bin.

1895 5.6 Results

1896 The expected and observed signal significance are shown in Table ?? for all the categories
1897 separately. Also, the observed signal strengths and the corresponding uncertainties are
1898 shown. The best fit signal strength obtained combining all the categories together is found
1899 to be $0.3^{+0.5}_{-0.5}$, corresponding to an observed significance of 0.7σ , to be compared with the
1900 expected significance of 2.0σ for a Higgs boson mass of 125 GeV.

1901 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 $e\mu$, categories.

Category	Expected significance	Observed significance	σ/σ_{SM}
0-jet μe	1.1	1.3	$1.13^{+0.9}_{-0.9}$
0-jet $e\mu$	1.3	0.4	$0.33^{+0.7}_{-0.7}$
1-jet μe	0.8	0	$-0.11^{+0.5}_{-1.7}$
1-jet $e\mu$	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

1902 **Search for high mass resonances 1903 decaying to a W boson pair with first 1904 13 TeV LHC data**

1905 **6.1 Introduction**

1906 In this chapter, a search for a high mass spin-0 particle (from now on denoted as X)
1907 in the $X \rightarrow WW \rightarrow \ell\nu\ell'\nu'$ decay channel is presented, where ℓ and ℓ' refer to an different
1908 flavour lepton pair, i.e. $e\mu$. The search is based upon proton-proton collision data samples
1909 corresponding to an integrated luminosity of up to 2.3 fb^{-1} at $\sqrt{s} = 13 \text{ TeV}$, recorded by
1910 the CMS experiment at the LHC during 2015. This analysis represents a general extension
1911 of the SM Higgs boson search presented in ?? and is performed in a range of heavy scalar
1912 masses from $M_X = 200 \text{ GeV}$ up to 1 TeV , extending the range studied in a similar analysis
1913 performed using Run 1 LHC data [97], which provided upper limits on the production cross
1914 section of new scalar resonances up to 600 GeV .

1915 Despite the discovery of a particle consistent with the SM Higgs boson in 2012, there is a
1916 possibility that this particle is only a part of a larger Higgs sector, and hence only partially
1917 responsible of the EW symmetry breaking. This can be achieved in different theoretical
1918 models that extends the SM, such as the two-Higgs-doublet models [98–100], or models in
1919 which the SM Higgs boson mixes with a heavy EW singlet, which predict the existence
1920 of an additional resonance at high mass, with couplings similar to those of the SM Higgs
1921 boson, as most recently described in [101, 102].

1922 This analysis reports a generic search for a scalar particle with different resonance decay
1923 widths hypothesis, produced via the ggH and VBF production mechanisms. The results
1924 can then be interpreted in terms of different theoretical models. This analysis is heavily
1925 based on the SM Higgs search described in ??, in terms of physics objects, selections and
1926 background estimation. The differences and similarities are discussed in this chapter.

1927 6.2 Data and simulated samples

1928 The data sets, triggers, pile up reweighting, lepton identification and isolation used in this
1929 analysis are the same as the SM Higgs search and are described in Sec. ??.

1930 Also, the same MC simulations are used for the background processes, the only exception
1931 being the DY background, for which the MG5_AMC@NLO generator is used with LO
1932 QCD accuracy, matching together events with up to four jets in addition to the vector
1933 boson with the MLM [103] matching scheme. Given that this analysis aims to probe regions
1934 of phase space where the DY contribution is very small, like in the high transverse mass
1935 region, the usage of a simulation of the inclusive DY process may lead to large uncertainties
1936 due to the limited simulation statistics in the sample. To partially overcome this issue,
1937 different DY samples are generated in restricted portions of the phase space defined by
1938 the H_T variable, i.e. the scalar sum of all the partons p_T in the event. For $H_T < 100$ GeV
1939 the inclusive simulation is used, while different samples are used for higher values of H_T .
1940 The samples are merged using the parton level information, and it has been verified that a
1941 smooth transition between different H_T regions is achieved, as shown in Fig. ???. The DY
1942 LO cross section obtained from the simulation is scaled using the LO to NNLO k-factor of
1943 1.23.

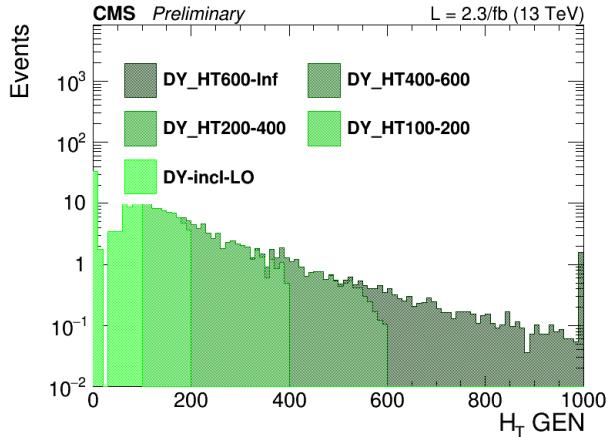


Figure 6.1: Generator level H_T distribution for the merged DY sample.

1944 In order to perform the resonance search in a large part of the mass spectrum, several
1945 signal samples for the gluon-gluon fusion and the vector boson fusion mechanisms have been
1946 generated corresponding to different Higgs boson masses in the range between 200 GeV
1947 and 1 TeV. The signal width for each mass point corresponds to the one expected for a
1948 SM Higgs boson at that mass. The samples are produced with a mass step of 50 GeV from
1949 250 to 800 GeV and of 100 GeV from 800 to 1000 GeV. A finer stepping is used between
1950 200 and 250 GeV. All the signal samples are generated with the POWHEG V2 generator,
1951 interfaced with the JHUGEN v6.2.8 generator, which handles the decay of the Higgs boson
1952 to $W^+W^- \rightarrow 2\ell 2\nu$.

The interference effects among $gg \rightarrow X \rightarrow WW$, $gg \rightarrow WW$ and $gg \rightarrow H \rightarrow WW$ are evaluated using the MCFM and JHUGEN generators, as implemented in the MELA (Matrix Element Likelihood Approach) framework [78]. Details about the interference effects are given in Sec. ??.

6.3 Analysis strategy

The analysis strategy for the first results on the high mass search in the $W^+W^- \rightarrow 2\ell 2\nu$ decay channel closely follows the strategy presented in the 13 TeV SM Higgs search in the $H \rightarrow W^+W^- \rightarrow 2\ell 2\nu$ 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 $\sigma_{VBF}/\sigma_{ggH}$ 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 $p_T > 10$ 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. ??.

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 $p_T > 30$ GeV, an invariant mass $m_{jj} > 500$ GeV and a gap in pseudorapidity $\Delta\eta_{jj} > 3.5$.

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

¹The 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_T^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 $\mathbf{E}_T^{\text{miss}} = (E_T^{\text{miss}}, \vec{p}_T^{\text{miss}})$, i.e.:

$$m_T^i = \sqrt{(p_{\ell\ell} + E_T^{\text{miss}})^2 - (\vec{p}_{\ell\ell} + \vec{p}_T^{\text{miss}})^2} . \quad (6.1)$$

This variable allows having a better sensitivity to different resonance mass hypothesis as shown in Fig. ??, where the shape of the m_T^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.

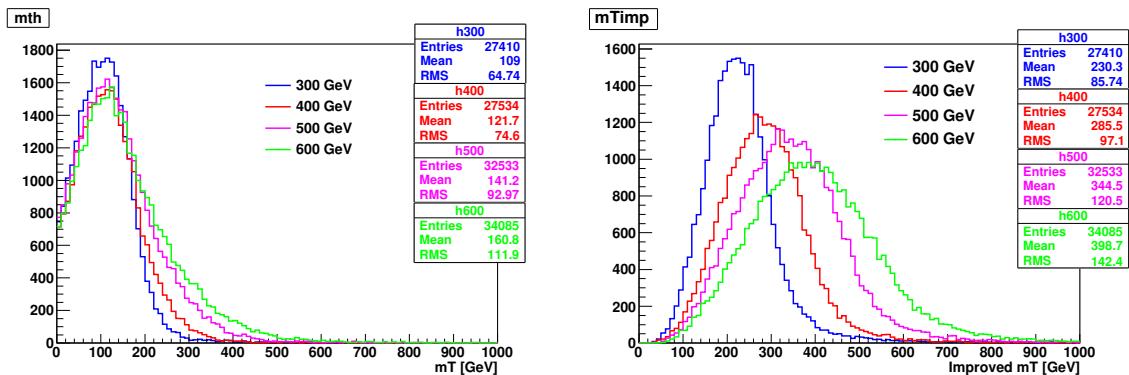


Figure 6.2: Distribution of the m_T and m_T^i variables at generator level for different resonance mass hypothesis.

