



FACULTÉ
DES SCIENCES

UNIVERSITÉ LIBRE DE BRUXELLES

Search for a heavy scalar boson in the $ZZ \rightarrow 2\ell 2\nu$ channel with
the CMS experiment at the LHC using 2016 data at 13 TeV

Thesis submitted by Hugo DELANNOY

in fulfillment of the requirements for the award of the degree
of Doctor in Science ("Docteur en sciences")

Academic year 2018-2019

Supervisor : Professor Pascal VANLAER

Université libre de Bruxelles

Service de Physique des particules élémentaires

Interuniversity Institute for High Energies

Thesis jury :

Prof. Michel TYTGAT (ULB, Belgium, Chair)

Prof. Laurent FAVART (ULB, Belgium, Secretary)

Dr. Erik BUTZ (KIT, Germany)

Prof. Krijn DE VRIES (VUB, Belgium)

Dr. Xavier JANSSEN (UAntwerp, Belgium)



Acknowledgments

This should be the longest chapter of my thesis. Four years worth of work, even longer accounting for all of the preparatory training, internships and studies that led me to having the pleasure of presenting this manuscript today. Obviously, a project so long in the making will not allow me to acknowledge every single person who helped along the way as much as I would want to — unless I try to outdo the sixteen pages of CMS collaboration contributors, that is. Therefore, I compiled below an all-too-short overview of you all who contributed to this manuscript, in one way or the other.

First and foremost, I would like to thank my supervisor, Prof. Pascal Vanlaer for the wealth of discussions, advice and guidance that he has provided me over the past five years. From supervising my very first data analysis internships, on to my Master’s thesis, and then my PhD, he has been following and shaping my journey as a researcher from the very beginning.

Then, I would like to extend my gratitude to the members of my jury, namely Prof. Michel Tytgat, Prof. Laurent Favart, Dr. Erik Butz, Prof. Krijn De Vries, and Dr. Xavier Janssen, for the interest they have shown for my research and for the careful corrections and clarifications that they have outlined.

I have been lucky to benefit from extraordinary working conditions at the Interuniversity Institute for High Energies (IIHE), of which I wish to thank all members, past and present. More specifically, I would like to thank the team of “old-timers”, PhD candidates from my early days, who will have received more than their share of technical questions until the very end: Alexandre Leonard, Laurent Thomas, Thomas Lenzi and Florian Zenoni.

One more invaluable asset at IIHE is our IT dream team. Always available, they have managed to turn each malfunction, each bug into a learning opportunity — so much that they had me on the verge of joining their team. Thank you Romain Rougny,

Samir Amary, Shkelzen Rugovac, Olivier Devroede, Adriano Scodrani, and Stéphane Gérard.

I would like to thank all IIHE permanent academic staff, and in particular Laurent Favart for his precious QCD advice, as well as Gilles De Lentdecker, Ioana Maris and Juan Antonio Aguilar Sánchez for the pleasure I had in working with them on the muon laboratories. During the last years of my PhD, Yannick Allard joined IIHE and these laboratories. I am grateful for his enthusiasm in sharing the details of his work and his thoughtful advice for my first job interviews.

The IIHE is lucky to be able to count on two incredible office managers: Marleen Goeman and Audrey Terrier. Efficient, always available — even on all-too-short deadlines — and gifted with bulletproof patience and kindness, they have set the bar very high for all future administrations I will cross paths with.

I would also like to thank my direct colleagues: David Vannerom for his technical and psychological support when bugs survived more than one day in my code; Louis Moureaux for his advice and detailed programming explanations; Bugra Bilin for his guidance and the never-ending atmosphere of joy that he managed to foster in the office. Together, the three of you succeeded in making me go back home way too late on too many installations.

More thanks go out to my working group, $2\ell 2\nu$, and in particular to Hugues Brun and Nicolas Postiau. Changing the world, one line of code at a time, while having enthralling conversations around a computer, be it at the CMS Center or in our offices, will be a memory that stays with me. Working with you was a real pleasure, and one facet of my IIHE years that I miss already dearly.

During my thesis, I had the opportunity to work for the CMS tracker data quality monitoring (Tracker DQM) — a wonderful experience, at the very core of operations. My gratitude goes to Tomas Hreus, who introduced me to this world and who has been of tremendous support for all my technical questions during the past four years. I would also like to thank Isabelle De Bruyn who, being junior co-convener of the Tracker DQM with me, also introduced me to the VUB on lunch breaks. She made the occasional frustrating Tracker DQM moment much more agreeable. Our discussions and plans to revolutionize the DQM are fond memories from my thesis. Finally, I would like to thank my supervisor, Prof. Pascal Vanlaer, one more time for allowing me to seize this fantastic opportunity and see it fully through.

During my many stays at CERN for the Tracker DQM, I seemed to always run into Jarne De Clercq. Thank you for countless conversations that made those stays less solitary. Were I to go back to CERN one day, I am sure I would run into you.

Finally, I would like to thank all of those who have been helping from the shadows: my friends and family. Always lending a listening ear when I needed to talk, in moments of passion or setbacks, about my research, my life. I would not stand where

I am today without your thoughtful guidance. While you are too many to name, you all belong here.

In particular, I would like to thank my parents, who are one of my life greatest chances.

And, last but not least, I want to thank my other greatest chance: Coraline. For everything.

Contents

1. Introduction	1
2. Theoretical introduction and motivations	3
2.1. A phenomenological description of the standard model of particle physics	3
2.1.1. The standard model Lagrangian	4
2.1.2. The scalar sector	8
2.1.3. Scalar boson production and decay modes	9
2.2. Motivations to search for BSM physics	11
2.2.1. Neutrino masses	12
2.2.2. Dark matter and dark energy	12
2.2.3. Matter-antimatter asymmetry	13
2.2.4. The number of free parameters in the SM	13
2.2.5. The hierarchy problems	13
2.2.6. Coupling unification	14
2.3. The scalar sector and its extension in BSM models	14
2.3.1. Electroweak singlet model	14
2.3.2. A class of BSM models: the two-Higgs-doublet models	15
2.4. Searches for BSM physics and constraints on 2HDM	17
2.4.1. Current constraints on 2HDM	20
3. The CMS experiment at the LHC	23
3.1. The Large Hadron Collider	23
3.1.1. The proton production, the injector complex and the LHC	24
3.1.2. Luminosity and data-taking condition at the LHC in 2016	26
3.2. The CMS experiment	28
3.2.1. Overview of the CMS detector	28
3.2.2. The coordinate system	30

3.2.3.	The superconducting solenoid magnet	30
3.2.4.	The tracker	31
3.2.5.	The calorimeters	40
3.2.6.	The muon detectors	47
3.2.7.	The trigger system	54
4.	Event generation, simulation and reconstruction	57
4.1.	Introduction	57
4.2.	Event generation	58
4.2.1.	Hard-scattering process and parton distribution functions	60
4.2.2.	Parton showering	61
4.2.3.	Hadronization	62
4.3.	Simulation of the detector	62
4.4.	Object reconstruction	63
4.4.1.	The particle-flow reconstruction	63
4.4.2.	Muon reconstruction	65
4.4.3.	Electron and photon reconstruction	67
4.4.4.	Jet reconstruction	72
4.4.5.	Missing transverse momentum reconstruction	77
4.4.6.	High-level object energy resolution summary	79
5.	Search for a heavy scalar boson in the $ZZ \rightarrow 2\ell 2\nu$ channel	85
5.1.	Foreword	85
5.2.	Signal simulation	86
5.2.1.	The MELA reweighting procedure	87
5.2.2.	Impact of interferences	88
5.3.	Trigger requirements	90
5.4.	Backgrounds	94
5.5.	Corrections to simulated samples	96
5.5.1.	Data-to-simulation scale factors	96
5.5.2.	Higher-order corrections to di-boson samples	98
5.6.	Analysis strategy	99
5.6.1.	Event selection	99
5.6.2.	Event categorization	101
5.6.3.	Choice of observable: the transverse mass	101
5.6.4.	Statistical analysis	101
5.7.	Data-to-simulation comparisons	103
5.8.	Data-driven estimations	104
5.8.1.	Z +jet background estimation	107
5.8.2.	Non-resonant background estimation	110

5.9. Uncertainties	113
5.9.1. Instrumental uncertainties	114
5.9.2. Theoretical uncertainties	115
5.10. Results	118
6. Data-driven estimation of the Z+jet background	127
6.1. Motivation	127
6.2. The data-driven method	128
6.3. Trigger requirement and event selection	129
6.4. Simulated samples for the photon control region	130
6.5. Photon control region	132
6.6. Reweighting procedure	133
6.7. Closure test	137
6.8. Data-driven estimation in the di-lepton region	137
6.8.1. Changes with respect to published analysis	138
6.8.2. Results and comparison with simulated samples in the di-lepton region	141
6.9. Uncertainties	141
6.9.1. Uncertainty from the reweighting procedure	142
6.9.2. Statistical uncertainties	142
6.9.3. Systematic uncertainties	144
6.10. Conclusion	149
7. Conclusion	151
Bibliography	155
A. Notions of statistics	167
A.1. Statistics and the likelihood at the LHC	167
A.2. Introducing uncertainties in the likelihood	168
A.3. Maximum likelihood estimation and nuisance parameter impacts . . .	170
B. Additional plots	173

Introduction

Developed in the latter half of the 20th century, the standard model (SM) of particle physics is the current best description of elementary particles and their interactions through electromagnetic, weak and strong forces. Since its finalization in the mid-1970s, a large number of experiments have observed results in agreement with the predictions of the SM, reinforcing confidence in this model over the years. Lately, the Large Hadron Collider (LHC) and its experiments allowed to probe many aspects of the SM, confirming many SM predictions up to a high level of precision. In particular, the discovery of a scalar boson compatible with SM predictions was the highlight of the first years of operation of the LHC at 7 and 8 TeV between 2010 and 2012 (run 1). This boson was introduced as a consequence of the electroweak symmetry breaking by Robert Brout, François Englert, Peter Higgs in 1964. In 2013, Englert and Higgs were awarded the Nobel Prize in Physics, acknowledging “the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN’s Large Hadron Collider” [1].

However, despite its success, the SM cannot explain everything on its own. Neutrino masses, dark matter and matter-antimatter asymmetry are experimental evidence that the SM of particle physics has limits. On the theory side, the SM shows conceptual problems such as a large number of free parameters, the so-called hierarchy problem or the unification of the coupling constants. Models going beyond the SM try to address such issues, and in particular classes of BSM theories require an extended scalar sector. Exploring BSM physics in the scalar sector can be done in two complementary ways: precision measurements of the observed 125 GeV scalar boson and search for direct evidence of new particles.

The scope of this thesis falls within the direct search for heavy BSM scalar boson (H) with the data of the second run of operations of the LHC at 13 TeV (run 2). In particular, the thesis focuses on $H \rightarrow ZZ \rightarrow \ell\bar{\ell}\nu\bar{\nu}$ processes, where ℓ is an electron or a muon and ν a neutrino. This choice is motivated by its sensitivity to a potential discovery of a new scalar. Indeed, the $ZZ \rightarrow \ell\bar{\ell}\nu\bar{\nu}$ decay has the advantage of having a clear signature from the $Z \rightarrow \ell\bar{\ell}$ decay and large missing energy due to neutrinos, while having a larger branching ratio than $ZZ \rightarrow 4\ell$ due to the $Z \rightarrow \nu\bar{\nu}$ decay.

At the beginning of this thesis, the last published results were produced using run 1 data [2]. This analysis was optimized separately in two production modes of the scalar boson: gluon fusion and vector boson fusion. The observed data was compatible with the expected background from SM predictions and a scalar boson with SM couplings was excluded for scalar masses between 248 GeV and 930 GeV at a 95% confidence level. This analysis also included a search for a narrow scalar resonance in the so-called electroweak singlet model. This model is based on the mixing of a SM-like scalar boson with an electroweak singlet. Limits have been set on the parameters of this model, such as the scalar decay width, the mixing parameter or the branching ratio of the new heavy scalar boson to new particles. These limits were more stringent than the indirect limits coming from precision measurements of the observed SM-like scalar boson for masses between 300 GeV and 700 GeV. Due to the increase in energy and luminosity for the run 2 of the LHC operations, the search for a heavy scalar boson decaying into a pair of Z bosons is expected to have a sensitivity at least 10 times greater than the run 1 published results. Also, since a scalar boson has been discovered in 2012, a search for an additional heavy scalar boson with SM-like couplings only makes sense in certain models such as the electroweak singlet model. A search for a generic H boson, varying its mass and width, is thus performed in the present thesis, in the gluon fusion and vector boson fusion production modes.

The present thesis is divided as follows. Chapter 2 presents the SM of particle physics as well as motivations to look beyond it. In addition, two classes of BSM models are presented: the electroweak singlet model and the two-Higgs-doublet models. The LHC and the CMS experiment with a focus on their status in 2016 are presented in Chapter 3. Chapter 4 presents the event generation, simulation and reconstruction as done by the CMS collaboration. A summary of high-level object energy resolution is then shown. Chapter 5 presents the search of a heavy BSM scalar boson in the $H \rightarrow ZZ \rightarrow \ell\bar{\ell}\nu\bar{\nu}$ channel with the 2016 data collected by the CMS collaboration. My main contribution is eventually detailed in Chapter 6 which describes the data-driven estimation of the Z +jet background, a major background of this thesis analysis.

Chapter 2

Theoretical introduction and motivations

This chapter starts by introducing the standard model (SM) of particle physics which has proven to be robust in the last 50 years. However, hints coming from both experimental observations and theory tend to show that physics going beyond this model exist and — since many models going beyond the SM predict an extended scalar sector — it motivates the search presented in this thesis . This chapter therefore describes motivations to look beyond the standard model (BSM) and focus on a class of BSM models: the two-Higgs-doublet models.

2.1. A phenomenological description of the standard model of particle physics

The standard model of particle physics is the current best description of elementary particles and their interactions. This model describes successfully three of the four known fundamental forces, namely the electromagnetic force, the weak force and the strong force. The SM does not include a description of the gravitational force.

Figure 2.1 shows the constituents of the SM. According to this model, matter is made of spin- $\frac{1}{2}$ fermions: quarks and leptons. Those fermions are divided into three generations, each generation being a heavier copy of the previous one. The quarks participate to all fundamental interactions and have a color charge, an electric charge and a weak charge. The leptons are divided into charged leptons and neutrinos. Charged

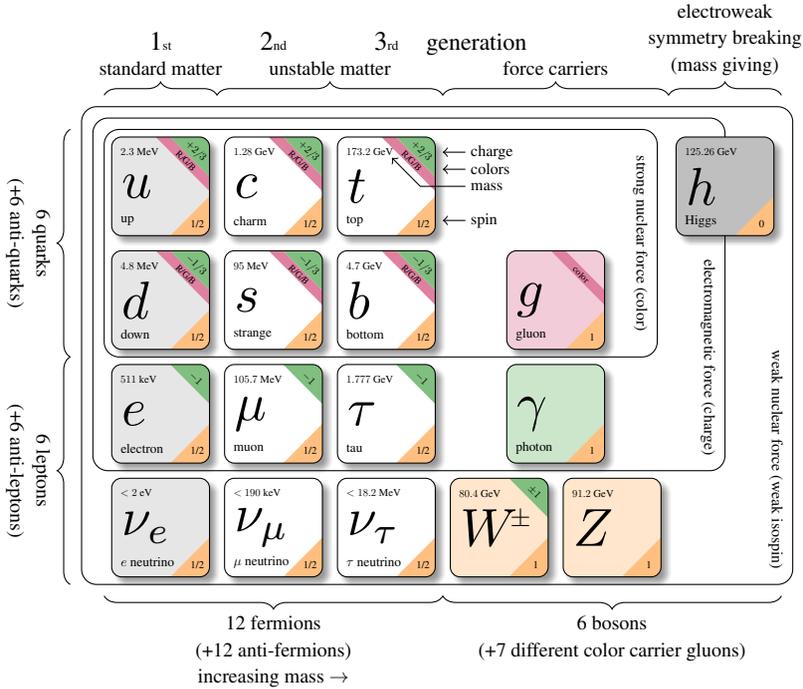


Figure 2.1.: Elementary particles of the SM. Inspired from [3].

leptons participate to the electromagnetic and weak interactions while neutrinos interact only through weak interactions.

Those interactions are described by the exchange of gauge bosons. The photon is the gauge boson for electromagnetic interactions, the W^\pm and Z are the gauge bosons for weak interactions and the gluons are the gauge bosons for strong interactions.

The last piece of the SM is the Higgs boson, a scalar boson without electric charge nor color charge. The interactions of particles (including the scalar boson itself) with the Higgs boson give rise to the mass of particles.

2.1.1. The standard model Lagrangian

The SM is a gauge invariant quantum field theory. This theory groups the electroweak theory together with quantum chromodynamics (QCD). Its fundamental particles have been described briefly above and their associated quantum fields are the following:

- the fermion field, noted ψ

- the electroweak boson fields¹, noted W^1, W^2, W^3 and B
- the eight gluon fields, noted \mathcal{A}^a with a going from 1 to 8
- the Higgs field, noted ϕ .

The SM Lagrangian has to be constructed in such a way that it contains a certain set of symmetries which represent physics conservation laws such as conservation of energy, momentum, angular momentum and charge. The SM Lagrangian density is therefore satisfying causality, gauge invariance and Lorentz invariance.

The gauge group of the SM is $SU(3)_C \times SU(2)_L \times U(1)_Y$. $SU(3)_C$ is the group of colour transformations for which quarks transform according to the fundamental representation and antiquarks according to its complex conjugate representation. $SU(2)_L \times U(1)_Y$ is the group for electroweak interactions. The subscript Y indicates that U(1) symmetry acts on the weak hypercharge, while the subscript L indicates that SU(2) symmetry is chiral and only acts on the left-handed states.

One way to decompose the SM Lagrangian density \mathcal{L}_{SM} is as follows:

$$\mathcal{L}_{SM} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_f + \mathcal{L}_{\text{Yuk}} + \mathcal{L}_\phi \quad (2.1)$$

for the gauge, fermion, Yukawa and scalar sectors respectively.

The gauge Lagrangian density $\mathcal{L}_{\text{gauge}}$ regroups the gauge fields of all three symmetry groups²:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}\mathcal{A}_{\mu\nu}^a\mathcal{A}^{a\mu\nu} - \frac{1}{4}W_{\mu\nu}^iW^{i\mu\nu} - \frac{1}{4}B_{\mu\nu}^iB^{i\mu\nu} \quad (2.2)$$

where the field strength tensors are:

$$\begin{aligned} \mathcal{A}_{\mu\nu}^a &= \partial_\mu\mathcal{A}_\nu^a - \partial_\nu\mathcal{A}_\mu^a - g_S f_{abc}\mathcal{A}_\mu^b\mathcal{A}_\nu^c, & \text{with } a, b, c = 1, \dots, 8; \\ W_{\mu\nu}^i &= \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g\varepsilon_{ijk}W_\mu^jW_\nu^k, & \text{with } i, j, k = 1, 2, 3; \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \end{aligned}$$

with g_S and g being the coupling constants associated to the $SU(3)_C$ and $SU(2)_L$ groups respectively. ε_{ijk} is a totally antisymmetric tensor and comes from the structure constants of the Lie algebra $\mathfrak{su}(2)$. f_{abc} are the structure constants of the Lie algebra $\mathfrak{su}(3)$. The third term of the expression for $\mathcal{A}_{\mu\nu}^a$ and $W_{\mu\nu}^i$ arises respectively from the non-Abelian nature of the $SU(3)$ and $SU(2)$ groups and indicates the self-interaction of the gauge fields \mathcal{A}^a and W^i . The U(1) group being Abelian the structure constants of its Lie algebra are zero and no self-interactions of the photon are allowed.

¹The W^\pm, Z and γ bosons are not the quanta directly corresponding to these fields but are the quanta of the mass eigenstates fields, result of the Brout-Englert-Higgs mechanism (see Sec. 2.1.2).

²This thesis is using the natural units system in which $\hbar = c = 1$.

The fermionic part of the Lagrangian density \mathcal{L}_f regroups the kinetic energy terms for fermions:

$$\mathcal{L}_f = i\bar{\psi}\not{D}\psi. \quad (2.3)$$

which includes an implicit summation on all the left-handed and right-handed fermion fields as described below. The distinction between right-handed and left-handed fields for fermions has been required after observing that only left-handed particles were undergoing weak interactions. Therefore, the SM requires that all right-handed fermion fields do not transform under the SU(2) group and are therefore SU(2) singlets. On the other hand, the left-handed fermion fields are forming SU(2) doublets. The fermion field ψ can therefore be written as follow:

$$\psi = \psi_R + \psi_L = \gamma_R\psi + \gamma_L\psi \quad (2.4)$$

where $\gamma_R = \frac{1}{2}(1 + \gamma^5)$ and $\gamma_L = \frac{1}{2}(1 - \gamma^5)$ are the projection operators, with $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ and γ^μ being the four Dirac matrices. The ψ_R and ψ_L correspond to:

$$\begin{aligned} \psi_R &= \gamma_R \ell, \gamma_R u, \gamma_R d \\ \psi_L &= \gamma_L \begin{pmatrix} \nu \\ \ell \end{pmatrix}, \gamma_L \begin{pmatrix} u \\ d \end{pmatrix} \end{aligned}$$

with ℓ representing charged leptons (e , μ and τ), ν neutrinos (ν_e , ν_μ and ν_τ), u quarks with an electric charge of $2/3$ (u , c and t) and d quarks with an electric charge of $-1/3$ (d , s , b). No right-handed neutrinos were observed to this day. Therefore they are not embedded in the basic description of the SM. However, even if the SM predicts that such particles would not interact at all, it is still possible to introduce them to build mass terms for neutrinos in the Lagrangian density. In the Eq. 2.3, the Feynman slash notation $\not{D} = \gamma_\mu D^\mu$ is used. D^μ is the gauge covariant derivative and is defined as follows:

$$\begin{aligned} D^\mu \gamma_L \begin{pmatrix} u \\ d \end{pmatrix} &= (\partial^\mu + \frac{i}{2}g_S\lambda_a\mathcal{A}_a^\mu + \frac{i}{2}g\sigma_i W_i^\mu + ig'YB^\mu)\gamma_L \begin{pmatrix} u \\ d \end{pmatrix} \\ D^\mu \gamma_L \begin{pmatrix} \nu \\ \ell \end{pmatrix} &= (\partial^\mu + \frac{i}{2}g\sigma_i W_i^\mu + ig'YB^\mu)\gamma_L \begin{pmatrix} \nu \\ \ell \end{pmatrix} \\ D^\mu \gamma_R u &= (\partial^\mu + \frac{i}{2}g_S\lambda_a\mathcal{A}_a^\mu + ig'YB^\mu)\gamma_R u \\ D^\mu \gamma_R d &= (\partial^\mu + \frac{i}{2}g_S\lambda_a\mathcal{A}_a^\mu + ig'YB^\mu)\gamma_R d \\ D^\mu \gamma_R \ell &= (\partial^\mu + ig'YB^\mu)\gamma_R \ell \end{aligned}$$

where g' is the coupling constant associated to the $U(1)_Y$ group. $\frac{1}{2}\sigma_i$ are the generators for the SU(2) group, with σ_i being the Pauli matrices. Similarly $\frac{1}{2}\lambda_a$ are the

fundamental representation of SU(3) color algebra with a running over the eight color degrees of freedom of the gauge group. The matrices λ_a are the Gell-Mann matrices. Y is the weak hypercharge and is the generator of U(1). It is defined as $Y = Q - T^3$ with Q the electric charge and T^3 the third component of the weak isospin in the SU(2) doublet.

At this stage, introducing a mass term for the gauge bosons would result in a non-invariant Lagrangian under local gauge transformations. Moreover, trying to add fermionic mass terms would also not leave the Lagrangian invariant under electroweak symmetry group transformations. However both fermions and the W^\pm and Z bosons are massive. The way these bosons and fermions acquire a mass in the SM is described by the Brout-Englert-Higgs mechanism (see Sec. 2.1.2) which adds a scalar field to the Lagrangian.

The Yukawa Lagrangian density \mathcal{L}_{Yuk} describes the interactions between the fermions and the scalar doublet ϕ which give rise to fermion masses. The scalar field is a SU(2) doublet ϕ composed of two complex scalar fields and can be written as $\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$. The Yukawa Lagrangian density is:

$$\mathcal{L}_{\text{Yuk}} = -\lambda_\ell \bar{\psi}_{\ell L} \phi \psi_{\ell R} - \lambda_d \bar{\psi}_{qL} \phi \psi_{dR} - \lambda_u \bar{\psi}_{qL} \tilde{\phi} \psi_{uR} + \text{h.c.} \quad (2.5)$$

where h.c. stands for hermitian conjugate and $\tilde{\phi} = i\sigma_2(\phi^\dagger)^t$. λ_f are the Yukawa couplings of the fermion f .

After electroweak symmetry breaking (see Sec. 2.1.2), the fermion mass is given by:

$$m_f = \frac{1}{\sqrt{2}} \lambda_f v. \quad (2.6)$$

where v is the vacuum expectation value (v.e.v.) of the scalar potential (see below) and corresponds to $v \simeq 246$ GeV. The Yukawa sector introduces many free parameters in the SM through the λ_f parameters.

The scalar sector and the electroweak symmetry breaking will be described in more details in the next section but one can already detail its Lagrangian density \mathcal{L}_ϕ . This Lagrangian density is made of a kinematic and a potential term:

$$\mathcal{L}_\phi = (D^\mu \phi)^\dagger D_\mu \phi - V(\phi), \quad (2.7)$$

where the covariant derivative is given by:

$$D_\mu \phi = \left(\partial_\mu + \frac{i}{2} g \sigma^i W_\mu^i + \frac{i}{2} g' B_\mu \right) \phi \quad (2.8)$$

and the Higgs potential is chosen to be:

$$V(\phi) = \lambda(\phi^\dagger \phi)^2 - \mu^2 \phi^\dagger \phi \quad (2.9)$$

with μ^2 and λ being positive real numbers.

2.1.2. The scalar sector

As aforementioned, mass terms for fermions and gauge fields are not present in $\mathcal{L}_{\text{gauge}}$ and \mathcal{L}_f because only singlets under $SU(3)_C \times SU(2)_L \times U(1)_Y$ could acquire a mass with an interaction of the type $m\bar{\psi}\psi$ without breaking gauge invariance. Electroweak symmetry breaking mechanism — also known as the *Brout-Englert-Higgs mechanism* — allows the introduction of masses for fermions and bosons and was developed simultaneously by R. Brout, F. Englert [4] and P. Higgs [5]. This mechanism implies the existence of a scalar boson that was observed on July 2012 at the LHC by both CMS and ATLAS collaborations [6, 7].

The electroweak symmetry breaking

An other way to write the scalar doublet is:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \varphi_1 + i\varphi_2 \\ \varphi_3 + i\varphi_4 \end{pmatrix} \quad (2.10)$$

with φ_i being real scalar fields.

Coming back at the potential for the scalar field given at Eq. 2.9, this potential has the shape of a “Mexican hat” and its lowest-energy configuration correspond to a non-zero field forming a dimension-3 circle of degenerate minima. The field can be developed around one of its degenerate minima in an arbitrary direction of the electroweak symmetry breaking:

$$\phi_v = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}. \quad (2.11)$$

With this choice of configuration the second derivative of the potential along the dimension-3 circle is zero and the second derivative in the radial direction is positive. This leads to the existence of three massless particles — called Goldstone bosons — and one massive particle.

Mass terms for gauge bosons appear from the interaction terms with the scalar field when replacing the field by its v.e.v. in:

$$|D_\mu \phi_v|^2 = \frac{g^2 v^2}{8} \left((W_\mu^1)^2 + (W_\mu^2)^2 + (-W_\mu^3 + \frac{g'}{g} B_\mu)^2 \right). \quad (2.12)$$

If we define two new complex fields as:

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp iW_\mu^2) \quad (2.13)$$

and two new real fields as:

$$Z_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(gW_\mu^3 - g'B_\mu)$$

$$A_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(gW_\mu^3 + g'B_\mu)$$

then Eq. 2.12 becomes:

$$|D_\mu \phi_v|^2 = \frac{g^2 v^2}{4} W_\mu^+ W^{-\mu} + \frac{1}{2} \frac{(g^2 + g'^2)v^2}{4} Z_\mu Z^\mu. \quad (2.14)$$

From this equation, we see that the result of the spontaneous symmetry breaking is a set of three massive vector bosons — the W^\pm vector bosons with a mass $m_W = \frac{gv}{2}$ and the Z boson with a mass $m_Z = \frac{v}{2} \sqrt{g^2 + g'^2}$ — and the massless photon A .

Regarding mass terms for fermions, they are built by replacing ϕ by its expectation value in \mathcal{L}_{Yuk} as shown in Eq. 2.6.

2.1.3. Scalar boson production and decay modes

The four largest production modes of a SM-like scalar boson at LHC [8] are shown on Fig. 2.2 and are:

- a) the **gluon-gluon fusion** (ggF or ggH). This process is mediated by top and bottom quark loops. Due to the large size of the top and bottom Yukawa couplings and the gluon densities it has the largest cross section of the SM-like scalar boson production at the LHC [9].
- b) the **vector boson fusion** (VBF or qqH). It has a cross section an order of magnitude below the ggH production at low mass and a cross section comparable to the ggH one at high mass. The VBF production mode is characterized by two high-momentum quarks in the final state. The topology of this pair of quarks, such as a large invariant mass and a big angular gap between the quarks, allows to reduce significantly the background associated to this process and makes it an interesting process to tag and study.
- c) the **associated production with a vector boson** (VH), also called *Higgsstrahlung*. It consists in the production of a virtual vector boson splitting into a real boson and a scalar boson. Its cross section is lower than the VBF one. In particular, in the search for a heavy SM-like scalar boson, the cross section of such a process decreases rapidly with the mass of the scalar boson for kinematic reasons, as shown on Fig. 2.3.

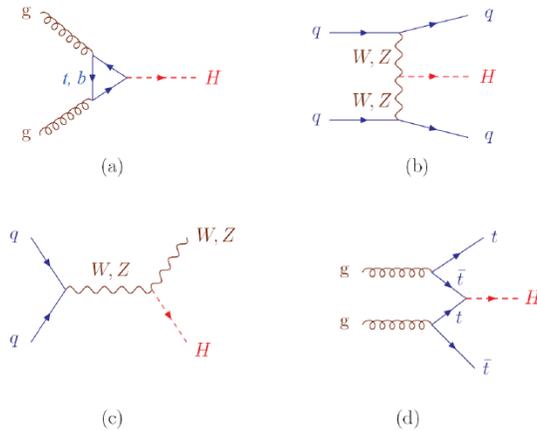


Figure 2.2.: Lowest order Feynman diagrams for the four largest SM-like scalar boson production mode at the LHC: a) gluon-gluon fusion, b) vector boson fusion, c) associated production with a vector boson and d) top quark fusion.

- d) the **associated production with a pair of top quarks** ($t\bar{t}H$). It has the same initiating particles (i.e. gluons) than the ggH production but has a two order of magnitudes lower cross section compared to the loop process.

The search presented in this thesis is a model independent search for a heavy scalar boson. The production modes described above can be affected by new physics, for example with the introduction of new particles coupling to the scalar boson or with a new scalar boson not coupling to any fermions. It is expected that for a new heavy scalar boson coupling to fermions, the ggF production mode will be the dominant process for the production processes involving coupling to fermions. Indeed, the introduction of new fermions would just enhance the cross section due to the loop and the hierarchy between ggF and other processes will be conserved [9]. If the scalar does not couple to fermions, the VBF production mode becomes the dominant process. Indeed, other processes such as VH production are kinematically less favored when seeking for a heavy new scalar boson. Therefore, the production modes studied in this thesis are these two dominant processes in their own category: the ggF and the VBF production.

The branching fractions of a SM-like scalar boson are shown on Fig. 2.4. A SM-like scalar boson can decay in:

- **fermions:** the couplings of fermions to the scalar boson being directly proportional to the fermion mass, the decay width of a scalar boson into fermions is

proportional to the square of the fermion mass. At low mass of the scalar boson, the $b\bar{b}$ decay dominates while at high mass (around $2m_t$) the $t\bar{t}$ decay becomes important.

- **bosons:** once the boson decay modes are kinematically favored, the branching fractions of the scalar boson into a pair of W or Z boson are dominant and these decay modes are therefore good channels to search for a heavy SM-like scalar boson. It should be noted that the scalar boson does not couple to photons or gluons, since they are massless. However, the decays of the scalar boson into gg , $\gamma\gamma$ and $Z\gamma$ are allowed through massive particle loops, such as W boson or fermion loops.

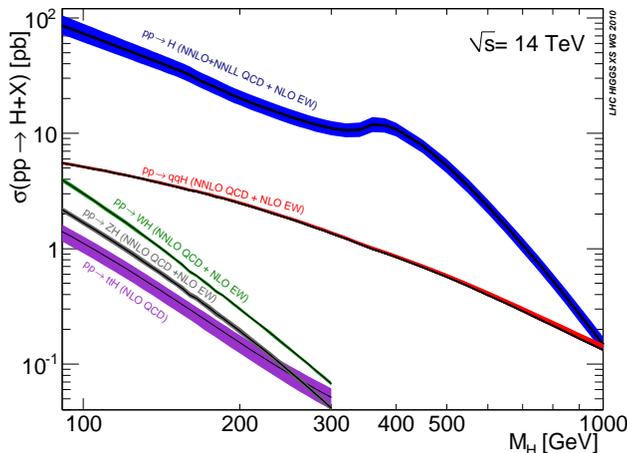


Figure 2.3.: SM-like scalar boson production cross sections as a function of the scalar boson mass in proton-proton collisions at a center-of-mass energy of 14 TeV for the ggH , qqH , VH and $t\bar{t}H$ production mechanisms [10].

Those couplings are for a SM-like scalar boson. However, scenarii going beyond the Standard Model predict new scalar bosons with different couplings such as the two-Higgs-doublet models discussed in Sec. 2.3.2.

2.2. Motivations to search for BSM physics

There are strong motivations to search for physics beyond the SM [11, 12]. Some of them are based on experimental observations such as neutrino masses, dark matter and matter-antimatter asymmetry while others are coming from conceptual problems

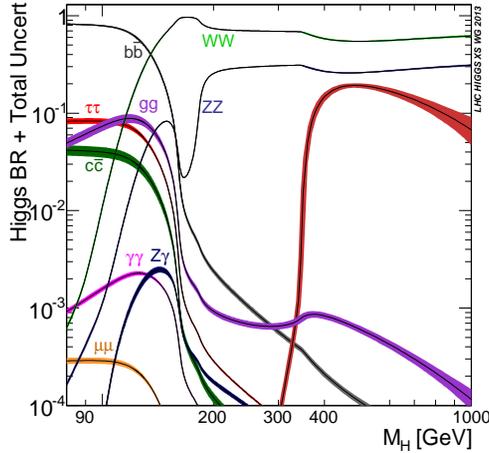


Figure 2.4.: Branching fractions (BR) of a SM-like scalar boson as a function of its mass [10].

in the SM theory, such as a large number of free parameters, the so-called *hierarchy problem* or the unification of the coupling constants.

2.2.1. Neutrino masses

Experiments looking at neutrinos from solar, atmospheric, reactor and accelerator sources observed that neutrinos oscillate and change their flavor in flight [13, 14]. Such oscillations are only possible if neutrinos are massive. If they are, flavor eigenstate neutrinos (ν_e , ν_μ and ν_τ) are linear combinations of the fields of (at least) three mass eigenstate neutrinos (ν_1 , ν_2 and ν_3). Upper limits have been set on neutrinos masses (with the most stringent being $m_{\bar{\nu}_e} < 2.05$ eV at 95% CL [15]) and the differences between the neutrino squared masses have been measured along with the mixing angles appearing in the Pontecorvo-Maki-Nakagawa-Sakata matrix relating the flavor eigenstate neutrinos to the mass eigenstate neutrinos.

2.2.2. Dark matter and dark energy

The search for dark matter is very well motivated by several cosmological observations. In particular, latest measurements of the temperature and polarization anisotropies of the cosmic microwave background by the Planck and WMAP experiments indicate that the matter described in the SM would only represent approximately 5% of the energy content of the universe while the dark matter would contribute to approximately 31% of this content, the rest being attributed to *dark energy* [16]. Dark

energy is one of the main suspects to explain the accelerated expansion of the universe [17].

2.2.3. Matter-antimatter asymmetry

Still looking at cosmological arguments, it is clear that we are surrounded by matter while it is believed that matter and antimatter were equally produced at the time of the Big Bang. One of the three conditions to obtain such a global matter-antimatter asymmetry [18] is CP-violation. The SM already includes sources of CP-violation through the CKM matrix³. However, they are not enough to explain the magnitude of the matter-antimatter asymmetry observed.

2.2.4. The number of free parameters in the SM

The next subsections are describing motivations from the theory point of view. First, the SM contains 19 free parameters including the CKM matrix parameters, the gauge coupling constants and the free parameters linked to the scalar and Yukawa sectors. This large number of free parameters, and in particular for the scalar and Yukawa sectors could be an indication for the existence of a more general theory than the SM.

2.2.5. The hierarchy problems

The hierarchy problems refer to the huge differences appearing in the scalar and Yukawa sectors of the SM [19]. The fermion mass hierarchy problem refers to the six orders of magnitude difference in the mass of the electron and the top quark which sounds unnatural. This difference is even bigger if one considers neutrinos. The gauge hierarchy problem refers to the huge energy difference between the weak and the Planck scale. The weak scale is given by the vev of the scalar field and is approximately 246 GeV while the Planck scale is around 10^{19} GeV. The radiative corrections to the scalar boson mass come from its couplings to gauge bosons, Yukawa couplings and its self-couplings. Therefore, if the SM was valid up to the Planck scale, then the scalar mass and its vev would be driven to the Planck scale by the radiative corrections... unless extraordinary cancellation of terms happen in the Lagrangian, which is quite unnatural. Similarly, if new heavy particles happened to exist at the Planck scale, they should couple to the scalar boson and they would therefore induce loop corrections to its physical mass that would have to cancel in an extraordinary fine-tuned way.

³The Cabibbo-Kobayashi-Maskawa matrix is the matrix describing the mixing of the flavour of the quarks through flavour changing charged currents. The matrix can be parametrized with three mixing angles and one CP-violating phase.

2.2.6. Coupling unification

The three forces/couplings (strong, weak and electromagnetic) arising from the symmetry groups of the SM have very different intensity around the electroweak scale. However, those three coupling strengths tend to become comparable at higher energies. In particular, the electromagnetic and weak forces can be unified to form the so-called electroweak interaction at high energies. Nevertheless the strong coupling does not meet the two other couplings at such energies and no grand unification can be reached unless new BSM particles modify the running of the couplings.

2.3. The scalar sector and its extension in BSM models

In order to try to solve the problems aforementioned, many extensions to the SM were proposed in the last decades, some of them including new scalar particles. In particular, this thesis focus on an extended scalar sector in BSM. It can be shown [15] that the parameter $\rho = \frac{m_Z}{m_W \cos \theta_W}$, measured to be close to one ($\rho = 1.00039 \pm 0.00019$), can also be written as it follows in a $SU(2) \times U(1)$ gauge theory, at tree level, if there are n scalar multiplets ϕ_i of weak isospin I_i , weak hypercharge Y_i and v.e.v. v_i :

$$\rho = \frac{\sum_{i=1}^n [I_i(I_i + 1) - \frac{1}{4}Y_i^2]v_i}{\sum_{i=1}^n \frac{1}{2}Y_i^2v_i}. \quad (2.15)$$

Therefore, according to the experimental value of ρ , the simplest extension of the SM compatible with $\rho = 1$ are models introducing $SU(2)$ singlets and $SU(2)$ doublets. Models with larger multiplets can also be compatible with $\rho = 1$ but tend to be more complex.

2.3.1. Electroweak singlet model

The simplest extension to the scalar sector consists in adding an extra electroweak singlet [20, 21]. After electroweak symmetry breaking, this singlet S mixes with the doublet ϕ predicted by the SM to form two mass states as follows:

$$H_1 = \phi \cos \alpha + S \sin \alpha \quad (2.16)$$

$$H_2 = -\phi \sin \alpha + S \cos \alpha \quad (2.17)$$

where $H_{1,2}$ have mass $m_{1,2}$. By convention, H_1 is taken as the observed scalar boson with a mass of 125 GeV. H_2 could be a lighter or a heavier scalar boson. However, in this search we only focus on $m_2 > m_1 = 125$ GeV.

By definition, the singlet does not interact of its own and therefore H_1 and H_2 only interact through their ϕ component. In other words, in this model, the expected couplings of the 125 GeV mass scalar boson are the same than the one predicted by the SM reduced by a factor $\cos^2 \alpha$.

This simple model can be used as a benchmark model due to its simplicity. It can also easily be extended by adding, for example, new decay channels to H_2 to non-SM particles, making it a candidate for a Higgs-portal mechanism [22], i.e. where the scalar sector would be the only one to interact with new physics.

2.3.2. A class of BSM models: the two-Higgs-doublet models

Still in the simple allowed cases by the measurement of ρ , the two-Higgs-doublet models (2HDM) [23], as the name suggests, introduce two doublets of scalar fields:

$$\phi_1 = \begin{pmatrix} \phi_1^+ \\ \phi_1^0 \end{pmatrix} \quad \phi_2 = \begin{pmatrix} \phi_2^+ \\ \phi_2^0 \end{pmatrix} \quad (2.18)$$

which after symmetry breaking lead to five physical states: two charged scalars H^\pm , one pseudoscalar A and two neutral scalars h and H . In addition to be allowed by the $\rho = 1$ constraint and knowing that there are no strong motivations against an extended scalar sector, the 2HDM are also motivated by the fact that they include the minimal supersymmetric extension of the SM (MSSM) which addresses the hierarchy problem and the coupling unification.

The minima of ϕ_i are $\begin{pmatrix} 0 \\ v_i/\sqrt{2} \end{pmatrix}$ with $\frac{v_2}{v_1} = \tan \beta$ and $v = \sqrt{v_1^2 + v_2^2} = 246$ GeV.

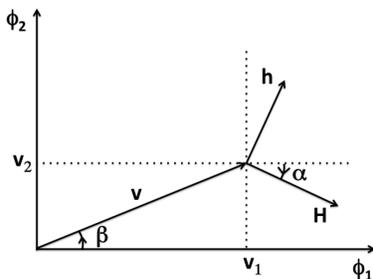


Figure 2.5.: Illustration of the two mixing angles α and β in 2HDM [24].

The angle β is also the rotation angle which diagonalizes the mass-squared matrices of the charged scalars and of the pseudoscalars. Similarly the angle α is the

rotation angle performing the diagonalization of mass-squared matrix of the scalars h and H . Figure 2.5 illustrates the mixing angles α and β between the two scalar fields and their links with the vacuum expectation values.

The 2HDM are divided into models with natural flavor conservation (Type I, Type II, Lepton-specific and Flipped models) and models with tree-level flavor-changing neutral currents (Type III). Focus will be set on models with natural flavor conservation. Such models are built with an additional condition asking that fermions of a certain type (up-type, down-type or leptons) couple to a same scalar doublet. With this condition, one can define four models depending on to which scalar field the fermions are coupled.

By convention, up-type quarks always couple to ϕ_2 in all four models. Type I model is the most SM-like: leptons ℓ , up-type quarks u and down-type quarks d all couple to the same doublet ϕ_2 . Type II model predicts leptons and down-type quarks to couple to ϕ_1 . The lepton-specific model predicts that all quarks couple to ϕ_2 while leptons couple to ϕ_1 . Ultimately, the flipped model predicts leptons and up-type quarks to couple to ϕ_2 while down-type quarks couple to ϕ_1 . The couplings of the scalar doublets with the fermions are summarized in Table 2.1.

	u	d	ℓ
Type I	ϕ_2	ϕ_2	ϕ_2
Type II	ϕ_2	ϕ_1	ϕ_1
Lepton-specific	ϕ_2	ϕ_2	ϕ_1
Flipped	ϕ_2	ϕ_1	ϕ_2

Table 2.1.: 2HDM types with natural flavor conservation and their couplings with fermions.

The intensity of couplings for fermions and bosons to neutral Higgs bosons (h , H and A) are described in Table 2.2. One can see that in all models, the coupling of neutral (pseudo) scalars to vector bosons (V) are the same. In particular, the coupling hVV is the same as the SM one times $\sin(\beta - \alpha)$ while the coupling HVV is the same as the SM one times $\cos(\beta - \alpha)$. There is no tree-level AVV coupling in the CP conserving 2HDM.

As one can see from these couplings, the search for a heavy scalar decaying into two Z bosons presented in this thesis cannot correspond to a pseudoscalar Higgs predicted by those types of CP-conserving 2HDM. It can however correspond to one of the neutral scalar boson h or H . In the SM, only one neutral scalar boson is predicted and a particle compatible with the SM hypotheses within uncertainties was discovered in 2012 with a mass of about 125 GeV. To comply with this discovery, the 2HDM

Particle	Coupling	Type I	Type II	Lepton-specific	Flipped
h	hVV	$\sin(\beta - \alpha)$	$\sin(\beta - \alpha)$	$\sin(\beta - \alpha)$	$\sin(\beta - \alpha)$
	$hu\bar{u}$	$\cos \alpha / \sin \beta$	$\cos \alpha / \sin \beta$	$\cos \alpha / \sin \beta$	$\cos \alpha / \sin \beta$
	$hd\bar{d}$	$\cos \alpha / \sin \beta$	$-\sin \alpha / \cos \beta$	$\cos \alpha / \sin \beta$	$-\sin \alpha / \cos \beta$
	$h\ell\bar{\ell}$	$\cos \alpha / \sin \beta$	$-\sin \alpha / \cos \beta$	$-\sin \alpha / \cos \beta$	$\cos \alpha / \sin \beta$
H	HVV	$\cos(\beta - \alpha)$	$\cos(\beta - \alpha)$	$\cos(\beta - \alpha)$	$\cos(\beta - \alpha)$
	$Hu\bar{u}$	$\sin \alpha / \cos \beta$	$\sin \alpha / \cos \beta$	$\sin \alpha / \cos \beta$	$\sin \alpha / \cos \beta$
	$Hd\bar{d}$	$\sin \alpha / \cos \beta$	$\cos \alpha / \sin \beta$	$\sin \alpha / \cos \beta$	$\cos \alpha / \sin \beta$
	$H\ell\bar{\ell}$	$\sin \alpha / \cos \beta$	$\cos \alpha / \sin \beta$	$\cos \alpha / \sin \beta$	$\sin \alpha / \cos \beta$
A	AVV	0	0	0	0
	$Au\bar{u}$	$\cot \beta$	$\cot \beta$	$\cot \beta$	$\cot \beta$
	$Ad\bar{d}$	$-\cot \beta$	$\tan \beta$	$-\cot \beta$	$\tan \beta$
	$A\ell\bar{\ell}$	$-\cot \beta$	$\tan \beta$	$\tan \beta$	$-\cot \beta$

Table 2.2.: Couplings of fermions and bosons to neutral (pseudo) scalar bosons h , H , A in the four CP-conserving 2HDM compared to the SM ones. Yukawa couplings to the charged scalars can be determined from the couplings to the neutral pseudoscalar.

have to lie in a region where one of the neutral scalar tends to be SM-like. And since no other scalar boson has been discovered below 125 GeV, one assumes that h should be SM-like. Two scenarii exist: the decoupling and the alignment limits [25].

The decoupling limit is the region of parameter space in which H , A and H^\pm are all much heavier than h . Therefore it is possible to integrate out those heavy fields, resulting in an effective theory which is SM-like with corrections due to the various couplings from the heavy sector.

The alignment limit requires the whole vacuum expectation value v to lie in the neutral component of only one of the scalar doublets, for example $v \simeq v_1$ and $v_2 \simeq 0$. In this case, the mixing between the neutral scalar bosons h and H states disappears.

From Table 2.2, one can see that requiring the decoupling or the alignment limit where one neutral scalar boson (here h) is close to the observed SM-like scalar boson means that the W and Z bosons dominantly acquire their masses through their couplings with h . Therefore, to be SM-like, $\cos(\beta - \alpha)$ needs to be small.

2.4. Searches for BSM physics and constraints on 2HDM

There are several complementary ways to explore BSM physics in the scalar sector:

- **precision measurements:** the 125 GeV scalar boson discovered at LHC could not be the SM one but a BSM one, coming from example from 2HDM. Precision measurements of the production cross section and branching ratios of the observed scalar boson are therefore needed. From those measurements, we already know that only the alignment and decoupling limits are allowed in the 2HDM since the observed scalar boson was found to be SM-like within uncertainties. The state-of-the-art constraints on 2HDM parameters will be summarized in next section.
- **direct discovery of new scalar bosons:** the discovery of a new scalar particle would be a direct evidence of BSM physics. This is the goal of this thesis: the search for a BSM scalar boson, heavier than 125 GeV. In particular, in the $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ group, we searched for a scalar boson decaying into two Z bosons: one of them decaying into a pair of charged leptons $\ell^+\ell^-$ and the other in a pair of neutrinos $\nu\bar{\nu}$.

To illustrate the complementary of the direct search and the precision measurements, Fig. 2.6 shows the exclusion limits at 95% CL⁴ as a function of the new heavy scalar mass on the parameter $C'^2 = \sin^2 \alpha$ of the extra electroweak singlet model [2]. These results are under the assumption that the new heavy scalar does not decay to any new particles and that its ggF and VBF production modes contribute similarly to the total production cross section as in the SM case. This analysis was done with data collected during the run 1 (2010-2012) of the LHC by CMS, it corresponds to the combination of collected data at a colliding center-of-mass energy of $\sqrt{s} = 7$ TeV (5.0 fb^{-1}) and at $\sqrt{s} = 8$ TeV (19.6 fb^{-1}).

Still using run 1 data, Fig. 2.7 shows the exclusion contours at 95% CL in the Type-I (left) and Type-II (right) 2HDM scenarii [26]. The exclusion contour coming from the precision measurements of the scalar boson with a mass of 125 GeV is shown in pink. Due to the couplings of the 2HDM Type-I case, its exclusion power is less stringent than in the 2HDM Type-II case and it is not visible in the scale chosen on the left figure. Other colored contours represent exclusion coming from direct searches for a 2HDM heavy scalar.

The $H \rightarrow ZZ \rightarrow \ell\bar{\ell}\nu\bar{\nu}$ decay channel studied in this thesis presents several advantages:

- if the BSM scalar searched for is described by 2HDM, then its couplings to vector bosons are SM-like reduced by a factor $\cos(\beta - \alpha)$. Therefore, under this assumption, one can see on Fig. 2.4 that $H \rightarrow WW$, $H \rightarrow ZZ$ and $H \rightarrow t\bar{t}$ are

⁴A more detailed description of setting up exclusion limits is given in 5.6.4.

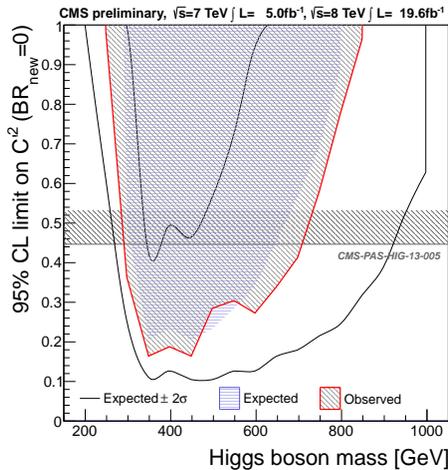


Figure 2.6.: Exclusion limits at 95% CL as a function of the new heavy scalar mass on the parameter $C'^2 = \sin^2 \alpha$ of the extra electroweak singlet model using LHC run 1 data collected by CMS. Indirect limits obtained from the precise measurements of the scalar boson with a mass of 125 GeV are represented by an horizontal line. Expected and observed limits coming from direct search of a new heavy scalar in the $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ channel are represented in blue and red respectively [2].

likely to be the dominant decays of a heavy scalar boson, even when $\cos(\beta - \alpha)$ is small as in the alignment limit.

- the choice of decay channel of the two Z into two charged leptons and two neutrinos present the advantage to have a clear signature for the $Z \rightarrow \ell\bar{\ell}$ decay while having a bigger branching ratio than $ZZ \rightarrow 4\ell$ due to the $Z \rightarrow \nu\bar{\nu}$ decay. This channel is sensitive in an intermediate mass region (between around 300 GeV and 1000 GeV). Below, there are too many backgrounds and a signal with a clearer signature such as $ZZ \rightarrow 4\ell$ is better, while at higher mass the $ZZ \rightarrow 2\ell 2q$ signal which has a less clean signature but a higher cross section gives stringent limits since there are less to no backgrounds in this region. In the end, all those channels are combined together as shown in [27] — which is the paper combining all three channels for data collected by CMS at the LHC in 2016, including the $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ channel on which I worked on and that is presented in this thesis.

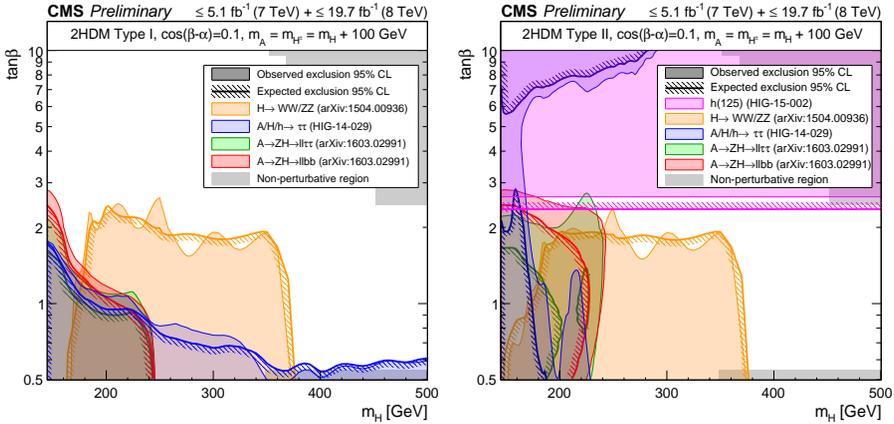


Figure 2.7.: Exclusion contours at 95% CL in the Type-I (left) and Type-II (right) 2HDM scenarios in the $\tan\beta - m_H$ plane. The colored filled areas correspond to the observed exclusion contours while the colored lines with hatches correspond to the expected exclusion contours. The exclusion contour coming from the precision measurements of the scalar boson with a mass of 125 GeV is shown in pink. The gray shaded areas correspond to the regions where each corresponding scenario is either non-perturbative or unstable. The other colored contours represent exclusion coming from direct searches for a 2HDM heavy scalar [26].

