

1

2

3

Measurements of the Higgs boson decay to W^+W^- with the CMS detector

4

5

Lorenzo Viliani
of University of Florence

6

7

PhD Thesis

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 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.

²⁴ Contents

1. Physics at the LHC	3
1.1. The Standard Model of particle physics	3
1.2. Beyond the Standard Model	4
1.3. The strong interaction	6
1.3.1. Proton-proton interactions	6
1.3.2. Hadron collider kinematics	8
1.4. Higgs boson phenomenology	8
1.4.1. Higgs boson production modes	9
2. The CMS experiment at the LHC	11
2.1. The Large Hadron Collider	11
2.2. The <i>Compact Muon Solenoid</i> experiment	15
2.2.1. The solenoid	18
2.2.2. The tracker	18
2.2.3. The electromagnetic calorimeter (ECAL)	23
2.2.4. The hadron calorimeter (HCAL)	25
2.2.5. The muon system	26
2.3. The CMS trigger system	28
2.3.1. The Level-1 trigger	29
2.3.2. The high level trigger (HLT)	30
3. Reconstruction and identification of physics objects	33
3.1. The Particle Flow technique	33
3.2. Leptons reconstruction and identification	35
3.2.1. Muon reconstruction and identification	35
3.2.2. Muon isolation	36
3.2.3. Muon momentum scale and resolution	37
3.2.4. Electron reconstruction and identification	38
3.2.5. Electron isolation	38
3.2.6. Electron momentum scale and resolution	39
3.2.7. Lepton identification and isolation efficiency	40
3.2.8. Lepton trigger efficiency	41
3.3. Jets reconstruction and identification	42
3.3.1. Jet reconstruction in CMS	43
3.3.2. Jet energy correction	45

58	3.4. Jet b tagging	48
59	3.5. Missing transverse energy	50
60	3.5.1. E_T^{miss} scale and resolution measurement	52
61	4. Measurement of the Higgs boson transverse momentum at 8 TeV using $H \rightarrow WW \rightarrow 2\ell 2\nu$ decays	55
62	4.1. Introduction	55
63	4.2. Data sets, triggers and MC samples	57
64	4.2.1. Data sets and triggers	57
65	4.2.2. Monte Carlo samples	58
66	4.3. Analysis Strategy	59
67	4.3.1. Event reconstruction and selections	60
68	4.3.2. Simulation efficiencies and scale factors	63
69	4.3.3. Fiducial phase space	67
70	4.3.4. Binning of the p_T^H distribution	68
71	4.4. Background estimation	69
72	4.4.1. Top quark background	69
73	4.4.2. WW background	71
74	4.4.3. W+jets background	72
75	4.4.4. $Z/\gamma^* \rightarrow \tau^+\tau^-$ -background	75
76	4.5. Systematic uncertainties	76
77	4.5.1. Background predictions uncertainties	77
78	4.5.2. Experimental uncertainties	79
79	4.5.3. Theoretical uncertainties	80
80	4.5.4. Statistics uncertainty of the simulated samples	80
81	4.5.5. Treatment of systematic uncertainties in the analysis	81
82	4.6. Signal extraction	81
83	4.6.1. Fitting procedure	81
84	4.6.2. Signal and background yields	84
85	4.7. Unfolding	89
86	4.7.1. Treatment of systematic uncertainties	92
87	4.8. Results	95
88	5. Search for the SM Higgs boson in the $H \rightarrow WW$ channel with the first 13 TeV LHC data	99
89	5.1. Introduction	99
90	5.2. Data and simulated samples	99
91	5.3. Analysis strategy	104
92	5.3.1. Event reconstruction	104
93	5.3.2. Event selection and background rejection	109
94	5.3.3. Signal extraction	111
95	5.4. Background estimation	116
96	5.4.1. WW background	116

99	5.4.2. Top quark background	116
100	5.4.3. Jet-induced (or Fake) background	118
101	5.4.4. DY background	118
102	5.4.5. Other backgrounds	119
103	5.5. Systematic uncertainties	120
104	5.6. Results	122
105	6. Search for high mass resonances decaying to a W boson pair with first 13 TeV LHC data	125
107	6.1. Introduction	125
108	6.2. Data and simulated samples	125
109	6.3. Analysis strategy	127
110	6.4. Background estimation	130
111	6.5. Systematic uncertainties	133
112	6.6. Signal extraction and limit setting	133
113	6.7. Results	137
114	7. Conclusions	143
115	A. Fiducial region definition and optimization	145
116	Bibliography	153

Chapter 1.

¹¹⁸ Physics at the LHC

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

¹²⁵ 1.1. The Standard Model of particle physics

¹²⁶ The Standard Model of Particle Physics (SM) is the theory that describes all fundamental
¹²⁷ constituents of matter and their interactions [1]. It is a renormalisable quantum field theory
¹²⁸ based on a $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ local gauge symmetry, and is capable to provide a
¹²⁹ quantitative description of three of the four interactions in nature: electromagnetism, weak
¹³⁰ interaction and strong nuclear force.

¹³¹ According to the SM, the ordinary matter is made up of spin-1/2 particles, denoted
¹³² as fermions. The fermions are subdivided into two classifications of elementary particles:
¹³³ leptons and quarks. Both classes consist of six particles, grouped into three doublets, called
¹³⁴ generations. Additional three doublets for each class are composed of leptons and quarks
¹³⁵ antiparticles. A charged particle with electric charge $Q = -1$, either the electron e , the
¹³⁶ muon μ or the tauon τ , and a neutral particle, the corresponding neutrino, compose the
¹³⁷ following lepton generations, ordered according to an increasing mass hierarchy:

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

¹³⁸ Charged leptons can interact via the electromagnetic and weak force, while neutrinos,
¹³⁹ that are assumed to be massless, can interact only through the weak interaction.

¹⁴⁰ Similarly, the quarks are organized in pairs composed of a particle with $Q = +2/3$, *up*
¹⁴¹ (*u*), *charm* (*c*) and *top* (*t*) quarks, and another particle with $Q = -1/3$, *down* (*d*), *strange*
¹⁴² (*s*) and *bottom* (*b*) quarks:

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

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

148 In the SM the interaction between elementary particles occurs through the exchange of
149 spin-1 particles, known as bosons, which identify the fundamental forces. The photon γ is
150 the mediator of the electromagnetic interaction, the W^\pm and Z bosons are the mediators of
151 the weak interaction, while the strong force is mediated by eight gluons g . Electromagnetic
152 and weak interactions are actually the manifestations of the same fundamental interaction,
153 the electroweak force.

154 Local gauge symmetries naturally lead to the presence of gauge bosons, the exchange
155 particles mediators of the fundamental interactions. The symmetry requires these gauge
156 bosons to be massless, which is unproblematic for photons and gluons, but in drastic contrast
157 to the known masses of the Z and W^\pm bosons, which are $m_Z = 91.1876 \pm 0.0021$ GeV and
158 $m_W = 80.385 \pm 0.015$ GeV, respectively. Moreover, the maximally parity violating structure
159 of the weak charged currents also breaks local gauge invariance for all massive fermions, due
160 to their coupling to the W boson. This leads to the apparent antagonism that, while the
161 $SU(2)_L \otimes U(1)_Y$ gauge symmetry does describe the coupling structure of the electroweak
162 force, at the same time it seems to contradict the fact that the W and Z bosons, and all
163 fermions have a non-vanishing mass.

164 The proposed solution to this problem is the mechanism of *spontaneous symmetry*
165 *breaking*, where the gauge symmetry is still intrinsic to the Lagrangian density of the theory,
166 but not manifest in its energy ground state, which in this case is the quantum vacuum.
167 The spontaneous symmetry breaking of the $SU(2)_L \otimes U(1)_Y$ symmetry group requires the
168 introduction of a self-interacting scalar field, the *Higgs field*.

169 During the past decades the predictions of the SM have been confirmed by experimental
170 results with outstanding precision, and in 2012 the discovery of a new boson with a mass
171 $m_H = 125.09 \pm 0.21$ (stat) ± 0.11 (syst) GeV [2], consistent with the predicted Higgs boson
172 was announced by the ATLAS and CMS experiments at CERN.

173 1.2. Beyond the Standard Model

174 The discovery of the new boson in accordance with the Higgs boson predicted by the SM
175 has been a major breakthrough in the contemporary particle physics. The Higgs boson mass
176 is a free parameter in the SM and its measurement fixes all the other parameters related to
177 the Higgs field, such as the coupling strengths with bosons and fermions. The current quest

is to establish whether the properties of the discovered boson are consistent with the SM predictions, or it is only a component of a more entangled Higgs sector. Moreover, there are still several aspects that are not explained by the SM, such as the hierarchy problem, the nature of dark matter and others.

Several theoretical models have been proposed to explain the deficiencies of the SM. One of the simplest extension of the SM Higgs sector requires the existence of an additional singlet scalar field, S , which is neutral under all quantum numbers of the SM gauge group [3]. In general the singlet field mixes with the SM Higgs boson, H , allowing it to couple to the same states as the SM Higgs boson itself. If the mass of the scalar singlet was more than twice that of the SM Higgs boson, the the S branching ratios would be reduced with respect to the H ones, because of the opening of the new $S \rightarrow HH$ decay channel.

The mixing of the two states S and H would manifest as a suppression of the production cross section of the two states and a suppression of the heavy mass Higgs boson decay modes to SM particles, if the $S \rightarrow HH$ decay is kinematically accessible. In particular, identifying as H the observed Higgs boson with $m_H = 125$ GeV, and supposing that the new scalar singlet S is heavier than H , one can introduce the scale factors of the low and high mass state couplings, \mathcal{C} and \mathcal{C}' , respectively. These factors are related by the unitarity condition $\mathcal{C}^2 + \mathcal{C}'^2 = 1$. The singlet cross section and width are consequently modified by the factors μ' and Γ' , respectively:

$$\begin{aligned}\mu' &= \mathcal{C}'^2 \cdot (1 - \mathcal{B}_{\text{new}}) \quad , \\ \Gamma' &= \Gamma_{\text{SM}} \cdot \frac{\mathcal{C}'^2}{1 - \mathcal{B}_{\text{new}}} \quad ,\end{aligned}\tag{1.3}$$

where \mathcal{B}_{new} is the singlet branching fraction to non-SM-like decay modes.

Other models, such as the *two-Higgs-doublet model* (2HDM) [4], extend the minimal Higgs content requiring the introduction of a second Higgs doublet. The generalization of the SM Lagrangian with two complex scalar fields, which are $SU(2)_L$ doublets, eventually gives rise to five physical Higgs bosons: a charged pair (H^\pm); two neutral CP -even scalars (H and h , where $m_H > m_h$ by convention); and a neutral CP -odd scalar (A) [5]. The parameter space of these 2HDM models can accommodate a wide range of variations in the production and decay modes of the SM-like Higgs boson. Nevertheless, tight constraints on flavour-changing neutral currents disfavour 2HDM with tree-level flavour violation. Similarly, limits on additional sources of CP violation favour 2HDM with a CP -conserving potential. These assumptions significantly reduce the parameter space of 2HDM models. Moreover, if the h boson is identified with the observed 125 GeV boson, the experimental measurements further constraint the possible production and decay modes of the other predicted particles. Examples of possible decay channels in this framework are: the heavy CP -even Higgs may decay to two light CP-even Higgs, $H \rightarrow hh$; the CP -odd pseudoscalar Higgs may decay to a light CP -even Higgs and a Z boson, $A \rightarrow Zh$; the charged Higgs bosons may decay to a SM-like Higgs and a W^\pm boson, $H^\pm \rightarrow W^\pm h$.

214 In order to search for new particles that could be ascribable to the simple models
215 depicted above or even to more complicated theories, it is of utmost importance to provide
216 precise measurements of the Higgs boson couplings and kinematics, as well as its spin and
217 parity properties. A complementary strategy is to perform direct searches for additional
218 Higgs bosons in the full mass range accessible to current and future experiments.

219 1.3. The strong interaction

220 Quantum Chromo-Dynamics (QCD) is the theory that describes the strong interactions [6].
221 It is an unbroken gauge non-abelian theory based on the group $SU(3)$ of colour ($SU(3)_c$).
222 The mediators of the interaction are eight massless gluons and the elementary particles of
223 matter are colour triplets of quarks, with different flavours. In fact, as shown in Eq. (1.2),
224 six types (flavours) of quark exist and each quark possesses a colour charge that can assume
225 three values, namely red, green and blue.

226 The physical vertices in QCD include the gluon-quark-antiquark vertex, analogous to
227 the QED photon-fermion-antifermion coupling, but also the three-gluon and four-gluon
228 vertices, i.e. gluon themselves carry colour charge, which have no analogue in an abelian
229 theory like QED. Quark and gluons are the only particles that interact through the strong
230 interaction.

231 The non-abelian nature of the theory leads to two important characteristics:

- 232 • *colour confinement*: the QCD coupling constant $\alpha_s = g_s^2/4\pi$ is a function of the scale
233 of the interaction Q . At low energy (corresponding to large distances of the order of
234 1 fm) the α_s value is large and a perturbative approach is not applicable. When a
235 quark-antiquark pair begins to separate, the colour field generated by the exchanged
236 gluons increases its intensity and, at some point, becomes more energetically favourable
237 to create a new quark-antiquark pair from the vacuum than increasing further the
238 interaction strength. This explains why free quarks are not observed and the final
239 state particles are made of colourless quark bound states (hadrons). This is also the
240 cause of the hadronization process which causes the formation of jets.
- 241 • *asymptotic freedom*: the coupling constant decreases at large scales Q approaching
242 to zero, meaning that quarks can be asymptotically considered as free particles. The
243 small value of the coupling constant at large scales justifies the usage of a perturbative
244 approach to describe hard processes.

245 1.3.1. Proton-proton interactions

246 The fundamental difference between hadron and lepton collisions is the fact that hadrons,
247 differently from leptons, are not elementary particles but have an internal structure, which
248 can be described in terms of the QCD-improved parton model. The basic idea of this model
249 is to represent the inelastic scattering as quasi-free scattering of point-like constituents
250 within the proton, the partons [7]. Hadrons, along with the valence quarks that contribute
251 to their quantum numbers (*uud* for protons), contain virtual quark-antiquark pairs known

as sea quarks. Sea quarks arise from gluon splitting; a pair of quarks can in turn annihilate producing a gluon. In addition, gluons are present in the sea also owing to the three-gluon and four-gluon vertices.

In proton-proton collisions the interaction generally involves a pair of partons and any of the partons in the sea can interact with a given likelihood, making possible several types of interaction, such as qq , qq' , $q\bar{q}$, $q\bar{q}'$, gq , $g\bar{q}$ or gg .

At a hadron collider the partons entering the hard scattering carry an event-by-event variable fraction x of the proton four-momentum, also known as Bjorken's scaling variable. Therefore the centre-of-mass energy of the hard scattering is given by $\sqrt{\hat{s}} = \sqrt{x_1 x_2 s}$, where \sqrt{s} is the centre-of-mass energy of the incoming protons, and x_1 , x_2 are the four-momentum fractions carried by the two interacting partons¹. Since generally x_1 and x_2 have different values, the centre-of-mass frame of the interaction is boosted along the beam direction. While this represents an experimental difficulty, on the other hand it allows to explore a wider range of energies with respect to an electron-positron collider.

In order to evaluate cross sections in hadron collisions, the calculation can be factorized into long-distance and short-distance components according to the QCD factorization theorem [8]. Therefore, a typical cross section calculation for an inclusive process $pp \rightarrow X$ consists of a term that describes the partonic hard scattering, which can be calculated using perturbative QCD, and factors that describe the incoming flux of partons, the *parton distribution functions* (PDF) f_i , as shown in the following equation [9]:

$$\sigma(pp \rightarrow X) = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij \rightarrow X}(x_1 x_2 s, \mu_R^2, \mu_F^2) . \quad (1.4)$$

In this expression the sum runs over all the initial-state partons with longitudinal momentum fractions x_1 and x_2 , where the subscripts 1 and 2 refers to the two incoming protons. The factorization scale μ_F is an arbitrary parameter that represents the scale at which the separation between the hard perturbative interaction and the long distance, non-perturbative, evolution of the produced partons occurs. The $\hat{\sigma}_{ij \rightarrow X}$ term corresponds to the partonic cross section evaluated at the scales μ_F and μ_R , where μ_R is the renormalization scale, an additional scale introduced in perturbative QCD to treat the ultraviolet divergences. The PDF $f_{i,j}$ represents the probability density for a parton i, j to be found within the incoming proton and to carry a fraction $x_{1,2}$ of its momentum. The PDFs are obtained performing global fits to data at different scales Q^2 and their evolution with scale is governed by the DGLAP equation [10]. The global PDFs fits are provided by three main collaborations: CTEQ [11], MSTW 2008 [12] and NNPDF [13].

¹Considering $\sqrt{s} = 14$ TeV and $x_1, x_2 \approx 0.15\text{--}0.20$, the partonic centre-of-mass energy is of the order of 1–2 TeV.

²⁸⁴ **1.3.2. Hadron collider kinematics**

²⁸⁵ As described before, at hadron colliders the centre-of-mass energy of the parton hard
²⁸⁶ scattering is generally boosted along the beam direction. It is therefore useful to describe
²⁸⁷ the final state in terms of variables that are invariant under Lorentz transformations along
²⁸⁸ that direction. A convenient set of kinematic variables is the transverse momentum p_T , the
²⁸⁹ rapidity y and the azimuthal angle ϕ . In term of these variables, the four-momentum of a
²⁹⁰ particle of mass m can be written as:

$$p^\mu = (E, p_x, p_y, p_z) = (m_T \cosh y, p_T \sin \phi, p_T \cos \phi, m_T \sinh y) \quad , \quad (1.5)$$

²⁹¹ where p_x , p_y and p_z are the components of the momentum \vec{p} (p_z is directed along the
²⁹² beam direction) and the transverse mass is defined as $m_T = \sqrt{p_T^2 + m^2}$. The rapidity y is
²⁹³ defined by the following formula:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \quad . \quad (1.6)$$

²⁹⁴ The rapidity is not invariant under boosts along the beam direction but it transforms
²⁹⁵ according to the law:

$$y \longrightarrow y + \frac{1}{2} \ln \left(\frac{1 + \beta}{1 - \beta} \right) \quad , \quad (1.7)$$

²⁹⁶ where β is the boost velocity. According to this definition the rapidity differences Δy
²⁹⁷ are Lorentz invariant. Experimentally it is more convenient to use the pseudorapidity η ,
²⁹⁸ defined as:

$$\eta = -\ln \tan \frac{\theta}{2} \quad , \quad (1.8)$$

²⁹⁹ where θ is the polar angle between the particle momentum and the beam direction
³⁰⁰ ($\cos \theta = p_z / |\vec{p}|$). For ultra-relativistic particles the pseudorapidity coincides with the
³⁰¹ rapidity.

³⁰² **1.4. Higgs boson phenomenology**

³⁰³ In this section the Higgs boson production modes and decay channels are described, spending
³⁰⁴ some time on the description of the $H \rightarrow WW$ channel, which is the channel considered
³⁰⁵ in this work. Afterwards, a brief review of the Monte Carlo (MC) generators used for the
³⁰⁶ simulation of Higgs boson processes is given. Finally, a description of the effects due to

307 higher order QCD corrections on variables like the Higgs boson transverse momentum are
 308 shown.

309 1.4.1. Higgs boson production mechanisms and decay channels

310 The main processes contributing to the Higgs boson production at hadron colliders are
 311 represented by the Feynman diagrams shown in Fig. 1.1.

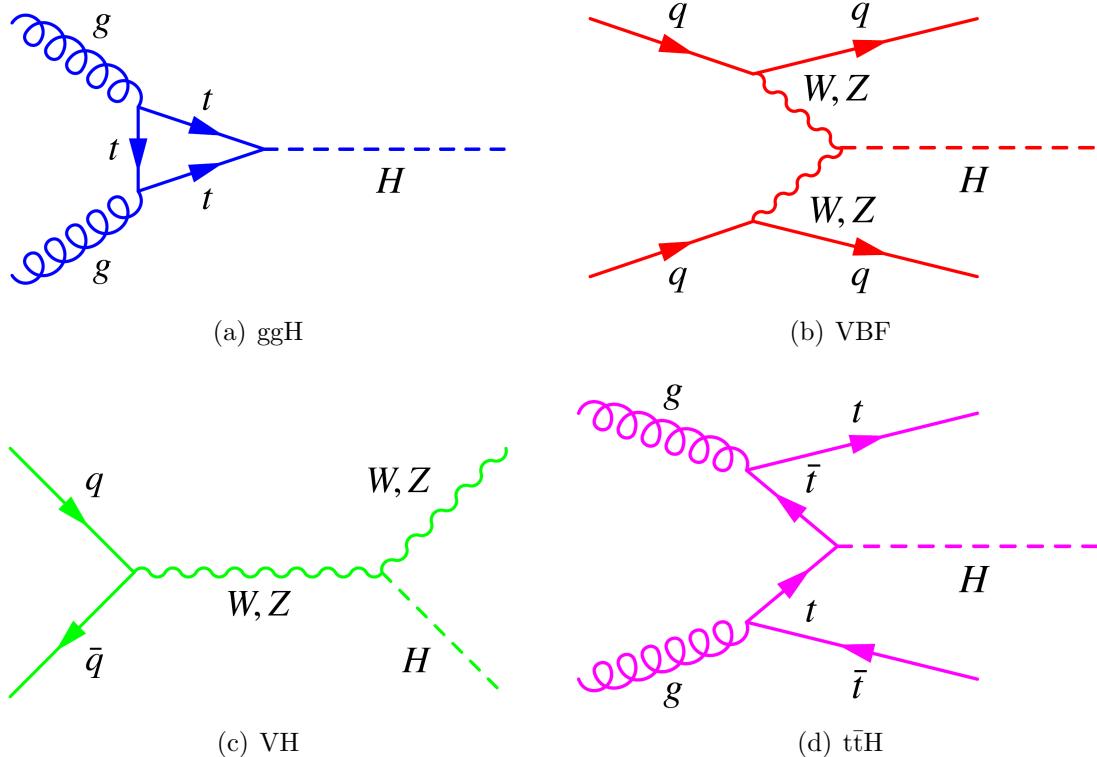


Figure 1.1.: Main Higgs boson production processes at LHC.

312 In order of decreasing cross section, the Higgs boson production modes are:

- 313 • *Gluon fusion (ggH)*: this is the main Higgs boson production mode at LHC over the
 314 whole mass spectrum. The process involves the fusion of two incoming gluons that
 315 give rise to the Higgs boson through a heavy quark loop, whose main contribution
 316 comes from the top quark, as shown in Fig. 1.1(a).
- 317 • *Vector Boson Fusion (VBF)*: each of the two interacting quarks emit a W or Z boson
 318 which, in turn, interact to produce the Higgs boson, as shown in Fig. 1.1(b). Quarks
 319 deriving from the incoming partons after the emission of vector bosons proceed in the
 320 forward direction and represent the peculiar signature of this production mode, i.e.
 321 two high energy forward jets separated by a large pseudorapidity gap. This process

has a cross section which is one order of magnitude lower than ggH for a large range of m_H values and it becomes comparable to ggH only for masses of the order of 1 TeV.

- *Vector boson associated production* (VH): also known as *Higgsstrahlung*, this process is characterized by the emission of a Higgs boson from a W^\pm or Z boson produced by two incoming quarks, as depicted in Fig. 1.1(c). The VH cross section is several orders of magnitude lower than the ggH and VBF cross sections.
- *Top quark associated production* ($t\bar{t}H$): a pair of top quarks, originated from the splitting of two incoming gluons, interacts to give rise to a Higgs boson, as illustrated in Fig. 1.1(d).

Another production mechanism analogous to the $t\bar{t}H$ process and with a similar cross section is the b quark associated production.

The SM Higgs boson production cross section for the various production modes depends on the Higgs boson mass and on the centre-of-mass energy, as shown in Fig. 1.2. In general, the production cross section of all processes decreases with increasing the Higgs boson mass, while the raise of the centre-of-mass energy reflects in an increase of the cross section over the whole mass range.

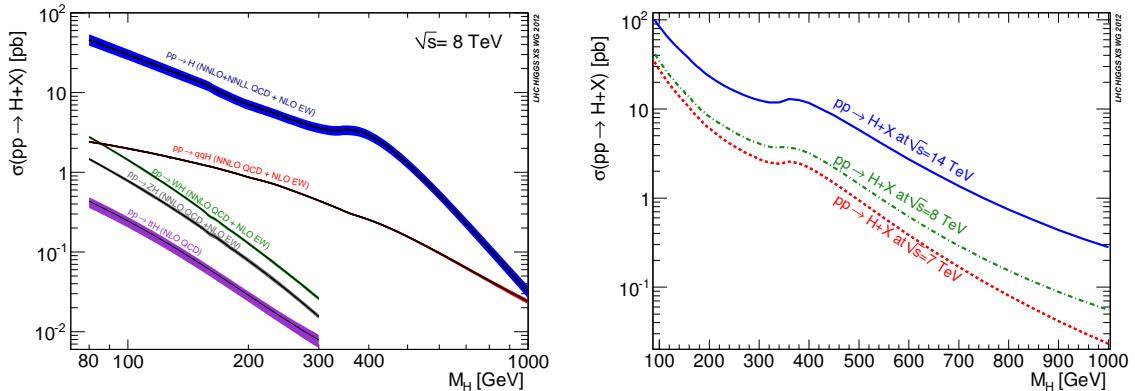


Figure 1.2.: Higgs boson cross section as a function of m_H for the various production mechanisms (left) and for different centre-of-mass energies (right).

The Higgs boson can decay to a variety of final states that can be divided in bosonic channels, like $\gamma\gamma$, ZZ or W^+W^- , and fermionic channels, like $\tau\tau$, $b\bar{b}$, etc. Its branching ratio also depends on the Higgs boson mass, as illustrated in Fig. ??, where different decay channels are compared over the whole mass spectrum.

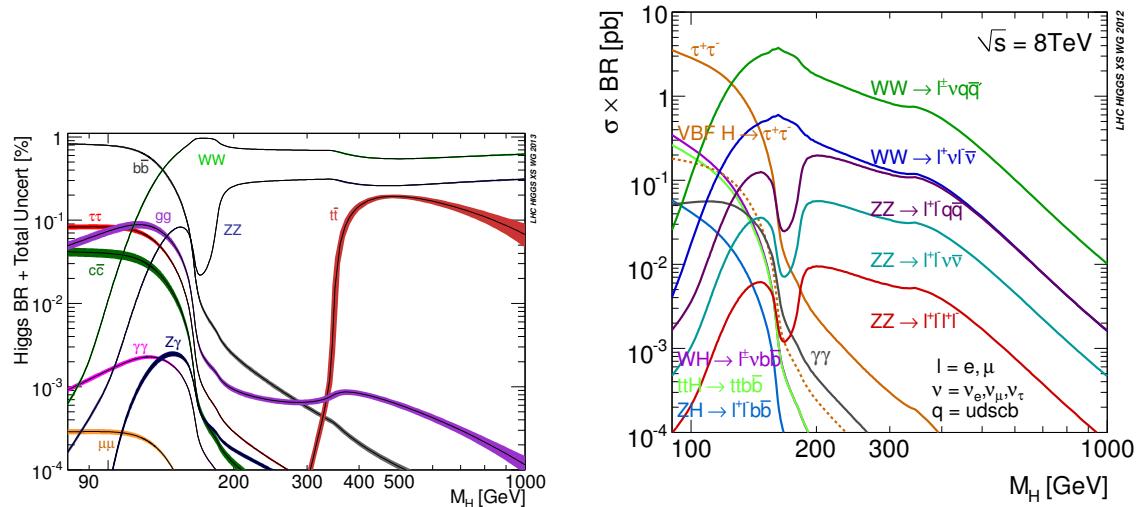


Figure 1.3.: Higgs boson branching ratio (left) and cross section times branching ratio (right) for all the decay channels as a function of m_H .

Chapter 2.

³⁴² The CMS experiment at the LHC

³⁴³ In this chapter, the main characteristics of the Large Hadron Collider (LHC particle
³⁴⁴ accelerator and Compact Muon Solenoid (CMS) experiment are described.

³⁴⁵ 2.1. The Large Hadron Collider

³⁴⁶ The LHC [14–17] at CERN, officially inaugurated on 21st October 2008, is the largest
³⁴⁷ and most powerful hadron collider ever built. Installed in the underground tunnel which
³⁴⁸ hosted the Large Electron Positron Collider (LEP) [18–20], the leptonic accelerator in
³⁴⁹ operation until 2nd November 2000, the LHC accelerator has the shape of a circle with
³⁵⁰ a length of about 27 km and is located underground at a depth varying between 50 m to
³⁵¹ 175 m, straddling the Franco-Swiss border near Geneva. It is designed to collide two 7 TeV
³⁵² counter-circulating beams of protons resulting in a center-of-mass energy of 14 TeV, or
³⁵³ two beams of heavy ions, in particular lead nuclei at an energy of 2.76 TeV/nucleon in the
³⁵⁴ center-of-mass frame.

³⁵⁵ The transition from a leptonic collider to a hadronic collider entailed the following
³⁵⁶ advantages: first, it has been possible to build a machine that having the same size of the
³⁵⁷ previous one (and therefore accommodated in the same LEP tunnel, substantially reducing
³⁵⁸ the cost and time of construction), could reach a higher energy in the center-of-mass
³⁵⁹ frame. This is due to the much lower amount of energy loss through synchrotron radiation
³⁶⁰ emitted by the accelerated particles, that is proportional to the fourth power of the ratio
³⁶¹ E/m between their energy and their mass. Secondly, the composite structure of protons
³⁶² compared to the elementary structure of electrons allows LHC to be able to simultaneously
³⁶³ access a wider energy spectrum, despite the production of many low energies particles in a
³⁶⁴ complex environment. This is a particularly important feature for a machine dedicated to
³⁶⁵ the discovery of “new” physics.

³⁶⁶ In Fig. 2.1 a schematic description of the accelerator complex installed at CERN is shown.
³⁶⁷ The acceleration is performed in several stages [17]. The protons source is a *Duoplasmatron*:
³⁶⁸ the protons are obtained by removing electrons from a source of hydrogen gas and then
³⁶⁹ sent to the LINAC2, a 36 m long linear accelerator which generates a pulsed beam with
³⁷⁰ an energy of 50 MeV using Radio Frequency Quadrupoles (RFQ) and focusing quadrupole
³⁷¹ magnets. The beam is subsequently sent to the Proton Synchrotron Booster (PSB), a
³⁷² circular accelerator consisting of four superimposed synchrotron rings with a circumference

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

- CMS (Compact Muon Solenoid) [21] and ATLAS (A Toroidal LHC ApparatuS) [22] are two general-purpose detectors designed to investigate the largest possible spectrum of physics. In particular, they have been devoted to the detection of particles produced by a Higgs boson decay and to look for any evidence of possible new physics. The use of two detectors chasing the same objectives but designed independently is crucial for a cross-check of any possible new discovery;
- LHCb (LHC beauty) [23] is an experiment primarily designed to study CP (combined Charge conjugation and Parity symmetry) violation in electroweak interactions and to study asymmetries between matter and antimatter through the analysis of rare decays of hadrons containing b quarks. The detector is also able to perform measurements in the forward region, at small polar angles with respect to the beam line;
- ALICE (A Large Ion Collider Experiment) [24] is an experiment studying heavy ions collisions, through the production of a new state of matter called quark-gluon plasma.

Two other smaller experiments are located along the circumference of the LHC accelerator, TOTEM and LHCf, which focus on particles emitted in the forward direction. TOTEM (TOTal Elastic and diffractive cross section Measurement) [25] measures the proton-proton interaction cross section and accurately monitors the luminosity of the LHC using detectors positioned on either side of the CMS interaction point. LHCf (LHC forward) [26] is made up of two detectors which sit along the LHC beamline, at 140 m either side of the ATLAS collision point. It makes use of neutral particles thrown forward by LHC collisions as a source to simulate the interaction with the atmosphere of very high energy cosmic rays (between 10^{17} TeV and 10^{20} TeV) in laboratory conditions.

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

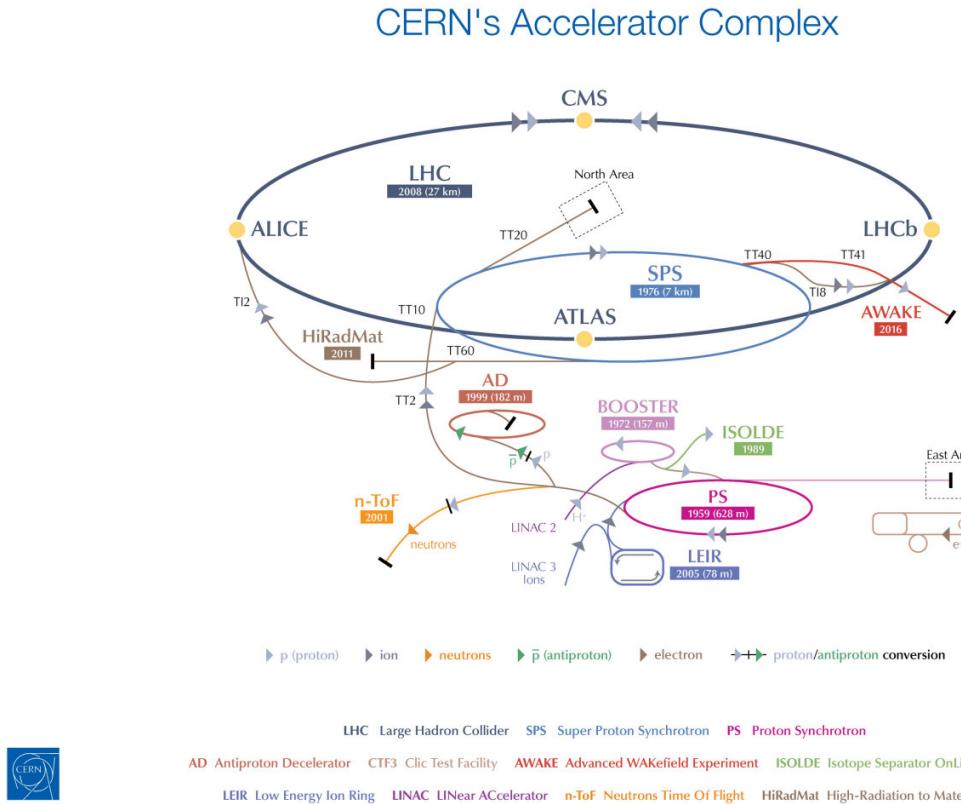


Figure 2.1.: Schematic description of the accelerator complex installed at CERN.

beam direction. A radio frequency acceleration system, consisting of 16 superconducting 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 σ , 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.

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

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

429 where n_1 and n_2 are the number of particles contained in the two bunches colliding at 430 a frequency f , and σ_x and σ_y are the beam sizes in the transverse plane. At LHC, the 431 bunches collide with $f = 40$ MHz and the transverse size of the beam can be squeezed down 432 to around 15 μm . Then, the integrated luminosity L is defined as the time integral of the 433 instantaneous luminosity:

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

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

435 The LHC started to be operative in September 2008 but, due to a faulty interconnection 436 between two magnets which caused a helium leakage in the tunnel, the operation was 437 stopped and restarted in March 2010. During 2010 and 2011 LHC ran successfully and 438 provided proton proton collisions at a center-of-mass energy of 7 TeV, delivering a total 439 integrated luminosity of about 6.1 fb^{-1} . The encouraging results in the Higgs boson search 440 provided by the ATLAS and CMS Collaborations led to the decision of extending the data 441 taking period to the end of 2012, and to increase the center-of-mass energy up to 8 TeV. 442 During 2012, LHC delivered to the experiments an integrated luminosity of 23.3 fb^{-1} . After 443 the first long shutdown (LS1), a two years period started in the early 2013 where the LHC 444 operation stopped for maintenance and upgrade, the LHC started again delivering proton 445 proton collisions on 3rd June 2015, at the new record center-of-mass energy of 13 TeV. 446 During the 2015 the LHC delivered an integrated luminosity of 4.2 fb^{-1} . Nowadays, LHC is 447 still colliding bunches of protons at $\sqrt{s} = 13$ TeV, reaching unprecedented instantaneous 448 luminosities and delivering a total integrated luminosity of 31 fb^{-1} . The cumulative delivered 449 luminosity versus time for the different LHC data taking periods is shown in Fig. 2.2.

450 As the instantaneous luminosity increases, the probability of multiple proton proton 451 interactions to occur in a single bunch crossing grows higher as well. In this instance, the 452 main goal is the identification and reconstruction of a single primary collision where the 453 physics event of interest occurs among the background of the additional proton proton 454 interactions. Such backgrounds are due to processes occurring with very high probability, 455 like the production of low- p_T jets. These additional collisions are known as pile up (PU). 456 During the LHC current run the average number of pile up events is 23, with some event 457 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

458 2.2. The *Compact Muon Solenoid* experiment

459 The CMS apparatus is a general purpose detector situated in one of the four LHC interaction
 460 points¹. The detector is designed to investigate a wide range of physics, from the search of
 461 the Higgs boson, to SM measurements and BSM physics searches. To achieve this goal, the
 462 detector is able to identify and reconstruct all the physics objects that may be produced
 463 in the proton proton collisions: electrons, muons, photons and jets. The main feature of
 464 the CMS detector is a superconducting solenoidal magnet which is capable to produce a

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

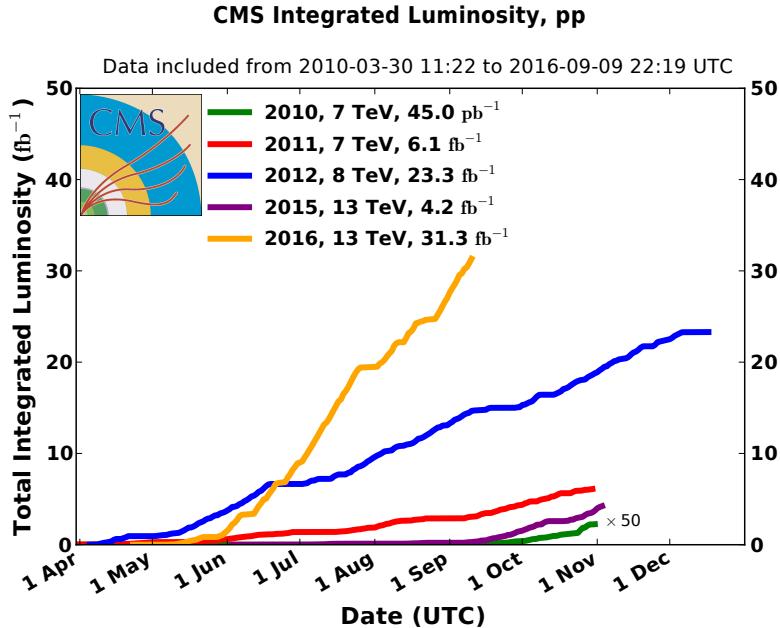


Figure 2.2.: Cumulative luminosity versus day delivered to CMS during proton proton collisions.

465 3.8 T magnetic field. Such a strong magnetic field is the key aspect which permits to have
 466 a compact design of the detector. The detector has a cylindrical structure, which is typical
 467 of general purpose detectors, which consists of several cylindrical detecting layers, coaxial
 468 with the beam direction (*barrel* region), closed at both ends with detecting disks (*endcap*
 469 region), in such a way to ensure the hermetic closure of the apparatus.

470 The coordinate system used by CMS is a right-handed Cartesian system, with the origin
 471 at the center of the detector, in the nominal beam collision point. The *x*-axis is chose
 472 to point radially towards the center of the LHC circumference and the *y*-axis is directed
 473 upwards along the vertical. The *z*-axis is oriented along the beam direction, according
 474 to the anticlockwise direction of the LHC ring if seen from above. The CMS cylindrical
 475 symmetry and the Lorentz invariant description of the proton proton collisions, suggest the
 476 use of a pseudo-angular reference frame, described by the triplet of coordinates (r, ϕ, η) ,
 477 where r is the distance from the *z*-axis, ϕ is the azimuthal angle, measured starting from
 478 the *x*-axis positive direction, and η is the pseudorapidity, defined in Sec. 1.3.2.

479 The schematic view of the CMS detector, which has a length of 21.5 m, a diameter of
 480 15 m and a weight of about 14000 tons, is shown in Fig. 2.3. From the inner region to the
 481 outer one, the various CMS sub-detectors are:

- 482 • **Silicon tracker:** it occupies the region $r < 1.2$ m and $|\eta| < 2.5$. It is composed of an
 483 inner silicon pixel vertex detector and a surrounding silicon microstrip detector, with
 484 a total active area of about 215 m². It is used to reconstruct charged particle tracks
 485 and vertices;

- 486 • **Electromagnetic calorimeter (ECAL)**: placed in the region $1.2 \text{ m} < r < 1.8 \text{ m}$
 487 and $|\eta| < 3$, it consists of many scintillating crystals of lead tungstate (PbWO_4). It is
 488 used for the measurement of the trajectory and the energy released by electrons and
 489 photons;
- 490 • **Hadronic calorimeter (HCAL)**: it is placed in the region $1.8 \text{ m} < r < 2.9 \text{ m}$ and
 491 $|\eta| < 5$. It is made up of brass layers alternated with plastic scintillators and it is
 492 used to measure the direction and energy deposited by the hadrons produced in the
 493 interactions;
- 494 • **Superconducting solenoidal magnet**: it occupies the region $2.9 \text{ m} < r < 3.8 \text{ m}$
 495 and $|\eta| < 1.5$ and generates an internal uniform magnetic field with an intensity of
 496 3.8 T , pointing along the direction of the beams. The magnetic field is necessary to
 497 bend the trajectories of charged particles, in order to allow the measurement of their
 498 momentum through the curvature observed in the tracking system. The magnetic field
 499 lines are closed by an external 21 m long iron yoke, that has a diameter of 14 m.
 500 Outside the return yoke, a residual 1.8 T magnetic field is present, pointing in the
 501 opposite direction with respect to the internal field;
- 502 • **Muon system**: the outermost system, which is placed in the region $4 \text{ m} < r < 7.4 \text{ m}$
 503 and $|\eta| < 2.4$, has the purpose of reconstructing the tracks of muons passing through
 504 it. It consists of Drift Tubes (DT) in the barrel region and Cathode Strip Chambers
 505 (CSC) in the endcaps. A complementary system of Resistive Plate Chambers (RPC)
 506 is used both in the barrel and endcaps. The muon chambers are housed inside the
 507 iron structure of the return yoke.

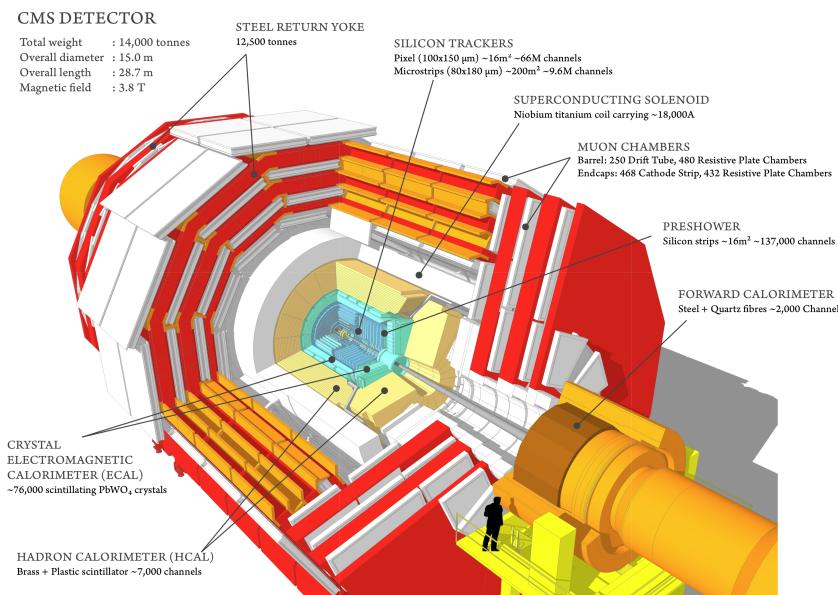


Figure 2.3.: Schematic view of the CMS detector showing its sub-detectors.

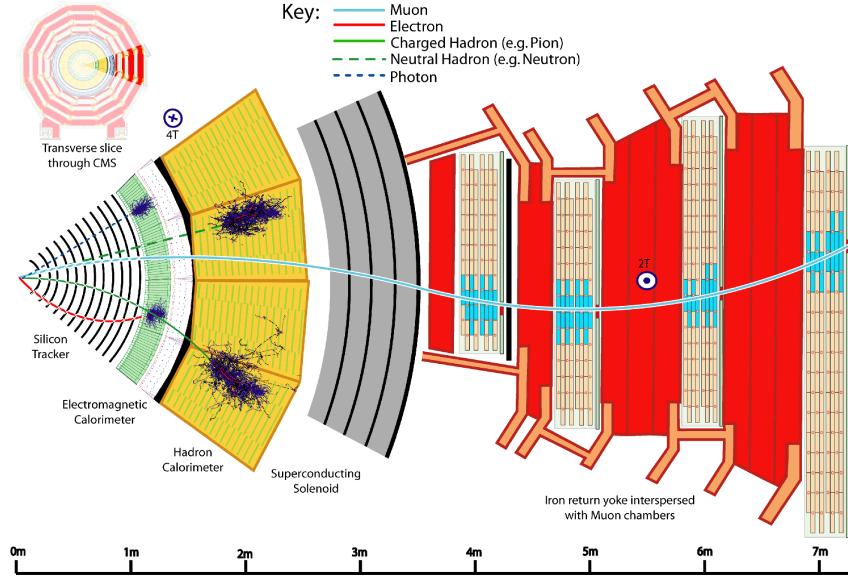


Figure 2.4.: Schematic view of a slice of the CMS detector, showing the sub-detectors response to the passage of different types of particles.

