

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

Lorenzo Viliani  
of University of Florence

PhD Thesis



## Abstract

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



<sup>25</sup> **Contents**







# Chapter 1

<sup>27</sup> Electroweak and QCD physics at  
<sup>28</sup> LHC



# Chapter 2

## <sup>29</sup> The CMS experiment at the LHC

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

### <sup>32</sup> 2.1 The Large Hadron Collider

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

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

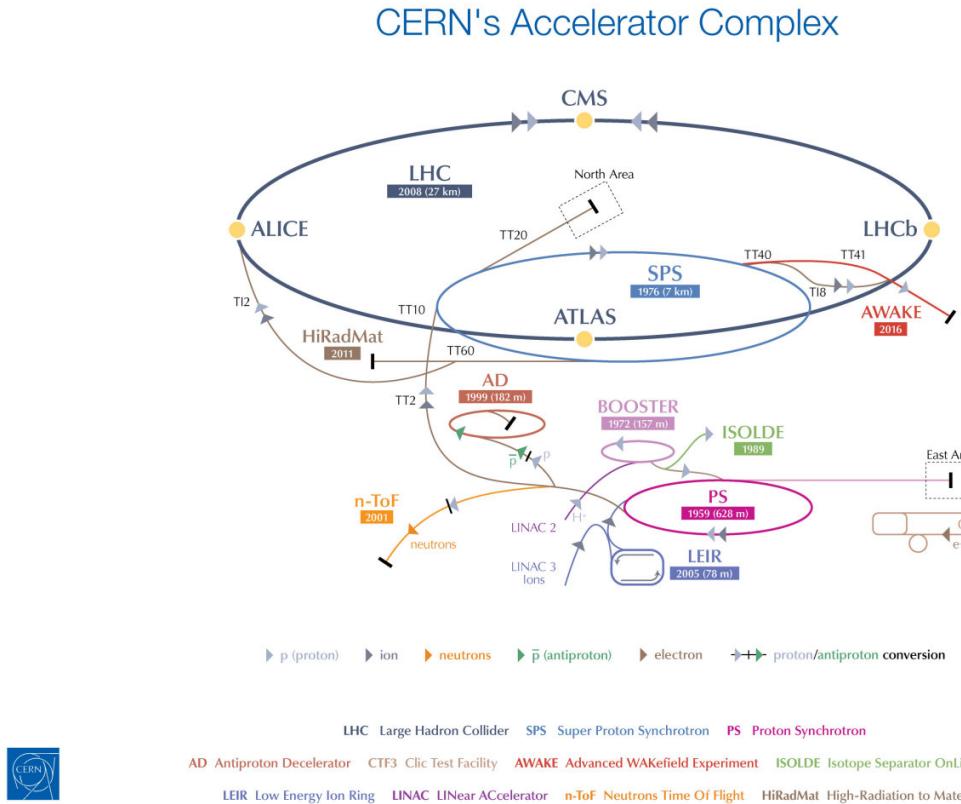
<sup>52</sup> In Fig. 2.1 a schematic description of the accelerator complex installed at CERN is shown.  
<sup>53</sup> The acceleration is performed in several stages [4]. The protons source is a *Duoplasmatron*:  
<sup>54</sup> the protons are obtained by removing electrons from a source of hydrogen gas and then  
<sup>55</sup> sent to the LINAC2, a 36 m long linear accelerator which generates a pulsed beam with  
<sup>56</sup> an energy of 50 MeV using Radio Frequency Quadrupoles (RFQ) and focusing quadrupole  
<sup>57</sup> magnets. The beam is subsequently sent to the Proton Synchrotron Booster (PSB), a  
<sup>58</sup> circular accelerator consisting of four superimposed synchrotron rings with a circumference  
<sup>59</sup> of about 160 m, which increases the proton energy up to 1.4 GeV. Then, protons are

60 injected into the Proton Synchrotron (PS), a single synchrotron ring with a circumference  
 61 of about 600 m where the energy is increased to 25 GeV. The sequential combination of  
 62 these two synchrotrons also allows to create a series of protons bunches interspersed by  
 63 25 ns as required for the correct operation of LHC. The final proton injection stage is the  
 64 Super Proton Synchrotron (SPS), a synchrotron with a circumference of approximately  
 65 7 km where protons reach an energy value of 450 GeV. Subsequently, protons are extracted  
 66 and injected into the LHC ring via two transmission lines, to generate two beams running  
 67 in opposite directions in two parallel pipes and which are accelerated up to the energy of  
 68 interest. In the two pipes an ultrahigh vacuum condition is maintained (about  $10^{-10}$  Torr)  
 69 to avoid the spurious proton interactions with the gas remnants. At full intensity, each  
 70 proton beam consists of 2808 bunches and each bunch contains around  $10^{11}$  protons. The  
 71 beams are squeezed and collide for a length of about 130 m at four interaction points where  
 72 the four main experiments (ALICE, ATLAS, CMS and LHCb) are placed:

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

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

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



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

radio-frequency resonant cavities, is used to increase the proton energy by 0.5 MeV with each beam revolution. The 7 TeV per-beam-energy limit on the LHC is not determined by the electric field generated by the radiofrequency cavity but by the magnetic field necessary to maintain the protons in orbit, given the current technology for the superconducting magnets, which is about 5.4 T on average.

One of the most important parameters of an accelerator is the instantaneous luminosity  $\mathcal{L}$ , which gives a measure of the rate of events one can expect given the process cross section. In fact, for a given physics process with cross section  $\sigma$ , producing  $N$  events for unit of time, the instantaneous luminosity is defined by the following equation:

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

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

113      The instantaneous luminosity is a parameter which depends on the construction charac-  
114    teristics of the accelerator, and can be expressed by the following approximated formula:

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

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

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

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

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

136    As the instantaneous luminosity increases, the probability of multiple proton proton  
137    interactions to occur in a single bunch crossing grows higher as well. In this instance, the  
138    main goal is the identification and reconstruction of a single primary collision where the  
139    physics event of interest occurs among the background of the additional proton proton  
140    interactions. Such backgrounds are due to processes occurring with very high probability,  
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  
143    exhibiting over 45 pile up collisions.

**Table 2.1:** LHC technical parameters for proton proton collisions.

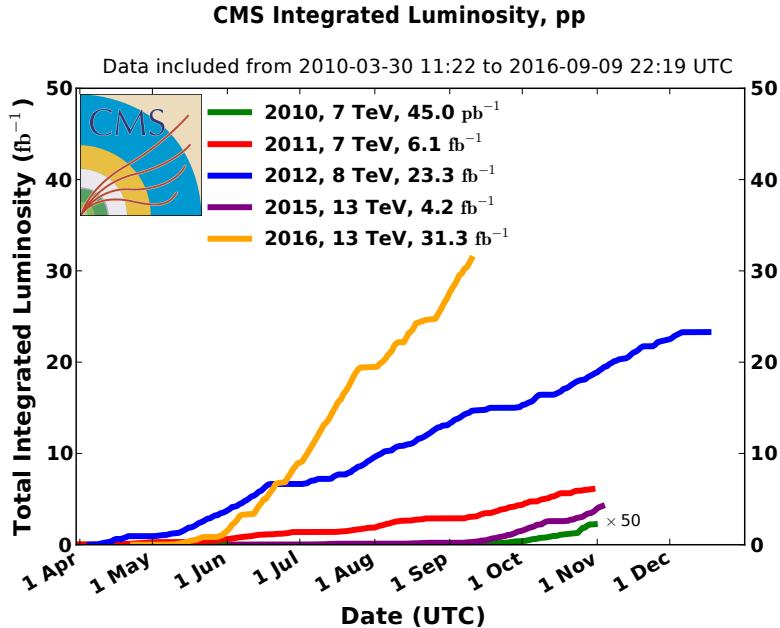
Parameter	Value
Maximum dipole magnetic field	8.33 T
Dipole operating temperature	1.9 K
Beam energy at injection	450 GeV
Beam energy at collision (nominal)	7 TeV
Beam energy at collision (2012)	4 TeV
Beam energy at collision (2015–2016)	6.5 TeV
Maximum instantaneous luminosity (nominal)	$10^{34} \text{ cm}^{-2}\text{s}^{-1}$
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

## <sup>144</sup> 2.2 The *Compact Muon Solenoid* experiment

<sup>145</sup> The CMS apparatus is a general purpose detector situated in one of the four LHC interaction  
<sup>146</sup> points<sup>1</sup>. The detector is designed to investigate a wide range of physics, from the search of  
<sup>147</sup> the Higgs boson, to SM measurements and BSM physics searches. To achieve this goal, the  
<sup>148</sup> detector is able to identify and reconstruct all the physics objects that may be produced  
<sup>149</sup> in the proton proton collisions: electrons, muons, photons and jets. The main feature of  
<sup>150</sup> the CMS detector is a superconducting solenoidal magnet which is capable to produce a

---

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



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

151 3.8 T magnetic field. Such a strong magnetic field is the key aspect which permits to have  
 152 a compact design of the detector. The detector has a cylindrical structure, which is typical  
 153 of general purpose detectors, which consists of several cylindrical detecting layers, coaxial  
 154 with the beam direction (*barrel* region), closed at both ends with detecting disks (*endcap*  
 155 region), in such a way to ensure the hermetic closure of the apparatus.

156 The coordinate system used by CMS is a right-handed Cartesian system, with the origin  
 157 at the center of the detector, in the nominal beam collision point. The  $x$ -axis is chose  
 158 to point radially towards the center of the LHC circumference and the  $y$ -axis is directed  
 159 upwards along the vertical. The  $z$ -axis is oriented along the beam direction, according  
 160 to the anticlockwise direction of the LHC ring if seen from above. The CMS cylindrical  
 161 symmetry and the Lorentz invariant description of the proton proton collisions, suggest the  
 162 use of a pseudo-angular reference frame, described by the triplet of coordinates  $(r, \phi, \eta)$ ,  
 163 where  $r$  is the distance from the  $z$ -axis,  $\phi$  is the azimuthal angle, measured starting from  
 164 the  $x$ -axis positive direction, and  $\eta$  is the pseudorapidity, defined by the following equation:

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

165 where  $\theta$  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  $y$ :

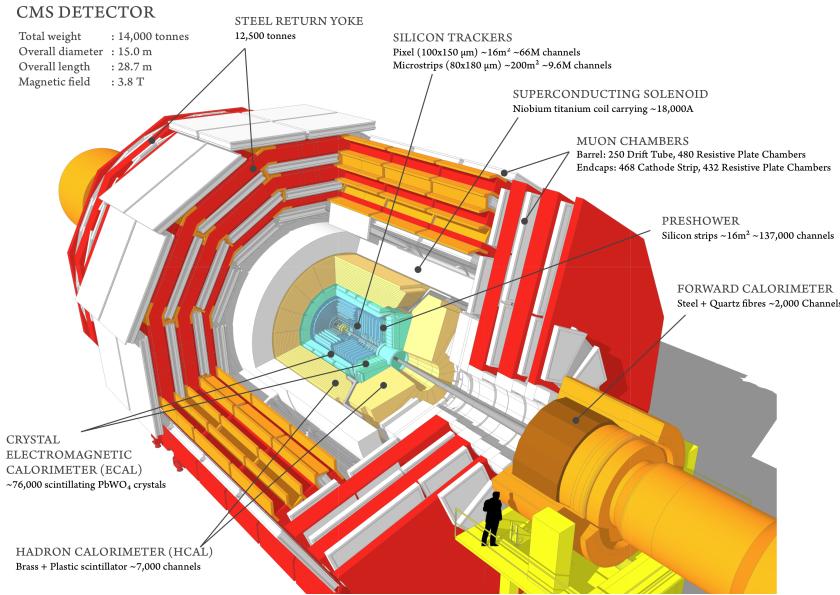
$$y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right) \quad , \quad (2.5)$$

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

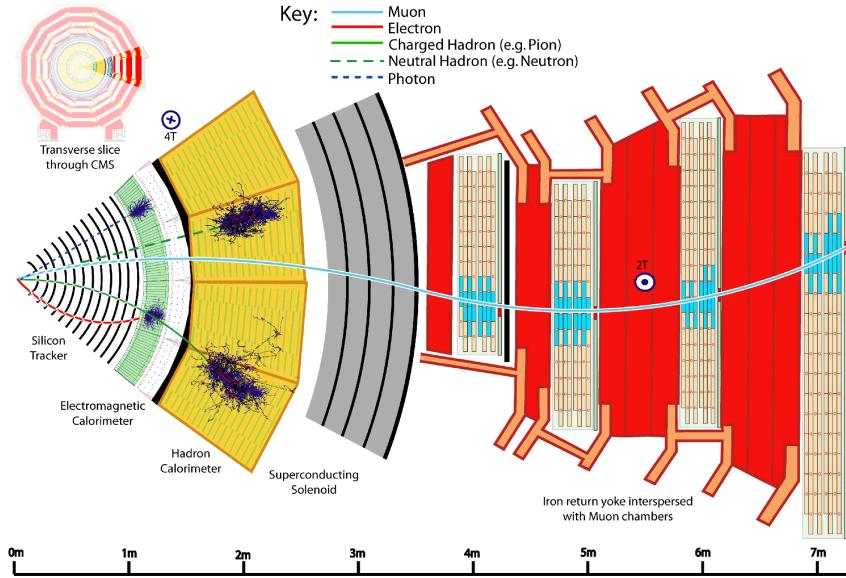
169 The schematic view of the CMS detector, which has a length of 21.5 m, a diameter of  
170 15 m and a weight of about 14000 tons, is shown in Fig. 2.3. From the inner region to the  
171 outer one, the various CMS sub-detectors are:

- 172 • **Silicon tracker:** it occupies the region  $r < 1.2$  m and  $|\eta| < 2.5$ . It is composed of an  
173 inner silicon pixel vertex detector and a surrounding silicon microstrip detector, with  
174 a total active area of about 215 m<sup>2</sup>. It is used to reconstruct charged particle tracks  
175 and vertices;
- 176 • **Electromagnetic calorimeter (ECAL):** placed in the region  $1.2 \text{ m} < r < 1.8$  m  
177 and  $|\eta| < 3$ , it consists of many scintillating crystals of lead tungstate (PbWO<sub>4</sub>). It is  
178 used for the measurement of the trajectory and the energy released by electrons and  
179 photons;
- 180 • **Hadronic calorimeter (HCAL):** it is placed in the region  $1.8 \text{ m} < r < 2.9$  m and  
181  $|\eta| < 5$ . It is made up of brass layers alternated with plastic scintillators and it is  
182 used to measure the direction and energy deposited by the hadrons produced in the  
183 interactions;
- 184 • **Superconducting solenoidal magnet:** it occupies the region  $2.9 \text{ m} < r < 3.8$  m  
185 and  $|\eta| < 1.5$  and generates an internal uniform magnetic field with an intensity of  
186 3.8 T, pointing along the direction of the beams. The magnetic field is necessary to  
187 bend the trajectories of charged particles, in order to allow the measurement of their  
188 momentum through the curvature observed in the tracking system. The magnetic  
189 field lines are closed by an external 21 m long iron yoke, that has a diameter of 14 m.  
190 Outside the return yoke, a residual 1.8 T magnetic field is present, pointing in the  
191 opposite direction with respect to the internal field;
- 192 • **Muon system:** the outermost system, which is placed in the region  $4 \text{ m} < r < 7.4$  m  
193 and  $|\eta| < 2.4$ , has the purpose of reconstructing the tracks of muons passing through  
194 it. It consists of Drift Tubes (DT) in the barrel region and Cathode Strip Chambers  
195 (CSC) in the endcaps. A complementary system of Resistive Plate Chambers (RPC)  
196 is used both in the barrel and endcaps. The muon chambers are housed inside the  
197 iron structure of the return yoke.

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.



**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.

### 2.2.1 The solenoid

The CMS magnet [14], which contains the tracker, the electromagnetic and the hadronic calorimeters, is the biggest superconducting solenoid ever built. The solenoid can generate 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  $\mu\text{m}$ ). 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.

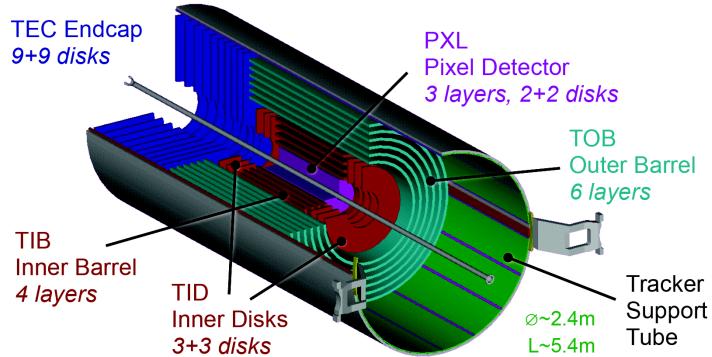
226 The first requirement is achieved building a detector with high granularity<sup>2</sup>. The redundancy  
 227 of the hits is instead achieved having several detecting layers, and is necessary to reduce the  
 228 ambiguity on the assignment of the hits to a given track. Nevertheless, the amount of tracker  
 229 material has to be as low as possible, in order to avoid compromising the measurement  
 230 of the particle trajectory. An excessive amount of material would indeed deteriorate the  
 231 measurement, mainly because of the increased probability of particle multiple scattering.  
 232 The outer detectors such as ECAL would be influenced by the material as well, for example  
 233 because of the increased probability for a photon to convert to an electron-positron pair  
 234 in the tracker material. For this reasons, the tracker layers are limited in number and  
 235 thickness.

236 The tracker comprises a large silicon strip detector with a small silicon pixel detector  
 237 inside it. In the central  $\eta$  region, the pixel tracker consists of three co-axial barrel layers at  
 238 radii between 4.4 cm and 10.2 cm and the strip tracker consists of ten co-axial barrel layers  
 239 extending outwards to a radius of 110 cm. Both sub-detectors are completed by endcaps on  
 240 either side of the barrel, each consisting of two disks in the pixel tracker, and three small  
 241 plus nine large disks in the strip tracker. The endcaps extend the acceptance of the tracker

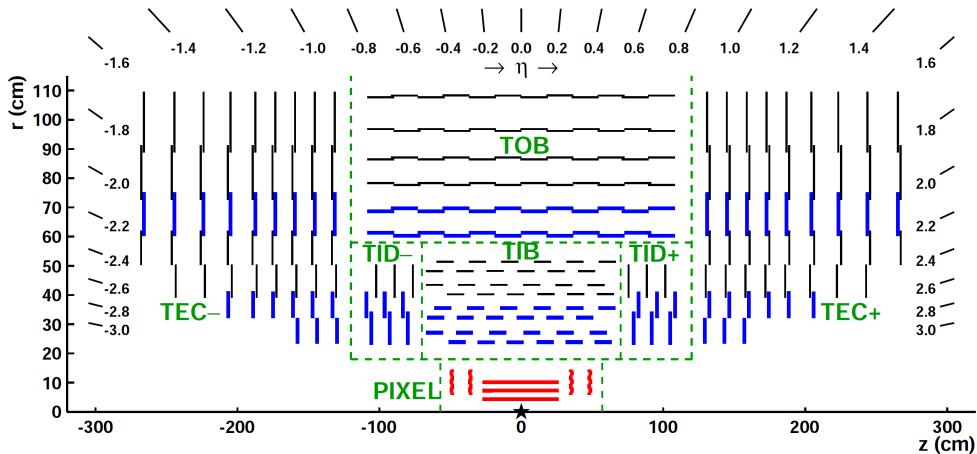
---

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

242 up to  $|\eta| < 2.5$ . A three-dimensional schematic view of the tracker is shown in Fig. 2.5,  
243 while in Fig. 2.6 a pictorial representation of a slice of the tracker is displayed, showing the  
244 various layers of the sub-detectors.



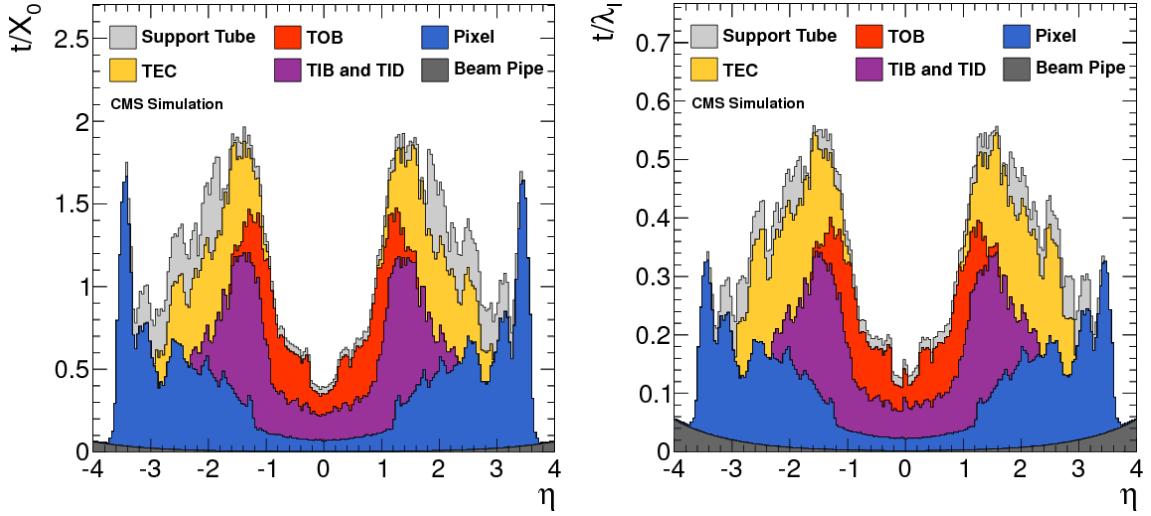
**Figure 2.5:** Three-dimensional schematic view of the CMS silicon tracker.



**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.

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

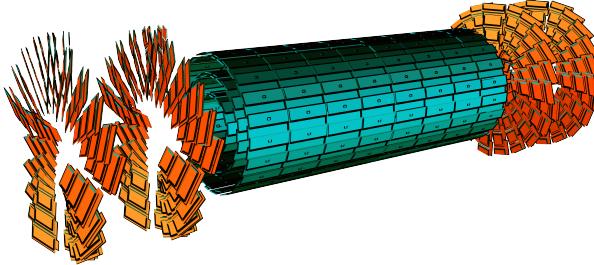
248 The material budget of the tracker as obtained from a simulation of the detector is  
249 shown in Fig. 2.7, reported both in units of radiation length  $t/X_0$  and in units of nuclear  
250 interaction length  $t/\lambda_I$ , as a function of  $\eta$ . The region  $1 < |\eta| < 2$  exhibits a larger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



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

## 252 The pixel detector

253 The pixel detector, shown in Fig. 2.8, is mainly used as starting point in the CMS track  
 254 reconstruction and is of fundamental importance for the reconstruction of primary and  
 secondary vertices. The pixel detector is placed in the closest position to the collision point,



**Figure 2.8:** Schematic view of the CMS pixel detector.

255 where the amount of radiation is larger. It is placed in the region  $|\eta| < 2.5$  and consists of  
 256 three cylindrical layers 53 cm long in the barrel region, located at  $r = 4.4, 7.3$  and  $10.2$  cm,  
 257 and two pairs of endcap disks with radii between 6 and 15 cm at  $z = \pm 34.5$  and  $\pm 46.5$  cm,  
 258 covering a total area of about  $1\text{ m}^2$ . The detector is composed of many modules, for a total  
 259 of 768 in the barrel and 672 in the endcaps. Each endcap is composed of 24 segments,  
 260 each one tilted with respect to the adjacent ones and containing 7 modules. Each module  
 261 consists of several units which contain a highly segmented silicon sensor with a thickness of  
 262 250  $\mu\text{m}$ . In order to achieve an optimal vertex position resolution in both the  $(r, \phi)$  and

264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303

$z$ -coordinates, a design with a rectangular pixel shape with an area of  $150 \times 100 \mu\text{m}^2$  was adopted, with the  $100 \mu\text{m}$  size oriented along the  $(r, \phi)$  direction in the barrel region, and along the  $z$ -direction in the endcap region. The achievable hit reconstruction resolution is about  $10 - 15 \mu\text{m}$  in the barrel and  $15 \mu\text{m}$  in the endcaps.

## The microstrip detector

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

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

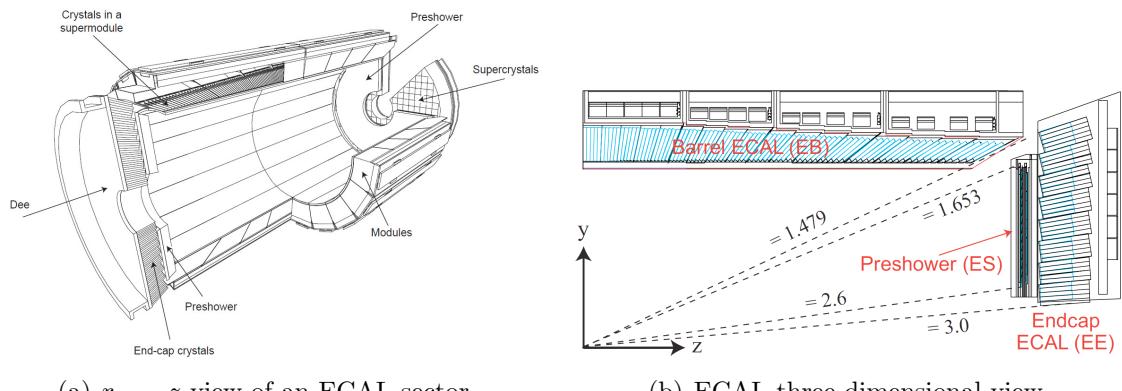
a number ranging from 4 to 7 modules are mounted in a ring shape, depending on the position along  $z$ . Also in this case the modules are alternately mounted on the internal and on the external side of the disk, with the strips radially oriented. On the two innermost rings and on the fifth one the stereo modules are installed to measure the  $z$  coordinate. The thickness of the sensors range between 300 and 500  $\mu\text{m}$  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)<sup>3</sup>.

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  $\mu\text{m}$ . 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  $J/\Psi$ ,  $\Upsilon$  or  $Z$  decays.

### 2.2.3 The electromagnetic calorimeter (ECAL)

The main function of an electromagnetic calorimeter is to identify electrons and photons and to measure accurately their energy. The CMS electromagnetic calorimeter (ECAL) [15, 16], shown in Fig. 2.9, is a hermetic homogeneous calorimeter with cylindrical geometry, composed of many scintillating crystals of lead tungstate ( $\text{PbWO}_4$ ) with a truncated pyramidal shape. As the other detectors it consists of two parts, the ECAL barrel (EB), which contains 61200 crystals, and two endcaps (EE) containing 7324 crystals each one.



**Figure 2.9:** Schematic representation of the CMS electromagnetic calorimeter.

<sup>3</sup>The 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.

The characteristics of the PbWO<sub>4</sub> crystals make them an appropriate choice for operation at LHC. The high density ( $\rho = 8.3 \text{ g/cm}^3$ ), short radiation length ( $X_0 = 0.89 \text{ cm}$ ) and small Molière radius<sup>4</sup> (2.2 cm) allow to build a compact and high granularity calorimeter. Another advantage of this material is the radiation hardness and the fast scintillation decay time ( $\tau = 10 \text{ ns}$ ), that permits to collect about 80% of the produced light within the 25 ns interval between two consecutive bunch crossings. The main drawbacks of this material are the low light yield ( $\sim 10 \text{ photoelectrons/MeV}$ ) and the strong dependence on the operating temperature, that makes it necessary to keep the crystals at a stabilized temperature (18°C).

The crystals are grouped into  $5 \times 5$  matrices called *towers*. The barrel has an inner radius of 129 cm, a length of 630 cm and extends in the region  $|\eta| < 1.479$ . The crystals in the barrel have the following dimensions:  $22 \times 22 \text{ mm}^2$  at the front face,  $26 \times 26 \text{ cm}^2$  at the rear face, and a length of 23 cm, corresponding to  $25.8X_0$ , and are mounted in a quasi-projective geometry, in order to have the long side tilted by 3° with respect to the direction pointing to the interaction point, both in the  $\eta$  and  $\phi$  coordinates. This is done to avoid the empty spaces between adjacent crystals to be aligned with the direction pointing to the interaction point. The granularity of the EB is about 1°. Avalanche photodiodes (APDs) are used as photodetectors connected with the crystals in the barrel region.

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

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

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

The stochastic term  $a$  dominates at low energies: it includes the contribution of statistical fluctuations in the number of generated and collected photoelectrons. This term takes into account the crystal light emission, the light collection efficiency and the photodetector

---

<sup>4</sup>The 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  $R_M$ .

<sup>362</sup> quantum efficiency<sup>5</sup>. The noise term  $b$  includes the contributions of pile up events and  
<sup>363</sup> electronic noise, both due to the photodetector and preamplifier. These contributions  
<sup>364</sup> depend on  $\eta$  and on the LHC operational luminosity. The constant term  $c$ , dominant at  
<sup>365</sup> high energies, takes into account several contributions. The most relevant are the non-  
<sup>366</sup> uniformity of the longitudinal light collection, the intercalibration errors and the leakage  
<sup>367</sup> of energy from the rear side of the crystal. The ECAL barrel resolution for electrons was  
<sup>368</sup> measured using test beams to be:

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

<sup>369</sup> where  $E$  is the energy measured in GeV.

### <sup>370</sup> 2.2.4 The hadron calorimeter (HCAL)

<sup>371</sup> The hadron calorimeter (HCAL) [17] is used together with ECAL to make a complete  
<sup>372</sup> calorimetric system for the jet energy and direction measurement. Moreover, thanks  
<sup>373</sup> to its hermetic structure, it can measure the energy imbalance in the transverse plane,  
<sup>374</sup>  $E_T^{\text{miss}}$ , a typical signature of non interacting particles, such as neutrinos. The HCAL is  
<sup>375</sup> a sampling calorimeter covering the region  $|\eta| < 5$ . As shown in Fig. 2.10, it is divided  
<sup>376</sup> in four sub-detectors: HB (*Barrel Hadronic Calorimeter*), located in the barrel region  
<sup>377</sup> inside the solenoid, extending up to  $|\eta| < 1.4$ ; HE (*Endcap Hadronic Calorimeter*), placed  
<sup>378</sup> in the endcaps region inside the magnet, covering the region  $1.3 < |\eta| < 3$  and partially  
<sup>379</sup> overlapping with the HB coverage; HO (*Outer Hadronic Calorimeter*), also known as  
<sup>380</sup> *tail-catcher*, placed along the inner wall of the magnetic field return yoke, just outside of  
<sup>381</sup> the magnet; HF (*Forward Hadronic Calorimeter*), a sampling calorimeter consisting of  
<sup>382</sup> quartz fibers sandwiched between iron absorbers, consisting of two units placed in the very  
<sup>383</sup> forward region ( $3 < |\eta| < 5$ ) outside the magnetic coil. The quartz fibers emit Cherenkov  
<sup>384</sup> light with the passage of charged particles and this light is detected by radiation resistant  
<sup>385</sup> photomultipliers. In order to maximize particle containment for a precise missing transverse  
<sup>386</sup> energy measurement, the amount of absorber material was maximized, reducing therefore  
<sup>387</sup> the amount of the active material. Since HCAL is mostly placed inside the magnetic coil,  
<sup>388</sup> a non-magnetic material like brass was chosen as absorber. HB and HE are therefore  
<sup>389</sup> made with 5 cm brass absorber layers interleaved with 3.7 mm plastic scintillators. The  
<sup>390</sup> scintillation light is collected by wavelength shifting (WLS) fibres and read out by hybrid  
<sup>391</sup> photodiodes (HPD). The granularity of the calorimeter is  $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$  for  
<sup>392</sup>  $|\eta| < 1.6$  and  $\Delta\eta \times \Delta\phi \approx 0.17 \times 0.17$  for  $|\eta| \geq 1.6$ . HO is made of 5 rings installed in  
<sup>393</sup> the wheel that compose the return yoke and is divided in 12 sectors, each one covering  
<sup>394</sup> a  $30^\circ$  angle in  $\phi$ . It consists of scintillating layers, with the same granularity as HB, and  
<sup>395</sup> the solenoid coil is used as an additional absorber to increase the effective depth of the

---

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



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

396 calorimeter in the barrel region, which is extended up to  $11.8 \lambda_I$ , thus improving the energy  
 397 resolution.

398 The energy resolution in the different regions of HCAL can be parametrized using a  
 399 stochastic and a constant term, as follows:

$$\begin{aligned} \left(\frac{\sigma_E}{E}\right)^2 &= \left(\frac{90\% \text{GeV}^{1/2}}{\sqrt{E}}\right)^2 + (4.5\%)^2 && \text{in the barrel/endcap ,} \\ \left(\frac{\sigma_E}{E}\right)^2 &= \left(\frac{172\% \text{GeV}^{1/2}}{\sqrt{E}}\right)^2 + (9\%)^2 && \text{in the HF ,} \end{aligned} \quad (2.8)$$

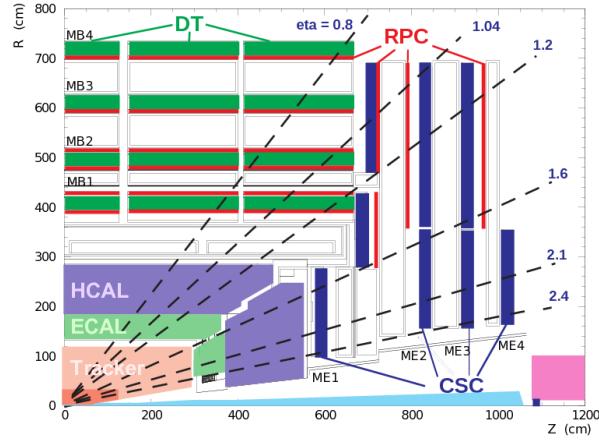
400 where  $E$  is expressed in GeV.

## 401 2.2.5 The muon system

402 The CMS muon system [18] is dedicated to the identification and measure of high  $p_T$  muons,  
 403 in combination with the tracker. The system is placed outside the magnetic coil, embedded  
 404 in the return yoke, to fully exploit the 1.8 T return flux. As shown in Fig. 2.11, the system  
 405 consists of three types of independent gaseous particle detectors:

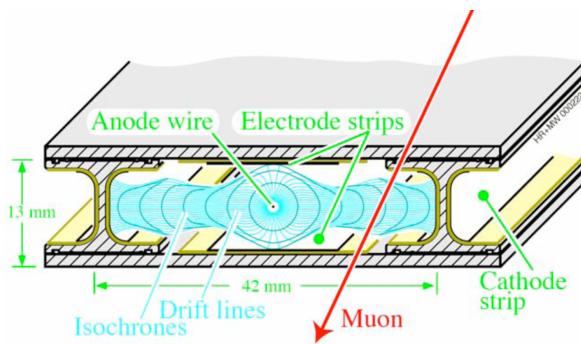
- 406 • *Drift Tubes* (DT) are placed in the barrel region, where the occupancy is relatively  
 407 low ( $< 10 \text{ Hz/m}^2$ );
- 408 • *Cathode Strip Chambers* (CSC) are installed in the endcaps, where the occupancy is  
 409 higher ( $> 100 \text{ Hz/m}^2$ );
- 410 • *Resistive Plate Chambers* (RPC) are placed both in the barrel and endcaps.

411 The DT system is placed in the region of the barrel with  $|\eta| < 1.2$ , where the magnetic  
 412 field is sufficiently weak and homogeneous. Along the longitudinal direction, the barrel



**Figure 2.11:** Schematic view of a quadrant of the CMS muon system.

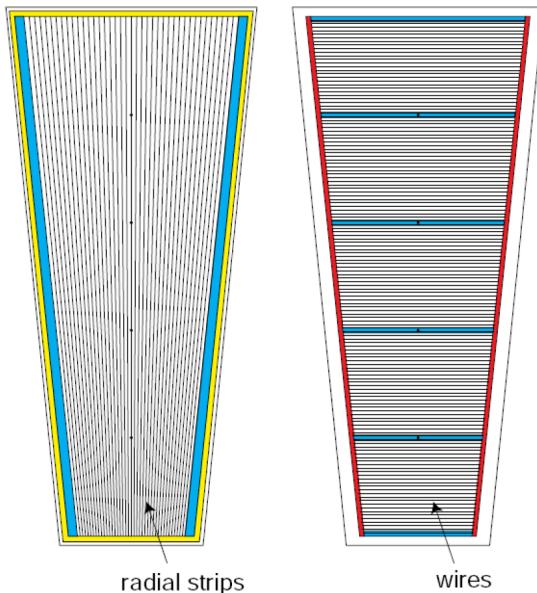
region is divided in 5 wheels, which are subdivided in 12 sectors covering a  $30^\circ$  azimuthal angle each. The wheels are composed of 4 concentric rings of chambers, called *stations*, interspersed in the layers of the iron return yoke, and each one formed by 12 DT chambers. The basic element of the DT system is a rectangular drift tube cell with a transverse size of  $13 \times 42 \text{ mm}^2$  and a variable length from 2 to 4 m. The chambers are filled with a gas mixture of Ar (85%) and CO<sub>2</sub> (15%) and are grouped in the radial direction to form detection layers. Groups of four layers form a *superlayer*. In each superlayer two chambers have anode wires parallel to the beam axis and two have perpendicular wires, thus providing two measurements of the  $(r, \phi)$  coordinate and two measurements of the  $z$  coordinate of the track hit positions. As shown in Fig. 2.12, each chamber is made of a stainless steel anode



**Figure 2.12:** Schematic representation of a drift tube chamber, showing the drift lines in presence of magnetic field.

wire between two parallel aluminium plates with “T” shaped spacer cathodes, isolated from the aluminium plates with polycarbonate plastic, and the hit resolution is about  $100 \mu\text{m}$  in both  $(r, \phi)$  and  $(r, z)$  directions.

In the endcaps, the high and non-uniform magnetic field and the particle rate do not allow to use drift tubes detectors to perform measurements. Therefore, a solution based on the CSC detector has been adopted. CSC are multi-wire proportional chambers with the cathodes segmented into strips oriented radially and transversely with respect to the anode wires (see Fig. 2.13), allowing a simultaneous measurement of two coordinates ( $r$  through the wires and  $\phi$  using the strips). The CSC chambers are filled with a gas mixture of Ar (40%), CO<sub>2</sub> (50%) and CF<sub>4</sub> (10%) and provide a spatial resolution of about 80–85  $\mu\text{m}$ . The drift path of the charge carriers is shorter with respect to the drift tubes, therefore these detectors can be placed in regions with higher flows of charged particles and less homogeneous magnetic fields. The CSC coverage is  $0.8 < |\eta| < 2.4$ .



**Figure 2.13:** Schematic representation of CSC cathode (left) and anode (right) panels.

RPCs are used both in barrel and endcaps, complementing DT and CSC systems, in order to ensure robustness and redundancy to the muon spectrometer. RPCs are gaseous detectors characterized by a coarse spatial resolution, but are able to perform precise time measurements, comparable with the ones provided by scintillators. These chambers are made of 4 bakelite planes, with a bulk resistivity of  $10^{10}$ – $10^{11} \Omega\text{cm}$ . The 2 mm gap between the plates is filled with a mixture of C<sub>2</sub>H<sub>2</sub>F<sub>4</sub> (94.5%) and Isobutane. The central part of the chamber is equipped with insulated aluminum strips, used to collect the signal generated by crossing particles. In the barrel the strips are rectangularly segmented and run along the beam axis, whereas the endcaps are equipped with trapezoidal shaped strips. The detector operates in avalanche mode, and covers the region  $|\eta| < 2.1$ .

## <sup>446</sup> 2.3 The CMS trigger system

<sup>447</sup> The LHC can provide proton-proton interactions at a crossing frequency of 40 MHz and,  
<sup>448</sup> for each bunch crossing, several collisions can occur (approximately 20 at the nominal  
<sup>449</sup> instantaneous luminosity). Since it is impossible to store and process the large amount  
<sup>450</sup> of data associated with the resulting large number of events, a drastic rate reduction has  
<sup>451</sup> to be achieved. In fact the speed at which data can be written to mass storage is limited  
<sup>452</sup> and, moreover, the vast majority of events produced is not interesting for physics analyses,  
<sup>453</sup> because it involves low transverse momentum interactions (also called *minimum bias events*).  
<sup>454</sup> The task of reducing this rate is accomplished by the CMS trigger system, which is the start  
<sup>455</sup> of the physics event selection. CMS makes use of a two-stage trigger system, consisting of a  
<sup>456</sup> *Level-1* trigger (L1) [19] and a *High Level Trigger* (HLT) [20].

<sup>457</sup> Level-1 trigger runs on dedicated processors, and accesses coarse level granularity  
<sup>458</sup> information from calorimetry and muon system. A L1 Trigger decision has to be taken for  
<sup>459</sup> each bunch crossing within  $3.2\ \mu\text{s}$ . Its task is to reduce the data flow from 40 MHz to about  
<sup>460</sup> 100 kHz.

<sup>461</sup> The High Level Trigger is responsible for reducing the L1 output rate down to a maximum  
<sup>462</sup> rate of the order of 1 kHz. The HLT code runs on a farm of commercial processors and can  
<sup>463</sup> access the full granularity information of all the sub-detectors.

<sup>464</sup> The main characteristics of the CMS trigger system are described in the following.

