

1

# Measurements of the Higgs boson 2 decay to $W^+W^-$ with the CMS 3 detector

4 Lorenzo Viliani  
5 of University of Florence

6 PhD Thesis  
7



---

## Abstract

8

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



# <sup>24</sup> Contents

<b>25</b>	<b>1. Electroweak and QCD physics at LHC</b>	<b>3</b>
26	1.1. The Standard Model of particle physics . . . . .	3
27	1.2. The Higgs mechanism . . . . .	4
28	1.3. The strong interaction . . . . .	4
29	1.4. Phenomenology of proton proton interactions . . . . .	4
30	1.5. Monte Carlo simulations . . . . .	4
31	1.6. Experimental results of the Higgs boson properties . . . . .	4
<b>32</b>	<b>2. The CMS experiment at the LHC</b>	<b>5</b>
33	2.1. The Large Hadron Collider . . . . .	5
34	2.2. The <i>Compact Muon Solenoid</i> experiment . . . . .	9
35	2.2.1. The solenoid . . . . .	12
36	2.2.2. The tracker . . . . .	13
37	2.2.3. The electromagnetic calorimeter (ECAL) . . . . .	17
38	2.2.4. The hadron calorimeter (HCAL) . . . . .	19
39	2.2.5. The muon system . . . . .	20
40	2.3. The CMS trigger system . . . . .	23
41	2.3.1. The Level-1 trigger . . . . .	23
42	2.3.2. The high level trigger (HLT) . . . . .	25
<b>43</b>	<b>3. Reconstruction and identification of physics objetc</b> s	<b>27</b>
44	3.1. The Particle Flow technique . . . . .	27
45	3.2. Leptons reconstruction and identification . . . . .	29
46	3.2.1. Muon reconstruction and identification . . . . .	29
47	3.2.2. Muon isolation . . . . .	30
48	3.2.3. Muon momentum scale and resolution . . . . .	31
49	3.2.4. Electron reconstruction and identification . . . . .	32
50	3.2.5. Electron isolation . . . . .	32
51	3.2.6. Electron momentum scale and resolution . . . . .	33
52	3.2.7. Lepton identification and isolation efficiency . . . . .	34
53	3.2.8. Lepton trigger efficiency . . . . .	35
54	3.3. Jets reconstruction and identification . . . . .	36
55	3.3.1. Jet reconstruction in CMS . . . . .	37
56	3.3.2. Jet energy correction . . . . .	39
57	3.4. Jet b tagging . . . . .	42

---

58	3.5. Missing transverse energy . . . . .	44
59	3.5.1. $E_T^{\text{miss}}$ scale and resolution measurement . . . . .	46
60	<b>4. Measurement of the Higgs boson transverse momentum at 8 TeV using <math>H \rightarrow WW \rightarrow 2\ell 2\nu</math> decays</b>	<b>49</b>
61	4.1. Introduction . . . . .	49
62	4.2. Data sets, triggers and MC samples . . . . .	51
63	4.2.1. Data sets and triggers . . . . .	51
64	4.2.2. Monte Carlo samples . . . . .	52
65	4.3. Analysis Strategy . . . . .	53
66	4.3.1. Event reconstruction and selections . . . . .	54
67	4.3.2. Simulation efficiencies and scale factors . . . . .	57
68	4.3.3. Fiducial phase space . . . . .	61
69	4.3.4. Binning of the $p_T^H$ distribution . . . . .	62
70	4.4. Background estimation . . . . .	63
71	4.4.1. Top quark background . . . . .	63
72	4.4.2. WW background . . . . .	65
73	4.4.3. W+jets background . . . . .	66
74	4.4.4. $Z/\gamma^* \rightarrow \tau^+\tau^-$ background . . . . .	70
75	4.5. Systematic uncertainties . . . . .	71
76	4.5.1. Background normalization uncertainties . . . . .	72
77	4.5.2. Experimental uncertainties . . . . .	72
78	4.5.3. Theoretical uncertainties . . . . .	73
79	4.5.4. Statistics uncertainty of the simulated samples . . . . .	76
80	4.5.5. Treatment of systematic uncertainties in the shape analysis . . . . .	76
81	4.6. Signal extraction . . . . .	77
82	4.6.1. Fitting procedure . . . . .	77
83	4.6.2. Signal and background yields . . . . .	80
84	4.7. Unfolding . . . . .	83
85	4.7.1. Treatment of systematic uncertainties . . . . .	87
86	4.8. Results . . . . .	89
88	<b>5. Search for the SM Higgs boson in the <math>H \rightarrow WW</math> channel with the first 13 TeV LHC data</b>	<b>93</b>
89	5.1. Introduction . . . . .	93
90	5.2. Data and simulated samples . . . . .	93
91	5.3. Analysis strategy . . . . .	98
92	5.3.1. Event reconstruction . . . . .	98
93	5.3.2. Event selection and background rejection . . . . .	103
94	5.3.3. Signal extraction . . . . .	105
95	5.4. Background estimation . . . . .	110
96	5.4.1. WW background . . . . .	110
97	5.4.2. Top quark background . . . . .	110

99	5.4.3. Jet-induced (or Fake) background . . . . .	112
100	5.4.4. DY background . . . . .	112
101	5.4.5. Other backgrounds . . . . .	113
102	5.5. Systematic uncertainties . . . . .	114
103	5.6. Results . . . . .	116
104	<b>6. Search for high mass resonances decaying to a W boson pair with first 13 TeV LHC data</b>	<b>119</b>
106	6.1. Introduction . . . . .	119
107	6.2. Data and simulated samples . . . . .	119
108	6.3. Analysis strategy . . . . .	121
109	6.4. Background estimation . . . . .	124
110	6.5. Systematic uncertainties . . . . .	127
111	6.6. Signal extraction and limit setting . . . . .	127
112	6.7. Results . . . . .	131
113	<b>7. Conclusions</b>	<b>137</b>
114	<b>A. Fiducial region definition and optimization</b>	<b>139</b>
115	Bibliography . . . . .	147







# Chapter 1.

## <sup>117</sup> Electroweak and QCD physics at <sup>118</sup> LHC

<sup>119</sup> In this chapter the Standard Model of particle physics is briefly described, in particular  
<sup>120</sup> the main characteristics of the electroweak and strong interactions are discussed, as well as  
<sup>121</sup> the mechanism of the electroweak symmetry breaking. In addition, the phenomenology  
<sup>122</sup> of the proton-proton interactions is described and the main features of the Monte Carlo  
<sup>123</sup> simulation techniques are given. Eventually, an overview of the Higgs boson physics from  
<sup>124</sup> an experimental point of view is reported.

### <sup>125</sup> 1.1. The Standard Model of particle physics

<sup>126</sup> The Standard Model of Particle Physics (SM) is the theory that describes all fundamental  
<sup>127</sup> constituents of matter and their interactions. It is a renormalisable quantum field theory  
<sup>128</sup> based on a  $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$  local gauge symmetry, and is capable to provide  
<sup>129</sup> a quantitative description of three of the four interactions in nature: electromagnetism,  
<sup>130</sup> weak interaction and strong nuclear force. The aforementioned symmetry holds only if  
<sup>131</sup> the fermion fields are massless, in contrast with the experimental observations of massive  
<sup>132</sup> fermions. A mechanism, known as *spontaneous symmetry breaking*, is introduced in the SM  
<sup>133</sup> allowing the elementary particles to acquire mass. This mechanism requires the presence  
<sup>134</sup> of a new field, known as Higgs field. During the past decades the predictions of the SM  
<sup>135</sup> have been confirmed by experimental results with outstanding precision, and in 2012 the  
<sup>136</sup> existence of a new boson consistent with the predicted Higgs boson was announced by the  
<sup>137</sup> ATLAS and CMS experiments at CERN.

<sup>138</sup> According to the SM, the ordinary matter is made up of spin-1/2 particles, denoted  
<sup>139</sup> as fermions. The fermions are subdivided into two classifications of elementary particles:  
<sup>140</sup> leptons and quarks. Both classes consist of six particles, grouped into three doublets, called  
<sup>141</sup> generations. Additional three doublets for each class are composed of leptons and quarks  
<sup>142</sup> antiparticles. A charged particle with electric charge  $Q = -1$ , either the electron  $e$ , the  
<sup>143</sup> muon  $\mu$  or the tauon  $\tau$ , and a neutral particle, the corresponding neutrino, compose the  
<sup>144</sup> following lepton generations, ordered according to an increasing mass hierarchy:

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix}, \quad \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix}, \quad \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix} . \quad (1.1)$$

145 Charged leptons can interact via the electromagnetic and weak force, while neutrinos,  
146 that are assumed to be massless, can interact only through the weak interaction.

147 Similarly, the quarks are organized in pairs composed of a particle with  $Q = +2/3$ , *up*  
148 (*u*), *charm* (*c*) and *top* (*t*) quarks, and another particle with  $Q = -1/3$ , *down* (*d*), *strange*  
149 (*s*) and *bottom* (*b*) quarks:

$$\begin{pmatrix} u \\ d \end{pmatrix}, \quad \begin{pmatrix} c \\ s \end{pmatrix}, \quad \begin{pmatrix} t \\ b \end{pmatrix} . \quad (1.2)$$

150 As well as leptons, quarks can interact via the electromagnetic and weak forces, but  
151 also via the strong interaction, responsible of their confinement within hadrons. In fact,  
152 free quarks are not observed in nature, but they bind together forming two categories of  
153 hadrons: mesons, bound states of a quark *q* and an anti-quark  $\bar{q}$ , and baryons, bound states  
154 of three quarks.

155 In the SM the interaction between elementary particles occurs through the exchange of  
156 spin-1 particles, known as bosons, which identify the fundamental forces. The photon  $\gamma$  is  
157 the mediator of the electromagnetic interaction, the  $W^\pm$  and  $Z$  bosons are the mediators of  
158 the weak interaction, while the strong force is mediated by eight gluons *g*. Electromagnetic  
159 and weak interactions are actually the manifestations of the same fundamental interaction,  
160 the electroweak force.

## 161 1.2. The Higgs mechanism

## 162 1.3. The strong interaction

## 163 1.4. Phenomenology of proton proton interactions

## 164 1.5. Monte Carlo simulations

## 165 1.6. Experimental results of the Higgs boson 166 properties

# Chapter 2.

## <sup>167</sup> The CMS experiment at the LHC

<sup>168</sup> In this chapter, the main characteristics of the Large Hadron Collider (LHC particle  
<sup>169</sup> accelerator and Compact Muon Solenoid (CMS) experiment are described.

### <sup>170</sup> 2.1. The Large Hadron Collider

<sup>171</sup> The LHC [1–4] at CERN, officially inaugurated on 21<sup>st</sup> October 2008, is the largest and  
<sup>172</sup> most powerful hadron collider ever built. Installed in the underground tunnel which hosted  
<sup>173</sup> the Large Electron Positron Collider (LEP) [5–7], the leptonic accelerator in operation until  
<sup>174</sup> 2nd November 2000, the LHC accelerator has the shape of a circle with a length of about  
<sup>175</sup> 27 km and is located underground at a depth varying between 50 m to 175 m, straddling the  
<sup>176</sup> Franco-Swiss border near Geneva. It is designed to collide two 7 TeV counter-circulating  
<sup>177</sup> beams of protons resulting in a center-of-mass energy of 14 TeV, or two beams of heavy  
<sup>178</sup> ions, in particular lead nuclei at an energy of 2.76 TeV/nucleon in the center-of-mass frame.

<sup>179</sup> The transition from a leptonic collider to a hadronic collider entailed the following  
<sup>180</sup> advantages: first, it has been possible to build a machine that having the same size of the  
<sup>181</sup> previous one (and therefore accommodated in the same LEP tunnel, substantially reducing  
<sup>182</sup> the cost and time of construction), could reach a higher energy in the center-of-mass  
<sup>183</sup> frame. This is due to the much lower amount of energy loss through synchrotron radiation  
<sup>184</sup> emitted by the accelerated particles, that is proportional to the fourth power of the ratio  
<sup>185</sup>  $E/m$  between their energy and their mass. Secondly, the composite structure of protons  
<sup>186</sup> compared to the elementary structure of electrons allows LHC to be able to simultaneously  
<sup>187</sup> access a wider energy spectrum, despite the production of many low energies particles in a  
<sup>188</sup> complex environment. This is a particularly important feature for a machine dedicated to  
<sup>189</sup> the discovery of “new” physics.

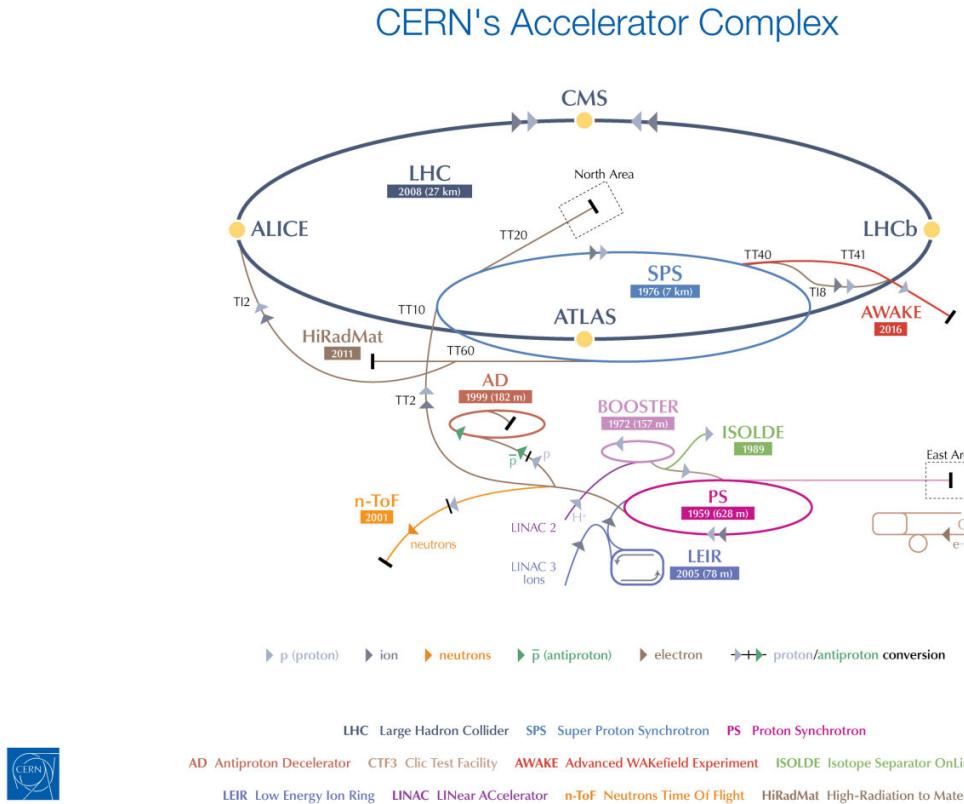
<sup>190</sup> In Fig. 2.1 a schematic description of the accelerator complex installed at CERN is shown.  
<sup>191</sup> The acceleration is performed in several stages [4]. The protons source is a *Duoplasmatron*:  
<sup>192</sup> the protons are obtained by removing electrons from a source of hydrogen gas and then  
<sup>193</sup> sent to the LINAC2, a 36 m long linear accelerator which generates a pulsed beam with  
<sup>194</sup> an energy of 50 MeV using Radio Frequency Quadrupoles (RFQ) and focusing quadrupole  
<sup>195</sup> magnets. The beam is subsequently sent to the Proton Synchrotron Booster (PSB), a  
<sup>196</sup> circular accelerator consisting of four superimposed synchrotron rings with a circumference  
<sup>197</sup> of about 160 m, which increases the proton energy up to 1.4 GeV. Then, protons are

198 injected into the Proton Synchrotron (PS), a single synchrotron ring with a circumference  
 199 of about 600 m where the energy is increased to 25 GeV. The sequential combination of  
 200 these two synchrotrons also allows to create a series of protons bunches interspersed by  
 201 25 ns as required for the correct operation of LHC. The final proton injection stage is the  
 202 Super Proton Synchrotron (SPS), a synchrotron with a circumference of approximately  
 203 7 km where protons reach an energy value of 450 GeV. Subsequently, protons are extracted  
 204 and injected into the LHC ring via two transmission lines, to generate two beams running  
 205 in opposite directions in two parallel pipes and which are accelerated up to the energy of  
 206 interest. In the two pipes an ultrahigh vacuum condition is maintained (about  $10^{-10}$  Torr)  
 207 to avoid the spurious proton interactions with the gas remnants. At full intensity, each  
 208 proton beam consists of 2808 bunches and each bunch contains around  $10^{11}$  protons. The  
 209 beams are squeezed and collide for a length of about 130 m at four interaction points where  
 210 the four main experiments (ALICE, ATLAS, CMS and LHCb) are placed:

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

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

233 A series of about 1200 magnetic dipoles bend the beams along the accelerator ring.  
 234 They are located along the “arc” structures of the circumference. The ring, in fact, can  
 235 be subdivided into octants, with eight curve regions (the “arcs”) separated by rectilinear  
 236 regions. In these straight regions, instead, almost 400 focusing and defocusing quadrupoles  
 237 are located, which maintain the beam stable along the orbit, and some other small multipolar  
 238 magnets (sextupoles and octupoles) are used to make additional minor corrections to the  
 239 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.

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

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

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

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

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

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

As the instantaneous luminosity increases, the probability of multiple proton proton interactions to occur in a single bunch crossing grows higher as well. In this instance, the main goal is the identification and reconstruction of a single primary collision where the physics event of interest occurs among the background of the additional proton proton interactions. Such backgrounds are due to processes occurring with very high probability, like the production of low- $p_T$  jets. These additional collisions are known as pile up (PU). During the LHC current run the average number of pile up events is 23, with some event exhibiting over 45 pile up collisions.

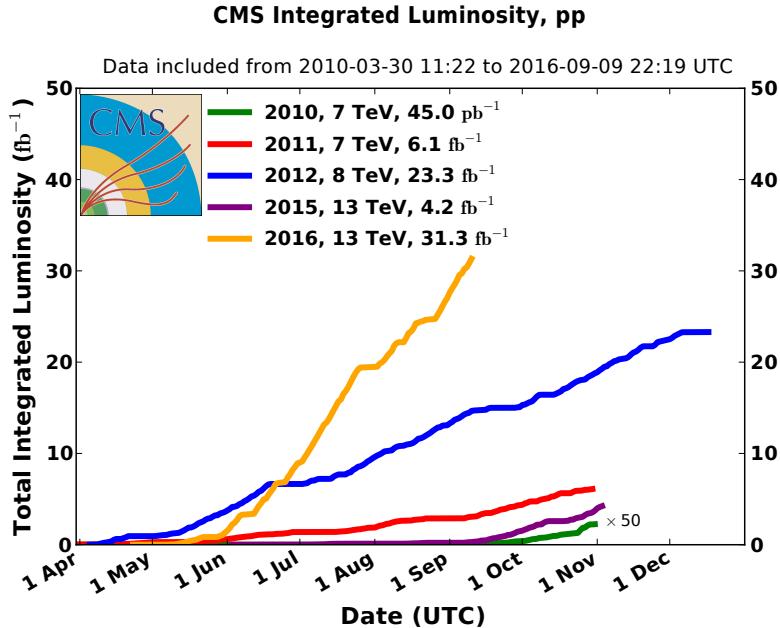
**Table 2.1.:** LHC technical parameters for proton proton collisions.

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

## 2.2. The *Compact Muon Solenoid* experiment

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

<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.

289 3.8 T magnetic field. Such a strong magnetic field is the key aspect which permits to have  
 290 a compact design of the detector. The detector has a cylindrical structure, which is typical  
 291 of general purpose detectors, which consists of several cylindrical detecting layers, coaxial  
 292 with the beam direction (*barrel* region), closed at both ends with detecting disks (*endcap*  
 293 region), in such a way to ensure the hermetic closure of the apparatus.

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

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

303 where  $\theta$  is the polar angle. The used of pseudorapidity is preferred over the polar angle  
 304 because differences in pseudorapidity are Lorentz invariant under boosts along the  $z$ -axis.

305 In the limit of ultrarelativistic particles the pseudorapidity coincides with the rapidity  $y$ :

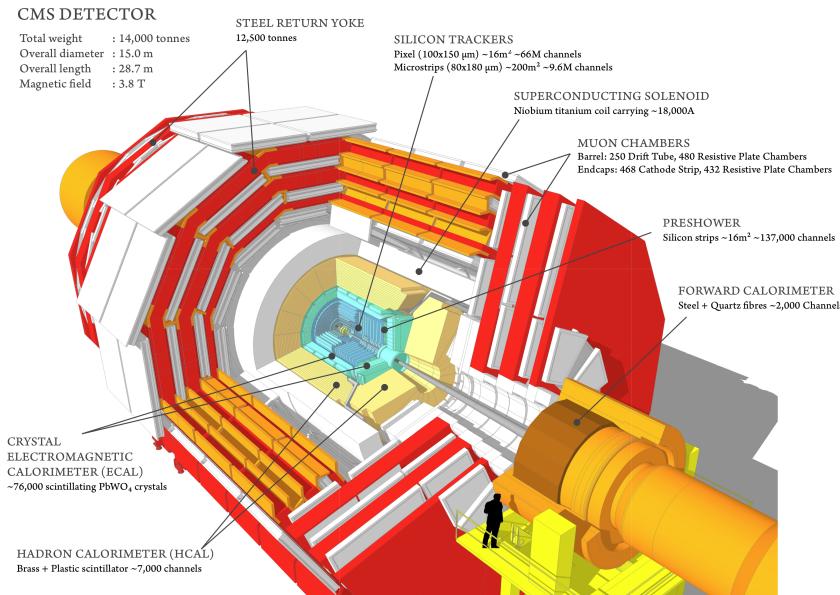
$$y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right) \quad , \quad (2.5)$$

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

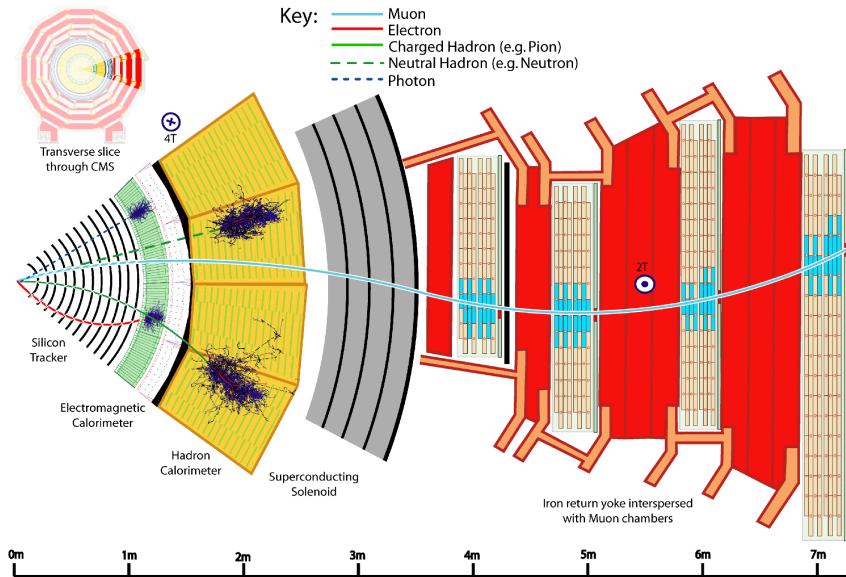
307 The schematic view of the CMS detector, which has a length of 21.5 m, a diameter of  
 308 15 m and a weight of about 14000 tons, is shown in Fig. 2.3. From the inner region to the  
 309 outer one, the various CMS sub-detectors are:

- 310 • **Silicon tracker:** it occupies the region  $r < 1.2$  m and  $|\eta| < 2.5$ . It is composed of an  
 311 inner silicon pixel vertex detector and a surrounding silicon microstrip detector, with  
 312 a total active area of about  $215 \text{ m}^2$ . It is used to reconstruct charged particle tracks  
 313 and vertices;
- 314 • **Electromagnetic calorimeter (ECAL):** placed in the region  $1.2 \text{ m} < r < 1.8$  m  
 315 and  $|\eta| < 3$ , it consists of many scintillating crystals of lead tungstate ( $\text{PbWO}_4$ ). It is  
 316 used for the measurement of the trajectory and the energy released by electrons and  
 317 photons;
- 318 • **Hadronic calorimeter (HCAL):** it is placed in the region  $1.8 \text{ m} < r < 2.9$  m and  
 319  $|\eta| < 5$ . It is made up of brass layers alternated with plastic scintillators and it is  
 320 used to measure the direction and energy deposited by the hadrons produced in the  
 321 interactions;
- 322 • **Superconducting solenoidal magnet:** it occupies the region  $2.9 \text{ m} < r < 3.8$  m  
 323 and  $|\eta| < 1.5$  and generates an internal uniform magnetic field with an intensity of  
 324 3.8 T, pointing along the direction of the beams. The magnetic field is necessary to  
 325 bend the trajectories of charged particles, in order to allow the measurement of their  
 326 momentum through the curvature observed in the tracking system. The magnetic  
 327 field lines are closed by an external 21 m long iron yoke, that has a diameter of 14 m.  
 328 Outside the return yoke, a residual 1.8 T magnetic field is present, pointing in the  
 329 opposite direction with respect to the internal field;
- 330 • **Muon system:** the outermost system, which is placed in the region  $4 \text{ m} < r < 7.4$  m  
 331 and  $|\eta| < 2.4$ , has the purpose of reconstructing the tracks of muons passing through  
 332 it. It consists of Drift Tubes (DT) in the barrel region and Cathode Strip Chambers  
 333 (CSC) in the endcaps. A complementary system of Resistive Plate Chambers (RPC)  
 334 is used both in the barrel and endcaps. The muon chambers are housed inside the  
 335 iron structure of the return yoke.

336 In Fig. 2.4 the response of the various CMS sub-detectors to the passage of different types  
 337 of particles is sketched. In the following sections a brief description of each sub-detector is  
 338 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.

### 339 2.2.1. The solenoid

340 The CMS magnet [14], which contains the tracker, the electromagnetic and the hadronic  
 341 calorimeters, is the biggest superconducting solenoid ever built. The solenoid can generate  
 342 a magnetic field of 3.8 T in the internal bore, which has a diameter of 6 m and a length of

343 12.5 m. The energy stored in the magnet is about 2.7 GJ at full current. The superconductor  
 344 is made of four Niobium-Titanium layers and it is cooled down to about 4 K through a  
 345 liquid Helium cooling plant. In case of a quench, when the magnet loses its superconducting  
 346 property, the energy is dumped to resistors within 200 ms. The magnet return yoke of  
 347 the barrel is composed with three sections along the  $z$ -axis; each one is split into 4 layers  
 348 (holding the muon chambers in the gaps). Most of the iron volume is saturated or nearly  
 349 saturated, and the field in the yoke is about the half (1.8 T) of the field in the central  
 350 volume.

### 351 2.2.2. The tracker

352 The silicon tracker is the detector closest to the beam collision point. Its goal is the  
 353 high resolution reconstruction of the trajectories of charged particles originating from the  
 354 interaction point and the identification of the position of secondary vertices produced by  
 355 particles with a short mean life time (in particular hadrons containing the b quark, that  
 356 decay after few hundred of  $\mu\text{m}$ ). The events produced in the proton proton collisions can  
 357 be very complex and track reconstruction is an entangled pattern recognition problem.  
 358 Indeed, at the nominal instantaneous luminosity of operation, an average of about 20 pile up  
 359 events overlapping to the event of interest are expected, leading to about 1000 tracks to be  
 360 reconstructed every 25 ns. In order to make the pattern recognition easier, two requirements  
 361 are fundamental:

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

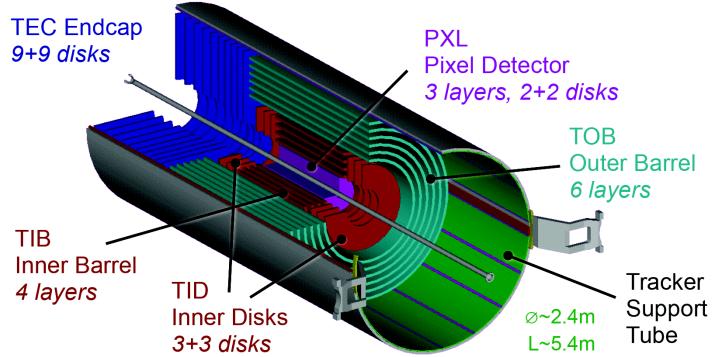
364 The first requirement is achieved building a detector with high granularity<sup>2</sup>. The redundancy  
 365 of the hits is instead achieved having several detecting layers, and is necessary to reduce the  
 366 ambiguity on the assignment of the hits to a given track. Nevertheless, the amount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

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

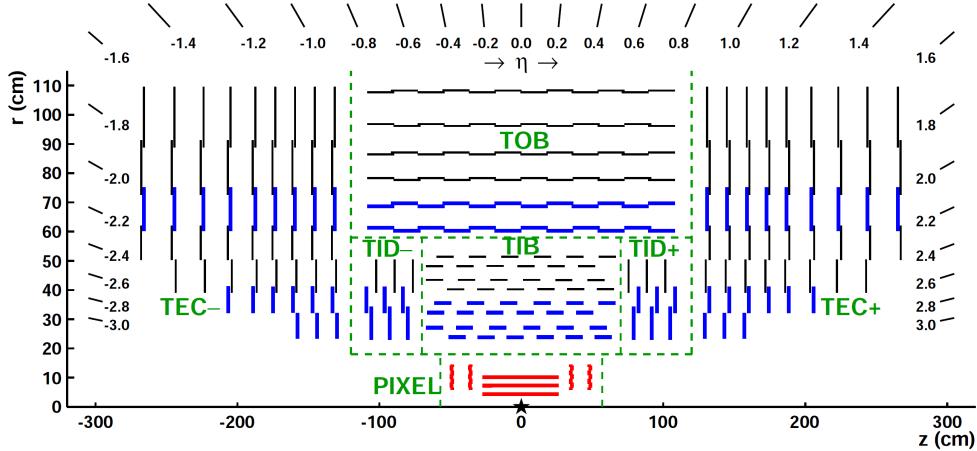
---

<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.

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



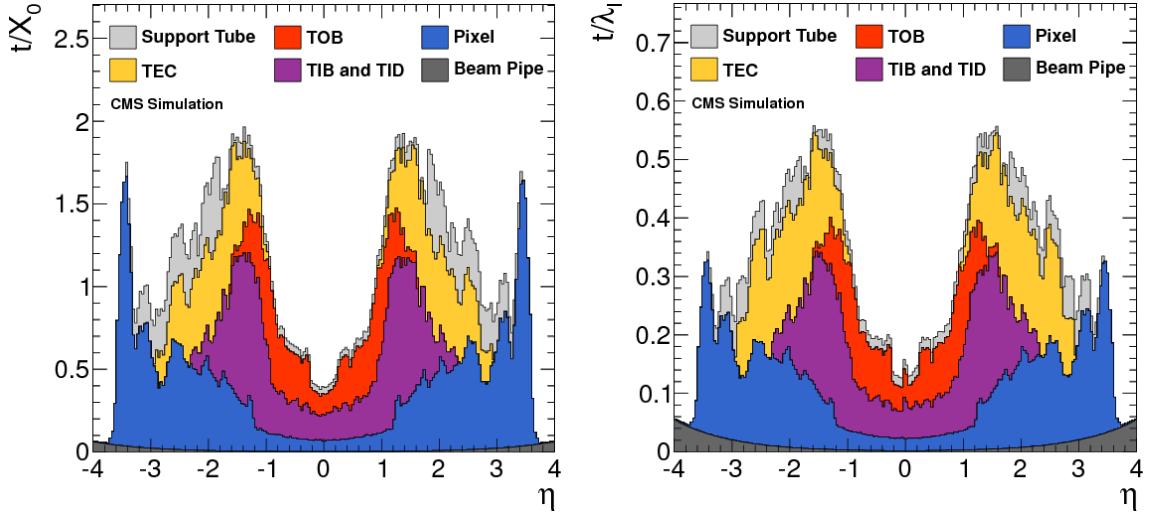
**Figure 2.5.:** Three-dimensional schematic view of the CMS silicon tracker.



**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.

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

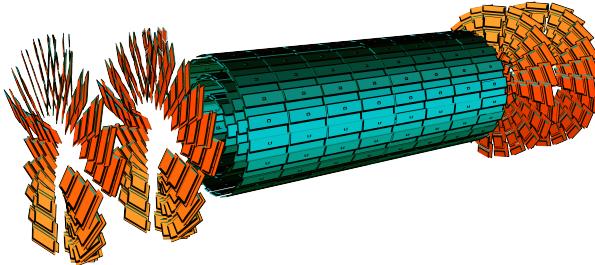
The material budget of the tracker as obtained from a simulation of the detector is shown in Fig. 2.7, reported both in units of radiation length  $t/X_0$  and in units of nuclear interaction length  $t/\lambda_I$ , as a function of  $\eta$ . The region  $1 < |\eta| < 2$  exhibits a larger material budget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

### 390 The pixel detector

391 The pixel detector, shown in Fig. 2.8, is mainly used as starting point in the CMS track  
 392 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.

393 where the amount of radiation is larger. It is placed in the region  $|\eta| < 2.5$  and consists of  
 394 three cylindrical layers 53 cm long in the barrel region, located at  $r = 4.4, 7.3$  and  $10.2$  cm,  
 395 and two pairs of endcap disks with radii between 6 and 15 cm at  $z = \pm 34.5$  and  $\pm 46.5$  cm,  
 396 covering a total area of about  $1\text{ m}^2$ . The detector is composed of many modules, for a total  
 397 of 768 in the barrel and 672 in the endcaps. Each endcap is composed of 24 segments,  
 398 each one tilted with respect to the adjacent ones and containing 7 modules. Each module  
 399 consists of several units which contain a highly segmented silicon sensor with a thickness of  
 400  $250\text{ }\mu\text{m}$ . In order to achieve an optimal vertex position resolution in both the  $(r, \phi)$  and

402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441

$z$ -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

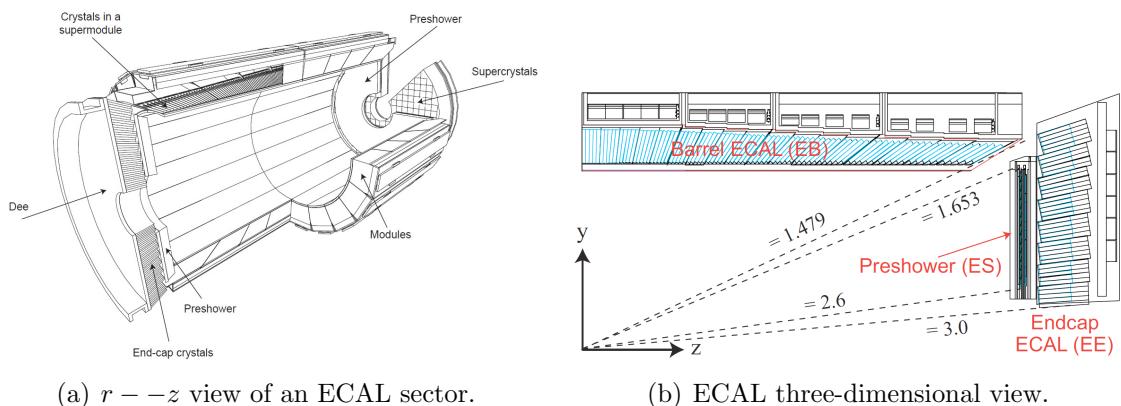
a number ranging from 4 to 7 modules are mounted in a ring shape, depending on the position along  $z$ . Also in this case the modules are alternately mounted on the internal and on the external side of the disk, with the strips radially oriented. On the two innermost rings and on the fifth one the stereo 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.



(a)  $r - z$  view of an ECAL sector.

(b) ECAL three-dimensional view.

**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  $R_M$ .

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

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

507 where  $E$  is the energy measured in GeV.

#### 508 2.2.4. The hadron calorimeter (HCAL)

509 The hadron calorimeter (HCAL) [17] is used together with ECAL to make a complete  
 510 calorimetric system for the jet energy and direction measurement. Moreover, thanks  
 511 to its hermetic structure, it can measure the energy imbalance in the transverse plane,  
 512  $E_T^{\text{miss}}$ , a typical signature of non interacting particles, such as neutrinos. The HCAL is  
 513 a sampling calorimeter covering the region  $|\eta| < 5$ . As shown in Fig. 2.10, it is divided  
 514 in four sub-detectors: HB (*Barrel Hadronic Calorimeter*), located in the barrel region  
 515 inside the solenoid, extending up to  $|\eta| < 1.4$ ; HE (*Endcap Hadronic Calorimeter*), placed  
 516 in the endcaps region inside the magnet, covering the region  $1.3 < |\eta| < 3$  and partially  
 517 overlapping with the HB coverage; HO (*Outer Hadronic Calorimeter*), also known as  
 518 *tail-catcher*, placed along the inner wall of the magnetic field return yoke, just outside of  
 519 the magnet; HF (*Forward Hadronic Calorimeter*), a sampling calorimeter consisting of  
 520 quartz fibers sandwiched between iron absorbers, consisting of two units placed in the very  
 521 forward region ( $3 < |\eta| < 5$ ) outside the magnetic coil. The quartz fibers emit Cherenkov  
 522 light with the passage of charged particles and this light is detected by radiation resistant  
 523 photomultipliers. In order to maximize particle containment for a precise missing transverse  
 524 energy measurement, the amount of absorber material was maximized, reducing therefore  
 525 the amount of the active material. Since HCAL is mostly placed inside the magnetic coil,  
 526 a non-magnetic material like brass was chosen as absorber. HB and HE are therefore  
 527 made with 5 cm brass absorber layers interleaved with 3.7 mm plastic scintillators. The  
 528 scintillation light is collected by wavelength shifting (WLS) fibres and read out by hybrid  
 529 photodiodes (HPD). The granularity of the calorimeter is  $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$  for  
 530  $|\eta| < 1.6$  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  
 531 the wheel that compose the return yoke and is divided in 12 sectors, each one covering  
 532 a  $30^\circ$  angle in  $\phi$ . It consists of scintillating layers, with the same granularity as HB, and  
 533 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.

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

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

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

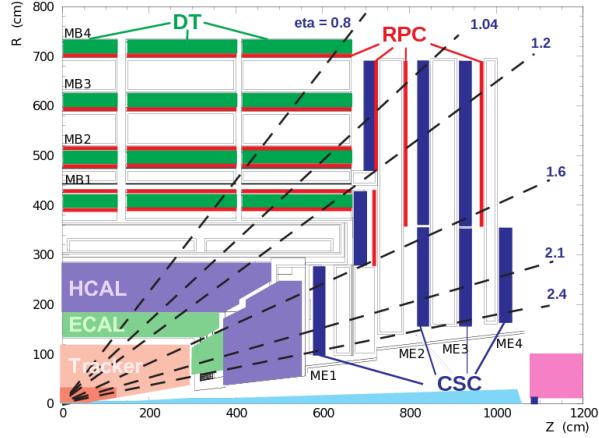
538 where  $E$  is expressed in GeV.

### 539 2.2.5. The muon system

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

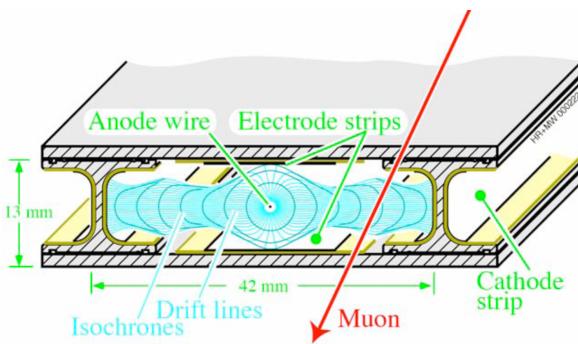
- 544 • *Drift Tubes* (DT) are placed in the barrel region, where the occupancy is relatively  
545 low ( $< 10 \text{ Hz/m}^2$ );
- 546 • *Cathode Strip Chambers* (CSC) are installed in the endcaps, where the occupancy is  
547 higher ( $> 100 \text{ Hz/m}^2$ );
- 548 • *Resistive Plate Chambers* (RPC)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

549 The DT system is placed in the region of the barrel with  $|\eta| < 1.2$ , where the magnetic  
550 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.

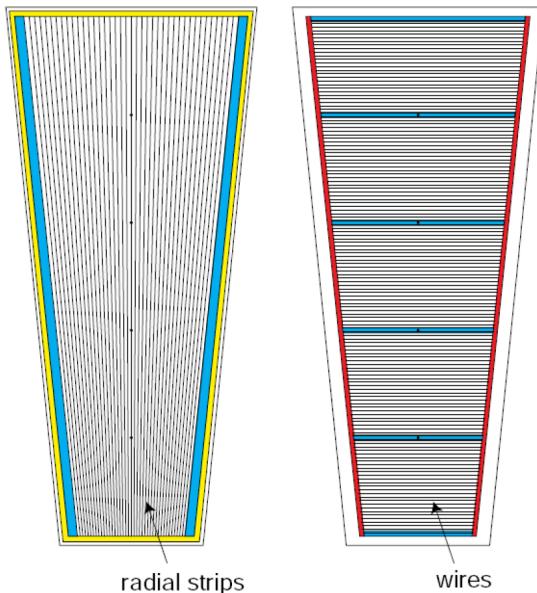
region is divided in 5 wheels, which are subdivided in 12 sectors covering a  $30^\circ$  azimuthal angle each. The wheels are composed of 4 concentric rings of chambers, called *stations*, interspersed in the layers of the iron return yoke, and each one formed by 12 DT chambers. The basic element of 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.

564 In the endcaps, the high and non-uniform magnetic field and the particle rate do not  
 565 allow to use drift tubes detectors to perform measurements. Therefore, a solution based  
 566 on the CSC detector has been adopted. CSC are multi-wire proportional chambers with  
 567 the cathodes segmented into strips oriented radially and transversely with respect to the  
 568 anode wires (see Fig. 2.13), allowing a simultaneous measurement of two coordinates ( $r$   
 569 through the wires and  $\phi$  using the strips). The CSC chambers are filled with a gas mixture  
 570 of Ar (40%), CO<sub>2</sub> (50%) and CF<sub>4</sub> (10%) and provide a spatial resolution of about 80–85  $\mu\text{m}$ .  
 571 The drift path of the charge carriers is shorter with respect to the drift tubes, therefore  
 572 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.

573  
 574     RPCs are used both in barrel and endcaps, complementing DT and CSC systems, in  
 575 order to ensure robustness and redundancy to the muon spectrometer. RPCs are gaseous  
 576 detectors characterized by a coarse spatial resolution, but are able to perform precise time  
 577 measurements, comparable with the ones provided by scintillators. These chambers are  
 578 made of 4 bakelite planes, with a bulk resistivity of  $10^{10}$ – $10^{11} \Omega\text{cm}$ . The 2 mm gap between  
 579 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  
 580 chamber is equipped with insulated aluminum strips, used to collect the signal generated by  
 581 crossing particles. In the barrel the strips are rectangularly segmented and run along the  
 582 beam axis, whereas the endcaps are equipped with trapezoidal shaped strips. The detector  
 583 operates in avalanche mode, and covers the region  $|\eta| < 2.1$ .

## 584 2.3. The CMS trigger system

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

595 Level-1 trigger runs on dedicated processors, and accesses coarse level granularity  
 596 information from calorimetry and muon system. A L1 Trigger decision has to be taken for  
 597 each bunch crossing within  $3.2\ \mu\text{s}$ . Its task is to reduce the data flow from 40 MHz to about  
 598 100 kHz.

599 The High Level Trigger is responsible for reducing the L1 output rate down to a maximum  
 600 rate of the order of 1 kHz. The HLT code runs on a farm of commercial processors and can  
 601 access the full granularity information of all the sub-detectors.

602 The main characteristics of the CMS trigger system are described in the following.

### 603 2.3.1. The Level-1 trigger

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

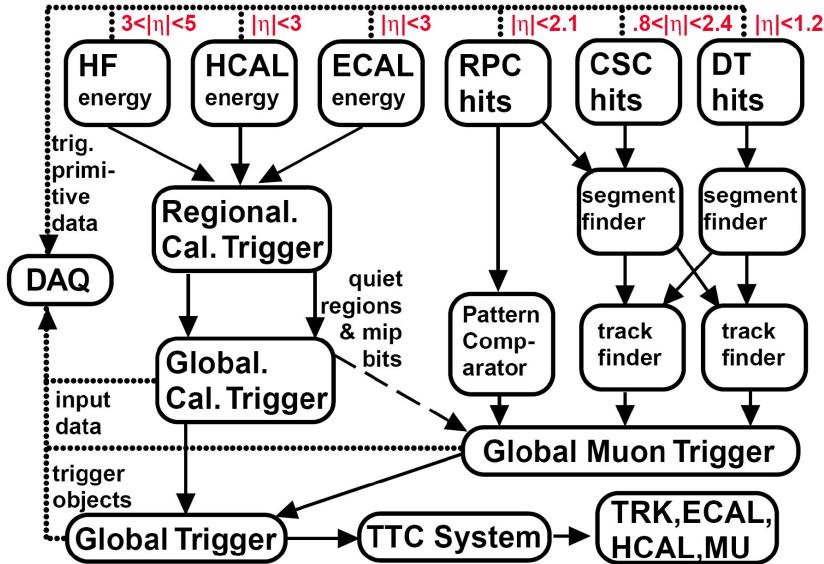
- 608 • L1 Calorimeter Trigger;
- 609 • L1 Muon Trigger;
- 610 • L1 Global Trigger.

611 The L1 Global Trigger is responsible for combining the output of L1 Calorimeter Trigger  
 612 and L1 Muon Trigger and for making the decision to either retain the event or discard it.  
 613 The organization of CMS L1 Trigger is schematically summarized in Fig. 2.14.

#### 614 L1 Calorimeter Trigger

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

616 The input for the L1 Calorimeter Trigger are calorimeter towers, which are clusters  
 617 of signals collected both from ECAL and HCAL. Towers are calculated by calorimeter  
 618 high level readout circuits, called Trigger Primitive Generators. The Regional Calorimeter  
 619 Trigger identifies electron, photon,  $\tau$  and jet candidates together with their transverse energy  
 620 and sends the information to the Global Calorimeter Trigger. The Global Calorimeter



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

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

### 623 L1 Muon Trigger

#### 624 Controllare se stato cambiato qualcosa nel Run2

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

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

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

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

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

644        **L1 Global Trigger**

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

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

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

653        **2.3.2. The high level trigger (HLT)**

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

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

666        Event by event, the HLT code is run on a single processor, and the time available to  
667        make a decision is about 300 ms. The real time nature of this selection imposes several  
668        constraints on the resources an algorithm can use. The reliability of HLT algorithms  
669        is of capital importance, because events not selected by the HLT are lost. In order to  
670        efficiently process events, the HLT code has to be able to quickly reject not interesting  
671        events; computationally expensive algorithms must be run only on good candidates for  
672        interesting events. In order to cope with this requirement the HLT code is organized in a  
673        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.