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

511 2.2.1. The solenoid

512 The CMS magnet [27], which contains the tracker, the electromagnetic and the hadronic
 513 calorimeters, is the biggest superconducting solenoid ever built. The solenoid can generate
 514 a magnetic field of 3.8 T in the internal bore, which has a diameter of 6 m and a length of
 515 12.5 m. The energy stored in the magnet is about 2.7 GJ at full current. The superconductor
 516 is made of four Niobium-Titanium layers and it is cooled down to about 4 K through a
 517 liquid Helium cooling plant. In case of a quench, when the magnet loses its superconducting
 518 property, the energy is dumped to resistors within 200 ms. The magnet return yoke of
 519 the barrel is composed with three sections along the z -axis; each one is split into 4 layers
 520 (holding the muon chambers in the gaps). Most of the iron volume is saturated or nearly
 521 saturated, and the field in the yoke is about the half (1.8 T) of the field in the central
 522 volume.

523 2.2.2. The tracker

524 The silicon tracker is the detector closest to the beam collision point. Its goal is the
 525 high resolution reconstruction of the trajectories of charged particles originating from the
 526 interaction point and the identification of the position of secondary vertices produced by

527 particles with a short mean life time (in particular hadrons containing the b quark, that
 528 decay after few hundred of μm). The events produced in the proton proton collisions can
 529 be very complex and track reconstruction is an entangled pattern recognition problem.
 530 Indeed, at the nominal instantaneous luminosity of operation, an average of about 20 pile up
 531 events overlapping to the event of interest are expected, leading to about 1000 tracks to be
 532 reconstructed every 25 ns. In order to make the pattern recognition easier, two requirements
 533 are fundamental:

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

536 The first requirement is achieved building a detector with high granularity². The redundancy
 537 of the hits is instead achieved having several detecting layers, and is necessary to reduce the
 538 ambiguity on the assignment of the hits to a given track. Nevertheless, the amount of tracker
 539 material has to be as low as possible, in order to avoid compromising the measurement
 540 of the particle trajectory. An excessive amount of material would indeed deteriorate the
 541 measurement, mainly because of the increased probability of particle multiple scattering.
 542 The outer detectors such as ECAL would be influenced by the material as well, for example
 543 because of the increased probability for a photon to convert to an electron-positron pair
 544 in the tracker material. For this reasons, the tracker layers are limited in number and
 545 thickness.

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

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

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

562 The pixel detector

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

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

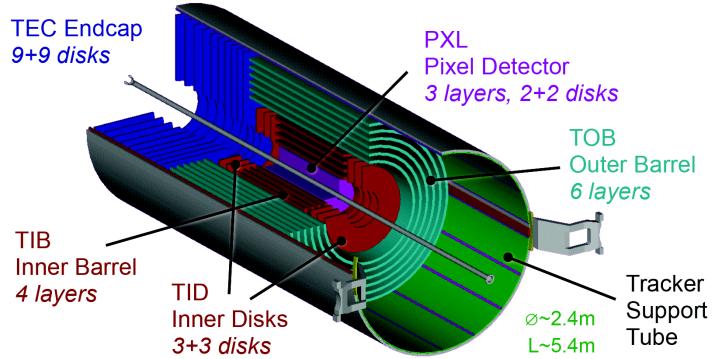


Figure 2.5.: Three-dimensional schematic view of the CMS silicon tracker.

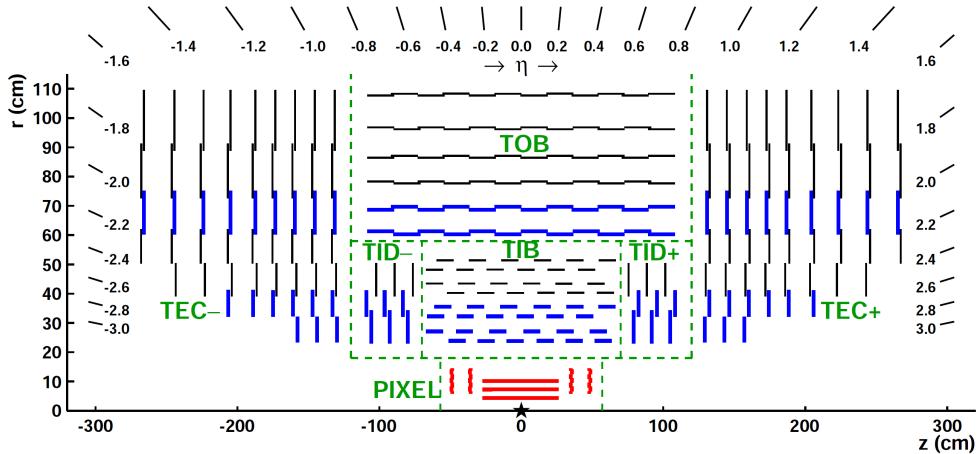


Figure 2.6.: Pictorial view of a tracker slice in the r - z plane. Pixel modules are shown in red, single-sided strip modules are depicted as black thin lines and strip stereo modules are shown as blue thick lines.

secondary vertices. The pixel detector is placed in the closest position to the collision point, where the amount of radiation is larger. It is placed in the region $|\eta| < 2.5$ and consists of three cylindrical layers 53 cm long in the barrel region, located at $r = 4.4, 7.3$ and 10.2 cm, and two pairs of endcap disks with radii between 6 and 15 cm at $z = \pm 34.5$ and ± 46.5 cm, covering a total area of about 1 m^2 . The detector is composed of many modules, for a total of 768 in the barrel and 672 in the endcaps. Each endcap is composed of 24 segments, each one tilted with respect to the adjacent ones and containing 7 modules. Each module consists of several units which contain a highly segmented silicon sensor with a thickness of 250 μm . In order to achieve an optimal vertex position resolution in both the (r, ϕ) and 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 μm size oriented along the (r, ϕ) direction in the barrel region, and

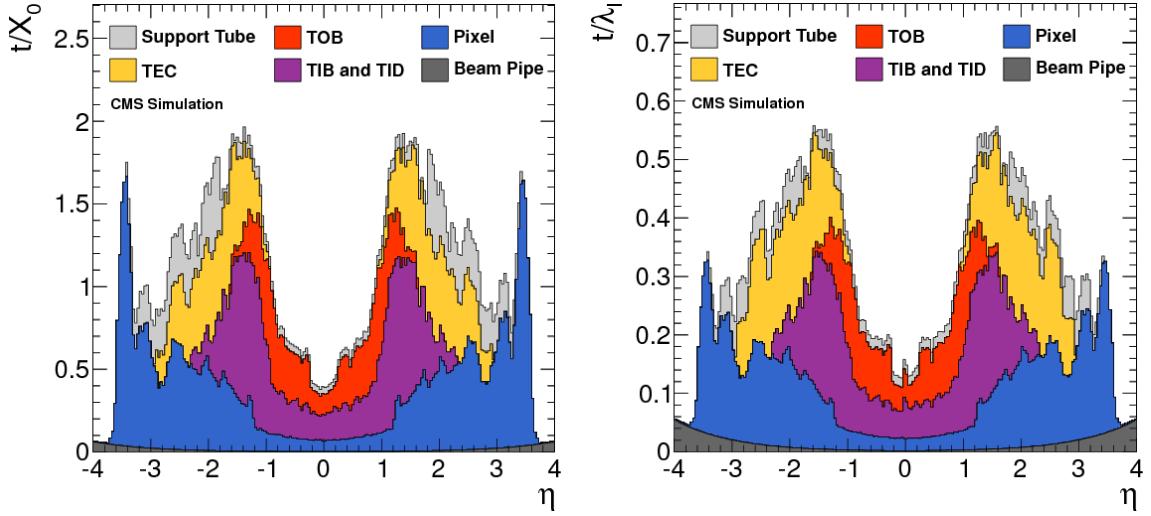


Figure 2.7.: Total thickness t of the tracker material expressed in units of X_0 (left) and λ_I (right), as a function of η . The contribution to the total material budget of each part of the detector is shown.

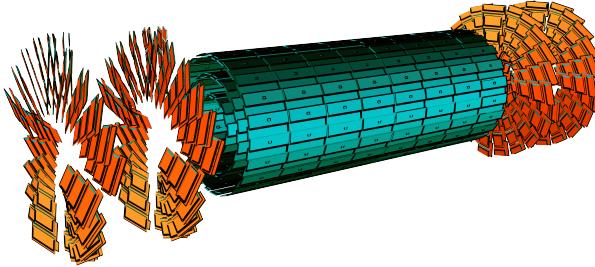


Figure 2.8.: Schematic view of the CMS pixel detector.

576 along the z -direction in the endcap region. The achievable hit reconstruction resolution is
577 about $10 - 15 \mu\text{m}$ in the barrel and $15 \mu\text{m}$ in the endcaps.

578 The microstrip detector

579 In this region of the detector the radiation flow is low enough to allow the use of a
580 less segmented device, such as the silicon microstrip detector. The microstrip tracker is
581 composed of 15148 silicon modules, covering a total area of about 193 m^2 with a total of
582 9.3 million strips. Two types of modules are installed: single sided modules consist of one
583 sensor stucked onto a carbon fiber support together with the readout electronics, with the
584 silicon strips laying along the z direction in the barrel and along the (r, ϕ) direction in the
585 endcaps. The other type of module, referred to as stereo-module, consists of two sensors
586 stucked together back to back and tilted of a relative angle of 100 mrad. This combination
587 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 $r = 1.1$ 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 $|z| < 65$ cm. In this region the detectors have a thickness of 300 μm and the strips are separated by a variable pitch between 80 and 120 μm . 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, ϕ) direction, about 23 – 34 μm , with respect to the z direction, where a resolution of about 230 μm 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 μm) and the strips are longer (190 mm). The spatial resolution varies in the range 25 – 52 μm in the (r, ϕ) direction, and is about 530 μm in the z coordinate in the stereo modules.

The endcaps are also made up of two parts:

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

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

The alignment of the tracker modules is very important to obtain a high spatial resolution. Deviations are caused by assembly inaccuracy, deformations due to cooling and stress from the magnetic field. Therefore, three methods are used for the tracker alignment. The

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

geometry was determined during the assembly to an accuracy of 80 to 150 μm . An infrared laser system is used for continuous monitoring of the position of selected tracker modules. The final alignment is done with tracks from well known physics processes, e.g. cosmic muons or muon pairs from the J/Ψ , Υ or Z decays.

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) [28, 29], shown in Fig. 2.9, is a hermetic homogeneous calorimeter with cylindrical geometry, composed of many scintillating crystals of lead tungstate (PbWO_4) with a truncated pyramidal shape. As the other detectors it consists of two parts, the ECAL barrel (EB), which contains 61200 crystals, and two endcaps (EE) containing 7324 crystals each one.

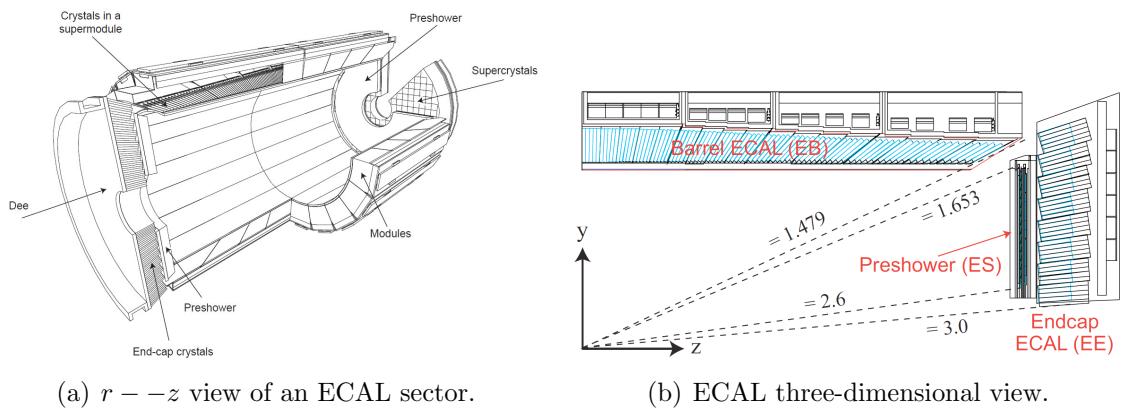


Figure 2.9.: Schematic representation of the CMS electromagnetic calorimeter.

The characteristics of the PbWO_4 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 radius⁴ (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 5×5 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, $26 \times 26 \text{ cm}^2$

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

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 η and ϕ 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 γ from those produced by forward emitted π^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:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2 \quad (2.4)$$

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 quantum efficiency⁵. The noise term b includes the contributions of pile up events and electronic noise, both due to the photodetector and preamplifier. These contributions depend on η and on the LHC operational luminosity. The constant term c , dominant at high energies, takes into account several contributions. The most relevant are the non-uniformity of the longitudinal light collection, the intercalibration errors and the leakage of energy from the rear side of the crystal. The ECAL barrel resolution for electrons was measured using test beams to be:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\% \text{ GeV}^{1/2}}{\sqrt{E}}\right)^2 + \left(\frac{12\% \text{ GeV}^{1/2}}{E}\right)^2 + (0.3\%)^2 , \quad (2.5)$$

where E is the energy measured in GeV.

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

680 2.2.4. The hadron calorimeter (HCAL)

681 The hadron calorimeter (HCAL) [30] is used together with ECAL to make a complete
 682 calorimetric system for the jet energy and direction measurement. Moreover, thanks
 683 to its hermetic structure, it can measure the energy imbalance in the transverse plane,
 684 E_T^{miss} , a typical signature of non interacting particles, such as neutrinos. The HCAL is
 685 a sampling calorimeter covering the region $|\eta| < 5$. As shown in Fig. 2.10, it is divided
 686 in four sub-detectors: HB (*Barrel Hadronic Calorimeter*), located in the barrel region
 687 inside the solenoid, extending up to $|\eta| < 1.4$; HE (*Endcap Hadronic Calorimeter*), placed
 688 in the endcaps region inside the magnet, covering the region $1.3 < |\eta| < 3$ and partially
 689 overlapping with the HB coverage; HO (*Outer Hadronic Calorimeter*), also known as
 690 *tail-catcher*, placed along the inner wall of the magnetic field return yoke, just outside of
 691 the magnet; HF (*Forward Hadronic Calorimeter*), a sampling calorimeter consisting of
 692 quartz fibers sandwiched between iron absorbers, consisting of two units placed in the very
 693 forward region ($3 < |\eta| < 5$) outside the magnetic coil. The quartz fibers emit Cherenkov
 694 light with the passage of charged particles and this light is detected by radiation resistant
 photomultipliers. In order to maximize particle containment for a precise missing transverse



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

696 energy measurement, the amount of absorber material was maximized, reducing therefore
 697 the amount of the active material. Since HCAL is mostly placed inside the magnetic coil,
 698 a non-magnetic material like brass was chosen as absorber. HB and HE are therefore
 699 made with 5 cm brass absorber layers interleaved with 3.7 mm plastic scintillators. The
 700 scintillation light is collected by wavelength shifting (WLS) fibres and read out by hybrid
 701 photodiodes (HPD). The granularity of the calorimeter is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for
 702 $|\eta| < 1.6$ 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
 703 the wheel that compose the return yoke and is divided in 12 sectors, each one covering
 704 a 30° angle in ϕ . It consists of scintillating layers, with the same granularity as HB, and
 705 the solenoid coil is used as an additional absorber to increase the effective depth of the
 706 calorimeter in the barrel region, which is extended up to $11.8 \lambda_I$, thus improving the energy
 707 resolution.

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

$$\begin{aligned} \left(\frac{\sigma_E}{E}\right)^2 &= \left(\frac{90\% \text{GeV}^{1/2}}{\sqrt{E}}\right)^2 + (4.5\%)^2 && \text{in the barrel/endcap ,} \\ \left(\frac{\sigma_E}{E}\right)^2 &= \left(\frac{172\% \text{GeV}^{1/2}}{\sqrt{E}}\right)^2 + (9\%)^2 && \text{in the HF ,} \end{aligned} \quad (2.6)$$

710 where E is expressed in GeV.

711 2.2.5. The muon system

712 The CMS muon system [31] is dedicated to the identification and measure of high p_T muons,
 713 in combination with the tracker. The system is placed outside the magnetic coil, embedded
 in the return yoke, to fully exploit the 1.8 T return flux. As shown in Fig. 2.11, the system

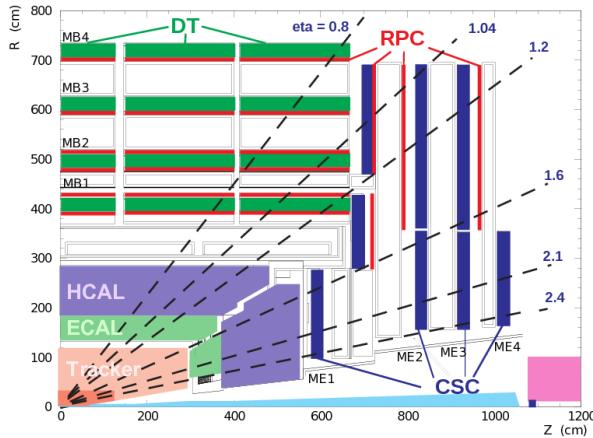


Figure 2.11.: Schematic view of a quadrant of the CMS muon system.

714
 715 consists of three types of independent gaseous particle detectors:

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

721 The DT system is placed in the region of the barrel with $|\eta| < 1.2$, where the magnetic
 722 field is sufficiently weak and homogeneous. Along the longitudinal direction, the barrel
 723 region is divided in 5 wheels, which are subdivided in 12 sectors covering a 30° 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 CO₂ (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, ϕ) 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

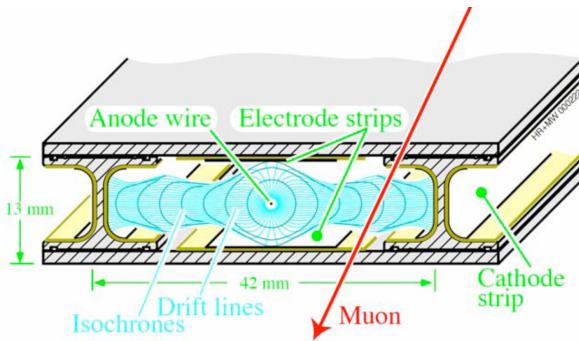


Figure 2.12.: Schematic representation of a drift tube chamber, showing the drift lines in presence of magnetic field.

wire between two parallel aluminium plates with ‘I’ shaped spacer cathodes, isolated from the aluminium plates with polycarbonate plastic, and the hit resolution is about 100 μm in both (r, ϕ) and (r, z) directions.

In the endcaps, the high and non-uniform magnetic field and the particle rate do not allow to use drift tubes detectors to perform measurements. Therefore, a solution based on the CSC detector has been adopted. CSC are multi-wire proportional chambers with the cathodes segmented into strips oriented radially and transversely with respect to the anode wires (see Fig. 2.13), allowing a simultaneous measurement of two coordinates (r through the wires and ϕ using the strips). The CSC chambers are filled with a gas mixture of Ar (40%), CO₂ (50%) and CF₄ (10%) and provide a spatial resolution of about 80–85 μm . The drift path of the charge carriers is shorter with respect to the drift tubes, therefore these detectors can be placed in regions with higher flows of charged particles and less homogeneous magnetic fields. The CSC coverage is $0.8 < |\eta| < 2.4$.

RPCs are used both in barrel and endcaps, complementing DT and CSC systems, in order to ensure robustness and redundancy to the muon spectrometer. RPCs are gaseous detectors characterized by a coarse spatial resolution, but are able to perform precise time measurements, comparable with the ones provided by scintillators. These chambers are made of 4 bakelite planes, with a bulk resistivity of 10^{10} – $10^{11} \Omega\text{cm}$. The 2 mm gap between the plates is filled with a mixture of C₂H₂F₄ (94.5%) and Isobutane. The central part of the chamber is equipped with insulated aluminum strips, used to collect the signal generated by

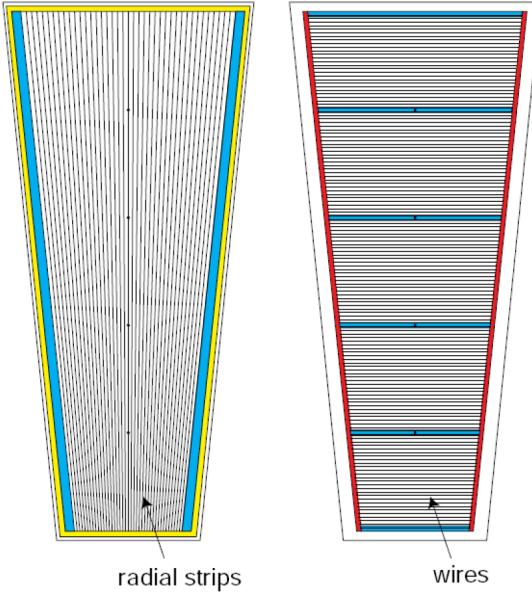


Figure 2.13.: Schematic representation of CSC cathode (left) and anode (right) panels.

crossing particles. In the barrel the strips are rectangularly segmented and run along the beam axis, whereas the endcaps are equipped with trapezoidal shaped strips. The detector operates in avalanche mode, and covers the region $|\eta| < 2.1$.

2.3. The CMS trigger system

The LHC can provide proton-proton interactions at a crossing frequency of 40 MHz and, for each bunch crossing, several collisions can occur (approximately 20 at the nominal instantaneous luminosity). Since it is impossible to store and process the large amount of data associated with the resulting large number of events, a drastic rate reduction has to be achieved. In fact the speed at which data can be written to mass storage is limited and, moreover, the vast majority of events produced is not interesting for physics analyses, because it involves low transverse momentum interactions (also called *minimum bias events*). The task of reducing this rate is accomplished by the CMS trigger system, which is the start of the physics event selection. CMS makes use of a two-stage trigger system, consisting of a *Level-1* trigger (L1) [32] and a *High Level Trigger* (HLT) [33].

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

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

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

775 2.3.1. The Level-1 trigger

776 The L1 trigger is responsible for the identification of electrons, muons, photons, jets and
 777 missing transverse energy. It is required to have a high and carefully understood efficiency.
 778 Its output rate and speed are limited by the readout electronics and by the performances
 779 of the data acquisition (DAQ) system [33]. It consists of three main subsystems:

- 780 • L1 Calorimeter Trigger;
- 781 • L1 Muon Trigger;
- 782 • L1 Global Trigger.

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

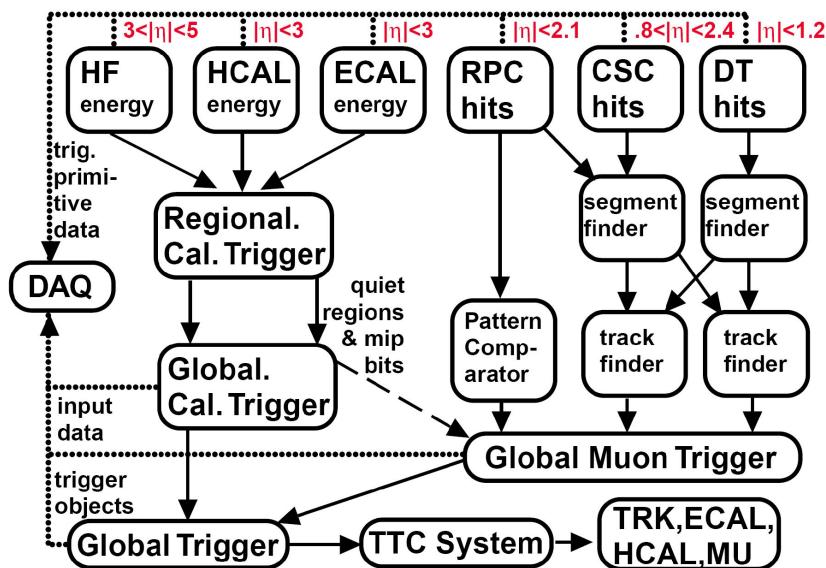


Figure 2.14.: Schematic representation of the Level-1 trigger components.

785

786 L1 Calorimeter Trigger

787 Controllare se stato cambiato qualcosa nel Run2

788 The input for the L1 Calorimeter Trigger are calorimeter towers, which are clusters
 789 of signals collected both from ECAL and HCAL. Towers are calculated by calorimeter
 790 high level readout circuits, called Trigger Primitive Generators. The Regional Calorimeter
 791 Trigger identifies electron, photon, τ and jet candidates together with their transverse energy
 792 and sends the information to the Global Calorimeter Trigger. The Global Calorimeter

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

795 L1 Muon Trigger

796 Controllare se stato cambiato qualcosa nel Run2

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

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

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

808 DTs are equipped with Track Identifier electronics, which is able to find groups of
809 aligned hits in the four chambers of a superlayer. Those Track Segments are sent to the
810 DT Track Correlator that tries to combine segments from two superlayers, measuring the
811 ϕ angle. The best two segments are sent to the DT Track Finder that builds tracks and
812 sends them to the Global Muon Trigger.

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

816 L1 Global Trigger

817 Controllare se stato cambiato qualcosa nel Run2

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

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

825 2.3.2. The high level trigger (HLT)

826 The High Level Trigger is designed to reduce the L1 output rate down to about 1000 events/s,
827 which is the amount that will be written to mass storage. HLT code runs on commercial

processors and performs reconstruction using the information from all sub-detectors. Events passing the HLT are stored on local disk or in CMS Tier 0⁶.

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

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

- 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.

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

⁶The 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.

Chapter 3.

854 Reconstruction and identification of 855 physics objects

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

860 3.1. The Particle Flow technique

861 The Particle Flow (PF) event reconstruction technique [34] aims at the reconstruction
862 and identification of all the stable particles in the event, i.e. electrons, muons, photons,
863 charged and neutral hadrons, with a thorough combination of the information from all CMS
864 sub-detectors, in order to determine their energy, direction and type. These individual
865 particles are then used, for example, to build jets, to measure the missing transverse energy
866 E_T^{miss} , to reconstruct the τ from their decay products, to quantify the charged lepton
867 isolation and to tag b-jets.

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

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

878 The good performance of the tracking system are achieved by means of an iterative
879 tracking strategy [35], based on the Kalman Filter algorithm [36]. The basic idea of iterative
880 tracking is that initial iterations search for tracks that are easiest to find, e.g. high p_T tracks
881 produced near the interaction region. After each iteration, hits associated to reconstructed
882 tracks are removed from the hit collection, thereby reducing the combinatorial complexity
883 and simplifying the subsequent iterations, which aim at finding more complicated set of

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

- 891 • *seeding*: initial track candidates are obtained using 2 or 3 hits in the innermost layers
 892 (these proto-tracks are called seeds);
- 893 • *pattern recognition*: this step is based on Kalman Filter and searches for hits in the
 894 outer layers that could be associated to the initial track candidate, reconstructing the
 895 particle trajectory;
- 896 • *track fitting*: in this step a fit of the trajectory is performed, using its associated hits
 897 and providing an estimate of the track parameters (p_T , η , ϕ , charge, etc.);
- 898 • *selection*: finally tracks are selected based on quality requirements.

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

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

- 913 • Muons: a *global muon* gives rise to a *PF muon* if its combined p_T measurement is
 914 compatible within 3 standard deviation with the one provided by the sole tracker. The
 915 corresponding track is removed from the block;
- 916 • Electrons: electrons tend to give rise to short tracks, and to lose energy by Bremsstrahlung
 917 in the tracker layers on their way to the calorimeter. The link between a charged-
 918 particle track (refitted with the Gaussian-Sum Filter (GSF) [37]) and one or more
 919 ECAL clusters identifies a *PF electron*. After the identification, the corresponding
 920 tracks and clusters are removed from the block.
- 921 • Charged hadrons: the remaining tracks give rise to *PF charged hadrons*. Tracks can be
 922 linked to ECAL and HCAL clusters, and the energy is determined taking into account
 923 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^{miss}* , which is described in Sec. 3.5. The missing transverse energy, E_T^{miss} , is defined as the modulus of \vec{p}_T^{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 [38]: 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*¹. 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

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

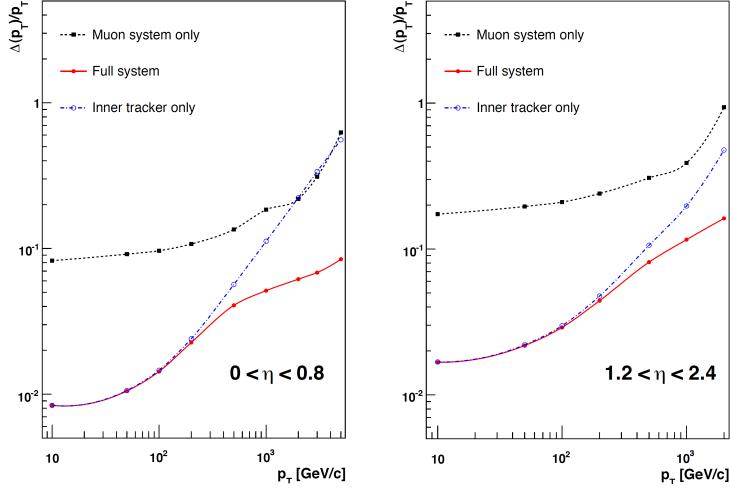


Figure 3.1.: Muon p_T resolution as a function of the muon p_T in the barrel (left) and in the endcap (right) regions. The resolution is provided for the measurement using the tracking system or the muon system only, as well as for the combination of the two methods.

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

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

963 3.2.2. Muon isolation

964 One of the most powerful requirements to select prompt muons, as the ones produced from
 965 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/Ψ 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 η and ϕ coordinates of the muon. After applying these corrections, the relative p_T resolution, $\sigma(p_T)/p_T$, is measured as a function of η and ϕ 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 η coordinate spread over the ϕ 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. [39, 40], 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 η and ϕ coordinates, respectively;
- $\sigma_{in,in}$: a variable related to the calorimeter shower shape, measuring the width of the ECAL supercluster along the η direction computed for all the crystals in the 5×5 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 η 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	≤ 2	≤ 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, ρ 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 ². 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. [40].

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

1053 3.2.7. Lepton identification and isolation efficiency

1054 The efficiency related to the identification and isolation selections applied to muons and
 1055 electrons are generally estimated both in data and simulation and the simulated events are
 1056 corrected for the observed differences by means of a scale factor (SF), defined as the ratio
 1057 of the efficiency measured in data and simulation, i.e. $SF = \varepsilon_{\text{data}}/\varepsilon_{\text{MC}}$.

1058 The identification and isolation efficiencies are measured using a Tag and Probe technique.
 1059 The Tag and Probe technique is a method to estimate the efficiency of a selection on data.
 1060 It can be applied whenever one has two objects in one event, by using one of the two, the
 1061 *tag*, to identify the process of interest, and using the second, the *probe*, to actually measure
 1062 the efficiency of the selection being studied. Concerning the electron and muon case, the
 1063 Tag and Probe method uses a known mass resonance (e.g. J/Ψ , Z) to select particles of the
 1064 desired type, and probe the efficiency of a particular selection criterion on these particles.
 1065 In general the *tag* is an object that passes a set of very tight selection criteria designed
 1066 to isolate the required particle type. Tags are often referred to as a “golden” electrons or
 1067 muons and the fake rate for passing tag selection criteria should be very small. A generic
 1068 set of the desired particle type (i.e. with potentially very loose selection criteria) known
 1069 as *probe* is selected by pairing these objects with tags such that the invariant mass of the
 1070 combination is consistent with the mass of the resonance. Combinatorial backgrounds may
 1071 be eliminated through any of a variety of background subtraction methods such as fitting,
 1072 or sideband subtraction. The definition of the probe objects depend on the specifics of
 1073 the selection criterion being examined. The simple expression to get the efficiency ε as a
 1074 function of p_T and η is given below:

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

1075 For the estimation of the electron or muon identification efficiency, the tag is chosen to
 1076 be a well identified and isolated electron or muon, while the probe is chosen as an electron
 1077 identified with loose selections. The invariant mass of the tag-probe pair is required to
 1078 be within a Z boson mass window (the effect of changing the Z mass window is included
 1079 as a systematic uncertainty). After that, the probe is required to pass the identification
 1080 selections discussed before for electrons and muons, and the efficiency is computed both
 1081 in data and simulation. A scale factor is then calculated by taking the ratio of the two
 1082 efficiencies and applied to reweight simulated events.

1083 There are two methods to measure the efficiencies: the counting method consists in
 1084 simply computing the ratio of probe events that pass the selections and total number of
 1085 probe events, as shown in Eq. (3.3). This method can be used when the tag requirement
 1086 selects a very pure set of events, with a small background contribution. The other approach
 1087 is the fitting method, which is used when the background contamination is not negligible.
 1088 In this latter case, which represents the commonly used method for estimating the lepton
 1089 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 η 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).

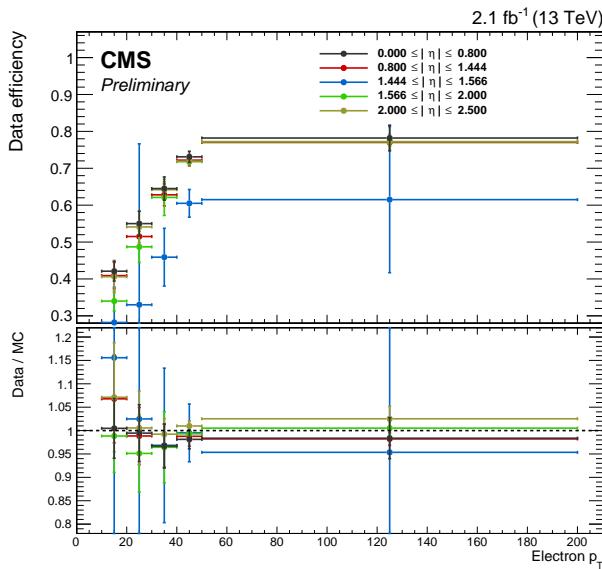


Figure 3.2.: Typical electron identification and isolation efficiencies in data (top panel) and data/simulation scale factor (bottom panel), as a function of the electron p_T and for different η bins.

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

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

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:

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

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:

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

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

1137 of particles, and play a predominant role in hadron colliders like the LHC, where the
1138 production cross section is very large. In this section, the jet reconstruction techniques
1139 used in CMS are described.

1140 3.3.1. Jet reconstruction in CMS

1141 The majority of physics analyses involving jets in the final state make use of particle flow
1142 jets. The PF jets are reconstructed using the technique described in Sec. 3.1, clustering all
1143 particles reconstructed with the PF algorithm, without any distinction of type and energy
1144 threshold. This method allows a remarkable improvement in the jet momentum and spatial
1145 resolutions with respect to the calorimeter jets, which are instead reconstructed using solely
1146 the information from the calorimeters, as the use of the tracker information provides a
1147 better p_T resolution for the charged particles constituting the jets³.

1148 Jets are defined through sequential, iterative clustering algorithms that combine the
1149 four-momenta of input particles until certain conditions are satisfied and jets are formed [41].
1150 Several algorithms are available for jet clustering, characterized by different features. From
1151 a theoretical point of view, an ideal jet clustering algorithm should fulfil the following
1152 requirements [42]:

- 1153 • *Infrared safety*: infrared singularities should not appear in the perturbative calculations
1154 and the solutions of the algorithm should be insensitive to soft radiation in the event;
- 1155 • *Collinear safety*: collinear singularities should not appear in the perturbative calculations
1156 and jets should be insensitive to collinear radiation in the event;
- 1157 • *Invariance under boosts*: the solutions of the algorithm should be the same indepen-
1158 dently of boosts in the longitudinal direction. This is particularly important for pp
1159 colliders, where the centre-of-mass of the individual proton proton collisions is typically
1160 boosted along the beam direction;
- 1161 • *Order independence*: the algorithm should find the same jets at parton, particle and
1162 detector level;
- 1163 • *Straightforward implementation*: the algorithm should be straightforward to implement
1164 in perturbative calculations.

1165 The ideal algorithm should also follow some experimental attributes. Among them, the
1166 performance of the algorithm should be as independent as possible of the detector that
1167 provides the data, the algorithm should not amplify the inevitable effects of resolution
1168 smearing and angle bias and should not be strongly affected by pile up and high beam
1169 luminosities. Furthermore, the algorithm should be easy to implement, efficient to identify
1170 all possible jet candidates and should keep at an acceptable level the necessary computing
1171 resources.

1172 Two main classes of jet clustering algorithms can be defined. The first one consists
1173 in the “cone” recombination, where jets are reconstructed associating together particles

³On 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 ΔR in the η - ϕ 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 ϕ_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 Δ_{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.

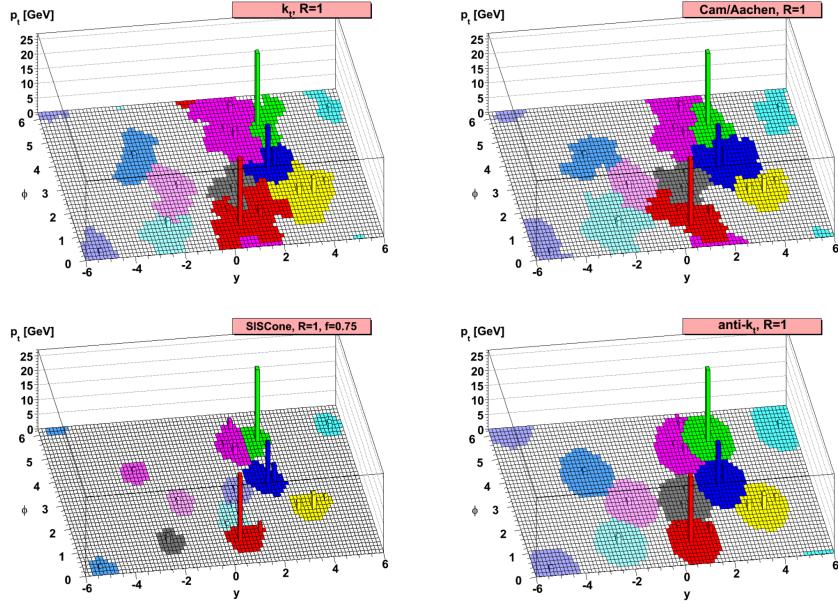


Figure 3.3.: Jets reconstructed with different algorithms starting from the same set of simulated particles. The jets reconstructed with the sequential recombination algorithms described in the text are shown, as well as with the SISCone algorithm.

1210 3.3.2. Jet energy correction

1211 The purpose of jet energy correction is to relate, on average, the jet energy measured in the
 1212 detector to the true energy of the corresponding final state particle or parton jet. The latter
 1213 is obtained in simulation by clustering, with the same algorithm used for jets in the detector,
 1214 all the stable particles, i.e. with $c\tau > 1$ cm, produced in the event excluding neutrinos.
 1215 This mismatch is mainly ascribable to the non uniform and linear response of the CMS
 1216 calorimeters, to the electronics noise and to pile up. For this reason, CMS has developed a
 1217 sequential procedure to calculate and apply the *jet energy corrections* (JEC) [43].

1218 The correction is applied as a multiplicative factor \mathcal{C} to each component of the raw jet
 1219 four-momentum p_μ^{raw} (components are indexed by μ in the following):

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

1220 where p_μ^{cor} is the corrected jet four-momentum. The correction factor is composed of the
 1221 offset correction C_{offset} , the MC calibration factor C_{MC} , and the residual calibrations C_{rel}
 1222 and C_{abs} for the relative and absolute energy scales, respectively. The offset correction
 1223 removes the extra energy due to noise and pile up, and the MC correction removes the bulk
 1224 of the non-uniformity in η and the non-linearity in p_T . Finally, the residual corrections
 1225 account for the small differences between data and simulation. The various components are

1226 applied in sequence as described by the equation below:

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

1227 where p'_{T} is the jet p_{T} after applying the offset correction and p''_{T} is the jet p_{T} after applying
1228 all previous corrections. Each component is briefly described in the following sections.

1229 Offset correction

1230 The offset correction purpose is to estimate and subtract, on average, the energy contribution
1231 that is not associated with the hard scattering in the event. The energy excess includes
1232 contributions from electronics noise and pile up. The approach followed for the estimation of
1233 the offset correction is known as *Jet Area Method*. For each event, an average p_{T} -density per
1234 unit area, ρ , is estimated, characterizing the soft jet activity. This p_{T} -density represents the
1235 combination of the underlying event, the electronics noise and the pile up effects. The two
1236 latter components contaminate the hard jet energy measurement and need to be corrected
1237 for with the offset correction. The key element for this approach is the jet area A_j . A
1238 very large number of infinitely soft four-momentum vectors (soft enough not to change
1239 the properties of the true jets) are artificially added in the event and clustered by the jet
1240 algorithm together with the true jet components. The extent of the region in the η - ϕ
1241 space occupied by the soft particles clustered in each jet defines the active jet area. The
1242 p_{T} -density ρ is calculated with the k_t algorithm with a distance parameter $R = 0.6$. The
1243 quantity ρ is estimated event by event as the median of the distribution of the variable
1244 $p_{\text{T}j}/A_j$, where j runs over all jets in the event, and is not sensitive to the presence of hard
1245 jets in the event. At the detector level, the measured density ρ is the convolution of the true
1246 particle-level activity (underlying event, pile-up) with the detector response to the various
1247 particle types. The event-by-event and jet-by-jet offset correction can thus be defined as:

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

1248 In the formula above, $\langle \rho_{\text{UE}} \rangle$ represents the average p_{T} -density component due to the
1249 underlying event and electronics noise, and is measured in events with exactly one recon-
1250 structed primary vertex, i.e. no pile up.

1251 An additional pile up subtraction method that is used in CMS is called *Charged Hadron*
1252 *Subtraction*. This method makes use of PF jets and exploits the excellent CMS tracking
1253 capabilities to identify and remove charged hadrons inside jets, which are known to originate
1254 from pile up vertices. This is a particle-by-particle method that is applied to jets before
1255 calculating the offset correction.

1256 **MC calibration correction**

1257 The MC calibration is based on the simulation and corrects the energy of the reconstructed
 1258 jets such that it is equal on average to the energy of the generated jets. In order to evaluate
 1259 this correction, simulated QCD events are generated and then processed through the CMS
 1260 detector simulation, based on the GEANT4 software. The jet reconstruction in simulation
 1261 is identical to the one applied to the data. Each reconstructed jet is spatially matched, in
 1262 the η - ϕ space, to a generated jet by requiring $\Delta R < 0.25$. In each bin of the generated
 1263 jet transverse momentum p_T^{gen} , the response variable $\mathcal{R} = p_T^{\text{reco}}/p_T^{\text{gen}}$ and the reconstructed
 1264 jet transverse momentum p_T^{reco} , are saved. The average correction in each bin is therefore
 1265 defined as:

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

1266 and is expressed as a function of the average reconstructed jet p_T , $\langle p_T^{\text{reco}} \rangle$.

1267 **Relative jet energy scale**

1268 The goal of the relative jet energy scale correction is to make the jet response flat versus η .
 1269 This is achieved by employing a Tag and Probe technique, selecting di-jet events in data.
 1270 The size of this residual correction is of the order of 2–3% in the central η region, while it
 1271 goes up to about 10% in the forward region.

1272 **Absolute jet energy scale**

1273 The goal of the absolute jet energy scale correction is to make the jet response at versus
 1274 p_T . The absolute jet energy response is measured in the reference region $|\eta| < 1.3$ with
 1275 the *Missing Transverse Energy Projection Fraction* (MPF) method [44], using $\gamma +$ jets and
 1276 $Z +$ jets events. The method is used to estimate the absolute jet energy correction and is
 1277 based on the fact that $\gamma +$ jets and $Z +$ jets events have no intrinsic E_T^{miss} and that, at parton
 1278 level, the γ and Z boson are perfectly balanced by the hadronic recoil in the transverse
 1279 plane.

1280 **Jet energy uncertainties**

1281 The uncertainties in the jet energy estimation arise from several different sources. Generally
 1282 these can be categorized as follows:

- 1283 • physics modelling in MC such as showering, underlying event, etc.;
- 1284 • MC modelling of true detector response and properties;
- 1285 • potential biases in the methodologies used to estimate the corrections.

1286 The sources are combined in different groups: absolute scale, relative scale, pile up, jet
 1287 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 η 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.

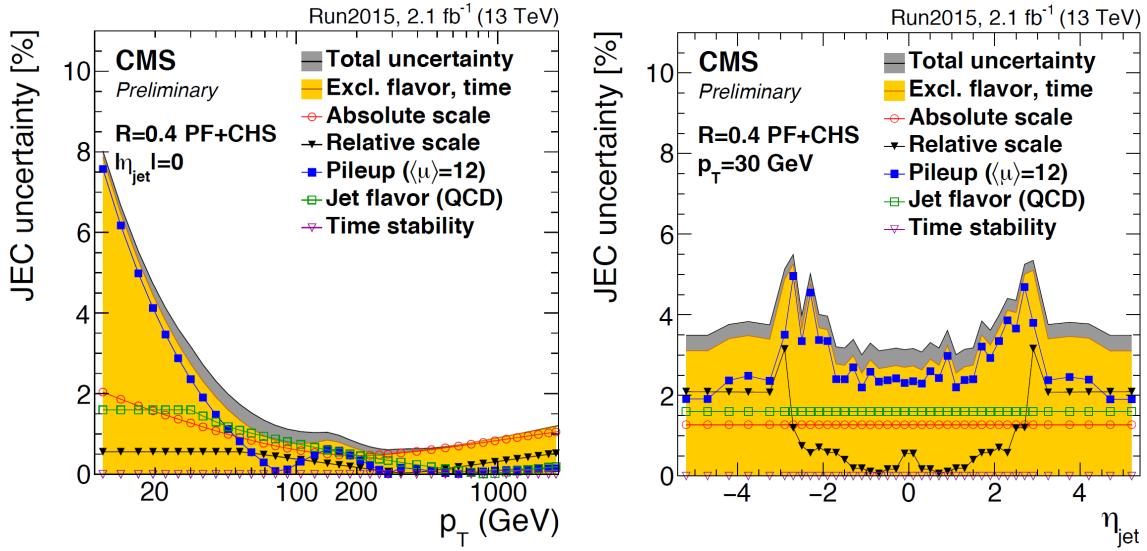


Figure 3.4.: JEC uncertainties as a function of p_T (left) for jets reconstructed with $\eta = 0$ and as a function of η (right) for jets with $p_T = 30$ GeV. All jets are reconstructed with the PF technique and using the anti- k_t algorithm with $R = 0.4$, after applying the CHS correction. Results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

1291 Jet energy resolution

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

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

jet identification?