### <sup>465</sup> 2.3.1 The Level-1 trigger

<sup>466</sup> The L1 trigger is responsible for the identification of electrons, muons, photons, jets and  
<sup>467</sup> missing transverse energy. It is required to have a high and carefully understood efficiency.  
<sup>468</sup> Its output rate and speed are limited by the readout electronics and by the performances  
<sup>469</sup> of the data acquisition (DAQ) system [20]. It consists of three main subsystems:

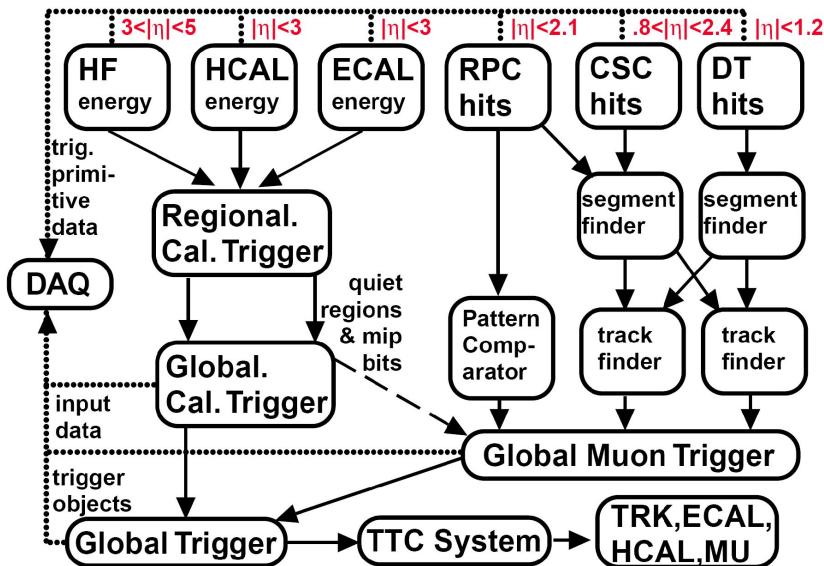
- <sup>470</sup> • L1 Calorimeter Trigger;
- <sup>471</sup> • L1 Muon Trigger;
- <sup>472</sup> • L1 Global Trigger.

<sup>473</sup> The L1 Global Trigger is responsible for combining the output of L1 Calorimeter Trigger  
<sup>474</sup> and L1 Muon Trigger and for making the decision to either retain the event or discard it.  
<sup>475</sup> The organization of CMS L1 Trigger is schematically summarized in Fig. 2.14.

#### <sup>476</sup> L1 Calorimeter Trigger

<sup>477</sup> Controllare se stato cambiato qualcosa nel Run2

<sup>478</sup> The input for the L1 Calorimeter Trigger are calorimeter towers, which are clusters  
<sup>479</sup> of signals collected both from ECAL and HCAL. Towers are calculated by calorimeter  
<sup>480</sup> high level readout circuits, called Trigger Primitive Generators. The Regional Calorimeter  
<sup>481</sup> Trigger identifies electron, photon,  $\tau$  and jet candidates together with their transverse energy  
<sup>482</sup> and sends the information to the Global Calorimeter Trigger. The Global Calorimeter



**Figure 2.14:** Schematic representation of the Level-1 trigger components.

483 Trigger sorts the candidates according to their transverse energy and sends the first four  
 484 objects to the L1 Global Trigger.

## 485 L1 Muon Trigger

### 486 Controllare se stato cambiato qualcosa nel Run2

487 The L1 Muon Trigger is actually a composite system itself: information from RPC, CSC  
 488 and DT specific triggers are combined in the so called L1 Global Muon Trigger.

489 The RPC trigger electronics builds Track Segments, gives an estimate of their  $p_T$  and  
 490 sends these segments to the Global Muon Trigger. It also provides the CSC logic unit with  
 491 information to solve hit position ambiguities in case of two or more muon tracks crossing  
 492 the same CSC chamber.

493 The CSC trigger builds Local Charged Tracks (LCT), that is track segments made out  
 494 of the cathode strips only, and assign a  $p_T$  value and a quality flag to the LCTs. The best  
 495 three LCTs in each sector of nine CSC chambers are passed to the CSC Track Finder, that  
 496 uses the full CSC information to build tracks, assigns them a  $p_T$  and a quality flag and  
 497 sends them to the Global Muon Trigger.

498 DTs are equipped with Track Identifier electronics, which is able to find groups of  
 499 aligned hits in the four chambers of a superlayer. Those Track Segments are sent to the  
 500 DT Track Correlator that tries to combine segments from two superlayers, measuring the  
 501  $\phi$  angle. The best two segments are sent to the DT Track Finder that builds tracks and  
 502 sends them to the Global Muon Trigger.

503        The Global Muon Trigger sorts the RPC, CSC and DT muon tracks and tries to combine  
504        them. The final set of muons is sorted according to the quality, and the best four tracks  
505        are passed to the L1 Global Trigger.

506        **L1 Global Trigger**

507        **Controllare se stato cambiato qualcosa nel Run2**

508        The L1 Global Trigger is responsible for collecting objects created from the Calorimeter  
509        and Muon Triggers and for making a decision whether to retain the event or not. In case  
510        the event is accepted, the decision is sent to the Timing Trigger and Control System, that  
511        commands the readout of the remaining subsystems.

512        In order to take the decision, the L1 Global Trigger sorts the ranked objects produced  
513        by calorimetry and muon system and checks if at least one of the thresholds in the L1  
514        trigger table is passed.

515        **2.3.2 The high level trigger (HLT)**

516        The High Level Trigger is designed to reduce the L1 output rate down to about 1000 events/s,  
517        which is the amount that will be written to mass storage. HLT code runs on commercial  
518        processors and performs reconstruction using the information from all sub-detectors. Events  
519        passing the HLT are stored on local disk or in CMS Tier 0<sup>6</sup>.

520        Data read from sub-detectors are assembled by a builder unit and then assigned to  
521        a switching network that dispatches events to the processor farm. The CMS switching  
522        network has a bandwidth of 1 Tbit/s. This simple design ensures maximum flexibility to  
523        the system, the only limitation being the total bandwidth and the number of processors.  
524        The system can be easily upgraded adding new processors or replacing the existing ones  
525        with faster ones as they become available. Since the algorithms have a fully software  
526        implementation, improvements to the algorithms can be easily implemented and do not  
527        require any hardware intervention.

528        Event by event, the HLT code is run on a single processor, and the time available to  
529        make a decision is about 300 ms. The real time nature of this selection imposes several  
530        constraints on the resources an algorithm can use. The reliability of HLT algorithms  
531        is of capital importance, because events not selected by the HLT are lost. In order to  
532        efficiently process events, the HLT code has to be able to quickly reject not interesting  
533        events; computationally expensive algorithms must be run only on good candidates for  
534        interesting events. In order to cope with this requirement the HLT code is organized in a  
535        virtually layered structure:

---

<sup>6</sup>The Worldwide LHC Computing Grid (WLCG) is composed of four levels, or “Tiers”, identified with numbers 0, 1, 2 and 3. Each Tier is made up of several computer centres and provides a specific set of services; they process, store and analyse all the data from the Large Hadron Collider (LHC). Tier 0 is the CERN Data Centre. All of the data from the LHC pass through this central hub. Tier 0 distributes the raw data and the reconstructed output to Tier 1’s, and reprocesses data when the LHC is not running.

- 536 • Level 2: uses only complete muon and calorimetry information;
- 537 • Level 2.5: uses also the pixel information;
- 538 • Level 3: makes use of the full information from all the tracking detectors.

539 Each step reduces the number of events to be processed in the following step. The most  
 540 computationally expensive tasks are executed in the Level 3; time consuming algorithms  
 541 such as track reconstruction are only executed in the region of interest. Besides, since the  
 542 ultimate precision is not required at a HLT level, track reconstruction is performed on a  
 543 limited set of hits, and is stopped once the required resolution is achieved.

## 544 2.4 Event reconstruction and particle identification

545 In CMS, the physics object reconstruction and identification is based on standard algorithms  
 546 developed by the collaboration and used by all the physics analyses. In this section, the  
 547 techniques used for the reconstruction and identification of the physics objects of interest  
 548 for  $H \rightarrow WW \rightarrow 2\ell 2\nu$  analyses are described.

### 549 2.4.1 The Particle Flow technique

550 The Particle Flow (PF) event reconstruction technique [21] aims at the reconstruction  
 551 and identification of all the stable particles in the event, i.e. electrons, muons, photons,  
 552 charged and neutral hadrons, with a thorough combination of the information from all CMS  
 553 sub-detectors, in order to determine their energy, direction and type. These individual  
 554 particles are then used, for example, to build jets, to measure the missing transverse energy  
 555  $E_T^{\text{miss}}$ , to reconstruct the  $\tau$  from their decay products, to quantify the charged lepton  
 556 isolation and to tag b-jets.

557 The CMS detector is well suited for this purpose. Indeed, the presence of a large internal  
 558 silicon tracker immersed in an intense solenoidal magnetic field allows the reconstruction of  
 559 charged particles with high efficiency and small fake rate, and provides a high precision  
 560 measurement of the particle  $p_T$  down to about 150 MeV, for  $|\eta| \leq 2.6$ . The high granularity  
 561 of the ECAL calorimeter is the additional key element for the feasibility of the PF technique,  
 562 allowing the reconstruction of photons and electrons with high energy resolution.

563 The first step of the PF technique consists in the reconstruction of the basic elements  
 564 from the various sub-detectors, such as charged-particle tracks, calorimeter clusters and  
 565 muon tracks. These elements, which are provided by the sub-detectors with high efficiency  
 566 and low fake rate, are then connected together with a link algorithm.

567 The good performance of the tracking system are achieved by means of an iterative  
 568 tracking strategy [22], based on the Kalman Filter algorithm [23]. The basic idea of iterative  
 569 tracking is that initial iterations search for tracks that are easiest to find, e.g. high  $p_T$  tracks  
 570 produced near the interaction region. After each iteration, hits associated to reconstructed  
 571 tracks are removed from the hit collection, thereby reducing the combinatorial complexity  
 572 and simplifying the subsequent iterations, which aim at finding more complicated set of

573 tracks, e.g. low  $p_T$  or displaced tracks. The *Iteration 0*, where the majority of tracks are  
 574 reconstructed, is designed to identify prompt tracks with  $p_T > 0.8$  GeV that have three  
 575 hits in the three layers of the pixel detector. *Iteration 1* is used to recover prompt tracks  
 576 that have only two pixel hits. *Iteration 2* aims at finding low- $p_T$  prompt tracks while  
 577 *Iterations 3–5* are intended to find tracks that originate outside the collision point, i.e.  
 578 tracks produced by a secondary vertex, and to recover undetected tracks in the previous  
 579 iterations. Each iteration proceeds according to four steps:

- 580 • *seeding*: initial track candidates are obtained using 2 or 3 hits in the innermost layers  
 581 (these proto-tracks are called seeds);
- 582 • *pattern recognition*: this step is based on Kalman Filter and searches for hits in the  
 583 outer layers that could be associated to the initial track candidate, reconstructing the  
 584 particle trajectory;
- 585 • *track fitting*: in this step a fit of the trajectory is performed, using its associated hits  
 586 and providing an estimate of the track parameters ( $p_T$ ,  $\eta$ ,  $\phi$ , charge, etc.);
- 587 • *selection*: finally tracks are selected based on quality requirements.

588 The high detection efficiency of the calorimeters is based on a specific calorimeter  
 589 clustering algorithm, which is performed separately in each sub-detector. The algorithm is  
 590 based on three steps: in the first step, “cluster seeds” are identified as local calorimeter cells  
 591 with an energy deposit above a given threshold. Then, “topological clusters” are grown  
 592 from the seeds by gathering cells with at least one side in common with a cell already in the  
 593 cluster, and with an energy above a given threshold. A topological cluster usually gives rise  
 594 to many “particle flow clusters” as seeds, which are identified sharing the energy of each  
 595 cell among the particle flow clusters, thereby allowing the determination of the particle flow  
 596 cluster energy and position.

597 These elements are then connected to each other using a link algorithm, which identifies  
 598 blocks of elements that are topologically compatible. For example, a charged-particle track  
 599 is linked to a calorimeter particle flow cluster if the extrapolated position from the track to  
 600 the calorimeter is compatible with the cluster boundaries. From these blocks, PF candidates  
 601 are identified according to the following order:

- 602 • Muons: a *global muon* gives rise to a *PF muon* if its combined  $p_T$  measurement is  
 603 compatible within 3 standard deviation with the one provided by the sole tracker. The  
 604 corresponding track is removed from the block;
- 605 • Electrons: electrons tend to give rise to short tracks, and to lose energy by Bremsstrahlung  
 606 in the tracker layers on their way to the calorimeter. The link between a charged-  
 607 particle track (refitted with the Gaussian-Sum Filter (GSF) [**Adam:815410**]) and  
 608 one or more ECAL clusters identifies a *PF electron*. After the identification, the  
 609 corresponding tracks and clusters are removed from the block.
- 610 • Charged hadrons: the remaining tracks give rise to *PF charged hadrons*. Tracks can be  
 611 linked to ECAL and HCAL clusters, and the energy is determined taking into account  
 612 information from calorimeters;

- Photons and neutral hadrons: ECAL clusters not linked with tracks give rise to *PF photons*, while the remaining HCAL clusters are identified as *PF neutral hadrons*.
- After the identification of all PF candidates in the event, *PF jets* are clustered as described in Sec. ???. The last step is the reconstruction of the *PF  $\vec{p}_T^{\text{miss}}$* , which is given by:

$$\vec{p}_T^{\text{miss}} = - \sum_{\text{PF obj}} \vec{p}_T^{\text{PF obj}} , \quad (2.9)$$

where the sum extends over all the PF objects. The  $E_T^{\text{miss}}$  is defined as the modulus of  $\vec{p}_T^{\text{miss}}$ .

## 2.4.2 Leptons reconstruction and identification

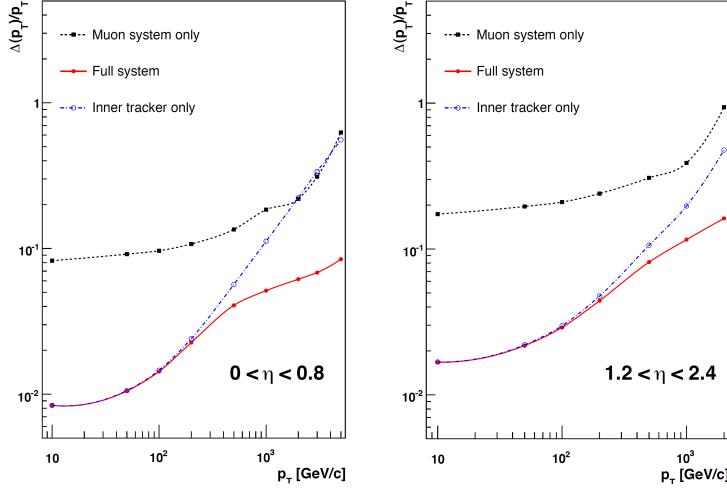
### Muon reconstruction and identification

Muons produced at the collision point can go through the entire detector with a negligible energy loss, thus reaching the detector outermost part where the muon chambers are installed (see Sec. 2.2.5). Muons interact through ionization with the layers of the silicon tracker, which is able to reconstruct their tracks (*tracker track*). The muon tracks are also reconstructed using the muon system (*standalone muon track*). Based on these objects, two reconstruction approaches are used [Chatrchyan:2012xi]: in the first method (outside-in), for each standalone muon tracks a tracker track is searched for by extrapolating the two tracks to a common surface. If a match is found, the hits associated to the two tracks are fitted together giving rise to a *Global Muon*. The second approach (inside-out) consists in considering all tracker tracks with  $p_T > 0.5$  GeV as potential muon candidates and are extrapolated to the muon system taking into account the magnetic field, the expected energy losses and the multiple scattering in the detector material. If at least one muon segment (a short track stub made of DT or CSC hits) matches the extrapolated tracks, the corresponding tracker track is identified as a *Tracker Muon*.

The matching with the muon system improves significantly the muon  $p_T$  resolution that can be obtained from the tracker only, especially in the region with  $p_T > 200$  GeV, as shown in Fig. ??.

Depending on the physics analysis, different muon definitions can be used by changing the selection on the muon identification variables, hence balancing between the muon identification efficiency and purity. The most widely used definition in physics analyses is the so-called *Tight muon selection*<sup>7</sup>. This selection requires the muon candidate to be reconstructed as a Global Muon and identified by the PF algorithm. The fit of the global track, which is required to include muon segments in at least two muon stations (this implies that the muon is also reconstructed as a Tracker Muon), must have a  $\chi^2/d.o.f.$  less than 10 and use more than 10 inner tracker hits. The transverse impact parameter with respect to the primary vertex is required to be  $|d_{xy}| < 2$  mm, significantly reducing the rate

<sup>7</sup>Small variations with respect to this baseline definition are adopted by the specific analyses.



**Figure 2.15:** Muon  $p_T$  resolution as a function of the muon  $p_T$  in the barrel (left) and in the endcap (right) regions. The resolution is provided for the measurement using the tracking system or the muon system only, as well as for the combination of the two methods.

of muons from decays in flight, i.e. non prompt muons. The requirements defining the Tight Muon identification are summarized in Table 2.2.

**Table 2.2:** Summary of the muon identification variables and the corresponding selections commonly used by physics analyses.

Observable	Cut
Is Global Muon	true
Is PF muon	true
Tracker layers with valid hits	> 5
Number of valid pixel hits	> 0
Number of valid muon hits	> 0
Number of matched muon stations	> 1
$\chi^2/d.o.f.$	< 10
$d_{xy}(PV)$	< 0.2 cm
$d_z(PV)$	< 0.5 cm

Another selection which is optimised for low- $p_{\text{T}}$  muons coming from in flight decays is called *Soft Muon selection*. This selection requires the muon to be reconstructed as a Tracker Muon with loose additional cuts on the transverse and longitudinal impact parameters. This selection is commonly used to identify muons coming from B hadron decays.

### 653 Muon isolation

One of the most powerful requirements to select prompt muons, as the ones produced from W or Z boson decays, and to reject muons produced by decays in flight, is the isolation. Indeed, prompt muons are expected to be isolated in the event, differently to non prompt muons that are generally produced within jets and characterized by many nearby particles.

Muons commonly used to reconstruct the W or Z decays are thus required to pass an isolation requirement, which includes a pile up mitigation correction called “ $\Delta\beta$  correction”. This correction is needed to obtain a robust isolation definition that is less sensitive to the pile up contribution. Indeed, simultaneous interactions manifest themselves as a mean energy deposited over all the detector acceptance, which is not due to the particles produced in the primary events, thus spoiling the isolation measurement. The relative isolation variable, usually called *PF relative isolation*, is defined as follows:

$$I_{\Delta\beta}^{\text{rel}} = \left[ \sum_{ChH} p_{\text{T}} + \max \left( 0, \sum_{NH} p_{\text{T}} + \sum_{Ph} p_{\text{T}} - 0.5 \sum_{ChHPU} p_{\text{T}} \right) \right] / p_{\text{T}}^{\mu\text{on}} . \quad (2.10)$$

The sums in Eq. (2.10) are performed in a cone of radius  $\Delta R < 0.4$  around the muon direction. The *ChH* subscript refers to charged hadrons, *NH* to neutral hadrons, *Ph* to photons and *ChHPU* to charged hadrons not arising from the primary vertex.

The cut applied on the isolation variable is analysis dependent, but a common value is  $I_{\Delta\beta}^{\text{rel}} < 0.15$ .

A different isolation definition is called *Tracker relative isolation*,  $I_{\text{trk}}^{\text{rel}}$ , which is calculated as the scalar sum of all the  $p_{\text{T}}$  of the tracker tracks reconstructed inside a cone of radius  $\Delta R < 0.3$  centred on the muon track direction.

### 673 Muon momentum scale and resolution

The measurement of the muon  $p_{\text{T}}$  is sensitive to the alignment of the tracker and the muon chambers, to the material composition and distribution inside the detector and to the knowledge of the magnetic field produced by the solenoid. The imperfect knowledge of the magnetic field and the effect of the material distribution introduce a relative bias in the muon  $p_{\text{T}}$  that is generally independent on the  $p_{\text{T}}$  itself, while the effect of the alignment is known to produce a bias that increases linearly with the  $p_{\text{T}}$ .

Different methods are used to estimate the muon  $p_{\text{T}}$  scale and resolution effects and to determine the corresponding uncertainties, depending on the  $p_{\text{T}}$  range. At low and intermediate  $p_{\text{T}}$  ( $< 100$  GeV), the di-muon events arising from the  $J/\Psi$  and Z resonance

683 decays are used to correct the  $p_T$  scale and to measure the  $p_T$  resolution. In the high  $p_T$   
 684 regime, the muon  $p_T$  scale and resolution are instead measured using cosmic ray muons.  
 685 One of the methods that is commonly used in the intermediate  $p_T$  range is the *MuScleFit*  
 686 (Muon momentum Scale calibration Fit), which provides the muon  $p_T$  scale corrections  
 687 by fitting the Z boson mass peak in data and simulation. These corrections are meant to  
 688 recover the bias of the Z mass peak with respect to the  $\eta$  and  $\phi$  coordinates of the muon.  
 689 After applying these corrections the relative  $p_T$  resolution,  $\sigma(p_T)/p_T$ , is measured as a  
 690 function of  $\eta$  and  $\phi$  and is found to be on average of the order of 2% in the barrel and up  
 691 to 6% in the endcaps, for muon  $p_T$  below 100 GeV.

## 692 Electron reconstruction and identification

693 The electron reconstruction is based on the combination of tracker and ECAL information.  
 694 The reconstruction technique starts by measuring the energy deposits in ECAL by electrons,  
 695 which form a “supercluster”. A supercluster is a group of one or more ECAL clusters  
 696 associated using an algorithm that takes into account the characteristic shape of the energy  
 697 deposited by electrons emitting Bremsstrahlung radiation in the tracker material. The  
 698 supercluster shape is characterized by a narrow width profile in the  $\eta$  coordinate spread over  
 699 the  $\phi$  direction. The superclusters are matched to tracks, reconstructed in the tracker with  
 700 the GSF algorithm, in order to obtain an electron candidate. An additional reconstruction  
 701 method, described in details in Ref. [CMS-PAS-EGM-10-004 ], is instead seeded by  
 702 electron tracks reconstructed in the inner tracker layers.

703 Several strategies are used in CMS to identify prompt isolated electrons (characteristic  
 704 of the signal processes of interest), and to separate them from background sources, mainly  
 705 originating from photon conversions, jets misidentified as electrons, or electrons from  
 706 semileptonic decays of b and c quarks. In order to achieve a good discrimination, several  
 707 identification variables are used:

- 708 •  $\Delta\eta_{\text{trk},\text{SC}}$  and  $\Delta\phi_{\text{trk},\text{SC}}$ : the variables measuring the spatial matching between the track  
 709 and the supercluster in the  $\eta$  and  $\phi$  coordinates, respectively;
- 710 •  $\sigma_{in,in}$ : a variable related to the calorimeter shower shape, measuring the width of the  
 711 ECAL supercluster along the  $\eta$  direction computed for all the crystals in the  $5 \times 5$   
 712 block of crystals centred on the highest energy crystal of the seed supercluster;
- 713 •  $H/E$ : the ratio between the energy deposited in the HCAL tower behind the ECAL  
 714 seed and the supercluster seed energy;
- 715 •  $|1/E - 1/p|$ : the difference of the inverse of energy  $E$  measured in ECAL and the  
 716 inverse of momentum  $p$  measured in the tracker;
- 717 • the number of missing hits in the back-propagation of the track to the interaction  
 718 point;
- 719 •  $d_{xy}$  and  $d_z$ : the transverse and longitudinal impact parameters with respect to the  
 720 primary vertex.
- 721 • a photon conversion veto ( $\gamma \rightarrow e^+e^-$ ) based on the primary vertex measurement.

722 Different working points are provided by CMS corresponding to different selections  
723 on the previously defined variables. One of the common working points used by several  
724 physics analyses, as the  $H \rightarrow WW$  analyses described in Secs.??, is the “tight working  
725 point”, summarised in Table ??.

**Table 2.3:** Electron identification selections corresponding to the tight working point.

Variable	Selection	
	$ \eta_{\text{SC}}  \leq 1.479$	$1.479 <  \eta_{\text{SC}}  \leq 2.5$
$\sigma_{in,in}$	0.01	0.028
$ \Delta\eta_{\text{trk,SC}} $	0.009	0.007
$ \Delta\phi_{\text{trk,SC}} $	0.03	0.09
$H/E$	0.06	0.06
$ 1/E - 1/p $	0.012	0.010
$ d_{xy} $	0.011 cm	0.035 cm
$ d_z $	0.047 cm	0.42 cm
missing inner hits	$\leq 2$	$\leq 1$
conversion veto	yes	yes

## 726 Electron isolation

727 Selected electrons are required to pass an isolation requirement that includes a pile up  
728 mitigation correction based on the electron effective catchment area, which is different in  
729 different  $\eta$  ranges. The isolation variable is given by the following formula:

$$I_{EA \text{ corrected}}^{\text{rel}} = \left[ \sum_{ChH} p_T + \max \left( 0, \sum_{Ph} p_T + \sum_{NH} p_T - \rho E A \right) \right] / p_T^{\text{electron}} \quad (2.11)$$

730 where  $ChH$  refers to charged hadrons,  $Ph$  to photons,  $NH$  to neutral hadrons,  $\rho$  is the  
731 energy density due to pile up events,  $E$  is the energy and  $A$  is an effective area. The sums  
732 are performed inside a cone of radius  $\Delta R < 0.4$  around the electron direction. The cut  
733 applied on this variable for the tight working point is  $I_{EA \text{ corrected}}^{\text{rel}} < 0.04$ .

<sup>734</sup> Lepton efficiency

<sup>735</sup> **2.4.3 Jets reconstruction and identification**

<sup>736</sup> Jet b tagging

<sup>737</sup> **2.5 The CMS framework**



# Chapter 3

## <sup>738</sup> Higgs boson properties in the <sup>739</sup> $H \rightarrow WW$ decay channel

### <sup>740</sup> 3.1 Higgs boson phenomenology at LHC

<sup>741</sup> The discovery of a new boson consistent with the standard model (SM) Higgs boson has  
<sup>742</sup> been reported by ATLAS and CMS Collaborations in 2012. The discovery has been followed  
<sup>743</sup> by a comprehensive set of studies of properties of this new boson in several production and  
<sup>744</sup> decay channels and no evidence of deviation from the SM expectation has been found so  
<sup>745</sup> far.

#### <sup>746</sup> 3.1.1 Higgs boson production mechanisms

#### <sup>747</sup> 3.1.2 Higgs boson decay channels

### <sup>748</sup> 3.2 The $H \rightarrow WW$ decay channel

<sup>749</sup> The CMS studies in the  $H \rightarrow WW \rightarrow 2\ell 2\nu$  decay channel include the measurement of the  
<sup>750</sup> Higgs properties, as well as constraints on the Higgs total decay width and gauge bosons  
<sup>751</sup> anomalous couplings.

#### <sup>752</sup> 3.2.1 The $H \rightarrow WW \rightarrow 2\ell 2\nu$ signature

#### <sup>753</sup> 3.2.2 Physics objects definition

#### <sup>754</sup> 3.2.3 Background processes

### <sup>755</sup> 3.3 Experimental results of the Higgs boson <sup>756</sup> measurements



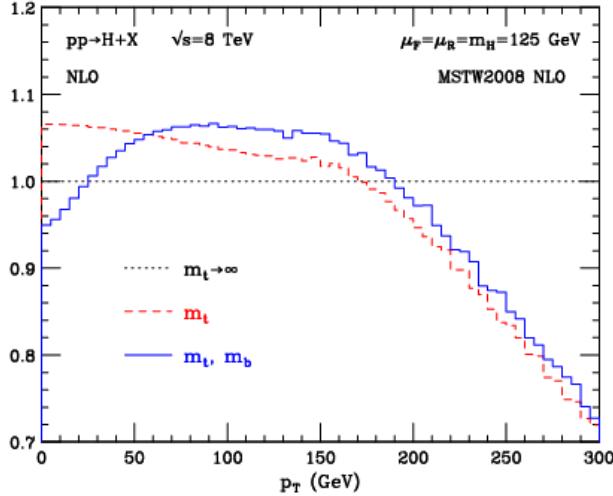
# Chapter 4

## 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  $\eta$ . The  $\eta$  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 [24, 25], and at NNLO using the so-called large- $m_t$  approximation [26–28], 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 [29].

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 bottom quark contribution in the fermionic loop can significantly modify the  $p_T^H$  shape [30], as shown in Fig. ???. Hence, a precise experimental measurement of the  $p_T^H$  spectrum is important to test the existing SM calculations.



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

Possible extensions of the SM predict a modification of the Higgs boson couplings to gluons and to the top quark. Many of these models actually predict the existence of new states that interact with the SM Higgs boson, but are beyond the direct production reach at the actual LHC energies. The effect of these new states could however show up as a deviation of the Higgs boson couplings with respect to the SM expectation. The modification of the couplings, as shown in Refs. [31, 32], can change the kinematics of the Higgs boson production and the effect can be particularly sizeable in the tail of the  $p_T^H$  distribution. Other models, such as Composite Higgs [33], predict the existence of top-partners, which are heavy resonances with the same quantum numbers as the top quark, that can interact with the Higgs boson in the ggH fermionic loop, changing the  $p_T^H$  shape with respect to what the SM predicts [34]. The measurement of the  $p_T^H$  spectrum is thus a useful tool for indirect searches of new particles predicted by theories beyond the SM.

Measurements of the fiducial cross sections and of several differential distributions, using the  $\sqrt{s} = 8$  TeV LHC data, have been reported by ATLAS [35–37] and CMS [38, 39] for the  $H \rightarrow ZZ \rightarrow 4\ell$  ( $\ell = e, \mu$ ) and  $H \rightarrow \gamma\gamma$  decay channels. In this chapter a measurement of the fiducial cross section times branching fraction ( $\sigma \times \mathcal{B}$ ) and  $p_T$  spectrum for Higgs boson production in  $H \rightarrow WW \rightarrow e^\pm \mu^\mp \nu\nu$  decays, based on  $\sqrt{s} = 8$  TeV LHC data, is reported.

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

Although the  $H \rightarrow WW \rightarrow 2\ell 2\nu$  channel has lower resolution in the  $p_T^H$  measurement compared to the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ \rightarrow 4\ell$  channels because of neutrinos in the final state, the channel has a significantly larger  $\sigma \times \mathcal{B}$ , exceeding those for  $H \rightarrow \gamma\gamma$  by a factor of 10 and  $H \rightarrow ZZ \rightarrow 4\ell$  by a factor of 85 for a Higgs boson mass of 125 GeV [40], and is

characterized by good signal sensitivity. Such sensitivity allowed the observation of a Higgs boson at the level of 4.3 (5.8 expected) standard deviations for a mass hypothesis of 125.6 GeV using the full LHC data set at 7 and 8 TeV [41].

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

The effect of the limited detector resolution, as well as the selection efficiency with respect to the fiducial phase space are corrected to particle level with an unfolding procedure [42], as explained in Sec. ??.

## 4.2 Data sets, triggers and MC samples

This analysis relies on the published  $H \rightarrow WW$  measurements [41] in terms of code, selections and background estimates for both the ggH and VBF production mechanisms.

### 4.2.1 Data sets and triggers

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

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

The trigger is not simulated in MC samples but the combined trigger efficiency is estimated from data and applied as a weight to all simulated events. The trigger efficiency for single and double lepton triggers is calculated using a Tag and Probe technique separately for muons and electrons, in bins of  $\eta$  and  $p_T$ . The Tag and Probe method uses a known mass resonance (e.g.  $J/\Psi$ , Z) to select particles of the desired type, and probe the efficiency of a particular selection criterion on these particles. In general the “tag” is an object that passes a set of very tight selection criteria designed to isolate the required particle type. Tags are often referred to as golden electrons or muons and the fake rate for passing tag selection criteria should be very small. A generic set of the desired particle type (i.e.

**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

with potentially very loose selection criteria) known as “probes” is selected by pairing these objects with tags such that the invariant mass of the combination is consistent with the mass of the resonance. Combinatorial backgrounds may be eliminated through any of a variety of background subtraction methods such as fitting, or sideband subtraction. The definition of the probe objects depend on the specifics of the selection criterion being examined. The simple expression to get the efficiency  $\epsilon$  as a function of  $p_T$  and  $\eta$  is given below:

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

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

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

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

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

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

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

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

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

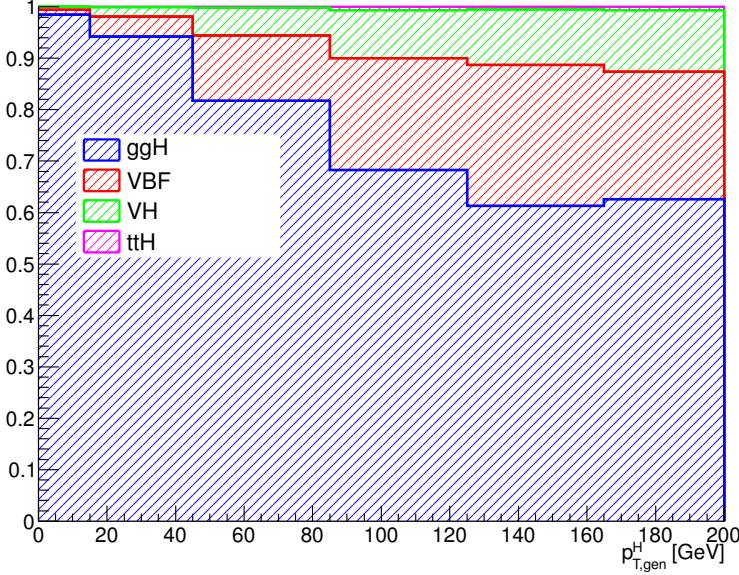
### 4.2.2 Monte Carlo samples

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

- The first version of the POWHEG program [43–47] (POWHEG V1) provides event samples for the  $H \rightarrow WW$  signal for the gluon fusion ( $ggH$ ) and VBF production mechanisms, as well as  $t\bar{t}$  and  $tW$  processes [48], with NLO accuracy.
- The  $qq \rightarrow W^+W^-$ , Drell-Yan, ZZ, WZ,  $W\gamma$ ,  $W\gamma^*$ , tri-bosons and  $W+jets$  processes are generated using the MADGRAPH 5.1.3 [49] event generator.
- The  $gg \rightarrow W^+W^-$  process is generated using the GG2WW 3.1 generator [50] and its cross section is scaled to the approximate NLO prediction [51, 52].
- The  $VH$  process is simulated using PYTHIA 6.426 [53].

For leading-order generators samples, the CTEQ6L [54] set of parton distribution functions (PDF) is used, while CT10 [55] is used for next-to-leading order (NLO) ones. Cross section calculations [56] 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 [40]. The reweighting of the  $p_T^H$  spectrum is achieved by tuning the POWHEG generator, as described in detail in Ref. [57]. Cross sections computed with NLO pQCD accuracy [40] 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.

For all processes, the detector response is simulated using a detailed description of the CMS detector, based on the GEANT4 package [58].

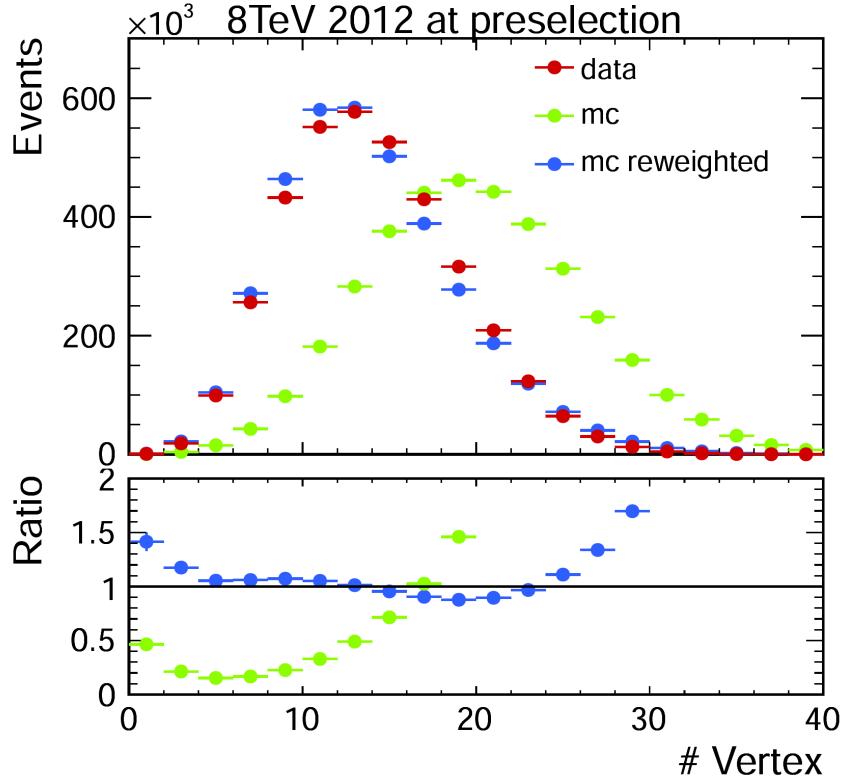


**Figure 4.2:** Relative fraction of ggH, VBF, VH and ttH in each bin of the Higgs boson transverse momentum.

900     Minimum bias events are superimposed on the simulated events to emulate the additional  
 901     pp interactions per bunch crossing (pileup). The number of pile-up events simulated in the  
 902     MC samples (in the same bunch crossing, in time, or in the previous or following one, out  
 903     of time pileup) have been generated poissonianly sampling from a distribution similar to  
 904     what is expected from data. For a given range of analyzed runs, the mean number of pileup  
 905     interactions per bunch crossing is estimated per luminosity block using the instantaneous  
 906     luminosity provided by the LHC, integrated over the entire run range and normalized. The  
 907     average number of pileup events per beam crossing in the 2011 data is about 10, and in the  
 908     2012 data it is about 20.

909     The simulated events are reweighted to correct for observed differences between data  
 910     and simulation in the number of pileup events, as shown in Fig. ??, trigger efficiency, and  
 911     lepton reconstruction and identification efficiencies [41].

912     For the comparison of the measured unfolded spectrum with the theoretical predictions,  
 913     two additional MC generators are used for simulating the SM Higgs boson production in  
 914     the ggH process: HRES 2.3 [30, 59] and the second version of the POWHEG generator  
 915     (POWHEG V2) [60]. HRES is a partonic level MC generator that computes the SM Higgs  
 916     boson cross section at NNLO accuracy in pQCD and performs the NNLL resummation  
 917     of soft-gluon effects at small  $p_T$ . The central predictions of HRES are obtained including  
 918     the exact top and bottom quark mass contribution to the gluon fusion loop, fixing the  
 919     renormalization and factorization scale central values at a Higgs boson mass of 125 GeV.  
 920     The cross section normalization is scaled, to take into account electroweak corrections, by a  
 921     factor of 1.05 and the effects of threshold resummation by a factor of 1.06 [61, 62]. The



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

upper and lower bounds of the uncertainties are obtained by scaling up and down both the renormalization and the factorization scales by a factor of two. The POWHEG V2 generator is a matrix element based generator that provides a NLO description of the ggH process in association with zero jets, taking into account the finite mass of the bottom and top quarks. The POWHEG prediction is tuned using the POWHEG damping factor *hdump* of 104.17 GeV, in order to match the  $p_T^H$  spectrum predicted by HRES in the full phase space. This factor reduces the emission of additional jets in the high  $p_T$  regime, and enhances the contribution from the Sudakov form factor in the limit of low  $p_T$ . The POWHEG generator is interfaced to the JHUGEN generator version 5.2.5 [63–65] for the decay of the Higgs boson to a W boson pair and interfaced with PYTHIA 8 [66] for the simulation of parton shower and hadronization effects.

### 4.3 Analysis Strategy

The analysis presented here is based on that used in the previously published  $H \rightarrow WW \rightarrow 2\ell 2\nu$  measurements by CMS [41], modified to be inclusive in the number of jets. This modifica-

tion significantly reduces the uncertainties related to the modelling of the number of jets produced in association with the Higgs boson.

### 4.3.1 Event reconstruction and selections

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

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

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

**This part would probably end up in the Object Reconstruction chapter**

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

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

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.

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

#### 1. Lepton preselection:

- at least two opposite-charge and opposite-flavour ( $e\mu$ ) isolated leptons reconstructed in the event;
- $|\eta| < 2.5$  for electrons and  $|\eta| < 2.4$  for muons;
- $p_T > 20 \text{ GeV}$  for the leading lepton. For the trailing lepton, the transverse momentum is required to be larger than 10 GeV.

#### 2. Extra lepton veto:

the event is required to have two and only two opposite-sign leptons passing the lepton selection.

#### 3. $E_T^{\text{miss}}$ preselection:

particle flow  $E_T^{\text{miss}}$  is required to be greater than 20 GeV.

#### 4. projected $E_T^{\text{miss}}$ selection:

minimum projected  $E_T^{\text{miss}}$  required to be larger than 20 GeV.

#### 5. Di-lepton mass cut:

$m_{\ell\ell} > 12 \text{ GeV}$  in order to reject low mass resonances and QCD backgrounds.

#### 6. Di-lepton $p_T$ cut:

$p_T^{\ell\ell} > 30 \text{ GeV}$ .