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

674           • Level 2: uses only complete muon and calorimetry information;  
675           • Level 2.5: uses also the pixel information;  
676           • Level 3: makes use of the full information from all the tracking detectors.  
677     Each step reduces the number of events to be processed in the following step. The most  
678     computationally expensive tasks are executed in the Level 3; time consuming algorithms  
679     such as track reconstruction are only executed in the region of interest. Besides, since the  
680     ultimate precision is not required at a HLT level, track reconstruction is performed on a  
681     limited set of hits, and is stopped once the required resolution is achieved.

# Chapter 3.

## 682 Reconstruction and identification of 683 physics objects

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

### 688 3.1. The Particle Flow technique

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

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

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

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

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

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

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

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

- 741 • Muons: a *global muon* gives rise to a *PF muon* if its combined  $p_T$  measurement is  
 742 compatible within 3 standard deviation with the one provided by the sole tracker. The  
 743 corresponding track is removed from the block;
- 744 • Electrons: electrons tend to give rise to short tracks, and to lose energy by Bremsstrahlung  
 745 in the tracker layers on their way to the calorimeter. The link between a charged-  
 746 particle track (refitted with the Gaussian-Sum Filter (GSF) [24]) and one or more  
 747 ECAL clusters identifies a *PF electron*. After the identification, the corresponding  
 748 tracks and clusters are removed from the block.
- 749 • Charged hadrons: the remaining tracks give rise to *PF charged hadrons*. Tracks can be  
 750 linked to ECAL and HCAL clusters, and the energy is determined taking into account  
 751 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 described in Sec. 3.5. The missing transverse energy,  $E_T^{\text{miss}}$ , is defined as the modulus of  $\vec{p}_T^{\text{miss}}$ .

## 3.2. Leptons reconstruction and identification

### 3.2.1. Muon reconstruction and identification

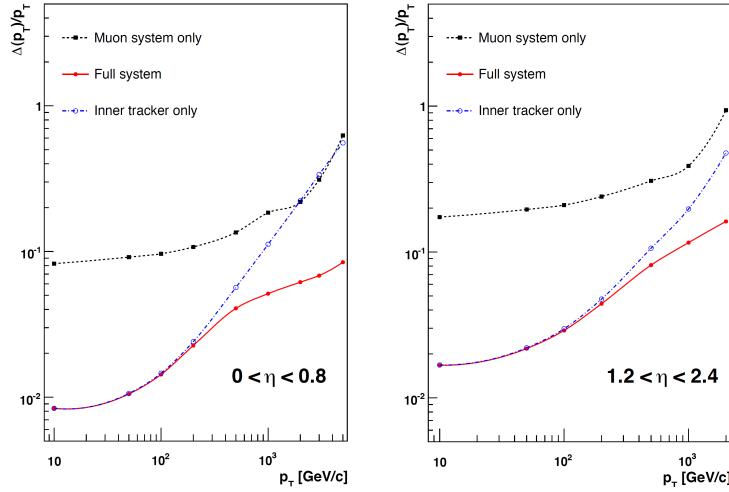
Muons produced at the collision point can go through the entire detector with a negligible energy loss, thus reaching the detector outermost part where the muon chambers are installed (see Sec. 2.2.5). Muons interact through ionization with the layers of the silicon tracker, which is able to reconstruct 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 [25]: 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. 3.1.

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>1</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 of muons from decays in flight, i.e. non prompt muons. The requirements defining the Tight Muon identification are summarized in Table 3.1.

Another selection which is optimised for low- $p_T$  muons coming from in flight decays is called *Soft Muon selection*. This selection requires the muon to be reconstructed as a Tracker

<sup>1</sup>Small variations with respect to this baseline definition are adopted by the specific analyses.



**Figure 3.1.:** 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.

**Table 3.1.:** 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

789 Muon with loose additional cuts on the transverse and longitudinal impact parameters.  
 790 This selection is commonly used to identify muons coming from B hadron decays.

### 791 3.2.2. Muon isolation

792 One of the most powerful requirements to select prompt muons, as the ones produced from  
 793 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:

$$I_{\Delta\beta}^{rel} = \left[ \sum_{ChH} p_T + \max \left( 0, \sum_{NH} p_T + \sum_{Ph} p_T - 0.5 \sum_{ChHPU} p_T \right) \right] / p_T^{\muon} . \quad (3.1)$$

The sums in Eq. (3.1) 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  $I_{\Delta\beta}^{rel} < 0.15$ .

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

### 3.2.3. Muon momentum scale and resolution

The measurement of the muon  $p_T$  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  $p_T$  that is generally independent on the  $p_T$  itself, while the effect of the alignment is known to produce a bias that increases linearly with the  $p_T$ .

Different methods are used to estimate the muon  $p_T$  scale and resolution effects and to determine the corresponding uncertainties, depending on the  $p_T$  range. At low and intermediate  $p_T$  ( $< 100$  GeV), the di-muon events arising from the  $J/\Psi$  and Z resonance decays are used to correct the  $p_T$  scale and to measure the  $p_T$  resolution. In the high  $p_T$  regime, the muon  $p_T$  scale and resolution are instead measured using cosmic ray muons. One of the methods that is commonly used in the intermediate  $p_T$  range is the *MuScleFit* (Muon momentum Scale calibration Fit), which provides the muon  $p_T$  scale corrections by fitting the Z boson mass peak in data and simulation. These corrections are meant to recover the bias of the Z mass peak with respect to the  $\eta$  and  $\phi$  coordinates of the muon. After applying these corrections, the relative  $p_T$  resolution,  $\sigma(p_T)/p_T$ , is measured as a function of  $\eta$  and  $\phi$  and is found to be on average of the order of 2% in the barrel and up to 6% in the endcaps, for muon  $p_T$  below 100 GeV.

### 3.2.4. Electron reconstruction and identification

The electron reconstruction is based on the combination of tracker and ECAL information. The reconstruction technique starts by measuring the energy deposits in ECAL by electrons, which form a “supercluster”. A supercluster is a group of one or more ECAL clusters associated using an algorithm that takes into account the characteristic shape of the energy deposited by electrons emitting Bremsstrahlung radiation in the tracker material. The supercluster shape is characterized by a narrow width profile in the  $\eta$  coordinate spread over the  $\phi$  direction. The superclusters are matched to tracks, reconstructed in the tracker with the GSF algorithm, in order to obtain an electron candidate. An additional reconstruction method, described in details in Refs. [26, 27], is instead seeded by electron tracks reconstructed in the inner tracker layers.

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

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

Different working points are provided by CMS corresponding to different selections on the previously defined variables. One of the common working points used by several physics analyses, as the  $H \rightarrow WW$  analyses described in Secs. 4, 5 and 6, is the “tight working point”, summarised in Table 3.2.

### 3.2.5. Electron isolation

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

**Table 3.2.:** Electron identification selections corresponding to the tight working point.

Variable	Selection	
	$ \eta_{\text{SC}}  \leq 1.479$	$1.479 <  \eta_{\text{SC}}  \leq 2.5$
$\sigma_{i\eta,i\eta}$	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

$$I_{EA \text{ corrected}}^{rel} = \left[ \sum_{ChH} p_T + \max \left( 0, \sum_{Ph} p_T + \sum_{NH} p_T - \rho EA \right) \right] / p_T^{\text{electron}} \quad (3.2)$$

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

### 3.2.6. Electron momentum scale and resolution

The electron momentum is estimated using a combination of the tracker and ECAL measurements. Before making the combination of the two measurements, the ECAL energy response is calibrated. Before doing the clustering, the energy response in individual crystals is calibrated and a correction factor is applied to take into account effects as energy leakage or changes in the crystal transparency induced by radiation <sup>2</sup>. Then the supercluster energy is also corrected using an MVA technique, selecting  $Z \rightarrow e^+e^-$  events in data and comparing to simulation. A detailed description of the techniques used to estimate the electron scale and resolution and the associated uncertainties is given in Ref. [27].

<sup>2</sup>The continuous monitoring of the crystals transparency is achieved by a laser-monitoring system.

### 881 3.2.7. Lepton identification and isolation efficiency

882 The efficiency related to the identification and isolation selections applied to muons and  
 883 electrons are generally estimated both in data and simulation and the simulated events are  
 884 corrected for the observed differences by means of a scale factor ( $SF$ ), defined as the ratio  
 885 of the efficiency measured in data and simulation, i.e.  $SF = \varepsilon_{\text{data}}/\varepsilon_{\text{MC}}$ .

886 The identification and isolation efficiencies are measured using a Tag and Probe technique.  
 887 The Tag and Probe technique is a method to estimate the efficiency of a selection on data.  
 888 It can be applied whenever one has two objects in one event, by using one of the two, the  
 889 *tag*, to identify the process of interest, and using the second, the *probe*, to actually measure  
 890 the efficiency of the selection being studied. Concerning the electron and muon case, the  
 891 Tag and Probe method uses a known mass resonance (e.g.  $J/\Psi$ , Z) to select particles of the  
 892 desired type, and probe the efficiency of a particular selection criterion on these particles.  
 893 In general the *tag* is an object that passes a set of very tight selection criteria designed  
 894 to isolate the required particle type. Tags are often referred to as a “golden” electrons or  
 895 muons and the fake rate for passing tag selection criteria should be very small. A generic  
 896 set of the desired particle type (i.e. with potentially very loose selection criteria) known  
 897 as *probe* is selected by pairing these objects with tags such that the invariant mass of the  
 898 combination is consistent with the mass of the resonance. Combinatorial backgrounds may  
 899 be eliminated through any of a variety of background subtraction methods such as fitting,  
 900 or sideband subtraction. The definition of the probe objects depend on the specifics of  
 901 the selection criterion being examined. The simple expression to get the efficiency  $\varepsilon$  as a  
 902 function of  $p_T$  and  $\eta$  is given below:

$$\varepsilon(p_T, \eta) = \frac{N_{\text{pass}}^{\text{probe}}}{N_{\text{pass}}^{\text{probe}} + N_{\text{fail}}^{\text{probe}}} \quad (3.3)$$

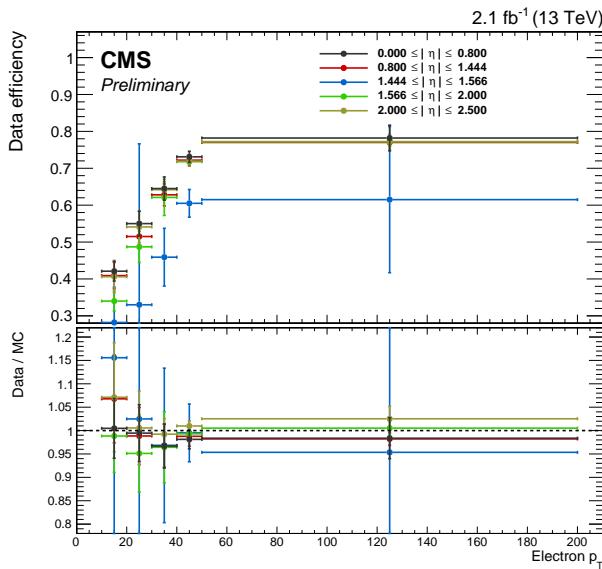
903 For the estimation of the electron or muon identification efficiency, the tag is chosen to  
 904 be a well identified and isolated electron or muon, while the probe is chosen as an electron  
 905 identified with loose selections. The invariant mass of the tag-probe pair is required to  
 906 be within a Z boson mass window (the effect of changing the Z mass window is included  
 907 as a systematic uncertainty). After that, the probe is required to pass the identification  
 908 selections discussed before for electrons and muons, and the efficiency is computed both  
 909 in data and simulation. A scale factor is then calculated by taking the ratio of the two  
 910 efficiencies and applied to reweight simulated events.

911 There are two methods to measure the efficiencies: the counting method consists in  
 912 simply computing the ratio of probe events that pass the selections and total number of  
 913 probe events, as shown in Eq. (3.3). This method can be used when the tag requirement  
 914 selects a very pure set of events, with a small background contribution. The other approach  
 915 is the fitting method, which is used when the background contamination is not negligible.  
 916 In this latter case, which represents the commonly used method for estimating the lepton  
 917 identification and isolation efficiencies, the invariant mass distribution of the tag-probe pair

for signal and background is fitted choosing proper functions. The signal plus background fit is performed simultaneously in two categories, corresponding to events in which the probe lepton pass or fail the identification requirements, and separately in bins of  $\eta$  and  $p_T$ .

A similar approach is used to estimate the lepton isolation efficiency, requiring the probe lepton to pass the isolation requirements instead of the identification ones and calculating the corresponding scale factor.

The identification and isolation efficiency and the scale factor are shown in Fig. 3.2 corresponding to the selections described in Sec. 3.2.4, for events of interest for a typical physics analysis of the  $H \rightarrow WW \rightarrow 2\ell 2\nu$  channel (in particular the analyses described in Sec. 5 and 6).



**Figure 3.2.:** Typical electron identification and isolation efficiencies in data (top panel) and data/simulation scale factor (bottom panel), as a function of the electron  $p_T$  and for different  $\eta$  bins.

### 3.2.8. Lepton trigger efficiency

Analyses that involves leptons in the final state generally select the interesting events using lepton triggers. For instance, the  $H \rightarrow WW \rightarrow 2\ell 2\nu$  channel is characterized by the presence of two leptons in the final state, thereby both single lepton and double lepton triggers are used. The lepton triggers at the HLT level are characterized by  $p_T$  thresholds, above which the trigger efficiency is very high (plateau region). Nevertheless, the trigger efficiency as a function of the lepton  $p_T$  is not a step function, but is characterized by a steep increase of the efficiency around the  $p_T$  threshold (turn-on region). The simulated samples thus need to be corrected in order to properly take into account the trigger efficiency. This can be achieved in two ways: including the HLT trigger in the event simulation or calculating the

trigger efficiency in data and then applying it on top of simulated events. Several analyses, such as those related to the  $H \rightarrow WW \rightarrow 2\ell 2\nu$  channel, opt for the second approach.

The trigger efficiency for single and double lepton triggers is calculated in bins of  $\eta$  and  $p_T$  using a Tag and Probe technique similar to the one described in Sec. 3.2.7, separately for muons and electrons. Since the triggered events arise from a mixture of two different triggers, the combined efficiency has to be computed and applied to simulated samples as an event weight. In the following, the approach used in the  $H \rightarrow WW \rightarrow 2\ell 2\nu$  analyses is described.

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

$$\varepsilon_{ev} = 1 - (1 - \varepsilon_{S,\ell 1}) \cdot (1 - \varepsilon_{S,\ell 2}) , \quad (3.4)$$

where  $\varepsilon_{S,\ell 1}$  and  $\varepsilon_{S,\ell 2}$  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 efficiency is calculated separately for each leg of the trigger. In the calculation of the efficiencies, the two trigger legs are considered 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 can be written as:

$$\varepsilon_{ev} = \varepsilon_{D,\ell 1}^{\text{lead}} \cdot \varepsilon_{D,\ell 2}^{\text{trail}} + (1 - \varepsilon_{D,\ell 1}^{\text{lead}} \cdot \varepsilon_{D,\ell 2}^{\text{trail}}) \cdot \varepsilon_{D,\ell 1}^{\text{trail}} \cdot \varepsilon_{D,\ell 2}^{\text{lead}} , \quad (3.5)$$

where  $\varepsilon_{D,\ell 1}^{\text{lead(trail)}}$  is the efficiency of the first lepton to pass the leading (trailing) leg of the double lepton trigger, and  $\varepsilon_{D,\ell 2}^{\text{lead(trail)}}$  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. (3.4) and (3.5), can be written as:

$$\begin{aligned} \varepsilon_{ev} = & 1 - (1 - \varepsilon_{S,\ell 1}) \cdot (1 - \varepsilon_{S,\ell 2}) + \\ & + (1 - \varepsilon_{S,\ell 1}) \cdot (1 - \varepsilon_{S,\ell 2}) \cdot \\ & \cdot [\varepsilon_{D,\ell 1}^{\text{lead}} \cdot \varepsilon_{D,\ell 2}^{\text{trail}} + (1 - \varepsilon_{D,\ell 1}^{\text{lead}} \cdot \varepsilon_{D,\ell 2}^{\text{trail}}) \cdot \varepsilon_{D,\ell 1}^{\text{trail}} \cdot \varepsilon_{D,\ell 2}^{\text{lead}}] . \end{aligned} \quad (3.6)$$

### 3.3. Jets reconstruction and identification

Jets are the experimental signature of quarks and gluons produced in high energy physics processes. They arise from the hadronization of partons, which forms collimated sprays

of particles, and play a predominant role in hadron colliders like the LHC, where the production cross section is very large. In this section, the jet reconstruction techniques used in CMS are described.

### 3.3.1. Jet reconstruction in CMS

The majority of physics analyses involving jets in the final state make use of particle flow jets. The PF jets are reconstructed using the technique described in Sec. 3.1, clustering all particles reconstructed with the PF algorithm, without any distinction of type and energy threshold. This method allows a remarkable improvement in the jet momentum and spatial resolutions with respect to the calorimeter jets, which are instead reconstructed using solely the information from the calorimeters, as the use of the tracker information provides a better  $p_T$  resolution for the charged particles constituting the jets<sup>3</sup>.

Jets are defined through sequential, iterative clustering algorithms that combine the four-momenta of input particles until certain conditions are satisfied and jets are formed [28]. Several algorithms are available for jet clustering, characterized by different features. From a theoretical point of view, an ideal jet clustering algorithm should fulfil the following requirements [29]:

- *Infrared safety*: infrared singularities should not appear in the perturbative calculations and the solutions of the algorithm should be insensitive to soft radiation in the event;
- *Collinear safety*: collinear singularities should not appear in the perturbative calculations and jets should be insensitive to collinear radiation in the event;
- *Invariance under boosts*: the solutions of the algorithm should be the same independently of boosts in the longitudinal direction. This is particularly important for pp colliders, where the centre-of-mass of the individual proton proton collisions is typically boosted along the beam direction;
- *Order independence*: the algorithm should find the same jets at parton, particle and detector level;
- *Straightforward implementation*: the algorithm should be straightforward to implement in perturbative calculations.

The ideal algorithm should also follow some experimental attributes. Among them, the performance of the algorithm should be as independent as possible of the detector that provides the data, the algorithm should not amplify the inevitable effects of resolution smearing and angle bias and should not be strongly affected by pile up and high beam luminosities. Furthermore, the algorithm should be easy to implement, efficient to identify all possible jet candidates and should keep at an acceptable level the necessary computing resources.

Two main classes of jet clustering algorithms can be defined. The first one consists in the “cone” recombination, where jets are reconstructed associating together particles

---

<sup>3</sup>On average, the typical jet energy fractions carried by charged particles, photons and neutral particles are 65%, 25% and 10%, respectively.

whose trajectories lie within a cone of radius  $\Delta R$  in the  $\eta$ - $\phi$  plane. The second class of algorithms uses the sequential recombination scheme, that iteratively recombine the closest pair of particles according to some distance measure. The standard algorithms used by CMS are the SIScone, which is a “cone” recombination algorithm, and the  $k_t$ , anti- $k_t$  and *Cambridge Aachen* (CA) algorithms, which instead belong to the sequential recombination class. All the analyses presented in Secs. 4, 5 and 6 make use of the sequential recombination scheme, in particular of the anti- $k_t$  algorithm with  $R = 0.4$ , which is briefly described in the following.

The  $k_t$ , anti- $k_t$  and CA algorithms are infrared and collinear safe algorithms characterized by the introduction of two definitions of distance:  $d_{ij}$ , the distance the two objects  $i$  and  $j$ , and  $d_{iB}$ , the distance between the object  $i$  and the beam. These distances are defined by the following equations:

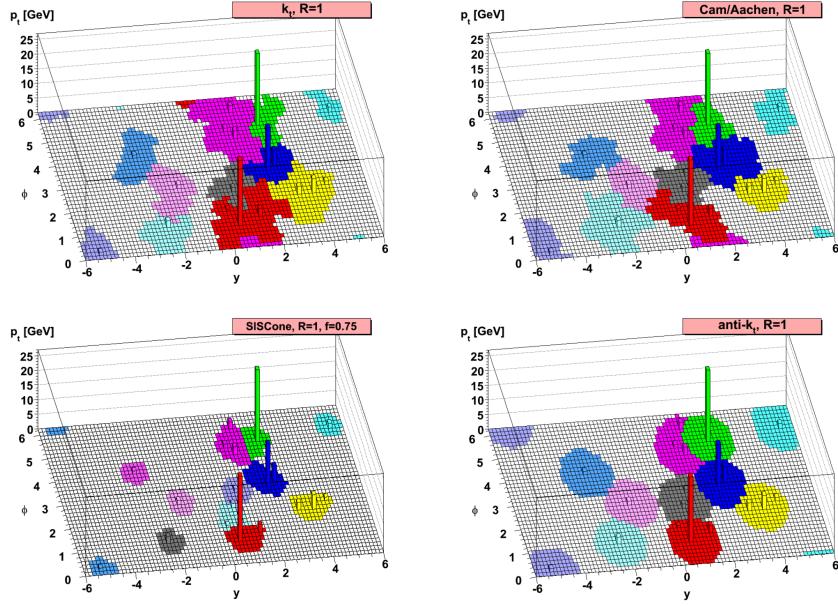
$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} \quad , \quad d_{iB} = k_{ti}^{2p} \quad , \quad (3.7)$$

where  $\Delta_{ij} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$  and  $k_{ti}$ ,  $y_i$  and  $\phi_i$  are the transverse momentum, rapidity and azimuthal angle of the particle  $i$ , respectively. In these formulas,  $R$  represents the radial parameter and  $p$  is a parameter that is 1 for  $k_t$ , 0 for CA and  $-1$  for anti- $k_t$  algorithm. The algorithm proceeds as follows:

- the distances  $d_{ij}$  are calculated for all pair of particle  $i, j$  and the distances  $d_{iB}$  are calculated for each particle  $i$ , according to Eq. (3.7);
- the smallest distance, which could be either of type  $d_{ij}$  or  $d_{iB}$ , is identified;
- if the smallest distance is a  $d_{ij}$ , the particles  $i$  and  $j$  are combined into a single new particle summing their four-momenta and the algorithm restarts from the first step;
- otherwise, if it is a  $d_{iB}$ ,  $i$  is declared to be a final state jet and the algorithm returns to the first step;
- the procedure is repeated until no particles are left.

The physical difference between the three algorithms is the momentum weighting. For the  $k_t$  algorithm, the weighting proportional to  $k_t^2$  implies that jets are reconstructed starting from particles with low transverse momentum. Moreover this algorithm produces jets with irregular borders, thereby complicating the correction for effects such as pile up. For the CA algorithm there is no transverse momentum weighting, and the particles are merged following just an angular approach, based on the distance  $\Delta_{ij}$ . Also this algorithm leads to jets with irregular borders. Finally, the anti- $k_t$  algorithm, uses a weighting proportional to  $1/k_t^2$ , favouring the merging of high transverse momentum particles. In this case the jets grow around the particles with highest transverse momenta and the jets have a circular shape.

Jets reconstructed with different algorithms starting from the same set of simulated particles are shown in Fig. 3.3.



**Figure 3.3.:** Jets reconstructed with different algorithms starting from the same set of simulated particles. The jets reconstructed with the sequential recombination algorithms described in the text are shown, as well as with the SIScone algorithm.

### 1038 3.3.2. Jet energy correction

1039 The purpose of jet energy correction is to relate, on average, the jet energy measured in the  
 1040 detector to the true energy of the corresponding final state particle or parton jet. The latter  
 1041 is obtained in simulation by clustering, with the same algorithm used for jets in the detector,  
 1042 all the stable particles, i.e. with  $c\tau > 1$  cm, produced in the event excluding neutrinos.  
 1043 This mismatch is mainly ascribable to the non uniform and linear response of the CMS  
 1044 calorimeters, to the electronics noise and to pile up. For this reason, CMS has developed a  
 1045 sequential procedure to calculate and apply the *jet energy corrections* (JEC) [30].

1046 The correction is applied as a multiplicative factor  $\mathcal{C}$  to each component of the raw jet  
 1047 four-momentum  $p_\mu^{\text{raw}}$  (components are indexed by  $\mu$  in the following):

$$p_\mu^{\text{cor}} = \mathcal{C} \cdot p_\mu^{\text{raw}} \quad , \quad (3.8)$$

1048 where  $p_\mu^{\text{cor}}$  is the corrected jet four-momentum. The correction factor is composed of the  
 1049 offset correction  $C_{\text{offset}}$ , the MC calibration factor  $C_{\text{MC}}$ , and the residual calibrations  $C_{\text{rel}}$   
 1050 and  $C_{\text{abs}}$  for the relative and absolute energy scales, respectively. The offset correction  
 1051 removes the extra energy due to noise and pile up, and the MC correction removes the bulk  
 1052 of the non-uniformity in  $\eta$  and the non-linearity in  $p_T$ . Finally, the residual corrections  
 1053 account for the small differences between data and simulation. The various components are

1054 applied in sequence as described by the equation below:

$$\mathcal{C} = C_{\text{offset}}(p_{\text{T}}^{\text{raw}}) \cdot C_{\text{MC}}(p'_{\text{T}}, \eta) \cdot C_{\text{rel}}(\eta) \cdot C_{\text{abs}}(p''_{\text{T}}) , \quad (3.9)$$

1055 where  $p'_{\text{T}}$  is the jet  $p_{\text{T}}$  after applying the offset correction and  $p''_{\text{T}}$  is the jet  $p_{\text{T}}$  after applying  
1056 all previous corrections. Each component is briefly described in the following sections.

1057 **Offset correction**

1058 The offset correction purpose is to estimate and subtract, on average, the energy contribution  
1059 that is not associated with the hard scattering in the event. The energy excess includes  
1060 contributions from electronics noise and pile up. The approach followed for the estimation of  
1061 the offset correction is known as *Jet Area Method*. For each event, an average  $p_{\text{T}}$ -density per  
1062 unit area,  $\rho$ , is estimated, characterizing the soft jet activity. This  $p_{\text{T}}$ -density represents the  
1063 combination of the underlying event, the electronics noise and the pile up effects. The two  
1064 latter components contaminate the hard jet energy measurement and need to be corrected  
1065 for with the offset correction. The key element for this approach is the jet area  $A_j$ . A  
1066 very large number of infinitely soft four-momentum vectors (soft enough not to change  
1067 the properties of the true jets) are artificially added in the event and clustered by the jet  
1068 algorithm together with the true jet components. The extent of the region in the  $\eta$ - $\phi$   
1069 space occupied by the soft particles clustered in each jet defines the active jet area. The  
1070  $p_{\text{T}}$ -density  $\rho$  is calculated with the  $k_t$  algorithm with a distance parameter  $R = 0.6$ . The  
1071 quantity  $\rho$  is estimated event by event as the median of the distribution of the variable  
1072  $p_{\text{T}j}/A_j$ , where  $j$  runs over all jets in the event, and is not sensitive to the presence of hard  
1073 jets in the event. At the detector level, the measured density  $\rho$  is the convolution of the true  
1074 particle-level activity (underlying event, pile-up) with the detector response to the various  
1075 particle types. The event-by-event and jet-by-jet offset correction can thus be defined as:

$$C_{\text{offset}}(p_{\text{T}}^{\text{raw}}, A_j, \rho) = 1 - \frac{(\rho - \langle \rho_{\text{UE}} \rangle) \cdot A_j}{p_{\text{T}}^{\text{raw}}} . \quad (3.10)$$

1076 In the formula above,  $\langle \rho_{\text{UE}} \rangle$  represents the average  $p_{\text{T}}$ -density component due to the  
1077 underlying event and electronics noise, and is measured in events with exactly one recon-  
1078 structed primary vertex, i.e. no pile up.

1079 An additional pile up subtraction method that is used in CMS is called *Charged Hadron  
1080 Subtraction*. This method makes use of PF jets and exploits the excellent CMS tracking  
1081 capabilities to identify and remove charged hadrons inside jets, which are known to originate  
1082 from pile up vertices. This is a particle-by-particle method that is applied to jets before  
1083 calculating the offset correction.

1084 **MC calibration correction**

1085 The MC calibration is based on the simulation and corrects the energy of the reconstructed  
 1086 jets such that it is equal on average to the energy of the generated jets. In order to evaluate  
 1087 this correction, simulated QCD events are generated and then processed through the CMS  
 1088 detector simulation, based on the GEANT4 software. The jet reconstruction in simulation  
 1089 is identical to the one applied to the data. Each reconstructed jet is spatially matched, in  
 1090 the  $\eta$ - $\phi$  space, to a generated jet by requiring  $\Delta R < 0.25$ . In each bin of the generated  
 1091 jet transverse momentum  $p_T^{\text{gen}}$ , the response variable  $\mathcal{R} = p_T^{\text{reco}}/p_T^{\text{gen}}$  and the reconstructed  
 1092 jet transverse momentum  $p_T^{\text{reco}}$ , are saved. The average correction in each bin is therefore  
 1093 defined as:

$$C_{\text{MC}}(p_T^{\text{reco}}) = \frac{1}{\langle R \rangle} \quad , \quad (3.11)$$

1094 and is expressed as a function of the average reconstructed jet  $p_T$ ,  $\langle p_T^{\text{reco}} \rangle$ .

1095 **Relative jet energy scale**

1096 The goal of the relative jet energy scale correction is to make the jet response flat versus  $\eta$ .  
 1097 This is achieved by employing a Tag and Probe technique, selecting di-jet events in data.  
 1098 The size of this residual correction is of the order of 2–3% in the central  $\eta$  region, while it  
 1099 goes up to about 10% in the forward region.

1100 **Absolute jet energy scale**

1101 The goal of the absolute jet energy scale correction is to make the jet response at versus  
 1102  $p_T$ . The absolute jet energy response is measured in the reference region  $|\eta| < 1.3$  with  
 1103 the *Missing Transverse Energy Projection Fraction* (MPF) method [31], using  $\gamma +$  jets and  
 1104  $Z +$  jets events. The method is used to estimate the absolute jet energy correction and is  
 1105 based on the fact that  $\gamma +$  jets and  $Z +$  jets events have no intrinsic  $E_T^{\text{miss}}$  and that, at parton  
 1106 level, the  $\gamma$  and  $Z$  boson are perfectly balanced by the hadronic recoil in the transverse  
 1107 plane.

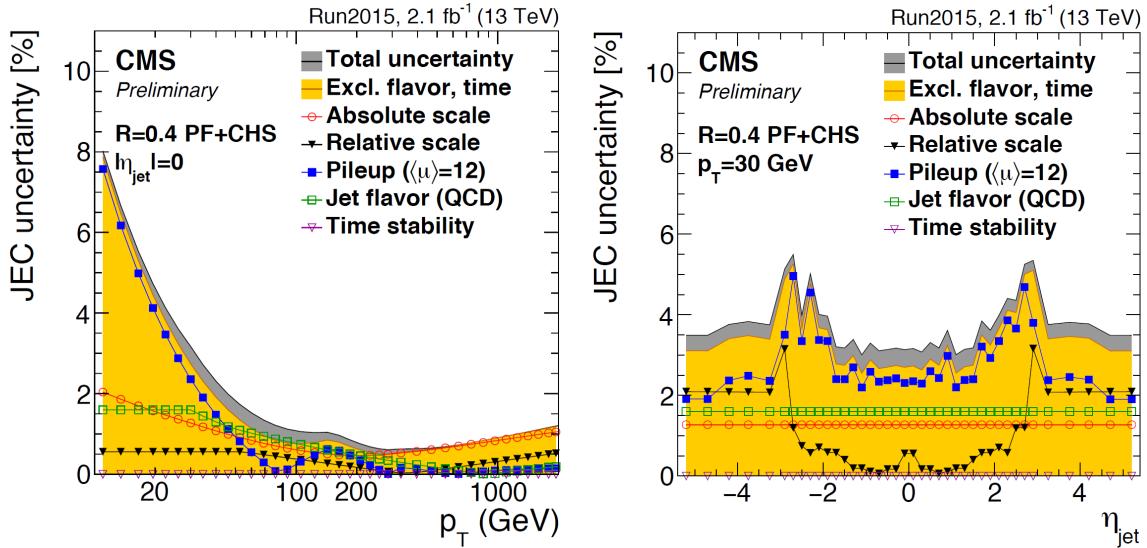
1108 **Jet energy uncertainties**

1109 The uncertainties in the jet energy estimation arise from several different sources. Generally  
 1110 these can be categorized as follows:

- 1111 • physics modelling in MC such as showering, underlying event, etc.;
- 1112 • MC modelling of true detector response and properties;
- 1113 • potential biases in the methodologies used to estimate the corrections.

1114 The sources are combined in different groups: absolute scale, relative scale, pile up, jet  
 1115 flavor and time stability. In Fig. 3.4 the effect of each group of uncertainties is shown

together with the total uncertainty obtained summing all sources in quadrature, both as a function of  $\eta$  and  $p_T$ . The pile up uncertainty dominates for low values of the jet  $p_T$  while the relative and absolute uncertainties are more important in the high  $p_T$  region.



**Figure 3.4.:** JEC uncertainties as a function of  $p_T$  (left) for jets reconstructed with  $\eta = 0$  and as a function of  $\eta$  (right) for jets with  $p_T = 30$  GeV. All jets are reconstructed with the PF technique and using the anti- $k_t$  algorithm with  $R = 0.4$ , after applying the CHS correction. Results are based on  $2.1 \text{ fb}^{-1}$  of data collected at 13 TeV.

### 1119 Jet energy resolution

1120 Measurements show that the jet energy resolution (JER) in data is worse than in the  
 1121 simulation, therefore the simulated jets need to undergo a smearing procedure in order to  
 1122 have a better description of the data.

1123 Reconstructed jets in simulated events are corrected for the jet energy resolution using  
 1124 a two step procedure. In the first step, the reconstructed jet  $p_T$  is scaled for the observed  
 1125  $p_T$  difference between reconstructed and generated jets. This method only works for  
 1126 reconstructed jets that are well matched to generated jets, where the matching is based  
 1127 on  $\Delta R$  and  $\Delta p_T$  requirements. For reconstructed jets that do not fulfil the matching  
 1128 requirements, a gaussian smearing of the  $p_T$  distribution is applied in order to obtain the  
 1129 desired resolution.

1130 **jet identification?**

## 1131 **3.4. Jet b tagging**

1132 Jets that arise from bottom quark hadronization (b-jets) are present in many physics  
 1133 processes, such as the decay of top quarks. The ability to accurately identify b-jets is crucial

1134 to reduce the otherwise overwhelming background from these processes to channels involving  
1135 jets from gluons ( $g$ ) and light-flavour quarks ( $u, d, s$ ), and from  $c$ -quark fragmentation.

1136 Algorithms for  $b$  jet identification (also known as  $b$  tagging algorithms) exploit the  
1137 long life time of  $b$  hadrons present in jets originating from the hadronization of  $b$  quarks.  
1138 This long life time results in a decay of the  $b$  hadron that is displaced with respect to the  
1139 primary interaction vertex. This displacement of a few millimetres results in the presence  
1140 of displaced tracks from which a secondary vertex may be reconstructed. In addition,  $b$   
1141 hadrons have a probability of around 20% to decay to a muon or electron. Hence, also the  
1142 presence of these charged leptons can be exploited for  $b$  jet identification techniques and  
1143 for measuring their performance with the collision data.

1144 A variety of reconstructed physics objects, as tracks, vertices and identified leptons,  
1145 can be used to build observables that discriminate between  $b$  and light-quark jets. Several  
1146  $b$  tagging algorithms have been developed by CMS, each one based on different input  
1147 information. A common feature of all the algorithms is that each one yields a single  
1148 discriminator value for each jet, which measures the likelihood that the jet has been produced  
1149 by the hadronization of a  $b$  quark. The minimum thresholds on these discriminators define  
1150 loose (“L”), medium (“M”), and tight (“T”) operating points with a misidentification  
1151 probability for light-parton jets close to 10%, 1%, and 0.1%, respectively, at an average  
1152 jet  $p_T$  of about 80 GeV. The misidentification probability, also known as mistag rate, is  
1153 defined as the probability to wrongly identify a light-parton jet as a  $b$ -jet.

1154 Some of the algorithms make use of the track impact parameters (IP) with respect  
1155 to the primary vertex, defined as the distance between the primary vertex and the track  
1156 at their point of closest approach, to distinguish the decay products of a  $b$  hadron from  
1157 prompt tracks. The impact parameter has the same sign as the scalar product of the vector  
1158 pointing from the primary vertex to the point of closest approach with the jet direction.  
1159 Tracks originating from the decay of particles travelling along the jet axis will tend to have  
1160 positive IP values. In contrast, the impact parameters of prompt tracks can have positive  
1161 or negative IP values. The impact parameter significance, defined as the ratio of the IP to  
1162 its estimated uncertainty, is used as an observable.

1163 The *Track Counting* (TC) algorithm sorts tracks inside a jet by decreasing values of the  
1164 IP significance. Although the ranking tends to bias the values for the first track to high  
1165 positive IP significances, the probability to have several tracks with high positive values is  
1166 low for light-parton jets. Therefore the two different versions of the algorithm use the IP  
1167 significance of the second and third ranked track as the discriminator value. These two  
1168 versions of the algorithm are called *Track Counting High Efficiency* (TCHE) and *Track*  
1169 *Counting High Purity* (TCHP), respectively.

1170 A general extension of the TC algorithm, i.e. the *Jet Probability* (JP), combines the  
1171 IP information of several tracks inside the jet, using an estimate of the likelihood that all  
1172 tracks associated to the jet come from the primary vertex as a discriminating variable. A  
1173 variant of the JP algorithm also exists in which the four tracks with the highest impact  
1174 parameter significance get a higher weight in the jet probability calculation. This algorithm  
1175 is referred to as *Jet B-Probability* (JBP).

1176 A different approach consists in using the secondary vertices and the related kinematic  
 1177 variables, together with displaced tracks information, to discriminate between b and non-b  
 1178 jets. This algorithm is known as *Combined Secondary Vertex* (CSV)<sup>4</sup>. The magnitude  
 1179 and direction of the vector connecting the primary and secondary vertices are used as  
 1180 a discriminating variables and quality requirements are imposed to secondary vertex  
 1181 candidates. In addition, the usage of displaced tracks information allows to increase the  
 1182 efficiency for events where no secondary vertex is found. Several variables related to  
 1183 secondary vertices and displaced tracks are used to build likelihood ratios that have a good  
 1184 discriminating power.

1185 Two algorithms for reconstructing secondary vertices are exploited. For the first  
 1186 algorithm, the tracks associated to jets and fulfilling some quality requirements are used in  
 1187 the adaptive vertex reconstruction (AVR) algorithm [32]. The AVR is the algorithm used  
 1188 for CMS analyses during the 8 TeV data taking. In contrast with this method, the Inclusive  
 1189 Vertex Finder (IVF) algorithm is not seeded from tracks associated to reconstructed jets, but  
 1190 instead makes use of all the tracks in the event, with appropriate selections, to reconstruct  
 1191 the secondary vertices. The latter is the default algorithm used to reconstruct secondary  
 1192 vertices for CMS analyses using 13 TeV data.

1193 A new b jet identification algorithm has been recently developed, combining the dis-  
 1194 criminating provided by the JP and CSV algorithms with a Boosted Decision Tree (BDT)  
 1195 technique. This combined multivariate algorithm (cMVA) is found to slightly improve the  
 1196 b jet identification efficiency.

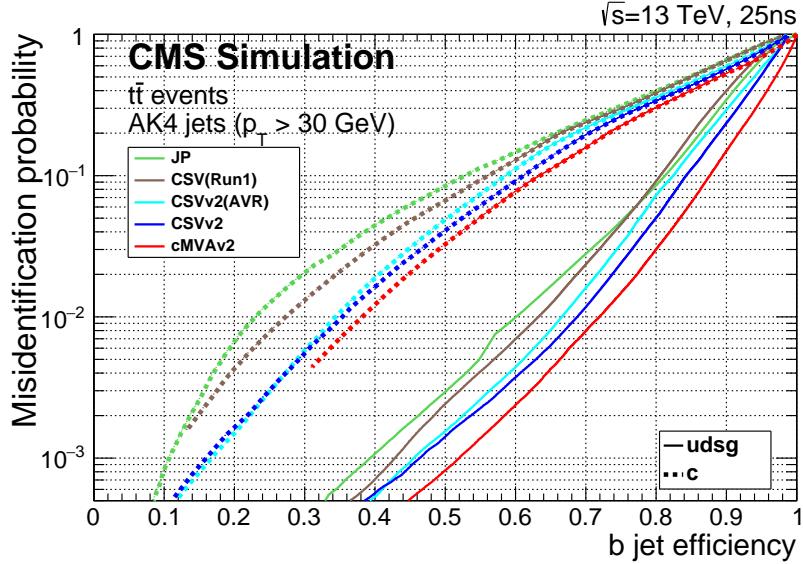
1197 The performance of these algorithms is determined using simulated  $t\bar{t}$  events, selecting  
 1198 events with at least one jet with  $p_T > 30$  GeV. This is shown in Fig. 3.5, where the b jet  
 1199 identification efficiency versus the misidentification probability is reported for the various  
 1200 algorithms. This figure serves as an illustration as the b tagging performance depend on  
 1201 the  $p_T$  and  $\eta$  distribution of the jets, and need to be checked for each analysis phase space.

### 1202 3.5. Missing transverse energy

1203 In hadron colliders the longitudinal momentum (along the beam axis) carried by the  
 1204 incoming partons is not known, preventing the possibility to measure the total missing  
 1205 energy. Nevertheless, the initial transverse momentum carried by the incoming partons  
 1206 is expected to be zero, thereby, for the conservation of the momentum components, also  
 1207 the net momentum of all the particles in the final state of collisions must be zero. The  
 1208 missing transverse momentum ( $\vec{p}_T^{\text{miss}}$ ) is the momentum imbalance in the transverse plane  
 1209 of all the visible particles in the event, and its modulus is the missing transverse energy  
 1210 ( $E_T^{\text{miss}}$ ). The  $\vec{p}_T^{\text{miss}}$  vector is defined as the negative vectorial sum of transverse momenta of  
 1211 all reconstructed PF objects, as shown in the following equation:

---

<sup>4</sup>An improved version of this algorithm, CSVv2, has been developed for Run 2 analyses.



**Figure 3.5.:** Performance of the b jet identification efficiency algorithms demonstrating the probability for non-b jets to be misidentified as b jet as a function of the efficiency to correctly identify b jets. The curves are obtained on simulated  $t\bar{t}$  events using anti- $k_t$  jets clustered with  $R = 0.4$  and requiring  $p_T > 30$  GeV.

$$\vec{p}_T^{\text{miss}} = - \sum_{\text{PF obj}} \vec{p}_T^{\text{PF obj}} , \quad (3.12)$$

where the sum extends over all the PF objects. A  $E_T^{\text{miss}}$  value different from zero is a potential signature of the presence of particles in the event that have not interacted with the detector, such as neutrinos or beyond the SM particles predicted by some models, but can also be ascribable to detector inefficiencies.

In addition to imperfect resolution of all detectable and reconstructed physics objects, the  $E_T^{\text{miss}}$  measurement is also sensitive to overlapping detector signals from additional pile up interactions (both in-time and out-of-time pile up **definire in-time e out-of-time PU la prima volta che parlo di PU**), particle misidentification, as well as detector malfunctions [33, 34]. The bias on the  $E_T^{\text{miss}}$  measurement is reduced by correcting the  $p_T$  of the jets with the jet energy corrections described in 3.3.2, and propagating the correction to the  $E_T^{\text{miss}}$  according to:

$$\vec{p}_T^{\text{miss Type-I}} = \vec{p}_T^{\text{miss}} - \sum_{\text{jets}} (\vec{p}_{T,\text{jet}}^{\text{JEC}} - \vec{p}_{T,\text{jet}}) , \quad (3.13)$$

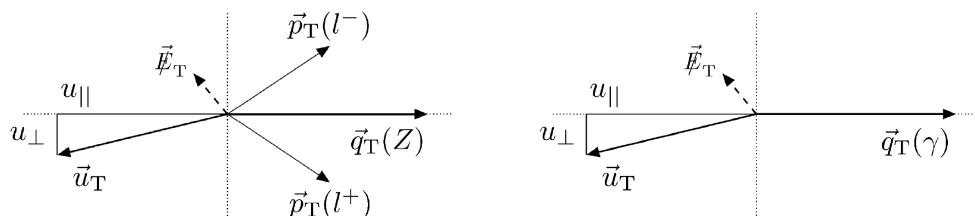
where the superscript JEC refers to corrected jets. This correction, called “Type-I” correction, uses the JEC for all jets with  $p_T > 15 \text{ GeV}$  that have less than 90% of their energy deposited in ECAL. Furthermore, if a muon is found inside a jet, its four-momentum is subtracted from the jet four-momentum before the correction, and added back to the corrected object.

Anomalous high- $E_T^{\text{miss}}$  events can be due to various phenomena. In the ECAL, spurious deposits may appear due to particles striking sensors in the ECAL photodetectors, or from real showers with non-collision origins such as those caused by beam halo particles<sup>5</sup>. ECAL dead cells can cause real energy to have been missed, again leading to a spurious imbalance. In the HCAL, spurious energy can arise due to noise in the hybrid photodiode and readout electronics, as well as direct particle interactions with the light guides and photomultiplier tubes of the forward calorimeter. The spurious  $E_T^{\text{miss}}$  produced by these effects is estimated using dedicated algorithms, and a cleaning procedure is applied to data in order to remove the affected events.

### 3.5.1. $E_T^{\text{miss}}$ scale and resolution measurement

The performance (scale and resolution) of  $E_T^{\text{miss}}$  can be studied in events with an identified Z boson or an isolated photon. Momenta of leptons and photons can be reconstructed with good resolutions, around 1–6%, while momenta of jets are reconstructed with less precision, with typical resolutions of 5–15%. As a consequence, the  $E_T^{\text{miss}}$  resolution in Z or  $\gamma + \text{jets}$  events is dominated by hadronic activity in the event.

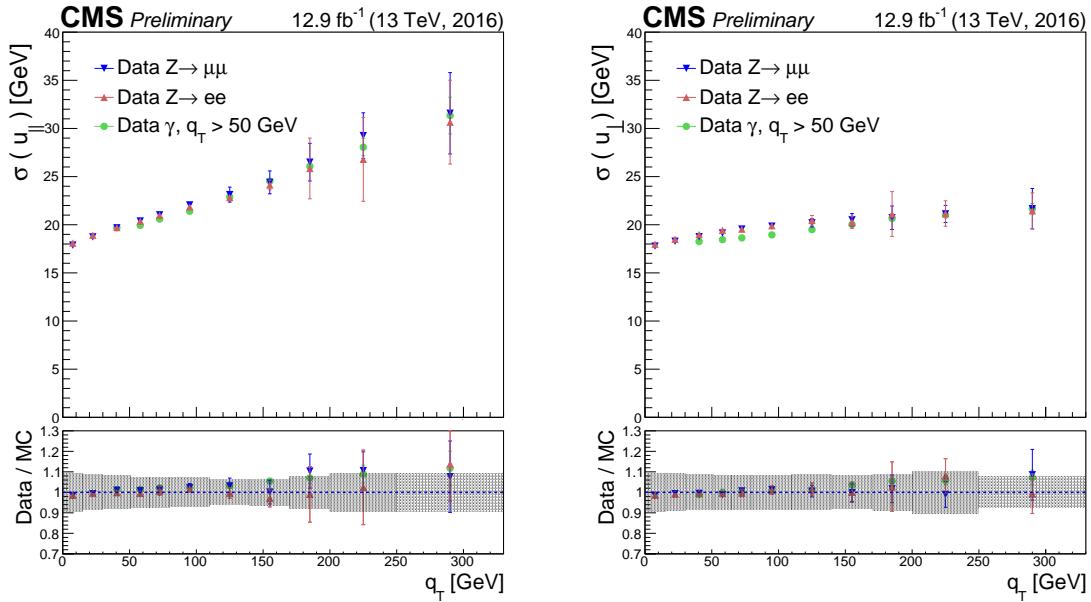
The comparison of the momenta of the vector boson with respect to the hadronic recoil system is used to measure the  $E_T^{\text{miss}}$  performance. In Fig. 3.6 the vector boson momentum in the transverse plane is shown as  $\vec{q}_T$ , and transverse momentum of the hadronic recoil, defined as the vectorial sum of the transverse momenta of all particles except the vector boson (or its decay products, in the case of Z bosons), is shown as  $\vec{u}_T$ . Momentum conservation in the transverse plane dictates that  $\vec{q}_T + \vec{u}_T + \vec{p}_T^{\text{miss}} = 0$ .