The signal extraction is based on a binned maximum likelihood fit using the m_T^i distribution for signal and background contributions as templates. The m_T^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: $\Gamma' = \Gamma_{\text{SM}}$, $\Gamma' = 0.49 \times \Gamma_{\text{SM}}$, $\Gamma' = 0.25 \times \Gamma_{\text{SM}}$ and $\Gamma' = 0.09 \times \Gamma_{\text{SM}}$. The reweighting is performed at generator level by computing the ratio of two relativistic

2007 Breit Wigner distributions with different decay widths, $f(E, \Gamma', M_X)/f(E, \Gamma_{\text{SM}}, M_X)$, where:

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

2008 Here, $f(E, \Gamma_{\text{SM}}, M_X)$ represents the distribution used for the simulation of the signal at
2009 a mass M_X , and $f(E, \Gamma', M_X)$ the distribution with the new decay width. Each event
2010 is multiplied by this ratio (which depends on the energy E of the event) to obtain the
2011 reweighted distribution.

2012 When a resonance with a non negligible width is considered, it is important to take
2013 into account the interference effects both with the $\text{gg} \rightarrow \text{WW}$ background and the SM Higgs
2014 boson off-shell tail. A study of the interference effects for a resonance X produced through
2015 the gluon fusion mechanism is performed within the MCFM+JHUGEN framework, and
2016 including NNLO corrections for cross section using HNNLO program [104] based on
2017 MCFM. The matrix element package MELA supports all of these processes and allows fast
2018 MC re-weighting and optimal discriminant calculation. The basic idea of this approach is to
2019 compute the matrix elements of the processes under study with the MCFM and JHUGEN
2020 generators, including the interference terms, and using these matrix elements to compute
2021 an event weight used to reweight the simulated samples. Using this approach the simulated
2022 events can be reweighted according to different scenarios, for instance including some or all
2023 the interference terms, allowing a detailed study of the interference contribution. The effect
2024 of the various interference terms for the M_X variable at generator level is shown in Fig. ??,
2025 after having applied the WW baseline selections. As can be observed the contribution of
2026 the interference of the scalar resonance with the $\text{gg} \rightarrow \text{WW}$ background and with SM Higgs
2027 boson have opposite sign and partially cancel out. This cancellation effect is different for
2028 different resonance masses and depends on the event selection. In particular the interference
2029 term with the SM Higgs off-shell tail is positive for values below M_X while it turns negative
2030 above M_X . The contribution of the interference with the $\text{gg} \rightarrow \text{WW}$ background is instead
2031 characterized by an opposite sign lineshape, thus leading to a partial cancellation when
2032 considering the total interference.

2033 The effect of the resulting interference contribution including all the different terms is
2034 shown in Fig. ?? for the m_T^i signal templates, in the three categories separately and for
2035 different M_X hypotheses.

2036 The interference contribution is thus not negligible, especially for large values of M_X ,
2037 and is included in the analysis as part of the signal contribution. More specifically, during
2038 the fit procedure the signal yield is scaled by the signal strength parameter μ (which is the
2039 parameter of interest of the fit), while the interference yield is scaled by $\sqrt{\mu}$.

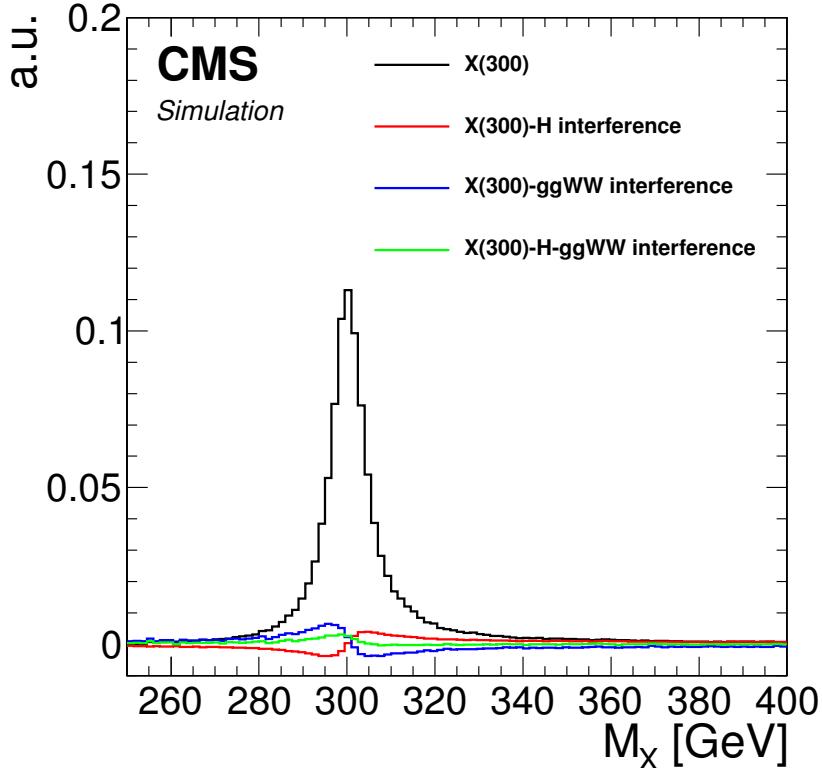


Figure 6.3: Distribution of the M_X variable for a resonance mass of 300 GeV, showing the various interference terms after the WW baseline selections.

2040 6.4 Background estimation

2041 The background processes affecting the analysis phase space are the same as the ones
 2042 contributing to the SM Higgs search described in Sec. ???. The techniques used for the
 2043 background estimation are the same as well.

2044 The most relevant difference is the addition of the 2 jets category. The WW, top
 2045 quark and DY background normalizations are estimated in this category using data driven
 2046 techniques, similarly to the other jet bins.

2047 Given the slightly different WW baseline selection with respect to the SM Higgs search,
 2048 also the control regions for the top quark and DY backgrounds estimation change, while the
 2049 WW background normalization is estimated from data in the three signal regions separately,
 2050 owing to the different m_T^j shapes for signal and background.

2051 For the estimation of the top quark background, three control regions enriched in b-jets
 2052 are defined by selecting events that pass the WW baseline selections and applying a b
 2053 tagging requirement which depends on the jet category as follows:

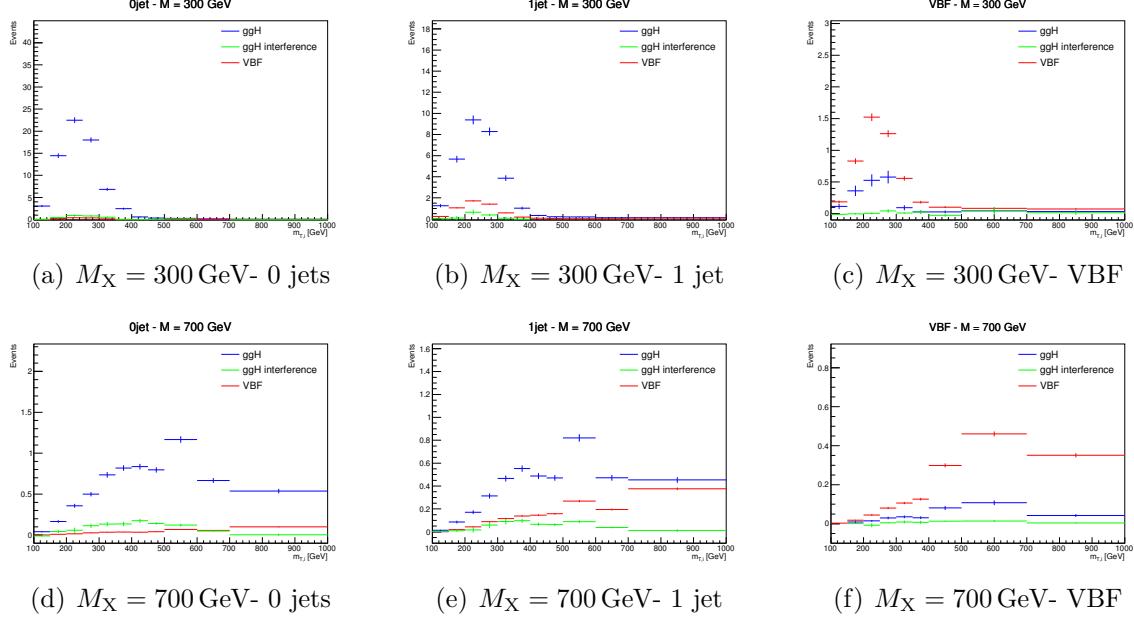


Figure 6.4: Distributions of the m_T^i variable for $M_X = 300$ and 700 GeV , showing the signal (both the ggH and VBF processes) and the interference contributions in the three jet categories.

• 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. ??.

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 ???. A crosscheck is performed selecting events passing the WW baseline selection but with an $e\mu$ pair with same charge. The m_T^i distributions for this phase space are shown in Fig. ?? 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 ??, 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 ??.