2.4.1. Current constraints on 2HDM

The free parameters of 2HDM are constrained by theory and experience as detailed below [28, 29].

Theoretical Constraints

- **Vacuum Stability:** this implies that the scalar potential should be bounded from below, putting various constraints on the quartic couplings and therefore on the scalar masses and mixing angles [30].
- **Unitarity:** in the SM, the scattering cross section for longitudinally polarized W bosons is unitary only if the scalar boson exchange diagrams are included. The modification of those couplings by 2HDM have still to comply with unitarity.
- **Perturbativity:** tools used to compute scalar couplings require to stay in the perturbative regime, translating into bounds on parameters.

Experimental Constraints

Figure 2.8 shows the available phase space region for a Type-II 2HDM scenario in the $\tan\beta - \sin(\beta - \alpha)$ region after applying those theoretical constraints combined with searches performed in multiple channels by the LEP, the Tevatron and the LHC collaborations [29]. Experimental constraints coming from the LHC are using run 1 data. The regions excluded by theoretical constraints are shown in dark blue. The light blue region corresponds to Z -pole precision measurements done at the LEP and in particular $\Delta\rho$ — the deviation from $\rho = \frac{m_W^2}{m_Z^2 \cos^2 \theta_W}$. Constraints coming from direct searches for scalar bosons are shown in green for the LEP, and in yellow for the Tevatron and the LHC collaborations. The orange region corresponds to precision measurements done at the LEP to study Zbb couplings, the later being influenced by potential charged scalar bosons. Eventually the regions passing all those exclusion limits are shown in red. However, this region can be considerably reduced if one considers the 125 GeV scalar boson and its properties as the light CP-even scalar boson from 2HDM. In this case the available region narrows down to the dark red area. Ultimately, the regions enclosed by black curves corresponds to the addition that no flavour changing neutral current are allowed.

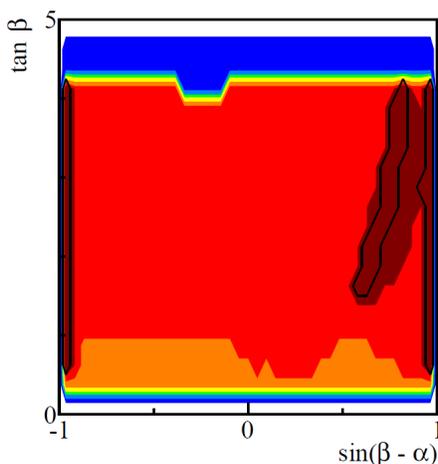


Figure 2.8.: Available phase space region for a Type-II 2HDM scenario in the $\tan\beta - \sin(\beta - \alpha)$ region after applying theoretical constraints combined with searches performed in multiple channels by the LEP, the Tevatron and the LHC (using run 1 data) collaborations [29]. The final allowed region is enclosed in black curves. More details are given in the text.

More recent results coming from the CMS collaboration [31] are shown on Fig. 2.9. This figure shows experimental constraints in the $\tan\beta - \cos(\beta - \alpha)$ plane coming from the precision measurement of the 125 GeV-mass scalar boson assuming it is the lighter scalar in the scenarii of Type-I and Type-II 2HDM. This analysis was performed by the CMS collaboration using data taken in 2016 at 13 TeV. The lobe feature which can be seen in Type-II model at $\cos(\beta - \alpha) > 0$ is due to negative values of the couplings $hd\bar{d}$ and $h\ell\bar{\ell}$ in this scenario which are not excluded with the current sensitivity.

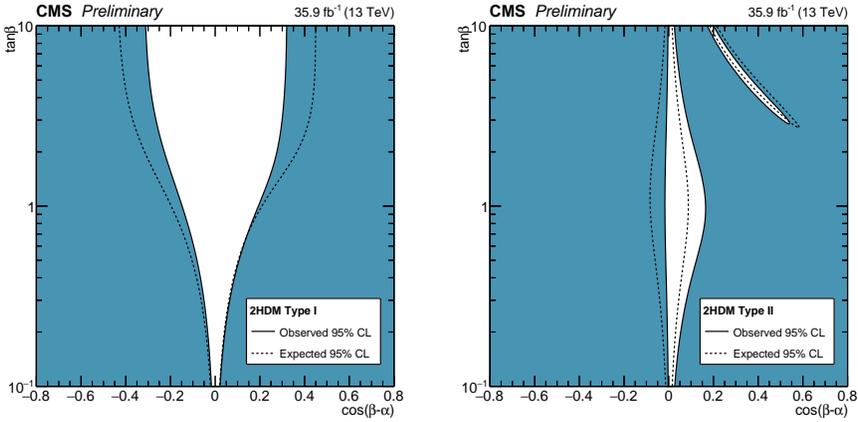


Figure 2.9.: Constraints in the $\cos(\beta - \alpha) - \tan\beta$ plane for Type-I (left) and Type-II 2HDM (right) from CMS 2016 data, obtained from the precision measurement of the 125 GeV-mass scalar boson assuming it is the lighter scalar of 2HDM. The white regions, bounded by the solid black lines, represent the regions of the parameter space which are allowed at 95% CL [31].

Chapter 3

The CMS experiment at the LHC

One of the possible ways to probe the SM and BSM scalar sector is by using data coming from collisions of particles at high energy. The European Organization for Nuclear Research (CERN) built the largest complex of particle accelerators to this day, the latest ring of this complex being the Large Hadron Collider (LHC). This complex and in particular the LHC are presented in the first part of this chapter. The second part of this chapter is dedicated to the CMS experiment: one of the experiments collecting data from collisions occurring at the LHC and providing the data analyzed in this thesis.

3.1. The Large Hadron Collider

The Large Hadron Collider (LHC) [32] is the largest circular particle collider to this day. Built close to Geneva by the CERN, it lies underground in the tunnel previously built for the LEP, at a depth of about 100 m and has a circumference of 26.7 km. Its first collisions happened in 2009. This thesis is presenting an analysis on data taken from proton-proton collisions in 2016 by the CMS experiment at the LHC, therefore focus will be set on this period and this type of collision.

3.1.1. The proton production, the injector complex and the LHC

The LHC is only the final accelerator for the protons. Before injection of the protons into this accelerator it is necessary to use an entire set of smaller accelerators of different kinds [33] as shown on Fig 3.1.

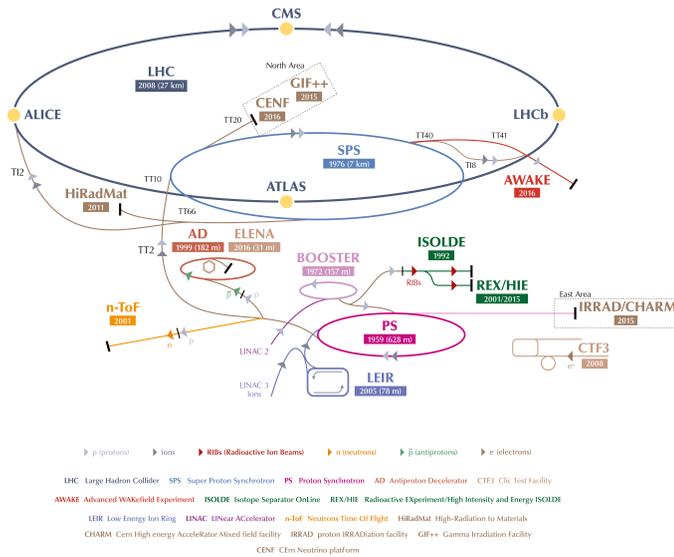


Figure 3.1.: Sketch of the CERN accelerator complex and the main experiments/area using the beams produced by those accelerators at different level of energy and intensity. Before reaching the LHC protons successively pass through the LINAC2, the Booster, the PS and the SPS. As one can see many experiments use the particle beams produced at CERN, the LHC experiments being the experiments with beams at the higher intensity and energy [34].

The very first step of the chain consists in the setup of a proton source. The source at the LHC is coming from a hydrogen gas surrounding a heated cathode filament. This heated filament emits electrons which ionize the hydrogen gas to form a electrons-protons plasma. The protons are then separated from electrons by applying an electric field, forming a first proton beam of an energy of about 100 keV. Those

protons are then accelerated and focalized in a Radio Frequency Quadripole (RFQ), increasing their energy from 100 keV to 750 keV.

After this first acceleration, protons are injected in the LINAC2: a 30 m-long linear accelerator made of radiofrequency cavities increasing progressively the energy of protons to 10, 30 and eventually 50 MeV. Once at this energy, protons are injected into the Proton Synchrotron Booster: a circular accelerator made of four superimposed synchrotron rings with a 25 m-radius. In the Booster, protons reach an energy of 1.4 GeV and a density of 1.38×10^{12} protons per ring. Those four proton bunches are then injected in the Proton Synchrotron (PS).

The PS is a circular accelerator with a 100 m-radius that increases the energy of proton bunches to 26 GeV. It receives two fills from the Booster for a total of eight bunches. This filling is part of a novelty for 2016 named the Batch Compression Merging and Splitting (BCMS) scheme [35]. The major improvement of this filling scheme is the reduction of the beam transverse size to about $2.5 \mu\text{m}$ (versus $3.5 \mu\text{m}$ before). This increase in beam density from the injectors leads to an increase in the number of collisions in the LHC and hence in an increase of about 20% in peak luminosity (see next section). The PS, while also increasing the energy of the proton bunches, groups the eight bunches by two and then splits them successively to reach 48 bunches of about 1.15×10^{11} protons (vs 72 bunches for the nominal filling scheme). Those 48 proton bunches are then injected into the Super Proton Synchrotron (SPS): a 6.9 km-long circular accelerator. This process is repeated twice to fill the SPS with a total of 96 bunches.

Under nominal condition, this cycle should have been repeated several times to fill completely the SPS with 288 bunches. However, in 2016, the presence of a vacuum leak in the SPS high energy beam dump limited the intensity that could be accelerated [36]. Those proton bunches are accelerated in the SPS to an energy of 450 GeV and then injected into the LHC where they undergo their final acceleration to reach an energy of 6500 GeV.

The LHC is made of two 26.7 km-long parallel rings containing proton bunches moving in opposite direction. To fill those rings, the full chain described above has to be reproduced 24 times per ring. In the end, the nominal number of colliding bunches recorded in 2016 in the LHC by the CMS experiment is 2208. Those bunches are split in batches containing 96 bunches separated by 25 ns. Therefore collisions happen every 25 ns except between two batches where the gap between collisions varies between 225 ns and $3 \mu\text{s}$.

3.1.2. Luminosity and data-taking condition at the LHC in 2016

Key specifications for a particle accelerator are the energy of collisions and the event rate — or luminosity \mathcal{L} — that can be reached. High energy collisions are necessary to be able to create new heavy particles while a high luminosity is needed to obtain enough collisions to observe events N with small cross sections σ :

$$\frac{dN}{dt} = \mathcal{L}\sigma. \quad (3.1)$$

Integrating this formula makes the link between theory expectation (through the cross section) and experiment (through the numbers of observed events), knowing the specifications of the LHC.

The luminosity is defined as follow from the machine point of view:

$$\mathcal{L} = \frac{f_r N^2 n_p}{4\pi\sigma_x\sigma_y} F, \quad (3.2)$$

where f_r is the beam revolution frequency, N the number of colliding bunches in one ring, n_p the number of protons in one bunch, $\sigma_{x,y}$ are the transverse size of the beam in the x and y direction and F a reduction factor coming from the crossing angle of the two beams. The product $\sigma_x\sigma_y$ can be rewritten as $\frac{\sqrt{\epsilon_x\epsilon_y\beta_x\beta_y}}{\gamma}$ with γ the Lorentz boost of the proton bunches in the laboratory, $\epsilon_{x,y}$ the normalized emittance in the x or y direction (which measures the average spread of the proton coordinates in position-momentum phase space inside a bunch) and $\beta_{x,y}$ the betatronic oscillation function parameter (which corresponds to the distance between the point where the bunches cross and the point where the transverse size of the bunches is twice as large).

The luminosity recorded in 2016 is shown on Fig. 3.2. The peak luminosity achieved is $1.53 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, going beyond the designed value of $1.0 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ firstly by using smaller beams from the injectors (BCMS scheme) and then via a reduction in the angle at which the beams cross at the interaction points of CMS. In addition, the LHC managed to reach about 50 % availability, leading to an integrated delivered luminosity to CMS of 40.82 fb^{-1} as shown on Fig. 3.3.

Over those 40.82 fb^{-1} delivered by the LHC in 2016, CMS recorded 37.76 fb^{-1} and 35.92 fb^{-1} were validated for physics analysis and are therefore used in this thesis [39].

An important parameter to monitor when running such high luminosity experiment is to look at the number of collisions occurring per bunch crossing (pileup). Figure 3.4 shows the distribution of pileup recorded by CMS in 2016. On average, a pileup of 27 was observed while a few events with more than 50 events of pileup were

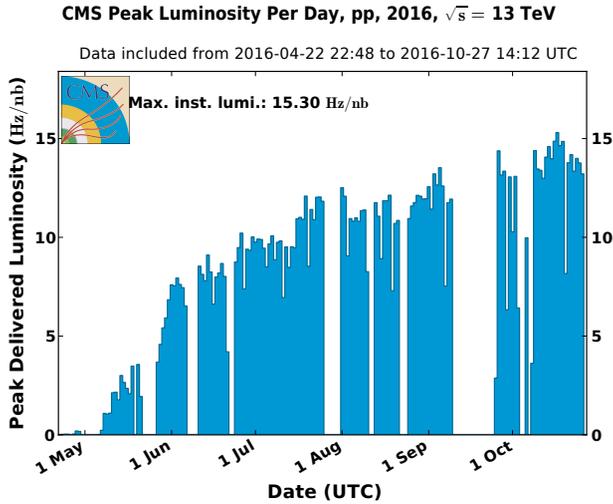


Figure 3.2.: LHC peak luminosity per day in 2016 [37].

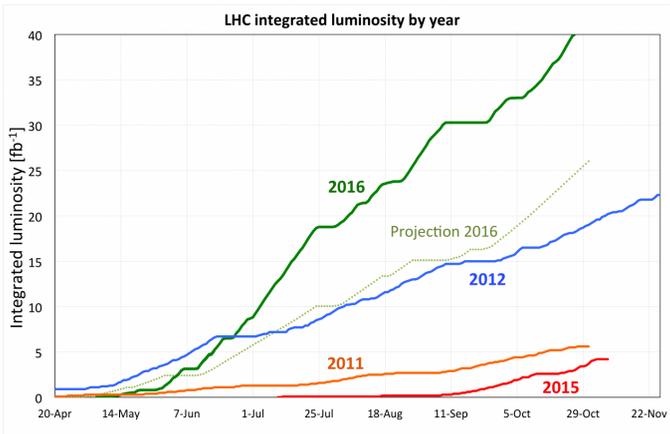


Figure 3.3.: LHC integrated luminosity in 2011, 2012, 2015 and 2016 and the prediction of the 2016 performance foreseen at the start of the year [38].

observed in 2016. The pileup is directly influence by the instantaneous luminosity and in particular by the luminosity per bunch, which raised by a factor 1.4 with the implementation of the BCMS scheme.

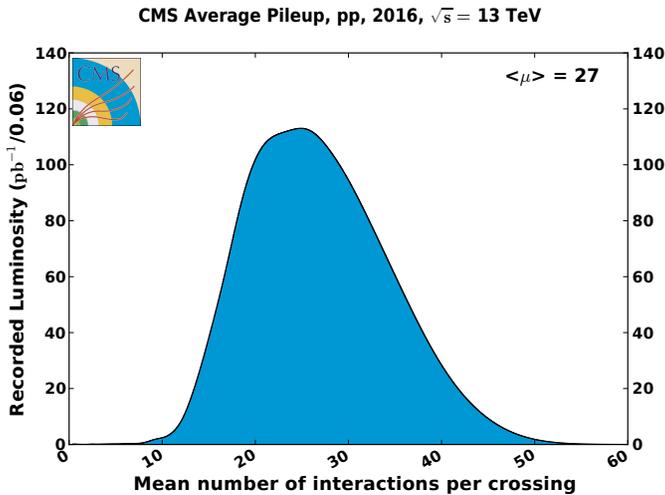


Figure 3.4.: Mean number of interactions (pileup) per bunch crossing recorded by CMS at the LHC in 2016 [39].

3.2. The CMS experiment

The Compact Muon Solenoid (CMS) detector [40] is one of the four main detectors installed along the ring of the LHC. It is a general purpose detector designed to perform a wide range of physics research, including both SM precision measurements and searches for new physics. To perform such research the CMS detector has been built to achieve a good muon identification and momentum resolution of a few percent, a good charged particle reconstruction efficiency and momentum resolution of a few percent, a good electromagnetic energy resolution of a few percent and the best possible jet energy resolution compatible with the budget constraints and the technology choices made to reach the aforementioned performance. The detector gets its name from its limited size considering its complexity, its sophisticated muon system, and its solenoid superconducting magnet which encompasses all the inner tracking system and calorimeters.

This section will describe the CMS detector, going through all its sub-detectors and their performance during data-taking at the LHC in 2016.

3.2.1. Overview of the CMS detector

The CMS detector is a 21.6 m long cylinder with a diameter of 15 m for an approximate weight of 14 000 t. Its overall layout, shown in Fig. 3.5, is divided into a barrel and two endcaps, where the sensitive elements are respectively placed parallel

and perpendicular to the beam pipe. The CMS detector consists of four sub-detectors, each of them designed for the detection of specific particle signatures. Those sub-detectors are, from the innermost to the outermost: the inner tracking system, the electromagnetic and hadronic calorimeters and the muon detection system. The inner tracking system lies within a radius of 1.25 m around the beam pipe and is made of layers of silicon pixel and silicon strip detectors. It is designed to provide the measurement of about 10 points on the trajectory of any charged particle with a spatial resolution ranging from about $10\ \mu\text{m}$ to $40\ \mu\text{m}$. Around the inner tracking system are the electromagnetic and hadronic calorimeters which measure the energy of electrons, photons and hadrons through the interaction of the latter within the calorimeters. These sub-detectors are placed inside a superconducting solenoid which produces an internal uniform magnetic field of 3.8 T oriented along the direction of the beams and which bends the trajectory of charged particles in the transverse plane allowing precise transverse momentum measurements. The flux of the magnetic field is closed by a steel yoke in which three different technologies of muon detectors are placed. The data analysis presented in this thesis, i.e. the study of $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ processes, depends on all those sub-detectors.

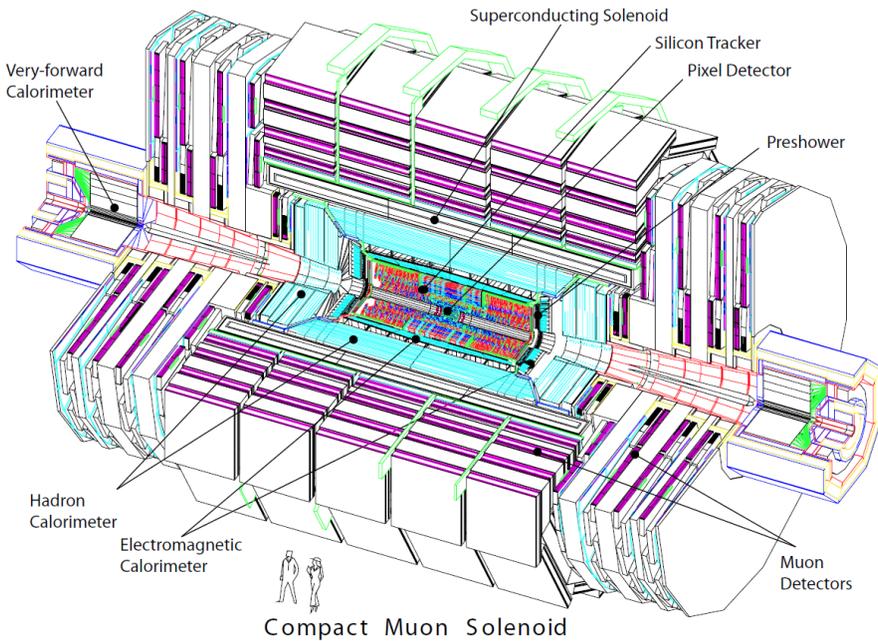


Figure 3.5.: A perspective view of the Compact Muon Solenoid [40].

3.2.2. The coordinate system

The origin of the coordinate system used in CMS is the nominal interaction point of the colliding beams. This coordinate system is a right-handed system where the y -axis is pointing upwards, the x -axis is pointing towards the center of the LHC, and the z -axis is aligned with the beam direction. The $x - y$ plane is referred to as the transverse plane. A spherical coordinate system (ρ, θ, ϕ) is often used with the azimuth angle ϕ defined in the transverse plane, starting at the x -axis, and the polar angle θ defined from the z -axis, in a plane orthogonal to the transverse plane and containing this z -axis. This polar angle θ is used to define the more widely used pseudorapidity η :

$$\eta = \log(\tan \theta/2). \quad (3.3)$$

The pseudorapidity is a good approximation of the rapidity y for ultra-relativistic particles with $E \gg m$:

$$y = \frac{1}{2} \log \left(\frac{E + p_z}{E - p_z} \right). \quad (3.4)$$

The difference in rapidity between two particles Δy is invariant under a Lorentz boost in the z -direction. This is an important property since the partons interacting during the proton-proton collision can have a big momentum difference along the z -axis (see section 4.2.1). Also the flux of particles produced during a collision is approximately uniform in rapidity [41].

The transverse plane is also defined in a polar coordinate system (r, ϕ) , with ϕ defined as above and $r = \sqrt{x^2 + y^2}$.

3.2.3. The superconducting solenoid magnet

A central element of the CMS detector is the superconducting solenoid magnet. It is made of a coil placed in the barrel of the CMS detector, between the inner tracking system and the calorimeters on the one hand, and the muon detectors on the other hand. This coil is 12.5 m long and has a diameter of 6 m. It provides a uniform constant magnetic field of 3.8 T inside the coil, pointing along the z -axis. The flow of the magnetic field is closed by a steel yoke in which an average magnetic field of 1.8 T is present. A simulation of the magnetic field in the components of the magnet is shown on Fig. 3.6. In this figure, the length scale is given in meters along the z -axis. The inner tracker ($r < 1.25$ m and $|z| < 2.9$ m) is immersed in a uniform field. The field intensity in the return yoke, of 1.8 T average value, is also visible in green. The coil is made of niobium-titanium wires, a superconducting alloy needed to allow electricity to flow without resistance and therefore allowing high intensity current of up to 19.14 kA (at the designed value of 4 T), giving a stored energy of

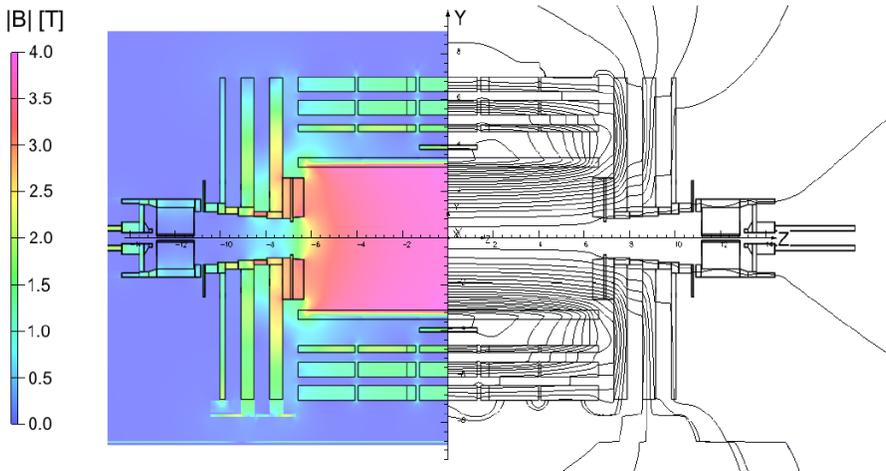


Figure 3.6.: Value of $|B|$ (left) and field lines (right) predicted on a longitudinal section of the CMS detector, at a central magnetic flux density of 3.8 T. Each field line represents a magnetic flux increment of 6 T m^2 [42].

2.6 GJ. To reach the superconducting state the magnet is cooled down to 4.5 K with liquid helium.

The coil contains the tracker and the calorimeters, which presents some advantages. As such, the particles need not go through the coil before interacting with the calorimeters, allowing for more precise measurements in the latter. Moreover, having the coil in front of the first muon detectors decreases the probability that electrons and hadrons reach them.

The 3.8 T magnetic field bends the trajectory of charged particles. The relation between the transverse momentum (p_T) of a charged particle and the radius of curvature as its track projected on the plane perpendicular to the magnetic field is:

$$p_T = 0.3BR \quad (3.5)$$

where p_T is expressed in GeV, B is the intensity of the magnetic field in Tesla and R the radius of curvature in meters.

3.2.4. The tracker

A key point in reconstructing events coming from a collision is the high-resolution reconstruction of the paths of charged particles coming from the primary interaction vertex, and the precise measurement of the latter. This justifies the need for a detector close to the beam pipe with high granularity and fast response time to unambiguously

assign particles to the correct collision. The CMS collaboration is achieving this by using a compromise between two kinds of silicon detectors. The silicon pixel detector is the closest detector installed around the beam pipe, making it subject to a high flux of particles (around 1 MHz mm^{-2} at 4 cm from the interaction point) and requiring it to be able to cope with a high-radiation environment while disentangling hits from charged particles. Further away from the interaction point, the rate is lower (between 60 kHz mm^{-2} at 22 cm and 3 kHz at 115 cm) and micro-strip detectors are used instead in order to reduce the budget.

The whole tracking system has a length of 5.8 m and a diameter of 2.5 m. It consists of a central barrel and two endcaps to close the cylinder. Its structure is optimized so as to detect 12–14 hits in average per track, ensuring both a high efficiency in reconstructing tracks and a low rate of misidentified tracks. Also, due to redundancy, the tracking system is less sensitive to sporadic failures of some of its components.

The top half tracker geometry in the r - z plane is shown on Fig. 3.7 both for the pixel detector and the micro-strips detector (here grouped under TIB, TID, TOB and TEC subsystems, defined in the “Silicon Strip Detector” sub-section below). In the barrel region, the tracker is composed of three layers of silicon pixel detectors, and of ten layers of silicon micro-strip detectors. The system is completed in the endcaps by two disks in the pixel detector and three plus nine disks in the strip detector, for pseudorapidities up to $|\eta| < 2.5$. In total, the inner tracker consists of 1440 silicon pixel and 15 148 silicon strip detector modules reaching a total active area of $\sim 200 \text{ m}^2$ which is the largest silicon detector ever built at the time of writing this thesis. The tracker’s simulated material budget is shown on Fig. 3.8 in units of radiation length χ_0 and nuclear interaction length λ_I . The radiation length is both the mean distance over which a high-energy electron loses all but $1/e$ — or 0.368 — of its energy by bremsstrahlung, and $7/9$ of the mean free path for pair production by a high-energy photon. The nuclear interaction length is the mean distance traveled by a hadronic particle before undergoing an inelastic nuclear interaction. Simulating the material budget of the detector is important in order to have a good modelling of the multiple scattering (for the track fitting) and of the interactions of electrons, photons and hadrons before they reach the calorimeters.

The Silicon Pixel Detector

The silicon pixel detector is used in CMS as the starting point for the reconstruction of the charged particle tracks and is essential for the reconstruction of the primary interaction point (primary vertex) and displaced decays (secondary vertices) from long-lived unstable particles, like bottom or charm mesons. It is the closest detector around the beam and covers the region $|\eta| < 2.5$. It consists of a central part

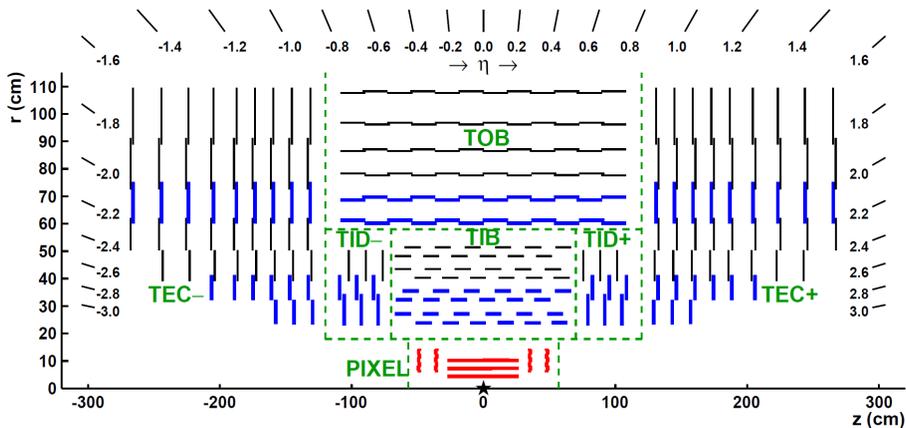


Figure 3.7.: Schematic cross section through the CMS tracker in the r - z plane. The tracker is symmetric about the horizontal line $r = 0$. The center of the tracker, corresponding to the approximate position of the proton-proton collision point, is indicated by a star. Green dashed lines help the reader understand which modules belong to each of the named tracker subsystems. Strip tracker modules that measure a single coordinate are shown by thin black lines, while those permitting the reconstruction of two coordinates are shown by thick blue lines. The pixel modules, shown by the red lines, also provide two coordinates [43].

made of three concentric cylindrical 53 cm-long layers (BPix, for Barrel Pixel) placed at a distance r from the beam axis of 4.4 cm, 7.3 cm and 10.2 cm, and four endcap disks (FPix, for Forward Pixel), two on each side at $z = \pm 34.5$ cm and $z = \pm 46.5$ cm covering radii between 6 cm and 15 cm. This geometry ensures that each particle emitted from the nominal interaction point will pass through three layers of the pixel detectors for $|\eta| < 2.1$, or two layers in the region $2.1 < |\eta| < 2.5$.

The barrel pixel tracker contains 768 modules for a total of 48 million pixels and covers an active area of 0.78 m^2 . The disks of the endcaps are divided into 24 segments (or blades) arranged in a turbine-like geometry as shown on Fig 3.9. Each FPix blade sums up to 672 modules and a total of 18 million pixels to cover an active area of 0.28 m^2 . The pixel sensors are silicon semiconductors with an area of $100 \times 150 \mu\text{m}^2$ and a thickness of $285 \mu\text{m}$. This geometry and granularity allows to keep the occupancy lower than 1% per pixel and per LHC bunch crossing.

During the data-taking at LHC in 2016 the pixel detector worked extremely well, especially considering the LHC delivered collisions with an instantaneous luminosity going beyond original design. The number of components presenting failures was

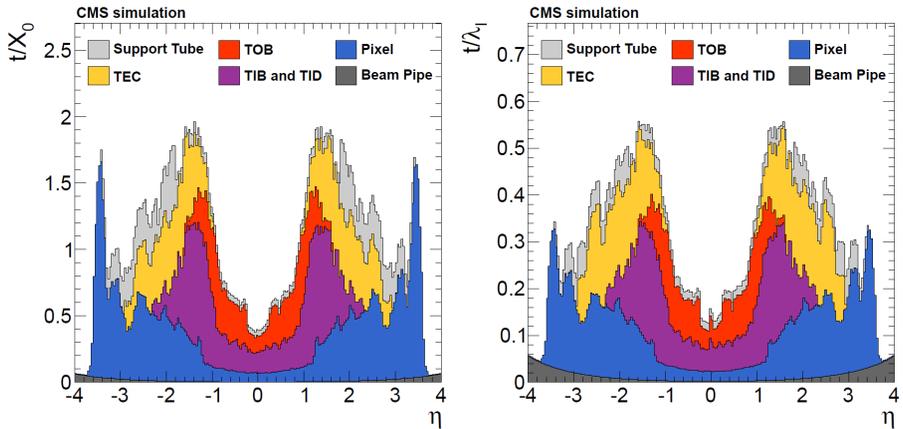


Figure 3.8.: Simulation of the total thickness t of the tracker material traversed by a particle produced at the nominal interaction point, as a function of pseudorapidity η , expressed in units of radiation length χ_0 (left) and nuclear interaction length λ_I (right). The contribution to the total material budget of each of the subsystems that comprise the CMS tracker is shown, together with contributions from the beam pipe and from the support tube that surrounds the tracker [43].

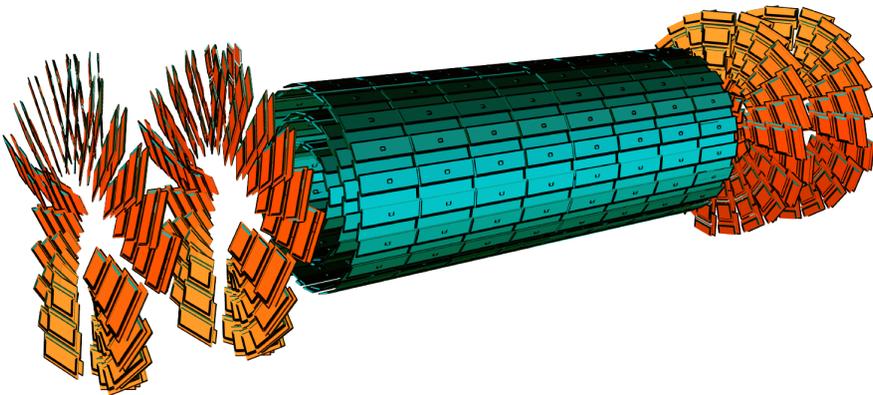


Figure 3.9.: Sketch of the CMS pixel detector. The three layers of the barrel pixel are shown in blue and the four forward disks in orange. Each box represents one module [44].

small and stable throughout the year and most of the issues were discrete and recovered quickly. The mean number of bad components in 2016 was 1.63% for the BPix and 0.22% for the FPix. The pixel hit efficiency stayed above 99% for all layers and disks except for the first layer of the BPix where dynamic inefficiency was expected at high instantaneous luminosity due to the design of the read-out chip (see Fig. 3.10). The hit resolution was of the order of $10\ \mu\text{m}$ in the transverse plane and $25\ \mu\text{m}$ along the beam axis in the BPix. In the FPix the hit resolution in the radial direction is of the order of $20\ \mu\text{m}$. Figure 3.11 shows the evolution of the hit resolution in the BPix and FPix as a function of the average instantaneous luminosity for two reconstruction algorithms [45].

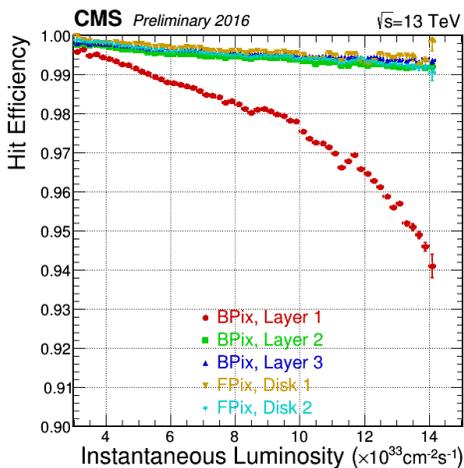


Figure 3.10.: Hit Efficiency as a function of instantaneous luminosity in 2016 in LHC filling schemes with more than 2000 colliding bunches [46].

The Silicon Strip Detector

Starting at a radius of 20 cm the flow of particles is getting low enough to allow the use of silicon micro-strip detectors instead of pixel sensors. The silicon strip detector is divided into three subsystems, ranging from an inner radius of 20 cm to an outer radius of 116 cm, and consists of almost 15 000 modules made of either 512 or 768 strips for a total of 9.6 million read-out channels. The Tracker Inner Barrel and Disks (TIB and TID) are composed of four layers in the barrel and three disks in each endcap, covering a region up to a radius of 55 cm (see Fig. 3.7). The sensors are $320\ \mu\text{m}$ thick and run parallel to the beam pipe in the barrel and radially in the disks.

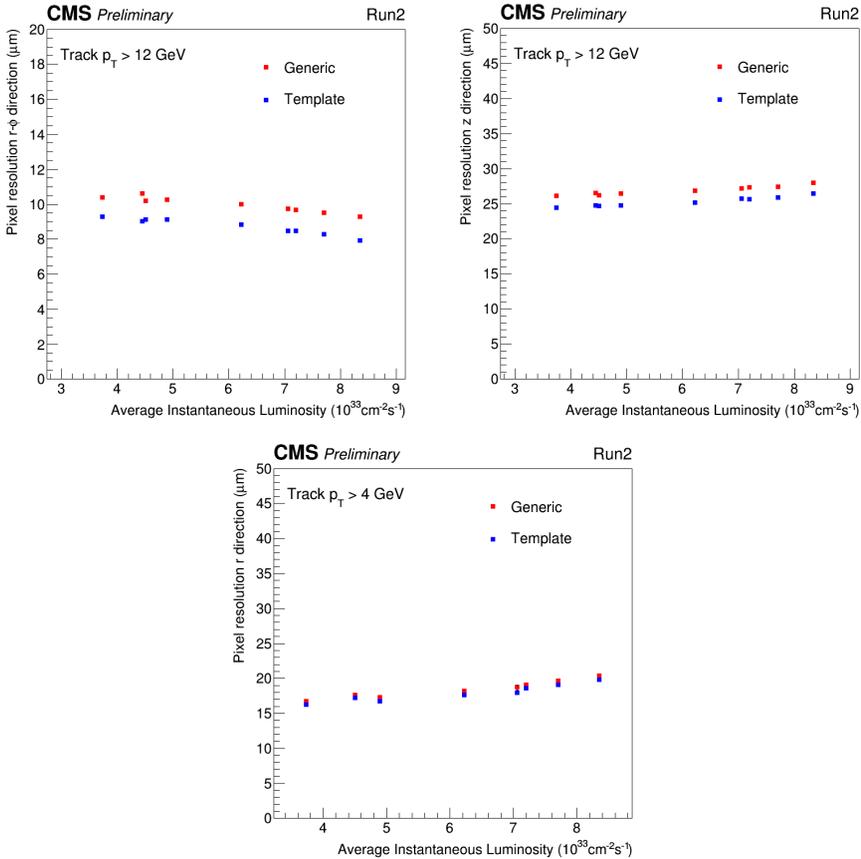


Figure 3.11.: Pixel detector hit resolution as a function of the average instantaneous luminosity for the BPix along $r - \phi$ direction (top left) and z (top right), and for the FPix along the r direction (bottom). The *generic* algorithm is a fast online algorithm while the *template* algorithm is a more detailed algorithm used by offline reconstruction [45].

In this region, a typical cell size is $10\text{ cm} \times 80\ \mu\text{m}$, leading to an occupancy of up to 2–3% per strip and LHC bunch crossing. The strip pitch is of $80\ \mu\text{m}$ in the two first layers of the TIB and $120\ \mu\text{m}$ in the two next ones, resulting in a spatial resolution respectively of the order of $18\ \mu\text{m}$ and $28\ \mu\text{m}$. Figure 3.12 shows the hit resolution in the inner and outer barrel detectors depending on the sensor thickness and strip pitch. In the TID the pitch varies between $100\ \mu\text{m}$ and $141\ \mu\text{m}$.

The TIB and TID are surrounded by the Tracker Outer Barrel (TOB) which consists of six layers of $500\ \mu\text{m}$ thick sensors extending to 116 cm away from the beam

pipe. Due to the larger areas that have to be instrumented in this region these sensors are longer (up to 25 cm) compared to the TIB/TID ones in order to limit the number of read-out channels and thicker to keep a good signal to noise ratio. They provide a resolution of the order of 40 μm for the four first layers and of the order 22 μm for the last two layers with an occupancy of about 1%. The longitudinal coverage of the TOB reaches $z = \pm 118\text{ cm}$, beyond which, closing the TOB on both sides, the Tracker Endcaps (TEC) are found. Each TEC is a collection of 9 disks covering the region $124 < |z| < 282\text{ cm}$ and $0.9 < |\eta| < 2.5$.

Additionally, as can be seen from Figure 3.7, the modules of the two inner layers of the TIB and TOB, the two inner rings of the TID and TEC, and the fifth ring of the TEC are equipped with a second strip detector mounted back-to-back with a crossing angle (also called ‘stereo angle’) of 100 mrad in order to measure the second coordinate (z in the barrel and r on the disks).

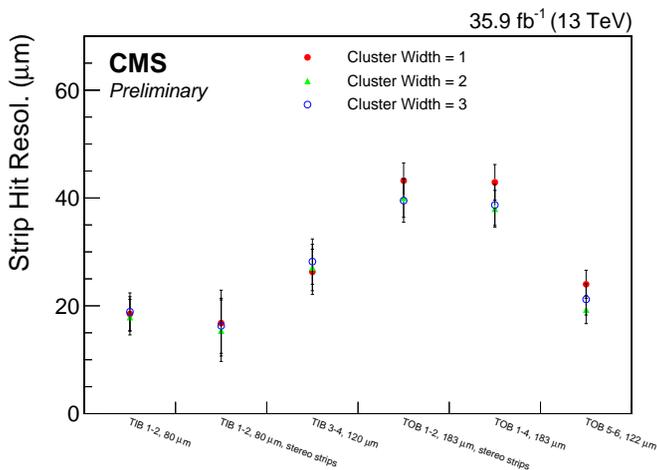


Figure 3.12.: Strip detector hit resolution derived with the pair method by selecting pairs of hits in different types of overlapping sensors and for different cluster widths expressed in units of number of strips [47].

Charged particle track reconstruction in the CMS tracker

Charged particles going through the tracker interact with the silicon material, generating specific detector hits that are recorded. The trajectory of those charged particles is reconstructed by combining the detector hits to form tracks. Due to the uniform axial magnetic field, the charged particles will have a helical trajectory inside the inner tracking system. Therefore an iterative helical track model is used for recon-

struction [43]. Starting from the reconstructed hits, the iterative track reconstruction procedure is composed of four logical steps: seed generation, pattern recognition (or trajectory building), ambiguity resolution and final track fit.

1. **Seed generation:** the first track segment is called the trajectory seed. It is constructed from pixel detector hits and requires at least three points in space. This requirement is met by either grouping three detector hits (hit triplets) or associating a pair of pixel hits to the nominal beam spot.
2. **Pattern recognition (or trajectory building):** the trajectory seed is extrapolated to the next detector layer by taking into account the estimated track parameters and their uncertainty. For each layer, new trajectory candidates are created by fitting the helix with a Kalman filter method [48] for every compatible detector hit. Moreover, a trajectory candidate is added in case the particle left no hits in that detector layer. In order not to skew the result, all trajectory candidates are grown in parallel while the total number of candidates is truncated at each layer in order to avoid an exponential increase of the number of candidates. This pattern recognition is repeated for each layer until reaching the outermost one or a stopping condition.
3. **Ambiguity resolution:** the previous steps can create ambiguities since a given track may be reconstructed starting from different seeds, or inversely a given seed may result in several trajectory candidates. To cleanse the trajectory candidates collection from double counting tracks coming from a single charged particle, it is required that each pair of trajectories has a fraction of shared detector hits lower than 50%. If a pair of trajectory candidates does not satisfy this condition, the candidate with the least number of reconstructed hits is removed. For cases with equal amount of hits, the track with the highest χ^2 value is discarded. This *ambiguity resolution* step is applied twice: once on all track candidates resulting from a single seed, and a second time on the complete set of track candidates coming from all seeds.
4. **Final track fit:** given that the full information on the trajectory is only available at the last hit of the trajectory and given that the first estimate of the track parameters can be skewed by constraints applied during the seeding stage, the final trajectory candidate needs to be refitted. This is done twice, using all associated detector hits at once. First, the track is refitted in an iterative way from the innermost hit estimated during seeding through the list of hits. Then a smoothing stage is applied by a second fit, taking the latter result and running backwards towards the interaction point.

Once a track has been reconstructed, the collection of detector hits is cleansed from the hits used in the reconstruction of the former, then the procedure is repeated

with looser criteria on the trajectory seed candidates and on the final track selection. This iterative procedure allows looser reconstruction criteria at each step, ensuring the minimization of the number of reconstructed fake tracks.

The reconstruction of electrons and muons and its efficiency are discussed in Sec. 4.4. The relative transverse momentum resolution of electron tracks and muon tracks is shown on is shown on Fig. 3.13. As mentioned in subsection 3.2.3, the decrease in resolution is expected at high- p_T due to the particle trajectories being less curved. Furthermore, the resolution decrease at higher pseudorapidity is mainly due to a shorter lever arm in the transverse plan and to an increase of the multiple scattering due to the material budget in this region of the detector. The momentum resolution for electrons measured in the tracker is much worse than for muons. For example, for $p_T = 100$ GeV at $\eta = 0$, the central interval containing 68% (90%) of the distribution corresponds to a relative momentum resolution of about 15% (40%) for electrons, while it corresponds to a resolution of about 2% (3%) for muons. This difference is due to the loss of an important fraction of the electron energy in a few discrete bremsstrahlung interactions in the tracker material, which make the helix model inadequate for electrons.

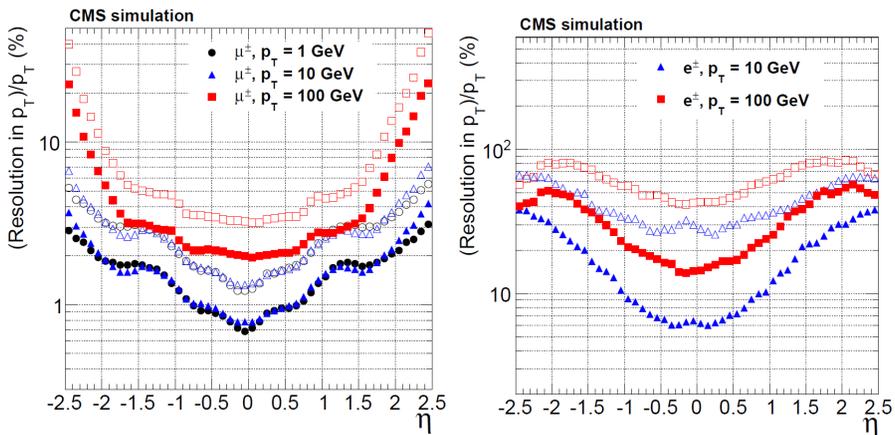


Figure 3.13.: Simulation of the relative transverse momentum resolution as a function of the pseudorapidity of the track for single isolated muon (left) and single isolated electrons (right) using the residual method. For each bin in pseudorapidity, the solid (open) symbols correspond to the width of a central interval containing 68% (90%) of the distribution of $\frac{p_T^{\text{rec}} - p_T^{\text{sim}}}{p_T^{\text{sim}}}$ [43].

Once all the clusters produced by a charged particle have been associated to a track, the reconstructed set of tracks is used together with the pixel hits in order to infer the positions of the interaction vertices. When all vertices have been found, they

are sorted by decreasing values of the sum of the squared tracks transverse momenta. The vertex with maximum value is the most of the times the one corresponding to the hardest collision. In the collision data recorded at CMS in 2016, 27 collisions per bunch crossing happened on average, with extrema of up to 55 interactions per bunch crossing. In those conditions, CMS reached a resolution on the primary vertex of $13\ \mu\text{m}$ and $19\ \mu\text{m}$ in the transverse plane and in the z direction respectively. Compared to 2015, a degradation of the resolutions by 10% was observed [49]. This is mainly caused by larger pixel dynamic inefficiency due to higher instantaneous luminosity in 2016, leading to larger fraction of tracks with hit missing from the innermost pixel barrel layer (see Fig. 3.10).

3.2.5. The calorimeters

Calorimeters are used to measure the energy of the particles produced during the collision. On the contrary of the tracker system which can only track charged particles, the calorimeters collect and measure the energy deposits of both charged and neutral particles. The only exceptions are neutrinos and muons: the former are passing through the whole detector without interacting and the latter just leave ionization deposits along their trajectory.

The calorimetry system of CMS is divided into an electromagnetic calorimeter (ECAL) specifically designed to measure the energy of photons and electrons, and a hadronic calorimeter (HCAL) part designed to measure the energy of hadrons. Since the requirements on the energy resolution for photons and electrons is much more stringent than for hadrons, the ECAL is placed before the HCAL.

The electromagnetic calorimeter

The purpose of the CMS electromagnetic calorimeter [50, 51] is to identify — with the help of the inner tracker system — electrons and photons and to measure accurately their energies. The CMS ECAL is a homogeneous calorimeter with a cylindrical geometry. It is composed of a barrel (EB) made of 61 200 scintillating lead tungstate (PbWO_4) crystals and two endcaps (EE) made of 7324 PbWO_4 crystals each. A layout of the CMS ECAL is shown on Fig. 3.14.

The barrel has an inner radius of 129 cm, a longitudinal length of 630 cm and has a pseudorapidity coverage up to $|\eta| < 1.479$. The barrel is segmented in 36 supermodules of half the longitudinal length of the barrel. These supermodules are themselves divided in four modules along the η direction. The module closest to the interaction point in η for each supermodule comprises 500 crystals while the three other modules comprise 400 crystals each.

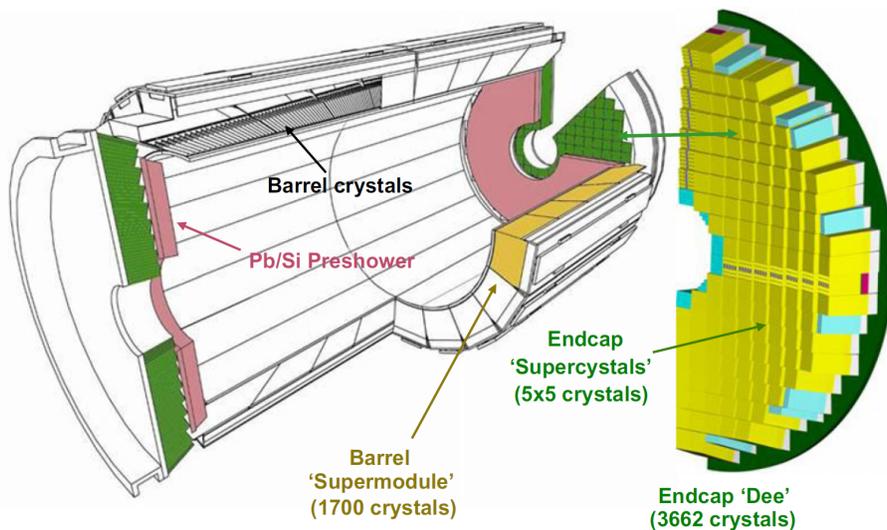


Figure 3.14.: Schematic layout of the CMS electromagnetic calorimeter [52].

The endcaps are covering the region $1.479 < |\eta| < 3$ and are made of two half-disks: the *dees*. Each of these *dees* is made of 3662 crystals.

The PbWO_4 crystals have a truncated pyramid-shaped geometry. In the barrel they have a length of 23 cm (corresponding to about $25.8 \chi_0$), a front area of $22 \times 22 \text{ mm}^2$ and rear area of $26 \times 26 \text{ mm}^2$. Each crystal covers a surface of $\Delta\eta \times \Delta\phi = 0.0175 \times 0.0175$ (i.e. of approximately 1° in ϕ). In order to avoid the situation in which a particle emitted from the interaction point passes mainly between two crystals, the crystals in the barrel are inclined by 3° in both η and ϕ directions with respect to the nominal point of interaction.

In the endcaps, the PbWO_4 crystals have a front area of $28.6 \times 28.6 \text{ mm}^2$ and a rear area of $30 \times 30 \text{ mm}^2$. Again to prevent particles to pass mainly between crystals, the axis of the crystals are oriented in such a way that they cross the axis of the beams at a distance of 130 cm beyond the nominal point of interaction.

Two preshower detectors are placed on the inner side of the endcaps to help discriminate isolated photons from pairs of photons coming from π^0 decays. Those detectors have smaller granularity compared to the ECAL granularity and cover the region $1.653 < |\eta| < 2.6$. They are sampling calorimeters formed by two disks of lead (of respective width of $2\chi_0$ and $1\chi_0$) that initiate the electromagnetic cascade, alternating with two layers of silicon micro strip detectors which also provide a measurement of the released energy.

The CMS ECAL requirements are a compact calorimeter with high granularity, fast response time and a high resistance to radiations. This is achieved by the choice

of PbWO_4 material. Indeed its high density ($\rho = 8.3 \text{ g/cm}^3$), short radiation length ($\chi_0 = 0.89 \text{ cm}$) and small Molière radius ($R_M = 2.2 \text{ cm}$) (which corresponds to the radius of a cylinder that would contain on average 90% of the shower's energy deposition) allow for a compact and fine granularity calorimeter. Furthermore, the decay time of the scintillation is of the order of 15 ns and it allows to gather $\sim 80\%$ of the light emitted within the 25 ns nominal time between two successive LHC beam bunches crossing. Finally, the crystals are resistant to high doses of radiations and therefore can operate for several years in a high radiation environment such as at the LHC. The main effect of those radiations is a loss of transparency of the crystals and is accounted for by special real time laser monitoring. Figure 3.15 shows the stability of the relative energy scale measured from π^0 decays in the ECAL barrel for a typical LHC fill in 2016. The effect of the light monitoring (LM) correction is also shown.

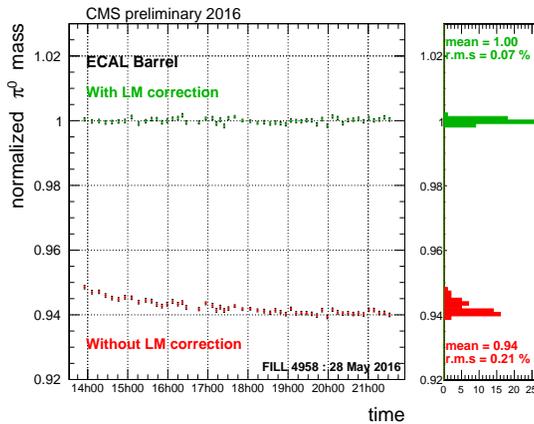


Figure 3.15.: Stability of the relative energy scale measured from the invariant mass distribution of π^0 decays in the ECAL barrel for a typical LHC fill in 2016. The energy scale is measured by fitting the invariant mass distribution of approximately 200 000 photon pairs in the mass range of the π^0 meson. The error bars represent the statistical errors on the fitted peak position. The plot shows the data with (green points) and without (red points) light monitoring (LM) corrections applied. The right-hand panel shows the projected relative energy scales [53].