1303 3.4. Jet b tagging

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

1306 to reduce the otherwise overwhelming background from these processes to channels involving
 1307 jets from gluons (g) and light-flavour quarks (u, d, s), and from c -quark fragmentation.

1308 Algorithms for b jet identification (also known as b tagging algorithms) exploit the
 1309 long life time of b hadrons present in jets originating from the hadronization of b quarks.
 1310 This long life time results in a decay of the b hadron that is displaced with respect to the
 1311 primary interaction vertex. This displacement of a few millimetres results in the presence
 1312 of displaced tracks from which a secondary vertex may be reconstructed. In addition, b
 1313 hadrons have a probability of around 20% to decay to a muon or electron. Hence, also the
 1314 presence of these charged leptons can be exploited for b jet identification techniques and
 1315 for measuring their performance with the collision data.

1316 A variety of reconstructed physics objects, as tracks, vertices and identified leptons,
 1317 can be used to build observables that discriminate between b and light-quark jets. Several
 1318 b tagging algorithms have been developed by CMS, each one based on different input
 1319 information. A common feature of all the algorithms is that each one yields a single
 1320 discriminator value for each jet, which measures the likelihood that the jet has been produced
 1321 by the hadronization of a b quark. The minimum thresholds on these discriminators define
 1322 loose (“L”), medium (“M”), and tight (“T”) operating points with a misidentification
 1323 probability for light-parton jets close to 10%, 1%, and 0.1%, respectively, at an average
 1324 jet p_T of about 80 GeV. The misidentification probability, also known as mistag rate, is
 1325 defined as the probability to wrongly identify a light-parton jet as a b -jet.

1326 Some of the algorithms make use of the track impact parameters (IP) with respect
 1327 to the primary vertex, defined as the distance between the primary vertex and the track
 1328 at their point of closest approach, to distinguish the decay products of a b hadron from
 1329 prompt tracks. The impact parameter has the same sign as the scalar product of the vector
 1330 pointing from the primary vertex to the point of closest approach with the jet direction.
 1331 Tracks originating from the decay of particles travelling along the jet axis will tend to have
 1332 positive IP values. In contrast, the impact parameters of prompt tracks can have positive
 1333 or negative IP values. The impact parameter significance, defined as the ratio of the IP to
 1334 its estimated uncertainty, is used as an observable.

1335 The *Track Counting* (TC) algorithm sorts tracks inside a jet by decreasing values of the
 1336 IP significance. Although the ranking tends to bias the values for the first track to high
 1337 positive IP significances, the probability to have several tracks with high positive values is
 1338 low for light-parton jets. Therefore the two different versions of the algorithm use the IP
 1339 significance of the second and third ranked track as the discriminator value. These two
 1340 versions of the algorithm are called *Track Counting High Efficiency* (TCHE) and *Track*
 1341 *Counting High Purity* (TCHP), respectively.

1342 A general extension of the TC algorithm, i.e. the *Jet Probability* (JP), combines the
 1343 IP information of several tracks inside the jet, using an estimate of the likelihood that all
 1344 tracks associated to the jet come from the primary vertex as a discriminating variable. A
 1345 variant of the JP algorithm also exists in which the four tracks with the highest impact
 1346 parameter significance get a higher weight in the jet probability calculation. This algorithm
 1347 is referred to as *Jet B-Probability* (JBP).

1348 A different approach consists in using the secondary vertices and the related kinematic
 1349 variables, together with displaced tracks information, to discriminate between b and non-b
 1350 jets. This algorithm is known as *Combined Secondary Vertex* (CSV)⁴. The magnitude
 1351 and direction of the vector connecting the primary and secondary vertices are used as
 1352 a discriminating variables and quality requirements are imposed to secondary vertex
 1353 candidates. In addition, the usage of displaced tracks information allows to increase the
 1354 efficiency for events where no secondary vertex is found. Several variables related to
 1355 secondary vertices and displaced tracks are used to build likelihood ratios that have a good
 1356 discriminating power.

1357 Two algorithms for reconstructing secondary vertices are exploited. For the first
 1358 algorithm, the tracks associated to jets and fulfilling some quality requirements are used in
 1359 the adaptive vertex reconstruction (AVR) algorithm [45]. The AVR is the algorithm used
 1360 for CMS analyses during the 8 TeV data taking. In contrast with this method, the Inclusive
 1361 Vertex Finder (IVF) algorithm is not seeded from tracks associated to reconstructed jets, but
 1362 instead makes use of all the tracks in the event, with appropriate selections, to reconstruct
 1363 the secondary vertices. The latter is the default algorithm used to reconstruct secondary
 1364 vertices for CMS analyses using 13 TeV data.

1365 A new b jet identification algorithm has been recently developed, combining the dis-
 1366 criminating provided by the JP and CSV algorithms with a Boosted Decision Tree (BDT)
 1367 technique. This combined multivariate algorithm (cMVA) is found to slightly improve the
 1368 b jet identification efficiency.

1369 The performance of these algorithms is determined using simulated $t\bar{t}$ events, selecting
 1370 events with at least one jet with $p_T > 30$ GeV. This is shown in Fig. 3.5, where the b jet
 1371 identification efficiency versus the misidentification probability is reported for the various
 1372 algorithms. This figure serves as an illustration as the b tagging performance depend on
 1373 the p_T and η distribution of the jets, and need to be checked for each analysis phase space.

1374 3.5. Missing transverse energy

1375 In hadron colliders the longitudinal momentum (along the beam axis) carried by the
 1376 incoming partons is not known, preventing the possibility to measure the total missing
 1377 energy. Nevertheless, the initial transverse momentum carried by the incoming partons
 1378 is expected to be zero, thereby, for the conservation of the momentum components, also
 1379 the net momentum of all the particles in the final state of collisions must be zero. The
 1380 missing transverse momentum (\vec{p}_T^{miss}) is the momentum imbalance in the transverse plane
 1381 of all the visible particles in the event, and its modulus is the missing transverse energy
 1382 (E_T^{miss}). The \vec{p}_T^{miss} vector is defined as the negative vectorial sum of transverse momenta of
 1383 all reconstructed PF objects, as shown in the following equation:

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

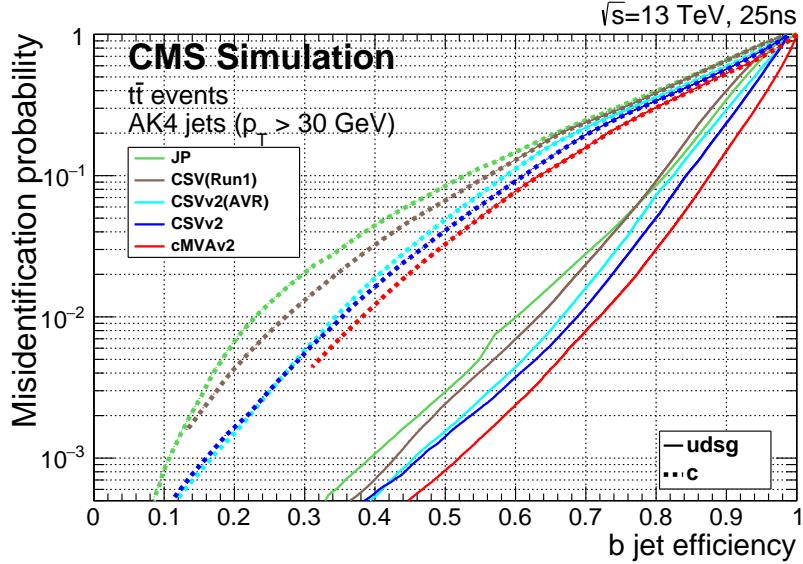


Figure 3.5.: Performance of the b jet identification efficiency algorithms demonstrating the probability for non-b jets to be misidentified as b jet as a function of the efficiency to correctly identify b jets. The curves are obtained on simulated $t\bar{t}$ events using anti- k_t jets clustered with $R = 0.4$ 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^{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^{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 [46, 47]. The bias on the E_T^{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^{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^{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 particles⁵. 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^{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^{miss} scale and resolution measurement

The performance (scale and resolution) of E_T^{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^{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^{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 $\vec{q}_T + \vec{u}_T + \vec{p}_T^{\text{miss}} = 0$.

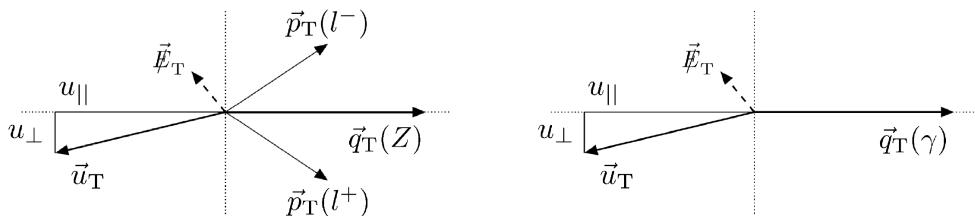


Figure 3.6.: Illustration of the $Z \rightarrow \ell^+ \ell^-$ (left) and photon (right) event kinematics in the transverse plane.

⁵These 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^{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^{miss} resolution. The resolutions $\sigma(u_{\parallel})$ and $\sigma(u_{\perp})$ obtained using recent 13 TeV data are shown in Fig. 3.7.

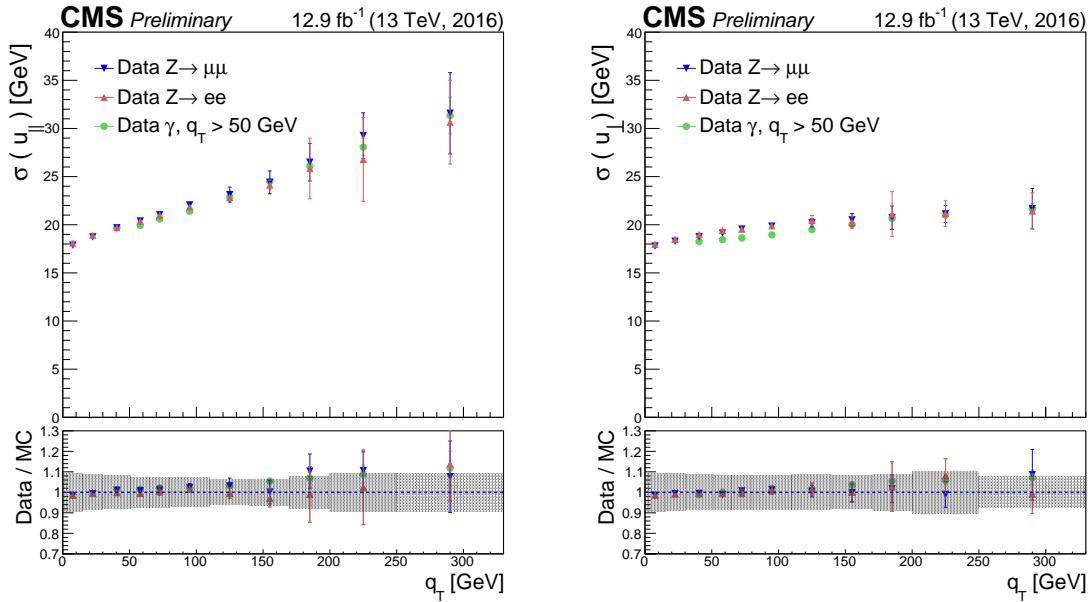


Figure 3.7.: Resolutions $\sigma(u_{\parallel})$ (left) and $\sigma(u_{\perp})$ (right) for $Z \rightarrow \mu^+\mu^-$, $Z \rightarrow e^+e^-$ and γ events as a function of the vector boson p_T . The upper panels show the resolution measured in data and the bottom panels the data to simulation ratio.

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, η . The η distribution is essentially driven by the PDF of the partons in the colliding hadrons, and it is only mildly sensitive to radiative corrections. The p_T^H distribution is instead sensitive to QCD radiative corrections. Considering the ggH production mode, at LO in perturbation theory, $\mathcal{O}(\alpha_s^2)$, the Higgs boson is always produced with p_T^H equal to zero. Indeed in order to have p_T different from zero, the Higgs boson has to recoil at least against one parton. Higher order corrections to the ggH process are numerically large and are known at NLO including full top quark mass dependence [48, 49], and at NNLO using the so-called large- m_t approximation [50–52], 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 [53].

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 [54, 55], 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 [56], 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.

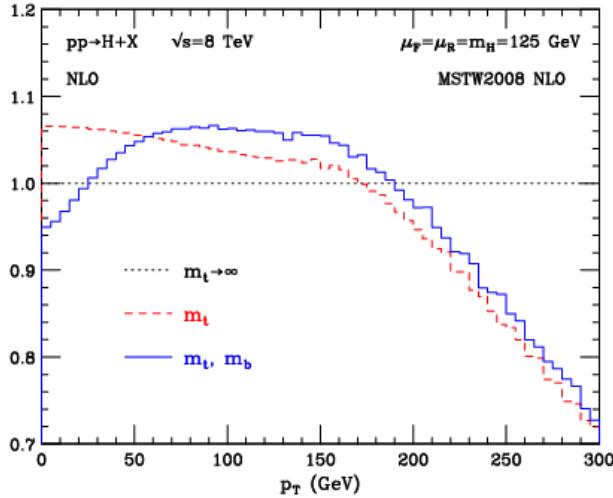


Figure 4.1.: Distribution of p_T^H computed at NLO (α_s^4) and divided by the calculation obtained in the large- m_t approximation. The red dashed line corresponds to the calculation including the top quark mass while the blue line refers to the calculation including also the bottom quark effects.

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. [57, 58], 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 [59], 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 [60]. 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 [61–63] and CMS [64, 65] 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 [66], and is

1482 characterized by good signal sensitivity. Such sensitivity allowed the observation of a Higgs
1483 boson at the level of 4.3 (5.8 expected) standard deviations for a mass hypothesis of 125.6
1484 GeV using the full LHC data set at 7 and 8 TeV [67].

1485 The measurement is performed in a fiducial phase space defined by kinematic requirements
1486 on the leptons that closely match the experimental event selection.

1487 The effect of the limited detector resolution, as well as the selection efficiency with respect
1488 to the fiducial phase space are corrected to particle level with an unfolding procedure [68],
1489 as explained in Sec. 4.7.

1490 4.2. Data sets, triggers and MC samples

1491 4.2.1. Data sets and triggers

1492 The data set used for the analysis corresponds to 19.4 fb^{-1} of proton-proton collisions at
1493 $\sqrt{s} = 8 \text{ TeV}$, collected by the CMS detector during 2012. Only data corresponding to good
1494 data taking quality are considered.

1495 Events are required to fire one of the unprescaled single-electron, single-muon or muon-
1496 electron triggers. Due the rather high LHC instantaneous luminosity the single-lepton
1497 triggers must have high HLT p_T thresholds, otherwise the rate of these triggers would be
1498 too large to be sustained. The double-lepton triggers allow to lower down the p_T thresholds
1499 while keeping a sustainable trigger rate, thus maintaining a good sensitivity to the Higgs
1500 boson signal, for which the lepton p_T can be rather small. A brief overview of the HLT
1501 p_T criteria on the leptons is given in Table 4.1. While the HLT lepton p_T thresholds of 17
1502 and 8 GeV for the double lepton triggers accommodate the offline lepton p_T selection of 20
1503 and 10 GeV, the higher p_T thresholds in the single lepton triggers help partially recovering
1504 double lepton trigger inefficiencies as a high p_T lepton is on average expected due to the
1505 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	$p_T > 17 \text{ and } 8 \text{ GeV}$

1506 The trigger is not simulated in MC samples but the combined trigger efficiency is
1507 estimated from data and applied as a weight to all simulated events, as described in
1508 Sec. 3.2.8.

1509 4.2.2. Monte Carlo samples

1510 Several Monte Carlo event generators are used to simulate the signal and background
 1511 processes:

- 1512 • The first version of the POWHEG program [69–73] (POWHEG V1) provides event
 1513 samples for the $H \rightarrow WW$ signal for the gluon fusion (ggH) and VBF production
 1514 mechanisms, as well as $t\bar{t}$ and tW processes [74], with NLO accuracy.
- 1515 • The $qq \rightarrow W^+W^-$, Drell-Yan, ZZ, WZ, $W\gamma$, $W\gamma^*$, tri-bosons and $W+jets$ processes
 1516 are generated using the MADGRAPH 5.1.3 [75] event generator.
- 1517 • The $gg \rightarrow W^+W^-$ process is generated using the GG2WW 3.1 generator [76] and its
 1518 cross section is scaled to the approximate NLO prediction [77, 78].
- 1519 • The VH process is simulated using PYTHIA 6.426 [79].

1520 For leading-order generators samples, the CTEQ6L [80] set of parton distribution functions
 1521 (PDF) is used, while CT10 [81] is used for next-to-leading order (NLO) ones. Cross section
 1522 calculations at next-to-next-to-leading order (NNLO) are used for the $H \rightarrow WW$ process [82].
 1523 The $H \rightarrow WW$ process simulation is reweighted so that the p_T^H spectrum and inclusive
 1524 production cross section closely match the SM calculations that have NNLO+NNLL pQCD
 1525 accuracy in the description of the Higgs boson inclusive production, in accordance with
 1526 the LHC Higgs Cross Section Working Group recommendations [66]. The reweighting of
 1527 the p_T^H spectrum is achieved by tuning the POWHEG generator, as described in detail in
 1528 Ref. [83]. Cross sections computed with NLO pQCD accuracy are used for the background
 1529 processes [66]. The contribution of the $t\bar{t}H$ production mechanism is checked to be negligible
 1530 in each bin of p_T^H (below 1%) and is not included among the different production mechanisms.
 1531 In Fig. 4.2 the relative fraction of the four production mechanisms is shown for each p_T^H bin.

1532 For all processes, the detector response is simulated using a detailed description of the
 1533 CMS detector, based on the GEANT4 package [84].

1534 Minimum bias events are superimposed on the simulated events to emulate the additional
 1535 proton-proton interactions per bunch crossing. The pile up multiplicity in simulated events
 1536 has been generated poissonianly sampling from a distribution similar to the one expected
 1537 from data. The simulated events are reweighted to correct for observed differences between
 1538 data and simulation in the number of pile up events, as shown in Fig. 4.3.

1539 For the comparison of the measured unfolded spectrum with the theoretical predictions,
 1540 two additional MC generators are used for simulating the SM Higgs boson production in
 1541 the ggH process: HRES 2.3 [55, 56] and the second version of the POWHEG generator
 1542 (POWHEG V2) [85]. HRES is a partonic level MC generator that computes the SM Higgs
 1543 boson cross section at NNLO accuracy in pQCD and performs the NNLL resummation
 1544 of soft-gluon effects at small p_T . The central predictions of HRES are obtained including
 1545 the exact top and bottom quark mass contribution to the gluon fusion loop, fixing the
 1546 renormalization and factorization scale central values at a Higgs boson mass of 125 GeV.
 1547 The cross section normalization is scaled, to take into account electroweak corrections, by a
 1548 factor of 1.05 and the effects of threshold resummation by a factor of 1.06 [86, 87]. The
 1549 upper and lower bounds of the uncertainties are obtained by scaling up and down both the

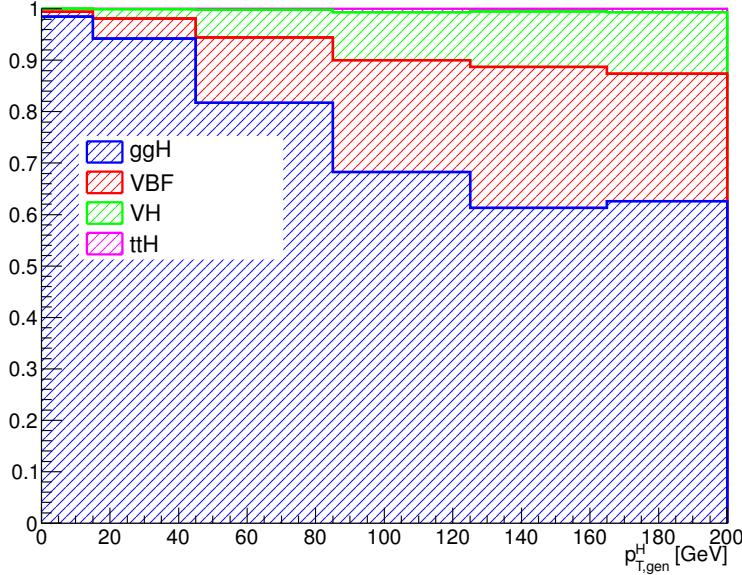


Figure 4.2.: Relative fraction of ggH, VBF, VH and ttH in each bin of the Higgs boson transverse momentum.

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 [88–90] for the decay of the Higgs boson to a W boson pair and interfaced with PYTHIA 8 [91] 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 [67], 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.

1568 controllare se $m_{\ell\ell}$ e m_T sono già definite

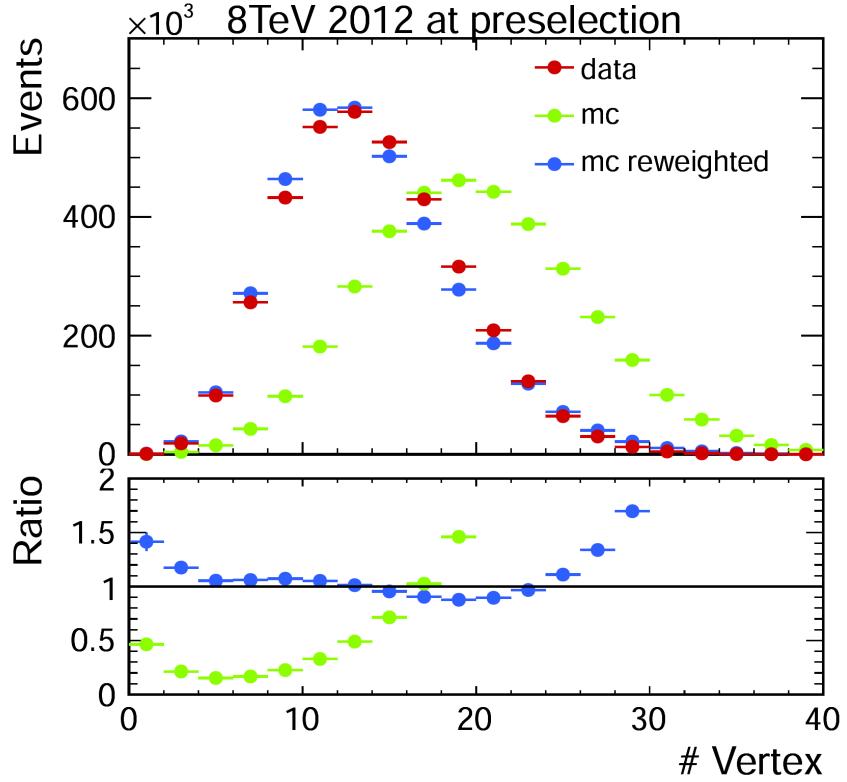


Figure 4.3.: Distribution of the number of vertices in data and in simulation, before and after applying the pile-up reweighting.

1569 4.3.1. Event reconstruction and selections

1570 Electrons and muons used in the analysis are reconstructed using the PF technique as
 1571 described in Sec. 3.2. In particular, muon candidates are required to be identified both as
 1572 Tracker Muons and Global Muons.

1573 Jets are reconstructed using the standard PF algorithm and using the anti- k_t clustering
 1574 algorithm with $R = 0.5$, as described in Sec. 3.3. If not specified otherwise, jets considered
 1575 for jet counting are the ones with $p_T > 30$ GeV.

1576 In addition to the standard CMS PF E_T^{miss} , in this analysis a *projected* E_T^{miss} variable
 1577 is also used. The *projected* E_T^{miss} is defined as the component of \vec{p}_T^{miss} transverse to the
 1578 nearest lepton if the lepton is situated within the azimuthal angular window of $\pm\pi/2$ from
 1579 the \vec{p}_T^{miss} direction, or the E_T^{miss} itself otherwise. Since the E_T^{miss} resolution is degraded
 1580 by pileup, the minimum of two projected E_T^{miss} variables is used: one constructed from
 1581 all identified particles (full projected E_T^{miss}), and another constructed from the charged
 1582 particles only (track projected E_T^{miss}).

1583 Background events from $t\bar{t}$ and tW production are rejected applying a soft-muon veto
 1584 and b tagging veto. The soft-muon algorithm is designed to identify muons from b quark

1585 decays. Events containing a muon satisfying the following requirements are rejected by the
 1586 soft-muon veto:

- 1587 • reconstructed as TrackerMuon;
- 1588 • number of hits in the Silicon Tracker greater than 10;
- 1589 • transverse impact parameter less than 0.2 cm;
- 1590 • relative isolation greater than 0.1 for muons with $p_T > 20$ GeV.

1591 The b tagging veto rejects events that contain jets identified as b-jets using two different
 1592 algorithms for high and low p_T jets (see Sec. 3.4). For jets with p_T between 10 and 30 GeV,
 1593 the TCHE algorithm is applied. Low- p_T jets passing the TCHE discriminant threshold of
 1594 2.1 are tagged as b-jets. For jets with $p_T > 30$ GeV, a better performing algorithm, JP, is
 1595 used. Jets are identified as b-jets by the JP algorithm if the discriminating variable has
 1596 a value above 1.4. In the following, a b tagged jet is defined as a jet, within $|\eta| < 2.4$
 1597 (b-tagging requires the tracker information), and with a value of the discriminating variable
 1598 above the mentioned thresholds for the two algorithms.

1599 The event selection consists of several steps. The first step is to select WW-like events
 1600 applying a selection that consists of the following set of cuts:

1601 1. **Lepton preselection:**

- 1602 • two opposite charge and different flavour ($e\mu$) isolated leptons reconstructed in
 the event;
- 1604 • $|\eta| < 2.5$ for electrons and $|\eta| < 2.4$ for muons;
- 1605 • $p_T > 20$ GeV for the leading lepton. For the trailing lepton, the transverse
 momentum is required to be larger than 10 GeV.

1607 2. **Extra lepton veto:** the event is required to have two and only two leptons with
 1608 opposite charge passing the lepton selection.

1609 3. **E_T^{miss} preselection:** particle flow E_T^{miss} is required to be greater than 20 GeV.

1610 4. **projected E_T^{miss} selection:** minimum projected E_T^{miss} required to be larger than
 1611 20 GeV.

1612 5. **Di-lepton mass cut:** $m_{\ell\ell} > 12$ GeV in order to reject low mass resonances and QCD
 1613 backgrounds.

1614 6. **Di-lepton p_T cut:** $p_T^{\ell\ell} > 30$ GeV to reduce the contribution of W+jets and DY to
 1615 $\tau\tau$ backgrounds.

1616 7. **Transverse mass:** $m_T > 60$ GeV to reject DY to $\tau\tau$ events.

1617 The requirement of different flavour leptons in the final state is important in order to
 1618 suppress the sizeable contribution of backgrounds containing a same flavour lepton pair
 1619 originating from Z boson decay.

1620 Events surviving these requirements are dominantly those where a top quark-antiquark
 1621 pair is produced and both W bosons, which are part of the top quark decay chain, decay
 1622 leptonically (dileptonic $t\bar{t}$). Two different selections are used depending on the number of

1623 jets in the event. This is done to suppress the top quark background both in the low p_T^H
 1624 region, where 0-jets events have the largest contribution, and for higher p_T^H values where
 1625 also larger jet multiplicity events are important. The selection for 0-jets events relies on the
 1626 soft-muon veto and on a soft jet (with $p_T < 30$ GeV) b tagging veto. The latter requirement
 1627 exploits the TCHE algorithm to reject soft jets that are likely to come from b quarks
 1628 hadronization.

1629 For events with a jet multiplicity greater or equal than one, a different selection is
 1630 applied. In this case we exploit the good b tagging performances of the JP tagger to reject
 1631 all the jets with $p_T > 30$ GeV that are likely to come from b quarks hadronization. The
 1632 analysis selection requires to have no events containing b-tagged jets with $p_T > 30$ GeV.

1633 A cut-flow plot is reported in Fig. 4.4, showing the effect of each selection using signal
 1634 and background simulations. In the first bin, labelled as “No cut”, no selection is applied
 1635 and the bin content corresponds to the total expected number of events with a luminosity
 1636 of 19.4 fb^{-1} . All the events in this bin have at least two leptons with a loose transverse
 1637 momentum cut of 8 GeV. In the following bin the lepton cuts are applied, including the
 1638 requirement to have two opposite sign and different flavour leptons and the extra lepton
 1639 veto. Then all the other selections are progressively reported, showing the effect of each cut
 1640 on the background and signal yields. For each selection the expected signal over background

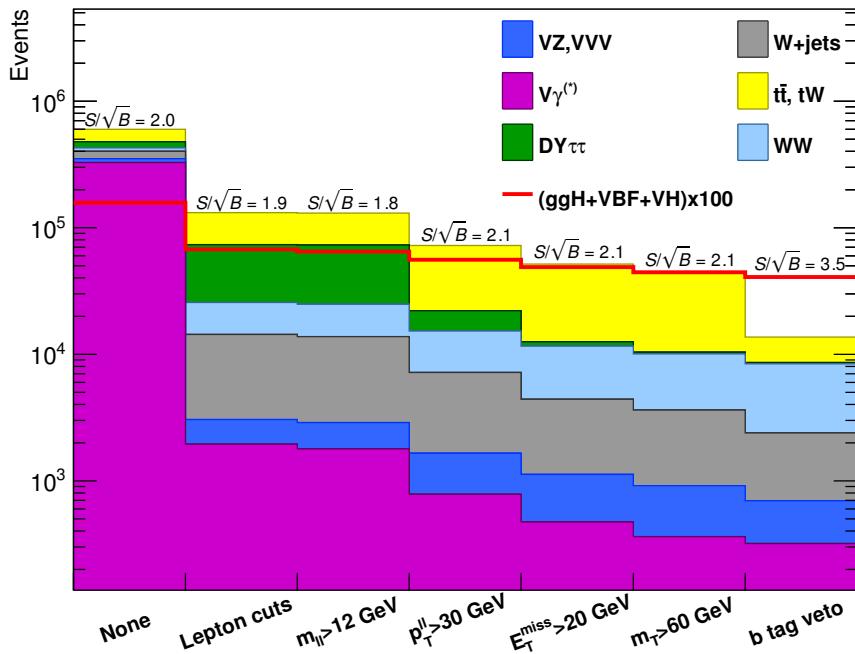


Figure 4.4.: Effect of selection cuts on simulated samples. The signal (red line) is multiplied by 100 and superimposed on stacked backgrounds. In each bin, corresponding to a different selection, is reported the expected number of events in MC at a luminosity of 19.46 fb^{-1} .

1641 ratio is also shown, which, after the full selection requirements, reaches a maximum value
1642 of about 3%.

1643 4.3.2. Simulation efficiencies and scale factors

1644 The efficiencies for the identification and isolation of the electrons and muons are measured
1645 in data and simulation selecting a pure sample of leptons coming from the $Z \rightarrow \ell\ell$ decay,
1646 and using the Tag and Probe technique described in Sec. 3.2.7. The efficiencies for data
1647 and simulation are used as scale factors to correct the simulated events to precisely model
1648 the data.

1649 The trigger efficiency is measured in data and applied to simulation as explained in
1650 Sec. 3.2.8.

1651 The efficiency of b tagging algorithms is not well simulated by MC generators and
1652 discrepancies can occur with respect to the data. For this reason is important to measure
1653 the b tagging efficiency and the misidentification probability for the given algorithms both
1654 in data and simulation, and to correct the simulated events using scale factors. This affects
1655 not only the top quark background estimation, but also the other backgrounds and the
1656 signal. As an example, if a light-parton jet in a signal event was misidentified as a b-jet,
1657 this event would be rejected by the b-jet veto.

1658 In this analysis, the b tagging efficiency and the misidentification probability are
1659 measured both in data and simulation, selecting a control sample enriched in b-jets, and
1660 using a Tag and Probe technique similar to the one described in Sec. 3.2.7. Below is
1661 described the method used to estimate the efficiency of the JP b tagging algorithm, but it
1662 is extendible to any other algorithm.

1663 The control sample is defined selecting the events that pass the selections listed in
1664 Sec. 4.3.1, and have at least two jets with p_T greater than 30 GeV. If the leading jet
1665 has a JP discriminator values above the threshold of 0.5, it is considered a *tag*, and the
1666 sub-leading jet is the *probe*. In order to avoid any bias that could arise from the probe
1667 being always the sub-leading jet, the pair is tested also in reverse order, i.e. sub-leading jet
1668 is tested against the *tag* selection, and in case it passes, then the leading jet is used as *probe*
1669 forming an independent *tag-probe* pair. If the *probe* jet has a discriminator value above the
1670 threshold used in the analysis, i.e. > 1.4 , then the *tag-probe* pair is called a *tag-pass-probe*
1671 pair. Otherwise it is identified as a *tag-fail-probe* pair.

1672 If the *tag* selection was sufficient to suppress any non top quark event, one could estimate
1673 the efficiency by dividing the number of *tag-probe* pairs in which the *probe* passes the analysis
1674 JP requirement by the total number of *tag-probe* pairs. However this is not the case, since
1675 the contamination due to other background sources is not negligible. In order to estimate
1676 the efficiency in the presence of background, a variable that discriminates between true
1677 b-jets and other jets in a $t\bar{t}$ sample is needed. This variable is the p_T of the *probe* jet. For
1678 real b-jets this variable has a peak around 60 GeV, while it has a broad distribution for
1679 other types of jets.

1680 The efficiencies are estimated performing a χ^2 simultaneous fit of the *probe* p_T spectrum
1681 in two different categories: one containing events with a *tag-pass-probe* pair and the other

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

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

where:

- N_{TPP} is the number of *tag-pass-probe* pairs;
- N_{TFP} is the number of *tag-fail-probe* pairs;
- N_s is the number of *tag-probe* pairs in which the *probe* is a b-jet;
- N_b is the number of *tag-probe* pairs in which the *probe* is not a b-jet;
- ε_s is the efficiency to identify a b-jet, i.e. the b tagging efficiency;
- ε_b is the probability to misidentify a non b-jet as a b-jet, i.e. the misidentification probability¹.

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

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

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

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

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

¹In 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

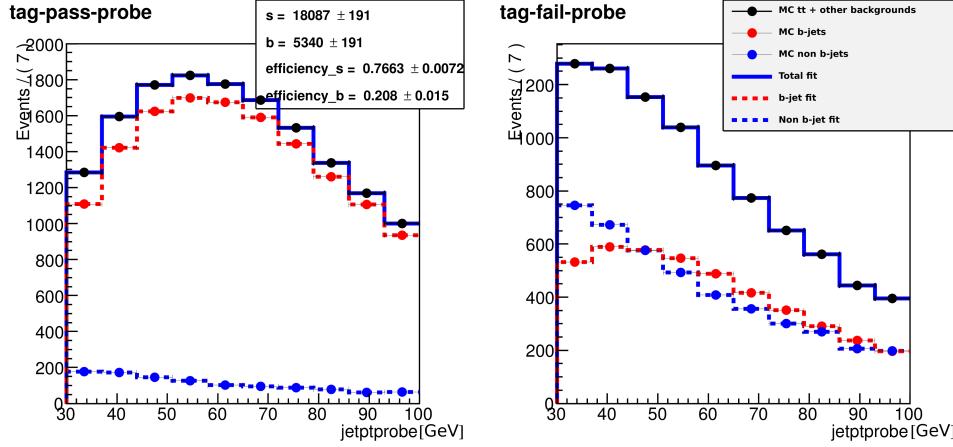


Figure 4.5.: Simultaneous fit of the *tag-pass-probe* and *tag-fail-probe* pairs in the MC.

1707 where $\sigma(\varepsilon_{s(b)}^i)$ is the uncertainty on the efficiency extracted from the fit. The pull
 1708 distributions are centred on zero and have σ close to one, as expected.

1709 Before running the fit on data, the shapes used in the fit have been validated. To do so,
 1710 a very pure phase space enriched in b jets has been defined by selecting events containing
 1711 exactly two jets with a JP discriminator greater than 1.5 and no additional b-tagged jets,
 1712 rejecting also events containing jets with p_T smaller than 30 GeV. On this very pure sample,
 1713 data have been compared against the shape used to fit the true b-jets in the *tag-pass-probe*
 1714 distribution. The result is shown in Fig. 4.6 and shows good agreement within uncertainties.

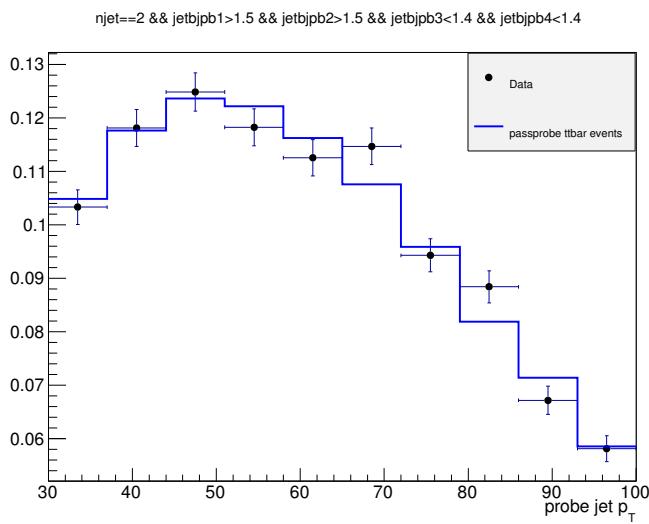


Figure 4.6.: Shape comparison for the p_T spectrum of the *probe* jet in data and simulation in a very pure phase space enriched in b-jets.

1715 Finally the fit has been performed on data, as shown in Fig. 4.7, providing the following
1716 efficiencies:

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

1717 Further studies have been performed to assess the effect of the relative uncertainty on
1718 the $t\bar{t}$ and tW event fractions. The same procedure described above has been applied to
1719 different simulation templates obtained varying the $t\bar{t}$ and tW fractions within theoretical
1720 uncertainties, and the effect on the parameters extracted with the fit procedure is found to
1721 be well below the fit uncertainties.

1722 The ratio of the efficiency measured in data and simulation represents a per-jet scale
1723 factor that can be used to reweight the simulated events. The weights to be applied
1724 event-by-event depend on the particular jet configuration in the events itself. For the signal
1725 region (SR), in which a b tagging veto is required, the event weight to be applied is given
1726 by:

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

1727 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
1728 in the simulated event, respectively. This weight is valid if the a b tagging veto is applied.
1729 If instead the b tagging veto is reverted, also the event weight has to be modified. This is
1730 done, for example, when one wants to define a $t\bar{t}$ enriched control region ($CR_{t\bar{t}}$) for the
1731 purpose of measuring the contribution of this background in a phase space orthogonal to
1732 the signal region. One simple way to define this control region is to require the leading jet

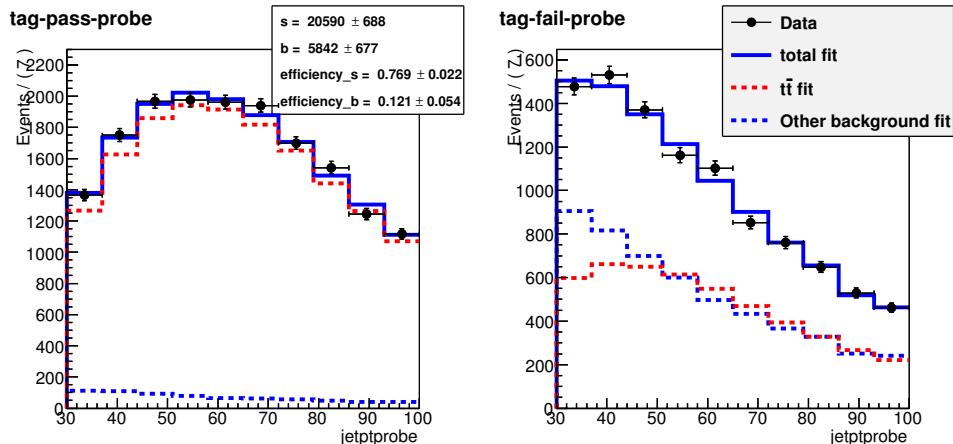


Figure 4.7.: Simultaneous fit of the *tag-pass-probe* and *tag-fail-probe* pairs in data.

1733 in the event to be b-tagged. Therefore, the simulated events falling in this category must
1734 be reweighted using the following weight:

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

1735 **4.3.3. Fiducial phase space**

1736 The Higgs boson transverse momentum is measured in a fiducial phase space, whose
1737 definition is chosen in order to minimize the dependence of the measurements on the
1738 underlying model of the Higgs boson production and decay properties.

1739 The exact requirements are determined by considering the two following correlated
1740 quantities: the reconstruction efficiency for signal events originating from within the fiducial
1741 phase space (fiducial signal efficiency ε_{fid}), and the ratio of the number of reconstructed
1742 signal events that are from outside the fiducial phase space (“out-of-fiducial” signal events)
1743 to the number from within the fiducial phase space. The requirement of having a small
1744 fraction of out-of-fiducial signal events, while at the same time preserving a high value of
1745 the fiducial signal efficiency ε_{fid} , leads to fiducial requirements at the generator level on the
1746 low-resolution variables, E_T^{miss} and m_T , that are looser with respect to those applied in the
1747 reconstructed event selection.

1748 The fiducial phase space used for the cross section measurements is defined at the
1749 particle level by the requirements given in Table 4.2. The leptons are defined as Born-level
1750 leptons, i.e. before the emission of final-state radiation (FSR), and are required not to
1751 originate from leptonic τ decays. The effect of including FSR is found to modify ε_{fid} at
1752 most of about 5%. For the VH signal process, the two leptons are required to originate
1753 from the $H \rightarrow WW \rightarrow 2\ell 2\nu$ decays in order to avoid including leptons coming from the
1754 associated W or Z boson.

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

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

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

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

Given these aspects, the criterion that 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^{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–∞] GeV.

The fiducial signal efficiency ε_{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 ε_{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 η 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 ε_{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 ε_{fid} keeping $f_{\text{out-of-fid}}$ constant. Instead, the effect of including low-resolution variables, such as m_T , is to increase both ε_{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 τ 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.

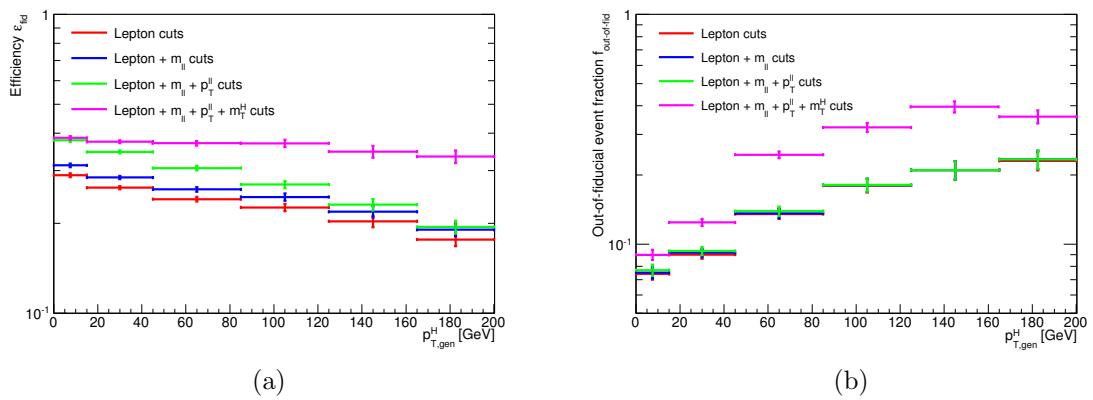


Figure 4.8.: Fiducial signal efficiency ε_{fid} and fraction of out-of-fiducial signal events $f_{\text{out-of-fid}}$ in each bin of the generator level p_T^H .

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

4.4. Background estimation

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 α 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 .

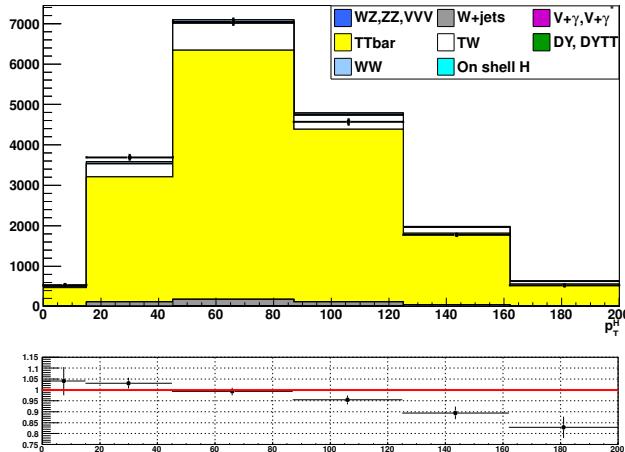


Figure 4.9.: Distribution of p_T^H in the $CR_{t\bar{t}}$ region, comparing data and simulation.