**Figure 3.6.:** Illustration of the  $Z \rightarrow \ell^+ \ell^-$  (left) and photon (right) event kinematics in the transverse plane.

<sup>5</sup>These background, also known as *Machine-Induced Background*, originates mainly from interactions of the beam protons with the final set of collimators before the CMS experiment and from proton gas interactions.

The  $E_T^{\text{miss}}$  characteristics are evaluated using two components of  $\vec{u}_T$ , one parallel ( $u_{\parallel}$ ) and one perpendicular ( $u_{\perp}$ ) to the axis defined by  $\vec{q}_T$ . The distributions of these variables are parametrized using a convolution of a Breit-Wigner and a Gaussian distribution, i.e. a Voigtian distribution, which is found to provide a good description of the observables and is used to measure the resolution in  $u_{\parallel}$  and  $u_{\perp}$ ,  $\sigma(u_{\parallel})$  and  $\sigma(u_{\perp})$ , respectively. These resolutions are closely related to the  $E_T^{\text{miss}}$  resolution. The resolutions  $\sigma(u_{\parallel})$  and  $\sigma(u_{\perp})$  obtained using recent 13 TeV data are shown in Fig. 3.7.



**Figure 3.7.:** Resolutions  $\sigma(u_{\parallel})$  (left) and  $\sigma(u_{\perp})$  (right) for  $Z \rightarrow \mu^+\mu^-$ ,  $Z \rightarrow e^+e^-$  and  $\gamma$  events as a function of the vector boson  $p_T$ . The upper panels show the resolution measured in data and the bottom panels the data to simulation ratio.



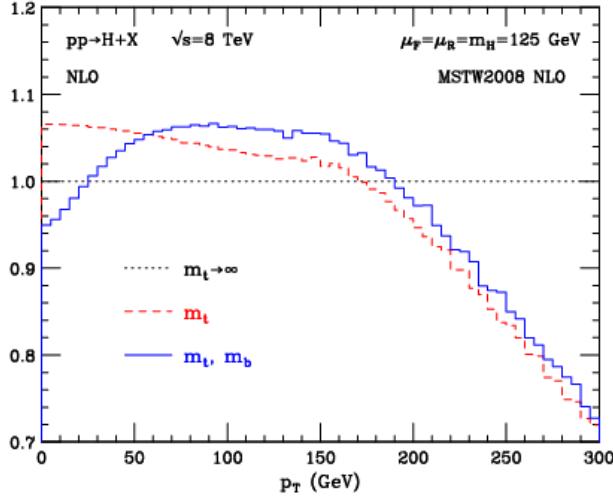
# Chapter 4.

## 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 the Higgs boson transverse momentum,  $p_T^H$ , and its pseudorapidity,  $\eta$ . 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 [35, 36], and at NNLO using the so-called large- $m_t$  approximation [37–39], 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 [40].

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 [41, 42], 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 [43], as shown in Fig. 4.1. Hence, a precise experimental measurement of the  $p_T^H$  spectrum is important to test the existing SM calculations.



**Figure 4.1.:** Distribution of  $p_T^H$  computed at NLO ( $\alpha_s^4$ ) and divided by 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. [44, 45], 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 [46], 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 [47]. 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 [48–50] and CMS [51, 52] 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 [53], and is

1310 characterized by good signal sensitivity. Such sensitivity allowed the observation of a Higgs  
1311 boson at the level of 4.3 (5.8 expected) standard deviations for a mass hypothesis of 125.6  
1312 GeV using the full LHC data set at 7 and 8 TeV [54].

1313 The measurement is performed in a fiducial phase space defined by kinematic requirements  
1314 on the leptons that closely match the experimental event selection.

1315 The effect of the limited detector resolution, as well as the selection efficiency with respect  
1316 to the fiducial phase space are corrected to particle level with an unfolding procedure [55],  
1317 as explained in Sec. 4.7.

## 1318 4.2. Data sets, triggers and MC samples

### 1319 4.2.1. Data sets and triggers

1320 The data set used for the analysis corresponds to  $19.4 \text{ fb}^{-1}$  of proton-proton collisions at  
1321  $\sqrt{s} = 8 \text{ TeV}$ , collected by the CMS detector during 2012. Only data corresponding to good  
1322 data taking quality are considered.

1323 Events are required to fire one of the unprescaled single-electron, single-muon or muon-  
1324 electron triggers. Due the rather high LHC instantaneous luminosity the single-lepton  
1325 triggers must have high HLT  $p_T$  thresholds, otherwise the rate of these triggers would be  
1326 too large to be sustained. The double-lepton triggers allow to lower down the  $p_T$  thresholds  
1327 while keeping a sustainable trigger rate, thus maintaining a good sensitivity to the Higgs  
1328 boson signal, for which the lepton  $p_T$  can be rather small. A brief overview of the HLT  
1329  $p_T$  criteria on the leptons is given in Table 4.1. While the HLT lepton  $p_T$  thresholds of 17  
1330 and 8 GeV for the double lepton triggers accommodate the offline lepton  $p_T$  selection of 20  
1331 and 10 GeV, the higher  $p_T$  thresholds in the single lepton triggers help partially recovering  
1332 double lepton trigger inefficiencies as a high  $p_T$  lepton is on average expected due to the  
1333 kinematic of the Higgs decay.

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

Trigger Path	8 TeV
Single-Electron	$p_T > 27 \text{ GeV}$
Single-Muon	$p_T > 24 \text{ GeV}$
Muon-Electron	$p_T > 17 \text{ and } 8 \text{ GeV}$
Electron-Muon	$p_T > 17 \text{ and } 8 \text{ GeV}$

1334 The trigger is not simulated in MC samples but the combined trigger efficiency is  
1335 estimated from data and applied as a weight to all simulated events, as described in  
1336 Sec. 3.2.8.

### 1337 4.2.2. Monte Carlo samples

1338 Several Monte Carlo event generators are used to simulate the signal and background  
 1339 processes:

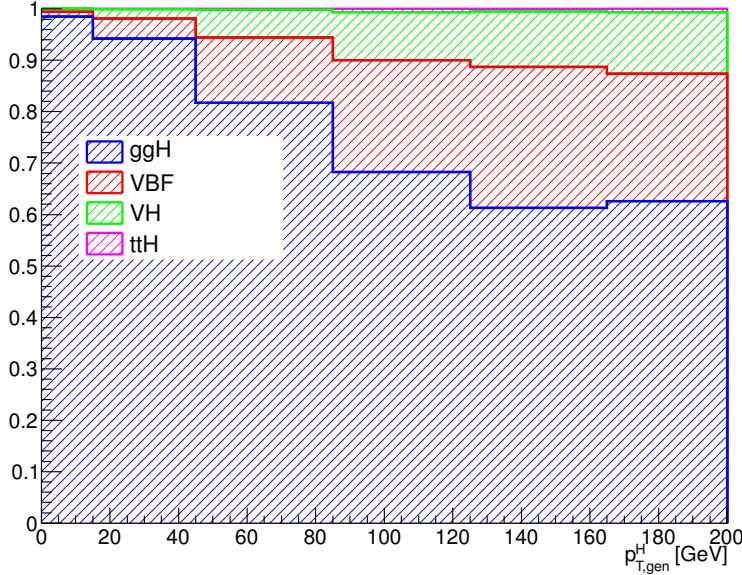
- 1340 • The first version of the POWHEG program [56–60] (POWHEG V1) provides event  
 1341 samples for the  $H \rightarrow WW$  signal for the gluon fusion (ggH) and VBF production  
 1342 mechanisms, as well as  $t\bar{t}$  and  $tW$  processes [61], with NLO accuracy.
- 1343 • The  $qq \rightarrow W^+W^-$ , Drell-Yan, ZZ, WZ,  $W\gamma$ ,  $W\gamma^*$ , tri-bosons and  $W+jets$  processes  
 1344 are generated using the MADGRAPH 5.1.3 [62] event generator.
- 1345 • The  $gg \rightarrow W^+W^-$  process is generated using the GG2WW 3.1 generator [63] and its  
 1346 cross section is scaled to the approximate NLO prediction [64, 65].
- 1347 • The VH process is simulated using PYTHIA 6.426 [66].

1348 For leading-order generators samples, the CTEQ6L [67] set of parton distribution functions  
 1349 (PDF) is used, while CT10 [68] is used for next-to-leading order (NLO) ones. Cross section  
 1350 calculations at next-to-next-to-leading order (NNLO) are used for the  $H \rightarrow WW$  process [69].  
 1351 The  $H \rightarrow WW$  process simulation is reweighted so that the  $p_T^H$  spectrum and inclusive  
 1352 production cross section closely match the SM calculations that have NNLO+NNLL pQCD  
 1353 accuracy in the description of the Higgs boson inclusive production, in accordance with  
 1354 the LHC Higgs Cross Section Working Group recommendations [53]. The reweighting of  
 1355 the  $p_T^H$  spectrum is achieved by tuning the POWHEG generator, as described in detail in  
 1356 Ref. [70]. Cross sections computed with NLO pQCD accuracy are used for the background  
 1357 processes [53]. The contribution of the  $t\bar{t}H$  production mechanism is checked to be negligible  
 1358 in each bin of  $p_T^H$  (below 1%) and is not included among the different production mechanisms.  
 1359 In Fig. 4.2 the relative fraction of the four production mechanisms is shown for each  $p_T^H$  bin.

1360 For all processes, the detector response is simulated using a detailed description of the  
 1361 CMS detector, based on the GEANT4 package [71].

1362 Minimum bias events are superimposed on the simulated events to emulate the additional  
 1363 proton-proton interactions per bunch crossing. The pile up multiplicity in simulated events  
 1364 has been generated poissonianly sampling from a distribution similar to the one expected  
 1365 from data. The simulated events are reweighted to correct for observed differences between  
 1366 data and simulation in the number of pile up events, as shown in Fig. 4.3.

1367 For the comparison of the measured unfolded spectrum with the theoretical predictions,  
 1368 two additional MC generators are used for simulating the SM Higgs boson production in  
 1369 the ggH process: HRES 2.3 [42, 43] and the second version of the POWHEG generator  
 1370 (POWHEG V2) [72]. HRES is a partonic level MC generator that computes the SM Higgs  
 1371 boson cross section at NNLO accuracy in pQCD and performs the NNLL resummation  
 1372 of soft-gluon effects at small  $p_T$ . The central predictions of HRES are obtained including  
 1373 the exact top and bottom quark mass contribution to the gluon fusion loop, fixing the  
 1374 renormalization and factorization scale central values at a Higgs boson mass of 125 GeV.  
 1375 The cross section normalization is scaled, to take into account electroweak corrections, by a  
 1376 factor of 1.05 and the effects of threshold resummation by a factor of 1.06 [73, 74]. The  
 1377 upper and lower bounds of the uncertainties are obtained by scaling up and down both the



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

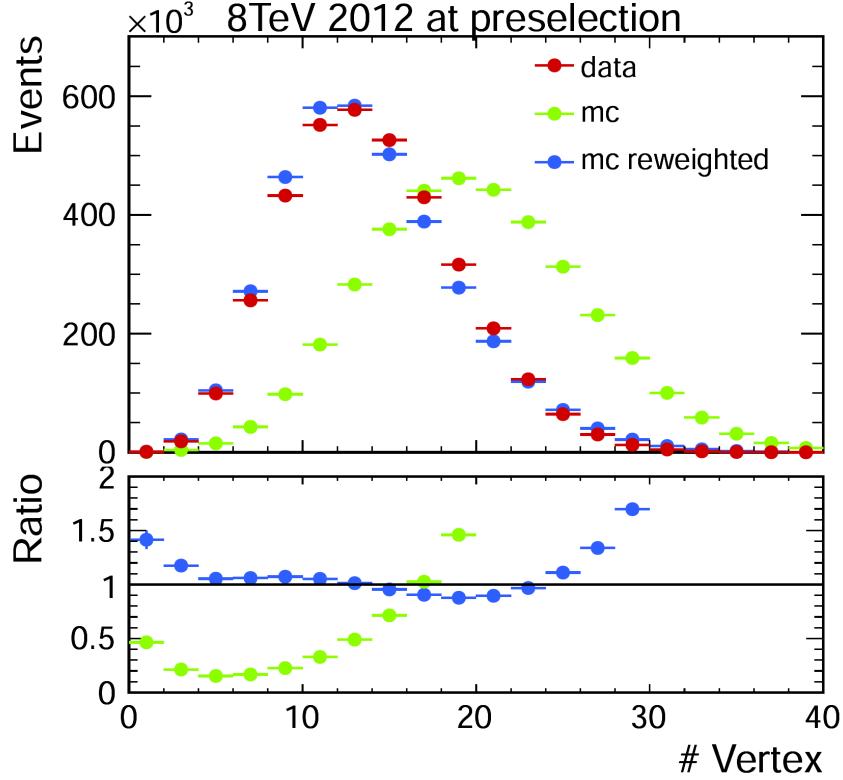
renormalization and the factorization scales by a factor of two. The POWHEG V2 generator is a matrix element based generator that provides a NLO description of the ggH process in association with zero jets, taking into account the finite mass of the bottom and top quarks. The POWHEG prediction is tuned 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 [75–77] for the decay of the Higgs boson to a W boson pair and interfaced with PYTHIA 8 [78] 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 [54], modified to be inclusive in the number of jets. This modification significantly reduces the uncertainties related to the modelling of the number of jets produced in association with the Higgs boson.

The signal contribution is extracted performing a template binned likelihood fit, using the two-dimensional  $(m_{\ell\ell}, m_T)$  shape for each background and signal process, as described in Sec. 4.6.

1396 controllare se  $m_{\ell\ell}$  e  $m_T$  sono già definite



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

### 1397 4.3.1. Event reconstruction and selections

1398 Electrons and muons used in the analysis are reconstructed using the PF technique as  
 1399 described in Sec. 3.2. In particular, muon candidates are required to be identified both as  
 1400 Tracker Muons and Global Muons.

1401 Jets are reconstructed using the standard PF algorithm and using the anti- $k_t$  clustering  
 1402 algorithm with  $R = 0.5$ , as described in Sec. 3.3. If not specified otherwise, jets considered  
 1403 for jet counting are the ones with  $p_T > 30$  GeV.

1404 In addition to the standard CMS PF  $E_T^{\text{miss}}$ , in this analysis a *projected*  $E_T^{\text{miss}}$  variable  
 1405 is also used. The *projected*  $E_T^{\text{miss}}$  is defined as the component of  $\vec{p}_T^{\text{miss}}$  transverse to the  
 1406 nearest lepton if the lepton is situated within the azimuthal angular window of  $\pm\pi/2$  from  
 1407 the  $\vec{p}_T^{\text{miss}}$  direction, or the  $E_T^{\text{miss}}$  itself otherwise. Since the  $E_T^{\text{miss}}$  resolution is degraded  
 1408 by pileup, the minimum of two projected  $E_T^{\text{miss}}$  variables is used: one constructed from  
 1409 all identified particles (full projected  $E_T^{\text{miss}}$ ), and another constructed from the charged  
 1410 particles only (track projected  $E_T^{\text{miss}}$ ).

1411 Background events from  $t\bar{t}$  and  $tW$  production are rejected applying a soft-muon veto  
 1412 and b tagging veto. The soft-muon algorithm is designed to identify muons from b quark

1413 decays. Events containing a muon satisfying the following requirements are rejected by the  
1414 soft-muon veto:

- 1415 • reconstructed as TrackerMuon;
- 1416 • number of hits in the Silicon Tracker greater than 10;
- 1417 • transverse impact parameter less than 0.2 cm;
- 1418 • relative isolation greater than 0.1 for muons with  $p_T > 20$  GeV.

1419 The b tagging veto rejects events that contain jets identified as b-jets using two different  
1420 algorithms for high and low  $p_T$  jets (see Sec. 3.4). For jets with  $p_T$  between 10 and 30 GeV,  
1421 the TCHE algorithm is applied. Low- $p_T$  jets passing the TCHE discriminant threshold of  
1422 2.1 are tagged as b-jets. For jets with  $p_T > 30$  GeV, a better performing algorithm, JP, is  
1423 used. Jets are identified as b-jets by the JP algorithm if the discriminating variable has  
1424 a value above 1.4. In the following, a b tagged jet is defined as a jet, within  $|\eta| < 2.4$   
1425 (b-tagging requires the tracker information), and with a value of the discriminating variable  
1426 above the mentioned thresholds for the two algorithms.

1427 The event selection consists of several steps. The first step is to select WW-like events  
1428 applying a selection that consists of the following set of cuts:

1429 1. **Lepton preselection:**

- 1430 • two opposite charge and different flavour ( $e\mu$ ) isolated leptons reconstructed in  
1431 the event;
- 1432 •  $|\eta| < 2.5$  for electrons and  $|\eta| < 2.4$  for muons;
- 1433 •  $p_T > 20$  GeV for the leading lepton. For the trailing lepton, the transverse  
1434 momentum is required to be larger than 10 GeV.

1435 2. **Extra lepton veto:** the event is required to have two and only two leptons with  
1436 opposite charge passing the lepton selection.

1437 3.  **$E_T^{\text{miss}}$  preselection:** particle flow  $E_T^{\text{miss}}$  is required to be greater than 20 GeV.

1438 4. **projected  $E_T^{\text{miss}}$  selection:** minimum projected  $E_T^{\text{miss}}$  required to be larger than  
1439 20 GeV.

1440 5. **Di-lepton mass cut:**  $m_{\ell\ell} > 12$  GeV in order to reject low mass resonances and QCD  
1441 backgrounds.

1442 6. **Di-lepton  $p_T$  cut:**  $p_T^{\ell\ell} > 30$  GeV to reduce the contribution of W+jets and DY to  
1443  $\tau\tau$  backgrounds.

1444 7. **Transverse mass:**  $m_T > 60$  GeV to reject DY to  $\tau\tau$  events.

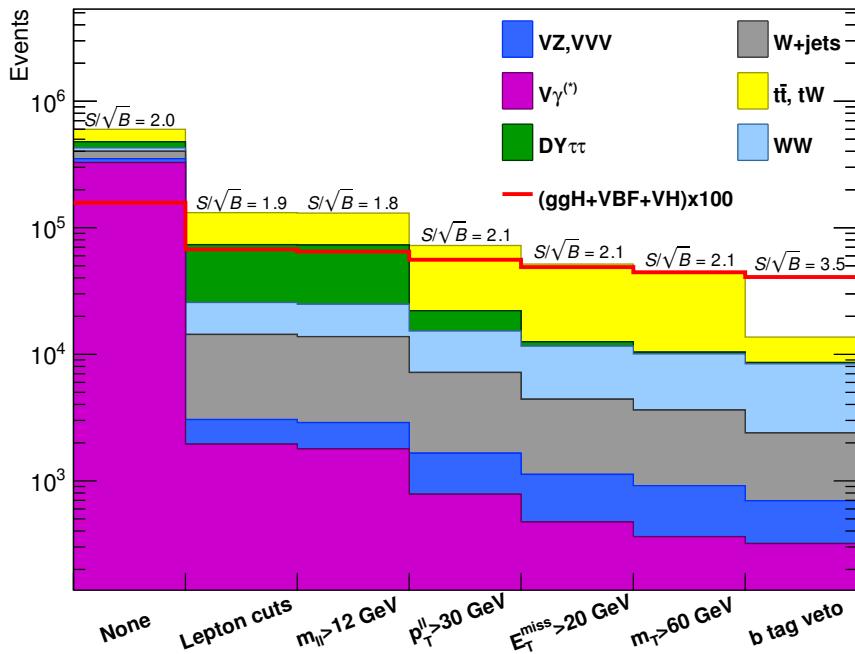
1445 The requirement of different flavour leptons in the final state is important in order to  
1446 suppress the sizeable contribution of backgrounds containing a same flavour lepton pair  
1447 originating from Z boson decay.

1448 Events surviving these requirements are dominantly those where a top quark-antiquark  
1449 pair is produced and both W bosons, which are part of the top quark decay chain, decay  
1450 leptonically (dileptonic  $t\bar{t}$ ). Two different selections are used depending on the number of

1451 jets in the event. This is done to suppress the top quark background both in the low  $p_T^H$   
1452 region, where 0-jets events have the largest contribution, and for higher  $p_T^H$  values where  
1453 also larger jet multiplicity events are important. The selection for 0-jets events relies on the  
1454 soft-muon veto and on a soft jet (with  $p_T < 30$  GeV) b tagging veto. The latter requirement  
1455 exploits the TCHE algorithm to reject soft jets that are likely to come from b quarks  
1456 hadronization.

1457 For events with a jet multiplicity greater or equal than one, a different selection is  
1458 applied. In this case we exploit the good b tagging performances of the JP tagger to reject  
1459 all the jets with  $p_T > 30$  GeV that are likely to come from b quarks hadronization. The  
1460 analysis selection requires to have no events containing b-tagged jets with  $p_T > 30$  GeV.

1461 A cut-flow plot is reported in Fig. 4.4, showing the effect of each selection using signal  
1462 and background simulations. In the first bin, labelled as “No cut”, no selection is applied  
1463 and the bin content corresponds to the total expected number of events with a luminosity  
1464 of  $19.4 \text{ fb}^{-1}$ . All the events in this bin have at least two leptons with a loose transverse  
1465 momentum cut of 8 GeV. In the following bin the lepton cuts are applied, including the  
1466 requirement to have two opposite sign and different flavour leptons and the extra lepton  
1467 veto. Then all the other selections are progressively reported, showing the effect of each cut  
1468 on the background and signal yields. For each selection the expected signal over background



**Figure 4.4.:** Effect of selection cuts on simulated 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  $19.46 \text{ fb}^{-1}$ .

1469 ratio is also shown, which, after the full selection requirements, reaches a maximum value  
1470 of about 3%.

### 1471 4.3.2. Simulation efficiencies and scale factors

1472 The efficiencies for the identification and isolation of the electrons and muons are measured  
1473 in data and simulation selecting a pure sample of leptons coming from the  $Z \rightarrow \ell\ell$  decay,  
1474 and using the Tag and Probe technique described in Sec. 3.2.7. The efficiencies for data  
1475 and simulation are used as scale factors to correct the simulated events to precisely model  
1476 the data.

1477 The trigger efficiency is measured in data and applied to simulation as explained in  
1478 Sec. 3.2.8.

1479 The efficiency of b tagging algorithms is not well simulated by MC generators and  
1480 discrepancies can occur with respect to the data. For this reason is important to measure  
1481 the b tagging efficiency and the misidentification probability for the given algorithms both  
1482 in data and simulation, and to correct the simulated events using scale factors. This affects  
1483 not only the top quark background estimation, but also the other backgrounds and the  
1484 signal. As an example, if a light-parton jet in a signal event was misidentified as a b-jet,  
1485 this event would be rejected by the b-jet veto.

1486 In this analysis, the b tagging efficiency and the misidentification probability are  
1487 measured both in data and simulation, selecting a control sample enriched in b-jets, and  
1488 using a Tag and Probe technique similar to the one described in Sec. 3.2.7. Below is  
1489 described the method used to estimate the efficiency of the JP b tagging algorithm, but it  
1490 is extendible to any other algorithm.

1491 The control sample is defined selecting the events that pass the selections listed in  
1492 Sec. 4.3.1, and have at least two jets with  $p_T$  greater than 30 GeV. If the leading jet  
1493 has a JP discriminator values above the threshold of 0.5, it is considered a *tag*, and the  
1494 sub-leading jet is the *probe*. In order to avoid any bias that could arise from the probe  
1495 being always the sub-leading jet, the pair is tested also in reverse order, i.e. sub-leading jet  
1496 is tested against the *tag* selection, and in case it passes, then the leading jet is used as *probe*  
1497 forming an independent *tag-probe* pair. If the *probe* jet has a discriminator value above the  
1498 threshold used in the analysis, i.e.  $> 1.4$ , then the *tag-probe* pair is called a *tag-pass-probe*  
1499 pair. Otherwise it is identified as a *tag-fail-probe* pair.

1500 If the *tag* selection was sufficient to suppress any non top quark event, one could estimate  
1501 the efficiency by dividing the number of *tag-probe* pairs in which the *probe* passes the analysis  
1502 JP requirement by the total number of *tag-probe* pairs. However this is not the case, since  
1503 the contamination due to other background sources is not negligible. In order to estimate  
1504 the efficiency in the presence of background, a variable that discriminates between true  
1505 b-jets and other jets in a  $t\bar{t}$  sample is needed. This variable is the  $p_T$  of the *probe* jet. For  
1506 real b-jets this variable has a peak around 60 GeV, while it has a broad distribution for  
1507 other types of jets.

1508 The efficiencies are estimated performing a  $\chi^2$  simultaneous fit of the *probe*  $p_T$  spectrum  
1509 in two different categories: one containing events with a *tag-pass-probe* pair and the other

1510 containing events with a *tag-fail-probe* pair. The normalisations in the two categories are  
 1511 linked by the following formulas:

$$\begin{aligned} N_{\text{TPP}} &= N_s \varepsilon_s + N_b \varepsilon_b \\ N_{\text{TFP}} &= N_s (1 - \varepsilon_s) + N_b (1 - \varepsilon_b) \quad , \end{aligned} \quad (4.1)$$

1512 where:

- 1513 •  $N_{\text{TPP}}$  is the number of *tag-pass-probe* pairs;
- 1514 •  $N_{\text{TFP}}$  is the number of *tag-fail-probe* pairs;
- 1515 •  $N_s$  is the number of *tag-probe* pairs in which the *probe* is a b-jet;
- 1516 •  $N_b$  is the number of *tag-probe* pairs in which the *probe* is not a b-jet;
- 1517 •  $\varepsilon_s$  is the efficiency to identify a b-jet, i.e. the b tagging efficiency;
- 1518 •  $\varepsilon_b$  is the probability to misidentify a non b-jet as a b-jet, i.e. the misidentification  
 1519 probability<sup>1</sup>.

1520 The  $p_T$  shapes of the *probe* jet used in the fit are taken from simulation, where the real  
 1521 flavour of the jet is known, both for the *tag-pass-probe* and *tag-fail-probe* categories. To  
 1522 check the consistency of the fitting procedure, a closure test fitting the simulation itself has  
 1523 been performed. The result of the fit on MC simulation is shown in Fig. 4.5. The relevant  
 1524 efficiencies are:

$$\begin{aligned} \varepsilon_s^{\text{MC}} &= 0.766 \pm 0.007 \\ \varepsilon_b^{\text{MC}} &= 0.208 \pm 0.015 \quad . \end{aligned} \quad (4.2)$$

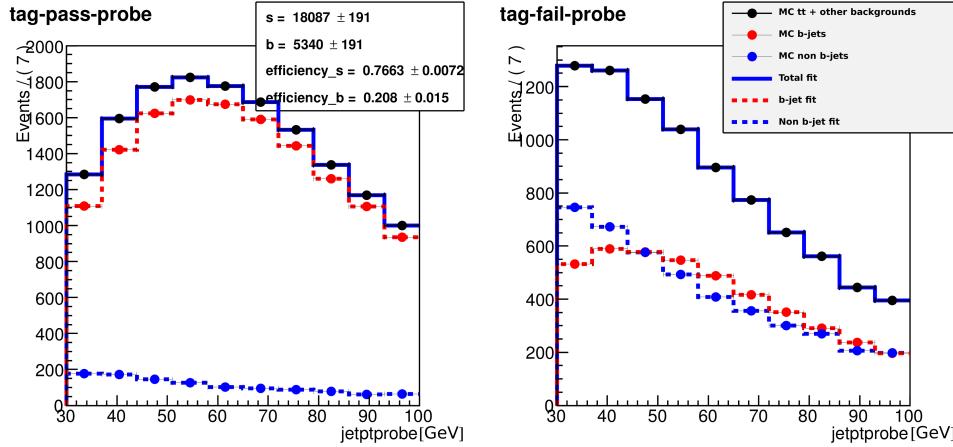
1525 These values are consistent with the true value of the b tagging efficiency in simulation. The  
 1526 true value is computed by selecting jets that are matched within a cone of  $\Delta R < 0.5$  with  
 1527 a generator level b quark, and counting the fraction of those that have a JP discriminator  
 1528 above threshold of 1.4. This check also assures that the *tag-probe* method does not introduce  
 1529 any bias within the simulation statistic accuracy.

1530 In order to assess the robustness of the fit, 5000 toy simulated samples have been  
 1531 generated with a statistics equivalent to the one expected in data and the same fit is  
 1532 performed. All the 5000 fit succeeded, and the pull distributions for  $\varepsilon_s$  and  $\varepsilon_b$  parameters  
 1533 are shown in Fig. ???. The distributions represent the *pull* of the efficiencies measured in  
 1534 the fit, where the pull variable for each toy  $i$  is defined as:

$$\text{pull}(\varepsilon_{s(b)}) = \frac{\varepsilon_{s(b)}^{\text{true}} - \varepsilon_{s(b)}^i}{\sigma(\varepsilon_{s(b)}^i)} \quad , \quad (4.3)$$

---

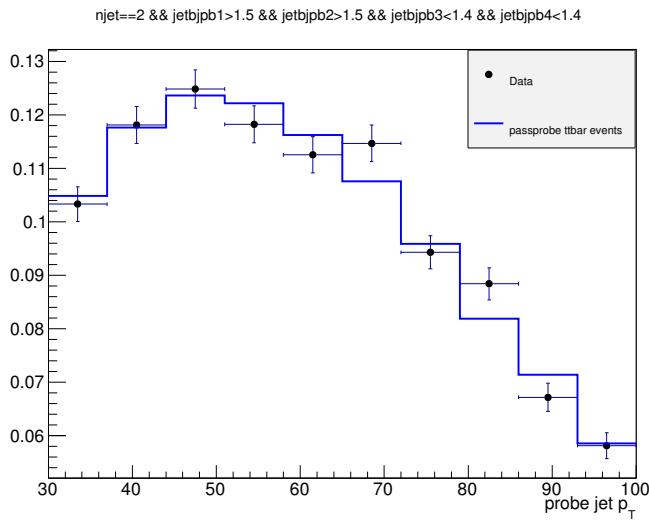
<sup>1</sup>In these naming convention, the subscript “s” stays for “signal”, since the b-jets represent the signal in this method. Similarly, the “b” subscript stays for “background”, identifying the cases where the *probe* is not a b-jet



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

where  $\sigma(\varepsilon_{s(b)}^i)$  is the uncertainty on the efficiency extracted from the fit. The pull distributions are centred on zero and have  $\sigma$  close to one, as expected.

Before running the fit on data, the shapes used in the fit have been validated. To do so, a very pure phase space enriched in b jets has been defined by selecting events containing exactly two jets with a JP discriminator greater than 1.5 and no additional b-tagged jets, rejecting also events containing jets with  $p_T$  smaller than 30 GeV. On this very pure sample, data have been compared against the shape used to fit the true b-jets in the *tag-pass-probe* distribution. The result is shown in Fig. 4.6 and shows good agreement within uncertainties.



**Figure 4.6.:** Shape comparison for the  $p_T$  spectrum of the *probe* jet in data and simulation in a very pure phase space enriched in b-jets.

Finally the fit has been performed on data, as shown in Fig. 4.7, providing the following efficiencies:

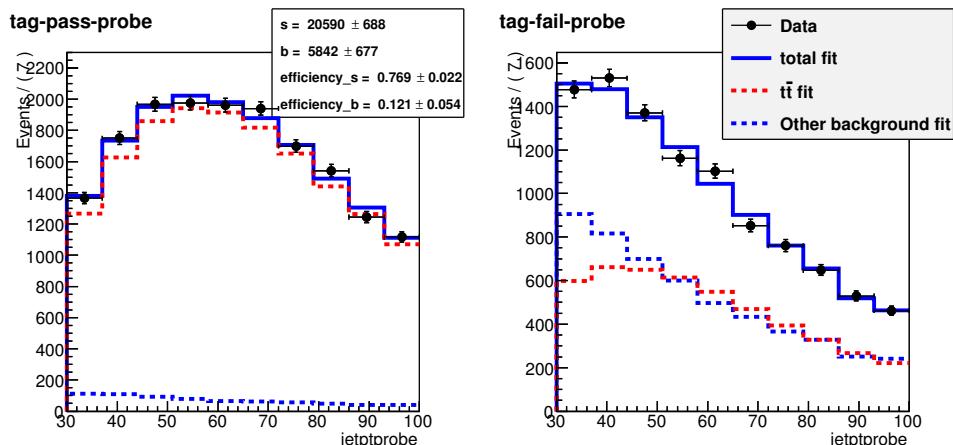
$$\begin{aligned}\varepsilon_s^{\text{Data}} &= 0.77 \pm 0.02 \\ \varepsilon_b^{\text{Data}} &= 0.12 \pm 0.05\end{aligned}\quad (4.4)$$

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

The ratio of the efficiency measured in data and simulation represents a per-jet scale factor that can be used to reweight the simulated events. The weights to be applied event-by-event depend on the particular jet configuration in the events itself. For the signal region ( $SR$ ), in which a b tagging veto is required, the event weight to be applied is given by:

$$w_{SR} = \prod_{N_{\text{b-jets}}} \left( \frac{1 - \varepsilon_s^{\text{Data}}}{1 - \varepsilon_s^{\text{MC}}} \right) \prod_{N_{\text{non b-jets}}} \left( \frac{1 - \varepsilon_b^{\text{Data}}}{1 - \varepsilon_b^{\text{MC}}} \right) , \quad (4.5)$$

where  $N_{\text{b-jets}}$  and  $N_{\text{non b-jets}}$  are the number of true b-jets and the number of non b-jets in the simulated event, respectively. This weight is valid if the a b tagging veto is applied. If instead the b tagging veto is reverted, also the event weight has to be modified. This is done, for example, when one wants to define a  $t\bar{t}$  enriched control region ( $CR_{t\bar{t}}$ ) for the purpose of measuring the contribution of this background in a phase space orthogonal to the signal region. One simple way to define this control region is to require the leading jet



**Figure 4.7.:** Simultaneous fit of the *tag-pass-probe* and *tag-fail-probe* pairs in data.

1561 in the event to be b-tagged. Therefore, the simulated events falling in this category must  
 1562 be reweighted using the following weight:

$$w_{CR_{t\bar{t}}} = \begin{cases} \varepsilon_s^{\text{Data}} / \varepsilon_s^{\text{MC}}, & \text{if the leading jet is a b-jet} \\ \varepsilon_b^{\text{Data}} / \varepsilon_b^{\text{MC}}, & \text{if the leading jet is not a b-jet} \end{cases} \quad (4.6)$$

1563 **4.3.3. Fiducial phase space**

1564 The Higgs boson transverse momentum is measured in a fiducial phase space, whose  
 1565 definition is chosen in order to minimize the dependence of the measurements on the  
 1566 underlying model of the Higgs boson production and decay properties.

1567 The exact requirements are determined by considering the two following correlated  
 1568 quantities: the reconstruction efficiency for signal events originating from within the fiducial  
 1569 phase space (fiducial signal efficiency  $\varepsilon_{\text{fid}}$ ), and the ratio of the number of reconstructed  
 1570 signal events that are from outside the fiducial phase space (“out-of-fiducial” signal events)  
 1571 to the number from within the fiducial phase space. The requirement of having a small  
 1572 fraction of out-of-fiducial signal events, while at the same time preserving a high value of  
 1573 the fiducial signal efficiency  $\varepsilon_{\text{fid}}$ , leads to fiducial requirements at the generator level on the  
 1574 low-resolution variables,  $E_T^{\text{miss}}$  and  $m_T$ , that are looser with respect to those applied in the  
 1575 reconstructed event selection.

1576 The fiducial phase space used for the cross section measurements is defined at the  
 1577 particle level by the requirements given in Table 4.2. The leptons are defined as Born-level  
 1578 leptons, i.e. before the emission of final-state radiation (FSR), and are required not to  
 1579 originate from leptonic  $\tau$  decays. The effect of including FSR is found to modify  $\varepsilon_{\text{fid}}$  at  
 1580 most of about 5%. For the VH signal process, the two leptons are required to originate  
 1581 from the  $H \rightarrow WW \rightarrow 2\ell 2\nu$  decays in order to avoid including leptons coming from the  
 1582 associated W or Z boson.

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

Physics quantity	Requirement
Leading lepton $p_T$	$p_T > 20 \text{ GeV}$
Subleading lepton $p_T$	$p_T > 10 \text{ GeV}$
Pseudorapidity of electrons and muons	$ \eta  < 2.5$
Invariant mass of the two charged leptons	$m_{\ell\ell} > 12 \text{ GeV}$
Charged lepton pair $p_T$	$p_T^{\ell\ell} > 30 \text{ GeV}$
Invariant mass of the leptonic system in the transverse plane	$m_T^{\ell\ell\nu\nu} > 50 \text{ GeV}$
$E_T^{\text{miss}}$	$E_T^{\text{miss}} > 0$

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

#### 4.3.4. Binning of the $p_T^H$ distribution

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

$$\vec{p}_T^H = \vec{p}_T^{\ell\ell} + \vec{p}_T^{\text{miss}} \quad (4.7)$$

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

Given these aspects, the criterion that is used to define the  $p_T^H$  bin size is devised to keep under control the bin migrations due to the finite resolution. For any given bin  $i$ , the purity  $P_i$  of the signal sample is defined as the number events that are generated and also reconstructed in that bin, i.e.  $N_i^{\text{GEN|RECO}}$ , divided by the number of events reconstructed in the same bin,  $N_i^{\text{RECO}}$ :

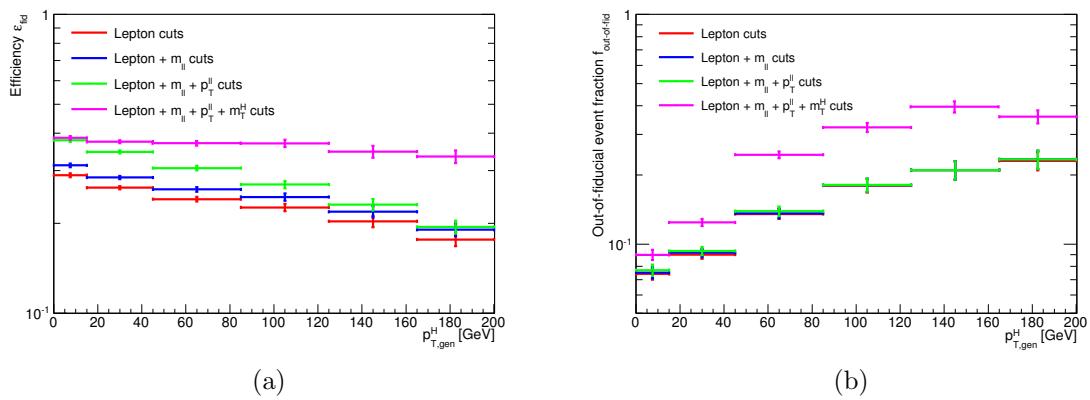
$$P_i = \frac{N_i^{\text{GEN|RECO}}}{N_i^{\text{RECO}}} \quad . \quad (4.8)$$

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

The fiducial signal efficiency  $\varepsilon_{\text{fid}}$  and the fraction of out-of-fiducial signal events,  $f_{\text{out-of-fid}}$ , are different in each  $p_T^H$  bin and depend on the definition of the fiducial phase space. In Fig. 4.8 the  $\varepsilon_{\text{fid}}$  and  $f_{\text{out-of-fid}}$  parameters are shown in each  $p_T^H$  bin for different definitions of the fiducial phase space. In particular, they have been evaluated adding the requirements reported in Table 4.2 in sequence, starting from a fiducial phase space defined just by the lepton  $p_T$  and  $\eta$  selections, together with the different flavour requirement, and adding each time an additional selection until the full fiducial phase space is obtained. In this way, the effect of every single selection (or group of selections) on  $\varepsilon_{\text{fid}}$  and  $f_{\text{out-of-fid}}$  can be assessed. Since the variables related to leptons are measured with good resolution, the effect of including the related selections in the fiducial phase space is to increase  $\varepsilon_{\text{fid}}$  keeping  $f_{\text{out-of-fid}}$  constant. Instead, the effect of including low-resolution variables, such as  $m_T$ , is to increase both  $\varepsilon_{\text{fid}}$  and  $f_{\text{out-of-fid}}$ . Nevertheless, the  $f_{\text{out-of-fid}}$  parameter is

different from zero even if only lepton cuts are taken into account. This is ascribable to two different aspects: the first one is that in the fiducial definition electrons and muons are required not to originate from  $\tau$  decays; the second one is instead related to the VH production mechanism, i.e. to the fact that leptons coming from the associated boson are not included.

The overall values integrating over  $p_T^H$  are  $\varepsilon_{\text{fid}} = 0.362 \pm 0.005$  and  $f_{\text{out-of-fid}} = 0.126 \pm 0.004$  respectively, where only statistical uncertainties are taken into account.



**Figure 4.8.:** Fiducial signal efficiency  $\varepsilon_{\text{fid}}$  and fraction of out-of-fiducial signal events  $f_{\text{out-of-fid}}$  in each bin of the generator level  $p_T^H$ .

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

## 4.4. Background estimation

The signal extraction procedure requires the determination of the normalization and the  $(m_{\ell\ell}, m_T)$  shape for each background source. In this section, the methods used to estimate these quantities are described.

### 4.4.1. Top quark background

In this analysis the top quark background is divided into two different categories depending on the number of jets in the event. Different selections are applied in the two categories, especially concerning the b tagging requirements, as explained in Sec. 4.3.1. A top quark enriched control region,  $CR_{t\bar{t}}$ , is defined for each of the two categories in order to estimate the process cross section directly from data. For the category with 0 counted jets, the control region is defined selecting events containing at least one soft jet, i.e. with  $p_T < 30 \text{ GeV}$ , that is identified as a b-jet by the TCHE and soft-muon algorithms. In the category with more than 0 counted jets, a similar control region is defined requiring events to contain at

least one jet with  $p_T > 30$  GeV identified by the JP algorithm. The control regions defined in this way are very pure and can be used to normalize the simulation prediction to the data (data-driven method).

Since the  $CR_{t\bar{t}}$  control region and the signal region ( $SR$ ) are orthogonal, a factor  $\alpha$  connecting the number of events in the two has to be evaluated using simulation:

$$\alpha = \frac{N_{MC}^{SR}}{N_{MC}^{CR_{t\bar{t}}}} , \quad (4.9)$$

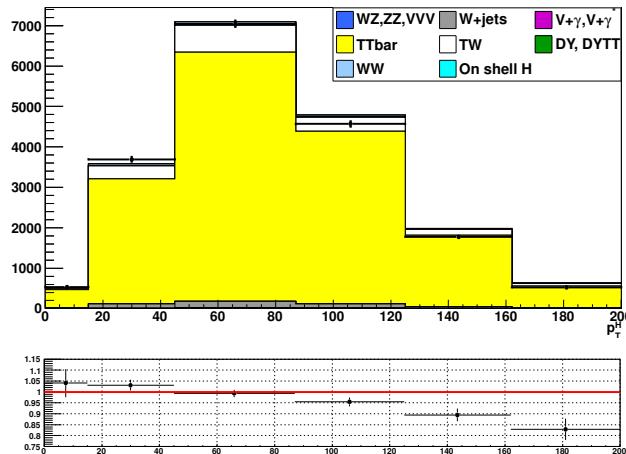
where  $N_{MC}^{SR}$  and  $N_{MC}^{CR_{t\bar{t}}}$  are the number of  $t\bar{t}$  events obtained from simulation in the  $SR$  and  $CR_{t\bar{t}}$  regions, respectively. The number of  $t\bar{t}$  events in data in the  $CR_{t\bar{t}}$  region,  $N_{Data}^{CR_{t\bar{t}}}$ , can be estimated subtracting the expected number of non  $t\bar{t}$  events. Finally, the number of expected events in the  $SR$  region  $N_{Data}^{SR}$ , can be obtained using the following formula:

$$N_{Data}^{SR} = \alpha N_{Data}^{CR_{t\bar{t}}} . \quad (4.10)$$

Before applying this procedure, the simulated events have been reweighted using the b tagging efficiency scale factors described in Sec. 4.3.2.

In the 0 jets category the residual top quark background is very small, and its normalization is estimated inclusively in  $p_T^H$ . For the other category, the aforementioned procedure is applied independently in each  $p_T^H$  bin, rather than inclusively, because an overall normalization factor would not be able to cover the discrepancies between data and simulation from bin to bin. This can be understood looking at the  $p_T^H$  distribution in the  $CR_{t\bar{t}}$  region, shown in Fig. 4.9, where the data and simulation are compared.

The results of the method discussed before are listed in Tab. 4.3 for each bin of  $p_T^H$ .



**Figure 4.9.:** Distribution of  $p_T^H$  in the  $CR_{t\bar{t}}$  region, comparing data and simulation.

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

$p_T^H$ [GeV]	$N_{\text{Data}}^{CR_{t\bar{t}}}$	$N_{\text{MC}}^{CR_{t\bar{t}}}$	$N_{\text{MC}}^{SR}$	$\alpha$	$\Delta\alpha$
[0–15]	406.7	358.8	117.8	0.33	0.08
[15–45]	2930.1	2703.4	859.1	0.32	0.07
[45–85]	5481.0	5207.5	1506.1	0.29	0.07
[85–125]	4126.4	4032.6	861.2	0.21	0.05
[125–165]	1612.6	1654.3	304.7	0.18	0.06
[165– $\infty$ ]	647.5	760.4	201.7	0.27	0.15

A comparison of the  $m_{\ell\ell}$  distributions in the  $CR_{t\bar{t}}$  region for data and simulation is shown in Fig. 4.10, separately for each  $p_T^H$  bin and after the application of the data-driven factors. The agreement between data and simulation is found to be satisfactory within uncertainties.