The main drawback of these crystals is the poor light collection (~ 10 photoelectrons/MeV) leading to a need to amplify the light signal. This is achieved through the use of avalanche photodiodes (APD) in the barrel and photo vacuum triodes (VPT)

in the endcaps, which are designed to be resistant to radiation and to operate under a strong magnetic field.

The energy resolution of the ECAL can be expressed by the sum of three terms:

$$\frac{\sigma}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C. \quad (3.6)$$

The first term S corresponds to the stochastic term and takes into account random fluctuations in the number of photoelectrons emitted. The noise term N represents the contribution of the electronic noise. This contribution varies depending on the colliding conditions delivered by the LHC and more specifically on the pile-up. The constant term C , dominant at energies above 100 GeV, includes various contributions such as the stability of the operating conditions (in particular temperature and voltage), the presence of inert material in the crystal, the non uniformity of the collection of light along the crystal, inter-calibration errors between crystals and damage from radiations.

An estimate value of S , N and C have been obtained from test beam measurements [54]. A typical resolution was found to be :

$$\frac{\sigma}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.30\%, \quad (3.7)$$

where E is given in GeV.

The hadronic calorimeter

The hadronic calorimeter of CMS (HCAL), in complement to the tracker and the ECAL, allows the measurement of the energy and direction of hadrons [55, 56]. The HCAL also helps in the identification of electrons, photons and muons in conjunction with the ECAL and the muon system.

The HCAL has to be compact — so it can be surrounded by the magnet coil of CMS — while providing good containment of hadronic shower and hermeticity to allow a precise energetic balance measurement in the transverse plane. Indeed, colliding beams are collinear and have an initial momentum in the transverse plane (p_T) close to zero. Therefore summing vectorially the total transverse energy in one event allows to attribute the unbalanced transverse energy to non interacting particles in this event, such as neutrinos. This is called the missing transverse energy, E_T^{miss} , and is a major element in the final state of this thesis.

Following those requirements the CMS HCAL is a hermetic sampling calorimeter with a pseudorapidity coverage up to 5. Figure 3.16 shows a quarter view of the HCAL and its sub-detectors in the $y-z$ plane. The HCAL is divided into four sub-detectors: the hadronic barrel calorimeter (HB), located in the region of the barrel and inside the magnet; the hadronic endcap calorimeter (HE), located in the region of the endcaps, also inside the magnet; the hadronic outer calorimeter (HO), placed along

the inner part of the return yoke of the magnetic field, just outside of the magnet; and the hadronic forward calorimeter (HF), located in the forward region, outside of the endcaps. The gap between the barrel and the endcap of the HCAL, through which the cable services of the ECAL and inner tracker pass, is inclined at 53° and points away from the center of the detector.

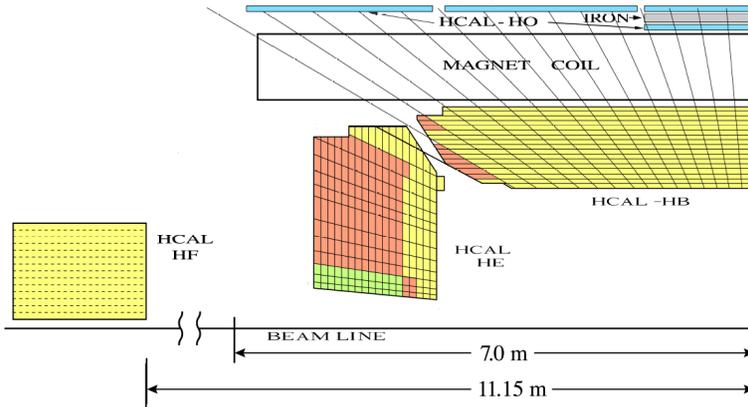


Figure 3.16.: Quarter view of the CMS hadronic calorimeter. The colours indicate the optical grouping of scintillator layers into different longitudinal read-outs [57].

The absorbing material is a dense, non-magnetic absorber made of brass and the active material is made of plastic scintillator tiles with wavelength-shifting (WLS) fibers to transport the light to hybrid photo-diodes.

As mentioned above, the HCAL is divided into two parts in the barrel region of CMS: the barrel part (HB) and the outer barrel part (HO). The HB is 9 m long and fills the remaining space between the ECAL and the magnet coil ($178 < r < 288$ cm and $|\eta| < 1.4$). The HB consists of two half-barrels, each of which contains 18 identical azimuthal wedges covering an angle $\Delta\phi = 20^\circ$ per wedge. Each wedge is segmented in four sectors along ϕ , covering an angle $\Delta\phi = 5^\circ$. The wedges are bolted together in a staggered geometry, avoiding a configuration with a projective passive material for the full radial extent of a wedge. Each sector is divided in 16 towers along η for a total of 2304 calorimeter towers of granularity $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. Each tower is made of a pile of 17 layers of 3.7 mm scintillator tiles (except for the innermost one which is 9 mm thick) interspersed with 15 ~50 mm-brass plates absorber and 2 stainless steel at the extremities to ensure the solidity of the

structure. All the fibers relative to the same tower are sent to the same photodetector which provides the integration of the signal.

Due to the limited space available inside the solenoid volume, an extension of the HB is installed outside the solenoid: the hadronic outer calorimeter (HO). The outer calorimeter improves the measurement of the energy of hadron cascades of higher energy that can overcome the depth of the HB and the solenoid. This extension is made only of scintillating detectors, in order to detect tails of hadronic cascade showers initiated inside the HB. The HO consists of 5 rings — mounted on the 5 return yokes of the magnet system — divided in 12 sectors of $\Delta\phi = 30^\circ$. Each of those segments has a granularity similar to the one in HB, leading to a one-to-one correspondence between towers of HB and segments of HO. The HO is made of two layers of plastic scintillators except for the ring at $\eta = 0$ which has an additional 15 cm of steel absorber plate as shown on Fig. 3.16. The light from the scintillators is collected via WLS fibers and transported to the photodetectors located on the return yoke.

The hadronic endcap calorimeter extends the HB coverage between $1.3 < |\eta| < 3.0$, creating a small overlap in pseudorapidity with the barrel region. The HE is divided along ϕ into 18 sectors of $\Delta\phi = 20^\circ$ and along r into 14 rings. The sectors are further divided into 72 sections of $\Delta\phi = 5^\circ$ each except for the 8 innermost sections whose segmentation is $\Delta\phi = 10^\circ$ to allow the passage of WLS fibers. The segmentation in η varies from 0.35 for the innermost towers to 0.87 for the outermost ones. The HE towers are using the same technology than the one in HB with a total of 19 layers of 3.7 mm thick plastic scintillators alternating with layers of 78 mm thick brass absorber.

The last part of the HCAL is the forward calorimeter (HF). These calorimeters are installed 11.15 m away from the nominal interaction point, covering a region $3.0 < |\eta| < 5.0$. This calorimeter enhances the identification of processes producing forward jets. Since those forward region receive a high flux of particle radiation, the use of scintillators is not allowed and the HF is a sampling calorimeter made of steel absorbers surrounding quartz fibers covered with plastic. Those quartz fibers emit Cerenkov light with the passage of charged particles which is detected by radio-resistant photomultipliers. The two forward calorimeters are cylindrical sub-detector of 1.65 m long with an external radius of 130 cm and an internal one of 12.5 cm and are made of 18 sections. Each section covers an angle of $\Delta\phi = 20^\circ$ and is made of 24 towers. The segmentation in η varies from 0.1 to 0.3 from the innermost to the outermost towers.

To maintain a good energy resolution, the HCAL has to be monitored and calibrated through the full data-taking period for radiation damage. Figure 3.17 shows the radiation damage in the HE for the 2016 data-taking period. The radiation damage is defined by the ratio of the recorded energy by HCAL cells after a certain amount of integrated luminosity divided by the recorded energy by HCAL cells at the beginning

of the data-taking period. This value is then normalized to the one of the cell receiving the minimal dose. As expected one can see that the closest the cells from the beam pipe (i.e. at the highest $i\eta$ on Fig. 3.17), the more radiation damage is measured and the higher the correction needed. Figure 3.18 shows the stability over the 2016 data-taking period of a fitted reconstructed Z boson mass peak position with one electron in the ECAL and one electron candidate in the HF after correcting HF for radiation damage. The reconstructed Z boson peak position is below Z boson mass due to a shower leakage into HF non instrumented region between HF wedges.

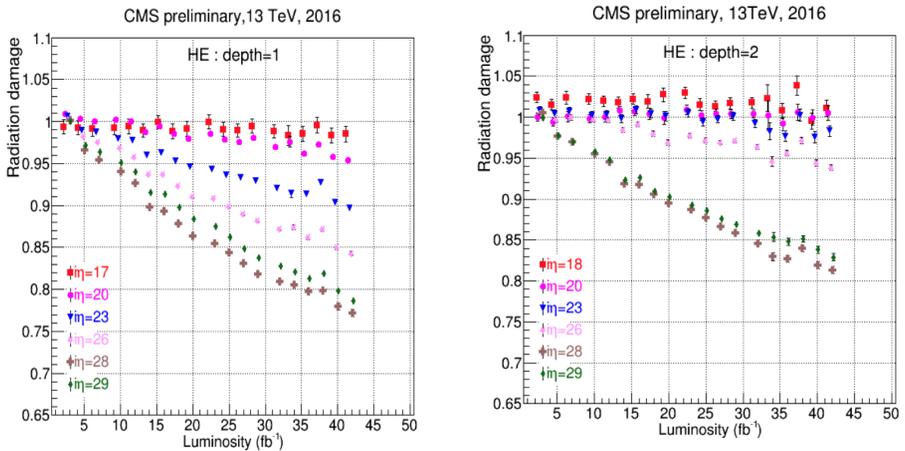


Figure 3.17.: HE radiation damage (see text) as a function of delivered luminosity in 2016 for depth 1 (left panel) and depth 2 (right panel) cell. The depths 1 and 2 corresponds respectively to yellow and red cells on Fig. 3.16 [58].

The resolution of the HCAL was measured by the CMS Collaboration in 2008 using test beams of electrons, pions, protons and muons [59]. An ECAL module was also included in the setup. The hadronic energy resolution for a combination of ECAL and HCAL is parameterized as:

$$\frac{\sigma}{E} = \frac{S}{\sqrt{E}} \oplus C, \quad (3.8)$$

where S corresponds to a stochastic term, C to a constant term and E is expressed in GeV. In the barrel it was found that $S = 0.847\sqrt{\text{GeV}}$ and $C = 0.074$. The corresponding values for HF are $S = 1.98\sqrt{\text{GeV}}$ and $C = 0.09$. The stochastic term is higher in the HF than in other calorimeters, however since forward jets typically have very high energies the aimed energy resolution is in general achieved. For hadrons with an energy of 100 GeV in the barrel, the energy resolution is of the order of 11%.

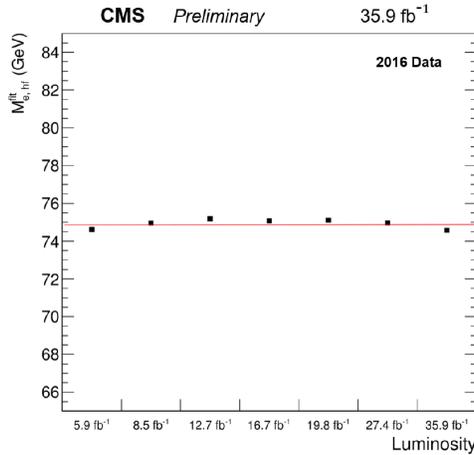


Figure 3.18.: Stability of fitted reconstructed Z boson mass peak position vs time of data-taking for 2016 data after correcting for radiation damage. The Z boson is required to be reconstructed with one electron in the ECAL and one electron candidate in the HF after correcting HF [58].

3.2.6. The muon detectors

Muons are the only particles that should cross the full CMS detector without being absorbed — with the exception of neutrinos which are not interacting with the detector at all — leaving almost nothing but a trajectory in the tracker. The muon detection system is located at the outermost region of the CMS detector — after the calorimeters and the magnet coil — where one can expect that any signal collected there is likely to come from the passage of a muon.

The muon detection system [60,61] — or muon spectrometer — is made of three different gaseous detection technologies, forming a total of 1846 muon chambers. It is made of 250 chambers composed of drift tubes (DT), 540 cathode strip chambers (CSC) and 1056 (480 in the barrel region and 576 in the endcaps) resistive plate chambers (RPC). DTs are placed only in the barrel region while CSCs are placed only in the endcaps. As for the RPCs, they are installed in both regions, adding redundancy to the system and, due to their excellent time resolution, additional trigger information to manage to associate detected muons to individual LHC bunch crossings. Figure 3.19 shows a quadrant of CMS emphasizing the placement of the various muon chambers.

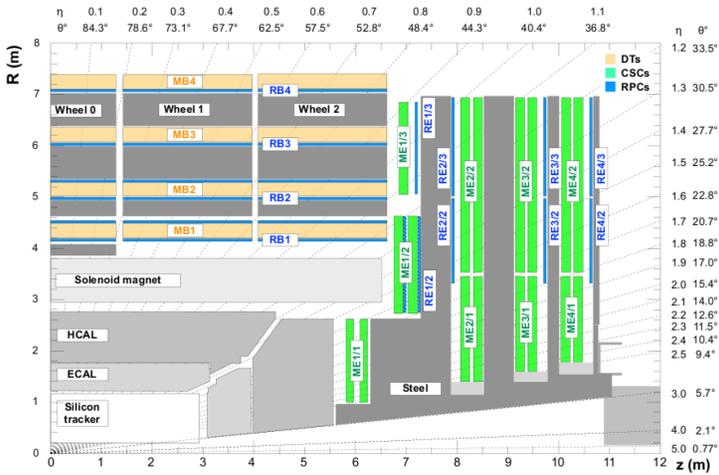


Figure 3.19.: Layout of the muon spectrometer of CMS depicting the three gaseous detectors in use: Drift Tubes (DTs) in orange labeled MB_x , Cathode Strip Chambers (CSCs) in green labeled $ME_{x/y}$, and Resistive Plate Chambers (RPCs) in blue labeled RB_x and $RE_{x/y}$. The index x represents the station number and y the ring number [62].

Drift tubes

As mentioned above, the muon drift tubes chambers are only present in the barrel region. Indeed the forward region is under an intense magnetic field and under a high neutron-induced background, making it a non ideal environment for drift tube technology. They cover a pseudorapidity region up to 1.2. They are mounted on five wheels, each segmented into 12 sectors of $\Delta\phi = 30^\circ$. Each sector is organized into 4 stations (labelled MB_1 – 4 on Fig. 3.19) alternating with steel plates from the return yoke. The three innermost stations are composed of 3 superlayers (SL) as shown on Fig. 3.20. A superlayer is made of 4 layers of rectangular drift cells staggered by half a cell. Inside those stations, the central superlayer measures muon coordinates in the r - z plane, while the two other layers — separated as much as possible to achieve the best angular resolution — are sensitive in the r - ϕ plane. The outermost DT station (MB_4) is only made of two r - ϕ superlayers.

A drift cell is 2.4 m long with a section of $13 \times 42 \text{ mm}^2$ in which a wire is stretched over the whole length as sketched on Fig. 3.21. Drift cells are filled with a gas mixture of Ar and CO_2 , providing the position of the muon by measuring the drift time of the ionization electrons to an anode wire within a shaped electric field. This electric field is created by setting 3 different potential respectively to the anode,

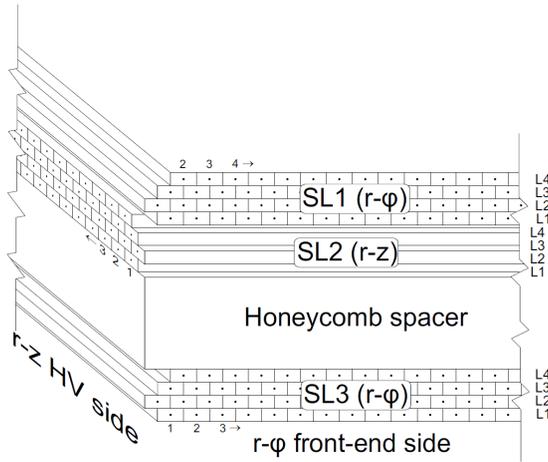


Figure 3.20.: Schematic layout of one of the innermost DT chamber , showing the arrangement of the three superlayers (SL) [63].

the electrode strips and the cathode strips, in such a way that the drift velocity is approximately constant in the whole cell.

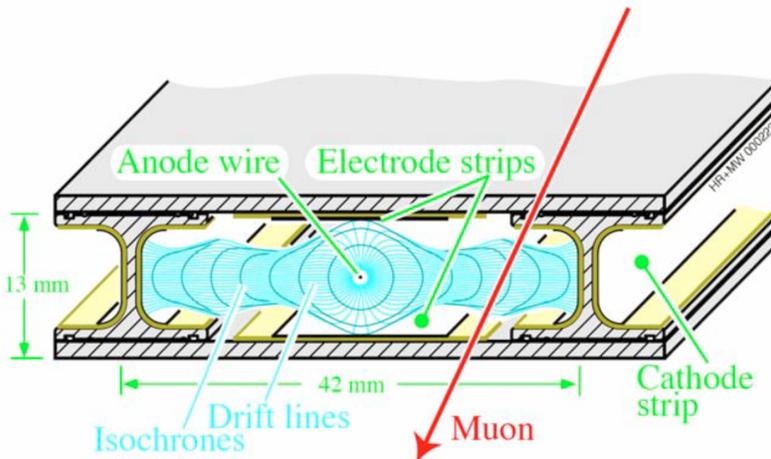


Figure 3.21.: Sketch of a DT cell showing drift lines and isochrones. The plates at the top and bottom of the cell are at ground potential. The voltages applied to the electrodes are 3600 V for wires, 1800 V for strips, and -1200 V for cathodes [40].

The flux of particles in this barrel region is of the order of 2 Hz cm^{-2} and permits the use of those long drift chambers which limits the number of active channels and therefore costs less while still providing a spatial resolution of the order of $100 \mu\text{m}$ and an angular resolution on the direction of the muon of the order of 1 mrad . The DTs are also relatively fast detectors with a time resolution of 5 ns , making them suitable to trigger events containing a muon candidate.

Figure 3.22 shows DT hit transverse spatial resolutions with 2016 data as a function of station and wheel both for the cells in the ϕ SLs and in the θ SLs. In the bending plane (i.e. in the ϕ SLs) the resolution is better than $250 \mu\text{m}$ in the three innermost station and better than $300 \mu\text{m}$ in the outermost station, which were the requirement to reach a resolution in a full DT station of the order of $100 \mu\text{m}$. In the θ SLs the resolution varies from about $250 \mu\text{m}$ to $600 \mu\text{m}$ except in the outer wheels of the first stations. The symmetric behaviour between the wheels with respect to the $z = 0$ plane is expected due to the detector geometry. In the central part of the detector, in wheel 0, the tracks are coming mostly perpendicular to all layers, yielding an equivalent resolution for the θ and ϕ superlayers. In the more forward regions, the inclination angle of the tracks has opposite effect in the θ and ϕ superlayers. In the ϕ superlayers in the forward regions, the tracks are closer to the direction of the wire, increasing the track path within the tube and therefore the ionization charge and the resolution. In the θ superlayers in the forward region, the inclination angle degrades the linearity of the distance–drift time relation, thus worsening the resolution.

Cathode strip chambers

Well suited for the uneven magnetic field and high particle rates in that region of the detector, the cathode strip chambers covers the region $0.9 < |\eta| < 2.4$, overlapping with the DTs until a pseudorapidity of 1.2.

The CSCs are trapezoidal multi-wire proportional chambers with a finely segmented cathode strip readout. In particular they are made of arrays of negatively-charged copper cathode strips directed radially from the beam direction and positively-charged anode wires orthogonal to both the strips and the beam direction, the whole system laying in a volume of gas. Figure 3.23 shows a schematic view of a CSC module. A CSC contains 6 layers of gas gaps with wires and strips. As in the DT, ionising processes occur in the CSC when muons pass through. However, not only the freed electrons induce a signal, the ions resulting from the ionisation induce an electric signal on the cathode stripes, yielding an accurate measurement of the position in the bending plane (r - ϕ).

Taking advantage of the multiple detection layers of the chambers, CSCs can provide a detection efficiency better than 99%, $90 \mu\text{m}$ resolution in the r - ϕ coordinate

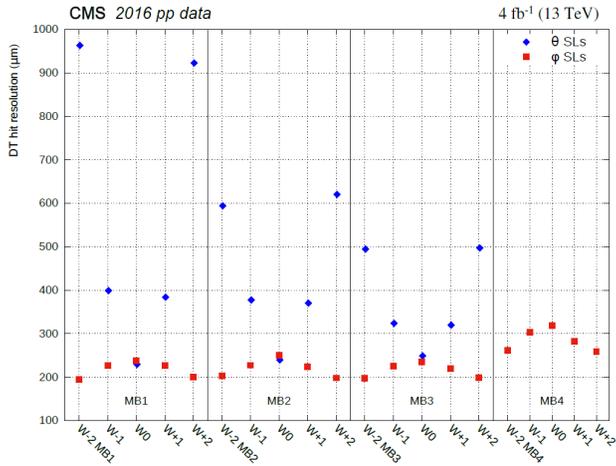


Figure 3.22.: DT hit transverse spatial resolutions with 2016 data, plotted as a function of station and wheel. Squares show the resolution in ϕ SLs and diamonds the resolution in θ SLs. The uncertainties in these values are smaller than the marker size in the figure. On average, eight hits are combined to give the overall resolution of a DT station [61].

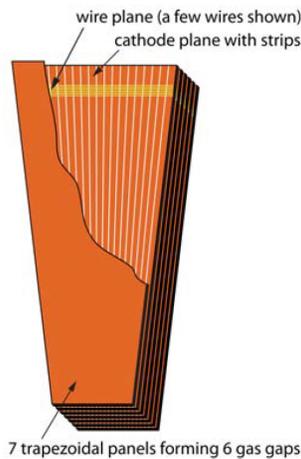


Figure 3.23.: Layout of a CSC made of 7 trapezoidal panels. The panels form 6 gas gaps with planes of sensitive anode wires. The cut-out in the top panel reveals anode wires and cathode strips. Only a few wires are shown to indicate their azimuthal direction. Strips of constant $\Delta\phi$ run lengthwise (radially). [40].

for ME1/1 and ME1/2 and about $150\ \mu\text{m}$ elsewhere, and a bunch crossing assignment efficiency of 99%. The time resolution of the CSCs is on the order of 5 ns.

Table 3.1 summarizes the mean spatial resolution in each CSC station and ring with 2016 data.

Station	Spatial Resolution (μm)
ME1/1a	45
ME1/1b	52
ME1/2	90
ME1/3	105
ME2/1	125
ME2/2	134
ME3/1	120
ME3/2	135
ME4/1	123
ME4/2	139

Table 3.1.: CSC transverse spatial resolution per station (6 hits) measured for all chamber types with 2016 data. The uncertainties on these values are of the order of 5% [61].

Resistive plate chambers

Resistive plate chambers are double-gap chambers operated in avalanche mode and are used for their excellent time resolution of less than 3 ns which allows them to unambiguously assign events to their corresponding bunch crossing while also providing a spatial resolution of the order of the centimeter. Figure 3.24 shows a sketch of a RPC module. The module consists of two gas filled gaps, with a common read-out strip layer in the middle. On each side of the gaps, an insulating Bakelite panel is covered with a conducting graphite layer on which a high voltage is applied. RPC are installed up to $|\eta| < 1.9$, leaving an empty region between $1.9 < |\eta| < 2.4$ where CSCs are the only detectors providing measurements.

Table 3.2 summarizes the mean spatial resolution in each RPC station and layer with 2016 data.

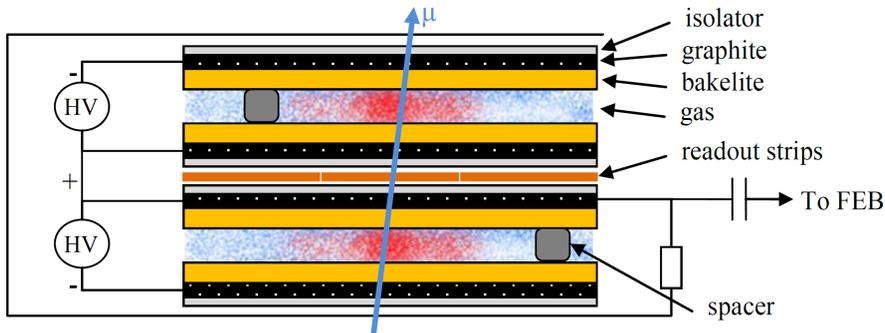


Figure 3.24.: Layout of a double-gap RPC chamber [64].

Barrel layer	σ (cm)	Endcap ring	σ (cm)	Endcap ring	σ (cm)
RB1in	0.78	RE1/2/A	0.98	RE(2,3,4)/2/A	1.19
RB1out	0.84	RE1/2/B	0.92	RE(2,3,4)/2/B	0.97
RB2in	1.02	RE1/2/C	0.99	RE(2,3,4)/2/C	0.89
RB2out	0.97			RE(1,2,3,4)/3/A	1.32
RB3	1.04			RE(1,2,3,4)/3/B	1.38
RB4	1.27			RE(1,2,3,4)/3/C	1.19

Table 3.2.: RPC transverse spatial resolution with 2016 data for each station and layer. The numbers have been corrected for small misalignment effects. The uncertainties on these values are of the order of a few % [61].

Trasverse momentum resolution

For muons with $p_T < 200$ GeV, the momentum measurement is dominated by the performance of the inner tracker. However at higher p_T , combining information from the muon system and information from the inner tracker improves the momentum resolution.

Figure 3.25 shows an estimate of the momentum resolution of cosmic ray muons by comparing the measured muon momentum using only the upper and only the lower half of the detector. In particular it shows the RMS of the relative q/p_T residual ($R(q/p_T)$) as a function of p_T for cosmic rays recorded in 2015 for two type of fits: one using only inner tracker information and one combining inner tracker and muon system information. $R(q/p_T)$ is defined as:

$$R(q/p_T) = \frac{1}{\sqrt{2}} \frac{(q/p_T)_{upper} - (q/p_T)_{lower}}{(q/p_T)_{lower}}, \quad (3.9)$$

where q is the muon charge, and *upper* and *lower* refer to the muon tracks reconstructed in the upper and lower halves of the CMS detector, respectively. The factor $\frac{1}{\sqrt{2}}$ takes into account the fact that the q/p_T measurements of the two tracks are independent.

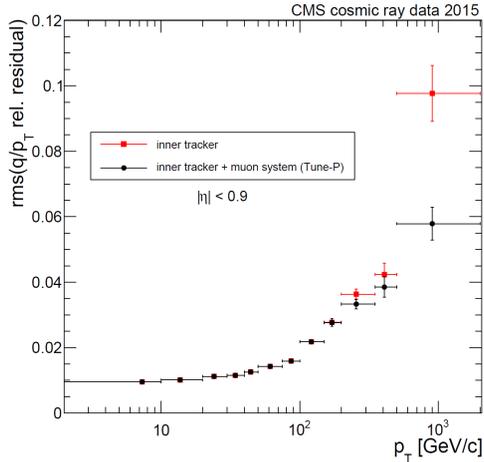


Figure 3.25.: RMS of $R(q/p_T)$ as a function of p_T for cosmic rays recorded in 2015, using the inner tracker fit only (squares) and including the muon system (circles) [61].

3.2.7. The trigger system

The nominal bunch crossing rate at LHC is 40 MHz, which corresponds to one bunch crossing every 25 ns. Each bunch crossing leads to multiple proton-proton collisions. During 2016, an average of 27 collisions per bunch crossing was observed [37]. However all events are not recorded. Indeed the maximum acceptable rate for data acquisition and storage is of the order of 1 kHz with the technology installed in CMS. It is the role of the trigger system to accept the highest possible cross section of interesting physics events while discarding low energy hadronic processes and elastic collisions [65,66].

CMS is using two trigger stages: the Level-1 Trigger (L1 Trigger), a hardware trigger used to reduce the 40 MHz rate of collisions to 100 kHz, and the High Level Trigger (HLT), a computer farm used to reduce the rate further down to the order of 1 kHz.

The Level-1 Trigger

The Level-1 Trigger runs on dedicated electronics using coarse level granularity information given by the calorimeters and the muon system. The tracker information is not used at this level. Knowing that sub-detectors electronics have a buffer memory size of 128 events, the L1 has to process the information from those sub-detectors and to take the decision to keep or reject an event for every bunch crossing within $3.2\ \mu\text{s}$.

The Level-1 Trigger purpose is to identify high-energy electrons, muons, taus, photons, jets and missing transverse energy. To achieve that, three main subsystems are used, as shown on Fig 3.26:

1. **L1 Calorimeter Trigger:** the energy deposits in the HCAL and ECAL are collected and sent to the Calo Trigger Layer 1. The base concept of the Calorimeter Trigger is the reduction of input data volume through several stages. Indeed at each stage objects are identified and sorted and the best candidates are forwarded to the next stage. In the latest step, the Calo Trigger Layer 2 finds the twelve highest transverse energy jet, tau and electron/photon candidates and computes global energy sums.
2. **L1 Muon Trigger:** it is divided in three sections: the barrel, the endcaps, and the overlap region between CSC and DT sub-detectors. Hits from the different chambers are combined in the Muon Track-Finder Layer which uses track extrapolation and pattern recognition algorithms to reconstruct tracks. These reconstructed tracks are sent to the Sorting/Merging Layer which selects the best candidates according to quality parameters. The four best tracks are sent to the Global Muon Trigger which uses information from the calorimeters to compute isolation values for each muon. This combined information is then sent to the Global Trigger along with the data of the calorimeter.
3. **L1 Global Trigger:** with the reconstructed electrons/photons, muons, jets and missing energy, the L1 Global Trigger performs selections and make the decision to accept or reject events based on a menu of triggers. Once the event accepted, this decision is sent to the Trigger Control System which in turn sends the command to read the corresponding event data from all CMS sub-detectors. Otherwise, the event is definitely lost.

The High Level Trigger

The High Level Trigger is a farm of computers processing events passing the L1 Trigger using the full CMS sub-detectors and more complex algorithms than at L1.

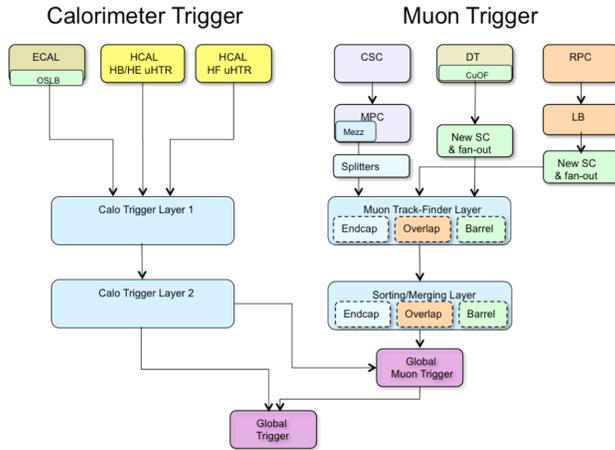


Figure 3.26.: Architecture of the Level-1 Trigger [65].

The HLT reduces the 100 kHz rate from L1 to about 1 kHz, taking up to 50 ms to reconstruct events. To achieve this, the HLT typically executes reconstruction algorithms in stages of increasing complexity, but this may vary from one HLT path to another. For example it can start by reconstructing clusters from ECAL and HCAL energy deposits taking into account possible bremsstrahlung radiation and by building muon tracks from muon chambers. In the second step, the hits found in the pixel sub-detector are usually combined with the clusters reconstructed at the previous stage in order to restrict the region of the tracker interesting for this event. In this region, track segments are then built. In the last step, the full tracker information is used to complete track reconstruction of the track segments previously made and an analysis of the full event is performed leading to the final decision of keeping or rejecting the event.

In order to save CPU time, each reconstruction step is followed by a filter in order to avoid running time-consuming code if it is already clear it will not be needed. This ensures a smaller and smaller rate of events needing more and more complex reconstruction algorithm.

Finally, events passing the HLT requirements are sent to the Storage Manager which saves the entire event raw data on disk, typically taking 1–3 MB/event. Then full off-line reconstruction of the event can start.

In total, the HLT is made of ~ 550 trigger paths. Some of them have loose thresholds and therefore lead to a high event rate. To reduce such trigger rates, the corresponding trigger paths are *prescaled*, i.e. the paths are actually triggered every number of events defined by the prescales, which evolve as a function of the luminosity.

Chapter 4

Event generation, simulation and reconstruction

In order to search for new physics, observed data has to be compared with theoretical predictions of SM processes (backgrounds) as well as BSM processes (signals) modelled by Monte Carlo (MC) simulations. The first part of this chapter will cover the basis of the event generation and simulation processes. The second part will cover the reconstruction of the observed — or generated — event signatures inside the detector into physics objects.

4.1. Introduction

The production of Monte Carlo events can be divided in four general steps. The first three steps are described in the next section and are specific to simulations while the last step touches the reconstruction procedure and is identical for both data and generated events. Those four steps can be summarised as follows:

- **Generation of the event:** the first step describes the proton-proton collision and in particular the hard-scattering process, the showering and hadronization of the partons, the beam remnants and the underlying events.
- **Simulation of the detector:** the second step consists in modelling the experimental setup described in the previous chapter. This consists in a detailed description of the detector sensitive volumes and material, the interaction of particles with the detector and their energy deposits in the detector material, and the spread of the proton-proton interaction point along the beam axis.

- **Digitization:** the response of detector electronic readouts has to be simulated. This simulation takes into account the effect of pile-up interactions, consisting in an additional set of soft inelastic collisions, and emulates the L1 and HLT triggers. As it is difficult to predict in advance the distribution of the number of pile-up interactions in data and since resources are limited, Monte Carlo datasets are generated for a scenario with a higher number of vertices and then reweighted to match the observed distribution of pile-up interactions during the data-taking period.
- **Reconstruction:** the last step is a step common to both data and simulations. Described in Sec. 4.4, it's a set of algorithms taking as input event signatures inside the detector and reconstructing and identifying physics objects.

4.2. Event generation

The big picture of Monte Carlo event simulation in particle physics is sketched on Fig. 4.1. Those steps are described below and more details are given in the following sub-sections [67]:

- **Hard-scattering process and parton distribution functions:** the very first step of the simulation is the generation of the hard-scattering of the two protons, or more precisely, of their partons. The probability for a parton to carry a fraction x of the proton momentum at an energy scale Q^2 is given by Parton Distribution Functions (PDF) and is discussed in Sec. 4.2.1. Given the large energy scale, the hard-scattering process is computed with the matrix element (ME) formalism in perturbative QCD, at a fixed order in α_S depending on the generator (typically to the next-to-leading order, NLO). The hard process includes the interaction between the incoming partons but also the decay of particles with a short lifetime resulting from the interaction, such as Z bosons decaying to leptons.
- **Parton showering:** the second step consists in describing the radiation of quarks and gluons in the initial and final states, creating shower of partons as described in Sec.4.2.2.
- **Underlying events and multiple partons interaction:** underlying events (UE) are coming from the interactions of the remnants of the incoming protons, consisting mainly in soft QCD interactions. Due to the unknown distribution of the energy in the beam remnants and its colour connection to the parton participating in the hard interaction, the description of underlying events is more ambiguous than for the parton showering and hadronization. In addition to UE,

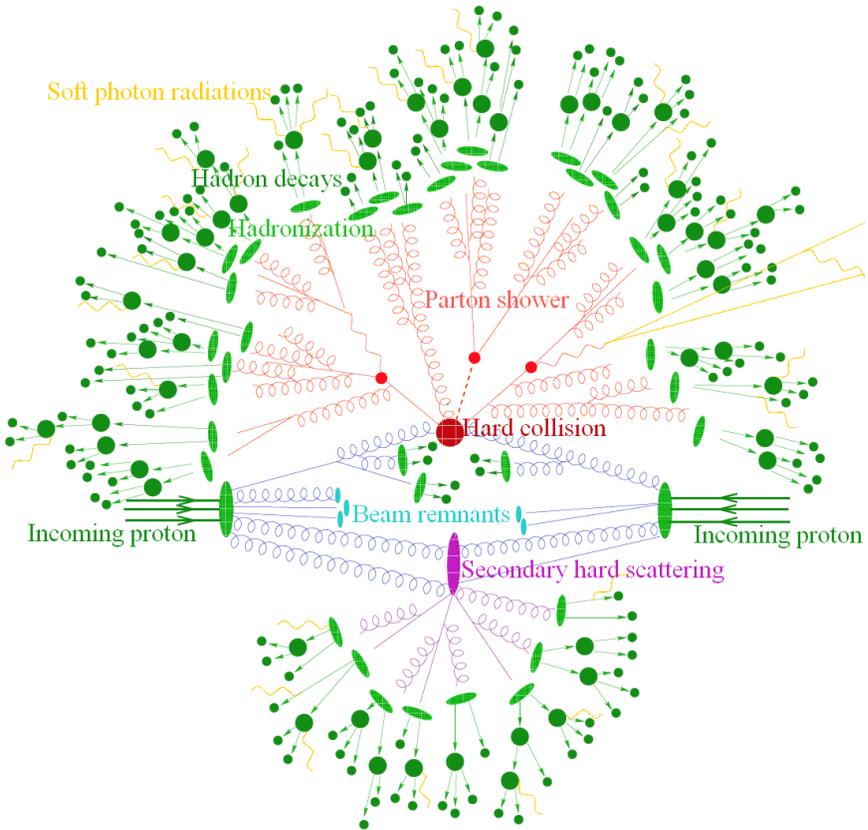


Figure 4.1.: Sketch of a proton-proton collision as simulated by a typical Monte Carlo event generator. The incoming protons are represented with three green lines coming from left and right in the middle of the picture. From those incoming protons, two gluons interact and represent (with a red blob) the hard collision which is surrounded by a tree-like structure. From those incoming partons, additional radiations are attached (initial state radiation) and, similarly, additional radiations are attached to final state partons (final state radiation). A cascade of soft gluon emission follows, called parton shower. The purple blob indicates a (softer) secondary hard scattering event while the cyan represents the fragments of the initial protons, also called beam remnants. Parton-to-hadron transitions are represented by light green ovals, dark green circles indicate hadron decays, while yellow lines signal soft photon radiation and resulting electrons [68].

there is also a small possibility of having an extra hard interaction during the collision, such events being labelled as multiple partons interactions.

- **Hadronization:** due to colour confinement, partons fragment and group into hadrons. This process called hadronization is described by a non-perturbative regime, using phenomenological models introduced in Sec. 4.2.3. The decay of short-lived hadrons and leptons is then simulated.

The Monte-Carlo generators used in the analysis to compute matrix elements are Madgraph [69], Powheg [70–72] and aMC@NLO [73] for the backgrounds while Powheg 2.0 [74] and JHUGen [75–78] are used for the signals. The aMC@NLO and Powheg generators can compute up to NLO matrix elements. All generated samples are then interfaced to PYTHIA, configured with the CUETP8M1 tune [79,80] for simulation of parton showers, hadronization and underlying event effects. All simulated events are further processed with a GEANT4-based description [81] of the CMS detector (see Sec. 4.3) and reconstructed with the same algorithms as used for data (see Sec. 4.4).

4.2.1. Hard-scattering process and parton distribution functions

The Parton Distribution Functions (PDF) are determined by global fits to data from deep inelastic scattering, Drell-Yan and jet processes performed by several collaborations. In this thesis, the set of PDF used is determined by the NNPDF collaboration [82, 83]. Figure 4.2 shows examples of PDF at two different energy scales for the various flavours of partons: $Q^2 = (2 \text{ GeV})^2$ and $Q^2 = (1000 \text{ GeV})^2$. The bands represent the uncertainty at 68% confidence level. From this figure, one can see that gluons are dense in protons, especially at low x , and therefore are the dominant colliding partons in the LHC. The asymmetry in the density of up and down quarks with their anti-particles is due to the fact that those quarks are also valence quarks for the proton. These two plots are not coming from different measurements at different energy scales. PDF are measured at specific scales and then extrapolated by perturbative QCD to other scales using the so called DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) equations [84–86].

The cross section $\sigma_{pp \rightarrow X(Q)}$ of a proton-proton collision producing a final state X of mass Q can be described by the convolution of the partonic cross section $\hat{\sigma}_{ab \rightarrow X}$ with the PDF of the involved partons [89] as follows:

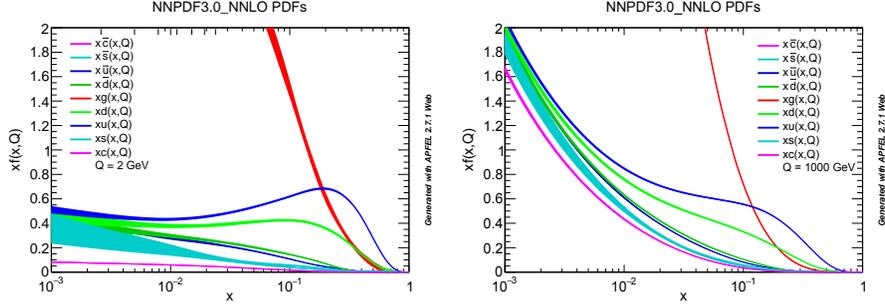


Figure 4.2.: NNPDF 3.0 parton distribution functions at energy scales $Q^2 = (2 \text{ GeV})^2$ (left) and $Q^2 = (1000 \text{ GeV})^2$ (right) generated with APFEL WEB [87, 88]. The bands represent the uncertainty at 68% confidence level.

$$\frac{d\sigma_{pp \rightarrow X}(Q)}{dQ^2} = \sum_{a,b} \int_0^1 dx_a \int_0^1 dx_b f_a(x_a, Q^2) f_b(x_b, Q^2) \times \hat{\sigma}_{ab \rightarrow X}(x_a, x_b, \alpha_S(Q^2)) \quad (4.1)$$

where the sum runs over all possible partons (taking also into account spins and colours). The differential partonic cross section is given by:

$$d\hat{\sigma}_{ab \rightarrow X} = \frac{|\mathcal{M}_{ab \rightarrow X}|^2}{64\pi^2 x_a x_b s} d\cos\theta d\phi \quad (4.2)$$

where \mathcal{M} is the matrix element of the process, which can be derived from the Lagrangian density. This matrix element depends on the strong coupling constant α_S and if the latter is small enough (i.e. if the energy scale of the interaction is high enough), it can be computed to a certain order perturbatively.

The running coupling α_S presents logarithmic ultraviolet (UV) divergence arising from vacuum polarization loop integration [90]. The UV divergence is eliminated when one normalizes the coupling to the measured value at a specific scale Q_0 . This renormalisation scale is labelled μ_R .

As aforementioned the PDF are functions of the momentum fraction of the interacting partons. Therefore, infrared divergences appear following the emission of low-energy gluons. To eliminate these divergences, the evolution of the PDF through the DGLAP equations is performed until what is called the factorization scale μ_F .

4.2.2. Parton showering

Partons can radiate softer gluons or split into two mostly collinear partons, who can again radiate or split into more particles creating a shower of partons. This pro-

cess is referred as Initial State Radiation (ISR) when it occurs with incoming partons participating to the hard process, and is called Final State Radiation (FSR) for partons coming out of the hard interaction. Knowing that a parton may branch into two daughter partons, like $q \rightarrow qg$, $g \rightarrow q\bar{q}$ or $g \rightarrow gg$, the parton showering consist in a succesion of perturbative treatments of QCD to find the probability of such branchings to appear. This succesion of splitting of partons goes on until a virtuality of the order of Λ_{QCD} is reached (i.e. where α_S is close to one) and where non-perturbative effects emerge. At this stage, the hadronisation model described in next section is taking care of combining partons into hadrons. While this procedure is followed for FSR, ISR are modelled differently for efficiency purposes. Indeed the parton showering process for ISR is reversed until the energy scale of the parton, emerging from the proton, is reached.

4.2.3. Hadronization

When the energy scale is of the order of Λ_{QCD} , due to colour confinement, the multiple partons produced by the hard scattering and the parton showering recombine into colourless state: hadrons. The hadronization is described by several phenomenological models. PYTHIA, the generator used in this analysis for hadronization, is using the Lund string model [91]. The gist of this model is that partons moving apart from each other are connected by a QCD color string. The further the partons are apart, the more stretched is the string and the more potential energy is built up. This potential energy is assumed to be proportional to the length of the string, i.e. to the distance between the partons, and of the order of 1 GeV/fm. In this approach, the probability to produce a new pair of quarks $q\bar{q}$ of a mass m increase with the distance r (and therefore the potential energy) and is proportional to:

$$\exp\left(\frac{-\pi(m^2 + p_T^2)}{\kappa r}\right) \quad (4.3)$$

where p_T is the transverse momentum of the quarks in the pair and $\kappa \approx 1$ GeV/fm. When this happen, one new $q\bar{q}$ pair is formed, lowering the available energy for a next splitting. Then the process eventually repeats and gives rise to the formation of hadrons until only colourless hadrons with on-shell mass remain. In this model, charm, bottom and top quarks are not considered since the probability to create a heavy quark pair is heavily suppressed.

4.3. Simulation of the detector

Once the generation steps described above have been completed, the simulated events can be used in analyses. They can either be directly compared to unfolded

(i.e. corrected for detector effects) data or, as it's done in this analysis, given as input to simulate the detector effects on such events in order to be able to compare them directly with data. This choice to not unfold the data is motivated by the fact that the analysis presented in this thesis is a model independent search for a new scalar boson.

The simulation of the interactions of particles with the detector material is done thanks to GEANT4 software [81, 92, 93]. This simulation is very detailed and very CPU intensive, taking up to several minutes per event depending on the number of particles in the final state. In particular, the GEANT4 toolkit includes active layers of the CMS detectors as well as dead zones created by support structures, cables and the magnet. In this simulation, both the energy losses of the particles going through the materials of CMS and the trajectory of these particles based on a complete description of the magnetic field of CMS is computed. The detector response and its electric signals are simulated by dedicated CMS offline software (CMSSW). Eventually pile-up interactions are simulated by generating extra low-energy proton-proton interactions.

4.4. Object reconstruction

4.4.1. The particle-flow reconstruction

The key elements of particle-flow (PF) reconstruction [94, 95] are its iterative-tracking strategy and its calorimeter clustering algorithm. Those algorithms take as inputs reconstructed hits and calorimeter clusters respectively. The iterative-tracking strategy provides high detection efficiencies and low misidentification rates while the clustering algorithm provides high detection efficiencies even for low-energy particles and help to separate close energy deposits. To fully reconstruct each particle with a more precise momentum resolution, the tracks and clusters are linked together. Those three key points of the PF are described in more details below.

Iterative-tracking algorithm

The iterative-tracking algorithm first reconstructs tracks with very tight criteria. This first step therefore has a small misidentification rate and a moderate efficiency. Then, the hits assigned to those tight reconstructed tracks are removed from the hit collection and the procedure to reconstruct tracks starts again but with looser criteria on the remaining hits. This iterative procedure goes on, loosening more and more the seeding criteria. More details were already given in Sec. 3.2.4).

Calorimeter clustering algorithm

Using calorimeter information, the clustering algorithm takes care of reconstructing neutral particles and complements the tracking information for charged par-

ticles. This algorithm is divided in three steps. First, *cluster seeds* are identified by taking calorimeter cells with an energy presenting a local maximum and above a given threshold. Those cluster seeds are then turned into *topological clusters* by adding adjacent cells with an energy above a threshold set to twice the noise level. Particle-flow clusters are then formed from topological clusters. Their energy and size are determined iteratively based on their distance to each cell in order to accordingly share the energy of the corresponding cells among all PF clusters that share the same cells.

Link algorithm

Once tracking and clustering elements have been computed in the previous steps of the PF algorithm, these are linked together in order to fully reconstruct particles while limiting double counting since a single particle can create multiple PF elements.

The algorithm links PF elements into blocks, which then get identified as muons, electrons, photons, charged hadrons or neutral hadrons. For example, an electron can be identified by linking a PF track to a PF calorimeter cluster in the ECAL. The quality and the likelihood of such a combination is characterized by a distance in the (η, ϕ) plane. When a block is found, its components are removed from the event and a next link is sought for until no element remains.

First, all charged particle tracks are extrapolated throughout the detector and matching PF elements are sought for in the muon tracker and in the ECAL in order to first reconstruct muons and electrons.

Then, once all muons and electrons have been reconstructed and their information removed from the event, the identification and reconstruction of charged hadrons is done by comparing the remaining tracks and ECAL and HCAL calorimeter clusters. The matching is evaluated by comparing their respective momentum measurement. If these measurements are compatible then the link is established and a charged hadron is identified. However if, on one hand, the momentum measured from the track is much lower than the one measured in the calorimeters, the link gets identified as a charged hadron with an additional neutral hadron or photon energy deposit depending on whether the calorimeter excess is measured in the HCAL or ECAL respectively. If, on the other hand, the momentum measured in the calorimeters is much lower than the one coming from the track then the PF algorithm tries to assign a muon with looser selection criteria to it, since such a signature cannot be due to charged hadrons. Eventually, when all blocks matching a muon, electron or charged hadron have been assigned, only ECAL and HCAL calorimeter clusters should remain in the events. These clusters are then assigned to photons or neutral hadrons respectively.

In the next sections the object reconstruction of the main objects used in this thesis is detailed.

4.4.2. Muon reconstruction

Muon reconstruction [61] is based on tracks from the inner tracker (tracker tracks) and from the muon system (muon segments) [96]. The CMS collaboration reconstructs *tracker muon* candidates by taking all tracks in the inner tracker with a $p_T > 0.5 \text{ GeV}$ and $p > 2 \text{ GeV}$ and extrapolating them to the muon system by taking into account Coulomb scattering, magnetic field and expected energy losses in the detector. A match is said to be found when the distance between the extrapolated tracker track and a segment (or hits) in the muon tracker system is smaller than 3 cm. Since several segments can be compatible with a single tracker track, the best matching segment is chosen for each tracker tracks forming what is defined as an *arbitrated tracker muon*.

Standalone muon tracks, on the other hand, are reconstructed from muon segments only. They in turn serve to reconstruct *global muons* by searching the best matching tracker track for each standalone muon track. For each of those tracker track and standalone muon pairs, a new fit using all hits in both tracks is then performed based on the Kalman filter (KF) technique to finally form a *global muon*.

Tracker muon reconstruction is more efficient at identifying muons with low p_T ($p_T \lesssim 5 \text{ GeV}$) while the global muon reconstruction is especially efficient for higher p_T muons leaving hits in several muon stations and giving therefore a higher p_T resolution. However, tracker muon reconstruction algorithm is better at disentangling two close high- p_T muons coming from a boosted Z decay as it is the case in this thesis and tracker muons are therefore used to tag the two leading leptons in the studied final state.

Muon identification

In this thesis three muon identification (ID) working points are used. Two loose working points (soft ID and loose ID) are used to reject additional muons in the event while one tight working point (tracker high- p_T ID) is used to identify muons coming from the Z decay in the final state.

The *soft muon ID* is tuned for muons present in jets and allows to seek for muons at very low p_T . It requires:

- an arbitrated tracker track matched ($< 3\sigma$) with at least one muon segment
- a cut on the number of tracker layers with hits above five in order to guarantee a good p_T measurement, for which some minimal number of measurement points in the tracker is needed. This cut also suppresses muons from decays in flight
- at least one hit in the pixel detector to further suppress muons from decays in flight

- a track high-purity flag to reject bad quality tracks
- a loose transverse and longitudinal impact parameter cuts, $d_{xy} < 0.3$ cm and $d_z < 20$ cm w.r.t. the primary vertex (PV) to ensure a loose compatibility with the PV (or rather with the beamspot).

The *loose muon ID* is tuned to have a high efficiency for prompt muons, as well as for muons from heavy and light quark decays. It is however less efficient for soft muons within jets hence the need of a soft muon ID. Its efficiency is close to 100% for muons coming from Z boson decays. The loose muon ID requires:

- the particle to be identified as a muon by the PF reconstruction
- the muon candidate to be a global muon or an arbitrated tracker muon.

The *tracker high- p_T muon ID* is tuned to efficiently identify high- p_T muon ($p_T \gtrsim 200$ GeV) without relying on external information. It has an efficiency of more than 95% for muons coming from Z boson decays. As mentioned in the previous chapter, this high- p_T region is where the muon detectors can significantly improve the momentum resolution. However the signal studied in this thesis is characterized by two charged leptons coming from the decay of a boosted Z . Therefore, those two charged leptons are close-by and using candidates reconstructed by the PF algorithm and global muons is less efficient. In order to recover the reconstruction efficiency of such cases, tracker muons are used instead, with the drawback that they have a poorer momentum resolution than global muons. The tracker high- p_T muon ID requires:

- an arbitrated tracker muon with muon segments in at least two muon stations
- a relative p_T error of the muon best track of less than 30%
- a tracker track with a transverse impact parameter $d_{xy} < 2$ mm w.r.t. the primary vertex, in order to suppress cosmic muons and suppress muons from decays in flight
- a longitudinal distance of the tracker track w.r.t. the primary vertex is $d_z < 5$ mm, in order to further suppress cosmic muons, muons from decays in flight and tracks coming from PU interactions
- at least one pixel hit to further suppress muons from decays in flight
- a cut on the number of tracker layers with hits above five in order to guarantee a good p_T measurement, for which some minimal number of measurement points in the tracker is needed. This cut also suppresses muons from decays in flight.

Muon isolation

The isolation criterium is requiring muon tracks to be isolated from other tracks and energy deposits in a cone $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ in order to distinguish between prompt muons and muons coming from weak decays within jets. Three relative isolations are used in this thesis. The first one is the standard loose relative isolation [97] and is a PF-based isolation defined as follows:

$$I_{rel}^\mu = \frac{1}{p_T^\mu} [I_{\text{charged hadrons}} + \max(0, I_{\text{neutral hadrons}} + I_{\text{photons}} - 0.5I_{\text{charged hadrons from PU}})] \quad (4.4)$$

where $I_{\text{charged hadrons}}$ corresponds to the scalar sum of the transverse momenta of all charged hadrons originating from the primary vertex and located in a cone $\Delta R = 0.4$ around the muon direction. To estimate the contribution of neutral particles from the hard event, the sum of the momenta of all neutral hadrons $I_{\text{neutral hadrons}}$ and photons I_{photons} is corrected by an estimate of the contribution of neutral particles from pile-up. This correction is estimated as half of the sum of the momenta of charged particles not associated to the primary vertex. This 0.5 factor corresponds approximately to the ratio of neutral to charged hadron production in the hadronization process of pile-up interactions. This isolation is used with a loose working point ($I_{rel}^\mu < 0.25$).