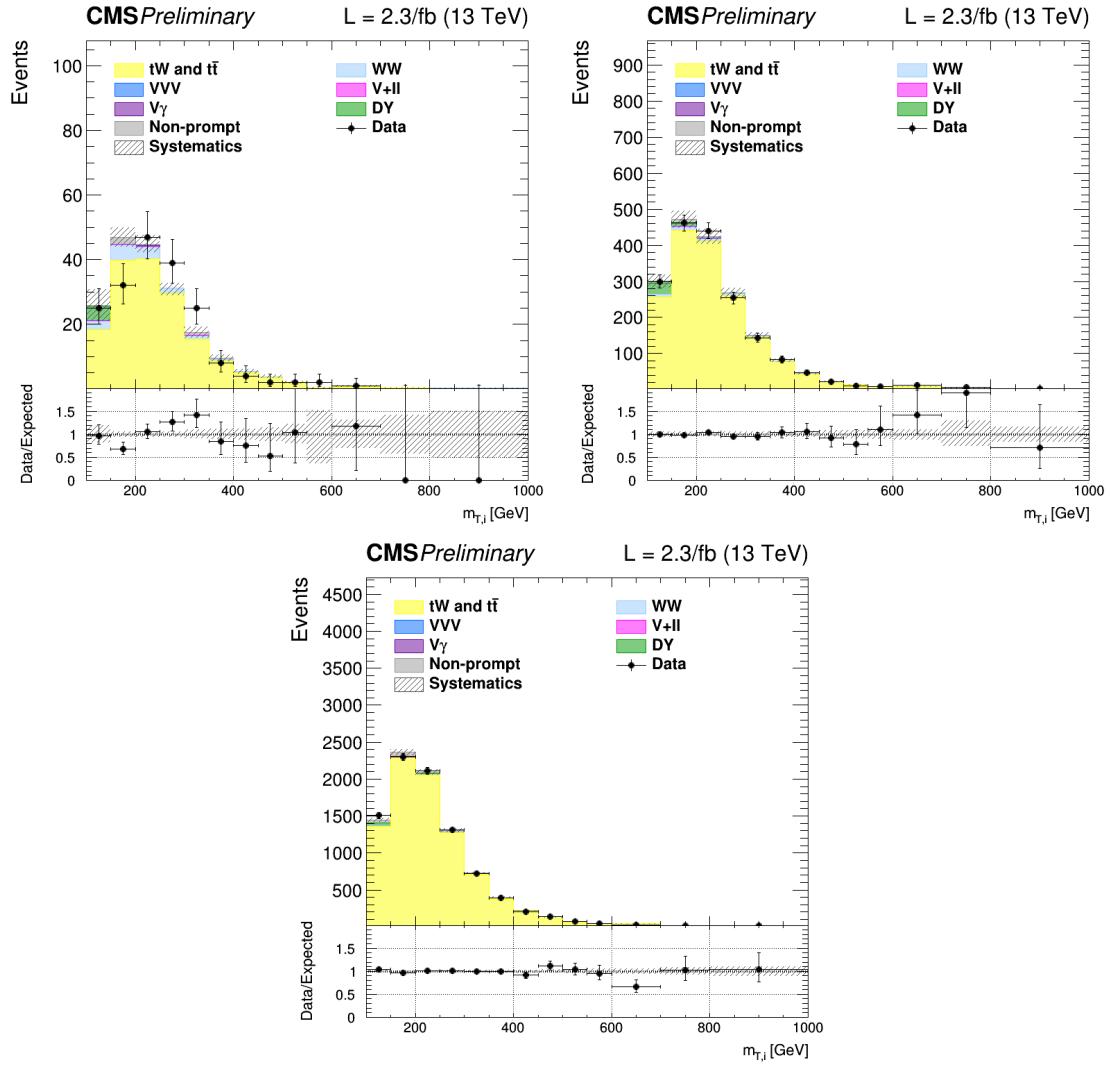


Figure 6.5: Distributions of $m_{T,i}^i$ for events with 0 jet (top left), 1 jet (top right) and 2 jets (bottom) in top enriched control region. Scale factors estimated from data are not applied in the plots.

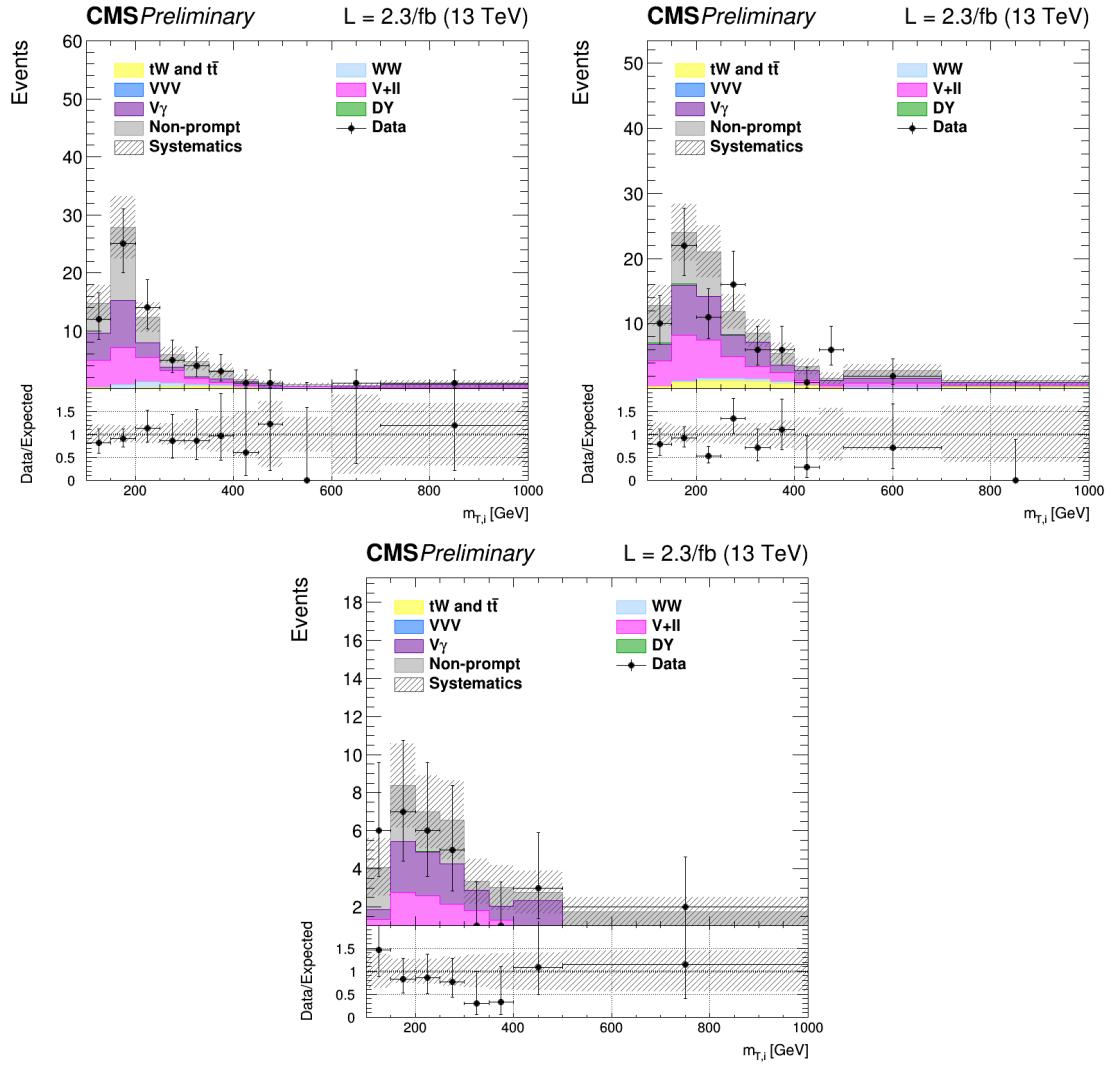


Figure 6.6: Distributions of $m_{T,i}^i$ for events with 0 jet (top left), 1 jet (top right) and 2 jets (bottom) in the same-charge dilepton control region. The last bin of the histograms includes overflows.

2072 6.5 Systematic uncertainties

2073 The systematic uncertainties affecting this analysis are the same discussed in Sec. ???. The
 2074 differences with respect to the SM Higgs boson search are described below.

2075 The PDF and α_s uncertainties on the signal cross sections are taken from the compu-
 2076 tations performed by the LHC cross section working group [95], and are included for all
 2077 the mass points. The value of these uncertainties depends on the resonance mass and vary
 2078 from 3 and 5% for ggH and from 2 and 3% for VBF production modes. The PDFs and α_s
 2079 uncertainties on the signal selection are evaluated for every resonance mass and are found
 2080 to be less than 1% for both ggH and VBF.

2081 The theoretical uncertainties in the signal yields due to the jet categorization are
 2082 evaluated for all the ggH signals following the prescription described in Sec. ???.