#### 7. Transverse mass:

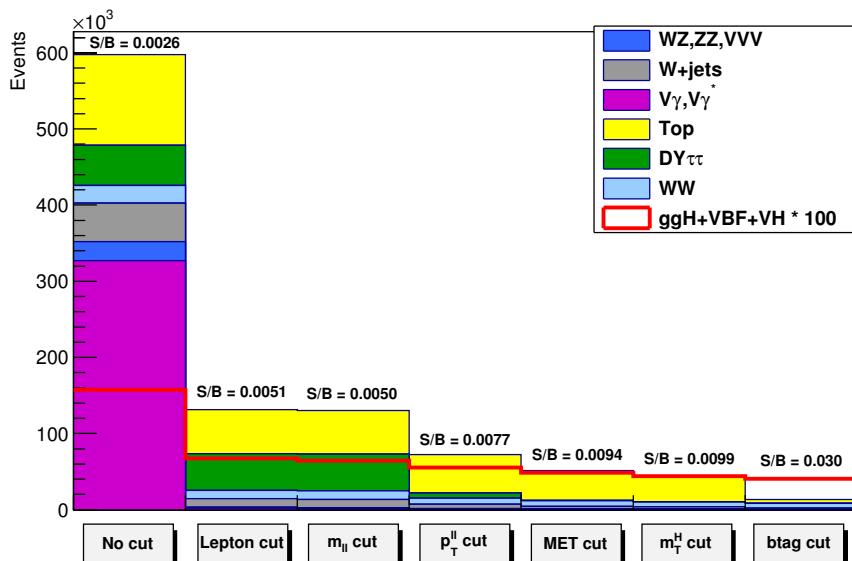
$m_T > 60 \text{ GeV}$  to reject Drell-Yan to  $\tau\tau$  events.

In addition to the WW-like preselection other cuts are applied in order to reduce the top quark background (both  $t\bar{t}$  and  $tW$ ), which is one of the main backgrounds in this final state. Two different selections are used depending on the number of jets with  $p_T > 30 \text{ GeV}$  in the event. This is done to suppress the top quark background both in the low  $p_T^H$  region, where 0-jets events have the largest contribution, and for higher  $p_T^H$  values where also larger jet multiplicity events are important. The selection for 0-jets events relies on the

soft-muon veto and on a soft jets (with  $p_T < 30$  GeV) anti b-tagging requirement. The latter requirement exploits the TCHE algorithm to reject soft jets that are likely to come from b quarks hadronization.

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

A cut-flow plot is reported in Fig. ??, showing the effect of each selection using signal and background simulations. In the first bin, labelled as “No cut”, no selection is applied and the bin content corresponds to the total expected number of events with a luminosity of  $19.4 \text{ fb}^{-1}$ . All the events in this bin have at least two leptons with a loose transverse momentum cut of 8 GeV. In the following bin the lepton cuts are applied, including the requirement to have two opposite-sign and opposite-flavour leptons and the extra lepton veto. Then all the other selections are progressively reported, showing the effect of each cut on the background and signal yields. For each selection the expected signal over background ratio is also shown, which, after the full selection requirements, reach a maximum value of about 3%.



**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  $19.46 \text{ fb}^{-1}$ .

### 1030 4.3.2 Fiducial phase space

1031 The Higgs boson transverse momentum is measured in a fiducial phase space, whose  
 1032 requirements are chosen in order to minimize the dependence of the measurements on the  
 1033 underlying model of the Higgs boson properties and its production mechanism.

1034 The exact requirements are determined by considering the two following correlated  
 1035 quantities: the reconstruction efficiency for signal events originating from within the fiducial  
 1036 phase space (fiducial signal efficiency  $\epsilon_{\text{fid}}$ ), and the ratio of the number of reconstructed  
 1037 signal events that are from outside the fiducial phase space (“out-of-fiducial” signal events)  
 1038 to the number from within the fiducial phase space. The requirement of having a small  
 1039 fraction of out-of-fiducial signal events, while at the same time preserving a high value of the  
 1040 fiducial signal efficiency  $\epsilon_{\text{fid}}$ , leads to a loosening of the requirements on the low-resolution  
 1041 variables,  $E_{\text{T}}^{\text{miss}}$  and  $m_{\text{T}}$ , with respect to the analysis selection.

1042 The fiducial phase space used for the cross section measurements is defined at the  
 1043 particle level by the requirements given in Table ???. The leptons are defined as Born-level  
 1044 leptons, i.e. before the emission of final-state radiation (FSR), and are required not to  
 1045 originate from leptonic  $\tau$  decays. The effect of including FSR is evaluated to be of the  
 1046 order of 5% in each  $p_{\text{T}}^{\text{H}}$  bin. For the VH signal process, the two leptons are required to  
 1047 originate from the  $H \rightarrow WW \rightarrow 2\ell 2\nu$  decays in order to avoid including leptons coming  
 1048 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 $p_{\text{T}}$	$p_{\text{T}} > 20 \text{ GeV}$
Subleading lepton $p_{\text{T}}$	$p_{\text{T}} > 10 \text{ GeV}$
Pseudorapidity of electrons and muons	$ \eta  < 2.5$
Invariant mass of the two charged leptons	$m_{\ell\ell} > 12 \text{ GeV}$
Charged lepton pair $p_{\text{T}}$	$p_{\text{T}}^{\ell\ell} > 30 \text{ GeV}$
Invariant mass of the leptonic system in the transverse plane	$m_{\text{T}}^{\ell\ell\nu\nu} > 50 \text{ GeV}$
$E_{\text{T}}^{\text{miss}}$	$E_{\text{T}}^{\text{miss}} > 0$

1049 A detailed description of the fiducial region definition and its optimization is given in  
 1050 appendix ???.

### 1051 4.3.3 Binning of the $p_{\text{T}}^{\text{H}}$ distribution

1052 Experimentally, the Higgs boson transverse momentum is reconstructed as the vector sum  
 1053 of the lepton momenta in the transverse plane and  $E_{\text{T}}^{\text{miss}}$ .

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

1054 Compared to other differential analysis of the Higgs cross section, such as those in the ZZ  
 1055 and  $\gamma\gamma$  decay channels, this analysis has to cope with the limited resolution due to the  $E_T^{\text{miss}}$   
 1056 entering the transverse momentum measurement. The effect of the limited  $E_T^{\text{miss}}$  resolution  
 1057 has two main implications on the analysis strategy. The first one is that the choice of  
 1058 the binning in the  $p_T^H$  spectrum needs to take into account the detector resolution. The  
 1059 second implication is that migrations of events across bins are significant and an unfolding  
 1060 procedure needs to be applied to correct for selection efficiencies and bin migration effects.

1061 Given these aspects the criterion that was used to define the  $p_T^H$  bin size is devised to  
 1062 keep under control the bin migrations due to the finite resolution. For any given bin  $i$  we  
 1063 can define the purity  $P_i$  on a signal sample as the number events that are generated and  
 1064 also reconstructed in that bin,  $N_i^{\text{GEN|RECO}}$ , divided by the number of events reconstructed  
 1065 there,  $N_i^{\text{RECO}}$ :

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

1066 The bin width is chosen in such a way as to make the smallest bins able to ensure a purity  
 1067 of about 60% on a ggH signal sample. Following this prescription we have divided the whole  
 1068  $p_T^H$  range in the following six bins: [0-15 GeV], [15-45 GeV], [45-85 GeV], [85-125 GeV],  
 1069 [125-165 GeV], [165- $\infty$  GeV].

1070 The efficiency of the analysis selection with respect to the fiducial phase space is reported  
 1071 in Fig. ?? (a) for each  $p_T^H$  bin. The efficiency denominator is the number of events that  
 1072 are inside the fiducial phase space, while the numerator is the number of events that pass  
 1073 both the analysis and the fiducial phase space selections. The fake rate, defined by the  
 1074 ratio of signal events that pass the analysis selection but are not within the fiducial phase  
 1075 space, divided by the total number of events passing both the analysis and the fiducial  
 1076 phase space selections is shown in Fig. ?? (b). For both the selection efficiency and the fake  
 1077 rate, all the signal production mechanisms are included. The overall efficiency and fake rate  
 1078 are  $\epsilon = 0.362 \pm 0.005$  and *fake rate* =  $0.126 \pm 0.004$  respectively, where only statistical  
 1079 uncertainties are taken into account.

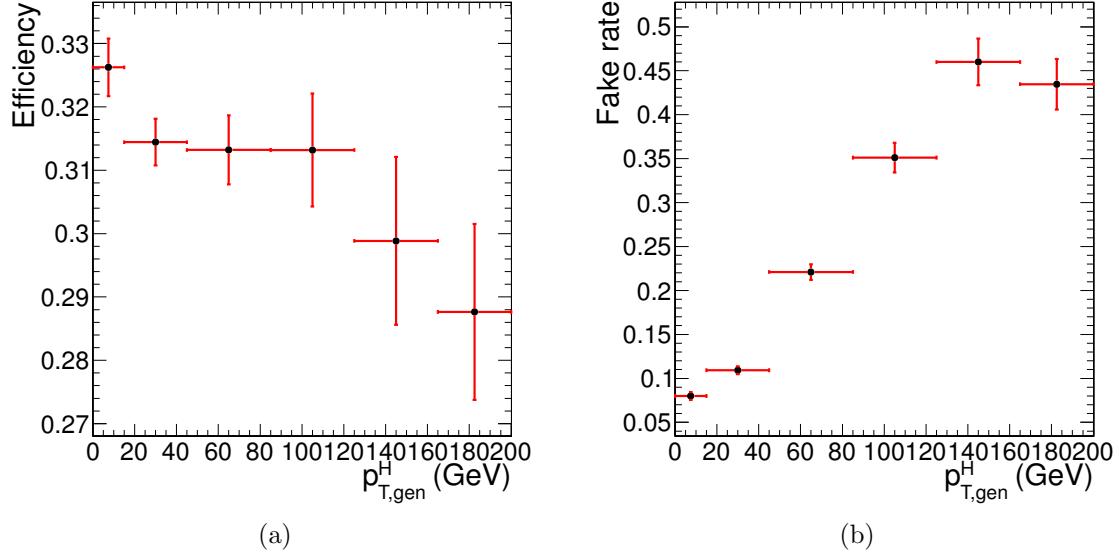
1080 If a  $4\pi$  acceptance is defined, requiring just that the Higgs decays to WW and then to  
 1081  $2\ell 2\nu$ , the efficiency becomes  $\epsilon = 0.0396 \pm 0.0003$  and the fake rate is zero.

## 1082 4.4 Background estimation

### 1083 4.4.1 Top quark background

1084 In this analysis the top quark background is divided into two different categories depending  
 1085 on the number of jets in the event. In the two categories different selections are applied,  
 1086 especially concerning the b-tagging requirements.

1087 The general strategy for determining the residual top events in the signal region is to  
 1088 first measure the top tagging efficiencies from an orthogonal region of phase space in data.



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

1089 The orthogonal phase space is defined inverting the b-veto requirement of the signal region,  
 1090 in such a way to have a control region enriched in top quark events. Then, using this  
 1091 efficiency, the number of events with the associated uncertainty is propagated from the  
 1092 control region to the signal region. The number of surviving top events in the signal region  
 1093 would then be:

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

1094 where  $N_{\text{btag}}^{\text{control}}$  is the number of events in the control region and  $\epsilon_{\text{top}}$  is the efficiency as  
 1095 measured in data.

1096 The methods to estimate the top background contribution in the two jet categories are  
 1097 different and are explained below.

### 1098 0-jets category

1099 Most of the top background, composed of  $t\bar{t}$  and  $tW$  processes, is rejected in the 0-jet bin  
 1100 by the jet veto. The top-tagging efficiency in the zero jet bin,  $\epsilon_{\text{tag}}^{0\text{-jet}}$ , is the probability for a  
 1101 top event to fail one of either the b-tagging veto or the soft muon veto, and is defined as:

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

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

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

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

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

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

## Category with more than 0 jets

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

## Tag&Probe

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

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  $\text{JetBProbability} > 1.4$  (*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:

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

1152

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

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,  $\epsilon_s$  is the b-tagging efficiency,  $\epsilon_b$  is the probability of identifying as b-jet a non-b-jets, i.e. the mistag rate.

A  $\chi^2$  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  $N_s$ ,  $N_b$ ,  $\epsilon_s$  and  $\epsilon_b$  from the fit. The result of the fit on simulation is shown in Fig. ???. The relevant efficiencies are:

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

1161

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

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  $\epsilon_s$  and  $\epsilon_b$  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  $\sigma$  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.

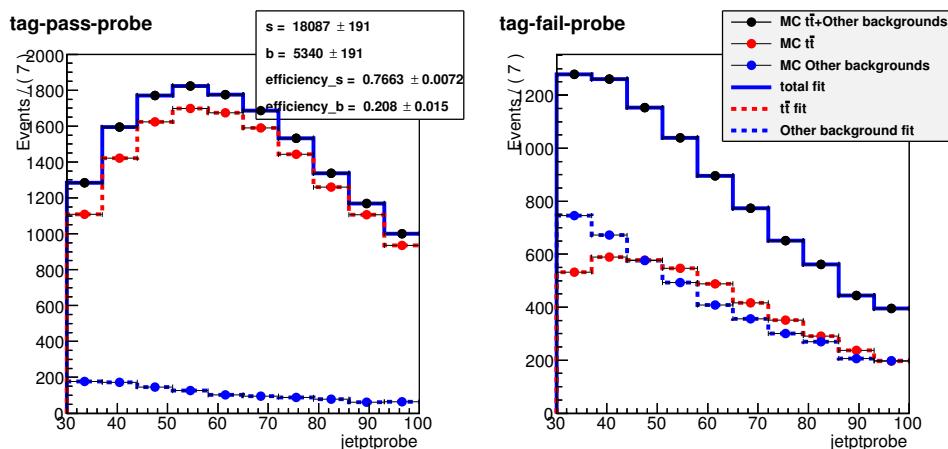
Finally the fit has been performed on data, as shown in Fig. ??, providing the following efficiencies:

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

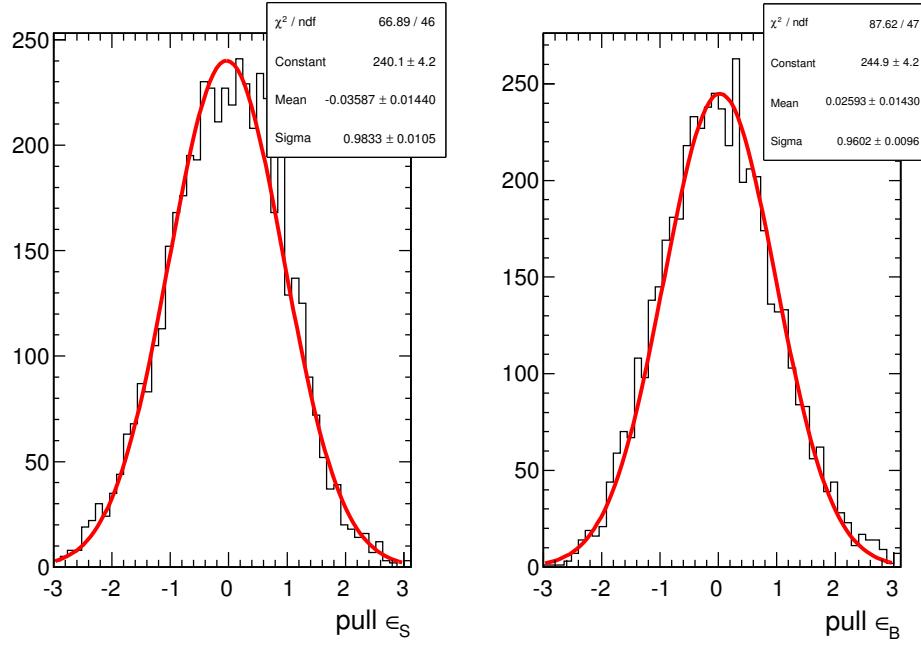
1182

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

1183 Further studies have been performed to assess the effect of the relative uncertainty on  
1184 the  $t\bar{t}$  and  $tW$  event fractions. The same procedure described above has been applied to

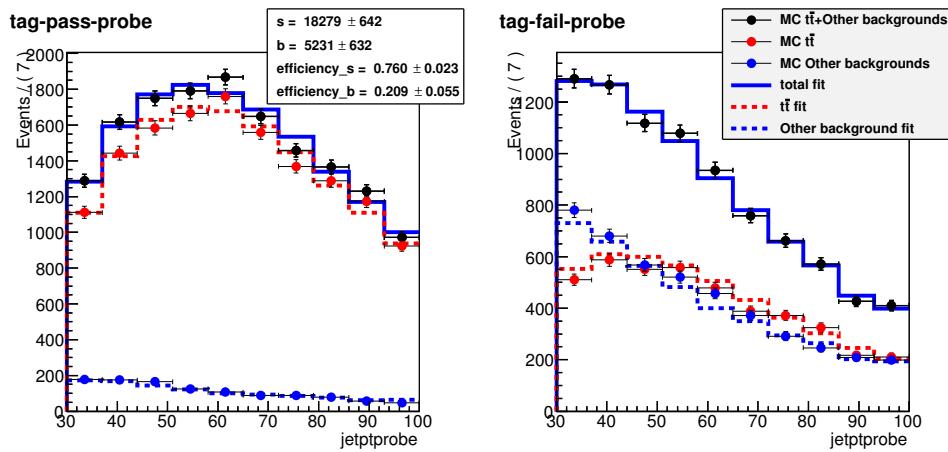


**Figure 4.6:** Simultaneous fit of the *tag-pass-probe* and *tag-fail-probe* pairs in the MC.

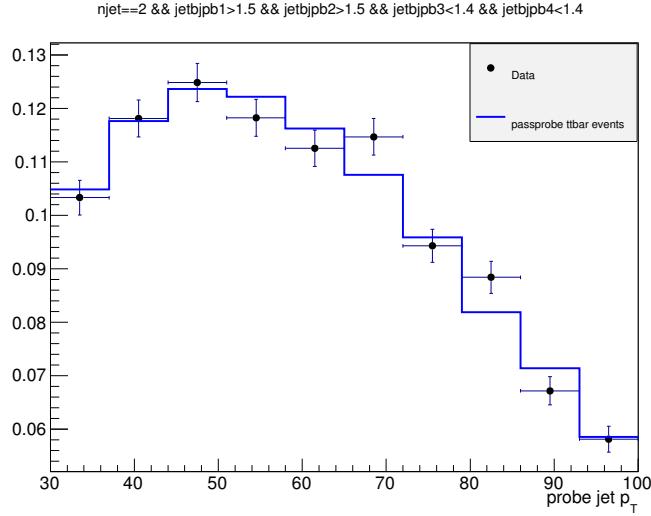


**Figure 4.7:** Pulls of the  $\epsilon_s$  and  $\epsilon_b$  parameters in 5000 toy MC.

1185 different simulation templates obtained varying the  $t\bar{t}$  and  $tW$  fractions within theoretical  
 1186 uncertainties, and the effect on the parameters extracted with the fit procedure is found to  
 1187 be well below the fit uncertainties.



**Figure 4.8:** Fit of a toy MC sample.

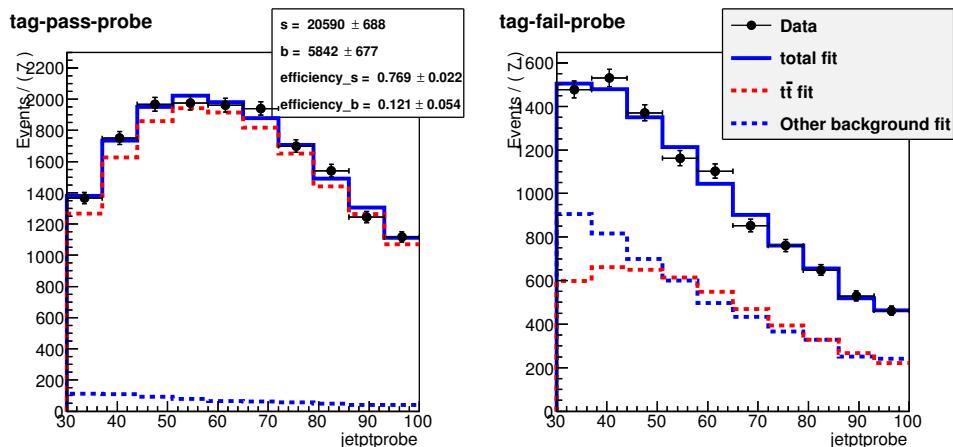


**Figure 4.9:** Shape comparison for the *probe*  $p_T$  spectrum in data and in MC in a very pure  $t\bar{t}$  sample.

### 1188 Data driven estimation

1189 In addition to the b-tagging efficiency, the other ingredient to estimate the  $t\bar{t}$  background is  
 1190 the process cross section. The idea is to measure the cross section in a  $t\bar{t}$  enriched control  
 1191 region, that is called CtrlDD. CtrlDD is defined according to the lepton preselection cuts  
 1192 defined in Sec. ??, and requiring in addition at least one jet with *JetBProbability* score  
 1193 higher than 1.4.

1194 From the simulation we derive the factor  $\alpha$  that connects CtrlDD to the signal region,  
 1195 calculating the ratio of  $t\bar{t}$  events in the two regions:



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

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

1196     The number of events in the CtrlDD region in data is counted, subtracting the expected  
 1197     number of events from non- $t\bar{t}$  backgrounds, and obtaining  $N_{\text{tt} Data}^{\text{CtrlDD}}$ . Finally the number of  
 1198     expected  $t\bar{t}$  events in the signal region ( $N_{\text{tt} Data}^{\text{SIG}}$ ) is obtained as:

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

1199     In evaluating  $\alpha$  and its error the b-tagging efficiencies determined in Sec. ?? are used.  
 1200     For each event an efficiency scale factor and a mistag rate scale factor are derived, depending  
 1201     on whether the event falls in the signal or CtrlDD region.

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

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

1202     where  $n_{b-jets}$  is the number of true b-jets in the event and  $n_{\text{non}-b-\text{jets}}$  is the number of  
 1203     non-b-jets in the event. The writing  $jet1 == b-jet$  ( $jet1 == \text{non}-b-\text{jets}$ ) is a boolean  
 1204     flag that is true when the leading jet, the one used for the CtrlDD selection, is (not) a true  
 1205     b-jet.

1206     Since the efficiency and mistag rate that have been measured on data are close to the one  
 1207     in the simulation, it was decided to assume a scale factor of 1 for both b-tagging efficiency  
 1208     and mis-tag rate. This means that the central values of the scale factors defined in Eq. ??  
 1209     and Eq. ?? is 1, but these numbers have an error that is derived assuming an uncertainty on  
 1210      $\epsilon_s^{\text{Data}}$  and  $\epsilon_b^{\text{Data}}$  that covers both the statistical error from the fit of the two quantities and  
 1211     the difference with respect to the simulation. This results in an up and a down variation of  
 1212     the scale factors in the signal and CtrlDD regions, that is used to derive an error on  $\alpha$ .

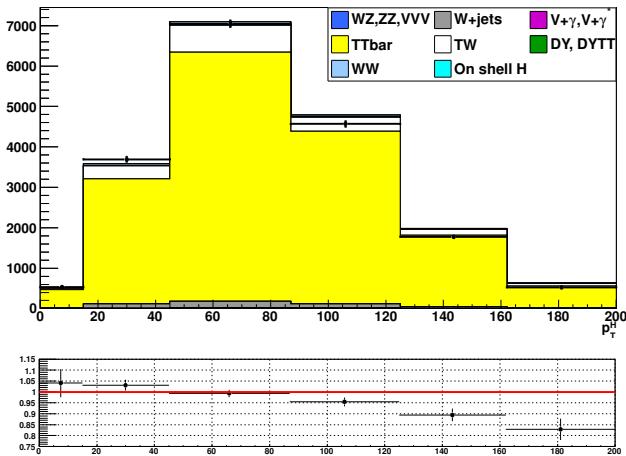
1213     A data driven estimation of the top quark background with the method described above  
 1214     is performed in each of the  $p_T^H$  bins independently. The reason to make this estimation  
 1215     in  $p_T^H$  bins, rather than inclusively is explained in Fig. ??, where the  $p_T^H$  distribution is  
 1216     shown in the CtrlDD region normalized to the cross section measured by a specific CMS  
 1217     analysis [67]. As shown in the ratio plot, an overall normalization factor would not be able  
 1218     to accommodate for the variations of the data/simulation ratio from bin to bin.

1219     The  $\alpha$  factors for each bin and the number of events in signal, CtrlDD regions in MC as  
 1220     well as in data are listed in Tab. ??.

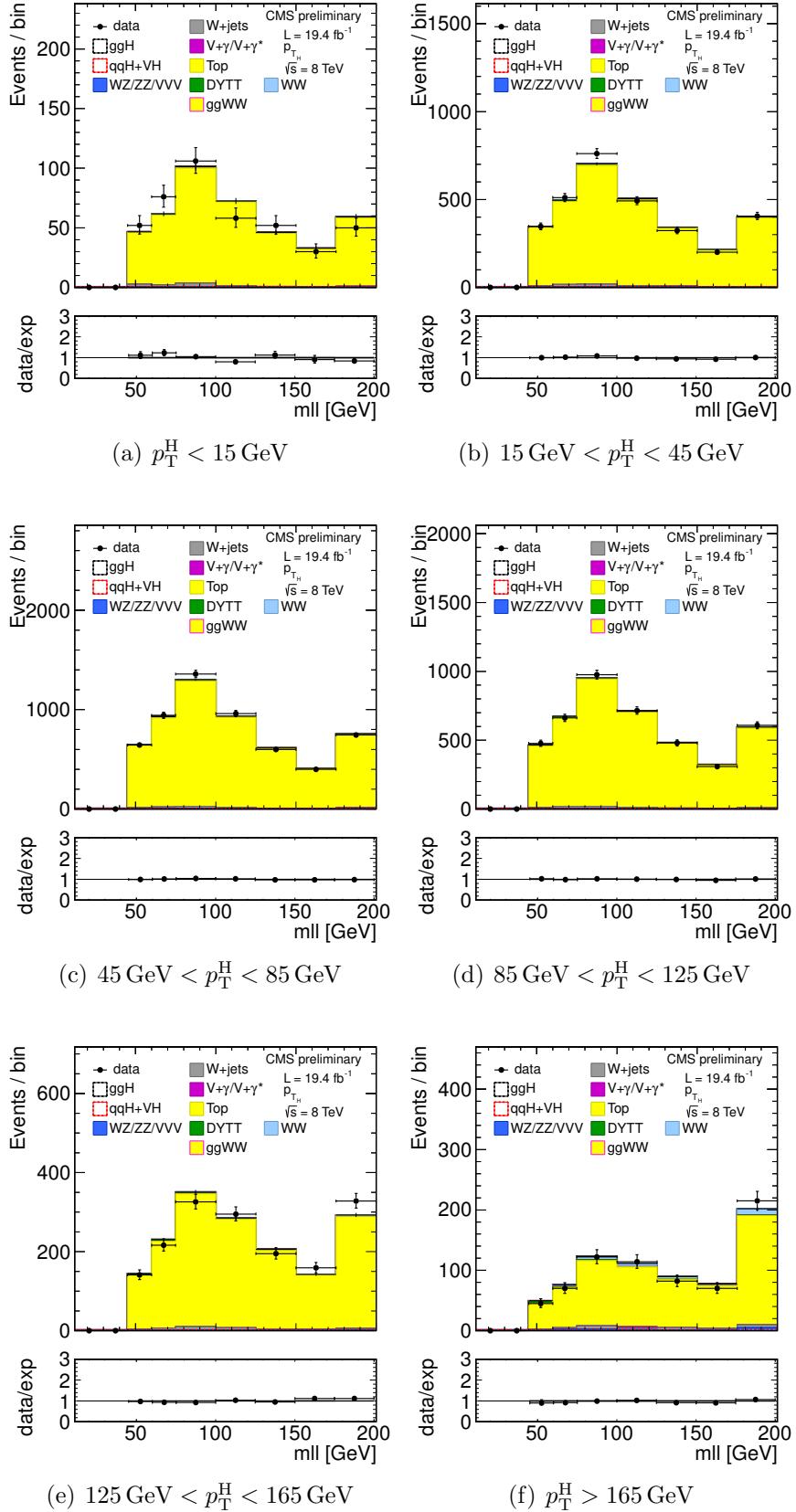
$p_T^H$ [ GeV ]	$N_{CTRL}^{DATA}$	$N_{CTRL}^{TOP}$	$N_{SIG}^{TOP}$	$\alpha$	$\Delta\alpha$
[0–15]	406.71	358.78	117.83	0.328	0.075
[15–45]	2930.14	2703.44	859.08	0.318	0.071
[45–85]	5481.02	5207.48	1506.05	0.289	0.065
[85–125]	4126.35	4032.56	861.22	0.214	0.052
[125–165]	1612.64	1654.27	304.69	0.184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



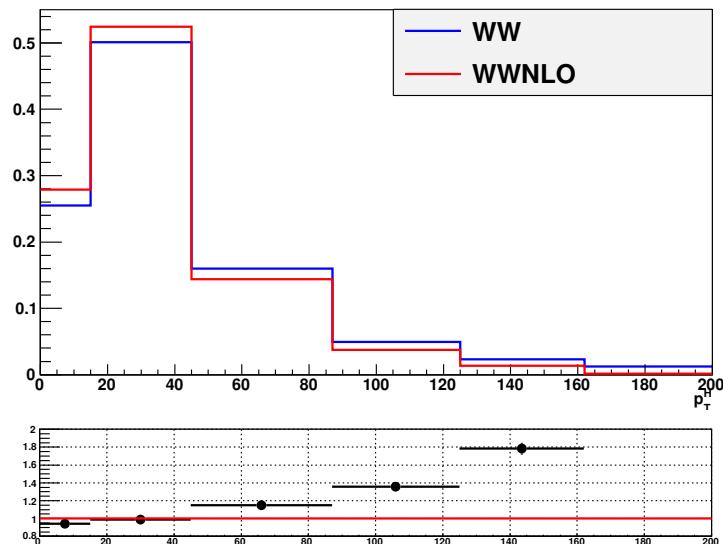
**Figure 4.12:**  $m_{\ell\ell}$  distributions in the CtrlDD region for the different  $p_T^H$  bins.

1223 **4.4.2 WW background**

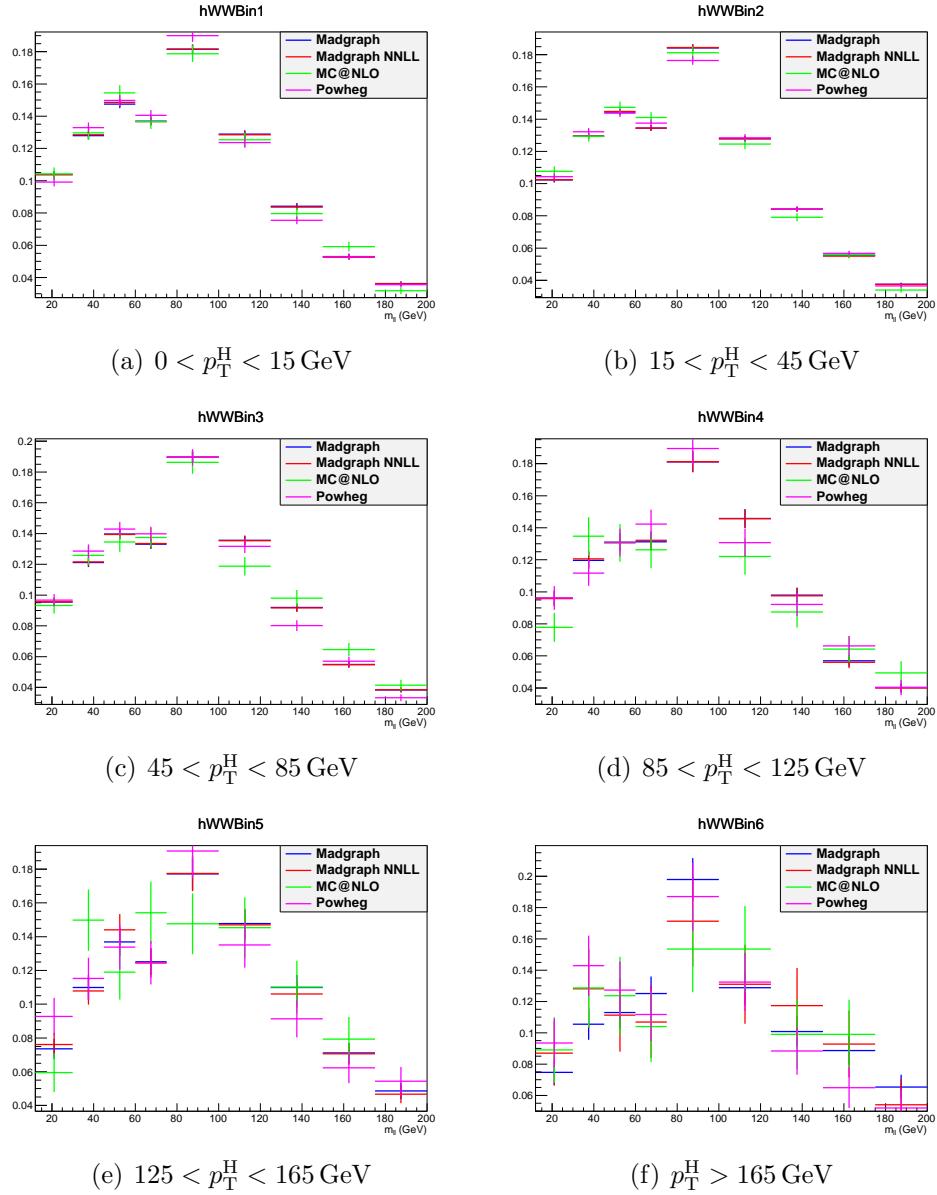
1224 For what the  $qq \rightarrow W^+W^-$  background shape is concerned, the prediction from the simula-  
1225 tion is used. This background is divided into six different parts, corresponding to the six  
1226  $p_T^H$  bins defined in the analysis. The normalization of the  $qq \rightarrow W^+W^-$  background is left  
1227 free to float in each bin, in such a way to adjust it in order to match the data during the fit  
1228 procedure. In this way the shape difference between the  $p_T^{WW}$  theory prediction and the  
1229 distribution provided by the simulation, which is obtained with the MADGRAPH generator,  
1230 is minimized.

1231 In figure ?? a comparison is shown between the  $p_T^{WW}$  spectra of two different  $qq \rightarrow W^+W^-$   
1232 samples: one obtained with the MADGRAPH generator and the other after applying to the  
1233 same distribution a reweighting in order to match the theoretical prediction at NLO+NNLL  
1234 precision.

1235 A shape discrepancy can be clearly observed and the effect becomes larger at high values  
1236 of  $p_T^H$ . In order to assess the effect of this discrepancy on the shapes of the variables used for  
1237 the signal extraction,  $m_{\ell\ell}$  and  $m_T$ , the shapes have been checked in all  $p_T^H$  bins, comparing  
1238 different MC samples. The MADGRAPH sample used for the nominal shape is compared  
1239 to the MADGRAPH sample with NLO+NNLL reweighting, a POWHEG sample with NLO  
1240 accuracy and an AMC@NLO sample. The results of this comparison are shown in figures  
1241 ?? and ?? . The shape discrepancy among the different models is included as an additional  
1242 systematic uncertainty.



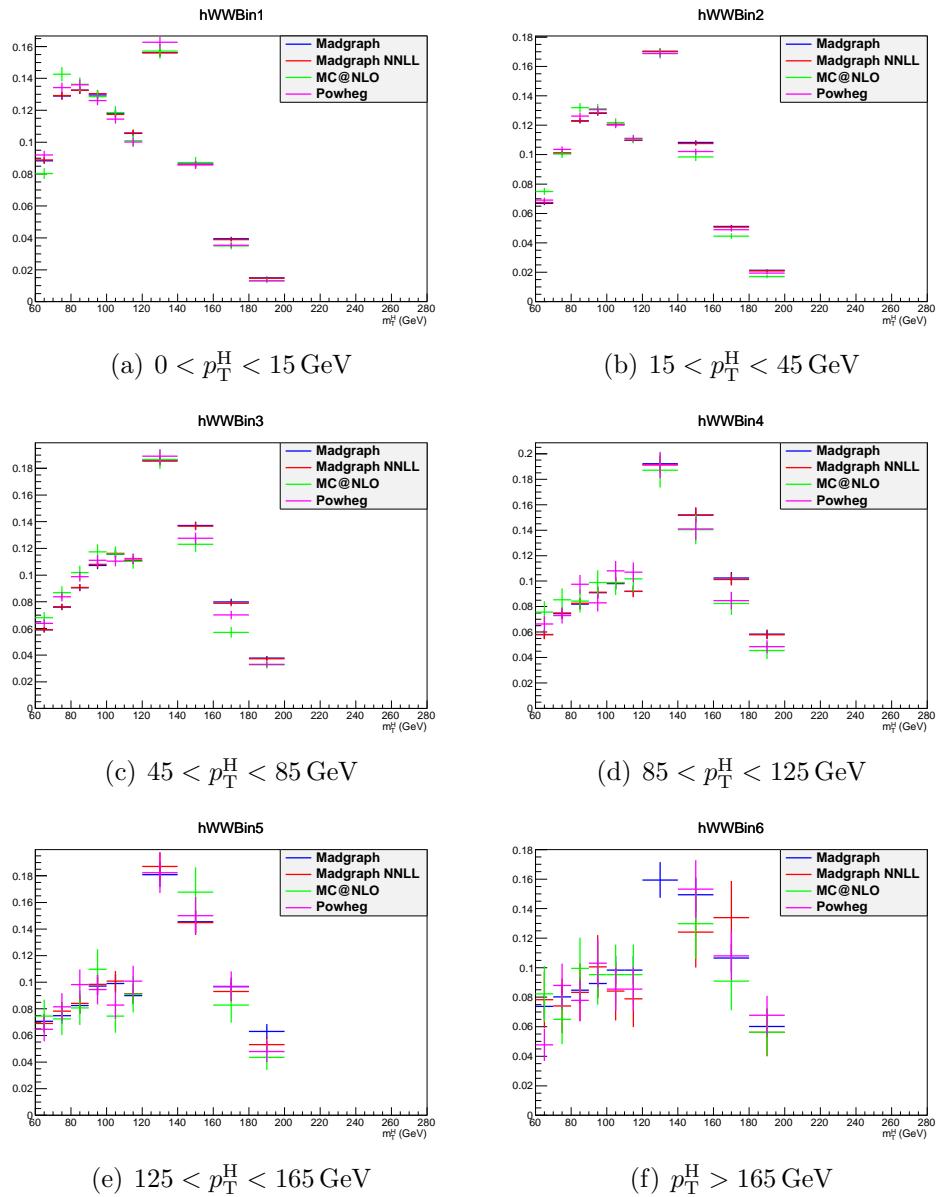
**Figure 4.13:** Comparison between the  $p_T^{WW}$  distributions obtained with two different MC generators: the blue line corresponds to the MADGRAPH generator and the red line refers to the same sample in which a reweighting has been applied in order to match the theoretical prediction at NLO+NNLL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

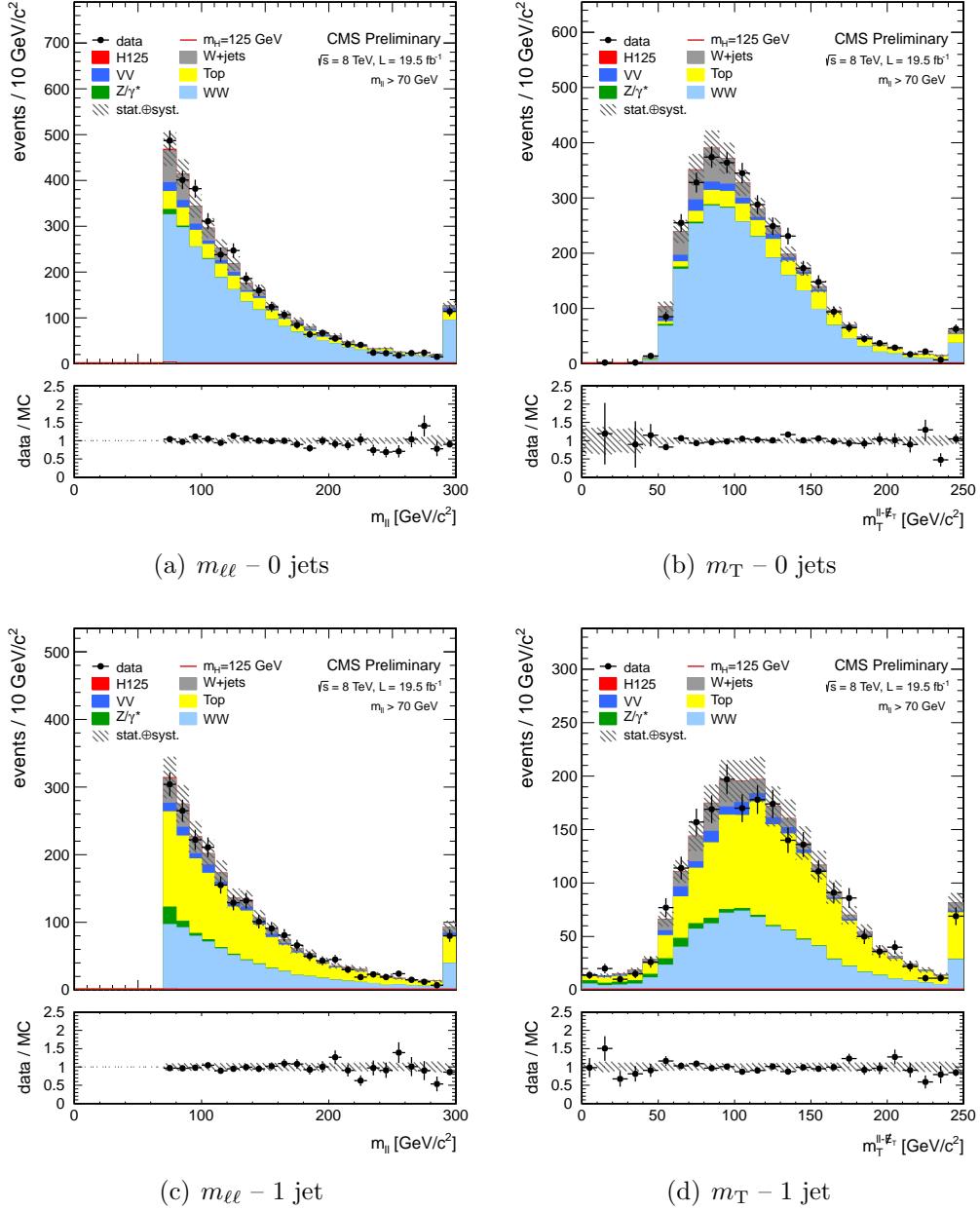
The gluon-induced WW process, i.e.  $gg \rightarrow W^+W^-$ , has a sub-dominant contribution with respect to the quark-induced process, being the cross section ratio between the two of about 5%. The  $m_{\ell\ell}$  and  $m_T$  shapes for this background are taken from simulation while the cross section is scaled to the approximate NLO calculation [51, 52].