As aforementioned, the signal studied in this thesis is characterized by two charged leptons coming from the decay of a boosted Z . Therefore, those two charged leptons are close-by and enter in the isolation cone of each other. To deal with such cases, two specific isolations have been used. First, the relative isolation I_{rel}^μ described above is re-computed but with the subtraction of the particle momenta of charged hadrons, photons and neutral hadrons compatible (i.e. in a cone of $\Delta R = 0.1$) with an other muon identified as a *tracker high- p_T* muon within the isolation cone. The tight cut on this *cleaned relative isolation* is 0.15. The other relative isolation is a tracker based isolation. It computes the sum of all tracks momenta compatible with the primary vertex of this event in an isolation cone of $\Delta R = 0.3$ around the muon. From this sum, the momenta of other *tracker high- p_T* muons found in the cone are again removed to avoid contamination. The tight cut on this *cleaned tracker relative isolation* is 0.10.

4.4.3. Electron and photon reconstruction

The reconstruction of electrons and photons is based on information from energy deposits in the ECAL and tracks (or lack of tracks in the case of photons reconstruction) in the tracker [98].

Due to the combined effect of the magnetic field and bremsstrahlung, the energy deposit in the ECAL of an electron is spread, mainly in the ϕ -direction due to the direction of the magnetic field. On average 33% of the electron energy is lost before

reaching the ECAL at $\eta = 0$ and up to 86% at $\eta = 1.4$ where the material budget is large.

Although electron tracks could be reconstructed from tracker information using a Kalman Filter reconstruction, the large energy loss caused by photon radiation lowers the efficiency and precision of such a method and a custom reconstruction algorithm has been developed. This algorithm is made of two steps — seeding and tracking — and uses tracker PF tracks and ECAL PF clusters as inputs. The seeding consists of two complementary algorithms — the Tracker-Driven and the ECAL-Driven seedings — that are then combined through a fit.

The Tracker-Driven seeding is based on tracks (reconstructed with general algorithms for charged particles) that are matched to a supercluster after extrapolation towards the ECAL. A supercluster is defined as the result of clustering several PF clusters using the so-called "Mustache" algorithm which takes into account correlations with sub-cluster energy as a function of η and uses a dynamic $\Delta\phi$ window to reject clusters that are too energetic with respect to their deflection from the seed cluster.

The ECAL-Driven seeding starts from such a supercluster and selects electron seeds to extrapolate the trajectory towards the PV. The second step consist in the tracking and fitting step. Since the energy loss of electrons in the tracker does not follow a Gaussian distribution, as assumed by the KF algorithm, but a Bethe-Heitler distribution which has a larger tail, the Gaussian Sum Filter (GSF) algorithm is used instead. The GSF algorithm is designed to follow the track curvature accounting for the bremsstrahlung loss up to the ECAL and uses the hit collection obtained with a KF algorithm by approximating the Bethe-Heitler distribution with a sum of Gaussian distributions.

Tracks and superclusters are then matched to each other forming Tracker-Driven GSF Tracks and ECAL-Driven GSF Tracks seeds. Those seeds are afterwards linked to potential photon radiations and to their conversion into a e^+e^- pair and the related ECAL clusters. After linking as many PF objects as possible, clusters and tracks with bad E/p or matched to HCAL deposits are eventually removed. With this final list of objects, a refined supercluster is reconstructed and stored as an electron or photon candidate depending on its properties.

While the electron charge can be evaluated from the sign of the GSF track curvature, this leads to a charge misidentification of up to 10% for electrons at large pseudorapidity, because of bremsstrahlung followed by photon conversions. To reduce the charge misidentification rate, two other charge estimates are computed; the final electron charge is then the one given by at least two of the methods. The first alternative method uses the sign of the KF track associated to a GSF track if they share at least one innermost hit, whereas the second one defines the charge sign as the sign of the differences between the vector joining the beam spot to the supercluster position, and the vector joining the beam spot and the first hit of the electron GSF tracks.

The combination of the three charge estimates reduces the charge misidentification rate to 1.5% for reconstructed electrons from Z boson decays. The momentum of the electron is calculated from a weighted combination of the measurements from track parameters and from supercluster parameters. The first one is dominant for low energy candidates and the latter is dominant for high energy candidates.

The reconstruction described above is conducted for both electron and isolated photon at the same time and is deeply linked to the PF algorithm to form what is called a "Global Event Description" where each object is only assigned once to a particle by the PF algorithm. In this perspective, a photon candidate is seeded from an ECAL supercluster when the transverse energy measured in this supercluster is greater than 10 GeV and if no link to a GSF track has been found. On the other hand, an electron candidate is seeded from an GSF track provided that the corresponding ECAL supercluster is not linked to three or more additional tracks. This already allows to separate clear photon signature from electron signature. However, since photons can convert into a e^+e^- pair, further identification criteria have been defined and are detailed below along with criteria to discriminate electrons and photons from jets.

Electron Identification

Variables used to identify electrons could be categorized into three categories: those measuring the level of agreement between ECAL and tracker measurements, those based on calorimeters information only and those based on tracking information.

The variables used in the electron identification in this analysis are the following [99]:

- $\Delta\phi_{\text{in}}$ which represents the distance in ϕ between the ECAL supercluster and the track direction extrapolated from the PV,
- $\Delta\eta_{\text{seed}}$ which represents the distance in η between the seed of the ECAL supercluster and the track direction extrapolated from the PV,
- the ratio between hadronic (H) and electromagnetic (E) energy around the seed cluster (i.e. H/E),
- the width $\sigma_{i\eta i\eta}$ of the electron shower in the η direction, calculated from an array of 5×5 crystals around the energy deposit in the ECAL,
- the isolation I_{rel}^{EA} defined in next sub-section,
- $1/E - 1/p$, where E is the energy deposit in the ECAL and p the momentum measured from the track curvature,
- the number of missing hits in the reconstructed electron track,

- and a conversion veto filter to cut electrons coming from a photon conversion.

The conversion veto filter used in this analysis considers the GSF track of the electron candidate, and vetoes the electron candidate if its track matches to at least one track coming from a photon conversion candidate. Photon conversion candidate are sought for by performing vertex fits on pair of tracks, checking if one of these fits present the properties of a vertex due to photon conversion. In particular the goodness of the fit is checked as well as the number of hits in the pixel detector between the vertex and the beam spot. Eventually, the transverse decay length is required to be larger than a given value.

In this thesis, two working points are used: the tight working point, used to identify electron candidates coming from a Z decay, and the loose working point used to reject additional leptons in the event. The tight working point has a signal efficiency of about 70% for electrons from Z boson decays and rejects more than 99.5% of the background coming $t\bar{t}$ events. The loose working point has an efficiency of about 90% and reject 99% of the background. The cut value are described in Table 4.1.

Barrel ($ \eta_{\text{supercluster}} \leq 1.479$)		Loose ID	Tight ID
$\sigma_{i\eta i\eta}$	<	0.011	0.00998
$\Delta\eta_{\text{seed}}$	<	0.00477	0.00308
$\Delta\phi_{\text{in}}$	<	0.222	0.0816
H/E	<	0.298	0.0414
I_{rel}^{EA}	<	0.0994	0.0588
$ E-1/p $	<	0.241	0.0129
expected missing inner hits	\leq	1	1
pass conversion veto	=	yes	yes
Endcaps ($ \eta_{\text{supercluster}} > 1.479$)		Loose ID	Tight ID
$\sigma_{i\eta i\eta}$	<	0.0314	0.0292
$\Delta\eta_{\text{seed}}$	<	0.00868	0.00605
$\Delta\phi_{\text{in}}$	<	0.213	0.0394
H/E	<	0.101	0.0641
I_{rel}^{EA}	<	0.107	0.0571
$ E-1/p $	<	0.14	0.0129
expected missing inner hits	\leq	1	1
pass conversion veto	=	yes	yes

Table 4.1.: Details of the electron identification (ID) working points. The definition of the variables are given in the text [99].

Electron Isolation

The relative isolation used for electrons is similar to the one used for muons (see Eq. 4.5) but with a different estimation of pileup contribution. The relative isolation is defined as follows:

$$I_{rel}^{EA} = \frac{1}{p_T^e} [I_{\text{charged hadrons}} + \max(0, I_{\text{neutral hadrons}} + I_{\text{photons}} - \rho A_{\text{eff}})] \quad (4.5)$$

where $I_{\text{charged hadrons}}$ corresponds to the scalar sum of the transverse momenta of all charged hadrons originating from the primary vertex and located in a cone $\Delta R = 0.3$ around the electron direction. To estimate the contribution of neutral particles from the hard event, the sum of the momenta of all neutral hadrons $I_{\text{neutral hadrons}}$ and photons I_{photons} is corrected by an estimate of the contribution of neutral particles from pile-up. This correction is performed using effective areas A_{eff} and the transverse momentum density ρ of the event. The effective area [100] is an estimate of the area in the isolation cone that does not originate from the electron footprint. This type of pileup subtraction is similar to the one applied in the L1 part of jet energy corrections (see Section 4.4.4).

Photon Identification

The photon identification [101] is based on calorimeter only information (σ_{inin} as defined above and H/E in a single tower) and on ρ -corrected isolations from charged hadrons $I_{\text{charged hadrons}}^{EA}$, neutral hadrons $I_{\text{neutral hadrons}}^{EA}$ and photons I_{photon}^{EA} , which are defined in next sub-section. An additional criterium is defined to remove the fake rate coming from electrons by asking the photon candidate to not have a track seed in the pixel. This analysis only uses photons in the barrel passing the tight ID requirement. The signal efficiency of this tight ID is 71% and its background rejection efficiency is 89%. The signal efficiency is evaluated from Z boson decaying in ee where one of the electron showers is reconstructed as a photon using the tight ID without the conversion veto and matched to one of the electron from the decay. The background rejection efficiency is evaluated from $\gamma + \text{jets}$ simulated samples [102]. The tight ID working point cuts are described in Table 4.2.

Photon Isolation

The photon isolation is divided into three variables: one for charged hadrons, one for neutral hadrons and one for photons. Each of them is computed by summing the transverse momentum of PF objects (charged hadrons, or neutral hadrons, or photons) matching with the PV and in a cone $\Delta R = 0.3$ around the photon candidate. To take into account contributions from pile-up particles, those isolations are ρ -corrected

Barrel		Tight ID
single tower H/E	<	0.0269
$\sigma_{i\eta i\eta}$	<	0.00994
$I_{\text{charged hadrons}}^{EA}$	<	0.202
$I_{\text{neutral hadrons}}^{EA}$	<	$0.264 + 0.0148 * p_T + 0.000017 * p_T^2$
I_{photons}^{EA}	<	$2.362 + 0.0047 * p_T$

Table 4.2.: Details of the photons identification (ID) tight working point. The definition of the variables are given in the text [101].

using effective areas as described for the electron isolation. Therefore, the ρ -corrected isolation for a type X of particle is defined as follows:

$$I_X^{EA} = \max(0, I_X - \rho A_{\text{eff}}^X). \quad (4.6)$$

4.4.4. Jet reconstruction

As mentioned in Section 4.2, quarks and gluons coming from proton-proton collision hadronize and produce showers of collimated hadrons and other particles: this is referenced as *jets*. Due to this signature, the reconstruction of jets consists in correctly clustering PF-level particles into a jet. However when clustering hadrons into jets, several sources may affect the reconstructed energy of the jet and corrections have to be applied. Furthermore, the energy resolution of jets measured in data differs from the one in simulations, which need to be calibrated. Those three points (clustering, energy scale corrections and energy resolution correction) are detailed in the sub-sections below.

The jet anti- k_T clustering algorithm

The clustering algorithm used to cluster PF-level particles into a jet in this thesis is the anti- k_T algorithm [103, 104]. The anti- k_T algorithm is a sequential recombination algorithm made of the following steps:

1. list all the distances, d_{ij} , between particles (or pseudo-jets) i and j , given by:

$$d_{ij} = \min \left(\frac{1}{k_{T,i}^2}, \frac{1}{k_{T,j}^2} \right) \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2} \quad (4.7)$$

where k_T , y_i and ϕ_i are, respectively, the transverse momentum, rapidity and azimuthal angle of particle i . R is a cone (in $\eta - \phi$) parameter chosen to be 0.4 and represents the size of the reconstructed jets.

2. list all the distances, d_{iB} , between the particle (or pseudo-jet) i and the beam B , given by:

$$d_{iB} = \frac{1}{k_{T,i}^2} \quad (4.8)$$

3. find the minimum of all the distances computed in steps 1 and 2:
 - a) if the minimum is of type d_{iB} the corresponding particle (or pseudo-jet) forms a jet and is removed from the list of particles (or pseudo-jets)
 - b) if the minimum is of type d_{ij} the corresponding particles (or pseudo-jets) i and j are grouped together into a new pseudo-jet by adding the two corresponding four-vectors. This new pseudo-jet is then added to the list while its constituents i and j are removed from the list.
4. repeat the steps above until all particles have been assigned to a jet

By construction, this algorithm will cluster soft particles around hard ones before soft particles cluster among themselves, resulting in a jet of conical shape of radius R lead by hard particles but still including low- p_T particles.

Jet energy scale corrections

The reconstruction of the jet energy is perturbed by detector effect (dead zones, low- p_T particles trapped in the magnetic field, non-linearity of the detector response) and the presence of pile-up particles. Therefore and in order to have a proper mapping between the measured jet energy deposition and the particle-level jet energy, the CMS collaboration has implemented a set of jet energy scale corrections (JEC) [105]. Those corrections are factorized: each level of correction is taking care of a different effect and is a scaling of the jet energy with a factor (correction) depending on various jet related quantities (p_T , η , flavor, etc.). Also, the levels of correction are applied sequentially (the output of each step is the input to the next) and with fixed order.

- **Pile-up offset correction:** the goal of the L1 correction is to remove energy coming from pile-up particles and electronic noise which is expected to increase the jet energy. The hybrid jet area method uses the effective area of the jets multiplied by the average energy density in the event to calculate the offset energy to be subtracted from the jets. The average pile-up offset correction are determined from simulation of QCD di-jet events processed with and without pile-up. These corrections are parameterized as a function of the offset energy density, jet area, jet pseudorapidity and jet p_T . Remaining residual corrections between data and simulations are evaluated as a function of η using the random cone method on randomly triggered events (*zero-bias events*). The random cone method consists of reconstructing many jets in each event, clustering particles

in randomly placed cones, effectively mapping all the space. The average p_T of these jets is a measurement, in each event, of the average energy density that gets clustered in a jet. In particular, on zero-bias events, this method gives a measure of the noise and pileup energy contribution to the jet. Once this correction applied, any dataset dependence on luminosity is in principle removed.

- **Simulated response corrections:** due to the non-uniformity of the response of the CMS detector, the data-to-MC comparison is not flat versus η and p_T . This effect is corrected by scaling the reconstructed jet p_T with a function binned in η and p_T as follows:

$$p_T^{L2L3} = \left(\left\langle \frac{p_T^{L1, \text{sim QCD}}}{p_T^{\text{gen QCD}}} \right\rangle [\eta, p_T^{L1}] \right)^{-1} p_T^{L1} \quad (4.9)$$

where $\langle (\dots) \rangle [\eta, p_T^{L1}]$ represents the average for pre-defined bins in η and p_T^{L1} . These binned correction factors are determined from a simulated QCD multijet sample corrected for pile-up offset by comparing the reconstructed jet p_T to the particle-level one.

- **Residual corrections for data :** These last corrections are meant to correct for small remaining differences (of the order of the %) in the jet energy response as a function of the jet p_T and η in data and simulations. They are separated into two corrections: an η -dependent correction and a correction to the jet absolute scale. The η -dependent correction is measured in QCD di-jet events where the two jets are back-to-back in the transverse plane with the additional requirement to have at least one of the jet in the central region of the detector ($|\eta| < 1.3$) where the detector response is uniform. The correction is computed by requiring a balance in p_T between the two jets, the difference in the energy between the two jets being used to correct the jet energy dependence in η . The jet absolute scale correction is obtained from $Z/\gamma(\ell\ell) + \text{jet}$ where the reconstructed Z/γ boson is back-to-back with the jet. The resolution on the transverse momentum of the lepton pair being better than the resolution on the jet energy, it is used to estimate the transverse momentum residual of the jets as a function of the jet p_T .

It should be noted that besides the corrections detailed above, the jets used in this analysis have been further mitigated for pile-up effects by using the so-called CHS jets. CHS stands for Charged Hadron Subtraction and corresponds to jets where charged hadrons associated to vertices associated to pile-up (i.e. not associated with the PV of the collision) are removed from the list of the PF particles on which the jet clustering algorithm is applied.

Jet energy resolution correction

Measurements show that the jet energy resolution (JER) in data is worse than in the simulation and therefore a combination of two methods is used to smear reconstructed jets in simulation so that their p_T resolution matches the one observed in data.

When a reconstructed jet matched a particle-level jet, the *scaling method* is used, otherwise the *stochastic smearing* is used instead.

The scaling method alters the jet p_T by rescaling it with

$$c_{\text{JER}} = \max \left(0, 1 + (s_{\text{JER}} - 1) \frac{p_T - p_T^{\text{gen}}}{p_T} \right), \quad (4.10)$$

where p_T is the reconstructed jet transverse momentum, p_T^{gen} is the transverse momentum of the corresponding jet clustered from generator-level particles, and s_{JER} is the data-to-simulation resolution scale factor which is determined in bins of η . The following requirements are imposed for the matching:

$$\Delta R < R_{\text{cone}}/2 \quad \text{and} \quad |p_T - p_T^{\text{gen}}| < 3\sigma_{\text{JER}}p_T \quad (4.11)$$

with $R_{\text{cone}} = 0.4$, the jet cone size parameter, and σ_{JER} the relative p_T resolution as measured in simulation.

In the stochastic smearing case, the scale factor c_{JER} is estimated by

$$c_{\text{JER}} = \max \left(0, 1 + \mathcal{N}(0, \sigma_{\text{JER}}) \sqrt{\max(0, s_{\text{JER}}^2 - 1)} \right), \quad (4.12)$$

with σ_{JEJERR} and s_{JER} defined as above and $\mathcal{N}(0, \sigma_{\text{JER}})$ denoting a random number sampled from a normal distribution with mean zero and variance σ_{JER}^2 .

Jet identification

The PF jet identification [106] is designed to reject fake jets originating from noise and reconstruction failures while keeping 98–99% of real jets. The criteria used to discriminate between a noise jet and a real jet are the number of PF candidates reconstructed in the jet, called constituents, and the jet energy fraction distribution among the reconstructed jet constituents. These jet constituents are grouped into charged ElectroMagnetic (for muons and electrons), neutral EM (for photons), charged hadron and neutral hadron candidates. Note that the variables related to charged candidates only extend up to $|\eta| < 2.4$ where the tracker coverage ends. Also the energy fractions have to be computed before any JEC is applied, otherwise the fractions don't add up to unity.

The tight working point used in this analysis is described in Table 4.3.

Jet variables	$ \eta $ range	Tight ID
Charged hadron fraction	$ \eta \leq 2.4$	> 0
Charged multiplicity	$ \eta \leq 2.4$	> 0
Charged EM fraction	$ \eta \leq 2.4$	< 0.99
Neutral hadron fraction	$ \eta \leq 2.7$	< 0.90
Neutral EM fraction	$ \eta \leq 2.7$	< 0.90
Number of constituents	$ \eta \leq 2.7$	> 1
Neutral EM fraction	$2.7 < \eta \leq 3.0$	> 0.01
Neutral hadron fraction	$2.7 < \eta \leq 3.0$	< 0.98
Number of neutral particles	$2.7 < \eta \leq 3.0$	> 2
Neutral EM fraction	$ \eta > 3.0$	< 0.90
Number of neutral particles	$ \eta > 3.0$	> 10

Table 4.3.: Details of the jets identification (ID) tight working point. The definition of the variables are given in the text [107].

Identification of b quark jets

The hadronization and decay characteristics of light quarks (u , d and s) and gluons are almost identical and so are the jets originating from them. Top quarks decay before hadronizing and are not directly subject to jet reconstruction. However, jets originating from intermediate heavy quarks (b and c) show distinct properties and can be differentiated from other jets. In particular, B mesons have a typical lifetime of $c\tau \approx 500 \mu\text{m}$ and leads to tracks from its decay products pointing to its decay position, forming a displaced *secondary vertex* with respect to the interaction point. The distance of the closest approach in the transverse plane of the track to the interaction point is called the impact parameter. The impact parameter significance of a track is often used instead and is defined as the impact parameter of the track divided by its uncertainty and gives a better estimation of the track displacement. This analysis used the so-called *Combined Secondary Vertex* (CSVv2) algorithm [108] to veto background events with b -jets (such as $t\bar{t} \rightarrow bW^+\bar{b}W^-$ events). This algorithm combines information from displaced tracks and secondary vertices in a multivariate analysis using a neural network [109] and reconstructing secondary vertices with the Inclusive Vertex Finding (IFV) algorithm [110, 111]. The CSVv2 neural network is trained on QCD multi-jets and $t\bar{t}$ events in three independent vertex categories and bins of jet p_T and η . The three vertex categories are the reco-vertex (at least one well reconstructed secondary vertex), the pseudo-vertex (jets without a well reconstructed secondary vertex but with displaced tracks with an impact parameter significance above 2) and the no-vertex categories. The variables entering in the neural network are, for the vertex related variables :

- the invariant mass of the first¹ reconstructed secondary vertex
- the flight distance significance in the transverse plane of the first reconstructed secondary vertex
- the number of secondary vertices in the jet

and for the tracking related variables:

- the number of tracks in the jet
- their impact parameters
- the ratio of the energy carried by tracks at the vertex with respect to all tracks in the jet

In this analysis, the loose working point of the CSVv2 discriminator is used and corresponds to a rate of misidentifying a light jet as a b jet of 10% while having a global efficiency of $\sim 83\%$ to correctly identify a b -jet. Figure 4.3 shows the performance of the b -jet identification for different algorithm.

High-level object disambiguation

It should be noted that charged lepton and photon candidates can also be reconstructed as jets in addition of being reconstructed as a charged lepton or photon. To resolve this ambiguity, the jet collection passing tight criteria is cleaned from jet candidates that have an axis closer than $\Delta R = 0.4$ from a charged lepton or photon passing tight identification and isolation criteria. Similarly, the collection of photons passing tight identification and isolation criteria is cleaned from photon candidates with a momentum vector closer than $\Delta R = 0.1$ from any charged lepton passing tight identification and isolation criteria.

4.4.5. Missing transverse momentum reconstruction

The PF missing transverse momentum (\vec{p}_T^{miss}) in an event is the imbalance in the sum of the transverse momenta $\vec{p}_{T,i}$ of all reconstructed particles by the PF algorithm [113] and is defined as follows:

$$\vec{p}_T^{\text{miss}} = - \sum_{i=1}^{N_{\text{particles}}^{\text{PF}}} \vec{p}_{T,i}. \quad (4.13)$$

¹The secondary vertices are sorted in ascending order of uncertainty on the flight distance.

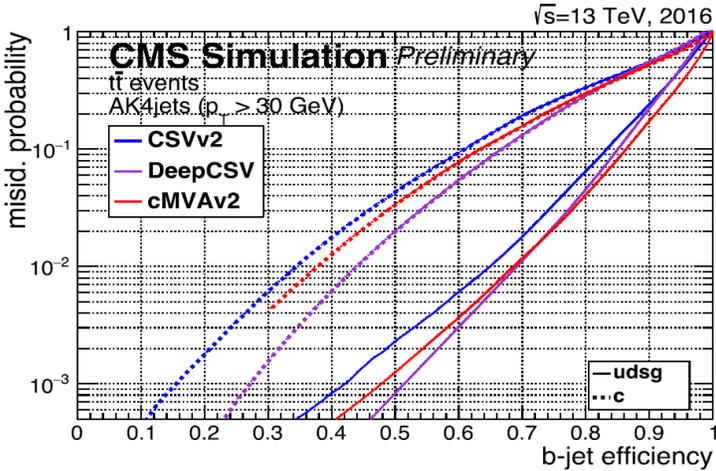


Figure 4.3.: Performance of the b -jet identification efficiency algorithms demonstrating the probability for non b -jets to be misidentified as b -jet as a function of the efficiency to correctly identify b -jets. The curves are obtained on simulated $t\bar{t}$ events using jets within tracker acceptance with $p_T > 30$ GeV, b -jets from gluon splitting to a pair of b -quarks are considered as b -jets. The lines shown are for CSVv2, DeepCSV, and cMVA2 algorithms. The performance in this figure serves as an illustration since the b -jet identification efficiency depends on the p_T and η distribution of the jets for the topology as well as the amount of b -jets from gluon splitting in the sample [112].

Due to momentum conservation in the transverse plane, \vec{p}_T^{miss} is the transverse momentum that must have been carried by something invisible to the CMS detector, namely neutrinos or other hypothetical neutral weakly interacting particles. However, the measure of \vec{p}_T^{miss} depends on the measurement and reconstruction of all other particles (i.e. leptons, photons and jets) in the event and is affected by other sources, such as mismeasurement or misidentification of particles, detector noise or malfunctions, pileup interactions...

In order to reduce the bias in the \vec{p}_T^{miss} measurement, several corrections are applied.

Type-I correction

The Type-I correction is a propagation of the jet energy corrections (JEC) to \vec{p}_T^{miss} to account for non-linearity of the response of the calorimeter, minimum energy thresholds in the calorimeters, etc. The Type-I correction replaces the vector sum of

transverse momenta of particles which can be clustered as jets with the vector sum of the transverse momenta of the jets with JEC applied.

Smearing correction

Similarly to jet energy corrections, jet energy smearing is also propagated to the missing transverse momentum by replacing the corresponding PF jet p_T by its smeared p_T .

ϕ modulation correction

The distribution of \vec{p}_T^{miss} is independent of ϕ for physics sources of missing transverse momentum such as neutrinos due to rotational symmetry of the collisions around the beam axis. However, it has been observed that the reconstructed \vec{p}_T^{miss} has modulation in ϕ (roughly looking like a sinusoidal curve with period of 2π). This can be due to many sources such as anisotropic detector responses, inactive calorimeter cells, misalignment, displacement of the beam spot... The ϕ modulation correction shifts the x and y value of the \vec{p}_T^{miss} as a function of the scalar sums of the p_T of charged hadrons, neutral hadrons and photons and in bins of pseudorapidity [114].

Anomalous high- \vec{p}_T^{miss}

Artificially large \vec{p}_T^{miss} can be measured due to spurious detector signals, such as particles striking sensors in the ECAL or HCAL, beam-halo² particles or ECAL dead cells. Such events are removed thanks to dedicated algorithm (filters) designed by the CMS collaboration.

4.4.6. High-level object energy resolution summary

This last section of the reconstruction chapter aims to give a quick overview of the energy resolution on the high-level objects discussed in the current chapter and that will be used extensively in the analysis presented in this thesis.

Muon momentum resolution

Already mentioned in chapter 3 and in particular on Fig. 3.25, the momentum resolution on reconstructed muons is of the order of 1% for low- p_T muons (< 50 GeV) and goes up to 6% for muons at the TeV scale.

²Protons from the beam can undergo a collision upstream of the detector which could produce muons which travel nearly parallel to the collision axis due to the magnetic field in the LHC. Such particles are often called *beam-halo particles*.

Electron momentum resolution

The effective momentum resolution on reconstructed electrons is shown on Fig. 4.4. On this figure, two categories of electrons are defined:

- *Golden* electrons are reconstructed electrons with a small fraction of their energy radiated by bremsstrahlung (less than 0.5) and consequently provide the most accurate estimation of momentum. They are also required to have a super-cluster (SC) made of a single cluster.
- *Showering* electrons on the other hand have a large fraction of their energy radiated by bremsstrahlung (more than 0.5) radiated all along the electron trajectory, and are defined by a SC containing several clusters.

For golden electrons in the barrel, the resolution is of the order of 2% while for showering electrons in the endcaps the resolution is of the order of 10% at a p_T of 10 GeV. The resolution improves quickly with increasing p_T to reach a resolution of the order of 4% at 100 GeV. One can also see that, as it was already hinted by Eq. 3.7 and Fig. 3.13, the resolution is dominated by the ECAL measurement for electrons with a p_T above 30 GeV.

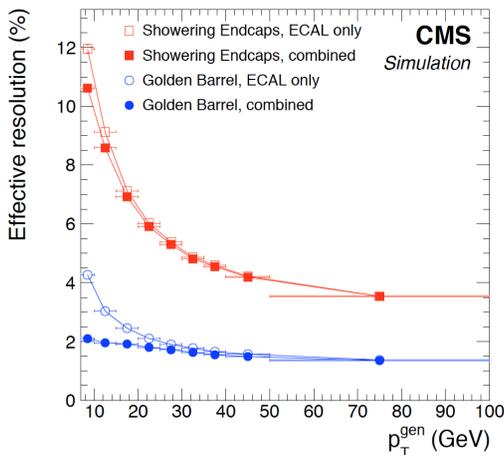


Figure 4.4.: Effective resolution in electron momentum after combining the ECAL and tracker informations (solid symbols), compared to that of the ECAL only (open symbols), as a function of the generated electron p_T for golden and showering electrons (see text) [115].

Photon energy resolution

Figure 4.5 shows the relative photon resolution computed in simulation using $H \rightarrow \gamma\gamma$ events passing a pre-selection as detailed in [102]. Events have been divided in two R_9 categories, defined as the energy sum of 3x3 crystals centered on the most energetic crystal in the supercluster, normalized by the energy of the supercluster. R_9 discriminates between unconverted, or late converted photons, and converted photons. The resolution is of the order of 1% in the barrel and 3% in the endcaps.

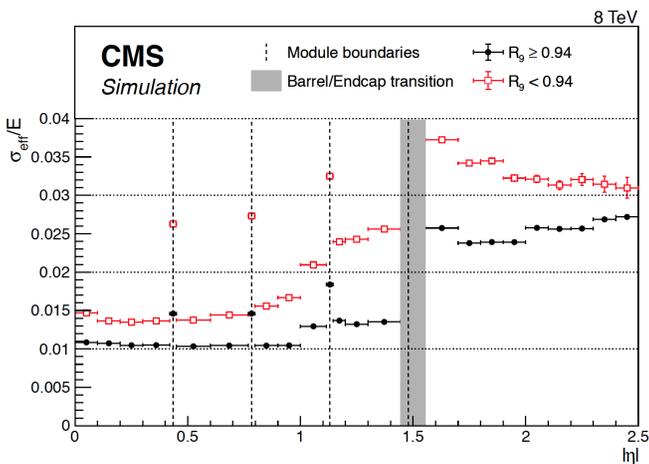


Figure 4.5.: Relative energy resolution, σ_{eff}/E , as a function of $|\eta|$, in simulated $H \rightarrow \gamma\gamma$ events, for photons with $R_9 \geq 0.94$ (solid circles) and photons with $R_9 < 0.94$ (open squares). The vertical dashed lines mark the module boundaries in the barrel, and the vertical grey band indicates the range of $|\eta|$, around the barrel/endcap transition, removed from the fiducial region [115].

Jet energy resolution

Figure 4.6 shows the relative jet energy resolution (JER) as a function of its p_T and the average number of pile-up interactions (μ). The JER is stable against μ for jets above 100 GeV while a degradation of up to 50% is observed at low- p_T (20 GeV) for very high pile-up.

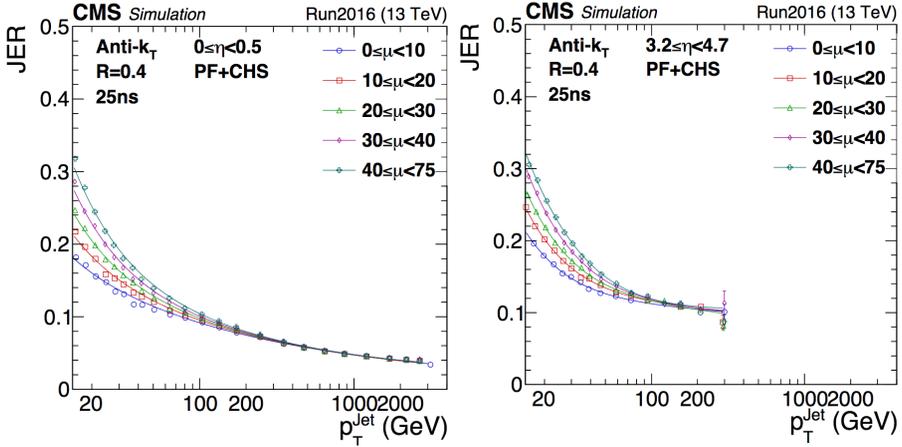


Figure 4.6.: Relative jet energy resolution (JER) as a function of the average number of pile-up interactions (μ) and the jet p_T , for jets in $0 \leq \eta < 0.5$ (left) and $3.2 \leq \eta < 4.7$ (right) [116].

Missing transverse momentum resolution

By definition of the \vec{p}_T^{miss} and as a consequence of the resolution on the objects defined above, the \vec{p}_T^{miss} in a $Z/\gamma + \text{jets}$ event is dominated by the hadronic activity in this event. In order to quantify the resolution of the \vec{p}_T^{miss} , its parallel u_{\parallel} and orthogonal u_{\perp} components with respect to the p_T of the Z boson are measured. Figure 4.7 shows the absolute resolution of this two components in data and simulation as a function of the p_T of a Z boson or photon.

The resolutions measured in the different samples are in good agreement and are found to be increasing with the p_T . The isotropic nature of energy fluctuations, such as detector noise and underlying event, causes the perpendicular component of the recoil energy to have a more stable resolution compared to the parallel component.

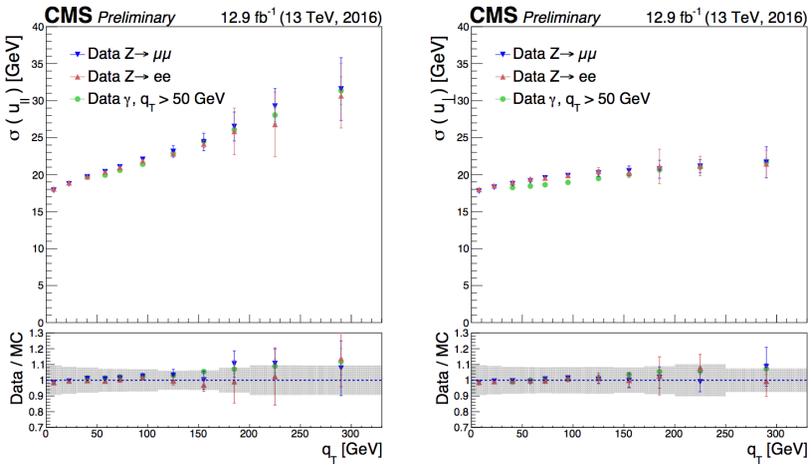


Figure 4.7.: Absolute resolution on the parallel component of the recoil energy u_{\parallel} (left) and its perpendicular component u_{\perp} (right) to the Z/γ boson. The upper frame shows the resolution in data; the lower frame shows the ratio of data to simulation with the error band displaying the systematic uncertainty of the simulation [113].

Chapter 5

Search for a heavy scalar boson in the $ZZ \rightarrow 2\ell 2\nu$ channel

This chapter presents the search for a heavy scalar boson in the $ZZ \rightarrow 2\ell 2\nu$ channel using 2016 data. It starts with a foreword putting my contributions into perspective before delving into the details of the analysis. First, the signal construction is presented. Second, the trigger requirements are shown. Then the backgrounds to this search are presented, followed by the corrections to the simulated signal and background samples. Next, the analysis strategy is detailed and data-to-simulation comparisons are shown followed by the estimations of the data-driven backgrounds and uncertainties. Eventually, limits are set on the $pp \rightarrow H \rightarrow ZZ$ cross section as a function of the scalar boson mass and width.

5.1. Foreword

This thesis consists in the search for a heavy scalar boson in the $ZZ \rightarrow 2\ell 2\nu$ channel using data collected by the CMS experiment at the LHC with proton-proton collisions at 13 TeV using a dataset corresponding to an integrated luminosity of 35.9 fb^{-1} collected in 2016. Performing such a search requires to master various subjects both on the technical, experimental and theoretical points of view. Therefore it has to be stressed that to achieve the results presented in this chapter a group effort was needed.

On a wide scope, one could say that the full CMS collaboration was needed in order to be able to take data successfully, to reconstruct physics objects, to investigate and fix issues observed during data taking and after object reconstruction, to provide corrections for physics objects used in analysis, etc. Each member of the CMS col-

laboration is performing those tasks for the full collaboration and without this global involvement performing any analysis would be impossible. In my personal case, I've worked on the monitoring of the data quality recorded by the tracker and also coordinated the related developments.

On a smaller scope, this work was done inside the $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ group which involves a dozen persons from the “Université libre de Bruxelles” (Belgium), the “Université catholique de Louvain” (Belgium), the “University of Delhi” (India) and the “Beihang University” (China). My work in this group could be divided in three periods:

- **Analysis of 2015 data:** an analysis was performed by the $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ group with 2.3 fb^{-1} of data recorded by CMS in 2015. Being still a newcomer at the time of this analysis, I have first taken care of applying corrections recommended by the CMS collaboration on reconstructed objects and technical aspects of the analysis. In parallel I have studied in more details the jets and the effect of the pileup in this analysis. More details can be found in [117].
- **Analysis of 2016 data:** my core contributions to the $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ group was for the analysis of 2016 data, hence this manuscript is focusing on it. In particular, I have contributed to the implementation of higher-order corrections to the ZZ background, have produced the observables that were published and have taken care of the estimation of a specific background of this search coming mainly from detector effects (Z +jet background). This analysis is presented in detail in the present chapter. A combination analysis with two other decay channels ($H \rightarrow ZZ \rightarrow 4\ell$ and $H \rightarrow ZZ \rightarrow 2\ell 2q$) was then performed [27].
- **New analysis framework:** the framework used to perform 2015 and 2016 analyses was inherited from the run 1 of the LHC and became hard to optimize and debug. Furthermore, it was not meant to run on such a big amount of data and performing a single update would take a week. After publishing the results based on the 2016 dataset, it was therefore decided to rewrite this framework, partly based on a framework developed by V +jet analysis: SHEARS. I have worked in the technical implementation of this framework and also took this opportunity to study into even more details the estimation of the background coming from detector effects. My study of this background with 2016 data in this new framework is shown in chapter 6.

5.2. Signal simulation

A model-independent search for a high-mass scalar ($J^P = 0^+; M_H > 125 \text{ GeV}$) with variable mass and total decay width, in the $H \rightarrow ZZ \rightarrow \ell^- \ell^+ \nu \bar{\nu}$ final

state (where ℓ is an electron e or a muon μ), is performed. This analysis is using the full proton-proton data recorded by CMS in 2016 which was validated for physics for a total of 35.9 fb^{-1} .

The signal is simulated starting from samples of events corresponding to heavy scalars with variable masses between 300 and 3000 GeV and SM-like couplings [118] fixing their width. These samples are simulated with the Powheg generator. Then, the distributions are reweighted thanks to the MELA package [75–78] to describe a signal with various width scenarii (5, 10 and 100 GeV), accounting also for the effect of interferences with the ZZ background continuum. This continuum includes the contribution of off-shell production and decay of the already-discovered scalar, that we will further denote h_{SM} . As explained in Chapter 2, two production processes are considered: gluon fusion and vector boson fusion. The fraction of signal events produced by each process is left free in the analysis or is set to be VBF-only.

5.2.1. The MELA reweighting procedure

The simulated samples production was in the hands of Alessio Magitteri (Université catholique de Louvain, Belgium) inside the $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ group. A more detailed description of the MELA reweighting and interferences can be found in his PhD thesis [119].

The Matrix Element Likelihood Approach (MELA) package based on the JHU Generator (for the SM and BSM scalar boson signal) and MCFM (for the ZZ background) can be used for event reweighting and building kinematic discriminants through the computing of LO matrix elements. In this analysis, it was used for reweighting signal events since having two neutrinos in the final state permits less kinematic discriminating power for the signal than in the $H \rightarrow ZZ \rightarrow \ell^+ \ell^- \ell'^+ \ell'^-$ final state. Mainly to avoid simulations of many signal samples, the reweighting procedure was used:

- to produce a signal sample with a given scalar boson width
- to take into account interferences between the signal and other processes.

The basis of the procedure is that for an event above the on-shell ZZ threshold, the probabilities for the signal $P_S(\vec{\Omega}|M_H; \Gamma_H)$, background $P_B(\vec{\Omega}|M_H; \Gamma_H)$ and signal-plus-background-and-interferences $P_{SBI}(\vec{\Omega}|M_H; \Gamma_H)$ for an event to occur giving its kinematical variables $\vec{\Omega}$ can be computed by MELA for a given scalar mass and width as described in [75, 76, 120, 121]. Therefore, these probabilities are computed event-by-event for both the SM scenarii and for the tested $(M_H; \Gamma_H)$ scenarii as a function of $\vec{\Omega}$. Then the ratios of these weights are computed and used to reweight the Powheg samples to the wanted signals. In the VBF case, the information of the additional jets in the final state is added to $\vec{\Omega}$.

The next and last step in the reweighting procedure is the normalizations of the signal and interference contributions. These normalizations are computed consistently by computing the scale factor that must be applied to the reweighted background-only sample to match the known $pp \rightarrow ZZ$ cross section above the ZZ threshold.

5.2.2. Impact of interferences

In the SM, four processes also have $2\ell 2\nu$ in the final state and cause interferences: $pp \rightarrow ZZ$, $pp \rightarrow WW$, $pp \rightarrow h_{\text{SM}} \rightarrow ZZ$ and $pp \rightarrow h_{\text{SM}} \rightarrow WW$. It has been shown that WW contributions can be neglected above $M_H \gtrsim 400$ GeV when applying analysis cuts and they won't be considered in this analysis [122].

Figure 5.1 shows the distribution of the invariant mass for a signal with $M_H = 800$ GeV and $\Gamma_H = 5$ GeV in red (X in the caption). The $pp \rightarrow ZZ$ (or continuum), the $pp \rightarrow h_{\text{SM}} \rightarrow ZZ$ and their interference are represented in blue (named Cont + h_1 + Interf_{Cont- h_1} in the caption) while the $pp \rightarrow ZZ$, the signal and their interference are shown in gray (Cont + X + Interf_{Cont-X} in the caption). Finally all the backgrounds, the signal and their interferences are shown in green (Cont + h_1 + X + Interf_{All}).

The bottom plots of these figures represent the absolute number of events due to interferences only. Interf_{Cont-X} corresponds to the interferences between the continuum and the heavy scalar boson and is computed by subtracting the continuum and the signal from the Cont + X + Interf_{Cont-X} line. Similarly, Interf_{All} corresponds to the interferences between the continuum, the signal and the $pp \rightarrow h_{\text{SM}} \rightarrow ZZ$ processes and is computed by subtracting the continuum, the signal and the $pp \rightarrow h_{\text{SM}} \rightarrow ZZ$ processes from the Cont + h_1 + X + Interf_{All} line.

Figures 5.2 and 5.3 represent the same distributions for respectively signals with $(M_H; \Gamma_H) = (800 \text{ GeV}; 100 \text{ GeV})$ and $(M_H; \Gamma_H) = (1500 \text{ GeV}; 100 \text{ GeV})$. Comparing the three figures, one can see the variation of the signal shape when varying its width and mass as well as the effects of interference. In particular one can see that increasing the width of the signal increases the region where interferences occur.

One can see on the bottom of this figure that the interference between the $pp \rightarrow ZZ$ background and the signal (in orange) increases the left tail of the peak while reducing the right tail. However, since the off-shell $pp \rightarrow h_{\text{SM}} \rightarrow ZZ$ background can reach high invariant M_{ZZ} , it can interfere with the signal. The interference between all the processes including the h_{SM} production (in green) is negative on the left of the peak and positive on the right when close to the peak, tending towards 0 at high invariant mass. This behaviour is compatible with what is reported on Figure 147 of [118] for the 4ℓ final state.

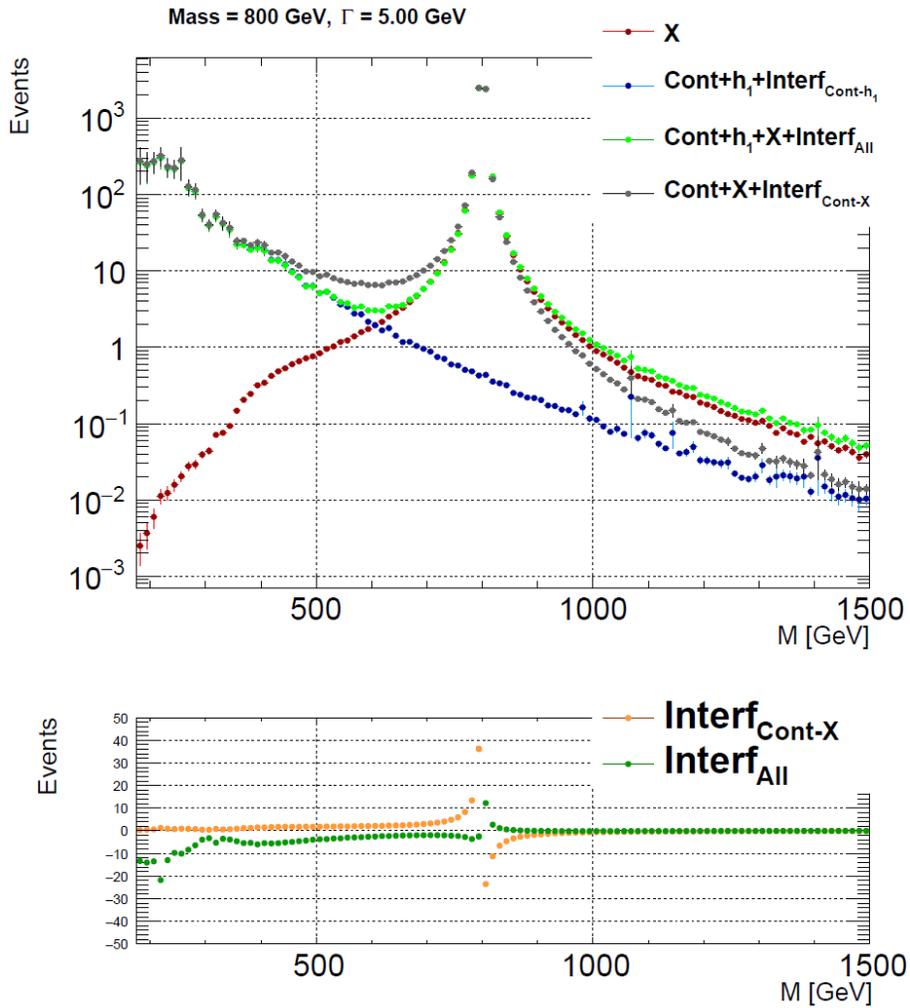


Figure 5.1.: Invariant mass distribution for a signal with $M_H = 800$ GeV and $\Gamma_H = 5$ GeV showing the various background contributions and interferences. Full description in the text [119].

Eventually, one can also see that the shape of the peak for signal only (in red) and the shape of the peak when accounting for all interferences (in green) are not very different. Also, note that the effect of the signal and its interferences with the other processes has been made more visible by an arbitrary large choice of cross section for the signal.

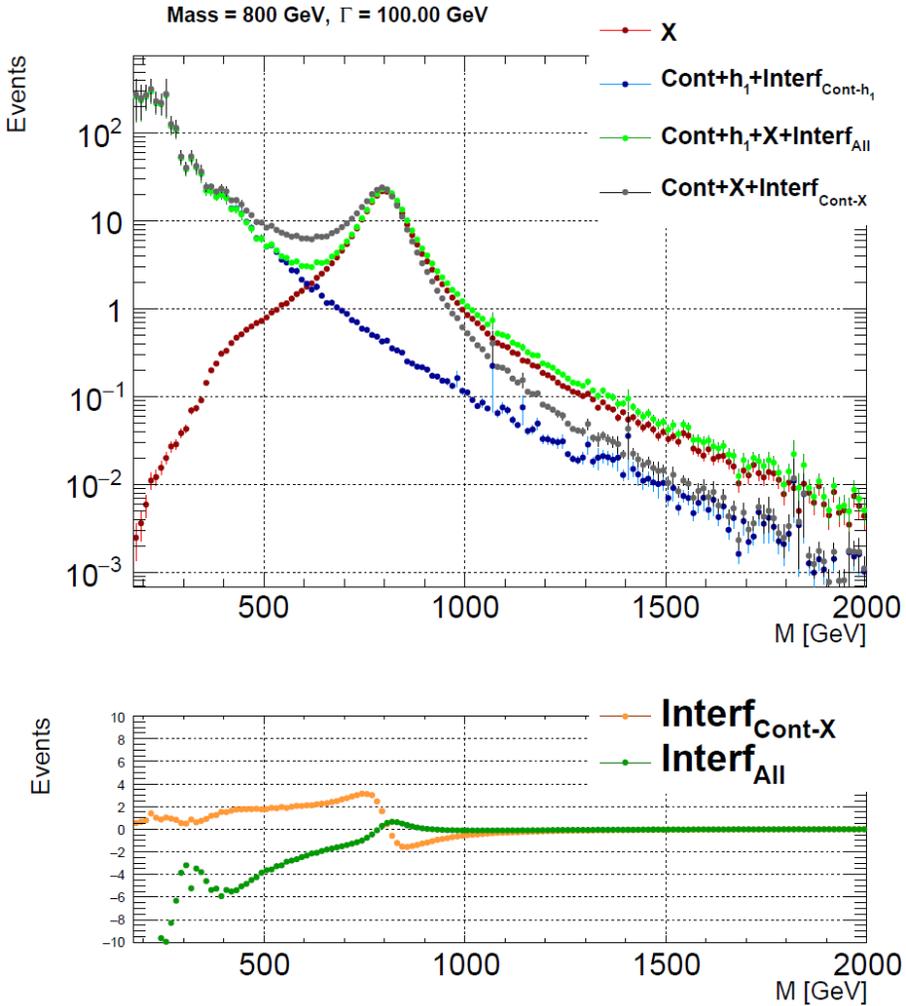


Figure 5.2.: Invariant mass distribution for a signal with $M_H = 800$ GeV and $\Gamma_H = 100$ GeV showing the various background contributions and interferences. Full description in the text [119].

5.3. Trigger requirements

Table 5.1 shows the lepton triggers used in this analysis. They can be split in two categories: double lepton triggers and single lepton triggers. Those double (single) triggers require two (one) HLT electron/muon with various p_T thresholds. In the case of double muon triggers, they require a global muon above 17 GeV and a global or

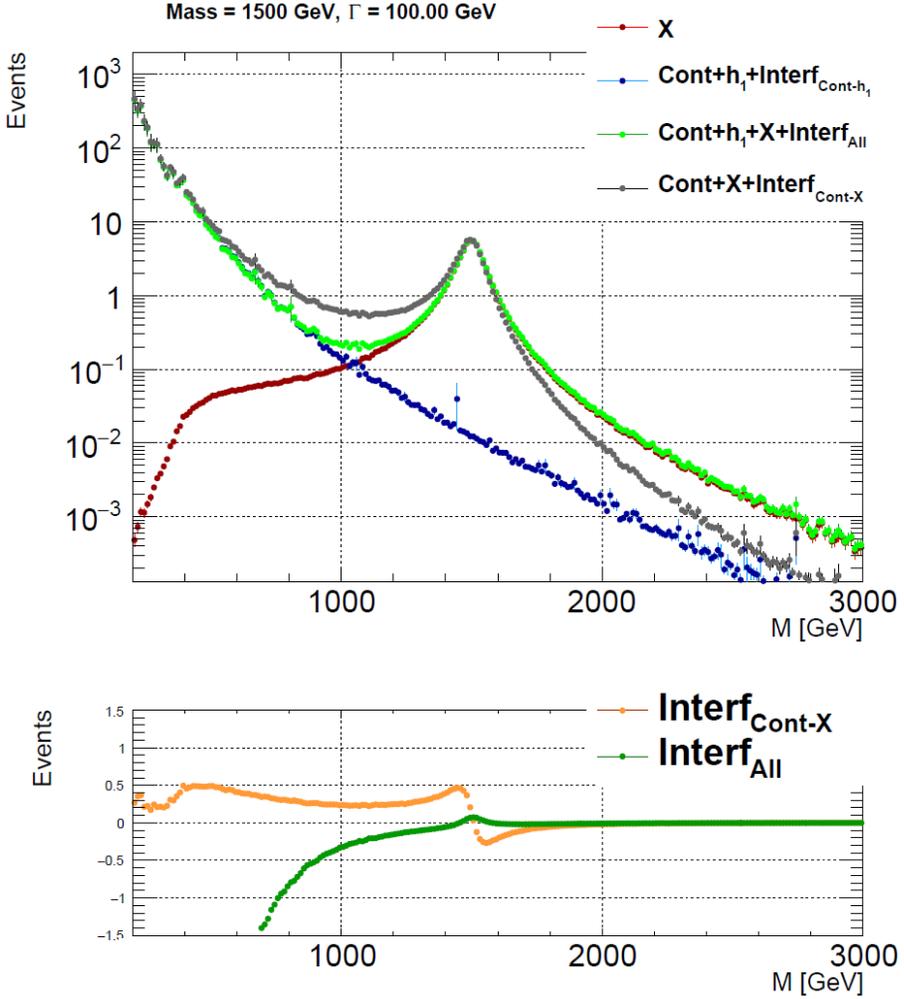


Figure 5.3.: Invariant mass distribution for a signal with $M_H = 1500$ GeV and $\Gamma_H = 100$ GeV showing the various background contributions and interferences. Full description in the text [119].

tracker muon above 8 GeV both passing a very very loose track isolation criteria. For some triggers it asks a loose compatibility between the longitudinal impact parameters d_z of the two lepton tracks. Single muon triggers require an isolated tracker or global muon above 22 or 24 GeV.

Double electron triggers require two electrons above 23 and 13 GeV respectively or two electrons above 33 GeV both passing loose calorimeter identification criteria. For the lowest p_T trigger, additional criteria are set asking for loose track identifica-

tion, very loose isolation and loose compatibility with the beamspot. Single electron trigger requires an electron above 27 GeV passing tight isolation and identification criteria. The DoubleEle33 trigger has been added to recover some events with electrons of $p_T > 400$ GeV [123].