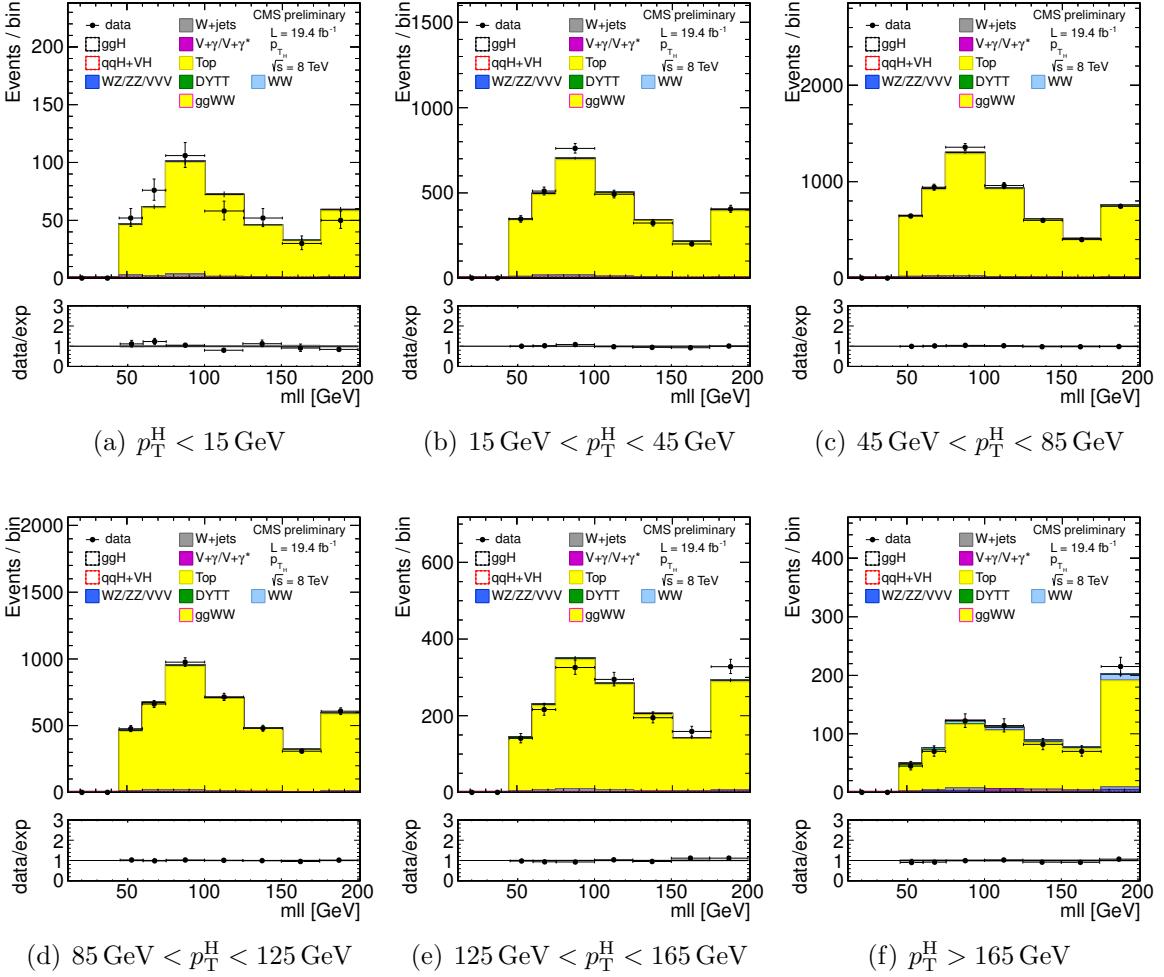
#### 4.4.2. WW background

First of all, the  $m_{\ell\ell}$  and  $m_T$  shapes of the  $qq \rightarrow W^+W^-$  background process have been compared to the data in a signal free phase space, as shown in Fig. 4.11. The signal-free region is defined requiring the preselections described in Sec. 4.3.1 with the additional cut  $m_{\ell\ell} > 70$  GeV. The comparison, which is performed inclusively in  $p_T^H$  in the 0 and 1 jet categories, shows a good data to simulation agreement within uncertainties.

In this analysis the  $qq \rightarrow W^+W^-$  background normalization is left free to float in each of the six  $p_T^H$  bins to match the data during the fitting procedure. This choice helps mitigating the  $p_T^H$  shape difference between data and simulation. This difference is due to missing higher order QCD corrections in the adopted simulation, obtained using the MADGRAPH generator. In fact, as shown in Fig. 4.12, the theoretical calculations for this process performed at NLO QCD accuracy and including soft-gluon resummation effects, i.e. NLO+NNLL accuracy, predict a rather different  $p_T^H$  spectrum.

To check the residual dependence of the  $m_{\ell\ell}$  and  $m_T$  shapes on the generator used for simulating the  $qq \rightarrow W^+W^-$  process, the shapes obtained using different generators have been compared in each  $p_T^H$  bin, as shown in Figs. 4.13 and 4.14. The usage of different generators only mildly affect the  $m_{\ell\ell}$  and  $m_T$  shapes. Nevertheless the observed differences are taken as shape uncertainties and propagated through the fit.

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 [64, 65].

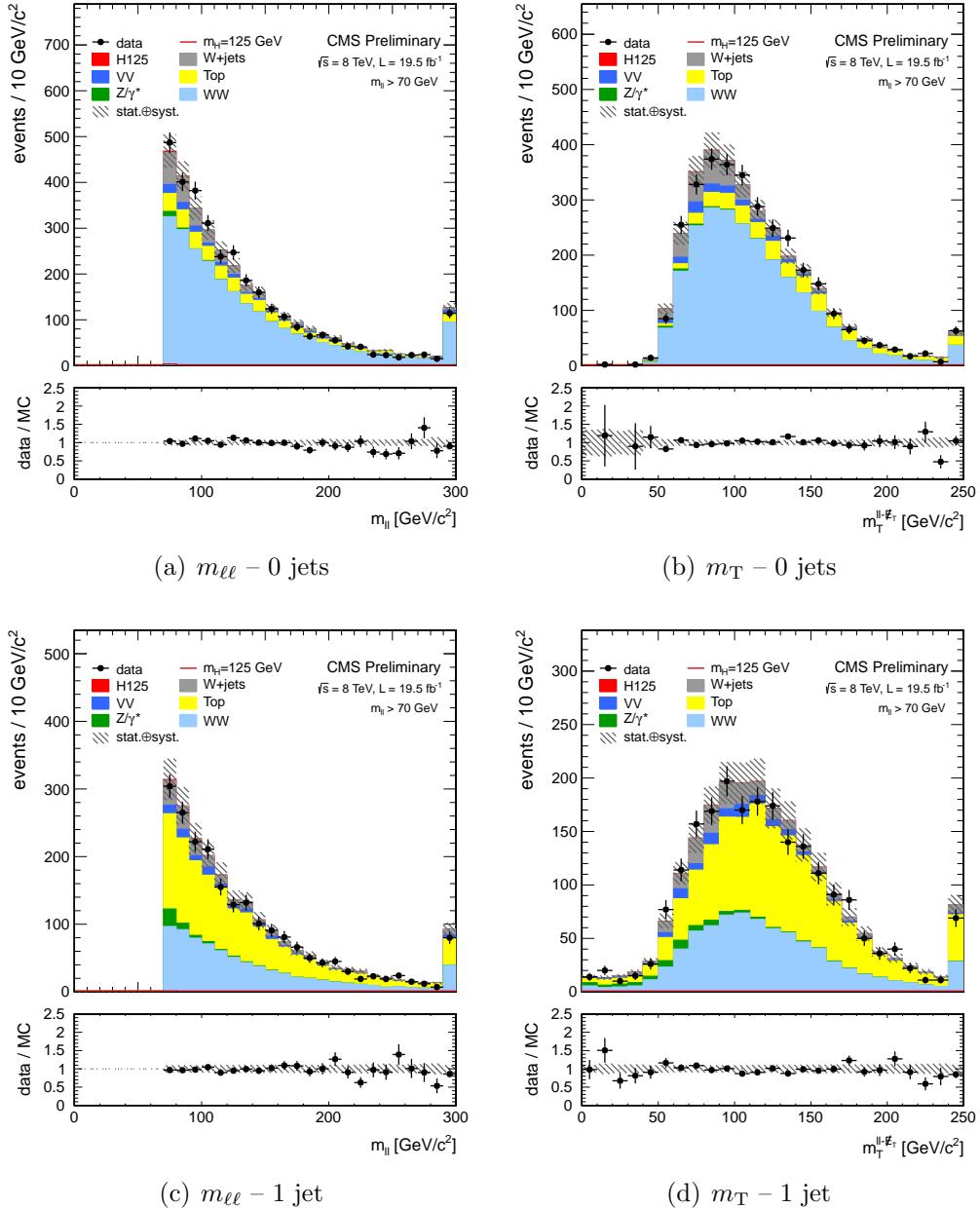


**Figure 4.10.:** Comparison of the  $m_{\ell\ell}$  distributions in the  $CR_{t\bar{t}}$  region for data and simulation in each  $p_T^H$  bin, after the application of the data-driven factors.

#### 4.4.3. W+jets background

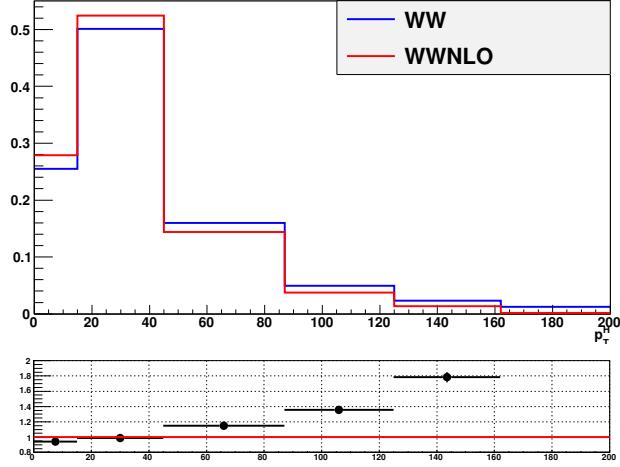
The non-prompt lepton background, originating from leptonic decays of heavy quarks, hadrons misidentified as leptons, and electrons from photon conversions in W+jets and QCD multijet production, is suppressed by the identification and isolation requirements on electrons and muons, as described in Sec. 3.2. The remaining contribution from the non-prompt lepton background is estimated directly from data and is ascribable especially to W+jets production. A control sample is defined by one lepton that passes the standard lepton selection criteria and another lepton candidate that fails the criteria, but passes a looser selection, resulting in a sample of “pass-fail” lepton pairs.

The efficiency,  $\varepsilon_{\text{pass}}$ , for a jet that satisfies the loose lepton requirements to pass the standard selection is determined using an independent sample dominated by events with

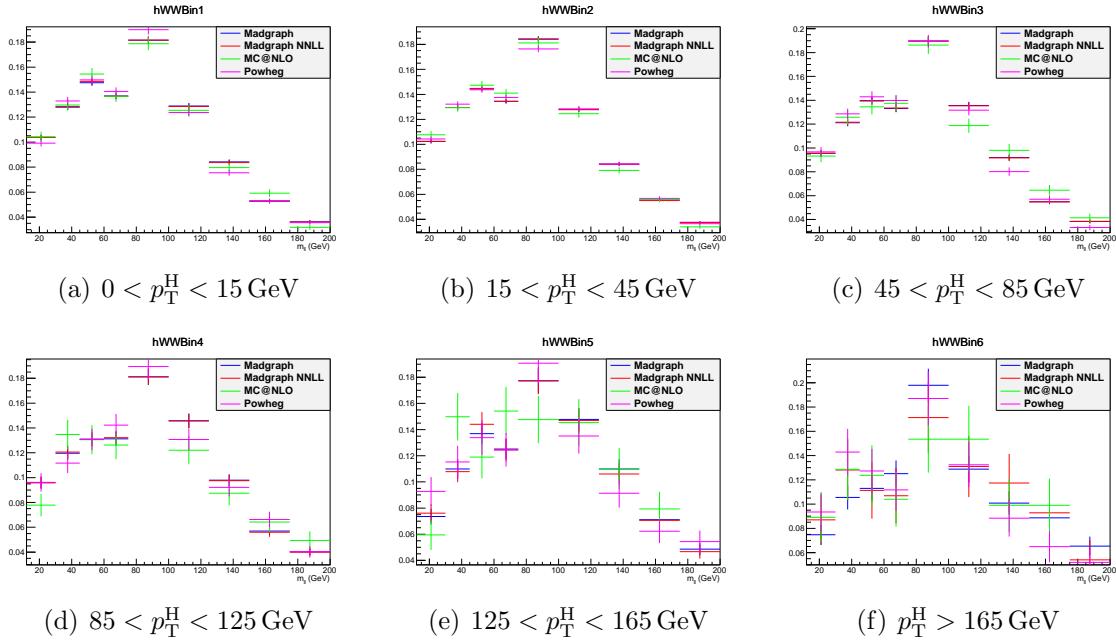


**Figure 4.11.:** Comparison of the  $m_{\ell\ell}$  and  $m_T$  shapes in data and simulation for events with 0 and 1 jets, inclusive in  $p_T^H$ . The events are required to pass the analysis requirements and, in order to define a signal-free control region, to have  $m_{\ell\ell} > 70$  GeV.

non-prompt leptons from QCD multijet processes. However, this sample is not a pure sample containing just non-prompt leptons, but may still contain prompt leptons coming from the W and Z boson decays. To reject muons from the W decay, the events are required to have  $E_T^{\text{miss}} < 20$  GeV and a W transverse mass below 20 GeV as well. Muons from the Z

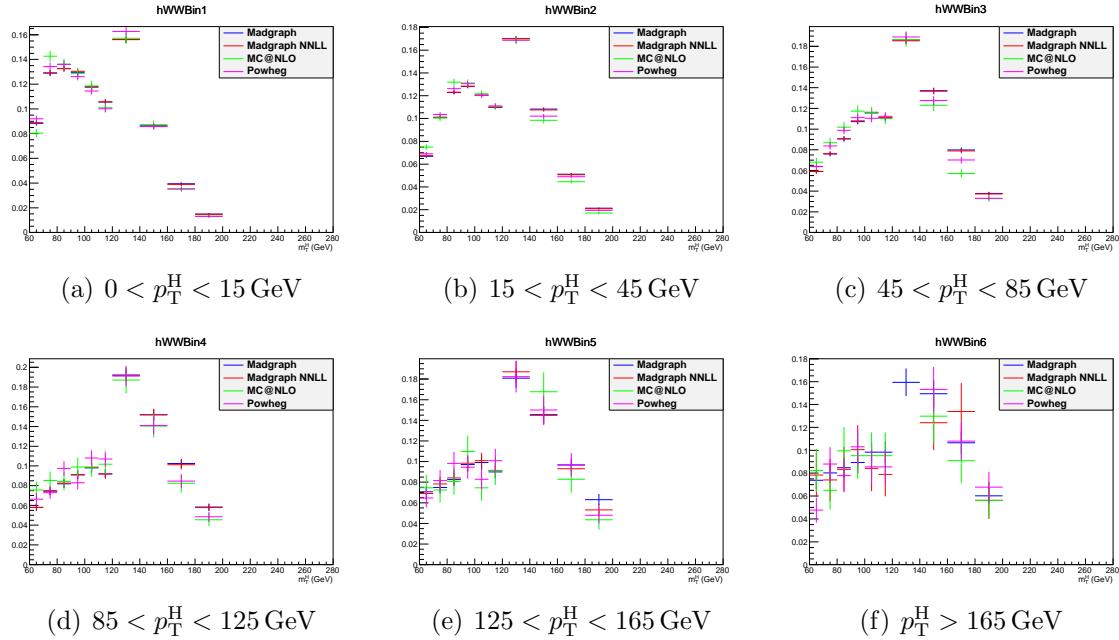


**Figure 4.12.:** Comparison between the  $p_T^{WW}$  distributions obtained with two different MC generators: the blue line corresponds to the MADGRAPH generator and the red line refers to the same sample in which a reweighting has been applied in order to match the theoretical prediction at NLO+NNLL precision.



**Figure 4.13.:** Comparison between the  $m_{\ell\ell}$  shape obtained with the default  $qq \rightarrow W^+W^-$  background simulation (MADGRAPH) and other theoretical models in every  $p_T^H$  bin.

decay are instead removed requiring  $m_{\mu\mu} \notin [76, 106]$  GeV. For electrons the Z mass peak veto is enlarged to  $m_{ee} \notin [60, 120]$  GeV. Finally, prompt electrons and muons are required to be isolated from the leading jet in the event, i.e.  $\Delta\phi(\ell, j) > 1$ . The residual prompt



**Figure 4.14.:** Comparison between the  $m_T$  shape obtained with the default  $\text{qq} \rightarrow \text{W}^+\text{W}^-$  background simulation (MADGRAPH) and other theoretical models in every  $p_T^H$  bin.

lepton contamination from EW processes such as  $\text{W}/\text{Z}+\text{jets}$  production, which can bias the fake rate measurement, is estimated using simulation and subtracted. This contribution is negligible for small values of the lepton  $p_T$  and increases at larger  $p_T$  values.

The  $\varepsilon_{\text{pass}}$  efficiency, parameterized as a function of  $p_T$  and  $\eta$  of the lepton, is then used to weight the events in the pass-fail sample by  $\varepsilon_{\text{pass}}/(1 - \varepsilon_{\text{pass}})$ , to obtain the estimated contribution from the non-prompt lepton background in the signal region. The systematic uncertainties from the determination of  $\varepsilon_{\text{pass}}$  dominate the overall uncertainty of this method. The systematic uncertainty has two sources: the dependence of  $\varepsilon_{\text{pass}}$  on the sample composition, and the method. The first source is estimated by modifying the jet  $p_T$  threshold in the QCD multijet sample, which modifies the jet sample composition. The uncertainty in the method is obtained from a closure test, where  $\varepsilon_{\text{pass}}$  is derived from simulated QCD multijet events and applied to simulated samples to predict the number of background events. The total uncertainty in  $\varepsilon_{\text{pass}}$ , including the statistical precision of the control sample, is of the order of 40%.

A validation of the estimate of this background is performed in a control sample obtained selecting events with two leptons with same charge, which is enriched in  $\text{W}+\text{jets}$  events. The results of this closure test show good agreement between data and the estimated background.

1718 **4.4.4.  $Z/\gamma^* \rightarrow \tau^+\tau^-$ background**

1719 The low  $E_T^{\text{miss}}$  threshold in the  $e\mu$  final state requires the consideration of the contribution  
1720 from  $Z/\gamma^* \rightarrow \tau^+\tau^-$ , that estimated from data. This is accomplished by selecting  $Z/\gamma^* \rightarrow$   
1721  $\mu^+\mu^-$ events in data and replacing both muons with a simulated  $\tau \rightarrow \ell\nu_\tau\bar{\nu}_\ell$  decay [54], thus  
1722 obtaining a “hybrid” event. The Z boson four-momentum is reconstructed in data from the  
1723 four-momenta of the daughter muons. Then a simulation step allows the replacement of the  
1724 muon objects with  $\tau$  leptons, in such a way to preserve the Z boson momentum direction  
1725 in its rest frame. The  $Z/\gamma^* \rightarrow \tau^+\tau^-$ decay is simulated with the TAUOLA package [79] to  
1726 correctly describe the  $\tau$ -polarization effects.

1727 After replacing muons from  $Z/\gamma^* \rightarrow \mu^+\mu^-$ decays with simulated  $\tau$  decays, the set  
1728 of pseudo- $Z/\gamma^* \rightarrow \tau^+\tau^-$ events undergoes the reconstruction step. Good agreement in  
1729 kinematic distributions for this sample and a simulated  $Z/\gamma^* \rightarrow \tau^+\tau^-$ sample is found. The  
1730 global normalization of pseudo- $Z/\gamma^* \rightarrow \tau^+\tau^-$ events is checked in the low  $m_T$  spectrum,  
1731 where a rather pure sample enriched in  $Z/\gamma^* \rightarrow \tau^+\tau^-$ events is expected.

1732 This method allows to avoid the simulation of very large MC samples that would be  
1733 needed for an accurate description of this process, given its large cross section.

1734 **Diboson backgrounds**

1735 The WZ and ZZ background events are largely rejected by requiring exactly two high  $p_T$   
1736 isolated leptons with opposite charge and different flavour in the event.

1737 The  $W\gamma^*$  electroweak process is included in standard CMS simulations as a part of  
1738 the WZ process using the MADGRAPH generator. Nevertheless, the low  $m_{\ell\ell}$  region is  
1739 not properly covered since the standard simulations have a generator-level requirement of  
1740  $m_{\gamma^*} > 12$  GeV and there could be a significant rate of events below that threshold passing  
1741 the selection criteria of the analysis. The low  $m_{\ell\ell}$  spectrum of the  $W\gamma^*$  process has been  
1742 produced using a dedicated simulation with MADGRAPH, requiring two leptons each with  
1743  $p_T > 5$  GeV and no restrictions on the third lepton. However, in order to have a reliable  
1744 prediction of the background cross section, the simulation needs to be validated using data  
1745 in a control region.

1746 A high purity  $W\gamma^*$  phase space is defined selecting events with three muons, where the  
1747 two muons with lowest invariant mass, which is required to be less than 12 GeV, are assumed  
1748 to originate from the  $\gamma^*$  decay. The top quark background contribution is suppressed using a  
1749 b tagging veto. The W+jets and multijet contributions are rejected requiring the minimum  
1750 transverse mass of each lepton and  $E_T^{\text{miss}}$  to be larger than 25 GeV, and the transverse mass  
1751 of the lepton associated with the W boson and  $E_T^{\text{miss}}$  to be larger than 45 GeV. Moreover,  
1752 the  $J/\Psi$  meson decay are rejected by requiring  $|m_{\mu^\pm\mu^\mp} - m_{J/\Psi}| > 0.1$  GeV. The measured  
1753 data/simulation scale factor in this control region is found to be  $1.5 \pm 0.5$ .

1754 The  $W\gamma$  background can also contribute to the signal phase space, because the photon  
1755 can interact with the tracker material converting to an  $e^+e^-$  pair. Its normalization is taken  
1756 from simulation while the  $m_{\ell\ell}$  and  $m_T$  shapes are checked selecting a data sample with one  
1757 lepton and one photon, finding a good agreement within uncertainties.

## 1758 4.5. Systematic uncertainties

1759 Systematic uncertainties play an important role in this analysis where no strong mass peak  
1760 is expected due to the presence of undetected neutrinos in the final state. One of the most  
1761 important sources of systematic uncertainty is the normalization of the backgrounds that  
1762 are estimated on data control samples whenever is possible.

1763 A summary of the main sources of systematic uncertainty and the corresponding estimate  
1764 is reported in Table 4.4. A detailed description of each source of systematic uncertainty is  
1765 discussed in the following sections.

**Table 4.4.:** 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
$t\bar{t}$ , $tW$	20–50%
$W + \text{jets}$	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%

### 1766 4.5.1. Background normalization uncertainties

1767 The signal extraction is performed subtracting the estimated backgrounds to the event  
 1768 counts in data. This uncertainty depends on the background:

- 1769 •  **$t\bar{t}$  and  $tW$  backgrounds:** The efficiency on jets b-tagging is estimated using the  
 1770 Tag&Probe technique in data and simulation control regions, as explained in ???. A  
 1771 per-jet scale factor, which takes into account the possibly different efficiency of the  
 1772 anti b-tagging selection in data and simulation, is computed by means of the efficiency  
 1773 measured with the Tag&Probe method. The Tag&Probe method has been used also  
 1774 to measure the mistag rates in data and simulation, which are the probability to b-tag  
 1775 a jet that is not produced by the hadronization of a b quark. These factors are used  
 1776 to reweigh the Top MC samples as explained in 4.4.1. The uncertainties provided by  
 1777 the Tag&Probe fit are then propagated to the factor  $\alpha$  that is used in the top data  
 1778 driven estimation ???. These uncertainties are embedded in a systematic error that  
 1779 affects the shape of the Top background in each  $p_T^H$  bin.

1780 Provided that the simulated samples include both  $t\bar{t}$  and  $tW$  processes, a systematic  
 1781 uncertainty related to the  $tW/t\bar{t}$  fraction has been included. In fact, a relative variation  
 1782 of the contribution of these two processes could modify the shape of the MC sample,  
 1783 and is thus included as a shape uncertainty affecting the top quark background shape  
 1784 in each  $p_T^H$  bin in a correlated way.

- 1785 •  **$W+jets$  background:** It is estimated with data control sample as described in  
 1786 Sec.4.4.3. With  $19.4 \text{ fb}^{-1}$  at 8 TeV, the uncertainty receives similar contributions from  
 1787 statistics and systematic error (mainly jet composition differences between the fake  
 1788 rate estimation sample and the application sample), the total error being about 40%,  
 1789 dominated by the closure test of the method on a same-sign control region.
- 1790 •  **$WZ, ZZ, W\gamma^{(*)}$  backgrounds:** those backgrounds, which are expected to give a  
 1791 small contribution, are estimated from simulation. Uncertainties on the cross sections  
 1792 reported in [80, 81] are 4% for  $WZ$  and 2.5% for  $ZZ$ . A 30% uncertainty is assigned to  
 1793 the  $W\gamma$  [82] yield and another 30% on  $W\gamma^{(*)}$  contribution according to the uncertainty  
 1794 on the normalization study (see Sec. 4.4.4).

### 1795 4.5.2. Experimental uncertainties

1796 The following experimental systematic sources have been taken into account:

- 1797 • **Luminosity:** Using the online luminosity monitoring CMS reached an uncertainty  
 1798 on the luminosity of 2.6% at 8 TeV.
- 1799 • **Trigger efficiency.** The uncertainties for both electrons and muons are at 1-2% level,  
 1800 which is added together to the lepton efficiency uncertainty.
- 1801 • **Lepton reconstruction and identification efficiency:** The lepton reconstruction  
 1802 and identification efficiencies are measured with the Tag&Probe method in data. To  
 1803 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 [83]) and compute the variation of the selection efficiency.
- **b jets mistag modeling:** A fraction of signal events is rejected because erroneously identified as b jets by the b-tagging algorithms. The mistag rate, as measured with the Tag&Probe technique described in Sec. 4.4.1, comes with an uncertainty due to different modeling of the b-tagging performance in data and simulation.
- **Pileup multiplicity:** Some of the variables used in the analysis are affected by the average number of pileup interactions. The simulated events have been reweighted according the instantaneous luminosity measured on data. The error in the average number of pileup interactions measured in data and the simulation of the modeling and physics aspects of the pileup simulation gives an uncertainty of 5% on the distribution used in the reweighting procedure. This uncertainty is propagated through all the analysis, and the estimated uncertainty on the efficiency is 2%.

### 4.5.3. Theoretical uncertainties

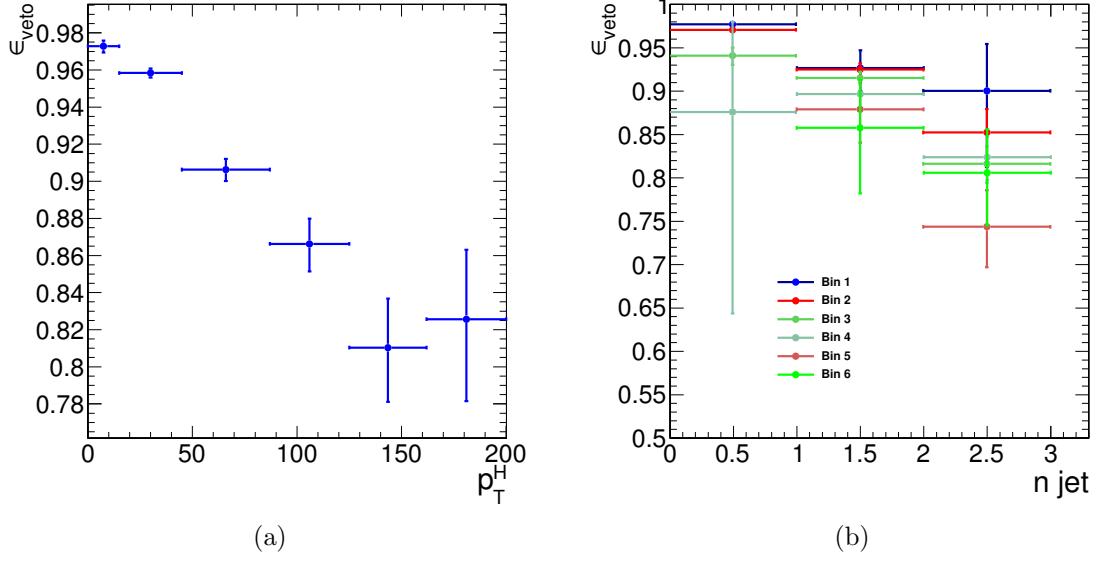
- **QCD scale uncertainties:** The uncertainties on the total cross sections due to the choice of the renormalization and factorization scale are assigned to MC-driven backgrounds. For the signal processes these uncertainties are separated in two categories: those affecting the selection efficiency and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

- **PDFs uncertainties:** The utilization of different PDF sets can affect both the normalization and the shapes of the signal contributions. The uncertainty related due to the variations in the choice of PDFs is considered following the PDF4LHC [84, 85] prescription, using CT10, NNPDF2.1 [86] and MSTW2008 [87] 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 4.4.2.

### 1857 Jet multiplicity uncertainty

1858 The jet bin uncertainty on the ggH production mode has been evaluated using the Stewart-  
 1859 Tackman method, following the recipe proposed in Refs. [53, 88]. Three independent  
 1860 nuisance parameters have to be associated with the inclusive ggH production cross sections  
 1861  $\sigma_{\geq 0}$ ,  $\sigma_{\geq 1}$  and  $\sigma_{\geq 2}$ , which corresponds to the cross sections with  $\geq 0$  jets,  $\geq 1$  jet and  $\geq 2$   
 1862 jets respectively. According to the agreement on the treatment of uncertainties in the  
 1863 combination of ATLAS and CMS results [89], these nuisance parameters are labelled as  
 1864  $QCDscale_{ggH}$ ,  $QCDscale_{ggH1in}$  and  $QCDscale_{ggH2in}$ . However, in case the analysis is  
 1865 split in exclusive jet multiplicity bins, the jet bin uncertainties can be evaluated taking into  
 1866 account the correct correlations among the three nuisances following the Stewart-Tackman  
 1867 prescription. Even though this analysis is inclusive in number of jets, the jet binning  
 1868 uncertainties must be included due to the presence of the b-jet veto, that introduces a  
 1869 dependency of the selection efficiency on the number of jets in the event. The veto efficiency  
 1870 has been evaluated in all the  $p_T^H$  bins defined in the analysis and as a function of jets  
 1871 multiplicity. The results are shown in Figs. 4.15(a) and 4.15(b). The drop of the veto  
 1872 efficiency at high values of  $p_T^H$  is due to the correlation with jets multiplicity.

1873 The first step of this procedure is to take the inclusive ggH cross section,  $\sigma_{ggH}$ , and  
 1874 to convert the relative QCD up/down scale uncertainties,  $\epsilon_+$  and  $\epsilon_-$ , to a log-normal  
 1875 uncertainty, i.e.  $\kappa = \sqrt{\exp(\epsilon_+) \cdot \exp(\epsilon_-)}$ . The exclusive cross sections,  $\sigma_0$ ,  $\sigma_1$  and  $\sigma_2$ , can  
 1876 be calculated starting from  $\sigma_{ggH}$  and using the selection efficiencies for the three jet bins.  
 1877 For every exclusive cross section the corresponding relative uncertainty is computed varying  
 1878 the renormalization ( $\mu_R$ ) and factorization ( $\mu_F$ ) scales independently of a factor 2 and  
 1879 1/2, and taking the cross section value corresponding to half of the maximum variation.  
 1880 The inclusive cross sections are then obtained summing the exclusive cross sections and  
 1881 propagating the uncertainties, i.e.  $\sigma_{\geq 0} = \sigma_0 + \sigma_1 + \sigma_2$ ,  $\sigma_{\geq 1} = \sigma_1 + \sigma_2$ ,  $\sigma_{\geq 2} = \sigma_2$ .



**Figure 4.15.:** (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.

1882      The three nuisance parameters, including all the proper correlations among the jet  
 1883      bins, are defined according to Table 4.5, where the  $f_n$  constants represent the exclusive  
 1884      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 4.5.:** 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		$\Delta_{\geq 2}^1 = (\kappa_{\geq 2})^{-\frac{f_2}{f_1}}$	$\Delta_{\geq 2}^2 = (\kappa_{\geq 2})$

1885 The nuisance parameters reported in table 4.5 have then been calculated for each  $p_T^H$   
1886 bin embedding the b-jet veto efficiency and using the following formulas:

$$QCDscale_{ggH} = \frac{\Delta_{\geq 0}^0 \cdot f_0 \cdot \varepsilon_0 + \Delta_{\geq 1}^0 \cdot f_1 \cdot \varepsilon_1}{\Delta_{\geq 0}^0 \cdot f_0 \cdot \varepsilon_0 + \Delta_{\geq 1}^0 \cdot f_1 \cdot \varepsilon_0} , \quad (4.11)$$

1887

$$QCDscale_{ggH1in} = \frac{\Delta_{\geq 1}^1 \cdot f_1 \cdot \varepsilon_1 + \Delta_{\geq 2}^1 \cdot f_2 \cdot \varepsilon_2}{\Delta_{\geq 1}^1 \cdot f_1 \cdot \varepsilon_1 + \Delta_{\geq 2}^1 \cdot f_2 \cdot \varepsilon_1} , \quad (4.12)$$

1888

$$QCDscale_{ggH2in} = 1 , \quad (4.13)$$

1889 where  $\varepsilon_0$ ,  $\varepsilon_1$  and  $\varepsilon_2$  are the selection efficiencies for the three jet categories. These nuisance  
1890 parameters are expected to be equal to one in case the efficiency is independent on the  
1891 number of jets, i.e if  $\varepsilon_0 = \varepsilon_1 = \varepsilon_2$ .

1892 The numerical values obtained following this procedure are reported in Table 4.6 for each  
1893  $p_T^H$  bin.

**Table 4.6.:** 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- $\infty$ ]
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

#### 1894 4.5.4. Statistics uncertainty of the simulated samples

1895 Due to the large range of weights used to correct the simulated distributions in order to  
1896 match those in data, the effective size of the MC samples are sometimes smaller than the  
1897 actual number of events in the sample. The statistical uncertainties of the event yields  
1898 estimated from MC samples are included as nuisance parameters in the fit and have a small  
1899 impact on the final result.

#### 1900 4.5.5. Treatment of systematic uncertainties in the shape 1901 analysis

1902 One can distinguish between normalization uncertainties, where a systematic effect is  
1903 changing the normalization of a given process assuming the shape is not affected, and shape  
1904 uncertainties where the actual change in the shape of the distribution is taken into account.

1905 The normalization uncertainties enter the shape analysis as a constant normalization factor,  
 1906 whereas for shape uncertainties the nominal and the  $+1\sigma$  and  $-1\sigma$  shapes enter the analysis  
 1907 in form of three histograms with the same normalization.

1908 For the W+jets background, the shape differences for different jet  $p_T$  thresholds in the  
 1909 di-jet control sample are considered separately for electron and muon fakes, while the other  
 1910 sources of systematics are taken as normalization uncertainties as in the cut-based analysis.

1911 Effects from experimental uncertainties are studied by applying a scaling and smearing  
 1912 of certain variables of the physics objects, followed by a subsequent recalculation of all the  
 1913 correlated variables. This is done for simulation, to account for possible systematic mis-  
 1914 measurements of the data. All experimental sources from Section 4.5.2 but luminosity are  
 1915 treated both as normalization and shape uncertainties. For background with a data-driven  
 1916 normalization estimation, only the shape uncertainty is considered.

1917 To account for statistical uncertainties, for each distribution going into the shape analysis,  
 1918 the  $+1\sigma$  and  $-1\sigma$  shapes were obtained by adding/subtracting the statistical error in each  
 1919 bin and renormalizing it to the nominal distribution. In addition to this procedure a  
 1920 constant normalization uncertainty due to the finite statistics of the MC sample used to  
 1921 extract the shape is assigned.

## 1922 4.6. Signal extraction

1923 According to the “blinding” policy of the CMS Collaboration, the strategy of the analysis  
 1924 has been scrutinized and approved by a selected committee of internal reviewers before  
 1925 looking at the data in the signal region. This approach prevents the analysts from being  
 1926 biased by the data in the developing phase of the analysis. Below are shown the results  
 1927 after having looked at the data.

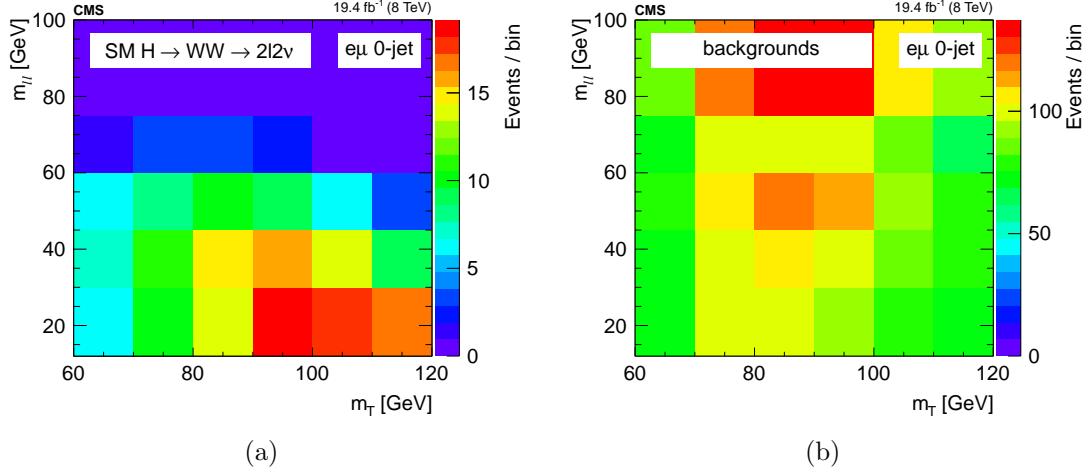
### 1928 4.6.1. Fitting procedure

1929 The signal, including ggH, VBF, and VH production mechanisms, is extracted in each bin  
 1930 of  $p_T^H$  by performing a binned maximum likelihood fit simultaneously in all  $p_T^H$  bins to a  
 1931 two-dimensional template for signals and backgrounds in the  $m_{\ell\ell}$ – $m_T$  plane. The variables  
 1932 used for the two-dimensional template are chosen for their power to discriminate signal  
 1933 and background contributions. This is shown in Fig. 4.16, where the two-dimensional MC  
 1934 distributions are shown for the signal and background processes in the 0-jets category.

1935 Six different signal strength parameters are extracted from the fit, one for each  $p_T^H$  bin.  
 1936 The relative contributions of the different Higgs production mechanisms in the signal  
 1937 template are taken to be the same as in the SM. The systematic uncertainty sources are  
 1938 considered as nuisance parameters in the fit.

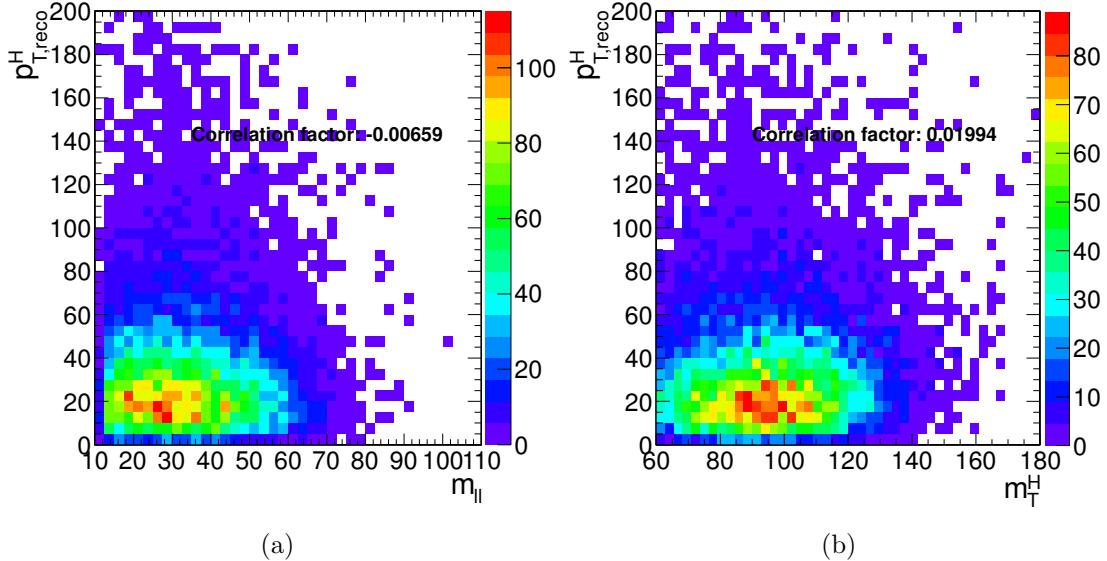
1939 The binning of the  $m_{\ell\ell}$  and  $m_T$  templates is chosen to be:

- 1940 •  $m_{\ell\ell}$ : [12, 30, 45, 60, 75, 100, 125, 150, 175, 200]
- 1941 •  $m_T$ : [60, 70, 80, 90, 100, 110, 120, 140, 160, 180, 200, 220, 240, 280]

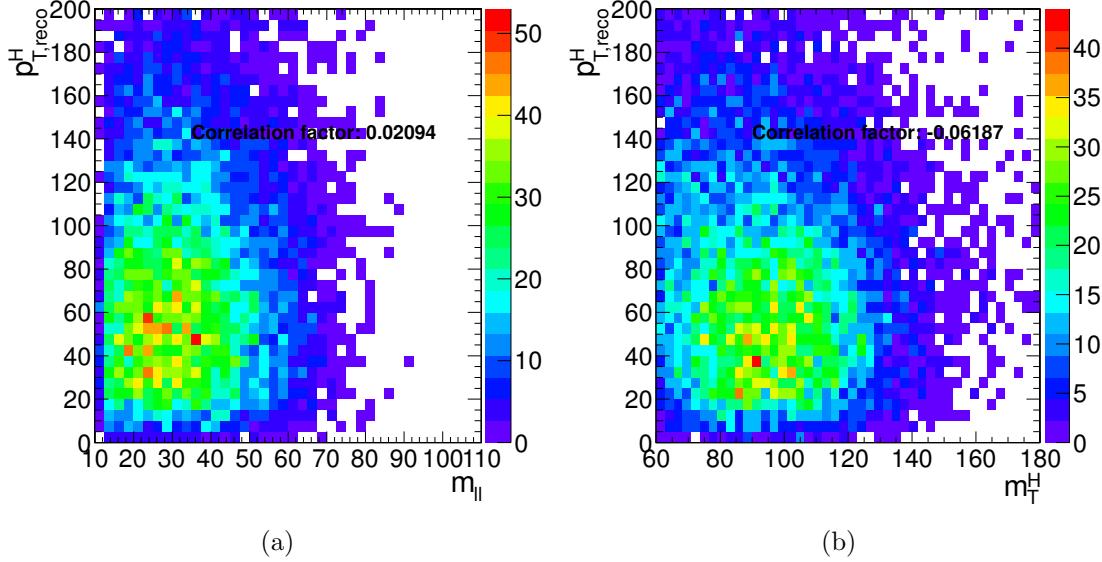


**Figure 4.16.:** Two-dimensional  $m_{\ell\ell}$ – $m_T$  distribution for signal (a) and background (b) processes in the 0-jets category.

1942 To avoid a dependence of the results on the variables used for the template fit,  $m_{\ell\ell}$  and  
 1943  $m_T$  need to be uncorrelated with respect to  $p_T^H$ . This has been verified and the correlation  
 1944 between the discriminating variables and  $p_T^H$  is shown in Fig. 4.17 and Fig. 4.18 for ggH  
 1945 and VBF production modes respectively.



**Figure 4.17.:** 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 4.18.:** 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.

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

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

$$N_i = \mu_i(s_i - f_i) \quad , \quad (4.14)$$

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!!**

$$\mathcal{L}(data|\mu_j, \theta) = \prod_{i=0}^{N_{bins}} \frac{(\mu_j s_i(\theta) + b_i(\theta))^{n_i}}{n_i!} e^{-\mu_j s_i(\theta) - b_i(\theta)} \cdot p(\tilde{\theta}|\theta) \quad , \quad (4.15)$$

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. 4.5.5. 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>2</sup>, which provides a simple method to estimate the signal sensitivity before looking at the data [90].

### 4.6.2. Signal and background yields

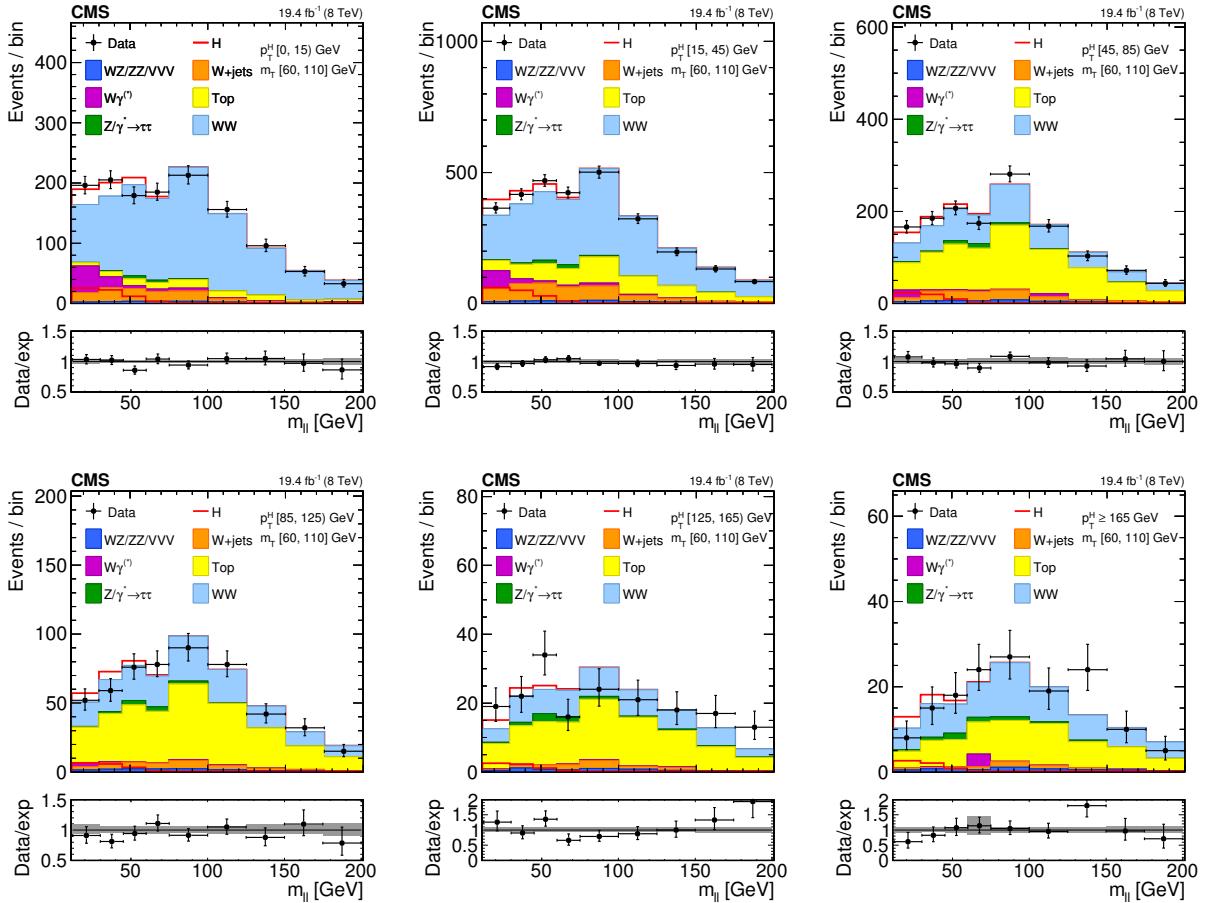
A comparison of data and background prediction is shown in Fig. 4.19, where the  $m_{\ell\ell}$  distribution is shown for the six  $p_T^H$  bins. Distributions correspond to the  $m_T$  window of [60, 110] GeV, in order to emphasize the signal contribution [54]. The  $m_T$  distributions are shown in Fig. 4.20 and correspond to the  $m_{\ell\ell}$  window of [12, 75] GeV.