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_{MC}^{SR}	α	$\Delta\alpha$
[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– ∞]	647.5	760.4	201.7	0.27	0.15

1828 A comparison of the $m_{\ell\ell}$ distributions in the $CR_{t\bar{t}}$ region for data and simulation is
 1829 shown in Fig. 4.10, separately for each p_T^H bin and after the application of the data-driven
 1830 factors. The agreement between data and simulation is found to be satisfactory within
 1831 uncertainties.

1832 4.4.2. WW background

1833 First of all, the $m_{\ell\ell}$ and m_T shapes of the $qq \rightarrow W^+W^-$ background process have been
 1834 compared to the data in a signal free phase space, as shown in Fig. 4.11. The signal-free
 1835 region is defined requiring the preselections described in Sec. 4.3.1 with the additional cut
 1836 $m_{\ell\ell} > 70$ GeV. The comparison, which is performed inclusively in p_T^H in the 0 and 1 jet
 1837 categories, shows a good data to simulation agreement within uncertainties.

1838 In this analysis the $qq \rightarrow W^+W^-$ background normalization is left free to float in each
 1839 of the six p_T^H bins to match the data during the fitting procedure. This choice helps
 1840 mitigating the p_T^H shape difference between data and simulation. This difference is due
 1841 to missing higher order QCD corrections in the adopted simulation, obtained using the
 1842 MADGRAPH generator. In fact, as shown in Fig. 4.12, the theoretical calculations for this
 1843 process performed at NLO QCD accuracy and including soft-gluon resummation effects, i.e.
 1844 NLO+NNLL accuracy, predict a rather different p_T^H spectrum.

1845 To check the residual dependence of the $m_{\ell\ell}$ and m_T shapes on the generator used for
 1846 simulating the $qq \rightarrow W^+W^-$ process, the shapes obtained using different generators have
 1847 been compared in each p_T^H bin, as shown in Figs. 4.13 and 4.14. The usage of different
 1848 generators only mildly affect the $m_{\ell\ell}$ and m_T shapes. Nevertheless the observed differences
 1849 are taken as shape uncertainties and propagated through the fit.

1850 The gluon-induced WW process, i.e. $gg \rightarrow W^+W^-$, has a sub-dominant contribution
 1851 with respect to the quark-induced process, being the cross section ratio between the two of
 1852 about 5%. The $m_{\ell\ell}$ and m_T shapes for this background are taken from simulation while
 1853 the cross section is scaled to the approximate NLO calculation [77, 78].

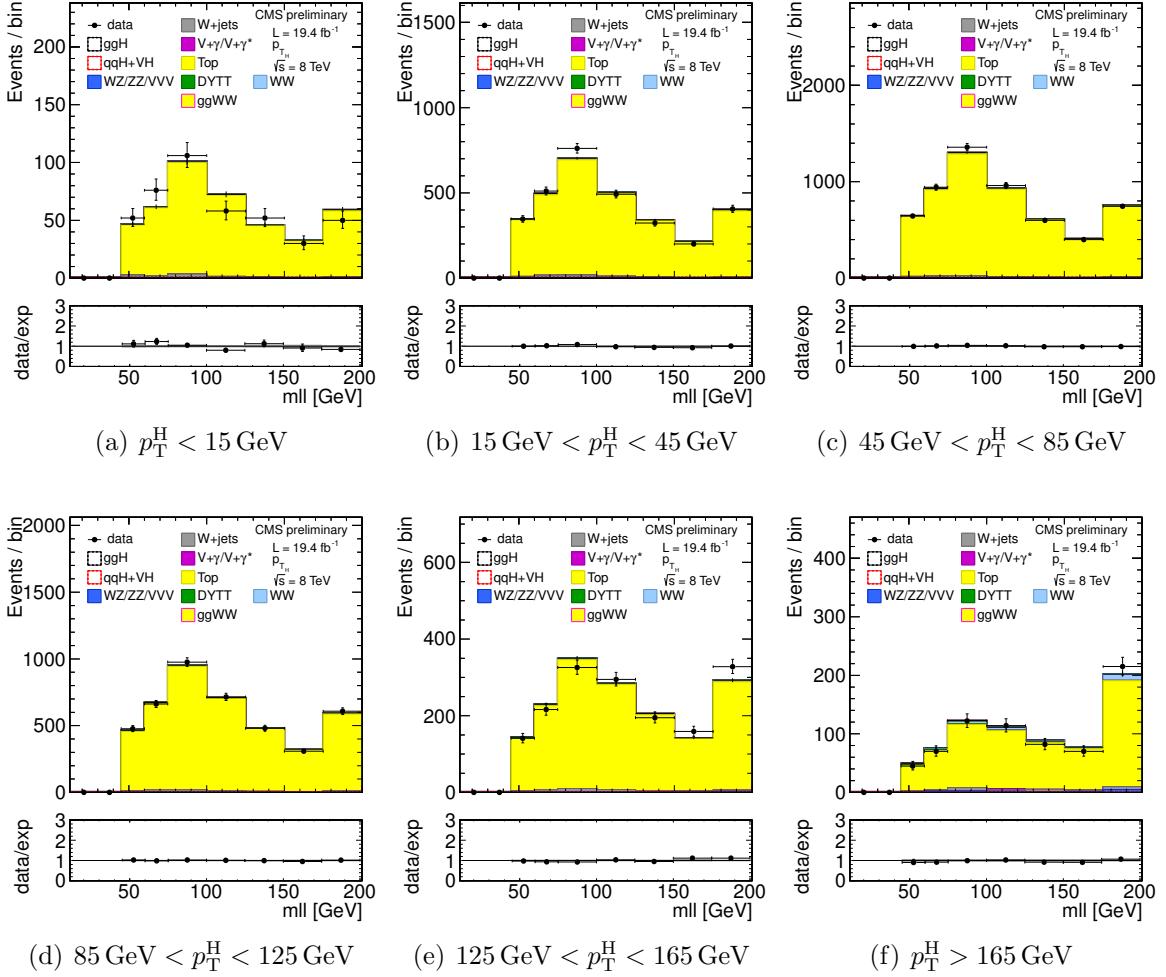


Figure 4.10.: Comparison of the $m_{\ell\ell}$ distributions in the $CR_{t\bar{t}}$ region for data and simulation in each p_T^H bin, after the application of the data-driven factors.

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

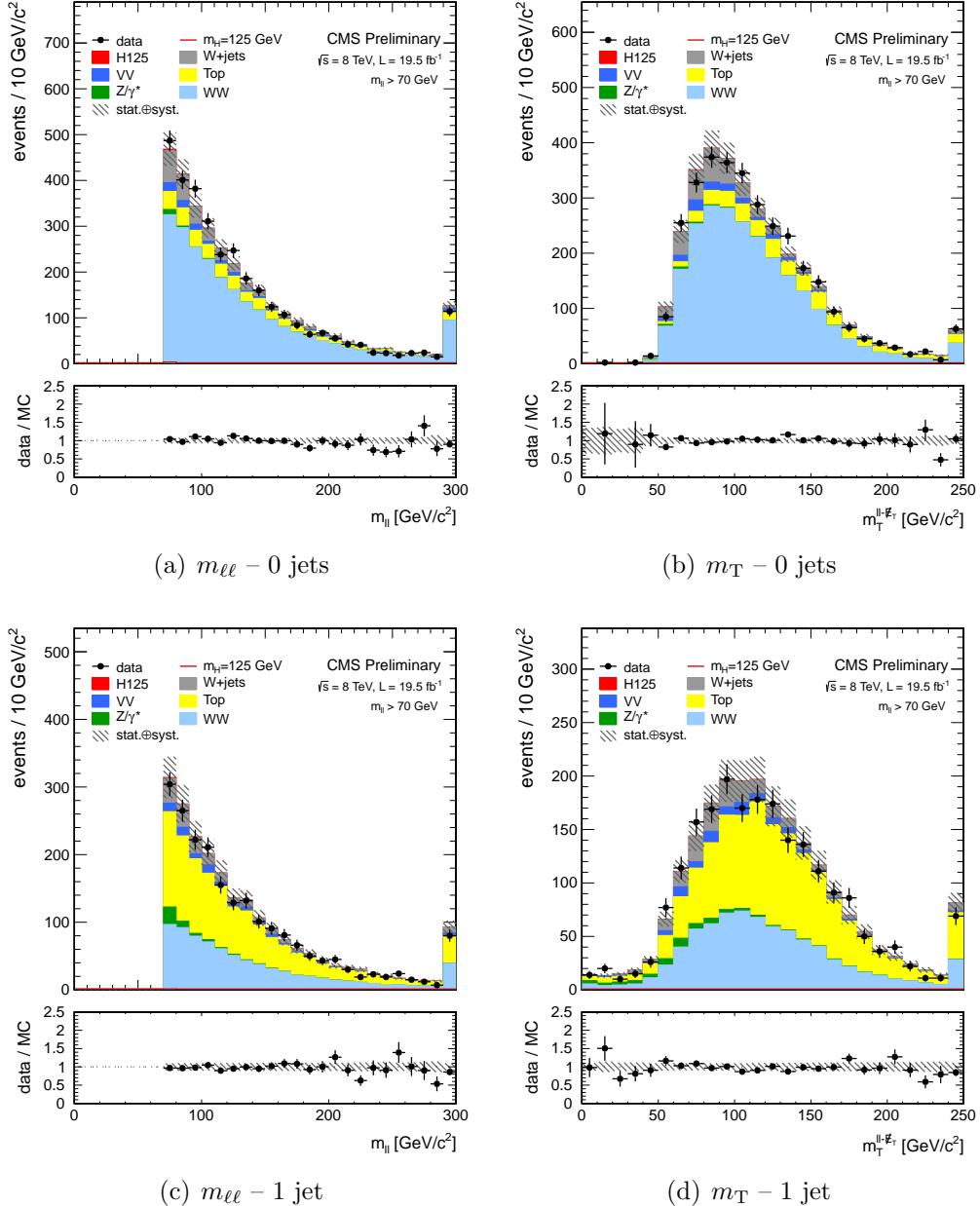


Figure 4.11.: Comparison of the $m_{\ell\ell}$ and m_T shapes in data and simulation for events with 0 and 1 jets, inclusive in p_T^H . The events are required to pass the analysis requirements and, in order to define a signal-free control region, to have $m_{\ell\ell} > 70$ GeV.

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

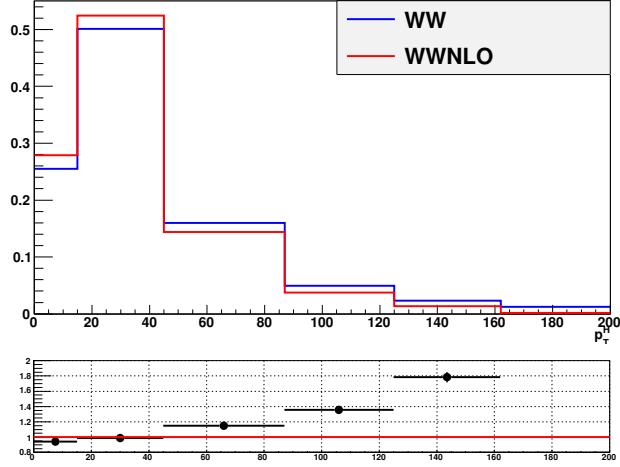


Figure 4.12.: Comparison between the p_T^{WW} distributions obtained with two different MC generators: the blue line corresponds to the MADGRAPH generator and the red line refers to the same sample in which a reweighting has been applied in order to match the theoretical prediction at NLO+NNLL precision.

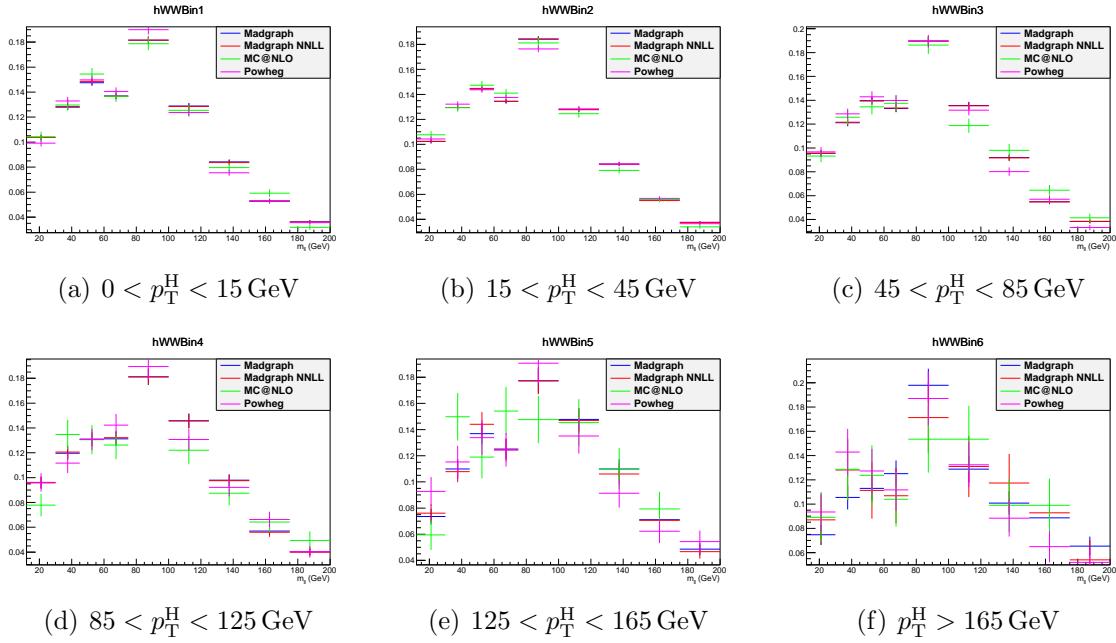


Figure 4.13.: Comparison between the $m_{\ell\ell}$ shape obtained with the default $qq \rightarrow W^+W^-$ background simulation (MADGRAPH) and other theoretical models in every p_T^H bin.

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

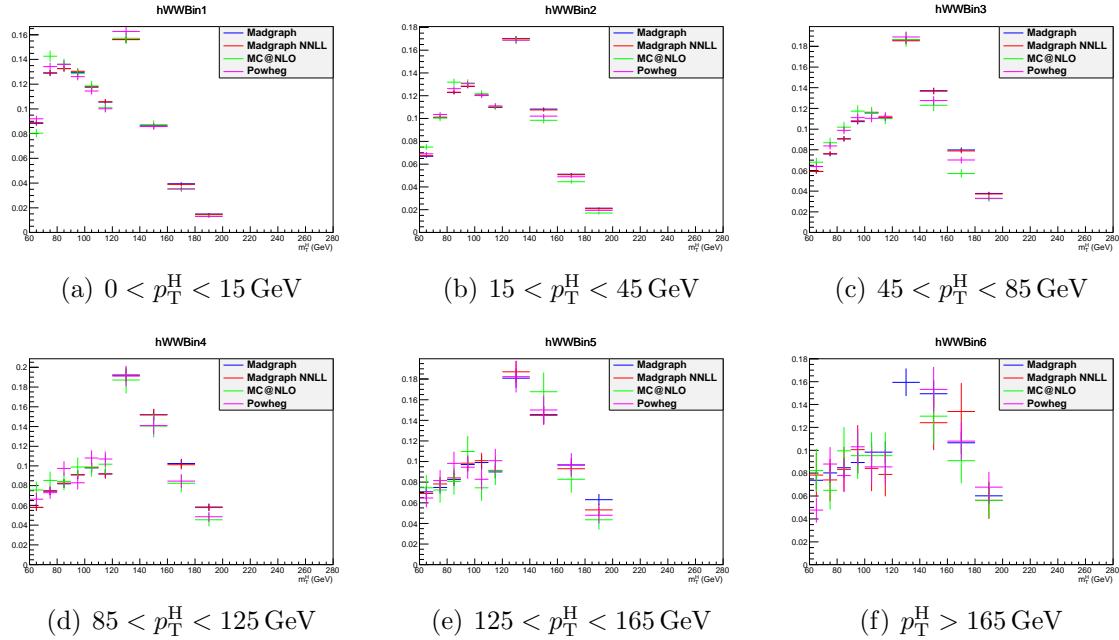


Figure 4.14.: Comparison between the m_T shape obtained with the default $qq \rightarrow W^+W^-$ background simulation (MADGRAPH) and other theoretical models in every p_T^H bin.

lepton contamination from EW processes such as $W/Z+jets$ 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, parametrized as a function of p_T and η 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.

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 $W+jets$ events. The results of this closure test show good agreement between data and the estimated background.

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

The low E_T^{miss} threshold in the $e\mu$ final state requires the consideration of the contribution from $Z/\gamma^* \rightarrow \tau^+\tau^-$, that estimated from data. This is accomplished by selecting $Z/\gamma^* \rightarrow \mu^+\mu^-$ events in data and replacing both muons with a simulated $\tau \rightarrow \ell\nu_\tau\bar{\nu}_\ell$ decay [67], thus obtaining a “hybrid” event. The Z boson four-momentum is reconstructed in data from the four-momenta of the daughter muons. Then a simulation step allows the replacement of the muon objects with τ leptons, in such a way to preserve the Z boson momentum direction

in its rest frame. The $Z/\gamma^* \rightarrow \tau^+\tau^-$ decay is simulated with the TAUOLA package [92] to correctly describe the τ -polarization effects.

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

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

1900 Diboson backgrounds

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

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

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

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

1924 4.5. Systematic uncertainties

Systematic uncertainties play an important role in this analysis where no strong mass peak is expected due to the presence of undetected neutrinos in the final state. They arise from three sources: background predictions, experimental measurements, and theoretical

1928 uncertainties. One of the most important sources is the normalization of the backgrounds
1929 that are estimated on data control samples whenever is possible.

1930 The systematic uncertainties can affect the measured signal strengths in different
1931 ways. The uncertainties on the background predictions can be divided in those affecting
1932 the background cross section, the ($m_{\ell\ell}$, m_T) shape or both. As an example, systematic
1933 uncertainties changing the background cross section are the ones related to the background
1934 data-driven estimation, while the b tagging uncertainties only have an effect on the ($m_{\ell\ell}$,
1935 m_T) shape. Uncertainties such as lepton energy scale can instead affect both normalization
1936 and shape. Also, uncertainties affecting the signal ($m_{\ell\ell}$, m_T) shape reflect on an uncertainty
1937 on the measured signal strength.

1938 A summary of the main sources of systematic uncertainty and the corresponding estimate
1939 is reported in Table 4.4. A brief description of each source of systematic uncertainty is
1940 discussed in the following sections.

1941 The uncertainties related to the unfolding procedure are treated separately and are
1942 discussed in Sec. ??.

1943 4.5.1. Background predictions uncertainties

1944 The signal extraction is performed fitting the estimated background contributions and
1945 subtracting them to the event counts in data. Therefore, the uncertainties on the background
1946 predictions indirectly reflect as uncertainties on the signal measurements. A list of the most
1947 important background uncertainties is given below.

- 1948 • **tt> and tW backgrounds:** the shapes of these backgrounds are corrected for different
1949 b tagging efficiency in data and simulation, and the normalization is taken from data
1950 in a top quark enriched control region independently in each p_T^H bin, as explained in
1951 Sec. 4.4.1. The uncertainties related to this procedure arise from the sample size in the
1952 control regions for each p_T^H bin, and are embedded in the α factors used to extrapolate
1953 the top quark background normalization from the control region to the signal region.
1954 They vary from 20% to 50% depending on the p_T^H bin.

1955 The simulated samples include both tt> and tW processes, a systematic uncertainty
1956 related to the tW/tt> fraction has been included. In fact, a relative variation of the
1957 contribution of these two processes could modify the shape of the simulated sample,
1958 and is thus included as a shape uncertainty affecting the ($m_{\ell\ell}$, m_T) shape in each p_T^H
1959 bin in a correlated way.

- 1960 • **WW background:** due to the fact that the WW background ($m_{\ell\ell}$, m_T) shape is
1961 entirely taken from simulation, the analysis is relying on theoretical models and can
1962 thus be affected by their uncertainties. Especially higher order QCD radiative effects
1963 have an influence on the generated WW shape. To study this impact, the shapes of
1964 the distributions produced with the MADGRAPH generator are compared to the ones
1965 produced with MC@NLO and other generators (see Sec. 4.4.2). The comparison is
1966 performed separately in each bin of p_T^H and the uncertainty includes shape differences

Table 4.4.: Main sources of systematic uncertainties and their estimate. The first category reports the uncertainties on the normalization of background contributions. The experimental and theoretical uncertainties refer to the effect on signal yields. A range is specified if the uncertainty varies across the p_T^H bins.

Uncertainties in backgrounds contributions	
Source	Uncertainty
t <bar>t>, tW</bar>	20–50%
W+jets	40%
WZ, ZZ	4%
W γ^* , W γ	30%

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

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

originating from the renormalization and factorization scale choice. The effect on the signal strengths is found to be of the order of 1%.

- **W+jets background:** the systematic uncertainties on the W+jets background arise from the estimation method explained in Sec. 4.4.3. This 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%.

- **Diboson backgrounds:** these backgrounds are expected to give a small contribution in the signal phase space. Uncertainties on the cross sections reported in [93, 94] are 4% for WZ and 2.5% for ZZ. A 30% uncertainty is assigned to the $W\gamma$ [95] yield and another 30% on $W\gamma^*$ contribution according to the uncertainty on the normalization study (see Sec. 4.4.4).

4.5.2. Experimental uncertainties

The following experimental systematic sources have been taken into account:

- **Luminosity:** using the CMS online luminosity monitoring system, the uncertainty on the integrated luminosity (19.4 fb^{-1}) collected during the 2012 data taking period is found to be of 2.6%.
- **Trigger efficiency:** the uncertainties for both electrons and muons, estimated as described in Sec. 3.2.8, are at 1-2% level.
- **Lepton reconstruction and identification efficiency:** this uncertainty is estimated with the Tag and Probe technique described in Sec. 3.2, resulting in a 4% uncertainty for electrons and 3% for muons.
- **Muon momentum and electron energy scale:** the momentum scale of leptons has relatively large uncertainties due to different detector effects, as explained in Sec. 3.2. 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 on the muon pseudorapidity, is assigned.
- **E_T^{miss} modelling:** the E_T^{miss} measurement is affected by the possible mis-measurement of individual particles addressed above, as well as the additional contributions from the pile-up interactions, as described in Sec. 3.5. 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. The effect is found to be around 2%.
- **Jet energy scale (JES) uncertainties:** JES uncertainties affect both the jet multiplicity and the jet kinematic variables, reflecting also on the $(m_{\ell\ell}, m_T)$ shape. This uncertainty is estimated varying the kinematics of the reconstructed jets within the uncertainties on the JES (which depend on η and p_T of the jet) described in Sec. 3.3, and recomputing all the correlated variables, like $m_{\ell\ell}$ and m_T .
- **b-jets misidentification modelling:** a fraction of signal events is rejected because erroneously identified as b-jets by the b tagging algorithms. The misidentification probability, as measured with the Tag and Probe technique described in Sec. 4.3.2, has an uncertainty related to different b tagging performance in data and simulation. This affects also non-top quark backgrounds.

- **Pileup multiplicity:** some of the variables used in the analysis are affected by the average number of pile up interactions. The simulated events have been reweighted according to the instantaneous luminosity measured on data. The error on the average number of pile up interactions measured in data and the simulation of the modelling and physics aspects of the pile up simulation provide an uncertainty of at most 5% on the distribution used in the reweighting procedure. This uncertainty is propagated through all the analysis, and the estimated uncertainty on the signal strengths do not exceed 2%.

4.5.3. Theoretical uncertainties

Theoretical uncertainties generally arise from missing higher-order corrections in QCD and PDF uncertainties. These uncertainties can affect both the cross section and the ($m_{\ell\ell}$, m_T) shape of the background predictions, as well as the shape of the signal model.

- **QCD scale uncertainties:** the uncertainties on the total cross sections due to the choice of the renormalization and factorization scale are assigned to simulation-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 scales variation on the jet bin migrations has to be taken into account because of the b tagging veto efficiency. Indeed, the b tagging veto efficiency is not flat as a function of jet multiplicity not p_T^H , as shown in Fig. ??, therefore introducing a dependence of the selection efficiency on the number of jets on the event. In order to take into account this effect, an uncertainty on the ggH production mode has been included according to the Stewart-Tackman method, following the recipe proposed in Refs. [66, 96]. The effect on the signal strengths is found to be of the order of 1–2%.
- **PDFs uncertainties:** the utilization of different PDF sets can affect the ($m_{\ell\ell}$, m_T) shapes of the signal contributions, as well as the normalization and shape of the background predictions. The uncertainty related due to the variations in the choice of PDFs is considered following the PDF4LHC [97, 98] prescription, using CT10, NNPDF2.1 [99] and MSTW2008 [100] PDF sets. The effect on the signal strengths is found to be of at most 1%.

4.5.4. Statistics uncertainty of the simulated samples

Due to the large range of weights used to correct the simulated distributions in order to match those in data, the effective size of the simulated samples are sometimes smaller than the actual number of events in the sample. The uncertainties due to the finite statistics of the simulated samples are taken into account and propagated through the final result. Their effect on the signal strengths is found to be negligible.

2052 4.5.5. Treatment of systematic uncertainties in the analysis

2053 As explained before, one can distinguish between normalization uncertainties, where a
 2054 systematic effect is changing the normalization of a given process assuming the ($m_{\ell\ell}$, m_T)
 2055 shape is not affected, and shape uncertainties where the actual change in the ($m_{\ell\ell}$, m_T)
 2056 shape of the process is taken into account. The normalization uncertainties enter the
 2057 analysis as constant normalization factors, whereas for shape uncertainties the nominal
 2058 and the $+1\sigma$ and -1σ shapes enter the analysis in form of three histograms with the same
 2059 normalization.

2060 Effects from experimental uncertainties are studied by applying a scaling and smearing
 2061 of certain variables of the physics objects, followed by a subsequent recalculation of all the
 2062 correlated variables. This is done for simulation, to account for possible systematic mis-
 2063 measurements of the data. All experimental sources from Section 4.5.2 but luminosity are
 2064 treated both as normalization and shape uncertainties. For background with a data-driven
 2065 normalization estimation, only the shape uncertainty is considered.

2066 4.6. Signal extraction

2067 According to the “blinding” policy of the CMS Collaboration, the strategy of the analysis
 2068 has been scrutinized and approved by a selected committee of internal reviewers before
 2069 looking at the data in the signal region. This approach prevents the analysts from being
 2070 biased by the data in the developing phase of the analysis. Below are shown the results
 2071 after having looked at the data.

2072 4.6.1. Fitting procedure

2073 The signal, including ggH, VBF, and VH production mechanisms, is extracted in each bin
 2074 of p_T^H by performing a binned maximum likelihood fit simultaneously in all p_T^H bins to a
 2075 two-dimensional template for signals and backgrounds in the $m_{\ell\ell}$ – m_T plane. The variables
 2076 used for the two-dimensional template are chosen for their power to discriminate signal and
 2077 background contributions. This is shown in Fig. 4.15, where the two-dimensional simulated
 2078 distributions are shown for the signal and background processes in the 0-jets category. As
 2079 can be observed, the signal contribution in the 0-jets category is mostly distributed in the
 2080 low- $m_{\ell\ell}$ region and for m_T values around 90–100 GeV. The background contribution, which
 2081 is mainly owed to the WW, W+jets and $Z/\gamma^* \rightarrow \tau^+\tau^-$ production, is instead distributed in
 2082 the high- $m_{\ell\ell}$ region and for intermediate values of m_T .

2083 Six different signal strength parameters are extracted from the fit, one for each p_T^H bin.
 2084 The relative contributions of the different Higgs production mechanisms in the signal
 2085 template are taken to be the same as in the SM. The systematic uncertainty sources are
 2086 considered as nuisance parameters in the fit.

2087 The binning of the $m_{\ell\ell}$ and m_T templates is chosen to be:

- 2088 • $m_{\ell\ell}$: [12, 30, 45, 60, 75, 100, 125, 150, 175, 200]

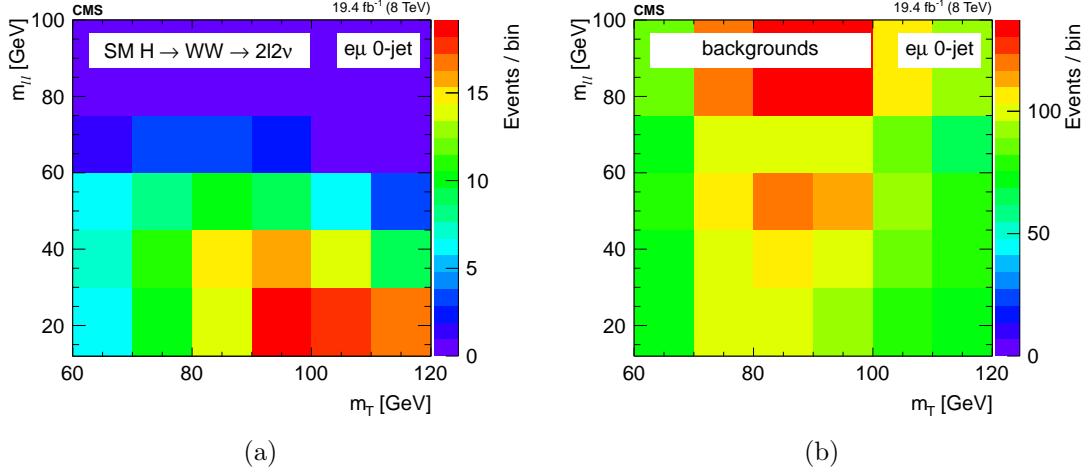


Figure 4.15.: Two-dimensional $m_{\ell\ell}$ – m_T distribution for signal (a) and background (b) processes in the 0-jets category.

- m_T : [60, 70, 80, 90, 100, 110, 120, 140, 160, 180, 200, 220, 240, 280]

To avoid a dependence of the results on the variables used for the template fit, $m_{\ell\ell}$ and m_T need to be uncorrelated with respect to p_T^H . This has been verified and the correlation between the discriminating variables and p_T^H is shown in Fig. 4.16 and Fig. 4.17 for ggH and VBF production modes, respectively.

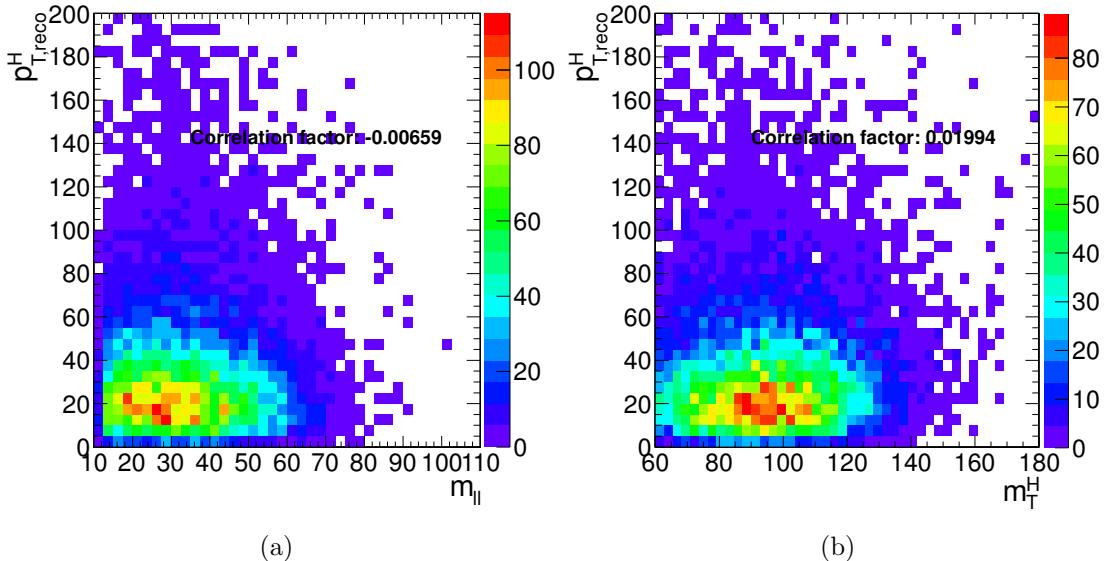


Figure 4.16.: Correlation between p_T^H and $m_{\ell\ell}$ (a) and between p_T^H and m_T (b) after the full selection for the ggH production mode.

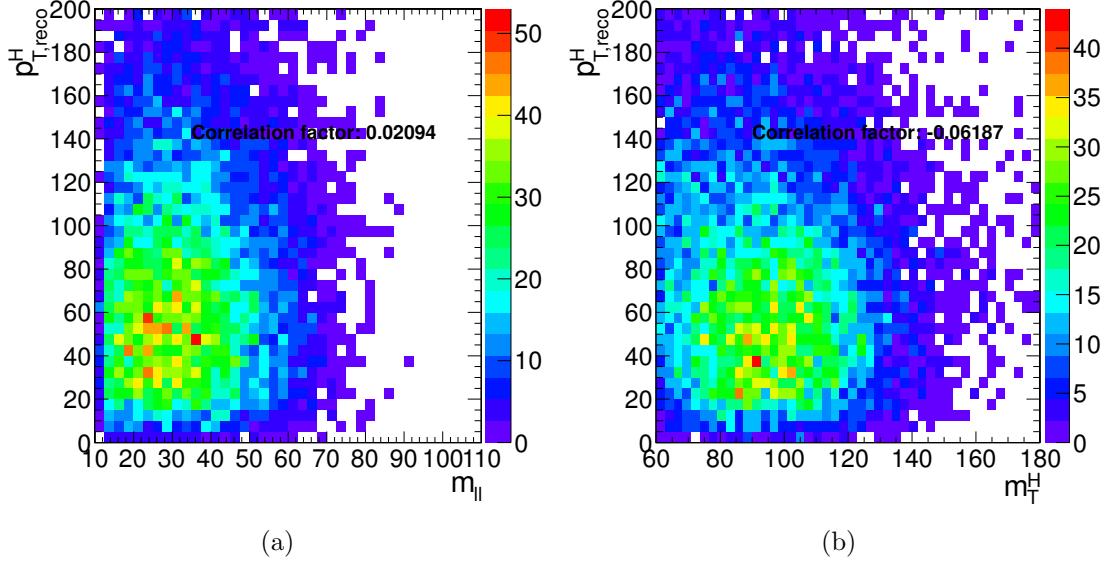


Figure 4.17.: Correlation between p_T^H and $m_{\ell\ell}$ (a) and between p_T^H and m_T (b) after the full selection for the VBF production mode.

2094 The signal strength μ in each bin, defined as the ratio between the measured cross section
 2095 and the SM one, $\mu = \sigma/\sigma_{\text{SM}}$, is allowed to float between -10 and +10, thus allowing negative
 2096 values. This is mainly intended to allow future combinations with similar measurements
 2097 without introducing any bias.

2098 Because of detector resolution effects, some of the reconstructed $H \rightarrow WW$ signal events
 2099 might originate from outside the fiducial phase space. These out-of-fiducial signal events
 2100 cannot be precisely handled by the unfolding procedure and must be subtracted from the
 2101 reconstructed spectrum. The p_T^H distribution of the out-of-fiducial signal events is taken
 2102 from simulation, and each bin is multiplied by the corresponding measured signal strength
 2103 before performing the subtraction.

2104 At the end, the number of events in each bin i of the measured spectrum is:

$$N_i = \mu_i(s_i - f_i) \quad , \quad (4.11)$$

2105 where μ_i is the measured signal strength, s_i and f_i are respectively the total number of
 2106 reconstructed signal events and the number of reconstructed out-of-fiducial signal events
 2107 expected from simulation.

2108 The fit makes use of the binned maximum likelihood approach. The likelihood function,
 2109 \mathcal{L} , restricted to the p_T^H bin j , can be written as:

$$\mathcal{L}(\mu_j, \theta) = \prod_{i=0}^{N_{\text{bins}}} \frac{(\mu_j s_i(\theta) + b_i(\theta))^{n_i}}{n_i!} e^{-\mu_j s_i(\theta) - b_i(\theta)} \cdot p(\theta|\tilde{\theta}) \quad , \quad (4.12)$$

where μ_j is the signal strength in the bin j , i.e. the parameter of interest of the fit, which multiplies the signal yield. The index i runs over the bins of the $m_{\ell\ell}$ - m_T two-dimensional histogram corresponding the p_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 θ represents the full suite of nuisance parameters used to incorporate the systematic uncertainties. Each nuisance parameter is constrained in the fit including the prior distributions functions $p(\theta|\tilde{\theta})$ in the likelihood, where $\tilde{\theta}$ is the set of default values for the θ parameters [101]. 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. This is the optimal choice to describe uncertainties on definite positive observables, like cross sections, efficiencies, luminosity, etc. The usage of a gaussian distribution, under certain circumstances, would indeed allow the observable value to fluctuate below zero. 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 uncertainties 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*², which provides a simple method to estimate the signal sensitivity before looking at the data [102].

4.6.2. Signal and background yields

A comparison of data and background prediction is shown in Fig. 4.18, 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 [67]. The m_T distributions are shown in Fig. 4.19 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 4.5.

²In 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.

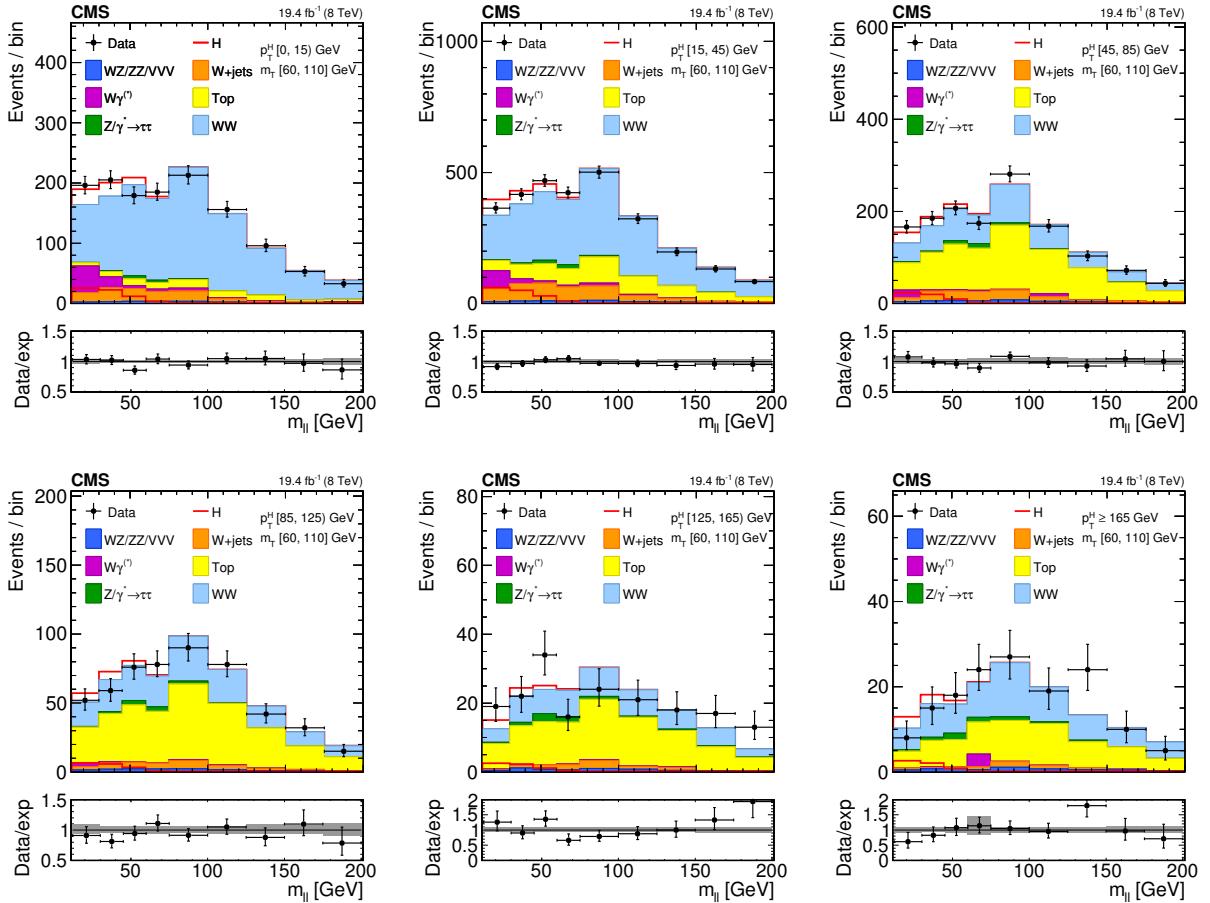


Figure 4.18.: Distributions of the $m_{\ell\ell}$ variable in each of the six p_T^H bins. Background normalizations correspond to the values obtained from the fit. Signal normalization is fixed to the SM expectation. The distributions are shown in an m_T window of $[60, 110]$ GeV in order to emphasize 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.

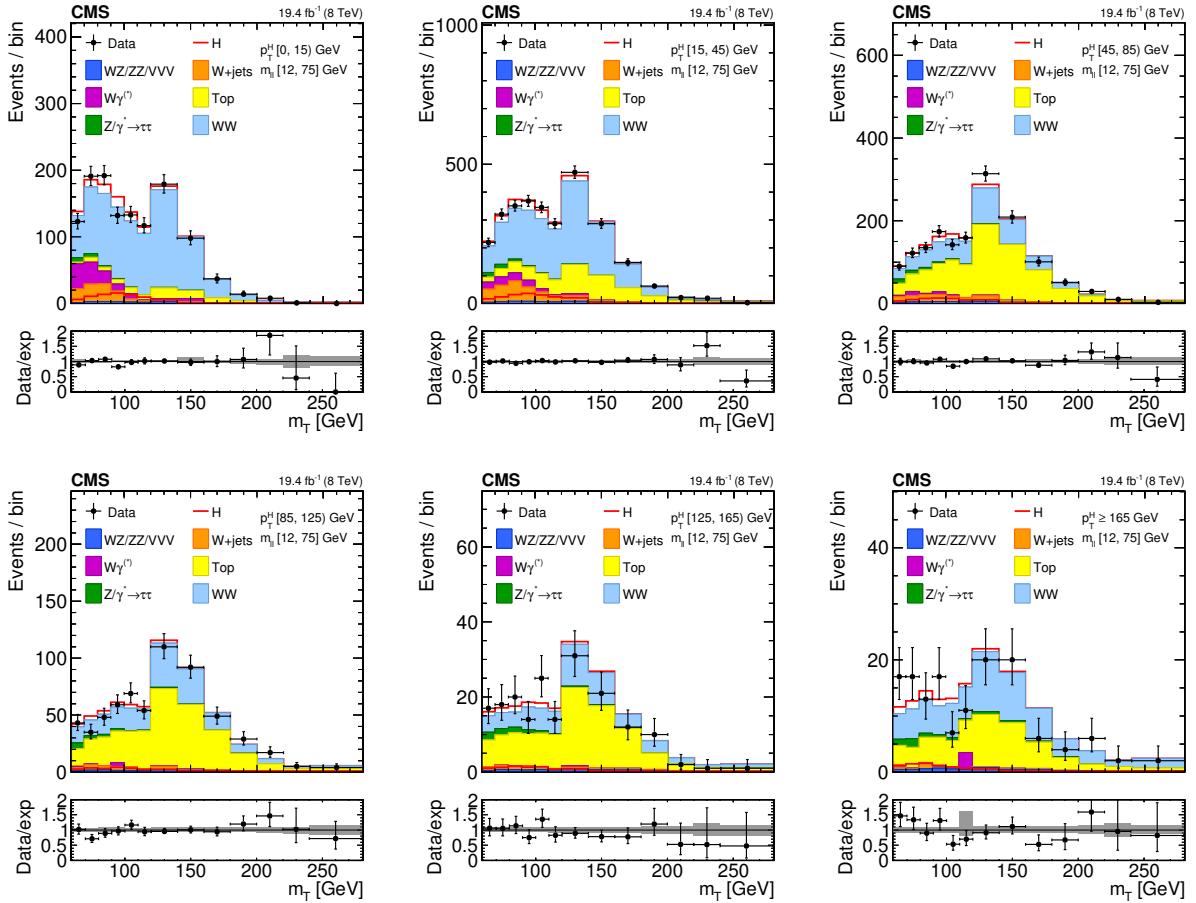


Figure 4.19.: Distributions of the m_T variable in each of the six p_T^H bins. Background normalizations correspond to the values obtained from the fit. Signal normalization is fixed to the SM expectation. The distributions are shown in an $m_{\ell\ell}$ window of $[12, 75]$ GeV in order to emphasize 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 4.5.: 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 .

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

²¹⁴² The signal strengths obtained performing the fit are shown in Table ???. In order to assess
²¹⁴³ the robustness of the fit, several toy MC samples have been generated, with a statistical
²¹⁴⁴ accuracy corresponding to the one expected in data, and fitted with the same procedure
²¹⁴⁵ described before. The distribution of the signal strengths extracted in each bin using the
²¹⁴⁶ toy MC samples and their pull distributions are shown in Fig. 4.20.