Transverse momentum requirements are tighter on electron triggers compared to muon triggers because electrons have a less clear signature than muons, especially at the level of the Level-1 trigger.

Trigger category	Trigger path
DoubleMu	Mu17_TrkIsoVVL_Mu8_TrkIsoVVL
	Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL
	Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ
	Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ
SingleMu	IsoMu24
	IsoTkMu24
	IsoMu22
	IsoTkMu22
DoubleElectron	Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
	DoubleEle33_CaloIdL
SingleElectron	Ele27_WPTight_Gsf

Table 5.1.: Double and single lepton paths used in the analysis. More details in the text.

Trigger efficiencies are estimated by a Tag&Probe method. This method is the standard method used to study efficiency for resonances in CMS. It requires one lepton to pass tight criteria (tag lepton) and then look at the other lepton (probe lepton) of the reconstructed resonance (here a Z boson), requiring looser criteria. The efficiency of the trigger relative to the offline selection is then assessed by counting the number of probe leptons passing the trigger requirement. In this way, the total efficiency is the product of the offline selection efficiency and the trigger efficiency relative to this offline selection. In the case of double electron triggers, this method is done twice — once for each electron required by the trigger — and the efficiencies for both electrons are multiplied. For double muon triggers, the efficiency of a looser single muon trigger is first estimated by Tag&Probe. Then, taking reconstructed Z bosons passing this trigger, the number of probes passing the double muon trigger is counted and the global efficiency is computed by multiplication.

Double muon (electron) triggers have an efficiency to tag di-muon (di-electron) events of about 90% (94%) as shown on Fig. 5.4.

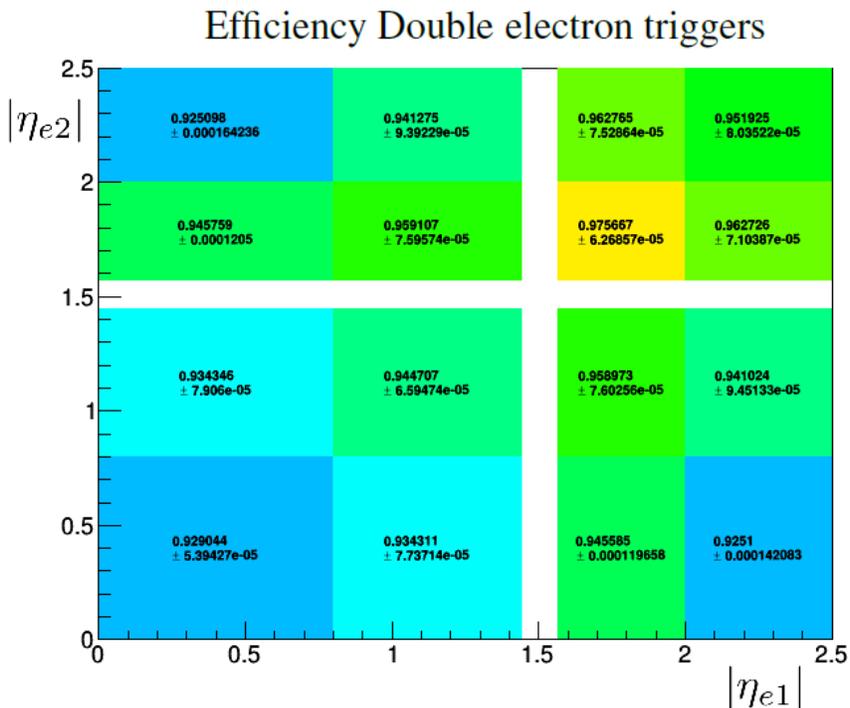
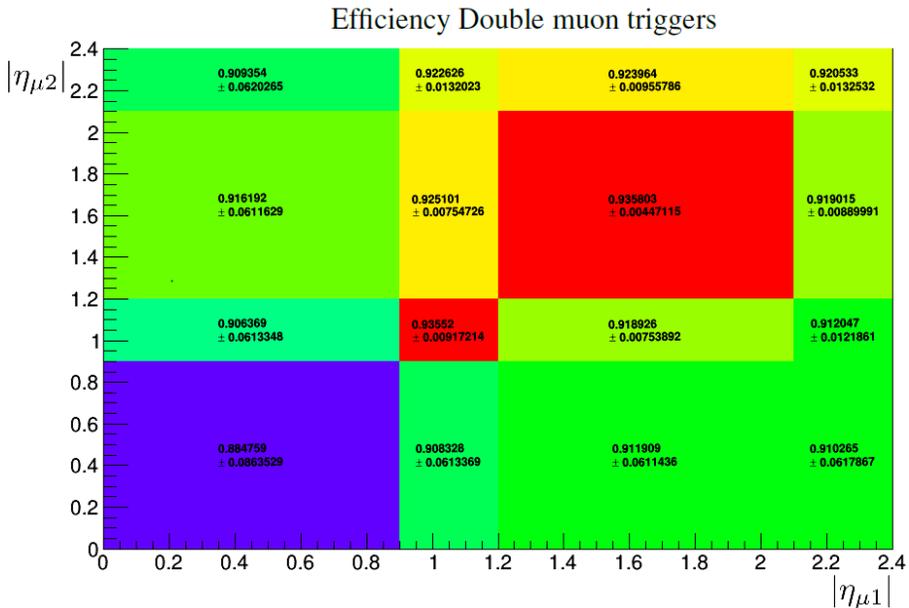


Figure 5.4.: Efficiency of the double muon (top) and electron (bottom) triggers estimated by the Tag&Probe method using the complete 2016 data as a function of the pseudorapidity of the leading and subleading leptons.

In order to reach an efficiency close to 100% to tag di-lepton events, events passing single lepton triggers have to be added to the di-lepton dataset. To avoid double counting events passing several trigger requirements, single lepton data sample events are required to not trigger double lepton triggers. Furthermore, while muon data sample events undergo the sole requirement of passing a muon trigger, electron sample events have to not pass any muon trigger in addition of passing an electron trigger. Figure 5.5 shows the global efficiency when taking events from double and single lepton triggers.

5.4. Backgrounds

The main backgrounds in the $2\ell 2\nu$ final state can be divided into three categories:

- $ZZ/Z\gamma^* \rightarrow 2\ell 2\nu$ and $WZ \rightarrow 3\ell\nu$: containing a Z boson and missing transverse energy coming from neutrinos, these two backgrounds are the main backgrounds of this analysis. Indeed, due to their final state similar to the signal they can hardly be disentangled from it. They are estimated from simulations, adding higher-order QCD and electroweak corrections. From now on, $ZZ/Z\gamma^*$ processes will be written as ZZ processes.
- $Z + \text{jets}$: containing two leptons in the final state coming from a Z boson decay, this background has a signature similar to the signal except that it does not contain a genuine source of missing transverse energy. It is therefore expected that applying a selection on \vec{p}_T^{miss} -related variables will suppress this background. However, its cross section is large and fake sources of missing transverse energy (such as mismeasurements of jets) are hard to be simulated with high accuracy which make this background an important one to control. The precise (data-driven) estimation of this background is my main contribution to this analysis and chapter 6 is dedicated to it.
- **Non-resonant background (NRB)**: containing processes with two leptons but not coming from a Z boson decay, such as $t\bar{t}$, tW and W^+W^- , these processes have a genuine source of missing transverse energy and can therefore contribute in the signal region in the rare case where their two leptons are compatible with leptons from a Z boson decay. The contribution coming from processes with top quarks are further suppressed since those quarks decay into b quarks that can be used to veto this kind of events. In the end, the fraction of NRB events passing the analysis selection is less than 0.1% of the initial background sample. However, the total cross section of NRB processes is large and the NRB contribution in the signal region is significant, especially at low $|\vec{p}_T^{\text{miss}}|$ value. Simulations for this background are available. However due to the low

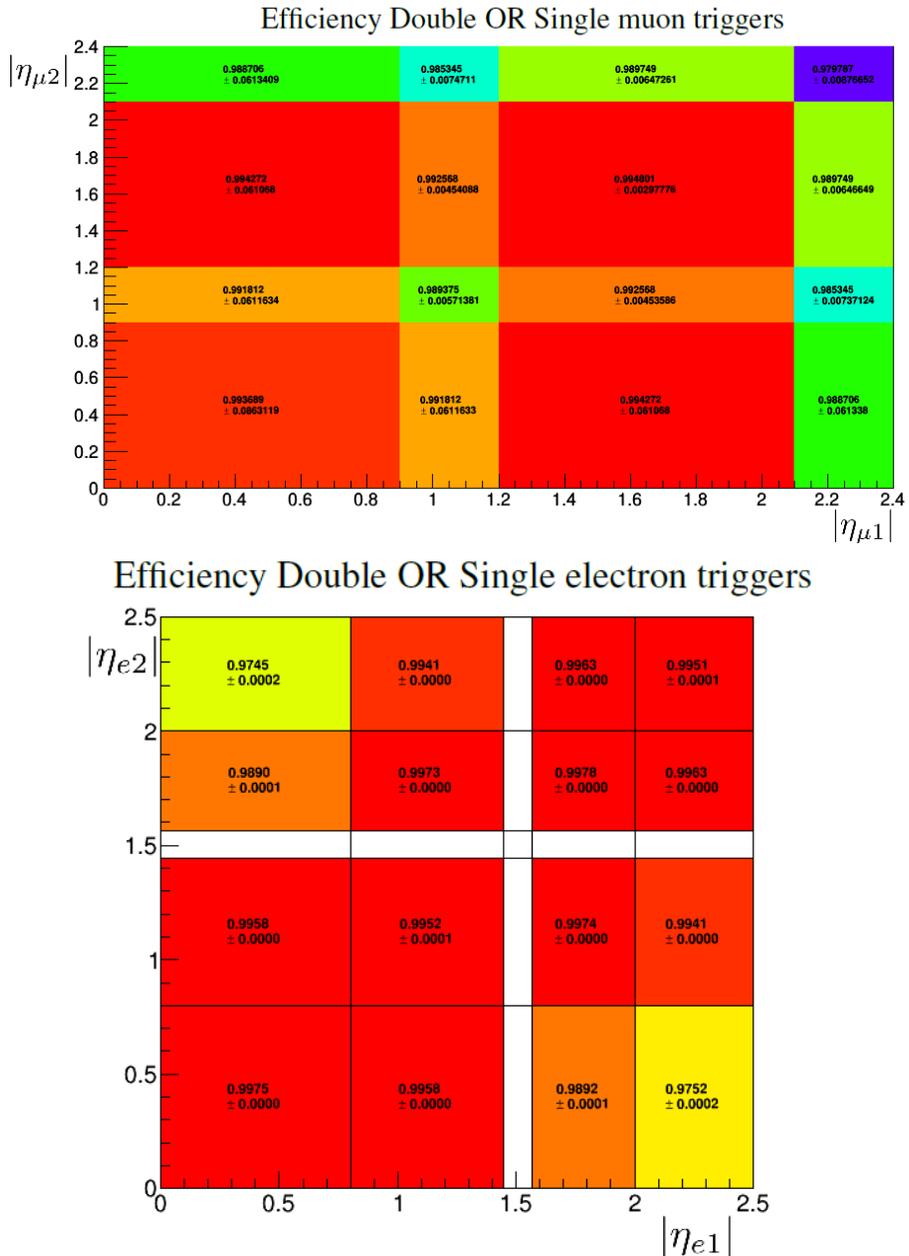


Figure 5.5.: Efficiency of the combination of the double and single muon triggers (top) and of the double and single electron triggers (bottom) estimated by the Tag&Probe method using the complete 2016 data as a function of the pseudorapidity of the leading and subleading leptons.

fraction of events passing the analysis cuts, it's preferable to use a data-driven approach to estimate this background in a more accurate way.

The simulated background samples are listed in Table 5.2 along with their generator and cross section. The order to which the cross section value is computed is given in the last column of the table. No simulated sample is used for the $gg \rightarrow WW$ background in this analysis since it does not contribute significantly and is anyway estimated by a data-driven method for NRB processes. The $gg(\rightarrow H) \rightarrow ZZ$ background represents the combination of the $gg \rightarrow ZZ$ process, the $gg \rightarrow H \rightarrow ZZ$ process and their interaction between themselves and the signal. It has been estimated by MELA as described in Sec. 5.2.1. Eventually, one should note that electroweak $ZZ + 2\text{jets}$ simulated samples are not included in this analysis. However, $ZZ + 2\text{jets}$ electroweak processes are non-negligible in the VBF category and represent about 10% of the background yields at $M_{ZZ} \gtrsim 400$ GeV in this category [124].

5.5. Corrections to simulated samples

5.5.1. Data-to-simulation scale factors

In order to be comparable with data, simulations need to be corrected for two types of effect. First, simulations have been generated with a generic pile-up distribution and need to be reweighted to the expected pile-up distribution knowing the LHC settings. Then, selection efficiencies differ between data and simulations for the trigger selection, the lepton identification, the lepton isolation and the b -veto requirement. Other corrections are also applied on the reconstructed objects of simulated samples to match — for example — the resolution of those objects in data but these have already been discussed in details in Sec 4.4.

Trigger, lepton reconstruction, identification and isolation scale factors

Computing efficiencies of those elements by the Tag&Probe method on both data and simulations, data-to-simulation scale factors are computed by taking the ratio between the efficiency values in $|\eta|$ and p_T bins between data and simulations. Then the total lepton selection scale factor SF_{tot} is computed as:

$$SF_{\text{tot}} = SF_{\text{reco}} \times SF_{\text{id}} \times SF_{\text{iso}} \times SF_{\text{trig}}. \quad (5.1)$$

The total scale factors for muons are of the order of 0.97, whereas total scale factors for electrons are of the order of 0.95. They are driven by the scale factors of the tight identification and isolation of the lepton.

Process	Generator	$\sigma \times BR$ (pb)	QCD precision
$W(\rightarrow \ell\nu) + \text{jets}$	MadGraph5_aMC@NLO	61 526.7	NNLO
$Z(\rightarrow \ell\ell) + \text{jets}$			
$10 \text{ GeV} < M_{\ell\ell} < 50 \text{ GeV}$	MadGraph5_aMC@NLO	18 610	NLO
$M_{\ell\ell} > 50 \text{ GeV}$	MadGraph5_aMC@NLO	5765.4	NNLO
Top			
$t\bar{t} + X$			
$t\bar{t} \rightarrow 2\ell 2\nu + \text{jets}$	Powheg	87.31	NNLO
$t\bar{t}W(\rightarrow \ell\nu)$	MadGraph5_aMC@NLO	0.2043	NLO
$t\bar{t}Z(\rightarrow \ell\bar{\ell})$	MadGraph5_aMC@NLO	0.2529	NLO
Single top			
$tW - \text{channel}(t)$	Powheg	35.85	NNLO
$tW - \text{channel}(\bar{t})$	Powheg	35.85	NNLO
$t - \text{channel}(t)$	Powheg	136.02	NLO
$t - \text{channel}(\bar{t})$	Powheg	80.95	NLO
$s - \text{channel}(\text{leptonic W-decay})$	MadGraph5_aMC@NLO	3.36	NLO
$qq \rightarrow WW$			
$qq \rightarrow WW \rightarrow \ell\nu q\bar{q}$	Powheg	49.997	NNLO
$qq \rightarrow WW \rightarrow 2\ell 2\nu$	Powheg	12.178	NNLO
WZ			
$WZ \rightarrow 2\ell 2q$	MadGraph5_aMC@NLO	5.595	NNLO
$WZ \rightarrow 3\ell\nu$	Powheg	4.4297	NLO
$qq \rightarrow ZZ$			
$qq \rightarrow ZZ \rightarrow 2\ell 2q$	MadGraph5_aMC@NLO	3.22	NLO
$qq \rightarrow ZZ \rightarrow 2\ell 2\nu$	Powheg	0.564	NLO
ZVV			
ZWW	MadGraph5_aMC@NLO	0.165 10	NLO
ZZW	MadGraph5_aMC@NLO	0.055 65	NLO
ZZZ	MadGraph5_aMC@NLO	0.013 98	NLO

Table 5.2.: List of the simulated background samples contributing in the di-lepton region along with their generator and cross section. The last column represents the QCD precision on the cross section.

b -veto scale factors

Prescriptions of the dedicated group of CMS on b -tagging are followed [125]. Discrepancies between data and simulations are corrected by randomly retagging a fraction of the jets that were not tagged as b -jets. The fraction of jets to be retagged is

given by:

$$\text{fraction to retag} = \frac{1 - SF}{1 - \frac{1}{\epsilon}} \quad (5.2)$$

where SF corresponds to the data-to-simulation scale factor produced by the CMS collaboration and ϵ to the efficiency computed on simulations of the loose working point used here (about 83%). The fraction of jets to be retagged is about 1% at low p_T and increases up to 6% at high p_T (above 200 GeV).

5.5.2. Higher-order corrections to di-boson samples

The ZZ and WZ backgrounds have been simulated at the NLO in QCD and correction factors are further applied to take into account missing QCD diagrams at the NNLO for ZZ , and missing electroweak diagrams at the NLO for ZZ and WZ .

NNLO QCD corrections on the ZZ samples have been provided by the $H \rightarrow ZZ \rightarrow 4\ell$ group for the $2e2\mu$ finale state, based on [126]. Figure 5.6 shows the M_{ZZ} distribution for the generated ZZ sample at NLO and its NNLO corrected distribution after applying the differential scale factor computed for the $2e2\mu$ final state. The bottom of this figure shows the scale factors applied. Above 500 GeV, the scale factor from the last bin is used.

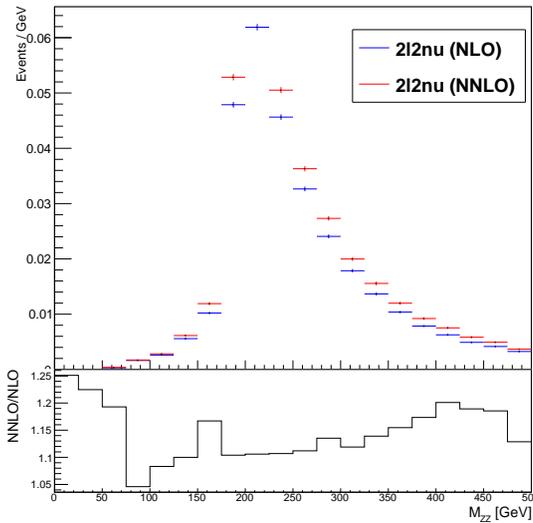


Figure 5.6.: M_{ZZ} distribution for the NLO generated ZZ sample and its corrected distribution for NNLO missing QCD diagrams.

NLO electroweak corrections are applied on the ZZ (except for the $gg \rightarrow ZZ$) and WZ production processes. These corrections come in three categories:

- emission of a real $\gamma/Z/W$ boson
- photon-induced ($\gamma q \rightarrow ZZ/WZ$) processes
- virtual corrections with a Z, γ or W in the loop.

References [127, 128] have shown that the emission of a real boson can be neglected for both ZZ and WZ cases. They have also shown that the photon-induced contribution is only non-negligible for WZ processes.

The virtual corrections are implemented on the basis of a table given by the authors of [127, 128]. This table gives correction scale factors as a function of the flavour of the quarks initiating the hard process and the Mandelstam variables \hat{s} and \hat{t} . These variables describe completely the kinematics of the process under the assumption that the p_T of complementary jets is small compared to the one of the bosons.

The correction for photon-induced contributions to the WZ background was computed by Nicolas Postiau using MadGraph and LUXqed PDFs which include precise PDF for the photon density in the proton [129].

Figure 5.7 shows the global NLO electroweak correction factors as a function of M_{ZZ} (top) and M_{WZ} (bottom). Virtual corrections have been applied in the case where the two bosons are on-shell and have been set to one for $M_{V_1 V_2} < M_{V_1} M_{V_2}$. Above this threshold, NLO electroweak corrections on the ZZ process are negative and of about 4% around 200 GeV and increase to reach 15% at 1 TeV. In the WZ case, virtual corrections are smaller and range between 1% at the on-shell threshold and increase up to 4% at high mass. In addition, photon-induced corrections are included and are positive, raising the total corrections between 1% at low-mass and 2% at high-mass for WZ .

5.6. Analysis strategy

5.6.1. Event selection

After the trigger requirements, events undergo the following selection:

- two well-identified electrons or muons coming from a Z boson decay:
 - $p_T > 25$ GeV for each lepton
 - $|\eta| < 2.5$ or 2.4 for each electron or muon respectively; in addition electrons in the barrel-endcap transition ($1.44 < |\eta| < 1.56$) are vetoed
 - tight identification and isolation criteria (see Chap. 4)

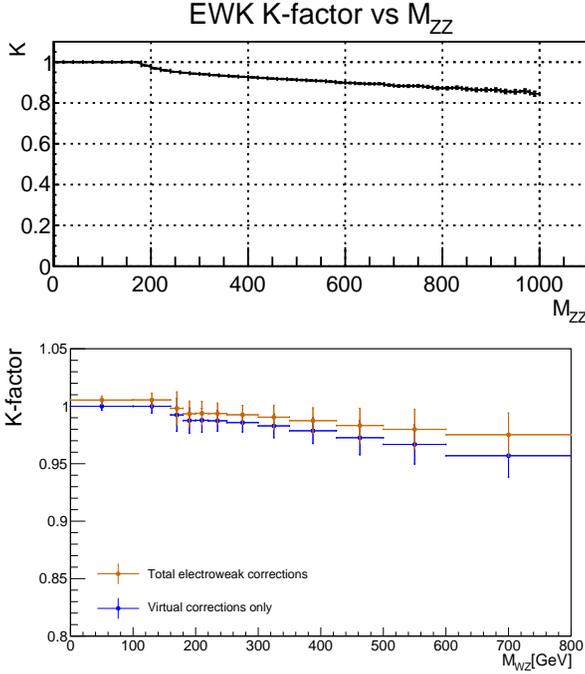


Figure 5.7.: NLO electroweak correction factors for the ZZ (top) and WZ (bottom) processes without gluon-induced interactions. Global uncertainties on these corrections are shown for WZ while only statistical uncertainties are shown for ZZ .

$$- |M_{\ell\ell} - 91.19 \text{ GeV}| < 15 \text{ GeV}, 91.19 \text{ GeV} \text{ being the mass of a } Z \text{ boson}$$

- $p_{T,Z} > 55 \text{ GeV}$ where Z represents the di-lepton system
- a veto on events with a third isolated lepton, mainly to reduce WZ background contamination
- a veto on events with at least one b -tagged jet in order to suppress backgrounds with top quarks
- $\Delta\phi(\text{jet}, \vec{p}_T^{\text{miss}}) > 0.5$ for each tight jet above 30 GeV in order to reduce events with \vec{p}_T^{miss} coming from mismeasurement of the jets
- $\Delta\phi(Z, \vec{p}_T^{\text{miss}}) > 0.5$ to reject a region badly modeled by the data-driven estimation of the $Z + \text{jet}$ background (see chapter 6)
- $|\vec{p}_T^{\text{miss}}| > 125 \text{ GeV}$ to suppress most of the $Z + \text{jet}$ background.

5.6.2. Event categorization

This event sample is split in lepton and jet categories. There are two lepton categories: $2e2\nu$ and $2\mu2\nu$ that are labelled ee and $\mu\mu$. Events are then split in three exclusive jet categories, defined on jets passing tight identification criteria and with $p_{T,\text{jet}} > 30$ GeV:

- = **0 jet**: no jet above 30 GeV in p_T
- \geq **1 jet**: at least one jet, except for events going into the VBF-tagged category
- **VBF-tagged**: at least two jets with the following conditions on the two jets j_1 and j_2 with the highest p_T :
 - $|\eta_{j_1} - \eta_{j_2}| > 4$
 - $M_{j_1 j_2} > 500$ GeV
 - no additional jets of $p_{T,\text{jet}} > 30$ GeV between j_1 and j_2 in η
 - the reconstructed Z boson has to be between j_1 and j_2 in η

This categorization is motivated by the fact that each category is affected differently by backgrounds and has different sensibility to signal. In particular, the VBF-tagged category is more sensible to VBF signal.

5.6.3. Choice of observable: the transverse mass

The search is performed by fitting the transverse mass spectrum with signal+background models where the signal strength is treated as a free parameter.

A common choice of observable when seeking for new particles is the reconstructed invariant mass of the final state since it allows to observe directly potential new resonances. However, the $2\ell2\nu$ final state contains neutrinos and the invariant mass cannot be computed. Instead, the transverse mass M_T is used:

$$M_T^2 = \left(\sqrt{p_{T,\ell\bar{\ell}}^2 + M_{\ell\bar{\ell}}^2} + \sqrt{|\vec{p}_T^{\text{miss}}|^2 + (91.19 \text{ GeV})^2} \right)^2 - (\vec{p}_{T,\ell\bar{\ell}} + \vec{p}_T^{\text{miss}})^2. \quad (5.3)$$

In the case of the signal, the transverse mass distribution has its most probable value at M_H but with a poor resolution.

5.6.4. Statistical analysis

The inputs to this analysis are six (for lepton and jet category) distributions of M_T both for background and signal models and data. The statistical analysis of those distributions can take two directions:

- **discovery:** there is an excess observed in data compared to background estimation. In this case, the significance of the excess is computed as the probability that background fluctuations can cause such a deviation.
- **exclusion limits:** upper limits on the cross section of signal processes are set. In particular this is done using the CL_s method which uses likelihood ratios¹. Since this is what is done in this analysis, the CL_s method is detailed below [130–132].

To compare the compatibility of the data events \vec{n} with the background-only and signal-plus-background hypotheses, where the signal \vec{s} is allowed to be scaled by the signal strength μ , the CL_s method is using a test statistic q_μ based on a profile likelihood ratio:

$$q_\mu = -2 \ln \frac{\mathcal{L}(\vec{n} | \mu \vec{s} + \vec{b}, \hat{\vec{\theta}}_\mu)}{\mathcal{L}(\vec{n} | \hat{\mu} \vec{s} + \vec{b}, \hat{\vec{\theta}})}, \quad \text{with } 0 \leq \hat{\mu} \leq \mu, \quad (5.4)$$

where \mathcal{L} is the likelihood of data \vec{n} for a certain signal distribution \vec{s} of strength μ , a background distribution \vec{b} and nuisance parameters $\vec{\theta}$. The vector symbol above \vec{n} , \vec{s} and \vec{b} denotes the number of events per bin in the M_T distribution. The vector symbol above $\vec{\theta}$ denotes the number of nuisance parameters. Floating parameters are noted with a hat and are fitted by maximizing the corresponding likelihood. In the numerator, only the nuisance parameters $\vec{\theta}_\mu$ are floating to maximize the likelihood for a given value of μ . In the denominator, both the signal strength $\hat{\mu}$ and the nuisance parameters are floating. The variation of $\hat{\mu}$ is bounded by zero since the signal rate searched for is positive. The upper bound $\hat{\mu} \leq \mu$ guarantees a one-side confidence interval. In other words, this means that upward fluctuations of the data such as $\hat{\mu} > \mu$ are not considered as evidence against a signal hypothesis of strength μ .

The observed value of this test statistic can be obtained as a function of $\mu > 0$ for the signal-plus-background hypothesis q_μ^{obs} and for $\mu = 0$ for the background-only hypothesis q_0^{obs} along with the nuisance parameters $\hat{\vec{\theta}}_\mu^{obs}$ and $\hat{\vec{\theta}}_0^{obs}$ maximizing the likelihood.

With those inputs, the probability density function of the test statistics for the signal-plus-background $f(q_\mu | \mu \vec{s} + \vec{b}, \hat{\vec{\theta}}_\mu^{obs})$ and background-only $f(q_0 | \vec{b}, \hat{\vec{\theta}}_0^{obs})$ can be constructed via simulated pseudo-data.

For the CL_s method, two probability values (p-values) are then be computed. P-values represent the probability for a given model to obtain the same or a more extreme value. The needed p-values are the one for the signal-plus-background hypothesis $p_{\mu \vec{s} + \vec{b}}$ and the one for the rejection of the background-only hypothesis $1 - p_{\vec{b}}$. They

¹Notion of statistics and in particular on the likelihood are given in appendix A.

are defined as follow:

$$p_{\mu\vec{s}+\vec{b}} = P(q_\mu \geq q_\mu^{obs} | signal - plus - background) = \int_{q_\mu^{obs}}^{\infty} f(q_\mu | \mu\vec{s} + \vec{b}, \hat{\theta}_\mu^{obs}) dq_\mu \quad (5.5)$$

and

$$1 - p_{\vec{b}} = P(q_\mu \geq q_\mu^{obs} | background - only) = \int_{q_\mu^{obs}}^{\infty} f(q_0 | \vec{b}, \hat{\theta}_0^{obs}) dq_\mu. \quad (5.6)$$

The CL_s value for a given signal strength is then given by the ratio of those p-values:

$$CL_s(\mu) = \frac{p_{\mu\vec{s}+\vec{b}}}{1 - p_{\vec{b}}} \quad (5.7)$$

One says that the signal strength $\mu = X$ is excluded with $(1-\alpha)$ CL_s confidence level (CL) if $CL_s(\mu = X) \leq \alpha$. In this analysis, results are quoted at 95% CL, i.e. μ is scanned until reaching a CL_s value of 0.05. The CL_s definition is known to produce rather conservative limits (see appendix A of [130] for more details). The use of a ratio of p-values makes it possible to treat cases where the signal is so small that both hypotheses are compatible with observation. In those cases, there is no sensitivity to the signal and the CL_s method avoids rejecting it.

Median expected limit as well as the one and two standard deviation (σ) bands are also computed and shown in this analysis. One method to produce them is to generate a large number of pseudo-experiments according to the background-only hypothesis. Then, their signal strength limit at 95% CL are computed as if they were real data.

The drawback of this method is that it is very CPU intensive. Therefore, when the number of expected events is large enough, asymptotic limits are used instead. In this method, the set of simulated pseudo-data can be replaced by a single representative dataset where all observed quantities are set equal to their expected (or observed for data) values. This dataset is called an Asimov dataset and more details can be found in [133].

5.7. Data-to-simulation comparisons

After applying all the corrections discussed in the previous sections to simulations, data-to-simulation comparisons can be done. Figure 5.8 summarizes the analysis flow, showing the number of events passing each step of the analysis selection both in the $\mu\mu$ and in the ee channels, including all jet categories. One can see that there are approximately twice less events in the ee category than in the $\mu\mu$. This is

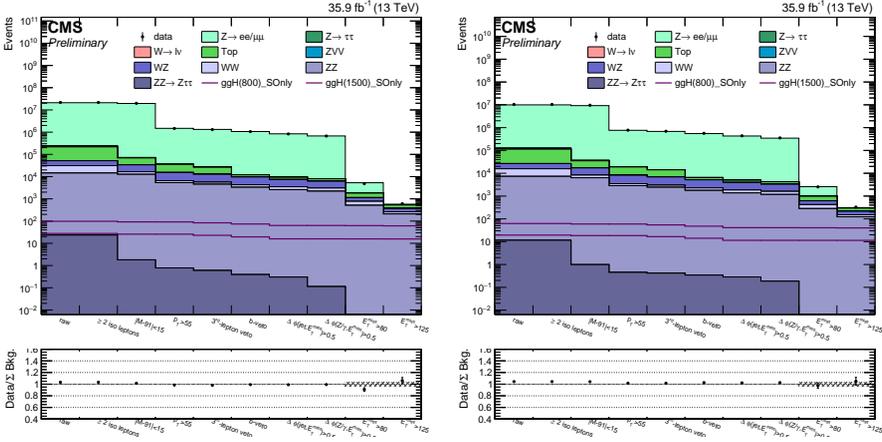


Figure 5.8.: Number of events passing each pre-selection cuts in the $\mu\mu$ (left) or ee (right) channels. Signal distributions for $m_H = 800$ and 1500 GeV produced by ggF are superimposed to all distributions. The first bin (raw) should be ignored.

mainly due to lower identification and isolation efficiencies for electrons than muons (see chapter 4).

Figures 5.9 and 5.10 — respectively for the $\mu\mu$ and in the ee channels — shows the main variables used in the analysis selection at the selection level at which they are cut on.

One can see that until the very last cut on the \vec{p}_T^{miss} , the analysis is dominated by the $Z + \text{jets}$ background. A fair agreement between data and simulations is observed for both lepton flavors. No corrections to cure the residual small discrepancies in the shape and normalization of the Z mass peak are applied since, as shown in the \vec{p}_T^{miss} distribution, the $Z + \text{jet}$ simulations do not model well this background. Therefore a data-driven approach is used.

5.8. Data-driven estimations

As mentioned in Sec. 5.4, the $Z + \text{jet}$ and the non-resonant backgrounds are estimated with data-driven methods.

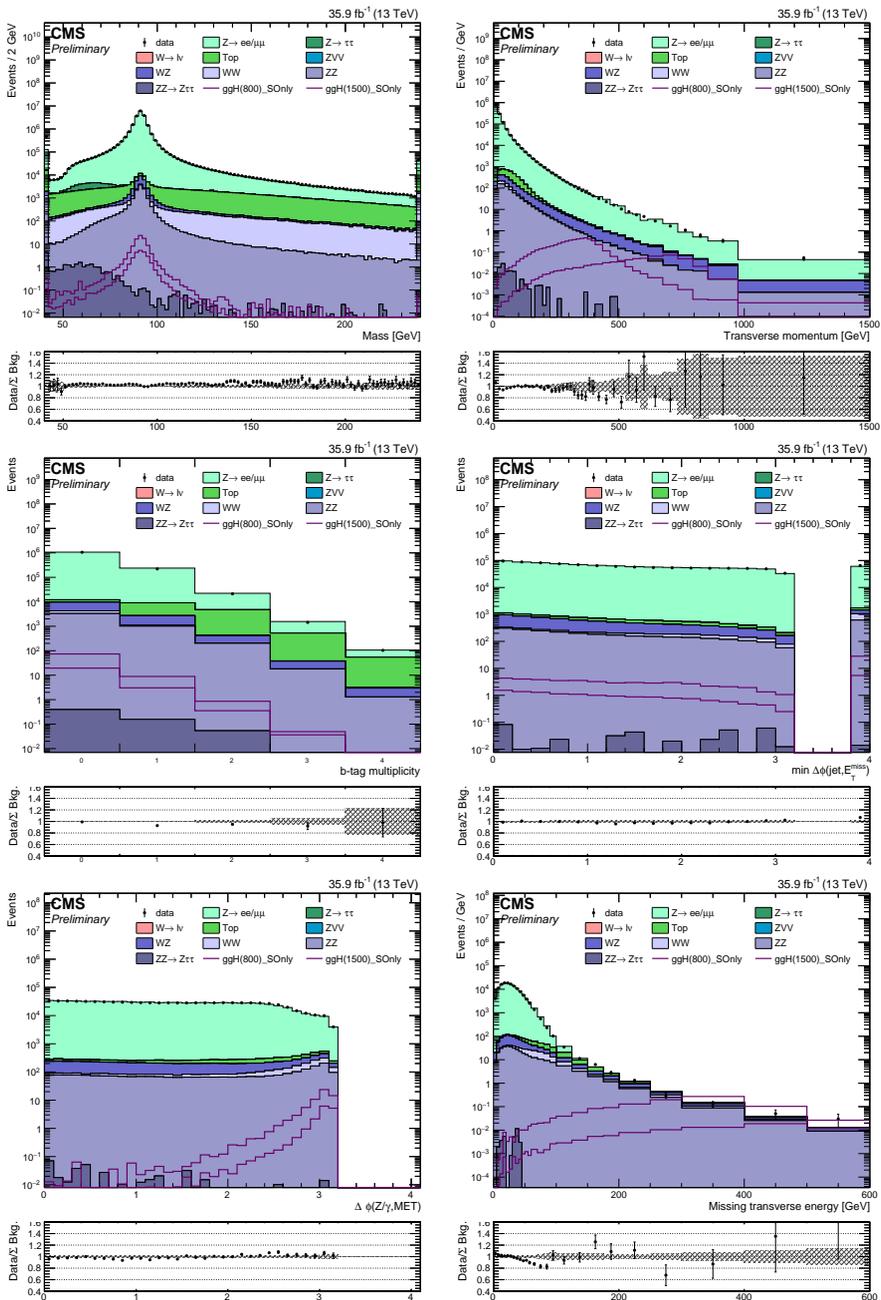


Figure 5.9.: Main variables used in the analysis selection presented at the selection level at which they are cut on in the $\mu\mu$ channel. **Top left:** di-lepton invariant mass. **Top right:** reconstructed Z boson p_T . **Center left:** number of jets identified as b jets. **Center right:** $\Delta\phi$ angle between \vec{p}_T^{miss} and the nearest jet with p_T above 30 GeV. **Bottom left:** $\Delta\phi$ angle between the reconstructed Z boson and \vec{p}_T^{miss} . **Bottom right:** $|\vec{p}_T^{\text{miss}}|$ distribution.

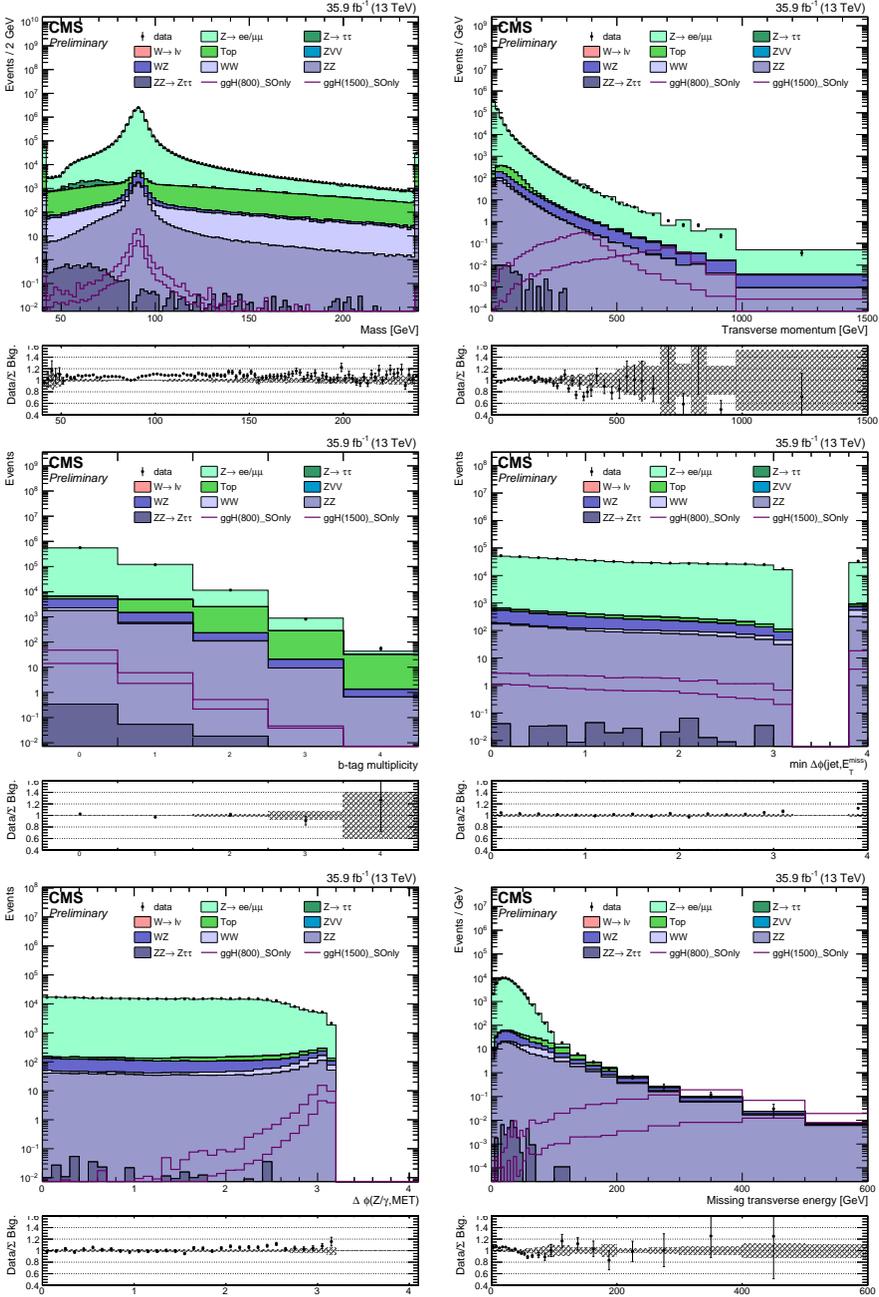


Figure 5.10.: Main variables used in the analysis selection presented at the selection level at which they are cut in the ee channel. **Top left:** di-lepton invariant mass. **Top right:** reconstructed Z boson p_T . **Center left:** number of jets identified as b jets. **Center right:** $\Delta\phi$ angle between \vec{p}_T^{miss} and the nearest jet with p_T^{jet} above 30 GeV. **Bottom left:** $\Delta\phi$ angle between the reconstructed Z boson and \vec{p}_T^{miss} . **Bottom right:** $|\vec{p}_T^{\text{miss}}|$ distribution.

5.8.1. Z +jet background estimation

Being a core subject of this thesis, Chapter 6 is entirely dedicated to the data-driven estimation of this background and only the results are presented here. Figures 5.11 and 5.12 show the comparison between the Z + jets background estimated from simulations (left) and from data-driven estimations (right) in the $\mu\mu$ and ee channels respectively. A good improvement is seen in the \vec{p}_T^{miss} distribution as well as in the M_T distributions. In particular, more Z + jets events are predicted at $|\vec{p}_T^{\text{miss}}| > 125$ GeV with the data-driven approach which populate the final M_T distribution.

The present thesis is the first time that the uncertainties on the data-driven estimation of the Z +jet background are estimated in detail. Uncertainties are divided into statistical uncertainties, systematic uncertainties and uncertainties inherent to the method, the latter being computed from a closure test of the method and evaluated to 10%. The relative statistical and systematic uncertainties are both of the order of 50%.

Using the final M_T distributions, exclusion limits are set. While this will be discussed later in this chapter, Fig. 5.13 shows the relative impact on these limits of the following scenarii for a heavy scalar of width 100 GeV and produced through ggF:

- **reference:** reference value, equal to one by definition;
- **noInstrMET:** limits computed when removing completely the Z +jet background;
- **noInstrMET_stat:** limits computed including the Z +jet background but without its statistical uncertainties;
- **noInstrMET_syst:** limits computed including the Z +jet background but without its systematic uncertainties;
- **noInstrMET_systAndStat:** limits computed including the Z +jet background but without any uncertainties on it;
- **noUncertaintyAtAll:** limits computed by removing all uncertainties on all processes.

One should note that the horizontal shift on this figure is artificial and has been set to improve readability.

It is expected that the relative impacts on the limit in these scenarii decrease with the mass since, at high mass the limits are lead by the number of observed data events and backgrounds play close to no role.

On this figure one can see that the relative impact on the exclusion limits of the statistical and systematic uncertainties on the data-driven estimation of the Z +jet background is between 10 and 15%, which make them non-negligible and support the importance of studying them into details, as shown in the next chapter.

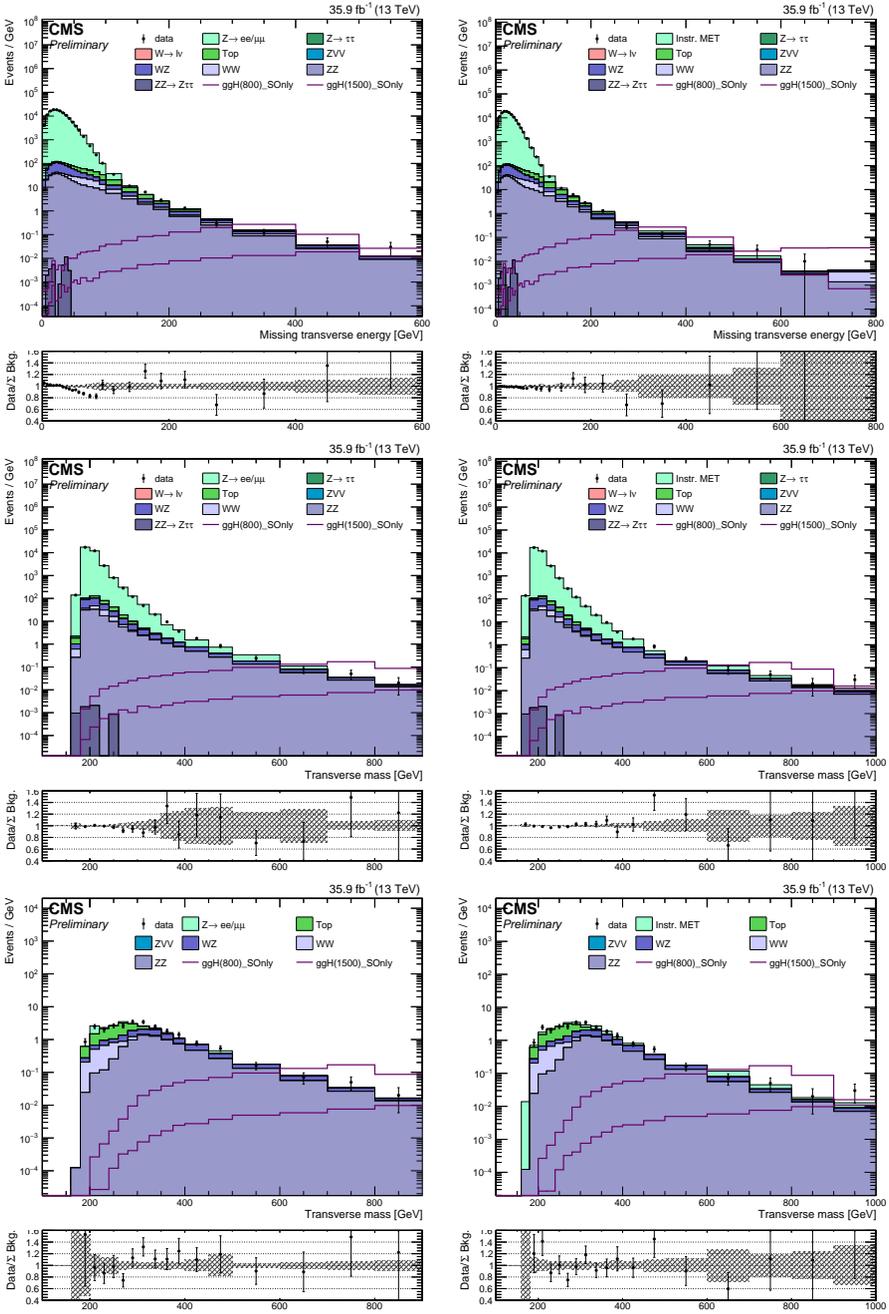


Figure 5.11.: Comparison in the $\mu\mu$ channel between the $Z + \text{jets}$ background estimated from simulations (left) and from data-driven estimation (right). Top and center histograms have all selection cuts applied except for the \vec{p}_T^{miss} cut. Bottom histograms include the $|\vec{p}_T^{\text{miss}}| > 125 \text{ GeV}$ cut. **Top:** $|\vec{p}_T^{\text{miss}}|$ distribution. **Center:** M_T distribution without any \vec{p}_T^{miss} cut. **Bottom:** M_T distribution after all selections applied.

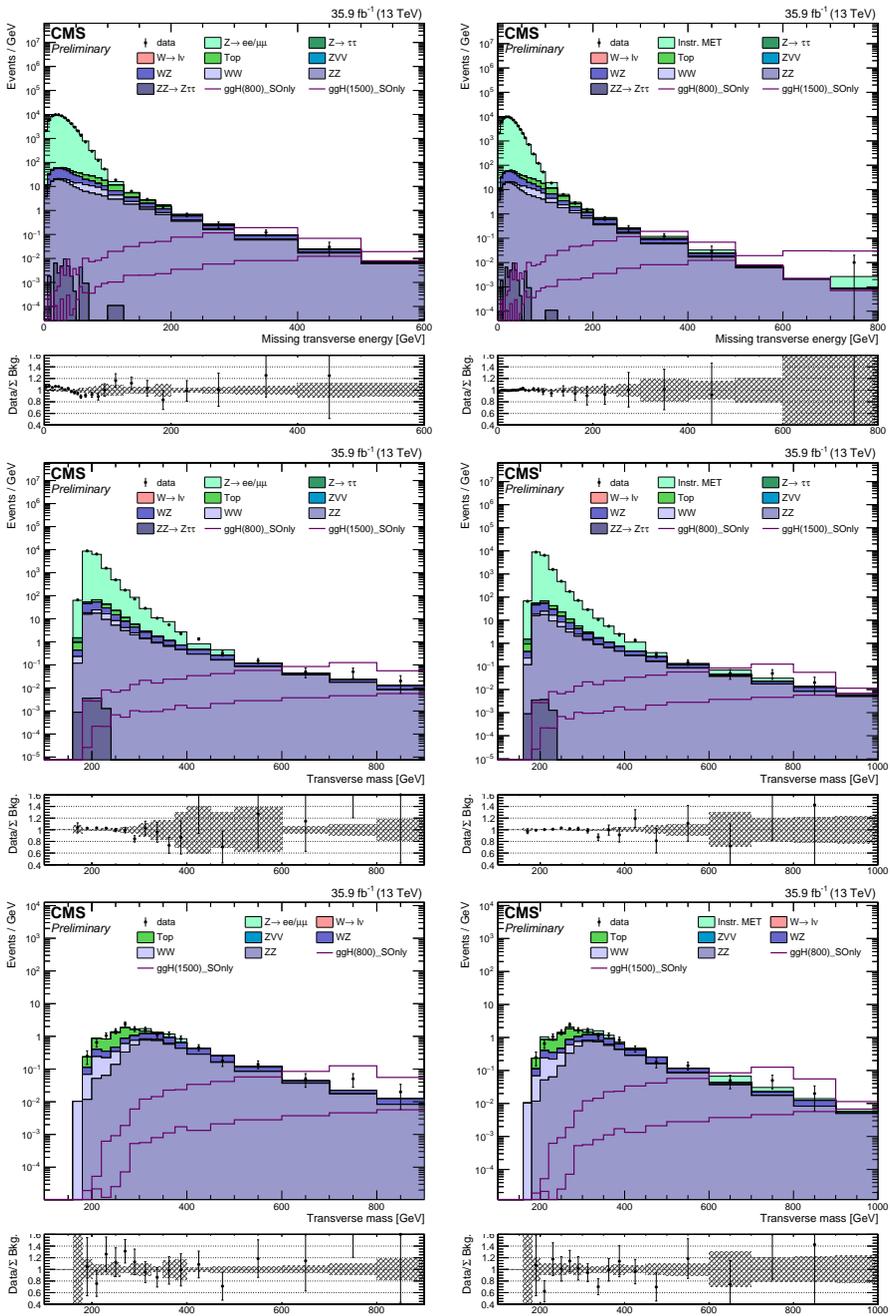


Figure 5.12.: Comparison in the ee channel between the Z +jets background estimated from simulations (left) and from data-driven estimation (right). Top and center histograms have all selection cuts applied except for the \bar{p}_T^{miss} cut. Bottom histograms include the $|\bar{p}_T^{\text{miss}}| > 125$ GeV cut. **Top:** $|\bar{p}_T^{\text{miss}}|$ distribution. **Center:** M_T distribution without any \bar{p}_T^{miss} cut. **Bottom:** M_T distribution after all selections applied.

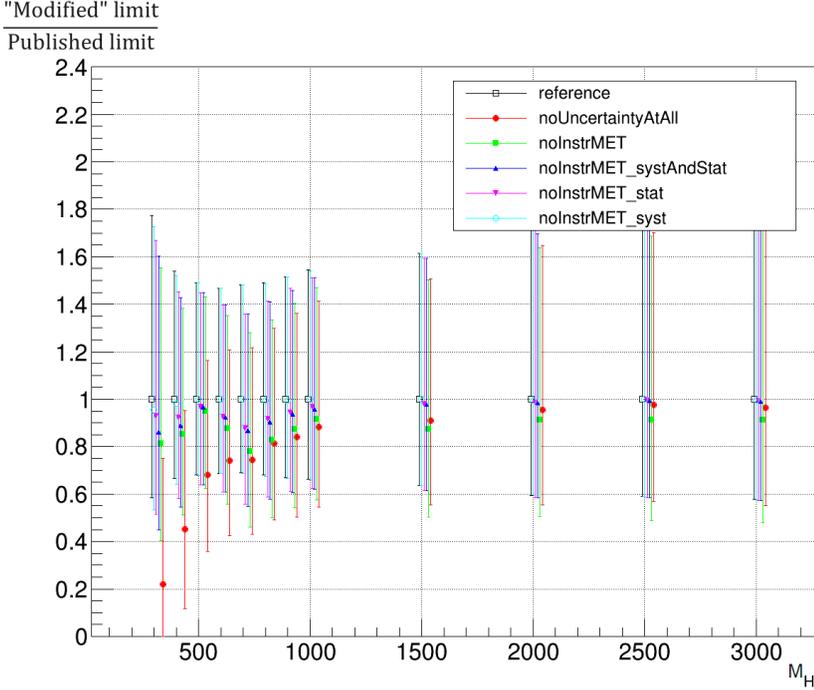


Figure 5.13.: Relative impact on the exclusion limits for a heavy scalar of width 100 GeV and produced through ggF in various scenarii as a function of M_H . The horizontal shift is artificial in order to improve readability. More details can be found in the text.

5.8.2. Non-resonant background estimation

As shown on figures 5.11 and 5.12, the top background is the dominant one at low M_T at the final selection level.

The NRB data-driven method estimates the contamination of this background, as well as the contamination of the WW , WWW and $W + \text{jet}$ backgrounds. Those backgrounds have a continuous spectrum of di-lepton invariant mass and the probability that they generate opposite flavor leptons (i.e. $e\mu$ final state) is exactly twice as large as the probability for same flavor leptons (ee or $\mu\mu$). Therefore, the data-driven method consists in taking data events that fired $e\mu$ triggers (see Table 5.3) and use them to predict the number of non-resonant $ee/\mu\mu$ events in the signal region.