2083 An additional uncertainty on the modelling of the top pair background is derived from
 2084 the observed discrepancy between data and POWHEG V2 plus PYTHIA 8.1 simulation
 2085 on the top quark p_T spectrum [105], which is particularly important in the tail of the m_T^i
 2086 distribution. Another uncertainty affecting the m_T^i tail for the top quark background is the
 2087 parton shower uncertainty. This is evaluated comparing the generator level m_T^i distributions
 2088 corresponding to two different simulations of the $t\bar{t}$ process: one obtained using PYTHIA
 2089 8.1 for the showering and hadronization of the simulated events, and the other using
 2090 HERWIG++. The difference between the two is used to extract a shape uncertainty, which
 2091 is less than 1% for low m_T^i values and reaches about 6% in the m_T^i tail.

2092 6.6 Signal extraction and limit setting

2093 The signal yield, including both the ggH and VBF production modes, is extracted performing
 2094 a combined fit of the three categories to the m_T^i simulation templates for backgrounds and
 2095 signal, and is repeated for each resonance mass hypothesis. Moreover, fixed the mass of
 2096 the resonance, the fit is performed again for the various hypotheses of the resonance decay
 2097 width. A single signal strength μ is extracted from the fit, which multiplies both the ggH
 2098 and VBF contributions. In other words it is assumed that the ratio of the two production
 2099 mechanism stays the same as the one predicted by the SM².

2100 The background yields expected from simulation corresponding to the three jet categories
 2101 and after the analysis event selection are shown in Table ???. The signal yields corresponding
 2102 to a selection of mass points and assuming $\Gamma' = \Gamma_{\text{SM}}$ are shown in Table ???.

²This is an approximation which limits the amount of models that can be tested with the provided results.
 A future development of 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 \rightarrow WW	501.93 ± 0.00 (0%)	198.72 ± 0.00 (0%)	4.54 ± 0.00 (0%)
gg \rightarrow 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%)
V γ	26.62 ± 0.72 (3%)	14.18 ± 0.38 (3%)	0.64 ± 0.02 (3%)
V γ^*	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 \rightarrow WW	6.04 ± 0.40 (7%)	3.10 ± 0.11 (5%)	0.34 ± 0.02 (7%)
SM H \rightarrow $\tau\tau$	0.50 ± 0.05 (9%)	0.43 ± 0.04 (9%)	0.04 ± 0.00 (9%)
Total background	876.5	625.9	34.5

2103 The strategy for computing the exclusion limits is based on the modified frequentist
 2104 approach, also referred to as CL_s, as described in [94]. The first step is to construct the
 2105 likelihood function $\mathcal{L}(\text{data}|\mu, \theta)$:

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

2106 where data represents the experimental observation, s and b are the expected signal and
 2107 background yields respectively and θ is the full set of nuisance parameters constrained by
 2108 the prior distribution functions $p(\tilde{\theta}|\theta)$. The default values for the nuisance parameters are
 2109 labelled as $\tilde{\theta}$.

2110 For a binned shape analysis, $\text{Poisson}(\text{data}|\mu \cdot s + b)$ is the product of the Poisson
 2111 probabilities to observe n_i events in bin i:

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

2112 In order to test the compatibility of the data with the signal plus background (or the
 2113 background only) hypothesis, the test statistic \tilde{q}_μ 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 $\Gamma' = \Gamma_{\text{SM}}$ 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%)

2114 likelihood ratio:

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

2115 where $\hat{\theta}_\mu$ refers to the conditional maximum likelihood estimators of θ , given the signal
2116 strength μ . The parameter estimators $\hat{\mu}$ and $\hat{\theta}$ correspond to the global maximum of the
2117 likelihood. The $0 \leq \hat{\mu}$ constraint is imposed to have a positive signal yield, e.g. background
2118 underfluctuations are forbidden, while $\hat{\mu} \leq \mu$ is imposed to have a one-sided confidence
2119 interval. The observed test statistic for the signal strength μ under test is referred to as
2120 $\tilde{q}_\mu^{\text{obs}}$. The values of the nuisance parameters obtained maximising the likelihood function are
2121 labelled as $\hat{\theta}_0^{\text{obs}}$ and $\hat{\theta}_\mu^{\text{obs}}$ for the background only and signal plus background hypotheses,
2122 respectively. The pdf of the test statistic is constructed by generating toy MC pseudo-data
2123 for both the background only and signal plus background hypotheses, i.e. $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}})$ and
2124 $f(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}})$. These distributions can be used to define two p-values corresponding to the

2125 two hypotheses, p_μ and p_b :

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

2126

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

2127 According to these definitions, p_μ and p_b can be identified with CL_{s+b} and $1 - \text{CL}_b$, 2128 respectively. The $\text{CL}_s(\mu)$ is calculated using the following ratio:

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

2129 If, for a given signal strength μ , $\text{CL}_s \leq \alpha$, then the hypothesis is excluded with a $(1 - \alpha)$ 2130 confidence level (CL). For instance, if one wants to quote the upper limit on μ with a 95% 2131 CL, the signal strength has to be adjusted until $\text{CL}_s = 0.05$.

2132 The expected median upper limit, as well as the $\pm 1\sigma$ (68% CL) and $\pm 2\sigma$ (95% CL) 2133 bands, are determined generating a large amount of pseudo-data in the background only 2134 hypothesis and calculating CL_s and the 95% CL upper limit for each of them, as if they 2135 were real data. Then the cumulative distribution of the 95% CL upper limits is built and 2136 the median expected value is identified as the value at which the cumulative distribution 2137 crosses the 50% quantile. The $\pm 1\sigma$ ($\pm 2\sigma$) band is defined by the values at which the 2138 cumulative distribution crosses the 16% (2.5%) and 84% (97.5%) quantiles.

2139 In order to assess the sensitivity of the analysis, the expected upper exclusion limits at 2140 95% CL on the signal strength are shown in Fig. ?? for the three jet categories separately. 2141 For a given mass of the resonance, the limits are derived assuming a signal decay width 2142 $\Gamma' = \Gamma_{\text{SM}}$ and a cross section equal to the one expected from a SM Higgs boson at that 2143 mass. The other decay width hypothesis have also been tested, showing a very similar 2144 expected exclusion limit, suggesting that this analysis is not strongly sensitive to variations 2145 of the resonance decay width.

2146 The 0 jets category is the most sensitive especially in the low mass region, while for 2147 very large masses of the resonance the 1 jet and VBF categories start being important. 2148 This is explained mainly by the fact that the VBF contribution increases, with respect to 2149 ggH, as the mass increases. The expected exclusion limit on the signal strength after the 2150 combination of the three categories is shown in Fig. ??.. Comparing the limits in the single 2151 categories with the combination of the three it is evident how the higher jet multiplicity 2152 categories help in improving the results for large values of M_X .

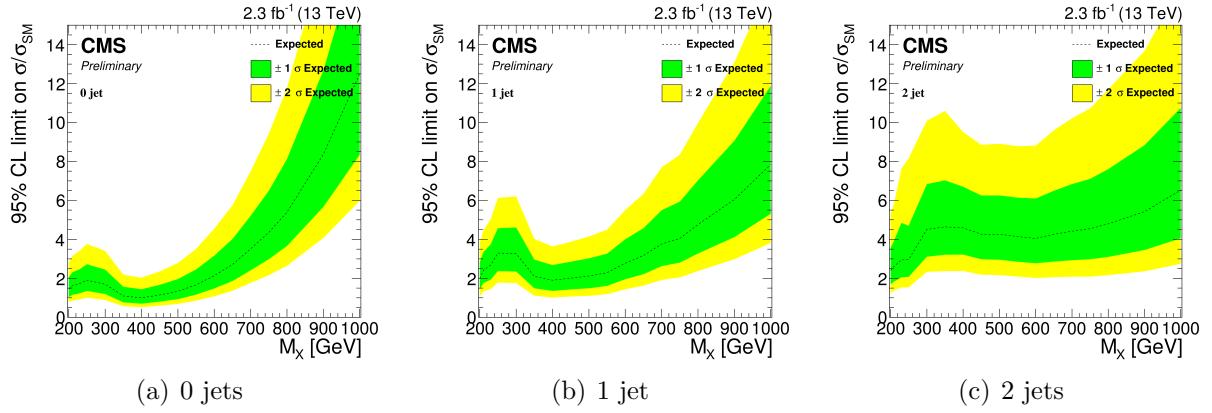


Figure 6.7: Expected exclusion upper limits at 95% CL on the signal strength in the three categories, as a function of the resonance mass. The dashed line corresponds to median upper limit, while the green and yellow regions represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands, respectively.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

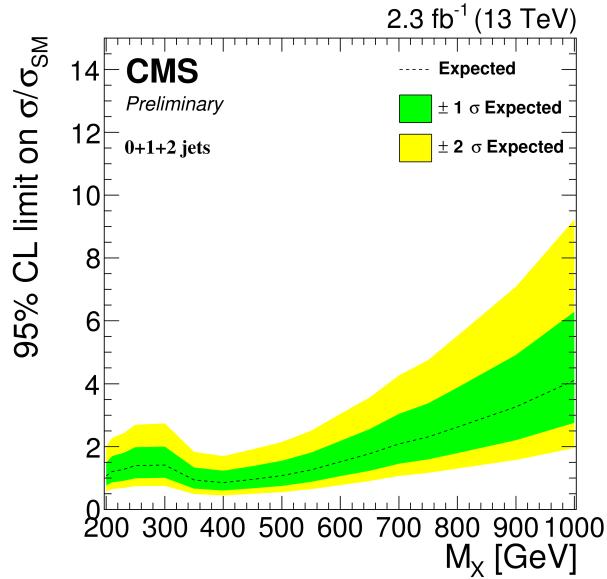


Figure 6.8: Expected exclusion upper limit at 95% CL on the signal strength for the combination of the three categories, as a function of the resonance mass. The dashed line corresponds to median upper limit, while the green and yellow regions represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty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

2153 6.7 Results

2154 The m_T^i distributions for the signal region after the full analysis selection are shown in Fig. ??
 2155 for the three jet categories. Two different signal hypotheses corresponding to $M_X = 400 \text{ GeV}$
 2156 and $M_X = 800 \text{ GeV}$ are shown superimposed on the background for comparison.