Table 4.6.: Signal strengths measured in data for each p_T^H bin with 68% CL uncertainties.

p_T^H [GeV]	μ	Uncertainty (68% CL)
0–15	+0.753	-0.424/+0.437
15–45	+0.716	-0.300/+0.308
45–85	+1.309	-0.445/+0.465
85–125	+0.165	-0.890/+0.898
125–165	+1.715	-1.103/+1.217
165–∞	+0.796	-0.913/+1.059

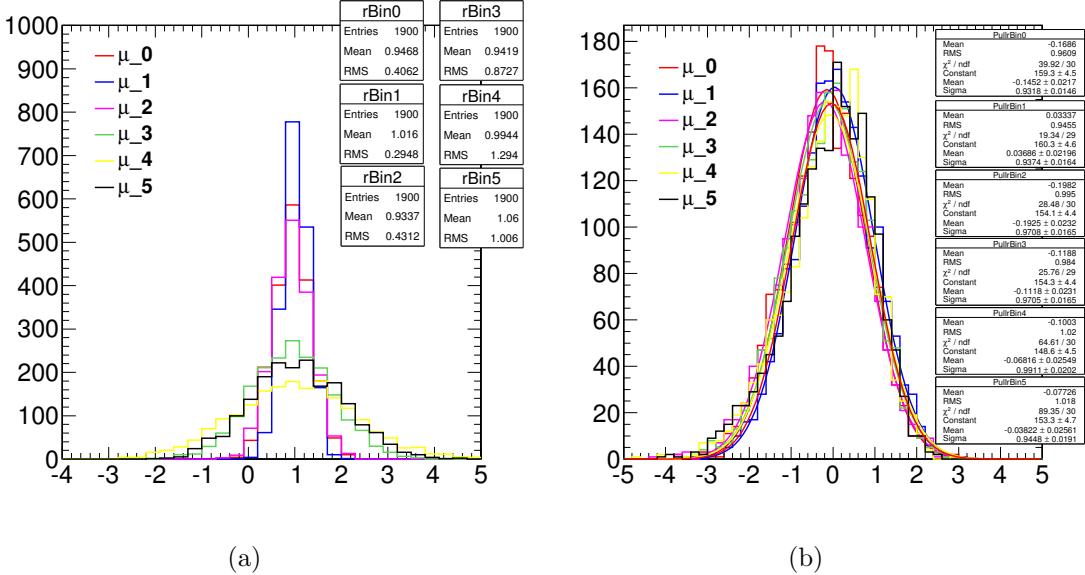


Figure 4.20.: Signal strength distribution as extracted from the fit of toy MC samples (a). Distribution of the pull of the signal strength parameters (b).

2147 The reconstructed spectrum is obtained starting from the signal yield N_i in each p_T^H bin
2148 i , obtained subtracting the out-of-fiducial events as shown in Eq. (??), and dividing it by
2149 the bin width w_i and integrated luminosity \mathcal{L} , i.e.:

$$\frac{d\sigma_i}{dp_{T,\text{reco}}^H} = \frac{N_i}{w_i \mathcal{L}} \quad . \quad (4.13)$$

2150 The spectrum shown in Fig. 4.21 is obtained after having performed the fit and after
2151 the subtraction of the out-of-fiducial signal events, but before undergoing the unfolding pro-
2152 cedure. The theoretical distribution after the detector simulation and event reconstruc-
2153 tion is also shown for comparison. Also, the expected distribution of the sub-dominant VBF
2154 and VH production mechanisms is shown.

2155 In order to measure the inclusive cross section times branching fraction in the fiducial
2156 phase space, the reconstructed differential spectrum of Fig. 4.21 is integrated over p_T^H . The
2157 contribution of the uncertainty in each bin is propagated to the inclusive measurement
2158 taking into account the correlation, i.e. using the covariance matrix, of the signal strengths.
2159 For the extrapolation of this result to the fiducial phase space, the unfolding procedure is not
2160 needed, and the inclusive measurement has only to be corrected for the fiducial phase space
2161 selection efficiency ϵ_{fid} . Dividing the measured number of events by the integrated luminosity
2162 and correcting for the overall selection efficiency, which is estimated in simulation to be
2163 $\epsilon_{\text{fid}} = 36.2\%$, the inclusive fiducial cross section times branching fraction σ_{fid} , is computed
2164 to be:

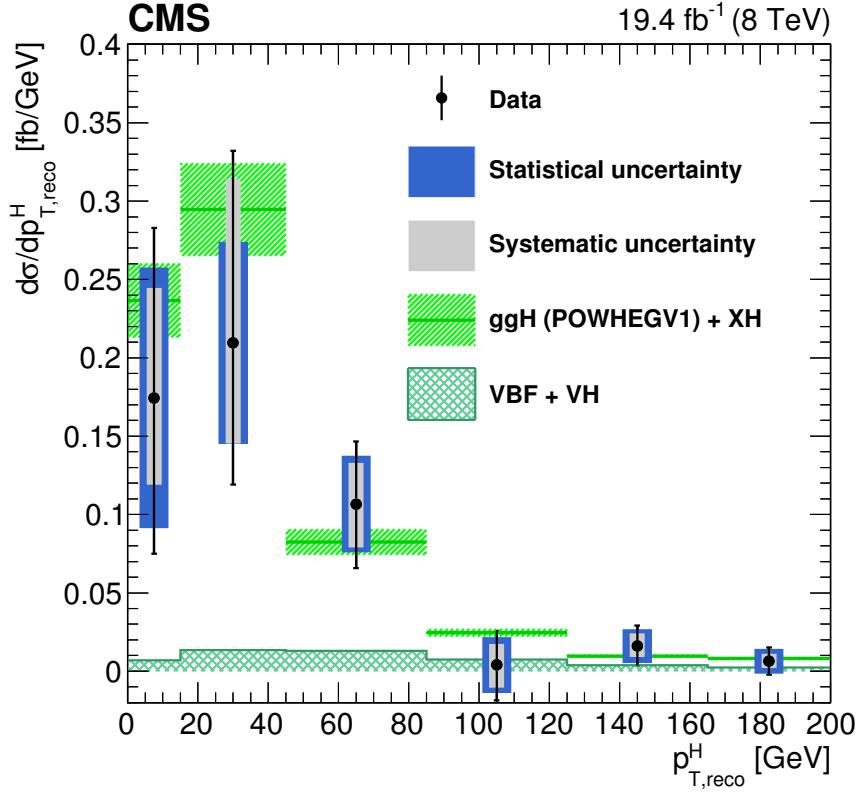


Figure 4.21.: Differential Higgs boson production cross section as a function of the reconstructed p_T^H , before applying the unfolding procedure. Data values after the background subtraction are shown together with the statistical and the systematic uncertainties, determined propagating the sources 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.

$$\sigma_{\text{fid}} = 39 \pm 8 \text{ (stat)} \pm 9 \text{ (syst)} \text{ fb} \quad , \quad (4.14)$$

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

4.7. Unfolding

To facilitate comparisons with theoretical predictions or other experimental results, the signal extracted performing the fit has to be corrected for detector resolution and efficiency

2171 effects and for the efficiency of the selection defined in the analysis. An unfolding procedure
 2172 is used relying on the RooUnfold package [103], which provides the tools to run various
 2173 unfolding algorithms.

2174 The basic principle behind the unfolding procedure in this analysis is to use MC signal
 2175 samples to make the “true” distribution of the variable of interest, which is obtained using
 2176 simulated events before particle interaction with the detector, and the same distribution
 2177 obtained using events reconstructed after the full GEANT4 simulation of the CMS detector
 2178 and event reconstruction.

2179 These two distributions are used to calculate the detector response matrix M :

$$R_i^{\text{MC}} = \sum_{j=1}^n M_{ij} T_j^{\text{MC}} , \quad (4.15)$$

2180 where T^{MC} and R^{MC} are two n -dimensional vectors representing the distribution before
 2181 and after event processing through CMS simulation and reconstruction. The dimension n
 2182 of the two vectors corresponds to the number of bins in the distributions, equal to six in
 2183 this analysis. The response matrix M includes all the effects related to the detector and
 2184 analysis selection that affect the R^{MC} distribution. The goal of the unfolding procedure
 2185 is to obtain the T^{truth} distribution starting from the measured R^{observed} distribution by
 2186 inverting the matrix M .

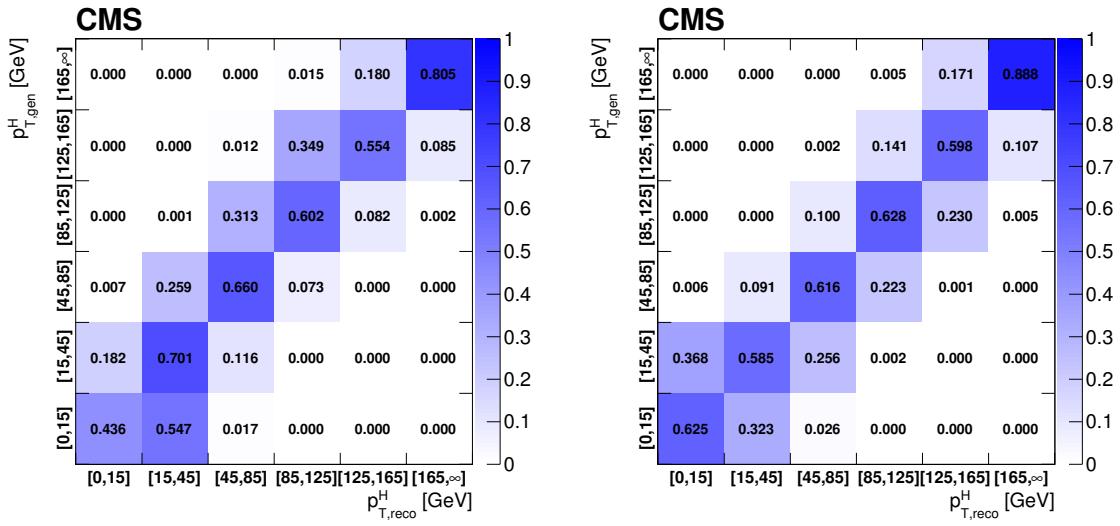
2187 Given the finite data statistical accuracy, a simple inversion could lead to large fluctua-
 2188 tions between bins in the unfolded result. In particular, if the off-diagonal elements of
 2189 the response matrix are sizeable, the unfolded distribution has large variance and strong
 2190 negative correlations between the neighbouring bins [68]. Several unfolding methods with
 2191 regularization are available in literature, such as a method based on the Bayes’ theorem,
 2192 which overcome the unfolding instability using an iterative procedure [104].

2193 The unfolding procedure in this analysis relies on the Singular Value Decomposition
 2194 (SVD) [105] method based on the Tikhonov regularization function. Such method introduces
 2195 a regularization function that controls the smoothness of the distribution and depends
 2196 generally on one regularization parameter, which can be controlled to achieve the desired
 2197 degree of smoothness. The choice of the regularization parameter is particularly critical,
 2198 and it should represent an optimal trade-off between taming the fluctuations in the unfolded
 2199 result, and biasing the unfolded distribution. The main feature of this method is the use
 2200 of the singular value decomposition of the response matrix, including an additional term
 2201 to suppress the oscillatory component of the solution, i.e. the regularization term, which
 2202 represents some *a priori* knowledge of the final solution. The regularization parameter k_{reg}
 2203 is chosen to obtain results that are robust against numerical instabilities and statistical
 2204 fluctuations, following the prescription described in Ref. [105]. This prescription consists in
 2205 the diagonalization of the response matrix using the SVD approach and in the subsequent
 2206 calculation of the vector \vec{d} , whose values d_i represent the measured distribution expressed
 2207 in a specific base defined by the SVD decomposition. Plotting $\log|d_i|$ as a function of i ,
 2208 where i is related to the amount of regularization, one should obtain a curve that flattens

out at some value of i . The regularization parameter corresponding to that value represents the optimal k_{reg} choice. The parameter obtained using this prescription is $k_{\text{reg}} = 3$.

The detector 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 matrix, including all signal sources and normalized by row, is shown in Fig. 4.22(a). The diagonal bins correspond to the purity P , defined as the ratio of the number of events generated and reconstructed in a given bin, to the number of events generated in that bin. The same matrix, normalized by column, is shown in Fig. 4.22(b). In this case the diagonal bins correspond to the stability S , 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 migration effects, the main source being the limited resolution in the measurement of E_T^{miss} .



(a) Response matrix normalized by row

(b) Response matrix normalized by column

Figure 4.22.: Response matrix normalized by row (a) and by column (b) including all signal processes. The matrices are normalized either by row or by column in order to show the purity or stability respectively in diagonal bins.

Several 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 to build the p_T^H spectra at generator and reconstructed level. The reconstructed spectrum obtained using POWHEG V1 is then unfolded using the response matrix built with JHUGEN, and the unfolded spectrum is compared to the POWHEG V1

Table 4.7.: 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 ± 0.016	0.704 ± 0.015	0.727 ± 0.015	0.755 ± 0.014
15–45	0.701 ± 0.015	0.665 ± 0.016	0.683 ± 0.015	0.733 ± 0.015
45–85	0.717 ± 0.015	0.706 ± 0.015	0.709 ± 0.015	0.716 ± 0.015
85–125	0.634 ± 0.016	0.681 ± 0.015	0.714 ± 0.015	0.739 ± 0.015
125–165	0.599 ± 0.016	0.650 ± 0.016	0.700 ± 0.015	0.751 ± 0.014
165– ∞	0.632 ± 0.016	0.674 ± 0.015	0.701 ± 0.015	0.722 ± 0.015

spectrum at generator level. The result of this test shows good agreement between the two distributions.

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 results are shown in Table 4.7 for different values of the regularization parameter: starting from $k_{\text{reg}} = 2$ (stronger regularization) up to $k_{\text{reg}} = 5$ (weaker regularization). The criterion for choosing the best k_{reg} value is to increase the regularization as much as possible without introducing a bias, i.e. until a 68% coverage is fulfilled. This criterion leads to the same result as the prescription described before, 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 divided into three categories, depending on whether the uncertainty affects only the signal yield (type A), both the signal yield and the response matrix (type B), or only the response matrix (type C). These three classes propagate differently through the unfolding procedure.

Type A uncertainties are extracted directly from the fit in the form of a covariance matrix, which is passed to the unfolding tool as the covariance matrix of the measured distribution. The nuisance parameters belonging to this category are the background shape and normalization uncertainties. To extract the effect of type A uncertainties a dedicated fit is performed, fixing to constant all the nuisance parameters in the model, but type A

2256 nuisance parameters. The correlation matrix among the six signal strengths corresponding
 2257 to the six p_T^H bins, including all type A uncertainties, is shown in Fig. 4.23. The correlation
 2258 $\text{cor}(i,j)$ of bins i and j is defined as:

$$\text{cor}(i,j) = \frac{\text{cov}(i,j)}{s_i s_j} , \quad (4.16)$$

2259 where $\text{cov}(i,j)$ is the covariance of bins i and j , and (s_i, s_j) are the standard deviations
 2260 of bins i and j , respectively.

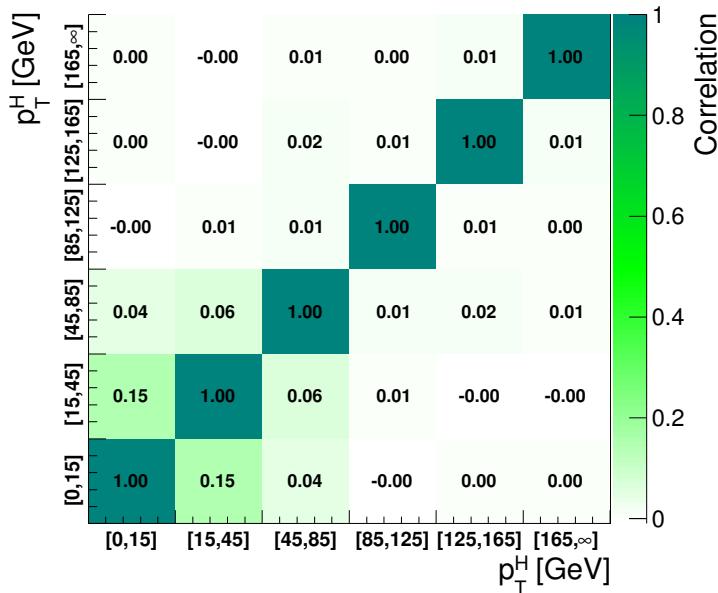


Figure 4.23.: Correlations among the signal strengths corresponding to the six p_T^H bins including all type A uncertainties.

2261 The nuisance parameters belonging to the Type B class are the ones related to:

- 2262 • b veto scale factor. It affects the signal and background templates by varying the
 2263 number of events with jets that enter the selection. It also affects the response matrix
 2264 because the reconstructed spectrum is harder or softer depending on the number of
 2265 jets, which in turn depends on the veto.
- 2266 • lepton efficiency scale factor. It affects the signal and background template shape and
 2267 normalization. It affects the response matrix by varying the reconstructed spectrum;
- 2268 • E_T^{miss} scale and resolution, which have an effect similar to the above;
- 2269 • lepton scale and resolution. The effect is similar to the above;
- 2270 • jet energy scale. It affects the signal and background template shape and normalization.
 2271 It also affects the response matrix because, by varying the fraction of events with jets,

Table 4.8.: 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. Uncertainties related to E_T^{miss} and lepton resolution are single-sided, i.e. only an up variation is implemented.

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

the b veto can reject more or fewer events, thus making the reconstructed spectrum harder or softer.

The effect of each type B uncertainty is evaluated separately, since each one changes the response matrix in a different way. In order to evaluate their effect on the signal strengths parameters, two additional fits are performed, each time fixing the nuisance parameter value to ± 1 standard deviation with respect to its nominal value. The results of the fits are then compared to the results of the full fit obtained by floating all the nuisance parameters, thus determining the relative uncertainty on the signal strengths due to each nuisance parameter, as shown in Tab. 4.8. 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 [106]. 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.

Table 4.9.: Differential cross section in each p_T^H bin, together with the total uncertainty and the separate components of the various sources of uncertainty.

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

2295 4.8. Results

2296 In order to unfold the spectrum, the procedure described in section 4.7 has been pursued.
 2297 The statistical plus type A systematic uncertainties are propagated by the unfolding
 2298 procedure into the final spectrum, taking into account the signal strengths covariance
 2299 matrix. The type B systematic uncertainty has been propagated using the following
 2300 procedure: for each p_T^H bin, we compute the upper bound of the systematic band computing
 2301 the square sum of all the signal strength variations that deviate in the up direction with
 2302 respect to the bin central value, whether or not this variation corresponds to the up or
 2303 down shift of the systematic uncertainty. The same is done for the lower bound of the
 2304 systematic band. If both the up and down shifts of a given nuisance parameter lead to a
 2305 same direction variation of the signal strength, only the larger variation is considered.

2306 The unfolded p_T^H spectrum is shown in Fig. 4.24. Statistical, systematic, and model
 2307 dependence uncertainties are shown as separate error bands in the plot. The model
 2308 dependence uncertainty corresponds to the effect of Type C errors described before.

2309 The unfolded spectrum is compared with the SM-based theoretical predictions where the
 2310 ggH contribution is modeled using the HRES and POWHEG V2 programs. The comparison
 2311 shows good agreement between data and theoretical predictions within the uncertainties.
 2312 The measured values for the differential cross section in each bin of p_T^H are reported together
 2313 with the total uncertainty in Table 4.9.

2314 Figure 4.25 shows the correlation matrix for the six bins of the differential spectrum,
 2315 where the correlation is defined as in Eq. (4.16). The correlation among unfolded bins is
 2316 mostly of statistical nature, arising from the unfolding procedure.

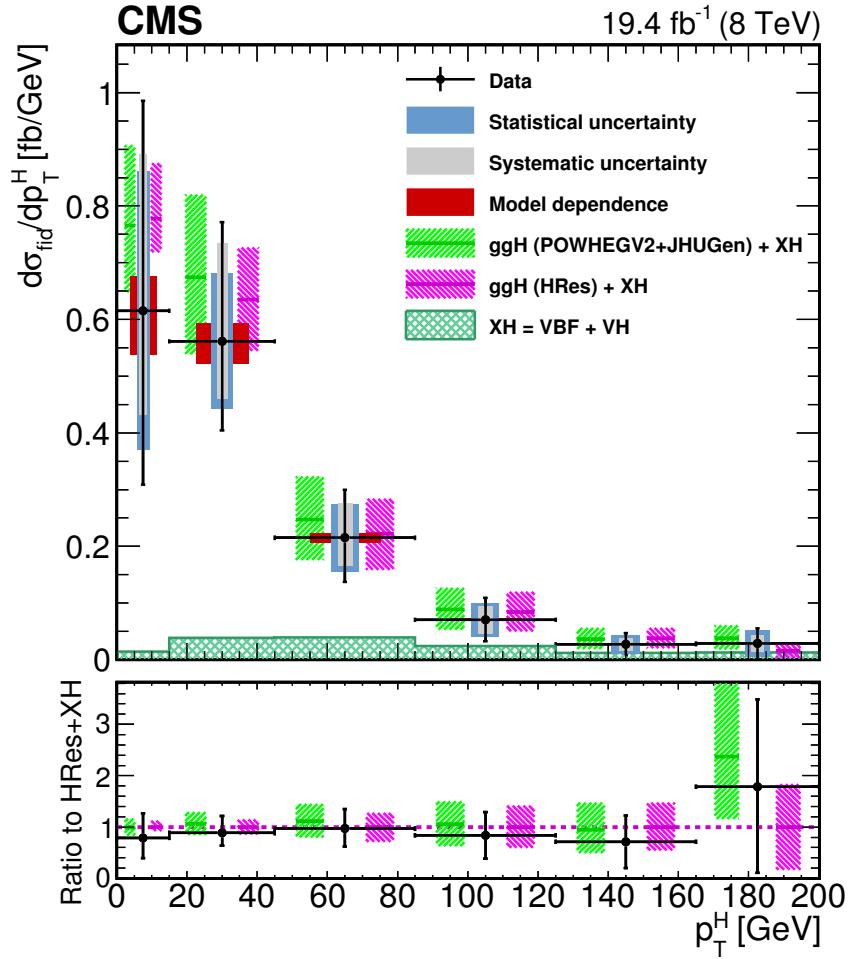


Figure 4.24.: Higgs boson production cross section as a function of p_T^H , after applying the unfolding procedure. Data points are shown, together with statistical and systematic uncertainties. The vertical bars on the data points correspond to the sum in quadrature of the statistical and systematic 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 modeled by HRES and POWHEG V2, respectively. The sub-dominant 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.

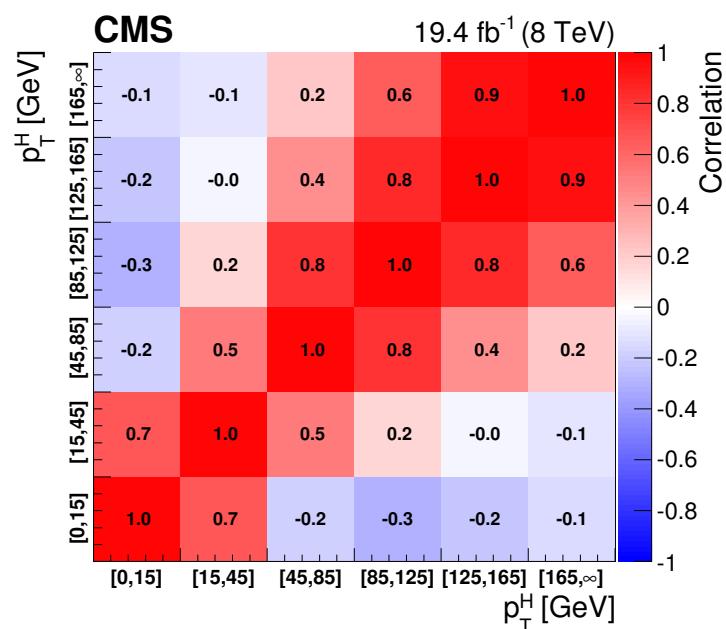


Figure 4.25.: Correlation matrix among the p_T^H bins of the differential spectrum.

Chapter 5.

2317 **Search for the SM Higgs boson in the 2318 $H \rightarrow WW$ channel with the first 2319 13 TeV LHC data**

2320 **5.1. Introduction**

2321 In this chapter, the first search for the SM Higgs boson decaying to a W boson pair at
2322 13 TeV is presented, using a total integrated luminosity of 2.3 fb^{-1} , collected during the
2323 2015 proton proton data taking period of the LHC.

2324 Final states in which the two W bosons decay leptonically are studied. Therefore,
2325 events with a pair of oppositely-charged leptons, exactly one electron and one muon, a
2326 substantial amount of missing transverse energy, E_T^{miss} , due to the presence of neutrinos in
2327 the final state, and either zero or one jet are selected. This signature is common to other
2328 processes, which enter the analysis as backgrounds. The main background comes from
2329 WW production, irreducible background that shares the same final states and can only be
2330 separated by the use of certain kinematic properties. Another important background is
2331 $W+jets$, where a jet can mimick a leptonic signature. Background coming from top quark
2332 events, i.e. $t\bar{t}$ and single top production, is also important, followed by other processes such
2333 as Drell-Yan, WZ , and other EWK production. The analysis strategy follows the one used
2334 during Run 1 in the same channel, described in Chapter 4, with a few different aspects that
2335 are described in the next sections.

2336 With respect to 8 TeV, the ggH production cross section at 13 TeV is expected to
2337 increase of a factor of 2, thus raising the number of expected signal events. In addition, the
2338 cross section for the background processes is increasing as well. The WW production cross
2339 section increases of a factor of 1.8 and the $t\bar{t}$ cross section of a factor of 3.5, due to the
2340 enhancement of the gluon PDFs at higher center of mass energies.

2341 **5.2. Data and simulated samples**

2342 Data recorded in proton proton collisions at 13 TeV during 2015 was used in the analysis,
2343 with a total integrated luminosity of 2.3 fb^{-1} . Single and double lepton triggers are used

²³⁴⁴ similarly to the same analysis at 8 TeV. The HLT paths and descriptions of the triggers used
²³⁴⁵ 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.

²³⁴⁶ The trigger efficiencies are measured in data and applied on simulated events as described
²³⁴⁷ in Sec. 4.2.1.

²³⁴⁸ Concerning the simulated samples, several different Monte Carlo (MC) generators were
²³⁴⁹ used. In the simulation, ‘lepton’ includes also τ . Higgs signal samples have been simulated
²³⁵⁰ in all channels with POWHEG v2 [70, 83, 107], designed to describe the full NLO properties
²³⁵¹ of these processes. In particular, for Higgs produced via gluon fusion [72], and vector-
²³⁵² boson-fusion (VBF) [73], the decay of the Higgs boson into two W boson and subsequently
²³⁵³ into leptons was done using JHUGEN v5.2.5 [108]. For associated production with a
²³⁵⁴ vector boson (W^+H , W^-H , ZH) [109], including gluon fusion produced ZH ($ggZH$), the
²³⁵⁵ Higgs decay was done via PYTHIA 8.1 [91]. Alternative signal samples were produced with
²³⁵⁶ AMC@NLO [75], or with POWHEG v2 but decayed via PYTHIA 8.1 for gluon fusion and
²³⁵⁷ VBF assuming a Higgs boson mass of 125 GeV. In the following, the mass of the SM Higgs
²³⁵⁸ boson is assumed to be 125 GeV.

²³⁵⁹ The WW production, irreducible background for the analysis, was simulated in different
²³⁶⁰ ways. POWHEG v2 [110] was used for $q\bar{q}$ produced WW in different decays. The cross
²³⁶¹ section used for normalizing WW processes produced via $q\bar{q}$ was computed at next-to-next-
²³⁶² to-leading order (NNLO) [111]. In order to control the top quark background processes,
²³⁶³ the analysis is performed with events that have no more than one high- p_T jet. The veto on
²³⁶⁴ high- p_T jets enhances the importance of logarithms of the jet p_T , spoiling the convergence
²³⁶⁵ of fixed-order calculations of the $q\bar{q} \rightarrow WW$ process and requiring the use of dedicated
²³⁶⁶ resummation techniques for an accurate prediction of differential distributions [112, 113].
²³⁶⁷ The p_T of the jets produced in association with the WW system is strongly correlated with
²³⁶⁸ 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%

2369 simulated $q\bar{q} \rightarrow WW$ events are reweighted to reproduce the p_T^{WW} distribution from the
 2370 p_T -resummed calculation.

2371 Gluon fusion produced WW was generated, with and without Higgs diagrams, using
 2372 MCFM v7.0 [114]. A $t\bar{t}$ sample dilepton sample was also generated using POWHEG v2. The
 2373 WW and $t\bar{t}$ samples produced specifically for this analysis are presented in Table 5.3. Other
 2374 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

2375 All processes are generated using the NNPDF2.3 [13, 115] parton distribution functions
 2376 (PDF) for NLO generators, while the LO version of the same PDF is used for LO generators.
 2377 All the event generators are interfaced to PYTHIA 8.1 [91] 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 [116]. 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 [117, 118]. For all processes, the detector response is simulated using a detailed description of the CMS detector, based on the GEANT4 package [84].

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 [119], 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. [66].

The cross section used for normalizing $q\bar{q}$ produced WW processes was computed at next-to-next-to-leading order (NNLO) [111]. 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 [120].

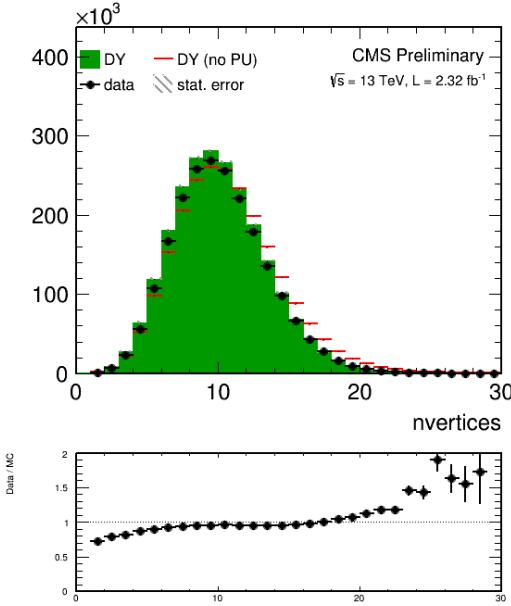


Figure 5.1.: Distributions of the number of vertices in a Drell-Yan enriched sample in data, together with the simulation before (red) and after (solid green) the pileup reweighting.

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 [121] at NLO. The $t\bar{t}$ cross section is also provided by the LHC Top Working group [122], and it is computed at NNLO, with NNLL soft gluon resummation.

Drell-Yan (DY) production of Z/γ^* is generated using AMC@NLO [75]. 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 [13, 115] 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 [116].

2420 5.3. Analysis strategy

2421 5.3.1. Event reconstruction

2422 Regarding the electrons, muons, jets and E_T^{miss} definition and reconstruction, the standard
 2423 CMS recommendations described in Chapter 2 are used. The specific selections used in
 2424 this analysis are briefly summarised below.

2425 Muons are identified according to the CMS recommendations for the medium working
 2426 point, with the addition of some extra cuts, as defined by the following selections:

- 2427 • identified by the standard medium muon selection described in Sec. ??; Not yet defined
 2428 :)
- 2429 • $p_T > 10 \text{ GeV}$;
- 2430 • $|\eta| < 2.4$;
- 2431 • $|d_{xy}| < 0.01 \text{ cm}$ for $p_T < 20 \text{ GeV}$ and $|d_{xy}| < 0.02 \text{ cm}$ for $p_T > 20 \text{ GeV}$, d_{xy} being the
 2432 transverse impact parameter with respect to the primary vertex;
- 2433 • $|d_z| < 0.1 \text{ cm}$, where d_z is the longitudinal distance of the muon track in the tracker
 2434 extrapolated along the beam direction.

2435 For the muon isolation, the CMS recommended particle flow isolation based on the
 2436 tight working point is used, corresponding to a requirement on the isolation variable of
 2437 $ISO_{\text{tight}} < 0.15$. In addition a tracker relative isolation is also applied.

2438 For the electron identification, the tight working point is used. In addition some
 2439 additional cuts to make the selection “trigger-safe” are included. This is done because the
 2440 electron triggers already include some identification and isolation requirements that are
 2441 based on the raw detector information, while the offline selections make use of particle flow
 2442 requirements. The “trigger-safe” selections are defined to make the the offline identification
 2443 and isolation requirements tighter with respect to the online triggers.

2444 The simulated events are corrected for the lepton trigger, identification and isolation
 2445 efficiencies measured in data using the same techniques described in Sec. 4.3.1.

2446 Jets are defined clustering the particle flow objects using the anti- k_t algorithm with a
 2447 distance parameter of 0.4. The CHS pileup mitigation technique is used. The L1, L2, L3
 2448 and L2L3 jet energy correction described in Sec. ?? are applied. The reject jets coming from
 2449 calorimeter or readout electronics noise, the loose working point for PF jet identification is
 2450 used.

2451 The b-tagging algorithm for this analysis is chosen comparing the performances of
 2452 different algorithms using simulations for signal and background contributions in the phase
 2453 space defined by the analysis kinematic requirements. More precisely, two MC samples are
 2454 used, one corresponding the the $H \rightarrow WW \rightarrow 2\ell 2\nu$ signal produced via the ggH production
 2455 mode and another corresponding to the $t\bar{t}$ process. In fact, the first sample is enriched
 2456 in light jets, i.e. originating by the hadronization of light quarks like u,d,c and s quarks,
 2457 while the second sample is enriched in b jets, coming from the top quark decay. The b-veto
 2458 efficiency, ϵ_{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 ≥ 2 jets are shown in Fig. 5.2. Events considered for this study are the ones passing the WW baseline selection.

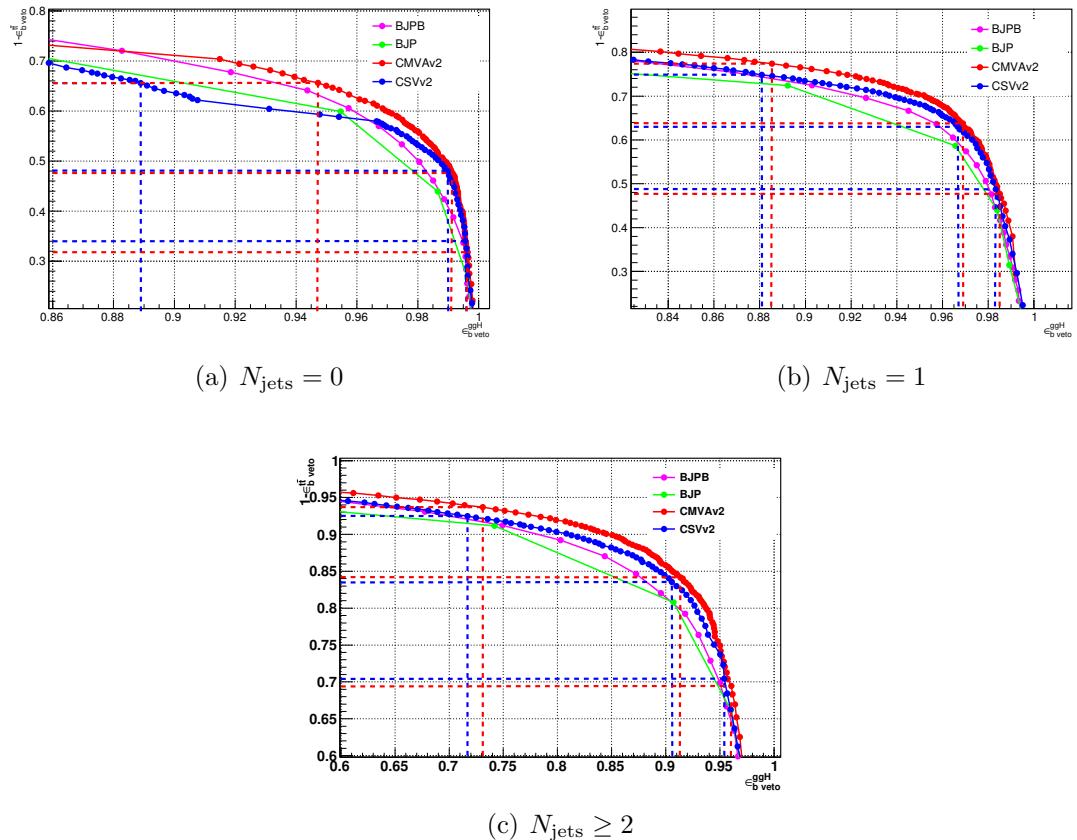


Figure 5.2.: ROC curve for the b veto efficiency on signal and background events. The blue and red lines point out the signal efficiency and the background rejection corresponding to the three working points considered for the CSVv2 and the cMVAv2 algorithms respectively.

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 rates¹ 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.

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

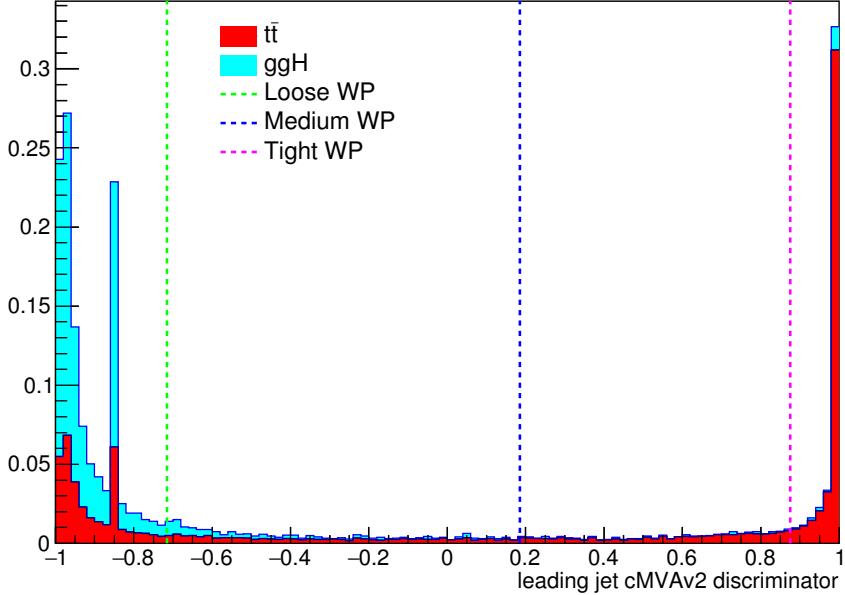


Figure 5.3.: cMVAv2 discriminator associated to the leading jet (with $p_T > 30$ GeV) both for the ggH and the $t\bar{t}$ processes. The two processes are normalized to unity and stacked. The vertical dashed lines show the discriminator value corresponding to the three working points.

2470 In order to determine the best working point for this analysis a preliminary significance
 2471 assessment is performed, using a complete analysis procedure in which only statistical
 2472 effects are taken into account (no systematics are included). The significance assessment
 2473 was performed using a two dimensional discriminating variable consisting of the dilepton
 2474 invariant mass versus the transverse mass. The assessment was performed with the following
 2475 leptonic selection:

- 2476 • two leptons, an electron and a muon with opposite charge, with leading lepton p_T
 2477 greater than 20 GeV and sub-leading lepton p_T greater than 13 GeV;
- 2478 • no other lepton (electron or muon) with p_T greater than 10 GeV;
- 2479 • $m_{\ell\ell}$ greater than 12 GeV;
- 2480 • PF type 1 corrected MET greater than 20 GeV;
- 2481 • $p_T^{\ell\ell}$ greater than 30 GeV.

2482 In addition to this global selection, two categories were identified:

- 2483 • 0 jets: no jets above 30 GeV, jets between 20 GeV and 30 GeV are b-vetoed with the
 2484 cMVAv2 WP under study;
- 2485 • 1 jet: exactly 1 jet above 30 GeV, no b-tagged jets above 30 GeV according to the
 2486 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

2487 The two categories were eventually combined together and the significance assessment was
 2488 repeated for the three working points. With these selection we find the significance values
 2489 listed in Table 5.5 for the three working points.

2490 The working point providing the best significance in the combined 0 + 1 jets category is
 2491 found to be the loose one.

2492 To correct for a possible different b tagging efficiency in data and simulation, the
 2493 simulated events are reweighted using scale factors computed in bins of the jet η and p_T .
 2494 These scale factors and the corresponding uncertainties are centrally calculated for each
 2495 working point, in such a way to be employable by all the CMS analyses. The prescription
 2496 to reweight the simulated events is the following. First of all one has to compute the b
 2497 tagging efficiency using the MC samples, $\varepsilon_{\text{MC}}(p_T, \eta, f)$, for the chosen working point in bins
 2498 of jet p_T and η . The efficiency has to be computed for different flavours f of the jets, b,
 2499 c and light (u,d,s), using the jet matching information² which is available in all the MC
 2500 samples. An MC-based event weight is then calculated computing the probability P_{MC} of a
 2501 given b tagging configuration to occur, e.g.:

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

2502 Afterwards, a similar probability is computed using data:

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

2503 where SF_i is the provided scale factor value for the relevant jet flavour, p_T and η . Products
 2504 in Eqs. 5.1 and 5.2 run over all jets. The event weight is finally given by the ration
 2505 $P_{\text{DATA}}/P_{\text{MC}}$.

2506 The b tagging efficiencies to be fed into Eq. 5.1 and Eq. 5.2 are derived using $t\bar{t}$ simulated
 2507 events and applying basic leptonic selections. These efficiencies are shown in Fig. 5.4 for

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

2508 light (a), c-jets (b) and b-jets (c), in bins of η and p_T . The uncertainties associated to the
 2509 efficiencies are representative of the statistics of the simulated $t\bar{t}$ sample, and are computed according to a binomial distribution.

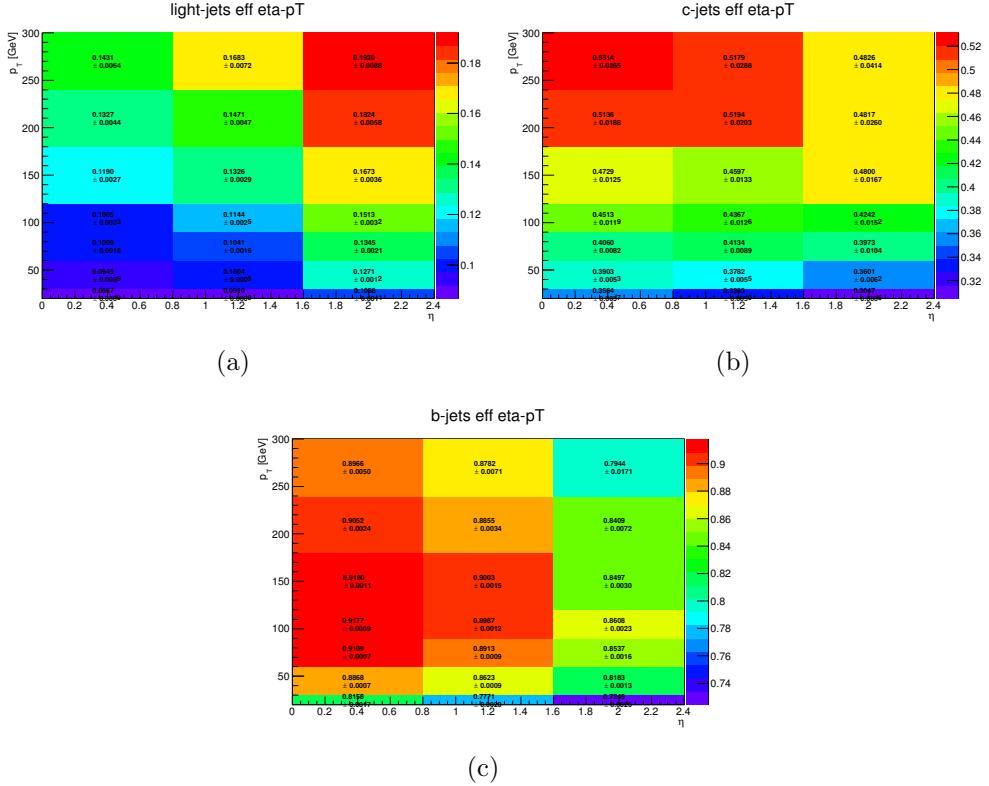


Figure 5.4.: B tagging efficiencies for light jets (a), c-jets (b) and b jets (c), as a function of η and p_T .

2510
 2511 The effect of the event reweighting is to correct the shape of the b tagging discriminator
 2512 in simulation, moving events from the b tag region (discriminator greater than > -0.715)
 2513 to the b veto region (discriminator < -0.715) and viceversa. A data/simulation comparison
 2514 of the b tagging discriminator for the leading and subleading jets is performed to check the
 2515 agreement after the application of the event weights. In order to evaluate the data/simulation
 2516 agreement for b-jets, the data and simulation are compared in a top enriched control region,
 2517 defined by the following requirements:

- 2518 • two leptons, an electron and a muon with opposite charge, with leading lepton p_T
 2519 greater than 20 GeV and sub-leading lepton p_T greater than 15 GeV;
- 2520 • no other lepton (electron or muon) with p_T greater than 10 GeV;
- 2521 • lepton invariant mass greater than 50 GeV;
- 2522 • at least two jets with p_T greater than 30 GeV;

- 2523 • at least one of the two leading jets with cMVAv2 btagging score greater than -0.715
 2524 (i.e. the loose working point).

2525 In order to evaluate the agreement for light jets, a second control region is defined, populated
 2526 by Z+light jet events, defined as follows:

- 2527 • two leptons, two electrons or two muons with opposite charge, with leading lepton p_T
 2528 greater than 20 GeV and sub-leading lepton p_T greater than 15 GeV.
 2529 • no other lepton (electron or muon) with p_T greater than 10 GeV.
 2530 • lepton invariant mass greater between 80 GeV and 110 GeV.
 2531 • at least two jets with p_T greater than 30 GeV.
 2532 • at least one jet above 30 GeV.
 2533 • no jets above 20 GeV with a TCHE score above 2.1.

2534 Although a Z+jets sample is dominated by light flavor jets, a b-veto on an alternative
 2535 algorithm (TCHE) is applied to reduce the contamination from b-jets, especially above the
 2536 cMVAv2 cut. This helps mitigating possible data/simulation discrepancies in the modeling
 2537 of the heavy/light flavour ratio. The comparison between data and simulation after the
 2538 event reweighting is shown in Figs. 5.5 and 5.6 for the b-jets and light jets enriched control
 2539 regions, respectively.