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.

<sup>1249</sup> greater than 70 GeV. A comparison of the  $m_{\ell\ell}$  and  $m_T$  shapes in data and simulation is  
<sup>1250</sup> 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.

---

### 1251 4.4.3 Other backgrounds

#### 1252 W+jets background

1253 Events in which W bosons are produced in association with jets, as well as multi-jet events,  
 1254 constitute a background for this analysis, because one or more jets can be misidentified as  
 1255 leptons. The rate at which jets are misidentified as leptons may be not accurately described  
 1256 in simulation, hence a data driven method is used to estimate this background.

1257 The idea is to estimate the background containing one or two fake leptons selecting  
 1258 events with relaxed lepton quality criteria, i.e. looser with respect to the selections used at  
 1259 the analysis level, and computing the efficiencies for real and fake leptons to pass the tight  
 1260 lepton quality requirements of the analysis. A data-driven approach is pursued to estimate  
 1261 this background. A set of loosely selected lepton-like objects, referred to as the “fakeable  
 1262 object” or “denominator” from here on, is defined in a data set of events dominated by  
 1263 dijet production. To measure the fake rate we count how many fakeable objects pass the  
 1264 full lepton selection of the analysis, parameterized as a function of the phase space of the  
 1265 fakeable lepton, therefore it is extracted in bins of  $\eta$  and  $p_T$ . The ratio of the fully identified  
 1266 lepton, referred as “numerator”, to the fakeable objects is taken as the probability for a  
 1267 fakeable object to fake a lepton:

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

1268 It is then used to extrapolate from the loose leptons sample to a sample of leptons  
 1269 satisfying the full selection.

1270 The definition of the denominator is of large impact in the systematic uncertainties  
 1271 related to this method. For the 2012 data taking period a summary of the selections used  
 1272 for the numerator and the denominator of Eq. (??) is shown below for electrons and muons  
 1273 respectively. For electrons the denominator is defined by the following requirements:

- 1274 •  $\sigma_{inj\eta} < 0.01(0.03)$  for barrel (endcap);
- 1275 •  $|\Delta\phi_{in}| < 0.15(0.10)$  for barrel (endcap);
- 1276 •  $|\Delta\eta_{in}| < 0.007(0.009)$  for barrel (endcap);
- 1277 •  $H/E < 0.12(0.10)$  for barrel (endcap);
- 1278 • electron conversion rejection;
- 1279 •  $|d_0| < 0.02$  cm;
- 1280 •  $\frac{\sum_{trk} E_T}{p_T^{ele}} < 0.2$ ;
- 1281 •  $\frac{\sum_{ECAL} E_T}{p_T^{ele}} < 0.2$ ;
- 1282 •  $\frac{\sum_{HCAL} E_T}{p_T^{ele}} < 0.2$ .

1283 For muons the selection are loosened with respect to the tight analysis selection requiring  
 1284 that:

- 1285     •  $|d_0| < 0.02 \text{ cm}$ ;
- 1286     • MVA isolation output  $> -0.6$ .

1287     The dijet enriched data set used for the fake rate measurement, which is selected using  
 1288     single lepton triggers with low  $p_T$  thresholds, it is not a pure sample containing just fake  
 1289     leptons, but may still contain prompt leptons coming from the W and Z boson decays.  
 1290     To reject muons from the W decay, the events are required to have  $E_T^{\text{miss}} < 20 \text{ GeV}$  and  
 1291     a W transverse mass below 20 GeV as well. Muons from the Z decay are instead remove  
 1292     requiring  $m_{\mu\mu} > 20 \text{ GeV}$  and  $m_{\mu\mu} \notin [76, 106] \text{ GeV}$ . For electrons the Z mass peak veto  
 1293     is enlarged to  $m_{ee} \notin [60, 120] \text{ GeV}$ . Finally both electrons and muons are required to be  
 1294     isolated from the leading jet in the event, i.e.  $\Delta\phi(\ell, j) > 1$ . The residual prompt lepton  
 1295     contamination from EW processes such as W/Z+jets production, which can bias the fake  
 1296     rate measurement, is estimated using simulation and subtracted from both the numerator  
 1297     and denominator. The contamination from EW processes is different for the numerator  
 1298     and denominator and is particularly important for relatively high lepton  $p_T$  values.

1299     In addition to the fake rate, also a prompt lepton rate is evaluated, defined as the  
 1300     probability of a prompt lepton passing the loose requirements to also pass the tight analysis  
 1301     selections. The prompt rate is also measured in data, defining a control region enriched in  
 1302      $Z \rightarrow \ell\ell$  events, selecting dilepton events with an invariant mass of the two leptons in the Z  
 1303     peak mass region.

1304     Both the fake and prompt rate are used to reweight the data samples used in the analysis  
 1305     in order to obtain directly from data the contribution of the fake lepton background. The  
 1306     method to apply those rates is explained below in the simple case of just one lepton in  
 1307     the data sample, i.e. data selected by single lepton triggers, but can be straightforwardly  
 1308     generalized to situations with more than one lepton. Suppose that the total number of  
 1309     leptons passing the loose requirements,  $N_\ell$ , is made up of  $N_p$  prompt and  $N_f$  fake leptons.  
 1310      $N_p$  and  $N_f$  cannot be directly measured but one can measure the number of events where  
 1311     no leptons,  $N_{t0}$ , or one lepton,  $N_{t1}$ , pass the tight analysis requirement. These numbers are  
 1312     related by the following equations:

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

1313     where  $p$  and  $f$  are the prompt and fake rates respectively. Equation (??) can be inverted  
 1314     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  $\eta$ ) 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.

**Table 4.4:** Measured prompt rate for electrons and muons in bins of  $\eta$ ,  $p_T$ . Only the statistical uncertainties are shown.

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

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

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

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

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

The low  $E_T^{\text{miss}}$  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  $Z/\gamma^* \rightarrow$

**Table 4.5:** Measured electrons and muons fake rates in bins of  $\eta$  and  $p_T$ , after the EWK correction. Only statistical uncertainties are shown.

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

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

<sup>1332</sup>  $\mu^+\mu^-$  events in data and replacing both muons with a simulated  $\tau \rightarrow \ell\nu_\tau\bar{\nu}_\ell$  decay [41], thus  
<sup>1333</sup> obtaining a “hybrid” event. The Z boson four-momentum is reconstructed in data from  
<sup>1334</sup> the four-momenta of the daughter muons. Then a simulation step allows the replacement  
<sup>1335</sup> of the muon objects with  $\tau$  leptons, in such a way to preserve the Z boson momentum  
<sup>1336</sup> direction is preserved in its rest frame. The  $Z/\gamma^* \rightarrow \tau^+\tau^-$  decay is simulated with the  
<sup>1337</sup> TAUOLA package [68] to correctly describe the  $\tau$ -polarization effects.

<sup>1338</sup> After replacing muons from  $Z/\gamma^* \rightarrow \mu^+\mu^-$  decays with simulated  $\tau$  decays, the set  
<sup>1339</sup> of pseudo-Z/ $\gamma^* \rightarrow \tau^+\tau^-$  events undergoes the reconstruction step. Good agreement in  
<sup>1340</sup> kinematic distributions for this sample and a MC based  $Z/\gamma^* \rightarrow \tau^+\tau^-$  sample is found. The  
<sup>1341</sup> global normalization of pseudo-Z/ $\gamma^* \rightarrow \tau^+\tau^-$  events is checked in the low  $m_T$  spectrum  
<sup>1342</sup> where a rather pure  $Z/\gamma^* \rightarrow \tau^+\tau^-$  sample is expected.

<sup>1343</sup> This method allows to avoid the simulation of very large MC samples that would be  
<sup>1344</sup> needed for an accurate description of this process.

### <sup>1345</sup> ZZ, WZ and $W\gamma$ backgrounds

<sup>1346</sup> The WZ and ZZ backgrounds are partially estimated from data when the two selected  
<sup>1347</sup> leptons come from the same Z boson. If the leptons come from different bosons the  
<sup>1348</sup> contribution is expected to be small. The WZ component is largely rejected by requiring  
<sup>1349</sup> only two high  $p_T$  isolated leptons in the event.

1350     The  $W\gamma^{(*)}$  background, where the photon decays to an electron-positron pair, is expected  
 1351     to be very small, thanks to the stringent photon conversion requirements. This background  
 1352     also includes events where a real photon is produced in association with the  $W$  boson.  
 1353     These events constitute a background for this analysis because the photon can interact with  
 1354     the tracker material converting to an electron-positron pair.

1355     Since the  $WZ$  simulated sample has a generation level cut on the di-lepton invariant mass  
 1356     ( $m_{\ell\ell} > 12$  GeV) and the cross-section raises quickly with the lowering of this threshold, a  
 1357     dedicated MADGRAPH sample has been produced with lower momentum cuts on two of  
 1358     the three leptons ( $p_T > 5$  GeV) and no cut on the third one. The surviving contribution  
 1359     estimated with this sample is still very small, and since the uncertainty on the cross-section  
 1360     for the covered phase space is large, a conservative 100% uncertainty has been given to it.  
 1361     A  $k$ -factor for  $W\gamma^*$  of  $1.5 \pm 0.5$  based on a dedicated measurement of tri-lepton decays,  
 1362      $W\gamma^* \rightarrow e\mu\mu$  and  $W\gamma^* \rightarrow \mu\mu\mu$ , is applied [41]. The contribution of  $W\gamma^{(*)}$  is also constrained  
 1363     by a closure test with same sign leptons on data, which reveals a good compatibility of the  
 1364     data with the expected background.

## 1365     4.5 Systematic uncertainties

1366     Systematic uncertainties play an important role in this analysis where no strong mass peak  
 1367     is expected due to the presence of undetected neutrinos in the final state. One of the most  
 1368     important sources of systematic uncertainty is the normalization of the backgrounds that  
 1369     are estimated on data control samples whenever is possible.

1370     A summary of the main sources of systematic uncertainty and the corresponding estimate  
 1371     is reported in Table ???. A detailed description of each source of systematic uncertainty is  
 1372     discussed in the following sections.

### 1373     4.5.1 Background normalization uncertainties

1374     The signal extraction is performed subtracting the estimated backgrounds to the event  
 1375     counts in data. This uncertainty depends on the background:

- 1376       •  **$t\bar{t}$  and  $tW$  backgrounds:** The efficiency on jets b-tagging is estimated using the  
 1377       Tag&Probe technique in data and simulation control regions, as explained in ???. A  
 1378       per-jet scale factor, which takes into account the possibly different efficiency of the  
 1379       anti b-tagging selection in data and simulation, is computed by means of the efficiency  
 1380       measured with the Tag&Probe method. The Tag&Probe method has been used also  
 1381       to measure the mistag rates in data and simulation, which are the probability to b-tag  
 1382       a jet that is not produced by the hadronization of a b quark. These factors are used  
 1383       to reweigh the Top MC samples as explained in ???. The uncertainties provided by the  
 1384       Tag&Probe fit are then propagated to the factor  $\alpha$  that is used in the top data driven  
 1385       estimation ???. These uncertainties are embedded in a systematic error that affects  
 1386       the shape of the Top background in each  $p_T^H$  bin.

**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.

<b>Uncertainties in backgrounds contributions</b>	
Source	Uncertainty
t <bar>t&gt;, tW</bar>	20–50%
W+ jets	40%
WZ, ZZ	4%
W $\gamma$ (*)	30%

<b>Effect of the experimental uncertainties on the signal and background yields</b>	
Source	Uncertainty
Integrated luminosity	2.6%
Trigger efficiency	1–2%
Lepton reconstruction and identification	3–4%
Lepton energy scale	2–4%
$E_T^{\text{miss}}$ modelling	2%
Jet energy scale	10%
Pileup multiplicity	2%
b mistag modelling	3%

<b>Effect of the theoretical uncertainties on signal yield</b>	
Source	Uncertainty
b jet veto scale factor	1–2%
PDF	1%
WW background shape	1%

1387      Provided that the simulated samples include both tt> and tW processes, a systematic  
 1388      uncertainty related to the tW/tt> fraction has been included. In fact, a relative variation  
 1389      of the contribution of these two processes could modify the shape of the MC sample,  
 1390      and is thus included as a shape uncertainty affecting the top quark background shape  
 1391      in each  $p_T^H$  bin in a correlated way.

- 1392    • **W+jets background:** It is estimated with data control sample as described in  
 1393      Sec.???. With  $19.4 \text{ fb}^{-1}$  at 8 TeV, the uncertainty receives similar contributions from  
 1394      statistics and systematic error (mainly jet composition differences between the fake  
 1395      rate estimation sample and the application sample), the total error being about 40%,  
 1396      dominated by the closure test of the method on a same-sign control region.

- **WZ, ZZ,  $W\gamma^{(*)}$  backgrounds:** those backgrounds, which are expected to give a small contribution, are estimated from simulation. Uncertainties on the cross sections reported in [69, 70] are 4% for WZ and 2.5% for ZZ. A 30% uncertainty is assigned to the  $W\gamma$  [71] yield and another 30% on  $W\gamma^{(*)}$  contribution according to the uncertainty on the normalization study (see Sec. ??).

#### 4.5.2 Experimental uncertainties

The following experimental systematic sources have been taken into account:

- **Luminosity:** Using the online luminosity monitoring CMS reached an uncertainty on the luminosity of 2.6% at 8 TeV.
- **Trigger efficiency.** The uncertainties for both electrons and muons are at 1-2% level, which is added together to the lepton efficiency uncertainty.
- **Lepton reconstruction and identification efficiency:** The lepton reconstruction and identification efficiencies are measured with the Tag&Probe method in data. To correct for the difference in the lepton identification efficiencies between data and MC, a scale factor is applied to MC. The uncertainties resulting from this procedure on the lepton efficiencies are 4% for electrons and 3% for muons.
- **Muon momentum and electron energy scale:** The momentum scale of leptons have relatively large uncertainties due to different detector effects. For electrons a scale uncertainty of 2% for the barrel, and 4% for the endcaps respectively, is assigned. For muons, a momentum scale uncertainty of 1.5%, independent of its pseudorapidity, is assigned.
- **$E_T^{\text{miss}}$  modeling:** The  $E_T^{\text{miss}}$  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  $\eta$  and  $p_T$  of the jet [72]) 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%.

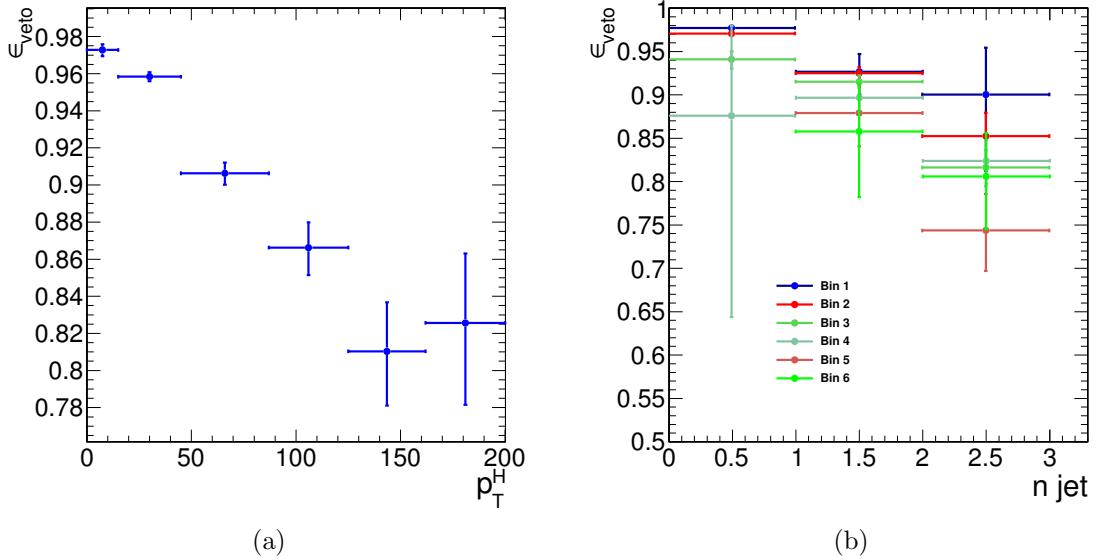
### 4.5.3 Theoretical uncertainties

- **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 [73, 74] prescription, using CT10, NNPDF2.1 [75] and MSTW2008 [76] 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 ??.

#### Jet multiplicity uncertainty

The jet bin uncertainty on the ggH production mode has been evaluated using the Stewart-Tackman method, following the recipe proposed in Refs. [40, 77]. Three independent nuisance parameters have to be associated with the inclusive ggH production cross sections  $\sigma_{\geq 0}$ ,  $\sigma_{\geq 1}$  and  $\sigma_{\geq 2}$ , which corresponds to the cross sections with  $\geq 0$  jets,  $\geq 1$  jet and  $\geq 2$  jets respectively. According to the agreement on the treatment of uncertainties in the combination of ATLAS and CMS results [78], these nuisance parameters are labelled as  $QCDscale_{ggH}$ ,  $QCDscale_{ggH1in}$  and  $QCDscale_{ggH2in}$ . However, in case the analysis is split in exclusive jet multiplicity bins, the jet bin uncertainties can be evaluated taking into

account the correct correlations among the three nuisances following the Stewart-Tackman prescription. Even though this analysis is inclusive in number of jets, the jet binning uncertainties must be included due to the presence of the b-jet veto, that introduces a dependency of the selection efficiency on the number of jets in the event. The veto efficiency has been evaluated in all the  $p_T^H$  bins defined in the analysis and as a function of jets multiplicity. The results are shown in Figs. ?? and ???. The drop of the veto efficiency at high values of  $p_T^H$  is due to the correlation with jets multiplicity.



**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.

The first step of this procedure is to take the inclusive ggH cross section,  $\sigma_{\text{ggH}}$ , and to convert the relative QCD up/down scale uncertainties,  $\epsilon_+$  and  $\epsilon_-$ , to a log-normal uncertainty, i.e.  $\kappa = \sqrt{\exp(\epsilon_+) \cdot \exp(\epsilon_-)}$ . The exclusive cross sections,  $\sigma_0$ ,  $\sigma_1$  and  $\sigma_2$ , can be calculated starting from  $\sigma_{\text{ggH}}$  and using the selection efficiencies for the three jet bins. For every exclusive cross section the corresponding relative uncertainty is computed varying the renormalization ( $\mu_R$ ) and factorization ( $\mu_F$ ) 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.  $f_0 = \sigma_0 / \sigma_{\geq 0}$ ,  $f_1 = \sigma_1 / \sigma_{\geq 0}$ ,  $f_2 = \sigma_2 / \sigma_{\geq 0}$ .

**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		$\Delta_{\geq 2}^1 = (\kappa_{\geq 2})^{-\frac{f_2}{f_1}}$	$\Delta_{\geq 2}^2 = (\kappa_{\geq 2})$

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:

$$QCDscale\_ggH = \frac{\Delta_{\geq 0}^0 \cdot f_0 \cdot \varepsilon_0 + \Delta_{\geq 1}^0 \cdot f_1 \cdot \varepsilon_1}{\Delta_{\geq 0}^0 \cdot f_0 \cdot \varepsilon_0 + \Delta_{\geq 1}^0 \cdot f_1 \cdot \varepsilon_0} , \quad (4.25)$$

$$QCDscale\_ggH1in = \frac{\Delta_{\geq 1}^1 \cdot f_1 \cdot \varepsilon_1 + \Delta_{\geq 2}^1 \cdot f_2 \cdot \varepsilon_2}{\Delta_{\geq 1}^1 \cdot f_1 \cdot \varepsilon_1 + \Delta_{\geq 2}^1 \cdot f_2 \cdot \varepsilon_1} , \quad (4.26)$$

$$QCDscale\_ggH2in = 1 , \quad (4.27)$$

where  $\varepsilon_0$ ,  $\varepsilon_1$  and  $\varepsilon_2$  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.

#### 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.

**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

### 1507 4.5.5 Treatment of systematic uncertainties in the shape analysis

1508 One can distinguish between normalization uncertainties, where a systematic effect is  
 1509 changing the normalization of a given process assuming the shape is not affected, and shape  
 1510 uncertainties where the actual change in the shape of the distribution is taken into account.  
 1511 The normalization uncertainties enter the shape analysis as a constant normalization factor,  
 1512 whereas for shape uncertainties the nominal and the  $+1\sigma$  and  $-1\sigma$  shapes enter the analysis  
 1513 in form of three histograms with the same normalization.

1514 For the W+jets background, the shape differences for different jet  $p_T$  thresholds in the  
 1515 di-jet control sample are considered separately for electron and muon fakes, while the other  
 1516 sources of systematics are taken as normalization uncertainties as in the cut-based analysis.

1517 Effects from experimental uncertainties are studied by applying a scaling and smearing  
 1518 of certain variables of the physics objects, followed by a subsequent recalculation of all  
 1519 the correlated variables. This is done for simulation, to account for possible systematic  
 1520 mis-measurements of the data. All experimental sources from Section ?? but luminosity are  
 1521 treated both as normalization and shape uncertainties. For background with a data-driven  
 1522 normalization estimation, only the shape uncertainty is considered.

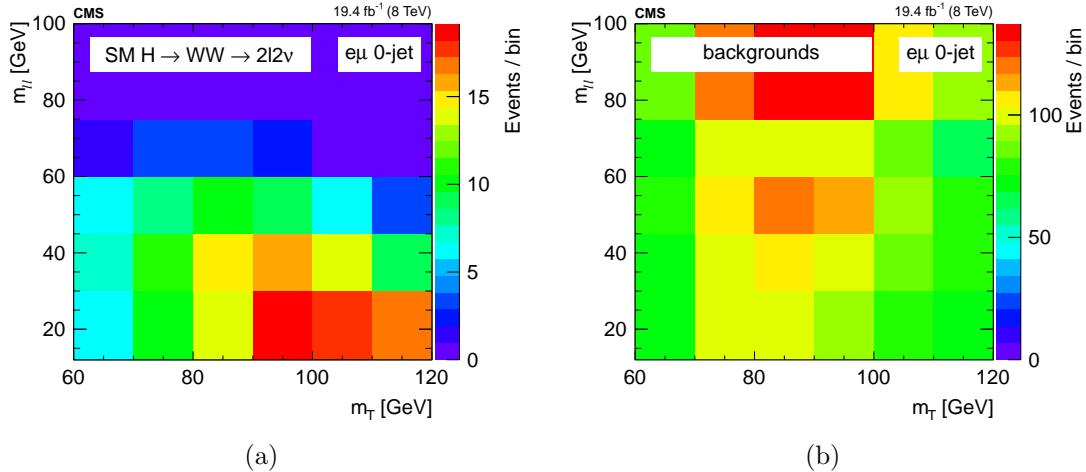
1523 To account for statistical uncertainties, for each distribution going into the shape analysis,  
 1524 the  $+1\sigma$  and  $-1\sigma$  shapes were obtained by adding/subtracting the statistical error in each  
 1525 bin and renormalizing it to the nominal distribution. In addition to this procedure a  
 1526 constant normalization uncertainty due to the finite statistics of the MC sample used to  
 1527 extract the shape is assigned.

## 1528 4.6 Signal extraction

1529 According to the “blinding” policy of the CMS Collaboration, the strategy of the analysis  
 1530 has been scrutinized and approved by a selected committee of internal reviewers before  
 1531 looking at the data in the signal region. This approach prevents the analysts from being  
 1532 biased by the data in the developing phase of the analysis. Below are shown the results  
 1533 after having looked at the data.

### 1534 4.6.1 Fitting procedure

1535 The signal, including ggH, VBF, and VH production mechanisms, is extracted in each bin  
 1536 of  $p_T^H$  by performing a binned maximum likelihood fit simultaneously in all  $p_T^H$  bins to a  
 1537 two-dimensional template for signals and backgrounds in the  $m_{\ell\ell}$ – $m_T$  plane. The variables  
 1538 used for the two-dimensional template are chosen for their power to discriminate signal  
 1539 and background contributions. This is shown in Fig. ??, where the two-dimensional MC  
 1540 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.

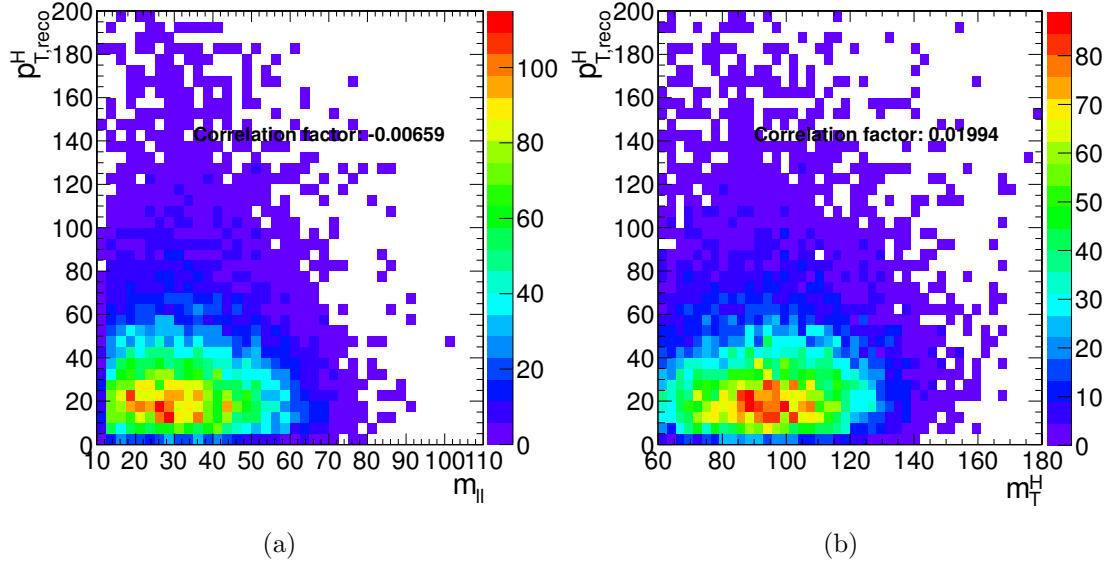
1541 Six different signal strength parameters are extracted from the fit, one for each  $p_T^H$  bin.  
 1542 The relative contributions of the different Higgs production mechanisms in the signal  
 1543 template are taken to be the same as in the SM. The systematic uncertainty sources are  
 1544 considered as nuisance parameters in the fit.

1545 The binning of the  $m_{\ell\ell}$  and  $m_T$  templates is chosen to be:

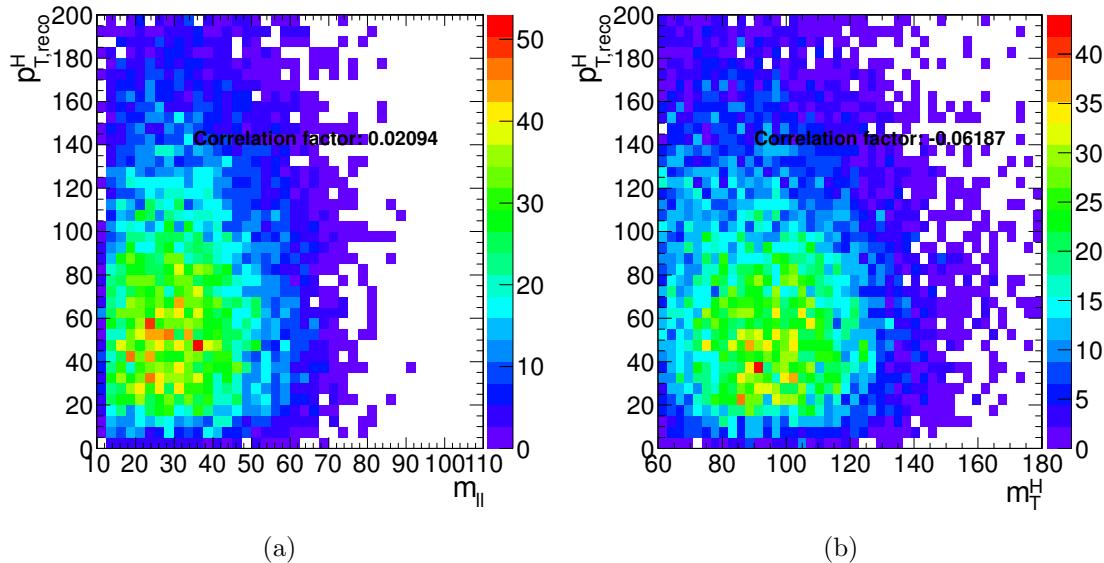
- 1546 •  $m_{\ell\ell}$ : [12, 30, 45, 60, 75, 100, 125, 150, 175, 200]
- 1547 •  $m_T$ : [60, 70, 80, 90, 100, 110, 120, 140, 160, 180, 200, 220, 240, 280]

1548 To avoid a dependence of the results on the variables used for the template fit,  $m_{\ell\ell}$  and  
 1549  $m_T$  need to be uncorrelated with respect to  $p_T^H$ . This has been verified and the correlation  
 1550 between the discriminating variables and  $p_T^H$  is shown in Fig. ?? and Fig. ?? for ggH and  
 1551 VBF production modes respectively.

1552 The relative contribution for different production mechanisms in the input signal template  
 1553 is taken to be the same as the SM. The signal strength  $\mu$  in each bin, i.e. the ratio between  
 1554 the measured cross section and the SM one,  $\mu = \sigma/\sigma_{SM}$ , is allowed to float between -10 and  
 1555 +10, thus allowing negative values. This is mainly intended to allow the error bars to float  
 1556 below 0.



**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.

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

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

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

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

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

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

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

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

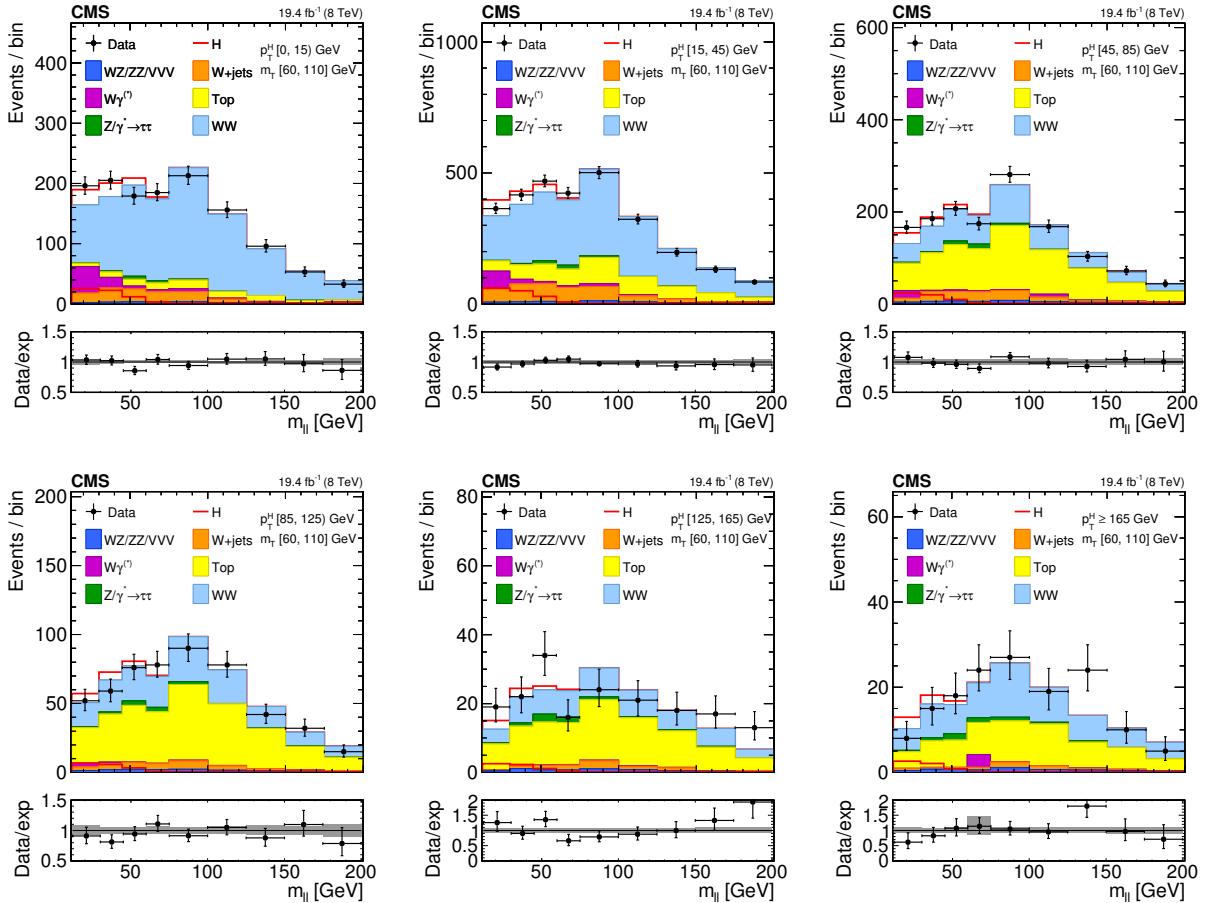
Before running the fit on the data, the same procedure has been applied to the so called *Asimov data set*<sup>1</sup>, which provides a simple method to estimate the signal sensitivity before looking at the data [79].

---

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

### 1591 4.6.2 Signal and background yields

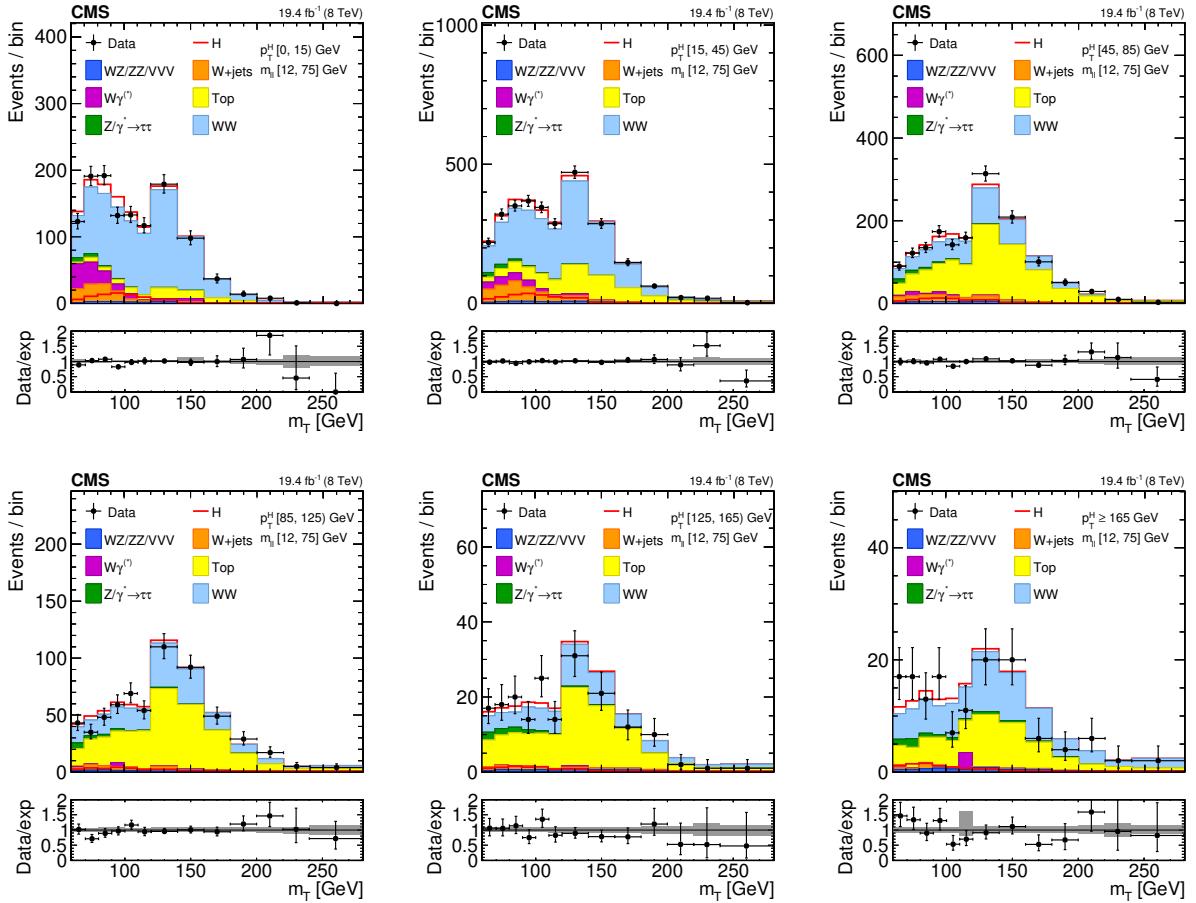
1592 A comparison of data and background prediction is shown in Fig. ??, where the  $m_{\ell\ell}$   
 1593 distribution is shown for the six  $p_T^H$  bins. Distributions correspond to the  $m_T$  window of  
 1594 [60, 110] GeV, in order to emphasize the signal contribution [41]. The  $m_T$  distributions are  
 1595 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.

1596 The signal and background yields after the analysis selection are reported in Table ??.

1597 The spectrum shown in Fig. ?? is obtained after having performed the fit and after the  
 1598 subtraction of the out-of-fiducial signal events, but before undergoing the unfolding proce-



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

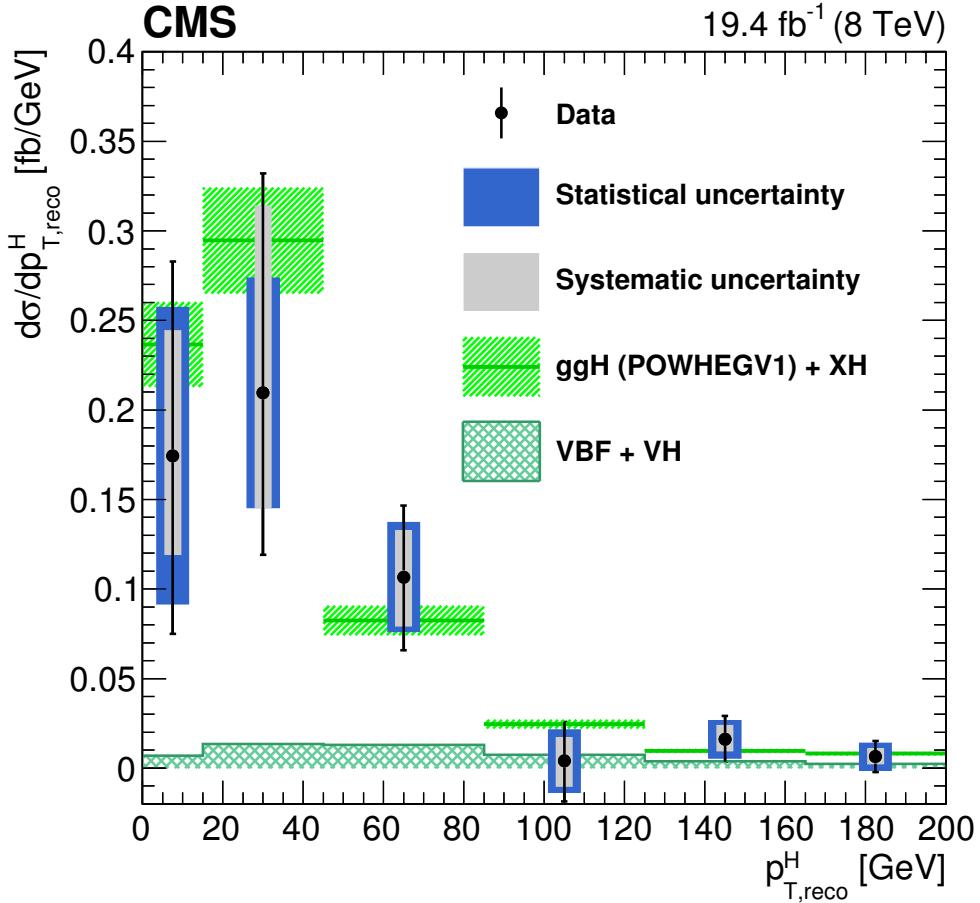
<sup>1599</sup> dure. The theoretical distribution after the detector simulation and event reconstruction is  
<sup>1600</sup> also shown for comparison.

<sup>1601</sup> In order to assess the robustness of the fit, several toy MC samples have been produced,  
<sup>1602</sup> with a statistical accuracy corresponding to the one expected in data. The distribution  
<sup>1603</sup> of the signal strengths extracted in each bin using the toy MC samples and the their pull  
<sup>1604</sup> distributions are shown in Fig. ??.

## <sup>1605</sup> 4.7 Unfolding

<sup>1606</sup> To facilitate comparisons with theoretical predictions or other experimental results, the  
<sup>1607</sup> signal extracted performing the fit has to be corrected for detector resolution and efficiency  
<sup>1608</sup> effects and for the efficiency of the selection defined in the analysis. An unfolding procedure  
<sup>1609</sup> is used relying on the ROOUNFOLD package [80], which provides the tools to run various  
<sup>1610</sup> unfolding algorithms.