The signal and background yields after the analysis selection are reported in Table 4.7.

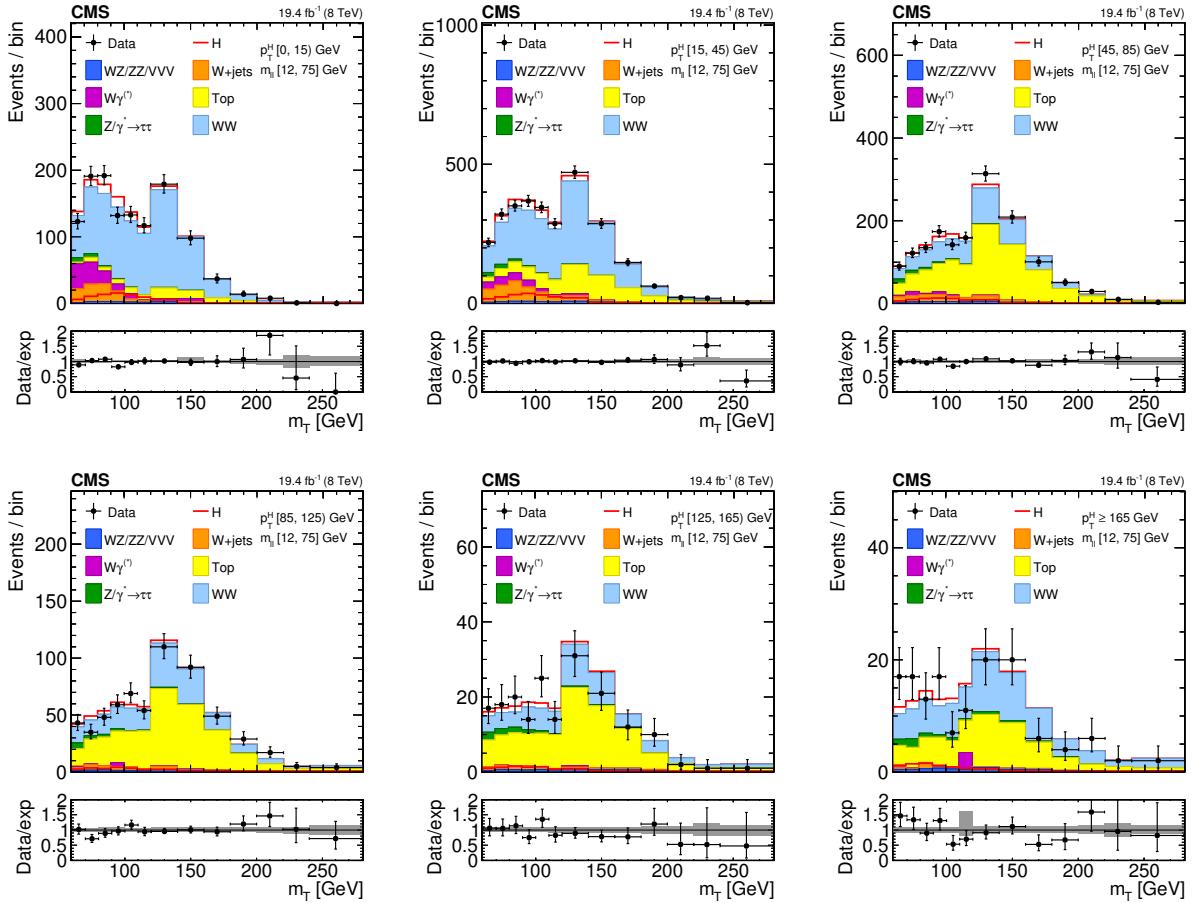
The spectrum shown in Fig. 4.21 is obtained after having performed the fit and after the subtraction of the out-of-fiducial signal events, but before undergoing the unfolding procedure. The theoretical distribution after the detector simulation and event reconstruction is also shown for comparison.

In order to assess the robustness of the fit, several toy MC samples have been produced, with a statistical accuracy corresponding to the one expected in data. The distribution

<sup>2</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.



**Figure 4.19.:** 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.



**Figure 4.20.:** 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 4.7.:** 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.

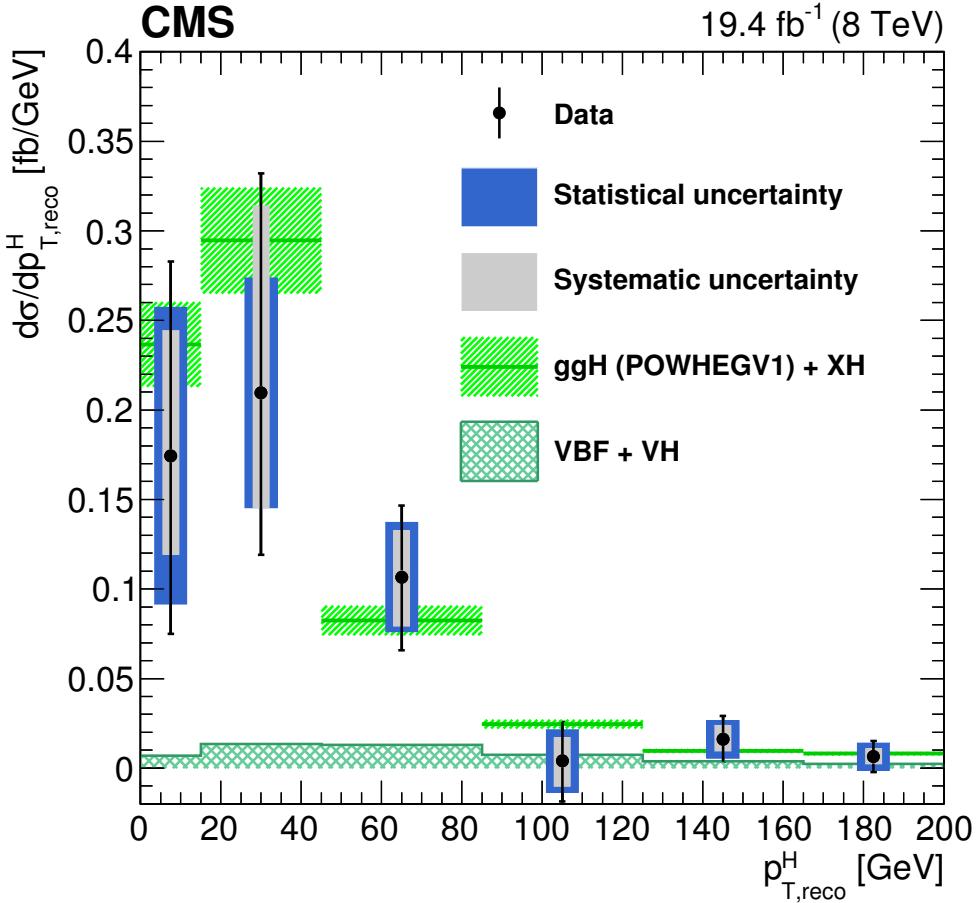
$p_T^H$ [GeV]	0-15	15-45	45-85	85-125	125-165	165- $\infty$
ggH	$73 \pm 3$	$175 \pm 5$	$59 \pm 3$	$15 \pm 2$	$5.1 \pm 1.5$	$4.9 \pm 1.4$
XH=VBF+VH	$4 \pm 2$	$15 \pm 4$	$16 \pm 4$	$8 \pm 2$	$3.8 \pm 1.1$	$3.0 \pm 0.8$
Out-of-fiducial	$9.2 \pm 0.5$	$19.9 \pm 0.7$	$11.4 \pm 0.6$	$4.4 \pm 0.3$	$1.6 \pm 0.2$	$2.4 \pm 0.2$
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>1997</sup> of the signal strengths extracted in each bin using the toy MC samples and the their pull  
<sup>1998</sup> distributions are shown in Fig. 4.22.

## <sup>1999</sup> 4.7. Unfolding

<sup>2000</sup> To facilitate comparisons with theoretical predictions or other experimental results, the  
<sup>2001</sup> signal extracted performing the fit has to be corrected for detector resolution and efficiency  
<sup>2002</sup> effects and for the efficiency of the selection defined in the analysis. An unfolding procedure  
<sup>2003</sup> is used relying on the ROOUNFOLD package [91], which provides the tools to run various  
<sup>2004</sup> unfolding algorithms.

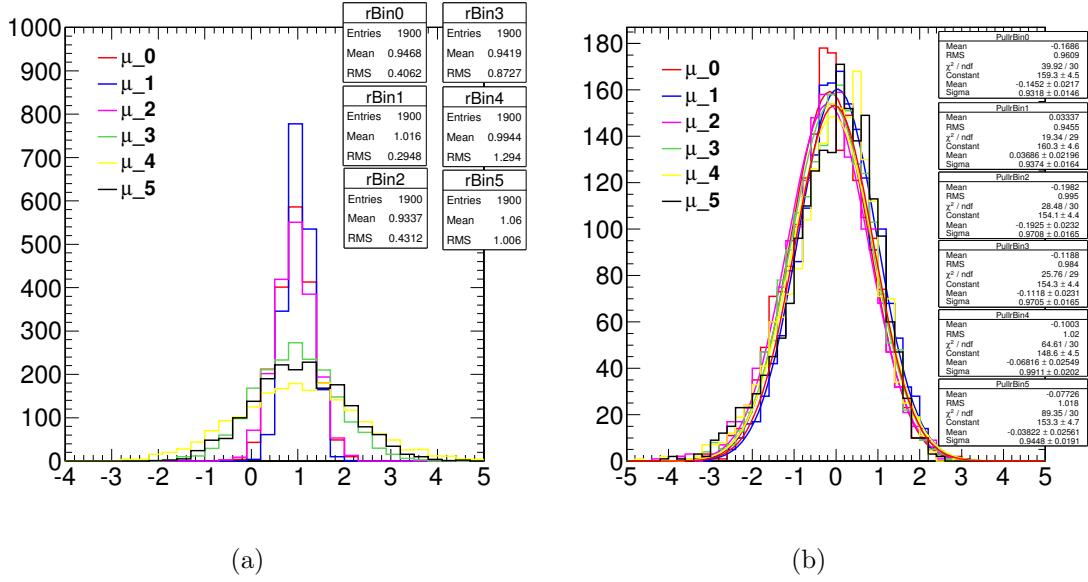
<sup>2005</sup> The basic principle behind the unfolding procedure in this analysis is to use MC signal  
<sup>2006</sup> samples to make the “true” distribution of the variable of interest, which is obtained using  
<sup>2007</sup> simulated events before particle interaction with the detector, and the same distribution  
<sup>2008</sup> obtained using events reconstructed after the full GEANT4 simulation of the CMS detector  
<sup>2009</sup> and event reconstruction. These two distributions are used to calculate the detector response  
<sup>2010</sup> matrix  $M$ :



**Figure 4.21.:** 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.

$$R_i^{\text{MC}} = \sum_{j=1}^n M_{ij} T_j^{\text{MC}} \quad , \quad (4.16)$$

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 4.22.:** 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 [55], the unfolding procedure in this analysis relies on the singular value decomposition [92] 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 [55]. 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 [93]. 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

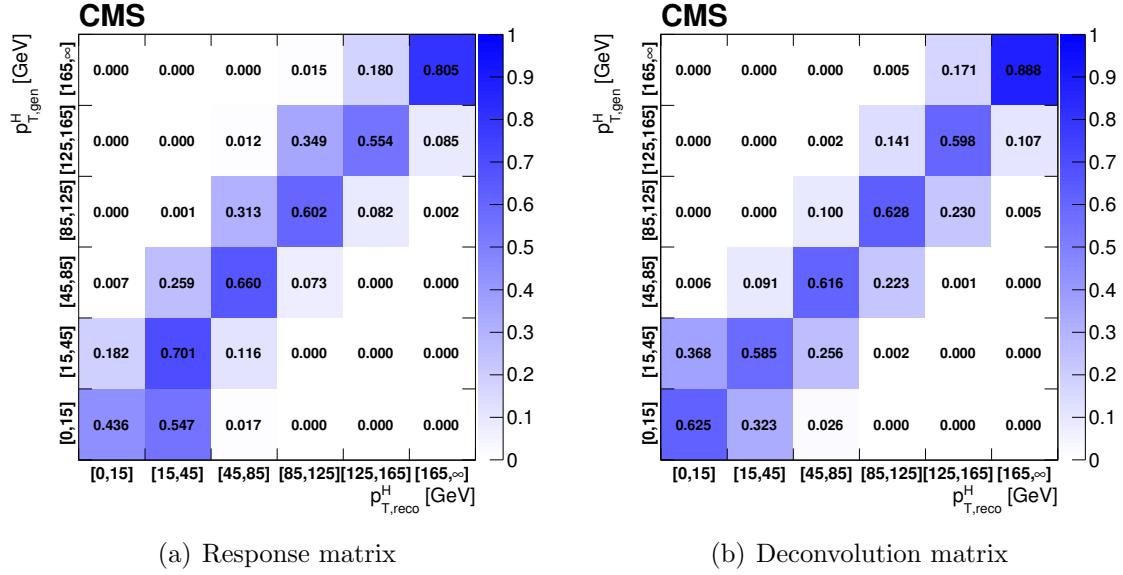
described in Ref. [92]. **Maybe I should add an appendix describing the SVD method in details**

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. 4.23(a). The value of the diagonal bins corresponds to the stability  $S$ . The same matrix, normalized by column, is shown in Fig. 4.23(b). In this case the diagonal bins correspond to the purity  $P$ . The  $S$  and  $P$  parameters, defined in Sec. 4.3, provide an estimate of the  $p_T^H$  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. 4.23(a). This representation shows the stability  $S$  in the diagonal bins, where  $S$  is defined as the ratio of the number of events generated and reconstructed in a given bin, and the number of events generated in that bin. In addition, a deconvolution matrix is constructed by normalizing each column to unity and is shown in Fig. 4.23(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. 4.23. 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 4.8 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 4.23.:** 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 4.8.:** 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– $\infty$	$0.632^{+0.016}_{-0.016}$	$0.674^{+0.015}_{-0.015}$	$0.701^{+0.015}_{-0.015}$	$0.722^{+0.014}_{-0.015}$

without introducing a bias, i.e. until a 68% coverage is fulfilled. This criterion leads to the same result as the prescription described in Ref. [92], strengthening the choice of  $k_{\text{reg}} = 3$ .

### 4.7.1. Treatment of systematic uncertainties

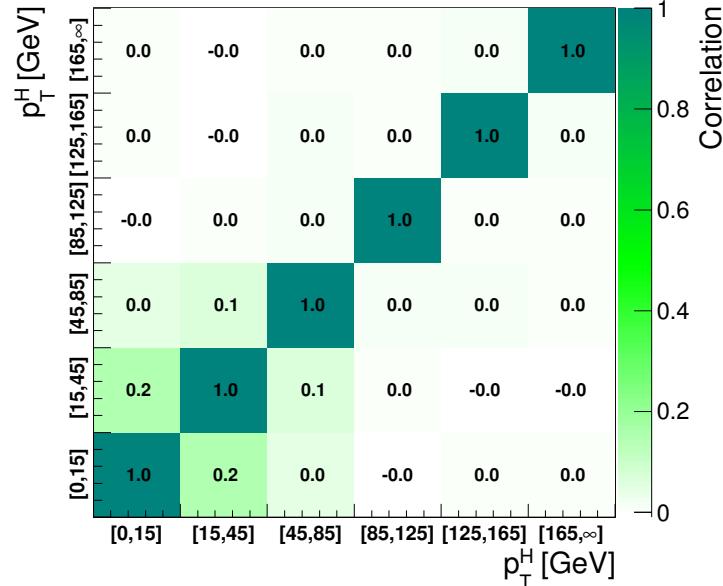
An important aspect of this analysis is the treatment of the systematic uncertainties and the error propagation through the unfolding procedure. The sources of uncertainty are

2086 divided into three categories, depending on whether the uncertainty affects only the signal  
 2087 yield (type A), both the signal yield and the response matrix (type B), or only the response  
 2088 matrix (type C). These three classes propagate differently through the unfolding procedure.

2089 Type A uncertainties are extracted directly from the fit in the form of a covariance  
 2090 matrix, which is passed to the unfolding tool as the covariance matrix of the measured  
 2091 distribution. The nuisance parameters belonging to this category are the background shape  
 2092 and normalization uncertainties. To extract the effect of type A uncertainties a dedicated  
 2093 fit is performed, fixing to constant all the nuisance parameters in the model, but type A  
 2094 nuisance parameters. The correlation matrix among the six signal strengths corresponding  
 2095 to the six  $p_T^H$  bins, including all type A uncertainties, is shown in Fig. 4.24. The correlation  
 2096  $\text{cor}(i,j)$  of bins  $i$  and  $j$  is defined as:

$$\text{cor}(i,j) = \frac{\text{cov}(i,j)}{s_i s_j} \quad , \quad (4.17)$$

2097 where  $\text{cov}(i,j)$  is the covariance of bins  $i$  and  $j$ , and  $(s_i, s_j)$  are the standard deviations  
 2098 of bins  $i$  and  $j$ , respectively.



**Figure 4.24.:** Correlations among the signal strengths corresponding to the six  $p_T^H$  bins including all type A uncertainties.

2099 The nuisance parameters falling in the type B class are:

- 2100 • the b veto scale factor. It affects the signal and background templates by varying the  
 2101 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. 4.9. 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 [94]. Three different response matrices are built, corresponding to the nominal, scaled up, and scaled down VBF/ggH ratio. The nominal matrix assumes the SM VBF/ggH ratio, while up- and down-scaled matrices are constructed by varying the SM signal strengths within the experimental constraints for VBF and ggH in such a way as to obtain the maximal variation of the VBF/ggH ratio allowed by the experimental constraints. These three matrices are used to unfold the reconstructed spectrum with the nominal VBF/ggH fraction, and obtain an uncertainty on the unfolded spectrum.

## 4.8. Results

In order to unfold the spectrum, the procedure described in section 4.7 has been pursued. The statistical plus type A systematic uncertainties are propagated by the unfolding procedure into the final spectrum, taking into account the signal strengths covariance matrix. The type B systematic uncertainty has been propagated using the following procedure: for each  $p_T^H$  bin, we compute the upper bound of the systematic band computing

**Table 4.9.:** 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.

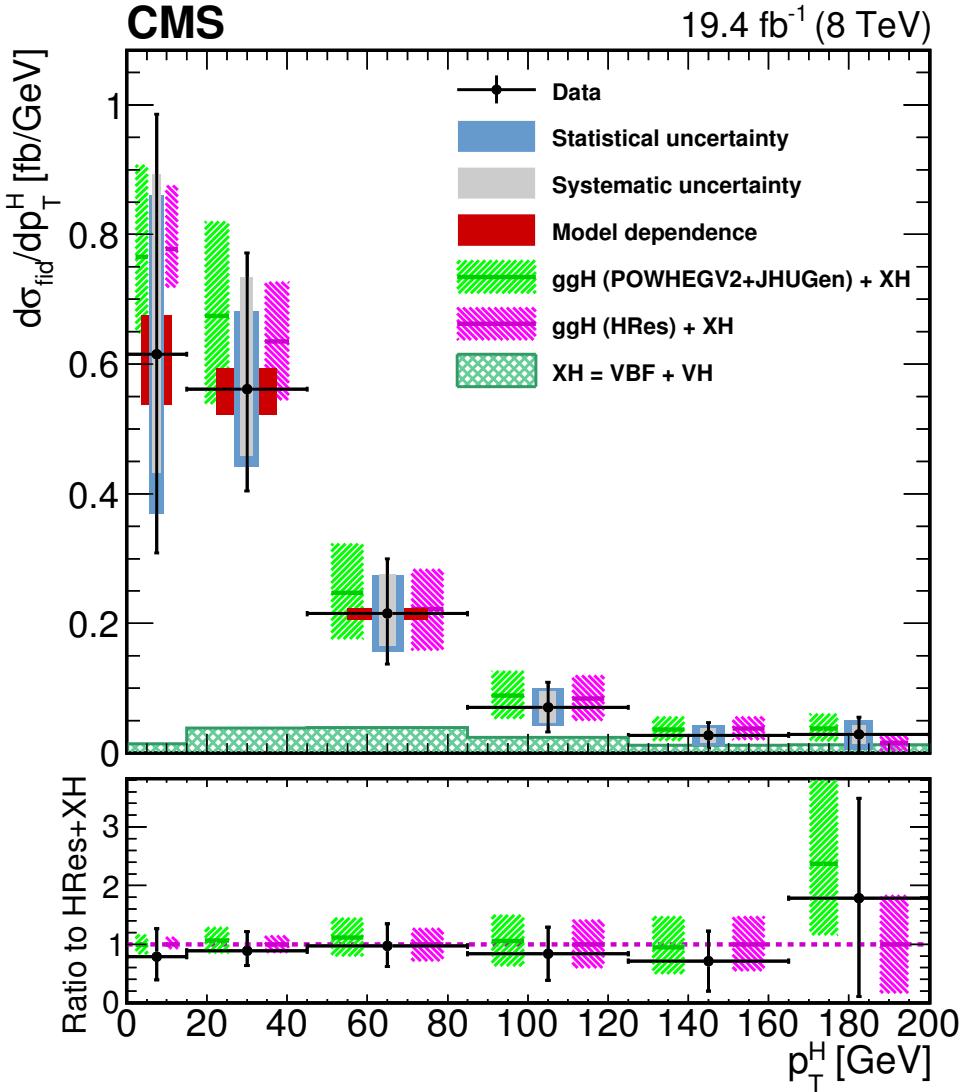
Type B uncertainty	Effect on signal strength ( $+1\sigma/-1\sigma$ [%])					
	[0–15]	[15–45]	[45–85]	[85–125]	[125–165]	[165– $\infty$ ]
b veto	-10.1/-8.8	7.3/12.2	-6.3/3.1	-14.4/-4.8	-5.4/14.5	-7.9/17.8
lepton efficiency	-14.7/-3.9	4.5/15.1	-5.7/2.5	-13.2/-5.3	-0.2/7.6	-0.1/6.8
$E_T^{\text{miss}}$ resolution	-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 4.10.:** 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

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

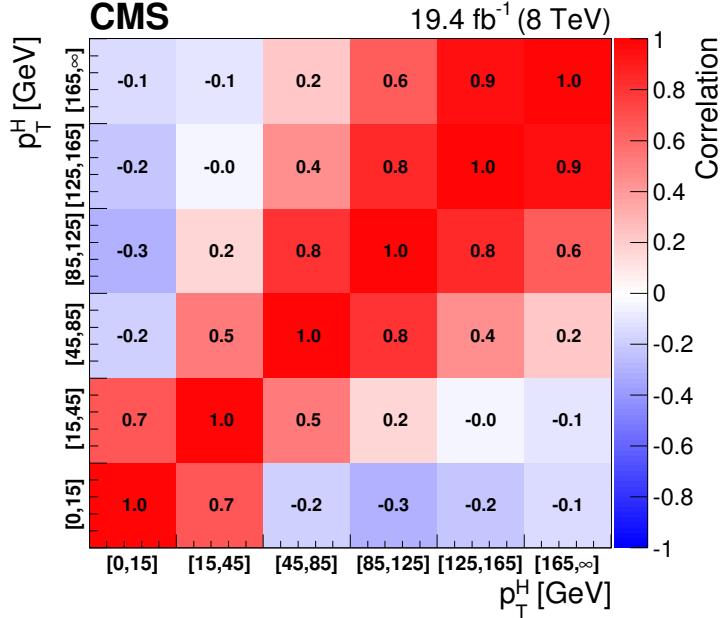
The unfolded  $p_T^H$  spectrum is shown in Fig. 4.25. Statistical, systematic, and theoretical uncertainties are shown as separate error bands in the plot. The unfolded spectrum is compared with the SM-based theoretical predictions where the ggH contribution is modelled using the HRES and POWHEG V2 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 4.10.



**Figure 4.25.:** Higgs boson production cross section as a function of  $p_T^H$ , after applying the unfolding procedure. Data points are shown, together with statistical and systematic uncertainties. The vertical bars on the data points correspond to the sum in quadrature of the statistical and systematic uncertainties. The model dependence uncertainty is also shown. The pink (and back-sashed filling) and green (and slashed filling) lines and areas represent the SM theoretical estimates in which the acceptance of the dominant ggH contribution is modelled by HRes and POWHEG V2, respectively. The subdominant component of the signal is denoted as  $\text{XH}=\text{VBF}+\text{VH}$  and it is shown with the cross filled area separately. The bottom panel shows the ratio of data and POWHEG V2 theoretical estimate to the HRes theoretical prediction.

Figure 4.26 shows the correlation matrix for the six bins of the differential spectrum.  
2152

The correlation of of bins is defined as in Eq. (4.17).  
2153



**Figure 4.26.:** 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  $\epsilon_{fid}$ . Dividing the measured number of events by the integrated luminosity and correcting for the overall selection efficiency, which is estimated in simulation to be  $\epsilon_{fid} = 36.2\%$ , the inclusive fiducial  $\sigma \times \mathcal{B}$ ,  $\sigma_{fid}$ , is computed to be:

$$\sigma_{fid} = 39 \pm 8 \text{ (stat)} \pm 9 \text{ (syst)} \text{ fb} \quad , \quad (4.18)$$

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.

## 2166 **Search for the SM Higgs boson in the 2167 $H \rightarrow WW$ channel with the first 2168 13 TeV LHC data**

### 2169 **5.1. Introduction**

2170 In this chapter, the first search for the SM Higgs boson decaying to a W boson pair at  
2171 13 TeV is presented, using a total integrated luminosity of  $2.3 \text{ fb}^{-1}$ , collected during the  
2172 2015 proton proton data taking period of the LHC.

2173 Final states in which the two W bosons decay leptonically are studied. Therefore,  
2174 events with a pair of oppositely-charged leptons, exactly one electron and one muon, a  
2175 substantial amount of missing transverse energy,  $E_T^{\text{miss}}$ , due to the presence of neutrinos in  
2176 the final state, and either zero or one jet are selected. This signature is common to other  
2177 processes, which enter the analysis as backgrounds. The main background comes from  
2178 WW production, irreducible background that shares the same final states and can only be  
2179 separated by the use of certain kinematic properties. Another important background is  
2180 W+jets, where a jet can mimick a leptonic signature. Background coming from top quark  
2181 events, i.e.  $t\bar{t}$  and single top production, is also important, followed by other processes such  
2182 as Drell-Yan, WZ, and other EWK production. The analysis strategy follows the one used  
2183 during Run 1 in the same channel, described in Chapter 4, with a few different aspects that  
2184 are described in the next sections.

2185 With respect to 8 TeV, the ggH production cross section at 13 TeV is expected to  
2186 increase of a factor of 2, thus raising the number of expected signal events. In addition, the  
2187 cross section for the background processes is increasing as well. The WW production cross  
2188 section increases of a factor of 1.8 and the  $t\bar{t}$  cross section of a factor of 3.5, due to the  
2189 enhancement of the gluon PDFs at higher center of mass energies.

### 2190 **5.2. Data and simulated samples**

2191 Data recorded in proton proton collisions at 13 TeV during 2015 was used in the analysis,  
2192 with a total integrated luminosity of  $2.3 \text{ fb}^{-1}$ . Single and double lepton triggers are used

2193 similarly to the same analysis at 8 TeV. The HLT paths and descriptions of the triggers used  
2194 in this analysis are described in Tables 5.1 and 5.2 for electrons and muons respectively.

**Table 5.1.:** HLT paths related to Electrons

HLT Path	Description
HLT_Ele23_WPLoose_Gsf_v*	Single Electron trigger. Best trigger to be used for 2015 data. In $WW$ , we are using “Trigger safe” Id. Turn on is at around $\text{Ele } p_T = 30 \text{ GeV}$
HLT_Ele17_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*	Double Electron Trigger. Best trigger to cover the turn on region from single electron trigger. “DZ” filter is also present. Its efficiency is also calculated separately.
HLT_Ele12_CaloIdL_TrackIdL_IsoVL_v*	This electron leg of $\text{HLT\_Mu17\_TrkIsoVVL\_Ele12\_CaloIdL\_TrackIdL\_IsoVL\_v^*}$ same as Ele12 leg of double electron trigger.
HLT_Ele17_CaloIdL_TrackIdL_IsoVL_v*	This electron leg of $\text{HLT\_Mu8\_TrkIsoVVL\_Ele17\_CaloIdL\_TrackIdL\_IsoVL\_v^*}$ same as Ele17 leg of double electron trigger.

2195 The trigger efficiencies are measured in data and applied on simulated events as described  
2196 in Sec. 4.2.1.

2197 Concerning the simulated samples, several different Monte Carlo (MC) generators were  
2198 used. In the simulation, ‘lepton’ includes also  $\tau$ . Higgs signal samples have been simulated  
2199 in all channels with POWHEG v2 [57, 70, 95], designed to describe the full NLO properties  
2200 of these processes. In particular, for Higgs produced via gluon fusion [59], and vector-boson-  
2201 fusion (VBF) [60], the decay of the Higgs boson into two W boson and subsequently into  
2202 leptons was done using JHUGEN v5.2.5 [96]. For associated production with a vector boson  
2203 ( $W^+H$ ,  $W^-H$ ,  $ZH$ ) [97], including gluon fusion produced  $ZH$  ( $ggZH$ ), the Higgs decay was  
2204 done via PYTHIA 8.1 [78]. Alternative signal samples were produced with AMC@NLO [62],  
2205 or with POWHEG v2 but decayed via PYTHIA 8.1 for gluon fusion and VBF assuming a  
2206 Higgs boson mass of 125 GeV. In the following, the mass of the SM Higgs boson is assumed  
2207 to be 125 GeV.

2208 The  $WW$  production, irreducible background for the analysis, was simulated in different  
2209 ways. POWHEG v2 [98] was used for  $q\bar{q}$  produced  $WW$  in different decays. The cross  
2210 section used for normalizing  $WW$  processes produced via  $q\bar{q}$  was computed at next-to-next-  
2211 to-leading order (NNLO) [99]. In order to control the top quark background processes, the  
2212 analysis is performed with events that have no more than one high- $p_T$  jet. The veto on  
2213 high- $p_T$  jets enhances the importance of logarithms of the jet  $p_T$ , spoiling the convergence  
2214 of fixed-order calculations of the  $q\bar{q} \rightarrow WW$  process and requiring the use of dedicated  
2215 resummation techniques for an accurate prediction of differential distributions [100, 101].  
2216 The  $p_T$  of the jets produced in association with the  $WW$  system is strongly correlated with  
2217 its 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%

2218 simulated  $q\bar{q} \rightarrow WW$  events are reweighted to reproduce the  $p_T^{WW}$  distribution from the  
 2219  $p_T$ -resummed calculation.

2220 Gluon fusion produced  $WW$  was generated, with and without Higgs diagrams, using  
 2221 MCFM v7.0 [102]. A  $t\bar{t}$  sample dilepton sample was also generated using POWHEG v2. The  
 2222  $WW$  and  $t\bar{t}$  samples produced specifically for this analysis are presented in Table 5.3. Other  
 2223 background samples are used, a list of the most relevant ones is presented in Table 5.4.

**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

2224 All processes are generated using the NNPDF2.3 [103, 104] parton distribution functions  
 2225 (PDF) for NLO generators, while the LO version of the same PDF is used for LO generators.  
 2226 All the event generators are interfaced to PYTHIA 8.1 [78] 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

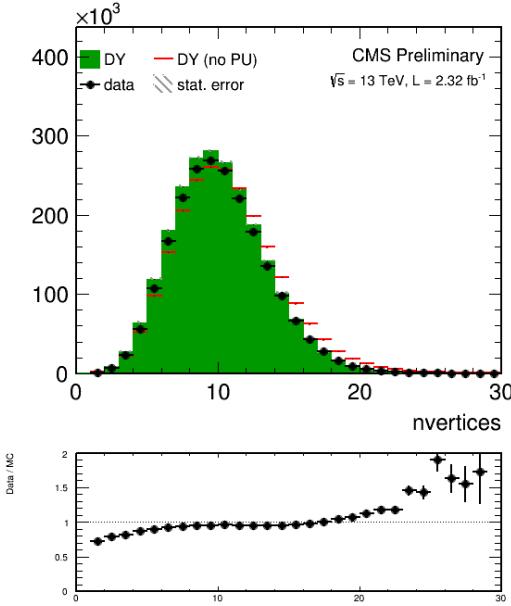
and hadronization, as well as including a simulation of the underlying event (UE) and multiple interaction (MPI) based on the CUET8PM1 tune [105]. To estimate the systematic uncertainties related to the choice of UE and MPI tune, the signal processes and the  $WW$  events are also generated with two 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 [106, 107]. For all processes, the detector response is simulated using a detailed description of the CMS detector, based on the GEANT4 package [71].

The simulated samples are generated with distributions for the number of pileup interactions that are meant to roughly cover, though not exactly match, the conditions expected for the different data-taking periods. In order to factorize these effects, the number of true pileup interactions from the simulation truth is reweighted to match the data. The re-weighting is propagated automatically to both the in-time pile up and the out-of-time one. In Fig. 5.1, the effect of this reweighting on a sample enriched in Drell-Yan events is shown. Before the reweighting the simulation is presented in the open red histogram; after the reweighting, it is represented by the solid green histogram that matched well the data. In order to select this sample, events with two leptons with  $p_T > 20 \text{ GeV}$ , 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 [108], 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. [53].

The cross section used for normalizing  $q\bar{q}$  produced  $WW$  processes was computed at next-to-next-to-leading order (NNLO) [99]. 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 [109].



**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.

For the LO simulation of the interference between  $gg \rightarrow WW$  and gluon fusion produced  $H \rightarrow WW$  a k-factor of 1.87 is applied. This k-factor is obtained as the average between LO to NNLO ggH scale factor and LO to NLO ggWW scale factor.

The cross sections of the different single top processes are estimated by the LHC Top Working group [110] at NLO. The  $t\bar{t}$  cross section is also provided by the LHC Top Working group [111], and it is computed at NNLO, with NNLL soft gluon resummation.

Drell-Yan (DY) production of  $Z/\gamma^*$  is generated using AMC@NLO [62]. 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 [103, 104] 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 [105].

## 2269 5.3. Analysis strategy

### 2270 5.3.1. Event reconstruction

2271 Regarding the electrons, muons, jets and  $E_T^{\text{miss}}$  definition and reconstruction, the standard  
 2272 CMS recommendations described in Chapter 2 are used. The specific selections used in  
 2273 this analysis are briefly summarised below.

2274 Muons are identified according to the CMS recommendations for the medium working  
 2275 point, with the addition of some extra cuts, as defined by the following selections:

- 2276 • identified by the standard medium muon selection described in Sec. ??; Not yet defined  
 2277 :)
- 2278 •  $p_T > 10 \text{ GeV}$ ;
- 2279 •  $|\eta| < 2.4$ ;
- 2280 •  $|d_{xy}| < 0.01 \text{ cm}$  for  $p_T < 20 \text{ GeV}$  and  $|d_{xy}| < 0.02 \text{ cm}$  for  $p_T > 20 \text{ GeV}$ ,  $d_{xy}$  being the  
 2281 transverse impact parameter with respect to the primary vertex;
- 2282 •  $|d_z| < 0.1 \text{ cm}$ , where  $d_z$  is the longitudinal distance of the muon track in the tracker  
 2283 extrapolated along the beam direction.

2284 For the muon isolation, the CMS recommended particle flow isolation based on the  
 2285 tight working point is used, corresponding to a requirement on the isolation variable of  
 2286  $ISO_{\text{tight}} < 0.15$ . In addition a tracker relative isolation is also applied.

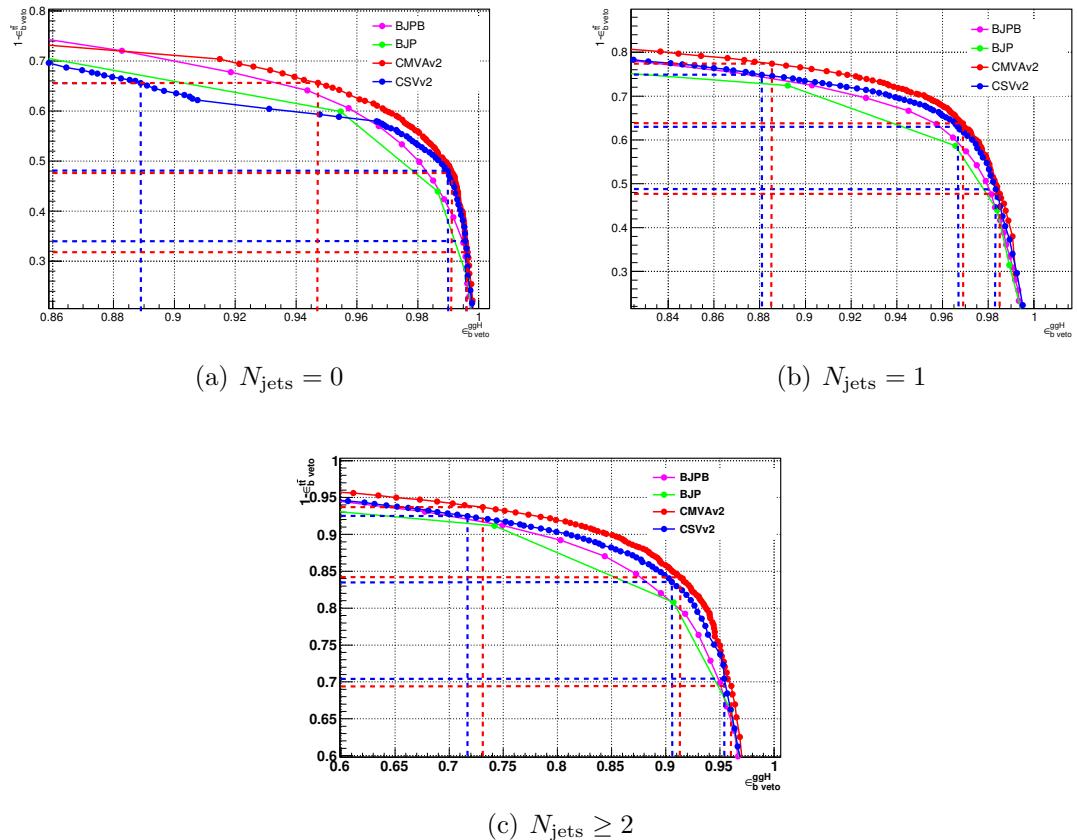
2287 For the electron identification, the tight working point is used. In addition some  
 2288 additional cuts to make the selection “trigger-safe” are included. This is done because the  
 2289 electron triggers already include some identification and isolation requirements that are  
 2290 based on the raw detector information, while the offline selections make use of particle flow  
 2291 requirements. The “trigger-safe” selections are defined to make the the offline identification  
 2292 and isolation requirements tighter with respect to the online triggers.

2293 The simulated events are corrected for the lepton trigger, identification and isolation  
 2294 efficiencies measured in data using the same techniques described in Sec. 4.3.1.

2295 Jets are defined clustering the particle flow objects using the anti- $k_t$  algorithm with a  
 2296 distance parameter of 0.4. The CHS pileup mitigation technique is used. The L1, L2, L3  
 2297 and L2L3 jet energy correction described in Sec. ?? are applied. The reject jets coming from  
 2298 calorimeter or readout electronics noise, the loose working point for PF jet identification is  
 2299 used.

2300 The b-tagging algorithm for this analysis is chosen comparing the performances of  
 2301 different algorithms using simulations for signal and background contributions in the phase  
 2302 space defined by the analysis kinematic requirements. More precisely, two MC samples are  
 2303 used, one corresponding the the  $H \rightarrow WW \rightarrow 2\ell 2\nu$  signal produced via the ggH production  
 2304 mode and another corresponding to the  $t\bar{t}$  process. In fact, the first sample is enriched  
 2305 in light jets, i.e. originating by the hadronization of light quarks like u,d,c and s quarks,  
 2306 while the second sample is enriched in b jets, coming from the top quark decay. The b-veto  
 2307 efficiency,  $\epsilon_{\text{bveto}}$ , is computed separately for the two samples and for the various b tagging

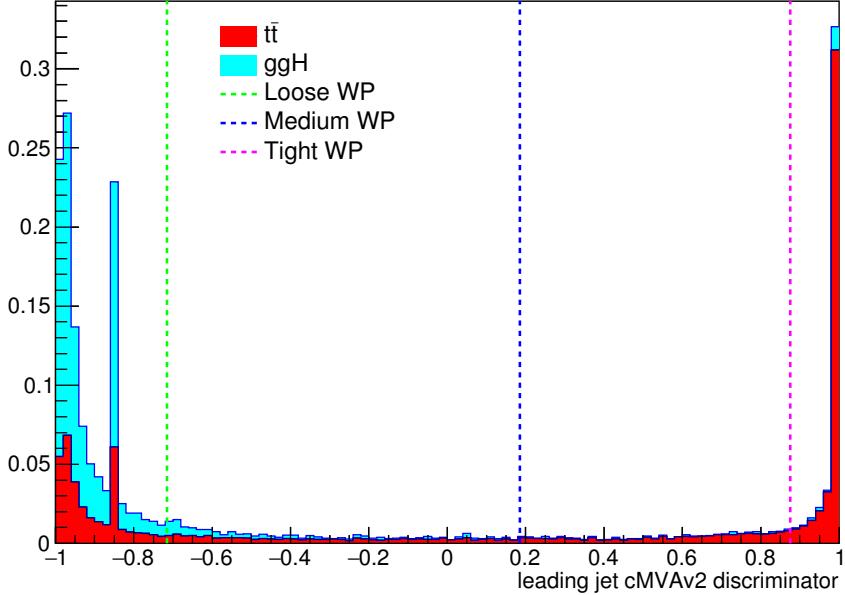
algorithms. To compare the b tagging performance  $\epsilon_{b\text{veto}}$  is computed for different working points, i.e. different selections on the specific b tagging discriminator, and the results are reported in the form of a ROC curve. The ROC curves corresponding to events with 0, 1 and  $\geq 2$  jets are shown in Fig. 5.2. 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 5.3.

<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

2336 The two categories were eventually combined together and the significance assessment was  
 2337 repeated for the three working points. With these selection we find the significance values  
 2338 listed in Table 5.5 for the three working points.

2339 The working point providing the best significance in the combined 0 + 1 jets category is  
 2340 found to be the loose one.

2341 To correct for a possible different b tagging efficiency in data and simulation, the  
 2342 simulated events are reweighted using scale factors computed in bins of the jet  $\eta$  and  $p_T$ .  
 2343 These scale factors and the corresponding uncertainties are centrally calculated for each  
 2344 working point, in such a way to be employable by all the CMS analyses. The prescription  
 2345 to reweight the simulated events is the following. First of all one has to compute the b  
 2346 tagging efficiency using the MC samples,  $\varepsilon_{\text{MC}}(p_T, \eta, f)$ , for the chosen working point in bins  
 2347 of jet  $p_T$  and  $\eta$ . The efficiency has to be computed for different flavours  $f$  of the jets, b,  
 2348 c and light (u,d,s), using the jet matching information<sup>2</sup> which is available in all the MC  
 2349 samples. An MC-based event weight is then calculated computing the probability  $P_{\text{MC}}$  of a  
 2350 given b tagging configuration to occur, e.g.:

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

2351 Afterwards, a similar probability is computed using data:

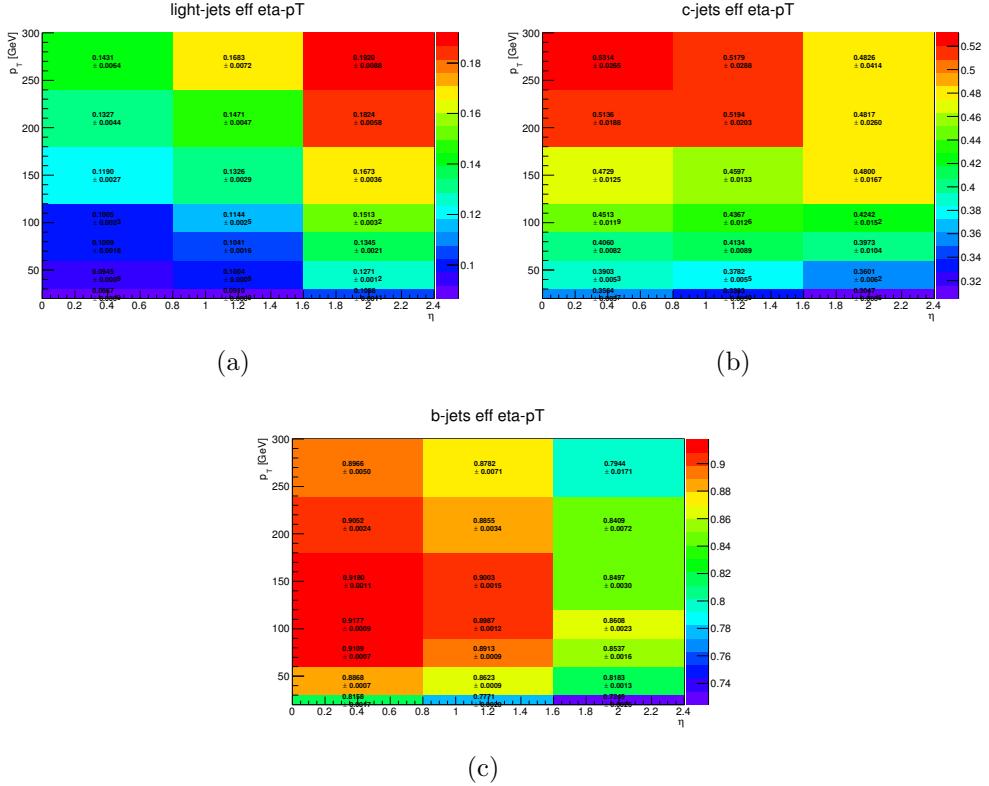
$$P_{\text{DATA}} = \prod_{i \in \text{b-tagged-jets}} SF_i \varepsilon_{\text{MC}_i} \prod_{j \in \text{non-b-tagged-jets}} (1 - SF_j \varepsilon_{\text{MC}_j}) , \quad (5.2)$$

2352 where  $SF_i$  is the provided scale factor value for the relevant jet flavour,  $p_T$  and  $\eta$ . Products  
 2353 in Eqs. 5.1 and 5.2 run over all jets. The event weight is finally given by the ration  
 2354  $P_{\text{DATA}}/P_{\text{MC}}$ .