However, the number of $ee/\mu\mu$ NRB events in the signal region is not exactly half the number of $e\mu$ NRB events in this region because the selection efficiencies are different for ee , $\mu\mu$ and $e\mu$ events. Therefore, a control region to compute ratios of

Trigger category	Trigger path
MuEG	Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
	Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL
	Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
	Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL
	Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
	Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL

Table 5.3.: Muon-electron trigger paths used by the NRB data-driven estimation.

selection efficiencies for the NRB background is used. This control region is defined where the contribution of the NRB is dominant and where the selection is as close as possible to the one in the signal region. The differences compared to the selection applied to the signal region are: the requirement of at least one b -tagged jets in order to suppress the contamination of $Z + \text{jets}$ events, a cut of $|\vec{p}_T^{\text{miss}}| > 50 \text{ GeV}$ instead of 125 GeV also to suppress $Z + \text{jets}$ contamination while keeping more NRB events (this cut choice is motivated in the discussion about uncertainties of this method) and the requirement to have a same-flavour lepton pair is removed. Figure 5.14 shows this control region in the $\mu\mu$, ee and $e\mu$ channels.

One can see that, outside the Z boson mass peak, the top background is dominant and contribution from ZZ and WZ backgrounds can be neglected in all three channels. Inside and outside regions around the Z peak are defined by $76 \text{ GeV} < M^{\text{in}} < 106 \text{ GeV}$ and $40 \text{ GeV} < M^{\text{out}} < 70 \text{ GeV} \cup 110 \text{ GeV} < M^{\text{out}} < 200 \text{ GeV}$. The number of $e\mu$ events in the signal region — and therefore inside the Z peak — $N_{e\mu}^{\text{SR}}$ and the number of $e\mu$ events outside the Z peak in the control region $N_{e\mu}^{\text{out, CR}}$ are counted. Similarly, the number of events in the ee and $\mu\mu$ channels outside the Z peak in the control region $N_{ee/\mu\mu}^{\text{out, CR}}$ are counted. Eventually, the number of $ee/\mu\mu$ events predicted inside the Z peak is given by:

$$N_{ee/\mu\mu}^{\text{SR, predicted}} = \alpha_{ee/\mu\mu} N_{e\mu}^{\text{SR}}, \quad (5.8)$$

where:

$$\alpha_{ee/\mu\mu} = \frac{N_{ee/\mu\mu}^{\text{out, CR}}}{N_{e\mu}^{\text{out, CR}}}. \quad (5.9)$$

This procedure is repeated for different \vec{p}_T^{miss} cuts: higher cuts meaning less contamination from $Z + \text{jets}$ but also more statistical uncertainties on the method. Figure 5.15 shows the value of α for different \vec{p}_T^{miss} cuts for both data and simulations. The agreement between data and simulations is good. α tends to be larger at low \vec{p}_T^{miss} cut since the contribution of the $Z + \text{jets}$ is less suppressed. Overall, the dependence

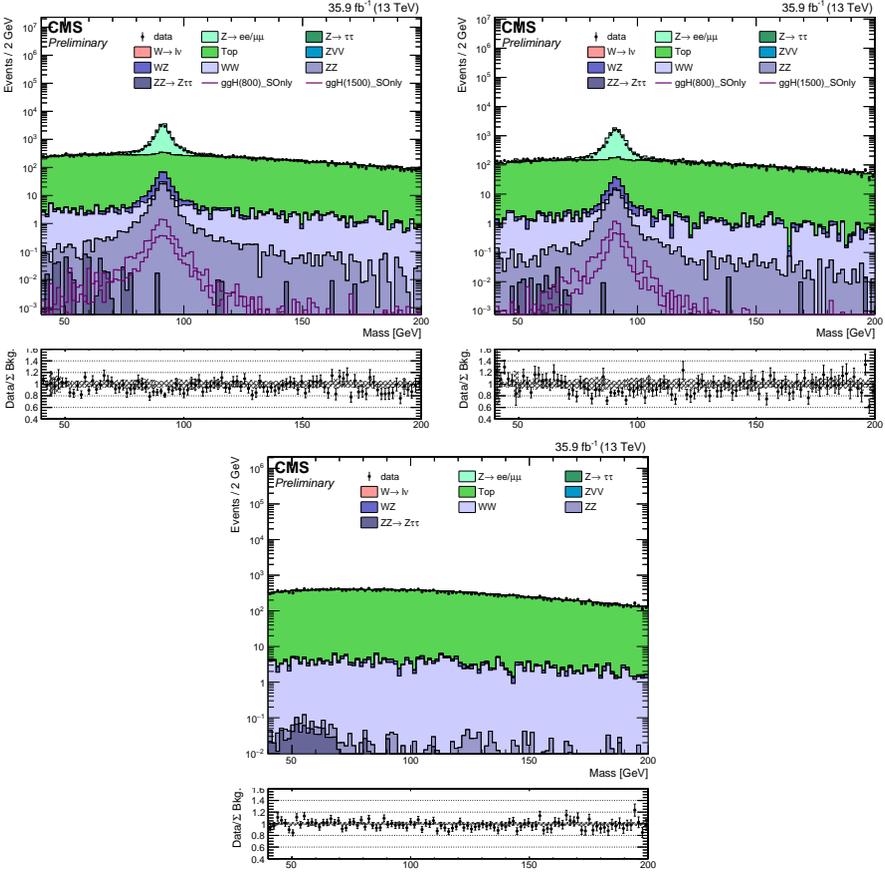


Figure 5.14.: Di-lepton invariant mass in the NRB control region (see text) for the $\mu\mu$ (top left), ee (top right) and $e\mu$ (bottom) channels.

of α with \vec{p}_T^{miss} is fairly flat and the final cut chosen is 70 GeV. The values of α at this cut are given in table 5.4.

Channel		α
ee	Data	0.369 ± 0.006
	MC	0.367 ± 0.002
$\mu\mu$	Data	0.683 ± 0.009
	MC	0.685 ± 0.003

Table 5.4.: Values of α for a \vec{p}_T^{miss} cut of 70 GeV.

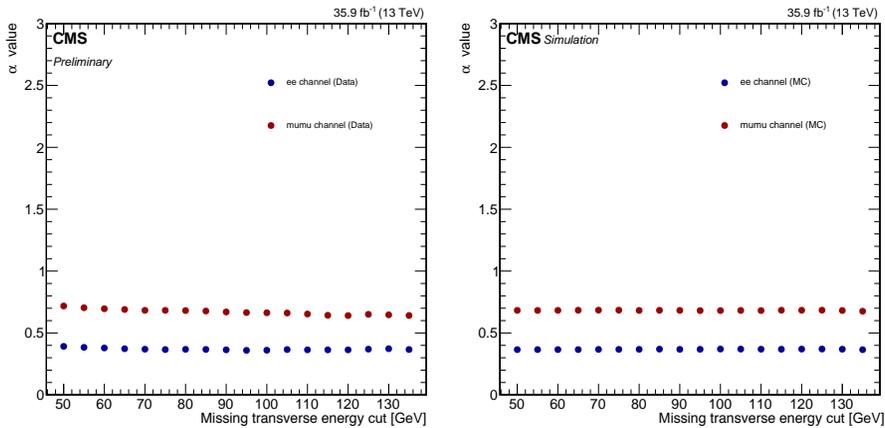


Figure 5.15.: Missing transverse energy dependence of α in data (left) and simulations (right). ee channel is represented in blue and $\mu\mu$ in red.

Tests have also been done to add back the veto on b -tagged jets or use only the upper outside region of the Z peak in the computation of α . No significant differences with the control region described above have been seen and since those tests decrease the number of events to be considered they are not chosen.

Comparing the number of events in the signal region estimated directly from simulations N_{expected} and the number of events predicted by this method applied on simulations $N_{\text{predicted}}$, a measure of the relative bias of the method has been done as a closure test:

$$\text{bias} = \frac{N_{\text{predicted}} - N_{\text{expected}}}{N_{\text{expected}}}. \quad (5.10)$$

The bias is less than 5% for both categories and is fairly flat with \bar{p}_T^{miss} . Figure 5.16 shows the total uncertainty on the NRB data-driven estimation. It combines the bias, statistical errors on α and the relative difference between α values at different \bar{p}_T^{miss} cut. Between 50 and ~ 70 GeV, uncertainties are bigger due to a greater contamination of $Z + \text{jets}$ events. At ~ 70 GeV, uncertainties are dominated by the bias and, at higher \bar{p}_T^{miss} , the statistical uncertainty increases.

5.9. Uncertainties

Now that a better description of the $Z + \text{jet}$ and non-resonant backgrounds have been discussed, the last point before going to results is the estimation of uncertainties. They come in three sources: statistical uncertainties, experimental uncertainties and theoretical uncertainties.

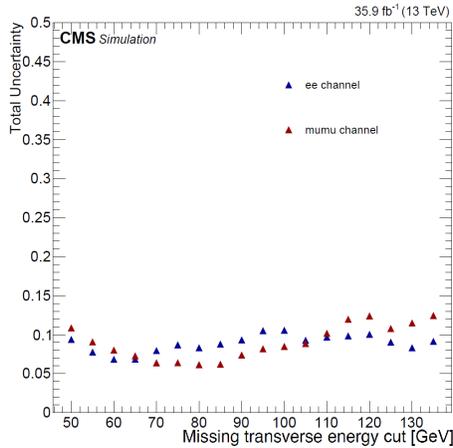


Figure 5.16.: Total uncertainties on the NRB data-driven estimation method for ee (blue) and $\mu\mu$ (red).

Uncertainties on the $Z + \text{jet}$ and non-resonant backgrounds are discussed in the dedicated sections of those data-driven estimations, namely Sec. 5.8.2 for the non-resonant background and Sec. 6.9 for the $Z + \text{jet}$ background.

5.9.1. Instrumental uncertainties

Instrumental uncertainties encompass all uncertainties linked to the experiment, such as the integrated luminosity, detector calibration, object reconstruction and selection.

Luminosity

Affecting the normalization of the simulated samples to data, the uncertainty on the luminosity recorded by CMS in 2016 is estimated to 2.5% [134].

Triggers, lepton identification and lepton isolation

Uncertainties on the Tag&Probe method are estimated by combining statistical uncertainties and relative differences when using different parameters (such as changing the tag definition, or the signal/background modeling). Finally, a total uncertainty on those scale factors is set: 6% for $\mu\mu$ events and 8% for ee .

Lepton momentum scale

The lepton momentum scale uncertainty is computed by varying the momentum of leptons by the uncertainties linked to their scale corrections and by checking the final M_T distribution. This can affect the normalization and the shape of distributions. However they have been found to be very small: 0.01% for muons and 0.3% for electrons.

Jet energy scale, resolution and \vec{p}_T^{miss}

As discussed in Sec. 4.4.4, jet energy and resolution need to be corrected as well as their impact on the \vec{p}_T^{miss} distribution. Their impact is evaluated by applying those corrections shifted by ± 1 standard deviation compared to their nominal value, correcting accordingly the \vec{p}_T^{miss} estimation and using the resulting M_T distributions as input to shape and normalization uncertainties. These corrections affect the jet p_T and therefore jet category migration occurs. Particles not clustered in jets but taken into account in the \vec{p}_T^{miss} undergo also an energy scale correction (unclustered \vec{p}_T^{miss}). Table 5.5 summarizes the systematic uncertainties on the ZZ background for its final M_T distributions.

Process	=0 jet			≥ 1 jet			VBF		
	JES	JER	$u\vec{p}_T^{\text{miss}}$	JES	JER	$u\vec{p}_T^{\text{miss}}$	JES	JER	$u\vec{p}_T^{\text{miss}}$
ZZ	2.728	0.182	1.902	4.886	0.164	0.744	34.949	2.581	4.497

Table 5.5.: Uncertainties (in %) from jet energy scale, jet energy resolution and unclustered \vec{p}_T^{miss} .

b -jet veto

The uncertainty on b -tagging scale factors is given by the CMS collaboration and the exact same method as described in Sec. 5.5.1 is applied with the upward and downward variation of those scale factors. The effect is evaluated on the final M_T distributions and is of the order of 3% for the normalization.

5.9.2. Theoretical uncertainties

Theoretical uncertainties are the ones related to signal and background modeling. It contains uncertainties on cross section computation and higher-order corrections.

PDF set choice uncertainty

PDF uncertainties arise from experimental uncertainties of the data used in their determination but also on theoretical uncertainties closely related to PDF, such as the strong coupling α_S . Uncertainties due to the PDF set are estimated by taking the RMS of all the variations by taking different PDF replicas of the default NNPDF set that were generated varying the parametrization of the PDF, following the PDF4LHC group prescriptions [135]. The impact of the PDF uncertainties ranges between 1 and 4%, depending on the jet category.

Renormalization and factorization scale uncertainties

Renormalization (μ_R) and factorization (μ_F) scale uncertainties represent the impact on the M_T distributions of the missing higher-order QCD contributions to the cross section and to PDF evolution of the simulated samples. These uncertainties are determined by varying these scales by 0.5 and 2, either one by one or both by the same factor at the same time. The combination giving the biggest shift is kept and the relative difference between this result and the nominal one is taken as a shape uncertainty.

Scale uncertainties on exclusive jet categories, i.e. the 0-jet category in this case, are hard to estimate since they contain compensating corrections coming from higher-order in QCD and from the jet veto. To estimate them as realistically as possible, the prescription of [136] was used. First, an auxiliary inclusive ≥ 0 jet category is defined and the exclusive = 0 jet category is defined as: $N_{=0j} = N_{\geq 0j} - N_{\geq 1j}$. Then, the uncertainty on the 0-jet category is simply given by:

$$\Delta\sigma^{=0j} = \sqrt{(\Delta\sigma^{\geq 0j})^2 + (\Delta\sigma^{\geq 1j})^2}. \quad (5.11)$$

Scale uncertainties are reported in Table 5.6. Such uncertainties affect both the global normalization of the process but also its division in jet bin category (migration effect). For the ZZ and WZ backgrounds, their values are found to be around 5–10%, except for the VBF category where it is about 35%. Possible hints for the difference between the VBF and other jet categories in the di-boson backgrounds are that the VBF events are in a limited number and that these events are already generated in a particular phase space requiring $ZZ \rightarrow 2\ell 2\nu + 2jets$ events.

Since the cross section of the signal is the parameter of interest of this search, the sum of the cross sections in the three jet categories obtained after scale variations is normalized to the nominal total cross section value in order to reflect only the jet category migration.

Process	=0 jet	≥ 1 jet	VBF
ZZ	+5.8%	+5.4%	+43%
	-6.3%	-5.3%	-29%
WZ	+9.2%	+4.6%	+44%
	-9.7%	-5.0%	-28%
ggH signal	-3.2%	+3.1%	+6.0%
$(M_H; \Gamma_H) = (800 \text{ GeV}; 100 \text{ GeV})$	+0.3%	-0.3%	+2.1%

Table 5.6.: Renormalization and factorization scale uncertainties (in %). For backgrounds, upper (lower) values represent upper (lower) variations of the cross section. For the signal, these are upper and lower shape variations representing jet bin migration.

The variation of μ_R and μ_F was also used to estimate the uncertainty on the QCD NNLO-to-NLO cross section correction described on Fig. 5.6 which is evaluated to 10% [126].

NLO-to-LO electroweak correction uncertainties

The uncertainty coming from the virtual component of the correction is computed in the same way for corrections on ZZ and WZ . Two event categories are defined according to the event recoil [128], which is defined as $\rho \equiv \frac{|\sum_{i=1}^4 \vec{p}_T^i|}{\sum_{i=1}^4 |\vec{p}_T^i|}$, with i a lepton coming from one of the two final bosons. The relative uncertainty on the corrected cross section is then given by:

$$\delta = \begin{cases} |(1 - K_{QCD}^{NLO})(1 - K_{EWK}^{NLO})|, & \text{if } \rho < 0.3 \\ |1 - K_{EWK}^{NLO}|, & \text{if } \rho \geq 0.3 \end{cases}$$

NLO EWK corrections have been computed to LO QCD diagrams. When $\rho < 0.3$, the event kinematics is well modeled by LO QCD, and the NLO QCD and EWK corrections are supposed to factorize. The uncertainty is then conservatively computed as their product, corresponding to the case where they both go in the same direction. The K_{QCD}^{NLO} -factor is taken from [137]: $K_{QCD}^{NLO} = 15.99/9.89$. When $\rho \geq 0.3$, an even more conservative approach is chosen, taking 100% of the EWK correction as an uncertainty. This case represents $\sim 25\%$ of our events.

Figure 5.17 shows the M_T distribution for the ZZ background at its nominal value of the electroweak corrections and when reweighting the events by $K_{EWK}^{NLO}(1 \pm \delta)$, the value of δ depending on ρ for each event.

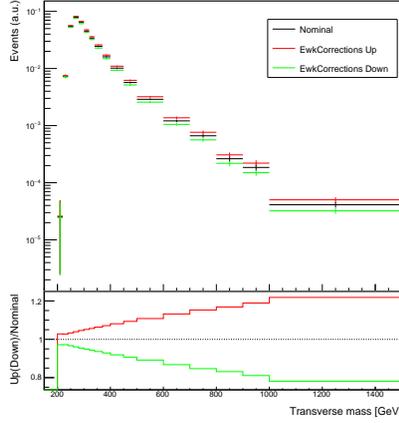


Figure 5.17.: Distribution of the transverse mass for the nominal, upper and lower values of the electroweak corrections for $pp \rightarrow ZZ$.

Concerning $pp \rightarrow WZ$ NLO electroweak corrections, the additional uncertainty due to the photon-induced component comes from the uncertainty on the photon PDF. However, this uncertainty is found to be very small (less than 1% in the whole spectrum) with the LUXqed photon PDF [138] and is neglected. Uncertainties on the WZ NLO electroweak K-factor are shown as error bars on the bottom plot from Fig. 5.7.

5.10. Results

The six final M_T distributions after applying all analysis cuts are shown on Fig. 5.18. Signals of $(M_H; \Gamma_H) = (800 \text{ GeV}; 100 \text{ GeV})$ produced by ggF (continuous line) and VBF (dashed line) are overlaid for a typical cross section of 0.05 pb to which this analysis is sensitive. The total uncertainties on the backgrounds are shown both for the systematic uncertainties (blue shading) and for the total statistical+systematic uncertainties (black shading).

One can see that the ZZ and WZ backgrounds indeed dominate at high M_T . The Z +jet instrumental background, in green, is small but non-zero at high M_T , which makes it interesting to study further. The statistical uncertainty on the backgrounds (and mainly on the Z +jet background) plays an important role in the sensitivity of this analysis as one can guess here and as already shown on Fig. 5.13. The source of this uncertainty is studied into more details in next chapter.

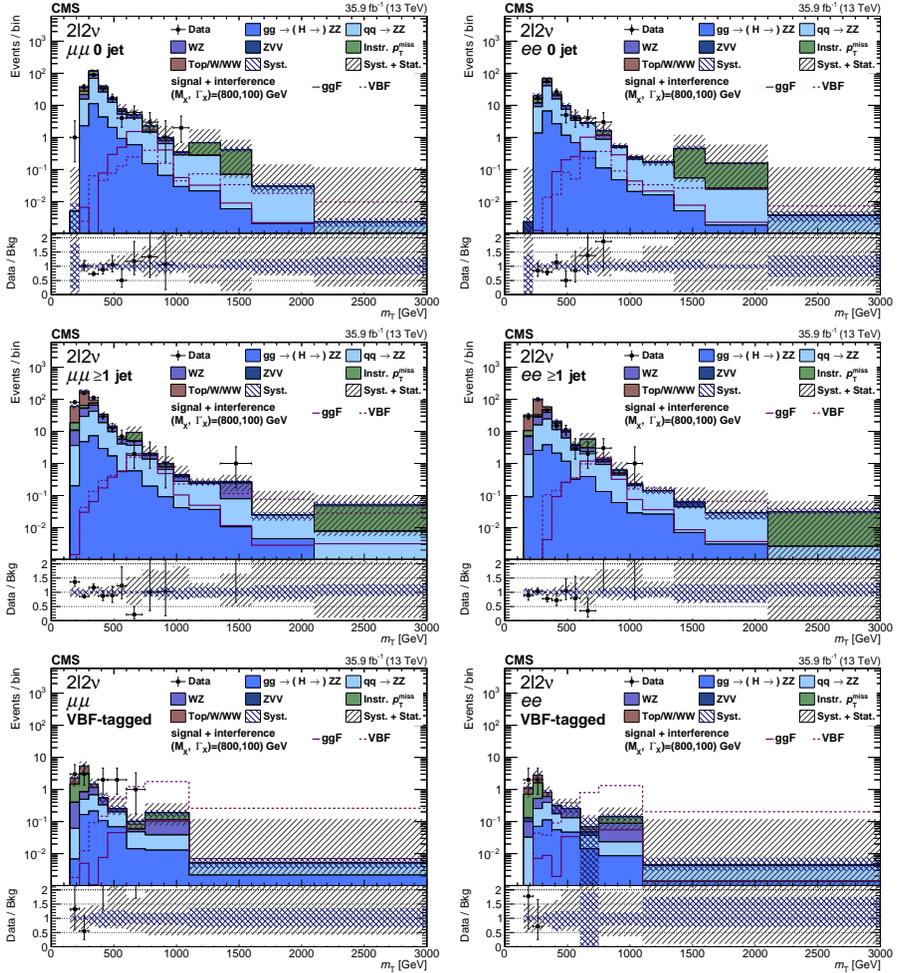


Figure 5.18.: Final M_T distributions for the $\mu\mu$ (left) and ee (right) channels and the 0 jet (top), ≥ 1 jet (center) and VBF (bottom) categories. See text for details.

The number of observed and expected events are summarized in Table 5.7 along with their uncertainties. A fair agreement is observed and no significant excess is seen. In particular, no concentration of high- M_T events is seen in data.

Upper limits using the CL_s method at 95% CL (see Sec. 5.6.4) have been set on the $pp \rightarrow H \rightarrow ZZ$ cross section as a function of M_H and Γ_H . In particular, M_H has been scanned from 300 to 3000 GeV for three Γ_H values equal to 5, 10 and 100 GeV.

Channel	Inclusive	$\mu\mu = 0$ jet	$\mu\mu \geq 1$ jet	$\mu\mu$ VBF	$ee = 0$ jet	$ee \geq 1$ jet	ee VBF
ZZ	$332.8 \pm 1.0^{+13.7}_{-13.6}$	$113.2 \pm 0.6^{+9.0}_{-8.4}$	$94.9 \pm 0.5^{+8.0}_{-7.7}$	$1.27 \pm 0.06^{+0.51}_{-0.35}$	$67.4 \pm 0.4^{+4.9}_{-5.6}$	$55.3 \pm 0.4^{+4.4}_{-4.9}$	$0.74 \pm 0.04^{+0.22}_{-0.22}$
WZ	$176.5 \pm 4.0^{+5.9}_{-5.4}$	$39.1 \pm 1.9^{+3.0}_{-2.5}$	$69.6 \pm 2.5^{+3.8}_{-4.0}$	$1.5 \pm 0.3^{+0.7}_{-0.1}$	$22.0 \pm 1.4^{+1.5}_{-1.2}$	$43.5 \pm 2.0^{+3.0}_{-2.3}$	$0.86 \pm 0.24^{+0.35}_{-0.09}$
ZVV	$6.8 \pm 0.4^{+2.1}_{-1.1}$	$0.45 \pm 0.10^{+0.20}_{-0.16}$	$3.8 \pm 0.3^{+1.8}_{-1.8}$	$0.06 \pm 0.03^{+0.03}_{-0.03}$	$0.22 \pm 0.05^{+0.06}_{-0.05}$	$2.1 \pm 0.2^{+1.0}_{-0.9}$	$0.04 \pm 0.03^{+0.06}_{-0.06}$
Instr. MET	$123.0 \pm 12.3^{+25.2}_{-18.4}$	$38.0 \pm 9.93^{+17.6}_{-11.9}$	$38.0 \pm 11.9^{+12.8}_{-10.7}$	$3.44 \pm 1.93^{+0.39}_{-0.04}$	$21.2 \pm 5.62^{+9.9}_{-6.8}$	$20.7 \pm 7.59^{+8.0}_{-6.2}$	$1.73 \pm 1.13^{+0.20}_{-0.02}$
Top/W/WW	$313.5 \pm 13.3 \pm 27.0$	$18.4 \pm 3.5 \pm 2.4$	$181.7 \pm 11.2 \pm 23.6$	$3.4 \pm 0.7 \pm 0.4$	$10.0 \pm 1.9 \pm 1.3$	$98.2 \pm 6.0 \pm 12.8$	$1.8 \pm 0.4 \pm 0.2$
Total Bkgd.	$952.7 \pm 23.0^{+39.4}_{-35.4}$	$209.2 \pm 10.7^{+20.0}_{-14.8}$	$388.0 \pm 16.5^{+28.0}_{-27.1}$	$9.7 \pm 2.1^{+0.8}_{-0.6}$	$120.8 \pm 6.1^{+11.1}_{-8.9}$	$219.8 \pm 9.9^{+15.7}_{-15.0}$	$5.2 \pm 1.2^{+0.4}_{-0.3}$
Data	938	194	405	11	115	209	4

Table 5.7.: Expected number of background events for an integrated luminosity of 35.9 fb^{-1} after applying the full selection compared to the number of observed events in data. Statistical uncertainties on these yields are shown along with upward and downward systematic uncertainties.

Figure 5.19 shows these upper limits for two production scenarios: ggF only (top) and VBF only (bottom).

Below 300 GeV, the acceptance of the analysis cuts and in particular of the $|\vec{p}_T^{\text{miss}}| > 125$ GeV becomes too small and limits are not set. The expected and observed limits on the pure VBF production cross section are better than the ggF ones, because backgrounds are smaller in the dedicated VBF categories. Limits are comparable in the three scalar widths cases, which is expected since the observable M_T is not very sensitive to the width.

As aforementioned, a combined analysis of $H \rightarrow ZZ \rightarrow 4\ell$, $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ and $H \rightarrow ZZ \rightarrow 2\ell 2q$ has been published by the CMS collaboration [27]. In this combined analysis, no significant excess of events over the SM expectation is observed. Limits are set on the product of the cross section and the branching fraction for a new scalar boson X to decay to ZZ for a wide range of masses and widths, and for different production mechanisms. The two dominant production mechanisms of a scalar boson are ggF and VBF production. The parameter f_{VBF} is defined as the fraction of the VBF production cross section with respect to the total cross section. Figure 5.20 shows the expected and observed upper limits at the 95% CL on the $pp \rightarrow X \rightarrow ZZ$ cross section as a function of m_X and for several width Γ_X and f_{VBF} values. The reported cross section corresponds to the signal only contribution in the absence of interference. The $H \rightarrow ZZ \rightarrow 4\ell$ channel is dominant below 500 GeV due to its best mass resolution, while the $H \rightarrow ZZ \rightarrow 2\ell 2q$ channel is dominant above 700 GeV due to its higher branching ratio. The $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ channel is more sensitive in the intermediate mass region between 500 GeV and 700 GeV but one can see that it helps to improve the limits also outside of this range.

A similar search was performed by the ATLAS collaboration which combined $H \rightarrow ZZ \rightarrow 4\ell$ and $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ channels for data collected by ATLAS at the LHC in 2015 and 2016 for a total of 36.1 fb^{-1} at 13 TeV [139]. No significant excess of events over the SM expectation is observed and limits are set, as shown on Fig. 5.21 and Fig 5.22, on the product of the cross section and the branching fraction of a narrow-width (NWA) or large-width (LWA) scalar boson to two Z . The interference between the heavy scalar and the SM scalar boson as well as the heavy scalar and the $gg \rightarrow ZZ$ continuum are taken into account.

The limits are similar compared to the CMS results for both $H \rightarrow ZZ \rightarrow 4\ell$ and $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ channels. The combined limits is however more stringent at high mass in the case of CMS thanks to the addition of the $H \rightarrow ZZ \rightarrow 2\ell 2q$ channel.

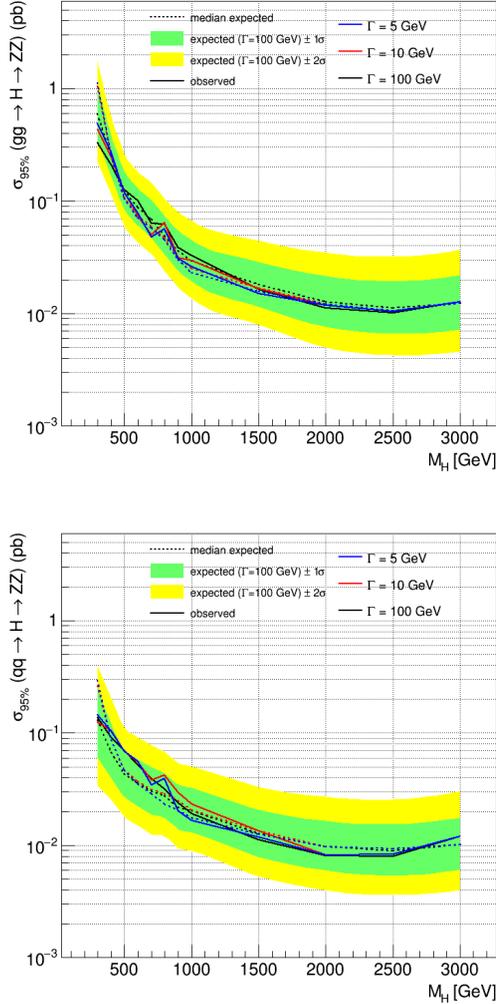


Figure 5.19.: Expected and observed upper limits at 95% CL set on the ggF (top) and VBF (bottom) production cross section times branching ratio of a scalar boson as function of its mass for various values of width. The reference expected limit and the $\pm 1\sigma$ and $\pm 2\sigma$ expected bands are the ones for a scalar width of 100 GeV.

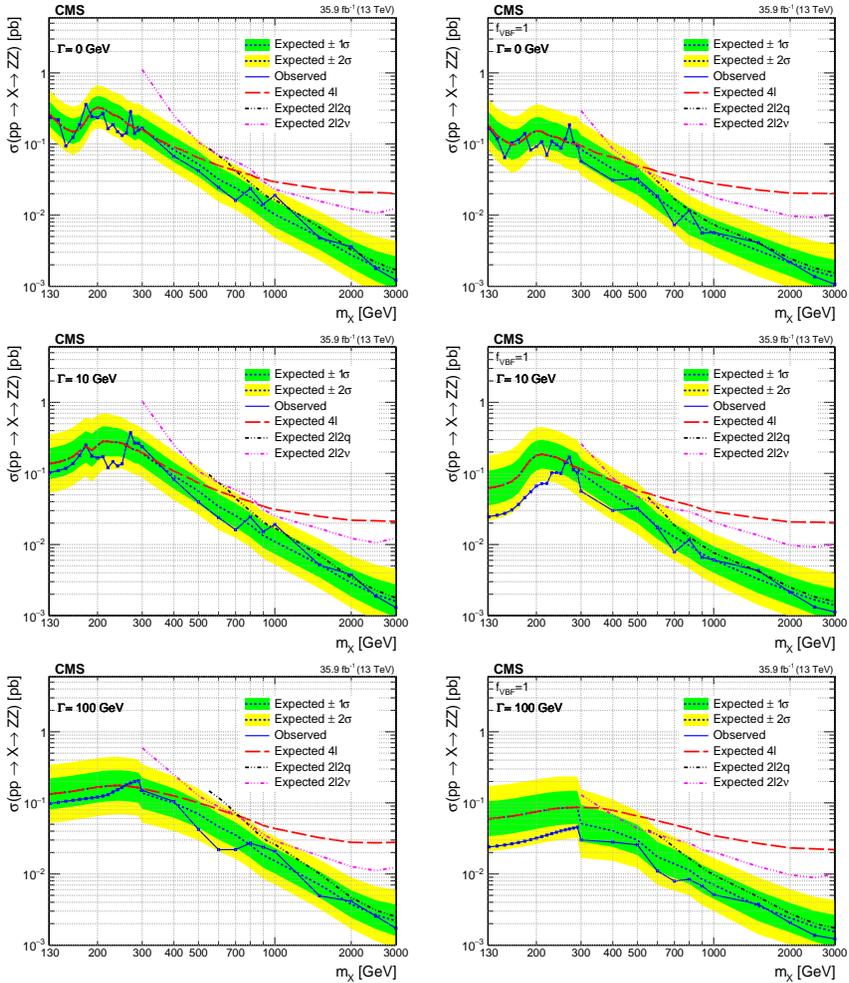


Figure 5.20.: Results of the combined $H \rightarrow ZZ$ searches by the CMS collaboration with 2016 data showing the expected and observed upper limits at the 95% CL on the $pp \rightarrow X \rightarrow ZZ$ cross section as a function of m_X and for several width Γ_X . Limits on the left have been produced by letting float f_{VBF} while limits on the right have been produced for $f_{\text{VBF}} = 1$. The reported cross section corresponds to the signal only contribution in the absence of interference [27].

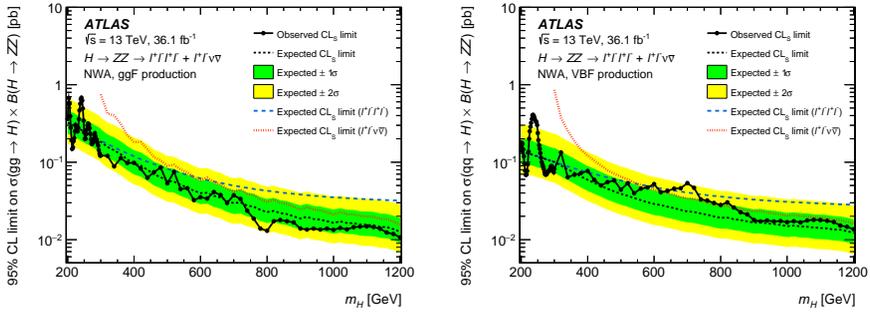


Figure 5.21.: Results of the combined $H \rightarrow ZZ$ searches by the ATLAS collaboration with 2015 and 2016 data showing the upper limits at 95% CL on the cross section times branching ratio as a function of the heavy resonance mass m_H for (left) the ggF production mode and (right) for the VBF production mode in the case of the narrow width approximation (NWA) [139].

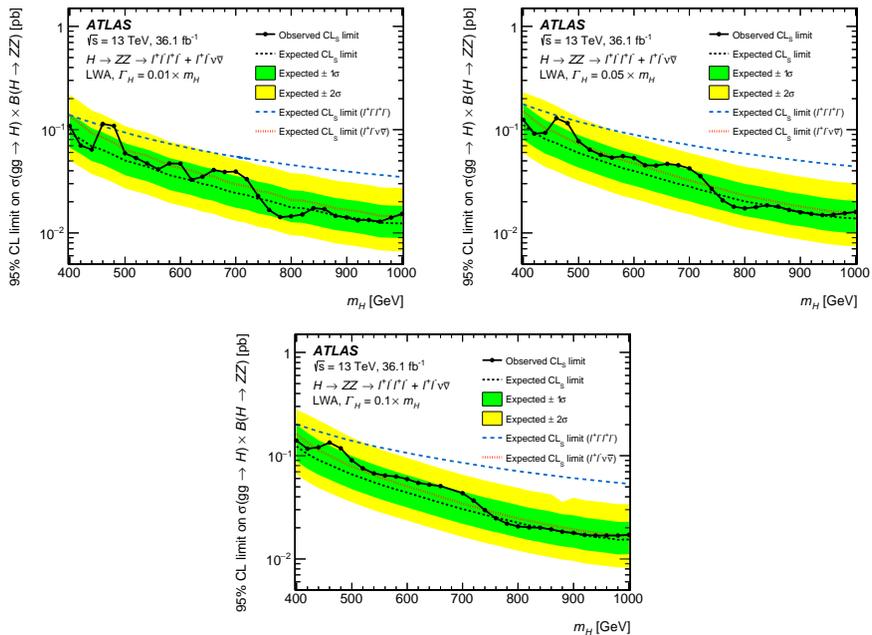


Figure 5.22.: Results of the combined $H \rightarrow ZZ$ searches by the ATLAS collaboration with 2015 and 2016 data showing the upper limits at 95% CL on the cross section for the ggF production mode times branching ratio as function of m_H for an additional heavy scalar assuming a large-width approximation (LWA), i.e. a width of (top left) 1%, (top right) 5%, and (bottom) 10% of m_H ([139]).

Data-driven estimation of the Z +jet background

This chapter presents the motivation and the method used to estimate the Z +jet background in this analysis. After going through the method, control distributions are shown for the γ +jet sample used to estimate the spectrum of the Z +jet background. Then, uncertainties are detailed. Figures and results shown in this chapter have been made with a new analysis framework. The latter has been validated by comparing event by event its results with the previous one. Its main improvement is its capability to perform multiple sub-analyses and checks in an efficient way.

6.1. Motivation

The Z +jet background does not contain genuine \vec{p}_T^{miss} and therefore its yield after applying the signal selection should be much smaller than the expected number of events for a $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ event in the signal region. However, sources of missing transverse energy due to detector energy resolution, jet energy mismeasurements, pileup energy fluctuations and instrumental noise can give a significant \vec{p}_T^{miss} value.

If reaching an artificially high $|\vec{p}_T^{\text{miss}}|$ is rare, the large cross section of Z +jet processes compared to the signal makes the contribution of these processes non-negligible even in high $|\vec{p}_T^{\text{miss}}|$ regions.

Simulating such a process is not expected to give accurate results for two reasons:

1. simulations do not fully reproduce detector and pileup effects on the \vec{p}_T^{miss} distribution, especially in the tails;
2. having high artificial $|\vec{p}_T^{\text{miss}}|$ is rare and too many events should be simulated.

Therefore, to estimate this background a data-driven method has been chosen.

6.2. The data-driven method

In order to construct a data-driven estimation of the Z +jet background, an independent control sample with a single prompt isolated photon associated with jets (γ +jets) is used. The advantage of this dataset is that the main sources of artificial \vec{p}_T^{miss} are expected to be the same than the ones for Z +jets processes. The statistical uncertainty from the yield of the selected γ +jet events was reduced by a suitable choice of prescale values for the single-photon triggers. Nevertheless, some corrections need to be applied to the γ +jet sample due to the difference in trigger selection efficiencies and in kinematics between Z +jet and γ +jet events. The construction of the data-driven estimation is done as follows:

1. select events triggered by single photon triggers and apply the selection detailed in next section;
2. reweight the number of vertices distribution in photon data to the corresponding distributions for di-electron and di-muon data;
3. once done, reweight the photon p_T distribution of γ +jet events to the corresponding di-lepton p_T distributions for di-electron and di-muon events in the three jet categories of the analysis described in the previous chapter;
4. assign a mass to the photon for each γ +jet event by picking randomly a mass value following the $M_{\ell\bar{\ell}}$ distribution (restricted between 76 GeV and 106 GeV) separately in di-electron and di-muon data;
5. then, contaminations of single photon data from processes with genuine \vec{p}_T^{miss} (such as $Z(\rightarrow \nu\bar{\nu})\gamma$, $W(\rightarrow \ell\nu)\gamma$ and $W(\rightarrow \ell\nu)$ +jets) are subtracted by using simulated samples that underwent the same reweightings. Since these processes have real sources of \vec{p}_T^{miss} the simulation of the latter is expected to be accurate. By definition the background cannot be negative. Therefore if statistical fluctuations in some bins lead to a negative estimation of the background, then these bins are set to zero;
6. eventually, the normalization of the data-driven background is found by requiring that the sum of all background processes including the data-driven estimation of the Z +jet process is equal to the number of di-lepton data events in the region $|\vec{p}_T^{\text{miss}}| < 125$ GeV.

6.3. Trigger requirement and event selection

First, events are required to pass photon triggers as described in Table 6.1.

Trigger category	Trigger path
SinglePhoton	Photon22_R9Id90_HE10_IsoM
	Photon30_R9Id90_HE10_IsoM
	Photon36_R9Id90_HE10_IsoM
	Photon50_R9Id90_HE10_IsoM
	Photon75_R9Id90_HE10_IsoM
	Photon90_R9Id90_HE10_IsoM
	Photon120_R9Id90_HE10_IsoM
	Photon165_R9Id90_HE10_IsoM
	HLT_Photon250
	HLT_Photon250_NoHE
	HLT_Photon300
	HLT_Photon300_NoHE

Table 6.1.: Single photon paths used to construct the data-driven estimation of the Z +jets background.

Due to the high cross section of such processes, the triggers are prescaled to have a rate of a few Hz per path, while providing at least ten times as many γ +jet events as di-lepton events in the same p_T range. The paths with a p_T threshold of 165 GeV and above are unrescaled.

Therefore, the very first step consists in applying a weight to the prescaled events before applying a selection close to the one of the di-lepton signal region. This selection requires:

- a single prompt photon γ :
 - in the barrel
 - passing tight photon identification and isolation criteria as described in Sec. 4.4.3
 - with the following additional cuts in order to remove anomalous noisy channels (called “spikes”) in the ECAL: $\sigma_{i\eta i\eta} > 0.001$ and $\sigma_{i\phi i\phi} > 0.001$ (see Sec. 4.4.3)
- $p_{T,\gamma} > 55$ GeV
- a veto on events with reconstructed electrons or muons passing loose or soft identification criteria (see Sec. 4.4)

- a veto on events with at least one b -tagged jet
- $\Delta\phi(\text{jet}, \vec{p}_T^{\text{miss}}) > 0.5$ for each jet above 30 GeV passing the tight jet ID as described in Sec. 4.4.4
- $\Delta\phi(\gamma, \vec{p}_T^{\text{miss}}) > 0.5$

The $\Delta\phi(\gamma, \vec{p}_T^{\text{miss}})$ cut was added after observing discrepancies between data and simulations in the control region. We investigated this issue but could not find the source of the discrepancy. Nevertheless, the signal is expected to have a $Z(\rightarrow \ell\bar{\ell})$ boson and the \vec{p}_T^{miss} back-to-back in $\Delta\phi$ as shown on Fig. 5.9. Therefore, this cut enhances the signal to background ratio but also solves a poorly modeled phase space of the photon region.

Taking photons only from the barrel does not introduce a bias in the method while accepting photons from the endcaps leads to an increase in fake photons — for example coming from beam halos. There exist filters for such events but since the uncertainty is not lead by the amount of γ + jet events and no bias is observed with photon in the barrel only, taking such photons insure a better purity.

6.4. Simulated samples for the photon control region

In order to check if everything is under control, we look at the observed γ +jets events and compare them to simulations. The simulated samples used are described in Table 6.2.

Since some of these simulated samples overlap in phase space, cleanings have to be performed. In particular, HT and p_T^γ are computed for each event and determine if the event should be kept knowing the simulated sample it came from.

An other cleaning that has to be performed is the one between QCD and γ + jet events. Knowing that those QCD samples have been generated allowing for the presence of a prompt photon with a p_T above 10 GeV while the γ + jet samples only allow for prompt photons above 25 GeV, the QCD samples are cleaned from all events with $p_T^\gamma > 25$ GeV.

Eventually, a p_T^γ dependent LO-to-NLO k-factor is applied for γ + jet events. We have estimated this k-factor by comparing NLO samples from the CMS central production to LO samples and comparing their ratio to a first order polynomial of equation 1.72 – $0.0012 * p_T^\gamma$ for $p_T^\gamma < 600$ GeV and equals to one above. The first order polynomial has been given from the JetMET CMS working group while to cut at 600 GeV has been decided to match the observed LO-to-NLO ratio. The NLO

Process	Generator	$\sigma \times BR$ (pb)	QCD precision
γ + jets			
40 GeV < HT < 100 GeV	MadGraph	20 790	LO
100 GeV < HT < 200 GeV	MadGraph	9238	LO
200 GeV < HT < 400 GeV	MadGraph	2305	LO
400 GeV < HT < 600 GeV	MadGraph	274.4	LO
600 GeV < HT	MadGraph	93.46	LO
QCD			
100 GeV < HT < 200 GeV	MadGraph	27 990 000	LO
200 GeV < HT < 300 GeV	MadGraph	1 712 000	LO
300 GeV < HT < 500 GeV	MadGraph	347 700	LO
500 GeV < HT < 700 GeV	MadGraph	32 100	LO
700 GeV < HT < 1000 GeV	MadGraph	6831	LO
1000 GeV < HT < 1500 GeV	MadGraph	1207	LO
1500 GeV < HT < 2000 GeV	MadGraph	119.9	LO
2000 GeV < HT	MadGraph	25.24	LO
$W(\rightarrow \ell\nu)\gamma$	MadGraph5_aMC@NLO	489	NLO
$Z(\rightarrow \nu\bar{\nu})\gamma$			
inclusive	MadGraph5_aMC@NLO	28.04	NLO
$p_T^\gamma > 130$ GeV	MadGraph5_aMC@NLO	0.2768	NLO
$W(\rightarrow \ell\nu) + \text{jets}$			
inclusive	MadGraph5_aMC@NLO	61 526.7	NNLO
100 GeV < HT < 200 GeV	MadGraph	1345 $\times 1.21$	LO
200 GeV < HT < 400 GeV	MadGraph	359.7 $\times 1.21$	LO
400 GeV < HT < 600 GeV	MadGraph	48.91 $\times 1.21$	LO
600 GeV < HT < 800 GeV	MadGraph	12.05 $\times 1.21$	LO
800 GeV < HT < 1200 GeV	MadGraph	5.501 $\times 1.21$	LO
1200 GeV < HT < 2500 GeV	MadGraph	1.329 $\times 1.21$	LO
2500 GeV < HT	MadGraph	0.032 16 $\times 1.21$	LO
$Z(\rightarrow \ell\bar{\ell})\gamma$	MadGraph5_aMC@NLO	117.864	NLO
Top+ γ			
$t\bar{t}\gamma$	MadGraph5_aMC@NLO	3.697	NLO
$t\gamma$	MadGraph5_aMC@NLO	2.967	NLO

Table 6.2.: List of the simulated background samples contributing in the photon control region along with their generator and cross section. For W +jets samples, the cross-section is multiplied by a LO-to-NLO k-factor of 1.21.

simulated samples are not used directly since they have a smaller number of events, especially at high HT.

Once all these corrections are applied, data-to-simulation scale factors are applied as described in Sec. 5.5.1. In particular, scale factors for photon selection efficiency as a function of photon p_T and η are shown on Fig. 6.1.

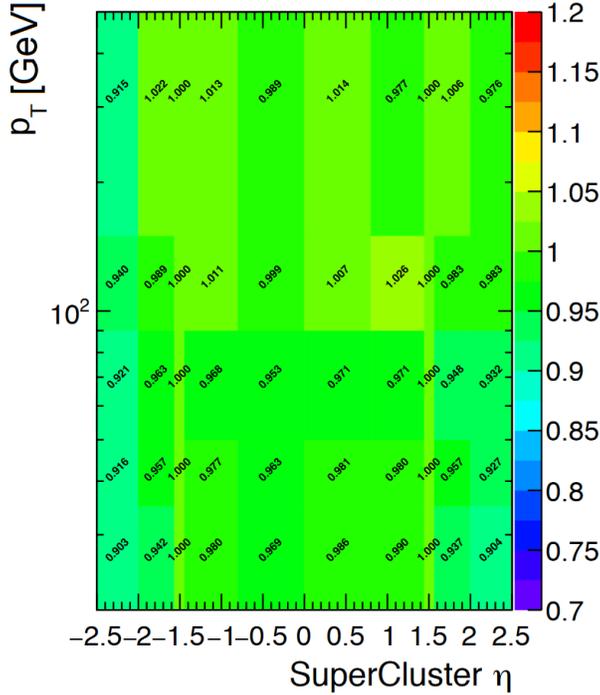


Figure 6.1.: Data-to-simulation scale factors for tightly-selected photons as prescribed by the CMS collaboration [140].

6.5. Photon control region

Figures 6.2 and 6.3 shows six control distributions of γ +jets events in jet categories after applying all corrections and selections described above. The goal of this control region is to check that everything needed for the data-driven method is under control and not to perform a single photon analysis. Therefore some discrepancies will not be investigated since they don't harm the data-driven estimation.

Overall, one can see that the ≥ 1 jet and VBF categories are showing a better agreement between data and simulations than the 0 jet category. The latter shows a lack of simulated events, which is expected since the main contributions are coming from γ +jets and QCD samples which have a minimal HT cut of 40 GeV and 100 GeV respectively and therefore are likely to have at least one jet with $p_T > 30$ GeV. This discrepancy does not harm the method since it comes from γ +jets and QCD simulated samples, which are not used in the data-driven construction of the Z +jets background.

The photon p_T distribution shows a good agreement between data and simulations at low p_T^γ (below 300 GeV). At higher p_T a discrepancy of the order of 10–

20% is observed. We attribute this discrepancy to γ +jet simulated samples where the LO-to-NLO k-factor we apply to this sample sounds like not taking everything into account. However and as aforementioned, simulated γ + jet events do not enter in the data-driven method presented in this chapter and no harm on the method is expected.

A reasonable agreement is also observed for the number of jets and same conclusions are drawn.

The number of vertices distribution shows a shift between data and all simulated samples. This shift is also observed between di-lepton data and simulations. One could reweight the simulations to data but we follow the CMS recommendations to correct simulations to match the average number of pile-up interactions (instead of the number of reconstructed vertices) which is found by multiplying the effective minimum bias cross section with the instantaneous luminosity [105].

The distributions of $\Delta\phi(\gamma, \vec{p}_T^{\text{miss}})$ and $\Delta\phi(\text{jet}, \vec{p}_T^{\text{miss}})$ are both cut at 0.5 where we observed non-negligible discrepancies between data and simulations. The decrease in yields when the \vec{p}_T^{miss} and the photon (jet) are back-to-back is a direct consequence of the cuts. A good agreement is observed.

Eventually, the \vec{p}_T^{miss} distribution shows a discrepancy up to 20% below 100 GeV where the γ +jets and QCD events are dominant. This is expected for the same reasons that lead us to a data-driven estimation of the Z +jets background (and the same trend was already observed on Fig. 5.9). Above 100 GeV, simulated samples with genuine sources of \vec{p}_T^{miss} are dominant and compatible with data.

6.6. Reweighting procedure

After the selection and corrections described above, γ +jets events are reweighted in bins of number of reconstructed vertices and then in bins of boson p_T such that these distributions match the di-lepton distributions in a region where the $|\vec{p}_T^{\text{miss}}|$ cut is inverted ($|\vec{p}_T^{\text{miss}}| < 125$ GeV). The $|\vec{p}_T^{\text{miss}}|$ cut is inverted to avoid taking events from the signal region into the reweighting procedure.

These weights are computed using photon and di-lepton data only and are applied both on photon data and simulations in the photon control region.

Figure 6.4 shows the weights as a function of the number of reconstructed vertices for a reweighting to di-electron and di-muon events. These weights have been computed after merging all jet categories together and normalizing to unity both photon and di-lepton distributions. One can see on this figure that the efficiency to reconstruct and tag a di-lepton event decreases with the increase of the number of reconstructed vertices. In particular, the isolation of electrons is more affected by the increase in pile-up.

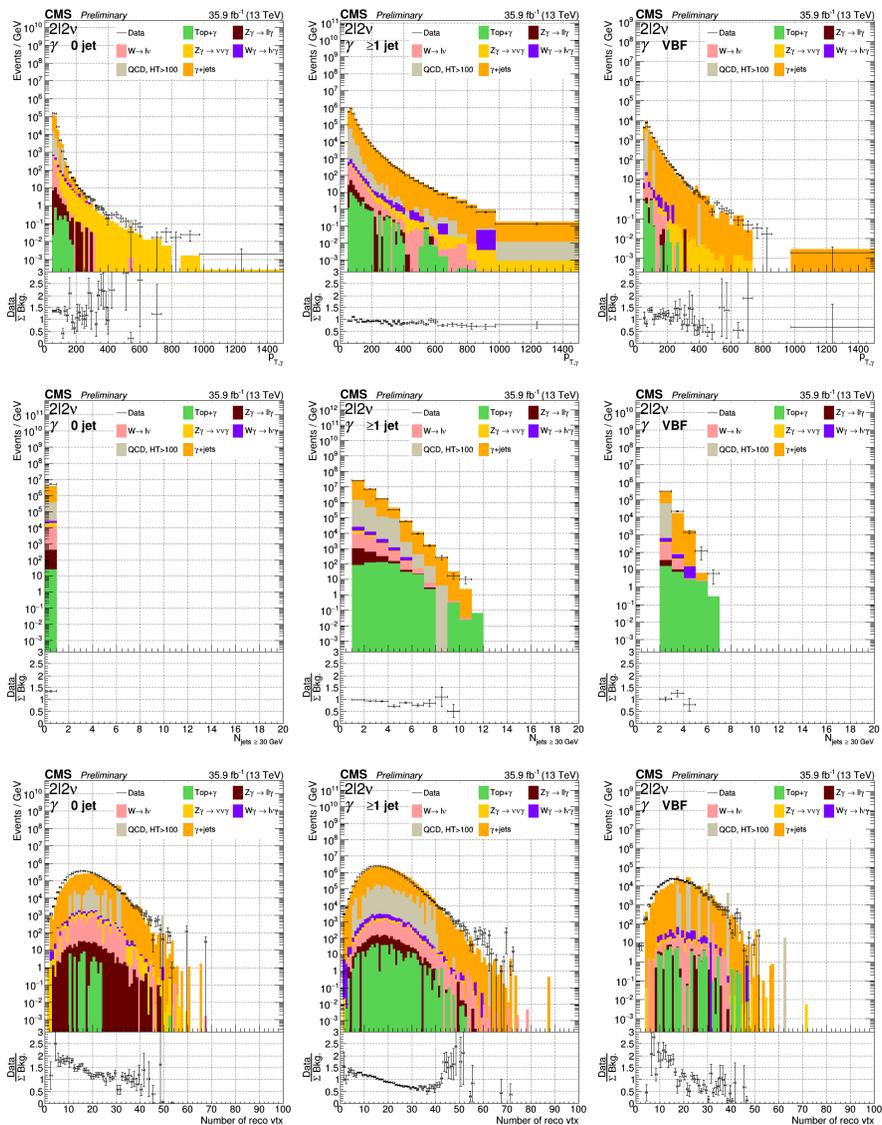


Figure 6.2.: Photon control region in the 0 jet (left), ≥ 1 jet (center) and VBF (right) categories. **Top:** photon p_T . **Center:** number of jets ($p_T > 30 \text{ GeV}$). **Bottom:** number of vertices.