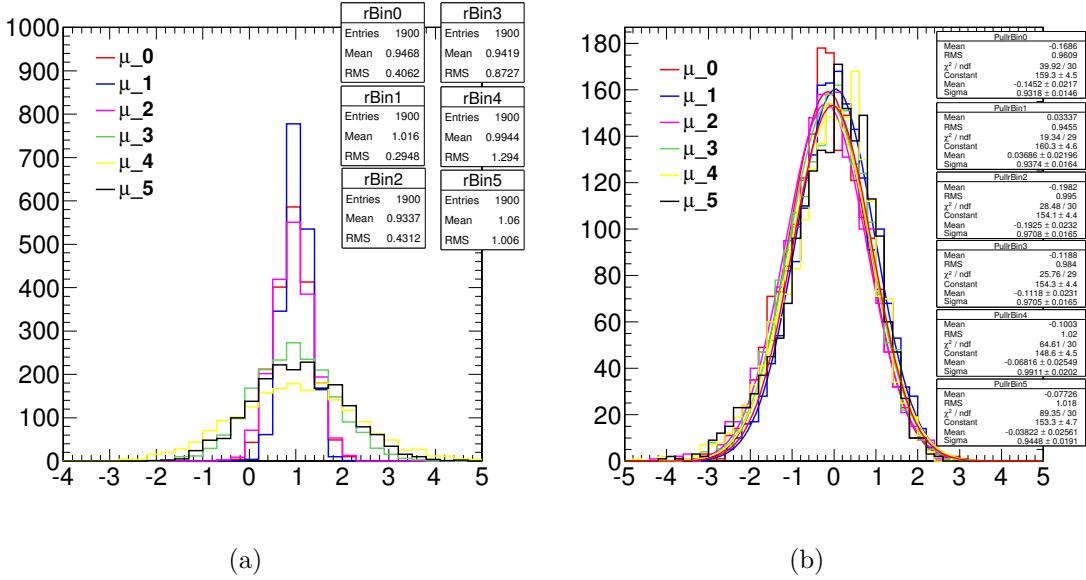
<sup>1611</sup> The basic principle behind the unfolding procedure in this analysis is to use MC signal  
<sup>1612</sup> samples to make the “true” distribution of the variable of interest, which is obtained using  
<sup>1613</sup> simulated events before particle interaction with the detector, and the same distribution  
<sup>1614</sup> obtained using events reconstructed after the full GEANT4 simulation of the CMS detector  
<sup>1615</sup> and event reconstruction. These two distributions are used to calculate the detector response  
<sup>1616</sup> matrix  $M$ :



**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.

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

where  $R^{\text{MC}}$  and  $T^{\text{MC}}$  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



**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).

analysis selection that affect the  $R^{\text{MC}}$  distribution. The goal of the unfolding procedure is to obtain the  $T^{\text{truth}}$  distribution starting from the measured  $R^{\text{observed}}$  distribution by inverting the matrix  $M$ . To avoid the large variance and strong negative correlation between the neighbouring bins [42], the unfolding procedure in this analysis relies on the singular value decomposition [81] 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 [42]. 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 [82]. One possible solution is the utilization of regularization methods. Such methods introduce a regularization function that controls the smoothness of the distribution and depends generally on one regularization parameter, which can be controlled to achieve the desired degree of smoothness. The choice of the regularization parameter is particularly critical, and it should represent an optimal trade-off between taming the fluctuations in the unfolded result, and biasing the unfolded distribution towards the one used to build the response matrix. The main feature of this method is the use of the singular value decomposition of the response matrix, including an additional term to suppress the oscillatory component of the solution, i.e. the regularization term, which represents some *a priori* knowledge of the final solution. The regularization parameter is chosen to obtain results that are robust against numerical instabilities and statistical fluctuations, following the prescription

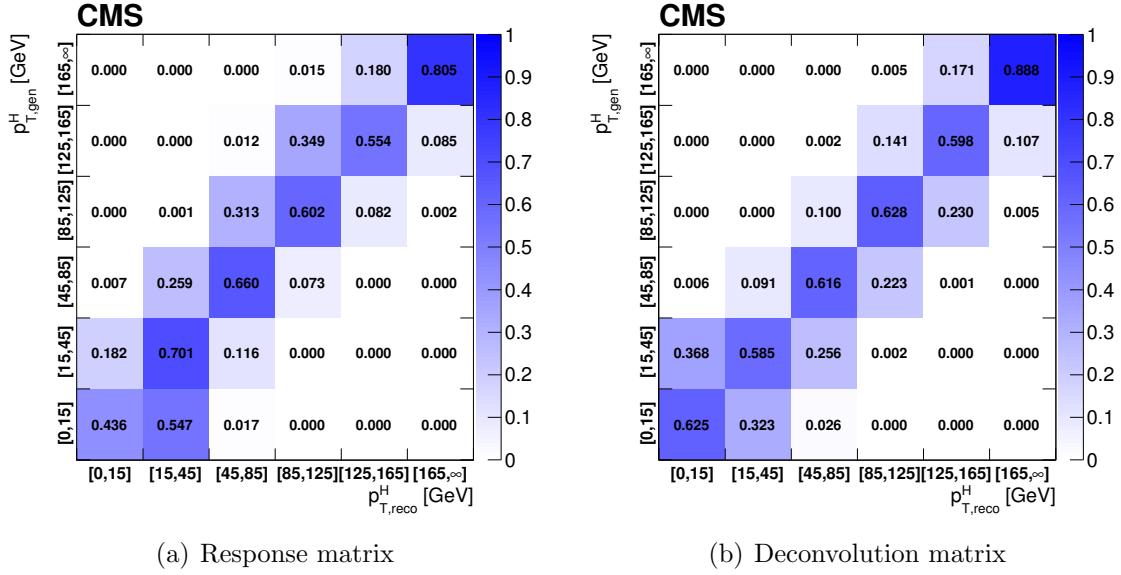
described in Ref. [81]. **Maybe I should add an appendix describing the SVD method in details**

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

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

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



**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.

**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– $\infty$	$0.632^{+0.016}_{-0.016}$	$0.674^{+0.015}_{-0.015}$	$0.701^{+0.015}_{-0.015}$	$0.722^{+0.014}_{-0.015}$

<sup>1687</sup> without introducing a bias, i.e. until a 68% coverage is fulfilled. This criterion leads to the  
<sup>1688</sup> same result as the prescription described in Ref. [81], strengthening the choice of  $k_{\text{reg}} = 3$ .

### <sup>1689</sup> 4.7.1 Treatment of systematic uncertainties

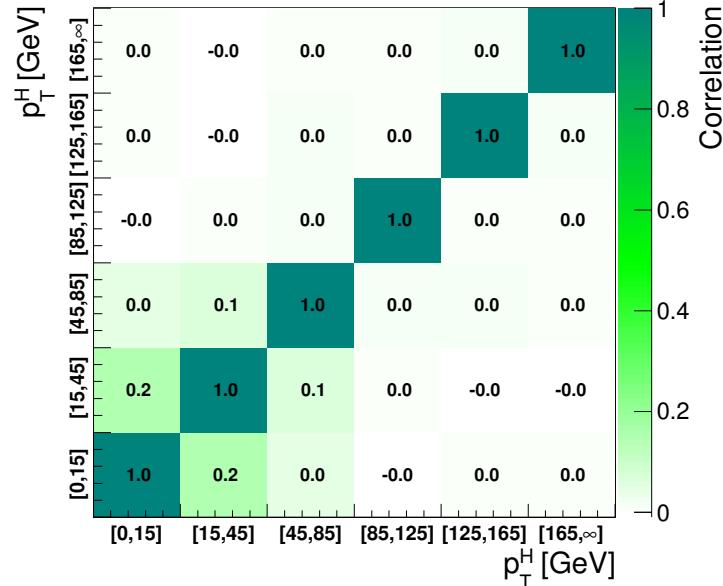
<sup>1690</sup> An important aspect of this analysis is the treatment of the systematic uncertainties and  
<sup>1691</sup> the error propagation through the unfolding procedure. The sources of uncertainty are

1692 divided into three categories, depending on whether the uncertainty affects only the signal  
 1693 yield (type A), both the signal yield and the response matrix (type B), or only the response  
 1694 matrix (type C). These three classes propagate differently through the unfolding procedure.

1695 Type A uncertainties are extracted directly from the fit in the form of a covariance  
 1696 matrix, which is passed to the unfolding tool as the covariance matrix of the measured  
 1697 distribution. The nuisance parameters belonging to this category are the background shape  
 1698 and normalization uncertainties. To extract the effect of type A uncertainties a dedicated  
 1699 fit is performed, fixing to constant all the nuisance parameters in the model, but type A  
 1700 nuisance parameters. The correlation matrix among the six signal strengths corresponding  
 1701 to the six  $p_T^H$  bins, including all type A uncertainties, is shown in Fig. ???. The correlation  
 1702  $\text{cor}(i,j)$  of bins  $i$  and  $j$  is defined as:

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

1703 where  $\text{cov}(i,j)$  is the covariance of bins  $i$  and  $j$ , and  $(s_i, s_j)$  are the standard deviations  
 1704 of bins  $i$  and  $j$ , respectively.



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

1705 The nuisance parameters falling in the type B class are:

- 1706 • the b veto scale factor. It affects the signal and background templates by varying the  
 1707 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.

- 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  $E_T^{\text{miss}}$  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  $\pm 1$  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.

Type C uncertainties are related to the underlying assumption on the Higgs boson production mechanism used to extract the fiducial cross sections. These are evaluated using alternative response matrices that are obtained by varying the relative fraction of the VBF and ggH components within the experimental uncertainty, as given by the CMS combined measurement [83]. Three different response matrices are built, corresponding to the nominal, scaled up, and scaled down VBF/ggH ratio. The nominal matrix assumes the SM VBF/ggH ratio, while up- and down-scaled matrices are constructed by varying the SM signal strengths within the experimental constraints for VBF and ggH in such a way as to obtain the maximal variation of the VBF/ggH ratio allowed by the experimental constraints. These three matrices are used to unfold the reconstructed spectrum with the nominal VBF/ggH fraction, and obtain an uncertainty on the unfolded spectrum.

## 4.8 Results

In order to unfold the spectrum, the procedure described in section ?? has been pursued. The statistical plus type A systematic uncertainties are propagated by the unfolding procedure into the final spectrum, taking into account the signal strengths covariance matrix. The type B systematic uncertainty has been propagated using the following procedure: for each  $p_T^H$  bin, we compute the upper bound of the systematic band computing the square sum

**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.

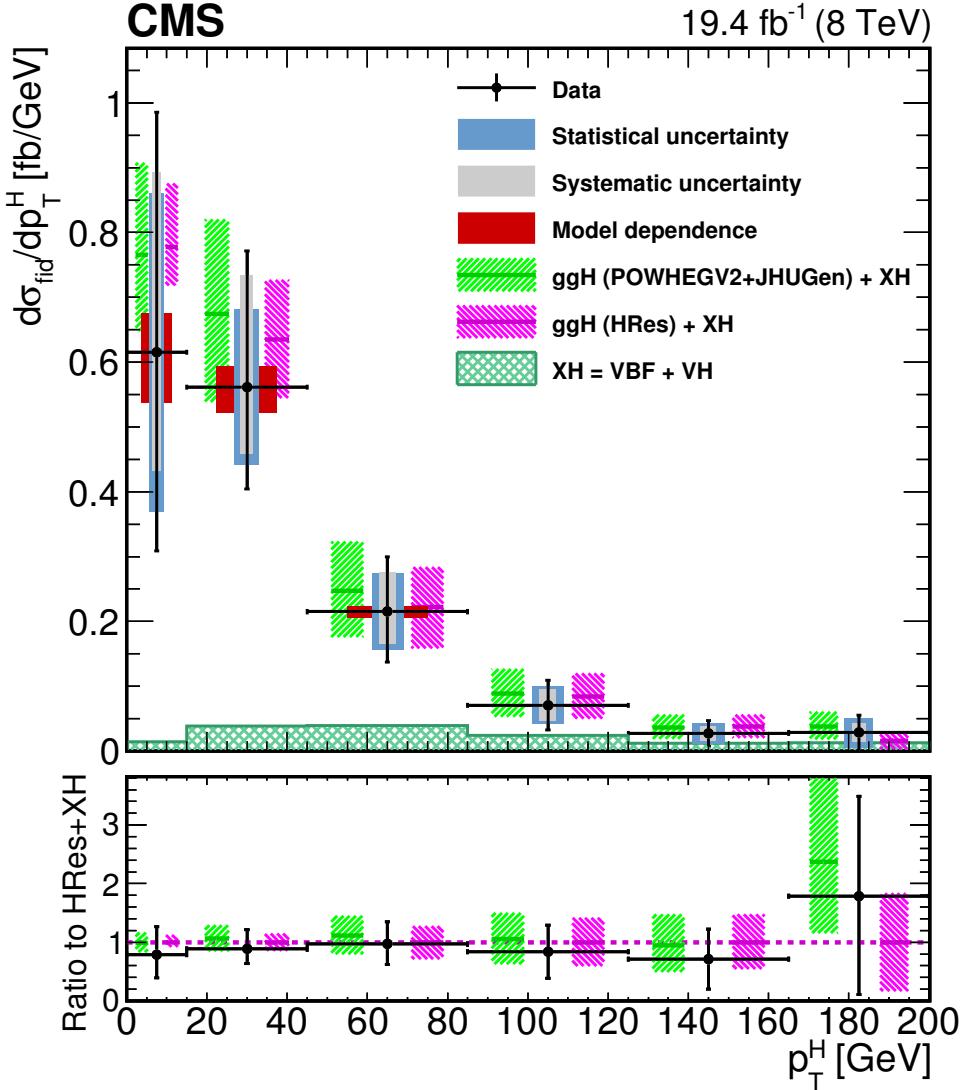
Type B uncertainty	Effect on signal strength ( $+1\sigma/-1\sigma$ [%])					
	[0–15]	[15–45]	[45–85]	[85–125]	[125–165]	[165– $\infty$ ]
b veto	-10.1/-8.8	7.3/12.2	-6.3/3.1	-14.4/-4.8	-5.4/14.5	-7.9/17.8
lepton efficiency	-14.7/-3.9	4.5/15.1	-5.7/2.5	-13.2/-5.3	-0.2/7.6	-0.1/6.8
$E_T^{\text{miss}}$ resolution	-12.5/0.0	15.4/-0.0	-12.8/-0.0	8.7/0.0	-20.9/-0.0	10.5/0.0
$E_T^{\text{miss}}$ 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

**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	$\pm 0.246$	$\pm 0.179$	+0.211/-0.038	+0.0782/-0.0608
15–45	0.561	+0.210/-0.157	$\pm 0.120$	$\pm 0.093$	+0.146/-0.041	+0.0395/-0.0327
45–85	0.215	+0.084/-0.078	$\pm 0.059$	$\pm 0.037$	+0.047/-0.034	+0.0089/-0.0084
85–125	0.071	+0.038/-0.038	$\pm 0.029$	$\pm 0.017$	+0.018/-0.017	+0.0018/-0.0022
125–165	0.027	+0.020/-0.019	$\pm 0.016$	$\pm 0.009$	+0.007/-0.007	+0.0003/-0.0006
165– $\infty$	0.028	+0.027/-0.027	$\pm 0.023$	$\pm 0.012$	+0.008/-0.007	+0.0002/-0.0006

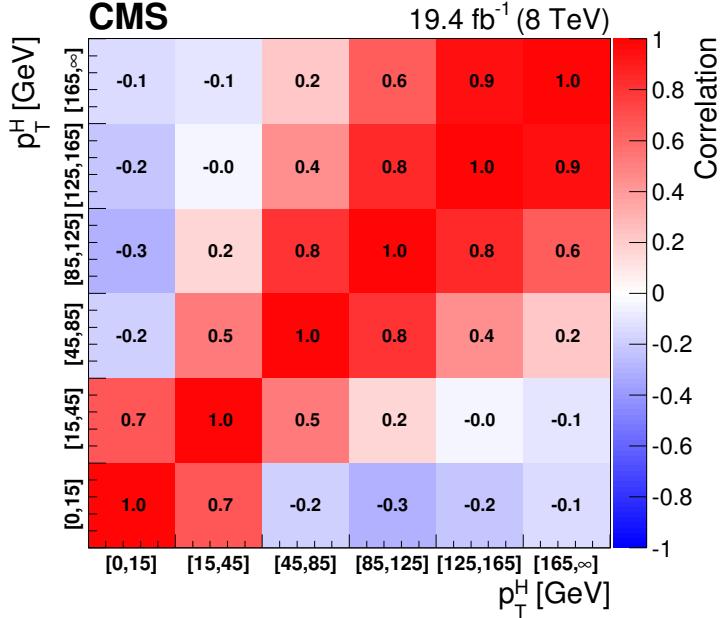
of all the signal strength variations that deviate in the up direction with respect to the bin central value, whether or not this variation corresponds to the up or down shift of the systematic uncertainty. The same is done for the lower bound of the systematic band. If both the up and down shifts of a given nuisance parameter lead to a same direction variation of the signal strength, only the larger variation is considered.

The unfolded  $p_T^H$  spectrum is shown in Fig. ???. Statistical, systematic, and theoretical uncertainties are shown as separate error bands in the plot. The unfolded spectrum is compared with the SM-based theoretical predictions where the ggH contribution is modelled using the HRES and POWHEG V2 programs. The comparison shows good agreement between data and theoretical predictions within the uncertainties. The measured values for the differential cross section in each bin of  $p_T^H$  are reported together with the total uncertainty in Table ??.



**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  $\text{ggH}$  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.

1758      Figure ?? shows the correlation matrix for the six bins of the differential spectrum. The  
 1759      correlation of of bins is defined as in Eq. (??).



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

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

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

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



# Chapter 5

## 1772 **Search for the SM Higgs boson in the 1773 $H \rightarrow WW$ channel with the first 1774 13 TeV LHC data**

### 1775 **5.1 Introduction**

1776 In this chapter, the first search for the SM Higgs boson decaying to a W boson pair at  
1777 13 TeV is presented, using a total integrated luminosity of  $2.3 \text{ fb}^{-1}$ , collected during the  
1778 2015 proton proton data taking period of the LHC.

1779 Final states in which the two W bosons decay leptonically are studied. Therefore,  
1780 events with a pair of oppositely-charged leptons, exactly one electron and one muon, a  
1781 substantial amount of missing transverse energy,  $E_T^{\text{miss}}$ , due to the presence of neutrinos in  
1782 the final state, and either zero or one jet are selected. This signature is common to other  
1783 processes, which enter the analysis as backgrounds. The main background comes from  
1784 WW production, irreducible background that shares the same final states and can only be  
1785 separated by the use of certain kinematic properties. Another important background is  
1786 W+jets, where a jet can mimick a leptonic signature. Background coming from top quark  
1787 events, i.e.  $t\bar{t}$  and single top production, is also important, followed by other processes such  
1788 as Drell-Yan, WZ, and other EWK production. The analysis strategy follows the one used  
1789 during Run 1 in the same channel, described in Chapter ??, with a few different aspects  
1790 that are described in the next sections.

1791 With respect to 8 TeV, the ggH production cross section at 13 TeV is expected to  
1792 increase of a factor of 2, thus raising the number of expected signal events. In addition, the  
1793 cross section for the background processes is increasing as well. The WW production cross  
1794 section increases of a factor of 1.8 and the  $t\bar{t}$  cross section of a factor of 3.5, due to the  
1795 enhancement of the gluon PDFs at higher center of mass energies.

### 1796 **5.2 Data and simulated samples**

1797 Data recorded in proton proton collisions at 13 TeV during 2015 was used in the analysis,  
1798 with a total integrated luminosity of  $2.3 \text{ fb}^{-1}$ . Single and double lepton triggers are used

<sup>1799</sup> similarly to the same analysis at 8 TeV. The HLT paths and descriptions of the triggers  
<sup>1800</sup> 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_WPLoose_Gsf_v*	Single Electron trigger. Best trigger to be used for 2015 data. In HWW, we are using “Trigger safe” Id. Turn on is at around Ele $p_T = 30$ GeV
HLT_Ele17_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*	Double Electron Trigger. Best trigger to cover the turn on region from single electron trigger. “DZ” filter is also present. Its efficiency is also calculated separately.
HLT_Ele12_CaloIdL_TrackIdL_IsoVL_v*	This electron leg of HLT_Mu17_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v* same as Ele12 leg of double electron trigger.
HLT_Ele17_CaloIdL_TrackIdL_IsoVL_v*	This electron leg of HLT_Mu8_TrkIsoVVL_Ele17_CaloIdL_TrackIdL_IsoVL_v* same as Ele17 leg of double electron trigger.

<sup>1801</sup> The trigger efficiencies are measured in data and applied on simulated events as described  
<sup>1802</sup> in Sec. ??.

<sup>1803</sup> Concerning the simulated samples, several different Monte Carlo (MC) generators were  
<sup>1804</sup> used. In the simulation, ‘lepton’ includes also  $\tau$ . Higgs signal samples have been simulated  
<sup>1805</sup> in all channels with POWHEG v2 [44, 57, 84], designed to describe the full NLO properties  
<sup>1806</sup> of these processes. In particular, for Higgs produced via gluon fusion [46], and vector-boson-  
<sup>1807</sup> fusion (VBF) [47], the decay of the Higgs boson into two W boson and subsequently into  
<sup>1808</sup> leptons was done using JHUGEN v5.2.5 [85]. For associated production with a vector boson  
<sup>1809</sup> ( $W^+H$ ,  $W^-H$ ,  $ZH$ ) [86], including gluon fusion produced  $ZH$  ( $ggZH$ ), the Higgs decay was  
<sup>1810</sup> done via PYTHIA 8.1 [66]. Alternative signal samples were produced with AMC@NLO [49],  
<sup>1811</sup> or with POWHEG v2 but decayed via PYTHIA 8.1 for gluon fusion and VBF assuming a  
<sup>1812</sup> Higgs boson mass of 125 GeV. In the following, the mass of the SM Higgs boson is assumed  
<sup>1813</sup> to be 125 GeV.

<sup>1814</sup> The  $WW$  production, irreducible background for the analysis, was simulated in different  
<sup>1815</sup> ways. POWHEG v2 [87] was used for  $q\bar{q}$  produced  $WW$  in different decays. The cross  
<sup>1816</sup> section used for normalizing  $WW$  processes produced via  $q\bar{q}$  was computed at next-to-next-  
<sup>1817</sup> to-leading order (NNLO) [88]. In order to control the top quark background processes, the  
<sup>1818</sup> analysis is performed with events that have no more than one high- $p_T$  jet. The veto on  
<sup>1819</sup> high- $p_T$  jets enhances the importance of logarithms of the jet  $p_T$ , spoiling the convergence  
<sup>1820</sup> of fixed-order calculations of the  $qq \rightarrow WW$  process and requiring the use of dedicated  
<sup>1821</sup> resummation techniques for an accurate prediction of differential distributions [89, 90]. The  
<sup>1822</sup>  $p_T$  of the jets produced in association with the  $WW$  system is strongly correlated with its  
<sup>1823</sup> transverse momentum,  $p_T^{WW}$ , especially in the case where only one jet is produced. The

**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%

<sup>1824</sup> simulated  $q\bar{q} \rightarrow WW$  events are reweighted to reproduce the  $p_T^{WW}$  distribution from the  
<sup>1825</sup>  $p_T$ -resummed calculation.

<sup>1826</sup> Gluon fusion produced  $WW$  was generated, with and without Higgs diagrams, using  
<sup>1827</sup> MCFM v7.0 [91]. A  $t\bar{t}$  sample dilepton sample was also generated using POWHEG v2. The  
<sup>1828</sup>  $WW$  and  $t\bar{t}$  samples produced specifically for this analysis are presented in Table ???. Other  
<sup>1829</sup> background samples are used, a list of the most relevant ones is presented in Table ???.

**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

<sup>1830</sup> All processes are generated using the NNPDF2.3 [92, 93] parton distribution functions  
<sup>1831</sup> (PDF) for NLO generators, while the LO version of the same PDF is used for LO generators.  
<sup>1832</sup> All the event generators are interfaced to PYTHIA 8.1 [66] for the showering of partons

**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

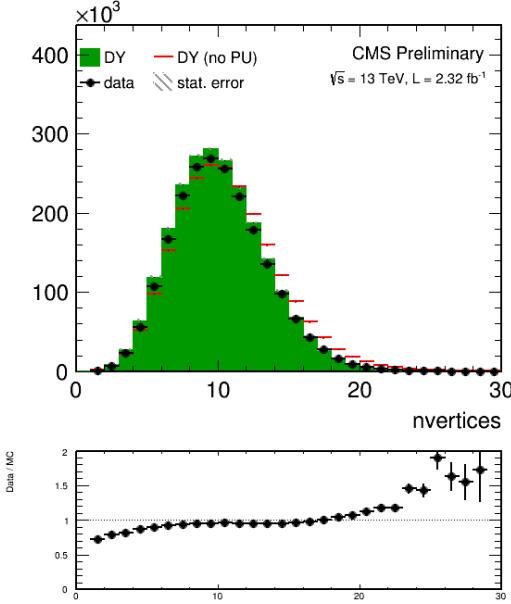
and hadronization, as well as including a simulation of the underlying event (UE) and multiple interaction (MPI) based on the CUET8PM1 tune [94]. To estimate the systematic uncertainties related to the choice of UE and MPI tune, the signal processes and the  $WW$  events are also generated with two alternative tunes which are representative of the errors on the tuning parameters. The showering and hadronization systematic uncertainty is estimated by interfacing the same MC samples with the HERWIG ++ 2.7 parton shower [95, 96]. For all processes, the detector response is simulated using a detailed description of the CMS detector, based on the GEANT4 package [58].

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

The average number of pileup is approximately 11.5.

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 [97], 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. [40].

The cross section used for normalizing  $q\bar{q}$  produced  $WW$  processes was computed at next-to-next-to-leading order (NNLO) [88]. 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 [98]. For



**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.

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 [99] at NLO. The  $t\bar{t}$  cross section is also provided by the LHC Top Working group [100], and it is computed at NNLO, with NNLL soft gluon resummation.

Drell-Yan (DY) production of  $Z/\gamma^*$  is generated using AMC@NLO [49]. Other multi-boson processes, such as  $WZ, ZZ$ , and  $VVV$  ( $V=W/Z$ ), are generated with AMC@NLO and normalized to the cross section obtained at NLO in generation.

All processes are generated using the NNPDF2.3 [92, 93] parton distribution functions (PDF) for NLO generators, while the LO version of the same PDF is used for LO generators. All the event generators are interfaced to PYTHIA 8.1 for the showering of partons and hadronization, as well as including a simulation of the underlying event (UE) and multiple interaction (MPI) based on the CUET8PM1 tune [94].

## 1875 5.3 Analysis strategy

### 1876 5.3.1 Event reconstruction

1877 Regarding the electrons, muons, jets and  $E_T^{\text{miss}}$  definition and reconstruction, the standard  
 1878 CMS recommendations described in Chapter 2 are used. The specific selections used in  
 1879 this analysis are briefly summarised below.

1880 Muons are identified according to the CMS recommendations for the medium working  
 1881 point, with the addition of some extra cuts, as defined by the following selections:

- 1882 • identified by the standard medium muon selection described in Sec. 2.4; **Not yet**  
 1883 **defined :)**
- 1884 •  $p_T > 10 \text{ GeV}$ ;
- 1885 •  $|\eta| < 2.4$ ;
- 1886 •  $|d_{xy}| < 0.01 \text{ cm}$  for  $p_T < 20 \text{ GeV}$  and  $|d_{xy}| < 0.02 \text{ cm}$  for  $p_T > 20 \text{ GeV}$ ,  $d_{xy}$  being the  
 1887 transverse impact parameter with respect to the primary vertex;
- 1888 •  $|d_z| < 0.1 \text{ cm}$ , where  $d_z$  is the longitudinal distance of the muon track in the tracker  
 1889 extrapolated along the beam direction.

1890 For the muon isolation, the CMS recommended particle flow isolation based on the  
 1891 tight working point is used, corresponding to a requirement on the isolation variable of  
 1892  $ISO_{\text{tight}} < 0.15$ . In addition a tracker relative isolation is also applied.

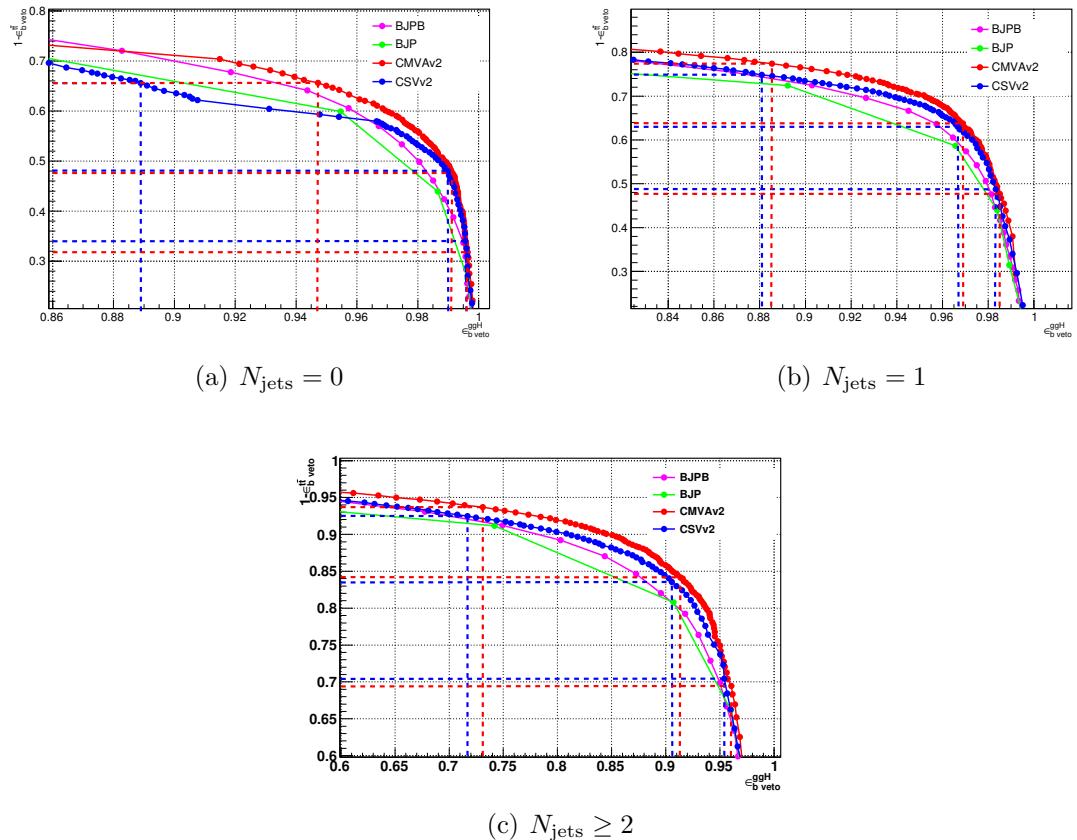
1893 For the electron identification, the tight working point is used. In addition some  
 1894 additional cuts to make the selection “trigger-safe” are included. This is done because the  
 1895 electron triggers already include some identification and isolation requirements that are  
 1896 based on the raw detector information, while the offline selections make use of particle flow  
 1897 requirements. The “trigger-safe” selections are defined to make the the offline identification  
 1898 and isolation requirements tighter with respect to the online triggers.

1899 The simulated events are corrected for the lepton trigger, identification and isolation  
 1900 efficiencies measured in data using the same techniques described in Sec. ??.

1901 Jets are defined clustering the particle flow objects using the anti- $k_t$  algorithm with a  
 1902 distance parameter of 0.4. The CHS pileup mitigation technique is used. The L1, L2, L3  
 1903 and L2L3 jet energy correction described in Sec. 2.4 are applied. The reject jets coming from  
 1904 calorimeter or readout electronics noise, the loose working point for PF jet identification is  
 1905 used.

1906 The b-tagging algorithm for this analysis is chosen comparing the performances of  
 1907 different algorithms using simulations for signal and background contributions in the phase  
 1908 space defined by the analysis kinematic requirements. More precisely, two MC samples are  
 1909 used, one corresponding the the  $H \rightarrow WW \rightarrow 2\ell 2\nu$  signal produced via the ggH production  
 1910 mode and another corresponding to the  $t\bar{t}$  process. In fact, the first sample is enriched  
 1911 in light jets, i.e. originating by the hadronization of light quarks like u,d,c and s quarks,  
 1912 while the second sample is enriched in b jets, coming from the top quark decay. The b-veto  
 1913 efficiency,  $\epsilon_{\text{bveto}}$ , is computed separately for the two samples and for the various b tagging

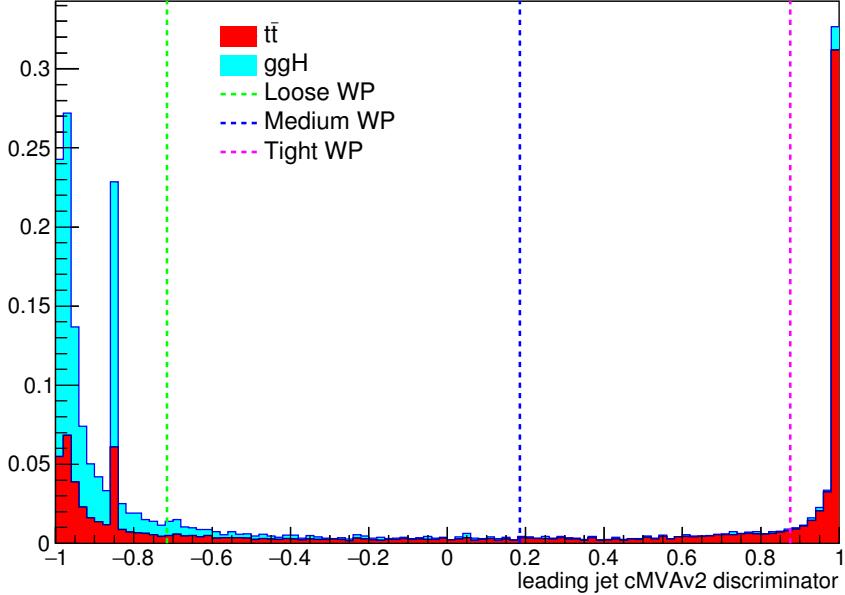
algorithms. To compare the b tagging performance  $\epsilon_{b\text{veto}}$  is computed for different working points, i.e. different selections on the specific b tagging discriminator, and the results are reported in the form of a ROC curve. The ROC curves corresponding to events with 0, 1 and  $\geq 2$  jets are shown in Fig. ???. Events considered for this study are the ones passing the WW baseline selection.



**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.

The ROC curves show that the cMVAv2 algorithm has the best performance for the analysis phase space among the algorithms taken into account. For both the CSVv2 and cMVAv2 algorithms, three working points are defined corresponding to the mistag rates<sup>1</sup> of 10% for the loose, 1% for the medium and 0.1% for the tight working point. The distribution of the cMVAv2 discriminator associated to the leading jet both for the ggH and the t̄t MC sample is shown in figure ??.

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



**Figure 5.3:** 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.

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

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

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

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

**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

1942 The two categories were eventually combined together and the significance assessment was  
1943 repeated for the three working points. With these selection we find the significance values  
1944 listed in Table ?? for the three working points.

1945 The working point providing the best significance in the combined 0 + 1 jets category is  
1946 found to be the loose one.

1947 To correct for a possible different b tagging efficiency in data and simulation, the  
1948 simulated events are reweighted using scale factors computed in bins of the jet  $\eta$  and  $p_T$ .  
1949 These scale factors and the corresponding uncertainties are centrally calculated for each  
1950 working point, in such a way to be employable by all the CMS analyses. The prescription  
1951 to reweight the simulated events is the following. First of all one has to compute the b  
1952 tagging efficiency using the MC samples,  $\varepsilon_{\text{MC}}(p_T, \eta, f)$ , for the chosen working point in bins  
1953 of jet  $p_T$  and  $\eta$ . The efficiency has to be computed for different flavours  $f$  of the jets, b,  
1954 c and light (u,d,s), using the jet matching information<sup>2</sup> which is available in all the MC  
1955 samples. An MC-based event weight is then calculated computing the probability  $P_{\text{MC}}$  of a  
1956 given b tagging configuration to occur, e.g.:

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

1957 Afterwards, a similar probability is computed using data:

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

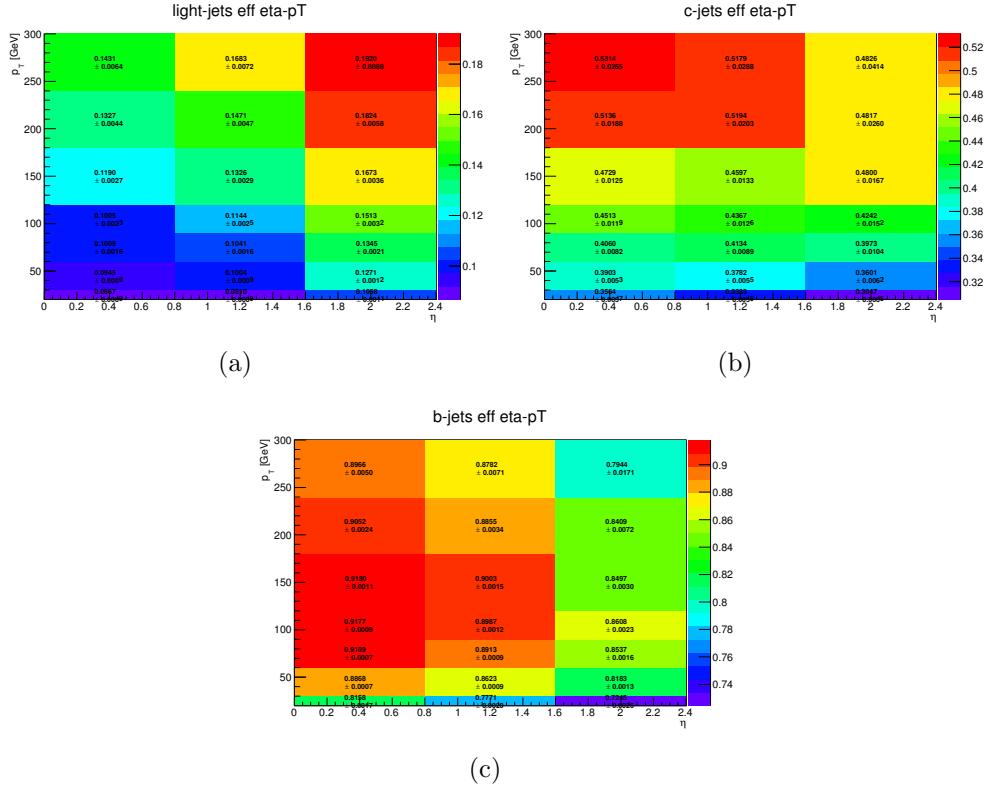
1958 where  $SF_i$  is the provided scale factor value for the relevant jet flavour,  $p_T$  and  $\eta$ . Products in  
1959 Eqs. ?? and ?? run over all jets. The event weight is finally given by the ration  $P_{\text{DATA}}/P_{\text{MC}}$ .

1960 The b tagging efficiencies to be fed into Eq. ?? and Eq. ?? are derived using  $t\bar{t}$  simulated  
1961 events and applying basic leptonic selections. These efficiencies are shown in Fig. ?? for  
1962 light (a), c-jets (b) and b-jets (c), in bins of  $\eta$  and  $p_T$ . The uncertainties associated to the

---

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

efficiencies are representative of the statistics of the simulated  $t\bar{t}$  sample, and are computed according to a binomial distribution.

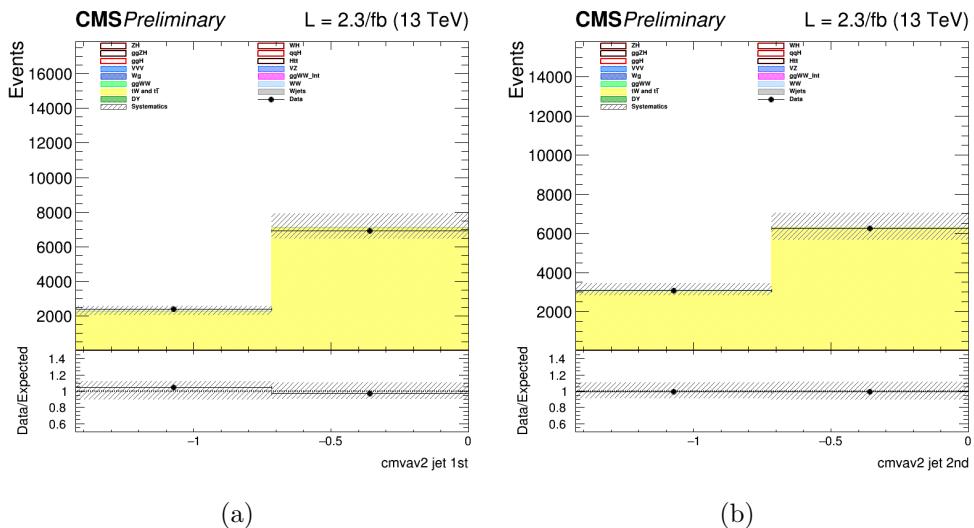


**Figure 5.4:** B tagging efficiencies for light jets (a), c-jets (b) and b jets (c), as a function of  $\eta$  and  $p_T$ .