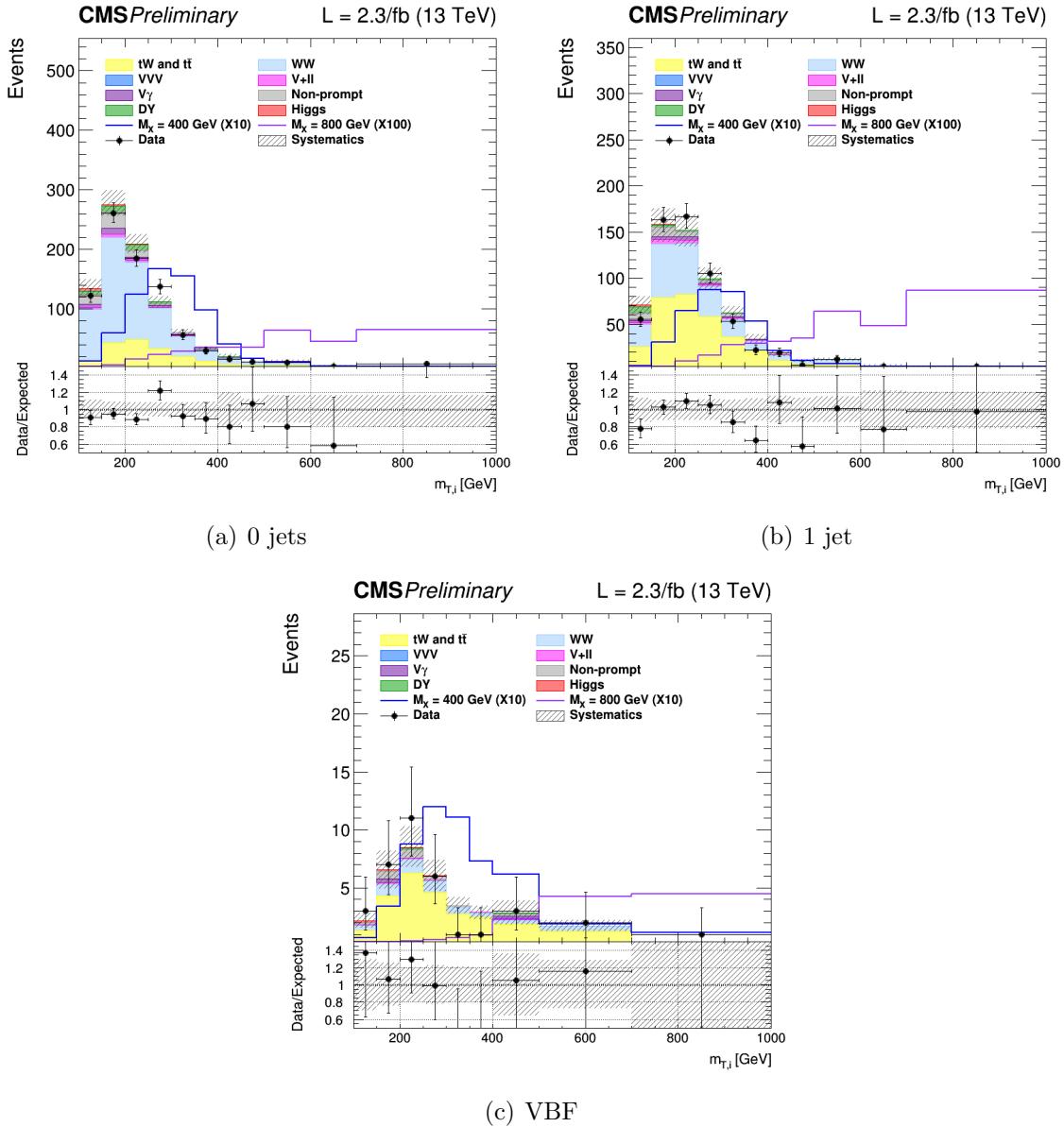


Figure 6.9: Distributions of m_T^i in the signal region for the 0 jets, 1 jet and VBF categories. Background normalisations correspond to the pre-fit value. Signal contributions for two mass hypotheses, $M_X = 400 \text{ GeV}$ and $M_X = 800 \text{ GeV}$, are shown superimposed on the background and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

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)

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 $\Gamma' = \Gamma_{SM}$ are shown in Fig. ??.

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

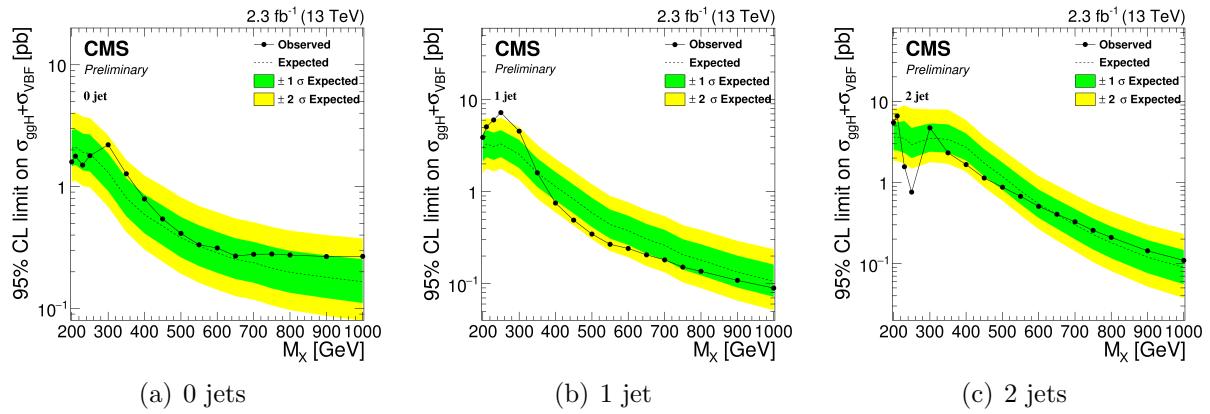


Figure 6.10: Expected and observed exclusion upper limits at 95% CL on $\sigma \times \mathcal{B}$ in the three categories, as a function of the resonance mass. The dashed line corresponds to median upper limit, while the green and yellow regions represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty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

mainly due to an underfluctuation of the background. This effect can be understood looking at the VBF shape in Fig. ??, 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. ??, 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. ?? is thus excluded with a 95% CL for masses ranging from 200 GeV up to 1 TeV.