2355 The b tagging efficiencies to be fed into Eq. 5.1 and Eq. 5.2 are derived using  $t\bar{t}$  simulated  
 2356 events and applying basic leptonic selections. These efficiencies are shown in Fig. 5.4 for

<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.

2357 light (a), c-jets (b) and b-jets (c), in bins of  $\eta$  and  $p_T$ . The uncertainties associated to the  
 2358 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  $p_T$ .

2359  
 2360 The effect of the event reweighting is to correct the shape of the b tagging discriminator  
 2361 in simulation, moving events from the b tag region (discriminator greater than  $> -0.715$ )  
 2362 to the b veto region (discriminator  $< -0.715$ ) and viceversa. A data/simulation comparison  
 2363 of the b tagging discriminator for the leading and subleading jets is performed to check the  
 2364 agreement after the application of the event weights. In order to evaluate the data/simulation  
 2365 agreement for b-jets, the data and simulation are compared in a top enriched control region,  
 2366 defined by the following requirements:

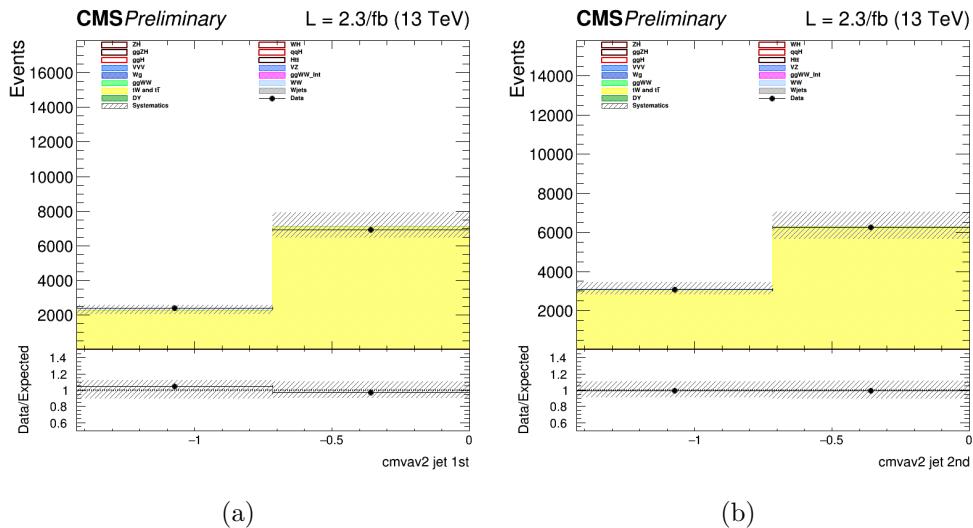
- 2367 • two leptons, an electron and a muon with opposite charge, with leading lepton  $p_T$   
   2368 greater than 20 GeV and sub-leading lepton  $p_T$  greater than 15 GeV;
- 2369 • no other lepton (electron or muon) with  $p_T$  greater than 10 GeV;
- 2370 • lepton invariant mass greater than 50 GeV;
- 2371 • at least two jets with  $p_T$  greater than 30 GeV;

- 2372     • at least one of the two leading jets with cMVAv2 btagging score greater than -0.715  
 2373       (i.e. the loose working point).

2374   In order to evaluate the agreement for light jets, a second control region is defined, populated  
 2375   by Z+light jet events, defined as follows:

- 2376     • two leptons, two electrons or two muons with opposite charge, with leading lepton  $p_T$   
 2377       greater than 20 GeV and sub-leading lepton  $p_T$  greater than 15 GeV.  
 2378     • no other lepton (electron or muon) with  $p_T$  greater than 10 GeV.  
 2379     • lepton invariant mass greater between 80 GeV and 110 GeV.  
 2380     • at least two jets with  $p_T$  greater than 30 GeV.  
 2381     • at least one jet above 30 GeV.  
 2382     • no jets above 20 GeV with a TCHE score above 2.1.

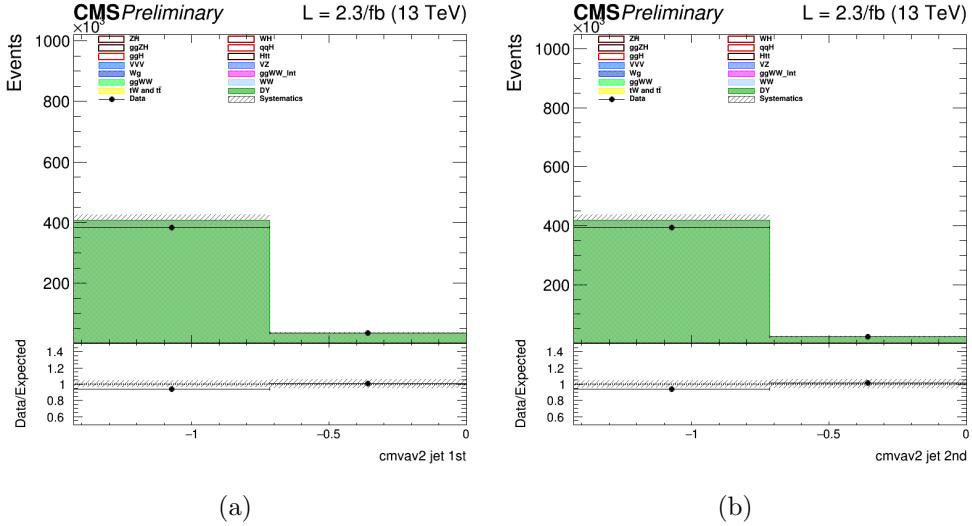
2383   Although a Z+jets sample is dominated by light flavor jets, a b-veto on an alternative  
 2384   algorithm (TCHE) is applied to reduce the contamination from b-jets, especially above the  
 2385   cMVAv2 cut. This helps mitigating possible data/simulation discrepancies in the modeling  
 2386   of the heavy/light flavour ratio. The comparison between data and simulation after the  
 2387   event reweighting is shown in Figs. 5.5 and 5.6 for the b-jets and light jets enriched control  
 2388   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.

### 2389 5.3.2. Event selection and background rejection

2390   Since the ggH production mechanism, which is the main production mode for a Higgs  
 2391   mass of around 125 GeV, is characterized by the emission of few jets arising from initial



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

or final state radiation, this analysis is limited to events with no jets or one jet. Due to the large DY background in di-electrons and di-muons events, only the  $e\mu$  final state is studied in this early Run 2 data analysis, including the indirect contribution from  $\tau$  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  $p_T > 30$  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.

2416 Higher jet multiplicity categories, which are sensitive to other Higgs production mechanisms,  
2417 such as VBF, are not included in this analysis, given the very low expected yield for other  
2418 production modes with the analysed integrated luminosity.

2419 Distributions of some variables of interest for the 0 and 1 jet categories separately,  
2420 but merging the  $e\mu$  and  $\mu e$  final states together, are shown in Figs. 5.7, 5.8 and 5.9 after  
2421 applying the WW baseline selections, with the addition of a cut on  $m_{\ell\ell}$  to remove the Higgs  
2422 signal contribution ( $m_{\ell\ell} > 80$  GeV), and a cut on  $m_T$  to be orthogonal to the  $Z\gamma^* \rightarrow \tau\tau$   
2423 background control region ( $m_T > 60$  GeV).

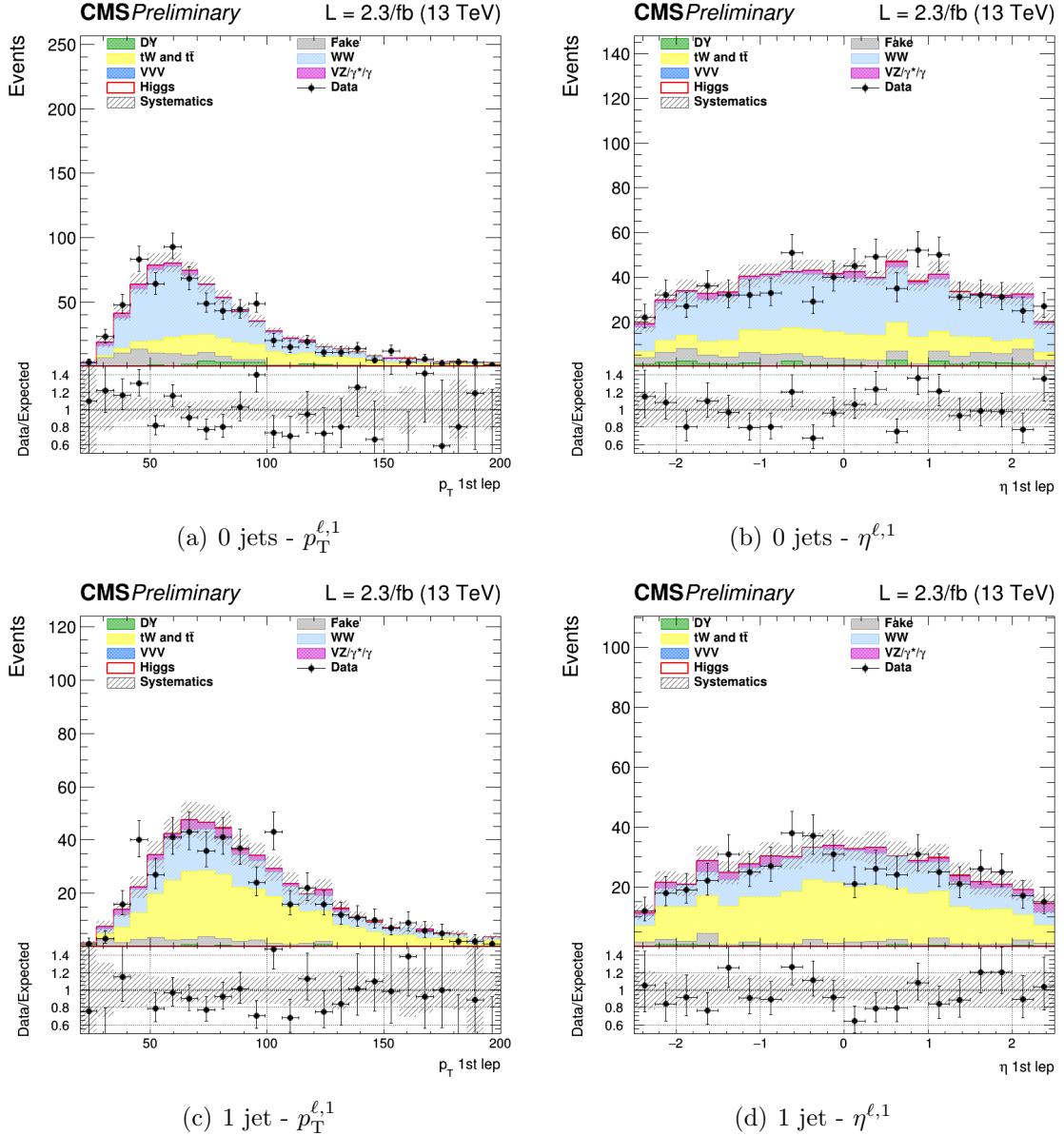
2424 The  $W+jets$  background, where one jet can be misidentified as a lepton, is a sub-  
2425 dominant background in the phase space defined by the analysis kinematic requirements.  
2426 The 0 ad 1 jets categories are further split according to the lepton flavour to  $e\mu$  and  $\mu e$ ,  
2427 where the first lepton refers to the leading one. In this way an improvement of about 10% in  
2428 terms of the signal significance can be achieved, exploiting the different  $W+jets$  background  
2429 contribution in the two categories. Indeed the probability for a jet to be misidentified as an  
2430 electron or a muon is not the same.

### 2431 5.3.3. Signal extraction

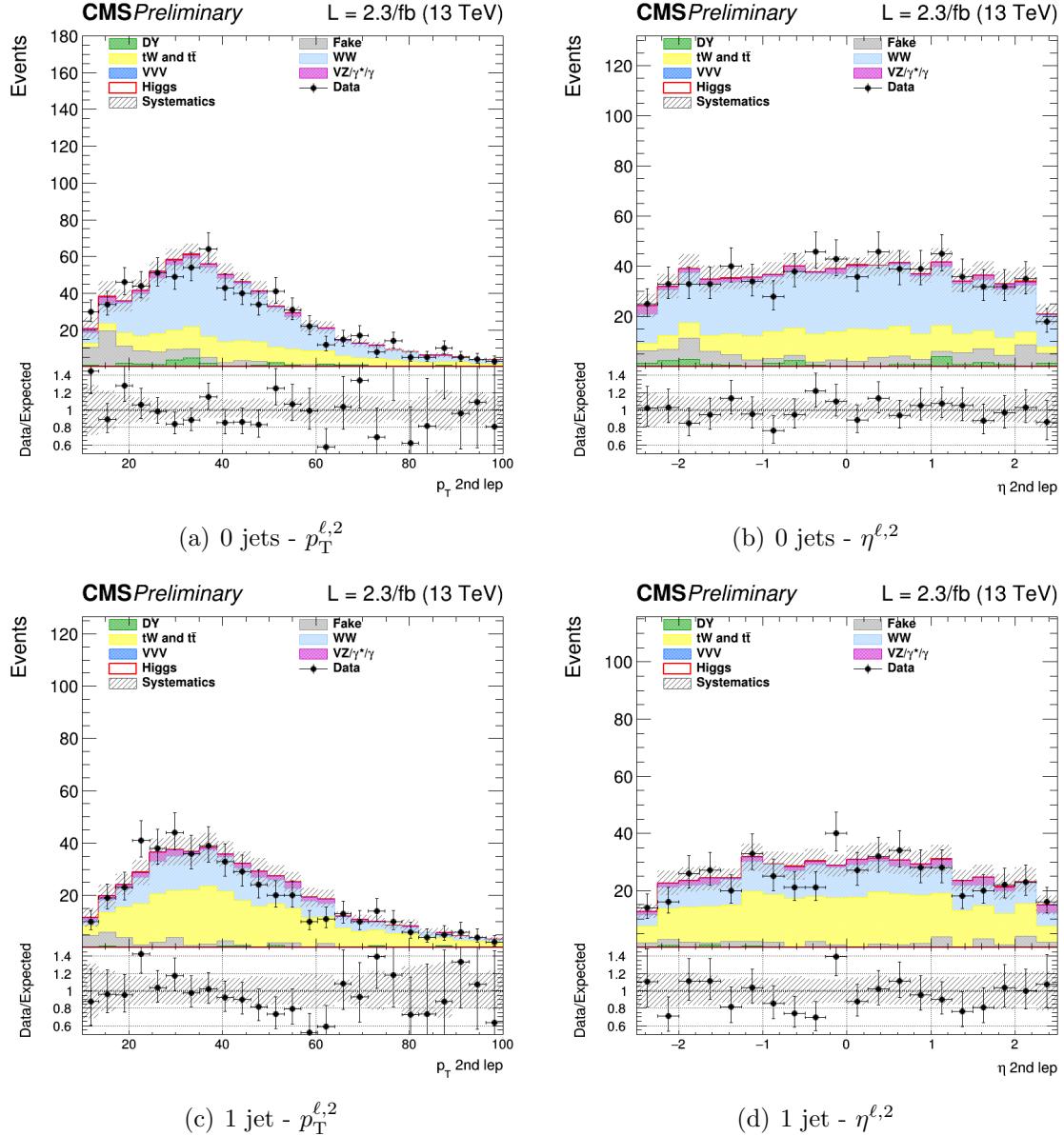
2432 To extract the Higgs boson signal contribution in the four previously mentioned categories,  
2433 a similar approach to the one used in the Run 1 analysis [54] is pursued. The analysis is  
2434 based on two-dimensional templates of  $m_{\ell\ell}$  versus  $m_T$  to discriminate signal and background  
2435 contributions. The  $m_{\ell\ell}$  template is defined using 5 bins from  $m_{\ell\ell} = 10$  GeV up to  $m_{\ell\ell} =$   
2436 110 GeV, while for the  $m_T$  template 7 bins are defined in the range  $60$  GeV  $< m_T < 200$  GeV.  
2437 The phase space with  $m_T < 60$  GeV is used as an orthogonal control region to extract the  
2438 normalization of the DY background. A binned maximum likelihood fit to the signal and  
2439 background two-dimensional templates is performed to extract the signal strength in the  
2440 four categories.

2441 Distributions of the  $m_{\ell\ell}$  and  $m_T$  variables after the WW level selection are shown in  
2442 Fig. 5.10 for the 0 and 1 jet categories separately, but merging the  $e\mu$  and  $\mu e$  final states  
2443 together.

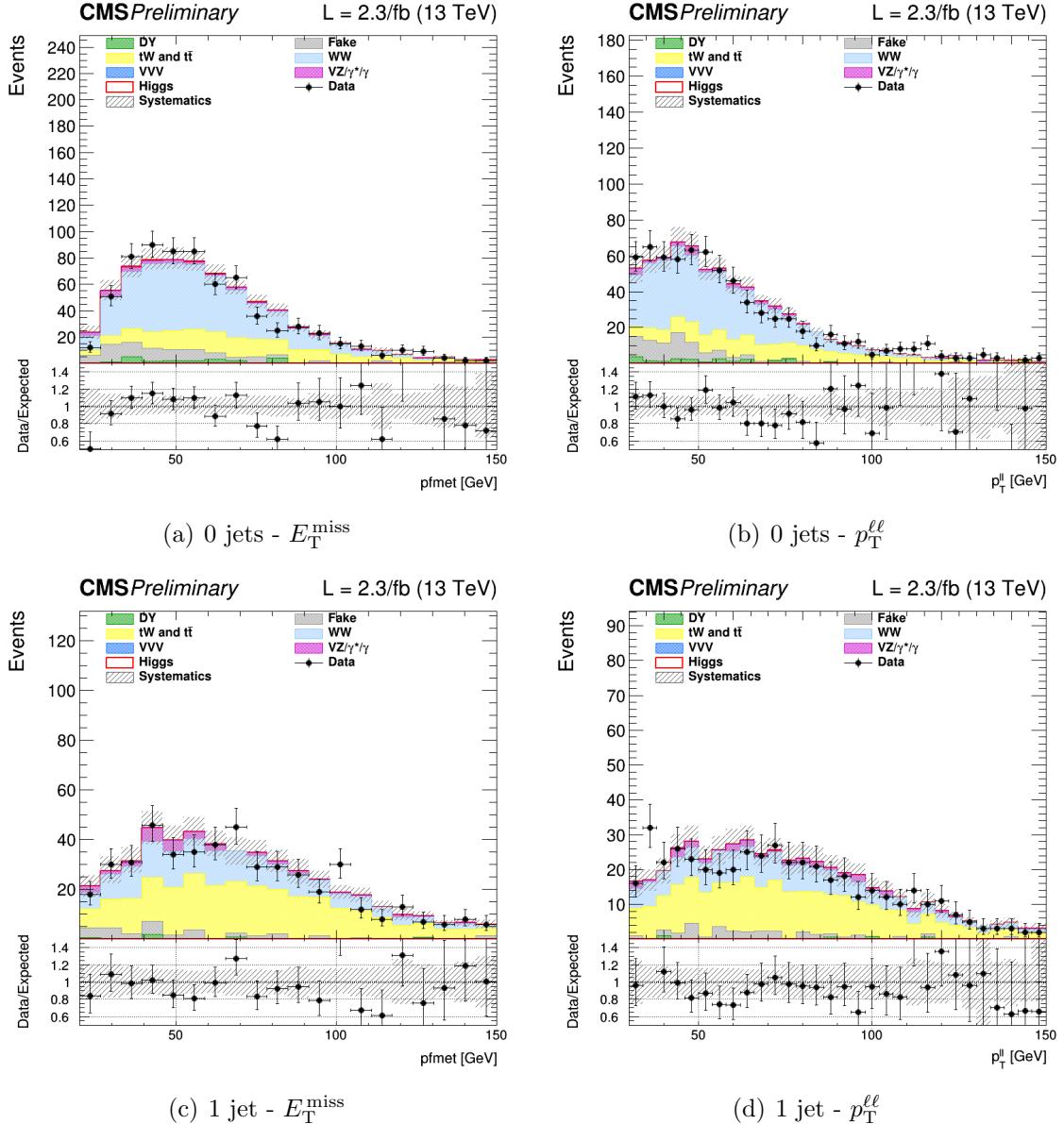
2444 The statistical methodology used to interpret the data and to combine the results from  
2445 the independent 0-jet and 1-jet categories in the  $e\mu$  and  $\mu e$  final states has been developed  
2446 by the ATLAS and CMS collaborations in the context of the LHC Higgs Combination  
2447 Group [94, 112]. The number of events in each category and in each bin of two-dimensional  
2448 template is modelled as a Poisson random variable, with a mean value given by the sum  
2449 of the contributions from all the processes under consideration. Systematic uncertainties  
2450 are represented by individual nuisance parameters with log-normal distributions. The  
2451 uncertainties affect the overall normalization of the signal and backgrounds as well as the  
2452 shape of the predictions across the distribution of the observables. Correlation between  
2453 systematic uncertainties in different categories are taken into account.



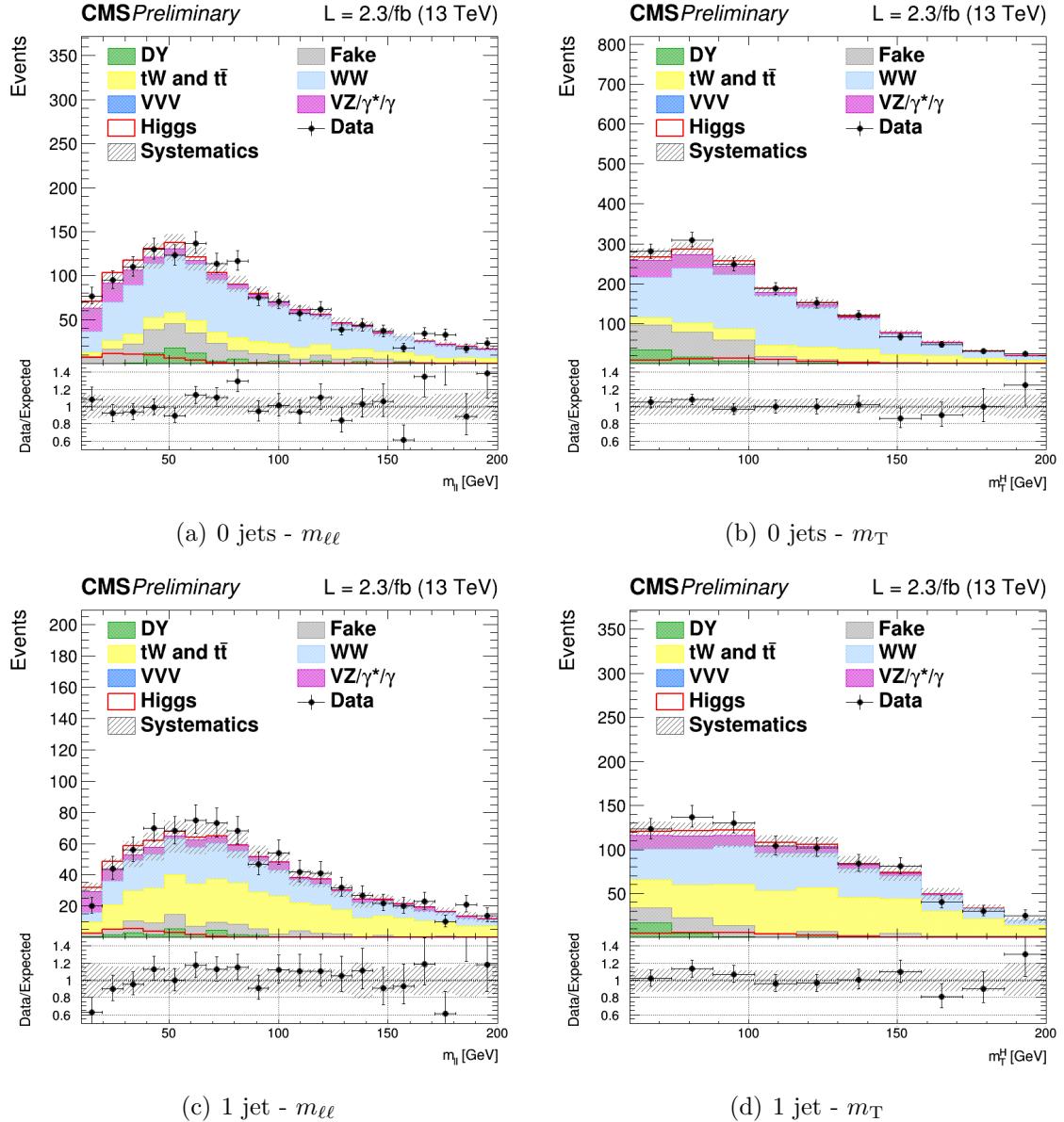
**Figure 5.7.:** Distributions of  $p_T$  (left) and  $\eta$  (right) of the leading lepton for events with 0 jet (upper row) and 1 jet (lower row), for the main backgrounds (stacked histograms), and for a SM Higgs boson signal with  $m_H = 125$  GeV (superimposed and stacked red histogram) at the  $WW$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

## 2454 5.4. Background estimation

2455 The main background processes affecting the analysis signature, non-resonant WW production and top quark processes, are estimated using data. Backgrounds arising from an 2456 experimental misidentification of the objects, such as W+jets (also called “Fake”), are 2457 estimated using data as well. The other minor backgrounds are generally estimated directly 2458 from simulation as described in the following subsections. 2459

### 2460 5.4.1. WW background

2461 The quark-induced WW background is simulated with NLO accuracy in perturbative 2462 QCD, and the transverse momentum of the diboson system is reweighted to match the 2463 NNLO+NNLL accuracy from theoretical calculations [100, 101]. However, given the large 2464 uncertainties on the jet multiplicity distribution associated to this process, the normalization 2465 of this background is measured from data separately for the 0 and 1 jet categories. The 2466 normalization k-factors are extracted directly from the fit together with the signal strengths, 2467 leaving the WW normalization free to float separately in the two jet multiplicity categories. 2468 An orthogonal control region for the WW background normalization estimation is not 2469 needed in this case, owing to the different  $m_{\ell\ell}$ - $m_T$  shape for signal and background.

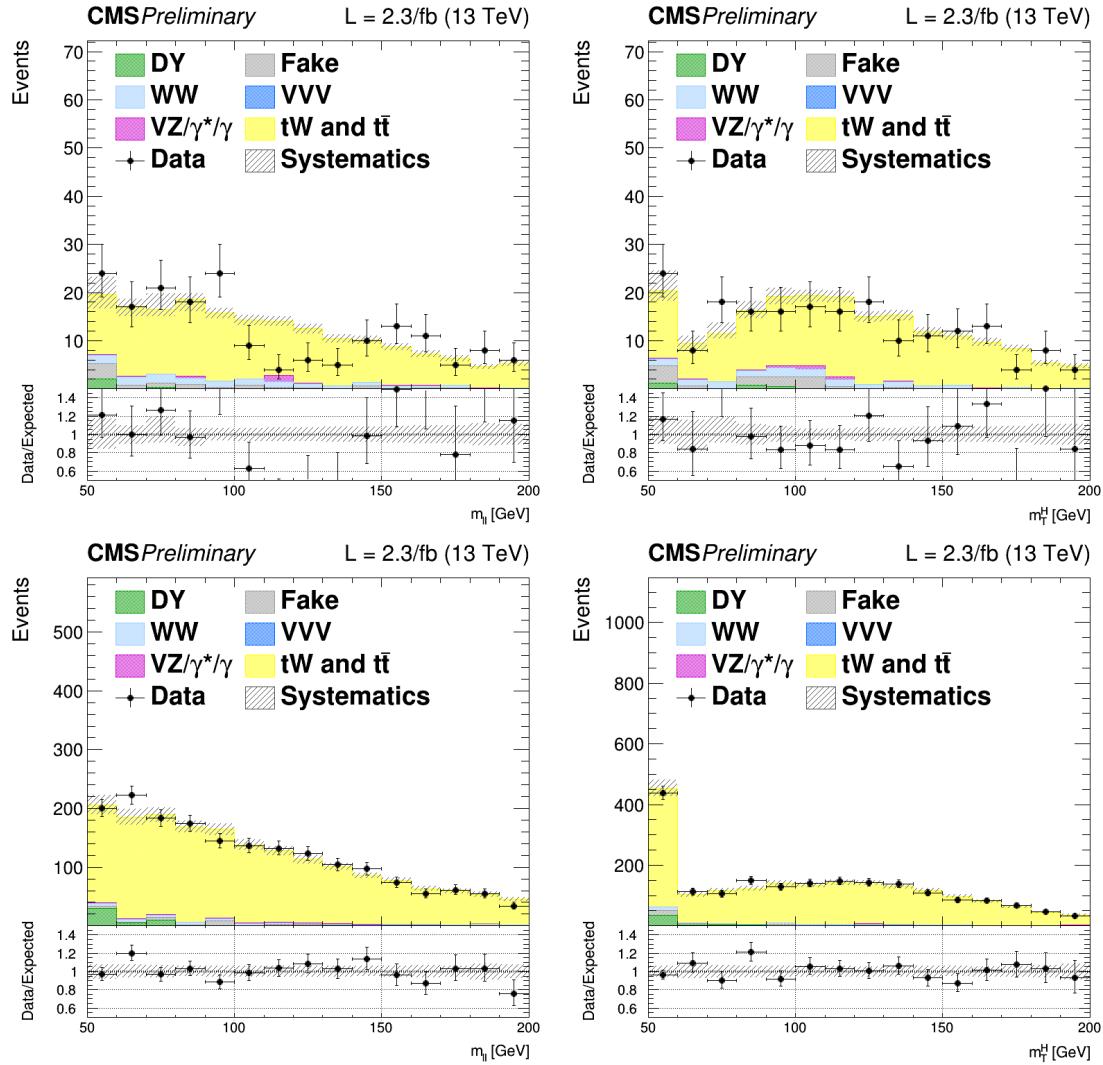
2470 The gluon-induced WW production is sub-dominant with respect to the quark-induced 2471 production, and its shape and normalization is fully taken from simulation, scaling the 2472 cross section to the theoretical prediction with NLO accuracy [109].

### 2473 5.4.2. Top quark background

2474 As explained in Sec. 5.3, the production of top quark pairs represents one of the dominant 2475 backgrounds in this analysis given its large cross section and a similar final state compared 2476 to the signal. A b-jet veto, based on the *cMVAv2* b tagging algorithm, is used to suppress 2477 this background and a reweighting procedure is applied on top of the simulated events to 2478 correct for different b tagging efficiency in data and simulation.

2479 The top quark background normalization is measured using data, defining a b-jets 2480 enriched control region by inverting the b-jet veto. More precisely, the b-jets enriched 2481 control region for the 0-jet category is defined with the same WW baseline selection but 2482 requiring at least one jet with  $20 < p_T < 30$  GeV to be identified as a b jet and no other 2483 jets with  $p_T > 30$  GeV. For the 1-jet category, the b-jets enriched region is defined requiring 2484 exactly one jet with  $p_T > 30$  GeV identified as a b-jet. To reduce other backgrounds in 2485 these two regions, the dilepton mass has to be greater than 50 GeV. Distributions of the 2486  $m_{\ell\ell}$  and  $m_T$  variables in the b-jets enriched control regions after applying the data driven 2487 estimation are shown in Figure 6.5, for the 0 and 1 jet categories separately.

2488 The top quark background normalization is constrained during the fit procedure sepa- 2489 rately in the two jet categories, by means of the control regions defined above, which are 2490 treated in the fit as two additional categories.

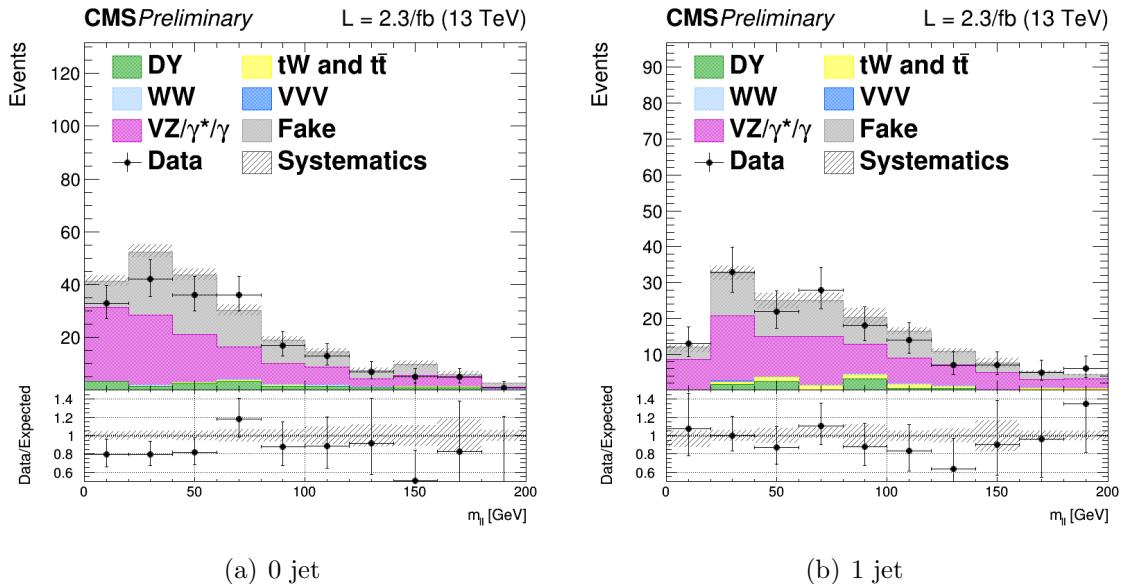


**Figure 5.11.:** Distributions of  $m_{\ell\ell}$  (left) and  $m_T$  (right) for events with 0 jet (top) and 1 jet (bottom) in top enriched phase space. Scale factors estimated from data are applied. The first (last) bin includes underflows (overflows).

### 2491 5.4.3. Jet-induced (or Fake) background

2492 One of the primary source belonging to this category arises from the misidentification of  
 2493 leptons in  $W+jets$  processes in the 0 jet category. Also, semileptonic  $t\bar{t}$  decays contribute  
 2494 especially for higher jet multiplicities. Multijet production and hadronic  $t\bar{t}$  decays are also  
 2495 taken into account, but have a much smaller contribution.

2496 This background is fully estimated using data, with the technique described in Sec. 4.4.3.  
 2497 To check the agreement of the background estimated in this way with data, a control sample  
 2498 enriched in jet-induced events is defined. The events in the control sample are selected  
 2499 applying the  $WW$  baseline requirements but requesting an  $e\mu$  pair with same charge, which  
 2500 significantly suppresses the  $WW$  and  $t\bar{t}$  processes. The  $m_{\ell\ell}$  distributions in this control  
 2501 region for the 0 and 1 jet categories are shown in Fig. 5.12. From the crosscheck in this  
 2502 control region, a global normalization factor of 0.8 is derived and applied to the jet-induced  
 2503 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.

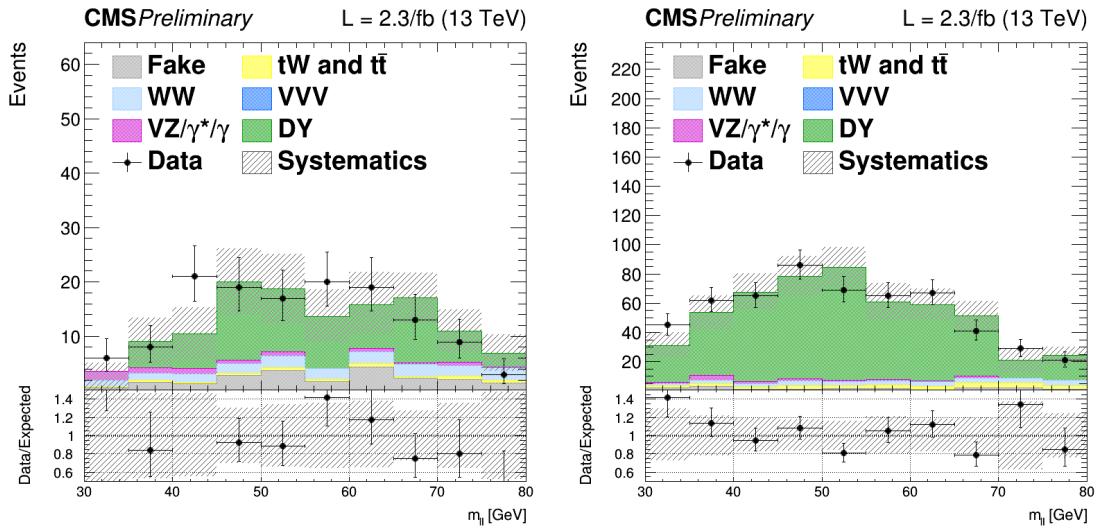
### 2504 5.4.4. DY background

2505 This background contributes to the analysis phase space because of the  $Z/\gamma^*$  decays to a  
 2506 pair of  $\tau$  leptons, which consequently decays to an  $e\mu$  pair. This background process is  
 2507 predominant in the low  $m_T$  region, which is used as an orthogonal control region to determine  
 2508 the background normalization in the 0 and 1 jet categories separately. In particular this  
 2509 control region is defined by selecting events with  $m_T < 60$  GeV and  $30$  GeV  $< m_{\ell\ell} < 80$  GeV.

2510 The  $m_{\ell\ell}$  distributions in these control regions for the 0 and 1 jet categories are shown in  
 2511 Fig. 5.13.

2512 As for the top quark background, the normalization of this background in the 0 and 1  
 2513 jet categories, is constrained directly in the fit by means of the control regions, which are  
 2514 treated as two additional categories.

2515 The kinematics of this background is taken from simulation, after reweighting the Z  
 2516 boson  $p_T$  spectrum to match the observed distribution measured in data. In fact, this  
 2517 variable is not well reproduced by the MC generator used for simulating this process,  
 2518 especially in the bulk of the distribution, the discrepancy being ascribed to the missing  
 2519 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.

#### 2520 5.4.5. Other backgrounds

2521 The  $W\gamma^*$  and the  $WZ$  electroweak processes can be gathered in the same physical process,  
 2522 although the final state kinematics is rather different. In particular, the invariant mass of  
 2523 the leptons arising from the  $\gamma^*$  decays is generally below 4 GeV, while the leptons from the  
 2524 Z boson decay are characterized by a larger invariant mass. Another background which can  
 2525 be experimentally identical to those is the  $W\gamma$  production, where a real photon is produced  
 2526 in association with a W boson and consequently undergoes a photon conversion to leptons  
 2527 due to the interaction with the material constituting the first layers of the silicon tracker.

2528 All these backgrounds may contribute to the signal phase space whenever one of the three  
 2529 leptons escape from the detector acceptance or is not identified. The shape and cross section  
 2530 of these backgrounds are taken from simulation. The only exception is the normalization of  
 2531 the  $W\gamma^*$  background, being this process dominant in the low  $m_{\ell\ell}$  region, which is scaled

2532 to data defining a proper control region. The control region is defined selecting events  
 2533 with three isolated muons, with  $p_T > 10,5$  and  $3\text{ GeV}$  for the first three leading muons  
 2534 respectively. The selection is further defined by  $E_T^{\text{miss}} < 25\text{ GeV}$  and  $E_T^{\text{miss}}$  projected to the  
 2535 leading muon  $< 45\text{ GeV}$ . The pair of muons with the smallest invariant mass is taken as  
 2536 coming from the  $\gamma^*$  decay. The k-factor measured in data for this background to be applied  
 2537 in the simulation is  $1.98 \pm 0.54$ .

2538 All remaining backgrounds from di-boson and tri-boson production, which are of mi-  
 2539 nor importance in the analysis phase space, are normalized according to their expected  
 2540 theoretical cross sections.

## 2541 5.5. Systematic uncertainties

2542 The systematic uncertainties affecting this measurement can be divided into three categories:  
 2543 the uncertainties on the background estimation, experimental uncertainties and theoretical  
 2544 uncertainties.

2545 The first category includes the uncertainties related to the background normalization  
 2546 and shape. For the non-resonant  $WW$  production the shape is taken from simulation.  
 2547 The input normalization to the fit is set to the expected value from simulation, and an  
 2548 unconstrained nuisance parameter with a flat distribution is associated to this number.  
 2549 This is done separately for the two jet categories.

2550 The top quark background shape is taken from simulation after correcting for the b  
 2551 tagging scale factors. An uncertainty due to these scale factors is included and affects  
 2552 both the normalization and the shape of the top quark background. The uncertainties on  
 2553 the normalization are treated similarly to the  $WW$  background case, but constraining the  
 2554 corresponding nuisances by means of the two control regions orthogonal to the signal phase  
 2555 space. A similar procedure is used for the DY background.

2556 Effects due to experimental uncertainties are studied by applying a scaling and smearing  
 2557 of certain variables related to the physics objects, e.g. the  $p_T$  of the leptons, followed by  
 2558 a subsequent recalculation of all the correlated variables. This is done for simulation, to  
 2559 account for possible systematic mismodeling.

2560 All experimental sources, except luminosity, are treated both as normalization and shape  
 2561 uncertainties, and are correlated among the signal and background processes and all the  
 2562 categories. The following experimental uncertainties are considered:

- 2563 • the uncertainty determined by the CMS online luminosity monitoring, 2.7% for the  
 2564 first data collected at  $\sqrt{s} = 13\text{ TeV}$ ;
- 2565 • the acceptance uncertainty associated with the combination of single and double lepton  
 2566 triggers, which is 2%;
- 2567 • the lepton reconstruction and identification efficiencies uncertainties, that are in the  
 2568 range 0.5-5% for electrons and 1-7% for muons depending on  $p_T$  and  $\eta$ ;
- 2569 • the muon momentum and electron energy scale and resolution uncertainties, that  
 2570 amount to 0.01-0.5% for electrons and 0.5-1.5% for muons depending on  $p_T$  and  $\eta$ ;

- 2571 • the jet energy scale uncertainties, that vary between 1-11% depending on the  $p_T$  and  
2572  $\eta$  of the jet;
- 2573 • the  $E_T^{\text{miss}}$  resolution uncertainty, that is taken into account by propagating the corre-  
2574 sponding uncertainties on the leptons and jets;
- 2575 • the scale factors correcting the b tagging efficiency and mistagging rate, that are varied  
2576 within their uncertainties. This systematic uncertainty is anticorrelated between the  
2577 top control regions and the other ones.

2578 The uncertainties in the signal and background production rates due to theoretical  
2579 uncertainties include several components, which are assumed to be independent: the PDFs  
2580 and  $\alpha_s$ , the underlying event and parton shower model, and the effect of missing higher-order  
2581 corrections via variations of the renormalization and factorization scales.

2582 The effects of the variation of PDFs,  $\alpha_s$  and renormalization/factorization QCD scales,  
2583 mainly affect the signal processes, being the most important backgrounds estimated using  
2584 data driven techniques. However, the uncertainties on minor backgrounds that are estimated  
2585 from simulation are taken into account. These uncertainties are split in the uncertainties  
2586 on the cross section, which are computed by the LHC cross section working group [113],  
2587 and on the selection efficiency [114]. The PDFs and  $\alpha_s$  signal cross section normalization  
2588 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  
2589 mechanism. The PDFs and  $\alpha_s$  acceptance uncertainties are less than 1% for gluon- and  
2590 quark-induced processes. The effect of the QCD scales variation on the selection efficiency is  
2591 around 1-3% depending on the specific process. To estimate these uncertainties, the events  
2592 are reweighted according to different QCD scales or different PDF sets and the selection  
2593 efficiency is recomputed each time. For the QCD scale uncertainty the maximum variation  
2594 with respect to the nominal value is taken as the uncertainty. For the case of PDF and  $\alpha_s$   
2595 uncertainties, the distribution of the selection efficiency is built taking into account all the  
2596 replicas in the NNPDF3.0 set and the uncertainty is estimated as the standard deviation of  
2597 that distribution.

2598 In addition, the categorization of events based on jet multiplicity introduces additional un-  
2599 certainties on the ggH production mode related to missing higher order corrections. These un-  
2600 certainties are evaluated following the prescription described in Sec. [subsec:stewart-tackman]  
2601 ] and correspond to 5.6% for the 0-jet and 13% for the 1-jet bin categories.

2602 The underlying event uncertainty is estimated by comparing two different PYTHIA  
2603 8 tunes, while parton shower modelling uncertainty is estimated by comparing samples  
2604 interfaced with the PYTHIA 8 and HERWIG++ parton shower programs. The effect on the  
2605 ggH (VBF) signal expected yield is about 5% (5%) for the PYTHIA 8 tune variation and  
2606 about 7% (10%) for the parton shower description.

2607 Other specific theoretical uncertainties are associated to some backgrounds. An uncer-  
2608 tainty on the ratio of the  $t\bar{t}$  and  $tW$  cross sections is included. Indeed, these two processes  
2609 are characterized by a different number of b-jets in the final state (2 b-jets for  $t\bar{t}$  and 1  
2610 for  $tW$ ) and the b-veto acts differently for the two. A variation of the relative ratio of the  
2611 cross sections can thus cause a migration of events from the 0 to the 1 jet categories and

2612 viceversa. The corresponding uncertainty is of 8%, according to the theoretical cross section  
2613 calculations [110, 111].

2614 The  $gg \rightarrow WW$  background LO cross section predicted by the MCFM generator is scaled  
2615 to the NLO calculation, applying a k-factor of 1.4 with an uncertainty of 15% [109]. The  
2616 interference term between the  $gg \rightarrow WW$  and the  $ggH$  signal is also included and simulated  
2617 with LO accuracy using MCFM. The k-factor to scale the interference term is 1.87, given by  
2618 the geometrical average of the LO to NNLO  $gg \rightarrow H \rightarrow WW$  scale factor (2.5) and the LO  
2619 to NLO  $gg \rightarrow WW$  scale factor (1.4). The uncertainty on this value is estimated as the  
2620 maximum variation with respect to the two scale factors mentioned above, and is found  
2621 to be of 25%. Anyway, with the current amount of integrated luminosity, the interference  
2622 contribution is found to be negligible.

2623 For what the  $qq \rightarrow WW$  background shape is concerned, an uncertainty related to  
2624 the diboson  $p_T$  reweighting is evaluated varying the renormalization, factorization and  
2625 resummation QCD scales.

2626 Finally, the uncertainties due to the limited statistical accuracy of the MC simulations  
2627 are also taken into account, including an independent uncertainty for each bin of the  
2628 two-dimensional distribution, and for each category. The uncertainty for a certain bin  
2629 and process is given by the standard deviation of the Poisson distribution with mean  
2630 corresponding to the number of MC events in that bin.

## 2631 5.6. Results

2632 The expected and observed signal significance are shown in Table 5.6 for all the categories  
2633 separately. Also, the observed signal strengths and the corresponding uncertainties are  
2634 shown. The best fit signal strength obtained combining all the categories together is found  
2635 to be  $0.3^{+0.5}_{-0.5}$ , corresponding to an observed significance of  $0.7\sigma$ , to be compared with the  
2636 expected significance of  $2.0\sigma$  for a Higgs boson mass of 125 GeV.