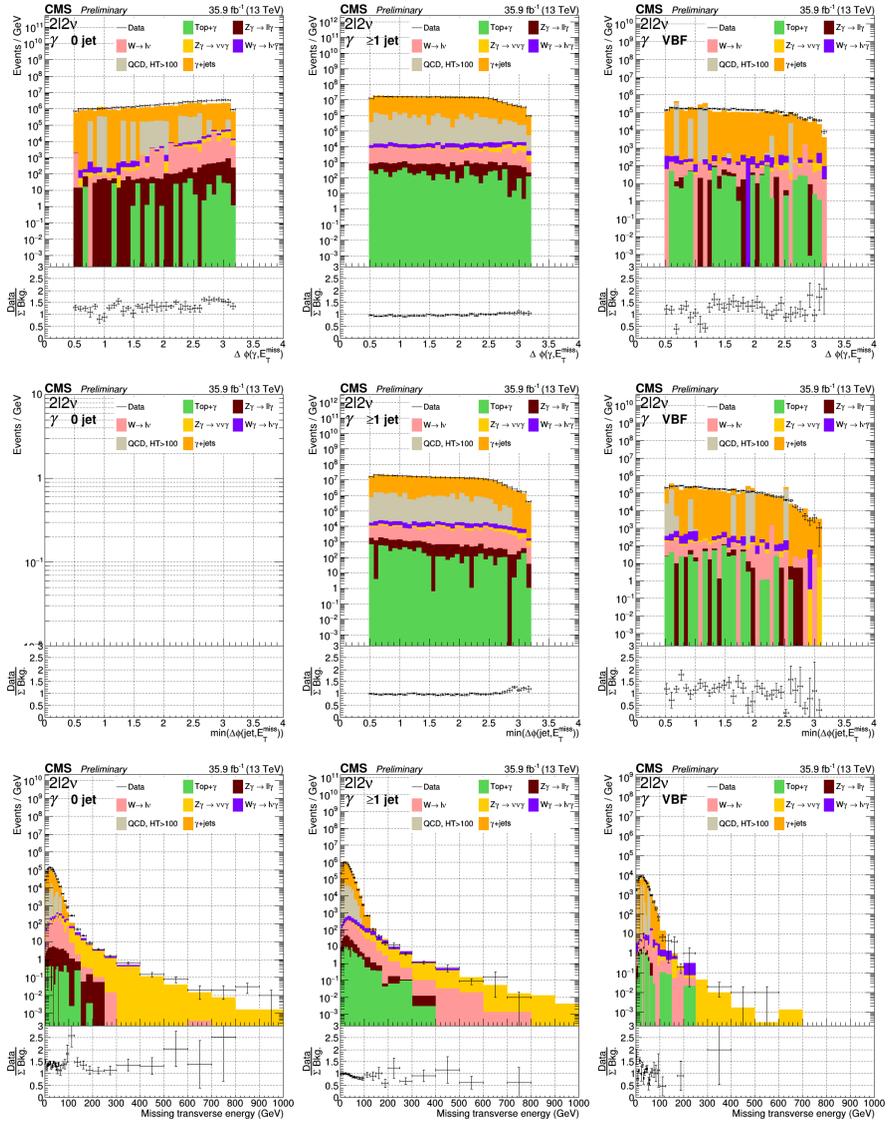


Figure 6.3.: Photon control region in the 0 jet (left), ≥ 1 jet (center) and VBF (right) categories. **Top:** $\Delta\phi$ angle between the photon and \vec{p}_T^{miss} . **Center:** $\Delta\phi$ angle between \vec{p}_T^{miss} and the nearest jet above 30 GeV. **Bottom:** $|\vec{p}_T^{\text{miss}}|$ distribution.

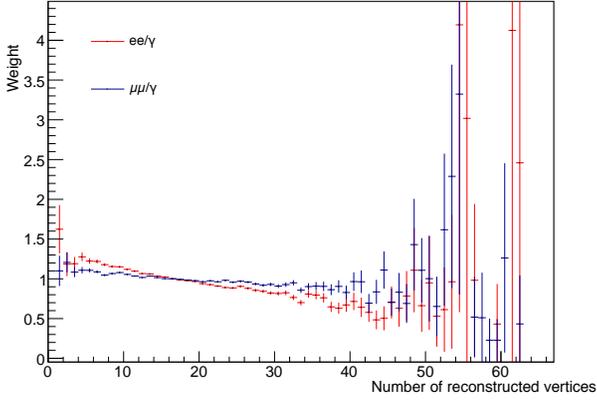


Figure 6.4.: Ratios of the normalized distributions of the number of reconstructed vertices in single photon and di-lepton events after a $|\vec{p}_T^{\text{miss}}| < 125$ GeV cut.

After applying these weights, the same procedure is repeated to compute the weights in bins of boson p_T . This time the ratio is not normalized in order to have a first estimation of the total yield of this background. Figure 6.5 shows these weights as a function of the boson p_T for a reweighting to di-electron and di-muon events. This weight is taking into account the difference in kinematics between the photon and Z boson. These differences decrease at high p_T (above 200 GeV).

Eventually, in order to construct the transverse mass distribution of the Z +jet background, a mass is assigned to the boson in each single photon event by taking a random mass value between 76 GeV and 106 GeV following the $M_{\ell\bar{\ell}}$ distributions separately for di-electron and di-muon data. Figure 6.6 shows the M_T distributions after all these three steps. Overall, a good agreement is found for the ≥ 1 jets and VBF categories both for reweightings to di-electron and di-muon data. As aforementioned, the 0 jet category is affected by the lack of simulated samples for this region. Furthermore, one can see that the distribution corresponding to a reweighting to di-electron data is cut at 800 GeV. This comes from the fact that photon+jet events with $M_T > 800$ GeV have p_T^γ values such that there are no more ee events in the same $p_T^{e\bar{e}}$ range. The weight of these photon+jet events is therefore set to zero.

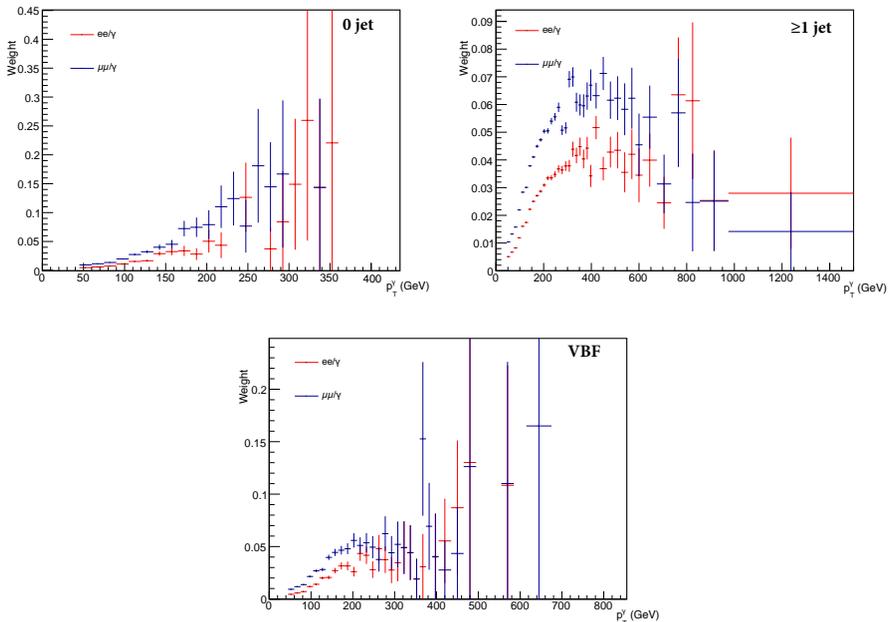


Figure 6.5.: Ratios between di-lepton data and single photon data for their distribution of boson p_T after a cut $|\vec{p}_T^{\text{miss}}| < 125$ GeV in the 0 jet (top left), ≥ 1 jet (top right) and VBF (bottom) categories.

6.7. Closure test

After checking the control region, we perform a closure test of the method using simulations. In particular, the full reweighting procedure is applied using only $\gamma + \text{jet}$ and $Z + \text{jet}$ simulated samples. Figures 6.7 and 6.8 show the comparison of the M_T distribution of $\gamma + \text{jet}$ and $Z + \text{jet}$ events without any $|\vec{p}_T^{\text{miss}}|$ cut since this would remove too many events to conclude. Overall, a good agreement within statistical uncertainties is observed.

6.8. Data-driven estimation in the di-lepton region

As shown in the previous section, the closure test is a success and we therefore use this data-driven method to estimate the $Z + \text{jet}$ background. The very last

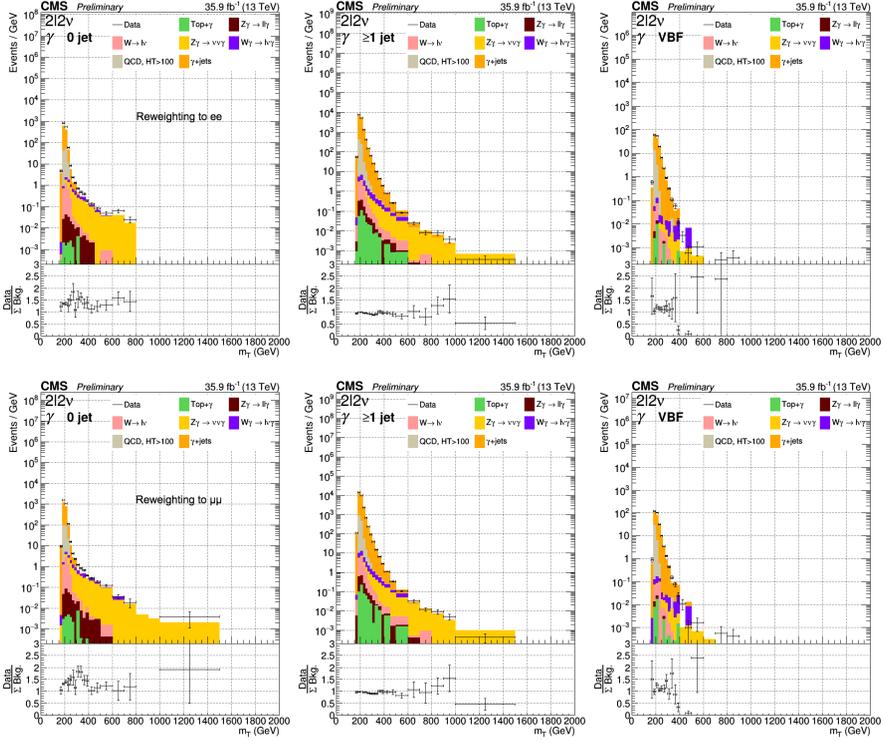


Figure 6.6.: Distributions of M_T for single photon events after all reweightings applied in the three jet categories for a reweighting to di-electron (top row) or di-muon (bottom row) data.

step of the procedure consists in subtracting the contamination of single photon data from simulated processes with genuine \vec{p}_T^{miss} (such as $Z(\rightarrow \nu\bar{\nu})\gamma$, $W(\rightarrow \ell\nu)\gamma$ and $W(\rightarrow \ell\nu)+\text{jets}$) that underwent the same reweightings. Eventually, the normalization of the data-driven background is found by requiring that the sum of the simulated background samples and the data-driven estimation events is equal to the number of di-lepton data events in the region $|\vec{p}_T^{\text{miss}}| < 125$ GeV.

This section presents a comparison of the final distributions of the $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ analysis with and without the data-driven estimation of the Z +jet background.

6.8.1. Changes with respect to published analysis

As aforementioned, the study presented in this chapter has been performed using a new analysis framework introduced after the publication of the 2016 data analysis [27]. It should be noted that, while in the process of building this framework, a few

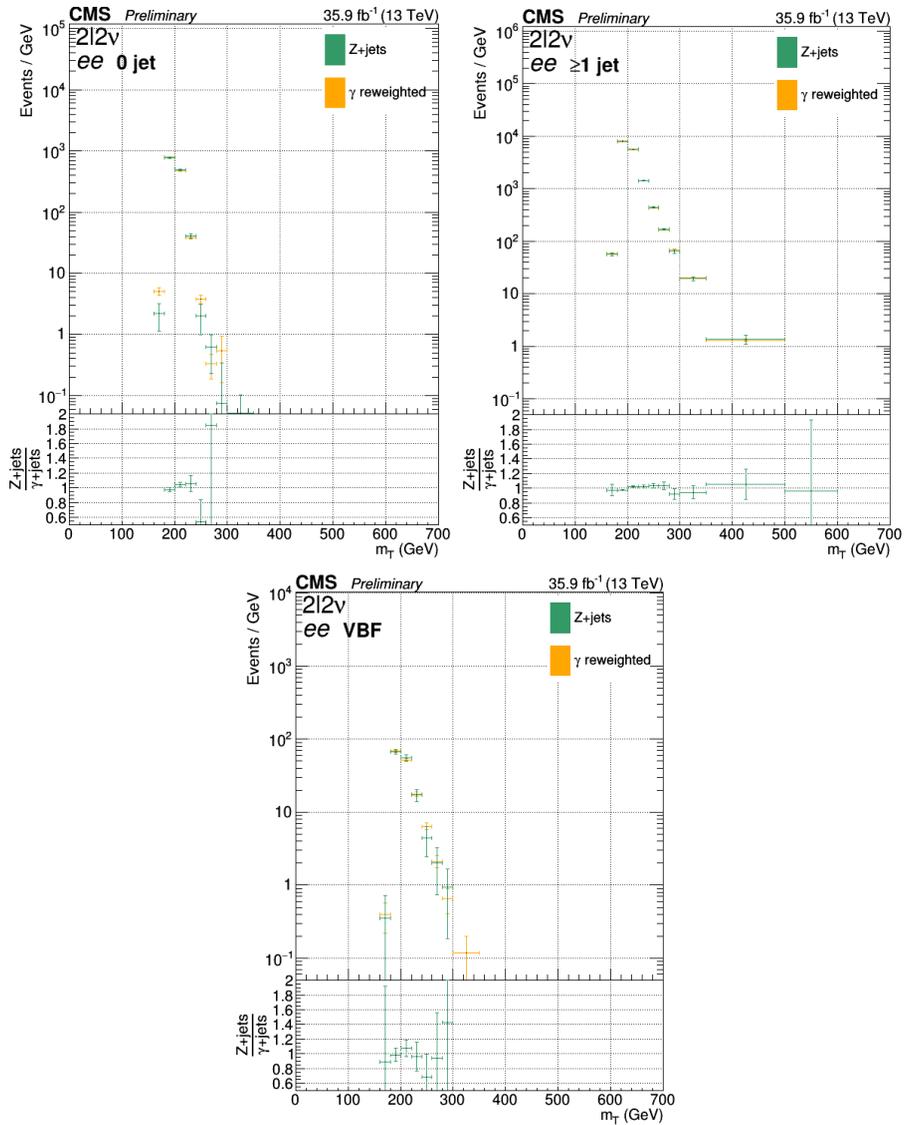


Figure 6.7.: M_T distributions in simulated Z +jet and γ +jet samples after γ +jet event reweighting in the three jet categories with no $|\vec{p}_T^{\text{miss}}|$ cut applied for a reweighting to di-electron events. $Z(\rightarrow ee) + \text{jets}$ events are shown in green while reweighted $\gamma + \text{jets}$ events are shown in orange.

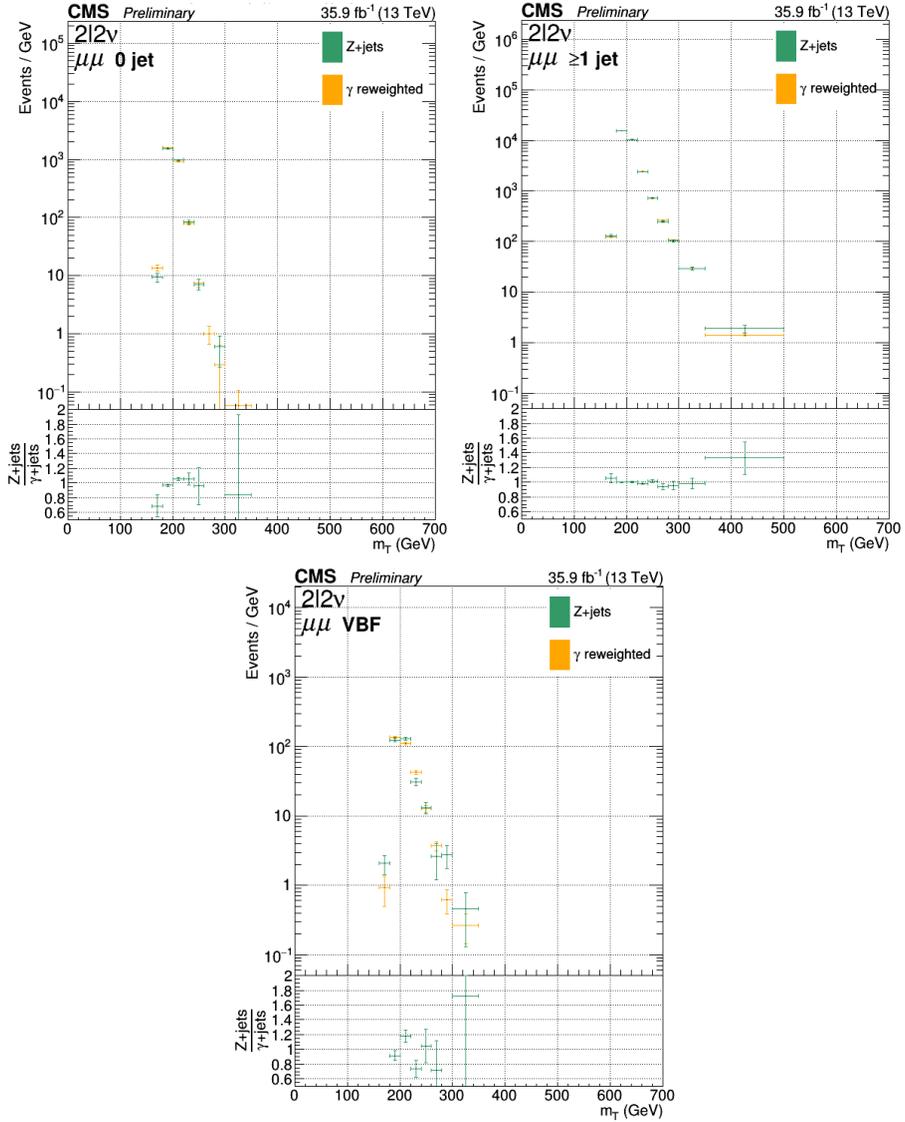


Figure 6.8.: M_T distributions in simulated Z +jet and γ +jet samples after γ +jet event reweighting in the three jet categories with no $|\vec{p}_T^{\text{miss}}|$ cut applied for a reweighting to di-muon events. $Z(\rightarrow \mu\mu) + \text{jets}$ events are shown in green while reweighted $\gamma + \text{jets}$ events are shown in orange.

changes have been made compared to the analysis presented in the previous chapter. In particular, one of the main sources of discrepancies is the addition of the ϕ modulation correction in the definition of the \vec{p}_T^{miss} in the new framework (see Sec. 4.4.5). We observe that the global effect of this correction is a shift of about -2 GeV in the $|\vec{p}_T^{\text{miss}}|$ distribution which leads to lower yields in the signal region of about 2%. An other source is the introduction of additional anomalous \vec{p}_T^{miss} filters and in particular filters for events coming from beam halo and HCAL noise, which also lowers the final yields in the signal region of about 3%. Eventually, while a data-driven estimation of the non-resonant background is used in the published analysis, it is not yet implemented in the new framework at the time of this thesis and a simulated $t\bar{t} \rightarrow 2\ell 2\nu + \text{jet}$ sample is used instead.

6.8.2. Results and comparison with simulated samples in the di-lepton region

Figures 6.9, 6.10, 6.11 and 6.12 show comparisons of the agreement between data and background estimations using $Z + \text{jet}$ simulations or its the data-driven estimation for the $|\vec{p}_T^{\text{miss}}|$ and M_T distributions before any $|\vec{p}_T^{\text{miss}}|$ cut in the di-lepton region. Additional plots can be found in Appendix B. Improvements are clearly visible in all distributions and di-lepton data and their estimations are in agreement once the data-driven estimation of the $Z + \text{jets}$ is used.

Figure 6.13 shows the M_T distributions in the three jet categories after requiring $|\vec{p}_T^{\text{miss}}| > 125$ GeV on top of all the selection. A good agreement between data and background estimations is observed and, as expected, most of the $Z + \text{jets}$ background is suppressed by the $|\vec{p}_T^{\text{miss}}| > 125$ GeV cut.

6.9. Uncertainties

The last step of the study consists in the estimation of uncertainties on this background. We have divided these uncertainties into three categories:

- uncertainties coming from the level of trust in the method;
- statistical uncertainties;
- systematic uncertainties.

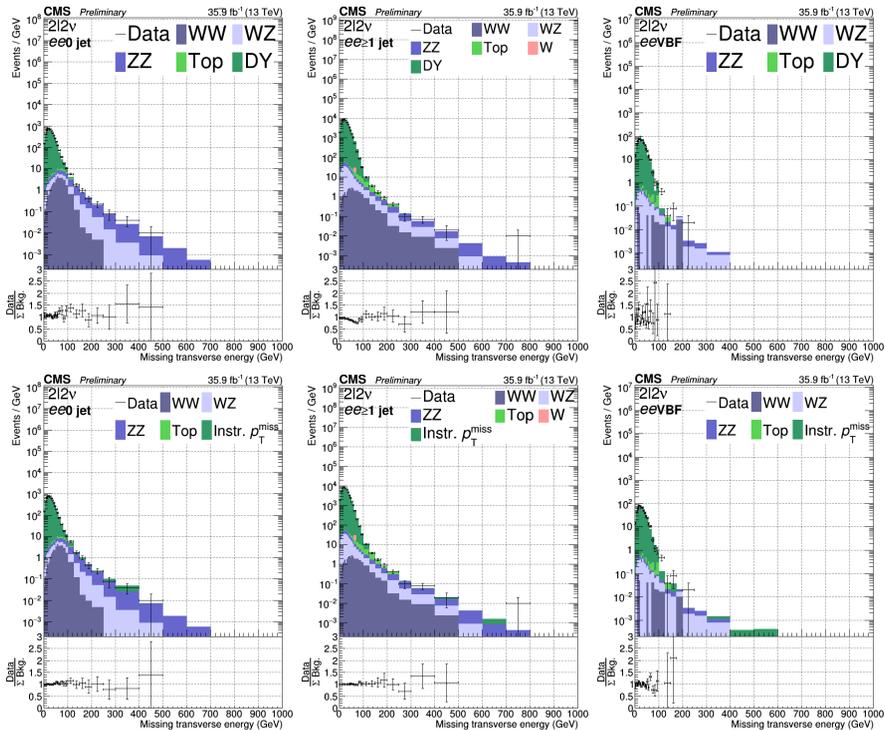


Figure 6.9.: Distributions of $|\vec{p}_T^{\text{miss}}|$ in the di-electron channel using simulated Z + jet samples (top row) and the data-driven estimation of this background (bottom row) before any $|\vec{p}_T^{\text{miss}}|$ cut. The three jet categories are shown: 0 jet (left), ≥ 1 jet (center) and VBF (right).

6.9.1. Uncertainty from the reweighting procedure

Following the results of the closure test, a global uncertainty of 10% is set for all lepton and jet categories. This value is estimated from the inclusive closure test as shown on Fig. 6.14. The choice to evaluate this uncertainty as a global uncertainty and not an uncertainty per category is done in order to improve the statistical precision of the closure test, which is the dominating source of difference between the Z +jet and γ +jet samples at $M_T > 300$ GeV.

6.9.2. Statistical uncertainties

Statistical uncertainties can be divided in two categories:

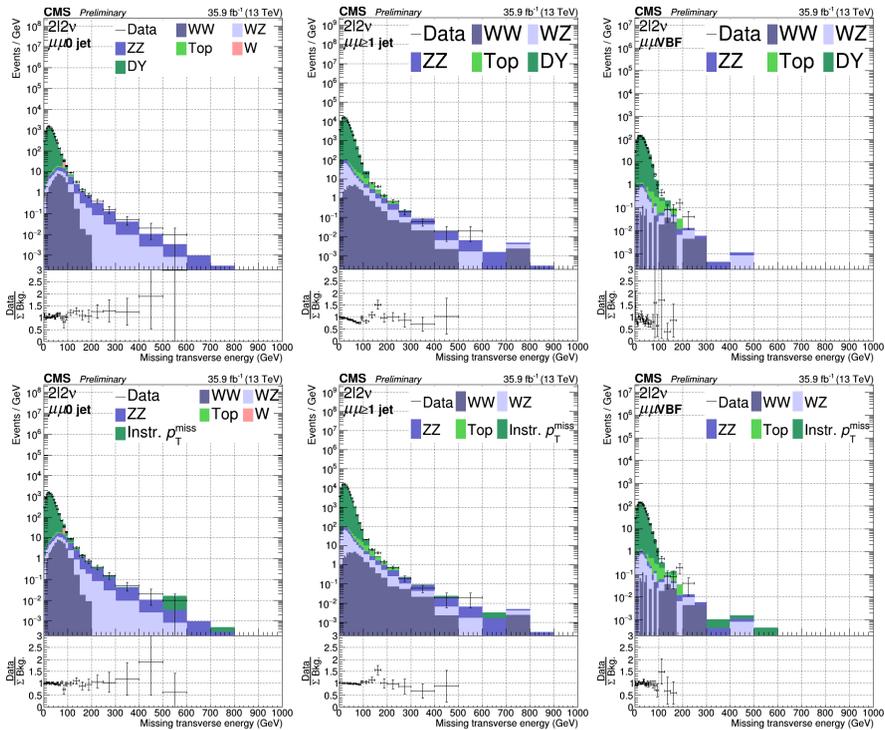


Figure 6.10.: Distributions of $|\vec{p}_T^{\text{miss}}|$ in the di-muon channel using simulated $Z + \text{jet}$ samples (top row) and the data-driven estimation of this background (bottom row) before any $|\vec{p}_T^{\text{miss}}|$ cut. The three jet categories are shown: 0 jet (left), ≥ 1 jet (center) and VBF (right).

- those due to the limited number of observed $\gamma + \text{jet}$ events in the signal region ($|\vec{p}_T^{\text{miss}}| > 125$ GeV), and due to the limited size of the subtracted simulated $Z\gamma$, $W\gamma$ and $W + \text{jet}$ samples
- those due to the number of observed events needed to compute the reweighting weights, hence in the region $|\vec{p}_T^{\text{miss}}| < 125$ GeV.

Figure 6.15 shows the statistical uncertainties in the signal region. Red bands represent statistical uncertainties in the signal region due to both simulations and data, while the blue crosses also include statistical uncertainties from the reweighting. Due to the very small number of events in the signal region compared to the reweighting region, it is expected that statistical uncertainties coming from the reweighting are negligible as shown on Fig. 6.15. However, one could notice that the reweighting uncertainties are more important in the 0 jet than in the VBF category at high- M_T . Actually, even if the 0 jet category has overall more events than the VBF one, there

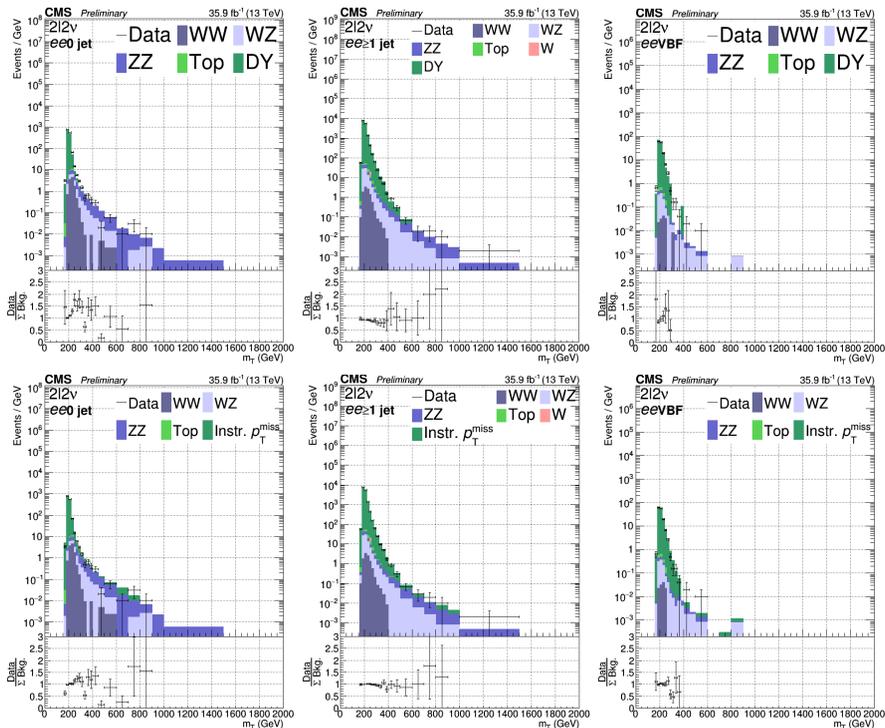


Figure 6.11.: Distributions of M_T in the di-electron channel using simulated Z + jet samples (top row) and the data-driven estimation of this background (bottom row) before any $|\vec{p}_T^{\text{miss}}|$ cut. The three jet categories are shown: 0 jet (left), ≥ 1 jet (center) and VBF (right).

are more events at high- p_T in the VBF category and hence at high- M_T . This is shown on Fig. 6.16.

6.9.3. Systematic uncertainties

The main sources of systematic uncertainties are coming from theoretical uncertainties on the cross sections of the two main simulated backgrounds, i.e. $Z(\rightarrow \nu\bar{\nu})\gamma$ and $W(\rightarrow \ell\nu)\gamma$. Varying scales and PDF choices as already described (see Sec. 5.9), the uncertainties on the total cross sections of these processes are reported in Table 6.3. These values are compatible with [141, 142].

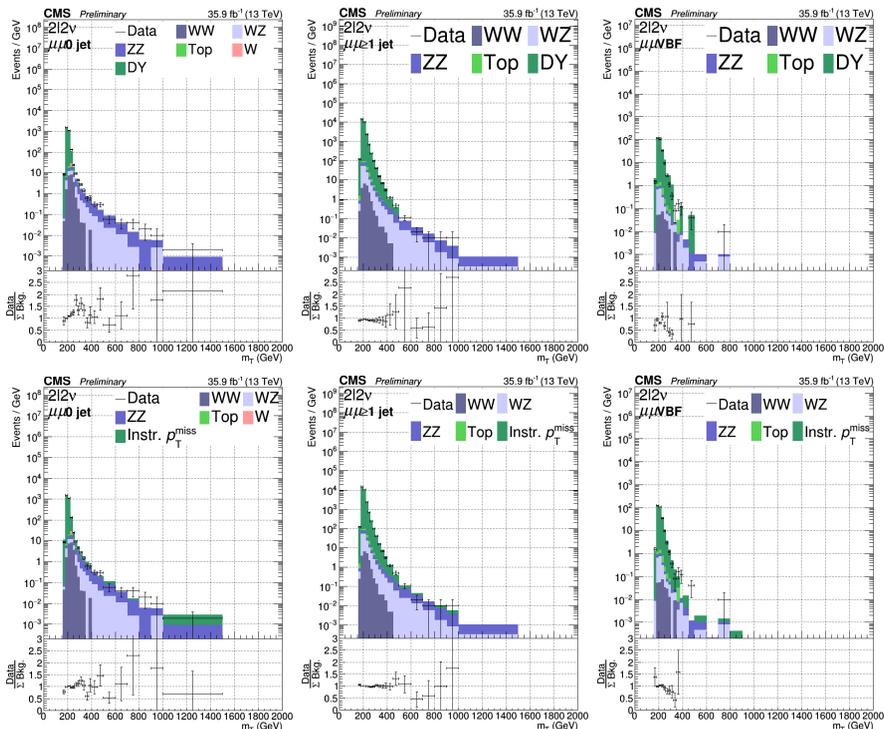


Figure 6.12.: Distributions of M_T in the di-muon channel using simulated Z + jet samples (top row) and the data-driven estimation of this background (bottom row) before any $|\vec{p}_T^{\text{miss}}|$ cut. The three jet categories are shown: 0 jet (left), ≥ 1 jet (center) and VBF (right).

Process	Renormalization and factorization scale	PDF
$Z\gamma$	17%	1%
$W\gamma$	13%	2%

Table 6.3.: Renormalization and factorization scale uncertainties and PDF uncertainties (in %) on the two main simulated samples used in the data-driven estimation in the signal region for all categories.

To estimate the effect of these scale and PDF choice variations on the Z +jet background itself, these theoretical uncertainties on the simulated $Z\gamma$ and $W\gamma$ samples are propagated to yields of the final data-driven estimation in jet category by recomputing the data-driven background but this time by subtracting the upward (or downward) variation of these simulated samples instead of their nominal values. Once again, bins with negative value are forced to zero which can lead to smaller downward uncertain-

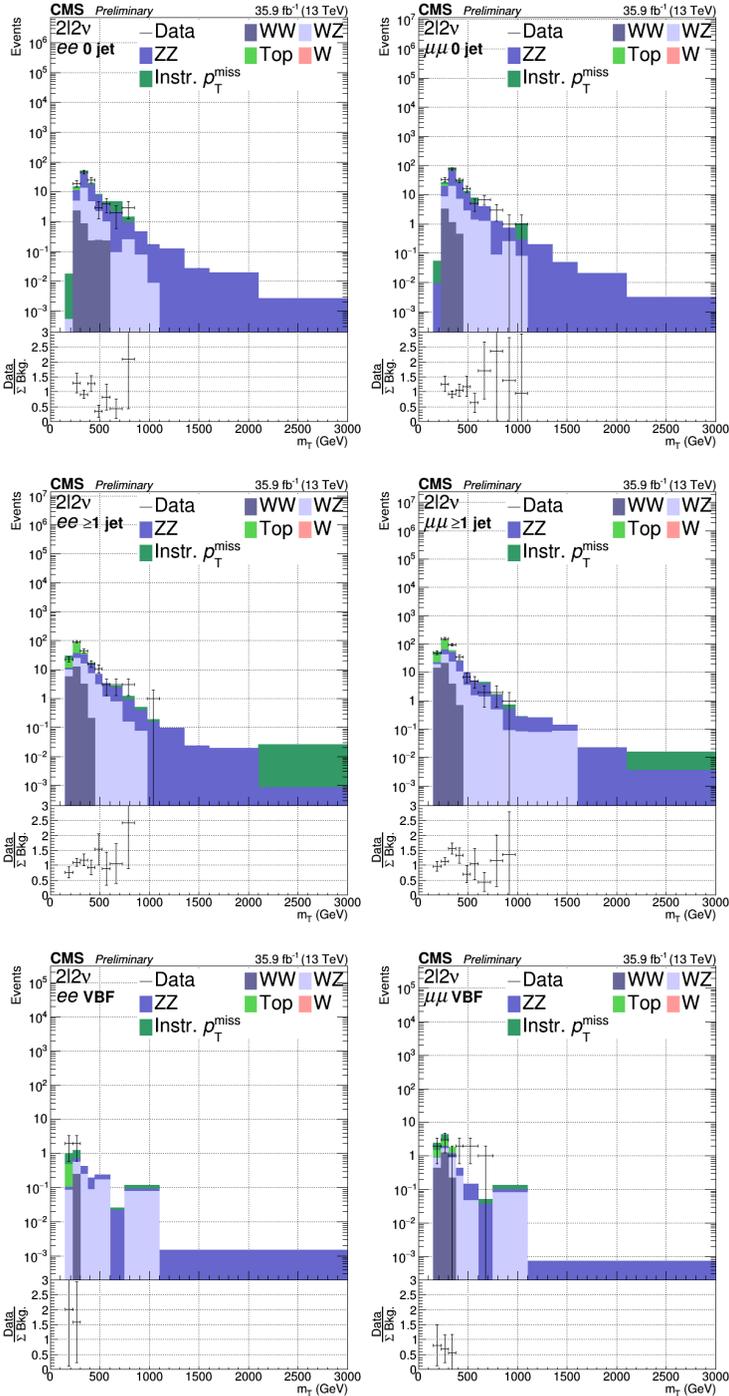


Figure 6.13.: Final M_T distributions for the $\mu\mu$ (left) and ee (right) channels and the 0 jet (top), ≥ 1 jet (center) and VBF (bottom) categories in the new framework, using a data-driven estimation of the Z + jets background.

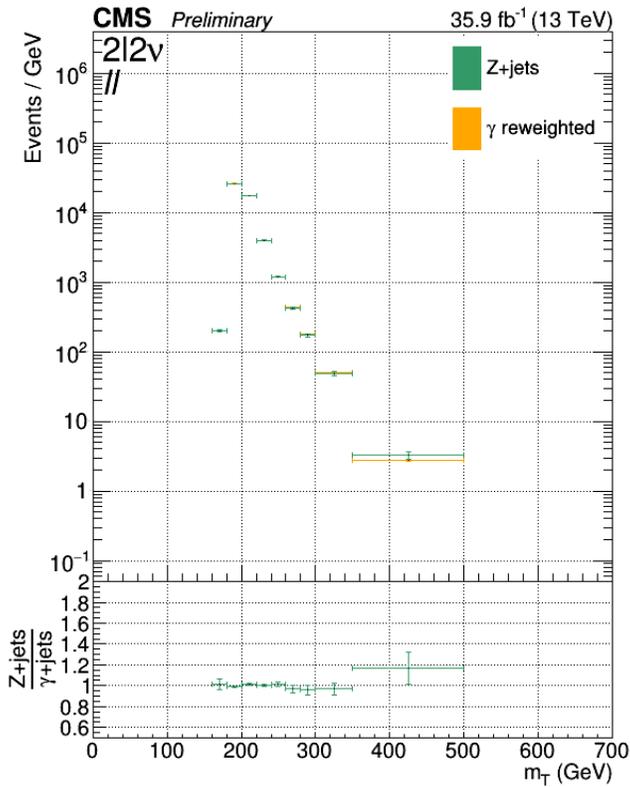


Figure 6.14.: Inclusive closure test. $Z(\rightarrow \ell\bar{\ell}) + \text{jets}$ events are shown in green while reweighted $\gamma + \text{jets}$ events are shown in orange.

ties. Table 6.4 shows the relative effect of these uncertainties propagated to final the data-driven estimation.

As the contribution from $Z\gamma$ and $W\gamma$ events is of the same order than the number of photon data events in the signal region in the 0 jet and ≥ 1 jet categories, the relative uncertainty on the background estimation is large. The absolute background in the signal region is however small compared to the main backgrounds (ZZ , WZ and NRB). The impact on the limit is therefore smaller and of their order of 15% (see Fig. 5.13).

One can see that the systematic uncertainties are smaller in the VBF category. This is due to the fact that, in this region, the subtracted part is close to zero events.

Eventually, Table 6.5 shows the number of predicted $Z+\text{jet}$ events with the new framework along with its uncertainties

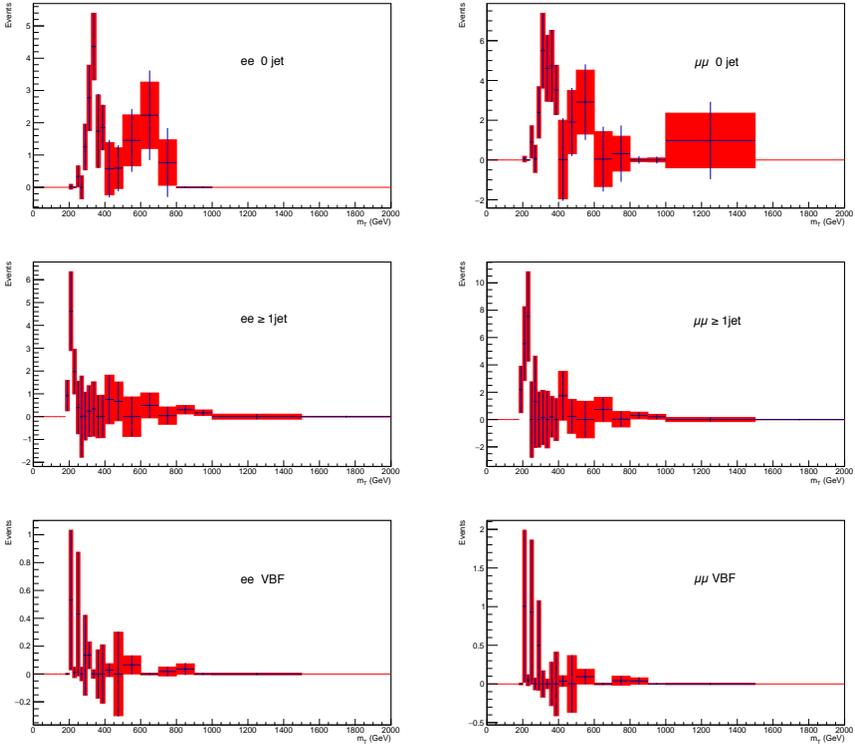


Figure 6.15.: Distributions of M_T for the Z + jets background estimation for the $\mu\mu$ (left) and ee (right) channels and the 0 jet (top), ≥ 1 jet (center) and VBF (bottom) categories. Red bands represent statistical uncertainties in the signal region due to both simulations and data, while the blue crosses also include statistical uncertainties from the reweighting.

Relative systematic uncertainty on the Z +jet background estimation

Source	Renormalization and factorization scale			PDF		
	=0 jet	≥ 1 jet	VBF	=0 jet	≥ 1 jet	VBF
$Z\gamma$	+59%	+47%	+1.4%	+3.5%	+3.2%	+0.3%
	-59%	-29%	-1.5%	-3.4%	-3.1%	-0.3%
$W\gamma$	+44%	+24%	+1.9%	+1.5%	+3.2%	+0.2%
	-44%	-13%	-2.8%	-1.5%	-3.2%	-0.2%

Table 6.4.: Uncertainties (in %) from the renormalization and factorization scale, and the PDF on the final estimation of the Z + jets background.

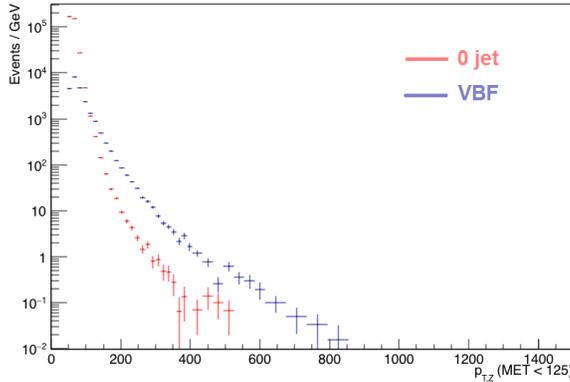


Figure 6.16.: Distributions of p_T^Z in the reweighting region for di-muon events in the 0 jet (red) and VBF (blue) categories.

One can see that the 0 jet category is dominated by systematic uncertainties, the ≥ 1 jet category has a similar contribution coming from statistical uncertainties and systematic uncertainties, while the VBF category is dominated by statistical uncertainties.

6.10. Conclusion

The method used for the data-driven estimation of the $Z + \text{jets}$ background is showing great results. All control plots ($|\vec{p}_T^{\text{miss}}| < 125 \text{ GeV}$) present a good improvement compared to the use of simulations and a good agreement is observed between data and estimations. Moreover, the large sample of $\gamma + \text{jet}$ events allows to control the tails of the \vec{p}_T^{miss} distribution coming from instrumental sources.

The absolute number of $Z + \text{jet}$ background events is expected to be small in the signal region ($|\vec{p}_T^{\text{miss}}| > 125 \text{ GeV}$). Therefore, it is expected that the relative values of the uncertainties on $Z + \text{jet}$ estimation are high. Indeed, by construction, it means that the number of photon data events and the number of subtracted simulated samples such as $Z\gamma$ and $W\gamma$ are of the same order, leading to a nominal value for the $Z + \text{jet}$ estimation close to zero and potentially high relative values due to uncertainties. The later being of the same order for both systematic and statistical uncertainties.

While it is true that raising a bit more the $|\vec{p}_T^{\text{miss}}|$ cut would suppress nearly completely this background, having such an accurate estimation of this background allows to play with the cuts and could be a next step of this analysis: retuning the p_T^Z and $|\vec{p}_T^{\text{miss}}|$ cuts.

Channel	Inclusive	$\mu\mu = 0$ jet	$\mu\mu \geq 1$ jet	$\mu\mu$ VBF	$ee = 0$ jet	$ee \geq 1$ jet	ee VBF
$Z + \text{jets}$	$75.85 \pm 11.35^{+27.89}_{-26.74}$	$27.35 \pm 5.79^{+21.79}_{-21.63}$	$17.64 \pm 8.03^{+8.59}_{-5.74}$	$2.45 \pm 1.62^{+0.25}_{-0.25}$	$17.92 \pm 3.17^{+12.41}_{-12.62}$	$9.39 \pm 4.19^{+5.68}_{-3.47}$	$1.1 \pm 0.85^{+0.12}_{-0.12}$

Table 6.5.: Expected number of Z +jet events for an integrated luminosity of 35.9 fb^{-1} after applying the full selection in the new framework. Statistic uncertainties on these yields are shown along with upward and downward systematic uncertainties.

Conclusion

The first years of operation of the LHC at 7 and 8 TeV (run 1) allowed to probe many aspects of the Standard Model (SM) of particle physics, confirming many predictions up to a high level of precision. An awaited event was the discovery of a scalar boson compatible with the SM in 2012. However, despite its success, the SM cannot explain everything on its own. Models going beyond the SM try to address such issues, and in particular classes of BSM theories require an extended scalar sector. Exploring BSM physics in the scalar sector can be done in two complementary ways: precision measurements of the observed 125 GeV scalar boson and search for direct evidence of new particles.

The scope of this thesis falls within the direct search for heavy BSM scalar boson (H) with the data of the second run of operation of the LHC at 13 TeV (run 2). In particular, the thesis focuses on $H \rightarrow ZZ \rightarrow \ell\bar{\ell}\nu\bar{\nu}$ processes, where ℓ is an electron or a muon and ν a neutrino. This choice is motivated by its sensitivity to a potential discovery of a new scalar. Indeed, the $ZZ \rightarrow \ell\bar{\ell}\nu\bar{\nu}$ decay has the advantage of having a clear signature from the $Z \rightarrow \ell\bar{\ell}$ decay and large missing energy due to neutrinos, while having a bigger branching ratio than $ZZ \rightarrow 4\ell$ due to the $Z \rightarrow \nu\bar{\nu}$ decay.

Together with the $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ group, I analyzed data collected by the CMS experiment during the operation of the LHC in 2015 and 2016. The analysis of the 2.3 fb^{-1} of data collected in 2015 showed no hint of new physics. This analysis is not presented in this thesis and more details can be found in [117].

This thesis therefore focuses on the analysis of the 35.9 fb^{-1} of data collected in 2016 by the CMS experiment. In particular, a statistical analysis of the transverse mass distributions of the reconstructed Z boson and the missing transverse energy is performed in categories of jet and lepton flavour after selecting events in order to enhance the number of potential signal events in comparison to background events. Various scenarii of masses, widths and production mechanisms (gluon fusion and vector

boson fusion) are tested and no excess is observed in data compared to background estimations for a scalar mass between 300 GeV and 3 TeV. Upper exclusion limits on the cross section of a potential signal are set using the CL_s method.

The results of this analysis are combined with two other channels studied by the CMS collaboration: $H \rightarrow ZZ \rightarrow 4\ell$ and $H \rightarrow ZZ \rightarrow 2\ell 2q$. This combination has been published [27] and shows upper exclusion limits on the cross section times branching fraction for a new scalar boson decaying into a pair of Z bosons. No significant hint for BSM physics is observed for a scalar mass between 130 GeV and 3 TeV. In the combined results, one can see that the $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ channel is sensitive in an intermediate mass region (between 300 GeV and 1000 GeV). Below, there are too many backgrounds and a signal with a clearer signature such as $ZZ \rightarrow 4\ell$ is expected to give better results, while at higher mass the $ZZ \rightarrow 2\ell 2q$ signal which has a less clean signature but a higher cross section gives stringent limits since there are less to no backgrounds in this region.

A critical background to control for this analysis is the Z +jet background. While this background does not have genuine missing transverse momentum (\vec{p}_T^{miss}) due to neutrinos, sources of \vec{p}_T^{miss} due to detector energy resolution, jet energy mismeasurements, pileup energy fluctuations and instrumental noise can give a significant \vec{p}_T^{miss} value. If reaching a high instrumental $|\vec{p}_T^{\text{miss}}|$ value is rare, the large cross section of the Z +jet processes compared to the expected signal makes the contribution of these processes non-negligible. Simulating such processes is not expected to give accurate results and a data-driven approach is used.

This data-driven method has been used in the previous analyses performed by the $H \rightarrow ZZ \rightarrow \ell\bar{\ell}\nu\bar{\nu}$ group, tuning it analysis after analysis. It shows great performance, improving clearly the agreement between data and background estimations compared to the use of simulations.

The present thesis is the first time that the uncertainties on the data-driven estimation of the Z +jet background are estimated in detail. Uncertainties are divided into statistical uncertainties, systematic uncertainties and uncertainties inherent to the method, the latter being computed from a closure test of the method and evaluated to 10%. The relative statistical and systematic uncertainties are both of the order of 50%.

Now that the run 2 is over, the logical continuation of this analysis would be to perform the same analysis with the 150 fb^{-1} gathered in the last three years. This would improve the description of the data-driven backgrounds, reducing the relative statistical uncertainties coming from these background estimations. This would also allow to test higher scalar mass scenarios in regions where less to no backgrounds are expected. However, systematic uncertainties bound to the data-driven background estimations, the ZZ irreducible background and the signal description have a non-negligible impact on the limits. Therefore limits in the intermediate scalar mass region between 300 GeV and 800 GeV are expected to improve slightly. An other source of

improvement would be to refine the event selection, taking advantage of the good data-driven description of the Z +jet background.

Not presented in this thesis, I've also worked on the monitoring of the data collected by the tracker of CMS (DQM). Starting as a code developer, I've performed shifts at CERN to monitor in real time data taking as a shifter and then shift leader and expert on-call. Eventually, I've been designated as junior co-convener for the 2016 and 2017 data taking. These two years have been intense for the tracker DQM where a significant drop in reconstruction efficiency was observed (and fixed) in 2016, while 2017 was marked by the scheduled replacement of the pixel detector of the tracker. For the strip detectors of the tracker, the fraction of working channels was stable over the run 2 and around 96.5%. For the pixel detectors in 2015 and 2016, the fraction of working channels was also stable and around 99%. In 2017, the new pixel detector showed an increasing number of faulty channels due to a progressive failure of about 5% of the DC-DC power converters, decreasing the fraction of working channels down to 88% at the end of the year. This had a marginal impact on the 2017 data quality and all DC-DC converters were replaced for 2018 data taking where no such issue was observed.

Bibliography

- [1] NobelPrize.org, “The Nobel Prize in Physics 2013 ”, 2019, <https://www.nobelprize.org/prizes/physics/2013/summary/>.
- [2] CMS Collaboration, “Search for a heavy Higgs boson in the H to ZZ to $2l2\nu$ channel in pp collisions at $\sqrt{s}= 7$ and 8 TeV”, CMS-PAS-HIG-13-014, 2013.
- [3] Galbraith, D. and Burgard, C., “Standard model of the Standard Model”, 2012, <http://davidgalbraith.org/portfolio/ux-standard-model-of-the-standard-model/>.
- [4] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.*, vol. 13, pp. 321–323, 1964, [157(1964)].
- [5] Peter W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.*, vol. 13, pp. 508–509, 1964, [160(1964)].
- [6] Serguei Chatrchyan et al., “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett.*, vol. B716, pp. 30–61, 2012.
- [7] Georges Aad et al., “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys. Lett.*, vol. B716, pp. 1–29, 2012.
- [8] Pran Nath et al., “The Hunt for New Physics at the Large Hadron Collider”, *Nucl. Phys. Proc. Suppl.*, vol. 200-202, pp. 185–417, 2010.
- [9] H. M. Georgi, S. L. Glashow, M. E. Machacek, and Dimitri V. Nanopoulos, “Higgs Bosons from Two Gluon Annihilation in Proton Proton Collisions”, *Phys. Rev. Lett.*, vol. 40, pp. 692, 1978.

- [10] CMS Collaboration, “LHC Higgs Cross Section Working Group”, accessed in 2018, <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>.
- [11] A.N. Schellekens, “Beyond the Standard Model”, 2017, Nikhef, lecture notes.
- [12] Frère, Jean-Marie, “Physics beyond the standard model”, 2017, ULB, lecture notes.
- [13] Royal Swedish Acad. Sci., “Scientific Background on the Nobel Prize in Physics 2015: Neutrino oscillations”, *The Universe*, vol. 3, no. 4, pp. 38–50, 2015.
- [14] Y. Farzan and M. Tortola, “Neutrino oscillations and Non-Standard Interactions”, *Front.in Phys.*, vol. 6, pp. 10, 2018.
- [15] M. Tanabashi et al. (Particle Data Group), “The Review of Particle Physics”, *Phys. Rev.*, vol. D 98, pp. 030001, 2018.
- [16] P. A. R. Ade et al., “Planck 2015 results. XIII. Cosmological parameters”, *Astron. Astrophys.*, vol. 594, pp. A13, 2016.
- [17] P. J. E. Peebles and Bharat Ratra, “The Cosmological constant and dark energy”, *Rev. Mod. Phys.*, vol. 75, pp. 559–606, 2003, [592(2002)].
- [18] A. D. Sakharov, “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe”, *Pisma Zh. Eksp. Teor. Fiz.*, vol. 5, pp. 32–35, 1967, [Usp. Fiz. Nauk161,no.5,61(1991)].
- [19] Gustavo Burdman, “New solutions to the hierarchy problem”, *Braz. J. Phys.*, vol. 37, pp. 506–513, 2007.
- [20] Cécile Caillol, Barbara Clerbaux, Jean-Marie Frère, and Simon Mollet, “Precision versus discovery: A simple benchmark”, *Eur. Phys. J. Plus*, vol. 129, pp. 93, 2014.
- [21] Tania Robens and Tim Stefaniak, “LHC Benchmark Scenarios for the Real Higgs Singlet Extension of the Standard Model”, *Eur. Phys. J.*, vol. C76, no. 5, pp. 268, 2016.
- [22] Ketevi Assamagan et al., “The Higgs Portal and Cosmology”, 2016, arXiv:1604.05324.
- [23] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, and Joao P. Silva, “Theory and phenomenology of two-Higgs-doublet models”, *Phys. Rept.*, vol. 516, pp. 1–102, 2012.