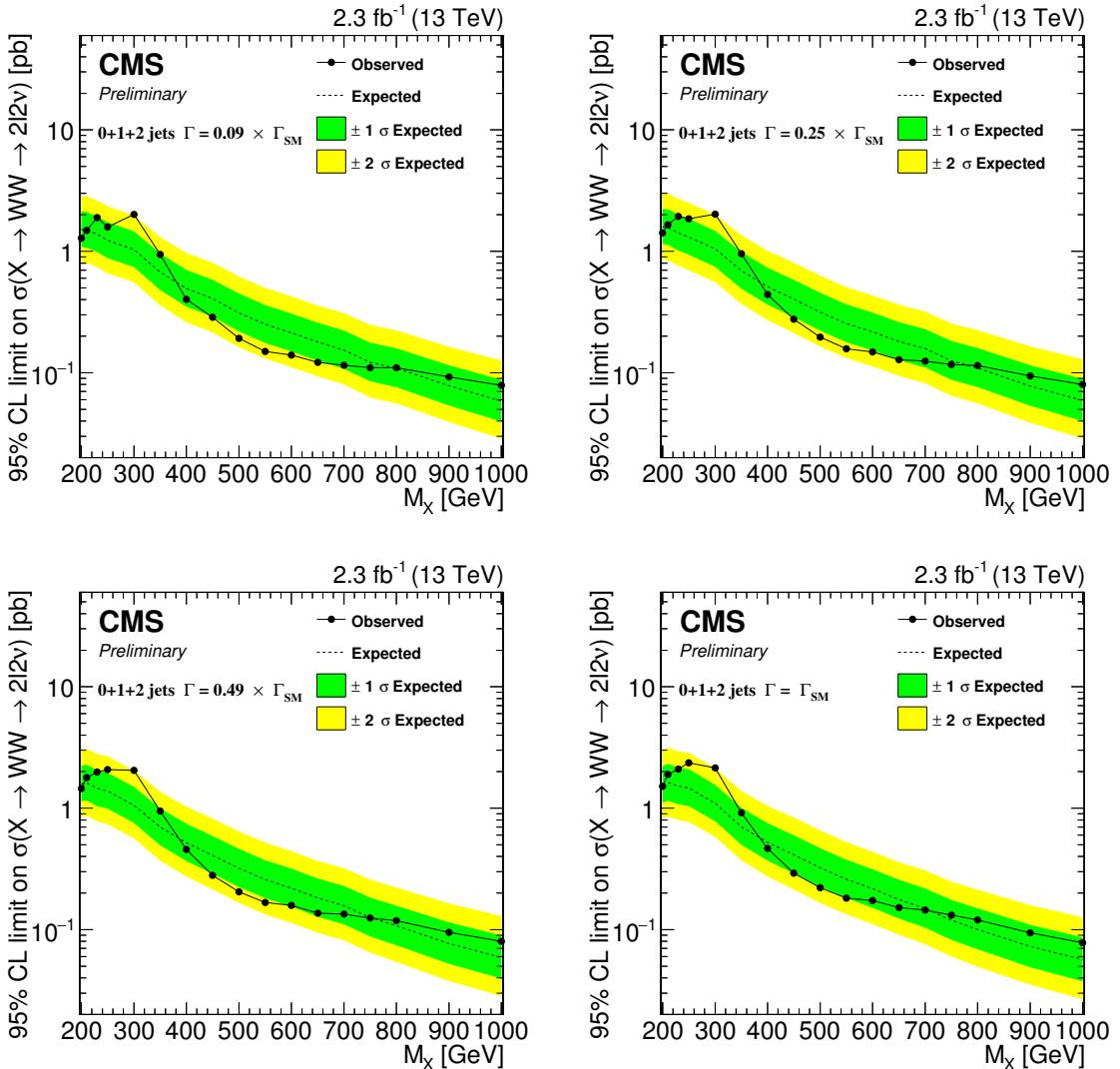


Figure 6.11: Expected and observed exclusion limits at 95% CL on $\sigma \times \mathcal{B}$ for the combination of the three jet categories as a function of the resonance mass. The black dotted line corresponds to the observed value while the yellow and green bands represent the $\pm 1\sigma$ and $\pm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

Chapter 7

Conclusions

2177

Appendix A

²¹⁷⁸ Fiducial region definition and ²¹⁷⁹ optimization

²¹⁸⁰ The fiducial region must be chosen in such a way to be as close as possible to the selections
²¹⁸¹ applied in the analysis, in order to reduce the model dependence in the extrapolation step.
²¹⁸² That means that for optimizing the fiducial volume definition, the efficiency has to be
²¹⁸³ maximized. Another parameter entering the game is the number of fake events, in other
²¹⁸⁴ words the number of reconstructed events which do not belong to the fiducial phase space.
²¹⁸⁵ This parameter should instead be as small as possible. Even if we have to observe the trend
²¹⁸⁶ of these two quantities as a function of p_T^H , we can maximize the ratio between the overall
²¹⁸⁷ efficiency and the overall fake rate as a proxy for establishing the “goodness” of the fiducial
²¹⁸⁸ region.

²¹⁸⁹ Several different fiducial region definitions were tested and the results show that:

- ²¹⁹⁰ • **of cut:** The fiducial region definition must include only the opposite flavor combination
²¹⁹¹ including one electron and one muon. If we include also the combinations involving
²¹⁹² τ 's the efficiency falls down.
- ²¹⁹³ • **Lepton cut:** Since the resolution on lepton transverse momentum is good, there is no
²¹⁹⁴ need to loosen the cuts related these variables, i.e. we can use the same cuts defined
²¹⁹⁵ in the analysis selection ($p_T^{\ell,1} > 20 \text{ GeV}$, $p_T^{\ell,2} > 10 \text{ GeV}$).
- ²¹⁹⁶ • **Di-lepton p_T cut:** As stated in the previous point, there is no need to loosen this
²¹⁹⁷ cut, so we kept the same value as the analysis selection, i.e. $p_T^{\ell\ell} > 30 \text{ GeV}$.
- ²¹⁹⁸ • **Di-lepton mass cut:** $m_{\ell\ell} > 12 \text{ GeV}$ as discussed before.
- ²¹⁹⁹ • **neutrino pair p_T cut:** Since the resolution on the measurement of the missing
²²⁰⁰ transverse energy is poor, the neutrino pair cut should not be included in the definition
²²⁰¹ of the fiducial region, because it would increase the fake rate without increasing the
²²⁰² efficiency, thus resulting in a lower ratio between overall efficiency and fake rate.
- ²²⁰³ • **m_T cut:** Also the m_T cut that we have in the analysis selection, i.e. $m_T > 60 \text{ GeV}$,
²²⁰⁴ 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_T^H after the optimization discussed before are shown in figure ???. 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.

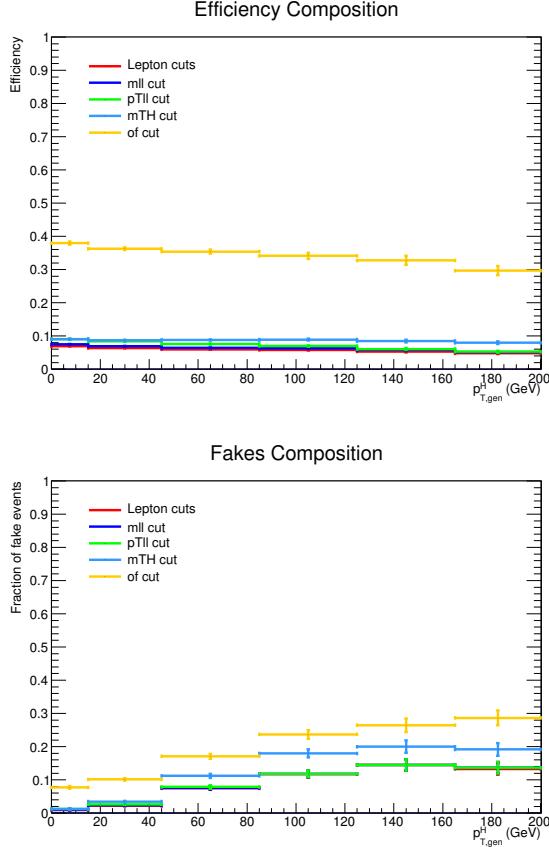


Figure A.1: Efficiency and fake rate as a function on Higgs transverse momentum. The plots correspond to the optimized fiducial region definition and show the effect of adding each of the mentioned cuts in sequence.

The efficiency and fraction of fake events have been measured also as a function of the E_T^{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. ?? 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_T^{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 , $\epsilon - f$

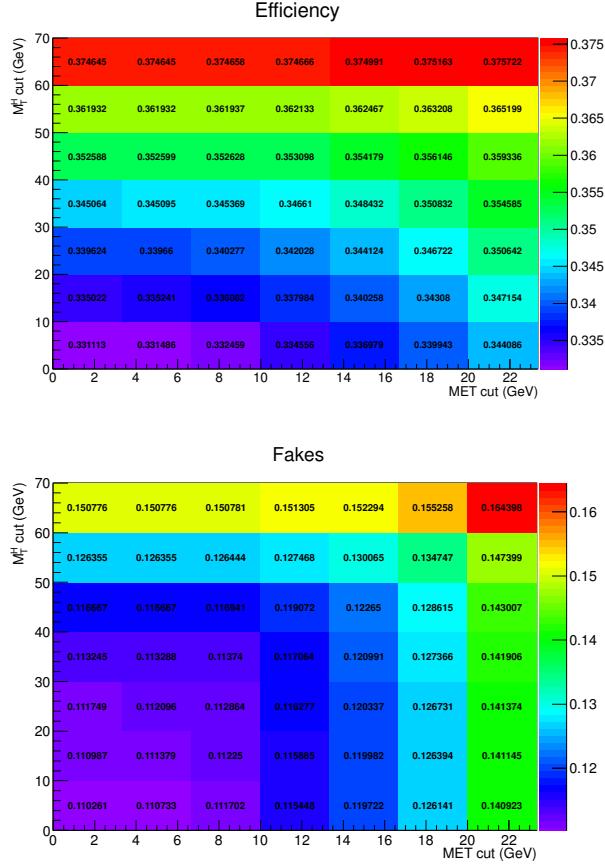


Figure A.2: Efficiency and fake rate as a function of E_T^{miss} and m_T cuts in the fiducial region.

and $(1 - f)/\epsilon$. The results for these three different figures of merit are shown in Fig. ?? as a function of the E_T^{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. ?? and Fig. ??, show a similar trend with respect to the model with nominal ggH/VBF ratio.