2637 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.

## 2638 **Search for high mass resonances 2639 decaying to a W boson pair with first 2640 13 TeV LHC data**

### 2641 **6.1. Introduction**

2642 In this chapter, a search for a high mass spin-0 particle (from now on denoted as  $X$ )  
2643 in the  $X \rightarrow WW \rightarrow \ell\nu\ell'\nu'$  decay channel is presented, where  $\ell$  and  $\ell'$  refer to an different  
2644 flavour lepton pair, i.e.  $e\mu$ . The search is based upon proton-proton collision data samples  
2645 corresponding to an integrated luminosity of up to  $2.3 \text{ fb}^{-1}$  at  $\sqrt{s} = 13 \text{ TeV}$ , recorded by  
2646 the CMS experiment at the LHC during 2015. This analysis represents a general extension  
2647 of the SM Higgs boson search presented in 5 and is performed in a range of heavy scalar  
2648 masses from  $M_X = 200 \text{ GeV}$  up to  $1 \text{ TeV}$ , extending the range studied in a similar analysis  
2649 performed using Run 1 LHC data [115], which provided upper limits on the production  
2650 cross section of new scalar resonances up to  $600 \text{ GeV}$ .

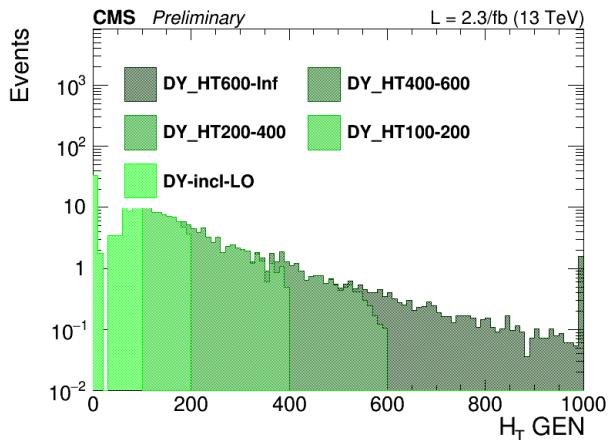
2651 Despite the discovery of a particle consistent with the SM Higgs boson in 2012, there is a  
2652 possibility that this particle is only a part of a larger Higgs sector, and hence only partially  
2653 responsible of the EW symmetry breaking. This can be achieved in different theoretical  
2654 models that extends the SM, such as the two-Higgs-doublet models [116–118], or models in  
2655 which the SM Higgs boson mixes with a heavy EW singlet, which predict the existence  
2656 of an additional resonance at high mass, with couplings similar to those of the SM Higgs  
2657 boson, as most recently described in [119, 120].

2658 This analysis reports a generic search for a scalar particle with different resonance decay  
2659 widths hypothesis, produced via the ggH and VBF production mechanisms. The results  
2660 can then be interpreted in terms of different theoretical models. This analysis is heavily  
2661 based on the SM Higgs search described in 5, in terms of physics objects, selections and  
2662 background estimation. The differences and similarities are discussed in this chapter.

### 2663 **6.2. Data and simulated samples**

2664 The data sets, triggers, pile up reweighting, lepton identification and isolation used in this  
2665 analysis are the same as the SM Higgs search and are described in Sec. 5.2.

Also, the same MC simulations are used for the background processes, the only exception being the DY background, for which the MG5\_AMC@NLO generator is used with LO QCD accuracy, matching together events with up to four jets in addition to the vector boson with the MLM [121] 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. 6.1. The DY LO cross section obtained from the simulation is scaled using the LO to NNLO k-factor of 1.23.



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

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

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 [96]. Details about the interference effects are given in Sec. 6.3.

## 2693 6.3. Analysis strategy

2694 The analysis strategy for the first results on the high mass search in the  $W^+W^- \rightarrow 2\ell 2\nu$   
 2695 decay channel closely follows the strategy presented in the 13 TeV SM Higgs search in the  
 2696  $H \rightarrow W^+W^- \rightarrow 2\ell 2\nu$  channel regarding the 0 and 1 jet categories. In addition a dedicated  
 2697 category to the VBF production mechanism is added, given that this production mode is  
 2698 particularly important in the high mass region. Indeed, assuming a SM Higgs boson, the  
 2699 ratio of cross sections  $\sigma_{\text{VBF}}/\sigma_{\text{ggH}}^1$  increases with the Higgs boson mass, making the VBF  
 2700 production mechanism more and more important as the mass of the resonance approaches  
 2701 to high values.

2702 This analysis is affected essentially by the same background processes as the SM Higgs  
 2703 boson search, with the difference that in this case the SM Higgs boson processes, including  
 2704 all production modes, are treated as backgrounds.

2705 In addition to requiring the events to pass the single or double lepton triggers, exactly  
 2706 one electron and one muon are required to be reconstructed in the event with opposite  
 2707 charges and a minimum  $p_T$  of 20 GeV for both the muon and electron. Both leptons are  
 2708 required to be well identified and isolated to reject fake leptons and leptons coming from  
 2709 decays in flight. To suppress background processes with three or more leptons in the final  
 2710 state, such as diboson or triboson production, events with any additional identified and  
 2711 isolated lepton with  $p_T > 10$  GeV are rejected. To suppress the contribution of the SM  
 2712 production of the Higgs boson at 125 GeV,  $m_{\ell\ell}$  is requested to be higher than 50 GeV. The  
 2713 other event requirements are identical to the 125 GeV Higgs boson search and are described  
 2714 in Sec. 5.3.2.

2715 In addition to the 0 and 1 jet categories, a specific category sensitive to the VBF  
 2716 production mode is defined exploiting the characteristic signature of this process, where  
 2717 two energetic jets are emitted in the forward region of the detector and with large  $\Delta\eta$   
 2718 gap. Events belonging to the VBF-enriched category are selected by requiring at least  
 2719 two jets with  $p_T > 30$  GeV, an invariant mass  $m_{jj} > 500$  GeV and a gap in pseudorapidity  
 2720  $\Delta\eta_{jj} > 3.5$ .

2721 In addition to the transverse mass variable  $m_T$ , which is used in the analysis selection  
 2722 to define the DY background control region, an additional variable is defined, that from  
 2723 now on will be labelled as “improved transverse mass”  $m_T^i$ . This variable is defined as  
 2724 the invariant mass of the four momentum resulting from the sum of the two leptons four  
 2725 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.:

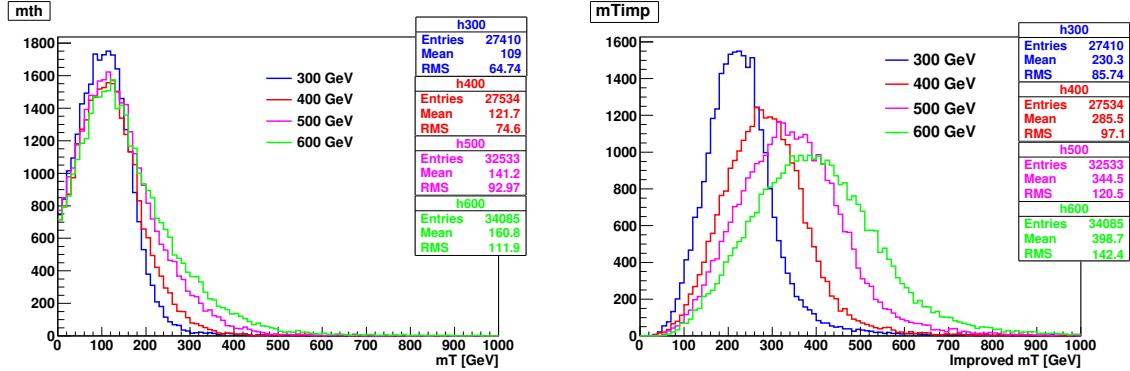
$$m_T^i = \sqrt{(p_{\ell\ell} + E_T^{\text{miss}})^2 - (\vec{p}_{\ell\ell} + \vec{p}_T^{\text{miss}})^2} \quad . \quad (6.1)$$

2726 This variable allows having a better sensitivity to different resonance mass hypothesis  
 2727 as shown in Fig. 6.2, where the shape of the  $m_T^i$  variable is shown for different SM Higgs  
 2728 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.

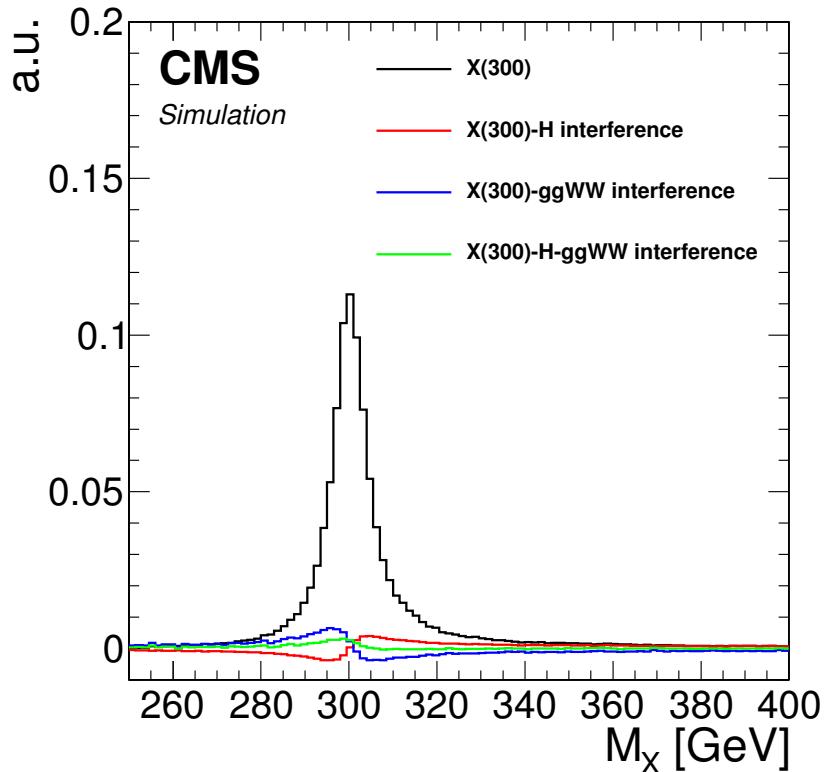
In order to test different resonance decay widths hypotheses, the signal samples, which are generated with a decay width corresponding to the expected value for a SM Higgs boson at that mass ( $\Gamma_{\text{SM}}$ ), are reweighted to obtain the desired width value ( $\Gamma'$ ). In particular the following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

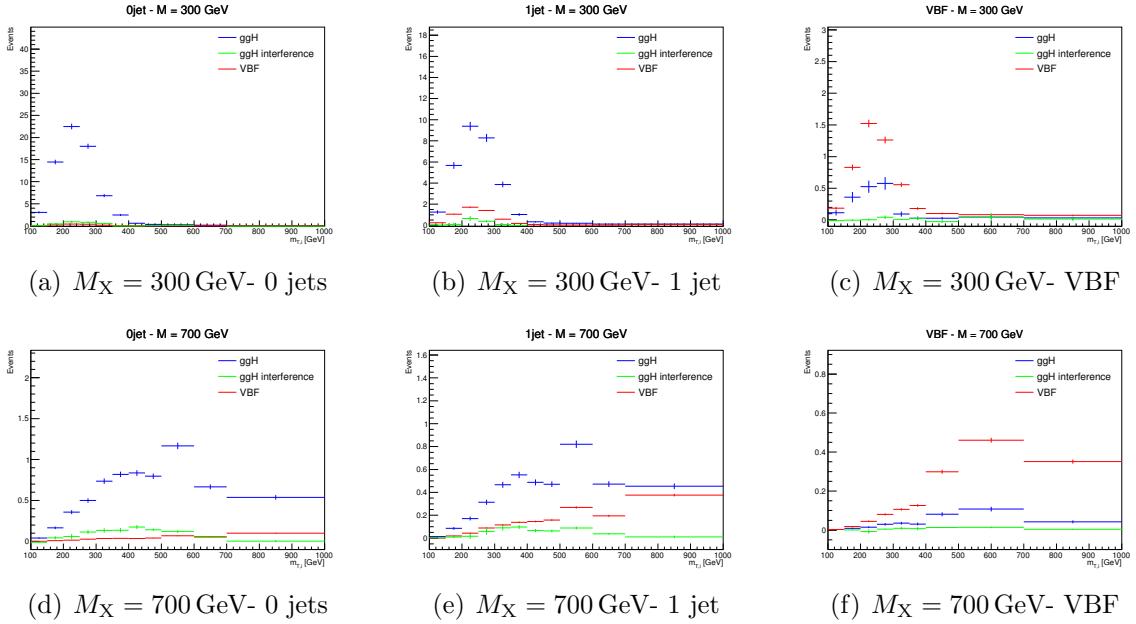
When a resonance with a non negligible width is considered, it is important to take into account the interference effects both with the  $gg \rightarrow WW$  background and the SM Higgs boson off-shell tail. A study of the interference effects for a resonance  $X$  produced through the gluon fusion mechanism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

including NNLO corrections for cross section using HNNLO program [122] based on MCFM. The matrix element package MELA supports all of these processes and allows fast MC re-weighting and optimal discriminant calculation. The basic idea of this approach is to compute the matrix elements of the processes under study with the MCFM and JHUGEN generators, including the interference terms, and using these matrix elements to compute an event weight used to reweight the simulated samples. Using this approach the simulated events can be reweighted according to different scenarios, for instance including some or all the interference terms, allowing a detailed study of the interference contribution. The effect of the various interference terms for the  $M_X$  variable at generator level is shown in Fig. 6.3, after having applied the WW baseline selections. As can be observed the contribution of the interference of the scalar resonance with the  $gg \rightarrow WW$  background and with SM Higgs boson have opposite sign and partially cancel out. This cancellation effect is different for different resonance masses and depends on the event selection. In particular the interference term with the SM Higgs off-shell tail is positive for values below  $M_X$  while it turns negative above  $M_X$ . The contribution of the interference with the  $gg \rightarrow WW$  background is instead characterized by an opposite sign lineshape, thus leading to a partial cancellation when 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.

2769 The effect of the resulting interference contribution including all the different terms is  
 2770 shown in Fig. 6.4 for the  $m_T^i$  signal templates, in the three categories separately and for  
 2771 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.

2772 The interference contribution is thus not negligible, especially for large values of  $M_X$ ,  
 2773 and is included in the analysis as part of the signal contribution. More specifically, during  
 2774 the fit procedure the signal yield is scaled by the signal strength parameter  $\mu$  (which is the  
 2775 parameter of interest of the fit), while the interference yield is scaled by  $\sqrt{\mu}$ .

## 2776 6.4. Background estimation

2777 The background processes affecting the analysis phase space are the same as the ones  
 2778 contributing to the SM Higgs search described in Sec. 5.4. The techniques used for the  
 2779 background estimation are the same as well.

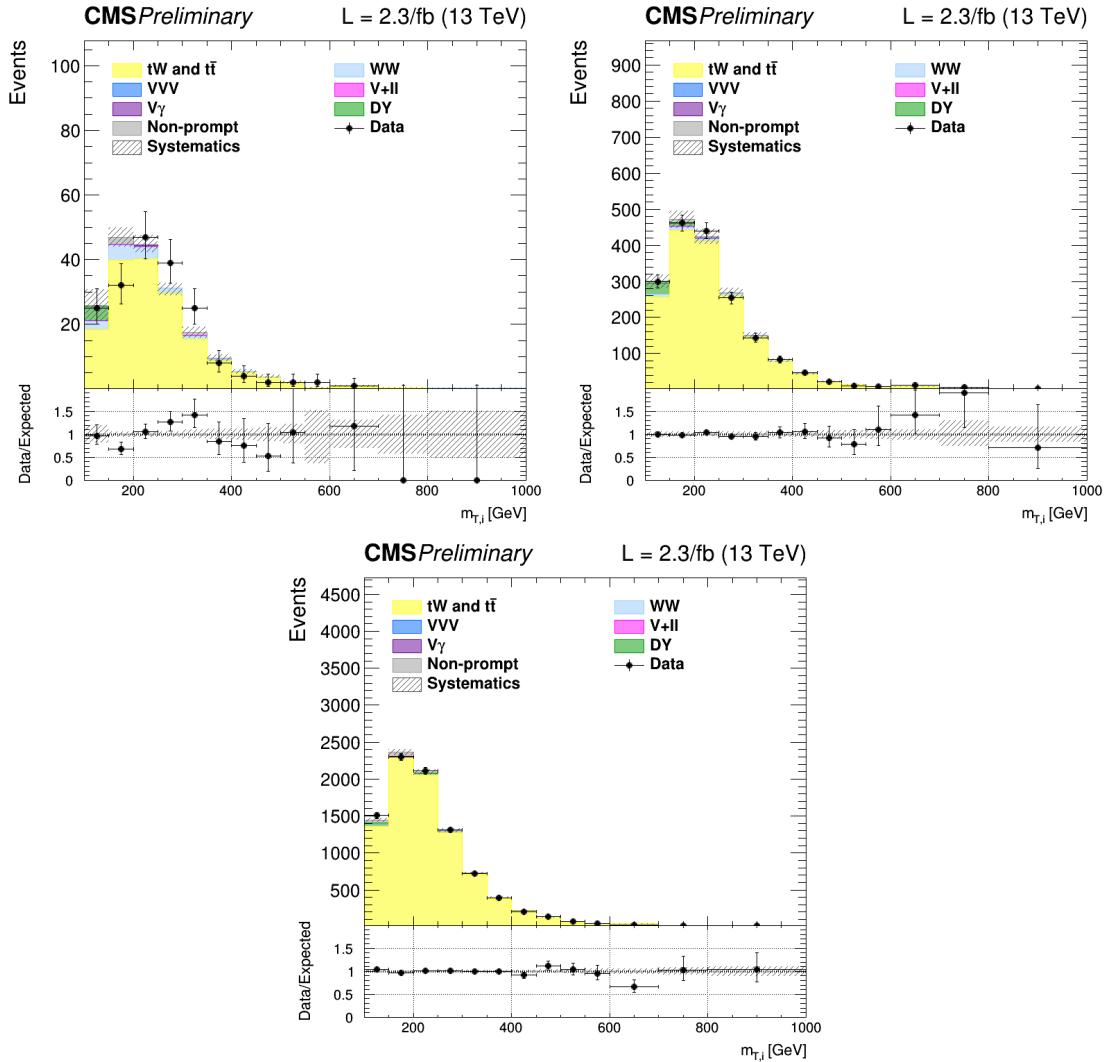
2780 The most relevant difference is the addition of the 2 jets category. The WW, top  
 2781 quark and DY background normalizations are estimated in this category using data driven  
 2782 techniques, similarly to the other jet bins.

2783 Given the slightly different WW baseline selection with respect to the SM Higgs search,  
 2784 also the control regions for the top quark and DY backgrounds estimation change, while the  
 2785 WW background normalization is estimated from data in the three signal regions separately,  
 2786 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:

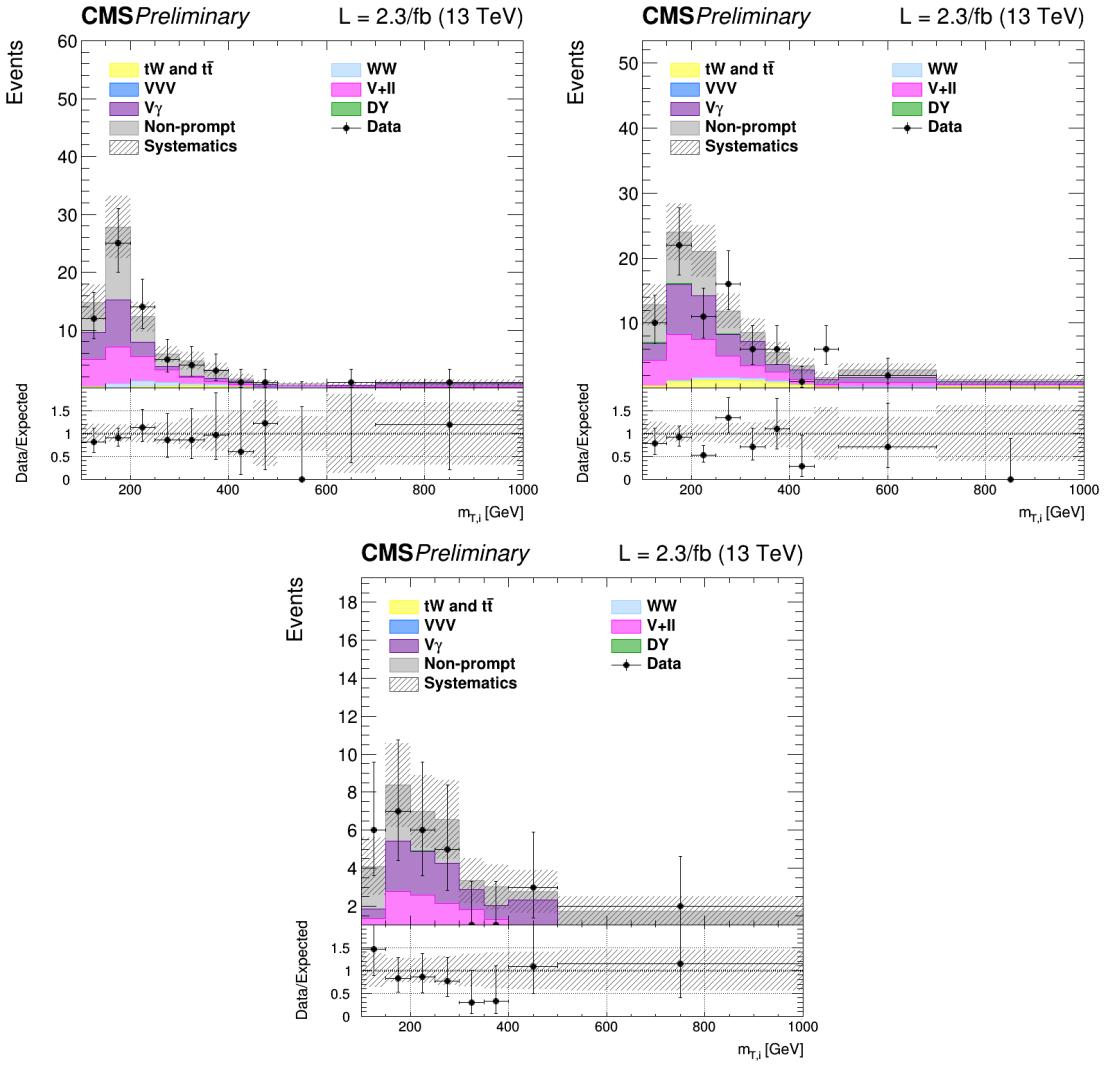
- 0 jets category: at least one b-tagged jet with  $20 < p_T < 30$  GeV is required;
- 1 jet category: exactly one b-tagged jet with  $p_T$  above 30 GeV is required;
- 2 jets category: at least one b-tagged jet with  $p_T$  above 30 GeV is required.

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. 6.5.



**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.

2795     The jet induced background, also labelled as “non-prompt” background, so as to highlight  
 2796     that these events do not contain prompt leptons, is estimated using the same fake rate  
 2797     method described in 4.4.3. A crosscheck is performed selecting events passing the WW  
 2798     baseline selection but with an  $e\mu$  pair with same charge. The  $m_T^i$  distributions for this  
 2799     phase space are shown in Fig. 6.6 for the three jet categories separately, showing agreement  
 2800     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.

2801     Due to the cuts on the leptons  $p_T$  and on  $m_{\ell\ell}$  in the WW baseline requirements, the  
 2802     contribution of the DY background decaying to a pair of  $\tau$  leptons is very small in the  
 2803     signal regions, especially in the VBF phase space. The normalization of this background is  
 2804     estimated from a control region in data, defined in the same way as explained in 5.4.4, for

2805 the 0 and 1 jet categories. In the VBF category, given the very small number of expected  
 2806 events, the normalization of this background is taken from simulation.

2807 Other minor background processes are estimated as described in 5.4.5.

## 2808 6.5. Systematic uncertainties

2809 The systematic uncertainties affecting this analysis are the same discussed in Sec. 5.5. The  
 2810 differences with respect to the SM Higgs boson search are described below.

2811 The PDF and  $\alpha_s$  uncertainties on the signal cross sections are taken from the computa-  
 2812 tions performed by the LHC cross section working group [113], and are included for all  
 2813 the mass points. The value of these uncertainties depends on the resonance mass and vary  
 2814 from 3 and 5% for ggH and from 2 and 3% for VBF production modes. The PDFs and  $\alpha_s$   
 2815 uncertainties on the signal selection are evaluated for every resonance mass and are found  
 2816 to be less than 1% for both ggH and VBF.

2817 The theoretical uncertainties in the signal yields due to the jet categorization are  
 2818 evaluated for all the ggH signals following the prescription described in Sec. 4.5.3.

2819 An additional uncertainty on the modelling of the top pair background is derived from  
 2820 the observed discrepancy between data and POWHEG V2 plus PYTHIA 8.1 simulation  
 2821 on the top quark  $p_T$  spectrum [123], which is particularly important in the tail of the  $m_T^i$   
 2822 distribution. Another uncertainty affecting the  $m_T^i$  tail for the top quark background is the  
 2823 parton shower uncertainty. This is evaluated comparing the generator level  $m_T^i$  distributions  
 2824 corresponding to two different simulations of the  $t\bar{t}$  process: one obtained using PYTHIA  
 2825 8.1 for the showering and hadronization of the simulated events, and the other using  
 2826 HERWIG++. The difference between the two is used to extract a shape uncertainty, which  
 2827 is less than 1% for low  $m_T^i$  values and reaches about 6% in the  $m_T^i$  tail.

## 2828 6.6. Signal extraction and limit setting

2829 The signal yield, including both the ggH and VBF production modes, is extracted performing  
 2830 a combined fit of the three categories to the  $m_T^i$  simulation templates for backgrounds and  
 2831 signal, and is repeated for each resonance mass hypothesis. Moreover, fixed the mass of  
 2832 the resonance, the fit is performed again for the various hypotheses of the resonance decay  
 2833 width. A single signal strength  $\mu$  is extracted from the fit, which multiplies both the ggH  
 2834 and VBF contributions. In other words it is assumed that the ratio of the two production  
 2835 mechanism stays the same as the one predicted by the SM<sup>2</sup>.

2836 The background yields expected from simulation corresponding to the three jet categories  
 2837 and after the analysis event selection are shown in Table 6.1. The signal yields corresponding  
 2838 to a selection of mass points and assuming  $\Gamma' = \Gamma_{SM}$  are shown in Table 6.2.

---

<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

2839     The strategy for computing the exclusion limits is based on the modified frequentist  
 2840     approach, also referred to as  $CL_s$ , as described in [112]. The first step is to construct the  
 2841     likelihood function  $\mathcal{L}(data|\mu, \theta)$ :

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

2842     where  $data$  represents the experimental observation,  $s$  and  $b$  are the expected signal and  
 2843     background yields respectively and  $\theta$  is the full set of nuisance parameters constrained by  
 2844     the prior distribution functions  $p(\tilde{\theta}|\theta)$ . The default values for the nuisance parameters are  
 2845     labelled as  $\tilde{\theta}$ .

2846     For a binned shape analysis,  $Poisson(data|\mu \cdot s + b)$  is the product of the Poisson  
 2847     probabilities to observe  $n_i$  events in bin  $i$ :

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

2848     In order to test the compatibility of the data with the signal plus background (or the  
 2849     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%)

2850 likelihood ratio:

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

2851 where  $\hat{\theta}_\mu$  refers to the conditional maximum likelihood estimators of  $\theta$ , given the signal  
2852 strength  $\mu$ . The parameter estimators  $\hat{\mu}$  and  $\hat{\theta}$  correspond to the global maximum of the  
2853 likelihood. The  $0 \leq \hat{\mu}$  constraint is imposed to have a positive signal yield, e.g. background  
2854 underfluctuations are forbidden, while  $\hat{\mu} \leq \mu$  is imposed to have a one-sided confidence  
2855 interval. The observed test statistic for the signal strength  $\mu$  under test is referred to as  
2856  $\tilde{q}_\mu^{\text{obs}}$ . The values of the nuisance parameters obtained maximising the likelihood function are  
2857 labelled as  $\hat{\theta}_0^{\text{obs}}$  and  $\hat{\theta}_\mu^{\text{obs}}$  for the background only and signal plus background hypotheses,  
2858 respectively. The pdf of the test statistic is constructed by generating toy MC pseudo-data  
2859 for both the background only and signal plus background hypotheses, i.e.  $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}})$  and  
2860  $f(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}})$ . These distributions can be used to define two p-values corresponding to the

2861 two hypotheses,  $p_\mu$  and  $p_b$ :

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

2862

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

2863 According to these definitions,  $p_\mu$  and  $p_b$  can be identified with  $\text{CL}_{s+b}$  and  $1 - \text{CL}_b$ , 2864 respectively. The  $\text{CL}_s(\mu)$  is calculated using the following ratio:

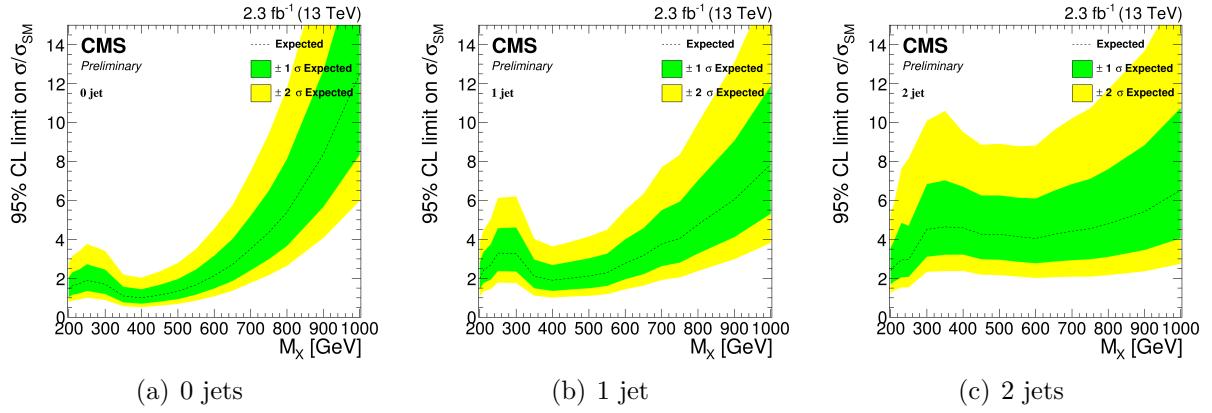
$$\text{CL}_s(\mu) = \frac{\text{CL}_{s+b}}{\text{CL}_b} = \frac{p_\mu}{1 - p_b} \quad . \quad (6.8)$$

2865 If, for a given signal strength  $\mu$ ,  $\text{CL}_s \leq \alpha$ , then the hypothesis is excluded with a  $(1 - \alpha)$  2866 confidence level (CL). For instance, if one wants to quote the upper limit on  $\mu$  with a 95% 2867 CL, the signal strength has to be adjusted until  $\text{CL}_s = 0.05$ .

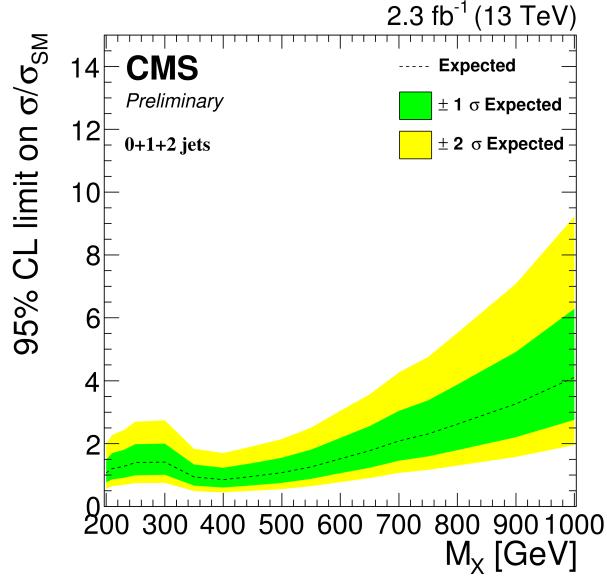
2868 The expected median upper limit, as well as the  $\pm 1\sigma$  (68% CL) and  $\pm 2\sigma$  (95% CL) 2869 bands, are determined generating a large amount of pseudo-data in the background only 2870 hypothesis and calculating  $\text{CL}_s$  and the 95% CL upper limit for each of them, as if they 2871 were real data. Then the cumulative distribution of the 95% CL upper limits is built and 2872 the median expected value is identified as the value at which the cumulative distribution 2873 crosses the 50% quantile. The  $\pm 1\sigma$  ( $\pm 2\sigma$ ) band is defined by the values at which the 2874 cumulative distribution crosses the 16% (2.5%) and 84% (97.5%) quantiles.

2875 In order to assess the sensitivity of the analysis, the expected upper exclusion limits at 2876 95% CL on the signal strength are shown in Fig. 6.7 for the three jet categories separately. 2877 For a given mass of the resonance, the limits are derived assuming a signal decay width 2878  $\Gamma' = \Gamma_{\text{SM}}$  and a cross section equal to the one expected from a SM Higgs boson at that 2879 mass. The other decay width hypothesis have also been tested, showing a very similar 2880 expected exclusion limit, suggesting that this analysis is not strongly sensitive to variations 2881 of the resonance decay width.

2882 The 0 jets category is the most sensitive especially in the low mass region, while for 2883 very large masses of the resonance the 1 jet and VBF categories start being important. 2884 This is explained mainly by the fact that the VBF contribution increases, with respect to 2885 ggH, as the mass increases. The expected exclusion limit on the signal strength after the 2886 combination of the three categories is shown in Fig. 6.8. Comparing the limits in the single 2887 categories with the combination of the three it is evident how the higher jet multiplicity 2888 categories help in improving the results for large values of  $M_X$ .



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

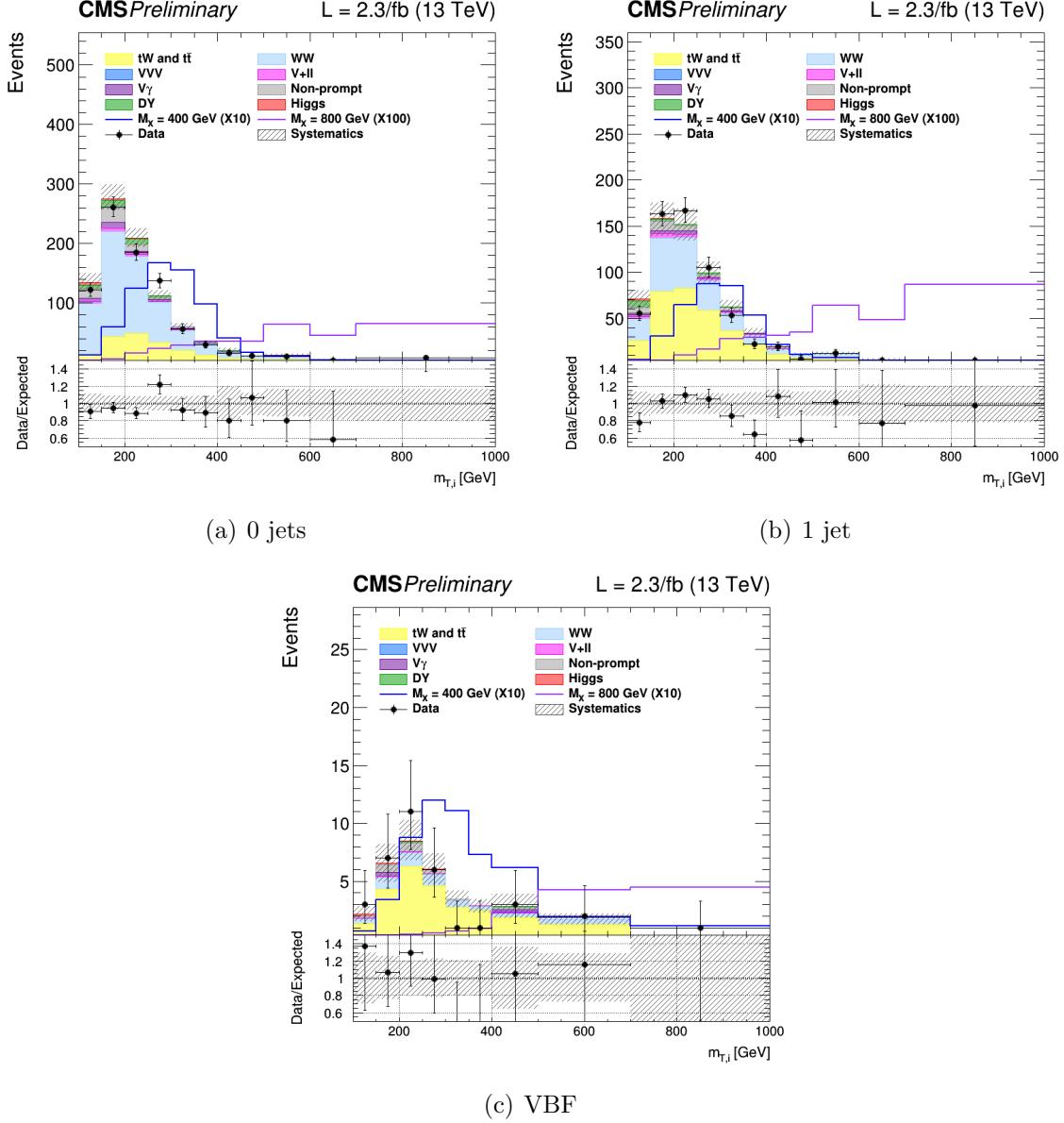


**Figure 6.8.:** Expected exclusion upper limit at 95% CL on the signal strength for the combination of the three categories, as a function of the resonance mass. The dashed line corresponds to median upper limit, while the green and yellow regions represent the  $\pm 1\sigma$  and  $\pm 2\sigma$  uncertainty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

## 2889 6.7. Results

2890 The  $m_T^i$  distributions for the signal region after the full analysis selection are shown  
 2891 in Fig. 6.9 for the three jet categories. Two different signal hypotheses corresponding

to  $M_X = 400$  GeV and  $M_X = 800$  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.

For every mass point from 200 GeV up to 1 TeV the observed p-value and the 95% CL upper exclusion limit are calculated for five hypothesis of the signal width. The observed

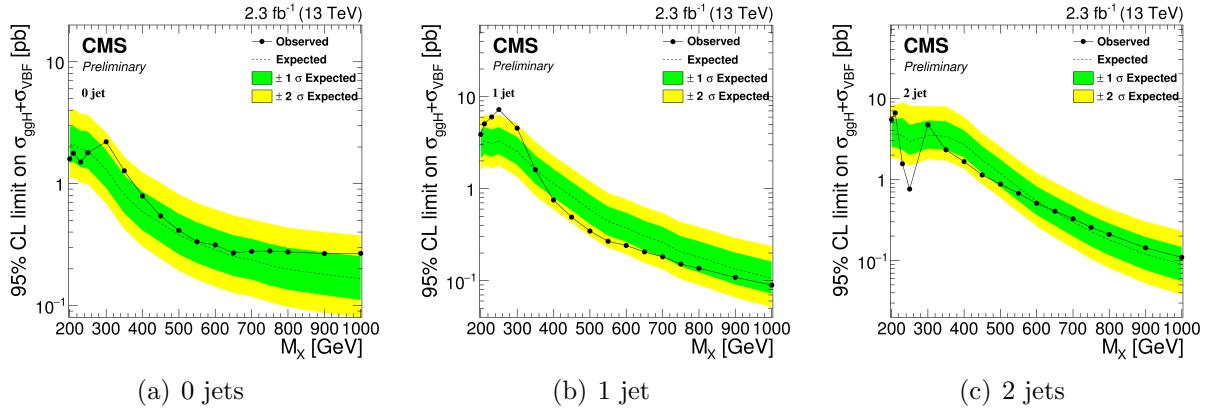
<sup>2896</sup> p-value as a function of the resonance mass for the combination of the three jet categories  
<sup>2897</sup> is shown in Table 6.3.

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

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

<sup>2898</sup> In order to be independent on the particular model assumed for the signal cross section,  
<sup>2899</sup> the results are interpreted as exclusion limits on  $\sigma \times \mathcal{B}$ , where  $\sigma$  stands for the sum of  
<sup>2900</sup> the ggH and VBF cross sections, and  $\mathcal{B}$  represents the  $X \rightarrow WW \rightarrow 2\ell 2\nu$  branching ratio  
<sup>2901</sup> including all lepton flavours. The expected and observed upper exclusion limits on  $\sigma \times \mathcal{B}$   
<sup>2902</sup> for  $\Gamma' = \Gamma_{SM}$  are shown in Fig. 6.10.

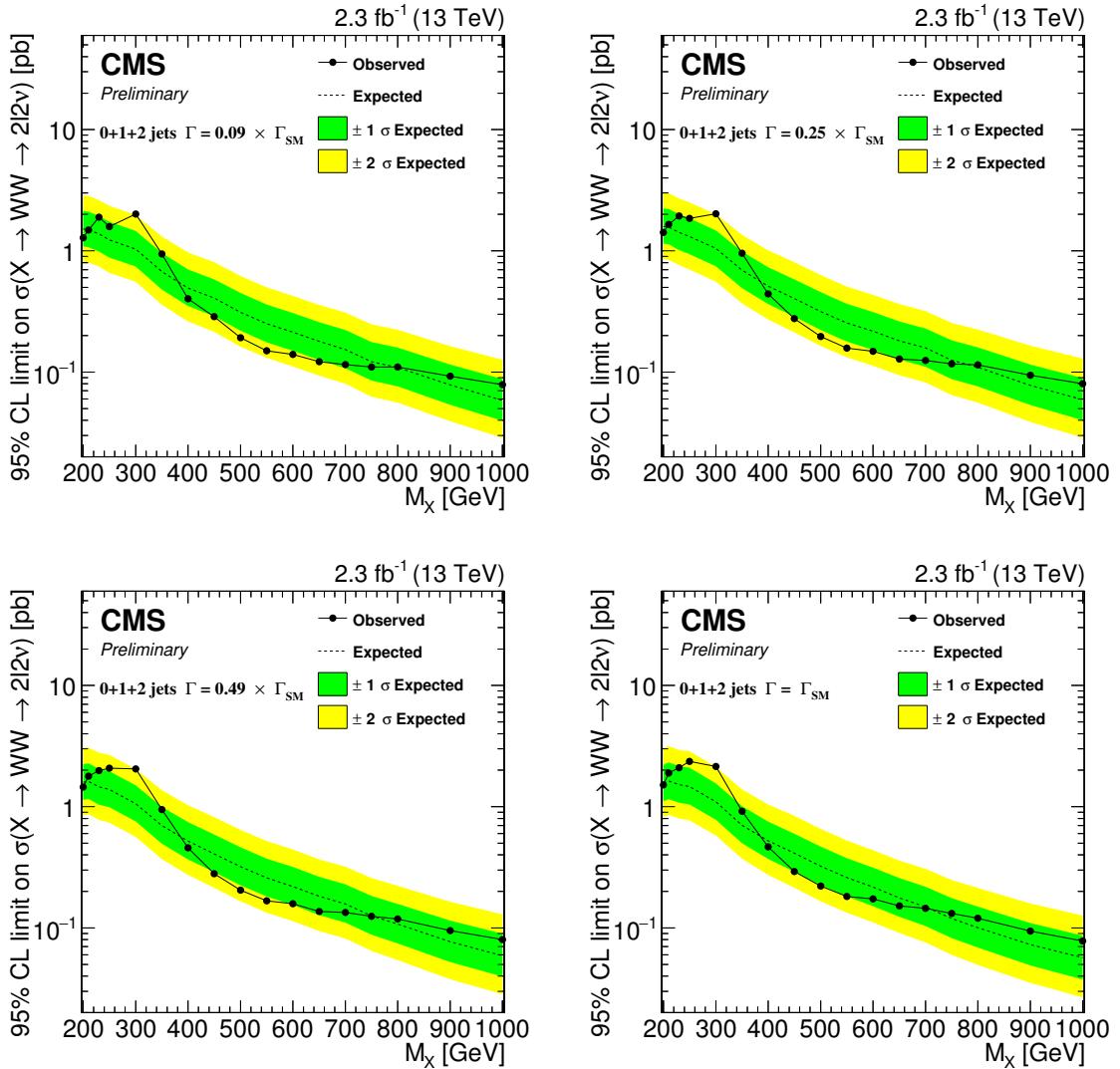
<sup>2903</sup> A mild excess is observed in the 0 jets category and, more evident, in the 1 jet category  
<sup>2904</sup> around 250-300 GeV. A deficit is instead in the VBF category around 250 GeV, which is  
<sup>2905</sup> mainly due to an underfluctuation of the background. This effect can be understood looking  
<sup>2906</sup> at the VBF shape in Fig. 6.9, where two adjacent data points, corresponding to the fifth



**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.

and sixth bins, clearly underfluctuate with respect to the background prediction, causing the dip in the observed limit.

The exclusion limit resulting from the combination of the three categories is shown in Fig. 6.11, for the five  $\Gamma'$  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. 6.11 is thus excluded with a 95% CL for masses ranging from 200 GeV up to 1 TeV.



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



# **Chapter 7.**

## **Conclusions**

2914



# Appendix A.

## <sup>2915</sup> **Fiducial region definition and <sup>2916</sup> optimization**

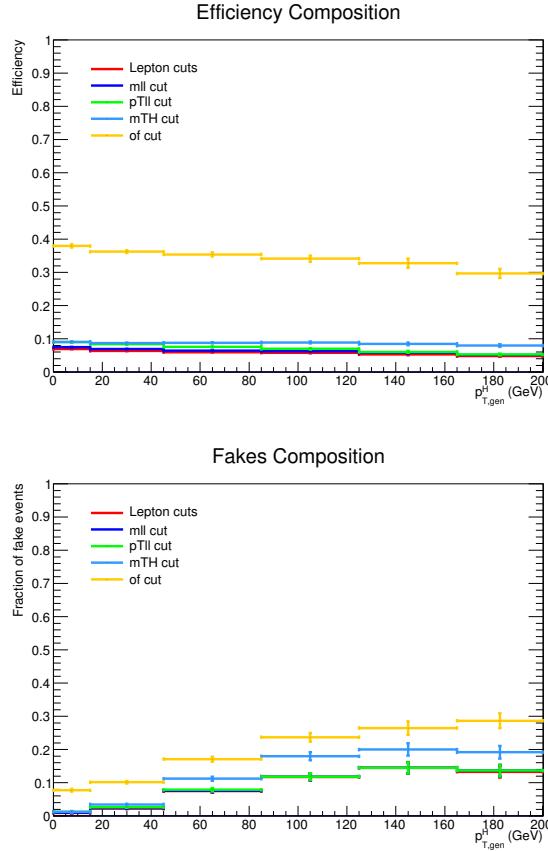
<sup>2917</sup> The fiducial region must be chosen in such a way to be as close as possible to the selections  
<sup>2918</sup> applied in the analysis, in order to reduce the model dependence in the extrapolation step.  
<sup>2919</sup> That means that for optimizing the fiducial volume definition, the efficiency has to be  
<sup>2920</sup> maximized. Another parameter entering the game is the number of fake events, in other  
<sup>2921</sup> words the number of reconstructed events which do not belong to the fiducial phase space.  
<sup>2922</sup> This parameter should instead be as small as possible. Even if we have to observe the trend  
<sup>2923</sup> of these two quantities as a function of  $p_T^H$ , we can maximize the ratio between the overall  
<sup>2924</sup> efficiency and the overall fake rate as a proxy for establishing the “goodness” of the fiducial  
<sup>2925</sup> region.