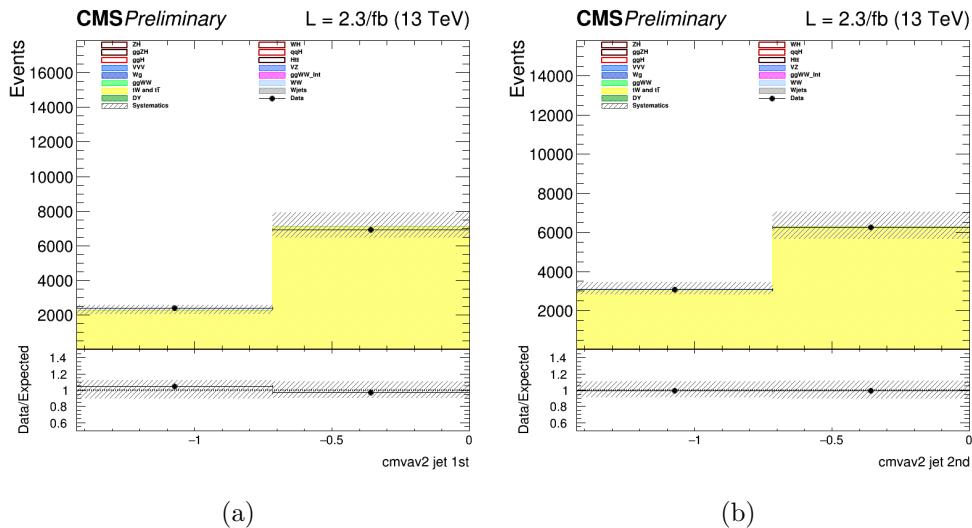


Figure 5.5.: B tagging cMVAv2 discriminator for the leading (a) and the subleading (b) jet in the b-jets enriched control region.

2540 5.3.2. Event selection and background rejection

2541 Since the ggH production mechanism, which is the main production mode for a Higgs
 2542 mass of around 125 GeV, is characterized by the emission of few jets arising from initial

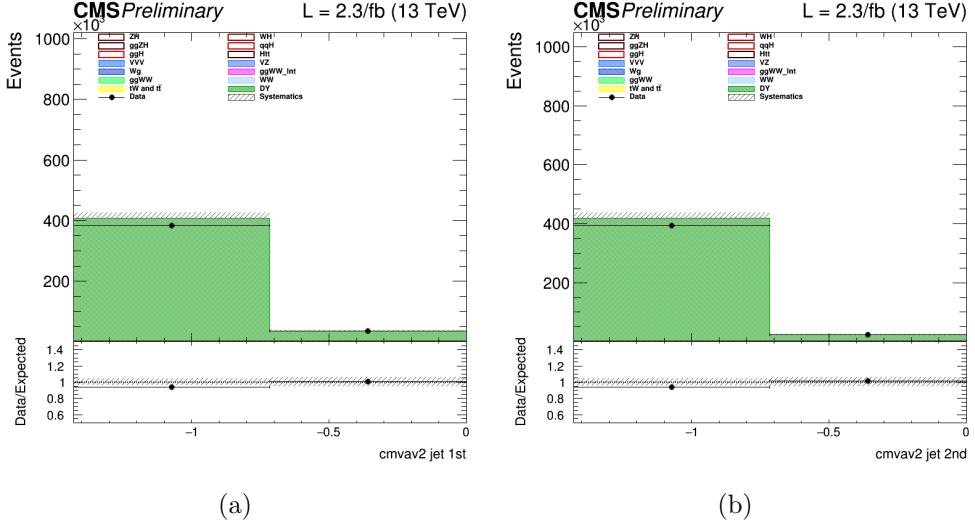


Figure 5.6.: B tagging cMVAv2 discriminator for the leading (a) and the subleading (b) jet in the light jets enriched control region.

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 τ leptons decaying to electron or muons. Exactly one electron and one muon are required to be reconstructed in the event with opposite charges and a minimum p_T of 10 (13) GeV for the muon (electron). One of the two leptons should also have a p_T greater than 20 GeV and both leptons are required to be well identified and isolated to reject fake leptons and leptons coming from QCD sources. To suppress background processes with three or more leptons in the final state, such as ZZ, WZ, $Z\gamma$, $W\gamma$, 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 τ lepton pair which subsequently decays to an $e\mu$ final state and suppress processes without genuine E_T^{miss} , a minimal E_T^{miss} of 20 GeV is required. The DY background is further reduced by requesting $p_T^{\ell\ell} > 30$ GeV. Finally the contribution from leptonic decays of single top and $t\bar{t}$ production is reduced by requesting that no jets with $p_T > 20$ GeV are identified by the b tagging algorithm as originating from a b quark in the event.

The requirements described above define the WW baseline selection. After those requirements the data sample is dominated by events arising from the non-resonant WW production and $t\bar{t}$ production. To further reduce the effect of these backgrounds on the signal sensitivity, the events are categorized depending on the jet multiplicity, counting jets with $p_T > 30$ GeV. Events with zero associated jets mainly arise from the WW production, while WW and $t\bar{t}$ productions have a similar contribution in the category with one jet.

2567 Higher jet multiplicity categories, which are sensitive to other Higgs production mechanisms,
2568 such as VBF, are not included in this analysis, given the very low expected yield for other
2569 production modes with the analysed integrated luminosity.

2570 Distributions of some variables of interest for the 0 and 1 jet categories separately,
2571 but merging the $e\mu$ and μe final states together, are shown in Figs. 5.7, 5.8 and 5.9 after
2572 applying the WW baseline selections, with the addition of a cut on $m_{\ell\ell}$ to remove the Higgs
2573 signal contribution ($m_{\ell\ell} > 80 \text{ GeV}$), and a cut on m_T to be orthogonal to the $Z\gamma^* \rightarrow \tau\tau$
2574 background control region ($m_T > 60 \text{ GeV}$).

2575 The $W+jets$ background, where one jet can be misidentified as a lepton, is a sub-
2576 dominant background in the phase space defined by the analysis kinematic requirements.
2577 The 0 ad 1 jets categories are further split according to the lepton flavour to $e\mu$ and μe ,
2578 where the first lepton refers to the leading one. In this way an improvement of about 10% in
2579 terms of the signal significance can be achieved, exploiting the different $W+jets$ background
2580 contribution in the two categories. Indeed the probability for a jet to be misidentified as an
2581 electron or a muon is not the same.

2582 5.3.3. Signal extraction

2583 To extract the Higgs boson signal contribution in the four previously mentioned categories,
2584 a similar approach to the one used in the Run 1 analysis [67] is pursued. The analysis is
2585 based on two-dimensional templates of $m_{\ell\ell}$ versus m_T to discriminate signal and background
2586 contributions. The $m_{\ell\ell}$ template is defined using 5 bins from $m_{\ell\ell} = 10 \text{ GeV}$ up to $m_{\ell\ell} =$
2587 110 GeV , while for the m_T template 7 bins are defined in the range $60 \text{ GeV} < m_T < 200 \text{ GeV}$.
2588 The phase space with $m_T < 60 \text{ GeV}$ is used as an orthogonal control region to extract the
2589 normalization of the DY background. A binned maximum likelihood fit to the signal and
2590 background two-dimensional templates is performed to extract the signal strength in the
2591 four categories.

2592 Distributions of the $m_{\ell\ell}$ and m_T variables after the WW level selection are shown in
2593 Fig. 5.10 for the 0 and 1 jet categories separately, but merging the $e\mu$ and μe final states
2594 together.

2595 The statistical methodology used to interpret the data and to combine the results from
2596 the independent 0-jet and 1-jet categories in the $e\mu$ and μe final states has been developed
2597 by the ATLAS and CMS collaborations in the context of the LHC Higgs Combination
2598 Group [101, 106]. The number of events in each category and in each bin of two-dimensional
2599 template is modelled as a Poisson random variable, with a mean value given by the sum
2600 of the contributions from all the processes under consideration. Systematic uncertainties
2601 are represented by individual nuisance parameters with log-normal distributions. The
2602 uncertainties affect the overall normalization of the signal and backgrounds as well as the
2603 shape of the predictions across the distribution of the observables. Correlation between
2604 systematic uncertainties in different categories are taken into account.

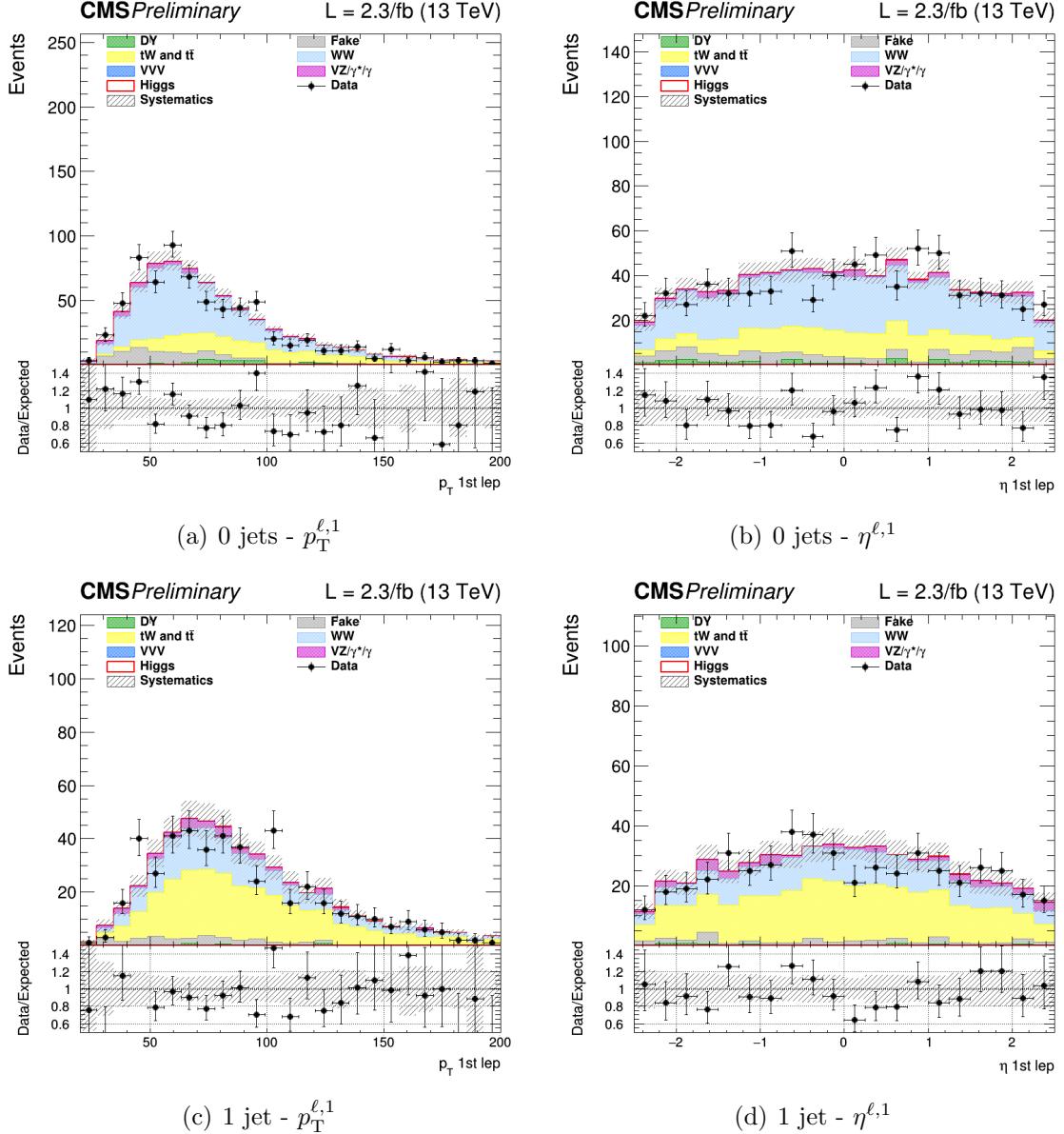


Figure 5.7.: Distributions of p_T (left) and η (right) of the leading lepton for events with 0 jet (upper row) and 1 jet (lower row), for the main backgrounds (stacked histograms), and for a SM Higgs boson signal with $m_H = 125$ GeV (superimposed and stacked red histogram) at the WW selection level.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

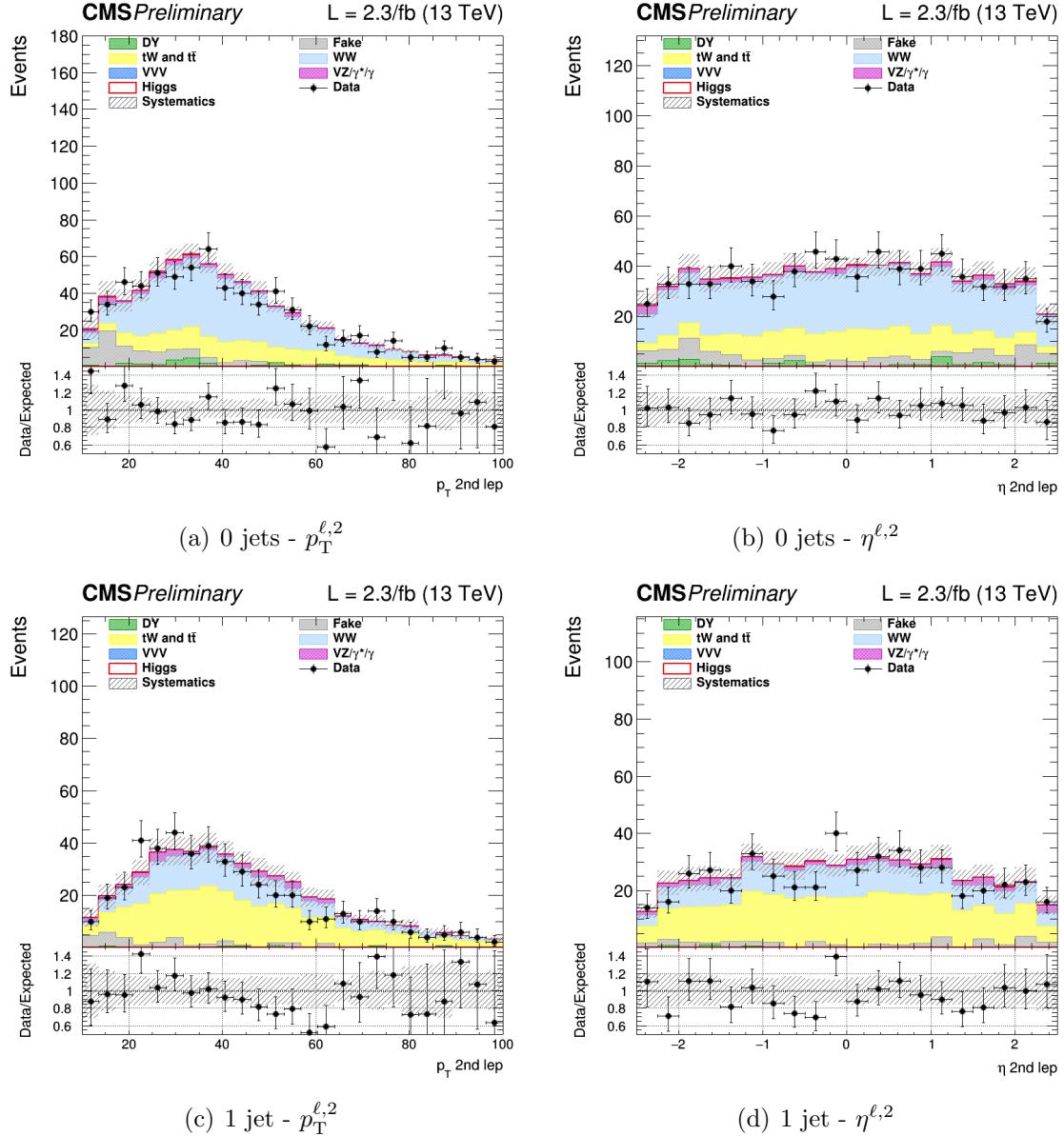


Figure 5.8.: Distributions of p_T (left) and η (right) of the subleading lepton for events with 0 jets (upper row) and 1 jet (lower row), for the main backgrounds (stacked histograms), and for a SM Higgs boson signal with $m_H = 125$ GeV (superimposed and stacked red histogram) at the WW selection level.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

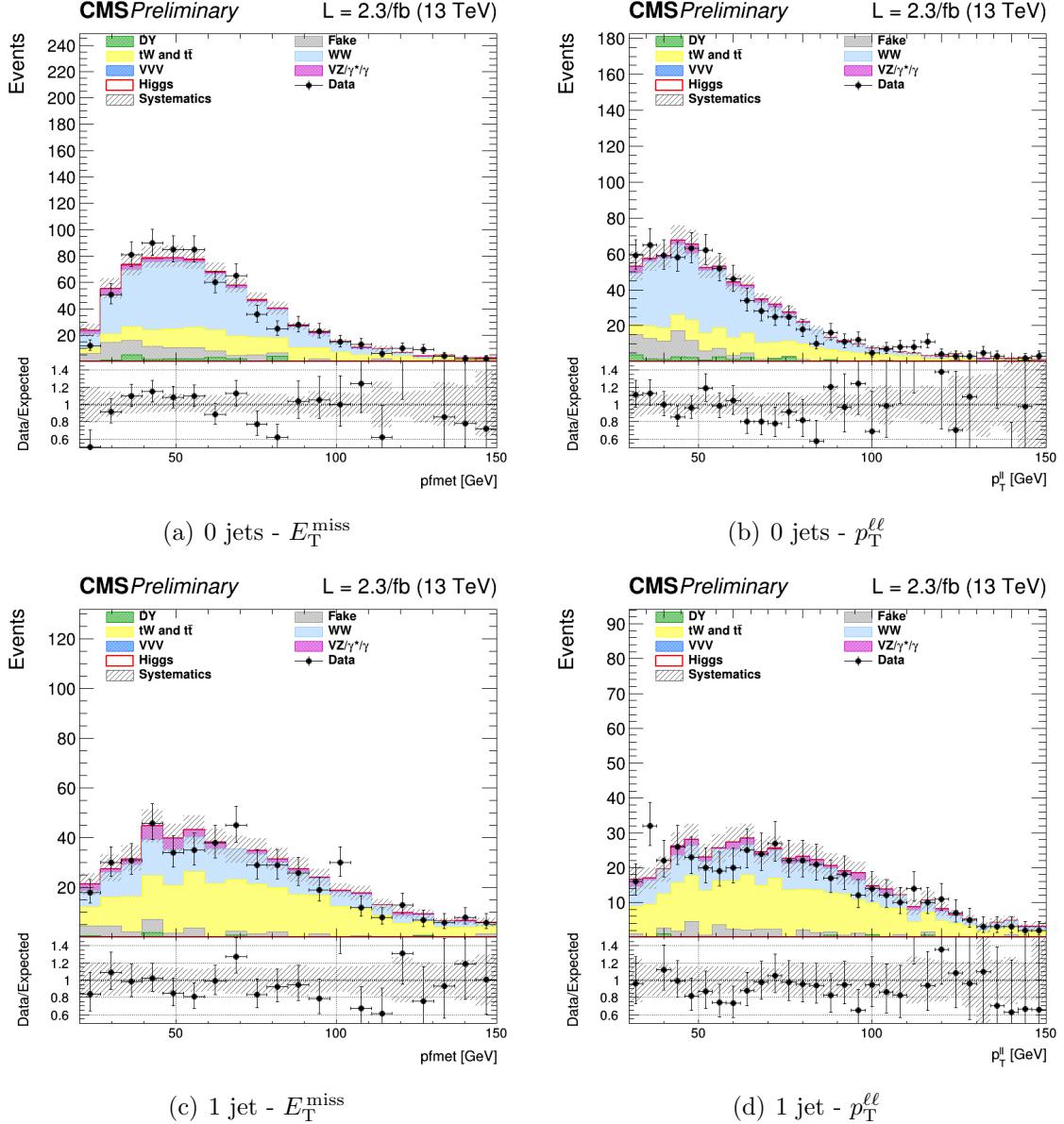


Figure 5.9.: Distributions of E_T^{miss} (left) and $p_T^{\ell\ell}$ (right) for events with 0 jets (upper row) and 1 jet (lower row), for the main backgrounds (stacked histograms), and for a SM Higgs boson signal with $m_H = 125$ GeV (superimposed and stacked red histogram) at the WW selection level.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

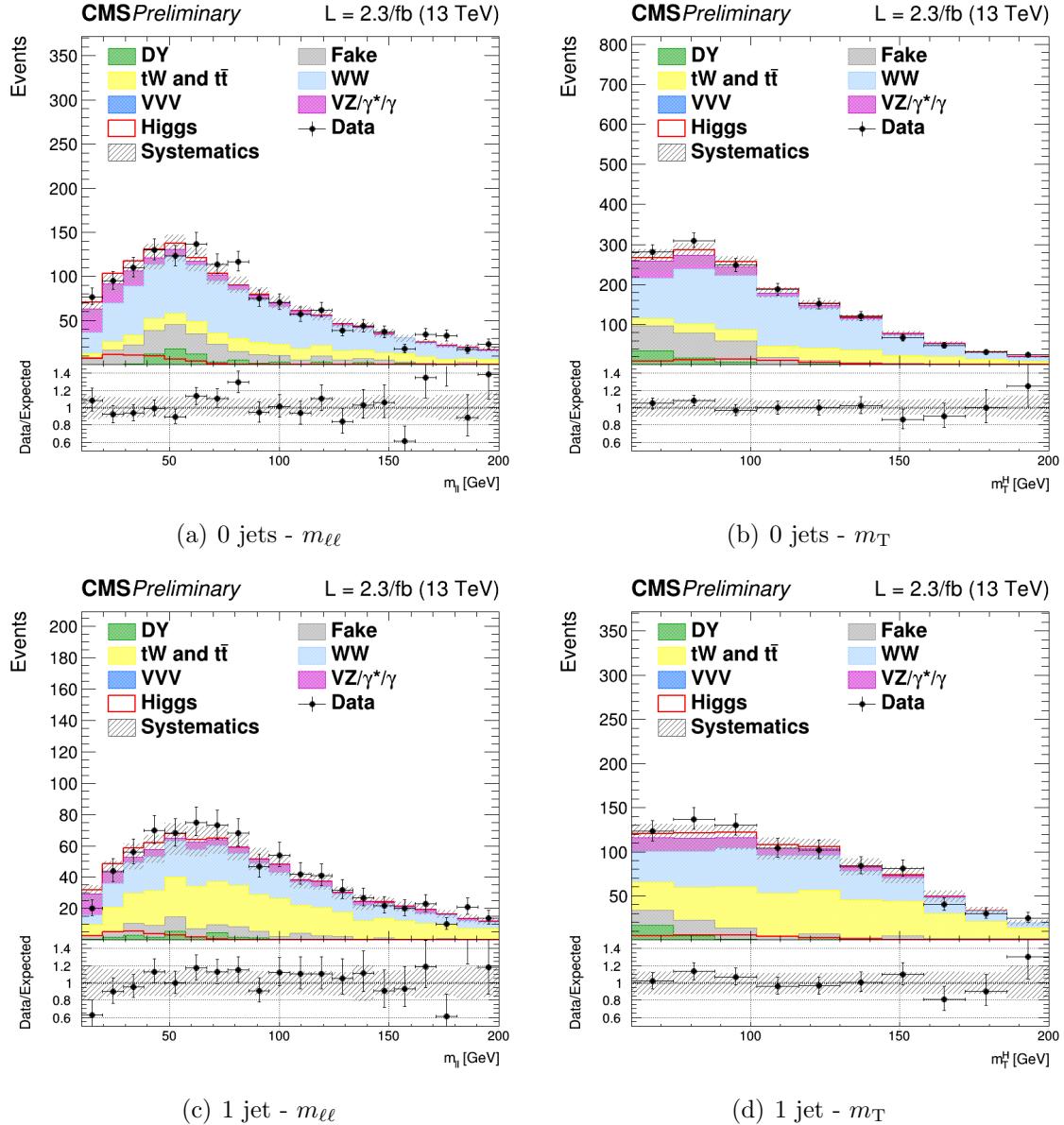


Figure 5.10.: Distributions of $m_{\ell\ell}$ (left) and m_T (right) for events with 0 jets (upper row) and 1 jet (lower row), for the main backgrounds (stacked histograms), and for a SM Higgs boson signal with $m_H = 125$ GeV (superimposed and stacked red histogram) at the WW selection level. The last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

2605 5.4. Background estimation

2606 The main background processes affecting the analysis signature, non-resonant WW pro-
 2607 duction and top quark processes, are estimated using data. Backgrounds arising from an
 2608 experimental misidentification of the objects, such as W+jets (also called “Fake”), are
 2609 estimated using data as well. The other minor backgrounds are generally estimated directly
 2610 from simulation as described in the following subsections.

2611 5.4.1. WW background

2612 The quark-induced WW background is simulated with NLO accuracy in perturbative
 2613 QCD, and the transverse momentum of the diboson system is reweighted to match the
 2614 NNLO+NNLL accuracy from theoretical calculations [112, 113]. However, given the large
 2615 uncertainties on the jet multiplicity distribution associated to this process, the normalization
 2616 of this background is measured from data separately for the 0 and 1 jet categories. The
 2617 normalization k-factors are extracted directly from the fit together with the signal strengths,
 2618 leaving the WW normalization free to float separately in the two jet multiplicity categories.
 2619 An orthogonal control region for the WW background normalization estimation is not
 2620 needed in this case, owing to the different $m_{\ell\ell}$ - m_T shape for signal and background.

2621 The gluon-induced WW production is sub-dominant with respect to the quark-induced
 2622 production, and its shape and normalization is fully taken from simulation, scaling the
 2623 cross section to the theoretical prediction with NLO accuracy [120].

2624 5.4.2. Top quark background

2625 As explained in Sec. 5.3, the production of top quark pairs represents one of the dominant
 2626 backgrounds in this analysis given its large cross section and a similar final state compared
 2627 to the signal. A b-jet veto, based on the *cMVAv2* b tagging algorithm, is used to suppress
 2628 this background and a reweighting procedure is applied on top of the simulated events to
 2629 correct for different b tagging efficiency in data and simulation.

2630 The top quark background normalization is measured using data, defining a b-jets
 2631 enriched control region by inverting the b-jet veto. More precisely, the b-jets enriched
 2632 control region for the 0-jet category is defined with the same WW baseline selection but
 2633 requiring at least one jet with $20 < p_T < 30$ GeV to be identified as a b jet and no other
 2634 jets with $p_T > 30$ GeV. For the 1-jet category, the b-jets enriched region is defined requiring
 2635 exactly one jet with $p_T > 30$ GeV identified as a b-jet. To reduce other backgrounds in
 2636 these two regions, the dilepton mass has to be greater than 50 GeV. Distributions of the
 2637 $m_{\ell\ell}$ and m_T variables in the b-jets enriched control regions after applying the data driven
 2638 estimation are shown in Figure 6.5, for the 0 and 1 jet categories separately.

2639 The top quark background normalization is constrained during the fit procedure sepa-
 2640 rately in the two jet categories, by means of the control regions defined above, which are
 2641 treated in the fit as two additional categories.

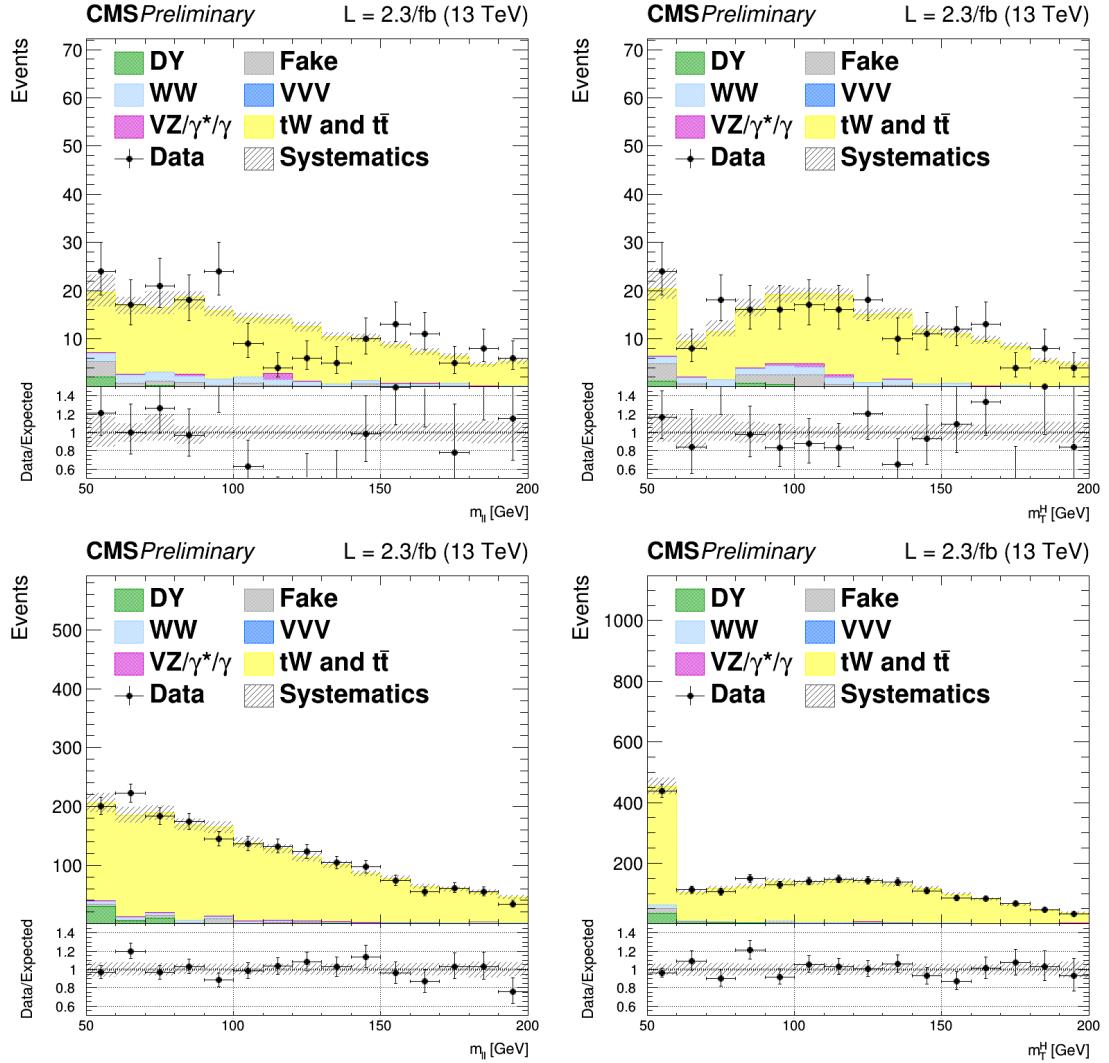


Figure 5.11.: Distributions of $m_{\ell\ell}$ (left) and m_T (right) for events with 0 jet (top) and 1 jet (bottom) in top enriched phase space. Scale factors estimated from data are applied. The first (last) bin includes underflows (overflows).

2642 5.4.3. Jet-induced (or Fake) background

2643 One of the primary source belonging to this category arises from the misidentification of
 2644 leptons in $W+jets$ processes in the 0 jet category. Also, semileptonic $t\bar{t}$ decays contribute
 2645 especially for higher jet multiplicities. Multijet production and hadronic $t\bar{t}$ decays are also
 2646 taken into account, but have a much smaller contribution.

2647 This background is fully estimated using data, with the technique described in Sec. 4.4.3.
 2648 To check the agreement of the background estimated in this way with data, a control sample
 2649 enriched in jet-induced events is defined. The events in the control sample are selected
 2650 applying the WW baseline requirements but requesting an $e\mu$ pair with same charge, which
 2651 significantly suppresses the WW and $t\bar{t}$ processes. The $m_{\ell\ell}$ distributions in this control
 2652 region for the 0 and 1 jet categories are shown in Fig. 5.12. From the crosscheck in this
 2653 control region, a global normalization factor of 0.8 is derived and applied to the jet-induced
 2654 background.

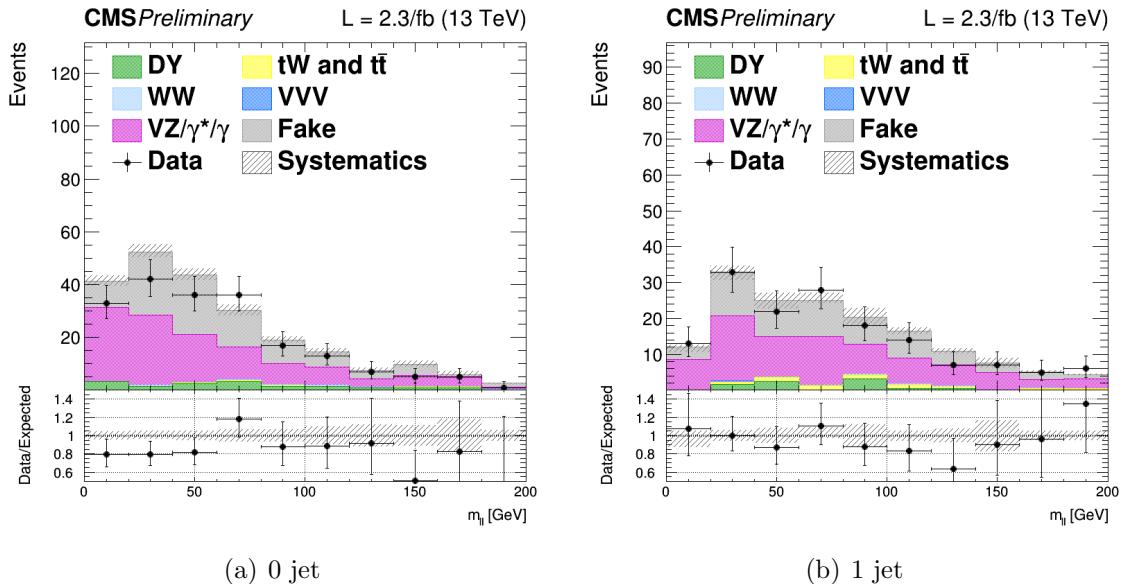


Figure 5.12.: Control plots for $m_{\ell\ell}$ in a fakes enriched phase space for events with 0 and 1 jet with $p_T > 30$ GeV, in $e\mu$ final state. Fake contribution has been scaled by 0.8 to match data.

2655 5.4.4. DY background

2656 This background contributes to the analysis phase space because of the Z/γ^* decays to a
 2657 pair of τ leptons, which consequently decays to an $e\mu$ pair. This background process is
 2658 predominant in the low m_T region, which is used as an orthogonal control region to determine
 2659 the background normalization in the 0 and 1 jet categories separately. In particular this
 2660 control region is defined by selecting events with $m_T < 60$ GeV and $30 \text{ GeV} < m_{\ell\ell} < 80$ GeV.

2661 The $m_{\ell\ell}$ distributions in these control regions for the 0 and 1 jet categories are shown in
 2662 Fig. 5.13.

2663 As for the top quark background, the normalization of this background in the 0 and 1
 2664 jet categories, is constrained directly in the fit by means of the control regions, which are
 2665 treated as two additional categories.

2666 The kinematics of this background is taken from simulation, after reweighting the Z
 2667 boson p_T spectrum to match the observed distribution measured in data. In fact, this
 2668 variable is not well reproduced by the MC generator used for simulating this process,
 2669 especially in the bulk of the distribution, the discrepancy being ascribed to the missing
 2670 contribution from resummed calculations.

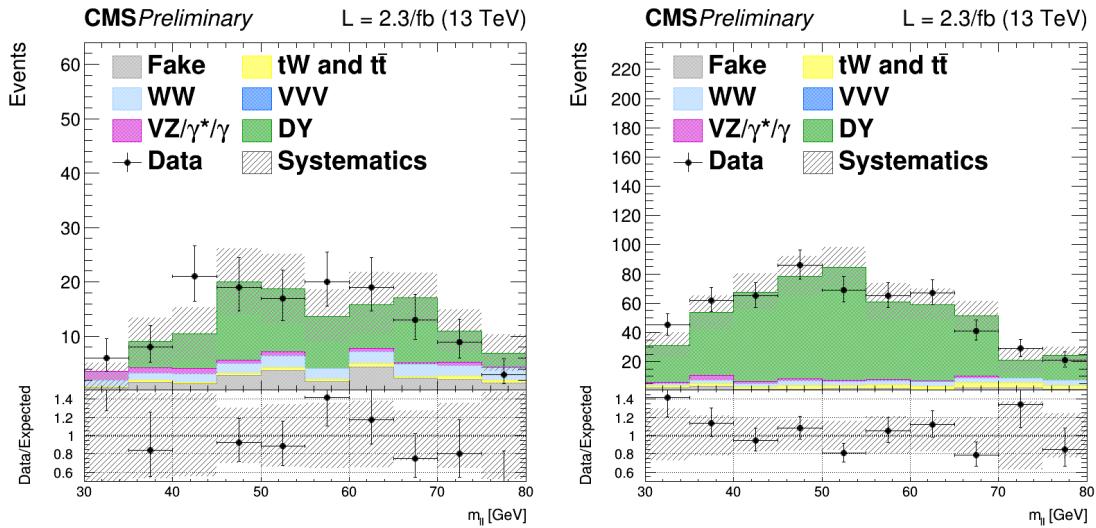


Figure 5.13.: Distributions of $m_{\ell\ell}$ for events with 0 jet (left) and 1 jet (right) in the $DY \rightarrow \tau\tau$ enriched control region. Scale factors estimated from data are applied.

2671 5.4.5. Other backgrounds

2672 The $W\gamma^*$ and the WZ electroweak processes can be gathered in the same physical process,
 2673 although the final state kinematics is rather different. In particular, the invariant mass of
 2674 the leptons arising from the γ^* decays is generally below 4 GeV, while the leptons from the
 2675 Z boson decay are characterized by a larger invariant mass. Another background which can
 2676 be experimentally identical to those is the $W\gamma$ production, where a real photon is produced
 2677 in association with a W boson and consequently undergoes a photon conversion to leptons
 2678 due to the interaction with the material constituting the first layers of the silicon tracker.

2679 All these backgrounds may contribute to the signal phase space whenever one of the three
 2680 leptons escape from the detector acceptance or is not identified. The shape and cross section
 2681 of these backgrounds are taken from simulation. The only exception is the normalization of
 2682 the $W\gamma^*$ background, being this process dominant in the low $m_{\ell\ell}$ region, which is scaled

2683 to data defining a proper control region. The control region is defined selecting events
 2684 with three isolated muons, with $p_T > 10.5$ and 3 GeV for the first three leading muons
 2685 respectively. The selection is further defined by $E_T^{\text{miss}} < 25$ GeV and E_T^{miss} projected to the
 2686 leading muon < 45 GeV. The pair of muons with the smallest invariant mass is taken as
 2687 coming from the γ^* decay. The k-factor measured in data for this background to be applied
 2688 in the simulation is 1.98 ± 0.54 .

2689 All remaining backgrounds from di-boson and tri-boson production, which are of mi-
 2690 nor importance in the analysis phase space, are normalized according to their expected
 2691 theoretical cross sections.

2692 5.5. Systematic uncertainties

2693 The systematic uncertainties affecting this measurement can be divided into three categories:
 2694 the uncertainties on the background estimation, experimental uncertainties and theoretical
 2695 uncertainties.

2696 The first category includes the uncertainties related to the background normalization
 2697 and shape. For the non-resonant WW production the shape is taken from simulation.
 2698 The input normalization to the fit is set to the expected value from simulation, and an
 2699 unconstrained nuisance parameter with a flat distribution is associated to this number.
 2700 This is done separately for the two jet categories.

2701 The top quark background shape is taken from simulation after correcting for the b
 2702 tagging scale factors. An uncertainty due to these scale factors is included and affects
 2703 both the normalization and the shape of the top quark background. The uncertainties on
 2704 the normalization are treated similarly to the WW background case, but constraining the
 2705 corresponding nuisances by means of the two control regions orthogonal to the signal phase
 2706 space. A similar procedure is used for the DY background.

2707 Effects due to experimental uncertainties are studied by applying a scaling and smearing
 2708 of certain variables related to the physics objects, e.g. the p_T of the leptons, followed by
 2709 a subsequent recalculation of all the correlated variables. This is done for simulation, to
 2710 account for possible systematic mismodeling.

2711 All experimental sources, except luminosity, are treated both as normalization and shape
 2712 uncertainties, and are correlated among the signal and background processes and all the
 2713 categories. The following experimental uncertainties are considered:

- 2714 • the uncertainty determined by the CMS online luminosity monitoring, 2.7% for the
 2715 first data collected at $\sqrt{s} = 13$ TeV;
- 2716 • the acceptance uncertainty associated with the combination of single and double lepton
 2717 triggers, which is 2%;
- 2718 • the lepton reconstruction and identification efficiencies uncertainties, that are in the
 2719 range 0.5-5% for electrons and 1-7% for muons depending on p_T and η ;
- 2720 • the muon momentum and electron energy scale and resolution uncertainties, that
 2721 amount to 0.01-0.5% for electrons and 0.5-1.5% for muons depending on p_T and η ;

- 2722 • the jet energy scale uncertainties, that vary between 1-11% depending on the p_T and
2723 η of the jet;
- 2724 • the E_T^{miss} resolution uncertainty, that is taken into account by propagating the corre-
2725 sponding uncertainties on the leptons and jets;
- 2726 • the scale factors correcting the b tagging efficiency and mistagging rate, that are varied
2727 within their uncertainties. This systematic uncertainty is anticorrelated between the
2728 top control regions and the other ones.

2729 The uncertainties in the signal and background production rates due to theoretical
2730 uncertainties include several components, which are assumed to be independent: the PDFs
2731 and α_s , the underlying event and parton shower model, and the effect of missing higher-order
2732 corrections via variations of the renormalization and factorization scales.

2733 The effects of the variation of PDFs, α_s and renormalization/factorization QCD scales,
2734 mainly affect the signal processes, being the most important backgrounds estimated using
2735 data driven techniques. However, the uncertainties on minor backgrounds that are estimated
2736 from simulation are taken into account. These uncertainties are split in the uncertainties
2737 on the cross section, which are computed by the LHC cross section working group [123],
2738 and on the selection efficiency [124]. The PDFs and α_s signal cross section normalization
2739 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
2740 mechanism. The PDFs and α_s acceptance uncertainties are less than 1% for gluon- and
2741 quark-induced processes. The effect of the QCD scales variation on the selection efficiency is
2742 around 1-3% depending on the specific process. To estimate these uncertainties, the events
2743 are reweighted according to different QCD scales or different PDF sets and the selection
2744 efficiency is recomputed each time. For the QCD scale uncertainty the maximum variation
2745 with respect to the nominal value is taken as the uncertainty. For the case of PDF and α_s
2746 uncertainties, the distribution of the selection efficiency is built taking into account all the
2747 replicas in the NNPDF3.0 set and the uncertainty is estimated as the standard deviation of
2748 that distribution.

2749 In addition, the categorization of events based on jet multiplicity introduces additional un-
2750 certainties on the ggH production mode related to missing higher order corrections. These un-
2751 certainties are evaluated following the prescription described in Sec. [subsec:stewart-tackman]
2752] and correspond to 5.6% for the 0-jet and 13% for the 1-jet bin categories.

2753 The underlying event uncertainty is estimated by comparing two different PYTHIA
2754 8 tunes, while parton shower modelling uncertainty is estimated by comparing samples
2755 interfaced with the PYTHIA 8 and HERWIG++ parton shower programs. The effect on the
2756 ggH (VBF) signal expected yield is about 5% (5%) for the PYTHIA 8 tune variation and
2757 about 7% (10%) for the parton shower description.

2758 Other specific theoretical uncertainties are associated to some backgrounds. An uncer-
2759 tainty on the ratio of the $t\bar{t}$ and tW cross sections is included. Indeed, these two processes
2760 are characterized by a different number of b-jets in the final state (2 b-jets for $t\bar{t}$ and 1
2761 for tW) and the b-veto acts differently for the two. A variation of the relative ratio of the
2762 cross sections can thus cause a migration of events from the 0 to the 1 jet categories and

2763 viceversa. The corresponding uncertainty is of 8%, according to the theoretical cross section
2764 calculations [121, 122].

2765 The $gg \rightarrow WW$ background LO cross section predicted by the MCFM generator is scaled
2766 to the NLO calculation, applying a k-factor of 1.4 with an uncertainty of 15% [120]. The
2767 interference term between the $gg \rightarrow WW$ and the ggH signal is also included and simulated
2768 with LO accuracy using MCFM. The k-factor to scale the interference term is 1.87, given by
2769 the geometrical average of the LO to NNLO $gg \rightarrow H \rightarrow WW$ scale factor (2.5) and the LO
2770 to NLO $gg \rightarrow WW$ scale factor (1.4). The uncertainty on this value is estimated as the
2771 maximum variation with respect to the two scale factors mentioned above, and is found
2772 to be of 25%. Anyway, with the current amount of integrated luminosity, the interference
2773 contribution is found to be negligible.

2774 For what the $qq \rightarrow WW$ background shape is concerned, an uncertainty related to
2775 the diboson p_T reweighting is evaluated varying the renormalization, factorization and
2776 resummation QCD scales.

2777 Finally, the uncertainties due to the limited statistical accuracy of the MC simulations
2778 are also taken into account, including an independent uncertainty for each bin of the
2779 two-dimensional distribution, and for each category. The uncertainty for a certain bin
2780 and process is given by the standard deviation of the Poisson distribution with mean
2781 corresponding to the number of MC events in that bin.

2782 5.6. Results

2783 The expected and observed signal significance are shown in Table 5.6 for all the categories
2784 separately. Also, the observed signal strengths and the corresponding uncertainties are
2785 shown. The best fit signal strength obtained combining all the categories together is found
2786 to be $0.3^{+0.5}_{-0.5}$, corresponding to an observed significance of 0.7σ , to be compared with the
2787 expected significance of 2.0σ for a Higgs boson mass of 125 GeV.