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

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

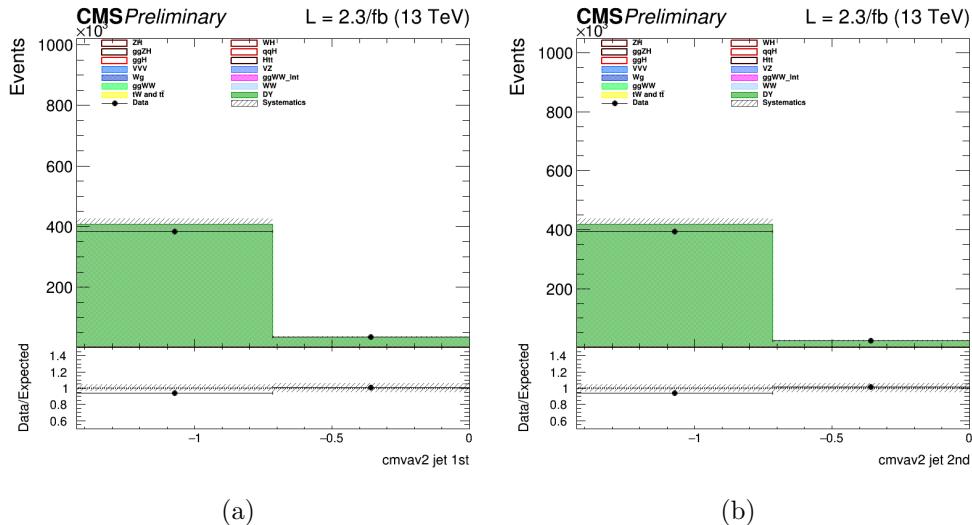
- In order to evaluate the agreement for light jets, a second control region is defined, populated by Z+light jet events, defined as follows:
- two leptons, two electrons or two muons with opposite charge, with leading lepton  $p_T$  greater than 20 GeV and sub-leading lepton  $p_T$  greater than 15 GeV.
  - no other lepton (electron or muon) with  $p_T$  greater than 10 GeV.
  - lepton invariant mass greater between 80 GeV and 110 GeV.
  - at least two jets with  $p_T$  greater than 30 GeV.
  - at least one jet above 30 GeV.
  - no jets above 20 GeV with a TCHE score above 2.1.
- Although a Z+jets sample is dominated by light flavor jets, a b-veto on an alternative algorithm (TCHE) is applied to reduce the contamination from b-jets, especially above the cMVAv2 cut. This helps mitigating possible data/simulation discrepancies in the modeling of the heavy/light flavour ratio. The comparison between data and simulation after the event reweighting is shown in Figs. ?? and ?? for the b-jets and light jets enriched control 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.

### 5.3.2 Event selection and background rejection

Since the ggH production mechanism, which is the main production mode for a Higgs mass of around 125 GeV, is characterized by the emission of few jets arising from initial or final state radiation, this analysis is limited to events with no jets or one jet. Due to the large DY background in di-electrons and di-muons events, only the  $e\mu$  final state is



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

studied in this early Run 2 data analysis, including the indirect contribution from  $\tau$  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 $\gamma$ , W $\gamma$ , 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  $\tau$  lepton pair which subsequently decays to an  $e\mu$  final state and suppress processes without genuine  $E_T^{\text{miss}}$ , a minimal  $E_T^{\text{miss}}$  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  $p_T > 30 \text{ GeV}$ . 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,

2022 such as VBF, are not included in this analysis, given the very low expected yield for other  
2023 production modes with the analysed integrated luminosity.

2024 Distributions of some variables of interest for the 0 and 1 jet categories separately,  
2025 but merging the  $e\mu$  and  $\mu e$  final states together, are shown in Figs. ??, ?? and ?? after  
2026 applying the WW baseline selections, with the addition of a cut on  $m_{\ell\ell}$  to remove the Higgs  
2027 signal contribution ( $m_{\ell\ell} > 80$  GeV), and a cut on  $m_T$  to be orthogonal to the  $Z\gamma^* \rightarrow \tau\tau$   
2028 background control region ( $m_T > 60$  GeV).

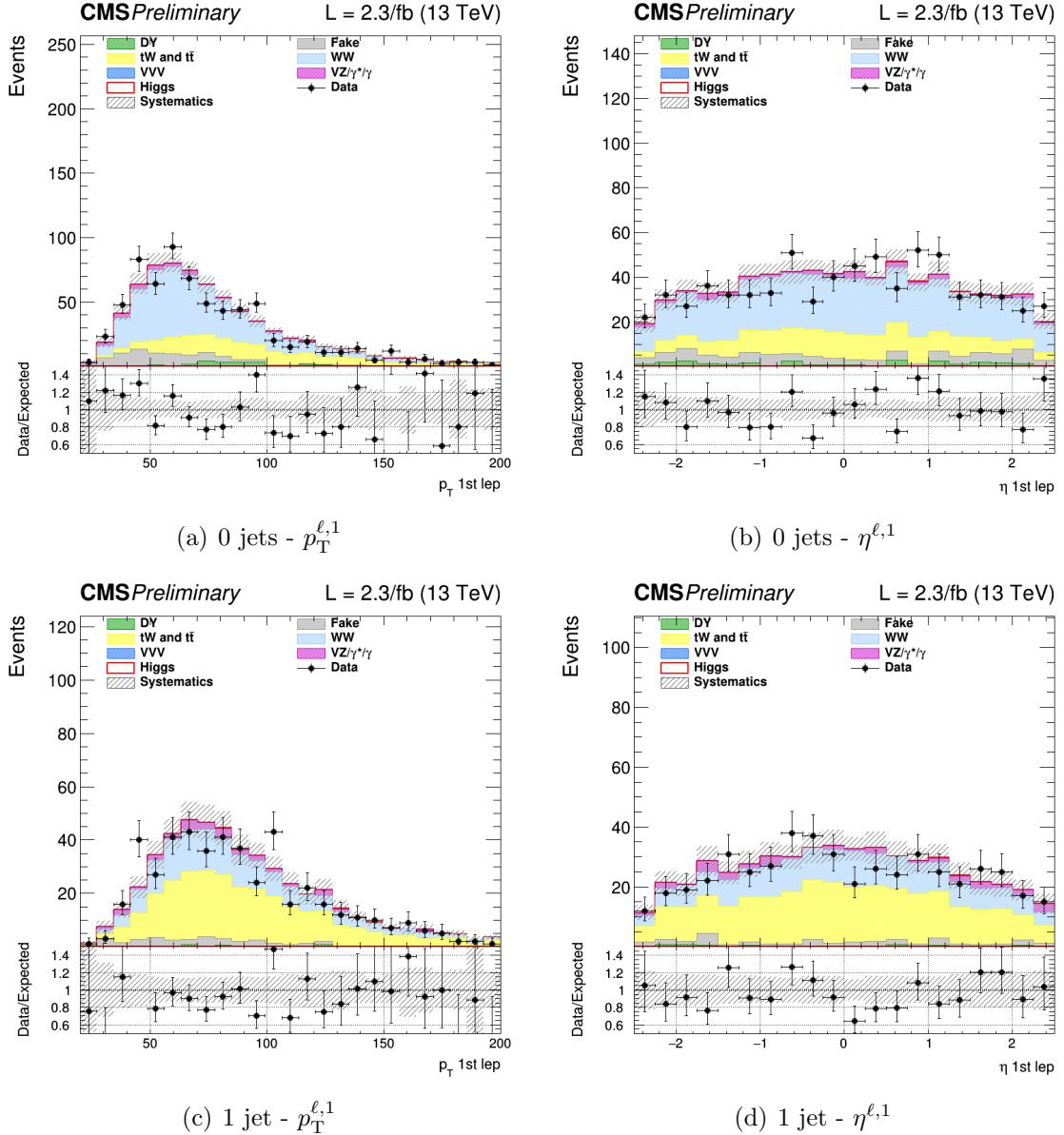
2029 The  $W+jets$  background, where one jet can be misidentified as a lepton, is a sub-  
2030 dominant background in the phase space defined by the analysis kinematic requirements.  
2031 The 0 ad 1 jets categories are further split according to the lepton flavour to  $e\mu$  and  $\mu e$ ,  
2032 where the first lepton refers to the leading one. In this way an improvement of about 10% in  
2033 terms of the signal significance can be achieved, exploiting the different  $W+jets$  background  
2034 contribution in the two categories. Indeed the probability for a jet to be misidentified as an  
2035 electron or a muon is not the same.

### 2036 5.3.3 Signal extraction

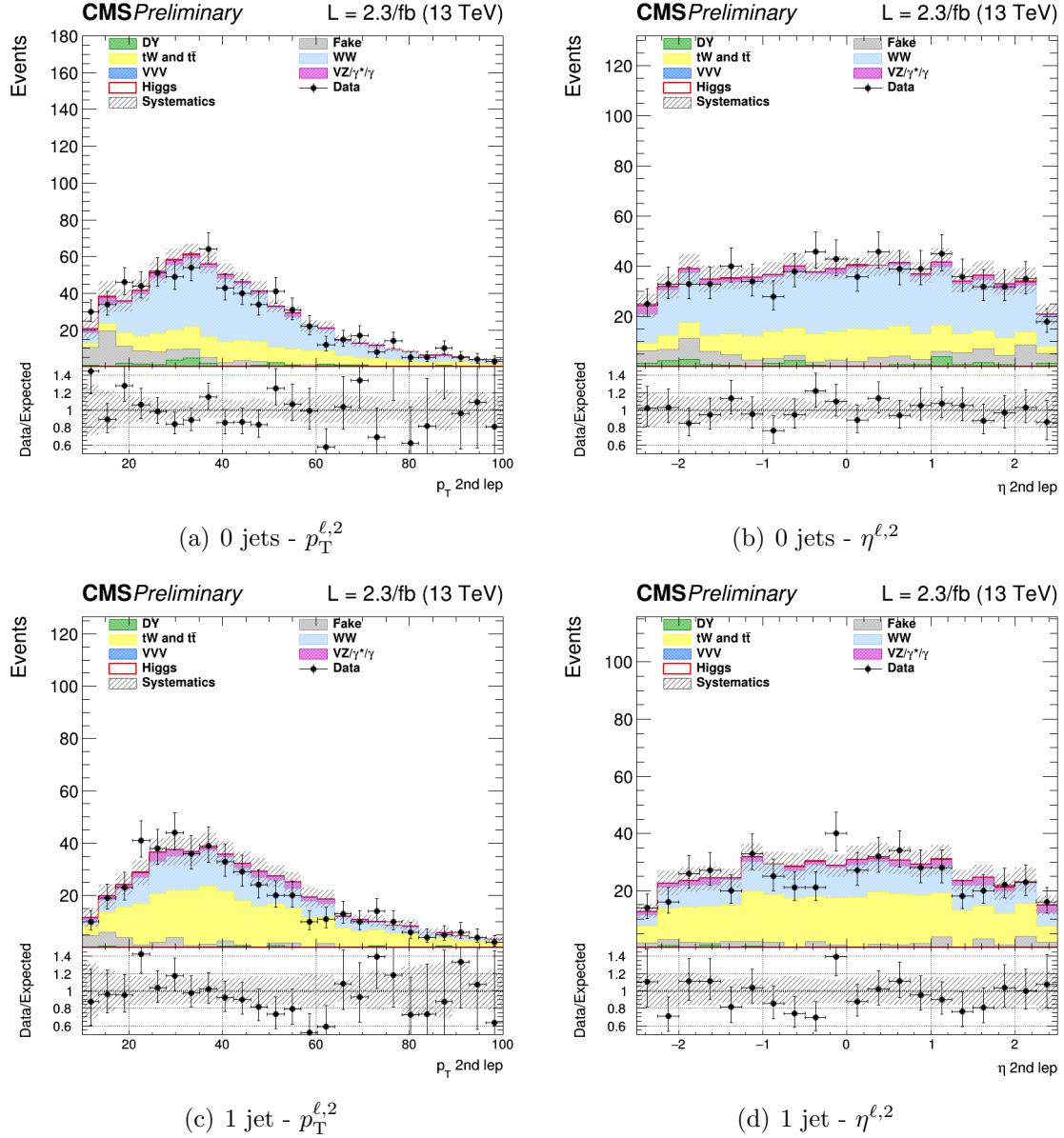
2037 To extract the Higgs boson signal contribution in the four previously mentioned categories,  
2038 a similar approach to the one used in the Run 1 analysis [41] is pursued. The analysis is  
2039 based on two-dimensional templates of  $m_{\ell\ell}$  versus  $m_T$  to discriminate signal and background  
2040 contributions. The  $m_{\ell\ell}$  template is defined using 5 bins from  $m_{\ell\ell} = 10$  GeV up to  $m_{\ell\ell} =$   
2041 110 GeV, while for the  $m_T$  template 7 bins are defined in the range  $60$  GeV  $< m_T < 200$  GeV.  
2042 The phase space with  $m_T < 60$  GeV is used as an orthogonal control region to extract the  
2043 normalization of the DY background. A binned maximum likelihood fit to the signal and  
2044 background two-dimensional templates is performed to extract the signal strength in the  
2045 four categories.

2046 Distributions of the  $m_{\ell\ell}$  and  $m_T$  variables after the WW level selection are shown in  
2047 Fig. ?? for the 0 and 1 jet categories separately, but merging the  $e\mu$  and  $\mu e$  final states  
2048 together.

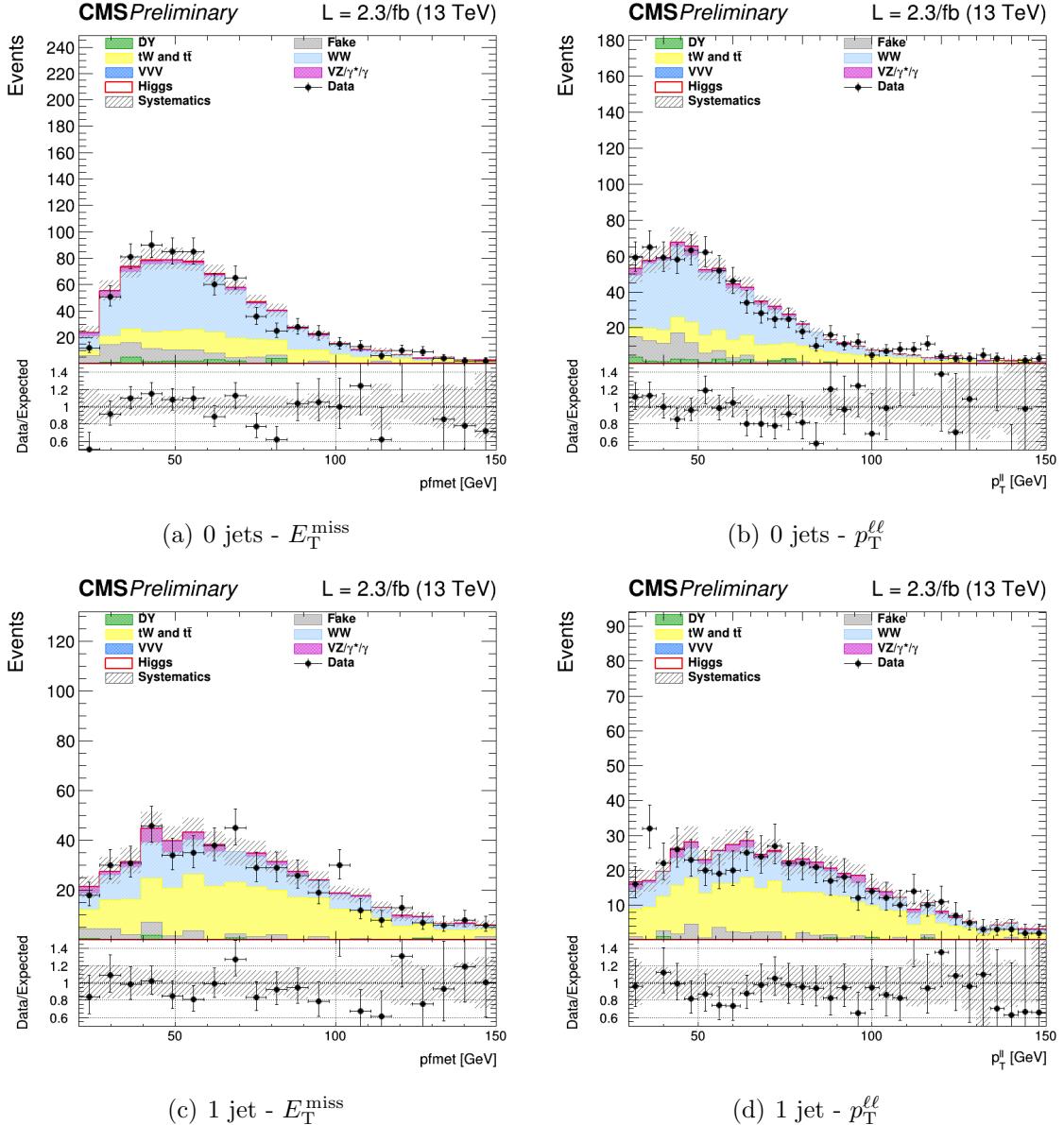
2049 The statistical methodology used to interpret the data and to combine the results from  
2050 the independent 0-jet and 1-jet categories in the  $e\mu$  and  $\mu e$  final states has been developed  
2051 by the ATLAS and CMS collaborations in the context of the LHC Higgs Combination  
2052 Group [83, 101]. The number of events in each category and in each bin of two-dimensional  
2053 template is modelled as a Poisson random variable, with a mean value given by the sum  
2054 of the contributions from all the processes under consideration. Systematic uncertainties  
2055 are represented by individual nuisance parameters with log-normal distributions. The  
2056 uncertainties affect the overall normalization of the signal and backgrounds as well as the  
2057 shape of the predictions across the distribution of the observables. Correlation between  
2058 systematic uncertainties in different categories are taken into account.



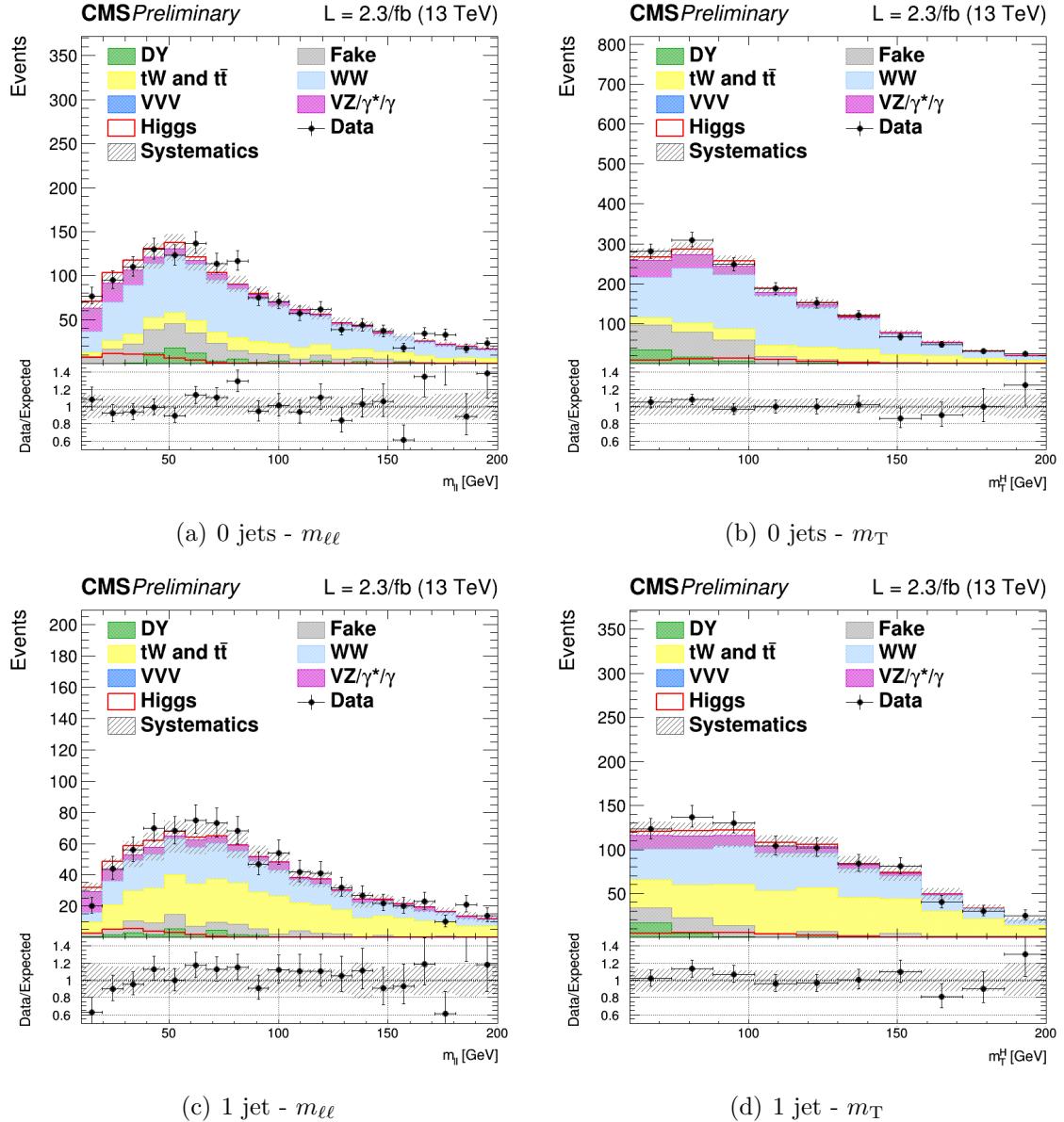
**Figure 5.7:** Distributions of  $p_T$  (left) and  $\eta$  (right) of the leading lepton for events with 0 jet (upper row) and 1 jet (lower row), for the main backgrounds (stacked histograms), and for a SM Higgs boson signal with  $m_H = 125$  GeV (superimposed and stacked red histogram) at the WW 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  $\eta$  (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  $m_H = 125$  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  $E_T^{\text{miss}}$  (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  $m_H = 125 \text{ GeV}$  (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 5.10:** 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  $m_H = 125 \text{ GeV}$  (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.

## 2059 5.4 Background estimation

2060 The main background processes affecting the analysis signature, non-resonant WW production  
 2061 and top quark processes, are estimated using data. Backgrounds arising from an  
 2062 experimental misidentification of the objects, such as W+jets (also called “Fake”), are  
 2063 estimated using data as well. The other minor backgrounds are generally estimated directly  
 2064 from simulation as described in the following subsections.

### 2065 5.4.1 WW background

2066 The quark-induced WW background is simulated with NLO accuracy in perturbative  
 2067 QCD, and the transverse momentum of the diboson system is reweighted to match the  
 2068 NNLO+NNLL accuracy from theoretical calculations [89, 90]. However, given the large  
 2069 uncertainties on the jet multiplicity distribution associated to this process, the normalization  
 2070 of this background is measured from data separately for the 0 and 1 jet categories. The  
 2071 normalization k-factors are extracted directly from the fit together with the signal strengths,  
 2072 leaving the WW normalization free to float separately in the two jet multiplicity categories.  
 2073 An orthogonal control region for the WW background normalization estimation is not  
 2074 needed in this case, owing to the different  $m_{\ell\ell}$ - $m_T$  shape for signal and background.

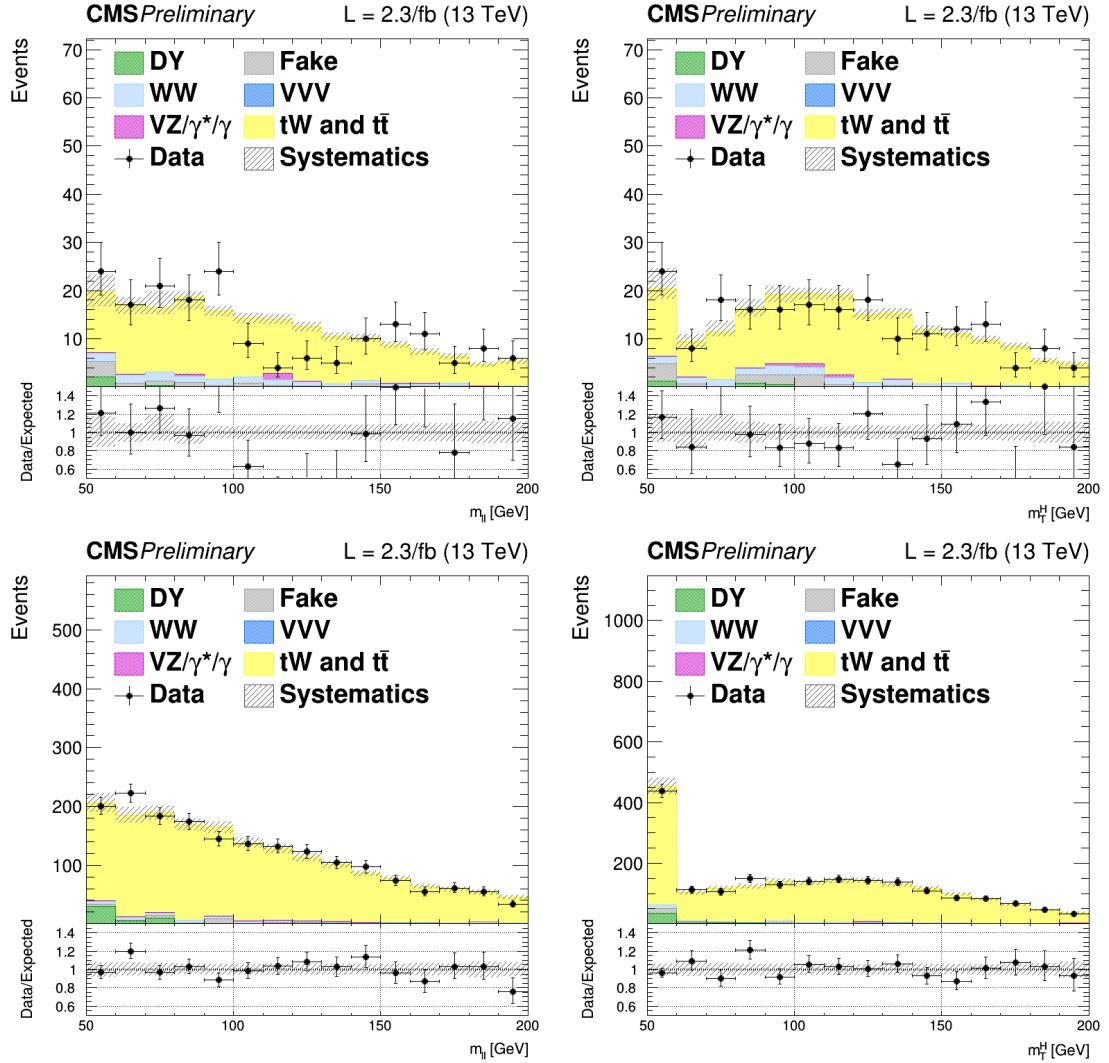
2075 The gluon-induced WW production is sub-dominant with respect to the quark-induced  
 2076 production, and its shape and normalization is fully taken from simulation, scaling the  
 2077 cross section to the theoretical prediction with NLO accuracy [98].

### 2078 5.4.2 Top quark background

2079 As explained in Sec. ??, the production of top quark pairs represents one of the dominant  
 2080 backgrounds in this analysis given its large cross section and a similar final state compared  
 2081 to the signal. A b-jet veto, based on the *cMVAv2* b tagging algorithm, is used to suppress  
 2082 this background and a reweighting procedure is applied on top of the simulated events to  
 2083 correct for different b tagging efficiency in data and simulation.

2084 The top quark background normalization is measured using data, defining a b-jets  
 2085 enriched control region by inverting the b-jet veto. More precisely, the b-jets enriched  
 2086 control region for the 0-jet category is defined with the same WW baseline selection but  
 2087 requiring at least one jet with  $20 < p_T < 30$  GeV to be identified as a b jet and no other  
 2088 jets with  $p_T > 30$  GeV. For the 1-jet category, the b-jets enriched region is defined requiring  
 2089 exactly one jet with  $p_T > 30$  GeV identified as a b-jet. To reduce other backgrounds in  
 2090 these two regions, the dilepton mass has to be greater than 50 GeV. Distributions of the  
 2091  $m_{\ell\ell}$  and  $m_T$  variables in the b-jets enriched control regions after applying the data driven  
 2092 estimation are shown in Figure ??, for the 0 and 1 jet categories separately.

2093 The top quark background normalization is constrained during the fit procedure sepa-  
 2094 rately in the two jet categories, by means of the control regions defined above, which are  
 2095 treated in the fit as two additional categories.

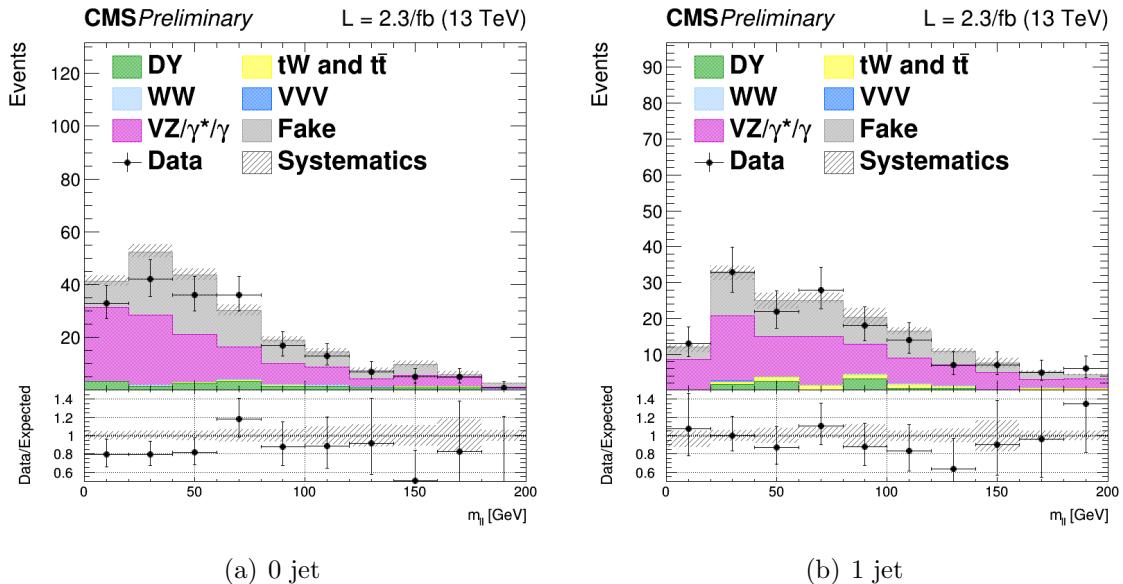


**Figure 5.11:** Distributions of  $m_{\ell\ell}$  (left) and  $m_T$  (right) for events with 0 jet (top) and 1 jet (bottom) in top enriched phase space. Scale factors estimated from data are applied. The first (last) bin includes underflows (overflows).

### 2096 5.4.3 Jet-induced (or Fake) background

2097 One of the primary source belonging to this category arises from the misidentification of  
 2098 leptons in  $W+jets$  processes in the 0 jet category. Also, semileptonic  $t\bar{t}$  decays contribute  
 2099 especially for higher jet multiplicities. Multijet production and hadronic  $t\bar{t}$  decays are also  
 2100 taken into account, but have a much smaller contribution.

2101 This background is fully estimated using data, with the technique described in Sec. ??.  
 2102 To check the agreement of the background estimated in this way with data, a control sample  
 2103 enriched in jet-induced events is defined. The events in the control sample are selected  
 2104 applying the  $WW$  baseline requirements but requesting an  $e\mu$  pair with same charge, which  
 2105 significantly suppresses the  $WW$  and  $t\bar{t}$  processes. The  $m_{\ell\ell}$  distributions in this control  
 2106 region for the 0 and 1 jet categories are shown in Fig. ???. From the crosscheck in this  
 2107 control region, a global normalization factor of 0.8 is derived and applied to the jet-induced  
 2108 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.

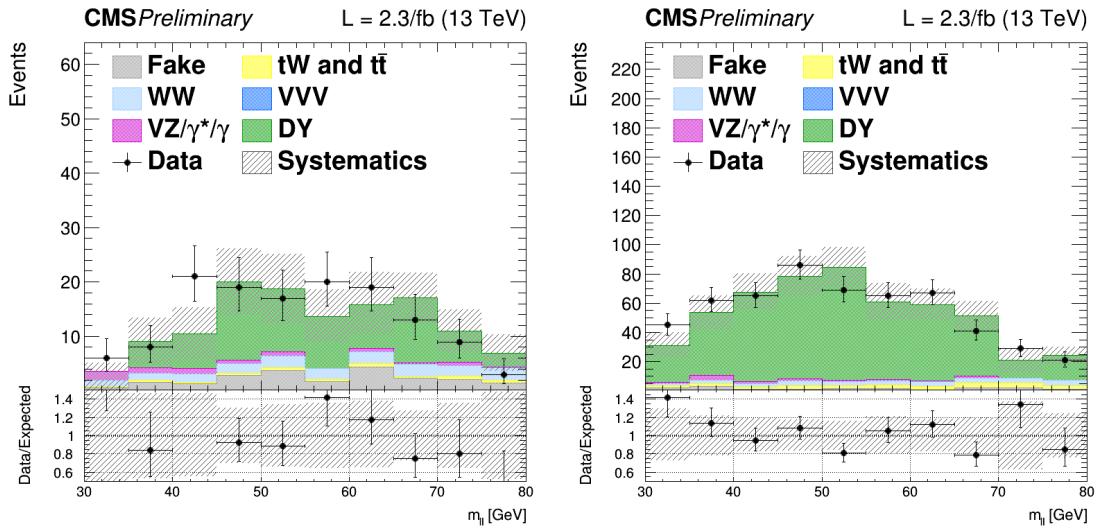
### 2109 5.4.4 DY background

2110 This background contributes to the analysis phase space because of the  $Z/\gamma^*$  decays to a  
 2111 pair of  $\tau$  leptons, which consequently decays to an  $e\mu$  pair. This background process is  
 2112 predominant in the low  $m_T$  region, which is used as an orthogonal control region to determine  
 2113 the background normalization in the 0 and 1 jet categories separately. In particular this  
 2114 control region is defined by selecting events with  $m_T < 60$  GeV and  $30$  GeV  $< m_{\ell\ell} < 80$  GeV.

2115 The  $m_{\ell\ell}$  distributions in these control regions for the 0 and 1 jet categories are shown in  
 2116 Fig. ??.

2117 As for the top quark background, the normalization of this background in the 0 and 1  
 2118 jet categories, is constrained directly in the fit by means of the control regions, which are  
 2119 treated as two additional categories.

2120 The kinematics of this background is taken from simulation, after reweighting the Z  
 2121 boson  $p_T$  spectrum to match the observed distribution measured in data. In fact, this  
 2122 variable is not well reproduced by the MC generator used for simulating this process,  
 2123 especially in the bulk of the distribution, the discrepancy being ascribed to the missing  
 2124 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.

#### 2125 5.4.5 Other backgrounds

2126 The  $W\gamma^*$  and the  $WZ$  electroweak processes can be gathered in the same physical process,  
 2127 although the final state kinematics is rather different. In particular, the invariant mass of  
 2128 the leptons arising from the  $\gamma^*$  decays is generally below 4 GeV, while the leptons from the  
 2129 Z boson decay are characterized by a larger invariant mass. Another background which can  
 2130 be experimentally identical to those is the  $W\gamma$  production, where a real photon is produced  
 2131 in association with a W boson and consequently undergoes a photon conversion to leptons  
 2132 due to the interaction with the material constituting the first layers of the silicon tracker.

2133 All these backgrounds may contribute to the signal phase space whenever one of the three  
 2134 leptons escape from the detector acceptance or is not identified. The shape and cross section  
 2135 of these backgrounds are taken from simulation. The only exception is the normalization of  
 2136 the  $W\gamma^*$  background, being this process dominant in the low  $m_{\ell\ell}$  region, which is scaled

2137 to data defining a proper control region. The control region is defined selecting events  
2138 with three isolated muons, with  $p_T > 10,5$  and  $3\text{ GeV}$  for the first three leading muons  
2139 respectively. The selection is further defined by  $E_T^{\text{miss}} < 25\text{ GeV}$  and  $E_T^{\text{miss}}$  projected to the  
2140 leading muon  $< 45\text{ GeV}$ . The pair of muons with the smallest invariant mass is taken as  
2141 coming from the  $\gamma^*$  decay. The k-factor measured in data for this background to be applied  
2142 in the simulation is  $1.98 \pm 0.54$ .

2143 All remaining backgrounds from di-boson and tri-boson production, which are of mi-  
2144 nor importance in the analysis phase space, are normalized according to their expected  
2145 theoretical cross sections.

## 2146 5.5 Systematic uncertainties

2147 The systematic uncertainties affecting this measurement can be divided into three categories:  
2148 the uncertainties on the background estimation, experimental uncertainties and theoretical  
2149 uncertainties.

2150 The first category includes the uncertainties related to the background normalization  
2151 and shape. For the non-resonant  $WW$  production the shape is taken from simulation.  
2152 The input normalization to the fit is set to the expected value from simulation, and an  
2153 unconstrained nuisance parameter with a flat distribution is associated to this number.  
2154 This is done separately for the two jet categories.

2155 The top quark background shape is taken from simulation after correcting for the b  
2156 tagging scale factors. An uncertainty due to these scale factors is included and affects  
2157 both the normalization and the shape of the top quark background. The uncertainties on  
2158 the normalization are treated similarly to the  $WW$  background case, but constraining the  
2159 corresponding nuisances by means of the two control regions orthogonal to the signal phase  
2160 space. A similar procedure is used for the DY background.

2161 Effects due to experimental uncertainties are studied by applying a scaling and smearing  
2162 of certain variables related to the physics objects, e.g. the  $p_T$  of the leptons, followed by  
2163 a subsequent recalculation of all the correlated variables. This is done for simulation, to  
2164 account for possible systematic mismodeling.

2165 All experimental sources, except luminosity, are treated both as normalization and shape  
2166 uncertainties, and are correlated among the signal and background processes and all the  
2167 categories. The following experimental uncertainties are considered:

- 2168 • the uncertainty determined by the CMS online luminosity monitoring, 2.7% for the  
2169 first data collected at  $\sqrt{s} = 13\text{ TeV}$ ;
- 2170 • the acceptance uncertainty associated with the combination of single and double lepton  
2171 triggers, which is 2%;
- 2172 • the lepton reconstruction and identification efficiencies uncertainties, that are in the  
2173 range 0.5-5% for electrons and 1-7% for muons depending on  $p_T$  and  $\eta$ ;
- 2174 • the muon momentum and electron energy scale and resolution uncertainties, that  
2175 amount to 0.01-0.5% for electrons and 0.5-1.5% for muons depending on  $p_T$  and  $\eta$ ;

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

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

The effects of the variation of PDFs,  $\alpha_s$  and renormalization/factorization QCD scales, mainly affect the signal processes, being the most important backgrounds estimated using data driven techniques. However, the uncertainties on minor backgrounds that are estimated from simulation are taken into account. These uncertainties are split in the uncertainties on the cross section, which are computed by the LHC cross section working group [102], and on the selection efficiency [103]. The PDFs and  $\alpha_s$  signal cross section normalization uncertainties are  $^{+7.4\%}_{-7.9\%}$  and  $^{+7.1\%}_{-6.0\%}$  for ggH and  $\pm 0.7\%$  and  $\pm 3.2\%$  for VBF Higgs production mechanism. The PDFs and  $\alpha_s$  acceptance uncertainties are less than 1% for gluon- and quark-induced processes. The effect of the QCD scales variation on the selection efficiency is around 1-3% depending on the specific process. To estimate these uncertainties, the events are reweighted according to different QCD scales or different PDF sets and the selection efficiency is recomputed each time. For the QCD scale uncertainty the maximum variation with respect to the nominal value is taken as the uncertainty. For the case of PDF and  $\alpha_s$  uncertainties, the distribution of the selection efficiency is built taking into account all the replicas in the NNPDF3.0 set and the uncertainty is estimated as the standard deviation of that distribution.

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

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

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

viceversa. The corresponding uncertainty is of 8%, according to the theoretical cross section calculations [99, 100].

The  $gg \rightarrow WW$  background LO cross section predicted by the MCFM generator is scaled to the NLO calculation, applying a k-factor of 1.4 with an uncertainty of 15% [98]. The interference term between the  $gg \rightarrow WW$  and the  $ggH$  signal is also included and simulated with LO accuracy using MCFM. The k-factor to scale the interference term is 1.87, given by the geometrical average of the LO to NNLO  $gg \rightarrow H \rightarrow WW$  scale factor (2.5) and the LO to NLO  $gg \rightarrow WW$  scale factor (1.4). The uncertainty on this value is estimated as the maximum variation with respect to the two scale factors mentioned above, and is found to be of 25%. Anyway, with the current amount of integrated luminosity, the interference contribution is found to be negligible.

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

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

## 5.6 Results

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

Maybe I should add the nuisances impact plots...

**Table 5.6:** Observed and expected significance and signal strength the SM Higgs boson with a mass of 125 GeV for the 0-jet and 1-jet,  $\mu e$  and  $e\mu$ , categories.

Category	Expected significance	Observed significance	$\sigma/\sigma_{SM}$
0-jet $\mu e$	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 $\mu e$	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

## 2243 **Search for high mass resonances 2244 decaying to a W boson pair with first 2245 13 TeV LHC data**

### 2246 **6.1 Introduction**

2247 In this chapter, a search for a high mass spin-0 particle (from now on denoted as  $X$ )  
2248 in the  $X \rightarrow WW \rightarrow \ell\nu\ell'\nu'$  decay channel is presented, where  $\ell$  and  $\ell'$  refer to an different  
2249 flavour lepton pair, i.e.  $e\mu$ . The search is based upon proton-proton collision data samples  
2250 corresponding to an integrated luminosity of up to  $2.3 \text{ fb}^{-1}$  at  $\sqrt{s} = 13 \text{ TeV}$ , recorded by  
2251 the CMS experiment at the LHC during 2015. This analysis represents a general extension  
2252 of the SM Higgs boson search presented in ?? and is performed in a range of heavy scalar  
2253 masses from  $M_X = 200 \text{ GeV}$  up to  $1 \text{ TeV}$ , extending the range studied in a similar analysis  
2254 performed using Run 1 LHC data [104], which provided upper limits on the production  
2255 cross section of new scalar resonances up to  $600 \text{ GeV}$ .