<sup>2926</sup> Several different fiducial region definitions were tested and the results show that:

- <sup>2927</sup> • **of cut:** The fiducial region definition must include only the opposite flavor combination  
<sup>2928</sup> including one electron and one muon. If we include also the combinations involving  
<sup>2929</sup>  $\tau$ 's the efficiency falls down.
- <sup>2930</sup> • **Lepton cut:** Since the resolution on lepton transverse momentum is good, there is no  
<sup>2931</sup> need to loosen the cuts related these variables, i.e. we can use the same cuts defined  
<sup>2932</sup> in the analysis selection ( $p_T^{\ell,1} > 20 \text{ GeV}$ ,  $p_T^{\ell,2} > 10 \text{ GeV}$ ).
- <sup>2933</sup> • **Di-lepton  $p_T$  cut:** As stated in the previous point, there is no need to loosen this  
<sup>2934</sup> cut, so we kept the same value as the analysis selection, i.e.  $p_T^{\ell\ell} > 30 \text{ GeV}$ .
- <sup>2935</sup> • **Di-lepton mass cut:**  $m_{\ell\ell} > 12 \text{ GeV}$  as discussed before.
- <sup>2936</sup> • **neutrino pair  $p_T$  cut:** Since the resolution on the measurement of the missing  
<sup>2937</sup> transverse energy is poor, the neutrino pair cut should not be included in the definition  
<sup>2938</sup> of the fiducial region, because it would increase the fake rate without increasing the  
<sup>2939</sup> efficiency, thus resulting in a lower ratio between overall efficiency and fake rate.
- <sup>2940</sup> •  **$m_T$  cut:** Also the  $m_T$  cut that we have in the analysis selection, i.e.  $m_T > 60 \text{ GeV}$ ,  
<sup>2941</sup> should be loosened or removed because it involves neutrinos and then increase the  
<sup>2942</sup> fake rate. We decided eventually to keep this cut, loosening it to 50 GeV, because in  
<sup>2943</sup> addition to increase the number of fake events, it increases the efficiency as well.

<sup>2944</sup> The fake rate and the efficiency as a function of  $p_T^H$  after the optimization discussed  
<sup>2945</sup> before are shown in figure A.1. To obtain these plots the fiducial region was modified

2946 adding in sequence the various cuts and computing the efficiency and the fake rate each  
 2947 time. In that way we can asses the composition of those distributions.

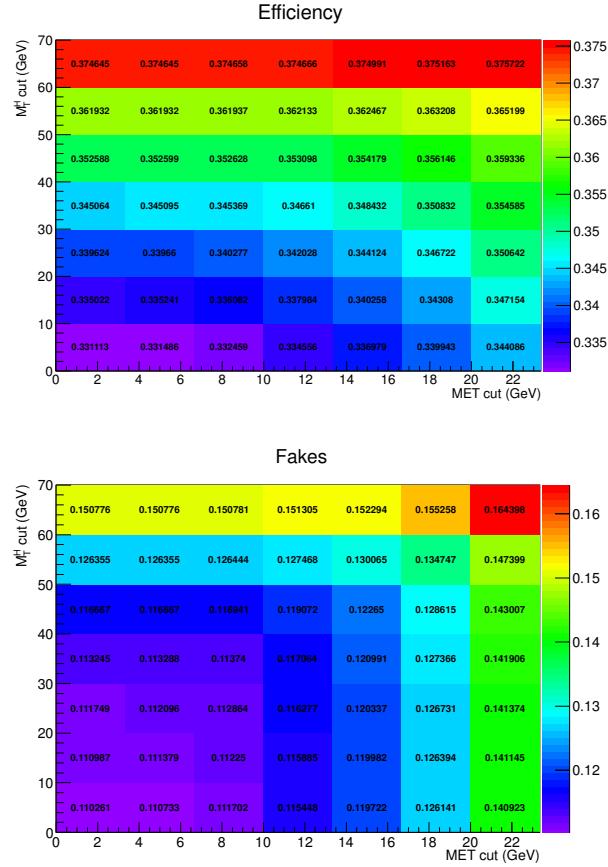


**Figure A.1.:** Efficiency and fake rate as a function on Higgs transverse momentum. The plots correspond to the optimized fiducial region definition and show the effect of adding each of the mentioned cuts in sequence.

2948 The efficiency and fraction of fake events have been measured also as a function of the  
 2949  $E_T^{\text{miss}}$  and  $m_T$  cuts in the fiducial region. Since these two variables are correlated, the  
 2950 results are reported as two-dimensional histograms. In Fig. A.2 are reported the efficiency  
 2951 and fraction of fake events for these two variables.

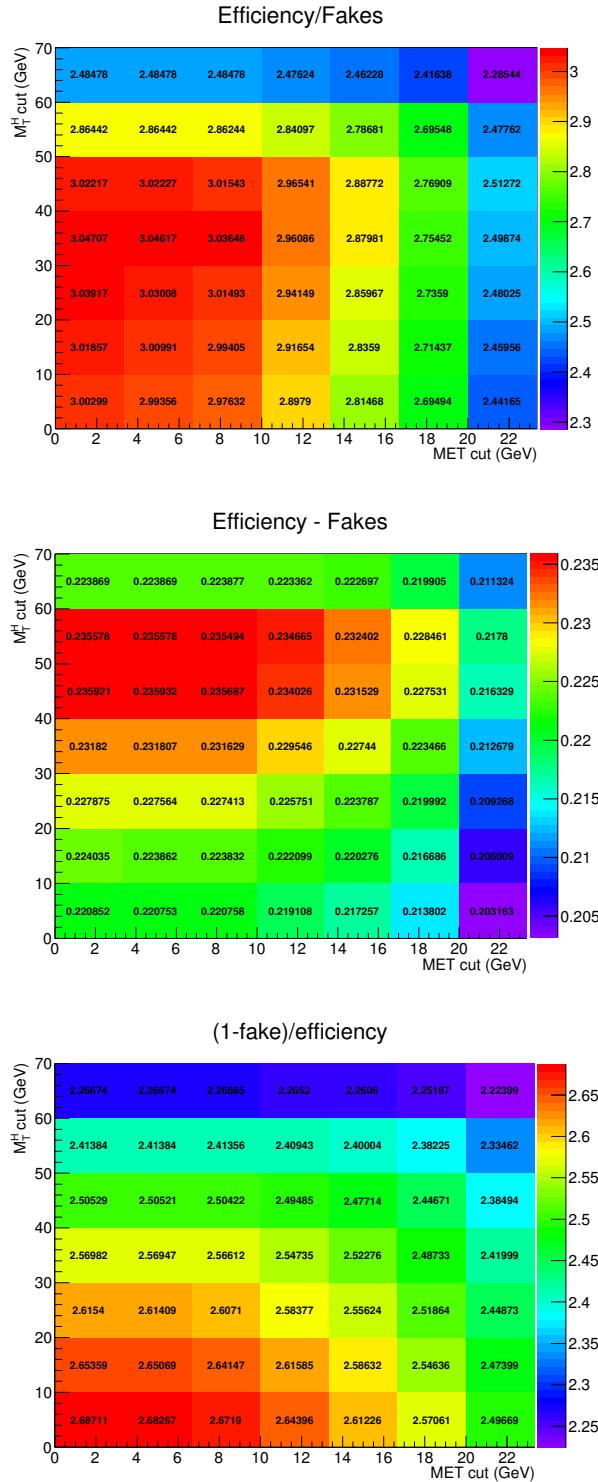
2952 The criterion adopted to define the fiducial region is a tradeoff between having a large  
 2953 efficiency and a small fraction of fake events. Especially when looking at the low resolution  
 2954 variables, such as  $E_T^{\text{miss}}$  and  $m_T$ , a suitable figure of merit has to be chosen for the estimation  
 2955 of the best cuts. Several different figures of merit have been checked, such as  $\epsilon/f$ ,  $\epsilon - f$   
 2956 and  $(1 - f)/\epsilon$ . The results for these three different figures of merit are shown in Fig. A.3 as  
 2957 a function of the  $E_T^{\text{miss}}$  and  $m_T$  cuts in the fiducial region.

2958 Following the same criterion, similar plots as above have been obtained for an alternative  
 2959 model, given by varying up the ggH/VBF ratio within the experimental uncertainties. The

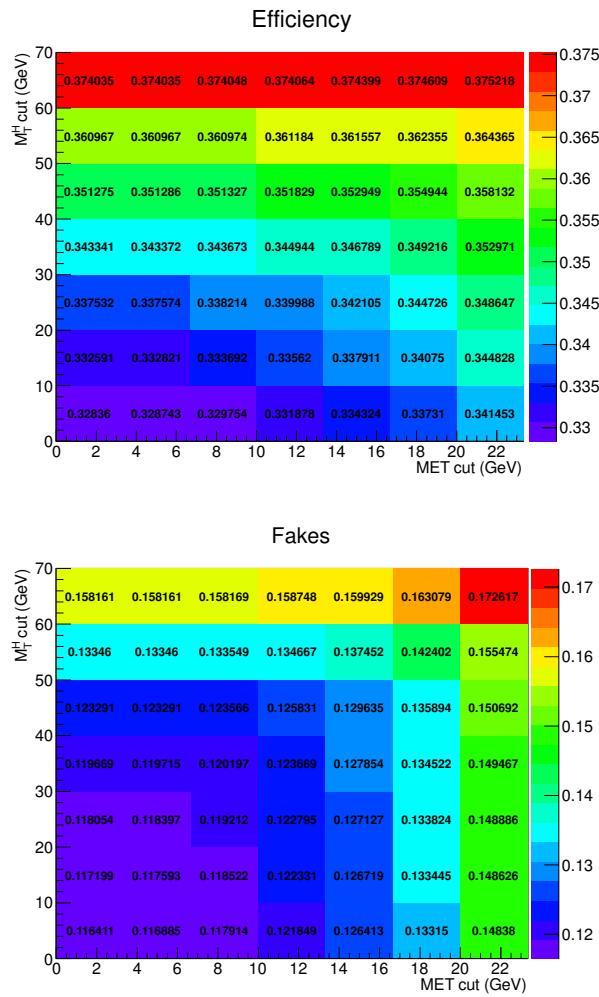


**Figure A.2.:** Efficiency and fake rate as a function of  $E_T^{\text{miss}}$  and  $m_T$  cuts in the fiducial region.

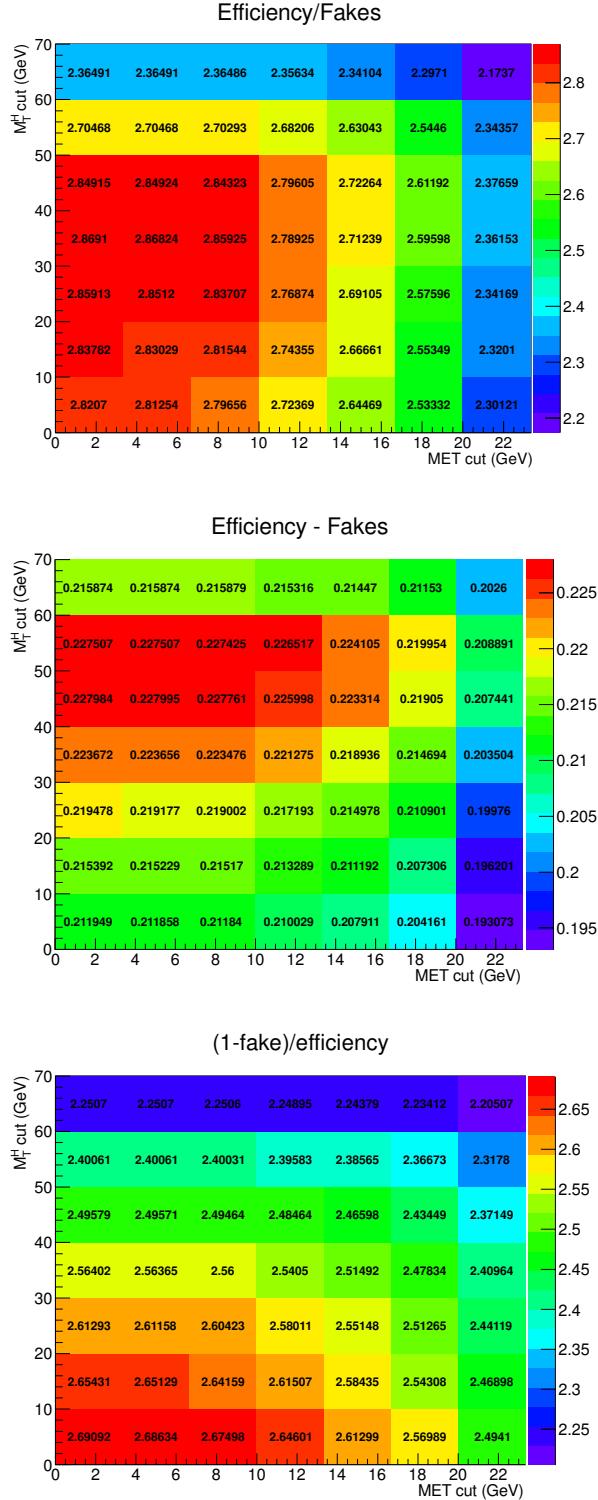
2960 results, shown in Fig. A.4 and Fig. A.5, show a similar trend with respect to the model  
 2961 with nominal ggH/VBF ratio.



**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.





2963

# Bibliography

- 2964 [1] Thomas Sven Pettersson and P Lefèvre. *The Large Hadron Collider: conceptual*  
2965 *design*. Tech. rep. CERN-AC-95-05-LHC. Oct. 1995. URL: <http://cds.cern.ch/record/291782>.
- 2966
- 2967 [2] Oliver Sim Brüning et al. *LHC Design Report*. Geneva: CERN, 2004. URL: <http://cds.cern.ch/record/782076>.
- 2968
- 2969 [3] Oliver Sim Brüning et al. *LHC Design Report*. Geneva: CERN, 2004. URL: <http://cds.cern.ch/record/815187>.
- 2970
- 2971 [4] Michael Benedikt et al. *LHC Design Report*. Geneva: CERN, 2004. URL: <http://cds.cern.ch/record/823808>.
- 2972
- 2973 [5] *LEP design report*. By the LEP Injector Study Group. Geneva: CERN, 1983. URL:  
2974 <http://cds.cern.ch/record/98881>.
- 2975 [6] *LEP design report*. Copies shelved as reports in LEP, PS and SPS libraries. Geneva:  
2976 CERN, 1984. URL: <http://cds.cern.ch/record/102083>.
- 2977 [7] Carlo Wyss. *LEP design report, v.3: LEP2*. Vol. 1-2 publ. in 1983-84. Geneva: CERN,  
2978 1996. URL: <http://cds.cern.ch/record/314187>.
- 2979 [8] S. Chatrchyan et al. “The CMS experiment at the CERN LHC”. In: *JINST* 3 (2008),  
2980 S08004. DOI: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- 2981 [9] G. Aad et al. “The ATLAS Experiment at the CERN Large Hadron Collider”. In:  
2982 *JINST* 3 (2008), S08003. DOI: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- 2983 [10] A. Augusto Alves Jr. et al. “The LHCb Detector at the LHC”. In: *JINST* 3 (2008),  
2984 S08005. DOI: [10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005).
- 2985 [11] K. Aamodt et al. “The ALICE experiment at the CERN LHC”. In: *JINST* 3 (2008),  
2986 S08002. DOI: [10.1088/1748-0221/3/08/S08002](https://doi.org/10.1088/1748-0221/3/08/S08002).
- 2987 [12] G. Anelli et al. “The TOTEM experiment at the CERN Large Hadron Collider”. In:  
2988 *JINST* 3 (2008), S08007. DOI: [10.1088/1748-0221/3/08/S08007](https://doi.org/10.1088/1748-0221/3/08/S08007).
- 2989 [13] O. Adriani et al. “The LHCf detector at the CERN Large Hadron Collider”. In:  
2990 *JINST* 3 (2008), S08006. DOI: [10.1088/1748-0221/3/08/S08006](https://doi.org/10.1088/1748-0221/3/08/S08006).
- 2991 [14] *The CMS magnet project: Technical Design Report*. Technical Design Report CMS.  
2992 Geneva: CERN, 1997. URL: <http://cds.cern.ch/record/331056>.
- 2993 [15] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical  
2994 Design Report CMS. Geneva: CERN, 1997. URL: <http://cds.cern.ch/record/349375>.
- 2995

- 2996 [16] Philippe Bloch et al. *Changes to CMS ECAL electronics: addendum to the Technical*  
2997 *Design Report*. Technical Design Report CMS. Geneva: CERN, 2002. URL: <http://cds.cern.ch/record/581342>.
- 2999 [17] *The CMS hadron calorimeter project: Technical Design Report*. Technical Design  
3000 Report CMS. Geneva: CERN, 1997. URL: <http://cds.cern.ch/record/357153>.
- 3001 [18] *The CMS muon project: Technical Design Report*. Technical Design Report CMS.  
3002 Geneva: CERN, 1997. URL: <http://cds.cern.ch/record/343814>.
- 3003 [19] S. Dasu et al. “CMS. The TriDAS project. Technical design report, vol. 1: The  
3004 trigger systems”. In: (2000).
- 3005 [20] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical*  
3006 *Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger*  
3007 *and data-acquisition project*. Technical Design Report CMS. Geneva: CERN, 2002.  
3008 URL: <http://cds.cern.ch/record/578006>.
- 3009 [21] CMS Collaboration. *Particle-Flow Event Reconstruction in CMS and Performance*  
3010 *for Jets, Taus, and  $E_T^{\text{miss}}$* . CMS Physics Analysis Summary CMS-PAS-PFT-09-001.  
3011 2009. URL: <http://cdsweb.cern.ch/record/1194487>.
- 3012 [22] Serguei Chatrchyan et al. “Description and performance of track and primary-  
3013 vertex reconstruction with the CMS tracker”. In: *JINST* 9.10 (2014), P10009. DOI:  
3014 [10.1088/1748-0221/9/10/P10009](https://doi.org/10.1088/1748-0221/9/10/P10009). arXiv: [1405.6569](https://arxiv.org/abs/1405.6569) [physics.ins-det].
- 3015 [23] Pierre Billoir and S. Qian. “Simultaneous pattern recognition and track fitting by  
3016 the Kalman filtering method”. In: *Nucl. Instrum. Meth. A* 294 (1990), p. 219. DOI:  
3017 [10.1016/0168-9002\(90\)91835-Y](https://doi.org/10.1016/0168-9002(90)91835-Y).
- 3018 [24] Wolfgang Adam et al. *Reconstruction of Electrons with the Gaussian-Sum Filter*  
3019 *in the CMS Tracker at the LHC*. Tech. rep. CMS-NOTE-2005-001. Geneva: CERN,  
3020 Jan. 2005. URL: <https://cds.cern.ch/record/815410>.
- 3021 [25] Serguei Chatrchyan et al. “Performance of CMS muon reconstruction in  $pp$  collision  
3022 events at  $\sqrt{s} = 7$  TeV”. In: *JINST* 7 (2012), P10002. DOI: [10.1088/1748-0221/7/10/P10002](https://doi.org/10.1088/1748-0221/7/10/P10002). arXiv: [1206.4071](https://arxiv.org/abs/1206.4071) [physics.ins-det].
- 3024 [26] *Electron reconstruction and identification at  $\sqrt{s} = 7$  TeV*. Tech. rep. CMS-PAS-  
3025 EGM-10-004. Geneva: CERN, 2010. URL: <http://cds.cern.ch/record/1299116>.
- 3026 [27] Vardan Khachatryan et al. “Performance of Electron Reconstruction and Selection  
3027 with the CMS Detector in Proton-Proton Collisions at  $\sqrt{s} = 8$  TeV”. In: *JINST*  
3028 10.06 (2015), P06005. DOI: [10.1088/1748-0221/10/06/P06005](https://doi.org/10.1088/1748-0221/10/06/P06005). arXiv: [1502.02701](https://arxiv.org/abs/1502.02701)  
3029 [physics.ins-det].
- 3030 [28] Gavin P. Salam. “Towards Jetography”. In: *Eur. Phys. J.* C67 (2010), pp. 637–686.  
3031 DOI: [10.1140/epjc/s10052-010-1314-6](https://doi.org/10.1140/epjc/s10052-010-1314-6). arXiv: [0906.1833](https://arxiv.org/abs/0906.1833) [hep-ph].

- 3032 [29] Gerald C. Blazey et al. “Run II jet physics”. In: *QCD and weak boson physics in*  
3033 *Run II. Proceedings, Batavia, USA, March 4-6, June 3-4, November 4-6, 1999.* 2000,  
3034 pp. 47–77. arXiv: hep-ex/0005012 [hep-ex]. URL: [http://lss.fnal.gov/cgi-bin/find\\_paper.pl?conf=00-092](http://lss.fnal.gov/cgi-bin/find_paper.pl?conf=00-092).
- 3035  
3036 [30] Serguei Chatrchyan et al. “Determination of jet energy calibration and transverse  
3037 momentum resolution in CMS”. In: *JINST* 6 (2011), P11002. DOI: 10.1088/1748-  
3038 0221/6/11/P11002. arXiv: 1107.4277 [physics.ins-det].
- 3039 [31] B. Abbott et al. “Determination of the absolute jet energy scale in the D0 calorimeters”. In: *Nucl. Instrum. Meth.* A424 (1999), pp. 352–394. DOI: 10.1016/S0168-  
3040 9002(98)01368-0. arXiv: hep-ex/9805009 [hep-ex].
- 3041  
3042 [32] Wolfgang Waltenberger. *Adaptive Vertex Reconstruction*. Tech. rep. CMS-NOTE-  
3043 2008-033. Geneva: CERN, July 2008. URL: <https://cds.cern.ch/record/1166320>.
- 3044 [33] *MET performance in 8 TeV data*. Tech. rep. CMS-PAS-JME-12-002. Geneva: CERN,  
3045 2013. URL: <https://cds.cern.ch/record/1543527>.
- 3046 [34] *Performance of missing energy reconstruction in 13 TeV pp collision data using*  
3047 *the CMS detector*. Tech. rep. CMS-PAS-JME-16-004. Geneva: CERN, 2016. URL:  
3048 <http://cds.cern.ch/record/2205284>.
- 3049 [35] M. Spira et al. “Higgs boson production at the LHC”. In: *Nucl. Phys.* B 453  
3050 (1995), pp. 17–82. DOI: 10.1016/0550-3213(95)00379-7. arXiv: hep-ph/9504378  
3051 [hep-ph].
- 3052 [36] Robert Harlander and Philipp Kant. “Higgs production and decay: Analytic results  
3053 at next-to-leading order QCD”. In: *JHEP* 12 (2005), p. 015. DOI: 10.1088/1126-  
3054 6708/2005/12/015. arXiv: hep-ph/0509189 [hep-ph].
- 3055 [37] V. Ravindran, J. Smith, and W. L. van Neerven. “NNLO corrections to the total  
3056 cross-section for Higgs boson production in hadron hadron collisions”. In: *Nucl.*  
3057 *Phys.* B 665 (2003), pp. 325–366. DOI: 10.1016/S0550-3213(03)00457-7. arXiv:  
3058 hep-ph/0302135 [hep-ph].
- 3059 [38] Stefano Catani and Massimiliano Grazzini. “An NNLO subtraction formalism in  
3060 hadron collisions and its application to Higgs boson production at the LHC”. In:  
3061 *Phys. Rev. Lett.* 98 (2007), p. 222002. DOI: 10.1103/PhysRevLett.98.222002.  
3062 arXiv: hep-ph/0703012 [hep-ph].
- 3063 [39] Charalampos Anastasiou et al. “Higgs Boson Gluon-Fusion Production in QCD at  
3064 Three Loops”. In: *Phys. Rev. Lett.* 114 (2015), p. 212001. DOI: 10.1103/PhysRevLett.  
3065 114.212001. arXiv: 1503.06056 [hep-ph].
- 3066 [40] X. Chen et al. “NNLO QCD corrections to Higgs boson production at large transverse  
3067 momentum”. In: (2016). arXiv: 1607.08817 [hep-ph].
- 3068 [41] Giuseppe Bozzi et al. “Transverse-momentum resummation and the spectrum of the  
3069 Higgs boson at the LHC”. In: *Nucl. Phys. B* 737 (2006), p. 73. DOI: 10.1016/j.  
3070 nuclphysb.2005.12.022. arXiv: hep-ph/0508068 [hep-ph].

- 3071 [42] D. de Florian et al. “Higgs boson production at the LHC: transverse momentum  
 3072 resummation effects in the  $H \rightarrow 2\gamma$ ,  $H \rightarrow WW \rightarrow l\nu l\nu$  and  $H \rightarrow ZZ \rightarrow 4l$  decay  
 3073 modes”. In: *JHEP* 06 (2012), p. 132. DOI: 10.1007/JHEP06(2012)132. arXiv:  
 3074 1203.6321 [hep-ph].
- 3075 [43] Massimiliano Grazzini and Hayk Sargsyan. “Heavy-quark mass effects in Higgs boson  
 3076 production at the LHC”. In: *JHEP* 09 (2013), p. 129. DOI: 10.1007/JHEP09(2013)  
 3077 129. arXiv: 1306.4581 [hep-ph].
- 3078 [44] Aleksandr Azatov and Ayan Paul. “Probing Higgs couplings with high  $p_T$  Higgs  
 3079 production”. In: *JHEP* 01 (2014), p. 014. DOI: 10.1007/JHEP01(2014)014. arXiv:  
 3080 1309.5273 [hep-ph].
- 3081 [45] Robert V. Harlander and Tobias Neumann. “Probing the nature of the Higgs-gluon  
 3082 coupling”. In: *Phys. Rev. D* 88 (2013), p. 074015. DOI: 10.1103/PhysRevD.88.  
 3083 074015. arXiv: 1308.2225 [hep-ph].
- 3084 [46] David Marzocca, Marco Serone, and Jing Shu. “General Composite Higgs Models”.  
 3085 In: *JHEP* 08 (2012), p. 013. DOI: 10.1007/JHEP08(2012)013. arXiv: 1205.0770  
 3086 [hep-ph].
- 3087 [47] Andrea Banfi, Adam Martin, and Veronica Sanz. “Probing top-partners in Higgs+jets”.  
 3088 In: *JHEP* 08 (2014), p. 053. DOI: 10.1007/JHEP08(2014)053. arXiv: 1308.4771  
 3089 [hep-ph].
- 3090 [48] Georges Aad et al. “Fiducial and differential cross sections of Higgs boson production  
 3091 measured in the four-lepton decay channel in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the  
 3092 ATLAS detector”. In: *Phys. Lett. B* 738 (2014), p. 234. DOI: 10.1016/j.physletb.  
 3093 2014.09.054. arXiv: 1408.3226 [hep-ex].
- 3094 [49] Georges Aad et al. “Measurements of fiducial and differential cross sections for Higgs  
 3095 boson production in the diphoton decay channel at  $\sqrt{s} = 8$  TeV with ATLAS”.  
 3096 In: *JHEP* 09 (2014), p. 112. DOI: 10.1007/JHEP09(2014)112. arXiv: 1407.4222  
 3097 [hep-ex].
- 3098 [50] Georges Aad et al. “Measurements of the Total and Differential Higgs Boson Production  
 3099 Cross Sections Combining the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4\ell$  Decay Channels at  
 3100  $\sqrt{s} = 8$  TeV with the ATLAS Detector”. In: *Phys. Rev. Lett.* 115 (2015), p. 091801.  
 3101 DOI: 10.1103/PhysRevLett.115.091801. arXiv: 1504.05833 [hep-ex].
- 3102 [51] Vardan Khachatryan et al. “Measurement of differential cross sections for Higgs  
 3103 boson production in the diphoton decay channel in  $pp$  collisions at  $\sqrt{s} = 8$  TeV”. In:  
 3104 *Eur. Phys. J. C* 76 (2016), p. 13. DOI: 10.1140/epjc/s10052-015-3853-3. arXiv:  
 3105 1508.07819 [hep-ex].
- 3106 [52] Vardan Khachatryan et al. “Measurement of differential and integrated fiducial cross  
 3107 sections for Higgs boson production in the four-lepton decay channel in  $pp$  collisions  
 3108 at  $\sqrt{s} = 7$  and 8 TeV”. In: *JHEP* 04 (2016), p. 005. DOI: 10.1007/JHEP04(2016)005.  
 3109 arXiv: 1512.08377 [hep-ex].

- 3110 [53] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs cross sections: 3.*  
3111 *Higgs properties*. CERN Report CERN-2013-004. 2013. DOI: 10.5170/CERN-2013-  
3112 004. arXiv: 1307.1347 [hep-ph].
- 3113 [54] Serguei Chatrchyan et al. “Measurement of Higgs boson production and properties  
3114 in the WW decay channel with leptonic final states”. In: *JHEP* 01 (2014), p. 096.  
3115 DOI: 10.1007/JHEP01(2014)096. arXiv: 1312.1129 [hep-ex].
- 3116 [55] G. Cowan. “A survey of unfolding methods for particle physics”. In: *Conf. Proc.*  
3117 C0203181 (2002), p. 248.
- 3118 [56] Michael Krämer, Stephen Mrenna, and Davison E. Soper. “Next-to-leading order  
3119 QCD jet production with parton showers and hadronization”. In: *Phys. Rev. D* 73  
3120 (2006), p. 014022. DOI: 10.1103/PhysRevD.73.014022. arXiv: hep-ph/0509127  
3121 [hep-ph].
- 3122 [57] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations  
3123 with Parton Shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070.  
3124 DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].
- 3125 [58] Nils Lavesson and Leif Lonnblad. “Extending CKKW-merging to one-loop matrix  
3126 elements”. In: *JHEP* 12 (2008), p. 070. DOI: 10.1088/1126-6708/2008/12/070.  
3127 arXiv: 0811.2912 [hep-ph].
- 3128 [59] Simone Alioli et al. “NLO Higgs boson production via gluon fusion matched with  
3129 shower in POWHEG”. In: *JHEP* 04 (2009), p. 002. DOI: 10.1088/1126-6708/2009/  
3130 04/002. arXiv: 0812.0578 [hep-ph].
- 3131 [60] Paolo Nason and Carlo Oleari. “NLO Higgs boson production via vector-boson  
3132 fusion matched with shower in POWHEG”. In: *JHEP* 02 (2010), p. 037. DOI:  
3133 10.1007/JHEP02(2010)037. arXiv: 0911.5299 [hep-ph].
- 3134 [61] Simone Alioli, Sven-Olaf Moch, and Peter Uwer. “Hadronic top-quark pair-production  
3135 with one jet and parton showering”. In: *JHEP* 01 (2012), p. 137. DOI: 10.1007/  
3136 JHEP01(2012)137. arXiv: 1110.5251 [hep-ph].
- 3137 [62] J. Alwall et al. “The automated computation of tree-level and next-to-leading order  
3138 differential cross sections, and their matching to parton shower simulations”. In:  
3139 *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301  
3140 [hep-ph].
- 3141 [63] T. Binoth et al. “Gluon-induced W-boson pair production at the LHC”. In: *JHEP*  
3142 12 (2006), p. 046. DOI: 10.1088/1126-6708/2006/12/046. arXiv: hep-ph/0611170  
3143 [hep-ph].
- 3144 [64] Marco Bonvini et al. “Signal-background interference effects for  $gg \rightarrow H \rightarrow W^+W^-$   
3145 beyond leading order”. In: *Phys. Rev. D* 88 (2013), p. 034032. DOI: 10.1103/  
3146 PhysRevD.88.034032. arXiv: 1304.3053 [hep-ph].
- 3147 [65] Giampiero Passarino. “Higgs CAT”. In: *Eur. Phys. J. C* 74 (2014), p. 2866. DOI:  
3148 10.1140/epjc/s10052-014-2866-7. arXiv: 1312.2397 [hep-ph].

- 3149 [66] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. “PYTHIA 6.4 physics and  
 3150 manual”. In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026.  
 3151 arXiv: hep-ph/0603175 [hep-ph].
- 3152 [67] Hung-Liang Lai et al. “Uncertainty induced by QCD coupling in the CTEQ global  
 3153 analysis of parton distributions”. In: *Phys. Rev. D* 82 (2010), p. 054021. DOI:  
 3154 10.1103/PhysRevD.82.054021. arXiv: 1004.4624 [hep-ph].
- 3155 [68] Hung-Liang Lai et al. “New parton distributions for collider physics”. In: *Phys. Rev.*  
 3156 *D* 82 (2010), p. 074024. DOI: 10.1103/PhysRevD.82.074024. arXiv: 1007.2241  
 3157 [hep-ph].
- 3158 [69] LHC Higgs Cross Section Working Group. “Handbook of LHC Higgs Cross Sections”.  
 3159 In: *arXiv:1101.0593* (2011).
- 3160 [70] Simone Alioli et al. “A general framework for implementing NLO calculations in  
 3161 shower Monte Carlo programs: the POWHEG BOX”. In: *JHEP* 06 (2010), p. 043.  
 3162 DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].
- 3163 [71] S. Agostinelli et al. “GEANT4: A simulation toolkit”. In: *Nucl. Instrum. Meth. A*  
 3164 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.
- 3165 [72] E. Bagnaschi et al. “Higgs production via gluon fusion in the POWHEG approach in  
 3166 the SM and in the MSSM”. In: *JHEP* 02 (2012), p. 088. DOI: 10.1007/JHEP02(2012)  
 3167 088. arXiv: 1111.2854 [hep-ph].
- 3168 [73] Stefano Actis et al. “NLO Electroweak Corrections to Higgs Boson Production at  
 3169 Hadron Colliders”. In: *Phys. Lett. B* 670 (2008), p. 12. DOI: 10.1016/j.physletb.  
 3170 2008.10.018. arXiv: 0809.1301 [hep-ph].
- 3171 [74] Stefano Catani et al. “Soft gluon resummation for Higgs boson production at hadron  
 3172 colliders”. In: *JHEP* 07 (2003), p. 028. DOI: 10.1088/1126-6708/2003/07/028.  
 3173 arXiv: hep-ph/0306211 [hep-ph].
- 3174 [75] Yanyan Gao et al. “Spin determination of single-produced resonances at hadron  
 3175 colliders”. In: *Phys. Rev. D* 81 (2010), p. 075022. DOI: 10.1103/PhysRevD.81.  
 3176 075022. arXiv: 1001.3396 [hep-ph].
- 3177 [76] Sara Bolognesi et al. “Spin and parity of a single-produced resonance at the LHC”.  
 3178 In: *Phys. Rev. D* 86 (2012), p. 095031. DOI: 10.1103/PhysRevD.86.095031. arXiv:  
 3179 1208.4018 [hep-ph].
- 3180 [77] Ian Anderson et al. “Constraining anomalous  $HVV$  interactions at proton and  
 3181 lepton colliders”. In: *Phys. Rev. D* 89 (2014), p. 035007. DOI: 10.1103/PhysRevD.  
 3182 89.035007. arXiv: 1309.4819 [hep-ph].
- 3183 [78] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. “A Brief Introduction to  
 3184 PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (2008), p. 852. DOI: 10.1016/j.  
 3185 cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

- 3186 [79] Stanislaw Jadach, Johann H. Kuhn, and Zbigniew Was. “TAUOLA: A Library of  
3187 Monte Carlo programs to simulate decays of polarized tau leptons”. In: *Comput.  
3188 Phys. Commun.* 64 (1990), p. 275. DOI: 10.1016/0010-4655(91)90038-M.
- 3189 [80] CMS Collaboration. “Standard Model Cross Sections for CMS at 7 TeV”. In: *CMS  
3190 Generator Group Twiki* (2010).
- 3191 [81] John M. Campbell, R. Keith Ellis, and Ciaran Williams. “Vector boson pair pro-  
3192 duction at the LHC”. In: *JHEP* 07 (2011), p. 018. DOI: 10.1007/JHEP07(2011)018.  
3193 arXiv: 1105.0020 [hep-ph].
- 3194 [82] J. Ohnemus. “Order alpha-s calculations of hadronic W+- gamma and Z gamma  
3195 production”. In: *Phys. Rev.* D47 (1993), pp. 940–955. DOI: 10.1103/PhysRevD.47.  
3196 940.
- 3197 [83] JetMET group. “Jet energy uncertainties, [https://twiki.cern.ch/twiki/bin/  
3198 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/>3199 ch/twiki/bin/view/CMS/JECUncertaintySources%5C#2012_JEC).
- 3200 [84] Sergey Alekhin et al. “The PDF4LHC Working Group Interim Report”. 2011.
- 3201 [85] Michiel Botje et al. “The PDF4LHC Working Group Interim Recommendations”.  
3202 2011.
- 3203 [86] Richard D. Ball et al. “Impact of Heavy Quark Masses on Parton Distributions  
3204 and LHC Phenomenology”. In: *Nucl. Phys. B* 849 (2011), p. 296. DOI: 10.1016/j.  
3205 nuclphysb.2011.03.021. arXiv: 1101.1300 [hep-ph].
- 3206 [87] A. D. Martin et al. “Parton distributions for the LHC”. In: *Eur. Phys. J. C* 63 (2009),  
3207 p. 189. DOI: 10.1140/epjc/s10052-009-1072-5. arXiv: 0901.0002 [hep-ph].
- 3208 [88] Iain W. Stewart and Frank J. Tackmann. “Theory uncertainties for Higgs and other  
3209 searches using jet bins”. In: *Phys. Rev. D* 85 (2012), p. 034011. DOI: 10.1103/  
3210 PhysRevD.85.034011. arXiv: 1107.2117 [hep-ph].
- 3211 [89] “Procedure for the LHC Higgs boson search combination in summer 2011”. In:  
3212 (2011).
- 3213 [90] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”.  
3214 In: *Eur. Phys. J. C* 71 (2011). [Erratum: *Eur. Phys. J.* C73,2501(2013)], p. 1554. DOI:  
3215 10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z. arXiv:  
3216 1007.1727 [physics.data-an].
- 3217 [91] Tim Adye. “Unfolding algorithms and tests using RooUnfold”. 2011.
- 3218 [92] Andreas Hocker and Vakhtang Kartvelishvili. “SVD approach to data unfolding”. In:  
3219 *Nucl. Instrum. Meth. A* 372 (1996), p. 469. DOI: 10.1016/0168-9002(95)01478-0.  
3220 arXiv: hep-ph/9509307 [hep-ph].
- 3221 [93] G. D’Agostini. “A Multidimensional unfolding method based on Bayes’ theorem”.  
3222 In: *Nucl. Instrum. Meth. A* 362 (1995), pp. 487–498. DOI: 10.1016/0168-9002(95)  
3223 00274-X.

- 3224 [94] Vardan Khachatryan et al. “Precise determination of the mass of the Higgs boson  
 3225 and tests of compatibility of its couplings with the standard model predictions using  
 3226 proton collisions at 7 and 8 TeV”. In: *Eur. Phys. J. C* 75 (2015), p. 212. DOI:  
 3227 10.1140/epjc/s10052-015-3351-7. arXiv: 1412.8662 [hep-ex].
- 3228 [95] Paolo Nason. “A New method for combining NLO QCD with shower Monte Carlo  
 3229 algorithms”. In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040.  
 3230 arXiv: hep-ph/0409146 [hep-ph].
- 3231 [96] S. Bolognesi, Y. Gao and A.V. Gritsan *et al.* “JHUGen”. In: URL <http://www.pha.jhu.edu/spin/>  
 3232 (2011).
- 3233 [97] Gionata Luisoni et al. “ $HW^\pm/HZ + 0$  and 1 jet at NLO with the POWHEG BOX  
 3234 interfaced to GoSam and their merging within MiNLO”. In: *JHEP* 10 (2013), p. 083.  
 3235 DOI: 10.1007/JHEP10(2013)083. arXiv: 1306.2542 [hep-ph].
- 3236 [98] Tom Melia et al. “W+W-, WZ and ZZ production in the POWHEG BOX”. In: *JHEP*  
 3237 11 (2011), p. 078. DOI: 10.1007/JHEP11(2011)078. arXiv: 1107.5051 [hep-ph].
- 3238 [99] T. Gehrmann et al. “ $W^+W^-$  Production at Hadron Colliders in Next to Next to  
 3239 Leading Order QCD”. In: *Phys. Rev. Lett.* 113.21 (2014), p. 212001. DOI: 10.1103/  
 3240 PhysRevLett.113.212001. arXiv: 1408.5243 [hep-ph].
- 3241 [100] Patrick Meade, Harikrishnan Ramani, and Mao Zeng. “Transverse momentum resum-  
 3242 mation effects in  $W^+W^-$  measurements”. In: *Phys. Rev.* D90.11 (2014), p. 114006.  
 3243 DOI: 10.1103/PhysRevD.90.114006. arXiv: 1407.4481 [hep-ph].
- 3244 [101] Prerit Jaiswal and Takemichi Okui. “Explanation of the  $WW$  excess at the LHC  
 3245 by jet-veto resummation”. In: *Phys. Rev.* D90.7 (2014), p. 073009. DOI: 10.1103/  
 3246 PhysRevD.90.073009. arXiv: 1407.4537 [hep-ph].
- 3247 [102] John M. Campbell, R. Keith Ellis, and Ciaran Williams. “Bounding the Higgs width  
 3248 at the LHC: Complementary results from  $H \rightarrow WW$ ”. In: *Phys. Rev.* D89.5 (2014),  
 3249 p. 053011. DOI: 10.1103/PhysRevD.89.053011. arXiv: 1312.1628 [hep-ph].
- 3250 [103] Richard D. Ball et al. “Parton distributions with QED corrections”. In: *Nucl. Phys.*  
 3251 B877 (2013), pp. 290–320. DOI: 10.1016/j.nuclphysb.2013.10.010. arXiv:  
 3252 1308.0598 [hep-ph].
- 3253 [104] Richard D. Ball et al. “Unbiased global determination of parton distributions and  
 3254 their uncertainties at NNLO and at LO”. In: *Nucl. Phys.* B855 (2012), pp. 153–221.  
 3255 DOI: 10.1016/j.nuclphysb.2011.09.024. arXiv: 1107.2652 [hep-ph].
- 3256 [105] Vardan Khachatryan et al. “Event generator tunes obtained from underlying event  
 3257 and multiparton scattering measurements”. In: (2015). arXiv: 1512.00815 [hep-ex].
- 3258 [106] Peter Richardson and Alexandra Wilcock. “Monte Carlo Simulation of Hard Radia-  
 3259 tion in Decays in Beyond the Standard Model Physics in Herwig++”. In: *Eur. Phys.*  
 3260 *J.* C74 (2014), p. 2713. DOI: 10.1140/epjc/s10052-014-2713-x. arXiv: 1303.4563  
 3261 [hep-ph].

- 3262 [107] J. Bellm et al. “Herwig++ 2.7 Release Note”. In: (2013). arXiv: 1310.6877 [hep-ph].
- 3263 [108] *SM Higgs production cross sections at  $\sqrt{s} = 13\text{--}14 \text{ TeV}$* . <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt1314TeV>.
- 3264
- 3265 [109] Fabrizio Caola et al. “QCD corrections to  $W^+W^-$  production through gluon fusion”. In: *Phys. Lett.* B754 (2016), pp. 275–280. DOI: 10.1016/j.physletb.2016.01.046. arXiv: 1511.08617 [hep-ph].
- 3266
- 3267
- 3268 [110] *NLO single-top channel cross sections*. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec>.
- 3269
- 3270 [111] *NNLO+NNLL top-quark-pair cross sections*. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO>.
- 3271
- 3272 [112] *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. Geneva: CERN, Aug. 2011.
- 3273
- 3274 [113] *LHC Higgs Cross Section Working Group*. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>.
- 3275
- 3276 [114] Jon Butterworth et al. “PDF4LHC recommendations for LHC Run II”. In: *J. Phys.* G43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].
- 3277
- 3278
- 3279 [115] Vardan Khachatryan et al. “Search for a Higgs Boson in the Mass Range from 145 to 1000 GeV Decaying to a Pair of W or Z Bosons”. In: *JHEP* 10 (2015), p. 144. DOI: 10.1007/JHEP10(2015)144. arXiv: 1504.00936 [hep-ex].
- 3280
- 3281
- 3282 [116] G. C. Branco et al. “Theory and phenomenology of two-Higgs-doublet models”. In: *Phys. Rept.* 516 (2012), p. 1. DOI: 10.1016/j.physrep.2012.02.002. arXiv: 1106.0034 [hep-ph].
- 3283
- 3284
- 3285 [117] N. Craig and T. Scott. “Exclusive signals of an extended Higgs sector”. In: *JHEP* 11 (2012), p. 083. DOI: 10.1007/JHEP11(2012)083. arXiv: 1207.4835 [hep-ph].
- 3286
- 3287 [118] H. Haber and O. Stal. “New LHC benchmarks for the CP-conserving two-Higgs-doublet model”. In: *Eur. Phys. J. C* 75 (2015), p. 491. DOI: 10.1140/epjc/s10052-015-3697-x. arXiv: 1507.04281 [hep-ph].
- 3288
- 3289
- 3290 [119] Suyong Choi, Sunghoon Jung, and P. Ko. “Implications of LHC data on 125 GeV Higgs-like boson for the Standard Model and its various extensions”. In: *JHEP* 1310 (2013), p. 225. DOI: 10.1007/JHEP10(2013)225. arXiv: 1307.3948.
- 3291
- 3292
- 3293 [120] Tania Robens and Tim Stefaniak. “Status of the Higgs singlet extension of the standard model after LHC run 1”. In: *Eur. Phys. J. C* 75 (2015), p. 105. DOI: 10.1140/epjc/s10052-015-3323-y. arXiv: 1501.02234 [hep-ph].
- 3294
- 3295
- 3296 [121] Johan Alwall et al. “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”. In: *Eur. Phys. J. C* 53 (2008), pp. 473–500. DOI: 10.1140/epjc/s10052-007-0490-5. arXiv: 0706.2569 [hep-ph].
- 3297
- 3298
- 3299

- 3300 [122] Massimiliano Grazzini. “NNLO predictions for the Higgs boson signal in the  $H \rightarrow WW \rightarrow$   
3301  $\ell\nu\ell\nu$  and  $H \rightarrow ZZ \rightarrow 4\ell$  decay channels”. In: *JHEP* 02 (2008), p. 043. DOI: 10.1088/  
3302 1126-6708/2008/02/043. arXiv: 0801.3232 [[hep-ph](#)].
- 3303 [123] Vardan Khachatryan et al. “Measurement of the differential cross section for top  
3304 quark pair production in pp collisions at  $\sqrt{s} = 8$  TeV”. In: *Eur. Phys. J. C* 75.11  
3305 (2015), p. 542. DOI: 10.1140/epjc/s10052-015-3709-x. arXiv: 1505.04480  
3306 [[hep-ex](#)].