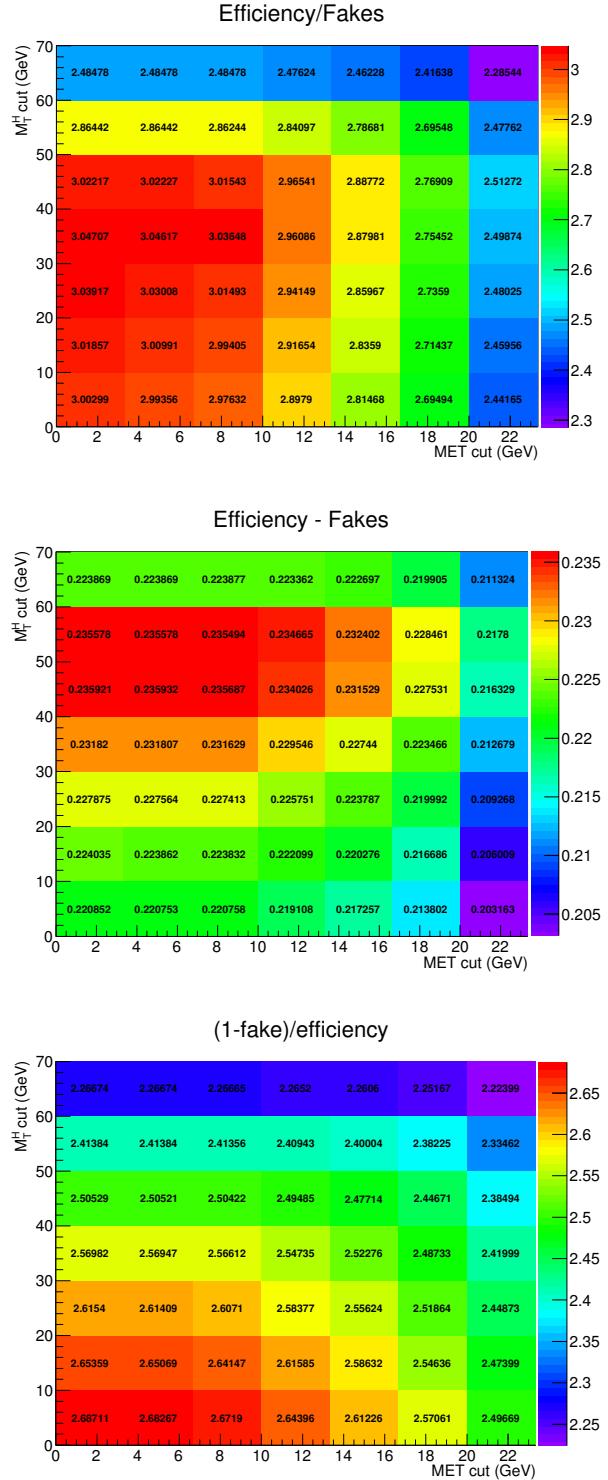


Figure A.3: Different figures of merit as a function of E_T^{miss} and m_T cuts in the fiducial region.

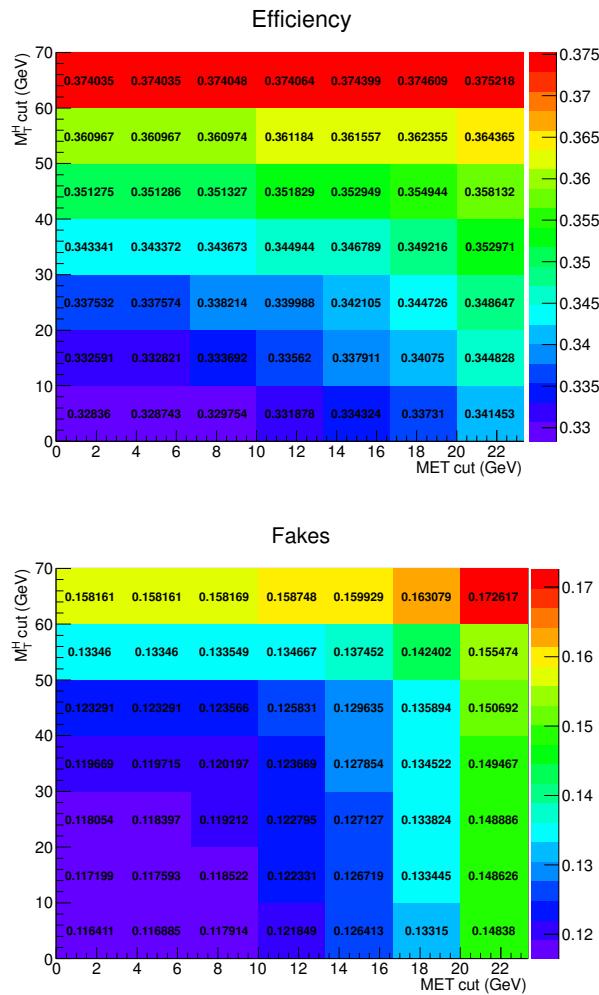


Figure A.4: Efficiency and fake rate as a function of E_T^{miss} and m_T cuts in the fiducial region, for the alternative model with an up variation of the ggH/VBF ratio.

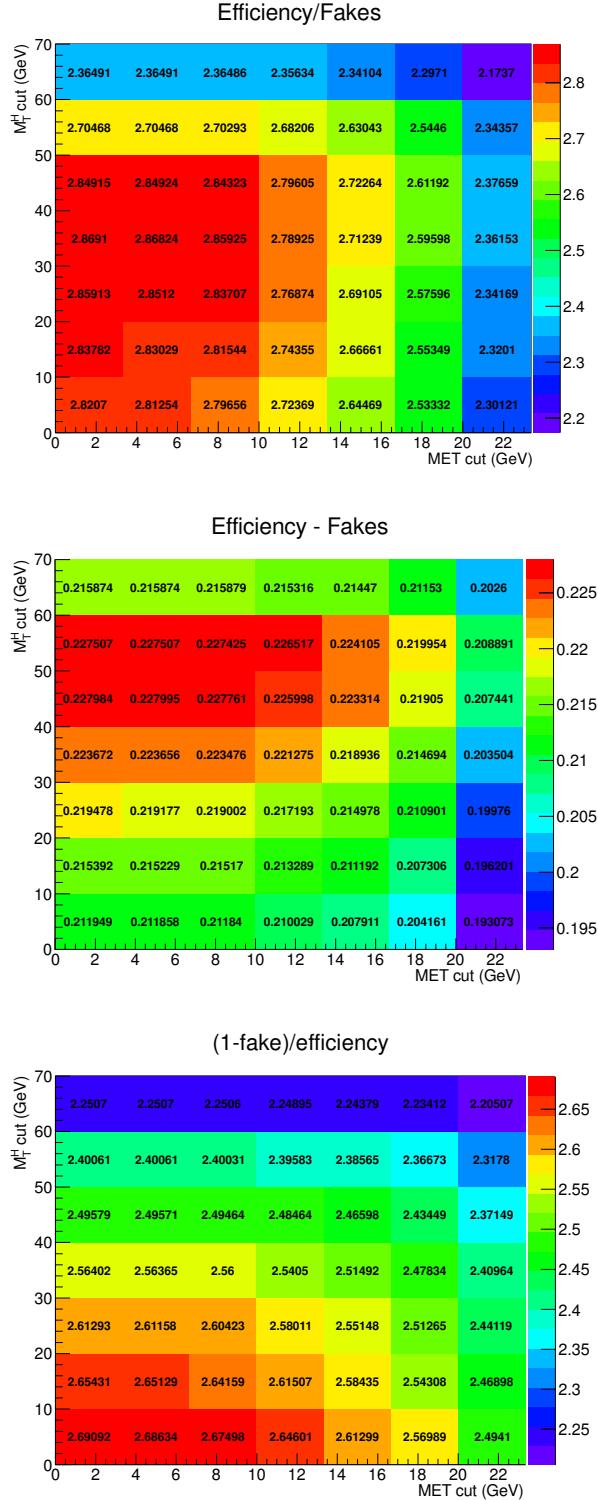


Figure A.5: Different figures of merit as a function of E_T^{miss} and m_T cuts in the fiducial region, for the alternative model with an up variation of the ggH/VBF ratio.