2788 Maybe I should add the nuisances impact plots...

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

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

Chapter 6.

2789 **Search for high mass resonances 2790 decaying to a W boson pair with first 2791 13 TeV LHC data**

2792 **6.1. Introduction**

2793 In this chapter, a search for a high mass spin-0 particle (from now on denoted as X)
2794 in the $X \rightarrow WW \rightarrow \ell\nu\ell'\nu'$ decay channel is presented, where ℓ and ℓ' refer to an different
2795 flavour lepton pair, i.e. $e\mu$. The search is based upon proton-proton collision data samples
2796 corresponding to an integrated luminosity of up to 2.3 fb^{-1} at $\sqrt{s} = 13 \text{ TeV}$, recorded by
2797 the CMS experiment at the LHC during 2015. This analysis represents a general extension
2798 of the SM Higgs boson search presented in 5 and is performed in a range of heavy scalar
2799 masses from $M_X = 200 \text{ GeV}$ up to 1 TeV , extending the range studied in a similar analysis
2800 performed using Run 1 LHC data [125], which provided upper limits on the production
2801 cross section of new scalar resonances up to 600 GeV .

2802 Despite the discovery of a particle consistent with the SM Higgs boson in 2012, there is a
2803 possibility that this particle is only a part of a larger Higgs sector, and hence only partially
2804 responsible of the EW symmetry breaking. This can be achieved in different theoretical
2805 models that extends the SM, such as the two-Higgs-doublet models [4, 126, 127], or models
2806 in which the SM Higgs boson mixes with a heavy EW singlet, which predict the existence
2807 of an additional resonance at high mass, with couplings similar to those of the SM Higgs
2808 boson, as most recently described in [3, 128].

2809 This analysis reports a generic search for a scalar particle with different resonance decay
2810 widths hypothesis, produced via the ggH and VBF production mechanisms. The results
2811 can then be interpreted in terms of different theoretical models. This analysis is heavily
2812 based on the SM Higgs search described in 5, in terms of physics objects, selections and
2813 background estimation. The differences and similarities are discussed in this chapter.

2814 **6.2. Data and simulated samples**

2815 The data sets, triggers, pile up reweighting, lepton identification and isolation used in this
2816 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 [129] 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.

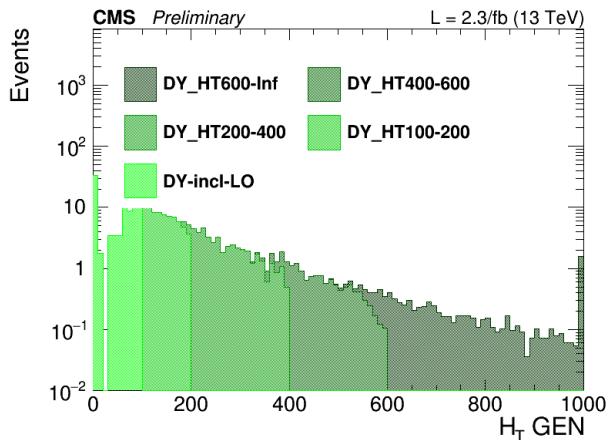


Figure 6.1.: Generator level H_T distribution for the merged DY sample.

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

2844 6.3. Analysis strategy

2845 The analysis strategy for the first results on the high mass search in the $W^+W^- \rightarrow 2\ell 2\nu$
 2846 decay channel closely follows the strategy presented in the 13 TeV SM Higgs search in the
 2847 $H \rightarrow W^+W^- \rightarrow 2\ell 2\nu$ channel regarding the 0 and 1 jet categories. In addition a dedicated
 2848 category to the VBF production mechanism is added, given that this production mode is
 2849 particularly important in the high mass region. Indeed, assuming a SM Higgs boson, the
 2850 ratio of cross sections $\sigma_{\text{VBF}}/\sigma_{\text{ggH}}^1$ increases with the Higgs boson mass, making the VBF
 2851 production mechanism more and more important as the mass of the resonance approaches
 2852 to high values.

2853 This analysis is affected essentially by the same background processes as the SM Higgs
 2854 boson search, with the difference that in this case the SM Higgs boson processes, including
 2855 all production modes, are treated as backgrounds.

2856 In addition to requiring the events to pass the single or double lepton triggers, exactly
 2857 one electron and one muon are required to be reconstructed in the event with opposite
 2858 charges and a minimum p_T of 20 GeV for both the muon and electron. Both leptons are
 2859 required to be well identified and isolated to reject fake leptons and leptons coming from
 2860 decays in flight. To suppress background processes with three or more leptons in the final
 2861 state, such as diboson or triboson production, events with any additional identified and
 2862 isolated lepton with $p_T > 10$ GeV are rejected. To suppress the contribution of the SM
 2863 production of the Higgs boson at 125 GeV, $m_{\ell\ell}$ is requested to be higher than 50 GeV. The
 2864 other event requirements are identical to the 125 GeV Higgs boson search and are described
 2865 in Sec. 5.3.2.

2866 In addition to the 0 and 1 jet categories, a specific category sensitive to the VBF
 2867 production mode is defined exploiting the characteristic signature of this process, where
 2868 two energetic jets are emitted in the forward region of the detector and with large $\Delta\eta$
 2869 gap. Events belonging to the VBF-enriched category are selected by requiring at least
 2870 two jets with $p_T > 30$ GeV, an invariant mass $m_{jj} > 500$ GeV and a gap in pseudorapidity
 2871 $\Delta\eta_{jj} > 3.5$.

2872 In addition to the transverse mass variable m_T , which is used in the analysis selection
 2873 to define the DY background control region, an additional variable is defined, that from
 2874 now on will be labelled as “improved transverse mass” m_T^i . This variable is defined as
 2875 the invariant mass of the four momentum resulting from the sum of the two leptons four
 2876 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.:

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

2877 This variable allows having a better sensitivity to different resonance mass hypothesis
 2878 as shown in Fig. 6.2, where the shape of the m_T^i variable is shown for different SM Higgs
 2879 mass hypothesis and it is compared to the standard m_T variable. The usage of this variable

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

also provide a good discriminating power between signal and background, which depends on the particular signal mass hypothesis.

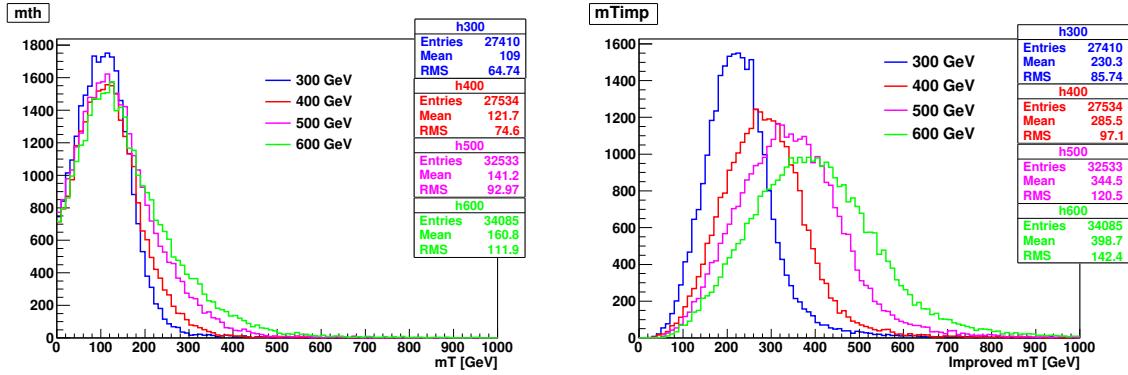


Figure 6.2.: Distribution of the m_T and m_T^i variables at generator level for different resonance mass hypothesis.

The signal extraction is based on a binned maximum likelihood fit using the m_T^i distribution for signal and background contributions as templates. The m_T^i template is defined using the following bin boundaries:

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

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

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

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

Here, $f(E, \Gamma_{SM}, M_X)$ 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 [130] 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.

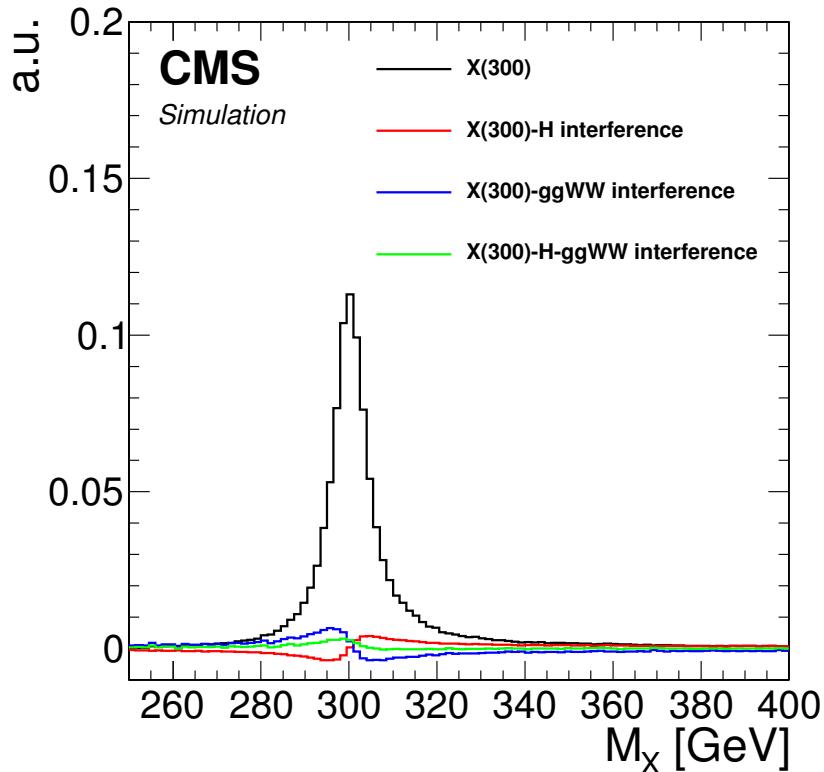


Figure 6.3.: Distribution of the M_X variable for a resonance mass of 300 GeV, showing the various interference terms after the WW baseline selections.

2920 The effect of the resulting interference contribution including all the different terms is
 2921 shown in Fig. 6.4 for the m_T^i signal templates, in the three categories separately and for
 2922 different M_X hypotheses.

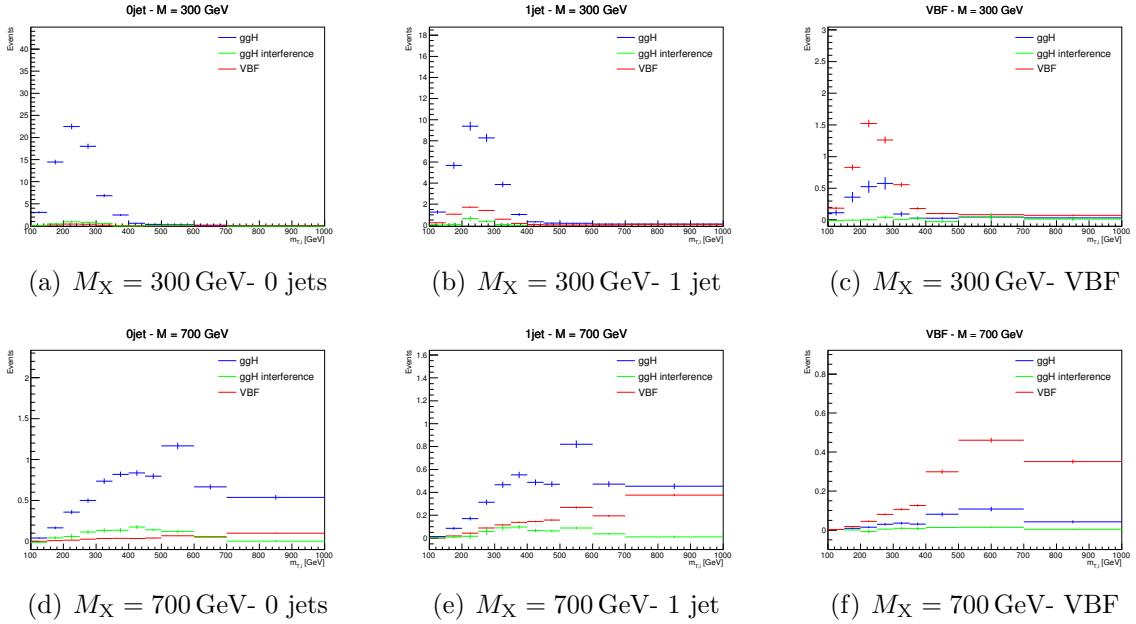


Figure 6.4.: Distributions of the m_T^i variable for $M_X = 300$ and 700 GeV , showing the signal (both the ggH and VBF processes) and the interference contributions in the three jet categories.

2923 The interference contribution is thus not negligible, especially for large values of M_X ,
 2924 and is included in the analysis as part of the signal contribution. More specifically, during
 2925 the fit procedure the signal yield is scaled by the signal strength parameter μ (which is the
 2926 parameter of interest of the fit), while the interference yield is scaled by $\sqrt{\mu}$.

2927 6.4. Background estimation

2928 The background processes affecting the analysis phase space are the same as the ones
 2929 contributing to the SM Higgs search described in Sec. 5.4. The techniques used for the
 2930 background estimation are the same as well.

2931 The most relevant difference is the addition of the 2 jets category. The WW, top
 2932 quark and DY background normalizations are estimated in this category using data driven
 2933 techniques, similarly to the other jet bins.

2934 Given the slightly different WW baseline selection with respect to the SM Higgs search,
 2935 also the control regions for the top quark and DY backgrounds estimation change, while the
 2936 WW background normalization is estimated from data in the three signal regions separately,
 2937 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.

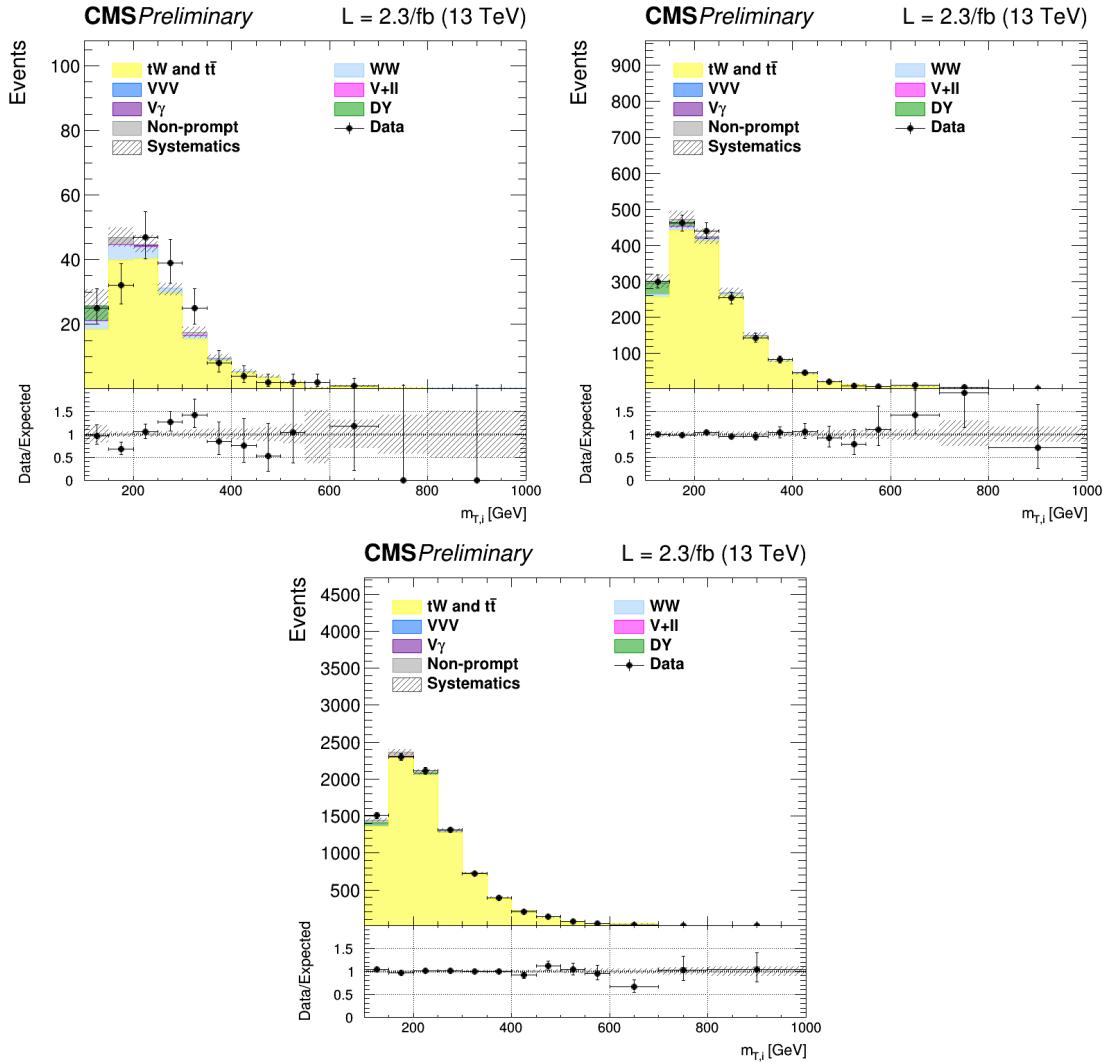


Figure 6.5.: Distributions of $m_{T,j}^i$ for events with 0 jet (top left), 1 jet (top right) and 2 jets (bottom) in top enriched control region. Scale factors estimated from data are not applied in the plots.

2946 The jet induced background, also labelled as “non-prompt” background, so as to highlight
 2947 that these events do not contain prompt leptons, is estimated using the same fake rate
 2948 method described in 4.4.3. A crosscheck is performed selecting events passing the WW
 2949 baseline selection but with an $e\mu$ pair with same charge. The m_T^i distributions for this
 2950 phase space are shown in Fig. 6.6 for the three jet categories separately, showing agreement
 2951 between data and simulation within the uncertainties.

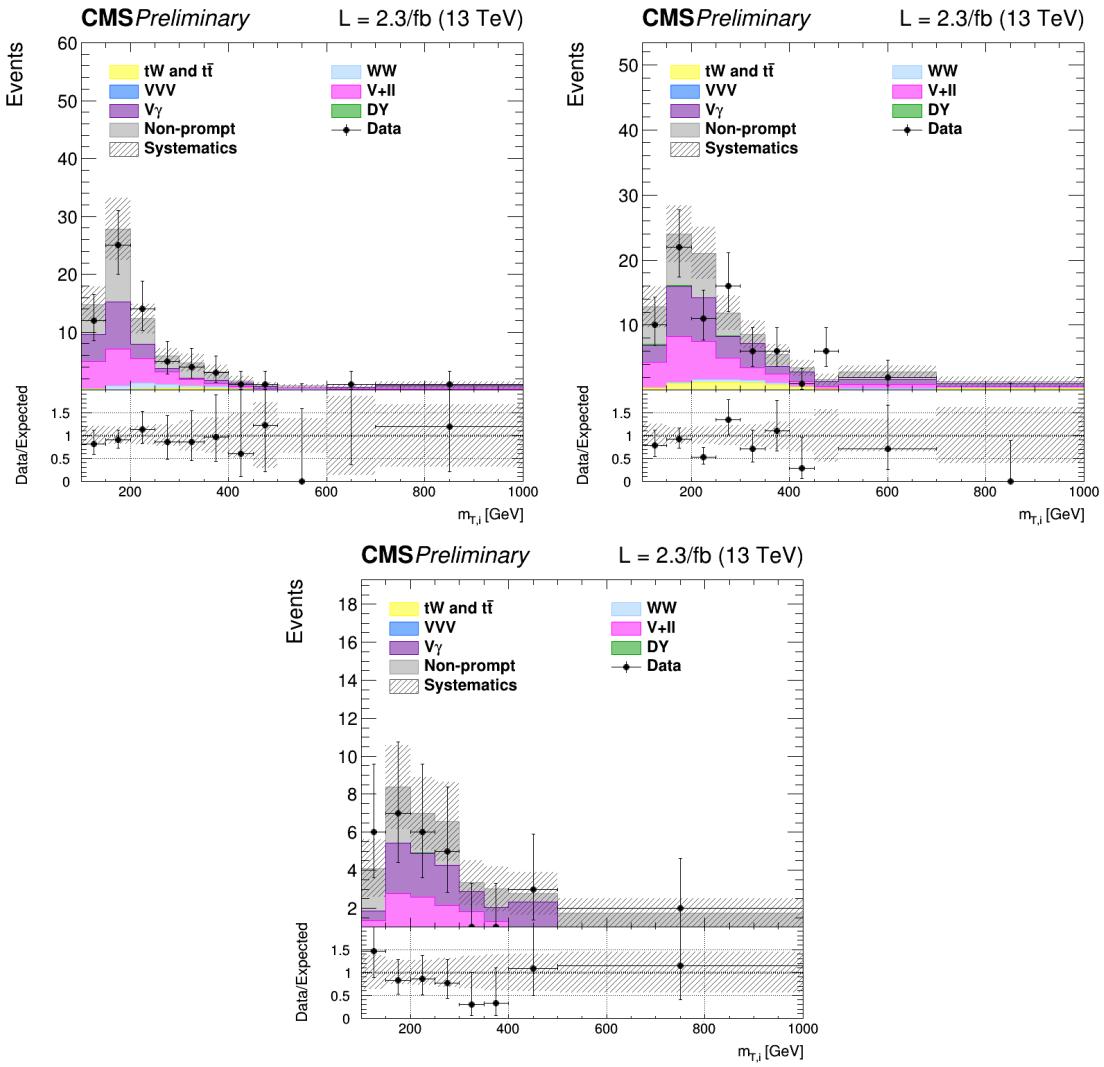


Figure 6.6.: Distributions of m_T^i for events with 0 jet (top left), 1 jet (top right) and 2 jets (bottom) in the same-charge dilepton control region. The last bin of the histograms includes overflows.

2952 Due to the cuts on the leptons p_T and on $m_{\ell\ell}$ in the WW baseline requirements, the
 2953 contribution of the DY background decaying to a pair of τ leptons is very small in the
 2954 signal regions, especially in the VBF phase space. The normalization of this background is
 2955 estimated from a control region in data, defined in the same way as explained in 5.4.4, for

2956 the 0 and 1 jet categories. In the VBF category, given the very small number of expected
2957 events, the normalization of this background is taken from simulation.

2958 Other minor background processes are estimated as described in 5.4.5.

2959 6.5. Systematic uncertainties

2960 The systematic uncertainties affecting this analysis are the same discussed in Sec. 5.5. The
2961 differences with respect to the SM Higgs boson search are described below.

2962 The PDF and α_s uncertainties on the signal cross sections are taken from the computations
2963 performed by the LHC cross section working group [123], and are included for all
2964 the mass points. The value of these uncertainties depends on the resonance mass and vary
2965 from 3 and 5% for ggH and from 2 and 3% for VBF production modes. The PDFs and α_s
2966 uncertainties on the signal selection are evaluated for every resonance mass and are found
2967 to be less than 1% for both ggH and VBF.

2968 The theoretical uncertainties in the signal yields due to the jet categorization are
2969 evaluated for all the ggH signals following the prescription described in Sec. ??.

2970 An additional uncertainty on the modelling of the top pair background is derived from
2971 the observed discrepancy between data and POWHEG V2 plus PYTHIA 8.1 simulation
2972 on the top quark p_T spectrum [131], which is particularly important in the tail of the m_T^i
2973 distribution. Another uncertainty affecting the m_T^i tail for the top quark background is the
2974 parton shower uncertainty. This is evaluated comparing the generator level m_T^i distributions
2975 corresponding to two different simulations of the $t\bar{t}$ process: one obtained using PYTHIA
2976 8.1 for the showering and hadronization of the simulated events, and the other using
2977 HERWIG++. The difference between the two is used to extract a shape uncertainty, which
2978 is less than 1% for low m_T^i values and reaches about 6% in the m_T^i tail.

2979 6.6. Signal extraction and limit setting

2980 The signal yield, including both the ggH and VBF production modes, is extracted performing
2981 a combined fit of the three categories to the m_T^i simulation templates for backgrounds and
2982 signal, and is repeated for each resonance mass hypothesis. Moreover, fixed the mass of
2983 the resonance, the fit is performed again for the various hypotheses of the resonance decay
2984 width. A single signal strength μ is extracted from the fit, which multiplies both the ggH
2985 and VBF contributions. In other words it is assumed that the ratio of the two production
2986 mechanism stays the same as the one predicted by the SM².

2987 The background yields expected from simulation corresponding to the three jet categories
2988 and after the analysis event selection are shown in Table 6.1. The signal yields corresponding
2989 to a selection of mass points and assuming $\Gamma' = \Gamma_{SM}$ are shown in Table 6.2.

²This is an approximation which limits the amount of models that can be tested with the provided results.

A future development of this analysis will also include the cases for which different ggH and VBF relative contributions are expected.

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

Background process	0 jets	1 jet	VBF
qq \rightarrow WW	501.93 \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 γ	26.62 \pm 0.72 (3%)	14.18 \pm 0.38 (3%)	0.64 \pm 0.02 (3%)
V γ^*	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

2990 The strategy for computing the exclusion limits is based on the modified frequentist
 2991 approach, also referred to as CL_s, as described in [101]. The first step is to construct the
 2992 likelihood function $\mathcal{L}(\text{data}|\mu, \theta)$:

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

2993 where data represents the experimental observation, s and b are the expected signal and
 2994 background yields respectively and θ is the full set of nuisance parameters constrained by
 2995 the prior distribution functions $p(\tilde{\theta}|\theta)$. The default values for the nuisance parameters are
 2996 labelled as $\tilde{\theta}$.

2997 For a binned shape analysis, $\text{Poisson}(\text{data}|\mu \cdot s + b)$ is the product of the Poisson
 2998 probabilities to observe n_i events in bin i:

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

2999 In order to test the compatibility of the data with the signal plus background (or the
 3000 background only) hypothesis, the test statistic \tilde{q}_μ is constructed based on the profile

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

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

3001 likelihood reatio:

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

3002 where $\hat{\theta}_\mu$ refers to the conditional maximum likelihood estimators of θ , given the signal
3003 strength μ . The parameter estimators $\hat{\mu}$ and $\hat{\theta}$ correspond to the global maximum of the
3004 likelihood. The $0 \leq \hat{\mu}$ constraint is imposed to have a positive signal yield, e.g. background
3005 underfluctuations are forbidden, while $\hat{\mu} \leq \mu$ is imposed to have a one-sided confidence
3006 interval. The observed test statistic for the signal strength μ under test is referred to as
3007 $\tilde{q}_\mu^{\text{obs}}$. The values of the nuisance parameters obtained maximising the likelihood function are
3008 labelled as $\hat{\theta}_0^{\text{obs}}$ and $\hat{\theta}_\mu^{\text{obs}}$ for the background only and signal plus background hypotheses,
3009 respectively. The pdf of the test statistic is constructed by generating toy MC pseudo-data
3010 for both the background only and signal plus background hypotheses, i.e. $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}})$ and
3011 $f(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}})$. These distributions can be used to define two p-values corresponding to the

3012 two hypotheses, p_μ and p_b :

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

3013

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

3014 According to these definitions, p_μ and p_b can be identified with CL_{s+b} and $1 - \text{CL}_b$, 3015 respectively. The $\text{CL}_s(\mu)$ is calculated using the following ratio:

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

3016 If, for a given signal strength μ , $\text{CL}_s \leq \alpha$, then the hypothesis is excluded with a $(1 - \alpha)$ 3017 confidence level (CL). For instance, if one wants to quote the upper limit on μ with a 95% 3018 CL, the signal strength has to be adjusted until $\text{CL}_s = 0.05$.

3019 The expected median upper limit, as well as the $\pm 1\sigma$ (68% CL) and $\pm 2\sigma$ (95% CL) 3020 bands, are determined generating a large amount of pseudo-data in the background only 3021 hypothesis and calculating CL_s and the 95% CL upper limit for each of them, as if they 3022 were real data. Then the cumulative distribution of the 95% CL upper limits is built and 3023 the median expected value is identified as the value at which the cumulative distribution 3024 crosses the 50% quantile. The $\pm 1\sigma$ ($\pm 2\sigma$) band is defined by the values at which the 3025 cumulative distribution crosses the 16% (2.5%) and 84% (97.5%) quantiles.

3026 In order to assess the sensitivity of the analysis, the expected upper exclusion limits at 3027 95% CL on the signal strength are shown in Fig. 6.7 for the three jet categories separately. 3028 For a given mass of the resonance, the limits are derived assuming a signal decay width 3029 $\Gamma' = \Gamma_{\text{SM}}$ and a cross section equal to the one expected from a SM Higgs boson at that 3030 mass. The other decay width hypothesis have also been tested, showing a very similar 3031 expected exclusion limit, suggesting that this analysis is not strongly sensitive to variations 3032 of the resonance decay width.

3033 The 0 jets category is the most sensitive especially in the low mass region, while for 3034 very large masses of the resonance the 1 jet and VBF categories start being important. 3035 This is explained mainly by the fact that the VBF contribution increases, with respect to 3036 ggH, as the mass increases. The expected exclusion limit on the signal strength after the 3037 combination of the three categories is shown in Fig. 6.8. Comparing the limits in the single 3038 categories with the combination of the three it is evident how the higher jet multiplicity 3039 categories help in improving the results for large values of M_X .

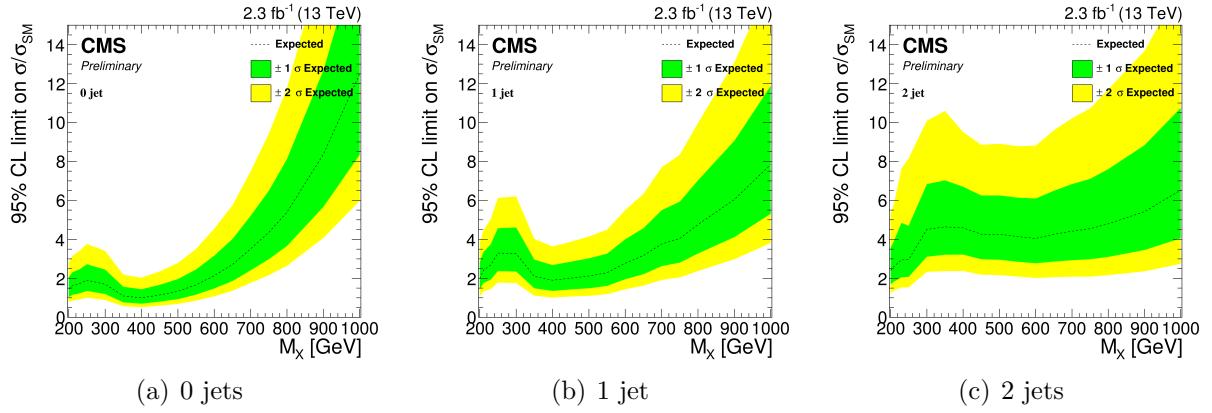


Figure 6.7.: Expected exclusion upper limits at 95% CL on the signal strength in the three categories, as a function of the resonance mass. The dashed line corresponds to median upper limit, while the green and yellow regions represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands, respectively.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

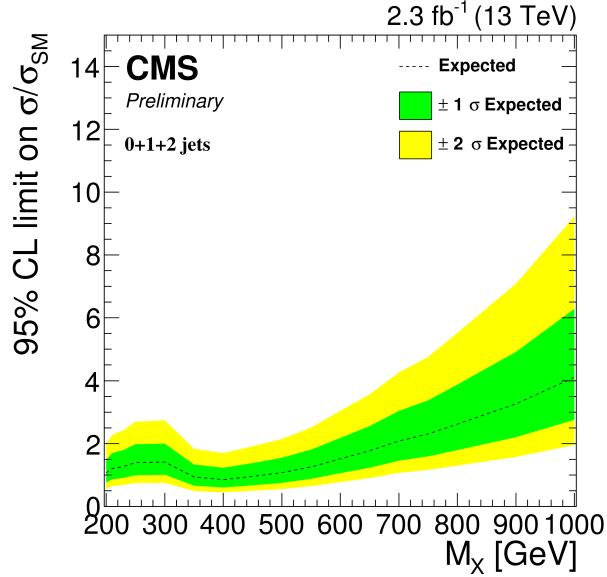


Figure 6.8.: Expected exclusion upper limit at 95% CL on the signal strength for the combination of the three categories, as a function of the resonance mass. The dashed line corresponds to median upper limit, while the green and yellow regions represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty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

6.7. Results

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

3043 to $M_X = 400$ GeV and $M_X = 800$ GeV are shown superimposed on the background for
3044 comparison.

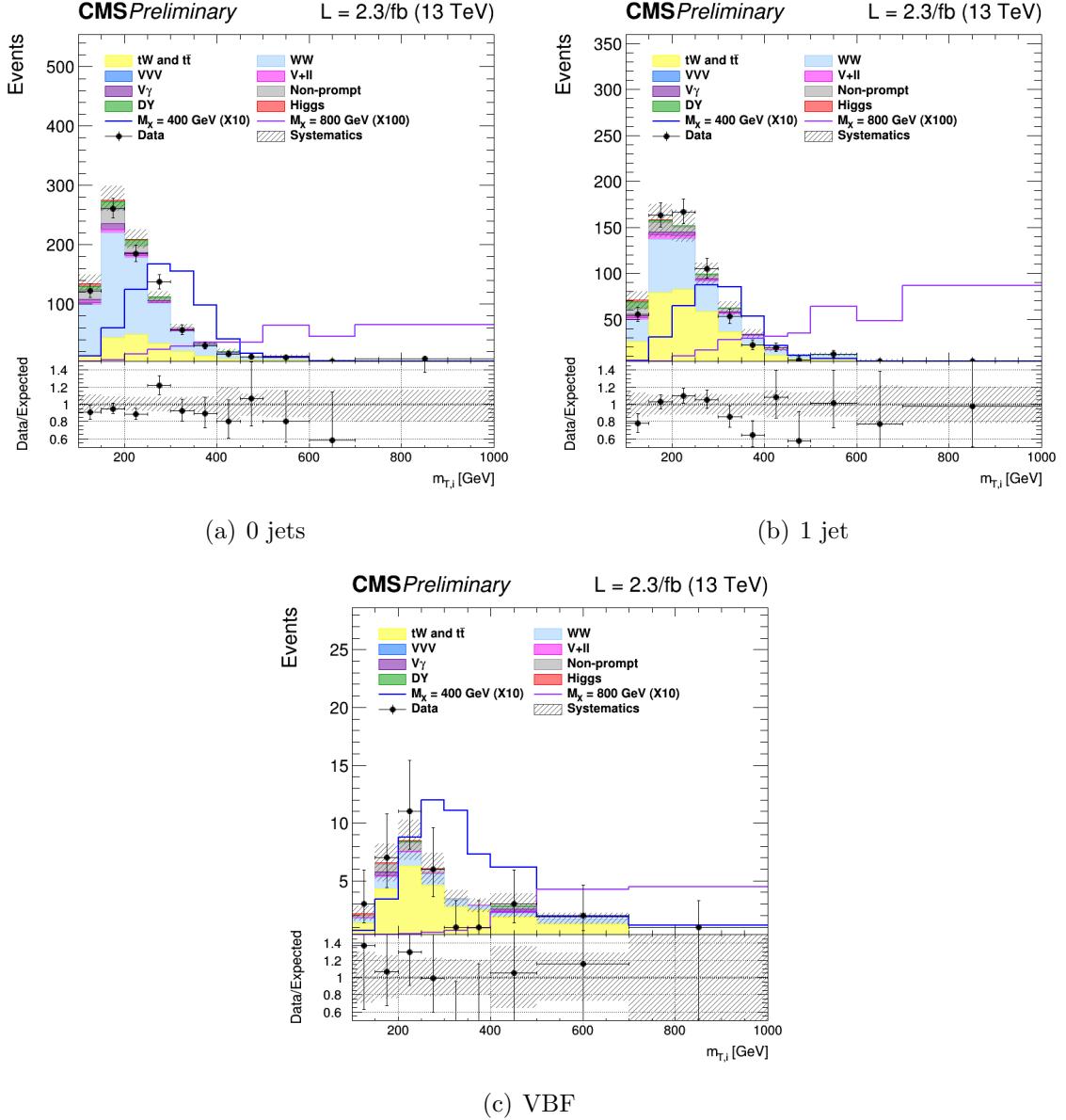


Figure 6.9.: Distributions of m_T^i in the signal region for the 0 jets, 1 jet and VBF categories. Background normalisations correspond to the pre-fit value. Signal contributions for two mass hypotheses, $M_X = 400$ GeV and $M_X = 800$ GeV, are shown superimposed on the background and scaled to facilitate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

3045 For every mass point from 200 GeV up to 1 TeV the observed p-value and the 95% CL
3046 upper exclusion limit are calculated for five hypothesis of the signal width. The observed

³⁰⁴⁷ p-value as a function of the resonance mass for the combination of the three jet categories
³⁰⁴⁸ 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)

³⁰⁴⁹ In order to be independent on the particular model assumed for the signal cross section,
³⁰⁵⁰ the results are interpreted as exclusion limits on $\sigma \times \mathcal{B}$, where σ stands for the sum of
³⁰⁵¹ the ggH and VBF cross sections, and \mathcal{B} represents the $X \rightarrow WW \rightarrow 2\ell 2\nu$ branching ratio
³⁰⁵² including all lepton flavours. The expected and observed upper exclusion limits on $\sigma \times \mathcal{B}$
³⁰⁵³ for $\Gamma' = \Gamma_{SM}$ are shown in Fig. 6.10.

³⁰⁵⁴ A mild excess is observed in the 0 jets category and, more evident, in the 1 jet category
³⁰⁵⁵ around 250-300 GeV. A deficit is instead in the VBF category around 250 GeV, which is
³⁰⁵⁶ mainly due to an underfluctuation of the background. This effect can be understood looking
³⁰⁵⁷ at the VBF shape in Fig. 6.9, where two adjacent data points, corresponding to the fifth

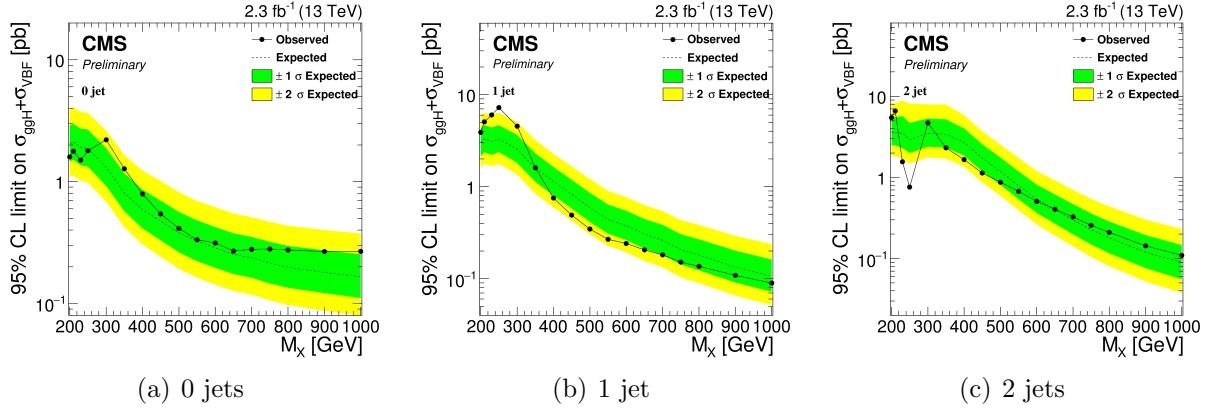


Figure 6.10.: Expected and observed exclusion upper limits at 95% CL on $\sigma \times \mathcal{B}$ in the three categories, as a function of the resonance mass. The dashed line corresponds to median upper limit, while the green and yellow regions represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands, respectively. The dotted line represents the observed limit. Limits are derived assuming $\Gamma' = \Gamma_{SM}$ 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 Γ' hypotheses discussed before. From the combined exclusion limits no significant evidence of a deviation from the SM prediction is observed. The presence of a scalar resonance with $\sigma \times \mathcal{B}$ higher than the values reported in Fig. 6.11 is thus excluded with a 95% CL for masses ranging from 200 GeV up to 1 TeV.

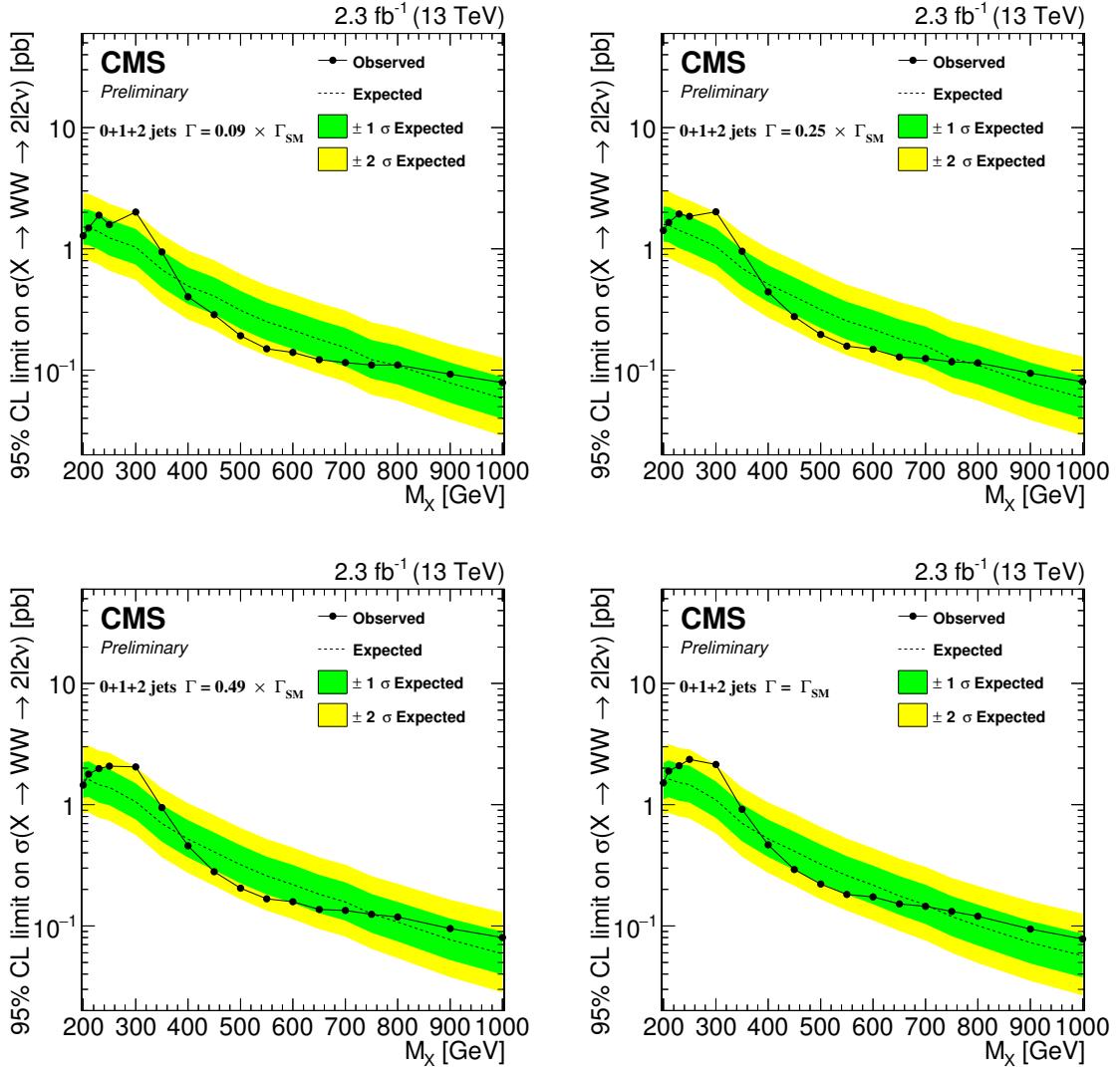


Figure 6.11.: Expected and observed exclusion limits at 95% CL on $\sigma \times \mathcal{B}$ for the combination of the three jet categories as a function of the resonance mass. The black dotted line corresponds to the observed value while the yellow and green bands represent the $\pm 1\sigma$ and $\pm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

Chapter 7.

Conclusions

Appendix A.

3066 Fiducial region definition and 3067 optimization

3068 The fiducial region must be chosen in such a way to be as close as possible to the selections
3069 applied in the analysis, in order to reduce the model dependence in the extrapolation step.
3070 That means that for optimizing the fiducial volume definition, the efficiency has to be
3071 maximized. Another parameter entering the game is the number of fake events, in other
3072 words the number of reconstructed events which do not belong to the fiducial phase space.
3073 This parameter should instead be as small as possible. Even if we have to observe the trend
3074 of these two quantities as a function of p_T^H , we can maximize the ratio between the overall
3075 efficiency and the overall fake rate as a proxy for establishing the “goodness” of the fiducial
3076 region.