- [24] Cécile Caillol, *Scalar boson decays to tau leptons : in the standard model and beyond*, PhD thesis, Université libre de Bruxelles, 2016.
- [25] Nathaniel Craig, Jamison Galloway, and Scott Thomas, “Searching for Signs of the Second Higgs Doublet”, 2013.
- [26] CMS Collaboration, “Summary results of high mass BSM Higgs searches using CMS run-I data”, CMS-PAS-HIG-16-007, 2016.
- [27] CMS Collaboration, “Search for a new scalar resonance decaying to a pair of Z bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP*, vol. 06, pp. 127, 2018.
- [28] Alessandro Broggio, Eung Jin Chun, Massimo Passera, Ketan M. Patel, and Sudhir K. Vempati, “Limiting two-Higgs-doublet models”, *JHEP*, vol. 11, pp. 058, 2014.
- [29] Baradhwaj Coleppa, Felix Kling, and Shufang Su, “Constraining Type II 2HDM in Light of LHC Higgs Searches”, *JHEP*, vol. 01, pp. 161, 2014.
- [30] Abdul Wahab El Kaffas, Wafaa Khater, Odd Magne OGREID, and Per Osland, “Consistency of the two Higgs doublet model and CP violation in top production at the LHC”, *Nucl. Phys.*, vol. B775, pp. 45–77, 2007.
- [31] CMS Collaboration, “Combined measurements of the Higgs boson’s couplings at $\sqrt{s} = 13$ TeV”, CMS-PAS-HIG-17-031, 2018.
- [32] Lyndon Evans and Philip Bryant, “LHC Machine”, *JINST*, vol. 3, pp. S08001, 2008.
- [33] Germain, P. and Dekkers, D. and Manglunki, D., “Introduction aux accélérateurs de particules”, 2014, <http://cds.cern.ch/record/199445/files/CERN-89-07.pdf>.
- [34] Esma Mobs, “The CERN accelerator complex. Complexe des accélérateurs du CERN”, Jul 2016, General Photo.
- [35] H. Damerau, A. Findlay, S. Gilardoni, and S. Hancock, “RF manipulations for higher brightness LHC-type beams”, Proceedings of IPAC2013, Shanghai, China.
- [36] C. Bracco, W. Bartmann, M. Calviani, M. I. Frankl, M. A. Fraser, S. Gilardoni, B. Goddard, V. Kain, A. Lechner, A. Perillo Marcone, M. Meddahi, F. X. Nuiry, and F. M. Velotti, “Operational limits from intercepting devices”, pp. 171–174. 4 p, 2017.

- [37] CMS Collaboration, “CMS Luminosity - Public Results”, accessed in 2017, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [38] by LHC Team and Stefania Pandolfi, “LHC Report: end of 2016 proton-proton operation”, Oct 2016.
- [39] CMS Collaboration, “CMS Data Quality Information - Public Results”, accessed in 2018, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/DataQuality>.
- [40] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST*, vol. 3, no. 08, pp. S08004, 2008.
- [41] Vardan Khachatryan et al., “Pseudorapidity distribution of charged hadrons in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Phys. Lett.*, vol. B751, pp. 143–163, 2015.
- [42] CMS Collaboration, “Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays”, *JINST*, vol. 5, pp. T03021, 2010.
- [43] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST*, vol. 9, no. 10, pp. P10009, 2014.
- [44] CMS Collaboration, “Commissioning and performance of the cms pixel tracker with cosmic ray muons”, *Journal of Instrumentation*, vol. 5, no. 03, pp. T03007, 2010.
- [45] CMS Internal, “CMS pixel detector performance results”, accessed in 2017, <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PixelOfflinePlots>.
- [46] CMS Internal, “CMS tracker detector performance results”, accessed in 2017, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK>.
- [47] CMS Internal, “CMS strip detector performance results”, accessed in 2017, <https://twiki.cern.ch/twiki/bin/viewauth/CMS/StripsOfflinePlots2016>.
- [48] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting”, *Nucl. Instrum. Meth.*, vol. A262, pp. 444–450, 1987.
- [49] CMS Collaboration, “Primary vertex resolution in 2016”, *CMS-DP-2016-041*, Jul 2016.

- [50] CMS Collaboration, *The CMS electromagnetic calorimeter project: Technical Design Report*, Technical Design Report CMS. CERN, Geneva, 1997.
- [51] CMS Collaboration, *Changes to CMS ECAL electronics: addendum to the Technical Design Report*, Technical Design Report CMS. CERN, Geneva, 2002.
- [52] R.M. Brown and D.J.A. Cockerill, “Electromagnetic calorimetry”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 666, no. Supplement C, pp. 47 – 79, 2012, Advanced Instrumentation.
- [53] CMS Collaboration, “CMS ECAL first results with 2016 data”, *CMS-DP-2016-031*, Jun 2016.
- [54] P. Adzic et al., “Reconstruction of the signal amplitude of the CMS electromagnetic calorimeter”, *Eur. Phys. J.*, vol. C46S1, pp. 23–35, 2006.
- [55] CMS Collaboration, *The CMS hadron calorimeter project: Technical Design Report*, Technical Design Report CMS. CERN, Geneva, 1997.
- [56] CMS Collaboration, “Performance of CMS hadron calorimeter timing and synchronization using test beam, cosmic ray, and LHC beam data”, *Journal of Instrumentation*, vol. 5, no. 03, pp. T03013, 2010.
- [57] CMS Collaboration, “Filtering noise in CMS hadron calorimeter”, vol. 5, pp. T03014, 03 2010.
- [58] CMS Collaboration, “HCAL Calibration in 2016”, *CMS-DP-2017-017*, May 2017.
- [59] CMS Collaboration, “Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data”, *Journal of Instrumentation*, vol. 5, no. 03, pp. T03012, 2010.
- [60] CMS Collaboration, *The CMS muon project: Technical Design Report*, Technical Design Report CMS. CERN, Geneva, 1997.
- [61] CMS Collaboration, “Performance of the CMS muon detector and reconstruction with proton-proton collisions at $\sqrt{s}=13$ TeV ”, *CMS-MUO-16-001, CERN-EP-2018-058*, 2018.
- [62] CMS Collaboration, “Web-based monitoring tools for Resistive Plate Chambers in the CMS experiment at CERN”, vol. 9, Oct 2014.

- [63] CMS Collaboration, “Performance of the CMS Drift Tube Chambers with Cosmic Rays”, *JINST*, vol. 5, no. arXiv:0911.4855. CMS-CFT-09-012, pp. T03015 . 45 p, Nov 2009.
- [64] Karol Bunkowski, *Optimization, Synchronization, Calibration and Diagnostic of the RPC PAC Muon Trigger System for the CMS detector*, PhD thesis, Warsaw U., 2009.
- [65] CMS collaboration, “CMS Technical Design Report for the Level-1 Trigger Upgrade”, Tech. Rep. CERN-LHCC-2013-011. CMS-TDR-12, Jun 2013, Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch.
- [66] Andrea Perrotta, “Performance of the cms high level trigger”, *Journal of Physics: Conference Series*, vol. 664, no. 8, pp. 082044, 2015.
- [67] Andy Buckley et al., “General-purpose event generators for LHC physics”, *Phys. Rept.*, vol. 504, pp. 145–233, 2011.
- [68] Stefan Höche, “Introduction to parton-shower event generators”, in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014): Boulder, Colorado, June 2-27, 2014*, 2015, pp. 235–295.
- [69] Johan Alwall et al., “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”, *Eur. Phys. J.*, vol. C53, pp. 473–500, 2008.
- [70] Stefano Frixione, Paolo Nason, and Carlo Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *JHEP*, vol. 11, pp. 070, 2007.
- [71] E. Bagnaschi, G. Degrossi, P. Slavich, and A. Vicini, “Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM”, *JHEP*, vol. 02, pp. 088, 2012.
- [72] Paolo Nason and Carlo Oleari, “NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG”, *JHEP*, vol. 02, pp. 037, 2010.
- [73] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP*, vol. 07, pp. 079, 2014.
- [74] Paolo Nason and Giulia Zanderighi, “ W^+W^- , WZ and ZZ production in the POWHEG-BOX-V2”, *Eur. Phys. J.*, vol. C74, no. 1, pp. 2702, 2014.

- [75] Yanyan Gao, Andrei V. Gritsan, Zijin Guo, Kirill Melnikov, Markus Schulze, and Nhan V. Tran, “Spin determination of single-produced resonances at hadron colliders”, *Phys. Rev.*, vol. D81, pp. 075022, 2010.
- [76] Sara Bolognesi, Yanyan Gao, Andrei V. Gritsan, Kirill Melnikov, Markus Schulze, Nhan V. Tran, and Andrew Whitbeck, “On the spin and parity of a single-produced resonance at the LHC”, *Phys. Rev.*, vol. D86, pp. 095031, 2012.
- [77] Ian Anderson et al., “Constraining anomalous HVV interactions at proton and lepton colliders”, *Phys. Rev.*, vol. D89, no. 3, pp. 035007, 2014.
- [78] Andrei V. Gritsan, Raoul Röntsch, Markus Schulze, and Meng Xiao, “Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques”, *Phys. Rev.*, vol. D94, no. 5, pp. 055023, 2016.
- [79] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands, “A Brief Introduction to PYTHIA 8.1”, *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008.
- [80] Vardan Khachatryan et al., “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J.*, vol. C76, no. 3, pp. 155, 2016.
- [81] S. Agostinelli et al., “GEANT4: A Simulation toolkit”, *Nucl. Instrum. Meth.*, vol. A506, pp. 250–303, 2003.
- [82] Stefano Forte, Lluís Garrido, Jose I. Latorre, and Andrea Piccione, “Neural network parametrization of deep inelastic structure functions”, *JHEP*, vol. 05, pp. 062, 2002.
- [83] Richard D. Ball et al., “Parton distributions for the LHC Run II”, *JHEP*, vol. 04, pp. 040, 2015.
- [84] Guido Altarelli and G. Parisi, “Asymptotic Freedom in Parton Language”, *Nucl. Phys.*, vol. B126, pp. 298–318, 1977.
- [85] Yuri L. Dokshitzer, “Calculation of the Structure Functions for Deep Inelastic Scattering and $e^+ e^-$ Annihilation by Perturbation Theory in Quantum Chromodynamics.”, *Sov. Phys. JETP*, vol. 46, pp. 641–653, 1977, [*Zh. Eksp. Teor. Fiz.*73,1216(1977)].
- [86] V. N. Gribov and L. N. Lipatov, “Deep inelastic $e p$ scattering in perturbation theory”, *Sov. J. Nucl. Phys.*, vol. 15, pp. 438–450, 1972, [*Yad. Fiz.*15,781(1972)].

- [87] Valerio Bertone, Stefano Carrazza, and Juan Rojo, “APFEL: A PDF Evolution Library with QED corrections”, *Comput. Phys. Commun.*, vol. 185, pp. 1647–1668, 2014.
- [88] Stefano Carrazza, Alfio Ferrara, Daniele Palazzo, and Juan Rojo, “APFEL Web”, *J. Phys.*, vol. G42, no. 5, pp. 057001, 2015.
- [89] John C. Collins, Davison E. Soper, and George F. Sterman, “Factorization of Hard Processes in QCD”, *Adv. Ser. Direct. High Energy Phys.*, vol. 5, pp. 1–91, 1989.
- [90] Alexandre Deur, Stanley J. Brodsky, and Guy F. de Teramond, “The QCD Running Coupling”, *Prog. Part. Nucl. Phys.*, vol. 90, pp. 1–74, 2016.
- [91] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, “Parton Fragmentation and String Dynamics”, *Phys. Rept.*, vol. 97, pp. 31–145, 1983.
- [92] John Allison et al., “Geant4 developments and applications”, *IEEE Trans. Nucl. Sci.*, vol. 53, pp. 270, 2006.
- [93] Makoto Asai, Andrea Dotti, Marc Verderi, and Dennis H. Wright, “Recent developments in Geant4”, *Annals Nucl. Energy*, vol. 82, pp. 19–28, 2015.
- [94] “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, 2009.
- [95] A. M. Sirunyan et al., “Particle-flow reconstruction and global event description with the CMS detector”, *JINST*, vol. 12, no. 10, pp. P10003, 2017.
- [96] Serguei Chatrchyan et al., “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”, *JINST*, vol. 7, pp. P10002, 2012.
- [97] CMS Internal, “Baseline muon selections for Run-II”, accessed in 2017, <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2>.
- [98] CMS Internal, “E/gamma Reconstruction and Particle-Flow Integration”, accessed in 2016, <https://twiki.cern.ch/twiki/bin/viewauth/CMS/EgReconstruction>.
- [99] CMS Internal, “CMS Cut Based Electron ID for Run 2”, accessed in 2018, <https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2>.

- [100] CMS Internal, “Summer16 effective areas for PF isolation for electrons”, 2017, https://indico.cern.ch/event/482673/contributions/2187022/attachments/1282446/1905912/talk_electron_ID_spring16.pdf.
- [101] CMS Internal, “CMS Cut Based Photon ID for Run 2”, accessed in 2018, <https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedPhotonIdentificationRun2>.
- [102] Vardan Khachatryan et al., “Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV”, *JINST*, vol. 10, no. 08, pp. P08010, 2015.
- [103] Gavin P. Salam, “Towards Jetography”, *Eur. Phys. J.*, vol. C67, pp. 637–686, 2010.
- [104] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, “The Anti-k(t) jet clustering algorithm”, *JHEP*, vol. 04, pp. 063, 2008.
- [105] Vardan Khachatryan et al., “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *JINST*, vol. 12, no. 02, pp. P02014, 2017.
- [106] CMS Collaboration, “Jet algorithms performance in 13 TeV data”, 2017.
- [107] CMS Internal, “CMS Jet Identification for the 13 TeV data Run2016”, accessed in 2018, <https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2016>.
- [108] The CMS collaboration, “Identification of b-quark jets with the CMS experiment”, *Journal of Instrumentation*, vol. 8, no. 04, pp. P04013, 2013.
- [109] Simon Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition, 1998.
- [110] Albert M Sirunyan et al., “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *Submitted to: JINST*, 2017.
- [111] R. Frühwirth, W. Waltenberger, and P. Vanlaer, “Adaptive Vertex Fitting”, Tech. Rep. CMS-NOTE-2007-008, CERN, Geneva, Mar 2007.
- [112] CMS Collaboration, “Heavy flavor identification at CMS with deep neural networks”, *CMS-DP-2017-005*, Mar 2017.
- [113] “Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector”, Tech. Rep. CMS-PAS-JME-16-004, CERN, Geneva, 2016.

- [114] CMS Internal, “Performance of missing transverse momentum in pp collision at 13 TeV using the full 2016 data”, 2018, CMS-AN-18-003.
- [115] Vardan Khachatryan et al., “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $s = 8$ TeV”, *JINST*, vol. 10, no. 06, pp. P06005, 2015.
- [116] CMS Collaboration, “Jet energy scale and resolution performances with 13TeV data”, *CMS-DP-2016-020*, Jun 2016.
- [117] CMS Collaboration, “Search for a heavy scalar boson decaying into a pair of Z bosons in the $2\ell 2\nu$ final state”, CMS-PAS-HIG-16-001, 2016.
- [118] D. de Florian et al., “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”, 2016.
- [119] Alessio Magitteri, *Search for a heavy scalar boson in the $ZZ \rightarrow l^+l^-\nu\bar{\nu}$ decay channel and reconstruction of high-energy muons in the CMS experiment*, PhD thesis, Université catholique de Louvain, 2018.
- [120] A. De Rujula, Joseph Lykken, Maurizio Pierini, Christopher Rogan, and Maria Spiropulu, “Higgs look-alikes at the LHC”, *Phys. Rev.*, vol. D82, pp. 013003, 2010.
- [121] James S. Gainer, Kunal Kumar, Ian Low, and Roberto Vega-Morales, “Improving the sensitivity of Higgs boson searches in the golden channel”, *JHEP*, vol. 11, pp. 027, 2011.
- [122] Nikolas Kauer, “Interference effects for $H \rightarrow WW/ZZ \rightarrow \ell\bar{\nu}_\ell\bar{\ell}\nu_\ell$ searches in gluon fusion at the LHC”, *JHEP*, vol. 12, pp. 082, 2013.
- [123] Keshri, Sumit, “Hight pt trigger issue and new scale factors”, 2017, https://github.com/cms212v/212v_fwkw/files/1033932/HightPtTriggerIssueNewSF.pdf.
- [124] Albert M Sirunyan et al., “Measurement of vector boson scattering and constraints on anomalous quartic couplings from events with four leptons and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Phys. Lett.*, vol. B774, pp. 682–705, 2017.
- [125] CMS Internal, “Methods to apply b-tagging efficiency scale factors ”, accessed in 2018, <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMethods>.

- [126] Massimiliano Grazzini, Stefan Kallweit, and Dirk Rathlev, “ZZ production at the LHC: fiducial cross sections and distributions in NNLO QCD”, *Phys. Lett.*, vol. B750, pp. 407–410, 2015.
- [127] Anastasiya Bierweiler, Tobias Kasprzik, and Johann H. Kühn, “Vector-boson pair production at the LHC to $\mathcal{O}(\alpha^3)$ accuracy”, *JHEP*, vol. 12, pp. 071, 2013.
- [128] Stefan Gieseke, Tobias Kasprzik, and Johann H. Kuhn, “Vector-boson pair production and electroweak corrections in HERWIG++”, *Eur. Phys. J.*, vol. C74, no. 8, pp. 2988, 2014.
- [129] Postiau, Nicolas, “ γ -induced electroweak corrections to the $pp \rightarrow W^\pm Z + X$ process at $\sqrt{s} = 13$ TeV, and associated uncertainties”, 2016, CMS Internal, AN-16-431.
- [130] “Procedure for the LHC Higgs boson search combination in Summer 2011”, Tech. Rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011.
- [131] Alexander L. Read, “Presentation of search results: The CL(s) technique”, *J. Phys.*, vol. G28, pp. 2693–2704, 2002, [,11(2002)].
- [132] A L Read, “Modified frequentist analysis of search results (the CL_s method)”, , no. CERN-OPEN-2000-205, 2000.
- [133] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells, “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J.*, vol. C71, pp. 1554, 2011, [Erratum: *Eur. Phys. J.*C73,2501(2013)].
- [134] “CMS Luminosity Measurements for the 2016 Data Taking Period”, Tech. Rep. CMS-PAS-LUM-17-001, CERN, Geneva, 2017.
- [135] Jon Butterworth et al., “PDF4LHC recommendations for LHC Run II”, *J. Phys.*, vol. G43, pp. 023001, 2016.
- [136] Iain W. Stewart and Frank J. Tackmann, “Theory Uncertainties for Higgs and Other Searches Using Jet Bins”, *Phys. Rev.*, vol. D85, pp. 034011, 2012.
- [137] John M. Campbell, R. Keith Ellis, and Ciaran Williams, “Vector boson pair production at the LHC”, *JHEP*, vol. 1107, pp. 018, 2011.
- [138] Aneesh Manohar, Paolo Nason, Gavin P. Salam, and Giulia Zanderighi, “How bright is the proton? A precise determination of the photon parton distribution function”, *Phys. Rev. Lett.*, vol. 117, no. 24, pp. 242002, 2016.

- [139] M. Aaboud et al., “Search for heavy ZZ resonances in the $\ell^+\ell^-\ell^+\ell^-$ and $\ell^+\ell^-\nu\bar{\nu}$ final states using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Eur. Phys. J.*, vol. C78, no. 4, pp. 293, 2018.
- [140] CMS Internal, “Measurement of 2016 Full dataset SFs of Moriond photon ID”, 2017, https://indico.cern.ch/event/604907/contributions/2452908/attachments/1401556/2139396/01_25_EGM.pdf.
- [141] Ansgar Denner, Stefan Dittmaier, Markus Hecht, and Christian Pasold, “NLO QCD and electroweak corrections to $Z + \gamma$ production with leptonic Z-boson decays”, *JHEP*, vol. 02, pp. 057, 2016.
- [142] CMS Collaboration, “Measurement of the $W\gamma$ and $Z\gamma$ inclusive cross sections in pp collisions at $\sqrt{s} = 7$ TeV and limits on anomalous triple gauge boson couplings”, *Phys. Rev. D*, vol. 89, pp. 092005, May 2014.
- [143] Glen Cowan, *Statistical data analysis*, Oxford science publications, 1998.
- [144] J.S. Conway, “Incorporating Nuisance Parameters in Likelihoods for Multi-source Spectra”, , no. arXiv:1103.0354, pp. 115–120. 6 p, Mar 2011, Comments: Presented at PHYSTAT 2011, CERN, Geneva, Switzerland, January 2011, to be published in a CERN Yellow Report.
- [145] Alexander L. Read, “Linear interpolation of histograms”, *Nucl. Instrum. Meth.*, vol. A425, pp. 357–360, 1999.
- [146] Roger J. Barlow and Christine Beeston, “Fitting using finite Monte Carlo samples”, *Comput. Phys. Commun.*, vol. 77, pp. 219–228, 1993.

Notions of statistics

Searches in particle physics rely heavily on statistical tools and interpretation. Therefore some notions of statistics are introduced here. In particular, the likelihood will be introduced and the embedding of uncertainties in the likelihood will be treated.

A.1. Statistics and the likelihood at the LHC

The likelihood \mathcal{L} describes the plausibility of an observed data set, under a certain hypothesis (a model parameter value, for instance) [143]. In collider experiments, the number of observed events for a certain process and a given integrated luminosity are following a Poisson distribution. This distribution is a discrete probability distribution giving the probability of a given number of events to occur in a fixed interval of time if these events occur with a known constant rate λ and independently of the time since the last event, which is exactly the case at the LHC. Its probability function $f(n)$ is given by

$$f(n) = \frac{e^{-\lambda} \lambda^n}{n!}, \quad (\text{A.1})$$

where n represents the number of occurrence in a given time and λ the rate of events.

Translating this to data taking at the LHC, the likelihood to observe n events while b events are expected (the model parameter value here) is given by:

$$\mathcal{L}(n|b) = \frac{e^{-b} b^n}{n!}. \quad (\text{A.2})$$

In the case of this thesis, data are binned in histograms for a given integrated luminosity and the N bins are independent and follow a Poisson distribution. Therefore

the likelihood becomes:

$$\mathcal{L}(\vec{n}|\vec{b}) = \prod_{i=1}^N \frac{e^{-b_i} b_i^{n_i}}{n_i!}. \quad (\text{A.3})$$

where \vec{n} and \vec{b} are the vectors of data and expectations in the bins i .

A.2. Introducing uncertainties in the likelihood

There are three sources of uncertainties in this analysis:

- **statistical**: coming from the number of observed events in a data control sample and from the number of MC events in simulation
- **theoretical**: coming from uncertainties on parton distribution functions and cross sections
- **experimental**: coming from the experimental set up, such as the luminosity measurement, trigger efficiencies, object resolutions...

The uncertainties can be embedded in the likelihood by introducing nuisance parameters $\vec{\theta}$ which influence the model but are not direct variables of interest [133]. The introduction of a nuisance parameter θ is done through a probability density function (pdf) $p(\theta)$ with $\bar{\theta}$ the best estimate of the nuisance (e.g. mean, median, peak) and other parameters characterizing the overall shape of the pdf. Those nuisance parameters can impact the number of expected events and the likelihood becomes:

$$\mathcal{L}(\vec{n}, \vec{\theta}_{obs} | \mu, \vec{\theta}) = \mathcal{L}(\vec{n} | \mu, \vec{\theta}) \mathcal{L}(\vec{\theta}_{obs} | \vec{\theta}) \quad (\text{A.4})$$

where μ is the parameter of interest and where the nuisance parameters $\vec{\theta}$ are constrained both by the data \vec{n} and by auxiliary observations $\vec{\theta}_{obs}$.

By convention, two pdf are used in this analysis:

- **the log-normal pdf**, to account for nuisance parameters corresponding to multiplicative factors on the signal or background yields and in particular to describe positively defined observables. A log-normal variable is a variable whose logarithm is normally distributed. By the central-limit theorem, the product of a large number of random variables a_i becomes log-normally distributed, since its logarithm is the sum of these random variables (i.e. $\log(\prod_i a_i) = \sum_i \log(a_i)$), and this sum becomes normally distributed for a large number of random variables. The log-normal pdf [130] is given by:

$$p(\theta) = \frac{1}{\sqrt{2\pi} \ln(\kappa)} \exp \left[-\frac{1}{2} \left(\frac{\ln(\theta/\bar{\theta})}{\ln(\kappa)} \right)^2 \right] \frac{1}{\theta} \quad (\text{A.5})$$

for $\theta > 0$ and goes to zero for $\theta = 0$. The width of the log-normal is characterized by κ . Compared to a Gaussian distribution, it is well suited to describe positively defined observables (like cross sections, cut efficiencies, luminosity...) and it is a more appropriate choice for very large relative uncertainties (i.e. κ). Left of Fig. A.1 shows log-normal pdf for various κ . One can see that, for small uncertainties it behaves like a Gaussian distribution while it extends to higher tails for large uncertainties.

- **the gamma distribution** [130], to account for statistical uncertainties in the signal region coming from the number of simulated events or from the number of observed events in a data control sample. In these cases, the number of events in the signal region can be small compared to the number of simulated events or to the number of events in the control region and the use of a Poisson distribution is not recommended. Instead, the number of events in the signal region n can be related to the total number N of events in the simulated samples or in the control sample via a factor α such as $n = \alpha N$. Ignoring uncertainties on α , the continuous distribution of the predicted event rate n given α and N is [130]:

$$p(n) = \frac{1}{\alpha} \frac{(n/\alpha)^N}{N!} \exp(-n/\alpha). \quad (\text{A.6})$$

Under this distribution, the most probable value for n is αN , its mean is $\alpha(N + 1)$ and its standard deviation is $\alpha\sqrt{N + 1}$. Let's note that $N = 0$ is a valid case and leads to an exponential distribution of parameter α for n . Right of Fig. A.1 shows gamma distributions varying N .

Nuisance parameters do not only affect the total number of events but also the shape of distributions of interest (M_T in this analysis). In practice, shape uncertainties are implemented in this analysis by computing two shapes per nuisance parameter, each of them corresponding to an upward or downward variation of one standard deviation of the studied nuisance parameter. This is added in the likelihood through a new parameter to interpolate smoothly between the upward and downward shapes by vertical morphing [144, 145]. The vertical morphing consists in turning the nominal, up and down value in a bin in the distribution of interest into a continuous estimate in each bin as a function of the ‘‘shape’’ nuisance parameters. This is done by introducing a morphing parameter f which follows a Gaussian distribution of mean 0 and standard deviation 1. For $f = 0$, the bin content assumes the nominal value and for $f = \pm 1$ the bin content assumes the upper or lower value. When $|f| < 1$ shapes are interpolated quadratically and when $|f| \geq 1$ shapes are extrapolated linearly.

A special kind of shape uncertainty also has to be added. The uncertainty on the number of simulated events and on the number of data events from a control region (non-resonant background and $Z + \text{jets}$ data-driven estimations here) used to estimate

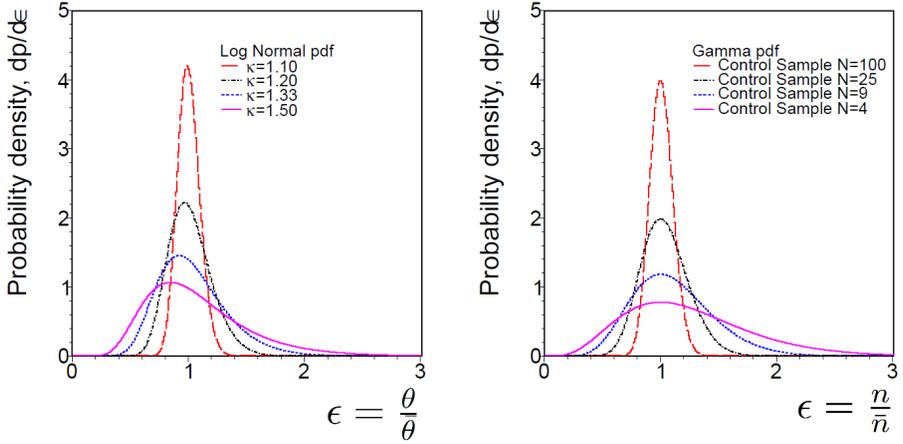


Figure A.1.: **Left:** log-normal (left) distributions for various κ value. **Right:** gamma distributions for various number of events N in a control sample. For small relative uncertainties (i.e. small κ or high N) both distributions are similar to a Gaussian distribution [130].

the background in the signal region has an impact on the shape of the distribution of interest (M_T). However, its behaviour in each bin is independent of the others. The method consists in introducing a separate nuisance parameter for each bin of each process which multiplies the number of expected events in this particular bin for this process [146]. Those numerous shape uncertainties are named “bin-by-bin uncertainties”. While there are many of them, it is possible to reduce their number by requiring a minimal number of events in a bin, discarding cases where this uncertainty is negligible.

A.3. Maximum likelihood estimation and nuisance parameter impacts

Taking all observed data, estimated backgrounds and signals and nuisances parameters, a maximum likelihood estimation (MLE) can be performed. Two scenarii are studied:

- **a background-only fit:** the nuisance parameters $\vec{\theta}$ acting on the expected background distributions \vec{b} are varied to maximize the likelihood $\mathcal{L}(\vec{n}|\vec{b}, \vec{\theta})$
- **a signal-plus-background fit:** the nuisance parameters and the freely floating signal strength μ of the expected signal distribution \vec{s} are varied to maximize

the likelihood $\mathcal{L}(\vec{n}|\mu\vec{s} + \vec{b}, \vec{\theta})$. The fraction between signal processes coming from ggF or VBF production can also be set as a free parameter.

The impact of a nuisance parameter on the parameter of interest can be estimated by performing a MLE, forcing the value of the nuisance parameter of interest to be at one standard deviation while letting the other nuisance parameters vary. The impact is then assessed by computing the difference between the value of the parameter of interest when performing the MLE with all nuisance parameters fitted but one for which the impact is assessed, and the result of the MLE fit without nuisance fitted.

Appendix **B**

Additional plots

This appendix contains additional distributions for the data-driven estimation of the $Z + \text{jets}$ background.

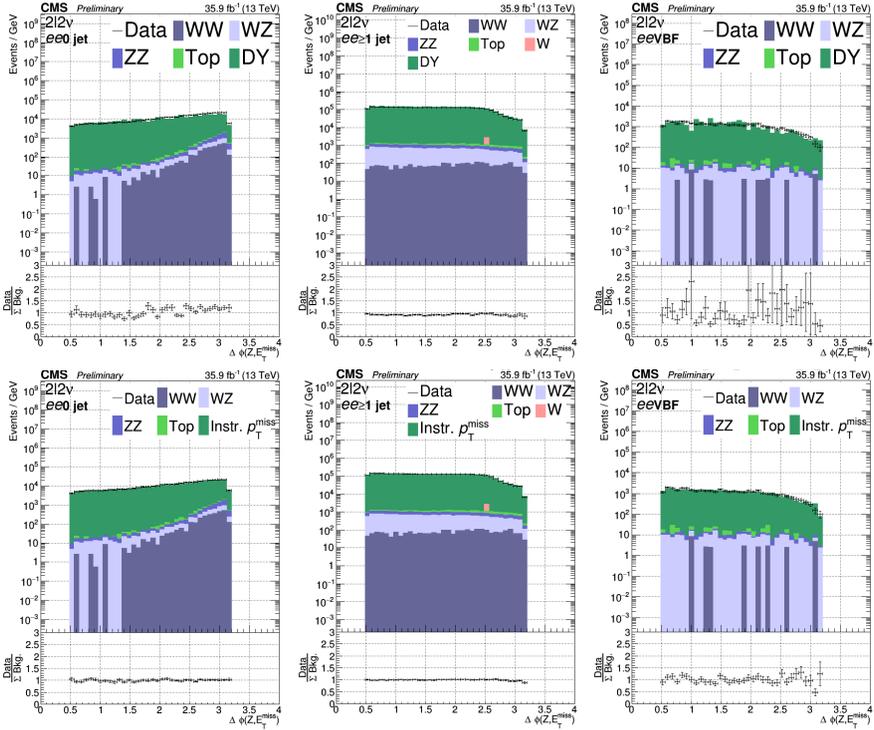


Figure B.1.: Azimuthal angle difference between the boson p_T vector and \vec{p}_T^{miss} in the di-electron channel using simulated $Z + \text{jets}$ samples (top row) and the data-driven estimation of this background (bottom row) before any $|\vec{p}_T^{\text{miss}}|$ cut. The three jet categories are shown: 0 jet (left), ≥ 1 jet (center) and VBF (right).

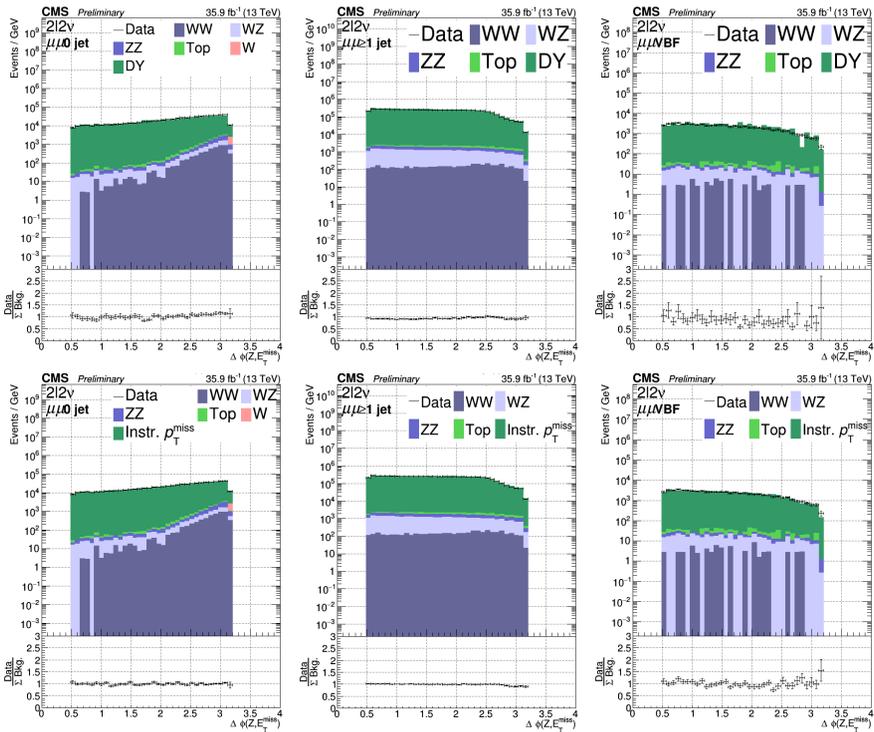


Figure B.2.: Azimuthal angle difference between the boson p_T vector and \vec{p}_T^{miss} in the di-muon channel using simulated Z +jets samples (top row) and the data-driven estimation of this background (bottom row) before any $|\vec{p}_T^{\text{miss}}|$ cut. The three jet categories are shown: 0 jet (left), ≥ 1 jet (center) and VBF (right).

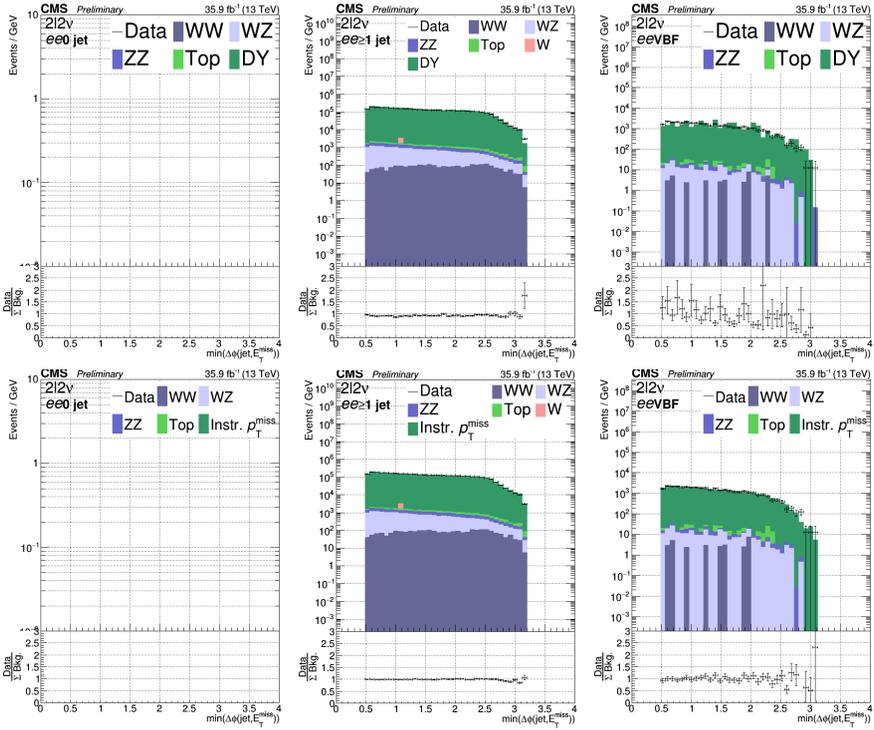


Figure B.3.: Azimuthal angle difference between \vec{p}_T^{miss} and the nearest jet p_T vector above 30 GeV in the di-electron channel using simulated Z + jets samples (top row) and the data-driven estimation of this background (bottom row) before any $|\vec{p}_T^{\text{miss}}|$ cut. The three jet categories are shown: 0 jet (left), ≥ 1 jet (center) and VBF (right).

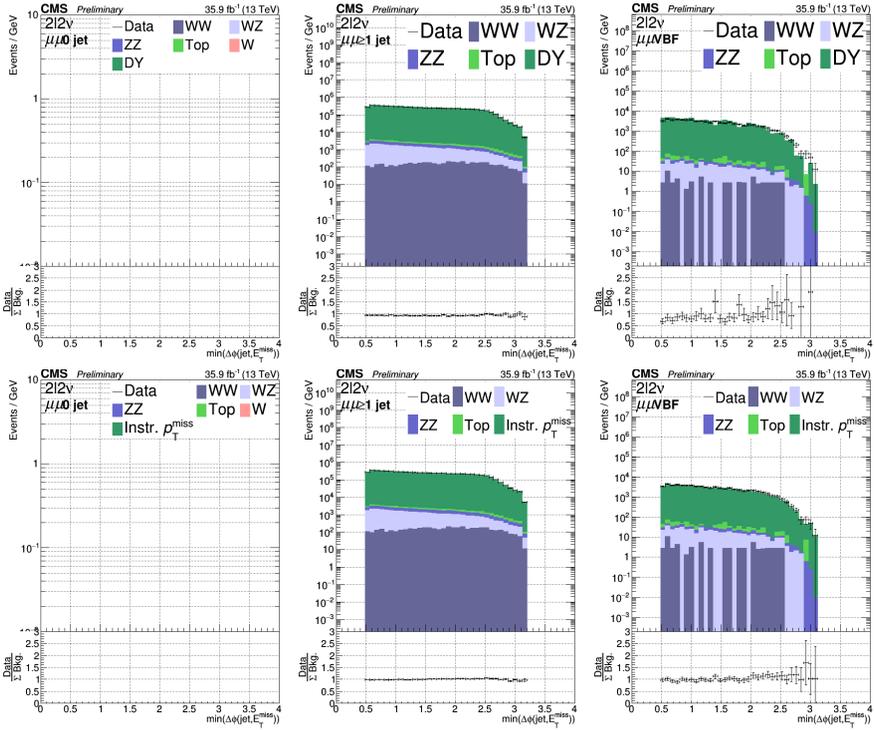


Figure B.4.: Azimuthal angle difference between \vec{p}_T^{miss} and the nearest jet p_T vector above 30 GeV in the di-muon channel using simulated $Z + \text{jets}$ samples (top row) and the data-driven estimation of this background (bottom row) before any $|\vec{p}_T^{\text{miss}}|$ cut. The three jet categories are shown: 0 jet (left), ≥ 1 jet (center) and VBF (right).

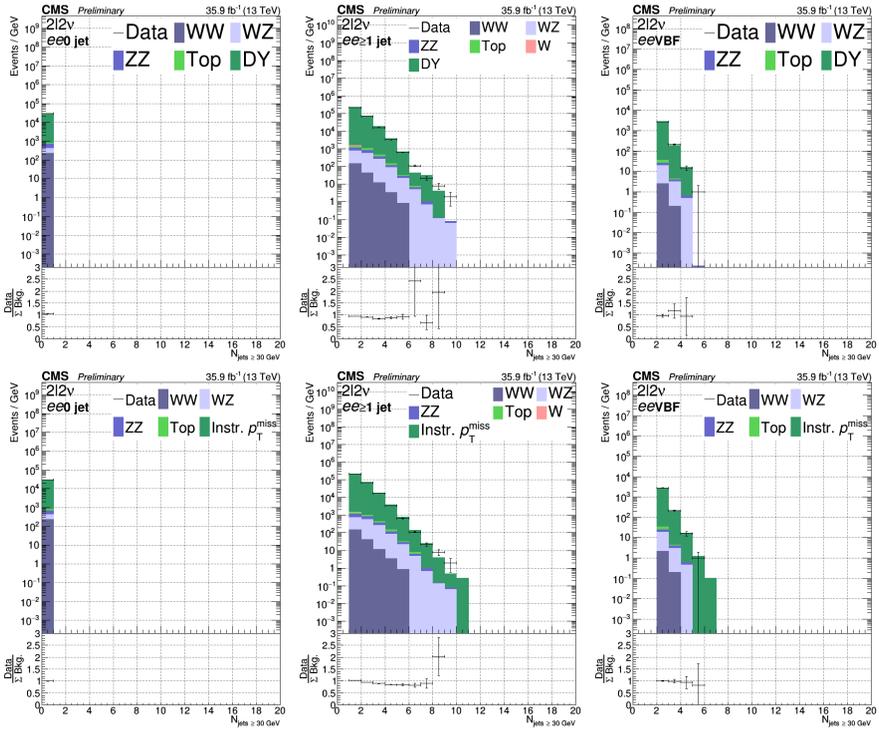


Figure B.5.: Number of jets above 30 GeV in the di-electron channel using simulated $Z + \text{jets}$ samples (top row) and the data-driven estimation of this background (bottom row) before any $|\vec{p}_T^{\text{miss}}|$ cut. The three jet categories are shown: 0 jet (left), ≥ 1 jet (center) and VBF (right).

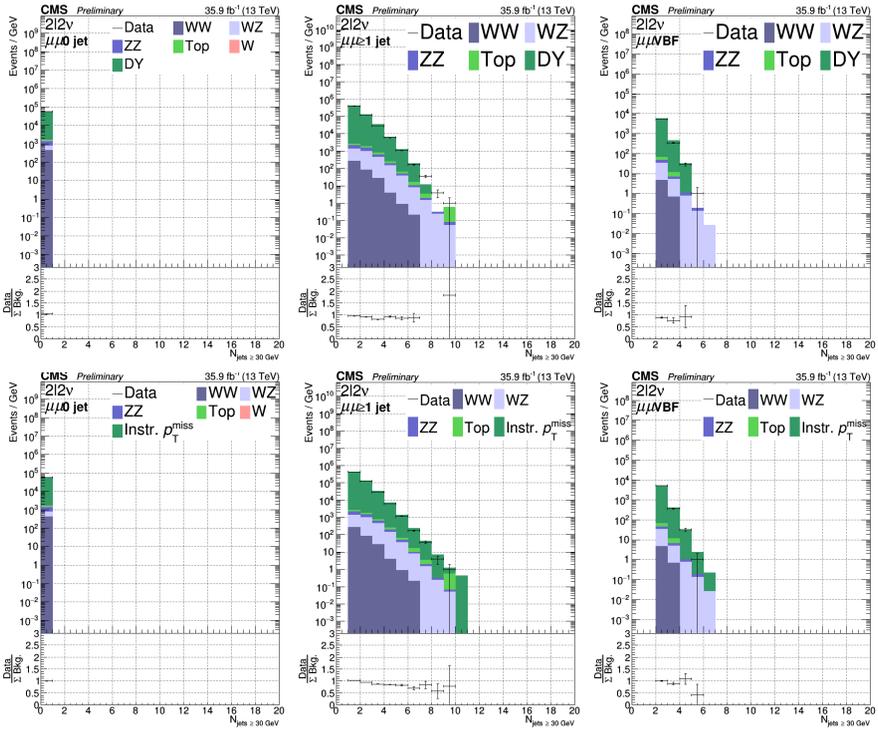


Figure B.6.: Number of jets above 30 GeV in the di-muon channel using simulated $Z + \text{jets}$ samples (top row) and the data-driven estimation of this background (bottom row) before any $|\vec{p}_T^{\text{miss}}|$ cut. The three jet categories are shown: 0 jet (left), ≥ 1 jet (center) and VBF (right).

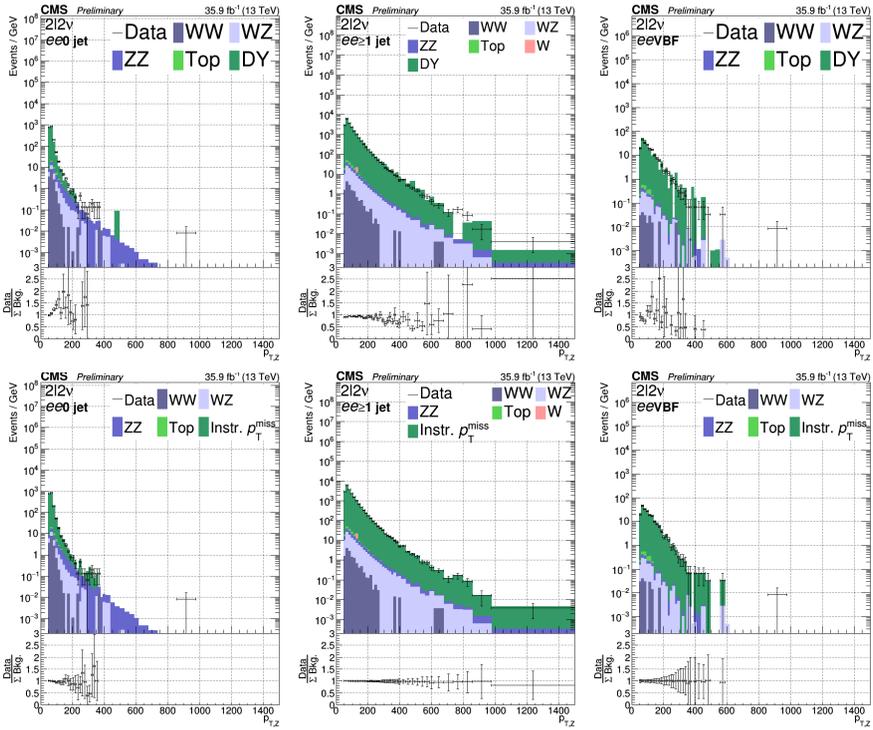


Figure B.7.: Distributions of the boson p_T in the di-electron channel using simulated $Z + \text{jets}$ samples (top row) and the data-driven estimation of this background (bottom row) before any $|\vec{p}_T^{\text{miss}}|$. The three jet categories are shown: 0 jet (left), ≥ 1 jet (center) and VBF (right).

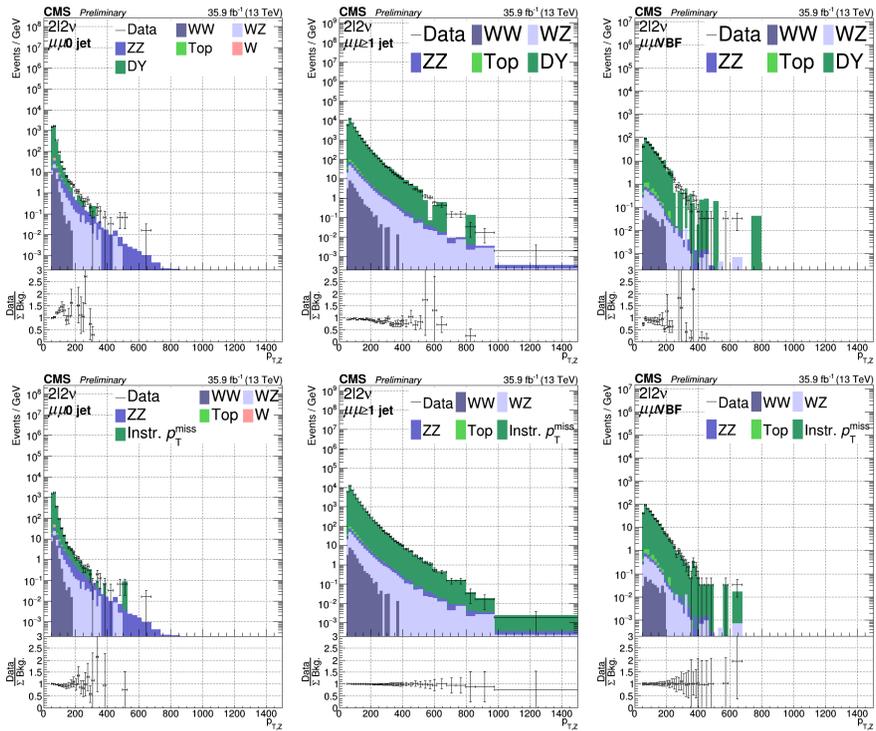


Figure B.8.: Distributions of the boson p_T in the di-muon channel using simulated $Z +$ jets samples (top row) and the data-driven estimation of this background (bottom row) before any $|\vec{p}_T^{\text{miss}}|$. The three jet categories are shown: 0 jet (left), ≥ 1 jet (center) and VBF (right).

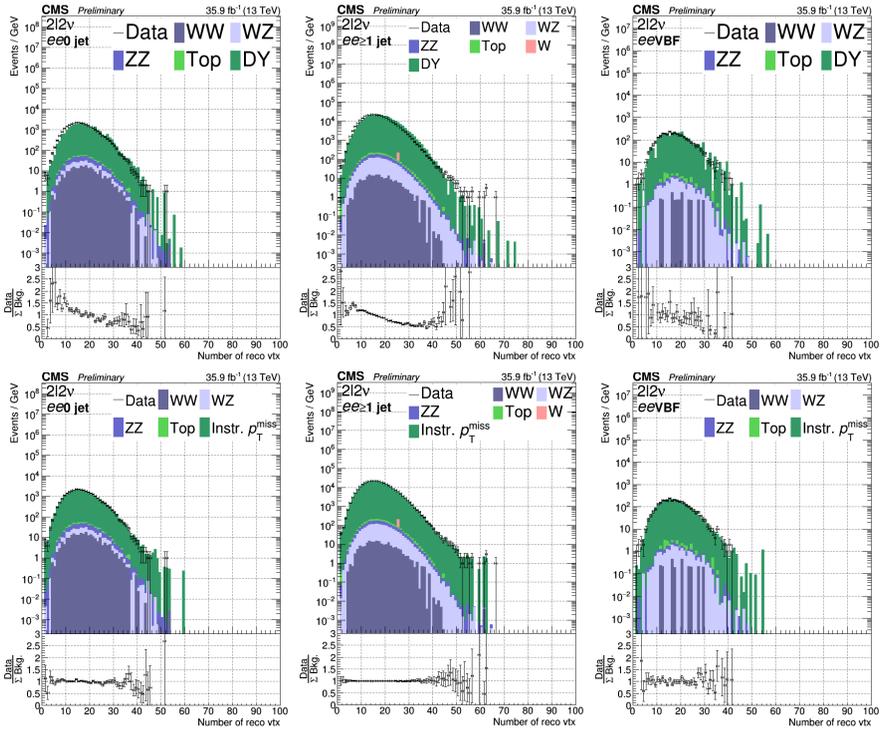


Figure B.9.: Number of reconstructed vertices in the di-electron channel using simulated $Z + \text{jets}$ samples (top row) and the data-driven estimation of this background (bottom row) before any $|p_T^{\text{miss}}|$. The three jet categories are shown: 0 jet (left), ≥ 1 jet (center) and VBF (right).

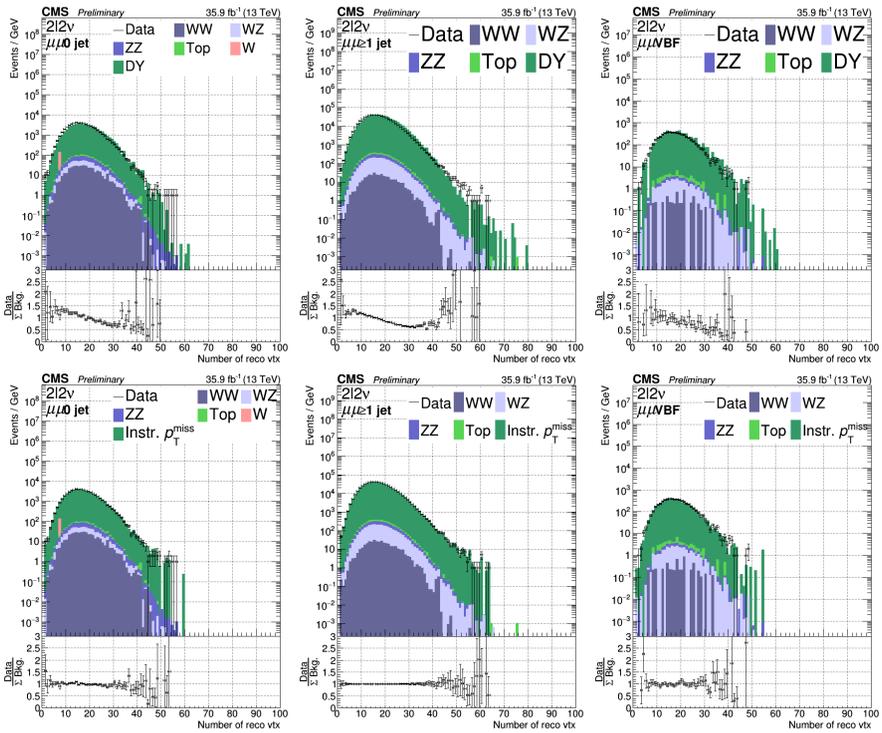


Figure B.10.: Number of reconstructed vertices in the di-muon channel using simulated $Z + \text{jets}$ samples (top row) and the data-driven estimation of this background (bottom row) before any $|\vec{p}_T^{\text{miss}}|$. The three jet categories are shown: 0 jet (left), ≥ 1 jet (center) and VBF (right).