2226

Bibliography

- 2227 [1] Thomas Sven Pettersson and P Lefèvre. *The Large Hadron Collider: conceptual*
2228 *design*. Tech. rep. CERN-AC-95-05-LHC. Oct. 1995. URL: <http://cds.cern.ch/record/291782>.
- 2230 [2] Oliver Sim Brüning et al. *LHC Design Report*. Geneva: CERN, 2004. URL: <http://cds.cern.ch/record/782076>.
- 2232 [3] Oliver Sim Brüning et al. *LHC Design Report*. Geneva: CERN, 2004. URL: <http://cds.cern.ch/record/815187>.
- 2234 [4] Michael Benedikt et al. *LHC Design Report*. Geneva: CERN, 2004. URL: <http://cds.cern.ch/record/823808>.
- 2236 [5] *LEP design report*. By the LEP Injector Study Group. Geneva: CERN, 1983. URL:
2237 <http://cds.cern.ch/record/98881>.
- 2238 [6] *LEP design report*. Copies shelved as reports in LEP, PS and SPS libraries. Geneva:
2239 CERN, 1984. URL: <http://cds.cern.ch/record/102083>.
- 2240 [7] Carlo Wyss. *LEP design report, v.3: LEP2*. Vol. 1-2 publ. in 1983-84. Geneva: CERN,
2241 1996. URL: <http://cds.cern.ch/record/314187>.
- 2242 [8] S. Chatrchyan et al. “The CMS experiment at the CERN LHC”. In: *JINST* 3 (2008),
2243 S08004. DOI: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- 2244 [9] G. Aad et al. “The ATLAS Experiment at the CERN Large Hadron Collider”. In:
2245 *JINST* 3 (2008), S08003. DOI: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- 2246 [10] A. Augusto Alves Jr. et al. “The LHCb Detector at the LHC”. In: *JINST* 3 (2008),
2247 S08005. DOI: [10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005).
- 2248 [11] K. Aamodt et al. “The ALICE experiment at the CERN LHC”. In: *JINST* 3 (2008),
2249 S08002. DOI: [10.1088/1748-0221/3/08/S08002](https://doi.org/10.1088/1748-0221/3/08/S08002).
- 2250 [12] G. Anelli et al. “The TOTEM experiment at the CERN Large Hadron Collider”. In:
2251 *JINST* 3 (2008), S08007. DOI: [10.1088/1748-0221/3/08/S08007](https://doi.org/10.1088/1748-0221/3/08/S08007).
- 2252 [13] O. Adriani et al. “The LHCf detector at the CERN Large Hadron Collider”. In:
2253 *JINST* 3 (2008), S08006. DOI: [10.1088/1748-0221/3/08/S08006](https://doi.org/10.1088/1748-0221/3/08/S08006).
- 2254 [14] *The CMS magnet project: Technical Design Report*. Technical Design Report CMS.
2255 Geneva: CERN, 1997. URL: <http://cds.cern.ch/record/331056>.

- [15] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: <http://cds.cern.ch/record/349375>.
- [16] Philippe Bloch et al. *Changes to CMS ECAL electronics: addendum to the Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 2002. URL: <http://cds.cern.ch/record/581342>.
- [17] 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 \[hep-ph\]](#).
- [18] 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 \[hep-ph\]](#).
- [19] 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 \[hep-ph\]](#).
- [20] 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 \[hep-ph\]](#).
- [21] 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 \[hep-ph\]](#).
- [22] X. Chen et al. “NNLO QCD corrections to Higgs boson production at large transverse momentum”. In: (2016). arXiv: [1607.08817 \[hep-ph\]](#).
- [23] 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 \[hep-ph\]](#).
- [24] 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 \[hep-ph\]](#).
- [25] 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 \[hep-ph\]](#).
- [26] 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 \[hep-ph\]](#).

- [27] 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].
- [28] Georges Aad et al. “Fiducial and differential cross sections of Higgs boson production measured in the four-lepton decay channel in pp collisions at $\sqrt{s} = 8$ 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].
- [29] Georges Aad et al. “Measurements of fiducial and differential cross sections for Higgs boson production in the diphoton decay channel at $\sqrt{s} = 8$ TeV with ATLAS”. In: *JHEP* 09 (2014), p. 112. DOI: 10.1007/JHEP09(2014)112. arXiv: 1407.4222 [hep-ex].
- [30] Georges Aad et al. “Measurements of the Total and Differential Higgs Boson Production Cross Sections Combining the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ 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].
- [31] Vardan Khachatryan et al. “Measurement of differential cross sections for Higgs boson production in the diphoton decay channel in pp collisions at $\sqrt{s} = 8$ TeV”. In: *Eur. Phys. J. C* 76 (2016), p. 13. DOI: 10.1140/epjc/s10052-015-3853-3. arXiv: 1508.07819 [hep-ex].
- [32] 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].
- [33] 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].
- [34] 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].
- [35] G. Cowan. “A survey of unfolding methods for particle physics”. In: *Conf. Proc. C*0203181 (2002), p. 248.
- [36] 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].
- [37] 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].

- [38] 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].
- [39] 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].
- [40] 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].
- [41] 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].
- [42] 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].
- [43] 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].
- [44] Marco Bonvini et al. “Signal-background interference effects for $gg \rightarrow H \rightarrow W^+W^-$ beyond leading order”. In: *Phys. Rev. D* 88 (2013), p. 034032. DOI: 10.1103/PhysRevD.88.034032. arXiv: 1304.3053 [hep-ph].
- [45] 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].
- [46] 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].
- [47] 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].
- [48] 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].
- [49] LHC Higgs Cross Section Working Group. “Handbook of LHC Higgs Cross Sections”. In: *arXiv:1101.0593* (2011).
- [50] 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].

- [51] 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.
- [52] D. de Florian et al. “Higgs boson production at the LHC: transverse momentum resummation effects in the $H \rightarrow 2\gamma$, $H \rightarrow WW \rightarrow l\nu l\nu$ and $H \rightarrow ZZ \rightarrow 4l$ decay modes”. In: *JHEP* 06 (2012), p. 132. DOI: 10.1007/JHEP06(2012)132. arXiv: 1203.6321 [hep-ph].
- [53] 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].
- [54] 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].
- [55] 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].
- [56] 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].
- [57] 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].
- [58] 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].
- [59] 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].
- [60] Vardan Khachatryan et al. “Measurement of the $t\bar{t}$ production cross section in the $e\mu$ channel in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV”. In: (2016). arXiv: 1603.02303 [hep-ex].
- [61] 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.
- [62] CMS Collaboration. “Standard Model Cross Sections for CMS at 7 TeV”. In: *CMS Generator Group Twiki* (2010).
- [63] 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].

- [64] 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.
- [65] 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.
- [66] Sergey Alekhin et al. “The PDF4LHC Working Group Interim Report”. 2011.
- [67] Michiel Botje et al. “The PDF4LHC Working Group Interim Recommendations”. 2011.
- [68] 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].
- [69] 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].
- [70] 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].
- [71] “Procedure for the LHC Higgs boson search combination in summer 2011”. In: (2011).
- [72] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *Eur. Phys. J. C* 71 (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].
- [73] Tim Adye. “Unfolding algorithms and tests using RooUnfold”. 2011.
- [74] 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].
- [75] 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.
- [76] 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].
- [77] 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].

- [78] S. Bolognesi, Y. Gao and A.V. Gritsan *et al.* “JHUGen”. In: URL <http://www.pha.jhu.edu/spin/> (2011).
- [79] 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].
- [80] 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].
- [81] 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].
- [82] 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].
- [83] 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].
- [84] John M. Campbell, R. Keith Ellis, and Ciaran Williams. “Bounding the Higgs width at the LHC: Complementary results from $H \rightarrow WW$ ”. In: *Phys. Rev.* D89.5 (2014), p. 053011. DOI: 10.1103/PhysRevD.89.053011. arXiv: 1312.1628 [hep-ph].
- [85] 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].
- [86] 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].
- [87] Vardan Khachatryan et al. “Event generator tunes obtained from underlying event and multiparton scattering measurements”. In: (2015). arXiv: 1512.00815 [hep-ex].
- [88] 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].
- [89] J. Bellm et al. “Herwig++ 2.7 Release Note”. In: (2013). arXiv: 1310.6877 [hep-ph].
- [90] *SM Higgs production cross sections at $\sqrt{s} = 13\text{--}14\text{ TeV}$* . <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt1314TeV>.
- [91] 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].

- [92] *NLO single-top channel cross sections*. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec>.
- [93] *NNLO+NNLL top-quark-pair cross sections*. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO>.
- [94] *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.
- [95] *LHC Higgs Cross Section Working Group*. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>.
- [96] 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].
- [97] 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].
- [98] 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].
- [99] 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].
- [100] 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].
- [101] 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.
- [102] 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].
- [103] 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].
- [104] Massimiliano Grazzini. “NNLO predictions for the Higgs boson signal in the $H \rightarrow WW \rightarrow \ell\nu\ell\nu$ and $H \rightarrow ZZ \rightarrow 4\ell$ decay channels”. In: *JHEP* 02 (2008), p. 043. DOI: 10.1088/1126-6708/2008/02/043. arXiv: 0801.3232 [hep-ph].
- [105] Vardan Khachatryan et al. “Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV”. In: *Eur. Phys. J. C* 75.11 (2015), p. 542. DOI: 10.1140/epjc/s10052-015-3709-x. arXiv: 1505.04480 [hep-ex].