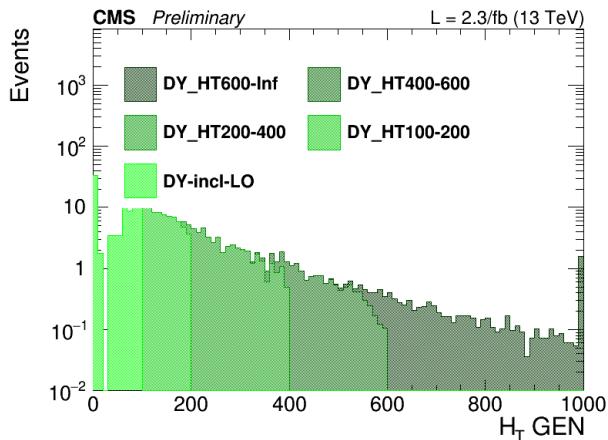
2256 Despite the discovery of a particle consistent with the SM Higgs boson in 2012, there is a  
2257 possibility that this particle is only a part of a larger Higgs sector, and hence only partially  
2258 responsible of the EW symmetry breaking. This can be achieved in different theoretical  
2259 models that extends the SM, such as the two-Higgs-doublet models [105–107], or models in  
2260 which the SM Higgs boson mixes with a heavy EW singlet, which predict the existence  
2261 of an additional resonance at high mass, with couplings similar to those of the SM Higgs  
2262 boson, as most recently described in [108, 109].

2263 This analysis reports a generic search for a scalar particle with different resonance decay  
2264 widths hypothesis, produced via the ggH and VBF production mechanisms. The results  
2265 can then be interpreted in terms of different theoretical models. This analysis is heavily  
2266 based on the SM Higgs search described in ??, in terms of physics objects, selections and  
2267 background estimation. The differences and similarities are discussed in this chapter.

### 2268 **6.2 Data and simulated samples**

2269 The data sets, triggers, pile up reweighting, lepton identification and isolation used in this  
2270 analysis are the same as the SM Higgs search and are described in Sec. ??.

Also, the same MC simulations are used for the background processes, the only exception being the DY background, for which the MG5\_AMC@NLO generator is used with LO QCD accuracy, matching together events with up to four jets in addition to the vector boson with the MLM [110] matching scheme. Given that this analysis aims to probe regions of phase space where the DY contribution is very small, like in the high transverse mass region, the usage of a simulation of the inclusive DY process may lead to large uncertainties due to the limited simulation statistics in the sample. To partially overcome this issue, different DY samples are generated in restricted portions of the phase space defined by the  $H_T$  variable, i.e. the scalar sum of all the partons  $p_T$  in the event. For  $H_T < 100$  GeV the inclusive simulation is used, while different samples are used for higher values of  $H_T$ . The samples are merged using the parton level information, and it has been verified that a smooth transition between different  $H_T$  regions is achieved, as shown in Fig. ???. The DY LO cross section obtained from the simulation is scaled using the LO to NNLO k-factor of 1.23.



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

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

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 [85]. Details about the interference effects are given in Sec. ??.

## 2298 6.3 Analysis strategy

2299 The analysis strategy for the first results on the high mass search in the  $W^+W^- \rightarrow 2\ell 2\nu$   
 2300 decay channel closely follows the strategy presented in the 13 TeV SM Higgs search in the  
 2301  $H \rightarrow W^+W^- \rightarrow 2\ell 2\nu$  channel regarding the 0 and 1 jet categories. In addition a dedicated  
 2302 category to the VBF production mechanism is added, given that this production mode is  
 2303 particularly important in the high mass region. Indeed, assuming a SM Higgs boson, the  
 2304 ratio of cross sections  $\sigma_{\text{VBF}}/\sigma_{\text{ggH}}^1$  increases with the Higgs boson mass, making the VBF  
 2305 production mechanism more and more important as the mass of the resonance approaches  
 2306 to high values.

2307 This analysis is affected essentially by the same background processes as the SM Higgs  
 2308 boson search, with the difference that in this case the SM Higgs boson processes, including  
 2309 all production modes, are treated as backgrounds.

2310 In addition to requiring the events to pass the single or double lepton triggers, exactly  
 2311 one electron and one muon are required to be reconstructed in the event with opposite  
 2312 charges and a minimum  $p_T$  of 20 GeV for both the muon and electron. Both leptons are  
 2313 required to be well identified and isolated to reject fake leptons and leptons coming from  
 2314 decays in flight. To suppress background processes with three or more leptons in the final  
 2315 state, such as diboson or triboson production, events with any additional identified and  
 2316 isolated lepton with  $p_T > 10$  GeV are rejected. To suppress the contribution of the SM  
 2317 production of the Higgs boson at 125 GeV,  $m_{\ell\ell}$  is requested to be higher than 50 GeV. The  
 2318 other event requirements are identical to the 125 GeV Higgs boson search and are described  
 2319 in Sec. ??.

2320 In addition to the 0 and 1 jet categories, a specific category sensitive to the VBF  
 2321 production mode is defined exploiting the characteristic signature of this process, where  
 2322 two energetic jets are emitted in the forward region of the detector and with large  $\Delta\eta$   
 2323 gap. Events belonging to the VBF-enriched category are selected by requiring at least  
 2324 two jets with  $p_T > 30$  GeV, an invariant mass  $m_{jj} > 500$  GeV and a gap in pseudorapidity  
 2325  $\Delta\eta_{jj} > 3.5$ .

2326 In addition to the transverse mass variable  $m_T$ , which is used in the analysis selection  
 2327 to define the DY background control region, an additional variable is defined, that from  
 2328 now on will be labelled as “improved transverse mass”  $m_T^i$ . This variable is defined as  
 2329 the invariant mass of the four momentum resulting from the sum of the two leptons four  
 2330 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.:

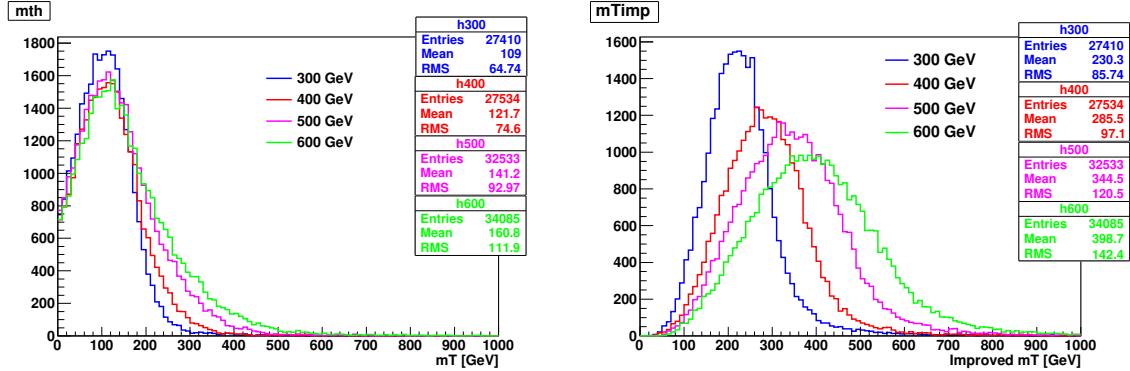
$$m_T^i = \sqrt{(p_{\ell\ell} + E_T^{\text{miss}})^2 - (\vec{p}_{\ell\ell} + \vec{p}_T^{\text{miss}})^2} \quad . \quad (6.1)$$

2331 This variable allows having a better sensitivity to different resonance mass hypothesis  
 2332 as shown in Fig. ??, where the shape of the  $m_T^i$  variable is shown for different SM Higgs  
 2333 mass hypothesis and it is compared to the standard  $m_T$  variable. The usage of this variable

---

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

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.

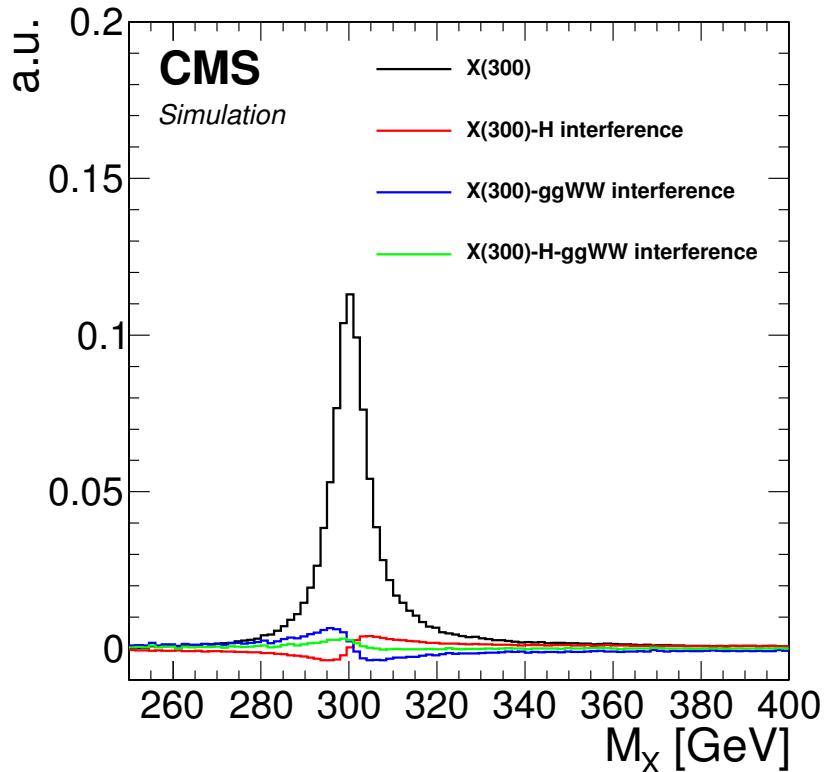
In order to test different resonance decay widths hypotheses, the signal samples, which are generated with a decay width corresponding to the expected value for a SM Higgs boson at that mass ( $\Gamma_{\text{SM}}$ ), are reweighted to obtain the desired width value ( $\Gamma'$ ). In particular the following 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 Breit Wigner distributions with different decay widths,  $f(E, \Gamma', M_X)/f(E, \Gamma_{\text{SM}}, M_X)$ , where:

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

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

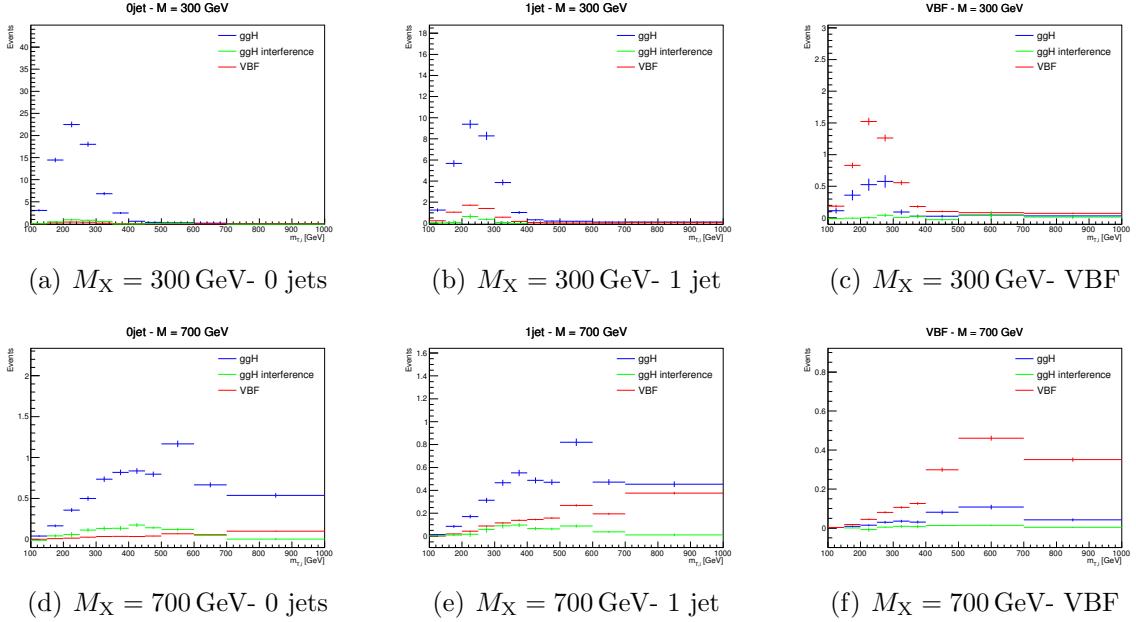
When a resonance with a non negligible width is considered, it is important to take into account the interference effects both with the  $gg \rightarrow WW$  background and the SM Higgs boson off-shell tail. A study of the interference effects for a resonance  $X$  produced through the gluon fusion mechanism is performed within the MCFM+JHUGEN framework, and

2357 including NNLO corrections for cross section using HNNLO program [111] based on  
 2358 MCFM. The matrix element package MELA supports all of these processes and allows fast  
 2359 MC re-weighting and optimal discriminant calculation. The basic idea of this approach is to  
 2360 compute the matrix elements of the processes under study with the MCFM and JHUGEN  
 2361 generators, including the interference terms, and using these matrix elements to compute  
 2362 an event weight used to reweight the simulated samples. Using this approach the simulated  
 2363 events can be reweighted according to different scenarios, for instance including some or all  
 2364 the interference terms, allowing a detailed study of the interference contribution. The effect  
 2365 of the various interference terms for the  $M_X$  variable at generator level is shown in Fig. ??,  
 2366 after having applied the WW baseline selections. As can be observed the contribution of  
 2367 the interference of the scalar resonance with the  $gg \rightarrow WW$  background and with SM Higgs  
 2368 boson have opposite sign and partially cancel out. This cancellation effect is different for  
 2369 different resonance masses and depends on the event selection. In particular the interference  
 2370 term with the SM Higgs off-shell tail is positive for values below  $M_X$  while it turns negative  
 2371 above  $M_X$ . The contribution of the interference with the  $gg \rightarrow WW$  background is instead  
 2372 characterized by an opposite sign lineshape, thus leading to a partial cancellation when  
 2373 considering the total interference.



**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.

2374 The effect of the resulting interference contribution including all the different terms is  
 2375 shown in Fig. ?? for the  $m_T^i$  signal templates, in the three categories separately and for  
 2376 different  $M_X$  hypotheses.



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

2377 The interference contribution is thus not negligible, especially for large values of  $M_X$ ,  
 2378 and is included in the analysis as part of the signal contribution. More specifically, during  
 2379 the fit procedure the signal yield is scaled by the signal strength parameter  $\mu$  (which is the  
 2380 parameter of interest of the fit), while the interference yield is scaled by  $\sqrt{\mu}$ .

## 2381 6.4 Background estimation

2382 The background processes affecting the analysis phase space are the same as the ones  
 2383 contributing to the SM Higgs search described in Sec. ???. The techniques used for the  
 2384 background estimation are the same as well.

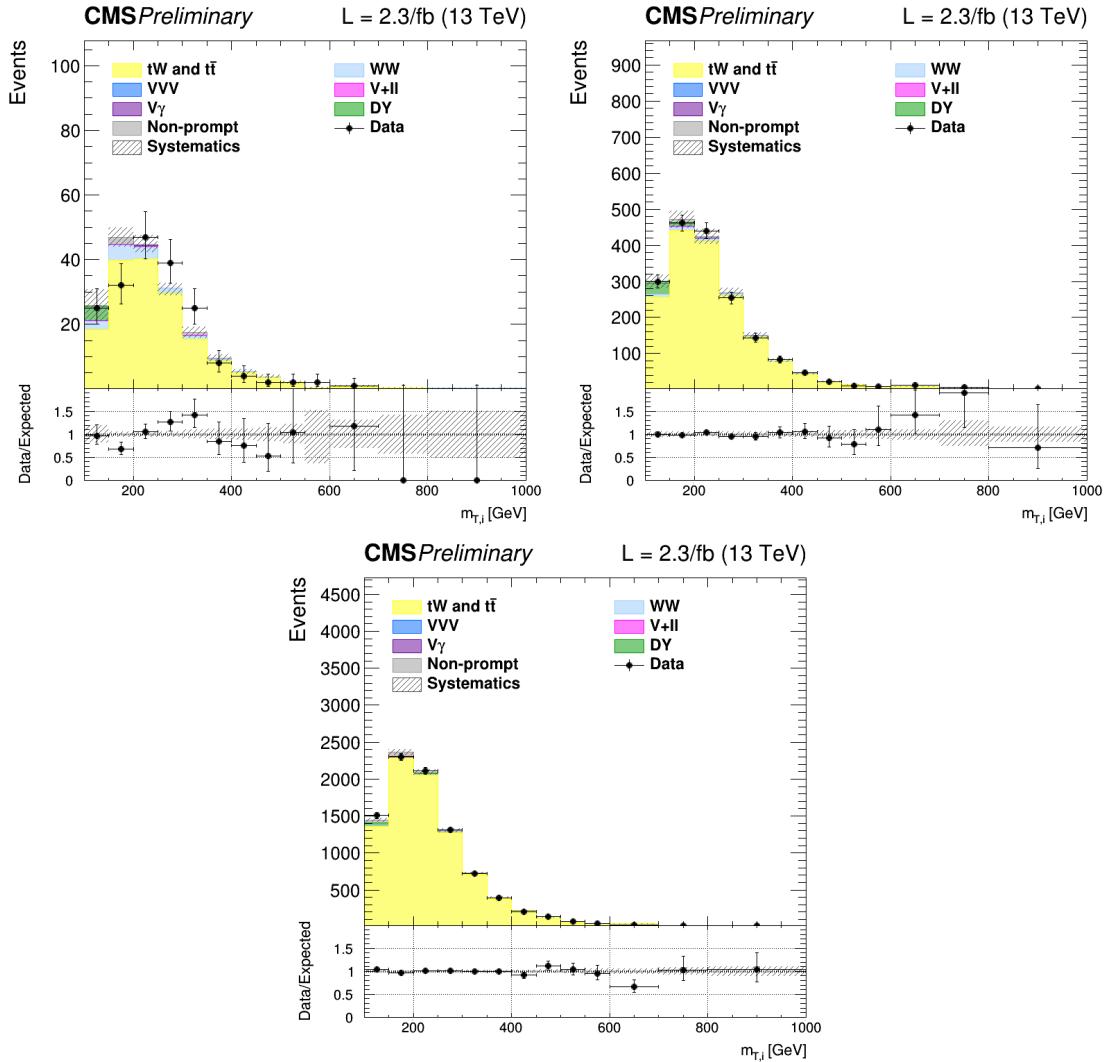
2385 The most relevant difference is the addition of the 2 jets category. The WW, top  
 2386 quark and DY background normalizations are estimated in this category using data driven  
 2387 techniques, similarly to the other jet bins.

2388 Given the slightly different WW baseline selection with respect to the SM Higgs search,  
 2389 also the control regions for the top quark and DY backgrounds estimation change, while the  
 2390 WW background normalization is estimated from data in the three signal regions separately,  
 2391 owing to the different  $m_T^i$  shapes for signal and background.

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

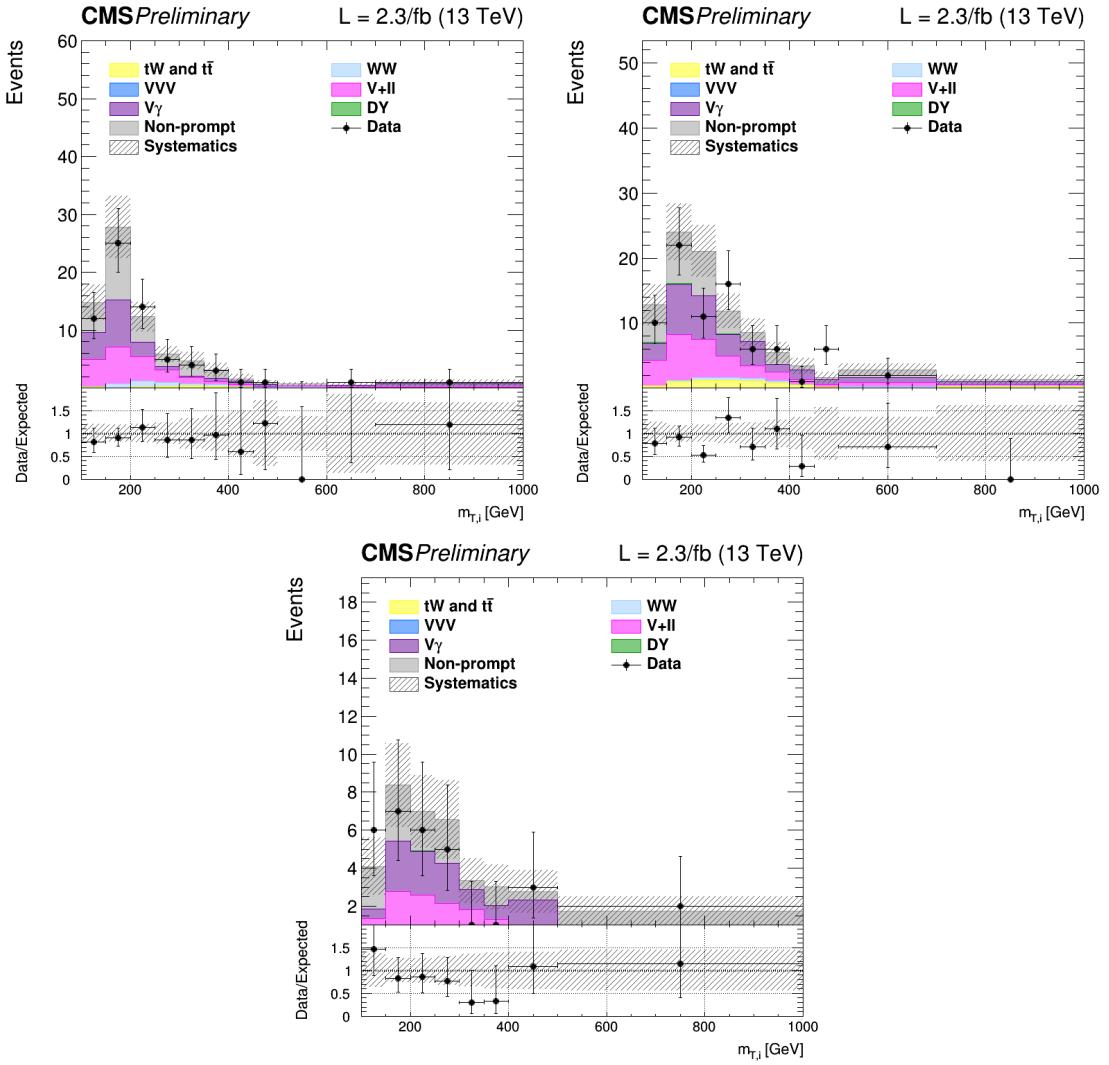
- 0 jets category: at least one b-tagged jet with  $20 < p_T < 30$  GeV is required;
- 1 jet category: exactly one b-tagged jet with  $p_T$  above 30 GeV is required;
- 2 jets category: at least one b-tagged jet with  $p_T$  above 30 GeV is required.

Distributions of the  $m_{T,j}^i$  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. ??.



**Figure 6.5:** Distributions of  $m_{T,j}^i$  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.

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.



**Figure 6.6:** Distributions of  $m_T^i$  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.

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  $\tau$  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

<sup>2410</sup> the 0 and 1 jet categories. In the VBF category, given the very small number of expected  
<sup>2411</sup> events, the normalization of this background is taken from simulation.

<sup>2412</sup> Other minor background processes are estimated as described in ??.

## <sup>2413</sup> 6.5 Systematic uncertainties

<sup>2414</sup> The systematic uncertainties affecting this analysis are the same discussed in Sec. ??.  
<sup>2415</sup> The differences with respect to the SM Higgs boson search are described below.

<sup>2416</sup> The PDF and  $\alpha_s$  uncertainties on the signal cross sections are taken from the computations  
<sup>2417</sup> performed by the LHC cross section working group [102], and are included for all  
<sup>2418</sup> the mass points. The value of these uncertainties depends on the resonance mass and vary  
<sup>2419</sup> from 3 and 5% for ggH and from 2 and 3% for VBF production modes. The PDFs and  $\alpha_s$   
<sup>2420</sup> uncertainties on the signal selection are evaluated for every resonance mass and are found  
<sup>2421</sup> to be less than 1% for both ggH and VBF.

<sup>2422</sup> The theoretical uncertainties in the signal yields due to the jet categorization are  
<sup>2423</sup> evaluated for all the ggH signals following the prescription described in Sec. ??.

<sup>2424</sup> An additional uncertainty on the modelling of the top pair background is derived from  
<sup>2425</sup> the observed discrepancy between data and POWHEG V2 plus PYTHIA 8.1 simulation  
<sup>2426</sup> on the top quark  $p_T$  spectrum [112], which is particularly important in the tail of the  $m_T^i$   
<sup>2427</sup> distribution. Another uncertainty affecting the  $m_T^i$  tail for the top quark background is the  
<sup>2428</sup> parton shower uncertainty. This is evaluated comparing the generator level  $m_T^i$  distributions  
<sup>2429</sup> corresponding to two different simulations of the  $t\bar{t}$  process: one obtained using PYTHIA  
<sup>2430</sup> 8.1 for the showering and hadronization of the simulated events, and the other using  
<sup>2431</sup> HERWIG++. The difference between the two is used to extract a shape uncertainty, which  
<sup>2432</sup> is less than 1% for low  $m_T^i$  values and reaches about 6% in the  $m_T^i$  tail.

## <sup>2433</sup> 6.6 Signal extraction and limit setting

<sup>2434</sup> The signal yield, including both the ggH and VBF production modes, is extracted performing  
<sup>2435</sup> a combined fit of the three categories to the  $m_T^i$  simulation templates for backgrounds and  
<sup>2436</sup> signal, and is repeated for each resonance mass hypothesis. Moreover, fixed the mass of  
<sup>2437</sup> the resonance, the fit is performed again for the various hypotheses of the resonance decay  
<sup>2438</sup> width. A single signal strength  $\mu$  is extracted from the fit, which multiplies both the ggH  
<sup>2439</sup> and VBF contributions. In other words it is assumed that the ratio of the two production  
<sup>2440</sup> mechanism stays the same as the one predicted by the SM<sup>2</sup>.

<sup>2441</sup> The background yields expected from simulation corresponding to the three jet categories  
<sup>2442</sup> and after the analysis event selection are shown in Table ??.  
<sup>2443</sup> The signal yields corresponding to a selection of mass points and assuming  $\Gamma' = \Gamma_{SM}$  are shown in Table ??.

---

<sup>2</sup>This 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	$501.93 \pm 0.00$ (0%)	$198.72 \pm 0.00$ (0%)	$4.54 \pm 0.00$ (0%)
gg $\rightarrow$ WW	$37.28 \pm 5.77$ (15%)	$19.63 \pm 3.04$ (15%)	$1.05 \pm 0.16$ (15%)
Top quark	$188.75 \pm 0.00$ (0%)	$330.05 \pm 0.00$ (0%)	$25.06 \pm 0.00$ (0%)
DY	$33.24 \pm 0.00$ (0%)	$12.99 \pm 0.00$ (0%)	$0.28 \pm 0.00$ (0%)
Non-prompt	$64.21 \pm 19.26$ (30%)	$31.69 \pm 9.51$ (30%)	$2.10 \pm 0.63$ (30%)
V $\gamma$	$26.62 \pm 0.72$ (3%)	$14.18 \pm 0.38$ (3%)	$0.64 \pm 0.02$ (3%)
V $\gamma^*$	$4.44 \pm 1.12$ (25%)	$3.39 \pm 0.85$ (25%)	$0.14 \pm 0.04$ (25%)
VZ	$13.51 \pm 0.76$ (6%)	$11.67 \pm 0.66$ (6%)	$0.28 \pm 0.02$ (6%)
VVV	$0.01 \pm 0.00$ (3%)	$0.02 \pm 0.00$ (3%)	$0.00 \pm 0.00$ (3%)
SM H $\rightarrow$ WW	$6.04 \pm 0.40$ (7%)	$3.10 \pm 0.11$ (5%)	$0.34 \pm 0.02$ (7%)
SM H $\rightarrow$ $\tau\tau$	$0.50 \pm 0.05$ (9%)	$0.43 \pm 0.04$ (9%)	$0.04 \pm 0.00$ (9%)
Total background	876.5	625.9	34.5

<sup>2444</sup> The strategy for computing the exclusion limits is based on the modified frequentist  
<sup>2445</sup> approach, also referred to as  $CL_s$ , as described in [101]. The first step is to construct the  
<sup>2446</sup> likelihood function  $\mathcal{L}(data|\mu, \theta)$ :

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

<sup>2447</sup> where  $data$  represents the experimental observation,  $s$  and  $b$  are the expected signal and  
<sup>2448</sup> background yields respectively and  $\theta$  is the full set of nuisance parameters constrained by  
<sup>2449</sup> the prior distribution functions  $p(\tilde{\theta}|\theta)$ . The default values for the nuisance parameters are  
<sup>2450</sup> labelled as  $\tilde{\theta}$ .

<sup>2451</sup> For a binned shape analysis,  $Poisson(data|\mu \cdot s + b)$  is the product of the Poisson  
<sup>2452</sup> probabilities to observe  $n_i$  events in bin i:

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

<sup>2453</sup> In order to test the compatibility of the data with the signal plus background (or the  
<sup>2454</sup> background only) hypothesis, the test statistic  $\tilde{q}_\mu$  is constructed based on the profile

**Table 6.2:** Expected signal yields for the ggH and VBF production modes estimated from simulation after the analysis event selection for different mass hypothesis assuming  $\Gamma' = \Gamma_{\text{SM}}$  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	$90.21 \pm 6.67$ (7%)	$37.47 \pm 1.81$ (5%)	$1.25 \pm 0.26$ (21%)
400	$66.35 \pm 4.90$ (7%)	$32.65 \pm 1.57$ (5%)	$2.04 \pm 0.42$ (21%)
600	$13.86 \pm 1.05$ (8%)	$8.56 \pm 0.44$ (5%)	$0.68 \pm 0.14$ (21%)
800	$3.20 \pm 0.25$ (8%)	$2.32 \pm 0.13$ (6%)	$0.22 \pm 0.05$ (21%)
1000	$0.88 \pm 0.07$ (8%)	$0.70 \pm 0.04$ (6%)	$0.07 \pm 0.02$ (21%)
VBF signal yields			
200	$1.54 \pm 0.06$ (4%)	$6.18 \pm 0.25$ (4%)	$5.05 \pm 0.20$ (4%)
400	$0.91 \pm 0.04$ (4%)	$3.42 \pm 0.14$ (4%)	$3.19 \pm 0.13$ (4%)
600	$0.50 \pm 0.02$ (4%)	$1.95 \pm 0.08$ (4%)	$1.88 \pm 0.08$ (4%)
800	$0.33 \pm 0.01$ (4%)	$1.21 \pm 0.05$ (4%)	$1.16 \pm 0.05$ (4%)
1000	$0.22 \pm 0.01$ (4%)	$0.79 \pm 0.03$ (4%)	$0.69 \pm 0.03$ (4%)

2455 likelihood reatio:

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

2456 where  $\hat{\theta}_\mu$  refers to the conditional maximum likelihood estimators of  $\theta$ , given the signal  
2457 strength  $\mu$ . The parameter estimators  $\hat{\mu}$  and  $\hat{\theta}$  correspond to the global maximum of the  
2458 likelihood. The  $0 \leq \hat{\mu}$  constraint is imposed to have a positive signal yield, e.g. background  
2459 underfluctuations are forbidden, while  $\hat{\mu} \leq \mu$  is imposed to have a one-sided confidence  
2460 interval. The observed test statistic for the signal strength  $\mu$  under test is referred to as  
2461  $\tilde{q}_\mu^{\text{obs}}$ . The values of the nuisance parameters obtained maximising the likelihood function are  
2462 labelled as  $\hat{\theta}_0^{\text{obs}}$  and  $\hat{\theta}_\mu^{\text{obs}}$  for the background only and signal plus background hypotheses,  
2463 respectively. The pdf of the test statistic is constructed by generating toy MC pseudo-data  
2464 for both the background only and signal plus background hypotheses, i.e.  $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}})$  and  
2465  $f(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}})$ . These distributions can be used to define two p-values corresponding to the

2466 two hypotheses,  $p_\mu$  and  $p_b$ :

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

2467

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

2468 According to these definitions,  $p_\mu$  and  $p_b$  can be identified with  $\text{CL}_{s+b}$  and  $1 - \text{CL}_b$ , 2469 respectively. The  $\text{CL}_s(\mu)$  is calculated using the following ratio:

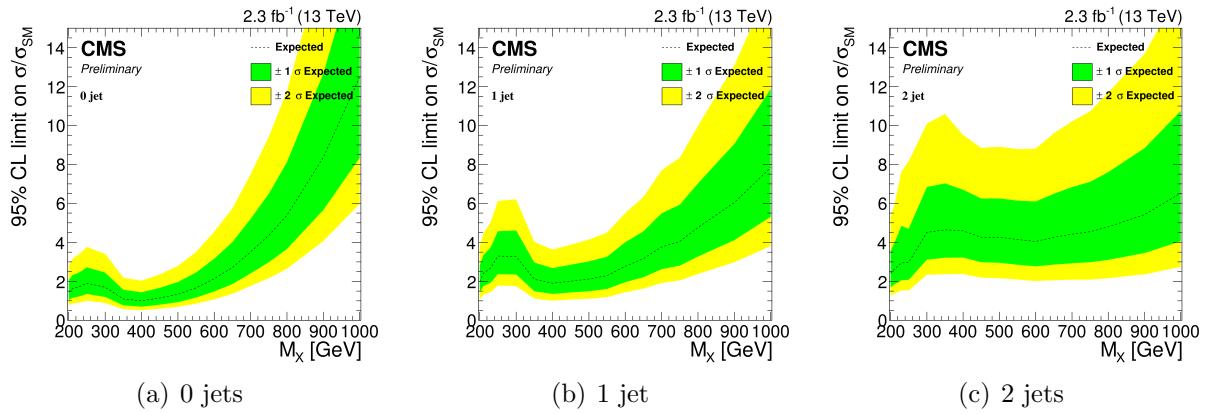
$$\text{CL}_s(\mu) = \frac{\text{CL}_{s+b}}{\text{CL}_b} = \frac{p_\mu}{1 - p_b} \quad . \quad (6.8)$$

2470 If, for a given signal strength  $\mu$ ,  $\text{CL}_s \leq \alpha$ , then the hypothesis is excluded with a  $(1 - \alpha)$  2471 confidence level (CL). For instance, if one wants to quote the upper limit on  $\mu$  with a 95% 2472 CL, the signal strength has to be adjusted until  $\text{CL}_s = 0.05$ .

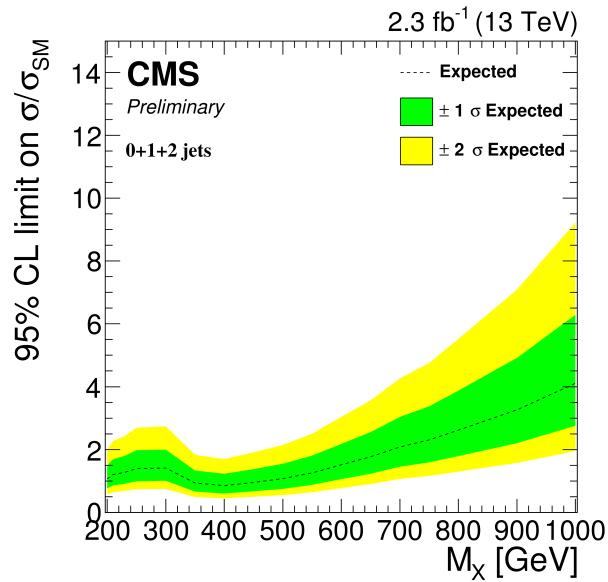
2473 The expected median upper limit, as well as the  $\pm 1\sigma$  (68% CL) and  $\pm 2\sigma$  (95% CL) 2474 bands, are determined generating a large amount of pseudo-data in the background only 2475 hypothesis and calculating  $\text{CL}_s$  and the 95% CL upper limit for each of them, as if they 2476 were real data. Then the cumulative distribution of the 95% CL upper limits is built and 2477 the median expected value is identified as the value at which the cumulative distribution 2478 crosses the 50% quantile. The  $\pm 1\sigma$  ( $\pm 2\sigma$ ) band is defined by the values at which the 2479 cumulative distribution crosses the 16% (2.5%) and 84% (97.5%) quantiles.

2480 In order to assess the sensitivity of the analysis, the expected upper exclusion limits at 2481 95% CL on the signal strength are shown in Fig. ?? for the three jet categories separately. 2482 For a given mass of the resonance, the limits are derived assuming a signal decay width 2483  $\Gamma' = \Gamma_{\text{SM}}$  and a cross section equal to the one expected from a SM Higgs boson at that 2484 mass. The other decay width hypothesis have also been tested, showing a very similar 2485 expected exclusion limit, suggesting that this analysis is not strongly sensitive to variations 2486 of the resonance decay width.

2487 The 0 jets category is the most sensitive especially in the low mass region, while for 2488 very large masses of the resonance the 1 jet and VBF categories start being important. 2489 This is explained mainly by the fact that the VBF contribution increases, with respect to 2490 ggH, as the mass increases. The expected exclusion limit on the signal strength after the 2491 combination of the three categories is shown in Fig. ???. Comparing the limits in the single 2492 categories with the combination of the three it is evident how the higher jet multiplicity 2493 categories help in improving the results for large values of  $M_X$ .



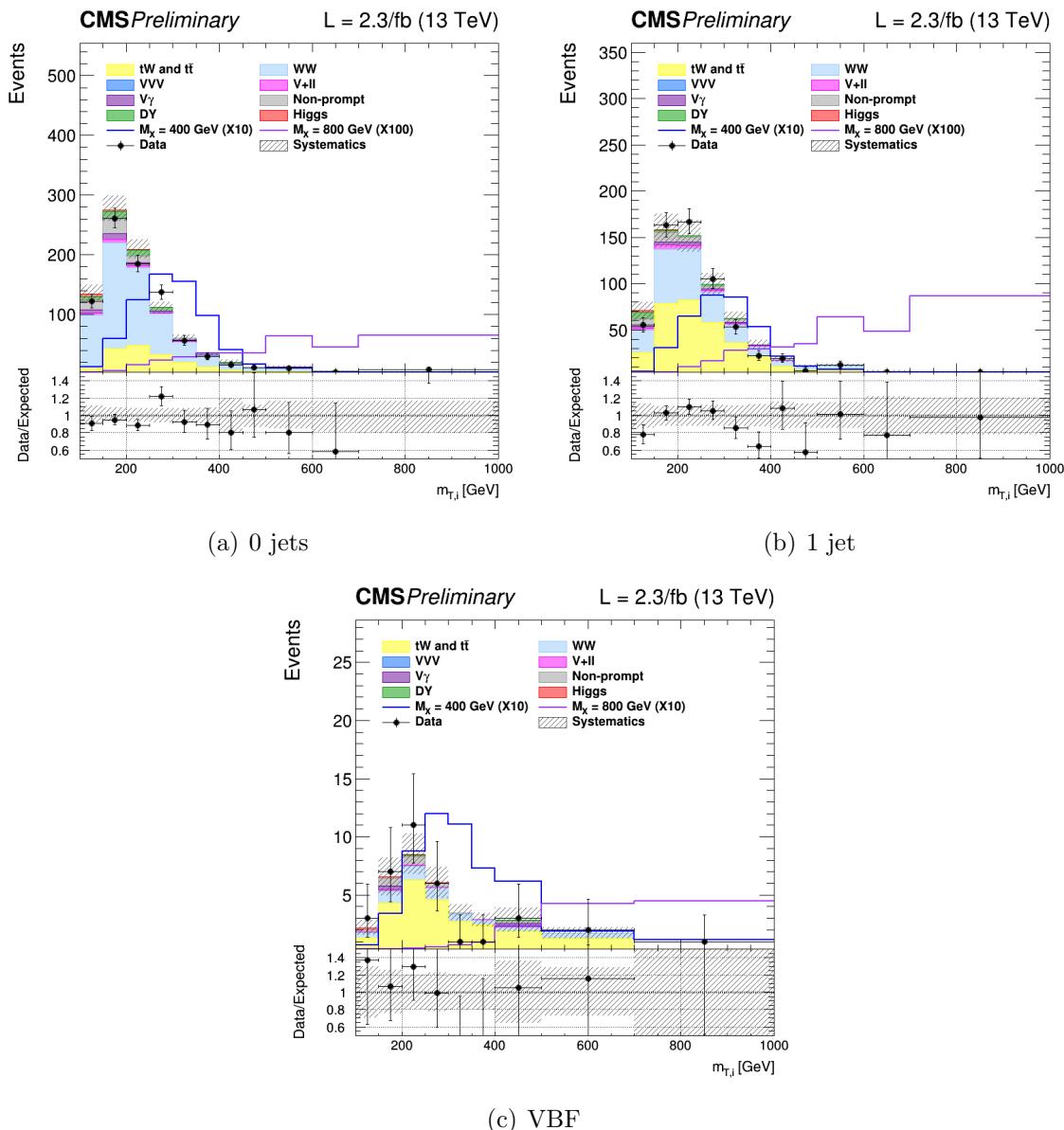
**Figure 6.7:** Expected exclusion upper limits at 95% CL on the signal strength in the three categories, as a function of the resonance mass. The dashed line corresponds to median upper limit, while the green and yellow regions represent the  $\pm 1\sigma$  and  $\pm 2\sigma$  uncertainty bands, respectively.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

2494 6.7 Results

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

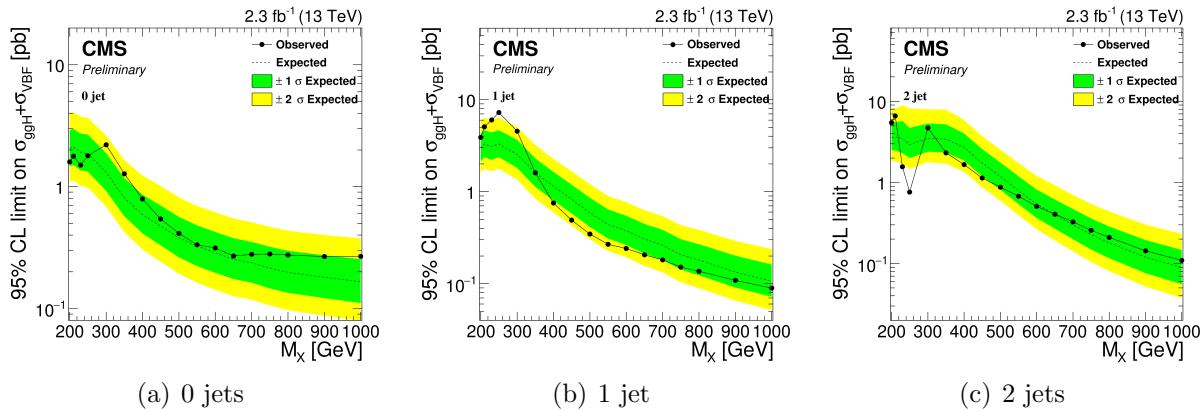
2498 For every mass point from 200 GeV up to 1 TeV the observed p-value and the 95% CL  
 2499 upper exclusion limit are calculated for five hypothesis of the signal width. The observed  
 2500 p-value as a function of the resonance mass for the combination of the three jet categories  
 2501 is shown 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)

2502 In order to be independent on the particular model assumed for the signal cross section,  
 2503 the results are interpreted as exclusion limits on  $\sigma \times \mathcal{B}$ , where  $\sigma$  stands for the sum of  
 2504 the ggH and VBF cross sections, and  $\mathcal{B}$  represents the  $X \rightarrow WW \rightarrow 2\ell 2\nu$  branching ratio  
 2505 including all lepton flavours. The expected and observed upper exclusion limits on  $\sigma \times \mathcal{B}$   
 2506 for  $\Gamma' = \Gamma_{SM}$  are shown in Fig. ??.