3077 Several different fiducial region definitions were tested and the results show that:

- 3078 • **of cut:** The fiducial region definition must include only the opposite flavor combination
3079 including one electron and one muon. If we include also the combinations involving
3080 τ 's the efficiency falls down.
- 3081 • **Lepton cut:** Since the resolution on lepton transverse momentum is good, there is no
3082 need to loosen the cuts related these variables, i.e. we can use the same cuts defined
3083 in the analysis selection ($p_T^{\ell,1} > 20 \text{ GeV}$, $p_T^{\ell,2} > 10 \text{ GeV}$).
- 3084 • **Di-lepton p_T cut:** As stated in the previous point, there is no need to loosen this
3085 cut, so we kept the same value as the analysis selection, i.e. $p_T^{\ell\ell} > 30 \text{ GeV}$.
- 3086 • **Di-lepton mass cut:** $m_{\ell\ell} > 12 \text{ GeV}$ as discussed before.
- 3087 • **neutrino pair p_T cut:** Since the resolution on the measurement of the missing
3088 transverse energy is poor, the neutrino pair cut should not be included in the definition
3089 of the fiducial region, because it would increase the fake rate without increasing the
3090 efficiency, thus resulting in a lower ratio between overall efficiency and fake rate.
- 3091 • **m_T cut:** Also the m_T cut that we have in the analysis selection, i.e. $m_T > 60 \text{ GeV}$,
3092 should be loosened or removed because it involves neutrinos and then increase the
3093 fake rate. We decided eventually to keep this cut, loosening it to 50 GeV, because in
3094 addition to increase the number of fake events, it increases the efficiency as well.

3095 The fake rate and the efficiency as a function of p_T^H after the optimization discussed
3096 before are shown in figure A.1. To obtain these plots the fiducial region was modified

3097 adding in sequence the various cuts and computing the efficiency and the fake rate each
 3098 time. In that way we can asses the composition of those distributions.

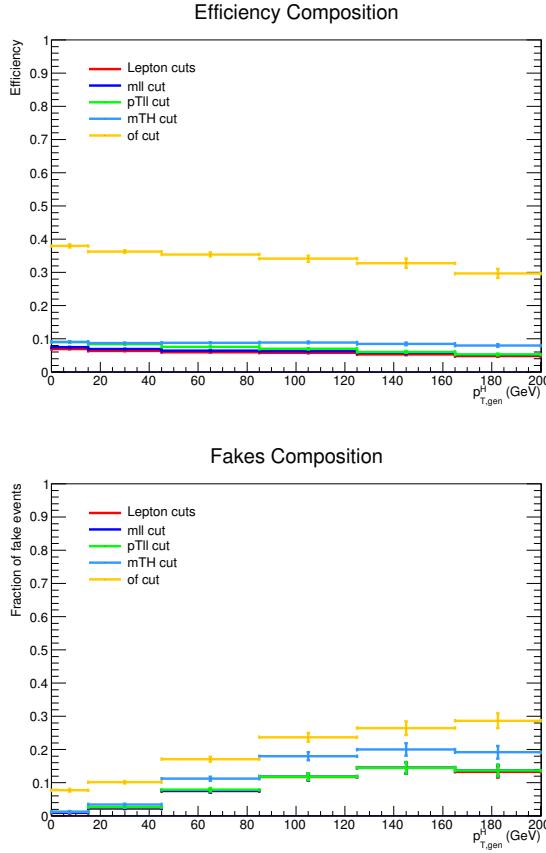


Figure A.1.: Efficiency and fake rate as a function on Higgs transverse momentum. The plots correspond to the optimized fiducial region definition and show the effect of adding each of the mentioned cuts in sequence.

3099 The efficiency and fraction of fake events have been measured also as a function of the
 3100 E_T^{miss} and m_T cuts in the fiducial region. Since these two variables are correlated, the
 3101 results are reported as two-dimensional histograms. In Fig. A.2 are reported the efficiency
 3102 and fraction of fake events for these two variables.

3103 The criterion adopted to define the fiducial region is a tradeoff between having a large
 3104 efficiency and a small fraction of fake events. Especially when looking at the low resolution
 3105 variables, such as E_T^{miss} and m_T , a suitable figure of merit has to be chosen for the estimation
 3106 of the best cuts. Several different figures of merit have been checked, such as ϵ/f , $\epsilon - f$
 3107 and $(1 - f)/\epsilon$. The results for these three different figures of merit are shown in Fig. A.3 as
 3108 a function of the E_T^{miss} and m_T cuts in the fiducial region.

3109 Following the same criterion, similar plots as above have been obtained for an alternative
 3110 model, given by varying up the ggH/VBF ratio within the experimental uncertainties. The

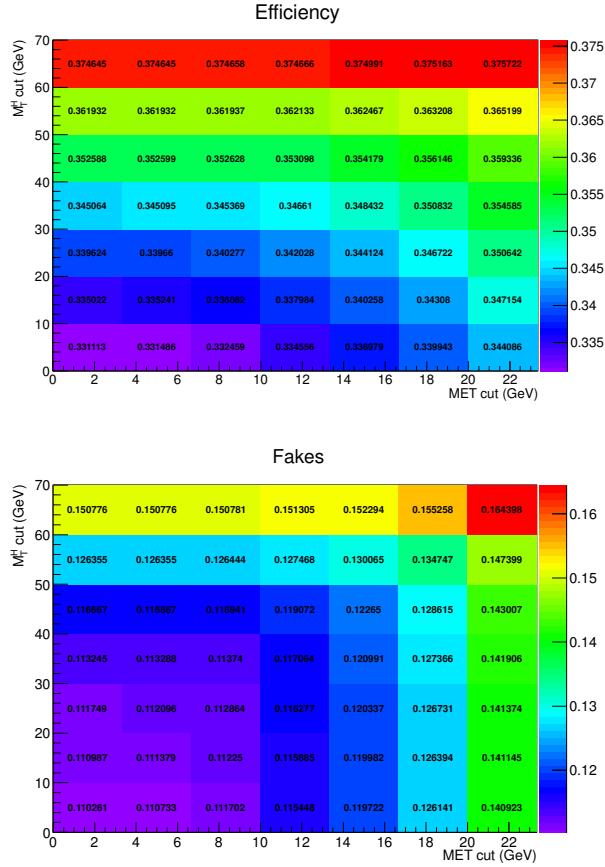


Figure A.2.: Efficiency and fake rate as a function of E_T^{miss} and m_T cuts in the fiducial region.

3111 results, shown in Fig. A.4 and Fig. A.5, show a similar trend with respect to the model
3112 with nominal ggH/VBF ratio.

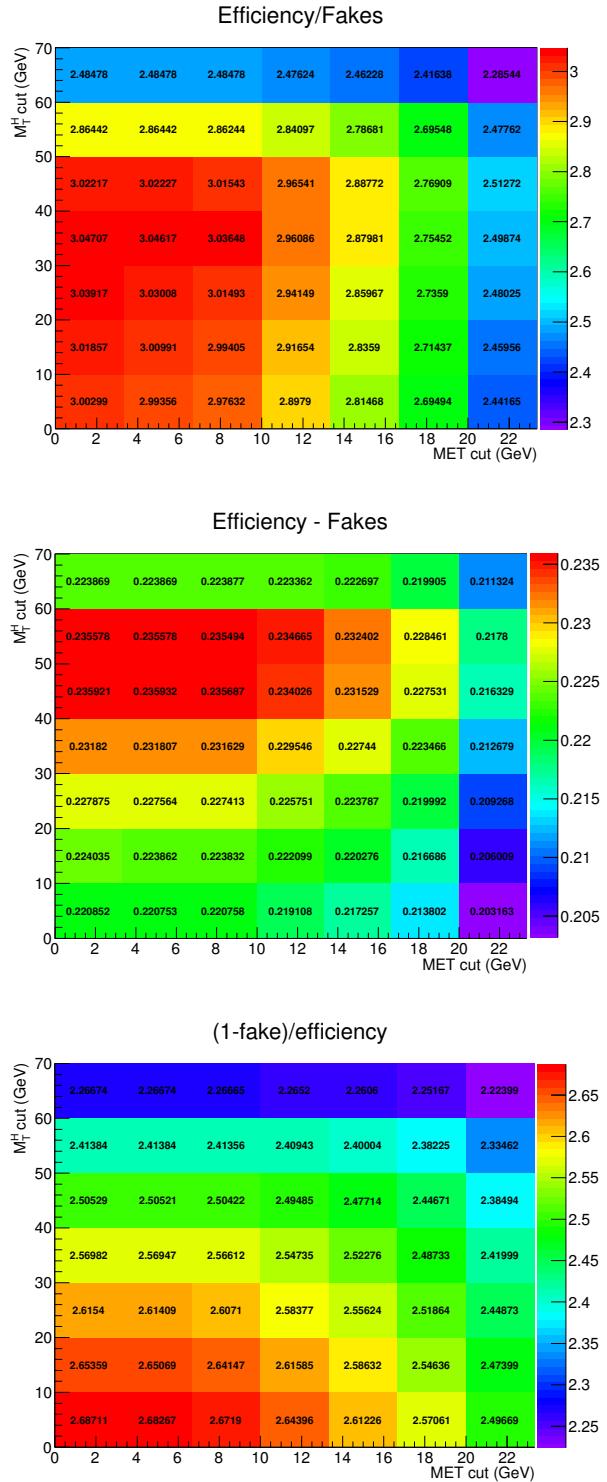


Figure A.3.: Different figures of merit as a function of E_T^{miss} and m_T cuts in the fiducial region.

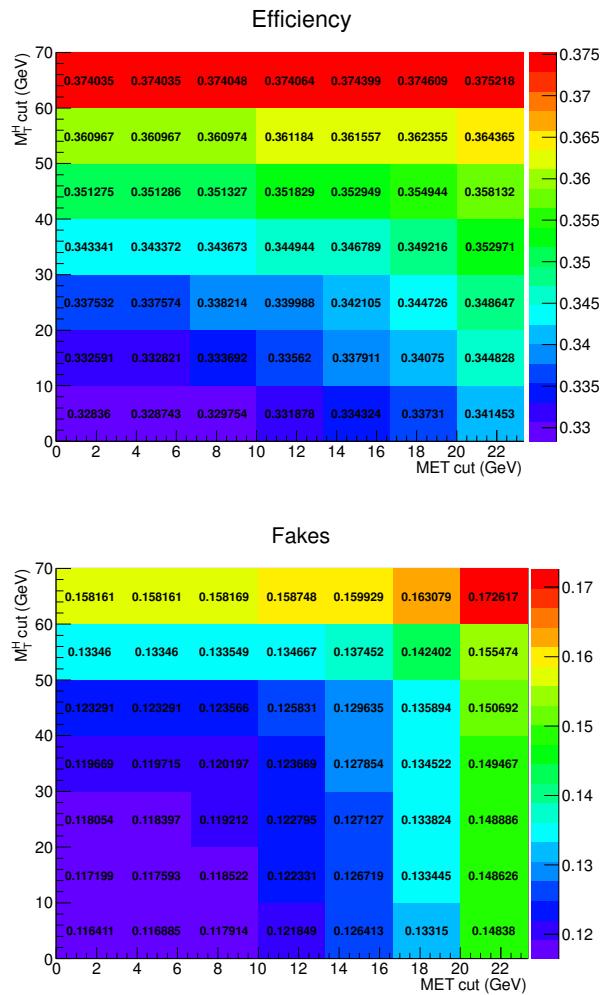


Figure A.4.: Efficiency and fake rate as a function of E_T^{miss} and m_T cuts in the fiducial region, for the alternative model with an up variation of the ggH/VBF ratio.

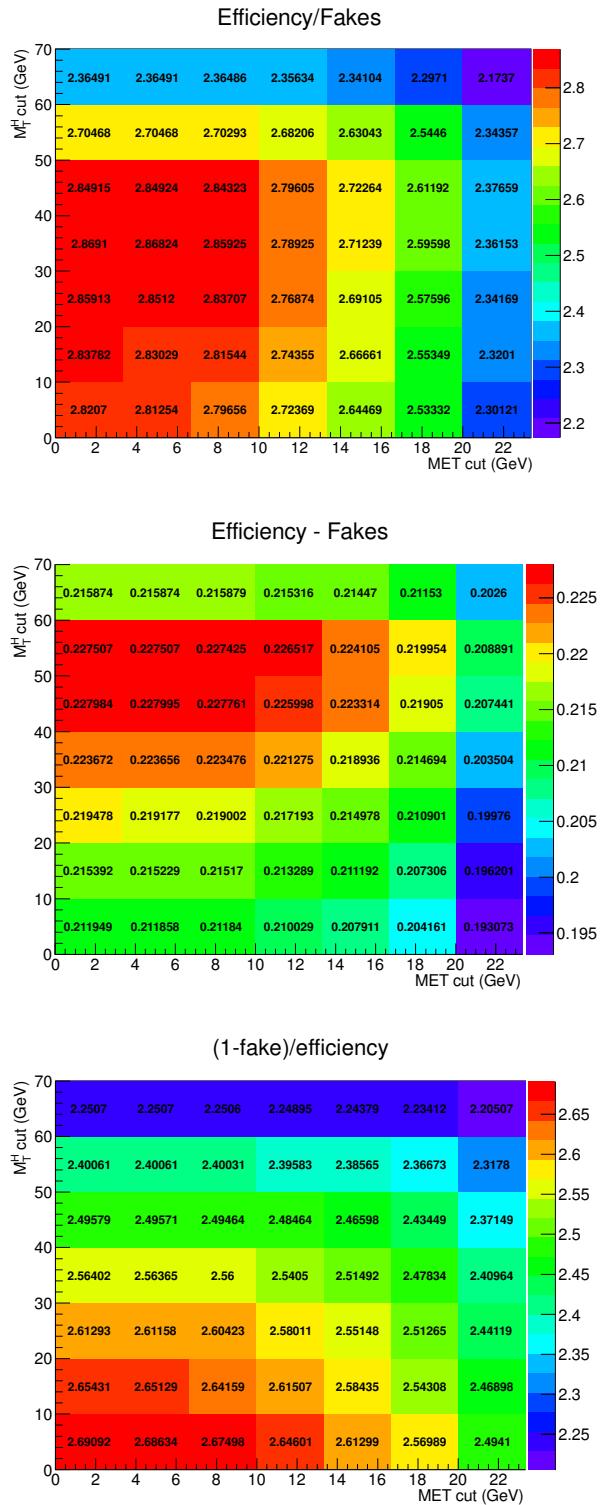


Figure A.5.: Different figures of merit as a function of E_T^{miss} and m_T cuts in the fiducial region, for the alternative model with an up variation of the ggH/VBF ratio.

₃₁₁₄

Bibliography

- ₃₁₁₅ [1] F. Halzen and Alan D. Martin. *QUARKS AND LEPTONS: AN INTRODUCTORY COURSE IN MODERN PARTICLE PHYSICS*. 1984. ISBN: 0471887412, 9780471887416.
- ₃₁₁₈ [2] ATLAS and CMS Collaborations. “Combined measurement of the Higgs boson mass in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS experiments”. In: *Phys. Rev. Lett.* 114 (2015), p. 191803. DOI: 10.1103/PhysRevLett.114.191803. arXiv: 1503.07589 [hep-ex].
- ₃₁₂₂ [3] 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].
- ₃₁₂₅ [4] 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].
- ₃₁₂₈ [5] Nathaniel Craig, Jamison Galloway, and Scott Thomas. “Searching for Signs of the Second Higgs Doublet”. In: (2013). arXiv: 1305.2424 [hep-ph].
- ₃₁₃₀ [6] R. Keith Ellis, W. James Stirling, and B. R. Webber. “QCD and collider physics”. In: *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* 8 (1996), pp. 1–435.
- ₃₁₃₂ [7] Guido Altarelli. “Collider Physics within the Standard Model: a Primer”. In: (2013). arXiv: 1303.2842 [hep-ph].
- ₃₁₃₄ [8] John C. Collins, Davison E. Soper, and George F. Sterman. “Factorization of Hard Processes in QCD”. In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313 [hep-ph].
- ₃₁₃₇ [9] Jonathan M. Butterworth, Guenther Dissertori, and Gavin P. Salam. “Hard Processes in Proton-Proton Collisions at the Large Hadron Collider”. In: *Ann. Rev. Nucl. Part. Sci.* 62 (2012), pp. 387–405. DOI: 10.1146/annurev-nucl-102711-094913. arXiv: 1202.0583 [hep-ex].
- ₃₁₄₁ [10] Guido Altarelli and G. Parisi. “Asymptotic Freedom in Parton Language”. In: *Nucl. Phys.* B126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4.
- ₃₁₄₃ [11] Pavel M. Nadolsky et al. “Implications of CTEQ global analysis for collider observables”. In: *Phys. Rev.* D78 (2008), p. 013004. DOI: 10.1103/PhysRevD.78.013004. arXiv: 0802.0007 [hep-ph].

- 3146 [12] G. Watt. “Parton distribution function dependence of benchmark Standard Model
3147 total cross sections at the 7 TeV LHC”. In: *JHEP* 09 (2011), p. 069. DOI: 10.1007/
3148 JHEP09(2011)069. arXiv: 1106.5788 [hep-ph].
- 3149 [13] Richard D. Ball et al. “Parton distributions with QED corrections”. In: *Nucl. Phys.*
3150 B877 (2013), pp. 290–320. DOI: 10.1016/j.nuclphysb.2013.10.010. arXiv:
3151 1308.0598 [hep-ph].
- 3152 [14] Thomas Sven Pettersson and P Lefèvre. *The Large Hadron Collider: conceptual*
3153 *design*. Tech. rep. CERN-AC-95-05-LHC. Oct. 1995. URL: <http://cds.cern.ch/record/291782>.
- 3155 [15] Oliver Sim Brüning et al. *LHC Design Report*. Geneva: CERN, 2004. URL: <http://cds.cern.ch/record/782076>.
- 3157 [16] Oliver Sim Brüning et al. *LHC Design Report*. Geneva: CERN, 2004. URL: <http://cds.cern.ch/record/815187>.
- 3159 [17] Michael Benedikt et al. *LHC Design Report*. Geneva: CERN, 2004. URL: <http://cds.cern.ch/record/823808>.
- 3161 [18] *LEP design report*. By the LEP Injector Study Group. Geneva: CERN, 1983. URL:
3162 <http://cds.cern.ch/record/98881>.
- 3163 [19] *LEP design report*. Copies shelved as reports in LEP, PS and SPS libraries. Geneva:
3164 CERN, 1984. URL: <http://cds.cern.ch/record/102083>.
- 3165 [20] Carlo Wyss. *LEP design report, v.3: LEP2*. Vol. 1-2 publ. in 1983-84. Geneva: CERN,
3166 1996. URL: <http://cds.cern.ch/record/314187>.
- 3167 [21] S. Chatrchyan et al. “The CMS experiment at the CERN LHC”. In: *JINST* 3 (2008),
3168 S08004. DOI: 10.1088/1748-0221/3/08/S08004.
- 3169 [22] G. Aad et al. “The ATLAS Experiment at the CERN Large Hadron Collider”. In:
3170 *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.
- 3171 [23] A. Augusto Alves Jr. et al. “The LHCb Detector at the LHC”. In: *JINST* 3 (2008),
3172 S08005. DOI: 10.1088/1748-0221/3/08/S08005.
- 3173 [24] K. Aamodt et al. “The ALICE experiment at the CERN LHC”. In: *JINST* 3 (2008),
3174 S08002. DOI: 10.1088/1748-0221/3/08/S08002.
- 3175 [25] G. Anelli et al. “The TOTEM experiment at the CERN Large Hadron Collider”. In:
3176 *JINST* 3 (2008), S08007. DOI: 10.1088/1748-0221/3/08/S08007.
- 3177 [26] O. Adriani et al. “The LHCf detector at the CERN Large Hadron Collider”. In:
3178 *JINST* 3 (2008), S08006. DOI: 10.1088/1748-0221/3/08/S08006.
- 3179 [27] *The CMS magnet project: Technical Design Report*. Technical Design Report CMS.
3180 Geneva: CERN, 1997. URL: <http://cds.cern.ch/record/331056>.
- 3181 [28] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical
3182 Design Report CMS. Geneva: CERN, 1997. URL: <http://cds.cern.ch/record/349375>.

- 3184 [29] Philippe Bloch et al. *Changes to CMS ECAL electronics: addendum to the Technical*
3185 *Design Report*. Technical Design Report CMS. Geneva: CERN, 2002. URL: <http://cds.cern.ch/record/581342>.
- 3187 [30] *The CMS hadron calorimeter project: Technical Design Report*. Technical Design
3188 Report CMS. Geneva: CERN, 1997. URL: <http://cds.cern.ch/record/357153>.
- 3189 [31] *The CMS muon project: Technical Design Report*. Technical Design Report CMS.
3190 Geneva: CERN, 1997. URL: <http://cds.cern.ch/record/343814>.
- 3191 [32] S. Dasu et al. “CMS. The TriDAS project. Technical design report, vol. 1: The
3192 trigger systems”. In: (2000).
- 3193 [33] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical*
3194 *Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and*
3195 *data-acquisition project*. Technical Design Report CMS. Geneva: CERN, 2002.
3196 URL: <http://cds.cern.ch/record/578006>.
- 3197 [34] CMS Collaboration. *Particle-Flow Event Reconstruction in CMS and Performance*
3198 *for Jets, Taus, and E_T^{miss}* . CMS Physics Analysis Summary CMS-PAS-PFT-09-001.
3199 2009. URL: <http://cdsweb.cern.ch/record/1194487>.
- 3200 [35] Serguei Chatrchyan et al. “Description and performance of track and primary-
3201 vertex reconstruction with the CMS tracker”. In: *JINST* 9.10 (2014), P10009. DOI:
3202 [10.1088/1748-0221/9/10/P10009](https://doi.org/10.1088/1748-0221/9/10/P10009). arXiv: 1405.6569 [physics.ins-det].
- 3203 [36] Pierre Billoir and S. Qian. “Simultaneous pattern recognition and track fitting by
3204 the Kalman filtering method”. In: *Nucl. Instrum. Meth. A* 294 (1990), p. 219. DOI:
3205 [10.1016/0168-9002\(90\)91835-Y](https://doi.org/10.1016/0168-9002(90)91835-Y).
- 3206 [37] Wolfgang Adam et al. *Reconstruction of Electrons with the Gaussian-Sum Filter*
3207 *in the CMS Tracker at the LHC*. Tech. rep. CMS-NOTE-2005-001. Geneva: CERN,
3208 Jan. 2005. URL: <https://cds.cern.ch/record/815410>.
- 3209 [38] Serguei Chatrchyan et al. “Performance of CMS muon reconstruction in pp collision
3210 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 [physics.ins-det].
- 3212 [39] *Electron reconstruction and identification at $\sqrt{s} = 7$ TeV*. Tech. rep. CMS-PAS-
3213 EGM-10-004. Geneva: CERN, 2010. URL: <http://cds.cern.ch/record/1299116>.
- 3214 [40] Vardan Khachatryan et al. “Performance of Electron Reconstruction and Selection
3215 with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV”. In: *JINST*
3216 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
3217 [physics.ins-det].
- 3218 [41] Gavin P. Salam. “Towards Jetography”. In: *Eur. Phys. J.* C67 (2010), pp. 637–686.
3219 DOI: [10.1140/epjc/s10052-010-1314-6](https://doi.org/10.1140/epjc/s10052-010-1314-6). arXiv: 0906.1833 [hep-ph].

- 3220 [42] Gerald C. Blazey et al. “Run II jet physics”. In: *QCD and weak boson physics in*
3221 *Run II. Proceedings, Batavia, USA, March 4-6, June 3-4, November 4-6, 1999*. 2000,
3222 pp. 47–77. arXiv: hep-ex/0005012 [hep-ex]. URL: http://lss.fnal.gov/cgi-bin/find_paper.pl?conf=00-092.
- 3223
3224 [43] Serguei Chatrchyan et al. “Determination of jet energy calibration and transverse
3225 momentum resolution in CMS”. In: *JINST* 6 (2011), P11002. DOI: 10.1088/1748-
3226 0221/6/11/P11002. arXiv: 1107.4277 [physics.ins-det].
- 3227 [44] 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-
3228 9002(98)01368-0. arXiv: hep-ex/9805009 [hep-ex].
- 3229
3230 [45] Wolfgang Waltenberger. *Adaptive Vertex Reconstruction*. Tech. rep. CMS-NOTE-
3231 2008-033. Geneva: CERN, July 2008. URL: <https://cds.cern.ch/record/1166320>.
- 3232 [46] *MET performance in 8 TeV data*. Tech. rep. CMS-PAS-JME-12-002. Geneva: CERN,
3233 2013. URL: <http://cds.cern.ch/record/1543527>.
- 3234 [47] *Performance of missing energy reconstruction in 13 TeV pp collision data using*
3235 *the CMS detector*. Tech. rep. CMS-PAS-JME-16-004. Geneva: CERN, 2016. URL:
3236 <http://cds.cern.ch/record/2205284>.
- 3237 [48] M. Spira et al. “Higgs boson production at the LHC”. In: *Nucl. Phys.* B 453
3238 (1995), pp. 17–82. DOI: 10.1016/0550-3213(95)00379-7. arXiv: hep-ph/9504378
3239 [hep-ph].
- 3240 [49] Robert Harlander and Philipp Kant. “Higgs production and decay: Analytic results
3241 at next-to-leading order QCD”. In: *JHEP* 12 (2005), p. 015. DOI: 10.1088/1126-
3242 6708/2005/12/015. arXiv: hep-ph/0509189 [hep-ph].
- 3243 [50] V. Ravindran, J. Smith, and W. L. van Neerven. “NNLO corrections to the total
3244 cross-section for Higgs boson production in hadron hadron collisions”. In: *Nucl.*
3245 *Phys.* B 665 (2003), pp. 325–366. DOI: 10.1016/S0550-3213(03)00457-7. arXiv:
3246 [hep-ph/0302135](http://arxiv.org/abs/hep-ph/0302135) [hep-ph].
- 3247 [51] Stefano Catani and Massimiliano Grazzini. “An NNLO subtraction formalism in
3248 hadron collisions and its application to Higgs boson production at the LHC”. In: *Phys.*
3249 *Rev. Lett.* 98 (2007), p. 222002. DOI: 10.1103/PhysRevLett.98.222002.
3250 arXiv: [hep-ph/0703012](http://arxiv.org/abs/hep-ph/0703012) [hep-ph].
- 3251 [52] Charalampos Anastasiou et al. “Higgs Boson Gluon-Fusion Production in QCD at
3252 Three Loops”. In: *Phys. Rev. Lett.* 114 (2015), p. 212001. DOI: 10.1103/PhysRevLett.
3253 114.212001. arXiv: 1503.06056 [hep-ph].
- 3254 [53] X. Chen et al. “NNLO QCD corrections to Higgs boson production at large transverse
3255 momentum”. In: (2016). arXiv: 1607.08817 [hep-ph].
- 3256 [54] Giuseppe Bozzi et al. “Transverse-momentum resummation and the spectrum of the
3257 Higgs boson at the LHC”. In: *Nucl. Phys. B* 737 (2006), p. 73. DOI: 10.1016/j.
3258 nuclphysb.2005.12.022. arXiv: hep-ph/0508068 [hep-ph].

- 3259 [55] D. de Florian et al. “Higgs boson production at the LHC: transverse momentum
3260 resummation effects in the $H \rightarrow 2\gamma$, $H \rightarrow WW \rightarrow l\nu l\nu$ and $H \rightarrow ZZ \rightarrow 4l$ decay
3261 modes”. In: *JHEP* 06 (2012), p. 132. DOI: 10.1007/JHEP06(2012)132. arXiv:
3262 1203.6321 [hep-ph].
- 3263 [56] Massimiliano Grazzini and Hayk Sargsyan. “Heavy-quark mass effects in Higgs boson
3264 production at the LHC”. In: *JHEP* 09 (2013), p. 129. DOI: 10.1007/JHEP09(2013)
3265 129. arXiv: 1306.4581 [hep-ph].
- 3266 [57] Aleksandr Azatov and Ayan Paul. “Probing Higgs couplings with high p_T Higgs
3267 production”. In: *JHEP* 01 (2014), p. 014. DOI: 10.1007/JHEP01(2014)014. arXiv:
3268 1309.5273 [hep-ph].
- 3269 [58] Robert V. Harlander and Tobias Neumann. “Probing the nature of the Higgs-gluon
3270 coupling”. In: *Phys. Rev. D* 88 (2013), p. 074015. DOI: 10.1103/PhysRevD.88.
3271 074015. arXiv: 1308.2225 [hep-ph].
- 3272 [59] David Marzocca, Marco Serone, and Jing Shu. “General Composite Higgs Models”.
3273 In: *JHEP* 08 (2012), p. 013. DOI: 10.1007/JHEP08(2012)013. arXiv: 1205.0770
3274 [hep-ph].
- 3275 [60] Andrea Banfi, Adam Martin, and Veronica Sanz. “Probing top-partners in Higgs+jets”.
3276 In: *JHEP* 08 (2014), p. 053. DOI: 10.1007/JHEP08(2014)053. arXiv: 1308.4771
3277 [hep-ph].
- 3278 [61] Georges Aad et al. “Fiducial and differential cross sections of Higgs boson production
3279 measured in the four-lepton decay channel in pp collisions at $\sqrt{s} = 8$ TeV with the
3280 ATLAS detector”. In: *Phys. Lett. B* 738 (2014), p. 234. DOI: 10.1016/j.physletb.
3281 2014.09.054. arXiv: 1408.3226 [hep-ex].
- 3282 [62] Georges Aad et al. “Measurements of fiducial and differential cross sections for Higgs
3283 boson production in the diphoton decay channel at $\sqrt{s} = 8$ TeV with ATLAS”.
3284 In: *JHEP* 09 (2014), p. 112. DOI: 10.1007/JHEP09(2014)112. arXiv: 1407.4222
3285 [hep-ex].
- 3286 [63] Georges Aad et al. “Measurements of the Total and Differential Higgs Boson Production
3287 Cross Sections Combining the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ Decay Channels at
3288 $\sqrt{s} = 8$ TeV with the ATLAS Detector”. In: *Phys. Rev. Lett.* 115 (2015), p. 091801.
3289 DOI: 10.1103/PhysRevLett.115.091801. arXiv: 1504.05833 [hep-ex].
- 3290 [64] Vardan Khachatryan et al. “Measurement of differential cross sections for Higgs
3291 boson production in the diphoton decay channel in pp collisions at $\sqrt{s} = 8$ TeV”. In:
3292 *Eur. Phys. J. C* 76 (2016), p. 13. DOI: 10.1140/epjc/s10052-015-3853-3. arXiv:
3293 1508.07819 [hep-ex].
- 3294 [65] Vardan Khachatryan et al. “Measurement of differential and integrated fiducial cross
3295 sections for Higgs boson production in the four-lepton decay channel in pp collisions
3296 at $\sqrt{s} = 7$ and 8 TeV”. In: *JHEP* 04 (2016), p. 005. DOI: 10.1007/JHEP04(2016)005.
3297 arXiv: 1512.08377 [hep-ex].

- 3298 [66] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs cross sections: 3.*
 3299 *Higgs properties*. CERN Report CERN-2013-004. 2013. DOI: 10.5170/CERN-2013-
 3300 004. arXiv: 1307.1347 [hep-ph].
- 3301 [67] Serguei Chatrchyan et al. “Measurement of Higgs boson production and properties
 3302 in the WW decay channel with leptonic final states”. In: *JHEP* 01 (2014), p. 096.
 3303 DOI: 10.1007/JHEP01(2014)096. arXiv: 1312.1129 [hep-ex].
- 3304 [68] G. Cowan. “A survey of unfolding methods for particle physics”. In: *Conf. Proc.*
 3305 C0203181 (2002), p. 248.
- 3306 [69] Michael Krämer, Stephen Mrenna, and Davison E. Soper. “Next-to-leading order
 3307 QCD jet production with parton showers and hadronization”. In: *Phys. Rev. D* 73
 3308 (2006), p. 014022. DOI: 10.1103/PhysRevD.73.014022. arXiv: hep-ph/0509127
 3309 [hep-ph].
- 3310 [70] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations
 3311 with Parton Shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070.
 3312 DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].
- 3313 [71] Nils Lavesson and Leif Lonnblad. “Extending CKKW-merging to one-loop matrix
 3314 elements”. In: *JHEP* 12 (2008), p. 070. DOI: 10.1088/1126-6708/2008/12/070.
 3315 arXiv: 0811.2912 [hep-ph].
- 3316 [72] Simone Alioli et al. “NLO Higgs boson production via gluon fusion matched with
 3317 shower in POWHEG”. In: *JHEP* 04 (2009), p. 002. DOI: 10.1088/1126-6708/2009/
 3318 04/002. arXiv: 0812.0578 [hep-ph].
- 3319 [73] Paolo Nason and Carlo Oleari. “NLO Higgs boson production via vector-boson
 3320 fusion matched with shower in POWHEG”. In: *JHEP* 02 (2010), p. 037. DOI:
 3321 10.1007/JHEP02(2010)037. arXiv: 0911.5299 [hep-ph].
- 3322 [74] Simone Alioli, Sven-Olaf Moch, and Peter Uwer. “Hadronic top-quark pair-production
 3323 with one jet and parton showering”. In: *JHEP* 01 (2012), p. 137. DOI: 10.1007/
 3324 JHEP01(2012)137. arXiv: 1110.5251 [hep-ph].
- 3325 [75] J. Alwall et al. “The automated computation of tree-level and next-to-leading order
 3326 differential cross sections, and their matching to parton shower simulations”. In:
 3327 *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301
 3328 [hep-ph].
- 3329 [76] T. Binoth et al. “Gluon-induced W-boson pair production at the LHC”. In: *JHEP*
 3330 12 (2006), p. 046. DOI: 10.1088/1126-6708/2006/12/046. arXiv: hep-ph/0611170
 3331 [hep-ph].
- 3332 [77] Marco Bonvini et al. “Signal-background interference effects for $gg \rightarrow H \rightarrow W^+W^-$
 3333 beyond leading order”. In: *Phys. Rev. D* 88 (2013), p. 034032. DOI: 10.1103/
 3334 PhysRevD.88.034032. arXiv: 1304.3053 [hep-ph].
- 3335 [78] Giampiero Passarino. “Higgs CAT”. In: *Eur. Phys. J. C* 74 (2014), p. 2866. DOI:
 3336 10.1140/epjc/s10052-014-2866-7. arXiv: 1312.2397 [hep-ph].

- 3337 [79] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. “PYTHIA 6.4 physics and
 3338 manual”. In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026.
 3339 arXiv: hep-ph/0603175 [hep-ph].
- 3340 [80] Hung-Liang Lai et al. “Uncertainty induced by QCD coupling in the CTEQ global
 3341 analysis of parton distributions”. In: *Phys. Rev. D* 82 (2010), p. 054021. DOI:
 3342 10.1103/PhysRevD.82.054021. arXiv: 1004.4624 [hep-ph].
- 3343 [81] Hung-Liang Lai et al. “New parton distributions for collider physics”. In: *Phys. Rev.*
 3344 *D* 82 (2010), p. 074024. DOI: 10.1103/PhysRevD.82.074024. arXiv: 1007.2241
 3345 [hep-ph].
- 3346 [82] LHC Higgs Cross Section Working Group. “Handbook of LHC Higgs Cross Sections”.
 3347 In: *arXiv:1101.0593* (2011).
- 3348 [83] Simone Alioli et al. “A general framework for implementing NLO calculations in
 3349 shower Monte Carlo programs: the POWHEG BOX”. In: *JHEP* 06 (2010), p. 043.
 3350 DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].
- 3351 [84] S. Agostinelli et al. “GEANT4: A simulation toolkit”. In: *Nucl. Instrum. Meth. A*
 3352 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.
- 3353 [85] E. Bagnaschi et al. “Higgs production via gluon fusion in the POWHEG approach in
 3354 the SM and in the MSSM”. In: *JHEP* 02 (2012), p. 088. DOI: 10.1007/JHEP02(2012)
 3355 088. arXiv: 1111.2854 [hep-ph].
- 3356 [86] Stefano Actis et al. “NLO Electroweak Corrections to Higgs Boson Production at
 3357 Hadron Colliders”. In: *Phys. Lett. B* 670 (2008), p. 12. DOI: 10.1016/j.physletb.
 3358 2008.10.018. arXiv: 0809.1301 [hep-ph].
- 3359 [87] Stefano Catani et al. “Soft gluon resummation for Higgs boson production at hadron
 3360 colliders”. In: *JHEP* 07 (2003), p. 028. DOI: 10.1088/1126-6708/2003/07/028.
 3361 arXiv: hep-ph/0306211 [hep-ph].
- 3362 [88] Yanyan Gao et al. “Spin determination of single-produced resonances at hadron
 3363 colliders”. In: *Phys. Rev. D* 81 (2010), p. 075022. DOI: 10.1103/PhysRevD.81.
 3364 075022. arXiv: 1001.3396 [hep-ph].
- 3365 [89] Sara Bolognesi et al. “Spin and parity of a single-produced resonance at the LHC”.
 3366 In: *Phys. Rev. D* 86 (2012), p. 095031. DOI: 10.1103/PhysRevD.86.095031. arXiv:
 3367 1208.4018 [hep-ph].
- 3368 [90] Ian Anderson et al. “Constraining anomalous HVV interactions at proton and
 3369 lepton colliders”. In: *Phys. Rev. D* 89 (2014), p. 035007. DOI: 10.1103/PhysRevD.
 3370 89.035007. arXiv: 1309.4819 [hep-ph].
- 3371 [91] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. “A Brief Introduction to
 3372 PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (2008), p. 852. DOI: 10.1016/j.
 3373 cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

- 3374 [92] Stanislaw Jadach, Johann H. Kuhn, and Zbigniew Was. “TAUOLA: A Library of
 3375 Monte Carlo programs to simulate decays of polarized tau leptons”. In: *Comput.
 3376 Phys. Commun.* 64 (1990), p. 275. DOI: 10.1016/0010-4655(91)90038-M.
- 3377 [93] CMS Collaboration. “Standard Model Cross Sections for CMS at 7 TeV”. In: *CMS
 3378 Generator Group Twiki* (2010).
- 3379 [94] John M. Campbell, R. Keith Ellis, and Ciaran Williams. “Vector boson pair pro-
 3380 duction at the LHC”. In: *JHEP* 07 (2011), p. 018. DOI: 10.1007/JHEP07(2011)018.
 3381 arXiv: 1105.0020 [hep-ph].
- 3382 [95] J. Ohnemus. “Order alpha-s calculations of hadronic W+- gamma and Z gamma
 3383 production”. In: *Phys. Rev. D* 47 (1993), pp. 940–955. DOI: 10.1103/PhysRevD.47.
 3384 940.
- 3385 [96] Iain W. Stewart and Frank J. Tackmann. “Theory uncertainties for Higgs and other
 3386 searches using jet bins”. In: *Phys. Rev. D* 85 (2012), p. 034011. DOI: 10.1103/
 3387 PhysRevD.85.034011. arXiv: 1107.2117 [hep-ph].
- 3388 [97] Sergey Alekhin et al. “The PDF4LHC Working Group Interim Report”. 2011.
- 3389 [98] Michiel Botje et al. “The PDF4LHC Working Group Interim Recommendations”.
 3390 2011.
- 3391 [99] Richard D. Ball et al. “Impact of Heavy Quark Masses on Parton Distributions
 3392 and LHC Phenomenology”. In: *Nucl. Phys. B* 849 (2011), p. 296. DOI: 10.1016/j.
 3393 nuclphysb.2011.03.021. arXiv: 1101.1300 [hep-ph].
- 3394 [100] A. D. Martin et al. “Parton distributions for the LHC”. In: *Eur. Phys. J. C* 63 (2009),
 3395 p. 189. DOI: 10.1140/epjc/s10052-009-1072-5. arXiv: 0901.0002 [hep-ph].
- 3396 [101] *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep.
 3397 CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. Geneva: CERN, Aug. 2011.
- 3398 [102] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”.
 3399 In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur. Phys. J.C73,2501(2013)], p. 1554. DOI:
 3400 10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z. arXiv:
 3401 1007.1727 [physics.data-an].
- 3402 [103] Tim Adye. “Unfolding algorithms and tests using RooUnfold”. 2011.
- 3403 [104] G. D’Agostini. “A Multidimensional unfolding method based on Bayes’ theorem”.
 3404 In: *Nucl. Instrum. Meth. A* 362 (1995), pp. 487–498. DOI: 10.1016/0168-9002(95)
 3405 00274-X.
- 3406 [105] Andreas Hocker and Vakhtang Kartvelishvili. “SVD approach to data unfolding”. In:
 3407 *Nucl. Instrum. Meth. A* 372 (1996), p. 469. DOI: 10.1016/0168-9002(95)01478-0.
 3408 arXiv: hep-ph/9509307 [hep-ph].
- 3409 [106] Vardan Khachatryan et al. “Precise determination of the mass of the Higgs boson
 3410 and tests of compatibility of its couplings with the standard model predictions using
 3411 proton collisions at 7 and 8 TeV”. In: *Eur. Phys. J. C* 75 (2015), p. 212. DOI:
 3412 10.1140/epjc/s10052-015-3351-7. arXiv: 1412.8662 [hep-ex].

- 3413 [107] Paolo Nason. “A New method for combining NLO QCD with shower Monte Carlo
3414 algorithms”. In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040.
3415 arXiv: hep-ph/0409146 [hep-ph].
- 3416 [108] S. Bolognesi, Y. Gao and A.V. Gritsan *et al.* “JHUGen”. In: URL <http://www.pha.jhu.edu/spin/>
3417 (2011).
- 3418 [109] Gionata Luisoni et al. “ $HW^\pm/HZ + 0$ and 1 jet at NLO with the POWHEG BOX
3419 interfaced to GoSam and their merging within MiNLO”. In: *JHEP* 10 (2013), p. 083.
3420 DOI: 10.1007/JHEP10(2013)083. arXiv: 1306.2542 [hep-ph].
- 3421 [110] Tom Melia et al. “W+W-, WZ and ZZ production in the POWHEG BOX”. In: *JHEP*
3422 11 (2011), p. 078. DOI: 10.1007/JHEP11(2011)078. arXiv: 1107.5051 [hep-ph].
- 3423 [111] T. Gehrmann et al. “ W^+W^- Production at Hadron Colliders in Next to Next to
3424 Leading Order QCD”. In: *Phys. Rev. Lett.* 113.21 (2014), p. 212001. DOI: 10.1103/
3425 PhysRevLett.113.212001. arXiv: 1408.5243 [hep-ph].
- 3426 [112] Patrick Meade, Harikrishnan Ramani, and Mao Zeng. “Transverse momentum resum-
3427 mation effects in W^+W^- measurements”. In: *Phys. Rev.* D90.11 (2014), p. 114006.
3428 DOI: 10.1103/PhysRevD.90.114006. arXiv: 1407.4481 [hep-ph].
- 3429 [113] Prerit Jaiswal and Takemichi Okui. “Explanation of the WW excess at the LHC
3430 by jet-veto resummation”. In: *Phys. Rev.* D90.7 (2014), p. 073009. DOI: 10.1103/
3431 PhysRevD.90.073009. arXiv: 1407.4537 [hep-ph].
- 3432 [114] John M. Campbell, R. Keith Ellis, and Ciaran Williams. “Bounding the Higgs width
3433 at the LHC: Complementary results from $H \rightarrow WW$ ”. In: *Phys. Rev.* D89.5 (2014),
3434 p. 053011. DOI: 10.1103/PhysRevD.89.053011. arXiv: 1312.1628 [hep-ph].
- 3435 [115] Richard D. Ball et al. “Unbiased global determination of parton distributions and
3436 their uncertainties at NNLO and at LO”. In: *Nucl. Phys.* B855 (2012), pp. 153–221.
3437 DOI: 10.1016/j.nuclphysb.2011.09.024. arXiv: 1107.2652 [hep-ph].
- 3438 [116] Vardan Khachatryan et al. “Event generator tunes obtained from underlying event
3439 and multiparton scattering measurements”. In: (2015). arXiv: 1512.00815 [hep-ex].
- 3440 [117] Peter Richardson and Alexandra Wilcock. “Monte Carlo Simulation of Hard Radia-
3441 tion in Decays in Beyond the Standard Model Physics in Herwig++”. In: *Eur. Phys.*
3442 *J.* C74 (2014), p. 2713. DOI: 10.1140/epjc/s10052-014-2713-x. arXiv: 1303.4563
3443 [hep-ph].
- 3444 [118] J. Bellm et al. “Herwig++ 2.7 Release Note”. In: (2013). arXiv: 1310.6877 [hep-ph].
- 3445 [119] *SM Higgs production cross sections at $\sqrt{s} = 13\text{--}14\text{ TeV}$.* <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt1314TeV>.
- 3446 [120] Fabrizio Caola et al. “QCD corrections to W^+W^- production through gluon fusion”.
3447 In: *Phys. Lett.* B754 (2016), pp. 275–280. DOI: 10.1016/j.physletb.2016.01.046.
3448 arXiv: 1511.08617 [hep-ph].

- 3450 [121] *NLO single-top channel cross sections*. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec>.
- 3451
- 3452 [122] *NNLO+NNLL top-quark-pair cross sections*. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO>.
- 3453
- 3454 [123] *LHC Higgs Cross Section Working Group*. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>.
- 3455
- 3456 [124] 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].
- 3457
- 3458
- 3459 [125] 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].
- 3460
- 3461
- 3462 [126] 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].
- 3463
- 3464 [127] 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].
- 3465
- 3466
- 3467 [128] Suyong Choi, Sungsoon 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.
- 3468
- 3469
- 3470 [129] 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].
- 3471
- 3472
- 3473
- 3474 [130] Massimiliano Grazzini. “NNLO predictions for the H \rightarrow WW $\rightarrow\ell\nu\ell\nu$ and H \rightarrow ZZ $\rightarrow 4\ell$ decay channels”. In: *JHEP* 02 (2008), p. 043. DOI: 10.1088/1126-6708/2008/02/043. arXiv: 0801.3232 [hep-ph].
- 3475
- 3476
- 3477 [131] Vardan Khachatryan et al. “Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV”. In: *Eur. Phys. J. C* 75.11 (2015), p. 542. DOI: 10.1140/epjc/s10052-015-3709-x. arXiv: 1505.04480 [hep-ex].
- 3478
- 3479
- 3480