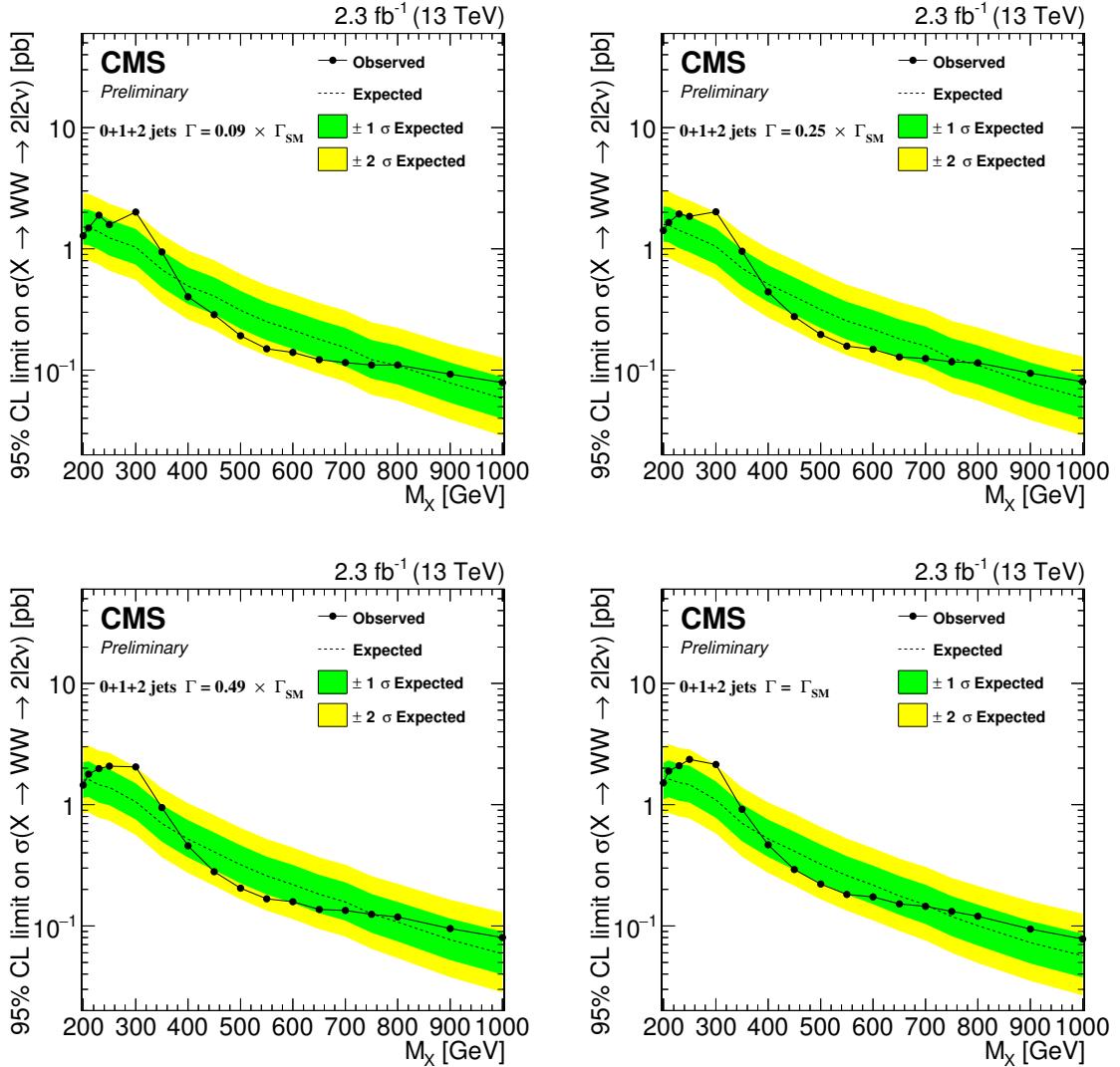
2507 A mild excess is observed in the 0 jets category and, more evident, in the 1 jet category  
 2508 around 250-300 GeV. A deficit is instead in the VBF category around 250 GeV, which is  
 2509 mainly due to an underfluctuation of the background. This effect can be understood looking



**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.

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  $\Gamma'$  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**

2518



# Appendix A

## <sup>2519</sup> Fiducial region definition and <sup>2520</sup> optimization

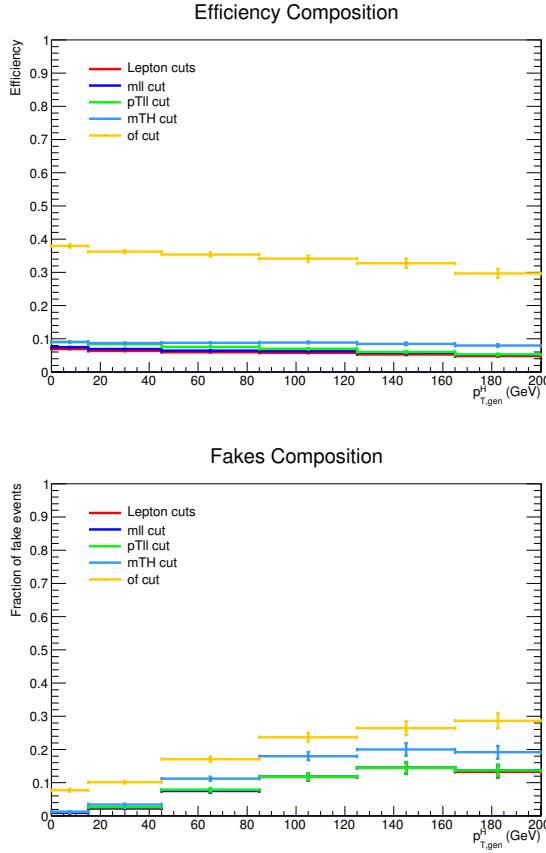
<sup>2521</sup> The fiducial region must be chosen in such a way to be as close as possible to the selections  
<sup>2522</sup> applied in the analysis, in order to reduce the model dependence in the extrapolation step.  
<sup>2523</sup> That means that for optimizing the fiducial volume definition, the efficiency has to be  
<sup>2524</sup> maximized. Another parameter entering the game is the number of fake events, in other  
<sup>2525</sup> words the number of reconstructed events which do not belong to the fiducial phase space.  
<sup>2526</sup> This parameter should instead be as small as possible. Even if we have to observe the trend  
<sup>2527</sup> of these two quantities as a function of  $p_T^H$ , we can maximize the ratio between the overall  
<sup>2528</sup> efficiency and the overall fake rate as a proxy for establishing the “goodness” of the fiducial  
<sup>2529</sup> region.

<sup>2530</sup> Several different fiducial region definitions were tested and the results show that:

- <sup>2531</sup> • **of cut:** The fiducial region definition must include only the opposite flavor combination  
<sup>2532</sup> including one electron and one muon. If we include also the combinations involving  
<sup>2533</sup>  $\tau$ 's the efficiency falls down.
- <sup>2534</sup> • **Lepton cut:** Since the resolution on lepton transverse momentum is good, there is no  
<sup>2535</sup> need to loosen the cuts related these variables, i.e. we can use the same cuts defined  
<sup>2536</sup> in the analysis selection ( $p_T^{\ell,1} > 20 \text{ GeV}$ ,  $p_T^{\ell,2} > 10 \text{ GeV}$ ).
- <sup>2537</sup> • **Di-lepton  $p_T$  cut:** As stated in the previous point, there is no need to loosen this  
<sup>2538</sup> cut, so we kept the same value as the analysis selection, i.e.  $p_T^{\ell\ell} > 30 \text{ GeV}$ .
- <sup>2539</sup> • **Di-lepton mass cut:**  $m_{\ell\ell} > 12 \text{ GeV}$  as discussed before.
- <sup>2540</sup> • **neutrino pair  $p_T$  cut:** Since the resolution on the measurement of the missing  
<sup>2541</sup> transverse energy is poor, the neutrino pair cut should not be included in the definition  
<sup>2542</sup> of the fiducial region, because it would increase the fake rate without increasing the  
<sup>2543</sup> efficiency, thus resulting in a lower ratio between overall efficiency and fake rate.
- <sup>2544</sup> •  **$m_T$  cut:** Also the  $m_T$  cut that we have in the analysis selection, i.e.  $m_T > 60 \text{ GeV}$ ,  
<sup>2545</sup> should be loosened or removed because it involves neutrinos and then increase the  
<sup>2546</sup> fake rate. We decided eventually to keep this cut, loosening it to 50 GeV, because in  
<sup>2547</sup> addition to increase the number of fake events, it increases the efficiency as well.

<sup>2548</sup> The fake rate and the efficiency as a function of  $p_T^H$  after the optimization discussed  
<sup>2549</sup> before are shown in figure ???. To obtain these plots the fiducial region was modified adding

2550 in sequence the various cuts and computing the efficiency and the fake rate each time. In  
 2551 that way we can asses the composition of those distributions.

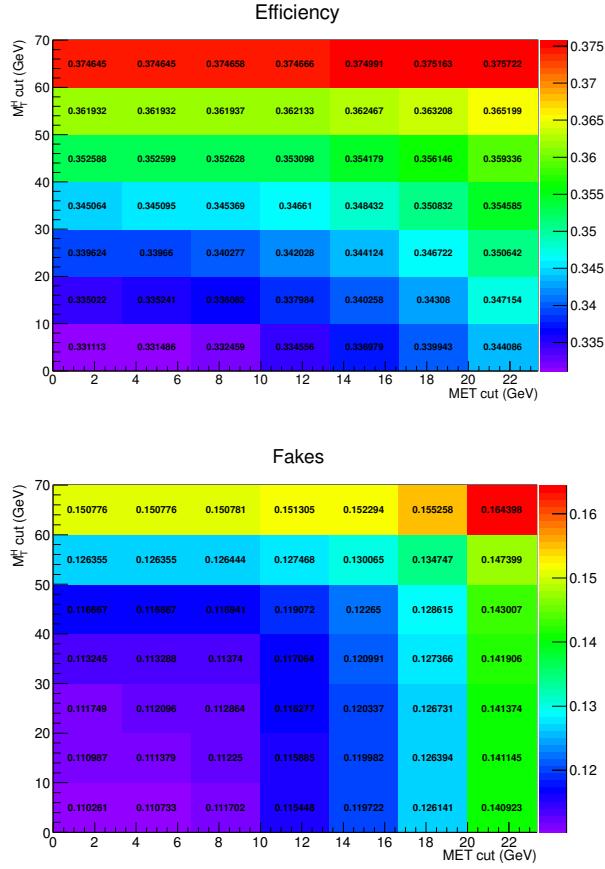


**Figure A.1:** Efficiency and fake rate as a function on Higgs transverse momentum. The plots correspond to the optimized fiducial region definition and show the effect of adding each of the mentioned cuts in sequence.

2552 The efficiency and fraction of fake events have been measured also as a function of the  
 2553  $E_T^{\text{miss}}$  and  $m_T$  cuts in the fiducial region. Since these two variables are correlated, the  
 2554 results are reported as two-dimensional histograms. In Fig. ?? are reported the efficiency  
 2555 and fraction of fake events for these two variables.

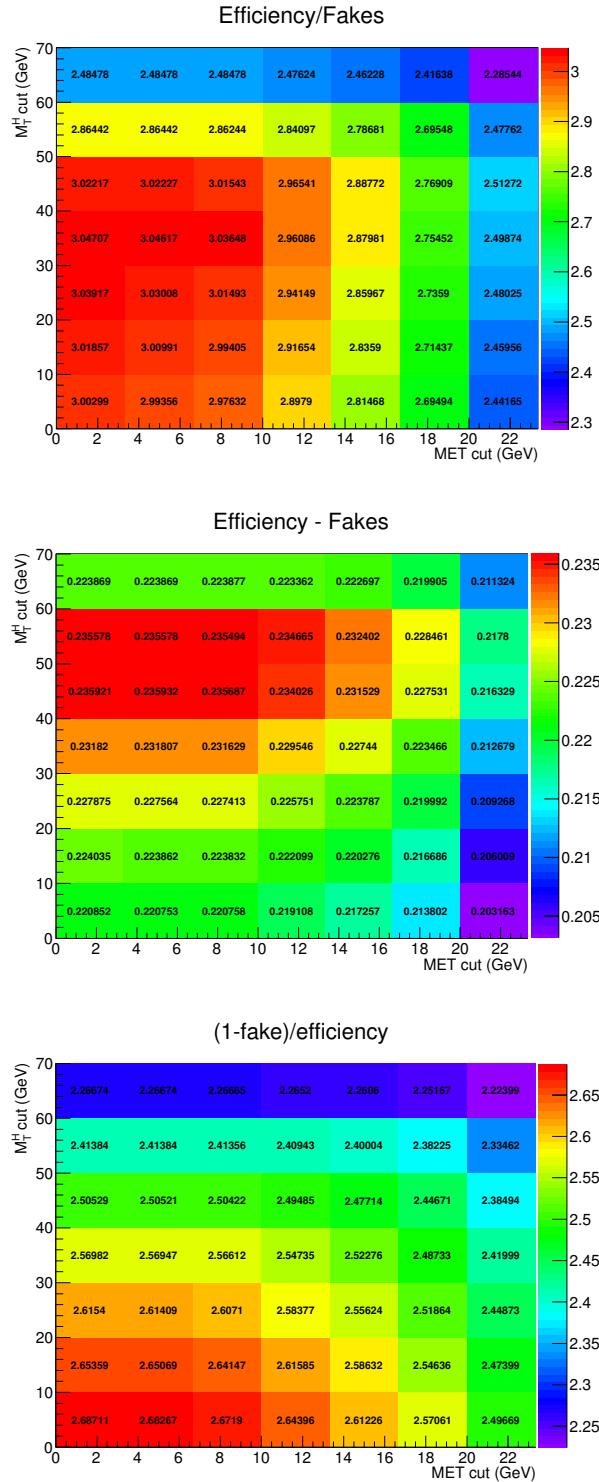
2556 The criterion adopted to define the fiducial region is a tradeoff between having a large  
 2557 efficiency and a small fraction of fake events. Especially when looking at the low resolution  
 2558 variables, such as  $E_T^{\text{miss}}$  and  $m_T$ , a suitable figure of merit has to be chosen for the estimation  
 2559 of the best cuts. Several different figures of merit have been checked, such as  $\epsilon/f$ ,  $\epsilon - f$   
 2560 and  $(1 - f)/\epsilon$ . The results for these three different figures of merit are shown in Fig. ?? as  
 2561 a function of the  $E_T^{\text{miss}}$  and  $m_T$  cuts in the fiducial region.

2562 Following the same criterion, similar plots as above have been obtained for an alternative  
 2563 model, given by varying up the ggH/VBF ratio within the experimental uncertainties. The

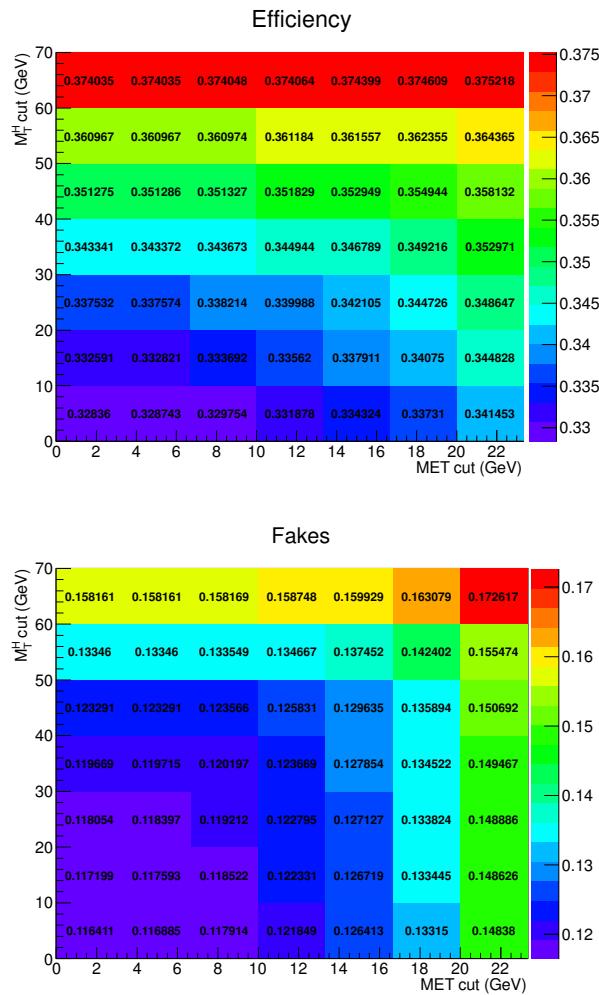


**Figure A.2:** Efficiency and fake rate as a function of  $E_T^{\text{miss}}$  and  $m_T$  cuts in the fiducial region.

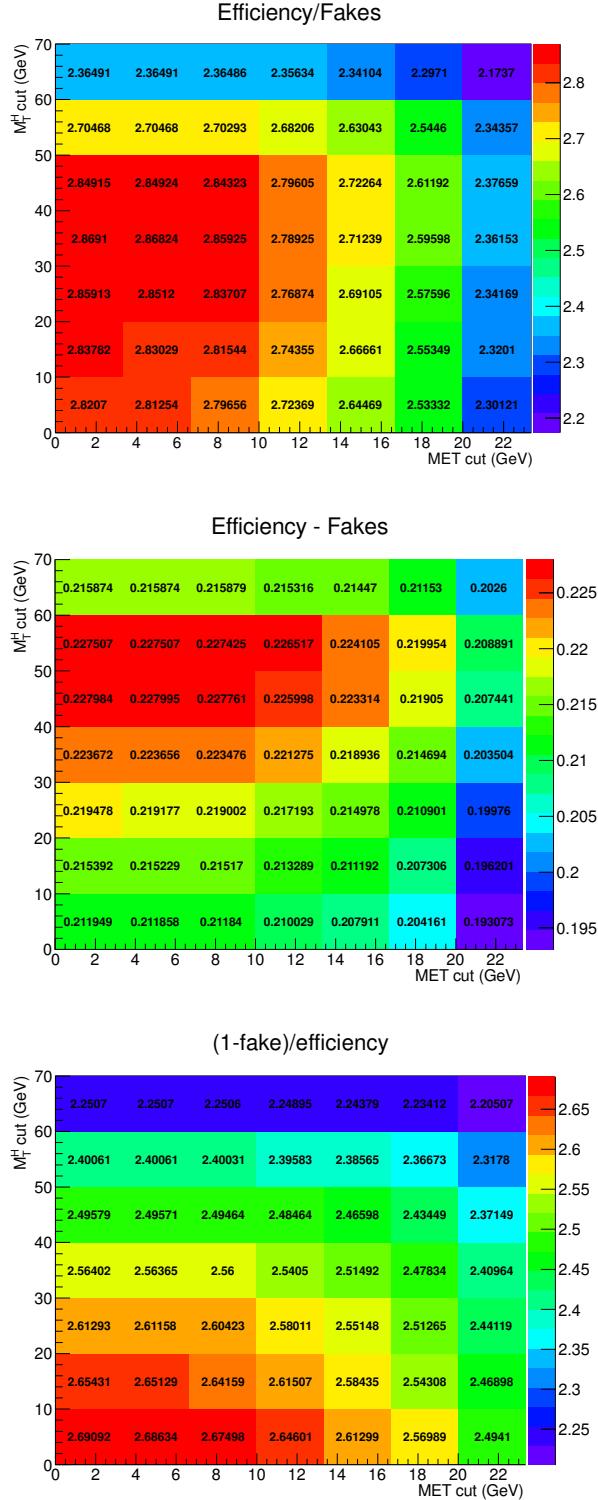
2564 results, shown in Fig. ?? and Fig. ??, show a similar trend with respect to the model with  
 2565 nominal ggH/VBF ratio.



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



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





2567

# Bibliography

- 2568 [1] Thomas Sven Pettersson and P Lefèvre. *The Large Hadron Collider: conceptual*  
2569 *design*. Tech. rep. CERN-AC-95-05-LHC. Oct. 1995. URL: <http://cds.cern.ch/record/291782>.
- 2571 [2] Oliver Sim Brüning et al. *LHC Design Report*. Geneva: CERN, 2004. URL: <http://cds.cern.ch/record/782076>.
- 2573 [3] Oliver Sim Brüning et al. *LHC Design Report*. Geneva: CERN, 2004. URL: <http://cds.cern.ch/record/815187>.
- 2575 [4] Michael Benedikt et al. *LHC Design Report*. Geneva: CERN, 2004. URL: <http://cds.cern.ch/record/823808>.
- 2577 [5] *LEP design report*. By the LEP Injector Study Group. Geneva: CERN, 1983. URL:  
2578 <http://cds.cern.ch/record/98881>.
- 2579 [6] *LEP design report*. Copies shelved as reports in LEP, PS and SPS libraries. Geneva:  
2580 CERN, 1984. URL: <http://cds.cern.ch/record/102083>.
- 2581 [7] Carlo Wyss. *LEP design report, v.3: LEP2*. Vol. 1-2 publ. in 1983-84. Geneva: CERN,  
2582 1996. URL: <http://cds.cern.ch/record/314187>.
- 2583 [8] S. Chatrchyan et al. “The CMS experiment at the CERN LHC”. In: *JINST* 3 (2008),  
2584 S08004. DOI: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- 2585 [9] G. Aad et al. “The ATLAS Experiment at the CERN Large Hadron Collider”. In:  
2586 *JINST* 3 (2008), S08003. DOI: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- 2587 [10] A. Augusto Alves Jr. et al. “The LHCb Detector at the LHC”. In: *JINST* 3 (2008),  
2588 S08005. DOI: [10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005).
- 2589 [11] K. Aamodt et al. “The ALICE experiment at the CERN LHC”. In: *JINST* 3 (2008),  
2590 S08002. DOI: [10.1088/1748-0221/3/08/S08002](https://doi.org/10.1088/1748-0221/3/08/S08002).
- 2591 [12] G. Anelli et al. “The TOTEM experiment at the CERN Large Hadron Collider”. In:  
2592 *JINST* 3 (2008), S08007. DOI: [10.1088/1748-0221/3/08/S08007](https://doi.org/10.1088/1748-0221/3/08/S08007).
- 2593 [13] O. Adriani et al. “The LHCf detector at the CERN Large Hadron Collider”. In:  
2594 *JINST* 3 (2008), S08006. DOI: [10.1088/1748-0221/3/08/S08006](https://doi.org/10.1088/1748-0221/3/08/S08006).
- 2595 [14] *The CMS magnet project: Technical Design Report*. Technical Design Report CMS.  
2596 Geneva: CERN, 1997. URL: <http://cds.cern.ch/record/331056>.
- 2597 [15] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical  
2598 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] *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: <http://cds.cern.ch/record/357153>.
- [18] *The CMS muon project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: <http://cds.cern.ch/record/343814>.
- [19] S. Dasu et al. “CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems”. In: (2000).
- [20] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS. Geneva: CERN, 2002. URL: <http://cds.cern.ch/record/578006>.
- [21] CMS Collaboration. *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and  $E_T^{\text{miss}}$* . CMS Physics Analysis Summary CMS-PAS-PFT-09-001. 2009. URL: <http://cdsweb.cern.ch/record/1194487>.
- [22] Serguei Chatrchyan et al. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].
- [23] Pierre Billoir and S. Qian. “Simultaneous pattern recognition and track fitting by the Kalman filtering method”. In: *Nucl. Instrum. Meth. A* 294 (1990), p. 219. DOI: 10.1016/0168-9002(90)91835-Y.
- [24] 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].
- [25] 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].
- [26] 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].
- [27] 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].
- [28] 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].

- 2639 [29] X. Chen et al. “NNLO QCD corrections to Higgs boson production at large transverse  
2640 momentum”. In: (2016). arXiv: 1607.08817 [hep-ph].
- 2641 [30] Massimiliano Grazzini and Hayk Sargsyan. “Heavy-quark mass effects in Higgs boson  
2642 production at the LHC”. In: *JHEP* 09 (2013), p. 129. DOI: 10.1007/JHEP09(2013)  
2643 129. arXiv: 1306.4581 [hep-ph].
- 2644 [31] Aleksandr Azatov and Ayan Paul. “Probing Higgs couplings with high  $p_T$  Higgs  
2645 production”. In: *JHEP* 01 (2014), p. 014. DOI: 10.1007/JHEP01(2014)014. arXiv:  
2646 1309.5273 [hep-ph].
- 2647 [32] Robert V. Harlander and Tobias Neumann. “Probing the nature of the Higgs-gluon  
2648 coupling”. In: *Phys. Rev. D* 88 (2013), p. 074015. DOI: 10.1103/PhysRevD.88.  
2649 074015. arXiv: 1308.2225 [hep-ph].
- 2650 [33] David Marzocca, Marco Serone, and Jing Shu. “General Composite Higgs Models”.  
2651 In: *JHEP* 08 (2012), p. 013. DOI: 10.1007/JHEP08(2012)013. arXiv: 1205.0770  
2652 [hep-ph].
- 2653 [34] Andrea Banfi, Adam Martin, and Veronica Sanz. “Probing top-partners in Higgs+jets”.  
2654 In: *JHEP* 08 (2014), p. 053. DOI: 10.1007/JHEP08(2014)053. arXiv: 1308.4771  
2655 [hep-ph].
- 2656 [35] Georges Aad et al. “Fiducial and differential cross sections of Higgs boson production  
2657 measured in the four-lepton decay channel in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the  
2658 ATLAS detector”. In: *Phys. Lett. B* 738 (2014), p. 234. DOI: 10.1016/j.physletb.  
2659 2014.09.054. arXiv: 1408.3226 [hep-ex].
- 2660 [36] Georges Aad et al. “Measurements of fiducial and differential cross sections for Higgs  
2661 boson production in the diphoton decay channel at  $\sqrt{s} = 8$  TeV with ATLAS”.  
2662 In: *JHEP* 09 (2014), p. 112. DOI: 10.1007/JHEP09(2014)112. arXiv: 1407.4222  
2663 [hep-ex].
- 2664 [37] Georges Aad et al. “Measurements of the Total and Differential Higgs Boson Production  
2665 Cross Sections Combining the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4\ell$  Decay Channels at  
2666  $\sqrt{s} = 8$  TeV with the ATLAS Detector”. In: *Phys. Rev. Lett.* 115 (2015), p. 091801.  
2667 DOI: 10.1103/PhysRevLett.115.091801. arXiv: 1504.05833 [hep-ex].
- 2668 [38] Vardan Khachatryan et al. “Measurement of differential cross sections for Higgs  
2669 boson production in the diphoton decay channel in  $pp$  collisions at  $\sqrt{s} = 8$  TeV”. In:  
2670 *Eur. Phys. J. C* 76 (2016), p. 13. DOI: 10.1140/epjc/s10052-015-3853-3. arXiv:  
2671 1508.07819 [hep-ex].
- 2672 [39] Vardan Khachatryan et al. “Measurement of differential and integrated fiducial cross  
2673 sections for Higgs boson production in the four-lepton decay channel in  $pp$  collisions  
2674 at  $\sqrt{s} = 7$  and 8 TeV”. In: *JHEP* 04 (2016), p. 005. DOI: 10.1007/JHEP04(2016)005.  
2675 arXiv: 1512.08377 [hep-ex].

- [40] 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].
- [41] 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].
- [42] G. Cowan. “A survey of unfolding methods for particle physics”. In: *Conf. Proc.* C0203181 (2002), p. 248.
- [43] 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].
- [44] 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].
- [45] 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].
- [46] 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].
- [47] 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].
- [48] 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].
- [49] 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].
- [50] 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].
- [51] 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].
- [52] 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].

- 2715 [53] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. “PYTHIA 6.4 physics and  
2716 manual”. In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026.  
2717 arXiv: hep-ph/0603175 [hep-ph].
- 2718 [54] Hung-Liang Lai et al. “Uncertainty induced by QCD coupling in the CTEQ global  
2719 analysis of parton distributions”. In: *Phys. Rev. D* 82 (2010), p. 054021. DOI:  
2720 10.1103/PhysRevD.82.054021. arXiv: 1004.4624 [hep-ph].
- 2721 [55] Hung-Liang Lai et al. “New parton distributions for collider physics”. In: *Phys. Rev.*  
2722 *D* 82 (2010), p. 074024. DOI: 10.1103/PhysRevD.82.074024. arXiv: 1007.2241  
2723 [hep-ph].
- 2724 [56] LHC Higgs Cross Section Working Group. “Handbook of LHC Higgs Cross Sections”.  
2725 In: *arXiv:1101.0593* (2011).
- 2726 [57] Simone Alioli et al. “A general framework for implementing NLO calculations in  
2727 shower Monte Carlo programs: the POWHEG BOX”. In: *JHEP* 06 (2010), p. 043.  
2728 DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].
- 2729 [58] S. Agostinelli et al. “GEANT4: A simulation toolkit”. In: *Nucl. Instrum. Meth. A*  
2730 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.
- 2731 [59] D. de Florian et al. “Higgs boson production at the LHC: transverse momentum  
2732 resummation effects in the  $H \rightarrow 2\gamma$ ,  $H \rightarrow WW \rightarrow l\nu l\nu$  and  $H \rightarrow ZZ \rightarrow 4l$  decay  
2733 modes”. In: *JHEP* 06 (2012), p. 132. DOI: 10.1007/JHEP06(2012)132. arXiv:  
2734 1203.6321 [hep-ph].
- 2735 [60] E. Bagnaschi et al. “Higgs production via gluon fusion in the POWHEG approach in  
2736 the SM and in the MSSM”. In: *JHEP* 02 (2012), p. 088. DOI: 10.1007/JHEP02(2012)  
2737 088. arXiv: 1111.2854 [hep-ph].
- 2738 [61] Stefano Actis et al. “NLO Electroweak Corrections to Higgs Boson Production at  
2739 Hadron Colliders”. In: *Phys. Lett. B* 670 (2008), p. 12. DOI: 10.1016/j.physletb.  
2740 2008.10.018. arXiv: 0809.1301 [hep-ph].
- 2741 [62] Stefano Catani et al. “Soft gluon resummation for Higgs boson production at hadron  
2742 colliders”. In: *JHEP* 07 (2003), p. 028. DOI: 10.1088/1126-6708/2003/07/028.  
2743 arXiv: hep-ph/0306211 [hep-ph].
- 2744 [63] Yanyan Gao et al. “Spin determination of single-produced resonances at hadron  
2745 colliders”. In: *Phys. Rev. D* 81 (2010), p. 075022. DOI: 10.1103/PhysRevD.81.  
2746 075022. arXiv: 1001.3396 [hep-ph].
- 2747 [64] Sara Bolognesi et al. “Spin and parity of a single-produced resonance at the LHC”.  
2748 In: *Phys. Rev. D* 86 (2012), p. 095031. DOI: 10.1103/PhysRevD.86.095031. arXiv:  
2749 1208.4018 [hep-ph].
- 2750 [65] Ian Anderson et al. “Constraining anomalous  $HVV$  interactions at proton and  
2751 lepton colliders”. In: *Phys. Rev. D* 89 (2014), p. 035007. DOI: 10.1103/PhysRevD.  
2752 89.035007. arXiv: 1309.4819 [hep-ph].

- 2753 [66] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. “A Brief Introduction to  
2754 PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (2008), p. 852. DOI: 10.1016/j.  
2755 cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].
- 2756 [67] Vardan Khachatryan et al. “Measurement of the  $t\bar{t}$  production cross section in the  
2757  $e\mu$  channel in proton-proton collisions at  $\sqrt{s} = 7$  and 8 TeV”. In: (2016). arXiv:  
2758 1603.02303 [hep-ex].
- 2759 [68] Stanislaw Jadach, Johann H. Kuhn, and Zbigniew Was. “TAUOLA: A Library of  
2760 Monte Carlo programs to simulate decays of polarized tau leptons”. In: *Comput.  
2761 Phys. Commun.* 64 (1990), p. 275. DOI: 10.1016/0010-4655(91)90038-M.
- 2762 [69] CMS Collaboration. “Standard Model Cross Sections for CMS at 7 TeV”. In: *CMS  
2763 Generator Group Twiki* (2010).
- 2764 [70] John M. Campbell, R. Keith Ellis, and Ciaran Williams. “Vector boson pair pro-  
2765 duction at the LHC”. In: *JHEP* 07 (2011), p. 018. DOI: 10.1007/JHEP07(2011)018.  
2766 arXiv: 1105.0020 [hep-ph].
- 2767 [71] J. Ohnemus. “Order alpha-s calculations of hadronic W+- gamma and Z gamma  
2768 production”. In: *Phys. Rev.* D47 (1993), pp. 940–955. DOI: 10.1103/PhysRevD.47.  
2769 940.
- 2770 [72] JetMET group. “Jet energy uncertainties, [https://twiki.cern.ch/twiki/bin/  
2771 view/CMS/JECUncertaintySources#2012\\_JEC](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](https://twiki.cern.<br/>2772 ch/twiki/bin/view/CMS/JECUncertaintySources%5C#2012_JEC).
- 2773 [73] Sergey Alekhin et al. “The PDF4LHC Working Group Interim Report”. 2011.
- 2774 [74] Michiel Botje et al. “The PDF4LHC Working Group Interim Recommendations”.  
2775 2011.
- 2776 [75] Richard D. Ball et al. “Impact of Heavy Quark Masses on Parton Distributions  
2777 and LHC Phenomenology”. In: *Nucl. Phys. B* 849 (2011), p. 296. DOI: 10.1016/j.  
2778 nuclphysb.2011.03.021. arXiv: 1101.1300 [hep-ph].
- 2779 [76] A. D. Martin et al. “Parton distributions for the LHC”. In: *Eur. Phys. J. C* 63 (2009),  
2780 p. 189. DOI: 10.1140/epjc/s10052-009-1072-5. arXiv: 0901.0002 [hep-ph].
- 2781 [77] Iain W. Stewart and Frank J. Tackmann. “Theory uncertainties for Higgs and other  
2782 searches using jet bins”. In: *Phys. Rev. D* 85 (2012), p. 034011. DOI: 10.1103/  
2783 PhysRevD.85.034011. arXiv: 1107.2117 [hep-ph].
- 2784 [78] “Procedure for the LHC Higgs boson search combination in summer 2011”. In:  
2785 (2011).
- 2786 [79] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”.  
2787 In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur. Phys. J.C73,2501(2013)], p. 1554. DOI:  
2788 10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z. arXiv:  
2789 1007.1727 [physics.data-an].
- 2790 [80] Tim Adye. “Unfolding algorithms and tests using RooUnfold”. 2011.

- [81] 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].
- [82] 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.
- [83] 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].
- [84] 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].
- [85] S. Bolognesi, Y. Gao and A.V. Gritsan *et al.* “JHUGen”. In: URL <http://www.pha.jhu.edu/spin/> (2011).
- [86] 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].
- [87] 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].
- [88] 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].
- [89] 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].
- [90] 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].
- [91] 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].
- [92] 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].
- [93] 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].

- 2829 [94] Vardan Khachatryan et al. “Event generator tunes obtained from underlying event  
2830 and multiparton scattering measurements”. In: (2015). arXiv: 1512.00815 [hep-ex].
- 2831 [95] Peter Richardson and Alexandra Wilcock. “Monte Carlo Simulation of Hard Radia-  
2832 tion in Decays in Beyond the Standard Model Physics in Herwig++”. In: *Eur. Phys.*  
2833 *J. C74* (2014), p. 2713. DOI: 10.1140/epjc/s10052-014-2713-x. arXiv: 1303.4563  
2834 [hep-ph].
- 2835 [96] J. Bellm et al. “Herwig++ 2.7 Release Note”. In: (2013). arXiv: 1310.6877 [hep-ph].
- 2836 [97] *SM Higgs production cross sections at  $\sqrt{s} = 13\text{-}14 \text{ TeV}$* . <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt1314TeV>.
- 2838 [98] Fabrizio Caola et al. “QCD corrections to  $W^+W^-$  production through gluon fusion”.  
2839 In: *Phys. Lett. B754* (2016), pp. 275–280. DOI: 10.1016/j.physletb.2016.01.046.  
2840 arXiv: 1511.08617 [hep-ph].
- 2841 [99] *NLO single-top channel cross sections*. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec>.
- 2843 [100] *NNLO+NNLL top-quark-pair cross sections*. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO>.
- 2845 [101] *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep.  
2846 CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. Geneva: CERN, Aug. 2011.
- 2847 [102] *LHC Higgs Cross Section Working Group*. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>.
- 2849 [103] Jon Butterworth et al. “PDF4LHC recommendations for LHC Run II”. In: *J. Phys.*  
2850 *G43* (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865  
2851 [hep-ph].
- 2852 [104] Vardan Khachatryan et al. “Search for a Higgs Boson in the Mass Range from 145  
2853 to 1000 GeV Decaying to a Pair of W or Z Bosons”. In: *JHEP* 10 (2015), p. 144.  
2854 DOI: 10.1007/JHEP10(2015)144. arXiv: 1504.00936 [hep-ex].
- 2855 [105] G. C. Branco et al. “Theory and phenomenology of two-Higgs-doublet models”.  
2856 In: *Phys. Rept.* 516 (2012), p. 1. DOI: 10.1016/j.physrep.2012.02.002. arXiv:  
2857 1106.0034 [hep-ph].
- 2858 [106] N. Craig and T. Scott. “Exclusive signals of an extended Higgs sector”. In: *JHEP*  
2859 11 (2012), p. 083. DOI: 10.1007/JHEP11(2012)083. arXiv: 1207.4835 [hep-ph].
- 2860 [107] H. Haber and O. Stal. “New LHC benchmarks for the CP-conserving two-Higgs-  
2861 doublet model”. In: *Eur. Phys. J. C* 75 (2015), p. 491. DOI: 10.1140/epjc/s10052-  
2862 015-3697-x. arXiv: 1507.04281 [hep-ph].
- 2863 [108] Suyong Choi, Sunghoon Jung, and P. Ko. “Implications of LHC data on 125 GeV  
2864 Higgs-like boson for the Standard Model and its various extensions”. In: *JHEP* 1310  
2865 (2013), p. 225. DOI: 10.1007/JHEP10(2013)225. arXiv: 1307.3948.

- 2866 [109] Tania Robens and Tim Stefaniak. “Status of the Higgs singlet extension of the  
2867 standard model after LHC run 1”. In: *Eur. Phys. J. C* 75 (2015), p. 105. DOI:  
2868 10.1140/epjc/s10052-015-3323-y. arXiv: 1501.02234 [hep-ph].
- 2869 [110] Johan Alwall et al. “Comparative study of various algorithms for the merging of  
2870 parton showers and matrix elements in hadronic collisions”. In: *Eur. Phys. J. C* 53  
2871 (2008), pp. 473–500. DOI: 10.1140/epjc/s10052-007-0490-5. arXiv: 0706.2569  
2872 [hep-ph].
- 2873 [111] Massimiliano Grazzini. “NNLO predictions for the Higgs boson signal in the  $H \rightarrow WW \rightarrow$   
2874  $\ell\nu\ell\nu$  and  $H \rightarrow ZZ \rightarrow 4\ell$  decay channels”. In: *JHEP* 02 (2008), p. 043. DOI: 10.1088/  
2875 1126-6708/2008/02/043. arXiv: 0801.3232 [hep-ph].
- 2876 [112] Vardan Khachatryan et al. “Measurement of the differential cross section for top  
2877 quark pair production in pp collisions at  $\sqrt{s} = 8$  TeV”. In: *Eur. Phys. J. C* 75.11  
2878 (2015), p. 542. DOI: 10.1140/epjc/s10052-015-3709-x. arXiv: 1505.04480  
2879 [hep-ex].

